Genomic investigations of acute hepatitis of unexplained aetiology in children

Sofia Morfopoulou[1,2] *, Sarah Buddle[1] *, Oscar Enrique Torres Montaguth[1], Laura Atkinson [3], José Afonso Guerra-Assunção [1], Mahdi Moradi Marjaneh [2,4], Riccardo Zenezini Chiozzi [5] , Nathaniel Storey[3], Luis Campos [6] , J Ciaran Hutchinson [6] , John R Counsell [7], Gabriele Pollara[8], Sunando Roy [1], Cristina Venturini [1], Juan F Antinao Diaz [7] , Ala'a Siam[7,9], Luke J Tappouni[7], Zeinab Asgarian[7], Joanne Ng[9], Killian S Hanlon[7] , Alexander Lennon [3], Andrew McArdle[2], Agata Czap[8], Joshua Rosenheim[8], Catarina Andrade [6], Glenn Anderson [6], Jack C D Lee [3], Rachel Williams [10], Charlotte A Williams [10], Helena Tutill [10], Nadua Bayzid [10], Luz Marina Martin Bernal [10], Hannah Macpherson[11], Kylie-Ann Montgomery [10,11], Catherine Moore [12], Kate Templeton [13] , Claire Neill [14], Matt Holden[15], Rory Gunson [16], Samantha J Shepherd [16], Priyen Shah[2], Samantha Cooray [2], Marie Voice[17], Michael Steele[17], Colin Fink [17] , Thomas E Whittaker[18], Giorgia Santilli [18], Paul Gissen[10], Benedikt B Kaufer[19], Jana Reich [19], Julien Andreani[20,21], Peter Simmonds [20], Dimah K. Alrabiah [10,22] , Sergi Castellano Hereza [10,23], Primmy Chikowore[24], Miranda Odam[24], Tommy Rampling [8, 25, 26], Catherine Houlihan [8, 25,27], Katja Hoschler [25], Tiina Talts [25], Cristina Celma [25], Suam Gonzalez [25], Eileen Gallagher [25], Ruth Simmons [25], Conall Watson [25], Sema Mandal [25], Maria Zambon [25], Meera Chand [25], James Hatcher [3], Surjo De [3], Kenneth Baillie [24], Malcolm Gracie Semple [28,29] , DIAMONDS, PERFORM and ISARIC consortia, Joanne Martin[30] , Ines Ushiro-Lumb[31] , Mahdad Noursadeghi [8], Maesha Deheragoda [32], Nedim Hadzic [32], Tassos Grammatikopoulos [32], Rachel Brown [33], Chayarani Kelgeri [34],Konstantinos Thalassinos [5,35,36], Simon N Waddington[9,37], Thomas S Jacques[6,38], Emma Thomson[39] , Michael Levin[2], Julianne R Brown [3], Judith Breuer [1, 3]

1 Infection, Immunity and Inflammation Department, GOS Institute of Child Health, University College London, London, UK

2 Section for Paediatrics, Department of Infectious Diseases, Faculty of Medicine, Imperial College London, London, UK

3 Great Ormond Street Hospital for Children NHS Foundation Trust, GOSH, Department of Microbiology, Virology and Infection Control, London, UK

4 Section of Virology, Department of Infectious Diseases, Faculty of Medicine, Imperial College London, London, UK

5 UCL Mass Spectrometry Science Technology Platform, Division of Biosciences, University College London, London, UK

6 Histopathology Department, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

7 Research Department of Targeted Intervention, UCL Division of Surgery and Interventional Science, Charles Bell House, 43-45 Foley Street, London, UK

8 Division of Infection and Immunity, University College London, London, UK

9 Gene Transfer Technology Group, EGA-Institute for Women's Health, University College London, London, UK

10 Genetics and Genomic Medicine Department, GOS Institute of Child Health, University College London, London, UK

11 Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, UK

12 Wales Specialist Virology Centre, Public Health Wales Microbiology Cardiff, University Hospital of Wales, Cardiff, UK

13 Department of Medical Microbiology, Royal Infirmary, Edinburgh, UK

14 Public Health Agency, Northern Ireland

15 School of Medicine, University of St Andrews, UK

16 West of Scotland Specialist Virology Centre

17 Micropathology Ltd., University of Warwick Science Park, Coventry, UK

18 Molecular and Cellular Immunology, GOS Institute of Child Health, University College London, London, UK

19 Institute of Virology, Freie Universität Berlin, Berlin, Germany

20 Nuffield Department of Medicine, University of Oxford, Oxford, UK

21 Centre Hospitalier Universitaire (CHU) Grenoble – Alpes, Grenoble, 38000, France

22 National Center for Biotechnology, King Abdulaziz City for Science and Technology, Riyadh 11461, Saudi Arabia

23 UCL Genomics, University College London, London, UK

24 The Roslin Institute, University of Edinburgh, Edinburgh, UK

25 UK Health Security Agency

26 Hospital for Tropical Diseases, University College London Hospitals NHS Foundation Trust, London, UK

27 Department of Clinical Virology University College London Hospitals, London, UK

28 The Pandemic Institute, University of Liverpool, Liverpool UK.

29 Respiratory Medicine, Alder Hey Children's Hospital NHS Foundation Trust, Liverpool UK.

30 Centre for Genomics and Child Health, The Blizard Institute, Queen Mary University London, UK

31 NHS Blood and Transplant, UKHSA

32 Kings College Hospital, NHS Trust

33 Department of Cellular Pathology - University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

34 Liver Unit, Birmingham Women's and Children's Hospital NHS Trust.

35 Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London, UK

36 Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, UK

37 MRC Antiviral Gene Therapy Research Unit, Faculty of Health Sciences, University of the Witswatersrand, Johannesburg, South Africa

38 Developmental Biology and Cancer Department, UCL GOS Institute of Child Health and Department of Histopathology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

39 Medical Research Council-University of Glasgow Centre for Virus Research, Glasgow, UK.

*These authors contributed equally to this work

Abstract

Since its first identification in Scotland, over 1000 cases of unexplained pediatric hepatitis in children have been reported worldwide, including 278 cases in the UK [1]. Here we report investigation of 38 cases, 66 age-matched immunocompetent controls and 21 immunocompromised comparator subjects, using a combination of genomic, transcriptomic, proteomic and immunohistochemical methods. We detected high levels of adeno-associated virus 2 (AAV2) DNA in liver, blood, plasma or stool from 27/28 cases. We found low levels of Adenovirus (HAdV) and Human Herpesvirus 6B (HHV-6B), in 23/31 and 16/23 respectively of the cases tested. In contrast, AAV2 was infrequently detected at low titre in blood or liver from control children with HAdV, even when profoundly immunosuppressed. AAV2, HAdV and HHV-6 phylogeny excluded emergence of novel strains in cases. Histological analyses of explanted livers showed enrichment for T-cells and B-lineage cells. Proteomic comparison of liver tissue from cases and healthy controls, identified increased expression of HLA class 2, immunoglobulin variable regions and complement proteins. HAdV and AAV2 proteins were not detected in the livers. Instead, we identified AAV2 DNA complexes reflecting both HAdV and HHV-6B-mediated replication. We hypothesize that high levels of abnormal AAV2 replication products aided by HAdV and in severe cases HHV-6B, may have triggered immune-mediated hepatic disease in genetically and immunologically predisposed children.

Introduction

The report, in March 2022, of five cases of severe hepatitis of unknown aetiology, led to the UK Health Security Agency (UKHSA) identifying 278 cases in total as of 30 September 2022[1]. Cases, defined as acute non-A-E hepatitis with serum transaminases >500IU in children under ten years of age, were found to have been occurring since January 2022[2]. In the UK, 196 cases required hospitalization, 69 were admitted to intensive care, and 13 required liver transplantation[1]. Case numbers have declined since April 2022[3].

UKHSA investigations identified HAdV to be commonly associated with the unexplained paediatric hepatitis, with 64.7% (156/241) testing positive in one or more samples from whole blood (the most sensitive sample-type[4]) or mucosal swabs. 35/77 HAdVs from blood were typed as F41. Seven of eight patients in England who required liver transplantation tested HAdV positive in blood, with F41 found in 5/5 genotyped [2]. SARS-CoV-2 infection was detected in 8.9% (15/169) of UK and 12.8% (16/125) of English cases[2].

Given the uncertainty around the aetiology of this outbreak, and the potential that HAdV-F41 if implicated (**Figure 1A**), could be a new or recombinant variant, we undertook untargeted metagenomic and metatranscriptomic sequencing, of liver biopsies from five liver transplant cases and whole blood from five non-transplanted cases (**Table 1, Figure 1B**). The results were further verified by confirmatory PCRs of liver, blood, stool and nasopharyngeal samples from a total of 38 cases for which there was sufficient residual material. We compared our results with those from 13 healthy children and 52 previously healthy children presenting to hospital with other febrile illness, including adenovirus, hepatitis unrelated to the current

outbreak or a critical illness requiring admission to the Intensive Care Unit. We also tested blood and liver biopsies from 17 profoundly immunosuppressed children with hepatitis who were not part of the current outbreak, in whom reactivation of latent infections might be expected.

Results

Cases

We received samples from 38 children meeting the case definition (**Table 1**). All cases were aged less than ten years old and 22/23 previously tested were positive by adenovirus PCR (**Supplementary Table 1, Table 2, Extended Data Table 1**). A summary of the samples received from these cases and investigations carried out on them are shown in **Figure 1B&C**.

Clinical details

Pre-existing conditions, autoimmune, toxic and other infectious causes of hepatitis were excluded in 12 transplanted (cases 1-5, 28, 29, 31-34, 36) and 4 non-transplanted (cases 30, 35, 37, 38) children, investigated at two liver transplant units, (**Supplementary Table 1**). The 12 transplanted cases reported gastrointestinal symptoms (nausea, vomiting, diarrhea) preceding transplant by a median of 20 days (range 8-42 days). All 12 transplanted children survived, while the four children who did not receive liver transplants recovered without sequelae or evidence of chronic liver-related conditions. Five of the remaining 22 cases referred by Health Security Agencies, for whom this information was available, recovered without sequelae (**Table 1, Supplementary Table 1**).

Metagenomic Sequencing

We performed metagenomic and metatranscriptomic sequencing on samples of frozen explanted liver tissue from five cases who received liver transplants (median age 3 years) and six blood samples from five non-transplanted hepatitis cases (median age 5 years) (**Table 1, Figure 1B**). The liver samples had uniform and consistently high sequencing depth both for DNA-seq and RNA-seq, while the blood samples had variable sequencing depth particularly for RNA-seq **(Supplementary Table 2**). We detected[5] abundant AAV2 reads in DNA-seq from 5/5 explanted livers and 4/5 blood samples from non-transplant cases (7-42 and 1.2-42 reads/million respectively) (**Table 2**). Lower levels of HHV-6B were present in DNA-seq of all explanted liver samples (0.09-4 reads/million) but not in the six blood samples (**Table 2**). HAdV was detected (five reads) in one blood sample (**Table 2**).

Evidence of AAV2 replication

Metatranscriptomics revealed AAV2, but not HHV-6B or HAdV, RNA reads, in liver and blood samples (0.7-10 and 0-7.8 reads per million respectively). Mapping liver RNA-seq data to the RefSeq AAV2 genome (NC_001401.2) identified high expression of the cap ORF, particularly at the 3' end of the capsid, suggesting viral replication[6] (**Extended Data Figure 1A**) while RT-PCR of two livers confirmed the presence of AAV2 mRNA from the cap ORF

(**Extended Data Figure 1C**). In the blood samples, which had not been treated to preserve RNA, we detected low levels of AAV2 RNA reads mapping throughout the genome. (**Extended Data Figure 1B**).

Nanopore sequencing of explanted livers

Ligation-based untargeted nanopore sequencing was applied to DNA from 4/5 frozen liver samples. All four samples were initially sequenced at a lower depth (Average N50: 8.37 kb). 6-16 AAV2 reads were obtained from each sample (5.57-22.24 million total reads, **Supplementary Table 3**). Mapping revealed concatenation of the 4kb genome, compatible with active AAV2 replication[7]. We observed alternating and head-to-tail concatemers which could be consistent with both HAdV and human herpesvirus-mediated rolling hairpin and rolling circle replication respectively[8]. Two of these samples were sequenced more deeply, resulting in 51 and 178 AAV2 reads in 82.9 and 122 million total (N50 4.40-8.52kb) (Supplementary Table 3). 42-51% of reads in the deeper sequences comprised randomly linked, truncated and rearranged genomes with few that were intact and full length (**Extended Data Figure 2**). The remaining reads were <3000 bp long and may represent sections of either monomeric genomes or of more complex structures.

Integration analysis

There was some evidence of AAV2 integration by deeper nanopore sequencing of explanted livers **(Supplementary Table 3),** however none of the integration sites were confirmed by Illumina metagenomic or targeted AAV2 sequencing. The results are likely to represent artefacts of this library preparation method, with chimeric reads described to occur in 1.7-3% of reads[9,10]. Given the number of human reads (72-120 million) we might expect to see this artefact occurring most commonly between AAV2 and human than between AAV2 reads.

Confirmatory real-time PCR

Where sufficient residual material was available, PCR tests were performed for AAV2 (28/38), HAdV (31/38), and HHV-6B (23/38). The results confirmed high levels (CTs: 17-21) of AAV2 DNA in all five frozen explanted livers that had undergone metagenomics (**Table 2, Figure 2D**) with lower levels of HHV-6B and HAdV DNA (CTs: 27-32 and 37-42 respectively). AAV2 DNA was also detected (CTs:19-25) in blood from 4/5 cases that had undergone metagenomics while HAdV, at levels too low to genotype and HHV-6B were detected in 2/4 and 3/4 respectively (one had insufficient material) (**Table 2**). One of the blood metagenomics cases (case 9, JBB1) with insufficient material to test for HAdV and HHV-6B, tested positive for both viruses in the referring laboratory. The AAV2-negative blood sample (case 10, JBB15) was also negative for HAdV but positive for HHV-6B (**Table 2**). A further 10/10 blood samples tested from cases were positive for HAdV by PCR. Sufficient material was available for AAV2 PCR in six of these (all positive; CTs: 20-23) and HHV-6B PCR in two (one positive CT: 37) (**Extended Data Table 1**).

AAV2 PCR was positive in nine formalin fixed paraffin embedded (FFPE) liver samples, including seven from transplanted (CTs: 23-25) and two from non-transplant cases (CTs:34-36, **Extended Data Table 1**). HHV-6B PCR was positive in 6/7 FFPE samples (not case 32) from transplanted (CTs: 30-37) and 0/2 (cases 30 & 35) from non-transplanted cases, with HAdV positive (CTs: 40-44) in 4/9. Three each transplanted (32, 34, 36) and non-transplanted (35, 37, 38) cases had serum available for testing. All were AAV2 positive (CTs: 27-32) and HHV-6B negative with one transplanted and one non-transplanted case testing HAdV positive (**Extended Data Table 1)**.

Taken together, 27/28 cases tested were AAV2 PCR positive, 23/31 HAdV positive and 16/23 HHV-6B positive. When results from referring laboratories were included, 33/38 were positive for HAdV and 19/26 for HHV-6B (**Table 2, Extended Data Table 1**).

Controls and comparators

To better contextualize the findings in cases with unexplained hepatitis, we selected control groups of children who were not part of the outbreak.

Blood from immunocompetent children

Whole blood from 65 immunocompetent children matched by age to cases (median age 3.8 years) **(Figure 1B**. **Extended Data Table 2A, Supplementary Table 4)** who were healthy, or had adenovirus infection, hepatitis, or critical illness, including requiring critical care, were selected from the PERFORM (Personalised Risk assessment in febrile illness to optimise Real-life Management,www.perform2020.org) and DIAMONDS (Diagnosis and Management of Febrile Illness using RNA Personalised Molecular Signature Diagnosis study, www.diamonds2020.eu) studies. Both studies recruited children presenting to hospital with an acute onset febrile illness between 2017 and 2020 (PERFORM) and July 2020 to October 2021, during the COVID-19 pandemic (DIAMOND) **(Supplementary Table 4**). Of the PERFORM/DIAMONDS control whole blood samples, 6/65 (9.2%) were AAV2 PCR positive **(Supplementary Table 5)**, as compared with 10/11 (91%) of whole blood samples from cases (**Figure 2A,** p= 8.466e-08, Fisher's exact test). AAV2 DNA levels were significantly higher in whole blood from cases as compared to controls (**Figure 2E,** p = 2.747e-11, Mann-Whitney Test**)**.

One subject with an HAdV-F4 positive blood sample, originally thought to have unexplained paediatric hepatitis, was later found to have a prior condition that explained the hepatitis and was therefore reclassified as a control, (referred to as "reclassified control" or CONB40, (**Supplementary Table 5**). This blood sample was negative for AAV2 by PCR (**Supplementary Table 5**).

Liver from immunocompromised children

Frozen liver biopsy material from four immunocompromised children, (median age 10 years) (CONL1-4) who had been investigated for other forms of hepatitis were also tested (**Figure 1B, Extended Data Table 2B**). In three, liver enzymes were raised (**Supplementary Table**

S6); no results were available for CONL4. AAV2 was detected in CONL3 (CT:39) and HHV-6B (CT:34), in CONL2, while HAdV was negative (**Figure 2D, Suppl. Table 5**).

Blood from immunocompromised comparators

We also tested immunocompromised children who are more likely to reactivate latent viruses. Whole blood from 17 immunocompromised children (median age 1 year) with raised liver transaminases (AST/ALT>500IU) and viraemia (HAdV or CMV), all sampled in 2022 (**Figure 1B**) were tested for AAV2, HHV-6B and HAdV (**Supplementary Table 5, Extended Table 2B**). The majority had received human stem cell or solid organ transplants, and none were linked to the recent hepatitis outbreak **(Extended Data Table 2B**). 5/15 (33%) were positive for HHV-6B while 6/17 (35%) were positive for AAV2, significantly fewer than in cases ($p = 0.005957$, Fisher's exact) and at significantly lower CT levels ($p = 6.517e-05$, Mann-Whitney) **(Figure 2, Supplementary Table 5**). One HAdV and AAV2-positive immunocompromised comparator (CONB23) was also positive for HHV-6B (**Supplementary Table 5**).

Four of the six AAV2 positive children from the DIAMONDS/PERFORM cohort (**Figure 2A, Supplementary Table 5**) and all six of the AAV2 positive immunocompromised children **(Figure 2A**, **Supplementary Table 5**) were also HAdV positive.

Whole viral genome sequencing

One full HAdV-F41 genome sequence from the stool of one case (OP174926, case 22) (**Supplementary Table 7**) clustered phylogenetically with the HAdV-F41 sequence obtained from the reclassified-control (CONB40) and with other HAdV-F41 sequences collected between 2015-2022, including 23 contemporaneous stool samples from children without the unexplained paediatric hepatitis (**Figure 3A, Figure 1C**). Sequencing and K-mer analysis[11] of HAdV from 13 cases with partial sequences, identified genotype HAdV-F41 in twelve (**Supplementary Tables 7, 8**). The partial sequences showed most similarity to control sequence OP047699 (**Supplementary Table 8**) mapping across the entire viral genome, thus further excluding a recombinant virus.

Single nucleotides polymorphisms (SNPs) were largely shared between the single HAdV positive case from stool (OP174926) and control whole genome sequences (**Extended Data Figure 3A).** Given reported mutation rates for HAdV-F41 and other adenoviruses[12,13], any differences are likely to have arisen before the outbreak. No new or unique amino acid substitutions were noted in HAdV sequences from cases with only two substitutions overall (**Extended Data Figure 2D**) and none in proteins critical for AAV2 replication.

AAV2 sequences from 15 cases, including five from the explanted livers and ten from whole blood from non-transplanted cases, clustered phylogenetically with control AAV2 sequences obtained from four immunocompromised HAdV positive children with elevated ALT in the comparator group (**Extended Data Table 2B**) and two healthy children with recent HAdV-F41 diarrhoea (**Figure 3B, Supplementary Table 9**). The degree of diversity and lack of a

unique common ancestor between case AAV2 genomes suggest these are not specific to the hepatitis outbreak, but instead reflect the general population's current viral diversity. While comparison of the AAV2 sequences showed no difference between cases and controls, contemporary AAV2s showed changes in the capsid compared to historic AAV2 (**Extended Data Figure 3C)**. None of these changes were shared with the hepatotropic AAV7 and AAV8 viruses (**Extended Data Figure 3B**). The majority of the contemporary AAV2 genomes in cases and controls (20/21) contained a stop codon in the X gene, which is involved in viral replication[14], while historic AAV2 genomes contained this less frequently (11/35). The significance, if any, of this is currently unknown.

While mean read depths for four HHV-6B genomes recovered from explanted livers were low (x5-x10) (**Supplementary Table S12**), phylogeny (**Figure 3C**) confirmed that all were different.

Transduction of AAV2 capsid mutants

Using a recombinant AAV2 (rAAV2) vector with a VP1 sequence (**Extended Data Figure 4A**) containing the consensus amino acid sequence from AAV2 cases (**Extended Data Figure 3B**) (AAV2Hepcase), we generated functional rAAV particles that transduced Huh-7 cells with comparable efficacy to both canonical AAV2 and the synthetic liver-tropic LK03 AAV vector[15]. Unlike canonical AAV2, AAV2Hepcase capsid, which contains mutations (R585S and R588T) that potentially affect the heparin sulfate proteoglycan (HSPG) binding domain, was unaffected by heparin competition, a feature that is associated with increased hepatotropism (**Extended Data Figure 4B&C)** [16,17].

Histology and Immunohistochemistry

Histological examination of the 12 liver explants and two liver biopsies showed non-specific features of acute hepatitis with ballooning hepatocytes, disrupted liver architecture with varying degrees of perivenular, bridging or pan acinar necrosis. There was no evidence of fibrosis suggestive of an underlying chronic liver disease. The appearances were similar to historic cases of seronegative hepatitis of unknown cause in children. There were no typical histological features of autoimmune hepatitis (AIH), notably no evidence of portal-based plasma cell rich infiltrates. A cellular infiltrate was present in all cases which on staining appeared to be predominantly of CD8 positive T-cells but also included CD20 positive B-cells. More widespread staining with the CD79a pan-B cell lineage which also identifies plasma cells was also observed (**Extended Data Figure 5**). Macrophage lineage cells showed some C4d complement staining, while staining for immunoglobulins was non-specific with disruption of the normal canalicular staining seen in controls due to the architectural collapse. MHC Class I and II staining although increased in cases, was non-specific and associated with sinusoid-containing blood cells and necrotic tissue (**Extended Data Figure 6A**). No viral inclusions were observed and there were no features suggestive of direct viral cytopathic effect.

Immunohistochemistry was negative for adenovirus. Staining of the five explanted livers with AAV2 antibodies demonstrated evidence of non-specific ingested debris but not the nuclear staining seen in the positive AAV2 infected cell lines and murine infected tissue (**Extended Data Figure 6B**). All five liver explants showed positive staining of macrophage derived cells with antibody to HHV-6B, with no staining of negative control serial sections (**Extended Data Figure 6B**). No specific HHV-6B staining was observed in 13 control liver biopsies from patients (including three children <18 years) with other viral hepatitis, toxic liver necrosis, autoimmune and other hepatitis, and normal liver. The control set was also negative for HAdV and AAV2 by IHC.

Liver sections were morphologically suboptimal for electron microscopy, but no viral particles were identified in hepatocytes, blood vessel endothelial cells and Kupffer cells.

Transcriptomic analysis

We quantified functional cytokine activity by expression of independently derived cytokine-inducible transcriptional signatures of cell mediated immunity (**Supplementary Table 11**) in bulk genome-wide transcriptional profiles from four of the frozen explanted livers. Results were compared to published data from normal adult livers (n=10) and adult hepatitis B-associated acute liver failure (n=17) (GSE96851)[18]. Data from the unexplained hepatitis cases revealed increased expression of diverse cytokines and pathways compared to normal liver. These pathways included prototypic cytokines associated with T cell responses including IFNγ, IL2, CD40LG, IL4, IL5, IL7, IL13 and IL15 (**Figure 4A, Supplementary Table 12**) as well as some evidence of innate immune type 1 interferon (IFN) responses. Many of these responses showed substantially greater activity in unexplained hepatitis compared to fulminant hepatitis B virus disease. The most striking enrichment was for TNF expression, and included other canonical pro-inflammatory cytokines including IL1 and IL-6 (**Extended Data Figure 7**). These data are consistent with an inflammatory process involving multiple pathways.

Proteomics

Proteomic analysis of the five frozen explanted livers did not detect AAV2 or HAdV proteins. Expression of the HHV-6B U4, a protein of unknown function, was found in 4/5 cases, U43, part of the helicase primase complex in 2/5 and U84, a homologue of cytomegalovirus UL117, implicated in HHV-6B nuclear replication, in 2/5 (**Extended Data Figure 8**).

The human proteome from the five frozen liver explants was compared with publicly available data from 7 control "normal" livers, taken from two different studies[19,20]. Both protein and peptide analyses (**Figure 4B &C, Supplementary Table 13&14**) found increased expression in unexplained hepatitis cases of HLA class 11 proteins and peptides (e.g. HLADRB1 and 4), multiple peptides from variable regions of the heavy and light chains of immunoglobulin, complement proteins (such as C1q) and intracellular and extracellular released proteins from neutrophils and macrophages (MMP8 and MPO).

There was no evidence of HAdV, AAV2 or HHV-6B in any of the control livers.

Discussion

Despite reports implicating HAdV-F41 as causing the recent outbreak of unexplained paediatric hepatitis, we found very low levels of HAdV DNA, no proteins, inclusions or viral particles, including in explanted liver tissue from affected cases and no evidence of a change in the virus. In contrast, metagenomic and PCR analysis of liver tissue and blood identified high levels of DNA from adeno-associated virus 2 (AAV2), a member of the *Dependoparvovirus* genus, which has not previously associated with clinical disease, in 27/28 cases. Replication of AAV2 requires coinfection with a helper virus, such as HAdV, herpesviruses, or papillomavirus[21] and can also be triggered in the laboratory by cellular damage[22], raising the possibility that the AAV2 detected was a bystander of previous HAdV-F41 infection and/or liver damage. Against this, we found little or no AAV2 in blood from age-matched immunocompetent, children including those with adenovirus infection, hepatitis or critical illness (**Figure 2D**). AAV2 has been reported to establish latency in liver[23], however, even in critically ill immunosuppressed children with hepatitis in whom reactivation might occur, we detected AAV2 infrequently and at significantly lower levels in blood or liver biopsies (**Figure 2D, Figure 2G).**

RNA transcriptomic and rt-PCR data from explanted livers point to active AAV2 infection, although we did not detect AAV2 proteins by immunohistochemistry (**Extended Data Figure 6B)** or proteomics (**Extended Data Figure 8**) and no viral particles. The abundant AAV2 genomes in the explanted liver are concatenated with many complex and abnormal configurations. AAV genome concatenation may occur during AAV2 replication[8], while abnormal AAV2 DNA complexes and rearrangements have been observed in the liver following AAV gene therapy[7,44]. Hepatitis following AAV gene therapy is well described [24–26] with deaths, albeit rarely [27] . The pattern of complexes typify both HAdV and herpesvirus (including HHV-6B)-mediated AAV2 DNA replication[6]. The presence of HHV-6B DNA in 11/12 explanted livers, but not in livers (0/2) of non-transplanted children, or control livers as well as the expression, in 5/5 cases tested, of HHV-6B proteins, including U43, a homologue of the HSV1 helicase primase UL52 which is known to aid AAV2 replication, highlight a possible role for HHV-6B as well as HAdV, in the pathogenesis of AAV2 hepatitis, particularly in severe cases. While AAV2 is also capable of chromosomal integration[28 29 30] we found little evidence of this by long read sequencing, computational analysis of metagenomics data or examination of unmapped reads, although further confirmatory studies may be required.

Although the pathogenesis of unexplained paediatric hepatitis and the role of AAV2, remain to be determined, our results point strongly to an immune-mediated process. Transcriptomic and proteomic data from the five explant livers identified significant immune dysregulation involving genes and proteins that are strongly associated with activation of B and T cells, neutrophils and macrophages as well as innate pathways. The findings are supported by immunohistochemical staining showing infiltration into liver tissue of CD8+, B cell and B

cell lineage cells. Upregulation of canonical proinflammatory cytokines including lL15, which has also been seen in a mouse model of AAV hepatitis[45], Il4 and TNF occurred at levels greater even than are seen in fulminant liver failure following hepatitis B virus. Increases in the same immunoglobulin variable region peptides and corresponding proteins from both immunoglobulin heavy and light chains across all five livers points to specific antibody involvement[31]. HLA DRB1*04:01 (12/13 tested) (**Supplementary Table 1**) among children in our study supports the same genetic predisposition as mooted in a sister Scottish study[32].

An immune mediated process is consistent with studies of hepatitis following AAV gene therapy, where raised AAV2 IgG and capsid specific CTLs are observed in the affected patients, although whether these directly mediate the hepatitis remains unclear [26,33]. While we did not find that AAV2 sequences in cases differed from those in AAV2 occurring as coinfections in HAdV-F41-positive stool collected from control children during the contemporary HAdV -F41 gastroenteritis outbreak (**Figure 3B**), rAAV capsid expressing consensus capsid sequence from the unexplained hepatitis cases (AAV2Hepcase), showed reduced HSPG dependency, compared to canonical AAV2 (**Extended Data Figure 4**, whilst retaining hepatocyte transduction ability. This points to likely greater *in vivo* hepatotropism of currently circulating AAV2 than has hitherto been assumed from data on canonical AAV2 [17]. Another member of the parvovirus family, Equine Parvovirus-Hepatitis (EqPV-H) has also been associated with acute hepatitis in horses (Theiler's disease) [34].

There are a number of limitations to our study. While other known infectious, autoimmune, toxic and metabolic aetiologies [3] have been excluded including by other studies[35,36], numbers of cases investigated here are small, the study is retrospective, the immunocompromised controls were not perfectly age-matched, and only one immunocompetent and 17 immunocompromised controls were sampled during exactly the same period as the outbreak. Age-matched DIAMONDS immunocompetent controls contemporaneous with the outbreak, although few in number, were however found to be AAV2 negative in a separate study carried out in Scotland[32].

Finally, our data alone are not sufficient on their own to rule out a contribution from SARS-CoV-2 Omicron, the appearance of which preceded the outbreak of unexplained hepatitis. (**Supplementary Table 1**). We did not detect SARS-CoV-2 metagenomically even in three subjects who tested positive on admission. Moreover, although seropositivity was higher in our cases (15/20) compared to controls (3/10), this was not the case for another UK cohort[35] (38%) or in preliminary data from a UKHSA case-control study [3], which showed similar SARS-CoV-2 antibody prevalence between unexplained hepatitis cases and population controls (<5y 60.5% versus 46.3% respectively, and 5-10y 66.7% versus 69.6%). In line with UK national recommendations at the time, none of the children had received a COVID vaccine.

While we find little evidence for SARS-CoV-2 directly causing the hepatitis outbreak, we cannot exclude the impact of the COVID-19 pandemic on child mixing and infection

patterns. The contemporaneous development of unexplained paediatric hepatitis with a national outbreak of HAdV-F41[2] and the finding of HAdV-F41 in many cases, suggests that the two are linked. Enteric adenovirus infection is most common in those aged under five[2] and infection is influenced by mixing and hygiene[37]. Few cases of HAdV-F41 occurred between 2020 and 2022 and no major outbreaks were recorded[2]. The current HAdV outbreak followed relaxation of restrictions due to the pandemic and represented one of many infections, including other enteric pathogens that occurred in UK children following return to normal mixing[38]. Under normal circumstances, AAV2 antibodies levels are high at birth, subsequently declining to reach their lowest point at 7-11 months, increasing thereafter through childhood and adolescence[39]. AAV2 is known to be spread with respiratory adenoviruses, infections which declined during the COVID-19 pandemic, and has not been detected by us in over 30 SARS-CoV-2 positive nasopharyngeal aspirates (data not shown). We also found AAV2 DNA to be present in HAdV-F41-positive stool from both cases and controls (**Supplementary Table 5**). With loss of child mixing during the COVID-19 pandemic, reduced spread of common respiratory and enteric viral infections and no evidence of AAV2 in SARS-CoV-2 positive nasal pharyngeal swabs, it is likely that immunity to both HAdV-F41 and AAV2 declined sharply in the age group affected by this unexplained hepatitis outbreak. Pre-existing antibody is known to reduce levels of AAV DNA in the liver of non-human primates following infusion of AAV gene therapy vectors[40]. The possibility that, in the absence of protective immunity, excessive replication of HAdV-F41 and AAV2 with accumulation of AAV2 DNA in the liver led to immune-mediated hepatic disease in genetically predisposed individuals now needs further investigation. Evaluation of drugs that inhibit TNF and other cytokines massively elevated in this condition may identify important therapeutic options for future cases.

Main Text References

1. Joint ECDC-WHO Regional Office for Europe Hepatitis of Unknown Origin in Children Surveillance Bulletin. https://cdn.ecdc.europa.eu/novhep-surveillance/.
2. Ukhsa. Investigation into acute hepatitis of unknown aetiology in children in England: Technical briefing 3.
3. Investigation into acute hepatitis of unknown aetiology in children in England: technical briefing 4. (2022).
4. Marsh, K. *et al.* Investigation into cases of hepatitis of unknown aetiology among young children, Scotland, 1 January 2022 to 12 April 2022. *Euro Surveill* **27**, 1–7 (2022).
5. Morfopoulou, S. & Plagnol, V. Bayesian mixture analysis for metagenomic community profiling. *Bioinformatics* **31**, 2930–2938 (2015).
6. Berns, K. I. Parvovirus replication. *Microbiol Rev* **54**, 316–329 (1990).
7. Sun, X. *et al.* Molecular Analysis of Vector Genome Structures After Liver Transduction by Conventional and Self-Complementary Adeno-Associated Viral Serotype Vectors in Murine and Nonhuman Primate Models. *https://home.liebertpub.com/hum* **21**, 750–762 (2010).
8. Meier, A. F. *et al.* Herpes simplex virus co-infection facilitates rolling circle replication of the adeno-associated virus genome. *PLoS Pathog* **17**, (2021).
9. Xu, Y. *et al.* Detection of viral pathogens with multiplex nanopore MinION sequencing: Be careful with cross-Talk. *Front Microbiol* **9**, 2225 (2018).
10. Eccles, D., White, R., Pellefigues, C., Ronchese, F. & Lamiable, O. Investigation of chimeric reads using the MinION. *F1000Research 2017 6:631* **6**, 631 (2017).
11. Guerra-Assunção, J. A., Goldstein, R. & Breuer, J. AYUKA: A toolkit for fast viral genotyping using whole genome sequencing. *bioRxiv* 2022.09.07.506755 (2022) doi:10.1101/2022.09.07.506755.
12. Risso-Ballester, J., Cuevas, J. M. & Sanjuán, R. Genome-Wide Estimation of the Spontaneous Mutation Rate of Human Adenovirus 5 by High-Fidelity Deep Sequencing. *PLoS Pathog* **12**, e1006013 (2016).
13. Liu, L. *et al.* Genetic diversity and molecular evolution of human adenovirus serotype 41 strains circulating in Beijing, China, during 2010–2019. *Infection, Genetics and Evolution* **95**, 105056 (2021).
14. Cao, M., You, H. & Hermonat, P. L. The X Gene of Adeno-Associated Virus 2 (AAV2) Is Involved in Viral DNA Replication. *PLoS One* **9**, e104596 (2014).
15. Lisowski, L. *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* **506**, 382–386 (2014).
16. Cabanes-Creus, M. *et al.* Novel human liver-tropic AAV variants define transferable domains that markedly enhance the human tropism of AAV7 and AAV8. *Mol Ther Methods Clin Dev* **24**, 88–101 (2022).
17. Cabanes-Creus, M. *et al.* Restoring the natural tropism of AAV2 vectors for human liver. *Sci Transl Med* **12**, (2020).
18. Chen, Z. *et al.* Role of humoral immunity against hepatitis B virus core antigen in the pathogenesis of acute liver failure. *Proc Natl Acad Sci U S A* **115**, E11369–E11378 (2018).
19. Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol* **15**, e8503 (2019).
20. Niu, L. *et al.* Dynamic human liver proteome atlas reveals functional insights into disease pathways. *Mol Syst Biol* **18**, e10947 (2022).

21. Timpe, J. M., Verrill, K. C. & Trempe, J. P. Effects of Adeno-Associated Virus on Adenovirus Replication and Gene Expression during Coinfection. *J Virol* **80**, 7807 (2006).
22. Yakobson, B., Koch, T. & Winocour, E. Replication of adeno-associated virus in synchronized cells without the addition of a helper virus. *J Virol* **61**, 972 (1987).
23. la Bella, T. *et al.* Adeno-associated virus in the liver: Natural history and consequences in tumour development. *Gut* **69**, 737–747 (2020).
24. Chowdary, P. *et al.* Phase 1–2 Trial of AAVS3 Gene Therapy in Patients with Hemophilia B. *New England Journal of Medicine* **387**, 237–247 (2022).
25. Chand, D. *et al.* Hepatotoxicity following administration of onasemnogene abeparvovec (AVXS-101) for the treatment of spinal muscular atrophy. *J Hepatol* **74**, 560–566 (2021).
26. Mullard, A. Gene therapy community grapples with toxicity issues, as pipeline matures. *Nat Rev Drug Discov* **20**, 804–805 (2021).
27. Morales, L., Gambhir, Y., Bennett, J. & Stedman, H. H. Broader Implications of Progressive Liver Dysfunction and Lethal Sepsis in Two Boys following Systemic High-Dose AAV. *Molecular Therapy* **28**, 1753 (2020).
28. Hüser, D. *et al.* Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome. *PLoS Pathog* **6**, 1–14 (2010).
29. Nault, J. C. *et al.* Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nature Genetics 2015 47:10* **47**, 1187–1193 (2015).
30. Dalwadi, D. A. *et al.* AAV integration in human hepatocytes. *Molecular Therapy* **29**, 2898–2909 (2021).
31. Nathwani, A. C. Gene therapy for hemophilia. *Hematology: the American Society of Hematology Education Program* **2019**, 1 (2019).
32. Ho, A. *et al.* Adeno-associated virus 2 infection in children with non-A-E hepatitis. *medRxiv* 2022.07.19.22277425 (2022) doi:10.1101/2022.07.19.22277425.
33. Perrin, G. Q., Herzog, R. W. & Markusic, D. M. Update on clinical gene therapy for hemophilia. *Blood* **133**, 407–414 (2019).
34. Divers, T. J., Tomlinson, J. E. & Tennant, B. C. The history of Theiler's disease and the search for its aetiology. *The Veterinary Journal* **287**, 105878 (2022).
35. Kelgeri, C. *et al.* Clinical Spectrum of Children with Acute Hepatitis of Unknown Cause. *https://doi.org/10.1056/NEJMoa2206704* (2022) doi:10.1056/NEJMOA2206704.
36. Sanchez, L. H. G. *et al.* A Case Series of Children with Acute Hepatitis and Human Adenovirus Infection. *https://doi.org/10.1056/NEJMoa2206294* (2022) doi:10.1056/NEJMOA2206294.
37. Yang, W. X. *et al.* Prevalence of serum neutralizing antibodies to adenovirus type 5 (Ad5) and 41 (Ad41) in children is associated with age and sanitary conditions. *Vaccine* **34**, 5579 (2016).
38. Ukhsa. National norovirus and rotavirus bulletin. Week 28 report: data to week 26 (3 July 2022).
39. Calcedo, R. *et al.* Adeno-Associated Virus Antibody Profiles in Newborns, Children, and Adolescents. *Clin Vaccine Immunol* **18**, 1586 (2011).
40. Nathwani, A. C. *et al.* Sustained high-level expression of human factor IX (hFIX) after liver-targeted delivery of recombinant adeno-associated virus encoding the hFIX gene in rhesus macaques. *Blood* **100**, 1662–1669 (2002).

Tables

Table 1: Characteristics of unexplained pediatric hepatitis cases and related specimens

CASE ID	Sex	Liver Transplant	Sender	Specim. 1	ID 1	Specim. 2	ID 2	Specim. 3	ID 3
1	M	Yes	BCH	Liver	JBL1				
2	M	Yes	BCH/PHW	Liver	JBL4	NPA	JBN1		
3	F	Yes	BCH	Liver	JBL3				
4	M	Yes	BCH/UKHSA	Liver	JBL2	Blood	JBB25		
5	F	Yes	BCH	Liver	JBL5				
6	F	No	UKHSA	Blood	JBB9	Blood	JBB14	Blood	JBB16
7	F	No	UKHSA	Blood	JBB11	Blood	JBB10		
8	F	No	UKHSA	Serum	JBPL1	Blood	JBB13		
9	M	No	UKHSA	Blood	JBB1				
10	M	No	UKHSA	Blood	JBB15				
11	NA	No	GRI	Blood	JBB2				
12	M	No	UKHSA	Blood	JBB12				
13	NA	No	GRI	Blood	JBB7				
14	NA	No	GRI	Blood	JBB8				
15	NA	No	GRI	Blood	JBB4	Blood	JBB3		
16	NA	No	GRI	Blood	JBB5				
17	F	No	UKHSA	Throat.S	JBB18	Stool	JBB17		
18	F	No	UKHSA	Blood	JBB19				
19	F	No	UKHSA	Blood	JBB20	Blood	JBB23		
20	M	No	UKHSA	Blood	JBB21				
21	NA	No	PHW	NPA	JBB26				
22	NA	No	GRI	Stool	JBB27				
23	NA	No	GRI	Throat.s	JBB28	Stool	JBB30		
24	NA	No	GRI	Stool	JBB29				
25	NA	No	NHSL	Blood	JBB31				
26	NA	No	NHSL	Stool	JBB32				
27	F	No	UKHSA	Blood	JBB24				
28	M	Yes	KCH	Liver	JBL6				
29	F	Yes	KCH	Liver	JBL7	Liver	JBL8		
30	F	No	KCH	Liver	JBL9				
31	F	Yes	KCH	Liver	JBL10				
32	M	Yes	KCH	Liver	JBL11	Serum	JBB34		
33	F	Yes	KCH	Liver	JBL12				
34	M	Yes	KCH	Liver	JBL13	Serum	JBB36		
35	F	No	KCH	Liver	JBL14	Serum	JBB35		
36	M	Yes	KCH	Liver	JBL15	Serum	JBB37		
37	F	No	KCH	Serum	JBB38				
38	M	No	KCH	Serum	JBB39				

The median age for the cases is 3 years old (age range: 1y-9y). **Case 10** was 9 years old. All other cases were aged 7 or under.

Cases 1-5 underwent liver transplant and had mNGS, PCR and viral WGS of their specimens. **Cases 28, 29, 31-34, 36** also underwent liver transplant and had PCR for all three viruses under investigation.
Cases 6-27, 30, 35, 37, 38 did not receive a liver transplant**. Cases 30 & 35** had liver biopsies**. Cases 6-10** had mNGS, PCR and viral WGS on their samples. **Cases 11-22** had PCR for 1-2 of the viruses under investigation and viral WGS of PCR positives**. Cases 23-27** only had HAdV WGS on their samples and there was no residual material for further testing**.** **Cases 31,36,38,39** had PCR for all three viruses under investigation.
NPA: Nasopharygeal aspirate BCH: Birmingham Children's Hospital, PHW: Public Health Wales, GRI: Glasgow Royal Infirmary, NHSL: NHS Lothian, KCH: King's College Hospital

Table 2: PCR, metagenomics and viral WGS results from cases where metagenomic sequencing was performed

Case ID	Sample ID	PCR CT values			Metagenomics reads						Viral WGS Coverage (10X)		
					DNA			RNA					
		AAV2	HAdV	HHV-6B	AAV2	HAdV	HHV-6B	AAV2	HAdV	HHV-6B	AAV2	HAdV	HHV-6B
Liver													
1	JBL1	17	37	29	1343	0	8	574	0	0	97	-	3
2	JBL4	21	42	32	360	0	8	49	0	0	93	-	2
3	JBL3	20	37	30	1189	0	4	95	0	0	98	-	2
4	JBL2	20	37	27	1564	0	203	42	0	0	98	-	94
5	JBL5	21	37	28	266	0	12	F	F	F	-	-	-
Blood													
6*	JBB14/ JBB16/ JBB9	24	36	37	151	0	0	77	0	0	95	35.5	-
7	JBB10/ JBB11	21	36	37	103	0	0	F	F	F	49	F	-
8	JBPL1/ JBB13	25	P/N	-/N	277	0	0	165	0	0	94	F	-
9	JBB1	19	P/-	P/-	1936	5	0	0	0	0	94	F	-
10	JBB15	-/N	N/N	37	0	0	0	F	F	F	-	F	-

- : Not tested (at GOSH due to insufficient residual material)

N: negative PCR result

P: Positive PCR result in referring laboratory

Where two results are shown, the first refers to the referring laboratory and the second to GOSH.

Where there was a discrepancy, the positive result is shown.

F: Failed

Where there is more than one sample for a single patient, CT values represent the mean across the samples that were tested.

*Metagenomics reads: the result of combining the datasets from two blood samples from the same case

De novo assembly of unclassified metagenomics reads was unremarkable

Figure Legends

Figure 1: HAdV Epidemiology and experimental outline

a, HAdV in all sample types; epidemiology since January 2022. Source: secondary Generation Surveillance system data, ie laboratory reports to UKHSA of a positive adenovirus result conducted by a laboratory in England, and includes any sample type. Dots represent the day of presentation for the 28/38 cases for which we had data, in green the liver-transplant cases and in red the non-transplant cases. **b,** Case and control specimens by source. **c,** Tests carried out by specimen type. More detail on samples tested and the results can be found in Tables 1 and 2. Not all tests were carried out on all samples due to lack of material. N refers to the total number of cases/controls. Numbers of each sample type may not sum to this total because samples of more than one type were sometimes taken from the same patient. For details, see Table 1.

Figure 2: Proportion of positive cases and viral loads (CT values) for cases and controls

* indicates immunocompromised comparators. Proportion of PCR positive and negative results for **a** AAV2, **b** HAdV and **c** HHV-6. CT values < 38 were defined as positive. CT >38 where the virus was detected within the maximum 45 cycles were defined as low-level positive (LLP). **d,** AAV2 in blood from cases, PERFORM /DIAMOND immunocompetent controls and immunocompromised comparators. Blue: HAdV infection, green: non-HAdV hepatitis, red: healthy. **e,** HAdV levels in whole blood from cases and immunocompromised comparators. **f,** HHV-6 in whole blood from cases and immunocompromised comparators. **g,** HAdV, AAV2 and HHV-6 levels in frozen liver tissue from cases and immunocompromised comparators. In the box plots, the bold middle line represents the median and the upper and lower horizontal lines represent the upper (75^{th} percentile) and lower (25^{th} percentile) quartiles respectively. Whiskers show maximum and minimum values. Each point represents one case or control. Where more than one sample for a case was tested, the midpoint of the CT has been plotted. All repeat tests had values <2CTs apart, ie within the limits of methodological error. The dotted line marked LLP indicates the low-level positive threshold (CT=38). Points below the second dotted line represent samples below the limit of PCR detection (CT=45). Wilcoxon non-parametric rank sum tests were conducted for **e & g** and a Kruskal-Wallis test followed by pairwise Wilcoxon tests with a Benjamini-Hochberg correction for multiple comparisons for **d & f**. All tests were two-tailed. Numbers show the p-value compared to cases. NS: not significant. tr: received liver transplant.

Figure 3: Phylogenetic trees for HAdV, AAV2 and HHV-6B

Maximum likelihood phylogenetic trees combining reference sequences from the RefSeq database, publicly available complete genomes from GenBank, UK non-outbreak controls (open squares) and unexplained hepatitis cases (black squares) for the different viruses involved: **a** HAdV **b** AAV2 and **c** HHV-6. HAdV and HHV-6B trees are mid-point rooted, while AAV2 is rooted the RefSeq sequence: NC_001401.2. Bootstrap values less than 90 are not shown.

Figure 4: Transcriptomic and proteomic analysis of case liver samples

Transcriptomic analysis was conducted for the five frozen case liver samples from transplanted patients. **a,** Expression of cytokine-inducible transcriptional modules in normal liver, and AAV2 (n=4) or HBV (n=17) associated hepatitis requiring transplantation are shown as DZ scores for the expression of each module, reflecting the difference from the average score from normal liver (n=10) data sets. Each point represents the score form a single data set/sample. **b & c**, Volcano plots of differentially expressed proteins **(b)** and peptides **(c)**. The volcano plots illustrate fold changes and corresponding p-values for the comparison between 5 liver explants and 7 control healthy livers. Each dot represents a protein/peptide. The p-values were calculated by applying two-tailed empirical Bayes moderated t-statistics on protein/peptide-wise linear models. Proteins **(b)** and peptides **(c)** differentially expressed (absolute log2(fold change) > 6 and P < 1e-07) are coloured as red (up-regulated) and blue (down-regulated). The p-values illustrated here are not adjusted for multiple comparisons. Full tables can be found in **Supplementary Tables 12-14**.

METHODS

Ethics

Metagenomic analysis and adenovirus sequencing were carried out by the routine diagnostic service at Great Ormond Street Hospital. Additional PCRs, Immunohistochemistry and proteomics on samples received for metagenomics are part of the Great Ormond Street Hospital (GOSH) protocol for confirmation of new and unexpected pathogens. The use for research of anonymised laboratory request data, diagnostic results and residual material from any specimen received in the GOSH diagnostic laboratory, including all cases received from Birmingham's Children Hospital UKHSA, Public Health Wales, Public health Scotland as well as non-case samples from UKHSA, Public Health Scotland and Great Ormond Street Hospital research was approved by UCL Partners Pathogen Biobank under ethical approval granted by the NRES Committee London-Fulham (REC reference: 17/LO/1530). Children undergoing liver transplant were consented for additional research under the International Severe Acute Respiratory and Emerging Infections Con Ethics sortium (ISARIC) WHO Clinical Characterisation Protocol UK (CCP-UK) [ISRCTN 66726260] (RQ3001-0591, RQ301-0594, RQ301-0596, RQ301-0597, RQ301-0598). Ethical approval for the ISARIC CCP-UK study was given by the South Central–Oxford C Research Ethics Committee in England (13/SC/0149), the Scotland A Research Ethics Committee (20/SS/0028), and the WHO Ethics Review Committee (RPC571 and RPC572).

The United Kingdom Health Security Agency (UKHSA) has legal permission, provided by Regulation 3 of The Health Service (Control of Patient Information) Regulations 2002, to process patient confidential information for national surveillance of communicable diseases and as such, individual patient consent is not required.

Control subjects from the EU horizon 2020 research and innovation program DIAMONDS/PERFORM (grant agreement No. 668303 and 848196) were recruited according to the approved enrolment procedures of each study, and with the informed consent of parents or guardians: DIAMONDS (London – Dulwich Research Ethics Committee: 20/HRA/1714); PERFORM (London – Central Research Ethics Committee: 16/LO/1684).

The sample IDs for the cases and controls are anonymised IDs that cannot reveal the identity of the study subjects and are not known to anyone outside the research group, such as the patients or the hospital staff.

Samples

Initial diagnostic testing by metagenomics and PCR was performed at Great Ormond Street Hospital Microbiology and Virology clinical laboratories. Further whole genome sequencing and characterization was performed at UCL.

Cases

Birmingham Children's Hospital provided us with explanted liver tissue from five biopsy sites from five cases, five whole blood 500ul from four cases and serum plasma from one case (**Table 1, Figure 1B**). These were used in metagenomics testing (**Table 2**), followed by HAdV, HHV-6 and AAV2 testing by PCR and, depending on CT value, whole genome sequencing **(Supplementary Table 7, 9, 10).** We subsequently received 25 additional specimens from UKHSA, Public Health Wales and Public Health Scotland / Edinburgh Royal Infirmary, including 16 additional blood samples, four respiratory specimens and five stool samples, for HAdV WGS and depending on residual material for AAV2 PCR testing followed by sequencing (**Table 1, Table 2**, **Figure 1B, Supplementary Table 7, 9, 10**). We also received 10 formalin fixed, paraffin embedded (FFPE) liver biopsy samples and 6 serum samples from 11 cases from King's College Hospital (**Table 1**). Of these cases, 7 had received liver transplants.

Controls from DIAMONDS and PERFORM

PERFORM (Personalised Risk assessment in Febrile illness to Optimise Real-life Management across the European Union) recruited children from 10 EU countries (2016-2020. PERFORM was funded by the European Union's Horizon 2020 program under GA No 668303.

DIAMONDS (Diagnosis and Management of Febrile Illness using RNA Personalised Molecular Signature Diagnosis) is funded by the European Union Horizon 2020 program grant number 848196. Recruitment commenced in 2020 and is ongoing. Both studies recruited children presenting with suspected infection or inflammation and assigned them to diagnostic groups according to a standardised algorithm.

Controls from GOSH for PCR

Blood samples from 17 patients not linked to the non-A-E hepatitis outbreak were tested by real-time PCR targeting AAV2 **(Extended Data Table 2B)**. These comparators were patients with ALT/AST >500 and HAdV or CMV viraemia. These were purified DNA from residual diagnostic specimens received in the GOSH Microbiology and Virology laboratory in the previous year. All residual specimens were stored at -80 ^{o}C prior to testing and pseudo-anonymised at the point of processing and analysis. Viraemia was initially detected using targeted real-time PCR during routine diagnostic testing with UKAS-accredited lab-developed assays that conform to ISO:15189 standards.

In addition to the blood samples, four residual liver biopsies from four control patients referred for investigation of infection were tested by AAV2 and HHV-6B PCR. The liver biopsies were submitted to the GOSH microbiology laboratory for routine diagnosis by bacterial broad-range 16S rRNA gene PCR or metagenomics testing in 2021 and 2022. 3/4 of the control patients were known to have elevated liver enzymes. Two adult frozen liver

samples previously tested by metagenomics were negative for AAV2 and positive for HHV6 (**Supplementary Table 5**).

Controls from UKHSA

We received a blood sample from one patient with raised liver enzymes and HAdV infection. We also received one control stool sample from Public Health Scotland/Edinburgh Royal Infirmary and 22 control stool samples for sequencing.

Controls from King's College Hospital

A single formalin fixed paraffin embedded (FFPE) liver biopsy control of normal marginal tissue from a hepatoblastoma from a child was negative for AAV2 and HAdV, but positive for HHV-6B (CT = 37).

Controls from QMUL

We received FFPE liver control samples from 10 adults and 3 children (under 18) with other viral hepatitis, toxic liver necrosis, autoimmune and other hepatitis, and normal liver, from Queen Mary University of London. PCR gave valid results for samples from 2 children and 8 adults, all of which were negative by PCR for AAV2 and HHV6, apart from one adult sample which was positive for HHV6 at high CT value (**Supplementary Table 5**).

Metagenomic sequencing

Nucleic acid purification

Frozen liver biopsies were infused overnight at -20°C with RNAlater-ICE. Up to 20 mg biopsy was lysed with 1.4mm ceramic, 0.1mm silica and 4mm glass beads, prior to DNA and RNA purification using the Qiagen AllPrep DNA/RNA Mini kit as per manufacturers' instructions, with a 30 µl elution volume for RNA and 50 µl for DNA.

Up to 400 µl whole blood was lysed with 0.5mm and 0.1 mm glass beads prior to DNA and RNA purification on a Qiagen EZ1 instrument with an EZ1 virus mini kit as per manufacturer's instructions, with a 60 µl elution volume.

For quality assurance, every batch of samples was accompanied by a control sample containing feline calicivirus RNA and cowpox DNA which was processed alongside clinical specimens, from nucleic acid purification through to sequencing. All specimens and controls were spiked with MS2 phage RNA internal control prior to nucleic acid purification.

Library preparation and sequencing

RNA from whole blood samples with RNA yield >2.5 ng/µl and from biopsies underwent ribosomal RNA depletion and library preparation with KAPA RNA HyperPrep kit with RiboErase, according to manufacturer's instructions. RNA from whole blood with RNA yield <2.5 ng/µl did not undergo rRNA depletion prior to library preparation.

DNA from whole blood samples with DNA yield >1 ng/µl and from biopsies underwent depletion of CpG-methylated DNA using the NEBNext® Microbiome DNA Enrichment Kit, followed by library preparation with NEBNext Ultra II FS DNA Library Prep Kit for Illumina, according to manufacturer's instructions. DNA from whole blood with DNA yield <1 ng/µl did not undergo depletion of CpG-methylated DNA prior to library preparation.

Sequencing was performed with a NextSeq High output 150 cycle kit with a maximum of 12 libraries pooled per run, including controls.

Metagenomics data analysis

Pre-processing pipeline

An initial quality control step was performed by trimming adapters and low-quality ends from the reads (Trim Galore![41]version 0.3.7). Human sequences were then removed using the human reference GRCH38 p.9 (Bowtie2 [42], version 2.4.1) followed by removal of low quality and low complexity sequences (PrinSeq[43], version 0.20.3). An additional step of human seq removal followed (megaBLAST[44], version 2.9.0). For RNA-seq, ribosomal RNA sequences were also removed using a similar 2 step-approach (Bowtie2 and megaBLAST). Finally, nucleotide similarity and protein similarity searches were performed (megaBLAST and DIAMOND[45] (version 0.9.30) respectively) against custom reference databases that consisted of nucleotide and protein sequences of the RefSeq collections (downloaded March 2020) for viruses, bacteria, fungi, parasites and human.

Taxonomic classification

DNA and RNA sequence data was analysed with metaMix[5] (version 0.4) nucleotide and protein analysis pipelines.

metaMix resolves metagenomics mixtures using Bayesian mixture models and parallel MCMC search of the potential species space to infer the most likely species profile.

metaMix considers all reads simultaneously to infer relative abundances and probabilistically assign the reads to the species most likely to be present. It uses an 'unknown' category to capture the fact that some reads cannot be assigned to any species. The resulting metagenomic profile includes posterior probabilities of species presence as well as Bayes factor for presence versus absence of specific species. There are two modes, metaMix-

protein, which is optimal for RNA virus detection and metaMix-nucl, which is best for speciation of DNA microbes. Both modes were used for RNA-seq while metaMix-nucl for DNA-seq.

For sequence results to be valid, MS2 phage RNA had to be detected in every sample and feline calicivirus RNA and cowpox DNA, with no additional unexpected organisms, detected in the controls.

Confirmatory mapping of AAV2

The RNA-Seq reads were mapped to the AAV2 reference genome (NCBI reference sequence NC_001401) using Bowtie2, with the –very-sensitive option. Samtools[46] version 1.9) and Picard (version 2.26.9, http://broadinstitute.github.io/picard/) were used to sort, deduplicate and index the alignments, and to create a depth file, which was plotted using a custom script in R.

de novo assembly of unclassified reads

We performed a *de novo* assembly step with metaSPADES[47](v3.15.5), using all the reads with no matches to the nucleotide database we used for our similarity search. A search using megaBLAST with the standard nucleotide collection was carried out on all resulting contigs over 1000bp in length. All of the contigs longer than 1000bp matched to human, except two which mapped to Torque Teno virus (TTV).

Nanopore Sequencing

DNA from up to 20 mg of liver was purified using the Qiagen DNeasy Blood & Tissue kit as per manufacturer's instructions. Samples with limited amount of DNA were fragmented to an average size of 10kb using a Megaruptor 3 (Diagenode) to reach an optimal molar concentration for library preparation. QC was perform using a Femto Pulse System (Agilent Technologies) and a Qubit fluorometer (Invitrogen). Samples were prepared for Nanopore sequencing using the Ligation Sequencing Kit SQK-LSK110. DNA was sequenced on a PromethION using R9.4.1 flowcells (Oxford Nanopore Technologies). Samples were run for 72 hours including a washing and reload step after 24 and 48 hours.

All library preparation and sequencing were performed by UCL Long Read Sequencing facility.

Passed reads were mapped to the reference AAV2 genome (NC_001401) using minimap2[48] using the default parameters. Reads that also mapped by minimap to the human genome (Ensemble GRCh38_v107), which could be ligation artefacts, were excluded from further analysis. The passed reads were also classified using Kraken2[49] with the PlusPF database (5/17/2021). The data relating to AAV2 reads in Supplementary Table 3 refer to reads that

were classified as AAV2 by both minimap2 and Kraken2, since the results from both methods were similar. Four reads across all four lower-depth samples were classified as HHV-6B by the EPI2ME WIMP [50]pipeline. No reads were classified as HAdV or HHV-6B by Kraken2 in the two higher-depth samples. Alignment dot plots were created for the AAV2 reads using redotable[51], with a window size of 20. These were manually classified into possible complex and monomeric structures.

Integration analysis of Illumina data

We investigated potential integrations of AAV2 and HHV-6 viruses into the genome using the Illumina metagenomics data for 5 liver transplant cases. We first processed the pair-end reads (average sequence coverage per genome=5x), first quality checking using FastQC[52], with barcode and adaptor sequence trimmed by TrimGalore (phred-score=20). Potential viral integrations were investigated with Vseq-Toolkit[53] (Mode 3 with default settings except for high stringency levels). Predicted genomic integrations were visualized with IGV[54], requiring at least 3 reads supporting an integration site, spanning both human and viral sequences. Predicted integrations were supported by only one read, thus not fulfilling the algorithm criteria. Sequencing was performed at a lower depth than optimal for integration analysis but no evidence was found for AAV2 or HHV-6B integration into cases' genomes.

PCR

Real-time PCR targeting a 62 nt region of the AAV2 inverted terminal repeat (ITR) sequence was performed using primers and probes previously described[55]. This assay is predicted to amplify AAV2 and AAV6. The Qiagen QuantiNova probe PCR kit (PERFORM and DIAMONDS controls) or Qiagen Quantifast probe PCR kit (all other samples) were used. Each 25 µl reaction consisted of 0.1 µM forward primer, 0.34 µM reverse primer, 0.1 µM probe with 5 µl template DNA.

Real-time PCR targeting a 74 bp region of the HHV6 DNA polymerase gene was performed using primers and probes previously described[56] multiplexed with an internal positive control targeting mouse (*mus*) DNA spiked into each sample during DNA purification, as previously described[57]. Briefly, each 25 µl reaction consisted of 0.5 µM each primer, 0.3 µM HHV-6 probe, 0.12 µM each *mus* primer, 0.08 µM *mus* probe and 12.5 µl Qiagen Quantifast Fast mastermix with 10 µl template DNA.

Real-time PCR targeting a 132 bp region of the Adenovirus hexon gene was performed using primers and probes previously described[58] multiplexed with an internal positive control targeting mouse (*mus*) DNA spiked into each sample during DNA purification, as previously described[57]. Briefly, each 25 µl reaction consisted of 0.6 µM each HHV6 primer, 0.4 µM HHV6 probe, 0.12 µM each *mus* primer, 0.08 µM *mus* probe and 12.5 µl Qiagen Quantifast Fast mastermix with 10 µl template DNA.

PCR cycling for all targets, apart from the controls from the PERFORM and DIAMONDS studies, was performed on an ABI 7500 Fast thermocycler and consisted of 95 ^{o}C for 5

minutes followed by 45 cycles of 95 °C for 30 seconds and 60 °C for 30 seconds. For the PERFORM and DIAMONDS controls, PCR was performed on a StepOnePlus™ Real-Time PCR System and consisted of 95 °C for 2 minutes followed by 45 cycles of 95 °C for 5 seconds and 60 °C for 10 seconds. Each PCR run included a no template control and a DNA positive control for each target.

Neat DNA extracts of the FFPE material were inhibitory to PCR so PCR results shown were performed following a 1 in 10 dilution,

AAV2 RT-qPCR

RNA samples were treated with Turbo-DNA free kit (Thermo) to remove residual genomic DNA. cDNA was synthesised using QuantiTect Reverse Transcription kit. Briefly, 12 µl of RNA were mixed with 2 µl of gDNA Wipeout buffer and incubated at 42 °C for 2 minutes and transferred to ice. 6 µl of reverse transcription mastermix and incubated at 42 °C for 20 min followed by 3 min at 95 °C.

Real-time PCR targeting a 120 nt region of the AAV2 *cap* ORF sequence was performed using primers *AAV2_cap _Fw*- ATCCTTCGACCACCTTCAGT, *AAV2_cap _Rv* – GATT CCAGCGTTTGCTGTT and probe *AAV2_cap _Pr* FAM-ACACAGTAT/ZEN/TCC ACGG GACAGGT-IBFQ. This assay is predicted to amplify AAV2 and AAV6. The Qiagen QuantiNova probe PCR kit was used. Each 25 µl reaction consisted of 0.1 µM forward primer, 0.1 µM reverse primer, 0.2 µM probe with 2.5 µl template cDNA.

PCR was performed on a StepOnePlus™ Real-Time PCR System and consisted of 95 °C for 2 minutes followed by 45 cycles of 95 °C for 5 seconds and 60 °C for 10 seconds. Each PCR run included a no template control, a DNA positive control and a RNA control from each sample to verify efficient removal of gDNA.

Immunohistochemistry (IHC)

All IHC was done on Formalin Fixed Paraffin Embedded tissue cut at 3µm thickness.

Adenovirus

Adenovirus immunohistochemistry was carried out using the Ventana Benchmark ULTRA, Optiview Detection Kit, PIER with Protease 1 for 4min, Ab incubation 32min (Adenovirus clone 2/6 & 20/11, Roche, 760-4870, pre-diluted). The positive control was a known Adenovirus positive gastrointestinal surgical case.

Preparation of AAV2 positive controls

The plasmid used for transfection was pAAV2/2 (addgene, Plasmid #104963, https://www.addgene.org/104963/) which expresses the Rep/Cap genes of AAV2. This was delivered by tail-vein hydrodynamic injection[59] into albino C57Bl/6 mice (5 microgrammes in 2 mls PBS). Negative controls received PBS alone. At 48 hours, mice were terminally

exsanguinated and perfused by PBS. Livers were collected into 10% Neutral Buffered Formalin (CellPath UK). This was performed under Home Office License PAD4E6357.

AAV2 immunohistochemistry was carried out with four commercially available antibodies:

- Leica Bond-III, Bond Polymer Refine Detection Kit with DAB Enhancer, HIER with Bond Epitope Retrieval Solution 1 (citrate based pH 6) for 30min, Ab incubation 30min (Anti-AAV VP1/VP2/VP3 clone B1, PROGEN, 690058S, 1:100).
- Leica Bond-III, Bond Polymer Refine Detection Kit with DAB Enhancer, HIER with Bond Epitope Retrieval Solution 1 (citrate based pH 6) for 40min, Ab incubation 30min (Anti-AAV VP1/VP2/VP3 rabbit polyclonal, OriGene, BP5024, 1:100)
- • Leica Bond-III, Bond Polymer Refine Detection Kit with DAB Enhancer, HIER with Bond Epitope Retrieval Solution 1 (citrate based pH 6) for 40min, Ab incubation 30min (Anti-AAV VP1 clone A1, OriGene, BM5013, 1:100).
- • Leica Bond-III, Bond Polymer Refine Detection Kit with DAB Enhancer, HIER with Bond Epitope Retrieval Solution 1 (citrate based pH 6) for 40min, Ab incubation 30min (Anti-AAV VP1/VP2 clone A69, OriGene, BM5014, 1:100)

HHV6 immunohistochemistry straining was carried out with:

- Leica Bond-III, Bond Polymer Refine Detection Kit with DAB Enhancer, PIER with Bond Enzyme 1 Kit 10min, Ab incubation 30min (Mouse monoclonal [C3108-103] to HHV6, ABCAM, ab128404, 1:100).

Negative reagent control slides were stained using the same antigen retrieval conditions and staining protocol incubation times using only BondTM Primary Antibody Diluent #AR9352 for the antibody incubation.

Electron Microscopy

Samples of liver were fixed in 2.5% glutaraldehyde in 0.1M cacodylate buffer followed by secondary fixation in 1.0% osmium tetroxide. Tissues were dehydrated in graded ethanol, transferred to an intermediate reagent, propylene oxide and then infiltrated and embedded in Agar 100 epoxy resin. Polymerisation was undertaken at 60 ºC for 48 hours. 90nm ultrathin sections were cut using a Diatome diamond knife on a Leica UC7 ultramicrotome. Sections were transferred to copper grids and stained with alcoholic urynal acetate and Reynold's lead citrate. The samples were examined using a JEOL 1400 transmission electron microscope. Images were captured on an AMT XR80 digital camera.

Whole genome sequencing

Bait Design

To produce the capture probes for hybridisation, biotinylated RNA oligonucleotides (baits) used in the SureSelectXT protocols for HAdV and HHV6 WGS were designed in-house using Agilent community design baits with part numbers 5191-6711 and 5191-6713 respectively. They were synthesised by Agilent Technologies, Santa Clara, California (Agilent Technologies, 2021) (available through Agilent's Community Designs programme: SSXT CD Pan Adenovirus and SSXT CD Pan HHV6 and used previously [60,61]).

Library preparation and sequencing

For whole genome sequencing of HAdV and HHV-6B, DNA (bulked with male human gDNA (Promega) if required) was sheared using a Covaris E220 focused ultra-sonication system (PIP 75, duty factor 10, cycles per burst 1000).End-repair, non-templated addition of 3′ poly A, adapter ligation, hybridisation, PCR (pre-capture cycles dependent on DNA input and post capture cycles dependent on viral load), and all post-reaction clean-up steps were performed according to either the SureSelectXT Low Input Target Enrichment for Illumina Paired-End Multiplexed Sequencing protocol (version A0), the SureSelectXT Target Enrichment for Illumina Paired-End Multiplexed Sequencing protocol (version C3)or SureSelectXTHS Target Enrichment using the Magnis NGS Prep System protocol (version A0) (Agilent Technologies). Quality control steps were performed on the 4200 TapeStation (Agilent Technologies). Samples were sequenced using the Illumina MiSeq platform. Base calling and sample demultiplexing were performed as standard for the MiSeq platform, generating paired FASTQ files for each sample. A negative control was included on each processing run. A targeted enrichment approach was used due to the predicted high variability of the HHV-6 and HAdV genomes.

For AAV2 WGS, an AAV2 primer scheme was designed using primalscheme[62] with 17 AAV2 sequences from NCBI and 1 AAV2 sequence provided by GOSH from metagenomic sequencing of a liver biopsy DNA extract as the reference material. These primers amplify 15 overlapping 400 bp amplicons. Primers were supplied by Merck. Two multiplex PCR reactions were prepared using Q5® Hot Start High-Fidelity 2X Master Mix, with a 65^{o}C, 3 min annealing/extension temperature. Pool 1 and 2 multiplex PCRs were run for 35 cycles. 10uL of each PCR reaction were combined and 20uL nuclease-free water added. Libraries were prepared either manually or on the Agilent Bravo NGS workstation option B, following a reduced-scale version of the Illumina DNA protocol as used in the CoronaHiT protocol[63]. Equal volumes of the final libraries were pooled, bead purified and sequenced on the Illumina MiSeq. A negative control was included on each processing run.

All library preparation and sequencing were performed by UCL Genomics.

AAV2 Sequence Analysis

The raw fastq reads were adapted, trimmed and low-quality reads removed. The reads were mapped to NC_001401 reference sequence and then the amplicon primers regions were trimmed using the location provided in a bed file. Consensus sequences were then called at a

minimum of 10X coverage. The entire processing of raw reads to consensus was carried out using nf-core/viralrecon pipeline (https://nf-co.re/viralrecon/2.4.1) (doi:https://doi.org/10.5281/zenodo.3901628). Basic quality metrics for the samples sequenced are in **Supplementary Table 9**. All samples that gave 10x genome coverage over 90% were then used for further phylogenetic analysis. Samples were aligned along with known reference strains from genbank using MAFFT[64] (version v7.271) and the trees were built with IQ-TREE[65] (multicore version 1.6.12) with 1000 rapid bootstraps and aLRT support. The samples were then labelled based on type and provider on the trees (Fig 3A).

For each AAV2 sample, we aligned the consensus nucleotide sequence to the AAV2 reference sequence. From these alignments, the exact coordinates of the sample capsid were determined. We then used the coordinates to extract the corresponding nucleotide sequence and translated it to find the amino acid sequence. We then compared each sample to the reference to identify amino acid changes. Amino acid sequences from AAV capsid sequences were retrieved from GenBank for AAV1 to AAV12. Amino acid sequences of capsid constructs designed to be more hepatotropic were retrieved from [16,66]. These sequence sets were then aligned to the AAV2 reference sequence using MAFFT[64] . We then compared each construct to the AAV2 reference to identify amino acid changes present, while retaining the AAV2 coordinate set.

HAdV and HHV-6B sequence analysis

Raw data quality control is performed using trim-galore (v.0.6.7) on the raw FASTQ files.

For HHV-6B, short reads were mapped with BWA mem[67] (0.7.17-r1188) using the RefSeq reference NC_000898.

For adenovirus, genotyping is performed using AYUKA[11](version 22-111). This novel tool is used to confidently assign one or more adenovirus genotypes to a sample of interest, assessing inter-genotype recombination if more than one genotype detected. The results from this screening step guide which downstream analyses are performed, and which reference genome(s) are used. If mixed infection is suspected, reads are separated using bbsplit (https://sourceforge.net/projects/bbmap/), and each genotype is analysed independently as normal. If recombination is suspected, a more detailed analysis is performed using RDP and the sample is excluded from phylogenetic analysis. After genotyping, the cleaned read data is mapped using BWA to the relevant reference sequence(s), single nucleotide polymorphisms and small insertions and deletions are called using bcftool (version1.15.1, https://github.com/samtools/bcftools) and a consensus sequence is generated also with bcftools, masking with Ns positions that do not have enough read support (15X by default). Consensus sequences generated with the pipeline are then concatenated to previously sequenced samples and a multiple sequence alignment is performed using the G-INS-I algorithm in the MAFFT software (MAFFT G-INS-I v7.481). The multiple sequence alignment is then used for phylogenetic analysis with IQ-TREE (IQ-TREE 2 2.2.0), using modelfinder and performing 1000 rapid bootstraps.

Proteomics Data generation

Liver explant tissue from cases was homogenized in lysis buffer, 100 mM Tris (pH 8.5), 5% Sodium dodecyl sulfate, 5 mM tris(2-carboxyethyl)phosphine, 20 mM chloroacetamide then heated at 95 degrees for 10 minutes and sonicated in ultrasonic bath for other 10. The lysed proteins were quantified with NanoDrop 2000 (Thermo Fisher Scientific). 100 µg were precipitated with Methanol/Chloroform protocol and then protein pellets were reconstituted in 100 mM tris (pH 8.5) and 4% sodium deoxycholate (SDC). The proteins were subjected to proteolysis with 1:50 trypsin overnight at 37°C with constant shaking. Digestion was stopped by adding 1% trifluoroacetic acid to a final concentration of 0.5%. Precipitated SDC was removed by centrifugation at 10,000g for 5 min, and the supernatant containing digested peptides was desalted on an SOLAµ HRP (Thermo Fisher Scientific). 50 µg of the desalted peptide were then fractionated on Vanquish HPLC (Thermo Fisher Scientific) using a Acquity BEH C18 column (2.1 x 50 mm with 1.7µm particles from Waters): buffer A was 10 mM ammonium formiate at pH 10, while buffer B was 80% Acetonitrile and the flow was set to 500µL/min. We used a gradient of 8 minutes to collect 24 fractions that were then concatenated to obtain 12. These 12 fractions were dried and dissolved in 2% formic acid before liquid chromatography–tandem mass spectrometry (MS/MS) analysis. An estimated total of 2000 ng from each fraction was analysed using an Ultimate3000 high-performance liquid chromatography system coupled online to an Eclipse mass spectrometer (Thermo Fisher Scientific). Buffer A consisted of water acidified with 0.1% formic acid, while buffer B was 80% acetonitrile and 20% water with 0.1% formic acid. The peptides were first trapped for 1 min at 30 µl/min with 100% buffer A on a trap (0.3 mm by 5 mm with PepMap C18, 5 µm, 100 Å; Thermo Fisher Scientific); after trapping, the peptides were separated by a 50-cm analytical column (Acclaim PepMap, 3 µm; Thermo Fisher Scientific). The gradient was 9 to 35% B in 103 min at 300 nl/min. Buffer B was then raised to 55% in 2 min and increased to 99% for the cleaning step. Peptides were ionized using a spray voltage of 2.1 kV and a capillary heated at 280°C. The mass spectrometer was set to acquire full-scan MS spectra (350 to 1400 mass/charge ratio) for a maximum injection time set to Auto at a mass resolution of 120,000 and an automated gain control (AGC) target value of 100%. For a second the most intense precursor ions were selected for MS/MS. HCD fragmentation was performed in the HCD cell, with the readout in the Orbitrap mass analyser at a resolution of 15,000 (isolation window of 3 Th) and an AGC target value of 200% with a maximum injection time set to Auto and a normalized collision energy of 30%. All raw files were analysed by MaxQuant[68] v2.1 software using the integrated Andromeda search engine and searched against the Human UniProt Reference Proteome (February release with 79,057 protein sequences) together with UniProt reported AAVs proteins and specific fasta created using EMBOSS Sixpack translating patient's virus genome. MaxQuant was used with the standard parameters with only the addition of deamidation (N) as variable modification. Data analysis was then carried out with Perseus[69] v2.05: Proteins reported in the file "proteinGroups.txt" were filtered for reverse and potential contaminants. Figures were created using Origin pro version 2022b.

Transduction of AAV2 capsid mutants

A transgene sequence containing enhanced green fluorescent protein (EGFP) was packaged into rAAV2 particles to track their expression in transduced cells, compared with rAAV capsids derived from canonical AAV2, AAV9, and a synthetic liver-tropic AAV vector called LK03[15].

rAAV vector particles were delivered to Huh-7 hepatocytes at a multiplicity of infection (MOI) of 100,000 vector genomes per cell before analysing EGFP expression by flow cytometry 72-hours later.

Recombinant AAV capsid sequence

The VP1 sequence was generated by generating a consensus sequence from a multiple sequence alignment of sequenced AAV2 genomes derived from patient samples, using Biopython[70] package AlignIO. The designed VP1 sequence was then synthesised as a 'gBlock' (Integrated DNA Technologies) and incorporated into an AAV2 RepCap plasmid (AAV2/2 a gift from Melina Fan, Addgene plasmid # 104963) between the SwaI and XmaI restriction sites, using InFusion cloning reagent (Clontech product 638948).

AAV vector production

rAAV particles were generated by transient transfection of HEK 293T cells as described previously[71]. Briefly, 1.8 x 10^7 cells were plated in 15cm dishes before transfecting the pAAV-CAG-EGFP transgene plasmid (a gift from Edward Boyden, Addgene plasmid # 37825), the relevant RepCap plasmid, and the pAdDeltaF6 helper plasmid (a gift from James M. Wilson, Addgene plasmid # 112867), at a ratio of 10.5 µg, 10.5 µg, and 30.5 µg, respectively, using PEIPro transfection reagent (PolyPlus) at a ratio of 1µL per 1µg DNA. 72-hours post-transfection, cell pellets and supernatant were harvested and rAAV particles were purified using an Akta HPLC platform. rAAV particle genome copy numbers were calculated by qPCR targeting the vector transgene region. The rAAV2 vector used in this study was purchased as ready-to-use AAV2 particles from Addgene (Addgene viral prep # 37825-AAV2).

Analysis of rAAV transduction

Huh-7 hepatocytes (a gift from Dr Julien Baruteau, UCL) were plated in DMEM medium supplemented with 10% Foetal Bovine Serum and 1% Penicillin Streptomycin supplement. Cells were plated at a density of 1.5 x 10^3 cells per cm^2 and transduced with 1 x 10^5 viral genomes per cell. Transductions were performed in the presence or absence of 400 µg/mL heparin which was supplemented directly to cell media. 72-hours after transduction, cells were analysed by microscopy using an EVOS Cell Imaging System (Thermo Fisher Scientific) before quantifying EGFP expression by flow cytometry using a Cytoflex Flow Cytometer (Beckman). EGFP positive cells were determined by gating the live cell population and quantifying the level of EGFP signal versus untransduced controls.

Human Short Read Data Analysis

Transcriptomics: cytokine analysis

Cytokine inducible gene expression modules were derived from previously published bulk tissue genome-wide transcriptomes of the tuberculin skin test that have been shown to reflect canonical human in vivo cell mediated immune pathways[72] using a validated bioinformatic approach[73]. Cytokine regulators of genes enriched in the tuberculin skin[72] test (ArrayExpress Accession Number E-MTAB-6816) were identified using Ingenuity Pathway Analysis (Qiagen, Venlo, The Netherlands). Average correlation of Log2 transformed transcripts per million (TPM) data for every gene-pair in each of the target gene modules was compared to 100 iterations of randomly selected gene modules of the same size, to select cytokine-inducible modules that showed significantly greater co-correlation (adjusted p value<0.05), representing co-regulated transcriptional networks for each 59 cytokines. We then used the average Log2 TPM expression of all the genes in each these co-regulated modules to quantify the biological activity of the associated upstream cytokine within bulk genome-wide transcriptional profiles from AAV2-associated hepatitis (n=4) obtained in the present study, compared to published Log2 transformed and normalised microarray data from normal adult liver (n=10) and hepatitis B adult liver (n=17)(Gene Expression Omnibus Accession Number GSE96851)[18] . To enable comparison across the data sets, we transformed average gene expression values for each cytokine-inducible module to standardised (Z scores) using mean and standard deviation of randomly selected gene sets of the same size within each individual data set. Statistical significant differences in Z scores between groups were identified by t-tests with multiple testing correction (adjusted p value<0.05).

Proteomics differential expression

To compare the proteomics data from our cases' explanted livers with data from healthy livers, we downloaded the raw files from 2 studies[19,20] from PRIDE. The raw files were searched together with our files using the same settings and databases.

We performed differential expression analyses at protein-level and peptide-level using a hybrid approach including statistical inference on the abundance (quantitative approach) as well as presence/absence (binary approach) of proteins/peptides. DEP R package version 1.18.0 was used for the quantitative analysis[74]. Proteins/peptides were filtered for those detected in all replicates of at least one group (case or control). The data were background corrected and variance normalized using variance stabilizing transformation. Missing intensity values were not distributed randomly and were biased to specific samples (either cases or controls). Therefore, for imputing the missing data, we applied random draws from a manually defined left-shifted Gaussian distribution using the DEP *impute* function with parameters *fun:"man"*, *shift:1.8*, and *scale:0.3*. The *test_diff* function based on linear models and empirical Bayes method was used for testing differential expressions between the case and control samples.

HLA typing methods

Typing was undertaken in the liver centre units. Next Generation Sequencing (Sequencing by synthesis (Illumina) using AllType kits (VHBio/OneLambda) – high resolution HLA typing method.

Statistical analysis

Fisher's exact test and two-sided Wilcoxon (Mann-Whitney) non-parametric rank sum test were used for differences between case and control groups. Where multiple groups were compared, Kruskal-Wallis tests followed by Wilcoxon pairwise tests using a Benjamini-Hochberg correction were performed. All analysis were performed in R version 4.2.0.

Data availability

The consensus genomes from viral WGS data are deposited in Genbank. IDs can be found in **Supplementary Table 7** (HAdV), **Supplementary Table 9** (AAV2) and **Supplementary Table 10** (HHV6).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD035925.

Code availability

Code for metagenomics and PCR analysis can be found at:

https://github.com/sarah-buddle/unknown-hepatitis

The transcriptomics analysis code is in

https://github.com/innate2adaptive/Bulk-RNAseq-analysis/tree/main/Zscore_gene_expression_module_analysis

The proteomics differential expression analysis code is in:

https://github.com/MahdiMoradiMarjaneh/proteomics_and_transcriptomics_of_hepatitis

Funding

The work was part funded by the National Institute for Health Research Blood and Transplant Research Unit in Genomics to Enhance Microbiology Screening (GEMS). S.M is funded by a W.T. Henry Wellcome fellowship (206478/Z/17/Z). S.B. and O.E.T.M. are funded by the NIHR Blood and Transplant Research Unit (GEMS). MM and ML are supported in part by the NIHR Biomedical Research Centre of Imperial College NHS Trust. JB receives NIHR Senior Investigator Funding.M.N and JB are supported by the Wellcome Trust (207511/Z/17/Z) and

(203268/Z/16/Z). M.N, J.B and G.P are supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. P.S is supported by NIHR (NIHR203338). T.J is grateful for funding from the Brain Tumour Charity, Children with Cancer UK, Great Ormond Street Hospital Children's Charity, Olivia Hodson Cancer Fund, Cancer Research UK and the National Institute of Health Research. DIAMONDS is funded by the European Union (Horizon 2020; grant 848196). PERFORM was funded by the European Union (Horizon 2020; grant 668303).

Acknowledgements

UKHSA for funding of the metagenomics and adenovirus sequencing.

Amit Nathwani for helpful discussions.

We acknowledge the considerable contribution from the Great Ormond Street Hospital NHS Foundation Trust (GOSH) microbiology laboratory. All research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Author Contributions

JBre, SM and SB conceived the study, analysed the data and wrote the manuscript. JRB, LA, NS, AL, JCDL, JH, SD coordinated samples and carried out the metagenomics and confirmatory PCRs. OETM, JAGA, SR, CV, LMMB, RW, CAW, HT, NB, HM, KAM, SCH DKA carried out genome sequencing and analyses. MMM, MN, GP, AC, AM, CV and ML analysed transcriptomic data, KT, ML, MMM, RZC generated and analysed proteomic data. SNW, JRC, JFAD, AS, LJT, ZA, JN, KSH carried out AAV2 tropism experiments. GS, PG, TEW, SNW JRC helped with AAV2 PCR development. LC, RB, MD, JM, JCH, CA, GA, TSJ carried out histology, immunohistochemistry and electron microscopy. BBK & JR provided control HHV6 material. PSh, JA provided control samples. ML, PSi, SC, MV, CF, MS provided PERFORM & DIAMONDS control samples. KB, MGS, PC, MO coordinated ISARIC consents and data collection TG, NH, CK provided data and samples from Kings and Birmingham Liver Units. IUL, MC, MZ, SM, CW, RS, EG, SG, CC, TT, KH, CH, TR, CM, KT, CN, MH, RG, SJS provided data and samples from UKHSA and devolved nations.ET provided reagents and contributed helpful discussions

Competing Interests Declaration

JB declares the following:

MHRA member of COVID Vaccines committee

Holder of Wellcome Trust, UKRI, NIHR funding

PI on the GSK LUNAR study to investigate SARS-CoV-2 sequences in patients treated with Sotrovimab. Commissioned by the MHRA

Additional Information

Supplementary information The online version contains supplementary information.

Correspondence should be addressed to Judith Breuer.

Methods References

41. Babraham Bioinformatics - Trim Galore! https://www.bioinformatics.babraham.ac.uk/projects/trim_galore//.
42. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. (2012) doi:10.1038/nmeth.1923.
43. Schmieder, R., Edwards, R. & Bateman, A. Quality control and preprocessing of metagenomic datasets. *BIOINFORMATICS APPLICATIONS NOTE* **27**, 863–864 (2011).
44. Morgulis, A. *et al.* Database indexing for production MegaBLAST searches. *Bioinformatics* **24**, 1757–1764 (2008).
45. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using diamond. **12**, (2015).
46. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
47. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. (2017) doi:10.1101/gr.213959.116.
48. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
49. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 1–13 (2019).
50. EPI2ME WIMP workflow: quantitative, real-time species identification from metagenomic samples. https://nanoporetech.com/resource-centre/epi2me-wimp-workflow-quantitative-real-time-species-identification-metagenomic.
51. Babraham Bioinformatics - reDOTable DotPlot tool. https://www.bioinformatics.babraham.ac.uk/projects/redotable/.
52. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
53. Afzal, S., Fronza, R. & Schmidt, M. VSeq-Toolkit: Comprehensive Computational Analysis of Viral Vectors in Gene Therapy. *Mol Ther Methods Clin Dev* **17**, 752–757 (2020).
54. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology 2011 29:1* **29**, 24–26 (2011).
55. Aurnhammer, C. *et al.* Universal Real-Time PCR for the Detection and Quantification of Adeno-Associated Virus Serotype 2-Derived Inverted Terminal Repeat Sequences. *https://home.liebertpub.com/hgtb* **23**, 18–28 (2011).
56. Watzinger, F. *et al.* Real-time quantitative PCR assays for detection and monitoring of pathogenic human viruses in immunosuppressed pediatric patients. *J Clin Microbiol* **42**, 5189–5198 (2004).
57. Tann, C. J. *et al.* Prevalence of bloodstream pathogens is higher in neonatal encephalopathy cases vs. controls using a novel panel of real-time PCR assays. *PLoS One* **9**, (2014).
58. Brown, J. R., Shah, D. & Breuer, J. Viral gastrointestinal infections and norovirus genotypes in a paediatric UK hospital, 2014–2015. *Journal of Clinical Virology* **84**, 1–6 (2016).
59. Karda, R. *et al.* Production of lentiviral vectors using novel, enzymatically produced, linear DNA. *Gene Ther* **26**, 86–92 (2019).

60. Myers, C. E. *et al.* Using Whole Genome Sequences to Investigate Adenovirus Outbreaks in a Hematopoietic Stem Cell Transplant Unit. *Front Microbiol* **12**, (2021).
61. Gaccioli, F. *et al.* Fetal inheritance of chromosomally integrated HHV-6 predisposes to preeclampsia in the mother Europe PMC Funders Group. **5**, 901–908 (2020).
62. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* **12**, 1261–1267 (2017).
63. Baker, D. J. *et al.* CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes. doi:10.1186/s13073-021-00839-5.
64. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492 (2018).
65. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).
66. Hsu, H. L. *et al.* Structural characterization of a novel human adeno-associated virus capsid with neurotropic properties. *Nat Commun* **11**, (2020).
67. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **25**, 1754–1760 (2009).
68. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology 2008 26:12* **26**, 1367–1372 (2008).
69. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods 2016 13:9* **13**, 731–740 (2016).
70. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
71. Ng, J. *et al.* Gene therapy restores dopamine transporter expression and ameliorates pathology in iPSC and mouse models of infantile parkinsonism. *Sci Transl Med* **13**, (2021).
72. Pollara, G. *et al.* Exaggerated IL-17A activity in human in vivo recall responses discriminates active tuberculosis from latent infection and cured disease. *Sci Transl Med* **13**, (2021).
73. Chandran, A. *et al.* Rapid synchronous type 1 IFN and virus-specific T cell responses characterize first wave non-severe SARS-CoV-2 infections. *Cell Rep Med* **3**, 100557 (2022).
74. Zhang, X. *et al.* Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nature Protocols 2018 13:3* **13**, 530–550 (2018).

Extended Data Figure Legends

Extended Data Figure 1: Evidence of AAV2 replication from meta-transcriptomics and RT-PCR

Mapping of AAV2 reads to the reference genome for **a** liver RNA-Seq from 4 cases, **b** blood RNA-Seq from 2 cases. The horizontal lines in the same colour as the coverage graph are the predicted transcripts for each case. The horizontal lines in purple and green are the AAV2 genes. **c,** RT-PCR results for liver cases. N: Negative PCR result

Extended Data Figure 2: Examples of AAV2 complexes

The y axis shows the coordinates of a full length AAV2 genome (rep gene in green and cap gene in yellow). X axis is the nanopore read with the length of the read indicated. Red dots indicate alignment to the forward strand and blue dots the reverse. **a,** indicative complexes based on literature[8] **b and c.** Examples of complex structures with both head to tail and alternating repeats, from a total of n=26 and n=75 such reads for cases 3 and 5 respectively. **b** shows the longest 2 reads for each case. **d.** Examples of truncated monomeric structures, from a total of n=25 and n=103 such reads for cases 3 and 5 respectively (Supplementary Table 3). The longest such read for each case is shown.

Extended Data Figure 3: HAdV and AAV2 sequence analysis

a, HAdV SNP plot: Visualisation of the multiple alignment of HAdV-F41 genomic sequences from the same clade as the single sequence from a case (highlighted in grey) (Figure 3A). Includes both contemporary controls and publicly available HAdV-F41 genomes from GenBank. Consensus-level mutations differing from the reference sequence (bottom) are highlighted across the genome. Genomic position of the mutation is shown at the top of the plot. **b,** Variants between stool complete HAdV genome from case JBB27 and combined blood partial genomes from other cases. **c,** Frequency table of capsid residues in cases and historical controls. There is no difference between the capsid sequences of cases and contemporaneously circulating controls. However, there are changes compared with historical controls in all contemporary sequences. None of the recently acquired capsid changes are shared with known hepatotrophic strains in AAV7, 8 and 9. **d,** Amino acid differences between AAV2 capsid sequences from cases, contemporaneously circulating controls and historical publicly available sequences compared with the AVV2 reference sequence NC_001401.2. Also shown are the capsid sequences from known AAV7,8 and 9 hepatotropic capsids compared to the reference sequence NC_001401.2.

Extended Data Figure 4: AAV2 capsid analysis

a, Amino acid sequence of novel AAV capsid variant. The consensus sequence of the VP1 sequence used for investigation of capsid transduction characteristics (AAVHepcase) is shown with alignment to canonical AAV2 VP1 (AAV2gp05). The alignment shows AAV2 amino acids that are different to the AAVHepcase sequence, with dots indicating matched amino acids sequence. **b,** In vitro analysis of AAV capsid transduction characteristics. Huh-7 hepatocytes were treated at MOI 100,000 with rAAV vectors containing capsid sequences derived from canonical AAV2, a consensus sequence derived from patient sequencing samples (Hepcase), LK03, or AAV9 (n=3 each treatment). Transduction efficiency was determined by flow

cytometry, based on the percentage of EGFP-positive cells, the EGFP fluorescence intensity in positive cells, and the 'relative activity' of EGFP expression (calculated by multiplying %GFP-positive cells by MFI/10070). Transductions were performed in the presence or absence of 400 µg/mL heparin to investigate the role of HSPG interaction. rAAV2 was significantly affected by heparin competition, whereas other capsids, including that derived from AAV Hepcase, were not. Heparin competition significantly affected rAAV2 transduction in terms of percentage of GFP-positive cells (P=0.0016), MFI (P=0.000008), and relative activity (P=0.000008), whereas other capsids, including that derived from AAV Hepcase, were not affected by heparin. All data were analysed by 2-sided t-test with Bonferroni post-hoc analysis. Error bars indicate standard deviation from the mean value. **c,** Images of Huh-7 cells treated with rAAV vectors *in vitro*. Images of transduced Huh-7 cells. Each cell population was treated with MOI 100,000 of the relevant viral vector, in the presence or absence of 400 µg/mL heparin and analysed by EGFP fluorescence 72-hours post-transduction. Scale bars = 300 µm.

Extended Data Figure 5: Representative histology of case livers

a & b, H&E sections x100 and x 200 showing a pattern of acute hepatitis with parenchymal disarray, there is a normal, uninflamed, portal tract lower left image **a.** Spotty inflammation and apoptotic bodies are shown in **b** along with perivenular hepatocyte loss/necrosis. Immunohistochemistry shows fewer mature B lymphocytes (CD20 panel **c**) than T lymphocytes (CD3, panel **d**, pan T cell marker) most of which are cytotoxic CD8 lymphocytes (panel **e**). In conclusion the livers of these children have a distinctive pattern of damage which does not indicate a specific aetiology, it does not exclude but does not offer positive support for either autoimmune hepatitis or a direct cytopathic effect of virus on hepatocytes. Each image shows a representative result from histology carried out on a minimum of five cases.

Extended Data Figure 6: Immunohistochemistry results for cases of unexplained hepatitis and control tissues

a, Inflammatory markers (IgG, C4d, HLA-ABC, HLA-DR) in acute hepatitis cases and control liver. IgG, HLA-ABC and HLA-DR show a canalicular pattern in the control liver. This pattern is disrupted in the acute hepatitis cases due to the architectural collapse. In addition, there is increased staining associated with inflammatory cell/macrophage infiltrates. C4d shows very weak staining in the acute hepatitis cases associated with macrophages but with without endothelial staining. All stains were undertaken on 5 affected cases and 13 control cases. **b,** Representative images of the immunohistochemistry (IHC). Acute hepatitis liver explant cases stained for HHV6, arrow shows staining of **A** representative cells, **B** adenovirus, AAV2 (**C** polyclonal antibody, **E** monoclonal antibody, clone A1). Paraffin embedded AAV2 transfected cell lines stained as positive controls for AAV2 (**D** polyclonal antibody, **F** monoclonal antibody, clone A1). All scale bars are 60 micrometres. HHV6, AAV2 (polyclonal) stains were undertaken on 15 affected cases and 13 controls. AAV2 (A1) stains were undertaken on 5 affected cases and 13 control cases. Staining for adenovirus was undertaken on 5 affected cases.

Extended Data Figure 7: Cytokine inducible transcriptional modules

Volcano plot of cytokine inducible transcriptional modules (n=52) comparing their Z score expression in AAV2-associated hepatitis (n=4) and HBV-associated hepatitis (n=17) requiring transplantation using two-tailed unpaired t tests with Holm Sidak multiple testing correction for adjusted p values. Each point represents a specific module listed in full in Supplementary Table 13. Labels for selected modules are shown.

Extended Data Figure 8: HLA and HHV-6B proteins in case livers

a & b Ranking of the quantified proteins using the log10 of iBAQ values for **a** JBL1, **b** JBL2, **c** JBL3, **d** JBL4, **e** JBL5. **f,** Scatter plot of quantified proteins in sample JBL4 versus JBL5. HLA proteins are highlighted in red. Red arrows denote HLA-DRB1 proteins. HHV6 proteins are highlighted in green and marked with green arrows.

Extended Data Table titles and footnotes

Extended Data Table 1: PCR and whole genome sequencing for samples from cases where metagenomic sequencing was not performed.

- : Not tested due to insufficient residual material

N: negative PCR result

P: Positive PCR result in referring laboratory

Where two results are shown, the first refers to the referring laboratory and the second to GOSH. Where there was a discrepancy, the positive result is shown.

F: Failed

Where there is more than one sample for a single patient, CT values represent the mean across the samples that were tested.

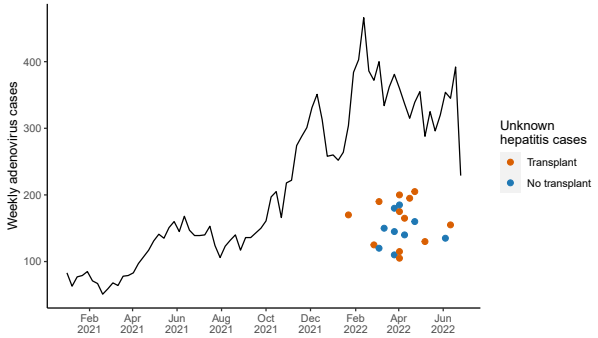
*Metagenomics reads: the result of combining the datasets from two blood samples from the same case

De novo assembly of unclassified metagenomics reads was unremarkable

Extended Data Table 2: Controls and comparators

a Summary of DIAMONDS and PERFORM immunocompetent controls. **b** immunocompromised comparators. **c** age distribution of blood comparator and control patients from GOSH, DIAMONDS and PERFORM

a



b

Cases

Liver (n=14)
5 Birmingham Children's Hospital (frozen, all received transplant)
9 King's College Hospital (FFPE, of which 7 received transplant)

Blood/serum (n=23)
11 UKHSA (of which 1 received transplant)
5 Glasgow Royal Infirmary
1 NHS Lothian
6 King's College Hospital (serum, of which 3 received transplants)

Stool (n=5)
3 Glasgow Royal Infirmary
1 NHS Lothian
1 UKHSA

Respiratory (n=4)
2 University Hospital Wales
1 UKHSA
1 Glasgow Royal Infirmary

Controls and comparators

Immunocompromised comparators from GOSH

GOSH Liver (n=4) (+2 adults)

GOSH Blood (n=17)
14 HAdV, raised ALT
3 CMV, raised ALT

Immunocompetent controls from PERFORM/ DIAMONDS cohorts

PERFORM/ DIAMONDS Blood (n=65)
13 Healthy
17 HAdV, normal ALT
8 HAdV, normal ALT (blood)
5 HAdV, raised ALT
11 Critical Illness, raised ALT
5 Non-HAdV, raised ALT
6 Other Hepatitis

Control from HAdV outbreak but no unknown hepatitis

UKHSA Blood (n=1)
1 HLH

FFPE liver controls from patients tested for other liver conditions

King's College Hospital Liver (n=1)

Queen Mary University of London (n=2) (+8 adults)

Other controls used for whole genome sequencing

UKHSA Stool (n=22)

GOSH Stool/Urine (n=11)
10 stool
1 urine

NHS Lothian Stool (n=1)

c

Cases

Metagenomics

DNA (n=10)
5 frozen liver
5 blood

RNA (n=7)
4 frozen liver
3 blood

PCR

AAV2 (n=28)
5 frozen liver (5 tr)
9 FFPE liver (7 tr)
11 blood
6 serum (3 tr)
1 stool
2 respiratory

HAdV (n=31)
5 frozen liver (5 tr)
9 FFPE liver (7 tr)
15 blood (1 tr)
6 serum (3 tr)
1 stool
3 respiratory

HHV6 (n=23)
5 frozen liver (5 tr)
9 FFPE liver (7 tr)
6 blood
6 serum (3 tr)
1 stool
2 respiratory

Whole genome sequencing

AAV2 (n=15)
10 blood
5 frozen liver
1 stool
2 respiratory

HAdV (n=16)
10 blood
4 stool
2 respiratory

HHV6 (n=4)
4 frozen liver

Controls and comparators

Immunocompetent controls

AAV2 (n=66)
65 blood (PERFORM/ DIAMONDS)
1 blood (UKHSA HLH)

HAdV/HHV6 (n=1)
1 blood (UKHSA HLH)

Immunocompromised comparators from GOSH

AAV2 (n=21)
17 blood
4 (+2 adult) liver

HAdV (n=20)
16 blood
4 (+2 adult) liver

HHV6 (n=19)
15 blood
4 (+2 adult) liver

AAV2 (n=6)
4 blood
2 stool

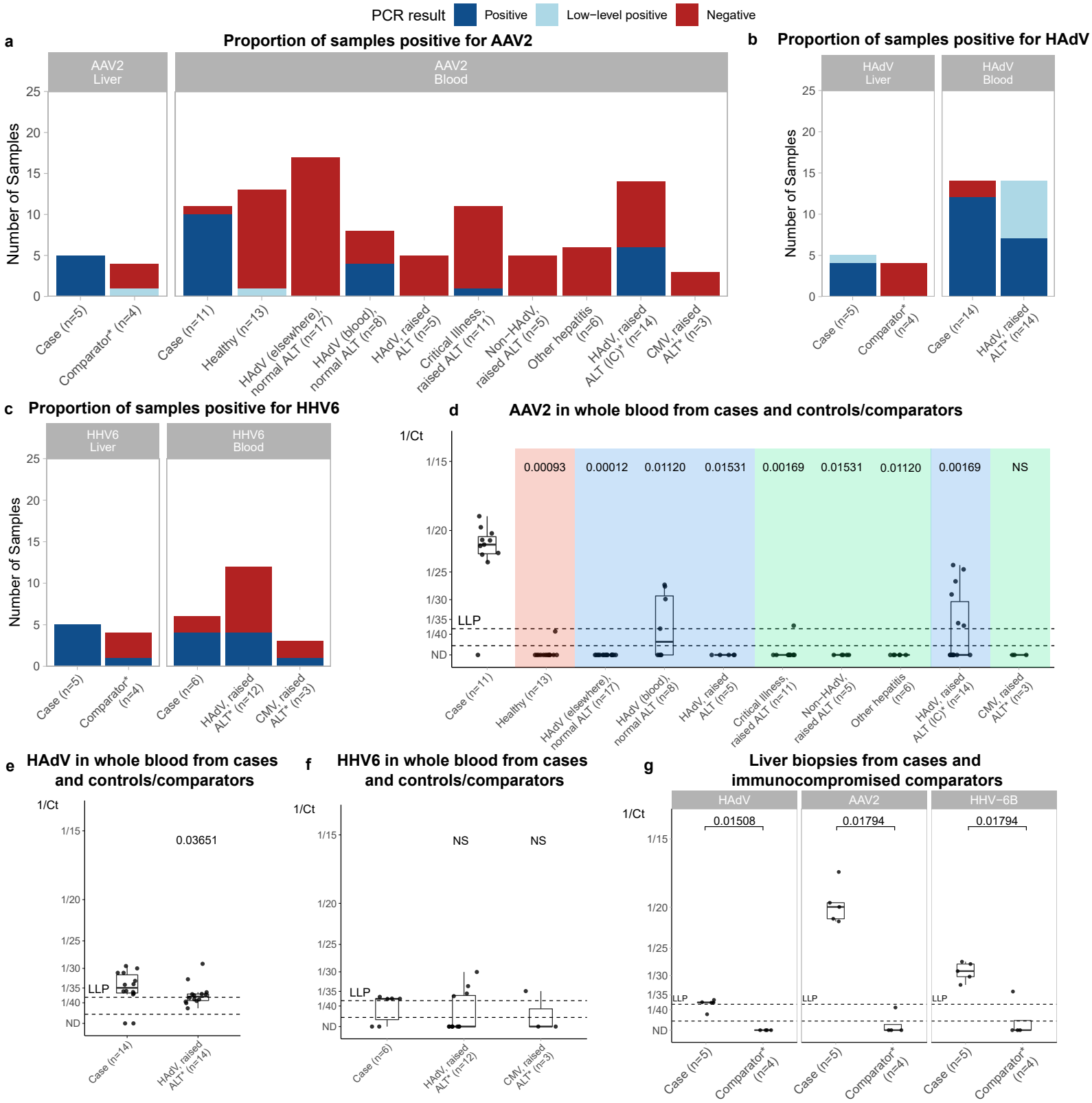
HAdV (n=35)
1 blood
33 stool
1 urine

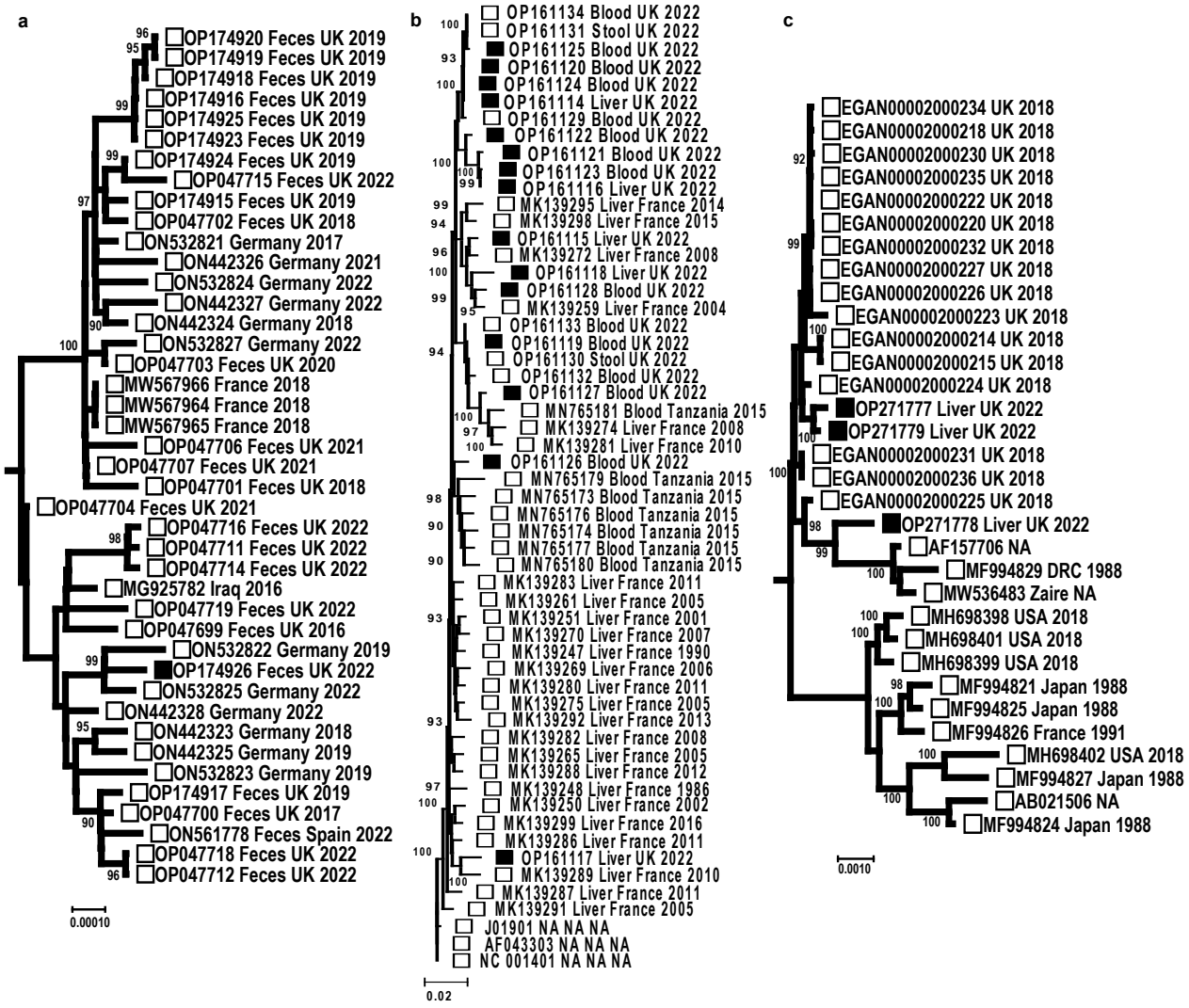
Immunocompetent FFPE liver comparators

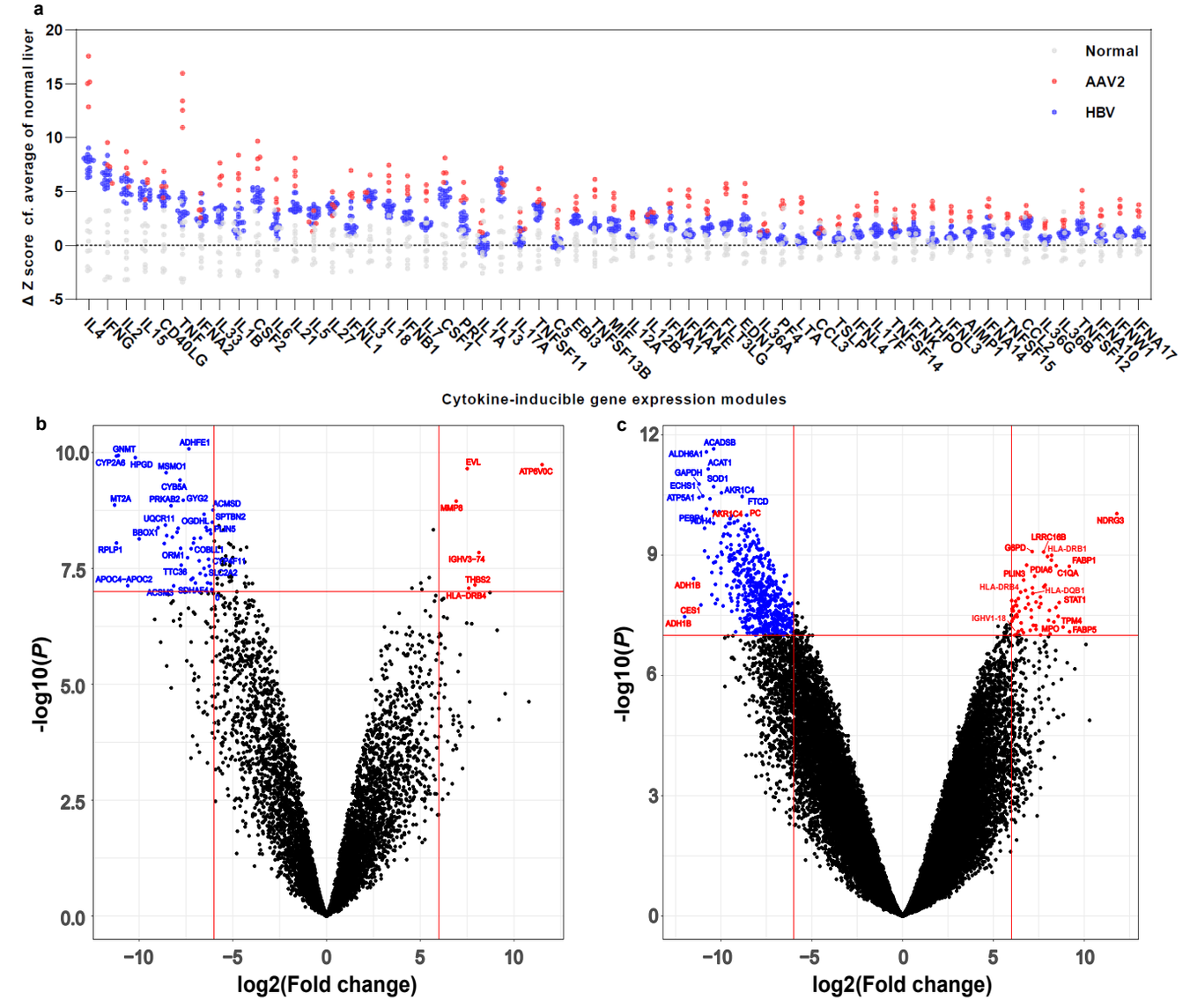
AAV2 (n=3) (+8 adults)
1 FFPE liver (King's College Hospital)
2 (+8 adults) FFPE liver (Queen Mary University of London)

HAdV (n=1)
1 FFPE liver (King's College Hospital)

HHV6 (n=3) (+8 adults)
1 FFPE liver (King's College Hospital)
2 (+8 adults) FFPE liver (Queen Mary University of London)


PCR result
Positive
Low−level positive
Negative
a
Proportion of samples positive for AAV2
AAV2 Liver
AAV2 Blood
Number of Samples
Case (n=5)
Comparator* (n=4)
Case (n=11)
Healthy (n=13)
HAdV (elsewhere), normal ALT (n=17)
HAdV (blood), normal ALT (n=8)
HAdV, raised ALT (n=5)
Critical Illness, raised ALT (n=11)
Non−HAdV, raised ALT (n=5)
Other hepatitis (n=6)
HAdV, raised ALT (IC)* (n=14)
CMV, raised ALT* (n=3)
b
Proportion of samples positive for HAdV
HAdV Liver
HAdV Blood
Case (n=14)
HAdV, raised ALT* (n=14)
c
Proportion of samples positive for HHV6
HHV6 Liver
HHV6 Blood
Case (n=6)
HAdV, raised ALT* (n=12)
d
AAV2 in whole blood from cases and controls/comparators
1/Ct
0.00093
0.00012
0.01120
0.01531
0.00169
0.01531
0.01120
0.00169
NS
LLP
ND
e
HAdV in whole blood from cases and controls/comparators
0.03651
f
HHV6 in whole blood from cases and controls/comparators
NS
g
Liver biopsies from cases and immunocompromised comparators
HAdV
AAV2
HHV−6B
0.01508
0.01794
0.01794


a
OP174920 Feces UK 2019
OP174919 Feces UK 2019
OP174918 Feces UK 2019
OP174916 Feces UK 2019
OP174925 Feces UK 2019
OP174923 Feces UK 2019
OP174924 Feces UK 2019
OP047715 Feces UK 2022
OP174915 Feces UK 2019
OP047702 Feces UK 2018
ON532821 Germany 2017
ON442326 Germany 2021
ON532824 Germany 2022
ON442327 Germany 2022
ON442324 Germany 2018
ON532827 Germany 2022
OP047703 Feces UK 2020
MW567966 France 2018
MW567964 France 2018
MW567965 France 2018
OP047706 Feces UK 2021
OP047707 Feces UK 2021
OP047701 Feces UK 2018
OP047704 Feces UK 2021
OP047716 Feces UK 2022
OP047711 Feces UK 2022
OP047714 Feces UK 2022
MG925782 Iraq 2016
OP047719 Feces UK 2022
OP047699 Feces UK 2016
ON532822 Germany 2019
OP174926 Feces UK 2022
ON532825 Germany 2022
ON442328 Germany 2022
ON442323 Germany 2018
ON442325 Germany 2019
ON532823 Germany 2019
OP174917 Feces UK 2019
OP047700 Feces UK 2017
ON561778 Feces Spain 2022
OP047718 Feces UK 2022
OP047712 Feces UK 2022
0.00010
b
OP161134 Blood UK 2022
OP161131 Stool UK 2022
OP161125 Blood UK 2022
OP161120 Blood UK 2022
OP161124 Blood UK 2022
OP161114 Liver UK 2022
OP161129 Blood UK 2022
OP161122 Blood UK 2022
OP161121 Blood UK 2022
OP161123 Blood UK 2022
OP161116 Liver UK 2022
MK139295 Liver France 2014
MK139298 Liver France 2015
OP161115 Liver UK 2022
MK139272 Liver France 2008
OP161118 Liver UK 2022
OP161128 Blood UK 2022
MK139259 Liver France 2004
OP161133 Blood UK 2022
OP161119 Blood UK 2022
OP161130 Stool UK 2022
OP161132 Blood UK 2022
OP161127 Blood UK 2022
MN765181 Blood Tanzania 2015
MK139274 Liver France 2008
MK139281 Liver France 2010
OP161126 Blood UK 2022
MN765179 Blood Tanzania 2015
MN765173 Blood Tanzania 2015
MN765176 Blood Tanzania 2015
MN765174 Blood Tanzania 2015
MN765177 Blood Tanzania 2015
MN765180 Blood Tanzania 2015
MK139283 Liver France 2011
MK139261 Liver France 2005
MK139251 Liver France 2001
MK139270 Liver France 2007
MK139247 Liver France 1990
MK139269 Liver France 2006
MK139280 Liver France 2011
MK139275 Liver France 2005
MK139292 Liver France 2013
MK139282 Liver France 2008
MK139265 Liver France 2005
MK139288 Liver France 2012
MK139248 Liver France 1986
MK139250 Liver France 2002
MK139299 Liver France 2016
MK139286 Liver France 2011
OP161117 Liver UK 2022
MK139289 Liver France 2010
MK139287 Liver France 2011
MK139291 Liver France 2005
J01901 NA NA NA
AF043303 NA NA NA
NC 001401 NA NA NA
0.02
c
EGAN00002000234 UK 2018
EGAN00002000218 UK 2018
EGAN00002000230 UK 2018
EGAN00002000235 UK 2018
EGAN00002000222 UK 2018
EGAN00002000220 UK 2018
EGAN00002000232 UK 2018
EGAN00002000227 UK 2018
EGAN00002000226 UK 2018
EGAN00002000223 UK 2018
EGAN00002000214 UK 2018
EGAN00002000215 UK 2018
EGAN00002000224 UK 2018
OP271777 Liver UK 2022
OP271779 Liver UK 2022
EGAN00002000231 UK 2018
EGAN00002000236 UK 2018
EGAN00002000225 UK 2018
OP271778 Liver UK 2022
AF157706 NA
MF994829 DRC 1988
MW536483 Zaire NA
MH698398 USA 2018
MH698401 USA 2018
MH698399 USA 2018
MF994821 Japan 1988
MF994825 Japan 1988
MF994826 France 1991
MH698402 USA 2018
MF994827 Japan 1988
AB021506 NA
MF994824 Japan 1988
0.0010


a
Δ Z score cf. average of normal liver
Normal
AAV2
HBV
IL4 IFNG IL2 IL15 CD40LG TNF IFNA2 IL33 IL1B CSF2 IL6 IL21 IL5 IL27 IFNL1 IL3 IL18 IFNB1 IL7 CSF1 PRL IL1A IL13 IL17A TNFSF11 C5 EBI3 TNFSF13B MIF IL12A IL12B IFNA1 IFNA4 IFNE FLT3LG EDN1 IL36A PF4 LTA CCL3 TSLP IFNL4 IL17F TNFSF14 THPO IFNL3 AIMP1 IFNA14 TNFSF15 CCL2 IL36G IL36B TNFSF12 IFNA10 IFNW1 IFNA17
Cytokine-inducible gene expression modules
b
-log10(P)
log2(Fold change)
c
-log10(P)
log2(Fold change)