Title: Use of whole-genome sequencing of adenovirus in immunocompromised paediatric patients to identify nosocomial transmission and mixed-genotype infection

Short title: Clinical adenovirus genomics & epidemiology

Authors

Charlotte J. Houldcroft 1,2α, Sunando Roy 2, Sofia Morfopoulou 2, Ben K. Margetts 1,3,4, Daniel P. Depledge 2†, Juliana Cudini 2, Divya Shah 4, Julianne R. Brown 4, Erika Yara Romero 2, Rachel Williams 2, Elaine Cloutman-Green 4, Kanchan Rao 4, Joseph F. Standing 1,4, John Hartley 4, Judith Breuer 1,2,4§

§Corresponding author

Professor Judith Breuer, Division of Infection & Immunity, Cruciform Building, London, WC1E 6BT, UK

j.breuer@ucl.ac.uk Tel: 020 3108 2130

Affiliations

1. Infection, Immunity & Inflammation section, UCL Great Ormond Street Institute of Child Health, University College London, UK.

2. Division of Infection and Immunity, University College London, Cruciform Building, Gower Street, London, UK.

3. Centre for Computation, Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX), University College London, London, UK.

4. Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK.

α Current address: Department of Medicine, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge, UK.

† Current address: Department of Microbiology, New York University, USA

Paper summary (36/40 words)

This study used whole-genome sequencing directly from clinical samples to analyse adenovirus from immunocompromised paediatric patients. We identified one nosocomial transmission event which included cases that occurred four years apart and one case of mixed-genotype adenoviraemia.

Abstract 197/200 words

Background: Adenoviruses are significant pathogens for the immunocompromised, arising from primary infection or reinfection. Serotyping is insufficient to support nosocomial transmission investigations. We investigate whether whole-genome sequencing (WGS) provides clinically relevant information on transmission among patients in a paediatric tertiary hospital.

Methods: We developed a target-enriched adenovirus WGS technique for clinical samples and retrospectively sequenced 107 adenovirus-positive residual diagnostic samples, including viraemias (>5x10^4 copies/ml), from 37 patients collected January 2011 – March 2016. WGS was used to determine genotype and for phylogenetic analysis.

Results: Adenovirus sequences were recovered from 105/107 samples. Full genome sequences were recovered from all 20 non-species C samples and from 36/85 species C viruses, with partial genome sequences recovered from the rest. Whole genome phylogenetic analysis suggested linkage of three genotype A31 cases and uncovered an unsuspected epidemiological link to an A31 infection first detected on the same ward four years earlier. In nine samples from one patient who died we identified a mixed genotype adenovirus infection.

Conclusions: Adenovirus WGS from clinical samples is possible and useful for genotyping and molecular epidemiology. WGS identified likely nosocomial transmission with greater resolution than conventional genotyping, and distinguished between adenovirus disease due to single or multiple genotypes.

Keywords: adenovirus; genotype; molecular epidemiology; whole-genome sequencing; nosocomial transmissions; genomics

Main text: 3486 words (max 3500 words)

Introduction

Human mastadenoviruses (HAdV) are double-stranded DNA viruses from the family Adenoviridae, which can cause morbidity and mortality through infection of a range of tissues [1]. HAdVs are classified into seven species, A-G. The seven species each contain various numbers of genetically distinct types, initially defined by serum neutralization methods as serotypes, but more recently distinguished based on whole genome sequencing and genomics as genotypes. Adenovirus infections are common in children, with outbreaks seen in a number of settings [2–4]. No vaccine is available for general use [5].

In immunocompromised paediatric patients, adenovirus presents a particular problem, accounting for 15% of diarrhoea in paediatric oncology patients [6]. Adenoviraemia is seen in 11% of paediatric human stem cell transplant (HSCT) recipients, associated with significant costs due to increased time spent in hospital and anti-viral drug costs [7]. There are limited treatment options (cidofovir [8], brincidofovir [9,10] and HAdV-specific T cells [11,12]), and thus increased mortality from HAdV in paediatric HSCT recipients. Immunocompromised individuals may also serve as a sentinel population for the detection of new strains, as these individuals may be simultaneously infected with more than one genotype, providing circumstances in which recombination may occur [13]. Identifying these patients sooner may be important for infection control to prevent onward transmission of new genotypes.

Studies of reported HAdV outbreaks in children's hospitals ([14–16]) have relied on serotyping or molecular typing to investigate the epidemiological linkage of cases. HAdV genotyping with small, single gene regions can be misleading, producing inconsistent molecular and serological typing profiles compared to the whole genome sequence [17,18]. Whole-genome sequencing (WGS) increases the resolution of molecular epidemiology and for other viruses has distinguished between patient-to-patient transmission [19] and cases of importation of virus from outside sources (eg family members) [20]. Few

hospitals undertake HAdV genotyping for routine diagnostics, making it difficult to compare data in different patient populations.

From 2010 we introduced selective molecular hexon genotyping of adenovirus infections as a means of identifying putative health care associated infections (HCAIs) particularly among patients undergoing HSCTs. However, the data were insufficient to resolve if infections were truly nosocomial. With methods that allow high-throughput WGS directly from clinical material [21–24] we examine the utility of WGS for the identification and management of adenovirus HCAIs. To do this, we optimised HAdV genome sequencing using three cultured isolates, with specific RNA-bait-based target enrichment [22]. Following this we extracted DNA directly from clinical samples and sequenced all available adenoviraemias and associated HAdV-positive samples occurring over a 15-month period between January 2015 and March 2016 (101 samples from 37 patients) in a tertiary care hospital in London, UK, and six archived DNA samples from five patients collected between January 2011 and July 2012. Many of the samples sequenced had been previously hexon genotyped; additional hexon genotyping data was available for other patients although the clinical samples had been discarded.

Methods

Ethics and samples

Samples were submitted to the UCL Infection DNA Bank for use in this study andsupplied in an anonymised form. The use of these specimens for research was approved by the NRES Committee London – Fulham (REC reference: 17/LO/1530).

These samples were collected from patients under 18 years of age as part of standard clinical care between January 2011 and July 2016. All samples were residual diagnostic specimens obtained from

patients with confirmed adenovirus infections [25]. Data on hexon genotype was available for patients who had undergone HSCT and were A31 positive between 2011 and mid-2016. Dates of hexon genotyping, WGS and ward stays are shown in figure 5. Clinical samples, DNA extraction and hexon genotyping are described in supplementary table 1 and supplementary methods.

Based on previous experience of the initial sensitivity of enrichment WGS, we selected predominately whole blood samples for this study with reported viral loads of >50,000 copies /ml. Six archived DNA samples (extracted from whole blood) from 2011 and 2012 were also available from five patients with adenoviraemia, with reported virus loads in the range 4292-1,000,000 copies/ml.

For optimisation, three typed strains D9 ATCC VR-1086 and E4 ATCC VR-1572 (both cultured in Vero cells), and F40 ATCC VR-931 (cultured in HEK-293 cells) that had each been passaged three times were sequenced. Cultured adenovirus was purified using OptiPrep density gradient medium (Sigma-Aldrich), following manufacturer's instructions.

Sure Select^{XT} bait design

120mer RNA baits were designed, using an in-house perl pipeline with a tiling factor of 6x (each position in a given genome is covered by six unique bait designs), against all whole human adenovirus sequences and supplementary hexon, penton and fiber gene sequences available in GenBank (accessed 29/10/2015). The bait design (Supplementary File 1) was uploaded to SureDesign and biotinylated RNA oligonucleotides (baits) synthesised by Agilent Technologies.

Sequencing

Hybridisation and library preparation was performed as previously described [23]. Briefly, extracted DNA was sheared by acoustic sonication (Covaris e220, Covaris Inc.). DNA fragments underwent end-repair, A'-tailing, and (Illumina) adaptor ligation. DNA libraries were hybridised with the biotinylated RNA baits

for 24 hrs at 65^{0}C and subsequently bound to MyOne™ Streptavidin T1 Dynabeads™ (ThermoFisher Scientific). Following washing, libraries were minimally amplified (12-22 cycles) to generate sufficient input material for Illumina sequencing. Samples were multiplexed to 48 samples per run. Paired end sequencing was performed on an Illumina MiSeq using 500 v2 Reagent Kit (Illumina, MS-102-2003). Base calling and sample demultiplexing were performed as standard on the MiSeq and paired FASTQs were generated for each sample. Where samples from different patients clustered, a second sequence, where additional clinical material was available, from each patient was analysed in a different run to ensure that the result was not due to sample mix up.

Genome mapping, assembly and phylogenetic analysis

Sequence data were analysed using CLC Genomics Workbench (Qiagen) version 8.5.1. Reads were quality trimmed and adapter sequences removed. Trimmed reads were mapped to a curated reference list of complete adenovirus genome sequences in Genbank as of 29/11/2015 (n=113). The CLC Genomics Workbench Microbial Genome Finishing Module was applied for *de novo* assembly of samples with predicted 100% genome coverage (based on mapping to reference genomes). *De novo* assemblies were accepted if a single contig of at least 34,000 bp could be generated. *De novo* assembly of two high coverage genomes produced identical consensus sequences to mapping to the closest reference sequence; we therefore mapped all sequences to the closest reference to produce consensus sequences to ensure consistency of mapping methodology between high and low coverage samples.

All reads mapping to the reference list (filtered reads) were taken forward for remapping to the best reference match (length fraction 0.8, similarity fraction 0.8) and a consensus sequence was generated. The best reference match was used to assign a genotype to each sample. Reads that did not map were assumed to be off-target (not adenovirus). Mapping parameters were as previously published [26], except for a minimum contig size of 500. The mapping mode was set to map reads back to contigs

(slow). Minority variants were called if: the base was sequenced at least six times; the variant was present in ≥six (including two forward and two reverse) reads; and it was present at a frequency ≥2%. The read direction filter significance was 0.05 and the relative read direction filter significance was 0.01. Full and partial consensus genomes were aligned using MAFFT [27] default settings, with manual correction in MEGA6 if required [28]. Phylogenetic trees were created using RAxML BlackBox [29].

Nucleotide diversity

Within-host nucleotide diversity was calculated for 12 samples from five patients with average read depths ≥1000X. For these twelve samples, reads were reference-mapped and base counts at each position were extracted to calculate within-host nucleotide diversity (π), defined as the average number of nucleotide differences between reads at a site [30]. Bases sequenced at ≥2X were used for imputation. Strand-bias and random error rates were estimated and corrected for using maximum likelihood methods [31].

To separate genotypes present in samples believed to contain mixed infections, all samples were first mapped to a panel of references, where 80% of each read mapped with minimum 80% identity. Samples with high coverage mapping to different genotypes were then re-mapped to the top two references where ≥95% of each read was mapped with a minimum of 95% sequence identity to reduce cross-mapping between the two reference genotypes. BAM files were generated containing reads mapping to either reference, and nucleotide diversity was calculated for each to determine single-genotype adenovirus diversity. Pairwise differences between genomes were calculated using UGENE v1.26.1 [32].

Statistical analysis

One-way single factor ANOVA with correction for multiple testing was used to compare the success of sequencing between genotypes. Figures were generated and statistical analyses performed in MATLAB 2015b.

Results

Success of the sequencing methodology

The WGS performance is shown in supplementary table 2. All three cultured samples gave genome coverage of 100% with mean read depths of 1000X. Near-complete genomes (>80% coverage) were recovered from 56/107 clinical samples all of which had >10X read depth. All 23 non-species C viruses produced full genomes. Two clinical samples failed, both with virus loads below 50,000 copies/ml. The minimum viral load needed for >80% coverage at any depth was found to be species-dependent (supplementary table 3). Despite similar average virus loads and mean read depths, species C viruses yielded lower percentage genome coverage than non-C viruses ($P = 0.00008$), with lower percentage on-target reads ($P = 0.016$, not significant) (figure 1, supplementary figure 1). Whole genomes were recovered from 36/85 species C samples. In nine out of 107 samples, all from a single patient, we identified mixed infections with at least two HAdV genotypes.

In total we identified seven genotypes among 105 clinical samples on the basis of similarity to reference HAdV genomes, the majority of which were species C (figure 2); in two clinical samples, there were no adenovirus-specific reads to use in molecular genotyping. Genotypes A31, B3, C C1, C2, C5 and E4 were detected in whole blood; genotype F41 was detected in stool or stool-contaminated urine only (figure 3). While genotype C2 was detected in 36.2% of the samples sequenced, it was only present in 7/27 (26%) patients with adenoviraemia. The large number of C2 samples from these seven patients reflects the long duration of viraemia during the year of sampling. No patients carried any of the previously reported polymerase mutations associated with drug resistance [33].

Molecular epidemiology and sample phylogenies

Focusing on sequences with ≥80% genome coverage (excluding identical sequences from sequentially sampled patients) we first constructed a maximum likelihood phylogenetic tree of 37 adenovirus genomes (supplementary figure 2). The tree confirmed genotyping based on hexon sequences. Two clusters defined as monophyletic groups comprising two or more samples from at least two patients with bootstrap support ≥90, were observed (figure 4). Analysis of these together with available GenBank genomes confirmed one A31 (*1) and one C1 (*2) cluster. The putative clusters and phylogenies are shown for each species (figure 4).

Figure 4A shows the phylogenetic tree of A31 sequences including Cluster *. This cluster comprised patients P8, P7, P18 and P25. All four were chronically immunosuppressed and had been under the same medical team and on the same ward at different times over a five year period (temporal relationship shown in figure 5). Patient 19 had also been on the same ward and under the same team during this period, but was infected with a genotype A31 virus that was distinct from the monophyletic cluster (figure 5). Patients 29 and 35 who also had distinct A31 viruses had had periods of time under the same team and briefly on the same ward as the others although neither were A31 positive during these periods. Hexon typing of blood and/or stool showing infection with A31 was available for 15 patients (including patients 19 and 8) who had been on the same ward as the patients in the cluster (figure 5), however for 13 of these patients, no samples were available for WGS. Overall, P18 was likely to have been the originator of the cluster. Although discharged in 2012, three years before P8, the next patient in the cluster, became positive, transmission of the A31 cluster may have been sustained by other patients during this period (figure 5). P8 became A31 hexon-positive (stool) in a different ward in early 2015, shortly after admission. However, our data confirms that at least three other A31 strains were present at various times in children on this ward and since no whole genome sequence data is

available for the other A31 infected patients or the January 2015 P8 samples, it is not possible to determine the exact route of transmission from P18 to P8; P8 may have been infected by an unknown route with this A31 strain at presentation, or they may have been infected with two different A31 strains during their inpatient stay (the second, falling in cluster 1, captured by WGS). The possibility of environmental transmission [34] remains since hexon typing of samples taken in 2011 from cubicles on the ward were A31 positive, although rooms were screened and re-cleaned until they became adenovirus-negative by PCR.

Within and between patient differences

To verify the cluster analyses, we calculated pairwise nucleotide differences for consensus adenovirus sequences from same-patient sequential samples, samples from identified clusters, newly sequenced samples which do not cluster and unrelated HAdV sequences (figure 6A). The pairwise differences between sequences in the A31 cluster were identical or nearly identical (≤3 nucleotide differences for cluster *1 sequences) to the within-patient differences (figure 6B). In contrast the lowest pairwise distance of samples in the cluster to sequences from patients 19, 35 and 29 were 55, 148 and 44 respectively.

The monophyletic cluster of C1 sequences comprising patients P23, P43 and a publicly available sequence collected in 2003 differed by 33-39 nucleotides as compared to 4 nucleotides for genotype C1 within-patient variation (figure 6B). Neither patient was on the same ward, nor were they looked after by the same team. The two genotype F41 patient samples (P27 and P28 (figure 4E)) shared a unique single nucleotide deletion at position 1074 located in an intron of gene E1A which distinguished them from other F41 sequences; however, they did not form a well-supported monophyletic cluster, with a pairwise distance of 20 and no history to link them. The close clustering in these examples is likely to

reflect the paucity of UK HAdV genome sequences and the probable existence of local variants that are more closely related than other non-UK viruses.

As shown in figure 4C, the species C sequences partition into genotypes 1, 2 and 5, except for sequences from patient P5. The polyphyletic sequences from patient 5 suggested a mixed-genotype infection, with some sequences clustering with C5 genomes (eg GOSH_2055, 2074, 2054 and 2051), two sequences with C1 genomes (GOSH_2049, 2052), and a further two sequences (GOSH_2005 and GOSH_2007) clustering in an intermediate position. By mapping reads specifically to individual adenovirus C1 or C5 genomes (supplementary figure 3) and creating a phylogeny, we confirmed from the allele frequencies for each genotype, that a mixture of C1 and C5 infection, rather than a recombinant virus was likely to be present. The pairwise differences between consensus sequences from patient 5 samples were intermediate in frequency between genomes from different patients, and genomes from different genotypes, consistent with the hypothesis that this patient had a mixed-genotype infection (figure 6A).

Finally, we calculated adenovirus within-host nucleotide diversity for eight samples from this patient. We also calculated within-host nucleotide diversity for samples of comparable depth from patients infected with single genotypes: A31 (P35, GOSH_2094), B3 (P3, GOSH_2004), C1 (P23, GOSH_2050) and C2 (P14, GOSH_2033). When total reads mapping to any adenovirus genome were considered, samples from patient 5, apart from the final sample, had high nucleotide diversity (figure 7). When reads were mapped stringently to either the C1 or C5 genome, the nucleotide diversity dropped to be comparable with the single genotype infections from other patients. This supported our inference of a mixed genotype infection in this patient, which resolved to a single genotype infection by the final sample (GOSH_2055). The proportion of reads mapping to C1 and C5 adenovirus fluctuated over time within patient 5 (Supplementary figure 4). By day 25 of PCR-detectable viraemia (sample GOSH_2055) the

sequence had returned to a C5 genotype and the secondary C1 genotype appears to have been purged from the population.

Discussion

Whole genome sequencing with phylogenetic analysis identified a monophyletic cluster of patients infected with adenovirus A31 which was potentially explained by nosocomial transmission. For patients P7, P8 and P25 standard IPC methods had already flagged up the likelihood of HCAI. The link of P18 to the other three was unexpected given that there was no overlap in the times that P18 was an inpatient with the other three. Previous hexon typing had shown that A31 infections were present among patients on this ward. However, the high degree of hexon sequence conservation makes it difficult to discriminate between related and unrelated strains. Patient P19, who was also A31-positive, was in the same ward as P18 in 2012 and was still A31 positive in 2015, but had a phylogenetically distinct strain that was not part of cluster 1. P19's virus differed from both P18 and P8 by 55 SNPs. Despite the passage of time, P18 and P8 sequences differed by only two SNPS. These data underline the superiority of WGS over hexon typing for HCAI investigation , with the latter showing low specificity for establishing relatedness. Without genome sequences from the many A31 cases that occurred on this ward between the inpatient stays of P18 and P8, we cannot be sure how the virus came to be transmitted over a three year period. There is evidence for persistence of adenovirus DNA on environmental surfaces for up to 3 months following infection [35,36], despite extensive cleaning [37]. This raises the possibility of transmission from the environment or fomites; A31 virus had been detected in two ward environments during suspected-outbreak management in 2011, and children with A31 were present on the ward (Figure 5) or in outpatient follow-up (data not shown) throughout. Community transmission also remains a possibility.

Two other putative clusters, the C1 cluster *2 and an F41 cluster were identified. In both cases, the pairwise distance between viruses was considerably higher than is seen for within-patient adenovirus variation. For cluster *2, an unrelated adenovirus sequence collected in the USA in 2003 [38] separated viruses P23 and P43. The F41 sequences, despite being genetically relatively distant, share a unique deletion that was not present in GenBank and may be a locally circulating variant. There are currently fewer than 10 species A, 10 genotype C1 and 15 species F adenovirus genomes publicly available. With additional sequences from both local and wider geographical areas, the granularity of data available (including the frequency of indels at particular loci) will increase and phylogeny will become more useful for identifying putative transmission events.

The availability of whole genomes and variant data also allowed us to identify a mixed genotype C1/C5 infection in patient 5. Adenoviraemia was detected in patient 5 within two weeks following HSCT and this patient ultimately died of overwhelming adenovirus infection. While mixed adenovirus infection of faecal [39] and urine [40] samples is commonly reported, this is the first reported case (to our knowledge) of a mixed adenoviraemia. Whether or not the dual infection contributed to the severity of outcome, is not clear. Further WGS should elucidate whether dual adenoviraemia is an indicator of poor prognosis.

The WGS approach presented here shows the utility of target-enriched adenovirus sequencing in a clinical setting directly from clinical samples [41], without culture [42] or specific PCR amplification [43]. While this first analysis reveals the need for re-design of the baits to better capture species C adenoviruses, this process is now well established and has been successfully applied to other DNA ([22,23,44]) and RNA ([45,46]) viruses.

Our data show that WGS was able to confirm nosocomial transmission of HAdV infection in immunosuppressed patients. We uncovered a linked case that had not previously been suspected and

which suggested sustained nosocomial transmission occurring over several years. While adenovirus is recognised to cause serious disease in children [7], better data are needed to determine whether, and from where these are acquired, as transmission within hospitals is a recognised phenomenon (including contact in in- and out-patient environments [36,47]). Our data suggest that the extent to which this happens may be greater than hitherto suspected and WGS data is necessary to identify the measures needed to interrupt spread of this virulent virus. HAdV genome deep-sequencing also allows rapid identification of potentially mixed infections, which demonstrate recognisably greater nucleotide diversity than single strain infections, which adenovirus detection and PCR alone cannot. The clinical implications of mixed infections are as yet unknown, but their identification may be important if, as with HCMV, they are associated with poorer prognosis [48]. With decreasing costs, improved methods and increasing automation, adenovirus genome sequencing offers a realistic potential for better understanding nosocomial transmission and pathogenesis.

Data availability

Genome assemblies are available from the European Nucleotide Archive, study accession PRJEB24711. Bait sequences are included as supplementary data.

Provisional data from this study was presented at the 19th Annual Meeting of the European Society for Clinical Virology, 14th-17th September 2016, Lisbon, Portugal[49]; and the Microbiology Society Annual Conference, 3rd-6th April 2017, Edinburgh, UK.

Author contributions

CJH, JFS and JBreuer conceived the study. DPD designed the sequencing baits. CJH, DS and ECG provided samples. ECG cultured type virus strains. BKM, JBrown, DS, KR, JH, JFS and JBreuer provided clinical data.

EYR and RW sequenced the samples. CJH, SM, SR and JC analysed the data. CJH, JH and JBreuer wrote the paper. All authors read and approved the final manuscript.

Funding

This work was funded by Action Medical Research grant GN2424 and supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. CJH was funded by Action Medical Research grant GN2424. DPD was supported by an MRF New Investigator Award. SM was funded by the Reuben Foundation. JB receives funding from the NIHR UCL/UCLH Biomedical Research Centre.

Conflicts of interest

Relevant conflicts of interest have been declared to the editor.

Acknowledgements

We acknowledge infrastructure support from the UCL Pathogen Genomics Unit (PGU) and UCL MRC Centre for Molecular Medical Virology. The clinical adenovirus samples were provided by Great Ormond Street Hospital for Children Departments of Microbiology, Virology and IPC to the UCL Infection DNA Bank. Virus load data was provided by the Clinical Microbiology and Virology laboratory at Great Ormond Street Hospital for Children NHS Foundation Trust. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The authors thank Dr Alison Niewiarowska for providing clinical details, and Dr Krishna Kumar and Dr James Taylor for advice on data analysis.

References

1. Echavarría M. Adenoviruses. eLS [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2015 [cited

2016 Dec 12]. Available from: http://doi.wiley.com/10.1002/9780470015902.a0000409.pub2

2. Kajon AE, Lamson DM, Bair CR, et al. Adenovirus Type 4 Respiratory Infections among Civilian Adults, Northeastern United States, 2011–20151. Emerg Infect Dis [Internet]. **2018** [cited 2018 Jan 22]; 24(2):201–209. Available from: http://wwwnc.cdc.gov/eid/article/24/2/17-1407_article.htm

3. Kajon AE, Hang J, Hawksworth A, et al. Molecular Epidemiology of Adenovirus Type 21 Respiratory Strains Isolated From US Military Trainees (1996–2014). J Infect Dis [Internet]. Oxford University Press; **2015** [cited 2018 Jan 28]; 212(6):871–880. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/jiv141

4. Kalimuddin S, Chan YFZ, Wu IQ, et al. A Report of Adult Human Adenovirus Infections in a Tertiary Hospital. Open Forum Infect Dis [Internet]. Oxford University Press; **2017** [cited 2018 Jan 28]; 4(2). Available from: https://academic.oup.com/ofid/article-lookup/doi/10.1093/ofid/ofx053

5. Echavarría M. Adenoviruses in Immunocompromised Hosts. Clin Microbiol Rev. **2008**; 21(4):704–715.

6. Mhaissen MN, Rodriguez A, Gu Z, et al. Epidemiology of Diarrheal Illness in Pediatric Oncology Patients. J Pediatric Infect Dis Soc [Internet]. **2016**; . Available from: http://www.ncbi.nlm.nih.gov/pubmed/27578209

7. Hiwarkar P, Gaspar HB, Gilmour K, et al. Impact of viral reactivations in the era of pre-emptive antiviral drug therapy following allogeneic haematopoietic SCT in paediatric recipients. Bone Marrow Transplant [Internet]. Nature Publishing Group; **2013** [cited 2016 Nov 8]; 48(6):803–808. Available from: http://www.nature.com/doifinder/10.1038/bmt.2012.221

8. Lugthart G, Oomen MA, Jol-van der Zijde CM, et al. The Effect of Cidofovir on Adenovirus Plasma

DNA Levels in Stem Cell Transplantation Recipients without T Cell Reconstitution. Biol Blood Marrow Transplant. **2015**; 21(2):293–299.

9. Florescu DF, Pergam SA, Neely MN, et al. Safety and Efficacy of CMX001 as Salvage Therapy for Severe Adenovirus Infections in Immunocompromised Patients. Biol Blood Marrow Transplant. **2012**; 18(5):731–738.

10. Voigt S, Hofmann J, Edelmann A, Sauerbrei A, Kühl J-S. Brincidofovir clearance of acyclovir-resistant herpes simplex virus-1 and adenovirus infection after stem cell transplantation. Transpl Infect Dis [Internet]. **2016** [cited 2017 Feb 21]; 18(5):791–794. Available from: http://doi.wiley.com/10.1111/tid.12582

11. Lugthart G, Albon SJ, Ricciardelli I, et al. Simultaneous Generation of Multivirus-specific and Regulatory T Cells for Adoptive Immunotherapy. J Immunother [Internet]. **2012** [cited 2016 Dec 12]; 35(1):42–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22130159

12. Chatziandreou I, Gilmour KC, McNicol A-M, et al. Capture and generation of adenovirus specific T cells for adoptive immunotherapy. Br J Haematol [Internet]. Blackwell Publishing Ltd; **2007** [cited 2016 Dec 12]; 136(1):117–126. Available from: http://doi.wiley.com/10.1111/j.1365-2141.2006.06386.x

13. Lion T. Adenovirus infections in immunocompetent and immunocompromised patients. Clin Microbiol Rev [Internet]. American Society for Microbiology; **2014** [cited 2016 Dec 12]; 27(3):441–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24982316

14. Mattner F, Sykora K-W, Meissner B, Heim A. An Adenovirus Type F41 Outbreak in a Pediatric Bone Marrow Transplant Unit. Pediatr Infect Dis J [Internet]. **2008** [cited 2016 Dec 12]; 27(5):419–424. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18382384

15. Swartling L, Allard A, Törlen J, Ljungman P, Mattsson J, Sparrelid E. Prolonged outbreak of adenovirus A31 in allogeneic stem cell transplant recipients. Transpl Infect Dis [Internet]. **2015** [cited 2016 Dec 12]; 17(6):785–794. Available from: http://doi.wiley.com/10.1111/tid.12443

16. Faden H, Wynn RJ, Campagna L, Ryan RM. Outbreak of adenovirus type 30 in a neonatal intensive care unit. J Pediatr. **2005**; 146(4):523–527.

17. Seto D, Jones MS, Dyer DW, Chodosh J. Characterizing, typing, and naming human adenovirus type 55 in the era of whole genome data. J. Clin. Virol. 2013. p. 741–742.

18. Singh G, Robinson CM, Dehghan S, et al. Overreliance on the Hexon Gene, Leading to Misclassification of Human Adenoviruses: Fig 1. J Virol [Internet]. American Society for Microbiology; **2012** [cited 2017 Feb 22]; 86(8):4693–4695. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22301156

19. Zhu Y, Zembower TR, Metzger KE, Lei Z, Green SJ, Qi C. Investigation of Respiratory Syncytial Virus Outbreak on an Adult Stem Cell Transplant Unit by Use of Whole-Genome Sequencing. J Clin Microbiol [Internet]. American Society for Microbiology; **2017** [cited 2018 Jan 24]; 55(10):2956–2963. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28747373

20. Kundu S, Lockwood J, Depledge DP, et al. Next-generation whole genome sequencing identifies the direction of norovirus transmission in linked patients. Clin Infect Dis [Internet]. Oxford University Press; **2013** [cited 2016 Dec 12]; 57(3):407–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23645848

21. Olp LN, Jeanniard A, Marimo C, West JT, Wood C. Whole-Genome Sequencing of Kaposi's Sarcoma-Associated Herpesvirus from Zambian Kaposi's Sarcoma Biopsy Specimens Reveals Unique Viral Diversity. J Virol [Internet]. American Society for Microbiology; **2015** [cited 2018 Jan

2]; 89(24):12299–308. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26423952

22. Depledge DP, Palser AL, Watson SJ, et al. Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. Jhaveri R, editor. PLoS One [Internet]. Public Library of Science; **2011** [cited 2016 Nov 7]; 6(11):e27805. Available from: http://dx.plos.org/10.1371/journal.pone.0027805

23. Houldcroft CJ, Bryant JM, Depledge DP, et al. Detection of Low Frequency Multi-Drug Resistance and Novel Putative Maribavir Resistance in Immunocompromised Pediatric Patients with Cytomegalovirus. Front Microbiol [Internet]. Frontiers; **2016** [cited 2016 Nov 7]; 7:1317. Available from: http://journal.frontiersin.org/Article/10.3389/fmicb.2016.01317/abstract

24. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature [Internet]. Nature Publishing Group; **2016** [cited 2018 Jan 2]; 530(7589):228–232. Available from: http://www.nature.com/articles/nature16996

25. Brown JR, Shah D, Breuer J. Viral gastrointestinal infections and norovirus genotypes in a paediatric UK hospital, 2014–2015. J Clin Virol. **2016**; 84:1–6.

26. Christiansen MT, Brown AC, Kundu S, et al. Whole-genome enrichment and sequencing of Chlamydia trachomatis directly from clinical samples. BMC Infect Dis [Internet]. **2014**; 14:591. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4233057&tool=pmcentrez&rendertype=abstract

27. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol [Internet]. Oxford University Press; **2013** [cited 2017 Feb 22]; 30(4):772–780. Available from: https://academic.oup.com/mbe/article-

lookup/doi/10.1093/molbev/mst010

28. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol [Internet]. Oxford University Press; **2013** [cited 2017 Feb 22]; 30(12):2725–2729. Available from: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst197

29. Stamatakis A, Hoover P, Rougemont J, Renner S. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Syst Biol [Internet]. Oxford University Press; **2008** [cited 2017 Feb 22]; 57(5):758–771. Available from: https://academic.oup.com/sysbio/article-lookup/doi/10.1080/10635150802429642

30. Nei M, Li W-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci [Internet]. **1979**; 76(10):5269–5273. Available from: http://www.ncbi.nlm.nih.gov/pubmed/291943

31. Giardina F, Romero-Severson EO, Albert J, Britton T, Leitner T. Inference of Transmission Network Structure from HIV Phylogenetic Trees. Tanaka MM, editor. PLOS Comput Biol [Internet]. Public Library of Science; **2017** [cited 2018 Jan 28]; 13(1):e1005316. Available from: http://dx.plos.org/10.1371/journal.pcbi.1005316

32. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics [Internet]. Oxford University Press; **2012** [cited 2017 Mar 16]; 28(8):1166–1167. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts091

33. Kinchington PR, Araullo-Cruz T, Vergnes J-P, Yates K, Gordon YJ. Sequence changes in the human adenovirus type 5 DNA polymerase associated with resistance to the broad spectrum antiviral

cidofovir. Antiviral Res. **2002**; 56(1):73–84.

34. Pankhurst L, Cloutman-Green E, Canales M, D'Arcy N, Hartley JC. Routine monitoring of adenovirus and norovirus within the health care environment. Am. J. Infect. Control. 2014.

35. Kramer A, Schwebke I, Kampf G, et al. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. BMC Infect Dis [Internet]. BioMed Central; **2006** [cited 2017 May 3]; 6(1):130. Available from: http://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-6-130

36. D'Arcy N, Cloutman-Green E, Klein N, Spratt DA. Environmental viral contamination in a pediatric hospital outpatient waiting area: Implications for infection control. Am J Infect Control. **2014**; 42(8):856–860.

37. Cloutman-Green E, Canales M, Pankhurst L, et al. Development and implementation of a cleaning standard algorithm to monitor the efficiency of terminal cleaning in removing adenovirus within a pediatric hematopoietic stem cell transplantation unit. Am J Infect Control. **2015**; 43(9):997–999.

38. Ismail AM, Cui T, Dommaraju K, et al. Genomic analysis of a large set of currently - And historically - Important human adenovirus pathogens. Emerg Microbes Infect [Internet]. Nature Publishing Group; **2018** [cited 2018 Feb 8]; 7(1):10. Available from: http://www.nature.com/articles/s41426-017-0004-y

39. Kroes ACM, Klerk EPA de, Lankester AC, et al. Sequential emergence of multiple adenovirus serotypes after pediatric stem cell transplantation. J Clin Virol. **2007**; 38(4):341–347.

40. Zheng X, Lu X, Erdman DD, et al. Identification of Adenoviruses in Specimens from High-Risk Pediatric Stem Cell Transplant Recipients and Controls. J Clin Microbiol [Internet]. American

Society for Microbiology; **2008** [cited 2016 Dec 12]; 46(1):317–320. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17989198

41. Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. Nat Rev Microbiol [Internet]. Nature Research; **2017** [cited 2017 Jan 25]; 15(3):183–192. Available from: http://www.nature.com/doifinder/10.1038/nrmicro.2016.182

42. Robinson CM, Singh G, Lee JY, et al. Molecular evolution of human adenoviruses. Sci Rep [Internet]. Nature Publishing Group; **2013** [cited 2017 Feb 22]; 3:1812. Available from: http://www.nature.com/articles/srep01812

43. Li X, Kong M, Su X, et al. An outbreak of acute respiratory disease in China caused by human adenovirus type B55 in a physical training facility. Int J Infect Dis. **2014**; 28:117–122.

44. Depledge DP, Gray ER, Kundu S, et al. Evolution of cocirculating varicella-zoster virus genotypes during a chickenpox outbreak in Guinea-Bissau. J Virol [Internet]. American Society for Microbiology; **2014** [cited 2017 Jan 25]; 88(24):13936–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25275123

45. Thomson E, Ip CLC, Badhan A, et al. Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes. J Clin Microbiol [Internet]. American Society for Microbiology; **2016** [cited 2017 Feb 22]; 54(10):2470–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27385709

46. Brown JR, Roy S, Ruis C, et al. Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment: a Robust and Sensitive Method. J Clin Microbiol [Internet]. American Society for Microbiology; **2016** [cited 2017 Sep 10]; 54(10):2530–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27487952

47. Cloutman-Green E, Hartley J. Effectiveness and efficiency of current cleaning protocols on Adenovirus in the environment-5 years data analysis in a UK children's hospital. J Hosp Infect [Internet]. **2010**; 76:S57–S58. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L70386754%5Cnhttp://ucelinks.cdlib.org:8888/sfx_ucsf?sid=EMBASE&issn=01956701&id=doi:&atitle=Effectiveness+and+efficiency+of+current+cleaning+protocols+on+Adenovirus+in+the+environme

48. Görzer I, Guelly C, Trajanoski S, Puchhammer-Stöckl E. Deep Sequencing Reveals Highly Complex Dynamics of Human Cytomegalovirus Genotypes in Transplant Patients over Time. J Virol. **2010**; 84(14):7195–7203.

49. Houldcroft CJ, Depledge DP, Williams R, Cloutman-Green E, Standing JF, Breuer J. Whole-genome sequencing of adenovirus in immunocompromised paediatric patients directly from clinical samples elucidates molecular epidemiology of suspected outbreaks. J Clin Virol. **2016**; 82.

Supplementary Table 1: sample and patient details

Patient	Sample(s)	Sample type	Genotype(s)	Multiple sample types?	Serial blood samples?	ENA accession
1	GOSH_2001_P1-1 GOSH_2003_P1-2	Blood Blood	C2	No	Yes	
2	GOSH_2002_P2-1 GOSH_2073_P2-2	Blood Blood	C1	No	Yes	
3	GOSH_2004_P3-1 GOSH_2006_P3-2	Blood Blood	B3	No	Yes	ERS2168202
5	GOSH_2049_P5-2 GOSH_2051_P5-3 GOSH_2052_P5-4 GOSH_2005_P5 GOSH_2053_P5-5 GOSH_2007_P5 GOSH_2054_P5-6	Blood Blood Urine Blood Urine Blood Blood	C1 and C5	Yes	Yes	ERS2412661 ERS2418440

	GOSH_2055_P5-7 GOSH_2074_P5-8	Blood Urine				
6	GOSH_2008_P6-1 GOSH_2010_P6-2	Blood Blood	E4	No	Yes	ERS2410635 ERS2410636
7	GOSH_2009_P7	Blood	A31	No	No	ERS2419547
8	GOSH_2011_P8-1 GOSH_2025_P8-3 GOSH_2079_P8-4 GOSH_2080_P8-5	Blood Blood Blood Blood	A31	No	Yes	 ERS2419551 ERS2419555 ERS2419556
9	GOSH_2012_P9	Blood	C2	No	No	
10	GOSH_2013_P10-1 GOSH_2014_P10-2	Blood Blood	C5	No	Yes	
11	GOSH_2015_P11	Blood	C2	No	No	
13	GOSH_2043_P13-2 GOSH_2044_P13-3 GOSH_2045_P13-4 GOSH_2046_P13-5	Blood Blood Blood Blood	C5	No	Yes	

	GOSH_2048_P13-6	Blood				
	GOSH_2069_P13-7	Blood				
	GOSH_2070_P13-8	Blood				
	GOSH_2072_P13-9	Blood				
14	GOSH_2034_P14-10	Blood	C2	Yes	Yes	
	GOSH_2035_P14-11	Blood				ERS2413007*
	GOSH_2036_P14-12	Blood				
	GOSH_2037_P14-13	Blood				
	GOSH_2038_P14-14	Ascitic Fluid				
	GOSH_2039_P14-15	Blood				
	GOSH_2040_P14-16	Swab				
	GOSH_2041_P14-17	Blood				
	GOSH_2042_P14-18	Blood				
	GOSH_2047_P14-19	Endo-tracheal aspirate				ERS2413009
	GOSH_2075_P14-20	Blood				
	GOSH_2076_P14-21	Blood				
	GOSH_2077_P14-22	Blood				

	GOSH_2078_P14-23 GOSH_2081_P14-24 GOSH_2082_P14-25 GOSH_2083_P14-26 GOSH_2085_P14-27 GOSH_2086_P14-28 GOSH_2026_P14-2 GOSH_2027_P14-3 GOSH_2028_P14-4 GOSH_2029_P14-5 GOSH_2030_P14-6 GOSH_2031_P14-7 GOSH_2032_P14-9 GOSH_2033_P14-9	Blood Blood Other Urine Blood Blood Urine Blood Blood Blood Blood Eye swab Blood Blood				
15	GOSH_2019_P15	Blood	C1	No	No	
16	GOSH_2019_P16	Blood	C1	No	No	
17	GOSH_2019_P17	Blood	C2	No	No	

18	GOSH_2022_P18-1 GOSH_2024_P18-2	Blood Blood	A31			ERS2419548 ERS2419550
19	GOSH_2023_P19	Blood	A31	No	No	ERS2419549
23	GOSH_2050_P23	Urine	C1	No	No	ERS2412662
24	GOSH_2056_P24	Throat swab	C1	No	No	
25	GOSH_2057_P25-1 GOSH_2059_P25-2	Urine Blood	A31	Yes	No	ERS2419552 ERS2419553
26	GOSH_2058_P26	Urine	C5	No	No	
27	GOSH_2060_P27	Faeces	F41	No	No	ERS2410637
28	GOSH_2061_P28	Urine	F41	No	No	ERS2410638
29	GOSH_2062_P29	Throat swab	A31	No	No	ERS2419554
30	GOSH_2063_P30-1 GOSH_2064_P30-2 GOSH_2065_P30-3 GOSH_2066_P30-4	Blood Blood Blood Blood	C2	No	Yes	
31	GOSH_2067_P31	BAL	C5	No	No	
32	GOSH_2068_P32	Faeces	C1	No	No	ERS2412663

33	GOSH_2071_P33	Blood	C5	No	No	
34	GOSH_2092_P34-1 GOSH_2093_P34-2 GOSH_2095_P34-3	Urine Blood Blood	C2	Yes	Yes	
35	GOSH_2094_P35-1 GOSH_2096_P35-2	Urine Urine	A31	No	No	ERS2419557 ERS2419558
36	GOSH_2097_P36-1 GOSH_2101_P36-2 GOSH_2102_P36-3 GOSH_2103_P36-4 GOSH_2104_P36-5	Urine Blood Blood Blood Blood	C5	Yes	Yes	
37	GOSH_2098_P37-1 GOSH_2099_P37-2	Urine Faeces	C5			
38	GOSH_2100_P38	Faeces	C1	No	No	
39	GOSH_2105_P39	Eye swab	C1	No	No	
40	GOSH_2106_P40-1 GOSH_2109_P40-2	NPA BAL	C1	Yes	No	

41	GOSH_2107_P41	NPA	B3	No	No	ERS2168203
42	GOSH_2108_P42	NPA	C2	No	No	
43	GOSH_2110_P43	NPA	C1	No	No	ERS2412664

*NB Consensus sequence from samples GOSH_2035 and GOSH_3036

Supplementary table 2: Sequencing statistics and sample metadata for each sample.

Sample ID	Original sample type	Diagnostic Pathogen load gc/ml or Ct	Mean read depth	Estimated % coverage of ref at 1x or greater	Genotype
GOSH_2025	Blood (EDTA)	2056900	2071.16	100	A31
GOSH_2079	Blood (EDTA)	120000	687.17	100	A31
GOSH_2080	Blood (EDTA)	582127	1814.77	100	A31
GOSH_2022	Blood (EDTA)	6116	2300.04	100	A31
GOSH_2024	Blood (EDTA)	1000000	2600.11	100	A31
GOSH_2023	Blood (EDTA)	152741	845.41	100	A31
GOSH_2057	Urine	16673	76.26	100	A31
GOSH_2059	Blood (EDTA)	55525	277.33	100	A31
GOSH_2062	Throat swab	20200	207.59	100	A31
GOSH_2094	Urine	30591600	4344.83	100	A31
GOSH_2096	Urine	70000000	5060.19	100	A31
GOSH_2009	Blood (EDTA)	167847	2712	100	A31
GOSH_2011	Blood (EDTA)	268045	1289.98	100	A31
GOSH_2004	Blood (EDTA)	20000000	2000	100	B3
GOSH_2107	Nasopharangeal aspirate	1000000	1882.37	100	B3
GOSH_2006	Blood (EDTA)	1233050	3050.72	100	B3
GOSH_2100	Faeces	90000	0.34	12	C1
GOSH_2019	Blood (EDTA)	4292	5.59	50	C1
GOSH_2106	Nasopharangeal aspirate	90000	6.26	52	C1
GOSH_2073	Blood (EDTA)	93957	17.08	67	C1
GOSH_2056	Throat swab	55129	15.83	71	C1
GOSH_2002	Blood (EDTA)	181734	39.84	73	C1
GOSH_2020	Blood (EDTA)	932130	62.63	78	C1
GOSH_2105	Eye swab	200000	139.14	84	C1
GOSH_2109	Bronchoalveolar lavage	1000000	500.88	89	C1
GOSH_2068	Faeces	Ct 33 ~20000 copies/ml	211.24	90	C1

GOSH_2049	Blood (EDTA)	10682000	1017.06	92	C1
GOSH_2052	Urine	16597100	3358	98	C1
GOSH_2110	Nasopharangeal aspirate	1000000	3673.32	99	C1
GOSH_2050	Urine	20000000	3996.97	100	C1
GOSH_2078	Blood (EDTA)	56325	2.27	37	C2
GOSH_2015	Blood (EDTA)	13699	3.41	45	C2
GOSH_2083	Urine	40844	4.05	45	C2
GOSH_2086	Blood (EDTA)	71985	4.39	48	C2
GOSH_2077	Blood (EDTA)	59267	4.76	49	C2
GOSH_2108	Nasopharangeal aspirate	200000	4.63	51	C2
GOSH_2092	Urine	5710	5.79	57	C2
GOSH_2012	Blood (EDTA)	488595	12.94	64	C2
GOSH_2003	Blood (EDTA)	133306	34.88	70	C2
GOSH_2042	Blood (EDTA)	67425	23.08	70	C2
GOSH_2075	Blood (EDTA)	70523	30.09	72	C2
GOSH_2076	Blood (EDTA)	158917	35.65	72	C2
GOSH_2081	Blood (EDTA)	765653	28.54	72	C2
GOSH_2082	Other	287760	26.51	73	C2
GOSH_2065	Blood (EDTA)	1134740	42.7	75	C2
GOSH_2001	Blood (EDTA)	1264060	147.86	79	C2
GOSH_2064	Blood (EDTA)	773056	56.85	79	C2
GOSH_2095	Blood (EDTA)	85561	59.54	79	C2
GOSH_2093	Blood (EDTA)	308628	105.48	81	C2
GOSH_2063	Blood (EDTA)	699635	134.07	82	C2
GOSH_2021	Blood (EDTA)	829881	361.02	83	C2
GOSH_2041	Blood (EDTA)	577808	144.33	87	C2
GOSH_2027	Blood (EDTA)	6000230	985.7	89	C2
GOSH_2028	Blood (EDTA)	34149600	718.64	89	C2
GOSH_2031	Eye swab	728939	323.81	90	C2
GOSH_2026	Urine	Ct 27	846.94	92	C2

GOSH_2029	Blood (EDTA)	34947600	2089.33	92	C2
GOSH_2047	Endo-tracheal aspirate	Ct 25	1243.7	94	C2
GOSH_2030	Blood (EDTA)	42970500	1966.07	94	C2
GOSH_2039	Blood (EDTA)	1650640	369.46	95	C2
GOSH_2032	Blood (EDTA)	85855600	2962.36	95	C2
GOSH_2040	Swab	Ct 24	2519.26	96	C2
GOSH_2038	Ascitic Fluid	12730700	2047.44	97	C2
GOSH_2037	Blood (EDTA)	20407000	1665.95	99	C2
GOSH_2033	Blood (EDTA)	26560200	2375.71	99	C2
GOSH_2034	Blood (EDTA)	75178500	3766.88	100	C2
GOSH_2035	Blood (EDTA)	5274900	3659.22	100	C2
GOSH_2036	Blood (EDTA)	63197600	2928.18	100	C2
GOSH_2067	Bronchoalveolar lavage	Ct 37 ~2000 copies/ml	0.49	20	C5
GOSH_2071	Blood (EDTA)	6401	0.7	25	C5
GOSH_2053	Urine	68486	0.84	29	C5
GOSH_2066	Blood (EDTA)	1044200	0.79	30	C5
GOSH_2101	Blood (EDTA)	93126	3.97	48	C5
GOSH_2102	Blood (EDTA)	95940	4.68	50	C5
GOSH_2013	Blood (EDTA)	40729	4.29	51	C5
GOSH_2058	Urine	97322	4.42	54	C5
GOSH_2099	Faeces	500000	4.85	55	C5
GOSH_2043	Blood (EDTA)	55406	5.52	56	C5
GOSH_2048	Blood (EDTA)	51527	12.73	61	C5
GOSH_2091	Blood (EDTA)	201184	21.91	66	C5
GOSH_2090	Blood (EDTA)	303814	23.14	67	C5
GOSH_2014	Blood (EDTA)	103556	25.2	69	C5
GOSH_2070	Blood (EDTA)	441445	26.47	69	C5
GOSH_2103	Blood (EDTA)	215793	28.91	69	C5
GOSH_2098	Urine	21640	9.05	69	C5
GOSH_2104	Blood (EDTA)	92183	25.57	70	C5

GOSH_2084	Blood (EDTA)	582874	30.39	72	C5
GOSH_2044	Blood (EDTA)	148412	23	72	C5
GOSH_2069	Blood (EDTA)	336548	29.65	73	C5
GOSH_2072	Blood (EDTA)	509127	37.45	74	C5
GOSH_2087	Blood (EDTA)	428489	76.09	78	C5
GOSH_2045	Blood (EDTA)	159334	103.85	79	C5
GOSH_2046	Blood (EDTA)	175152	97.97	81	C5
GOSH_2088	Blood (EDTA)	1727720	146.53	82	C5
GOSH_2089	Blood (EDTA)	489344	126.06	82	C5
GOSH_2051	Blood (EDTA)	20000000	2809.24	97	C5
GOSH_2074	Urine	20000000	2910.69	97	C5
GOSH_2005	Blood (EDTA)	1.79E+08	3291.33	98	C5
GOSH_2007	Blood (EDTA)	20000000	2479.69	99	C5
GOSH_2054	Blood (EDTA)	20000000	5589.31	100	C5
GOSH_2055	Blood (EDTA)	20000000	7632.37	100	C5
GOSH_2016	*Culture - cell lysate*	*1000000*	*2000*	*100*	*D9*
GOSH_2017	*Culture - cell lysate*	*1000000*	*1500*	*100*	*E4*
GOSH_2008	Blood (EDTA)	4010170	26.31	100	E4
GOSH_2010	Blood (EDTA)	20000000	1302.48	100	E4
GOSH_2018	*Culture - cell lysate*	*1000000*	*4000*	*100*	*F40*
GOSH_2060	Faeces	Ct 21	857.84	100	F41
GOSH_2061	Urine	613713	1092	100	F41
GOSH_2085	Blood (EDTA)	36933			
GOSH_2097	Urine	1279			

Supplementary table 3: Minimum virus load per ml of sample for success (>80% genome at 1x).

Genotype	**Virus load per ml**
A31	6116 (6.1 x 10e3)

B3*	1,000,000 (1 x 10e6)
C1	20,000 (2 x 10e4)
C2	308,628 (3.1 x 10e5)
C5	175,152 (1.8 x 10e5)
E4†	4,010,170 (4 x 10e6)
F41‡	613,713 (6.1 x 10e5)

*N = 3; † N = 2; ‡ N = 2

Figure 1: Percentage of the genome covered relative to calculated genome copies (log10) in each sequencing reaction. There is a relationship between log10 genome copies input and genome coverage for adenovirus species C (genotypes 1, 2 and 5), although not other genotypes.

Figure 2: Pie chart showing the percentage of samples sequenced from each genotype (N=105). The most common genotype sequenced was C2, largely reflecting the long duration of high-level viraemia (>50,000 copies/ml whole blood) in a single patient (P14). Patients 5 and 13 comprise the majority of C5-positive samples sequenced.

Figure 3: estimated genome coverage of adenovirus genotypes, comparing blood (top row) with other clinical samples such as urine, faeces and swabs (bottom row). The percentage of the genome recovered at 1x varied with genotype rather than sample type.

Figure 4: Phylogenies of adenovirus types sequenced as part of this study. Sequences were aligned using MAFFT, manually edited in MEGA7 if necessary, and ML trees were created using RAXML Black Box, 500 bootstraps. Branch support values higher than 70 are shown. Sequences were labelled as clusters if a branch contained at least two GOSH samples from at least two patients, and had bootstrap support greater than 90. **A** Phylogeny of adenovirus A31. Three of the eight A31 cases (patients P19, P29 and P35) can be excluded from possible patient-to-patient transmission phylogenetically. Nine A31 samples from four patients (P7, P8, P18 and P25 in cluster *1) cluster together phylogenetically and have low pairwise diversity to other another, but are separated temporally (P18 samples collected in 2011). **B** Phylogeny of genotype B3 sequenced from patient samples, which cluster separately and are thought to be unrelated. **C** Phylogeny of adenoviruses of genotypes C1, C2 and C5 from the same hospital. Sequences from patient 5 (highlighted in green) do not cluster monophyletically. Two patients (23 and 43) infected with adenovirus C1, cluster together (*2) and were shown by ward movement analysis to have had contact within the hospital. **D** Phylogeny of E4 sequences from a single patient, which are more similar to wild-type E4 sequences taken from GenBank than vaccine-strain genomes. **E** Phylogeny of F41 sequences from the hospital and GenBank.

Figure 5: Timeline of adenovirus A31 positive inpatient stays on a paediatric bone marrow transplant ward. All patients were confirmed as A31 positive by hexon molecular typing and/or WGS of at least one sample. Between 2011 and 2016, hexon sequencing was carried out as part of infection prevention and control measures for a suspected nosocomial adenovirus outbreak. Patients for whom only hexon molecular type data was available are labelled eg Hex70. Samples marked in red denote the period during which adenovirus was detected in stool or blood culture for each patient; samples marked in blue indicate samples in which adenovirus was not detected in stool or blood culture. Samples marked in green were available for whole-genome sequencing.

Figure 6A: pairwise differences between sequences of samples from GOSH and GenBank, divided by species. Bars marked in blue show pairwise differences between samples from the same patient. Red bars show pairwise differences between samples from different patients at GOSH, or between GOSH sequences and sequences available in GenBank. Orange bars show the number of differences between sequences of two closely related genotypes within the same species: the differences between A31 and A61, B3 and B7, C1 and C2, and F40 and F41. Bars in purple show the within-patient pairwise differences of sequences from patient 5, which are hypothesised to be a mixed genotype infection. Orange bars show the pairwise differences between samples from different patients in clusters *1 (A31) and *2 (C1) respectively (figure 4). For genotype E4, the between patient (including GenBank) difference reflects the change in circulating E4 sequence diversity since the vaccine strain was isolated from a 1950s field strain.

Figure 6B: Zoomed in view of data from 6A, showing pairwise differences up to 100 nucleotides between sequences of samples from GOSH and GenBank from species A and C, including samples from cluster *1 (A31) and putative cluster *2 (C1). The number of within-patient pairwise differences and between-patient pairwise differences for phylogenetically clustered samples (green bars) overlaps for patients infected with genotype A31 (cluster *1). However, for patients infected with genotype C1, the within-patient pairwise differences (<5 differences) are considerably fewer than the putatively clustered C1 sequences (cluster *2), which differ by 33 nucleotides (green bars).

Figure 7: Plot showing nucleotide diversity in samples from patient 5. Black dots show nucleotide diversity of all adenovirus reads within a sample. Blue dots show nucleotide diversity of reads mapped stringently to genotype C1. Red dots show nucleotide diversity within reads mapped stringently to genotype C5. Samples from other patients with single genotype infections, sequenced to comparable depth show much lower total-adenovirus reads nucleotide diversity than patient 5. When reads from patient 5 are mapped stringently to either adenovirus genotype C1 or genotype C5, the nucleotide diversity is more comparable to that seen in other patients with single genotype infections (A31, B3, C1 and C2, right of graph). This suggests that high nucleotide diversity within patient 5 genomes is accounted for by the presence of a mixed-genotype C1 and C5 co-infection.

Supplementary figure 1: Percentage of reads which are on-target relative to calculated genome copies used as input for each sequencing reaction. There is a positive relationship between the number of input genome copies and the percentage of reads which are map to adenovirus (on target read percentage; OTR).

Supplementary figure 2: Adenovirus whole-genome phylogeny showing species A, B, C, E and F, sequenced from 36 clinical samples and 18 representative strains from GenBank. Sequences were aligned using MAFFT and manually edited in MEGA7 if necessary. ML trees were created using RAXML Black Box, 500 bootstraps. Branch support values higher than 80 are shown.

Supplementary figure 3: Total adenovirus reads from patient 5 samples were mapped stringently (95% read length mapping and 95% sequence identity) to reference genomes for genotypes C1 (JX173085) and C5 (KF268199), reducing cross-mapping between genomes. Consensus genomes were then generated from the reads mapping to either genotype C1 or C5, aligned using MAFFT, and a whole-genome phylogeny created using RaXML. From eight samples, separate C1 and C5 consensus sequences could be generated.

Supplementary figure 4. The distribution of minority variants, by read proportion, in patient 5 over time. Total adenovirus reads were mapped to the C5 reference sequence, minority variants were called using CLC Genomics Workbench, and total numbers of variants per sample were binned by their frequency (up to 50%). On day 2 post-detection of adenovirus, the first available sample, minority variants are found predominantly at frequencies of less than 10%. On day 9, minority variants were both more numerous and found as a higher proportion of the sequence reads, the 30-50% frequency range, suggesting two adenovirus populations are present. The adenovirus extracted from day 11 samples from whole blood and urine contain a variable number of minority variants in the 40-50% (urine) and 10-30% (whole blood) range. This

mixed population is still detectable in whole blood at frequencies of below 20% on day 22 and is undetectable in blood by day 25. This reflects the evolution of the consensus sequence over time from a C5 to C1-like sequence, followed by version to a C5 consensus sequence (supplementary figure 3).

Supplementary methods

Virus load determination

Adenoviral loads were quantified as part of routine monitoring using an in-house diagnostic qPCR assay at Great Ormond Street Hospital, UK, previously described [1]. The standard curve for quantitation was constructed using a 10-fold dilution series of plasmid containing the target sequence.

DNA extraction

Samples selected for WGS were stored at -80C until DNA was extracted using the Qiagen EZ1 extraction system with the Virus Mini Kit, from 200ul (EDTA whole blood or clarified faeces) or 400ul (other sample types) starting material, with a 90ul elution volume. Samples which previously underwent hexon sequence typing were blood, faeces or nasopharyngeal aspirates tested as part of an ongoing infection prevention and control (IPC) investigation into suspected hospital acquired adenovirus infection.

Hexon typing

PCR detection of hyper variable region 7 (HVR-7) of the adenovirus hexon gene was carried out using published primers [2]. The reaction mixture was 1x Bioline Buffer (Bioline, London, UK), 1.5mM MgCl2 (Bioline, London, UK), 0.25µM of HVR-7 Forward, 0.25µM HVR-7 Reverse, 1mM dNTPs (Bioline, London, UK), 2.5 units of Bioline Taq (Bioline, London, UK) and 10µl DNA extract and molecular grade water to give a final volume of

50µl. Cycling conditions were as follows: 95˚C for 10 minutes, followed by 40 cycles of denaturation at 95˚C for 60 seconds, annealing at 51˚C for 60 seconds and extension at 72°C for 60 seconds, with a final elongation step at 72°C for 3 minutes.

1. Brown JR, Shah D, Breuer J. Viral gastrointestinal infections and norovirus genotypes in a paediatric UK hospital, 2014–2015. J Clin Virol. **2016**; 84:1–6.

2. Sarantis H, Johnson G, Brown M, Petric M, Tellier R. Comprehensive detection and serotyping of human adenoviruses by PCR and sequencing. J Clin Microbiol [Internet]. American Society for Microbiology; **2004** [cited 2018 Jan 12]; 42(9):3963–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15364976