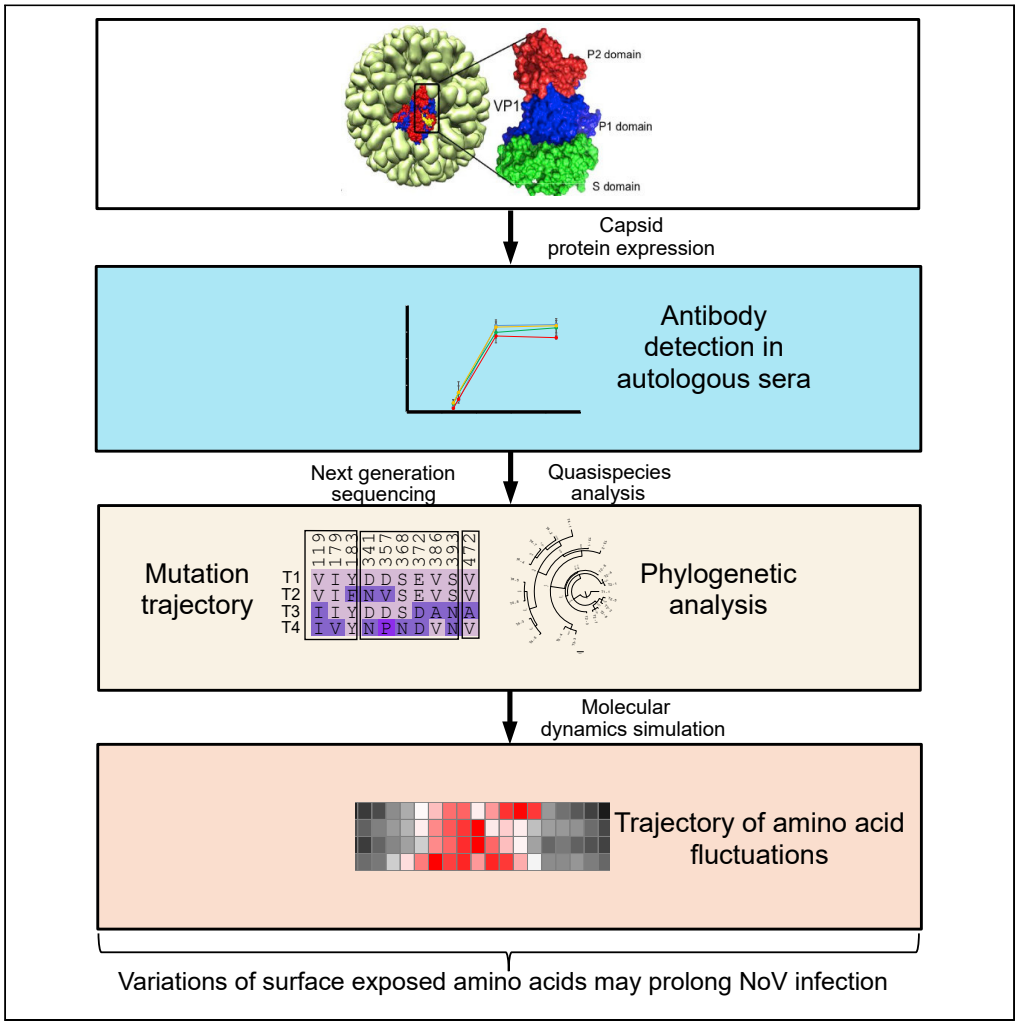
iScience

CellPress
OPEN ACCESS

Article

Prolonged norovirus infections correlate to quasispecies evolution resulting in structural changes of surface-exposed epitopes



Suliman Qadir Afridi, Zainab Usman, Sainitin Donakonda, ..., Ulrike Protzer, Hassan Moeini, Dieter Hoffmann

dieter.hoffmann@tum.de

Highlights

Noroviruses accumulated numerous mutations during chronic infection

After 1 month, intraindividual norovirus populations became genetically diverse

High genetic complexity and genetic distances were associated with prolonged infections

Afridi et al., iScience 24, 102802
July 23, 2021
https://doi.org/10.1016/j.isci.2021.102802

iScience

CellPress
OPEN ACCESS

Article

Prolonged norovirus infections correlate to quasispecies evolution resulting in structural changes of surface-exposed epitopes

Suliman Qadir Afridi,[1,6] Zainab Usman,[2] Sainitin Donakonda,[3,6] Jochen Martin Wettengel,[1] Stoyan Velkov,[1] Robert Beck,[4] Markus Gerhard,[5,6] Percy Knolle,[3,6] Dmitrij Frishman,[2] Ulrike Protzer,[1,6] Hassan Moeini,[1,6] and Dieter Hoffmann[1,6,7,*]

SUMMARY

In this study, we analyzed norovirus (NoV) evolution in sequential samples of six chronically infected patients. The capsid gene was amplified from stool samples, and deep sequencing was performed. The role of amino acid flexibility in structural changes and ligand binding was studied with molecular dynamics (MD) simulations. Concentrations of capsid-specific antibodies increased in sequential sera. Capsid sequences accumulated mutations during chronic infection, particularly in the surface-exposed antigenic epitopes A, D, and E. The number of quasispecies increased in infections lasting for >1 month. Interestingly, high genetic complexity and distances were followed by ongoing NoV replication, whereas lower genetic complexity and distances preceded cure. MD simulation revealed that surface-exposed amino acid substitutions of the P2 domain caused fluctuation of blockade epitopes. In conclusion, the capsid protein accumulates numerous mutations during chronic infection; however, only those on the protein surface change the protein structure substantially and may lead to immune escape.

INTRODUCTION

Noroviruses (NoV) are the main cause of viral gastroenteritis worldwide. They are small non-enveloped (27–40 nm), positive-sense RNA viruses with a 7.5 kb long genome comprising three overlapping open reading frames (ORFs). The first ORF encodes non-structural proteins (i.e. RNA polymerase), the second ORF encodes the major capsid protein (VP1), which helps the virus attach to the host cells, and the third ORF encodes a minor structural protein (VP2), which helps stabilize the capsid protein. Norovirus is highly diverse with more than 40 genotypes classified into seven genogroups (GI GVII), of which only GI, GII, and GIV are clinically associated with human disease. GII.4 has been the most prevalent genotype over the past years (Donaldson et al., 2010). Every 3–5 years, new GII.4 variants replace previous strains. Other genotypes with a lower prevalence do not show this epochal evolution. A natural NoV infection usually generates immunity for a couple of years to the respective genotype and possibly cross-immunity against other NoV strains. Immunity correlates to the presence of neutralizing antibodies (Bok et al., 2011; Malm et al., 2014). Previous studies described three blockade epitopes A (294, 296, 298), D (391–395), and E (407, 412, 413) within the P2 domain of the GII.4 capsid protein (Allen et al., 2009; Zakikhany et al., 2012). Of those particularly, epitope D plays a role in virus-host interactions (Debbink et al., 2012a; Zakikhany et al., 2012). Herd immunity against highly prevalent GII.4 strains may select new antigenically divergent variants. The emergence of such variants coincides with increased NoV prevalence and outbreaks (Siebenga et al., 2010). Since 1995, six different GII.4 variants caused pandemics worldwide: US 1995/96, Farmington Hills 2002, Hunter 2004, Den Haag 2006b, New Orleans 2009, and Sydney 2012 (Eden et al., 2014). NoV RNA polymerase introduces a large number of mutations due to poor proofreading mechanisms. Thus, closely related viruses, known as quasispecies (QS), arise within the infected host (Domingo et al., 1997, 1998; Eigen, 1993). Chronic NoV infections occur in immunocompromised individuals, such as transplant recipients and small/preterm infants with comorbidities (Angarone et al., 2016). Previously, we studied clonal Sanger sequences from chronically infected patients (Hoffmann et al., 2012). High intra-individual diversity and rapid evolution of NoV between sequential

[1]Institute of Virology, Technische Universität/Helmholtz Zentrum München, 81675 Munich, Germany

[2]Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, 85354 Freising, Germany

[3]Institute of Molecular Immunology and Experimental Oncology, Technische Universität München, 81675 Munich, Germany

[4]Institute of Medical Virology and Epidemiology of Viral diseases, Universitäts Klinikum Tübingen, 72076 Tübingen, Germany

[5]Institute of Medical Microbiology, Immunology and Hygiene, Technische Universität München, 81675 Munich, Germany

[6]German Center for Infection Research (DZIF), Munich, Germany

[7]Lead contact

*Correspondence: dieter.hoffmann@tum.de

https://doi.org/10.1016/j.isci.2021.102802

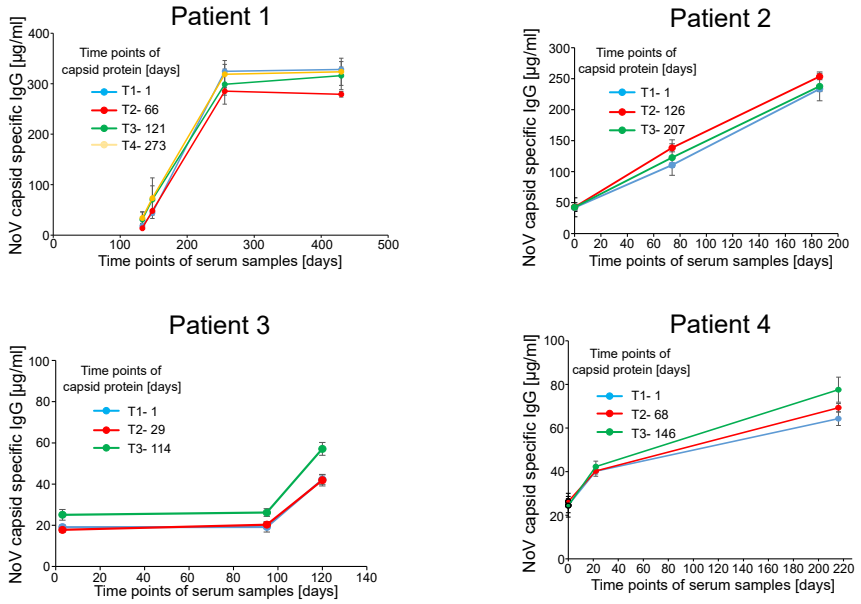


Figure 1. IgG antibodies directed at autologous capsid proteins in sera of patients 1–4

Capsid proteins were expressed from NoV sequences originating from sequential stool samples (T1−T4, labeled blue, red, green, and yellow). X axis: time points of sera drawn after the first positive PCR. Y axis: capsid-specific IgG concentration (means of triplicate measurements, bars indicate standard deviations). Antibody concentrations increase over time and are virtually identical between sequential capsid proteins.

samples prompted us to apply next-generation sequencing in this setting. We selected the PacBio platform as it provides long-read sequencing targeting the entire capsid gene as our gene of interest. Recently, we have developed a quantitative enzyme-linked immunosorbent assay (ELISA) for NoV-specific IgG and analyzed sera from immunocompromised patients before and after acute NoV infection (Afridi et al., 2019). In the present study, we complement capsid-specific antibody measurement in sequential seras with genotypic and protein structural analysis derived from autologous capsid sequences.

RESULTS

Capsid-specific antibody concentrations in patient sera

To monitor humoral immune responses over time, we measured binding of autologous sera samples to recombinant capsid proteins cloned from the respective patient's stool samples. Quantitative ELISA revealed sequential increase of NoV-specific antibodies between sera of patients chronically infected with GII.4 (patients 1–3) or GII.3 (patient 4) (Figure 1). NoV capsid gene sequences of patient 1 originated from day 0, 66, 121, and 27 stool samples. Serum samples were drawn at days 133, 148, 256, and 430 after first positive polymerase chain reaction (PCR). Day 133 and day 148 sera yielded low antibody concentrations (14–34 and 44–73 µg/mL, respectively) with all sequential antigens. In day 256 sera, antibody concentration levels increased markedly to 324, 285, 298, and 318 µg/mL, while levels remained unchanged in day 430 serum (328, 279, 315, and 323 µg/mL). Apparently, this patient mounted a robust antibody response between day 148 and 256. Day 0 serum of patient 2 yielded a low antibody concentration (around 42 µg/mL) directed at all sequential antigens. In day 74 serum, antibodies substantially increased (110, 138, and 122 µg/mL). Conversely, day 186 serum yielded high antibody concentrations of 233, 253, and 237.7 µg/mL, respectively. The capsid proteins of patient 3 derived from days 0, 29, and 114 were incubated with serum samples from days 3, 95, and 120. Low antibody concentration was detected in day 3 sera (19, 18, and 25 µg/mL) and day 95 sera (19, 20, and 26.2 µg/mL). Antibody concentration increased markedly in day 120 serum. Serum samples of patient 4 drawn at days 0, 22, and 216 were incubated with capsid antigens expressed from days 0, 68, and 146. Low antibody concentrations (25, 26, and 25 µg/mL) were detected in day 0 serum. They increased to 40, 40, and 42 µg/mL at day 22 and further to 62, 68, and 72 µg/mL at day 216. This analysis showed that antibodies directed at autologous capsid proteins

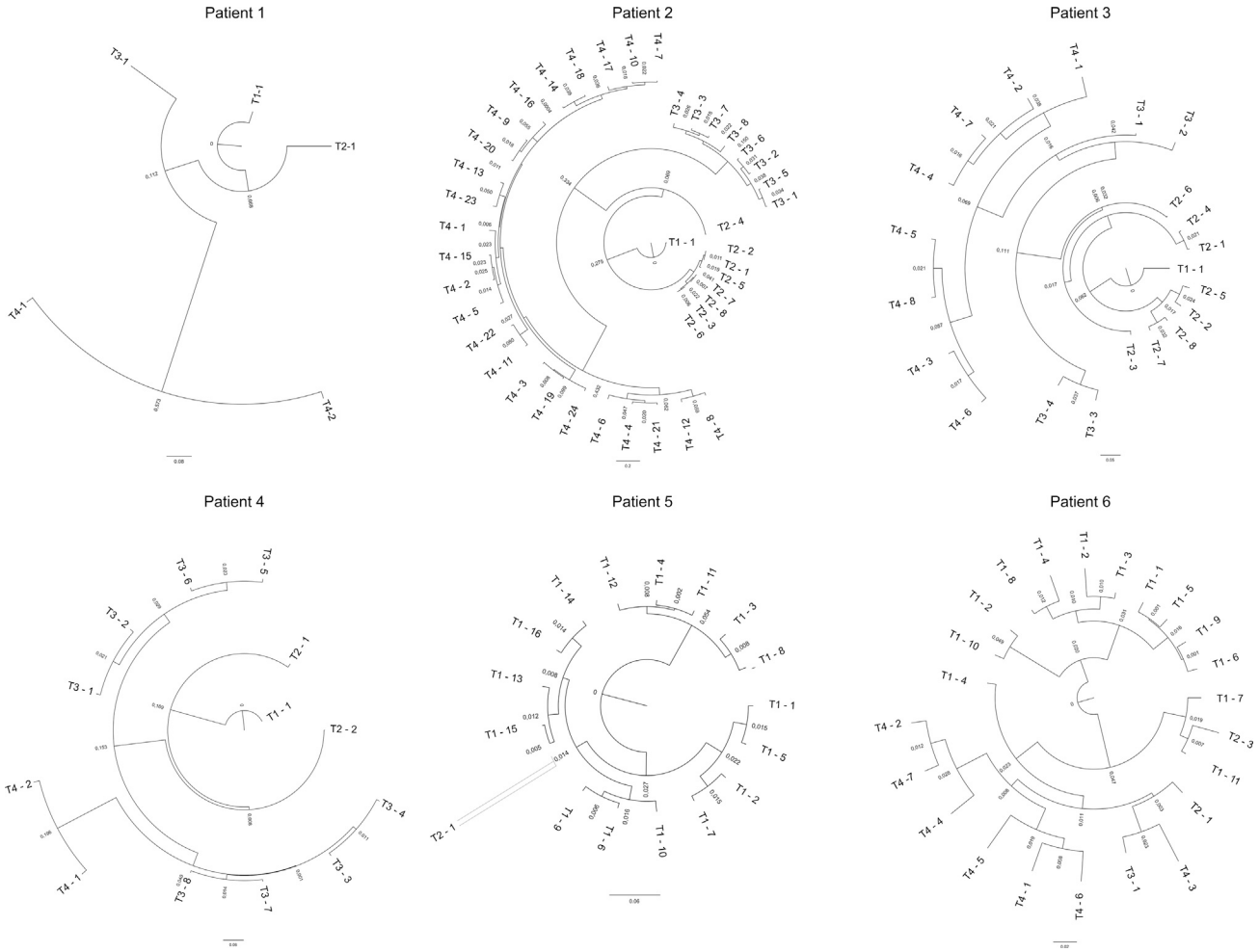


Figure 2. Phylograms of capsid sequences derived from six chronically infected patients

Trees illustrate the evolution of reconstructed quasispecies with >1% frequency. The numbers above the branches indicate posterior probabilities of branch lengths. Earlier sequences are located in the inner part and later sequences in the outer part of the phylograms. See also Table 1.

increased markedly over time in sequential sera of chronically infected patients. Of note, ELISA signals were constant between sequentially expressed capsid proteins.

Intraindividual evolution of NoV quasispecies

To examine the complex evolution of heterogeneous NoV populations, we reconstructed QS from next generation (NG) capsid sequences. Between 679 and 14,816 reads per sample were generated, encompassing the entire capsid gene and aligned with the respective reference sequences. For QS analysis, we considered all variants with a frequency >0.1%. In the phylograms (Figure 2), we only depict QS with a frequency >1% for better graphic representation. The number of QS was low in the first samples of all patients (Table 1). This implies that they originated from an early infection stage and reflects a genetic bottleneck during transmission. Overall, more QSs were reconstructed from later samples of patients 1–4 during the course of chronic infection. The QS number decreased from 149 to 32 between sample 3 and 4 of patient 4 and from 41 to 7 between the 2 samples of patient 5. This reduced diversity possibly reflects selection of fit variants supporting extended viral replication. Patient 5 was NoV positive for at least 17 months before our first sequenced sample, so a higher diversity had already been established. The ratio QS/read reflects these trends and is low for all samples of patient 6 because of the high read number per sample. We hypothesize that this patient was not infected long enough for the virus populations to get more diverse. Viral populations are

Table 1. Patient characteristics, reads, complexity, and quasispecies of sequential samples

Patient	Gender	Age(y)	Genotypes	Underlaying conditions	Samples	Time interval	No. of reads	Mapped reads	Shanon entropy	Total no. of quasispecies	QS/reads	Ct Value
1	M		GII.4	Bone marrow transplant	T1							
		58			T2	0	1965	1957	0.020	15	0.007	22.9
					T3	66	823	820	0.023	12	0.014	19.3
					T4	121	5888	5866	0.027	113	0.019	16.7
						273	5507	5487	0.025	125	0.022	17.5
2	M		GII.4	Bone marrow transplant	T1							
		2			T2	0	3108	3013	0.022	4	0.001	15.9
					T3	126	2218	2150	0.030	92	0.041	17.7
					T4	207	2775	2692	0.026	104	0.037	19.1
						458	7936	7834	0.038	130	0.0163	23.5
3	M		GII.4	Bone marrow transplant	T1							
		60			T2	0	2718	2619	0.022	2	0.001	20.5
					T3	29	978	940	0.022	179	0.183	20.1
					T4	114	890	862	0.027	161	0.180	21.7
						141	897	863	0.021	318	0.345	33.9
4	F		GII.3	Bone marrow transplant	T1							
		68			T2	0	905	838	0.049	8	0.008	28.6
					T3	68	5841	5723	0.041	58	0.009	25.9
					T4	146	1186	1156	0.033	149	0.125	32.9
						216	2304	2025	0.052	32	0.013	21.4
5	M		GII.6	HIV infection	T1							–
		46			T2	0	3249	3155	0.050	41	0.012	–
						176	3593	3493	0.043	7	0.001	
6	M		GII.4	Bone marrow transplant	T1							–
		59			T2	0	9257	9257	0.019	16	0.001	–
					T3	8	14,816	14,816	0.012	4	0.002	–
					T4	14	1194	1194	0.015	1	0.0008	–
						25	6670	6670	0.013	12	0.001	
7[a]	M		GII.4	–	T1				–		–	–
		45				0	506	506		2		
8[a]	M		GII.3	–	T1				–		–	–
		55				0	1106	1106		1		
9[a]	M		GII.2	–	T1				–		–	–
		52				0	3478	3478		1		

Related to Figure 2.

[a]Acutely infected patients.

Table 2. Genetic distance and dN/dS ratio between sequential samples

Patient	Time interval (days)	Genetic distance		
		Nucleotide	Amino acid	
1	63			
	54	0.040	0.089	2.15
	152	0.127	0.241	0.91
		0.041	0.073	0.92
2	124			
	206	0.005	0.015	0
	125	0.027	0.054	2.50
		0.060	0.149	3.76
3	22			
	62	0.076	0.151	1.90
	56	0.080	0.149	4.60
		0.011	0.034	0
4	65			
	78	0.076	0.141	1.74
	70	0.016	0.047	0
		3.786	4.074	2.5
5	175			
		0.028	0.052	2.11
6	7			
	9	0	0	0
	8	0	0	0
		0.077	0.232	0

Related to Figure 3.

virtually homogeneous in patients 7 9 with short virus replication. This illustrates that homogeneous viruses are transmitted to a new host, representing a genetic bottleneck. Over time, genetic diversity builds up because of high mutation rate.

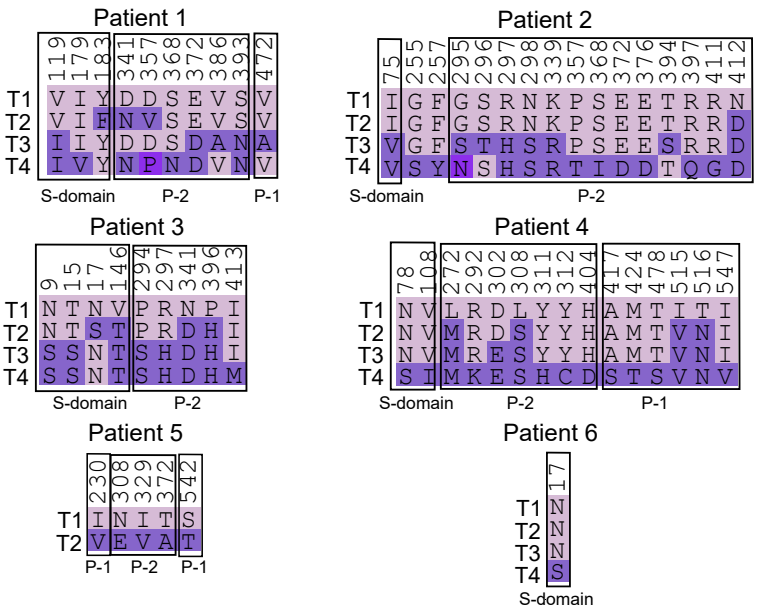
The phylogenetic trees (Figure 2) show only QS with >1% frequency. Samples T1-T4 of patient 1 yield only 5 QS with >1% frequency while 15, 12, 113, and 125 QSs with >0.1% frequency were reconstructed (Table 1). In comparison, patients 2–6 yielded more QS >1%. Phylograms of patients 4 with ongoing replication for at least 6 months and patient 5 without follow-up reflect homogeneous NoV populations of the respective last sample with only 2 and 1 QSs prevalent >1% as well as 32 and 7 > 0.1%, respectively.

Most branches that separate samples are reliable with a posterior probability >0.60. QS derived from the same sample mainly cluster together. As expected, variation within the respective samples was lower than that between them. QS of patients 1–5 reconstructed from a given sample evolved from a single ancestor of the previous sample. This finding was in line with positive selection, confirmed by ratios of divergence at nonsynonymous and synonymous sites, dN/dS. Four sequential samples of patient 6 taken within 25 days were separated by small evolutionary distances, reflected by small branch lengths and low posterior probabilities of the reconstructed QS. Reconstructed QS >0.1% (Table 1) and phylogenetic trees consisting of QS >1% illustrate how complex NoV populations emerge and evolve in chronically infected patients.

Higher genetic distance and complexity may predict ongoing viral replication

Diversity within and between samples was correlated to the duration of NoV replication. Genetic distances between consensus sequences of sequential samples were between 0.034 and 4.07 and generally lower at nucleic acid level (Table 2). This corresponds to dN/dS ratios greater or close to 1 between all sequential sequence pairs, indicating intra-individual positive selection (Table 2). Genetic distances seem to depend on the course of infection: They were higher (4.07) between sample 3 and 4 of patient 4 with ongoing NoV replication than between the respective samples 3 and 4 of patient 1 and 3 (0.073 and 0.034) who cleared NoV. Shannon entropy of patient-specific sequential samples reflecting genetic complexity ranged from 0.012 to 0.052 (Table 1). Samples of patient 6 with less than 30 days' course of infection were genetically less complex. Of note, values of

A



B

Patient	Genotype	Epitope A							Epitope D				Epitope E		
		294	296	297	298	368	372	373	391	393	394	395	407	412	413
1	GII.4					S	E			S					
						N	D			N					
2	GII.4		S	R	N	S	D				T	T		N	
			T	H	S	I	E				S	N		D	
3	GII.4	P		R											I
		S		H											M
6	GII.4														

Figure 3. Amino acid mutation trajectories of sequential consensus sequences

(A) All variable amino acid positions. Mutated amino acids are labeled dark purple. T1–T4 indicate time points of sequential samples.

(B) Mutations located in epitopes A, D, and E of GII.4-infected patients. See also Table 2.

patients with ongoing replication (4) or without follow-up (2, 5) were higher than those of patients who eliminated NoV shortly after our last sequenced samples (1, 3). Complex populations are expected to support chronic viral replication by providing numerous variants to adapt to the selection by the host's immune response. Successful adaptation may go in line with high genetic distances between consensus sequences and thus high evolutionary rates.

Amino acid mutations in consensus sequences of sequential samples

We aligned sample-specific consensus sequences to determine amino acid exchanges in the trajectory of each patient. Five to 16 amino acids mutated in patients infected >30 days (Figure 3A). Consensus sequences of patients 1 and 3 with documented cure accumulated 10 and 9 mutations. In comparison, patients 2 (no follow-up) and 4 (ongoing replication for >6 months) yielded 16 and 15 mutations, respectively. In patients 2 and 4, most mutations occurred between T3 and T4, whereas in patients 1 and 3, only few amino acids changed in the last samples. Amino acids within or neighboring the previously

Table 3. Positively selected amino acid sites with respective p values (likelihood-ratio test)

Patient	Genotype	Positively selected amino acid	p value
1	GII.4	119	0.662
		341	0.921
		357	0.996**
		368	1.000**
		372	0.993**
		393	0.955**
2	GII.4	297	1.00**
		357	0.996**
		372	0.993**
		412	0.960*
3	GII.4	146	0.736
		294	0.942*
		297	1.00**
		340	0.921
		341	0.921
		368	1.00**
		357	0.996**
		372	0.993**
		412	0.926
4	GII.3	78	0.985*
		108	0.997**
		272	0.985*
		292	0.958*
		302	0.985*
		308	1.00**
		311	0.985*
		312	0.985*
		404	0.985*
		417	0.99**
		424	0.998**
		478	0.985*
		515	0.997*
		516	0.985*
		547	0.985*

Related to Figure 4.
**p>99%.
*p>95%.

described epitopes A, D, and E mutated sequentially in all patients with virus replication >30 days (Figure 3B). Most mutations occurred in epitope A, followed by D and E. Seven of 16 total amino acid mutations of patient 2 were located in epitopes A, D, and E. A similar number of mutations occurred in patients 4 (GII.3 infected) and 5 (GII.6 infected). Most mutations were genotype specific. As expected, no mutations occurred at putative histo-blood group antigen (HBGA)-binding sites at positions 342–347 and 391–92 (Debbink et al., 2012a; Lindesmith et al., 2013).

Positively selected amino acids

Four to 15 amino acids were positively selected in patients 1–4 and none in patient 5 (Table 3). Positions 297, 341, and 357 were positively selected in 3 out of 4 GII.4-infected patients; positions 297 and 368 in epitope A were positively selected with the highest p values.

Molecular dynamics analysis reveals fluctuating surface-exposed epitopes

To assess the effects of amino acid perturbation at a structural level, we performed homology modeling of consensus sequences from patient samples (P1, P2, P3, and P6). We generated 3D structural models based on

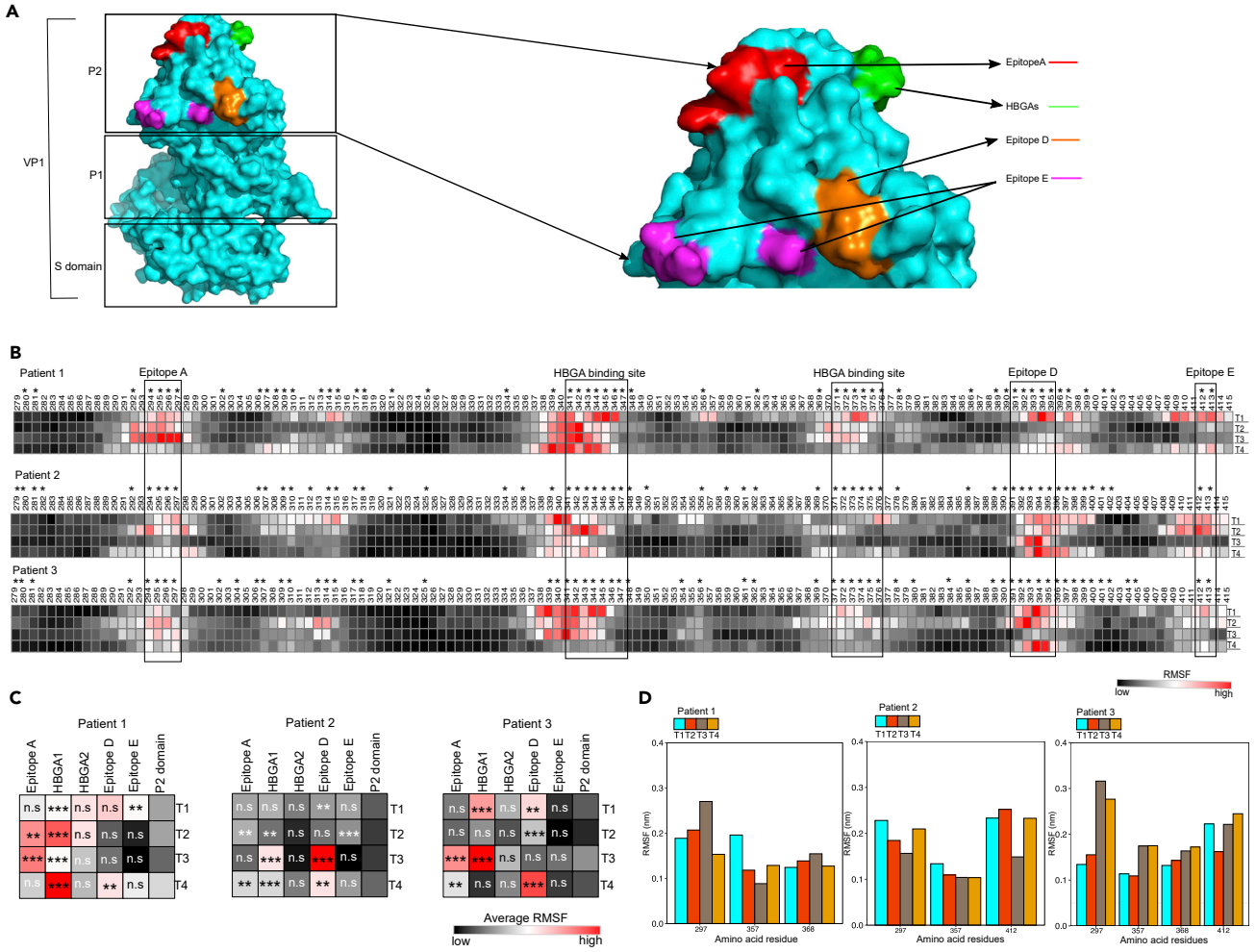


Figure 4. Root-mean-square fluctuation (RMSF) analysis of the P2 domain

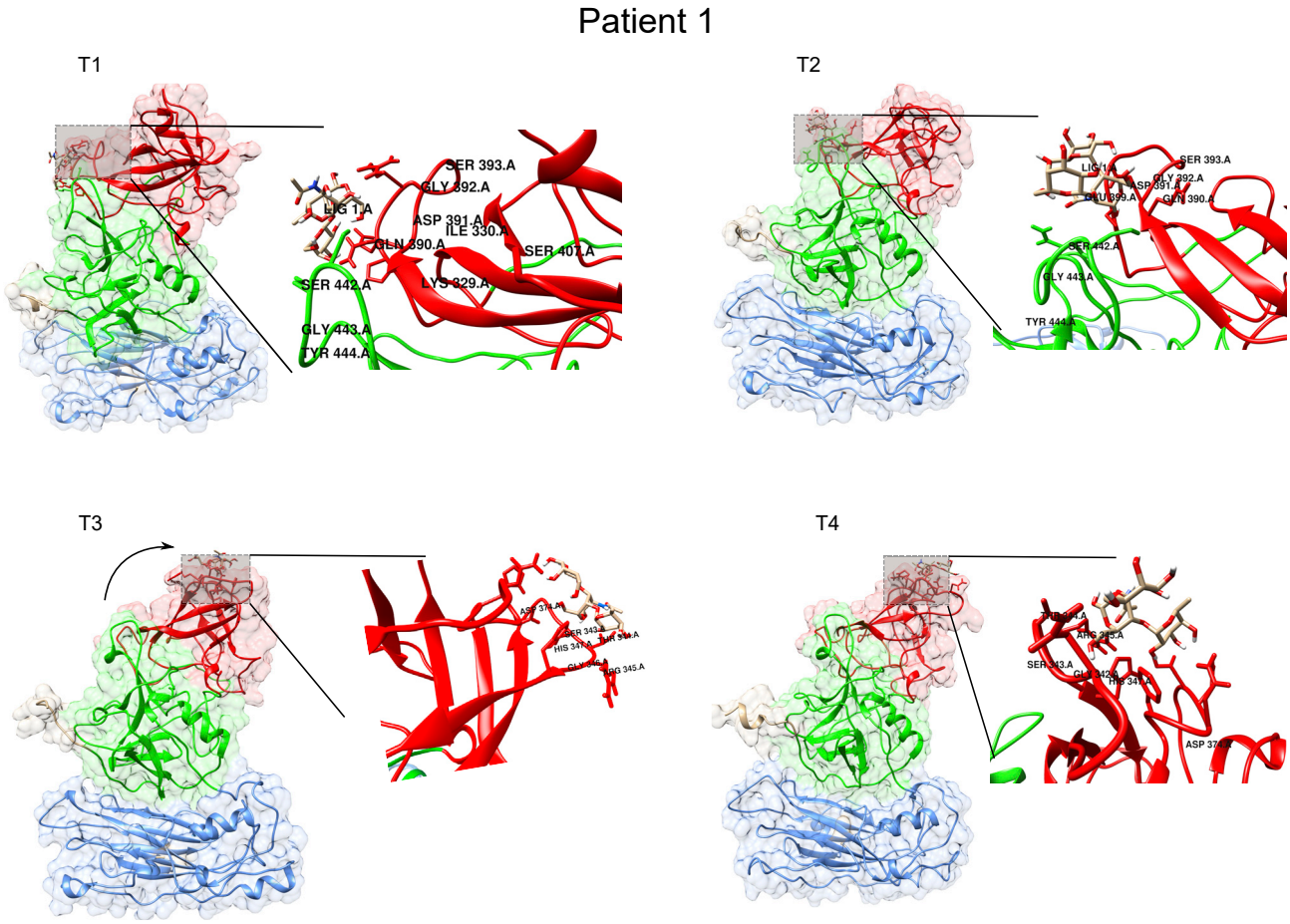
(A) Mapped epitopes of the capsid protein. See also Figures S1–S3.

(B) The heatmap represents the RMSF of the P2 domain. Fluctuation was simulated over 20 ns; asterisks highlight surface-exposed residues.

(C) RMSD values of epitopes and HBGA-binding sites compared to surface-exposed regions of the P2 domain. Statistical significance was analyzed with Wilcoxon rank-sum tests. (** p value <0.05, *** p value <0.001, n.s: not significant). Red color indicates high RMSF values and black color low RMSF values.

(D) RMSF of positively selected amino acid sites in three GII.4-infected patients, see also Table 3.

the crystal structure of capsid protein (PDBID: 1IHM). Predicted homology models overlapped well with experimental crystal structures (root-mean-square deviation [RMSD] <2 angstroms, Figure S1). ProSA analysis revealed that all models are in the range of native conformations (*Z* score: −5.98 to −6.57) (Figure S1). Protein geometry analysis with Ramachandran plots indicated that all the structures have > 93–97% favored regions (Figure S2). We also calculated *Z* scores to assess the distributions of main chain conformations in Ramachandran plots. *Z* scores <2 indicate a normal backbone geometry. Our analysis revealed *Z* scores <2 in all Ramachandran plot distributions, confirming trustworthy modeled structures (Figure S2). Evolutionary conservation analysis revealed that the P2 domain is the most variable part of all capsid proteins (Figure S3). To confirm the variability of the mapped epitopes (Figure 4A), we performed 20 ns MD simulations and explored the root-mean-square fluctuation (RMSF) for the patients with a long duration of infection (P1, P2, and P3). The heatmap (Figure 4B) illustrates that surface-exposed amino acids fluctuate most, particularly epitopes and HBGA-binding sites are more flexible than the remaining P2 domain of the VP1 capsid protein. Figure 4C illustrates significance levels of these differences, resulting from Wilcoxon tests. In P2 domain structures derived from patient 1, epitope A and HBGA-binding sites fluctuate significantly more than other surface-exposed amino acids. While HBGA-binding sites and epitope D fluctuate significantly stronger in P2 domain structures of patients 2 and 3, epitope A additionally fluctuates stronger than other surface-exposed amino acids of patient 3. Next, we analyzed the fluctuation of



Patient	Sample number	Predicted binding residues
1	T-1	390, 391, 392, 393, 442, 443, 444
	T-2	390, 391, 392, 393, 442, 443, 444
	T-3	343, 344, 345, 347, 374
	T-4	343, 344, 345, 347, 374

Figure 5. HBGA-binding sites of patient 1

Predicted HBGA-binding sites evolve between sequential samples. The binding pocket changes its position between T2 and T3, see also Figure S4.

positively selected amino acids (Figure 4D). In patient 1, amino acid 297 (epitope A) fluctuated more than amino acids 357 and 368; similarly, in patient 2, amino acids 297 and 412 (epitope E) fluctuated more than amino acid 357; finally, in patient 3, amino acids 297 and 412 fluctuated more than amino acids 357 and 368. In summary, positively selected amino acid sites located in epitopes A and E exhibited stronger fluctuation than positively selected sites outside of epitopes.

Molecular docking analysis indicates altered HBGA-binding pockets

HBGA binding of sequential GII.4 capsid protein structures derived from chronic infections was predicted by molecular docking. The results are summarized in Figure 5 (patient 1) and in Figure S4 (patients 2, 3, and 6). Altered binding sites correlate with the occurrence of amino acid substitutions near the HBGA-binding pocket. While identical amino acids were predicted to bind to HBGA in T1 and T2 of patient 1, the pattern changed in T3 and T4 when S394 mutated to N. In patient 2, identical sequences in T1 and T2 correlated to a stable HBGA-binding pocket. In sample T3, along with the T394S mutation and subsequent conformational changes, the binding pocket was predicted to shift from epitope D to the neighboring sites 344, 345, 347, and 374. Along with substitutions at positions 372, 376, and 397 in sample T4, the predicted binding sites evolved again and now shared a part of epitope D. For patient 3, the same amino acids in epitope

D and the surface-exposed part of the P1 domain were predicted to bind HBGA in samples T1 and T2. In sample T3, P294S and R297S and N341D substitutions near the binding pocket altered the HBGA-binding sites. In sample T4, along with mutation I413M, amino acids 392 and 393 replaced 397 as binding sites. As expected, the sequential samples of patient 6 with only one mutation in the trajectory showed a constant binding pattern. Taken together, mutations in the blockade epitopes located next to the HBGA-binding pocket alter the binding pattern.

DISCUSSION

Over the past few decades, the number of immunocompromised patients has greatly increased due to a rising number of allogeneic transplants, and NoV is recognized as an important cause of prolonged or chronic diarrhea in these patients (Echenique et al., 2016; Angarone et al., 2016). Infection in immunodeficient patients has been described as a chronic symptomatic infection with prolonged moderate diarrhea or as an acute symptomatic phase followed by extended asymptomatic shedding of NoVs (Echenique et al., 2016). In the present study, we monitored the evolution of NoV capsid sequences in chronically infected patients. The studied chronic infections differ substantially: Patients 1 and 3 overcame the infection shortly after our last NoV positive sample, whereas no cure was documented for patient 2. Patient 4 remained positive for at least 6 months after the last sample was collected. No negative sample was documented in the course of patient 5 with advanced HIV disease and poor immune function, making prolonged virus replication likely. Parallel to our previous study (Afridi et al., 2019) with immunocompromised patients who quickly cleared the infection, we detected increasing P2 domain-specific antibody concentrations in sequential sera. Antibody concentrations in both studies were in a similar range reaching up to 300 μg/mL (Afridi et al., 2019). While capsid-specific IgG concentrations vary between our chronically infected patients, they apparently do not correlate to the patient's outcome. Also, previous data show that HBGA blocking antibodies rather than total NoV-specific IgG protect individuals from infection (Atmar et al., 2015). Capsid proteins expressed from sequential stool samples yielded similar ELISA signals with autologous sera, showing no sign for immune escape. In accordance with our previous work, sequential capsid sequences clearly point at immune response as selection factor (Hoffmann et al., 2008, 2012).

As expected, QS increased in patients NoV positive for >1 month. Patient 5 was an exception because the patient was NoV positive for 1 year before we sequenced the first sample. While the QS number increased steadily in patients 1, 2, and 3, it dropped in the last sample T4 of patient 4 and patient 5 (Table 1). Numerous amino acid mutations were observed in T4 of patient 4. Only for this patient, a PCR-positive follow-up for >6 months after our last tested sample had been documented. Along with a high genetic distance between sequential consensus sequences, this could indicate that fitter new variants were selected, which may prolong infection. Genetic distances were particularly high between T3 and T4 of patient 4 and to a lesser extent in patient 2 (Table 2). Accordingly, 15 and 16 amino acid mutations occurred in these patients out of which 10 and 9 appeared in T4, respectively (Figure 3A). Shannon entropy for the entire capsid protein gene varied between 0.023 and 0.047 in chronic infections. Remarkably, virus populations of patient 4 with ongoing replication for >6 months and patient 5 with probable chronic course exhibited high entropy values in their last samples (Table 1). In addition, dN/dS calculated for the entire capsid was highest between T3 and T4 of patients 2 and 4, in line with positive selection (Table 2). In patient 5, capsid sequences evolved slower with fewer mutations, reflected by smaller genetic distances. Patients 6 and 7 were not infected long enough to establish diverse QS and reliable time-adjusted genetic distances between sequential samples. The last sequences sampled before cure in patients 1 and 3 were characterized by numerous reconstructed QS and low genetic distances between T3 and T4. The dN/dS ratio of complete capsid genes was >1 between most sequential samples, indicating positive selection (Table 2). Accordingly, genetic distances estimated between the sequential samples were higher for amino acids than for nucleic acids.

Phylograms based on next generation sequencing (NGS) (Figure 2) resemble their counterparts built from Sanger clonal sequences in chronically infected patients in that only one QS of a given sample evolves into QS of the next sequential sample (Hoffmann et al., 2012). Also, other authors constructed similar phylograms from the sequential samples of chronically infected patients (van Beek et al., 2017). Amino acid-specific analysis revealed more positively selected sites in patient 4 infected with GII.3 than patients 1, 2, and 3 (GII.4). In contrast to global setting, where GII.4 evolves faster than other II genotypes (Boon et al., 2011), intra-individual NoV populations seem to evolve dependent on their host's immune response, rather than

viral genotypes. Genetic distances between the QS remained low; however, they were higher on amino acid level than nucleotide level, reflecting positive selection. Particularly, in patient 3 with documented cure, the number of QS decreased sharply in the last sample, one week before cure.

Our analysis revealed accumulation of mutations across the VP1 gene including the antigenic epitopes A, D, and E in immunocompromised patients. Only few amino acid exchanges were located in other parts of the P domain (F257Y and G255S in patient 2) or in the S domain (V119I and I179V in patient 1; I75V in patient 2; N9S, T15S, and N17S in patient 3; N17S in patient 6). It seems that amino acid substitutions and subsequent evolution of blockade epitopes in the surface-exposed part of the P2 domain can lead to immune escape by decreasing antibody affinity while retaining HBGA binding (Nasheri et al., 2017). Amino acids at positions 294, 356, 368, 373, 376, 380, and 413 had also mutated in New Orleans (2009) and Sydney 2012 variants (Eden et al., 2014; Sabrià et al., 2018), suggesting also a global immunologic advantage. To study the effect of amino acid substitutions on the epitope and receptor conformation, 3D models of the capsid proteins were reconstructed, and the blockade epitopes were mapped (Debbink et al., 2012b). The 3D structure variation of proteins plays a vital role in biological activity and interaction with other molecules (Ponsel and Bruss, 2003). We applied MD simulations (Pronk et al., 2013) to assess the fluctuation of the important amino acid residues within the surface-exposed parts of the P2 domain (Figure 4A) over a period of time. RMSF analysis of the P2 domain revealed high fluctuation of the epitope sites, as shown in Figure 4B. Epitope sites fluctuate significantly higher than surface-exposed amino acids outside of epitopes (Figure 4C).

When taking positively selected amino acids into account, those within epitopes fluctuate stronger with higher RMSF values (Figure 4D). In patient 1, the amino acid substitutions at positions 372, 386, and 393 at T3 (Figure 3A) go in line with fluctuation at these positions (Figure 4B) and different predicted HBGA-binding sites (Figure 5). Similar changes were observed in patients 2 and 3 (Figure S4). Mutations in the surface-exposed amino acids 394, 397, and 412 in patient 2 and amino acids 297, 396, and 413 in patient 3 (Figure 3A) are fluctuating (Figure 4B) and are paralleled by altered HBGA-binding pockets (Figure S4). These data suggest that the amino acid substitutions within epitopes and neighboring sites coupled with conformational changes in the surface-exposed part of the P2 domain impact HBGA-binding patterns. This can lead to immune escape, as also recently shown by Lindesmith et al. (Lindesmith et al., 2019). We plan to confirm our HBGA binding predictions by *in vitro* assays as described previously (Moore et al., 2016). Amino acid residues binding to cellular receptors like HBGA are highly conserved and thus could be targeted effectively by neutralizing antibodies. Lindesmith et al. reported two mechanisms that NoV uses to escape from such neutralizing immune response: (1) amino acid mutations in epitope D that decrease antibody binding while retaining receptor binding (Lindesmith et al., 2019) and (2) mutations of epitope neighboring sites that cause conformational occlusion of blocking antibodies (Lindesmith et al., 2018). Both antibody evasion strategies likely imply conformational changes of epitopes, associated with mutations outside of the epitopes, as we described in the present study.

In conclusion, we describe how NoV strains evolve in chronically infected patients. We found most mutations in the P2 domain. Only substitutions of surface-exposed amino acids were paralleled by structural changes that may lead to immune escape. The data presented here will not only help to understand the genotypic and phenotypic changes in antigenic epitopes during NoV infection but also assist the development of NoV vaccines. So far, NoV transmission between immunocompromised patients has been reported only in an early stage of infection when the virus still maintains features necessary for transmission to new hosts (Kundu et al., 2013; Sukhrie et al., 2010; van Beek et al., 2017). In view of the substantial sequence and structural changes, we observed that the transmission of such heavily mutated strains to other individuals should be further studied. This could better define the role of chronically infected patients as a source of new variants with potential epidemic or even pandemic spread.

Limitations of the study

In this study, we have not generated the protein structures through experimental approaches (X-ray or NMR). To fully understand the biological structural alterations, experimental structures should be complemented by *in vivo* studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102802.

ACKNOWLEDGMENTS

We thank doctors, nurses, and patients for providing samples and clinical data. Technicians of our Clinical Virology Laboratory extracted RNA from patient samples. The German Academic Exchange Service (DAAD) supported S.Q.A. during his PhD studies. H.M. was supported by the German Center for Infection Research (DZIF) as a postdoctoral researcher (Grant numbers 06.804 and 06.817).

AUTHOR CONTRIBUTIONS

S.Q.A. performed the experiments and wrote the manuscript; Z.U. analyzed NGS data; S.D. contributed to structural modeling, validation of structural models, molecular dynamics simulations, and analysis of trajectories; J.M.W. helped clone the capsid gene for protein expression; S.V. helped in phylogenetic analysis; R.B. provided patients samples; D.F., P.K., and M.G. contributed to data analysis and manuscript; U.P. contributed to project design and manuscript; H.M. contributed to data analysis and manuscript; D.H. supervised the project, planned the experiments, and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 9, 2020
Revised: May 13, 2021
Accepted: June 24, 2021
Published: July 23, 2021

REFERENCES

Afridi, S.Q., Moeini, H., Kalali, B., Wettengel, J.M., Quitt, O., Semper, R., Gerhard, M., Protzer, U., and Hoffmann, D. (2019). Quantitation of norovirus-specific IgG before and after infection in immunocompromised patients. Braz. J. Microbiol. 1–5.

Allen, D.J., Noad, R., Samuel, D., Gray, J.J., Roy, P., and Iturriza-Gómara, M. (2009). Characterisation of a GII-4 norovirus variant-specific surface-exposed site involved in antibody binding. Virol. J. *6*, 1.

Angarone, M.P., Sheahan, A., and Kamboj, M. (2016). Norovirus in transplantation. Curr. Infect. Dis. Rep. *18*, 17.

Atmar, R.L., Bernstein, D.I., Lyon, G.M., Treanor, J.J., Al-Ibrahim, M.S., Graham, D.Y., Vinjé, J., Jiang, X., Gregoricus, N., and Frenck, R.W. (2015). Serological correlates of protection against a GII. 4 norovirus. Clin. Vaccin. Immunol. *22*, 923–929.

Bok, K., Parra, G.I., Mitra, T., Abente, E., Shaver, C.K., Boon, D., Engle, R., Yu, C., Kapikian, A.Z., and Sosnovtsev, S.V. (2011). Chimpanzees as an animal model for human norovirus infection and vaccine development. Proc. Natl. Acad. Sci. U S A *108*, 325–330.

Boon, D., Mahar, J.E., Abente, E.J., Kirkwood, C.D., Purcell, R.H., Kapikian, A.Z., Green, K.Y., and Bok, K. (2011). Comparative evolution of GII. 3 and GII. 4 norovirus over a 31-year period. J. Virol. *85*, 8656–8666.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569.

Debbink, K., Donaldson, E.F., Lindesmith, L.C., and Baric, R.S. (2012a). Genetic mapping of a highly variable norovirus GII. 4 blockade epitope: potential role in escape from human herd immunity. J. Virol. *86*, 1214–1226.

Debbink, K., Lindesmith, L.C., Donaldson, E.F., and Baric, R.S. (2012b). Norovirus immunity and the great escape. PLoS Pathog. *8*, e1002921.

Domingo, E., Baranowski, E., Ruiz-Jarabo, C.M., Martín-Hernández, A.M., Sáiz, J.C., and Escarmís, C. (1998). Quasispecies structure and persistence of RNA viruses. Emerg. Infect. Dis. *4*, 521.

Domingo, E., Martin, V., Perales, C., Grande-Perez, A., Garcia-Arriaza, J., and Arias, A. (2006). Viruses as Quasispecies: Biological Implications. Quasispecies: Concept and Implications for Virology (Springer).

Domingo, E., Menéndez-Arias, L., Quiñonesmateu, M.E., Holguín, A., Gutiérrez-Rivas, M., Martínez, M.A., Quer, J., Novella, I.S., and Holland, J.J. (1997). Viral Quasispecies and the Problem of Vaccine-Escape and Drug-Resistant Mutants. Progress in Drug Research/ Fortschritte der Arzneimittelforschung/Progrès des Recherches Pharmaceutiques (Springer).

Donaldson, E.F., Lindesmith, L.C., Lobue, A.D., and Baric, R.S. (2010). Viral shape-shifting: norovirus evasion of the human immune system. Nat. Rev. Microbiol. *8*, 231.

Echenique, I., Stosor, V., Gallon, L., Kaufman, D., Qi, C., and Zembower, T.R. (2016). Prolonged norovirus infection after pancreas transplantation: a case report and review of chronic norovirus. Transpl. Infect. Dis. *18*, 98–104.

Eden, J.-S., Hewitt, J., Lim, K.L., Boni, M.F., Merif, J., Greening, G., Ratcliff, R.M., Holmes, E.C., Tanaka, M.M., and Rawlinson, W.D. (2014). The emergence and evolution of the novel epidemic norovirus GII. 4 variant Sydney 2012. Virology *450*, 106–113.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., and Bettman, B. (2009). Real-time DNA sequencing from single polymerase molecules. Science *323*, 133–138.

Eigen, M. (1993). Viral quasispecies. Sci. Am. *269*, 42–49.

Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package), version 3.5 c (Joseph Felsenstein).

Hoffmann, D., Hutzenthaler, M., Seebach, J., Panning, M., Umgelter, A., Menzel, H., Protzer, U., and Metzler, D. (2012). Norovirus GII. 4 and GII. 7 capsid sequences undergo positive selection in chronically infected patients. Infect. Genet. Evol. *12*, 461–466.

Hoffmann, D., Seebach, J., Cosma, A., Goebel, F.D., Strimmer, K., Schatzl, H.M., and Erfle, V. (2008). Therapeutic vaccination reduces HIV sequence variability. FASEB J. *22*, 437–444.

Kundu, S., Lockwood, J., Depledge, D.P., Chaudhry, Y., Aston, A., Rao, K., Hartley, J.C., Goodfellow, I., and Breuer, J. (2013). Next-generation whole genome sequencing identifies the direction of norovirus transmission in linked patients. Clin. Infect. Dis. *57*, 407–414.

Lindesmith, L.C., Brewer-Jensen, P.D., Mallory, M.L., Yount, B., Collins, M.H., Debbink, K., Graham, R.L., and Baric, R.S. (2019). Human norovirus epitope D plasticity allows escape from antibody immunity without loss of capacity for binding cellular ligands. J. Virol. *93*, e01813–e01818.

Lindesmith, L.C., Costantini, V., Swanstrom, J., Debbink, K., Donaldson, E.F., Vinjé, J., and Baric, R.S. (2013). Emergence of a norovirus GII. 4 strain correlates with changes in evolving blockade epitopes. J. Virol. *87*, 2803–2813.

Lindesmith, L.C., Mallory, M.L., Debbink, K., Donaldson, E.F., Brewer-Jensen, P.D., Swann, E.W., Sheahan, T.P., Graham, R.L., Beltramello, M., and Corti, D. (2018). Conformational occlusion of blockade antibody epitopes, a novel mechanism of GII. 4 human norovirus immune evasion. MSphere *3*, e00518-17.

Malm, M., Uusi-Kerttula, H., Vesikari, T., and Blazevic, V. (2014). High serum levels of norovirus genotype–specific blocking antibodies correlate with protection from infection in children. J. Infect. Dis. *210*, 1755–1762.

Moore, M.D., Bobay, B.G., Mertens, B., and Jaykus, L.A. (2016). Human norovirus Aptamer Exhibits high degree of target conformation-dependent binding similar to that of receptors and discriminates Particle functionality. mSphere *1*, e00298-16.

Nasheri, N., Petronella, N., Ronholm, J., Bidawid, S., and Corneau, N. (2017). Characterization of the genomic diversity of norovirus in linked patients using a metagenomic deep sequencing approach. Front. Microbiol. *8*, 73.

Nishijima, N., Marusawa, H., Ueda, Y., Takahashi, K., Nasu, A., Osaki, Y., Kou, T., Yazumi, S., Fujiwara, T., and Tsuchiya, S. (2012). Dynamics of hepatitis B virus quasispecies in association with nucleos (t) ide analogue treatment determined by ultra-deep sequencing. PLoS One *7*, e35052.

Ponsel, D., and Bruss, V. (2003). Mapping of amino acid side chains on the surface of hepatitis B virus capsids required for envelopment and virion formation. J. Virol. *77*, 416–422.

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., Van Der Spoel, D., et al. (2013). Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics *29*, 845–854.

Rambaut, A., and Drummond, A. (2010). FigTree V1. 3.1 Institute of Evolutionary Biology (University of Edinburgh).

Sabrià, A., Pintó, R.M., Bosch, A., Quer, J., Garcia-Cehic, D., Gregori, J., Dominguez, A., Carol, M., Sala-Farré, M.-R., and Guix, S. (2018). Characterization of intra-and inter-host norovirus P2 genetic variability in linked individuals by amplicon sequencing. PLoS One *13*, e0201850.

Siebenga, J.J., Lemey, P., Pond, S.L.K., Rambaut, A., Vennema, H., and Koopmans, M. (2010). Phylodynamic reconstruction reveals norovirus GII. 4 epidemic expansions and their molecular determinants. PLoS Pathog. *6*, e1000884.

Sobolev, O.V., Afonine, P.V., Moriarty, N.W., Hekkelman, M.L., Joosten, R.P., Perrakis, A., and Adams, P.D. (2020). A global ramachandran score identifies protein structures with unlikely stereochemistry. Structure *28*, 1249–1258 e2.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. *4*, vey016.

Sukhrie, F.H., Siebenga, J.J., Beersma, M.F., and Koopmans, M. (2010). Chronic shedders as reservoir for nosocomial transmission of norovirus. J. Clin. Microbiol. *48*, 4303–4305.

Töpfer, A., Zagordi, O., Prabhakaran, S., Roth, V., Halperin, E., and Beerenwinkel, N. (2013). Probabilistic inference of viral quasispecies subject to recombination. J. Comput. Biol. *20*, 113–123.

van Beek, J., De Graaf, M., Smits, S., Schapendonk, C.M., Verjans, G.M., Vennema, H., Van Der Eijk, A.A., Phan, M.V., Cotten, M., and Koopmans, M. (2017). Whole-genome next-generation sequencing to study within-host evolution of norovirus (NoV) among immunocompromised patients with chronic NoV infection. J. Infect. Dis. *216*, 1513–1524.

Wu, Q., Peng, Z., Zhang, Y., and Yang, J. (2018). COACH-D: improved protein–ligand binding sites prediction with refined ligand-binding poses through molecular docking. Nucleic Acids Res. *46*, W438–W442.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nat. Methods *12*, 7.

Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

Zakikhany, K., Allen, D.J., Brown, D., and Iturriza-Gómara, M. (2012). Molecular evolution of GII-4 Norovirus strains. PLoS One *7*, e41625.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PCR		
RNA extraction kit	Promega, Abbott molecular, USA	AS8500
Superscript III	Invitrogen, USA	12574026
PCR master mix	New England Biolabs, Germany	M0531L
Deposited data		
Sample specific sequences	Individual patient samples	[GenBank]: [MW363174] to [MW363198]
Software and algorithms		
InDelFixer	https://github.com/cbg-ethz/InDelFixer	V 1.1
QuasiRecomb	https://github.com/cbg-ethz/QuasiRecomb/releases	V 1.2
PHYLIP	https://evolution.genetics.washington.edu/phylip.html	v3.698
PROTDIST module	https://evolution.genetics.washington.edu/phylip.html	V 3.698
CODEML module	http://abacus.gene.ucl.ac.uk/software/paml.html	V 4.7
Bayesian Markov Chain Monte Carlo (MCMC) beast	https://beast.community	V 1.8.4
FIGTREE	https://beast.community/figtree	V 1.3.1
ITASSER	https://zhanglab.dcmb.med.umich.edu/I-TASSER/	V 5.0
ModRefiner	https://zhanglab.dcmb.med.umich.edu/ModRefiner/	V 1.0
Chiron tools	https://dokhlab.med.psu.edu/chiron	V 1.0
PROSA-Web	https://prosa.services.came.sbg.ac.at/prosa.php	V 1.0
Mol probity	http://molprobity.biochem.duke.edu	V 1.0
ConSurf	https://consurf.tau.ac.il	V 1.0
GROMACS	http://www.gromacs.org	V 4.5.5
COACH-D	https://yanglab.nankai.edu.cn/COACH-D/	V 1.0
Chimera	https://www.cgl.ucsf.edu/chimera/	V 1.13.1
Pymol	www.pymol.org	V 2.4
Pubchem	https://pubchem.ncbi.nlm.nih.gov	Accessed in 2018
R software	https://www.r-project.org	V 3.5.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, dieter.hoffmann (dieter.hoffmann@tum.de)

Materials availability

We used the sequence data as mentioned below. This study has not generated new reagents.

Data and code availability

- Sample-specific consensus sequences [have been deposited at GenBank and are publicly available as of the date of publication. Accession numbers are GenBank: MW363174, GenBank: MW363175, GenBank: MW363176, GenBank: MW363177, GenBank: MW363178, GenBank: MW363179, GenBank: MW363180, GenBank: MW363181, GenBank: MW363182, GenBank: MW363183, GenBank: MW363184, GenBank: MW363185, GenBank: MW363186, GenBank: MW363187, GenBank: MW363188, GenBank: MW363189, GenBank: MW363190, GenBank: MW363191,GenBank:

MW363192, GenBank: MW363193, GenBank: MW363194, GenBank: MW363195, GenBank: MW363196, GenBank: MW363197, GenBank: MW363198.

- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Sequential stool samples from six immune-compromised patients tested positive for norovirus RNA by an accredited diagnostic RT-PCR assay with Ct < 34 were collected at the Institutes of Virology of the Technical University of Munich and the University Hospital Tübingen, Germany. Three additional stool samples from patients with acute NoV infection, serving as controls, were collected. Stool specimens were stored at −80°C. The patients' demographic and clinical characteristics are presented in Table 1. Patients NoV positive for longer than 30 days were considered chronically infected. Among these five patients, four had undergone a stem cell transplantation and were immunocompromised, one patient was HIV-infected (Table 1). After our last sequenced sample of patient 1 one more sample was low positive in PCR, thereafter 4 NoV negative stool samples are documented, indicating cure. In parallel 4 sequential stool samples of patient 3 were NoV negative, also indicating cure, and 3 cleared their infections shortly after the last samples were collected. We have no follow up of patients 2 and 5 after our last sequenced samples. Patient 5 tested NoV positive for more than a year before our first sequenced sample. His advanced HIV infection makes prolonged NoV replication after the second sample we sequenced likely. Patient 4 stayed NoV positive for at least 6 months after our last sequenced sample, documented by 6 sequential stool samples. Patient 6 tested positive for 26 days. Patients 7–9 with acute, self-limiting infection were only symptomatic for a few days with only one NoV positive sample available. Sequential serum samples from patients 1–4 were tested for NoV-specific antibodies, no sera were available from the other patients.

METHOD DETAILS

RNA extraction and PCR

Viral RNAs were extracted from 10% stool suspensions in PBS with the mSample Preparation System DNA and m2000sp liquid handler (Promega, Abbott molecular, USA). cDNA was reverse-transcribed with Superscript III (Invitrogen, USA) and the capsid gene was amplified with barcoded primers and the Phusion High-fidelity PCR master mix (New England Biolabs, Germany). Following an initial denaturation at 95°C for 30s, the samples were subjected to 35 cycles of 10s denaturation at 95°C, 30s annealing at 60°C, 30s extension at 72°C, and a final extension at 72°C for 10 min. The PCR products were purified and subjected to PacBio NGS at Eurofins GATC Biotech (Germany).

Antibody quantitation by ELISA

NoV-specific antibody levels in serum samples from chronically infected patients 1–4 were quantitated as described previously (Afridi et al., 2019). Briefly, NoV capsid gene sequences were amplified from sequential stool samples and individually expressed in *E. coli* to obtain recombinant NoV capsid protein. Individual recombinant capsid proteins were coated on ELISA plates and incubated with serial autologous serum samples collected from the same patient. After thorough washing, bound antibodies were detected with peroxidase-labeled protein A.

Sequencing and generation of consensus sequences

The PacBio DNA library was constructed starting with 1μg PCR product and the PacBio standard 2 kb template preparation protocol. Our 25 samples were sequenced on the PacBio RSII platform in 5 SMRT (single molecule real time sequencing) cells. The CCS (circular consensus sequencing) reaction driven by C2 polymerase was recorded in 240 min windows. Raw sequence data were processed with PacBio's SMRT analysis software version 5.3.0 (Eid et al., 2009). The CCS reads making three or greater full passes around the closed loop SMRT bell amplicon were retained (~99% accuracy). Subsequently, CCS read files were obtained in fastq format and loaded into the Quiver consensus calling framework (Chin et al., 2013). Error-connected sequences are generated for each patient sample.

Quasispecies reconstruction

Quality clipping at both ends of the CCS reads with a PHRED threshold of 30 was performed using InDelFixer software (https://github.com/cbg-ethz/InDelFixer). The adapters were removed, and approximately matching

regions were found in the reference genome using k-mer matching. Subsequently, the reads were pairwise aligned to the reference sequence using a sensitive Smith–Waterman algorithm. Reads with three or more consecutive N base calls and read length below 300 bp were discarded. Viral quasispecies were reconstructed using the assembly tool QuasiRecomb 1.2 (https://github.com/cbg-ethz/QuasiRecomb/releases) (Töpfer et al., 2013). QuasiRecomb implements a hidden Markov model to infer viral quasispecies from deep-coverage NGS data using expectation maximization (EM) algorithm for maximum a posteriori (MAP) parameter estimation.

Sequence analysis

Genetic complexity was analyzed using Shannon entropy (Sn) (Domingo et al., 2006; Nishijima et al., 2012). Sn can vary from 0 (no complexity) to 1 (maximum complexity). Genetic distance was determined with the DNADIST and PROTDIST modules from the PHYLIP Package v3.698 (Felsenstein, 1993). The resulting values were corrected for alignment length and elapsed days between sequential samples, thus representing substitutions/site/1000 days. The values were scaled up to per 1000 days for better comparability. The dN/dS ratios between nonsynonymous (dN) and synonymous (dS) mutations were computed using the Nei-Gojobori model. dN/dS > 1 indicates positive selection whereas dN/dS < 1 suggests negative selection. In addition, to determine positively selected sites within the VP1 region, Naive Empirical Bayes (NEB) and Bayes Empirical Bayes (BEB) methods available in the CODEML module of PAML 4.7 software package were employed (Yang, 2007).

Phylogenetic analysis of NoV sequences

We estimated intra-host viral evolutionary (divergence) rate using the Bayesian Markov Chain Monte Carlo (MCMC) method as implemented in BEAST 1.8.4 (Suchard et al., 2018). The dominant quasispecies with the prevalence >1% were analyzed under a strict molecular clock. The Hasegawa–Kishono–Yano (HKY) nucleotide substitution model with gamma-distributed rate heterogeneity among sites for each patient was employed. Simulations were run for at least 10 million generations, and trees were sampled and stored after every 1000 generations. The first 10% of each MCMC runs were discarded as burn-in. To compute and visualize the sampled trees we used the program FigTree version 1.3.1 (Rambaut and Drummond, 2010).

Homology modeling, consurf and molecular dynamics simulation analysis

Homology-based 3D structural models of the capsid proteins from the patient samples were generated from the reference FASTA sequences using the I-TASSER 5.0 server (Yang et al., 2015). The models were improved by repeating the minimization procedure for twenty times using minimization server ModeRefiner (https://zhanglab.ccmb.med.umich.edu/ModRefiner/) and corrected the side chains with Chiron tools (https://dokhlab.med.psu.edu/chiron). The quality of the structures were assessed by PROSA-Web (https://prosa.services.came.sbg.ac.at/prosa.php). Ramachandran plots were analyzed on the molprobity server (http://molprobity.biochem.duke.edu) and their distribution was assessed by computing z-scores (Sobolev et al., 2020) of torsion angels of each amino acids and to predict the stereochemical quality of the modeled 3D structures. Z-scores <2 indicates good plot quality. We computed RMSD between the homology models and crystal structure of norovirus capsid proteins (PDB ID: 1IHM) using Pymol v2.7 software. Protein structures were visualized with Pymol v2.7 and the previously described epitopes A (294, 296, 298), D (391–395), E (407, 412, 413) and HBGA binding sites were mapped. Evolutionary residue conservation of all the structures was determined by Consurf server (https://consurf.tau.ac.il). Molecular dynamic simulations (MD) were done using GROMACS 4.5.5. The CHARMM27 force field is adapted during the simulation (Pronk et al., 2013). The protein structures were engaged in a cubic water box using the explicit TIP3 water model at a buffering distance of 1.2 nm. These complexes were subjected to minimization procedure with the steepest descent method for 2000 steps. All simulations were run in periodic boundary conditions with the NPT ensemble. The Berendsen coupling algorithm was used for this process, and the temperature was kept at 300 K with pressure at 1 bar. Electrostatic interactions with an interpolation of order 4 and grid spacing of 0.12 nm were calculated with the Particle-Mesh Ewald method and all bonds were constrained using the LINCS algorithm. The time step for all simulations was set at 2 fs, and 20 ns of MD simulation was performed. MD trajectories were analyzed with built-in GROMACS tools (Pronk et al., 2013) Root-mean-square fluctuation (RMSF) was applied to quantify the fluctuation of residues within epitopes and HBGA binding pocket of the protein structures for 20ns. GII.4 sequences from chronic infections with l positively selected sites were analyzed. We extracted RMSF values of the epitopes and P2 domain of VP1 capsid protein. We considered if the residue >50% SASA, the residue is classified as surface accessible and extracted RMSF values for these residues.

Molecular docking of protein models to the HBGA oligosaccharides

Molecular docking between the protein models of human blood group antigen (HBGA) Lewis[a] molecule PubChem database (https://pubchem.ncbi.nlm.nih.gov) and 3D model of the capsid proteins was conducted with COACH-D web server (Wu et al., 2018).

QUANTIFICATION AND STATISTICAL ANALYSIS

NoV capsid-specific antibodies were quantitated as described previously (Afridi et al., 2019).

Positively selected amino acid sites were statistically assessed with the likelihood-ratio test. To certify the difference between the RMSF values of the P2 domain, SASA residues and epitopes we performed Wilcoxon rank-sum test in the R software. p values < 0.05 were considered significant.