Figure A1. Sequencing outcome proportions for each norovirus genotype. Pass, >90% genome coverage and >100-fold read depth; Sub-optimal, >90% genome coverage or >100-fold read depth; Fail, , <90% genome coverage and <100-fold read depth. Genotype refers to capsid genotype only.

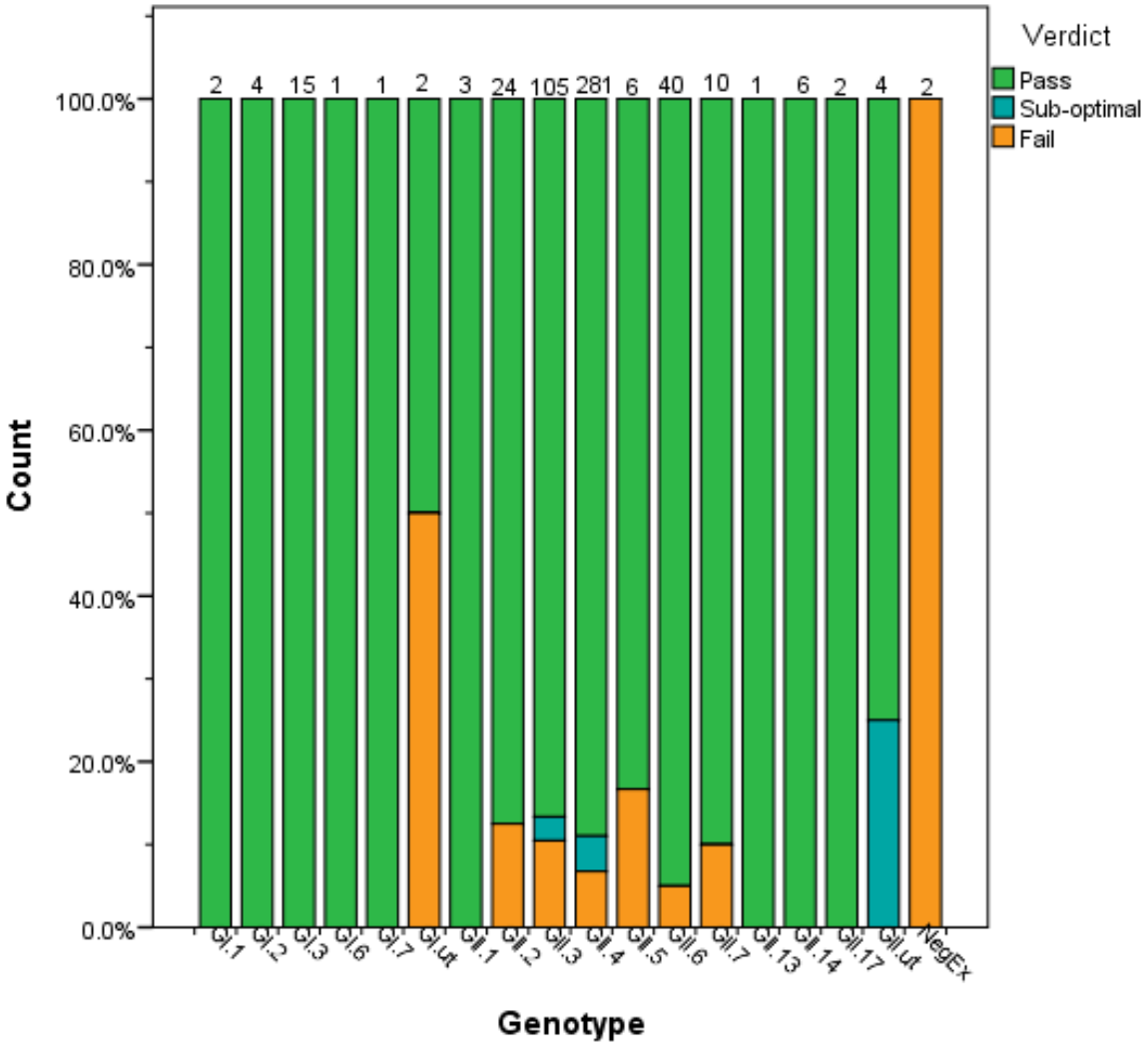


Figure A2. Correlation of read depth and % OTR across all samples (n = 507). R = 0.757, P <0.001 (Spearman's correlation)

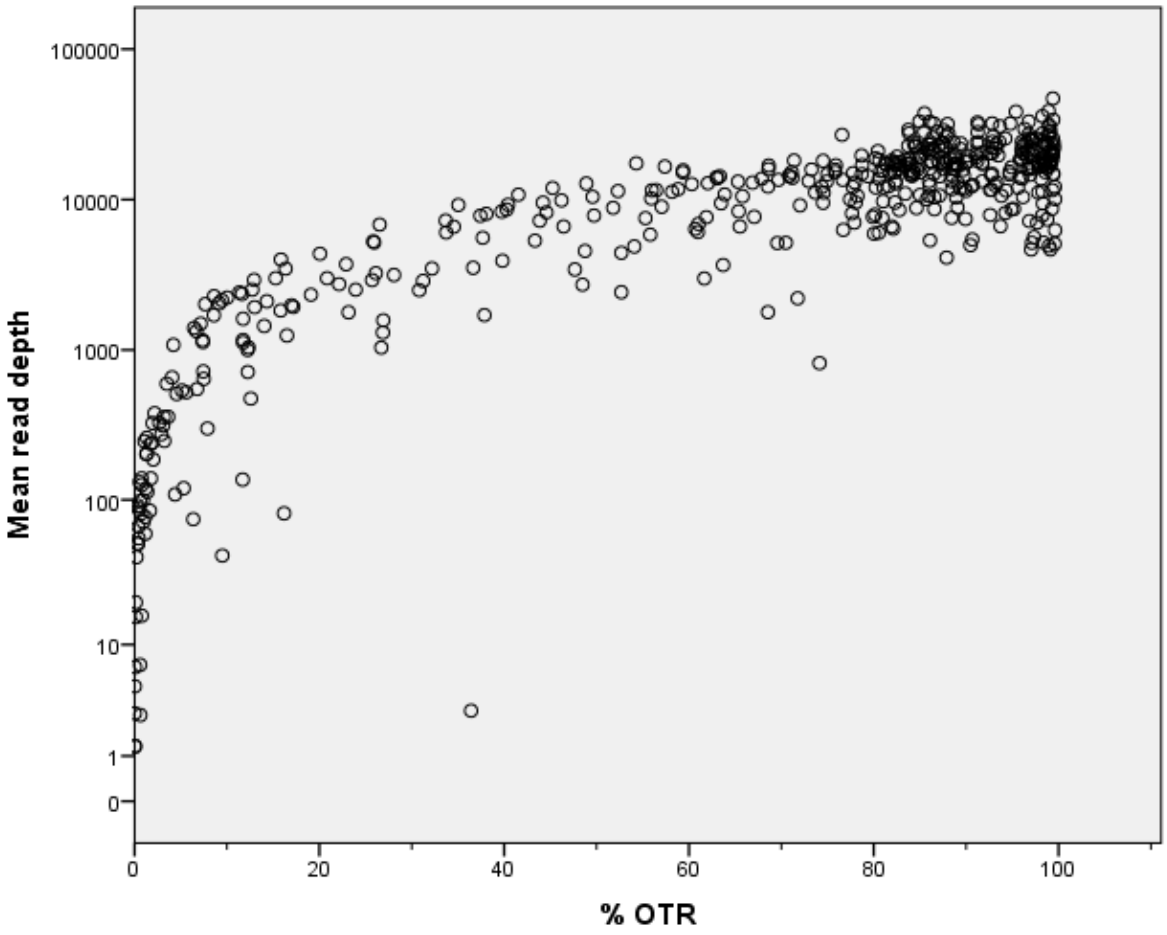


Figure A3. % genome coverage and Ct value, with sequencing runs 30 and 31 highlighted. Green circles highlight outliers that cannot be explained by Sequencing runs 30 and 31

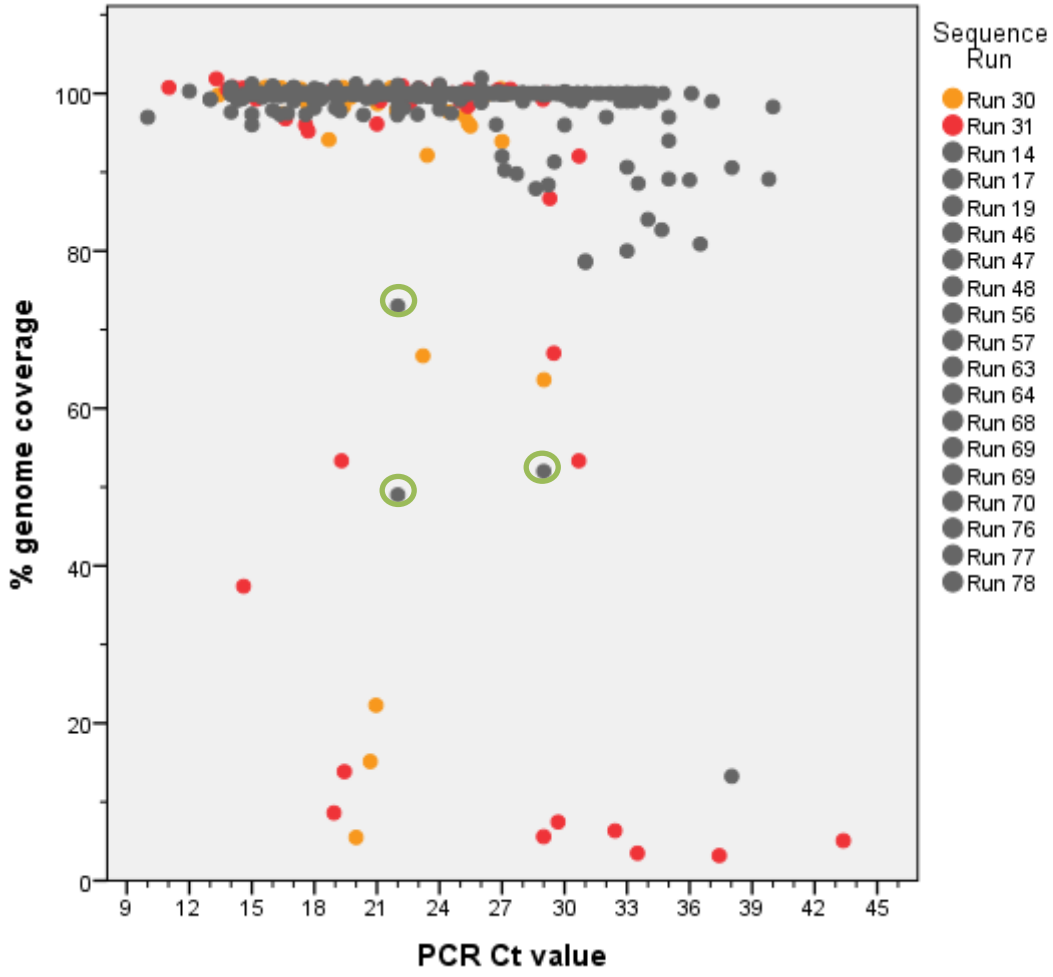


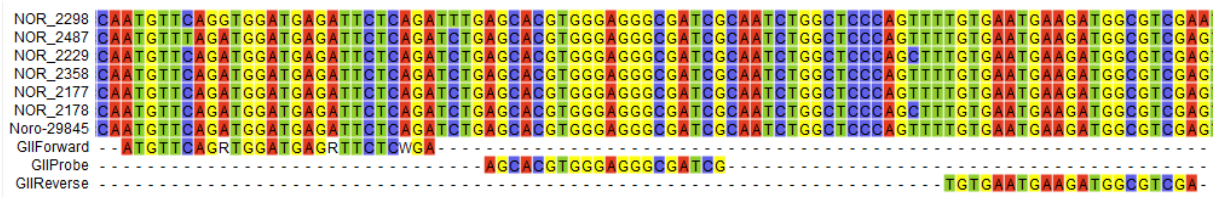
Table A1. Sequencing results of samples with RT-qPCR Ct value <30 and unexpectedly low % genome coverage (<80%), excluding samples from sequencing runs 30 and 31 (n = 3; highlighted in Supplementary Figure 2).

Sample name	Genotype	RT-qPCR Ct value	% genome coverage	% OTR*	Average read depth
NOR_2431	GII.5	22	49%	0.01%	1
NOR_2488	GII.4	29	52%	6.76%	137
NOR_2500	GII.4	22	73%	2.53%	120

*% OTR, % on-target-reads

Figure A4. Alignment of samples with RT-qPCR Ct value <30 and unexpectedly low % genome coverage (<80%), excluding samples from sequencing runs 30 and 31 (n = 3; highlighted in Supplementary Figure 2 and detailed in Supplementary Table 1).

Figure A5. Nucleotide alignment of all samples with real-time PCR Ct value ≥36 (i.e. apparently low titre) but >80% genome coverage (n = 7) by SureSelect target enrichment sequencing



Degenerate nucleotide code: R, A/G; W, A/T

Table A2. Genotyping success of samples processed in parallel by SureSelect target enrichment for full genome sequencing and PCR amplification of capsid shell domain followed by Sanger sequencing

	Full genome sequenced	Full genome failed**	Total
PCR amplified	158	0	158 (96%)
PCR failed*	6	0	6
Total	164 (100%)	0	164

* no amplification by PCR

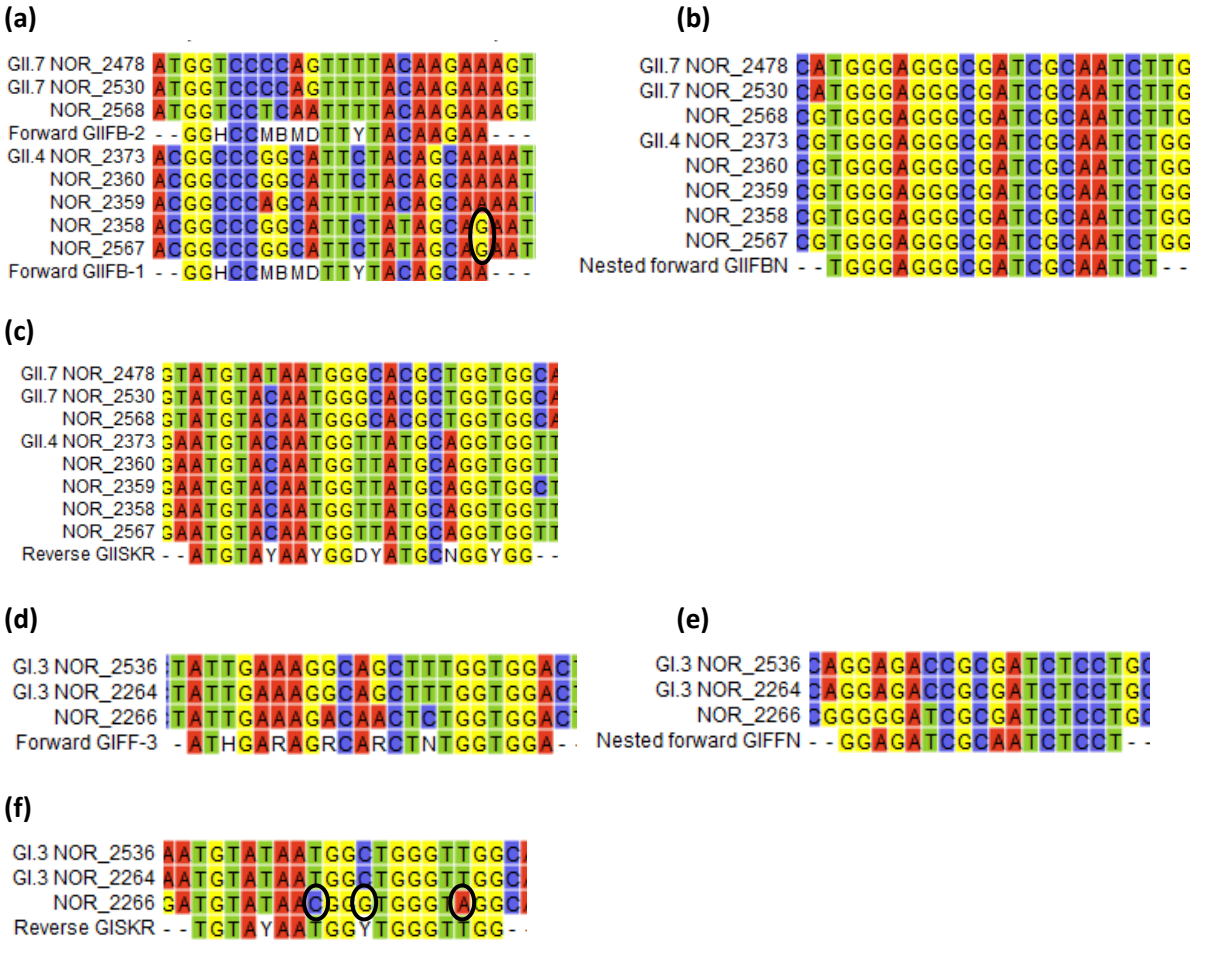
** <90% genome coverage and <100-fold read depth

Table A3. Sequencing results of samples that failed to amplify the capsid shell domain by PCR for genotyping, but were successfully sequenced (>99% genome coverage) by SureSelect target enrichment (n=6).

Sample name	Genotype	RT-qPCR Ct value	% genome coverage	% OTR*	Average read depth
NOR_2266	GI.3	20	100%	70%	13,957
NOR_2568	GII.7	31	99%	83%	11,235
NOR_2567	GII.4	27	100%	91%	23,068
NOR_2360	GII.4	33	100%	15%	3,995
NOR_2359	GII.4	33	100%	8%	2,281
NOR_2358	GII.4	37	99%	55%	11,740

*% OTR, % on-target-reads

Figure A6. Nucleotide alignment of all GII (n=5) and GI (n =1) samples that failed to amplify the capsid shell domain by PCR for genotyping, but were successfully sequenced (>99% genome coverage) by SureSelect target enrichment. Additional sequences (labelled with genotype) that were genotyped successfully are included for comparison. Mismatches predicted to cause amplification failure are circled. **(a)** GII forward; **(b)** GII nested forward; **(c)** GII reverse; **(d)** GI forward; **(e)** GI nested forward; **(f)** GI reverse



Degenerate nucleotide code: **H**, A/C/T; **M**, A/C; **B**, G/T/C; **D**, G/A/T; **Y**, C/T; **R**, A/G; **N**, G/C/A/T

Figure A7. Distribution of **(a)** mean read depth and **(b)** % genome coverage for stool samples and negative extracts. Red lines indicate median values

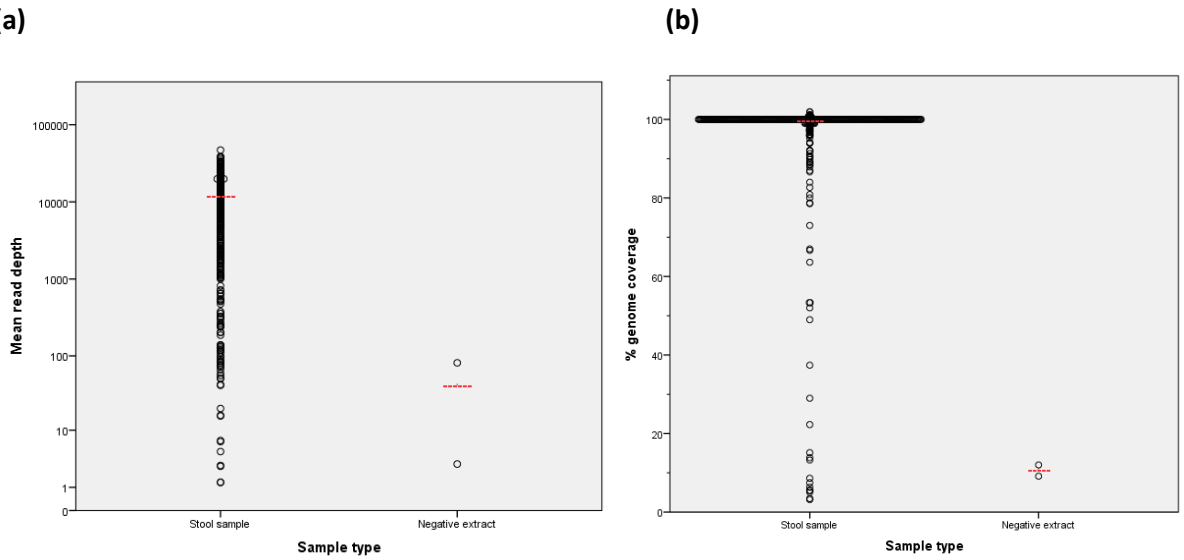


Figure A8. Contigs generated from Negative Extract mapping to norovirus full genome

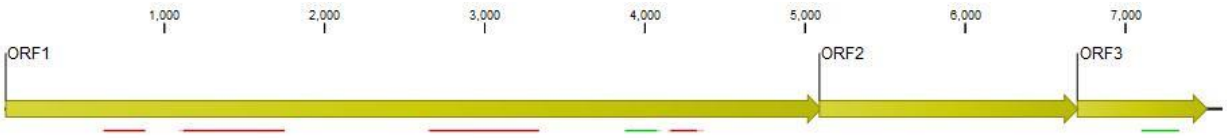


Table A4. Mixed infections in clinical specimens identified during assembly pipeline

Sample name	Genotypes present	Step in assembly pipeline that identified mixed infection	Number of reads mapping to each genotype in filtering step
NOR_2565	GII.3, GII.4 and GII.2	Filtering step (mapping to reference list)	GII.3: 792,264 GII.4: 113,068 GII.2: 859,899
NOR_2276	GII.3 and GII.6	Filtering step (mapping to reference list)	GII.3: 425,961 GII.6: 373,439
13V35152	GII.3 and GII.4	Align contigs to reference	GII.4: 470,377 GII.3: 1,155

Table A5. Comparison of consensus sequences generated from *de novo* assembly in single infections (original) and a simulated mixed infection (mixed).

	Samples data originates from	Number of reads mapping	Read depth	Consensus sequence length	% identity between consensus sequences	SNP difference between consensus sequences	SNPs in ORF1	SNPs in ORF2	SNPs in ORF3
GII.3 original	Noro-14069	413,812	3,154	7,555	97.61	178	163/178	15/178	0/178
GII.3 mixed	Noro-14069 & -28464	630,742	3,850	7,459					
GII.4 original	Noro-28464	431,263	2,917	7,489	95.53	332	284/332	22/332	26/332
GII.4 mixed	Noro-14069 & -28464	676,523	4,017	7,426					

Table A6. Summary of norovirus whole genome sequencing reports in the literature

Author	Country	Method	Sequencing platform	Assembly	Number of samples	Genotypes (% success*)	% OTR	Genome coverage	Read depth
Chhabra (2010)(1)	India	Overlapping PCR (9 amplicons)	Sanger sequencing	Not stated	3	GII.4 and GII.b_GII.3†	n/a**	100%	n/a††
Won (2013)(2)	South Korea	Overlapping PCR (10 amplicons)	Sanger sequencing	Not stated	1	GII.P12_GII.13†	n/a**	100%	n/a††
Eden (2013) (3)	Australia	Long PCR (7.6kb)	Sanger sequencing	Not applicable	25	GII.4†	n/a**	100%	n/a††
Kundu (2013)(4)	UK	Overlapping PCR (22 amplicons)	Roche 454	De novo	13	GII.4 †	n/a**	86–99%	Median 580 (423–951)
Cotten (2014) (5)	Vietnam	Overlapping PCR (3 amplicons)	Illumina MiSeq	De novo	265	GI (20%) GII.2 (40%) GII.3 (77%) GII.4 (92%) GII.6 (88%) GII.7 (0%) GII.9 (100%) GII.12 (50%) GII.13 (83%) Overall success 78%	n/a**	100%	Not stated
Nakamura (2012) (6)	Japan	Whole transcriptome amplification kit (70 cycles PCR)	Roche 454	Reference mapping	5	GII.4 (40%)	Median 3% (0.05–60%)	Median 84.5% (2.1–98%)	9–259

Author	Country	Method	Sequencing platform	Assembly	Number of samples	Genotypes (% success*)	% OTR	Genome coverage	Read depth
Batty (2013) (7)	UK	mRNA RNASeq	Illumina MiSeq and HiSeq	Reference mapping	3 (MiSeq) plus 77 (HiSeq)	GII.4 (99%)	MiSeq: median 1.8% (0.12–1.90) HiSeq: median 2.7% (0.01–97.98%)	MiSeq: 97–99% HiSeq: Mean 97% (59 – 99%)	MiSeq: Not stated HiSeq: Median 100 (10–1,000)
Wong (2013) (8)	UK	mRNA RNASeq	Illumina MiSeq	De novo	32	GII.4 (66%)	Not stated	21/32 >97%	Not stated
Bavelaar *et al.* (2015) (9)	Netherlands	Whole transcriptome amplification with ribosomal RNA depletion	Ion Torrent PGM (318 chip)	Reference mapping	10	GI.1 (100%) GI.6 (100%) GII.4 (100%) GII.6 (100%) GII.21 (100%) GII.2 (100%)	Median 28% (0.7–70.9%)	100%	Median 1309 (25–3607)

*>90% genome coverage

** not applicable for PCR amplicon sequencing

† success rate not reported

†† Read depth not applicable for capillary sequencing

1. **Chhabra P, Walimbe AM, Chitambar SD.** 2010. Complete genome characterization of Genogroup II norovirus strains from India: Evidence of recombination in ORF2/3 overlap. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **10:**1101-1109.
2. **Won YJ, Park JW, Han SH, Cho HG, Kang LH, Lee SG, Ryu SR, Paik SY.** 2013. Full-genomic analysis of a human norovirus recombinant GII.12/13 novel strain isolated from South Korea. PloS one **8:**e85063.

3. **Eden JS, Tanaka MM, Boni MF, Rawlinson WD, White PA.** 2013. Recombination within the pandemic norovirus GII.4 lineage. Journal of virology **87:**6270-6282.
4. **Kundu S, Lockwood J, Depledge DP, Chaudhry Y, Aston A, Rao K, Hartley JC, Goodfellow I, Breuer J.** 2013. Next-generation whole genome sequencing identifies the direction of norovirus transmission in linked patients. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **57:**407-414.
5. **Cotten M, Petrova V, Phan MV, Rabaa MA, Watson SJ, Ong SH, Kellam P, Baker S.** 2014. Deep sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting. Journal of virology **88:**11056-11069.
6. **Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T.** 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PloS one **4:**e4219.
7. **Batty EM, Wong TH, Trebes A, Argoud K, Attar M, Buck D, Ip CL, Golubchik T, Cule M, Bowden R, Manganis C, Klenerman P, Barnes E, Walker AS, Wyllie DH, Wilson DJ, Dingle KE, Peto TE, Crook DW, Piazza P.** 2013. A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. PloS one **8:**e66129.
8. **Wong TH, Dearlove BL, Hedge J, Giess AP, Piazza P, Trebes A, Paul J, Smit E, Smith EG, Sutton JK, Wilcox MH, Dingle KE, Peto TE, Crook DW, Wilson DJ, Wyllie DH.** 2013. Whole genome sequencing and de novo assembly identifies Sydney-like variant noroviruses and recombinants during the winter 2012/2013 outbreak in England. Virology journal **10:**335.
9. **Bavelaar HH, Rahamat-Langendoen J, Niesters HG, Zoll J, Melchers WJ.** 2015. Whole genome sequencing of fecal samples as a tool for the diagnosis and genetic characterization of norovirus. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **72:**122-125.