Title

Adeno-associated virus type 2 in US children with acute severe hepatitis

Permalink

https://escholarship.org/uc/item/83d7j1fd

Journal

Nature, 617(7961)

ISSN

0028-0836

Authors

Servellita, Venice
Sotomayor Gonzalez, Alicia
Lamson, Daryl M
et al.

Publication Date

2023-05-18

DOI

10.1038/s41586-023-05949-1

Peer reviewed

Article

Adeno-associated virus type 2 in US children with acute severe hepatitis

Check for updates

Venice Servellita[1,22], Alicia Sotomayor Gonzalez[1,22], Daryl M. Lamson[2,22], Abiodun Foresythe[1], Hee Jae Huh[1,3], Adam L. Bazinet[4], Nicholas H. Bergman[4], Robert L. Bull[5], Karla Y. Garcia[4], Jennifer S. Goodrich[4], Sean P. Lovett[4], Kisha Parker[4], Diana Radune[4], April Hatada[6], Chao-Yang Pan[6], Kyle Rizzo[6], J. Bradford Bertumen[6,7], Christina Morales[6], Paul E. Oluniyi[1], Jenny Nguyen[1], Jessica Tan[1], Doug Stryke[1], Rayah Jaber[8], Matthew T. Leslie[9], Zin Lyons[10], Hayden D. Hedman[7,11], Umesh Parashar[7], Maureen Sullivan[12], Kelly Wroblewski[12], M. Steven Oberste[7], Jacqueline E. Tate[7], Julia M. Baker[7], David Sugerman[7], Caelin Potts[7], Xiaoyan Lu[7], Preeti Chhabra[7], Pediatric Hepatitis of Unknown Etiology Working Group*, L. Amanda Ingram[13], Henry Shiau[14], William Britt[14], Luz Helena Gutierrez Sanchez[14], Caroline Ciric[15], Christina A. Rostad[15], Jan Vinjé[7], Hannah L. Kirking[7], Debra A. Wadford[6], R. Taylor Raborn[4], Kirsten St. George[2,16] & Charles Y. Chiu[1,17,18] ✉

As of August 2022, clusters of acute severe hepatitis of unknown aetiology in children have been reported from 35 countries, including the USA[1,2]. Previous studies have found human adenoviruses (HAdVs) in the blood from patients in Europe and the USA[3–7], although it is unclear whether this virus is causative. Here we used PCR testing, viral enrichment-based sequencing and agnostic metagenomic sequencing to analyse samples from 16 HAdV-positive cases from 1 October 2021 to 22 May 2022, in parallel with 113 controls. In blood from 14 cases, adeno-associated virus type 2 (AAV2) sequences were detected in 93% (13 of 14), compared to 4 (3.5%) of 113 controls ($P < 0.001$) and to 0 of 30 patients with hepatitis of defined aetiology ($P < 0.001$). In controls, HAdV type 41 was detected in blood from 9 (39.1%) of the 23 patients with acute gastroenteritis (without hepatitis), including 8 of 9 patients with positive stool HAdV testing, but co-infection with AAV2 was observed in only 3 (13.0%) of these 23 patients versus 93% of cases ($P < 0.001$). Co-infections by Epstein–Barr virus, human herpesvirus 6 and/or enterovirus A71 were also detected in 12 (85.7%) of 14 cases, with higher herpesvirus detection in cases versus controls ($P < 0.001$). Our findings suggest that the severity of the disease is related to co-infections involving AAV2 and one or more helper viruses.

Viral hepatitis is an inflammation of the liver most commonly caused by one of the main hepatitis viruses (A–E). Since October 2021, clusters of acute non-A–E severe hepatitis of unknown aetiology in children have been reported in 35 countries, including the USA[1,2]. As of 17 August 2022, 358 people under investigation have been reported in the USA, of whom 22 (6%) required a liver transplant and 13 (4%) died[8].

HAdVs are non-enveloped, double-stranded DNA viruses that cause a variety of infections in both adults and children, including respiratory tract infection, conjunctivitis and gastroenteritis (mainly from HAdV type 40 (HAdV-40) and HAdV-41)[9]. Although a potentially

life-threatening disease in immunocompromised patients[7,9], hepatitis from adenovirus infection has been thought to be rare in immunocompetent children without underlying comorbidities[10]. HAdVs, particularly HAdV-41, have been found in blood from clusters of acute severe hepatitis cases from Scotland, the UK and the USA[3–7], but whether the virus is causative remains unclear. Adeno-associated viruses (AAVs) are small, single-stranded DNA parvoviruses that are considered non-pathogenic in humans and for this reason have been widely used as vectors for gene therapy[11]. Importantly, AAVs require a helper virus, such as a herpesvirus or adenovirus, for productive infection of liver tissue[12]. AAVs have previously been reported in children with acute
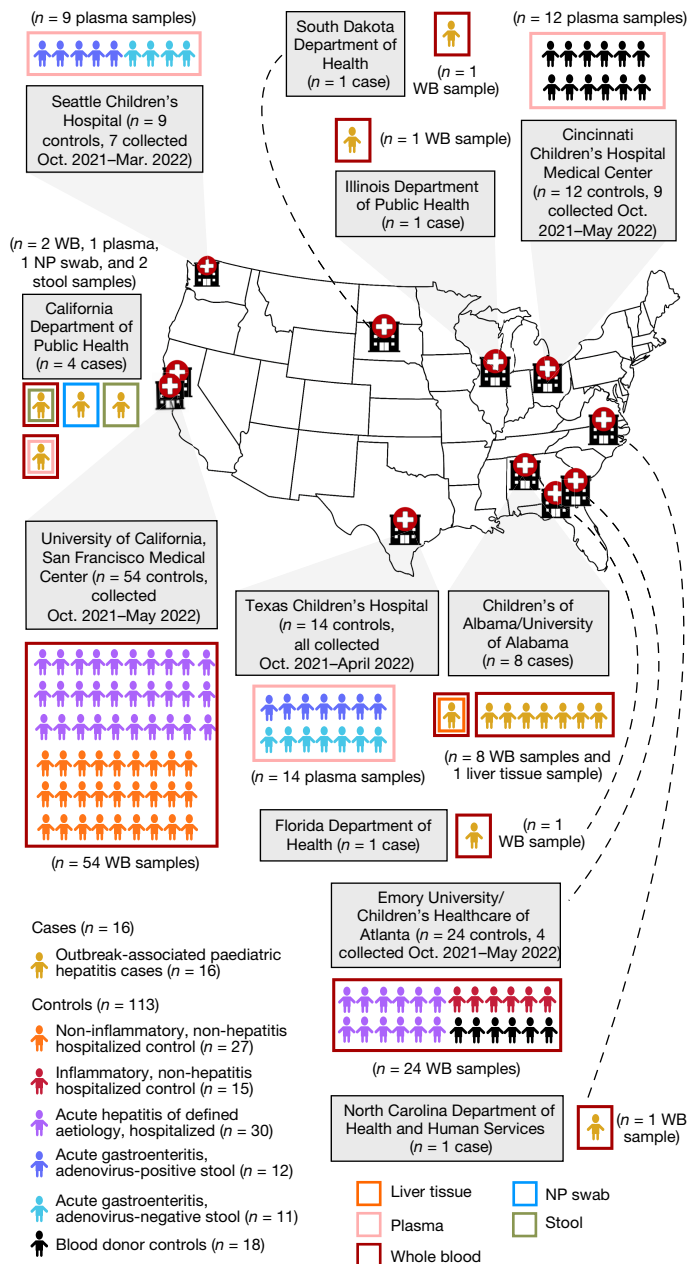
Fig. 1 | Epidemiology of cases and controls. Geographic distribution of the 16 acute severe hepatitis cases and 113 controls in the study, showing the hospital or public health laboratory sites providing samples and associated clinical data from cases and/or controls. WB, whole blood; NP, nasopharyngeal.

severe hepatitis in a study of nine patients from the UK[13], often in association with adenovirus or human herpesvirus 6 (HHV-6) infection[14].

In this study, 27 samples (21 whole blood, 2 plasma, 1 liver tissue, 1 nasopharyngeal swab and 2 stool sample(s)) from 16 children with acute severe hepatitis of unknown aetiology were analysed (Fig. 1). All children met the clinical case definition established by the Centers for Disease Control and Prevention, including lack of a confirmed aetiology, liver enzyme levels (aspartate aminotransferase (AST) or alanine aminotransferase (ALT)) >500 U l⁻¹, age < 10 years and onset on or after 1 October 2021 (refs. 2,15). Cases were enrolled from six states (Alabama, California, Florida, Illinois, North Carolina, and South Dakota) from 1 October 2021 to 22 May 2022. All 16 cases had positive testing for HAdV from blood, and thus HAdV infection was over-represented compared to the overall affected population, of which HAdV is detected in 45–90%

(refs. 3–7). The median age of affected children was 3 years; 56% were female and 44% male (Table 1). Mean elevations in ALT and AST were 2,293 ± 1,486 U l⁻¹ (normal range: 4–36 U l⁻¹) and 2,652 ± 1,851 U l⁻¹ (normal range: 8–33 U l⁻¹), respectively (Supplementary Table 1). A liver biopsy was carried out in eight patients; viral PCR testing of the liver biopsies for HAdVs, herpesviruses, enteroviruses and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was incomplete, but yielded positive results for HAdV in three patients with no other viruses detected (Fig. 2b and Supplementary Table 1). Two children underwent liver transplantation, but none died from complications of liver failure.

Controls (n = 113) consisted of 78 whole blood, 1 serum and 34 plasma samples. Many controls were enrolled from California (n = 54) and Georgia (n = 24) to be geographically similar to the cases, with the remaining controls enrolled from Ohio (n = 12), Texas (n = 14) and Washington (n = 9). Of the 113 controls, 69 (61.0%) were collected over the same time frame as that for the cases (that is, collected between 1 October 2021 and 22 May 22). The 113 controls consisted of 42 patients (37%) without hepatitis, 30 (26.5%) patients with acute hepatitis (ALT > 100 U l⁻¹) of defined aetiology, 23 (21%) patients with acute gastroenteritis (12 with positive HAdV stool testing) and 18 (16%) blood donors. Differences in age and sex between cases and controls were non-significant overall, although cases were significantly younger and older, respectively, than controls with hepatitis of defined aetiology and those with acute gastroenteritis (Table 1). Among reported symptoms, only jaundice was significantly more associated with cases than controls (Table 1). Mean elevations in ALT and AST in the 30 hepatitis controls were 291 ± 288 U l⁻¹ and 455 ± 833 U l⁻¹, respectively, and significantly lower than in cases (P < 0.0001; Table 1 and Supplementary Table 1).

Virus detection in hepatitis cases

Metagenomic sequencing for agnostic detection of all viruses[16], tiling multiplex PCR amplicon sequencing for AAV2 and HAdV-41 (ref. 17), metagenomic sequencing with probe capture viral enrichment[18] and virus-specific polymerase chain reaction (PCR) testing for HAdV with genotyping by hexon gene sequencing were carried out to identify AAV2 and HAdV viruses in clinical samples from cases (Fig. 2a). We detected AAV2 from available whole-blood or plasma samples in 13 (93%) of 14 cases and confirmed detection of HAdV in all 14 cases (Fig. 2b). Of the 14 cases, 10 (71%) were typed as HAdV-41, 1 (7%) was typed as HAdV-40, and 1 (7%) was typed as HAdV-2; 1 (7%), sample NC_14, was untypable. AAV2 and HAdV-41 enrichment were carried out using either tiled multiplex PCR amplicon sequencing targeting the genomes of HAdV-41 and AAV (n = 9 samples, with a mean 6.6 ± 1.5 (s.d.) million raw reads generated per sample) or probe capture viral enrichment targeting 3,153 species (n = 4, with a mean 128 ± 34 (s.d.) million raw reads generated per sample). Mean normalized viral counts expressed as reads per kilobase per million mapped reads for AAV2 ($\overline{x}$ = 4,063) were approximately 12.7 times greater than for HAdV-41 ($\overline{x}$ = 320; Supplementary Table 1). AAV2 was detected in both liver tissue and whole blood from one case and in one of two plasma samples, but neither AAV2 nor HAdV-41 were detected in nasopharyngeal swab and stool samples. No reads from AAVs other than AAV2 were detected, despite using methods that would enrich for other AAV subtypes. Among the four cases analysed using metagenomic sequencing with probe capture viral enrichment (13_FL, 14_NC, 15_IL and 16_SD), Epstein–Barr virus (EBV) was also detected in the blood sample from 13_FL.

Metagenomic sequencing was less sensitive than viral enrichment and targeted sequencing for detection of AAV2 (8 of 14 cases, 57%) and HAdV-41 (0 of 13 cases, 0%; Fig. 2b and Supplementary Table 1). However, reads from additional viruses were identified including EBV (n = 2), cytomegalovirus (CMV; n = 1), HAdV-2 (adenovirus type C; n = 1) and enterovirus A71 (EV-A71; n = 1) from whole blood, HAdV-1

Article

Table 1 | Demographic and clinical characteristics of cases and controls

Characteristic		Cases (n=16)		Controls, all (n=113)			Controls, acute hepatitis of defined aetiology (n=30)			Controls, acute gastroenteritis (n=23)		
		No.	Percentage	No.	Percentage	P value	No.	Percentage	P value	No.	Percentage	P value
Sex	Female	9	56.3%	56	49.6%	0.79	15	50.0%	0.76	10	43.5%	0.52
	Male	7	43.8%	57	50.4%		15	50.0%		13	56.5%	
Age (years) mean±s.d.		3.8±2.0		7.4±6.2		0.0239	10.7±5.8		<0.0001*	1.6±1.6		<0.0001*
Immunocompromised	Yes	0	0.0%	15	13.3%	0.21	6	20.0%	0.08	0	0.0%	1
	No	16	100.0%	98	86.7%		24	80.0%		23	100.0%	
ICU stay	Yes	2	12.5%	30	26.5%	0.50	15	50.0%	0.08	1	3.3%	0.27
	No	10	62.5%	71	62.8%		15	50.0%		22	73.3%	
	Unknown	4	25.0%	12	10.6%		0			0	0.0%	
ALT (U l⁻¹) mean±s.d.		2,293 ± 1,486		N/A			291 ± 288		<0.0001*	N//A		
AST (U l⁻¹) mean±s.d.		2,652 ± 1,851		N/A			455 ± 833		<0.0001*	N/A		
Fever	Yes	6	37.5%	40	35.4%	0.78	16	53.3%	0.10	13	43.3%	0.51
	No	9	56.3%	46	40.7%		7	23.3%		10	33.3%	
	Unknown	1	6.3%	27	23.9%		7	23.3%		0	0.0%	
Diarrhoea	Yes	8	50.0%	36	31.9%	0.58	10	33.3%	0.74	20	87.0%	0.03
	No	7	43.8%	45	39.8%		12	40.0%		3	13.0%	
	Unknown	1	6.3%	32	28.3%		8	26.7%		0	0.0%	
Nausea and/or vomiting	Yes	11	68.8%	38	33.6%	0.05	14	46.7%	0.50	17	73.9%	1
	No	0	0.0%	15	13.3%		2	6.7%		0	0.0%	
	Unknown	5	31.3%	60	53.1%		14	46.7%		6	26.1%	
Malaise and/or fatigue	Yes	10	62.5%	15	13.3%	0.016	9	30.0%	0.49	0	0.0%	N/A
	No	5	31.3%	35	31.0%		8	26.7%		0	0.0%	
	Unknown	1	6.3%	63	55.8%		13	43.3%		23	100.0%	
Abdominal pain	Yes	6	37.5%	22	19.5%	0.24	7	23.3%	0.71	2	8.7%	0.18
	No	10	62.5%	17	15.0%		7	23.3%		0	0.0%	
	Unknown	0	0.0%	74	65.5%		16	53.3%		21	91.3%	
Anorexia	Yes	10	62.5%	4	3.5%	0.08	3	10.0%	0.12	0	0.0%	N/A
	No	6	37.5%	10	8.8%		8	26.7%		0	0.0%	
	Unknown	0		99	87.6%		19	63.3%		23	100.0%	
Jaundice	Yes	13	81.3%	0	0.0%	<0.0001*	0	0.0%	<0.0001*	0	0.0%	N/A
	No	2	12.5%	25	22.1%		10	33.3%		0	0.0%	
	Unknown	1	6.3%	88	77.9%		20	66.7%		23	100.0%	
Cough	Yes	5	31.3%	13	11.5%	0.29	10	33.3%	0.74	0	0.0%	N/A
	No	8	50.0%	45	39.8%		12	40.0%		0	0.0%	
	Unknown	3	18.8%	55	48.7%		8	26.7%		23	100.0%	

Uncorrected P values were calculated using two-tailed Fisher's exact test for categorical variables and two-tailed unpaired t-test for continuous variables. A Bonferroni-corrected significance level of $P < 0.001$ was considered significant ($n = 50$ comparisons). Significant comparisons are denoted with an asterisk. ICU, intensive care unit; ALT, alanine transaminase; AST, aspartate transaminase.

(adenovirus type C; $n = 1$) from nasopharyngeal swab, and picobirnavirus ($n = 1$) from stool. EV-A71 was detected as a co-infection in an HAdV-41–AAV2 case, and HAdV-2 was detected in the single AAV2-negative case.

Among the 113 controls, AAV2 was detected in 4 (3.5%), including 2 (16.7%) of 12 children with acute HAdV-positive gastroenteritis, 1 (9.1%) of 11 hospitalized children with HAdV-negative gastroenteritis, and 1 donor control who was also positive for HAdV (Fig. 2b and Supplementary Table 1). The AAV2-positive child with HAdV-negative gastroenteritis was HAdV-negative from stool but both HAdV41-positive and CMV-positive from blood and had been discharged from the hospital with liver failure (although additional clinical details were not available). Of note, neither AAV2 nor HAdV-41 was detected among the 30 paediatric controls with acute hepatitis of defined aetiology and 42 hospitalized children without hepatitis.

We next carried out virus-specific PCR testing for EBV, CMV, HHV-6 and SARS-CoV-2 for all available cases and controls (Fig. 2b and Supplementary Table 1). EBV and HHV-6 were detected from blood in 11 (79%) of 14 and 7 (50%) of 14 cases, respectively, versus 1 (0.88%) of 113 controls for each virus. No tested cases were positive for CMV compared to 2 (1.8%) of 113 controls. SARS-CoV-2 was not detected in blood from cases or controls (Fig. 2b), or in liver biopsy tissue (Supplementary Table 1).

Virus associations with hepatitis cases

We used Fisher's exact test on cases ($n = 14$) and controls to investigate associations between detected viruses in blood and cases of acute severe hepatitis. Controls were stratified into four groups for comparison against cases (Fig. 2c and Supplementary Table 2). Three viruses,
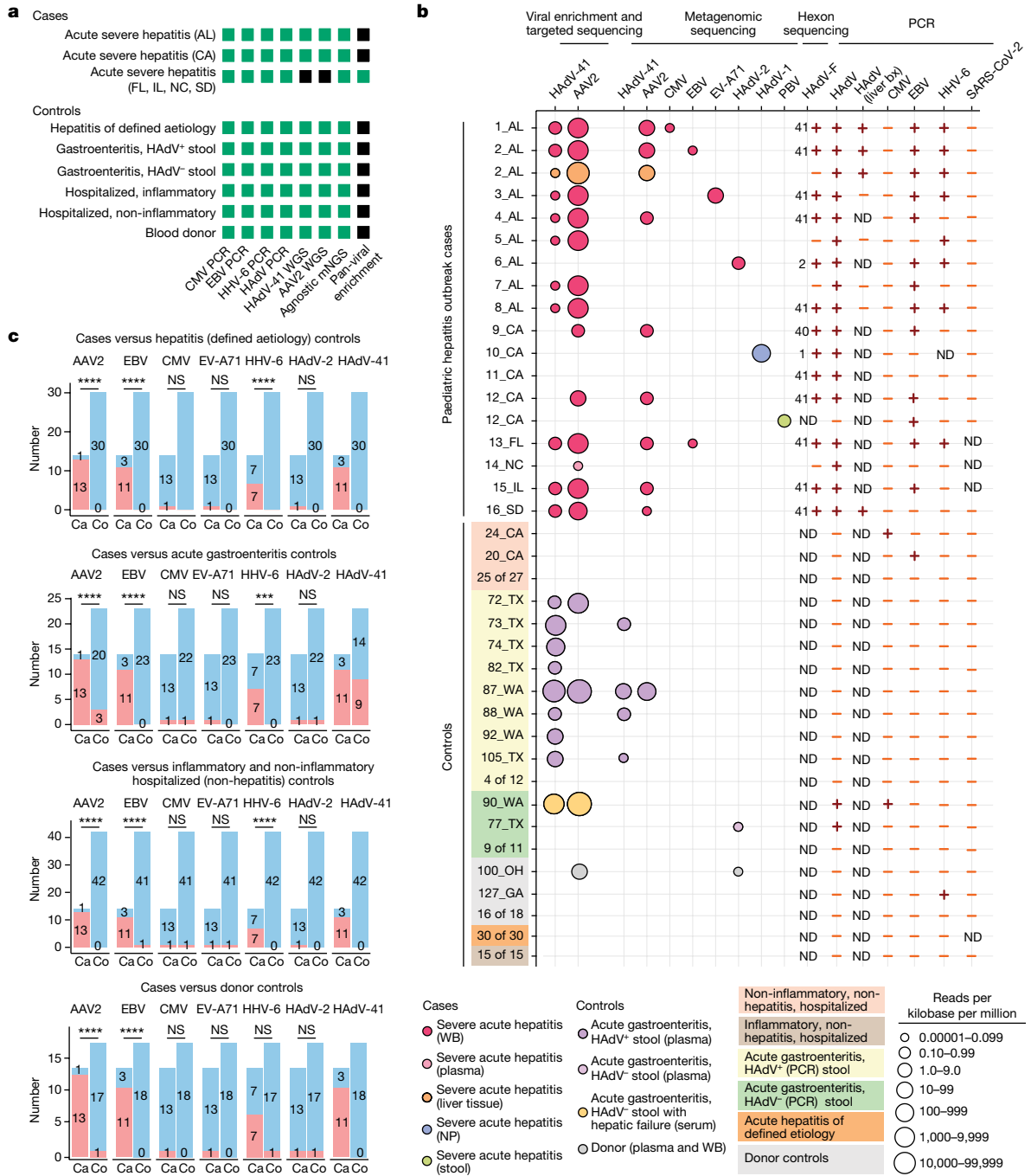
Fig. 2 | Sequencing and molecular-based testing for viruses carried out on cases and controls. a, Analyses carried out on the different sample groups. The specific assays carried out for each group are shown in green whereas assays that were not carried out are shaded in black. WGS, whole-genome sequencing. mNGS, metagenomic next-generation sequencing. **b**, Graphical chart showing results of sequencing or PCR-based assays for virus detection. Each circle represents a sequenced sample with at least three non-overlapping reads[29] aligning to a viral reference sequence. Circles are colour-coded on the basis of sample type and scaled according to normalized viral read counts. Viral PCR positive and negative results are denoted by plus and minus symbols, respectively. For the HAdV-F PCR, the HAdV subtype identified by hexon

sequencing is shown to the left of the viral PCR result. PBV, picobirnavirus; bx, biopsy; ND, not determined. **c**, Associations between viruses detected in blood and cases of acute severe hepatitis of unknown aetiology in children. Red and blue shading indicate positive and negative detection, respectively. Uncorrected *P* values were calculated using two-tailed Fisher's exact test. A Bonferroni-corrected significance level of *P* < 0.002 was considered significant (*n* = 24 comparisons). Exact *P* values are provided in Supplementary Table 2. Ca, cases; Co, controls; AL, Alabama; CA, California; FL, Florida; GA, Georgia; IL, Illinois; NC, North Carolina; OH, Ohio; SD, South Dakota; TX, Texas; WA, Washington; NS, not significant; ***P* < 0.001; ****P* < 0.0001.

AAV2, EBV and HHV-6, were significantly associated with cases when compared to each control group, except for HHV-6 detection between cases and donors (*P* = 0.010, Bonferroni-corrected significance level of *P* < 0.002). As all 14 cases were known to be HAdV-positive a priori, statistical analysis was not carried out for HAdV-41.

Phylogenetic and substitution analyses
We carried out multiple sequence alignment and phylogenetic analysis of 13 recovered AAV2 genomes from 12 cases with >25% breadth of coverage (Fig. 3 and Extended Data Fig. 1). Multiple
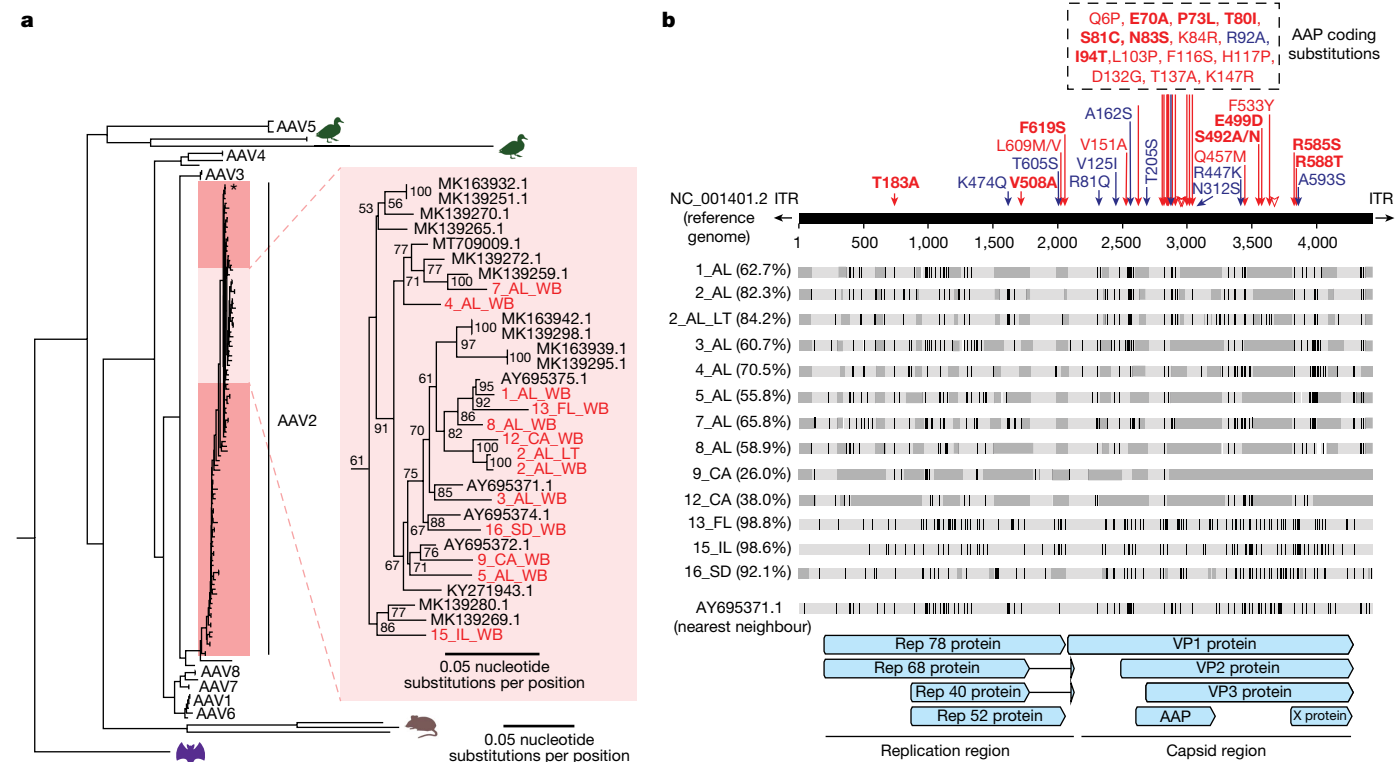
Fig. 3 | Phylogenetic and substitution analysis of AAV2 genomes.
a, Phylogenetic tree of 119 AAV2 genomes available from GenBank as of 18 August 2022. The 12 recovered genomes from this study with >25% coverage are denoted in red. The location of the AAV2 reference genome (NC_001401.2) is marked with a black asterisk. The phylogenetic tree was constructed by multiple sequence alignment of the AAV genomes or amino acid sequences using the MAFFT algorithm[30], followed by maximum-likelihood-based tree construction using IQ-TREE[31]. **b**, Multiple sequence alignment of 13 AAV2 genomes recovered from cases. Nucleotide mismatches are represented by black-coloured vertical lines, whereas areas of missing coverage are represented by dark grey rectangles. Amino acid variants with respect to the reference genome (NC_001401.2) are denoted in blue and red arrows; red arrows indicate shared substitutions that were reported in another study from the UK and Scotland[14]. Substitution sites that were identified in 100% of cases with sufficient coverage in both studies are highlighted in bold red text. ITR, inverted terminal repeat.

sequence alignment was carried out in parallel with all complete 119 AAV2 reference genomes deposited in GenBank as of 18 August 2022. A whole-genome nucleotide phylogenetic tree revealed that the 13 genomes were located within a distinct subgroup of a large human-infecting AAV2 clade (Fig. 3a). Other previously sequenced AAV2 genomes from France and the USA from patients without hepatitis were also found within this subgroup. Amino acid phylogenetic trees of the VP1 protein and the assembly-activating protein (AAP) showed slightly different topologies (Supplementary Fig. 1), but the 13 genomes were still positioned within a distinct subgroup.

To explore the underlying basis behind the observed phylogenetic clustering, we carried out multiple sequence alignment of the 13 genomes from cases. We then searched for shared coding substitutions relative to the AAV2 reference genome that were found in ≥50% of genomes with coverage at that substitution site. This analysis yielded 35 substitutions that were unevenly distributed across the viral genome, with 15 of 35 (42.9%) in the AAP, 14 of 35 (40%) in the VP1 protein, and 6 of 35 (17.1%) in the Rep78 protein (Fig. 3b and Supplementary Table 3). Clusters of substitutions were found to be located within hypervariable regions of the capsid VP1 and AAP proteins[13]. Of note, 25 (71.4%) of the 35 substitutions were shared with those identified in an independent study of severe paediatric hepatitis from the UK[14].

Discussion

Here we report virus findings from PCR, metagenomic and targeted sequencing of samples from 16 paediatric cases of acute severe hepatitis

of unknown aetiology that were first identified in the USA in October 2021. After combining these three diagnostic modalities, AAV2 was detected in 93% (13 of 14) of cases for whom whole blood or plasma samples were available (P < 0.001). Conversely, AAV2 was detected in only 4 (3.5%) of 113 controls and HAdV-41 was detected in 9 (8%) of 113. Notably, AAV2 was not detected in whole blood from 30 paediatric controls with hepatitis of defined aetiology. HAdV sequences were detected in all 14 cases, of which 11 (79%) were genotyped as HAdV-41. The 100% detection rate of HAdV is expected given that all 14 cases were known to be HAdV positive by clinical testing a priori. Other co-infecting viruses, including EBV and HHV-6, were detected by PCR in many cases but rarely in controls (P < 0.001); EV-A71 was also detected by metagenomic sequencing in one case but not in any control. Taken together, these findings show a significant association between co-infection by AAV2 and one or more hepatotropic viral pathogens and the clinical manifestations of severe acute hepatitis, although a direct causal link has yet to be confirmed.

Initial epidemiological investigation and molecular testing have found adenovirus in 45–90% of severe acute hepatitis cases[3–7]. In the UK, 116 of 179 (64.8%) reported cases as of May 2022 have tested positive for HAdV (ref. 1). This is in comparison to 8 of 9 (88.9%) positive cases in the USA corresponding to a cluster of cases who presented at a children's hospital in Alabama between October 2021 and March 2022 (ref. 5), for which there is a substantial overlap with cases from our study. An expanded national investigation in the USA detected HAdV in slightly less than half (44.6%) of cases[3]. Consistent with previous reports[5,14], HAdV genotyping of our cases by sequencing confirmed that most of the detected serotypes were HAdV-41 (11 of 14, 79%).

Our results suggest that co-infection with AAV2 may cause more severe liver disease than infection by an adenovirus and/or herpesvirus alone. This condition may be analogous to the fulminant liver failure that can occur when hepatitis D virus (HDV, 'delta') infection is superimposed on a chronic hepatitis B virus (HBV) infection[19]. Notably, AAV2 was seen in only 2 (16.7%) of 12 paediatric controls with HAdV-associated acute gastroenteritis in the absence of liver inflammation, suggesting that co-infection with AAV2 may predispose patients to more severe disease. Here we also found that in cases of AAV2–HAdV co-infection, viral loads for AAV2 were higher than for HAdV, with an approximately 12.7× increase in mean normalized read counts by targeted sequencing. This observation may be partially explained by previously published data showing that AAVs can suppress the replication of other hepatotropic viruses, including adenoviruses and herpesviruses[12]. Similarly, patients co-infected with HDV usually exhibit low HBV viral loads owing to suppression of HBV replication by HDV-induced interferons[20].

Among the AAVs, AAV2 is the most well-characterized AAV and has been shown to replicate to high titres in the liver and spleen[11]. Seroprevalence data from infants, children and adults demonstrate that 30–80% of the general population is seropositive and that natural infection with AAV can occur at all ages[21]. However, there is a peak of AAV2 infection between the ages of 1 to 5 years old[22], consistent with the observed distribution of ages of cases in the current study.

Phylogenetic analysis reveals that the recovered AAV2 genomes from all cases map to a distinct subclade. The positioning is driven by groups of substitutions located within hypervariable regions of the capsid VP1 and AAP proteins[13]. Notably, two of these substitutions in the VP1 protein, R585S and R588T, are arginine-to-serine and arginine-to-threonine substitutions that probably impact receptor binding, as these residues are necessary for the interaction of AAV2 with its heparan sulfate proteoglycan receptor[23]. These are not only shared among the US genomes in the current study but also overlap substantially with the substitutions found in an independent study of acute severe hepatitis cases in children from the UK[14]. Notably, several of these shared capsid substitutions (V151A, Q457M, S492A, E499D, F533Y, R585S and R588T) are also found in a sublineage of AAV2 (AAVv66) that exhibits increased replication, virion stability, central nervous system transduction and evasion of neutralizing antibodies relative to wild-type AAV2 (ref. 24). However, as other contemporary AAV2 genomes are not readily available, it is unclear whether the relatedness in AAV2 genomes across geographically dispersed regions and to AAVv66 merely represents detection of a predominant global circulating strain.

In the current study, hepatotropic viruses other than HAdV, including EBV, HHV-6 and EV-A71, were detected, albeit in a smaller proportion of cases. Infection by EBV or HHV-6 alone has been implicated in cases of liver failure requiring transplantation[25,26]. However, although we found substantial differences in the relative proportions of EBV and HHV-6 in cases compared to controls, we do not believe it likely that these viruses are the primary cause of acute severe hepatitis in these children. First, the viral loads for these herpesviruses were very low, with median PCR cycle threshold (C_t) values of 38.1 and 38 for EBV and HHV-6, respectively, and thus the borderline positive PCR results may represent detection of integrated proviral DNA rather than bona fide low-level herpesvirus viraemia. Second, herpesvirus reactivation can be seen in various cytokine-associated inflammatory conditions such as coronavirus disease 2019 (COVID-19)[27], and whether such reactivation contributes to disease pathogenesis is unclear. Third, herpesviruses were not detected in available liver biopsy tissue from cases, and the only viruses detected were HAdV (two of six, 33%) and AAV2 (one of one, 100%). Nevertheless, it is striking that among the 16 cases, triple or quadruple infections with AAV2, adenovirus, EBV and/or HHV-6 were detected in whole blood from at least 12 cases (75%). We postulate that the COVID-19 pandemic and more than 2 years of school and childcare closures and decreased social interactions with other children may

have generated a vulnerable population of young children who failed to develop broad immunity to common viral pathogens owing to lack of exposure. Notably, among the nine cases with available information, five of nine (55.6%) children had never attended school or a childcare centre, and the remaining four were probably hindered from interacting with other children by social isolation measures enacted at the onset of the pandemic. Decreased immunity from lack of exposure to common viral pathogens may have predisposed cases to infection by multiple viruses, thus increasing the likelihood of more severe disease manifestations such as hepatitis.

Limitations of the study include: lack of availability and/or incomplete testing of liver biopsies, particularly for the presence or absence of AAV2, making it difficult to ascertain the pathogenetic mechanisms underlying the viral infection; low titres associated with detected viral pathogens; and the retrospective study design and the inclusion of only cases that had previously tested positive for HAdV.

In summary, here we identify a distinct strain of AAV2 and co-infection with at least one helper virus in blood from US paediatric cases of acute severe hepatitis of unknown aetiology. These results are consistent with findings from independent and contemporary studies of acute severe hepatitis in children from Scotland and the UK[14,28]. AAV2 infection may contribute to the pathogenesis and/or severity of the hepatitis, or alternatively, may be a non-pathogenic marker of liver inflammation. Further studies including serologic surveillance, viral culture and animal models are needed to investigate the potential role that AAV2 infection may play in this disease.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-023-05949-1.

1. Khader, S., Foster, I., Dagens, A., Norton, A. & Sigfrid, L. Severe acute hepatitis of unknown aetiology in children-what is known? *BMC Med.* **20**, 280 (2022).
2. Severe acute hepatitis of unknown aetiology in children - multi-country. *WHO Disease Outbreak News* https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON400 (2022).
3. Cates, J. et al. Interim analysis of acute hepatitis of unknown etiology in children aged <10 years - United States, October 2021-June 2022. *MMWR Morb. Mortal. Wkly Rep.* **71**, 852–858 (2022).
4. Elsheikh, R. et al. Acute hepatitis of unknown origin in children: Behind the statistics. *Hepatology* https://doi.org/10.1002/hep.32682 (2022).
5. Gutierrez Sanchez, L. H. et al. A case series of children with acute hepatitis and human adenovirus infection. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2206294 (2022).
6. Kelgeri, C. et al. Clinical spectrum of children with acute hepatitis of unknown cause. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2206704 (2022).
7. 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**, 2200318 (2022).
8. *Technical Report: Acute Hepatitis of Unknown Cause* (CDC, 2022); https://www.cdc.gov/ncird/investigation/hepatitis-unknown-cause/technical-report.html.
9. Kager, J. et al. Fulminant adenoviral-induced hepatitis in immunosuppressed patients. *Viruses* **14**, 1459 (2022).
10. Hierholzer, J. C. Adenoviruses in the immunocompromised host. *Clin. Microbiol. Rev.* **5**, 262–274 (1992).
11. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).
12. Meier, A. F., Fraefel, C. & Seyffert, M. The interplay between adeno-associated virus and its helper viruses. *Viruses* **12**, 662 (2020).
13. Gao, G. et al. Adeno-associated viruses undergo substantial evolution in primates during natural infections. *Proc. Natl Acad. Sci. USA* **100**, 6081–6086 (2003).
14. Ho, A. et al. Adeno-associated virus 2 infection in children with non-A-E hepatitis. *Nature* https://doi.org/10.1038/s41586-023-05948-2 (2023).
15. Recommendations for adenovirus testing and reporting of children with acute hepatitis of unknown etiology. *CDC Health Alert Network* https://emergency.cdc.gov/han/2022/han00462.asp (2022).
16. Chiu, C. Y. & Miller, S. A. Clinical metagenomics. *Nat. Rev. Genet.* **20**, 341–355 (2019).
17. Deng, X. et al. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat. Microbiol.* **5**, 443–454 (2020).
18. Kim, K. W. et al. Respiratory viral co-infections among SARS-CoV-2 cases confirmed by virome capture sequencing. *Sci. Rep.* **11**, 3934 (2021).

Article

19. Negro, F. Hepatitis D virus coinfection and superinfection. *Cold Spring Harb. Perspect. Med.* **4**, a021550 (2014).

20. Urban, S., Neumann-Haefelin, C. & Lampertico, P. Hepatitis D virus in 2021: virology, immunology and new treatment approaches for a difficult-to-treat disease. *Gut* **70**, 1782–1794 (2021).

21. La Bella, T. et al. Adeno-associated virus in the liver: natural history and consequences in tumour development. *Gut* **69**, 737–747 (2020).

22. Calcedo, R. et al. Adeno-associated virus antibody profiles in newborns, children, and adolescents. *Clin. Vaccine Immunol.* **18**, 1586–1588 (2011).

23. Chen, C. L. et al. Molecular characterization of adeno-associated viruses infecting children. *J. Virol.* **79**, 14781–14792 (2005).

24. Hsu, H. L. et al. Structural characterization of a novel human adeno-associated virus capsid with neurotropic properties. *Nat. Commun.* **11**, 3279 (2020).

25. Chevret, L. et al. Human herpesvirus-6 infection: a prospective study evaluating HHV-6 DNA levels in liver from children with acute liver failure. *J. Med. Virol.* **80**, 1051–1057 (2008).

26. Mellinger, J. L. et al. Epstein-Barr virus (EBV) related acute liver failure: a case series from the US Acute Liver Failure Study Group. *Dig. Dis. Sci.* **59**, 1630–1637 (2014).

27. Brooks, B. et al. Epstein-Barr virus and human herpesvirus-6 reactivation in acute COVID-19 patients. *Viruses* **14**, 1872 (2022).

28. Morfopoulou, S. et al. Genomic investigations of unexplained acute hepatitis in children. *Nature* https://doi.org/10.1038/s41586-023-06003-w (2023).

29. Miller, S. et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res.* **29**, 831–842 (2019).

30. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

31. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

Pediatric Hepatitis of Unknown Etiology Working Group

Olivia Almendares[7], Neha Balachandran[7], Jordan E. Cates[7], Brandi Freeman[7], Anita Kambhampati[7], Rachel Marine[7], Everardo Vega[7], Mary E. Wiksow[7], Erin R. McKeever[7], David Kelly[14], Julie A. Boom[19], Leila C. Sahni[19], Janet A. Englund[20], Eileen J. Klein[20], Mary A. Staat[21], Maeve Marsh[9], Erin Ricketts[10], Joshua L. Clayton[11], Natalia Cano[8], Miriam Oseguera[1] & Robert Blattner[1]

[19]Baylor College of Medicine, Houston, TX, USA. [20]Seattle Children's Research Institute, Seattle, WA, USA. [21]Children's Hospital Medical Center, Cincinnati, OH, USA.

Methods

Ethics statement

Remnant clinical samples from cases with acute severe hepatitis were collected and analysed under 'no subject contact' protocols with waiver of informed consent approved by the institutional review boards (IRBs) of the University of Alabama, Birmingham, the California Department of Public Health, the New York State Department of Health and the Centers for Disease Control and Prevention (CDC). Whole-blood samples from paediatric controls (age < 18) from Children's Healthcare of Atlanta were prospectively collected and analysed under a protocol approved by the Emory IRB (STUDY00000723); parents or guardians of these children provided oral consent for study enrolment and collection and analysis of their samples. Remnant whole-blood samples from paediatric controls (age < 18) at University of California, San Francisco (UCSF) were collected, biobanked and analysed under a 'no subject contact' protocol with waiver of informed consent approved by the UCSF IRB (protocol no. 11-05519). A subset of the control samples was provided by the CDC from children enrolled in the National Vaccine Surveillance Network (NVSN) study. Approval for the NVSN study was obtained from the institutional review board at each participating site and from the CDC (protocol no. 6164). Parents or guardians of eligible children provided written informed consent for participant enrolment. Blood specimens were also collected as remnant samples from clinical procedures.

Participant recruitment and sample collection

This was a retrospective observational case–control study using all available samples from cases and controls. A severe acute hepatitis case enrolled in this study was a person under investigation by local, state or federal public health agencies, defined as a person <10 years of age with elevated (>500 U l^{-1}) AST or ALT, an unknown aetiology for the hepatitis, and onset on or after 1 October 2021 (ref. 8). All cases (n = 16) were hospitalized with acute elevation in liver enzymes, AST or ALT, and one or more of the following symptoms on presentation: nausea, vomiting, jaundice, generalized weakness and abdominal pain. Cases were selected for inclusion into the current study if blood samples had previously tested positive for adenovirus by clinical testing, resulting in a selection bias that made case–control comparisons related to adenovirus detection not meaningful. Blood samples from cases meeting the criteria for a person under investigation but testing negative for adenovirus were not available.

Controls from UCSF in California (n = 54) included children hospitalized with hepatitis of defined aetiology (>100 U l^{-1}) or another inflammatory or non-inflammatory condition. These controls were selected on the basis of availability of convenience samples from a study of biomarkers of acute inflammatory disease, including severe COVID-19, in hospitalized patients. Controls were selected to be geographically similar (located within the same state) to cases from California. Remnant whole-blood samples from controls from UCSF (n = 54) were retrospectively biobanked and aliquoted with addition of 2× DNA/RNA Shield (Zymo Research) in a 1:1 ratio by volume and stored at −80 °C until use.

Controls from Children's Healthcare of Atlanta (n = 24) in Georgia included children hospitalized with hepatitis of defined aetiology (>100 U l^{-1}) or another inflammatory or non-inflammatory condition (n = 18) or blood donors (n = 6). These controls were selected from available biobanked samples from consented participants enrolled in a prospective study of COVID-19 and multi-system inflammatory syndrome in children. Controls were selected to be geographically similar (located within a neighbouring state with similar demographic characteristics) to the cases from Alabama and Florida[32]. Collected samples were stored at −80 °C until use.

Serum or plasma samples from children enrolled in the NVSN study were also included as controls. Three groups of children were selected: children admitted for acute gastroenteritis who tested positive for adenovirus in the stool (n = 12), children admitted for acute gastroenteritis who tested negative for adenovirus in the stool (n = 11) and blood donors (n = 12). These children were enrolled from three sites (Seattle, Houston and Cincinnati) from June 2021 to May 2022 and enrolment criteria have previously been described[33]. Collected samples were stored at −80 °C until use.

Nucleic acid extraction

Whole-blood and plasma samples collected from cases from Alabama were extracted at the Wadsworth Center laboratory using the NucliSENSÒ easyMAG 'specific B' protocol (bioMerieux) according to the manufacturer's instructions. Samples from cases from Florida, Illinois, North Carolina and South Dakota were extracted using the Zymo Direct-zol DNA/RNA Miniprep Kit (Zymo Research) following the manufacturer's instructions. Briefly, 200 μl of sample was extracted and total nucleic acid was eluted in 60 μl and stored at −80 °C until use.

Samples from cases from California were extracted at the California Department of Public Health. Whole-blood samples (200 μl) were extracted using the Qiagen Blood Mini Kit (Qiagen) according to the manufacturer's instructions. Total nucleic acid was eluted in 100 μl and stored at −80 °C until use. Respiratory samples, serum, plasma and clarified stool suspensions were extracted using the NucliSENS easyMAG instrument (bioMerieux). Briefly, 300 μl of nasopharyngeal swab, serum or plasma sample or 140 μl of clarified stool suspension was lysed with 1 ml of lysis buffer and incubated for 10 min at room temperature, followed by addition of 100 μl magnetic silica and transfer to the instrument to begin the automated extraction. Total nucleic acid was eluted in 110 μl for nasopharyngeal swab, serum or plasma or 60 μl for stool and stored at −80 °C before use.

For control samples from California, Georgia or the NVSN study, total nucleic acid was extracted at UCSF from the original sample using two different protocols for whole blood, plasma or serum. Whole-blood samples (400 μl) that had been pretreated with DNA/RNA Shield (Zymo Research) were extracted using the Quick-RNA Whole Blood Kit (Zymo Research) according to the manufacturer's instructions. Total nucleic acid was eluted in 15 μl and stored at −80 °C until use. Plasma or serum samples (200 μl) were extracted using the Mag-Bind Viral DNA/RNA 96 Kit (Omega Bio-Tek) on a KingFisher Flex instrument (Thermo-Fisher) according to the manufacturer's instructions. Total nucleic acid was eluted in 100 μl and stored at −80 °C until use.

Viral PCR testing

Samples from California cases were screened for adenovirus using a pan-adenovirus PCR targeting all HAdVs[34] and/or a group F adenovirus (HAdV-40 or HAdV-41) real-time PCR[35]. A cycle threshold cutoff of ≤40 was used to call a positive result by PCR. PCR and Sanger sequencing of the HAdV hexon gene targeting hypervariable regions 1–6 (ref. 36) were carried out on all adenovirus-positive samples with purified PCR products sequenced in-house or sent to an outside laboratory (Sequetech, Mountain View, CA) for sequencing. Sanger sequences were assembled and edited in Sequencher 5.2.4 (Gene Codes) and analysed using the nucleotide BLAST aligner[37]. Samples from cases outside California and all control samples were screened for adenovirus using a pan-adenovirus PCR assay[34] and sequenced using a nested PCR assay targeting HAdV hexon hypervariable regions 1–6, followed by Sanger sequencing[38]. PCR testing for detection of CMV, EBV and HHV-6 was carried out as previously described[39].

Viral enrichment-based targeted sequencing

For cases from California and Alabama and all controls, viral enrichment followed by targeted sequencing of viral genomes was carried out using custom spiked primers designed to target HAdV-41 or AAV1–AAV8 genomes. Spiked primers were designed using the metagenomic sequencing with spiked primer enrichment (MSSPE) algorithm[17] as follows. For HAdV-41, 22 representative HAdV-41 genomes were aligned

Article

using MAFFT, and for AAV, 11 genomes representing AAV1–AAV8 were aligned using MAFFT. Next, automated primer design was carried out by running the MSSPE algorithm with the following parameters: kmer size = 25, segment window = 500, e-value = 0.1, dS = 2, dG = −9,000. The algorithm generated 219 HAdV and 150 AAV spiked primers (Supplementary Table 4). PCR amplification from extracted DNA was carried out as previously described[40], with the following modifications: the annealing temperature for PCR was 60 °C for HAdV and 53 °C for AAV. The HAdV primers were amplified in two different reactions using two different sets of primers (Supplementary Table 4, Set1 and Set2), whereas the AAV2 primers were amplified in a single reaction. NGS libraries were prepared using the NEBNext Ultra II DNA Kit (New England Biolabs) and Revelo DNA-Seq Mech Kit on the MagicPrep NGS system (Tecan Genomics). Final libraries were quantified using the Qubit Flex instrument (Invitrogen) with the dsDNA HS Assay Kit (Invitrogen). Libraries were pooled and sequenced on an Illumina NextSeq using 300-base-pair (bp) single-end sequencing. Negative template controls were included in every run to monitor for contamination. No contamination from HAdV-41 or AAV2 reads was detected in the negative template control libraries.

For cases from Florida, Illinois, North Carolina and South Dakota, probe capture target enrichment of viral genomes was carried out on metagenomic libraries using the Twist Comprehensive Viral Research Panel (Twist Biosciences), which covers reference genomes of 3,153 viruses and 15,488 different strains[18]. Individual DNA libraries or cDNA libraries from RNA were hybridized to Comprehensive Viral Research Panel probes according to the manufacturer's instructions. Libraries were then barcoded and sequenced on an Illumina MiSeq (250-bp paired-end sequencing) or an Illumina NextSeq (150-bp paired-end sequencing). For plasma sample 14_NC, nucleic acid carrier, either bacteriophage lambda DNA (New England Biolabs) or HeLa total RNA (Thermo-Fisher), was added during the library preparation step according to the manufacturer's instructions.

Viral metagenomic sequencing and analysis

Metagenomic DNA and RNA libraries were prepared using the Revelo DNA-Seq Mech Kit on the MagicPrep NGS system (Tecan Genomics), NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) and NEBNext Ultra II RNA Library Prep Kit (New England Biolabs), according to the manufacturer's instructions. Libraries were pooled and sequenced on a NextSeq 550 Sequencing System using 150-bp single-end sequencing. Potential contamination was monitored in each run by processing water and negative controls in parallel with samples.

Sequencing data from all cases and controls were analysed for viral nucleic acids using SURPI+ (v1.0.7-build.4)[41], an automated bioinformatics pipeline for pathogen detection and discovery from metagenomic data that has been modified to incorporate enhanced filtering and classification algorithms[29]. A threshold of ≥3 non-overlapping reads was used for calling a positive virus detection[29].

Phylogenetic analysis

Multiple sequence alignments were carried out using MAFFT algorithm (v7.388)[30] as implemented in Geneious (version 10.0.9)[42]. Nucleotide and amino acid phylogenetic trees were inferred using a maximum-likelihood method with ultrafast bootstrap approximation as implemented in IQ-TREE (version 1.6.1)[31] using 1,000 bootstrap replicates. Trees were visualized using FigTree (version 1.4.4).

Viral genome assembly and analysis

Binary base call (BCL) files generated by Illumina sequencers were simultaneously demultiplexed and converted to FASTQ files using bcl2fastq (version 2.20.0.422). Custom scripts were used to assemble AdV and AAV2 genomes as follows. Briefly, raw FASTQ reads were filtered using BBDuk (version 38.87)[43] for removal of adaptors, primer sequences and low-quality reads, and then HAdV-41 or AAV reads were identified by Bowtie2 (ref. 44) alignment (parameters: -D 20 -R 3 -L11 -N1) to a reference database consisting of 1,395 HAdV-41 or 3,600 AAV partial sequences or genomes, respectively. These aligned reads were then mapped to the HAdV-41 reference genome (accession DQ315364.2) or consecutively to AAV genomes 1–8. For all AAV genomes, the assembly with the highest breadth of coverage corresponded to the AAV2 reference genome (accession number NC_001401.2). Consensus assemblies were generated using iVar[45] (parameters: -t 0.5 -m 1). AAV consensus genomes were further analysed for shared alterations in the nucleotide and translated nucleotide (amino acid) sequences relative to the AAV2 reference by carrying out multiple sequence alignment using the MAFFT algorithm[30], followed by visualization of the alignment using Geneious software (version 10.0.9)[42].

Statistical analysis

Statistical analyses were carried out using the Python scipy package (version 1.5.2)[46] and rstatix package (version 0.7.0) in R (version 4.0.3)[47]. Uncorrected P values were calculated using two-tailed Fisher's exact test for categorical variables and two-tailed unpaired t-test for continuous variables, with the significance level determined after a Bonferroni correction for multiple comparisons.

Data visualization

Plots were generated using matplotlib (version 3.3.2), seaborn (version 0.11.0) and plotly (version 5.6.0) packages in Python software (version 3.7.12), Jupyter notebook (version 6.1.4), RStudio (version 1.4) and Adobe Illustrator (version 26.4.1) software.

Reporting summary

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

Data availability

Raw metagenomic sequencing data with human reads removed have been deposited in the NCBI Sequence Read Archive (Bioproject accession number PRJNA918667, under umbrella BioProject accession number PRJNA171119). Detailed protocols for spiked primer enrichment and metagenomic library preparation, consensus genome files, alignment files, phylogenetic tree files and supplementary tables have been uploaded to a Zenodo data repository (https://doi.org/10.5281/zenodo.7089581).

Code availability

Custom scripts for HAdV-41 and AAV genome assembly are available in a Zenodo data repository (https://doi.org/10.5281/zenodo.7089581).

32. QuickFacts: Alabama. US *Census Bureau* https://www.census.gov/quickfacts/AL (2021).
33. Esona, M. D. et al. Rotavirus genotype trends and gastrointestinal pathogen detection in the United States, 2014-2016: results from the New Vaccine Surveillance Network. *J. Infect. Dis.* **224**, 1539–1549 (2021).
34. Heim, A., Ebnet, C., Harste, G. & Pring-Akerblom, P. Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *J. Med. Virol.* **70**, 228–239 (2003).
35. Lyman, W. H. et al. Prospective study of etiologic agents of acute gastroenteritis outbreaks in child care centers. *J. Pediatr.* **154**, 253–257 (2009).
36. Lu, X. & Erdman, D. D. Molecular typing of human adenoviruses by PCR and sequencing of a partial region of the hexon gene. *Arch. Virol.* **151**, 1587–1602 (2006).
37. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
38. Okada, M., Ogawa, T., Kubonoya, H., Yoshizumi, H. & Shinozaki, K. Detection and sequence-based typing of human adenoviruses using sensitive universal primer sets for the hexon gene. *Arch. Virol.* **152**, 1–9 (2007).
39. Dupuis, M. et al. Molecular detection of viral causes of encephalitis and meningitis in New York State. *J. Med. Virol.* **83**, 2172–2181 (2011).
40. 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**, 1261–1276 (2017).
41. Naccache, S. N. et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.* **24**, 1180–1192 (2014).

42. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
43. Bushnell, B. BBMap: a fast, accurate, splice-aware aligner. https://www.osti.gov/servlets/purl/1241166 (2014).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
45. Grubaugh, N. D. et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).
46. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
47. R Core Team. *R: A Language and Environmental for Statistical Computing* http://www.R-project.org (R Foundation for Statistical Computing, 2013).

Author contributions C.Y.C., J.V., R.T.R. and K.S.G. conceived and designed the study. C.Y.C., V.S. and A.S.G. coordinated the study. V.S., A.S.G., A.F., H.J.H., P.E.O. and J.T. carried out metagenomic sequencing and targeted viral enrichment experiments. V.S. and D. Stryke ran the SURPI computational pipeline for detection of viruses from metagenomic data. V.S. and C.Y.C. analysed metagenomic and targeted viral enrichment data, assembled AAV2 genomes and carried out phylogenetic and substitution analyses. A.L.B., N.H.B., R.L.B., K.Y.G., J.S.G., S.P.L., K.P., D.R. and R.T.R. carried out viral probe capture enrichment experiments and assembled AAV2 genomes. R.J., M.T.L., Z.L. and H.D.H. investigated acute severe hepatitis cases and collected samples and clinical metadata. U.P., M.S., K.W., M.S.O., J.E.T., J.M.B., D. Sugerman, C.P., X.L., P.C., J.V., H.L.K. and the CDC Pediatric Hepatitis of Unknown Etiology Working Group investigated acute severe hepatitis cases for the CDC and collected samples and clinical metadata. L.A.I., H.S., W.B. and L.H.G.S. investigated acute severe hepatitis cases in Alabama and collected samples and clinical metadata. C.C. and C.A.R. provided samples and clinical metadata from paediatric controls from Georgia. V.S., A.S.G., J.N. and C.Y.C. provided samples and clinical metadata from paediatric controls from California. A.H., C.-Y.P., K.R., J.B.B., C.M. and D.A.W. investigated acute severe hepatitis cases and carried out viral PCR testing and HAdV genotyping at the California Department of Public Health. D.M.L. and K.S.G. carried out viral PCR testing and HAdV genotyping at the Wadsworth Center/New York State Department of Public Health. C.Y.C. and V.S. wrote the initial draft of the manuscript and prepared the figures. C.Y.C., V.S., A.S.G., U.P., J.E.T., J.M.B., C.P., X.L., J.V., H.L.K. and R.T.R. edited the manuscript. All authors read the manuscript and agreed to its contents.

Extended Data Fig. 1 | Amino acid phylogenetic analysis of AAV2 genomes from cases, related to Fig. 3. (**A**) Phylogenetic tree corresponding to the Assembly Activating Protein (AAP) protein. (**B**) Phylogenetic tree corresponding to the VP1 protein. Phylogenetic trees was constructed by multiple sequence alignment of the AAV genomes or amino acid sequences using the MAFFT algorithm[30], followed by maximum likelihood based tree construction using IQ-TREE[31].

nature portfolio

Corresponding author(s):	Charles Chiu
Last updated by author(s):	1/5/2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Viral PCR and hexon sequencing data were collected using manufacturer-provided software associated with the instrument. Metagenomic and whole-genome sequencing data were collected on Illumina sequencers (MiSeq or NextSeq Sequencing System) using manufacturer-provided software.
Data analysis	A cycle threshold cutoff of ≤40 was used to call a positive result by PCR. Sanger sequences were assembled and edited in Sequencher 5.2.4 (Gene Codes) and analyzed using the nucleotide BLAST aligner. Binary base call (BCL) files generated by Illumina sequencers were simultaneously demultiplexed and converted to FASTQ files using bcl2fastq (version 2.20.0.422). Metagenomic sequencing data from all cases and controls were analyzed for viral nucleic acids using SURPI+ (v1.0.7-build.4), an automated bioinformatics pipeline for pathogen detection and discovery from metagenomic data that has been modified to incorporate enhanced filtering and classification algorithms. A threshold of ≥ 3 non-overlapping reads was used for calling a positive virus detection. A custom script was used to assemble AdV and AAV2 genomes as follows. Briefly, raw FASTQ reads were filtered using BBDuk (version 38.87) for removal of adapters, primer sequences, and low-quality reads, and then HAdV-41 or AAV reads were identified by Bowtie2 alignment (parameters: -D 20 -R 3 -L11 -N 1) to a reference database consisting of 1,395 HAdV-41 or 3,600 AAV partial sequences / genomes, respectively. These aligned reads were then mapped to the HAdV-41 reference genome (accession DQ315364.2) or consecutively to AAV genomes 1-8. For all AAV genomes, the assembly with the highest breadth of coverage corresponded to the AAV2 reference genome (accession number NC_001401.2). Consensus assemblies were generated using iVar (parameters: -t 0.5 -m 1). AAV consensus genomes were further analyzed for shared mutations in the nucleotide and translated nucleotide (amino acid) sequences relative to the AAV2 reference by performing a multiple sequence alignment using the MAFFT algorithm (v7.388), followed by visualization of the alignment using Geneious

software (version 10.0.9).

Multiple sequence alignments were performed using MAFFT algorithm (v7.388) as implemented in Geneious (version 10.0.9). Nucleotide and amino acid phylogenetic trees were inferred using a maximum likelihood method with ultrafast bootstrap approximation as implemented in IQ-TREE (version 1.6.1). Trees were visualized using FigTree (version 1.4.4).

Statistical analyses were performed using the Python scipy package (version 1.5.2) and rstatix package (version 0.7.0) in R (version 4.0.3). Fisher's Exact Test was used to assess the association between variables. All statistical tests were conducted as two-sided at the 0.05 significance level.

Plots were generated using matplotlib (version 3.3.2), seaborn (version 0.11.0) and plotly (version 5.6.0) packages in Python software (version 3.7.12), Jupyter notebook (version 6.1.4), RStudio (version 1.4) and Adobe Illustrator (version 26.4.1) software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw metagenomic sequencing data with human reads removed have been deposited in the NCBI Sequence Read Archive (Bioproject accession number PRJNA918667, under umbrella BioProject accession number PRJNA171119). Consensus genome files, alignment files, phylogenetic tree files and supplementary tables have been uploaded to a Zenodo data repository (doi: 10.5281/zenodo.7089581).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Sex and gender were not considered in study design. Analysis was performed on aggregated data..
Population characteristics	In this study, 27 samples (21 whole blood, 2 plasma, 1 liver tissue, 1 nasopharyngeal swab, and 2 stool sample(s)) from 16 children with acute severe hepatitis of unknown etiology were analyzed. All children met the clinical case definition established by the CDC, including lack of a confirmed etiology, liver enzyme levels (aspartate aminotransferase (AST) or alanine aminotransferase (ALT)) >500 UL, age<10 years, and onset on or after October 1, 2021. Cases were enrolled from 6 states (Alabama, California, Florida, Illinois, North Carolina, and South Dakota) from October 1, 2021 to May 22, 2022. All 16 cases had positive testing for HAdV from blood, and thus HAdV infection was over-represented compared to the overall affected population, of which HAdV is detected in 45-90%. The median age of affected children was 3 years; 56% were female and 44% male. Controls (n=113) consisted of 78 whole blood, 1 serum, and 34 plasma samples. Many controls were enrolled from California (n=54) and Georgia (n=24) to be geographically similar to the cases, with the remaining controls enrolled from Ohio (n=12), Texas (n=14), and Washington (n=9). Sixty-nine of 113 controls (61.0%) were collected over the same time frame as the cases (i.e., collected between October 1, 2021 – May 22, 2022). Differences in age and gender between cases and controls were non-significant. The 113 controls consisted of 42 patients (37%) without hepatitis, 30 (26.5%) patients with acute hepatitis (ALT > 100 U/L) of defined etiology, 23 (21%) patients with acute gastroenteritis (12 with positive HAdV stool testing), and 18 (16%) blood donors.
Recruitment	This was a retrospective observational case-control study using all available samples from cases and controls. A severe acute hepatitis case enrolled in this study was a person under investigation (PUI) by local, state, or federal public health agencies, defined as a person <10 years of age with elevated (>500 U/L) aspartate aminotransferase (AST) or alanine aminotransferase (ALT), an unknown etiology for the hepatitis, and onset on or after October 1, 20218. All cases (n=16) were hospitalized with acute elevation in liver enzymes, asparate aminotransferase (AST) or alanine aminotransferase (ALT), and one or more of the following symptoms on presentation: nausea, vomiting, jaundice, generalized weakness, and abdominal pain. Cass were selected for inclusion into the current study if blood samples had previously tested positive for adenovirus by clinical testing, resulting in a selection bias that made case-control comparisons related to adenovirus detection not meaningful. Blood samples from cases meeting PUI criteria but testing negative for adenovirus were not available. Controls from University of California, San Francisco (UCSF) in California (n=54) included children hospitalized with hepatitis of defined etiology or another inflammatory or non-inflammatory condition. These controls were selected to be geographically similar (located within the same state) to cases from California. Remnant whole blood samples from controls from UCSF (n=54) were retrospectively biobanked and aliquoted with addition of 2X DNA/RNA Shield (Zymo Research) in a 1:1 ratio by volume and stored at -80°C until use. Controls from Children's Healthcare of Atlanta (n=24) in Georgia included children hospitalized with hepatitis of defined etiology or another inflammatory or non-inflammatory condition (n=18) or blood donors (n=6). These controls were selected to be geographically similar (located within a neighboring state with similar demographic characteristics) to the cases from

Alabama and Florida. Collected samples were stored at -80°C until use.

Serum or plasma samples from children enrolled in the New Vaccine Surveillance Network (NVSN) were also included as controls. Three groups of children were selected: children admitted for acute gastroenteritis who tested positive for adenovirus in the stool (n=12), children admitted for acute gastroenteritis who tested negative for adenovirus in the stool (n=11), and blood donors (n=12). These children were enrolled from three sites (Seattle, Houston, and Cincinnati) from June 2021 to May 2022. Collected samples were stored at -80°C until use.

Ethics oversight

Remnant clinical samples from cases with acute severe hepatitis were collected and analyzed under "no subject contact" protocols with waiver of informed consent approved by the institutional review boards (IRBs) of University of Alabama, Birmingham, California Department of Public Health, New York State Department of Health, and CDC. Whole blood samples from pediatric controls (age < 18) from Children's Healthcare of Atlanta were prospectively collected and analyzed under a protocol approved by the Emory IRB (STUDY00000723); parents or guardians of these children provided oral consent for study enrollment and collection and analysis of their samples. Remnant whole blood samples from pediatric controls (age < 18) at University of California, San Francisco (UCSF) were collected, biobanked, and analyzed under a "no subject contact" protocol with waiver of informed consent approved by the UCSF IRB (protocol no. 11-05519). A subset of the control samples was provided by the CDC from children enrolled in the National Vaccine Surveillance Network (NVSN) study. Approval for the NVSN study was obtained from the institutional review board at each participating site and from the Centers for Disease Control and Prevention (protocol no. 6164). Parents or guardians of eligible children provided written informed consent for participant enrollment. Blood specimens were also collected as leftover samples from clinical procedures.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No sample size was calculated. All available samples were processed. 27 samples (21 whole blood, 2 plasma, 1 liver tissue, 1 nasopharyngeal swab, and 2 stool sample(s)) from 16 pediatric patients with acute severe hepatitis of unknown etiology were analyzed. Affected children included in this analysis were admitted to tertiary care hospitals in 6 states (Alabama, California, Florida, Illinois, North Carolina, and South Dakota) from October 1, 2021 to May 22, 2022.
Controls (n=113) consisted of 78 whole blood, 1 serum, and 34 plasma samples. The majority of controls were enrolled from California (n=54) and Georgia (n=24) to be geographically similar to the cases, with the remaining controls enrolled from Ohio (n=12), Texas (n=14), and Washington (n=9). The controls consisted of a mix of remnant convenience samples (California) and samples from patients prospectively enrolled in a clinical (Georgia) or population-level surveillance (Ohio, Texas, and Washington) study.

Data exclusions

No data were excluded from the analysis.

Replication

Samples were processed only once but analyses were done on independent sample cohorts from 6 states (Alabama, California, Florida, Illinois, North Carolina, and South Dakota) and independent control cohorts from California, Georgia, Ohio, Texas and Washington state.

Randomization

Randomization is not applicable as this is a retrospective observational case-control study that used all available samples from cases (pediatric acute severe hepatitis) and controls.

Blinding

Researchers were blinded during processing as all samples were assigned a random unique ID. Cases and controls were analyzed in parallel. Samples were unblinded during analyses to interpret and visualize results and to determine associations between detection of virus and severe acute hepatitis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology and archaeology
☒	Animals and other organisms
☒	Clinical data
☒	Dual use research of concern

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging