Norovirus transmission dynamics in a paediatric hospital using full genome sequences

Julianne R Brown¹, Sunando Roy², Divya Shah¹, Charlotte A Williams², Rachel Williams², Helen Dunn¹, John Hartley³, Kathryn Harris¹ and Judy Breuer¹,²

¹ Microbiology, Virology and Infection Prevention and Control, Great Ormond Street Hospital NHS Foundation Trust, UK
² Infection and Immunity, University College London, UK

Corresponding author: Dr Julianne R Brown; Virology, Great Ormond Street Hospital NHS Foundation Trust, Great Ormond Street, London WC1N 3JH; julianne.brown@nhs.net

Summary: Norovirus genome sequencing identified 33% of patients whose sequences were linked phylogenetically to those of another patient in the study. Another 24% nosocomially infected patients had unlinked sequences suggesting infection from unsampled sources. Genome sequencing identifies unsuspected nosocomial norovirus infection.

© The Author(s) 2018. Published by Oxford University Press for the Infectious Diseases Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Abstract

Background

Norovirus is a leading cause of worldwide and nosocomial gastroenteritis. This study aimed to assess the utility of molecular epidemiology using full genome sequences, compared to routine Infection Prevention and Control (IPC) investigations.

Norovirus genomes were generated from new episodes of norovirus at a pediatric tertiary referral hospital over 19 months (n=182). Phylogeny identified clusters of related sequences which were verified using epidemiological and clinical data.

Results

Twenty four clusters of related norovirus sequences (“sequence clusters”) were observed, including eight previously identified by IPC investigations (“IPC outbreaks”). Seventeen sequence clusters (involving 77/182 patients) were corroborated by epidemiological data (“epidemiologically supported clusters”), suggesting transmission between patients. Linked infections were identified among 44 patients who were missed by IPC investigations.

33% of norovirus sequences were linked suggesting nosocomial transmission. 24% of patients had nosocomial infections from an unknown source. 43% were norovirus positive on admission.

Conclusions

We show that there are frequent introductions of multiple norovirus strains with extensive onward nosocomial transmission of norovirus in a paediatric hospital with a high proportion of immunosuppressed patients nursed in isolation. Phylogenetic analysis using full genome sequences is more sensitive than classical IPC investigations for identifying linked cases and should be considered when investigating norovirus nosocomial -transmission. Sampling of staff, visitors and
the environment may be required for complete understanding of the sources of infection and transmission routes in patients with nosocomial infections that are not linked to other patients and among patients with phylogenetically linked cases but no evidence of direct contact.

**Key words:** Norovirus; epidemiology; molecular epidemiology; sequencing; whole genome
Background

Norovirus is a leading cause of gastroenteritis worldwide, associated with large outbreaks in healthcare facilities with substantial clinical and economic implications[1-4]; the global financial burden is estimated at US$60.3 billion annually[2]. Norovirus infections are typically self-limiting with vomiting and diarrhoea lasting 1-2 days, however in immunocompromised patients chronic infections can develop lasting weeks to years with considerable associated morbidity such as severe weight loss and malnutrition [5, 6].

Norovirus infections are primarily caused by genogroups GI and GII, each categorised into 9 and 22 genotypes respectively (GI.1–GI.9 and GII.1–GII.22). Due to recombination between genotypes at the ORF1/ORF2 junction, norovirus has a dual typing system based on the polymerase (ORF1) and capsid (ORF2) sequences [7].

Partial genome sequencing of the hypervariable region of the capsid (P2) has previously identified nosocomial norovirus outbreaks in immunocompetent patients [8]. P2 sequences in transmission events are often identical [8, 9], with a 10% probability of 1–2 nucleotide changes in samples collected three weeks post-infection[10]. In immunocompromised patients, however, transmission may occur after in vivo evolution; several single nucleotide polymorphisms (SNPs) are observed in the capsid sequence between linked patients [11]. When the infected population includes a high proportion of immunocompromised patients, as is the case at the tertiary referral children’s hospital in which our study was conducted, whole genome sequencing (WGS) may provide increased resolution to identify routes of transmission [12].

In this study we evaluated whether WGS can be used to better understand the sources of norovirus infection and transmission dynamics in a paediatric population with a high prevalence of immunocompromised patients.
Methods

Study population and stool samples

The study was carried out in a pediatric tertiary referral hospital with 350 beds, 60% of which are in single isolation rooms. There is no accident and emergency department, therefore acute gastroenteritis is not the primary reason for admission unless it occurs in a patient already under the hospital’s care.

Stools from all symptomatic (diarrhoea and/or vomiting) children (inpatient or outpatient) are tested for gastrointestinal viruses by real-time polymerase chain reaction (RT-PCR), the methods for which are described elsewhere [13]. Norovirus positive PCR results are reported as either norovirus GI or GII. Enhanced surveillance, with screening on admission and weekly whilst in-patients, is performed for all children admitted for haematopoietic stem cell transplant or congenital immunodeficiencies (symptomatic or asymptomatic). Details of the usual management of patients who are symptomatic on admission and any patients found to be norovirus positive by PCR are given (Supplementary Figure 1 and Supplementary Methods).

Norovirus infections detected less than 48 hours after admission to hospital are considered positive on admission; those detected more than two days after admission are considered to have a nosocomial infection. Since the study hospital is a tertiary referral hospital, many patients have previously been admitted to local hospitals or had several outpatient visits prior to admission; earlier acquisition of infection in this or another healthcare facility cannot be excluded.

A nosocomial outbreak (“IPC outbreak”) is suspected when two or more cases of gastroenteritis or confirmed positive norovirus cases occur within 48-72 hours in patients, staff or visitors on the same ward or when an nosocomial infection occurs on the same ward as a chronically infected patient.
An outbreak meeting is called and standardised IPC outbreak measures are implemented (Supplementary Figure 1 and Methods). The index is presumed to be the person in whom norovirus was first detected, unless a point source is suspected.

In this study residual specimen from the first positive sample from all norovirus positive patients between 1st July 2014 and 17th February 2016 (19 months) was submitted for whole genome sequencing. A total of 205 norovirus PCR positive patients were identified during the study period. Ten had no residual specimens and six were PCR negative on re-extraction. The remaining 189 samples were whole genome sequenced. The median patient age was 2 years (range 1 month–16 years); 59% of patients were profoundly immunocompromised with primary immunodeficiency syndromes, solid or bone marrow transplants or receiving chemotherapy for malignancies.

**Norovirus whole genome sequencing**

RNA purification, cDNA synthesis and full genome sequencing using SureSelect target enrichment followed by de novo assembly was carried out as described previously[14]. A consensus sequence was generated. To verify correct genome assembly, open reading frames (ORF-1, ORF-2 and ORF-3) were identified for all sequences using the Find Open Reading Frames tool in CLC Genomics Workbench (v 9.0) [14].

**Genotyping and phylogenetic analysis**

In total, 184/189 samples generated greater than 90% genome coverage and >100-fold read depth; full genome consensus sequences were submitted to the Norovirus Genotyping Tool (http://www.rivm.nl/mpf/norovirus/typingtool)[15] to determine the genotype.

Two sequences were excluded from phylogenetic analysis as infection with a mixture of genotypes was detected therefore a robust consensus sequence could not be generated [14]. Phylogenies were
reconstructed in CLC Genomics Workbench (v 9.0) from the remaining 182 consensus sequences, as described in Supplementary Methods.

Data-mining to establish epidemiological support of sequence clusters

Patients whose norovirus genomes branched together in a phylogenetic tree from a single common ancestral node (monophyletic), with absent or short branch lengths and a high probability that the branching order was correct (high bootstrap value) were referred to as “sequence clusters”. To determine whether these could be due to nosocomial transmission, inpatient and outpatient records together with norovirus PCR laboratory result histories for each patient were retrieved. Patients within a sequence cluster were considered to be epidemiologically linked if they were norovirus positive with an inpatient or outpatient visit that overlapped by at least 24 hours; these are referred to as “epidemiologically supported clusters”. Sequence clusters that were not supported by epidemiological evidence are referred to as “epidemiologically unsupported clusters”.

Results

Norovirus infections and genotypes

During a 19 month period (1st July 2014 – 1st February 2016) we generated full genome sequences for 182 new norovirus episodes in a paediatric tertiary referral hospital with a high prevalence of immunocompromised patients; of these 84 were, based on routine Infection Prevention and Control (IPC) investigations, considered to be nosocomial infections. During this period eight IPC outbreaks of norovirus were identified (Table 1), which accounted for only 37/84 of the nosocomial infections. To confirm whether the declared outbreaks were truly linked and to better understand the sources of infection for the remaining episodes, we reconstructed phylogenetic trees from all 182 sequenced norovirus genomes. Phylogenetic trees reconstructed from GII.4 partial capsid sequences (the hypervariable P2 region used in routine typing methods), provided insufficient information to
identify two of the ten sequence clusters obtained using full genomes and missed 14 of the 37 (38%) patients linked by full genome sequences (Supplementary Figure 2). Moreover, whilst phylogenetic analysis using full genome sequences is well supported, with 77% (55/71) of internal nodes in the whole genome tree supported by bootstrap values ≥70, maximum likelihood phylogeny using the hyper-variable capsid P2 domain sequences (427 nt) generated a tree with low bootstrap support; only 34% (24/71) of internal nodes were supported by bootstrap values ≥70 (Supplementary Figure 3). Separate phyllogenies reconstructed for ORF1 and ORF2 sequences showed no difference in branching order suggesting recombination at the ORF1/ORF2 junction, the most frequently described breakpoint in norovirus recombination events[16, 17], did not occur within the sampled population. Therefore all further analysis was performed using full genomes.

A total of 11 capsid and 14 polymerase genotypes were identified, in 17 unique combinations (Figure 1, Supplementary Figure 4).

**IPC outbreaks and sequence clusters**

All (8/8) IPC outbreaks previously identified corresponded to monophyletic sequence clusters identified by phylogenetic analysis (Supplementary Figure 5). An additional 23 patients, for whom no source of infection had previously been identified, were linked by phylogenetic analysis to six of the IPC outbreaks (Table 1). Eight of the 23 patients were norovirus positive on admission but had had previous inpatient stays or outpatient appointments during which transmission could have occurred. Four patients previously identified as part of three of the outbreaks were shown to have sporadic infections with unrelated genotypes. All four patients had nosocomial infections of unknown origin.

Phylogenetic analysis identified 16 new sequence clusters involving 36 patients that had not previously been identified by IPC investigations, including six patients who were norovirus positive on admission but had had previous inpatient stays or outpatient appointments during which transmission could have occurred. In total, 24 sequence clusters were identified by phylogenetic
analysis comprising 2–17 sequences (median 2) per cluster (Table 1, Supplementary Figure 5, Supplementary figure 6, Supplementary Methods).

_Epidemiological support for sequence clusters_

Review of inpatient and outpatient histories together with the timing of norovirus shedding identified plausible links between patients within 17 of the 24 sequence clusters (including the eight IPC outbreaks originally identified by IPC investigations) (Supplementary Figures 6 and 7). These epidemiologically supported clusters confirm nosocomial transmission among 77/92 (84%) of the patients linked in sequence clusters, of whom only 33 had been previously identified by IPC methods (Table 2). The epidemiologically supported clusters were caused by genotypes GII.P21_GII.3 (7/17 clusters), GII.Pe_GII.4 (5/17), GII.P4_GII.4 (4/17) and by GII.P7_GII.6 (1/17). Of the 44 patients in epidemiologically supported clusters who were missed by IPC investigations, 11/44 were on the same ward but involved only two patients each, 7/44 were on a different ward with a shared clinical team and 21/44 occurred over a prolonged period of time.

For the remaining seven (of 24) sequence clusters, none of which were identified by IPC investigations, no evidence of an epidemiological link could be found between patients in each cluster; including when the residential postcode of patients were examined to assess whether community transmission could have occurred. These epidemiologically unsupported clusters were caused by GII.P21_GII.3 (3/7), GII.Pe_GII.4 (1/7), GII.P3_GII.3 (1/7), GII.P2_GII.2 (1/7) and GII.7_GII.6 (1/7).
Sources of infection

In total 103/182 patients had nosocomial infections, of whom 84 were identified based on routine Infection Prevention and Control (IPC) investigations. No source could be identified for 43/84 patients (table 3), including 9 whose infection occurred at the beginning of the study and who therefore may have been infected by a patient not included in the study. The remaining 19/103 patients with nosocomial infection, were originally classified as norovirus positive on admission (POA). However, all were linked by phylogenetic analysis to other patients in epidemiologically supported clusters; they were all subsequently found to have attended the hospital in the recent past. Seventy-nine patients (79/182) were norovirus positive on admission (POA) (table 3) and were not linked to any other sequences in the phylogenetic analysis. This included four patients who were index cases in subsequent onward transmissions.

Pairwise distances between genome sequences (see Supplementary Methods)

The pairwise distances (Supplementary Methods) between GII.4 norovirus genomes from each patient demonstrated a distinct population of sequences with ≤38 SNPs difference (Figure 2a), which corresponds to >99.5% sequence identity across the genome and falls within the range of previously described within-host SNP diversity seen in longitudinally sampled chronically infected patients [18] (Figure 2b). This equates to the cut-off of ≤2 SNPs (99.5% identity) value where the partial capsid P2 domain has been used to identify transmission events among norovirus GII.4 infections in immunocompetent patients [10, 11]. Consequently in this study sequence clusters were identified if groups of full genome sequences formed a monophyletic cluster with fewer than 38 pairwise SNPs between them. As controls, we included pairwise distances between sequences collected from repeatedly sampled chronically-infected individuals [18].

The mean pairwise distance for 62 GII.3 and 71 GII.4 sequences comprising the majority of the epidemiologically supported clusters was 7 SNPs for the former and 12 SNPs for the latter (range 0–
22 and 0–35, respectively). The mean pairwise distance in the GII.3 and GII.4 epidemiologically unsupported clusters was 25 SNPs for the former (range 17–36) and 12 SNPs for the only GII.4 epidemiologically unsupported cluster (Figure 2b). Thus pairwise differences were significantly higher for epidemiologically unsupported as compared with epidemiological supported GII.3 clusters (P<0.001, mean difference 18 SNPs, 95% CI of the difference 13–23 SNPs; 2-tailed T Test, SPSS v24).

The results were skewed by the epidemiologically supported cluster, GII.P21_GII.3 Cluster 10 (Supplementary Figure 6). Sequence cluster 10 was monophyletic with strong epidemiological support but higher pairwise distances (19–149 SNPs). However, all patients were located on the same or a linked ward, managed by the same clinical team and with overlapping admissions and norovirus shedding (Supplementary Figure 6c). The sequence cluster could not be disrupted by inclusion of all publically available GII.P21_GII.3 sequences in the phylogenetic analysis (Supplementary Figure 8), making it unlikely that this sequence cluster represents repeated introductions to the hospital from an external source. We have previously shown a linear relationship between the number of SNPs and duration of infection in chronically infected immunosuppressed patients [18]. The diversity within epidemiologically supported Cluster 10 is therefore likely to have arisen due to virus evolution in this patient group, the majority (8/9) whom were immunosuppressed and chronically infected with extended periods (38–388 days) between transmission events.

Conclusions

Information on the routes of norovirus transmission in a nosocomial setting is necessary for allocation of IPC resources and effective containment of infection. While capsid P2 sequencing is currently the standard method for norovirus molecular epidemiology, we show that it is not sufficiently discriminating for robust investigation of putative transmissions among immunosuppressed patients. Here we demonstrate that WGS is superior to both P2 and IPC
methods for ascertaining transmission among immunosuppressed patients identifying 44 patients not previously known from IPC investigations to be part of transmission chains.

For 60/103 nosocomial infections phylogenetic analysis showed linkage to another patient in the study. However for the remainder, including the index patients in 13/17 epidemiologically supported sequence clusters, the sources were unknown and potentially included unsampled staff, visitors or patients. Within epidemiologically supported clusters, 75% of patients were profoundly immunocompromised and managed in single isolation rooms with limited or no direct contact between patients making the route of transmission unclear. Additional sampling could help here and may also shed light on the 7/24 epidemiologically unsupported clusters which accounted for 14 of the 93 (15%) patients identified as linked by phylogenetic analysis. Of interest, for non-GII.4 norovirus genotypes which do not spread as pandemics, Parra et al [19] observed a “static” pattern of diversification, with only a few residue changes over several decades. Thus for the unsupported non-GII.4 clusters it is possible that, due to a lack of genome variability, independent episodes of norovirus infection appear linked based on phylogenetic analysis alone; this needs further investigation.

We further attempted to verify sequence clusters by calculating pairwise SNPs for sequences within a monophyletic sequence cluster. While epidemiologically supported GII.4 clusters all had <38 SNPs between sequences and most epidemiologically supported GII.3 clusters had <22 SNPs, the epidemiologically supported GII.3 Cluster 10 fell outside this range probably because of within-host evolution prior to transmission among chronically infected patients. This should be noted when using genomics to manage outbreaks (Supplementary Figure 9).

Norovirus is now the major cause of acute gastroenteritis worldwide[2] with morbidity and mortality seen in infants in low income countries and the elderly and immunosuppressed in middle and high income countries[6]. IPC is critical to managing disease, particularly in hospitals and other
institutions. However for immunosuppressed patients in particular, the sensitivity of routine IPC investigations alone for identifying linked transmission is 44% as compared with IPC plus WGS. (Table 2). While 33% of new norovirus cases in this study were acquired from another patient, despite isolation nursing and stringent IPC measures, the source of infection for 43% of nosocomial infections remains unknown even with WGS, pointing to the need for wider sampling of patients, staff, visitors and the environment. Nonetheless WGS could be a valuable tool with which to focus IPC interventions in areas of the hospital where nosocomial acquired infections are most problematic. With ever decreasing sequencing costs and technologies that allow rapid turnaround times [14] the possibility that norovirus genome sequencing, perhaps linked to electronic patient records, could be used routinely to control nosocomial infections is now a reality.
Acknowledgments

We acknowledge the work and support of the Great Ormond Street Hospital for Children NHS Foundation Trust Infection Prevention and Control team. We acknowledge the infrastructure support from the UCL Pathogen Genomics Unit (PGU), the NIHR UCL/UCLH BRC, the UCL MRC CMMV. All research at Great Ormond Street Hospital NHS Foundation Trust is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre.

The study was approved by the NRES Committee London - Brent (REC reference 14/LO/1331).

Funding

This work was supported by the PATHSEEK European Union’s Seventh Programme for research, technological development and demonstration (grant number 304875) and a National Institute for Health Research (NIHR) doctoral fellowship (grant number NIHR-HCS-D12-03-15) to JRB. JBreuer receives funding from the NIHR UCL/UCLH Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflicts of Interest

The authors declare no conflicts of interest.
References


Figure legends

Figure 1 Maximum likelihood phylogeny of full genome sequences from norovirus episodes over a 19 month period (2014–2016) a) colour coded by genotype and b) colour coded by sequence cluster number. Separate maximum likelihood phylogenies for each genotype, with sequence cluster; b) number annotated and displaying greater resolution, are shown in Supplementary Figure 5.

Figure 2 Pairwise distances in GII.3 and GII.4 epidemiologically supported clusters, epidemiologically unsupported clusters and longitudinally sampled chronically infected patients. Epidemiologically unsupported clusters are those identified by phylogenetic analysis but not supported by classical epidemiological evidence.

a) Pairwise differences between local and database WGS

b) Pairwise differences plotted by the category of cluster (epidemiologically supported or unsupported and longitudinally sampled individuals) into which they fall. The greater pairwise differences within cluster 10 are shown.
### Table 1. Sequence clusters identified by maximum likelihood phylogeny using full genome sequences. All GII.4 capsid genotypes are Sydney_2012. GII.4 polymerase genotypes are GII.Pe Sydney_2012 or GII.P4 New Orleans_2009

<table>
<thead>
<tr>
<th>Sequence cluster number</th>
<th>Genotype</th>
<th>Number of patients</th>
<th>Date range</th>
<th>Number of wards</th>
<th>Number of clinical specialties involved</th>
<th>Bootstrap support</th>
<th>Diversity within cluster †</th>
<th>Identified by infection control (IPC) investigations</th>
<th>Supported by classical epidemiology*</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>GII.P7_GII.6</td>
<td>3</td>
<td>7 days</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>GII.P21_GII.3</td>
<td>17</td>
<td>3 months</td>
<td>6</td>
<td>3</td>
<td>100</td>
<td>0–22</td>
<td>Partially</td>
<td>Yes (16/17)</td>
</tr>
<tr>
<td>6</td>
<td>GII.P21_GII.3</td>
<td>2</td>
<td>3 days</td>
<td>1</td>
<td>1</td>
<td>82</td>
<td>14</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>GII.P21_GII.3</td>
<td>6</td>
<td>1 month</td>
<td>3</td>
<td>2</td>
<td>70</td>
<td>0–10</td>
<td>Partially</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>GII.P21_GII.3</td>
<td>2</td>
<td>2 months</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>11</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>GII.P21_GII.3</td>
<td>2</td>
<td>2 days</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>12</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>GII.P21_GII.3</td>
<td>9</td>
<td>17 months</td>
<td>2</td>
<td>1</td>
<td>100</td>
<td>19–149</td>
<td>Partially</td>
<td>Yes</td>
</tr>
<tr>
<td>23</td>
<td>GII.P21_GII.3</td>
<td>2</td>
<td>3 months</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>29</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>GII.Pe_GII.4</td>
<td>8**</td>
<td>2 months</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>1–24</td>
<td>Partially</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>GII.Pe_GII.4</td>
<td>2</td>
<td>6 days</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>3</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>GII.Pe_GII.4</td>
<td>2</td>
<td>3 days</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>14</td>
<td>GII.Pe_GII.4</td>
<td>4</td>
<td>11 days</td>
<td>2</td>
<td>1</td>
<td>100</td>
<td>1–4</td>
<td>Partially</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>GII.Pe_GII.4</td>
<td>3</td>
<td>3 days</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>1–3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>16</td>
<td>GII.P4_GII.4</td>
<td>7</td>
<td>3 months</td>
<td>2</td>
<td>1</td>
<td>100</td>
<td>0–35</td>
<td>Partially</td>
<td>Yes</td>
</tr>
<tr>
<td>17</td>
<td>GII.P4_GII.4</td>
<td>2</td>
<td>25 days</td>
<td>2</td>
<td>1</td>
<td>100</td>
<td>14</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>GII.P4_GII.4</td>
<td>5</td>
<td>2.5 months</td>
<td>3</td>
<td>2</td>
<td>77</td>
<td>0–25</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>19</td>
<td>GII.P4_GII.4</td>
<td>2</td>
<td>19 days</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>6</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>GII.P4_GII.4</td>
<td>2</td>
<td>8 months</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>31</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>21</td>
<td>GII.P2_GII.2</td>
<td>2</td>
<td>2 months</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>7</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>22</td>
<td>GII.P7_GII.6</td>
<td>2</td>
<td>3 months</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>17</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>23</td>
<td>GII.P2_GII.2</td>
<td>2</td>
<td>5.5 months</td>
<td>2</td>
<td>2</td>
<td>100</td>
<td>12</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>24</td>
<td>GII.P21_GII.3</td>
<td>2</td>
<td>3 months</td>
<td>2</td>
<td>1</td>
<td>100</td>
<td>18</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
† expressed as the number of nucleotide differences across whole genome; * overlap in norovirus positive period and hospital attendance; ** including one parent (NORO/51, father of NORO/52)
Table 2 Comparison of norovirus transmission events identified by phylogenetic analysis (epidemiologically supported clusters) and classical Infection Prevention and Control (IPC) investigations (IPC outbreak). Non-underlined text indicates patients correctly assigned by IPC investigations; underlined indicated patients incorrectly assigned by IPC investigations.

<table>
<thead>
<tr>
<th>Part of IPC outbreak</th>
<th>Not part of IPC outbreak</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission inferred by molecular epidemiology</td>
<td>33</td>
<td>44**</td>
</tr>
<tr>
<td>No transmission inferred by molecular epidemiology</td>
<td>3*</td>
<td>102</td>
</tr>
<tr>
<td>TOTAL</td>
<td>36</td>
<td>146</td>
</tr>
</tbody>
</table>

* shown to be a different genotype to the rest of their respective IPC outbreak; **including one patient who was incorrectly assigned to an outbreak by IPC investigations (Cluster 11), but shown by molecular epidemiology to be a different genotype to the rest of the outbreak and linked to another patient in Cluster 23.
Table 3. Summary of sources of infection at GOSH during study period, July 2014–February 2016. Epidemiologically supported clusters are sequence clusters identified by phylogenetic analysis and supported by epidemiological evidence.

<table>
<thead>
<tr>
<th>Source of infection</th>
<th>Number of patients, n=182 (%)</th>
<th>Proportion of patients immunocompromised (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Another patient (part of an epidemiologically supported cluster)</td>
<td>60 (33%)</td>
<td>45/60 (75%)</td>
</tr>
<tr>
<td>Outside of the hospital (norovirus positive on admission)</td>
<td>79 (43%)</td>
<td>37/79 (49%)</td>
</tr>
<tr>
<td>Unknown (nosocomial infection but not transmission from another patient in the study*)</td>
<td>43 (24%)</td>
<td>27/43 (63%)</td>
</tr>
</tbody>
</table>

*not part of a monophyletic sequence cluster (with <38 SNPs within cluster) therefore not linked to any other virus sequences within the study cohort
Figure 2.