

DISSERTATION

Charakterisierung und klinische Relevanz von Virusvarianten  
für den chronischen Verlauf einer Hepatitis E

Characterization and clinical relevance of virus variants for a  
chronic infection with hepatitis E virus

zur Erlangung des akademischen Grades  
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Paula Kristina Elisabeth Biedermann

Erstbetreuung: Prof. Dr. Jörg Hofmann

Datum der Promotion: 28.02.2025



## Inhaltsverzeichnis

Tabellenverzeichnis.....	iii
Abbildungsverzeichnis.....	iv
Abkürzungsverzeichnis.....	v
Zusammenfassung .....	1
Abstract .....	2
1. Einleitung.....	3
1.1 Entdeckung und Verbreitung des Hepatitis-E-Virus .....	3
1.2 Aufbau des Hepatitis-E-Virus .....	5
1.3 Klinische Manifestation und Übertragungswege.....	6
1.4 Chronische Verläufe und Therapiemöglichkeiten .....	6
1.5 Stand der Forschung.....	8
1.6 Ziel der Arbeit und Fragestellung .....	9
2. Methodik.....	11
2.1 Patientenkohorte .....	11
2.2 Ethikvotum.....	11
2.3 RNA-Extraktion.....	12
2.4 Quantitative Echtzeit-Reverse Transkriptase-PCR.....	12
2.5 Synthese der cDNA .....	12
2.6 Genotypisierung .....	13
2.7 Amplifikation der hypervariablen Region .....	13
2.8 Bioinformatische Auswertung .....	15
2.9 Analyse der Sequenzen und statistische Auswertung .....	15
3. Ergebnisse .....	16
4. Diskussion.....	22
4.1 Kurze Zusammenfassung der Ergebnisse.....	22
4.2 Interpretation der Ergebnisse .....	22

---

4.3	Einbettung der Ergebnisse in den bisherigen Forschungsstand.....	23
4.4	Stärken und Schwächen der Studie .....	25
4.5	Implikationen für Praxis und zukünftige Forschung .....	26
5.	Schlussfolgerungen .....	28
	Literaturverzeichnis .....	29
	Eidesstattliche Versicherung .....	35
	Anteilerklärung an den erfolgten Publikationen.....	36
	Druckexemplar der Publikation.....	37
	Lebenslauf.....	48
	Komplette Publikationsliste.....	49
	Danksagung .....	50

## Tabellenverzeichnis

Tabelle 1: Erster Ansatz der „nested“ PCR zur Amplifikation der HVR.....	14
Tabelle 2: Zweiter Ansatz der „nested“ PCR zur Amplifikation der HVR .....	14
Tabelle 3: Analyse der Aminosäurezusammensetzung der Sequenzen und Vergleich zwischen Sequenzen ohne Insertionen oder Deletionen (Indels) und Sequenzen mit Indels.....	20
Tabelle 4: Analyse potenzieller posttranslationaler Modifikationsstellen und Aminosäureeigenschaften der Sequenzen sowie Vergleich zwischen Sequenzen ohne Insertionen oder Deletionen (Indels) und Sequenzen mit Indels. ....	21

## **Abbildungsverzeichnis**

Abbildung 1: Aminosäuresequenzen der Insertionen und Deletionen in den analysierten Sequenzen in einem Alignment. ....	18
------------------------------------------------------------------------------------------------------------------------	----

## Abkürzungsverzeichnis

cDNA	complementary deoxyribonucleic acid
DNA	deoxyribonucleic acid
GSTA1	Glutathione S-transferase alpha 1
HBV	Hepatitis-B-Virus
HCC	hepatozelluläres Karzinom (hepatocellular carcinoma)
HEV	Hepatitis-E-Virus
HVR	hypervariable region
IfSG	Infektionsschutzgesetz
IgG	Immunglobulin G
Indel	Insertion oder Deletion
NGS	Next-Generation Sequencing
ORF	open reading frame
PCR	polymerase chain reaction
RT-qPCR	quantitative real-time reverse-transcription PCR
RdRp	RNA-dependent RNA-Polymerase
RNA	ribonucleic acid
RNasen	Ribonukleasen
RPL18	Ribosomal Protein L18
UTR	untranslated region

## Zusammenfassung

Chronische Infektionen mit dem Hepatitis-E-Virus (HEV) stellen aufgrund eingeschränkter Therapiemöglichkeiten eine Herausforderung in der medizinischen Behandlung dar. Der genaue Entstehungsmechanismus hinter dem Übergang von einer akuten zu einer chronischen Infektion ist noch nicht bekannt. Das Ziel dieser Arbeit war es, Veränderungen im viralen Genom zu identifizieren, die möglicherweise die Entstehung einer chronischen Infektion begünstigen. Vorangegangene Arbeiten liefern bereits Hinweise, dass die hypervariable Region (HVR) des viralen Genoms und der Einbau von Genfragmenten (Insertionen) eine Rolle in dem Prozess der Chronifizierung spielen könnte.

Daher wurde die HVR des viralen Genoms aus den Blutproben von 16 HEV-infizierten Patienten und Patientinnen isoliert und mittels „Next-Generation Sequencing“ untersucht, um genetische Veränderungen in Form von Insertionen zu identifizieren und deren Eigenschaften zu analysieren. Dabei handelte es sich um neun akut und sieben chronisch HEV-infizierte Patienten und Patientinnen. Acht Insertionen in der hypervariablen Region von drei chronisch HEV-infizierten und einem akut HEV-infizierten Patienten und Patientinnen konnten dabei nachgewiesen werden. Die Insertionen waren sowohl viralen als auch humanen Ursprungs. Es ließen sich Insertionen nachweisen, welche die höchste Sequenzübereinstimmung mit Fragmenten der humanen Gene Ahnak nucleoprotein, Ribosomales Protein L18 und Glutathione S-Transferase Alpha zeigten. Des Weiteren fielen bei der Untersuchung Deletionen in der HVR auf.

Bei der weiteren Analyse der Sequenzen zeigten sich signifikante Unterschiede in der Aminosäurezusammensetzung der HEV-Varianten mit Insertionen oder Deletionen gegenüber Sequenzen ohne Insertionen oder Deletionen. Es zeigte sich zudem eine signifikante Erhöhung der potenziellen Stellen für posttranslationale Modifikationen wie Acetylierungen, Ubiquitinierungen und Methylierungen in Sequenzen mit Insertionen oder Deletionen in der HVR der HEV-Genome.

Durch diese Eigenschaften könnten Insertionen und Deletionen einen entscheidenden Vorteil im Prozess der Chronifizierung der HEV-Infektion darstellen. Daher könnte die Untersuchung hinsichtlich dieser Veränderungen im viralen HEV-Genom in der Zukunft eine Identifizierung von Individuen mit einem erhöhten Risiko für einen schweren Verlauf ermöglichen oder einen Angriffspunkt einer zielgerichteten, individualisierten Therapie darstellen.

## Abstract

The treatment of chronic infections caused by the Hepatitis E virus (HEV) poses a challenge for the medical provider due to limited treatment options. The mechanism behind the progress from an acute to a chronic infection is still unclear. The aim of this study was to possibly identify changes in the viral genome fostering chronic infection. Previous work indicates that the hypervariable region (HVR) of the viral genome and the incorporation of gene fragments into the viral genome plays a role in the process of chronification.

The HVR of the viral genome was isolated using blood samples of 16 acutely and chronically HEV-infected patients and analysed using Next-Generation Sequencing to identify genetic changes like insertions and analyse them regarding their characteristics. The study included nine acutely and seven chronically HEV-infected patients. Eight inserted gene fragments (insertions) in the hypervariable region of three chronically infected patients and one acutely infected patient were found. The insertions were of viral and human origin. The insertions of human origin derived from the human genes coding for Ahnak nucleoprotein, Ribosomal Protein L18 and Glutathione S-Transferase Alpha 1. In addition, deletions were found in the HVR.

A significant change in the composition of amino acids of HEV variants with insertions and deletions compared to variants without insertions or deletions was detected. Observation revealed a significant increase in potential sites for post-translational modifications like acetylation, ubiquitination, and methylation for sequences with insertions or deletions compared to sequences without insertions or deletions in the HVR of the HEV-genome.

These characteristics could represent a critical viral advantage in the process of chronification of an HEV-infection. The screening the viral HEV-genome regarding these viral changes could therefore be beneficial in the future to identify patients at an increased risk for a severe course of disease or could be a potential target for an individualized treatment approach.

## 1. Einleitung

### 1.1 Entdeckung und Verbreitung des Hepatitis-E-Virus

Hepatitisviren sind die wichtigsten Verursacher von viralen Leberentzündungen. Im engeren Sinne zählen dazu die Hepatitisviren A bis E. Die meisten Neuinfektionen weltweit verursacht mit über 100 Millionen Fällen jährlich das Hepatitis-A-Virus, von denen die meisten Infektionen allerdings asymptomatisch verlaufen (1). Eine chronische Hepatitis wird dagegen vorrangig durch das Hepatitis-B-Virus (HBV) und Hepatitis-C-Virus verursacht (2, 3). Gegenwärtig leiden weltweit über 300 Millionen Menschen an einer viralen, chronischen Hepatitis und mehr als 1,5 Millionen Menschen versterben pro Jahr an den Folgen ihrer Erkrankung (4). Als Folge der Chronifizierung treten vermehrt Leberzirrhosen und das hepatozelluläre Karzinom (HCC) auf. Weiterhin ist das Hepatitis-D-Virus bekannt. Eine Infektion mit dem Hepatitis-D-Virus tritt ausschließlich gemeinsam mit einer Infektion mit dem Hepatitis-B-Virus auf (5). Durch die zusätzliche Infektion mit dem Hepatitis-D-Virus erhöht sich insbesondere bei chronisch HBV-infizierten Personen das Risiko einer Leberzirrhose und eines HCCs.

Die Hepatitisviren A, B und C sind vergleichsweise gut erforscht. Das Hepatitis-E-Virus (HEV) spielte hingegen sowohl in der wissenschaftlichen Diskussion als auch in der öffentlichen Wahrnehmung lange eine untergeordnete Rolle. Entdeckt wurde das Hepatitis-E-Virus in den frühen 1980er Jahren. Im Jahr 1980 beschrieb der indische Gastroenterologe Mohammad S. Khuroo eine Hepatitisepidemie im Kaschmirtal in Indien zwischen 1978 und 1979 und mutmaßte, dass der zugrundeliegende Krankheitserreger ein neuartiges Hepatitisvirus sei (6). Dem Team um den russischen Virologen Mikhail S. Balayan gelang schließlich im Jahr 1983 der immunelektronenmikroskopische Nachweis des bis dahin unbekanntes Pathogens. Balayan war zuvor auf den Ausbruch von Leberentzündungen unbekannter infektiöser Ursache bei sowjetischen Truppen in Afghanistan aufmerksam geworden (7).

Wenngleich das Hepatitis-E-Virus seit gut vierzig Jahren bekannt ist, sind die Kenntnisse über dessen Verbreitung eher neueren Datums. Aktuellere Daten zeigen, dass die Anzahl der HEV-Infektionen lange unterschätzt wurde. So infizieren sich laut Weltgesundheitsorganisation ungefähr 20 Millionen Menschen weltweit jährlich (8). Eine Me-

taanalyse aus dem Jahr 2020 geht von einer weltweiten Seroprävalenz von anti-HEV Immunglobulin G (IgG) von etwa zwölf Prozent aus (9). Die Daten zeigen also, dass Infektionen mit dem Hepatitis-E-Virus deutlich häufiger auftreten als lange Zeit angenommen wurde.

In Deutschland handelt es sich bei einer HEV-Infektion um eine meldepflichtige Erkrankung. Dabei ist laut Robert Koch-Institut „der Krankheitsverdacht, die Erkrankung sowie der Tod an akuter Virushepatitis sowie gemäß § 7 Abs. 1 IfSG der direkte oder indirekte Nachweis von HAV, soweit er auf eine akute Infektion hinweist“ namentlich meldepflichtig (10). Für das Jahr 2022 betrug die übermittelte Zahl an HEV-Infektionen 3.513 Fälle. Im gleichen Jahr wurden 710 Hepatitis-A-Virus, 16.974 Hepatitis-B-Virus und 8.015 Hepatitis-C-Virus Nachweise gemeldet. Für das Jahr 2023 zeigte sich ein Anstieg der HEV-Infektionen auf 4.658 gemeldete Fälle (11). Allerdings ist von einer großen Untererfassung der HEV-Infektionen auszugehen, da die Prävalenz von HEV-Antikörpern in der Bevölkerung deutlich höher ist als von den Infektionszahlen her anzunehmen wäre. Eine Studie aus dem Jahr 2012 ergab im Mittel für Deutschland eine Seroprävalenz von 16,8 Prozent über alle Altersgruppen. Die untersuchte Gruppe umfasste Personen im Alter von 18 bis 79 Jahren, die ihren Wohnsitz in Deutschland hatten und fließend Deutsch sprachen. Personen mit Migrationshintergrund waren daher in der Untersuchung unterrepräsentiert (12). In einer weiteren, im Jahr 2019 veröffentlichten Metaanalyse wurde eine Prävalenz von anti-HEV IgG zwischen 11 und 28,3 Prozent in der deutschen Bevölkerung beschrieben. Bei Betrachtung der Prävalenz in anderen europäischen Ländern fallen teilweise noch höhere Werte auf, wobei die verschiedenen, eingesetzten Testsysteme unterschiedlich präzise sein können und daher die Zahlen eingeschränkt vergleichbar sind. In Frankreich zum Beispiel wurden landesweit Seroprävalenzen von 22,4 Prozent detektiert. In einigen Regionen zeigte sich sogar, dass bei mehr als der Hälfte der untersuchten Bevölkerung ein anti-HEV IgG nachweisbar war. Auch in anderen Ländern wie Polen und Italien zeigten sich regional sehr hohe Seroprävalenzen (13).

Das Hepatitis-E-Virus wird aktuell in acht Genotypen unterteilt, von denen primär die Genotypen 1 bis 4 für Erkrankungen beim Menschen verantwortlich sind. Genotyp 1 und 2 kommen überwiegend in Entwicklungsländern vor. In Europa werden nahezu ausschließlich HEV-Infektionen des Genotyps 3 nachgewiesen. Genotyp 4 findet sich in

asiatischen Ländern, vorwiegend in der Volksrepublik China (14). Von Genotyp 3 sind zahlreiche Subgenotypen bekannt. In Deutschland ist mit ca. 67,3 Prozent 3c der häufigste Subgenotyp gefolgt von 3f und 3e (15).

## 1.2 Aufbau des Hepatitis-E-Virus

Die bekannten Hepatitis-Viren unterscheiden sich im Aufbau grundlegend. Die gemeinsame Betrachtung dieser unterschiedlichen Viren erfolgt aufgrund ihrer primären Hepatotropie, d.h. sie infizieren in erster Linie die Hepatozyten. Während sich das Hepatitis-B-Virus aus einzel- und doppelsträngiger Desoxyribonukleinsäure („deoxyribonucleic acid“ – DNA) aufbaut, besteht das Genom von Hepatitis-A-Virus, Hepatitis-C-Virus und Hepatitis-E-Virus aus einzelsträngiger Ribonukleinsäure („ribonucleic acid“, RNA). Das Hepatitis-E-Virus ist ein unbehülltes Einzel(+)-Strang-RNA-Virus (16). Es gehört zur Familie der Hepeviridae und zur Gattung des Paslahepevirus. Das Genom des Hepatitis-E-Virus ist ca. 7,2 Kilobasenpaare lang und enthält drei offene Leserahmen („open reading frame“, ORF) sowie am 5'- und 3'-Ende zwei untranslatierte Regionen („untranslated regions“, UTRs). Am 3'-Ende befindet sich außerdem ein Poly-A-Schwanz. ORF1 kodiert u.a. für die Polymerase des Virus („RNA-dependent RNA polymerase“, RdRp) und enthält die hypervariable Region („hypervariable region“, HVR), die sich durch hohe genetische Diversität auszeichnet. ORF2 kodiert für das Kapsidprotein (17). ORF3 spielt mutmaßlich eine Rolle in der Virionenfreesetzung und Interaktion mit dem Wirtsimmunsystem (18).

RNA-Viren wie das Hepatitis-E-Virus existieren als Quasispezies in einem Wirt. Dies bedeutet, dass eine Vielzahl von Virusvarianten im Wirt gleichzeitig vorhanden ist. Diese Vielfalt an Varianten entsteht durch die fehlende Korrekturlesefunktion der viralen Polymerase, woraus sich eine hohe Mutationsrate ergibt. Die Fitness der einzelnen Varianten entscheidet über ihre Durchsetzung in der viralen Population (19). So kann das Virus sich an verschiedene Anforderungen anpassen, da sich bestimmte Varianten im Zusammenspiel mit dem Immunsystem des Wirts als besonders vorteilhaft herausstellen können. Dies könnte insbesondere bei prolongierten Verläufen eine Rolle spielen.

### 1.3 Klinische Manifestation und Übertragungswege

Eine HEV-Infektion verläuft meistens klinisch inapparent und ist in der Regel selbstlimitierend (20). Wenn sich eine symptomatische Hepatitis E entwickelt, geht diese mit typischen Symptomen einer viralen Hepatitis wie Ikterus, Übelkeit, Erbrechen, Fieber, Diarrhoe und abdominellen Schmerzen einher (21). Dabei beträgt die Inkubationszeit zwei bis acht Wochen (22). Bei einem akuten Verlauf besteht eine hohe Hospitalisierungsrate von 74,5 Prozent (23). Die Erkrankung wird laborchemisch von einer Erhöhung der Transaminasen und Cholestaseparameter begleitet (24). In seltenen Fällen (ca. 0,5 bis 4 Prozent) kann sich im Rahmen einer akuten Infektion ein akutes Leberversagen entwickeln (25).

Allerdings unterscheiden sich die Verläufe je nach Genotyp des Virus. Die HEV-Genotypen 1 und 2 werden fäkal-oral übertragen (14). Bei schwangeren Frauen ist eine Infektion mit diesen Genotypen besonders gefährlich, da ein akutes Leberversagen gehäuft auftritt und in Folge die Mortalitätsrate bis zu 30 Prozent beträgt (26) (27).

Die erhöhte Inzidenz und Schwere der Infektion bei schwangeren Frauen ist für Genotyp 3 nicht beschrieben. Die Verläufe bei Genotyp 3 verlaufen häufiger asymptomatisch, wobei 5 bis 33 Prozent der Patienten und Patientinnen Symptome in Form von Ikterus entwickeln oder sich laborchemisch erhöhte Leberenzyme zeigen (28).

Genotyp 3 wird vorwiegend zoonotisch übertragen und Schweine stellen das Hauptreservoir dar (24). Als Hauptübertragungsweg gilt der Verzehr von nicht ausreichend erhitzten Schweinefleischprodukten (28). Aber auch unter Vegetariern sind hohe Seroprävalenzen beschrieben worden. Als möglicher Übertragungsweg gilt u.a. die Düngung von Agrarflächen mit Gülle (29). Daneben sind Übertragungen über Blutprodukte bekannt (30). Eine Übertragung durch eine Lebertransplantation bei einem Spender mit okkulten HEV-Infektion ist ebenfalls dokumentiert (31).

### 1.4 Chronische Verläufe und Therapiemöglichkeiten

Bis zum Jahr 2008 waren keine chronischen Verläufe einer Hepatitis E publiziert worden. In jenem Jahr berichtete Kamar et al. (32) von acht immunsupprimierten, organtransplantierten Patienten und Patientinnen mit einer chronischen HEV-Infektion. Eine chronische HEV-Infektion wird durch den Nachweis von HEV-RNA im Blut für mehr als drei Monate definiert (33). Seit dem Jahr 2008 sind zunehmend chronische Verläufe bei

immunsupprimierten Patienten und Patientinnen detektiert worden (34). Dabei handelt es sich vorwiegend um Patienten und Patientinnen, die eine Organtransplantation erhalten haben. Eine Übersichtsarbeit aus dem „Journal of Hepatology“ aus dem Jahr 2022 geht davon aus, dass ungefähr zwei Drittel aller mit HEV-infizierten organtransplantierten Patienten und Patientinnen eine chronische Infektion entwickeln (34). Insbesondere eine immunsuppressive Medikation mit Tacrolimus stellt dabei einen Risikofaktor für die Entwicklung einer chronischen Hepatitis E dar (35). Weiterhin wurde auch zum Beispiel bei einem Patienten mit akuter lymphoblastischer Leukämie und Stammzelltransplantation die Reaktivierung und Persistenz einer HEV-Infektion beobachtet (36).

Chronische Verläufe zeigen sich klinisch häufig asymptomatisch und fallen nur durch leicht erhöhte Werte der Leberenzyme auf (32). Die chronische Entzündung kann die Entstehung einer Fibrose und im weiteren Verlauf eine Leberzirrhose begünstigen (35). Von den chronisch HEV-infizierten Personen entwickeln ca. 10 Prozent eine Leberzirrhose (32, 35). Weiterhin zeigte sich histopathologisch eine destruktive Cholangitis bei einigen Patientinnen und Patienten (37). Auch verschiedene extrahepatische Manifestationen insbesondere neurologische Erkrankungen wie das Guillain-Barré-Syndrom sind beobachtet worden (38, 39).

Als Erstlinientherapie ist eine Reduktion der immunsuppressiven Medikation empfohlen, soweit dies medizinisch vertretbar erscheint. In etwa einem Drittel der Fälle kann so die Elimination des Virus erreicht werden (40). Bei Versagen dieser Maßnahme steht bei chronischen Verläufen als off-label Therapie das Medikament Ribavirin zur Verfügung (41). Bei Ribavirin handelt es sich um ein Nukleosid-Analogon, welches u.a. virostatisch wirkt, indem es die virale Polymerase hemmt. Dabei liegt die Rate für anhaltendes Ansprechen auf die antivirale Therapie bei ca. 78 Prozent (42). Der Erfolg der Therapie wurde in dieser Metaanalyse mittels der „sustained virological response“ definiert. Hierfür durfte für mindestens drei Monate nach Therapieende keine HEV-RNA im Serum nachweisbar sein. Für therapie-refraktäre Patienten und Patientinnen sind die Behandlungsmöglichkeiten allerdings sehr eingeschränkt.

Die Häufung der chronischen Erkrankungsfälle in bestimmten Patientengruppen lässt die Frage nach den Faktoren aufkommen, die eine Chronifizierung begünstigen. Es ist bekannt, dass patientenbezogene Faktoren eine Rolle spielen. Insbesondere die T-

Zellantwort scheint entscheidend für die Entwicklung einer chronischen Infektion zu sein (43-45). Die unzureichende Antwort der CD8+ T-Zellen trägt mutmaßlich entscheidend zu einer fehlenden viralen Elimination bei (46).

## 1.5 Stand der Forschung

Es ist davon auszugehen, dass bestimmte virale Eigenschaften an der Entwicklung einer chronischen HEV-Infektion beteiligt sind. Diese sind jedoch noch nicht hinreichend erforscht.

Bekannt sind Veränderungen im HEV-Genom, welche mit dem Therapieversagen von Ribavirin in Verbindung gebracht werden. Dabei handelt es um die Punktmutationen Y1320H, K1383N, D1384G, K1398R, V1479I, Y1587F and G1634R, welche in der Polymeraseregion des HEV-Genoms liegen (47-49). Diese sorgen nach aktuellem Kenntnisstand nicht für eine Ribavirinresistenz *in vitro*, sondern verändern die virale Fitness (47).

Bezüglich der Entwicklung einer chronischen HEV-Infektion zeigte eine weitere Untersuchung, dass in der akuten Phase einer Infektion eine erhöhte Heterogenität in der für das Kapsidprotein kodierenden ORF2 Region mit der Entwicklung einer chronischen Infektion assoziiert ist (43). Bekannt ist auch, dass nicht-strukturelle Virusgene, die nicht an der Replikation beteiligt sind, normalerweise in die Modulation der Immunantwort involviert sind (50). Daher sind bestimmte virale Regionen wie die HVR und das „macro domain“ des HEV-Genoms besonders interessant. Eine Studie beschreibt einen Zusammenhang zwischen der Entwicklung einer chronischen Infektion und der erhöhten Heterogenität der Quasispezies im ORF1 bei organtransplantierten Patienten und Patientinnen. Dabei zeigte sich eine erhöhte Heterogenität in der „macro domain“ (X-domain) und in der HVR bei Patienten und Patientinnen, die im Verlauf eine chronische HEV-Infektion entwickelten (51).

Daher ist die HVR eine Region von besonderem Interesse im viralen Genom. Diese Region ist gekennzeichnet durch eine hohe Mutationsrate und die höchste Variabilität im Genom (52). Es wird vermutet, dass dieser Abschnitt im Genom in Prozesse der viralen Adaptation und Interaktion mit dem Wirt durch Protein-Protein Interaktionen invol-

viert ist und eine Rolle im Prozess der Chronifizierung einer Infektion spielen könnte (53). In den HEV-Sequenzen von chronisch infizierten Patienten und Patientinnen wurden in dieser Region Insertionen detektiert, d.h. Genfragmente, welche zusätzlich in diese Region eingebaut wurden. Dabei wurden sowohl Fragmente humanen als auch viralen Ursprungs gefunden (54-59). Einzelne Untersuchungen zeigten signifikante Veränderungen der Aminosäurezusammensetzung sowie eine Erhöhung der potenziellen Stellen für posttranslationale Modifikationen der HVR-Sequenzen mit Insertionen im Vergleich zu Sequenzen ohne Insertionen (57, 58). Einige Studien konnten auch *in vitro* einen Replikationsvorteil der Sequenzen mit Insertionen gegenüber Sequenzen ohne Insertionen zeigen (56, 57).

## 1.6 Ziel der Arbeit und Fragestellung

Ziel der Dissertation war es, Virusvarianten von akut und chronisch HEV-Infizierten zu charakterisieren und so Faktoren zu identifizieren, die relevant für den chronischen Verlauf einer Hepatitis E sein könnten. Diese Arbeit ordnet sich damit in den Bereich der Grundlagenforschung ein.

Zu Beginn dieses Promotionsvorhabens wurden Punktmutationen in der Polymeraseregion (RdRp) des HEV-Genoms untersucht. Diese Mutationen in der RdRp scheinen insbesondere bei therapieresistenten Verläufen unter Ribavirin eine Rolle zu spielen. In einer ersten Publikation im Rahmen des Promotionsvorhabens wurden Patienten und Patientinnen mit einer chronischen Hepatitis E nach Vorbehandlung mit Rituximab im Rahmen einer Grunderkrankung und Behandlung mit Ribavirin hinsichtlich dieser Mutationen analysiert (60).

Im Verlauf der Untersuchungen rückte insbesondere die HVR in das Zentrum des Interesses. Die Datenlage zu der Häufigkeit und der Charakteristik der oben genannten Insertionen bei chronisch und akut HEV-infizierten Patienten und Patientinnen kann als noch unzureichend bezeichnet werden und die klinische und virologische Relevanz der Insertionen für den Verlauf einer chronischen HEV-Infektion ist noch unklar. Durch die Analyse von Virusvarianten der HVR, die aus Isolaten von Blutproben akut und chronisch mit HEV-infizierter Personen gewonnen wurden, sollten genauere Erkenntnisse

über virale Einflussfaktoren in der HVR, die den chronischen Verlauf einer Infektion begünstigen könnten, erlangt werden.

Um möglichst detailliert das Vorkommen von HEV-Varianten zu erfassen, wurde in der Forschungsgruppe zunächst eine geeignete Methode mittels „Next-Generation Sequencing“ (NGS) entwickelt, bei welcher eine robuste Polymerase-Kettenreaktion („polymerase chain reaction“, PCR) und eine sensitive Sequenziermethode zur Untersuchung der HVR etabliert wurden (61). Mithilfe dieser Methode sollten die Sequenzen der hypervariablen Region von akuten und chronisch mit HEV-infizierten Patienten und Patientinnen verglichen und auf das Vorkommen von Insertionen untersucht werden, um potenziell neue Insertionen zu identifizieren. Weiterhin sollten die Charakteristika der Sequenzen analysiert werden. Insbesondere galt es, Untersuchungen bezüglich Veränderungen der Aminosäurezusammensetzung und potenzieller Stellen für posttranslationale Modifikationen anzustellen, um vorherige Untersuchungen zu verifizieren oder zu widerlegen.

Die Fragestellungen der Arbeit waren:

- 1) Kommen Insertionen bei akut HEV-infizierten erkrankten Personen vor?
- 2) Können neue, unbekannte Insertionen identifiziert werden?
- 3) Was charakterisiert Sequenzen mit Insertionen, insbesondere hinsichtlich ihrer Aminosäurezusammensetzung und Möglichkeiten der posttranslationalen Modifikation?

## **2. Methodik**

Zur Beantwortung der Fragestellung sollte die hypervariable Region des HEV-Genoms mittels „Next-Generation Sequencing“ analysiert werden. In einem ersten Schritt wurde die RNA aus den Proben isoliert. Anschließend erfolgte die Synthese der komplementären DNA. Mit dieser konnte eine PCR durchgeführt werden, um den spezifischen Abschnitt des HEV-Genoms zu amplifizieren. Nach Aufreinigung und Quantifizierung des PCR-Produktes erfolgte sowohl eine konventionelle Sanger-Sequenzierung als auch eine Sequenzierung mittels „Next-Generation Sequencing“. Anschließend wurden eine bioinformatische Auswertung sowie eine Analyse der Eigenschaften der Sequenzen durchgeführt. Die Sequenzierung mittels „Next-Generation Sequencing“ ermöglicht dabei eine bessere Annäherung an die virale Vielfalt im Wirtsorganismus als eine Sanger-Sequenzierung.

### **2.1 Patientenkohorte**

Aus der Routinediagnostik des Entnahmezeitraumes vom 7. Februar 2017 bis 2. April 2019 wurden 16 Proben ausgewählt und in den Laboren des Robert-Koch-Institut untersucht. Es wurde angestrebt, eine ungefähr gleiche Anzahl von Proben von jeweils chronisch und akut HEV-infizierten Personen zu untersuchen. Ausschlaggebend für die Auswahl der Proben war zum einen eine möglichst hohe Viruslast und zum anderen die Verfügbarkeit einer ausreichenden Probenmenge aus der Routinediagnostik. Die Viruslast ist wichtig, da bei niedrigerer Viruslast keine ausreichende Menge an PCR-Produkten für die Sequenzierung generiert werden kann. Es handelte sich um neun Proben von akut und sieben Proben von chronisch HEV-infizierten Patientinnen und Patienten. Die 16 Proben wurden von 14 männlichen Patienten und 2 weiblichen Patientinnen entnommen. Der jüngste Patient war 24 Jahre und der älteste 90 Jahre alt.

### **2.2 Ethikvotum**

Die Untersuchungen wurden nach den Richtlinien der Deklaration von Helsinki durchgeführt. Ein positives Votum (EA1/367/16) der zuständigen Ethikkommission der Charité – Universitätsmedizin Berlin liegt vor. Von allen Personen lag eine schriftliche Einwilligung vor.

### **2.3 RNA-Extraktion**

Die RNA-Extraktion aus den Plasmaproben erfolgte durch ein geschlossenes Robotiksystem. Dabei handelte es sich um den QIAcube (Qiagen, Hilden, Deutschland) unter Benutzung des QIAamp Viral RNA mini Kit (Qiagen, Hilden, Deutschland). Der Prozess ist weitgehend automatisiert. Von den Serum- oder Plasmaproben wurden jeweils 140 µl manuell in Kunststoffröhrchen pipettiert. Im anschließenden automatisierten Prozess wurde die RNA mit Hilfe der Spin-Säulen-Extraktionsmethode gewonnen. Die Proben wurden in einem ersten Schritt lysiert, um Ribonukleasen (RNasen) zu inaktivieren und die Isolierung der intakten viralen RNA zu ermöglichen. Die RNA bindet an eine Siliziumoxidmembran. Die unerwünschten, ungebundenen Substanzen wurden in zwei Waschschritten mittels zweier verschiedener Waschpuffer ausgewaschen. Im Anschluss wurde die gewonnene RNA in einem RNasen-freien Puffer gelöst.

### **2.4 Quantitative Echtzeit-Reverse Transkriptase-PCR**

Zur Quantifizierung der gewonnenen Menge an HEV RNA wurde eine quantitative Real-Time Reverse Transkriptase PCR (quantitative real-time reverse-transcription PCR, RT-qPCR) durchgeführt. Da herkömmliche Polymerasen keine RNA amplifizieren können, erfolgt im ersten Schritt die Umwandlung der RNA in komplementäre Desoxyribonukleinsäure (complementary deoxyribonucleic acid, cDNA) mittels der Reversen Transkriptase. Anschließend wird die cDNA amplifiziert. Hierfür wurde ein LightCycler 480 (Roche Diagnostics GmbH, Mannheim, Deutschland) und das LightCycler FastStart DNA Master Plus Kit (Roche Diagnostics GmbH, Mannheim, Deutschland) verwendet. Dabei wurde eine kurze Region in ORF 2/3 amplifiziert nach einem adaptierten PCR-Protokoll (62, 63).

### **2.5 Synthese der cDNA**

Im nächsten Schritt wurde die virale RNA mittels SuperScript IV Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) in cDNA transkribiert. Hierbei wurden als Primer sowohl zufällig generierte Hexanukleotide als auch oligo(dT)-Primer, die an den poly(A)-Schwanz der RNA binden, verwendet. Dabei wird die spezifische Wirkweise der oligo(dT)-Primer mit der hohen Ausbeute der zufällig generierte Hexanukleotide kombi-

niert. Weiterhin soll durch die Verwendung der zufällig generierten Hexanukleotide eine möglichst zufällige Amplifikation der Sequenzen erfolgen.

Die cDNA-Synthese wurde nach der Anleitung des Herstellers durchgeführt bis auf einen von 10 auf 20 Minuten verlängerten Inkubationsschritt bei 50°C und einem zusätzlichen Inkubationsschritt von 15 Minuten bei 55°C. Um überschüssige RNA zu entfernen, wurde das Produkt mit der Ribonuklease „RNase H“ gereinigt.

## 2.6 Genotypisierung

Alle Proben wurden genotypisiert und eine phylogenetische Analyse durchgeführt. Hierfür wurde ein 307 bp langes Fragment in der ORF1 des HEV-Genoms amplifiziert. Dabei wurde ein Protokoll nach Wang et al. angewandt (63). Anschließend wurden die Proben mittels Sanger-Sequenzierung sequenziert. Die Sequenzdaten wurden mit Geneious 11.1.5 (Biomatters, Auckland, New Zealand) bearbeitet. Die Primersequenzen wurden entfernt und eine Konsenssequenz erzeugt.

Im Anschluss wurden diese Sequenzen mit den Referenzsequenzen (nach Empfehlung der ICTV für Hepeviridae) (64), die ähnlichsten Sequenzen laut NCBI GenBank, sowie alle Sequenzen der HEV-Glue Sequence Database ([hev.glue.cvr.ac.uk/#/home](http://hev.glue.cvr.ac.uk/#/home)) zur Verwendung eines Kladogramms verwendet und angepasst mittels QIAGEN CLC Genomics Workbench 22.0 (<https://digitalinsights.qiagen.com/>).

## 2.7 Amplifikation der hypervariablen Region

Zur Amplifikation der HVR wurde eine „nested“ PCR durchgeführt wie in Papp et al. beschrieben (61). Bei einer „nested“ PCR handelt es sich um eine PCR, bei der zwei unterschiedliche Primerpaare verwendet werden, um möglichst spezifisch die Zielsequenz zu amplifizieren. Als Polymerase wurde die Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA) verwendet, die bei der Amplifikation eine hohe Genauigkeit aufweist. In den Tabellen 1 und 2 wird die Zusammensetzung des jeweiligen Ansatzes der PCR dargestellt. Nach der Gelelektrophorese wurden die PCR-Produkte aufgereinigt. Dazu wurden magnetische Partikel, sogenannte „magnetic beads“ MagSi-NGSPrep-Plus (Steinbrenner Laborsysteme GmbH, Wiesenbach, Deutschland), verwendet, die selektiv an PCR-Produkte einer bestimmten Größe bin-

den. Die Menge des PCR-Produktes wurde auf dem Invitrogen Qubit 4 Fluorometer mittels des Kits 1X dsDNA HS Assay Kit (Thermo Fischer Scientific, Waltham, MA, USA) gemessen. Die Sequenzierung erfolgte auf dem MiSeq Sequencer (Illumina Inc., San Diego, CA, USA). Dabei werden gepaarte „reads“ von einer 300 bp Länge generiert. Bei der DNA-Sequenzierung wird als ein „read“ eine abgeleitete Sequenz von Basenpaaren, die einem DNA-Fragment entspricht, bezeichnet.

**Tabelle 1: Erster Ansatz der „nested“ PCR zur Amplifikation der HVR**

1. PCR	Konzentration	20µl
Wasser		11,4
Primer HEV-255 [10µM]	500 nM	1
Primer HEV-257 [10µM]	500 nM	1
dNTPs [10mM]	200 µM	0,4
5x HF Buffer	1x	4
Phusion HotStart II [2U/µl]	0,25 U	0,2
Summe		18
Template (cDNA)		2

**Tabelle 2: Zweiter Ansatz der „nested“ PCR zur Amplifikation der HVR**

2. PCR	Konzentration	18µl
Wasser		11,92
Primer HEV-345 [10µM]	400 nM	0,72
Primer HEV-346 [10µM]	400 nM	0,72
dNTPs [10mM]	200 µM	0,36
5x HF Buffer	1x	3,6
Phusion HotStart II [2U/µl]	0,25 U	0,18
Summe		17,5
Template (1. PCR)		0,5

## 2.8 Bioinformatische Auswertung

Um die verschiedenen Muster zu erkennen, wurde aus den unter 2.7 erhaltenen Sequenzen zunächst verschiedene Cluster gebildet. Dafür wurde der UPARSE Algorithmus verwendet (65). Anschließend wurde die Sequenzen nach Länge sortiert und aus jeder Gruppe der Vertreter mit der größten Clustergröße ausgewählt. Anschließend wurden die Sequenzen in einer Pipeline weiterverarbeitet, die aus einem „merging“, „trimming“ und „mapping“ Schritt besteht (66-69). Aus den „alignments“ wurde eine Konsenssequenz für jede angebotene Referenz generiert anhand der häufigsten Kodons. Mithilfe von Samtools (Version 1.11) wurde die Verteilung der reads auf die Referenzsequenzen analysiert (70). Sequenzen mit mindestens 100 reads und einem Mindestanteil von 0,5 Prozent wurden aufgenommen. Für die weiteren Berechnungen wurden die Sequenzen in drei Gruppen aufgeteilt: Sequenzen mit Insertionen, Sequenzen mit Deletionen und Sequenzen ohne Insertionen oder Deletionen. Sequenzen, die sowohl eine Insertion als auch eine Deletion enthielten, wurden der Gruppe „Sequenzen mit Insertionen“ zugeordnet. Die Sequenzen sind in der NCBI-Genbank unter den Nummern OL469119–OL469148 hinterlegt.

## 2.9 Analyse der Sequenzen und statistische Auswertung

Die Analyse der Aminosäurezusammensetzung, die Visualisierung sowie die statistische Auswertung (unter Verwendung des „Wilcoxon rank sum test“) erfolgte mithilfe von Paketen der Programmiersprache R (71). Weiterhin wurde die Programmiersprache Julia Version 1.7.3 für die Datenverarbeitung genutzt (72). Die erweiterte Aminosäurecharakteristik der Sequenzen und die potenziellen Stellen für posttranslationale Modifikationen wie z.B. Acetylierungen wurden mittels verschiedener web-basierter Anwendungen ermittelt (73-78).

### 3. Ergebnisse

Es wurden 16 Proben analysiert. In der phylogenetischen Untersuchung wurden 15 Proben dem Subgenotyp 3c und eine Probe dem Subgenotyp 3e zugeordnet. In den untersuchten Proben der HEV-infizierten Patientinnen und Patienten wurden Insertionen und Deletionen (Indels) in den Sequenzen des viralen Genoms detektiert und zum Teil erstmalig beschrieben. Eine Insertion beschreibt dabei den zusätzlichen Einbau von Nukleotiden in eine Gensequenz und eine Deletion das Nichtvorhandensein einer Anzahl von Nukleotiden im Vergleich zur Referenzsequenz. Keine der detektierten Indels führten zu einer Verschiebung des Leserasters.

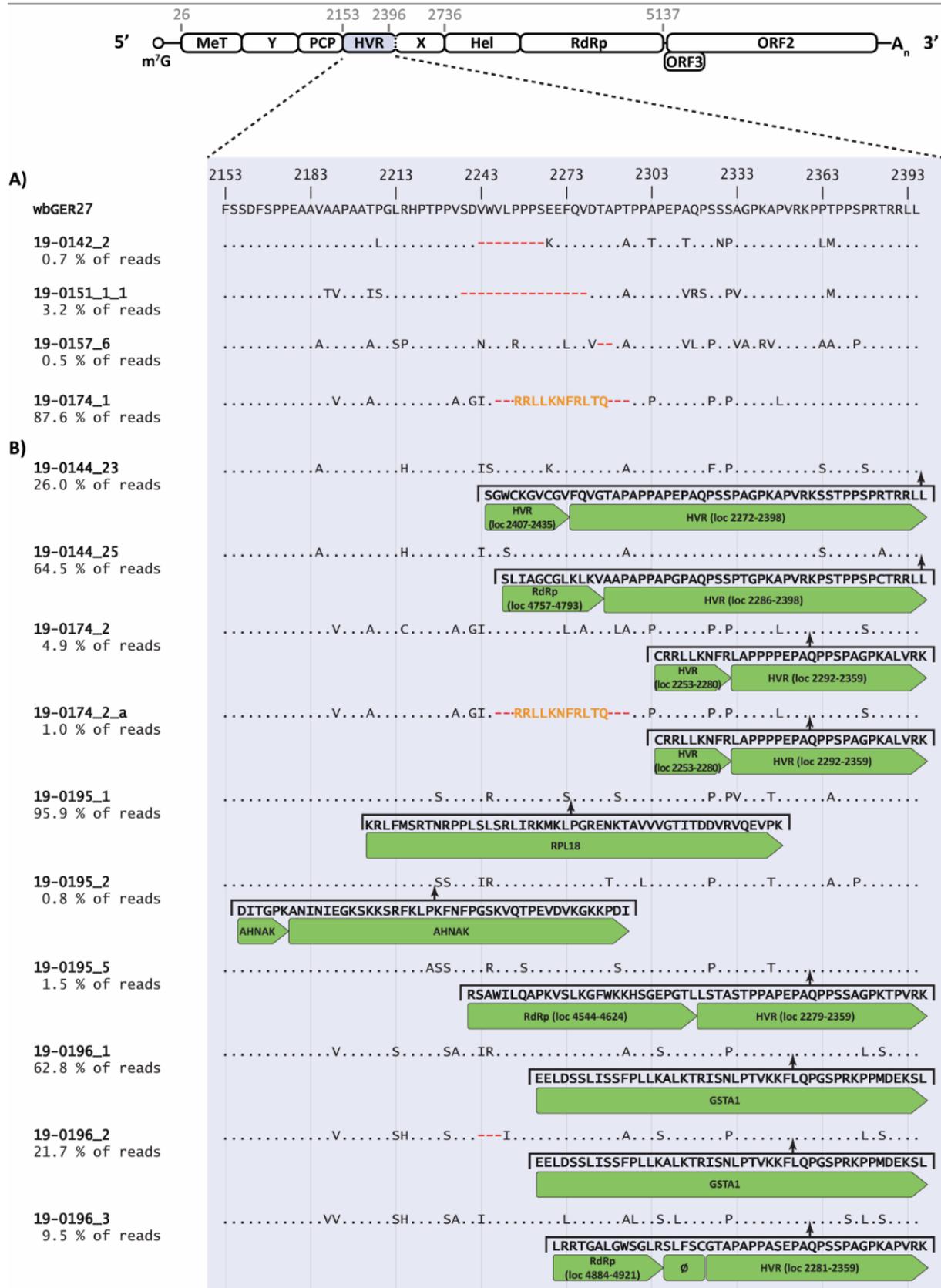
Weiterhin zeigten sich signifikante Unterschiede in der Aminosäurezusammensetzung und der Anzahl an posttranslationalen Modifikationen zwischen Sequenzen mit und ohne Indels. Konkret zeigten sich Indels in sieben von sechzehn untersuchten Proben. Dabei handelte es sich um drei akut und vier chronisch erkrankte Patienten und Patientinnen. Es wurden acht unterschiedliche Insertionen und fünf Deletionen detektiert. Die Insertionen wurden in den Proben von drei chronisch infizierten Personen und in der Probe einer akut HEV-infizierten Person festgestellt. Bei dem Vergleich mit der Datenbank „NCBI GenBank“ zeigten fünf Insertionen die höchste Sequenzübereinstimmung mit Abschnitten des HEV-Genoms und drei Insertionen die höchste Sequenzübereinstimmung mit den Abschnitten von drei unterschiedlichen humanen Genen. Es konnten weiterhin fünf verschiedene Deletionen identifiziert werden. Die Deletionen fanden sich bei zwei akut und drei chronisch HEV-infizierten Patienten und Patientinnen.

In Abbildung 1 sind die detektierten Indels detailliert dargestellt. Am oberen Rand der Abbildung findet sich eine vereinfachte Darstellung des HEV-Genoms mit den verschiedenen Regionen und eine Darstellung der Aminosäuresequenz der HVR der Referenzsequenz wbGER27. Im Teil A sind die Sequenzen, die nur Deletionen enthalten. Teil B stellt Sequenzen mit Insertionen und zum Teil auch mit zusätzlicher Deletion dar. Weiterhin sind die Sequenzen nach den ursprünglichen Patientenproben gruppiert. Die ersten sechs Ziffern des Sequenznamens entsprechen der Probennummer. Die Ziffer nach dem Unterstrich identifiziert die Variante innerhalb einer Probe. Varianten einer Probe ohne Indels sind in der Abbildung 1 nicht dargestellt.

Die kürzeste Insertion ist 96 Nukleotide lang (Abbildung 1: 19-0174\_2). Die längste Insertion umfasst 162 Nukleotide (Abbildung 1: 19-0195\_5). Die fünf Insertionen viralen Ursprungs zeigen sich nochmal aufgebaut in zwei Teile, welche unterschiedlichen Abschnitten des HEV-Genoms zuzuordnen sind (Abbildung 1: 19-0144\_23, 19-0144\_25, 19-0174\_2, 19-0195\_5, 19-0196\_3). Alle Abschnitte lassen sich entweder dem Bereich zuordnen, welcher für die virale RNA-abhängigen RNA-Polymerase (RdRp) kodiert oder einem Abschnitt aus der HVR selbst. Die Insertionen mit Ursprung aus dem humanen Genom ließen sich Teilbereichen der Gene des Ahnak nucleoprotein, Ribosomal Protein L18 (RPL18) und Glutathione S-Transferase Alpha 1 (GSTA1) zuordnen (Abbildung 1: 19-0195\_1, 19-0195\_2, 19-0196\_1). Es wurden auch drei verschiedene Insertionen innerhalb einer Probe detektiert (19-0195\_1, 19-0195\_2, 19-0195\_5).

Die Sequenzen mit Insertionen stellten häufig den Hauptteil der „reads“ einer Probe dar. Beispielweise lag bei der Probe 19-0195 der Anteil der Variante mit der Insertion RPL18 (19-0195\_1) bei 95,9 Prozent bezogen auf die Gesamtmenge der „reads“ der Patientenprobe. Für GSTA1 zeigten sich unter Berücksichtigung einer Variante mit Insertion und Deletion (19-0196\_1 und 19-0196\_2) ein Anteil von 84,5 Prozent an der Gesamtmenge. Die Sequenzen mit der Ahnak Insertion (19-0195\_2) machte hingegen nur 0,8 Prozent der „reads“ aus.

Weiterhin fielen in den Sequenzen der HVR in fünf Proben Deletionen auf, welche eine Länge von 6 bis 45 Nukleotiden besaßen. In einer Probe waren zwei Deletionen in einer Sequenz (Abbildung 1: 19-0174\_1). Es zeigten sich auch zwei HEV-Varianten mit einer Insertion und einer Deletion (Abbildung 1: 19-0174\_2A und 19-0196\_2). Es zeigte sich ein Anteil von 0,7 Prozent (19-0142\_2) bis zu 87,6 Prozent (19-0174\_1) an der Gesamtmenge der „reads“.



**Abbildung 1: Aminosäuresequenzen der Insertionen und Deletionen in den analysierten Sequenzen in einem Alignment.** Abschnitt A) zeigt die Sequenzen mit Deletionen. Abschnitt B) zeigt die Sequenzen mit Insertionen und Insertion plus Deletion. Links sind die Sequenznummern dargestellt und unterhalb derer der Anteil dieser Sequenz an der Gesamtmenge der „reads“ in einer Probe. In grün markiert sind die Insertionen und deren unterschiedliche Bestandteile. In den geschweiften Klammern ist die Aminosäuresequenz der Insertionen dargestellt. Mit roten Bindestrichen sind die Deletionen gekennzeichnet. Gelb markiert ist eine verän-

derte Aminosäuresequenz zwischen zwei Deletionen. Oben findet sich die Referenzsequenz *wbGER27* (FJ705359.1). *GSTA1* - Glutathione S-transferase alpha 1; *HVR* - hypervariable region; *He* - helicase region; *MeT* - methyltransferase region; *ORF* - open reading frame; *PCP* - papain-like cysteine protease; *RdRp* - RNA-dependent RNA-Polymerase; *RPL18* - Ribosomal Protein L18; *X* - X-domain; *Y* - Y-domain. Grafik aus Biedermann et al. (79)

In einem weiteren Schritt erfolgte eine statistische Analyse der Sequenzen hinsichtlich ihrer Aminosäurezusammensetzung. Hierfür wurden die Sequenzen mit Insertionen sowie die Sequenzen mit Deletionen mit den Sequenzen ohne Indels verglichen. In Tabelle 3 werden die Anteile der Aminosäuren aufgeführt, bei denen sich bei mindestens einem dieser Vergleiche ein signifikanter Unterschied feststellen ließ. In den Spalten zwei bis vier sind die verschiedenen Aminosäureanteile der Gruppen in Prozent angegeben. In den Spalten fünf und sechs ist dargestellt, bei welchem Vergleich sich ein signifikanter Unterschied zeigte mit dem jeweiligen P-Wert in Klammern. Aus den Daten in Tabelle 3 lässt sich ablesen, dass Sequenzen mit Insertionen einen signifikant höheren Anteil an Lysin und Glycin sowie einen geringeren Anteil an Prolin, Glutamat und Valin aufwiesen. Sequenzen mit Deletionen enthielten ebenfalls im Vergleich zu den Sequenzen ohne Indels mehr Lysin. Des Weiteren zeigten diese einen erhöhten Anteil an Arginin, Asparagin und Histidin.

**Tabelle 3: Analyse der Aminosäurezusammensetzung der Sequenzen und Vergleich zwischen Sequenzen ohne Insertionen oder Deletionen (Indels) und Sequenzen mit Indels.** Hinter den Aminosäurenamen in Klammern ist ihr Buchstabenkode aufgeführt. Die statistische Analyse wurde mittels "Wilcoxon Rank Sum test" durchgeführt. P-Werte <0,05 wurden als signifikant gewertet und sind in fett dargestellt. Tabelle modifiziert nach Biedermann et al. (79).

Aminosäure	Sequenzen ohne Indels (Anteil in Prozent)	Sequenzen mit Insertionen (Anteil in Prozent)	Sequenzen mit Deletionen (Anteil in Prozent)	Signifikanter Unterschied Insertionen/keine Indels	Signifikanter Unterschied Deletionen/ keine Indels
Arginin (R)	6,2	6,7	8,4	Nein (p-Wert: 1)	<b>Ja (p-Wert: 0,006)</b>
Asparagin (N)	0,3	0,8	1,0	Nein (p-Wert: 0,098)	<b>Ja (p-Wert: 0,017)</b>
Glutamat (E)	4,7	4,2	3,7	<b>Ja (p-Wert: 0,04)</b>	Nein (p-Wert: 0,166)
Glycin (G)	2,5	4,0	2,7	<b>Ja (p-Wert: 0,013)</b>	Nein (p-Wert: 0,234)
Histidin (H)	1,2	1,2	1,3	Nein (p-Wert: 1)	<b>Ja (p-Wert: 0,012)</b>
Lysin (K)	2,2	5,5	3,0	<b>Ja (p-Wert: &lt;0,001)</b>	<b>Ja (p-Wert: 0,048)</b>
Prolin (P)	28,5	23,3	27,8	<b>Ja (p-Wert: 0,001)</b>	Nein (p-Wert: 0,738)
Valin (V)	6,8	5,6	7,1	<b>Ja (p-Wert: 0,032)</b>	Nein (p-Wert: 0,962)

In Tabelle 4 werden die Aminosäureeigenschaften und potenziellen posttranslationaler Modifikationsstellen der Sequenzen aufgeführt, bei denen sich im Vergleich zwischen den Gruppen ein signifikanter Unterschied zeigte. Bezüglich der Aminosäureeigenschaften enthielten Sequenzen mit Insertionen mehr polare und basische Aminosäuren und weniger „small“ Aminosäuren und nicht-polare Aminosäuren. Sequenzen mit Deletionen enthielten mehr basische und weniger saure Aminosäure. Weiterhin wurden die Sequenzen bezüglich potenzieller Stellen für posttranslationale Modifikationen untersucht. Posttranslationale Modifikation bezeichnet dabei den Vorgang, bei welchem nach der Translation das Protein verändert wird. Häufig geschieht dies durch die zusätzliche Bindung anorganischer oder organischer Gruppen.

Es zeigten sich ein signifikant höherer Anteil an potenziellen Modifikationsstellen für Acetylierungen, Ubiquitinierungen und Methylierungen (Lysin) in Sequenzen mit Insertionen oder Deletionen im Vergleich zu Sequenzen ohne Indels.

**Tabelle 4: Analyse potenzieller posttranslationaler Modifikationsstellen und Aminosäureigenschaften der Sequenzen sowie Vergleich zwischen Sequenzen ohne Insertionen oder Deletionen (Indels) und Sequenzen mit Indels.** Die statistische Analyse wurde mittels "Wilcoxon Rank Sum test" durchgeführt. P-Werte <0,05 wurden als signifikant gewertet und sind in fett dargestellt. Tabelle modifiziert nach Biedermann et al. (79).

Eigenschaften	Sequenzen ohne Indels (Anteil in Prozent - Medianwert)	Sequenzen mit Insertionen (Anteil in Prozent - Medianwert)	Sequenzen mit Deletionen (Anteil in Prozent - Medianwert)	Signifikanter Unterschied Insertionen/keine Insertionen	Signifikanter Unterschied Deletionen/keine Deletionen
Acetylierungsstellen	2,4	4,6	2,8	<b>Ja (p-Wert: 0,01)</b>	<b>Ja (p-Wert: 0,048)</b>
Ubiquitinierungsstellen	2,4	4,6	2,8	<b>Ja (p-Wert: 0,01)</b>	<b>Ja (p-Wert: 0,048)</b>
Methylierungsstellen (Lysin)	2,4	3,2	3,4	<b>Ja (p-Wert: 0,0018)</b>	<b>Ja (p-Wert: 0,037)</b>
„Small“	72,0	65,8	71,0	<b>Ja (p-Wert: 0,0023)</b>	Nein (p-Wert: 0,47)
Nicht-polar	62,8	58,1	63,2	<b>Ja (p-Wert: 0,029)</b>	Nein (p-Wert: 1)
Polar	37,2	41,9	36,8	<b>Ja (p-Wert: 0,029)</b>	Nein (p-Wert: 0,89)
Basisch	9,8	13,0	12,8	<b>Ja (p-Wert: &lt; 0,0001)</b>	<b>Ja (p-Wert: 0,014)</b>
Sauer	8,5	6,5	6,7	Nein (p-Wert: 0,08)	<b>Ja (p-Wert: 0,0065)</b>

## **4. Diskussion**

### **4.1 Kurze Zusammenfassung der Ergebnisse**

In dieser Arbeit wurde erstens untersucht, ob Insertionen bei akut HEV-infizierten Personen vorkommen, zweitens ob neue, unbekannte Insertionen identifiziert werden können und drittens wodurch Sequenzen mit Insertionen, insbesondere hinsichtlich ihrer Aminosäurezusammensetzung und Möglichkeiten der posttranslationalen Modifikation, charakterisiert sind.

In den Untersuchungen der HVR des HEV-Genoms mittels NGS-Methoden konnten acht verschiedene Insertionen in vier Patientenproben detektiert werden. Eine Probe entstammte einem akut HEV-infizierten Patienten. Es konnte außerdem mit GSTA1 eine bisher noch nicht beschriebene Insertion humanen Ursprungs identifiziert werden. Diese Insertionen wiesen besondere Charakteristika auf. Es zeigte sich insbesondere ein erhöhter Anteil von Lysin und eine erhöhte Anzahl posttranslationaler Modifikationsstellen.

Daneben zeigten sich fünf Deletionen in fünf Patientenproben. In dieser Arbeit konnte erstmals nachgewiesen werden, dass auch Sequenzen mit Deletionen einen erhöhten Anteil von potenziellen Stellen für posttranslationale Modifikationen und eine spezifische Aminosäurezusammensetzung haben.

### **4.2 Interpretation der Ergebnisse**

Insgesamt zeigten sich Insertionen in der überwiegenden Zahl bei chronisch HEV-infizierten Patienten und Patientinnen. Insofern konnte diese Arbeit eine Korrelation zwischen dem Auftreten von Insertionen und dem Vorliegen einer chronischen HEV-Infektion bestätigen.

In den durchgeführten Analysen konnte eine Insertion humanen Ursprungs identifiziert werden, welche bisher noch nicht beschrieben worden war. Dabei handelt es sich um ein Fragment des Gens, welches die GSTA1 transkribiert. Dieses Enzym ist Teil der Biotransformation in der Leber und sorgt durch das Anfügen eines Glutathionrestes für die Wasserlöslichkeit von körpereigenen und -fremden Stoffen und ermöglicht so deren

Ausscheidung (80). Möglicherweise wurde ein Abschnitt des GSTA1 Gens eingebaut, da dieses ubiquitär in der Leber exprimiert wird. Weiterhin wurden Insertionen mit höchster Sequenzübereinstimmung zu den humanen Genen RPL18 sowie von Ahnak detektiert, welche zuvor in unserer Arbeitsgruppe gefunden und publiziert wurden (61). Dabei handelt es sich bei RPL18 um ein Gen, welches für ein ribosomales Protein kodiert.

In der weiteren Analyse der Sequenzen der HVR zeigte sich eine signifikant höhere Anzahl an potenziellen Stellen für posttranslationale Modifikationen für Acetylierungen, Ubiquitinierungen und Methylierungen (Lysin) in Sequenzen mit Insertionen im Vergleich zu Sequenzen ohne Insertionen. Die erhöhte Anzahl von Stellen für posttranslationale Modifikationen gibt einen Hinweis darauf, dass die Gensequenzen mit Insertionen in Proteine übersetzt werden und dass sie funktionell relevant sind für das Hepatitis-E-Virus.

### **4.3 Einbettung der Ergebnisse in den bisherigen Forschungsstand**

Zu Beginn dieses Forschungsprojektes im Jahr 2018 lagen zum Vorkommen und zum Charakter von Insertionen der HVR lediglich anekdotische Beobachtungen oder Untersuchungen mittels konventioneller Sanger-Sequenzierung vor. Auch die Untersuchung von Munoz et al. aus dem Jahr 2020 wurde mittels Sanger-Sequenzierung durchgeführt (59). Dagegen wurden in der im Jahr 2020 erschienenen Arbeit von Lhomme et al. ebenfalls Sequenzen mittels „Next-Generation Sequencing“ untersucht. Deren Untersuchungen wurden allerdings mittels einer anderen Sequenzierungsplattform, PacBio SMRT, durchgeführt, während in dieser Untersuchung das MiSeq System von Illumina verwendet wurde. In dieser Arbeit wurden wie in Lhomme et al. (58) Insertionen in einem akut HEV-infizierten Patientinnen und Patienten detektiert, in der überwiegenden Mehrheit aber bei chronisch HEV-infizierten Patientinnen und Patienten. Alle bisher detektierten Insertionen setzten sich aus vollständigen Codons zusammen, sodass keine Verschiebung des Leserahmens entstand.

Weiterhin wurde in dieser Arbeit untersucht, ob sich weitere neuartige Insertionen in den Virussequenzen nachweisen lassen. Hierzu lagen zu Beginn der Untersuchung die Arbeiten von Johne et al., Nguyen et al, Shukla et al. und Lhomme et al. vor (54-57). In

den ersten drei Arbeiten wurden in den Proben einzelner Personen Sequenzen mit Insertionen nachgewiesen. In der Arbeit von Lhomme et al. aus dem Jahr 2014 wurde die HVR von Patientinnen und Patienten untersucht, die im Rahmen einer Organtransplantation immunsupprimiert waren und bei denen eine HEV-Infektion vorlag. Bei Personen, die einen chronischen Verlauf aufwiesen, erfolgte eine erneute Sequenzierung. Bei 11 Prozent der untersuchten chronisch HEV-infizierten Patientinnen und Patienten zeigten sich Insertionen. In der Untersuchung im Rahmen dieser Dissertation zeigten sich dagegen bei über vierzig Prozent der untersuchten Proben der chronisch HEV-infizierten Personen Insertionen.

Bei den erstmals von Ngyuen et al. sowie Shukla et al. nachgewiesenen Insertionen in der HVR des HEV-Genoms handelte es sich um Abschnitte, welche die größte Sequenzübereinstimmung mit humanen ribosomalen Genen, dem ribosomalem Gen S17 und dem ribosomalen Gen S19, zeigten (54, 56). In der vorliegenden Arbeit wurde ebenfalls eine Insertion identifiziert, die die größte Sequenzübereinstimmung mit einem humanen ribosomalen Gen zeigte, und zwar dem Gen, das das ribosomale Protein L 18 kodiert. Möglicherweise sind diese Transkripte in der Zelle ubiquitär verfügbar und wurden daher in das Virusgenom eingebaut. Die genauen Mechanismen des Einbaus bzw. der Rekombination sind jedoch bislang unbekannt.

Weiterhin konnten in dieser Arbeit Insertionen beschrieben werden, deren Sequenzen die höchste Übereinstimmung mit Teilen der HVR selbst oder mit der viralen Polymerase, der RdRp, zeigten. Solche Duplikationen wurden auch bei anderen RNA-Viren der *Orthornavirae* beobachtet. Villordo et al. (81) untersuchten konservierte Duplikationen in den 3'-UTRs von Flavivirus-Genomen. In deren Untersuchung konnte gezeigt werden, dass konservierte Duplikationen eine wichtige Rolle bei dem Wechsel zwischen den Insekten- und Säugetierwirten spielen. Einzelne Fälle von Duplikationen wurden ebenfalls in Wildtypvarianten des Humanen Respiratorischen Synzytial-Virus (HRSV) beschrieben (82, 83). *In vitro* Analysen zeigten eine verbesserte Fitness von Virusvarianten mit Duplikation gegenüber Virusvarianten ohne Duplikation (84).

Hinsichtlich der Frage nach dem Charakter der Insertionen zeigte sich in dieser Arbeit eine Veränderung der Aminosäurezusammensetzung der Sequenzen mit Insertionen im Vergleich zu Sequenzen ohne Insertionen. Festgestellt wurden u.a. eine signifikante

Minderung des Prolinanteils sowie eine signifikante Steigerung des Lysinanteils bei den Sequenzen mit Insertionen im Vergleich zu Sequenzen ohne Insertionen. Dies entspricht den Beobachtungen von Lhomme et al. (58). Daher unterstützt diese Arbeit auch die Hypothese von Kenney et al., in der postuliert wurde, dass der erhöhte Anteil an Lysinen ein Schlüsselmerkmal der Insertionen in der HVR ist (85). Da Lysine posttranslational auf vielfältige Weise modifiziert werden können, ist dadurch auch die in dieser Arbeit bei Sequenzen mit Insertionen festgestellte erhöhte Anzahl potenzieller Positionen für posttranslationalen Modifikationen erklärbar.

Analog zu den Ergebnissen von Lhomme et al. (57, 86) und Muñoz-Chimeno et al. (59) wurde ebenfalls eine signifikante Zunahme potenzieller Stellen für Acetylierungen und Ubiquitinierungen in Sequenzen mit Insertionen festgestellt. Anders verhält es sich mit dem Anstieg potenzieller Stellen für Methylierungen (Lysin). Dieser Anstieg wurde in der vorliegenden Arbeit beobachtet, nicht jedoch bei den oben genannten Arbeiten. Umgekehrt verhält es sich in Hinblick auf die erhöhte Anzahl von potenziellen Phosphorylierungen (58) oder N-Glycolysierungen (59). Jene Beobachtungen konnten in dieser Untersuchung nicht bestätigt werden.

Deletionen im HEV-Genom wurden bisher nur von zwei Forschungsgruppen beschrieben (56, 59). Sequenzen mit Deletionen zeigten ebenfalls einen erhöhten Anteil von Lysin im Vergleich zu Sequenzen ohne Deletionen. In dieser Arbeit konnte zum ersten Mal der Nachweis erbracht werden, dass Sequenzen mit Deletionen ebenfalls einen erhöhten Anteil von potenziellen Stellen für posttranslationale Modifikationen aufweisen.

#### **4.4 Stärken und Schwächen der Studie**

Einerseits stellen die großen Datenmengen, die durch „Next-Generation Sequencing“ generiert werden, eine Herausforderung für die Auswertung dar. Es gilt im Auswertungsprozess die Frage zu beantworten, wie die biologische Realität möglichst genau abgebildet werden kann. Mögliche Fehlerquellen sind etwa die Selektion bestimmter Varianten durch die PCR sowie die richtige Interpretation. In dieser Untersuchung wurde diesen Herausforderungen durch die Wahl einer sehr präzisen Methode, der „Next-Generation Sequencing“, sowie strengen Qualitätsfiltern begegnet. Andererseits können

durch die große Anzahl der generierten „reads“ viele Virusvarianten rekonstruiert werden. In dieser Arbeit wurden Amplicons mit einer Länge von 300 bp sequenziert.

In der Ampliconsequenzierung können im Vergleich zur „Shot-Gun-Sequenzierung“ lange Abschnitte an einem Stück sequenziert werden und so die Rekonstruktion der Varianten erleichtert werden. Ein einfaches Mapping ist zur Abbildung aller Varianten mit Indels nicht zielführend. Daher wurde eine Clustering-Methode gewählt, um initial alle Varianten mit Indels zu identifizieren. Nach dem anschließenden Mapping der einzelnen Varianten konnte dann eine weitere Analyse der Sequenzen erfolgen.

Daher ist davon auszugehen, dass die Beobachtungen und Ergebnisse dieser Arbeit sehr robust und valide sind. Im Bereich der statistischen Auswertung zeigte sich eine außerordentlich hohe Signifikanz.

Eine Limitierung der vorliegenden Arbeit ist die geringe Anzahl an untersuchten Patientenproben. Aktuell wird bei immunsupprimierten Patientinnen und Patienten mit HEV-Infektion frühzeitig eine antivirale Therapie begonnen. Das hat zur Folge, dass zum definitionsgemäßen Übergang in eine chronische Infektion nach drei Monaten häufig eine geringere Viruslast vorliegt und so weniger Proben chronischer Patienten und Patientinnen mit ausreichend hoher Viruslast für eine Sequenzierung mittels NGS zur Verfügung stehen. Insgesamt ist aber aufgrund der präzisen Methodik und einer hohen statistischen Signifikanz davon auszugehen, dass die Ergebnisse dieser Arbeit zuverlässig sind, zu einem Zuwachs an Wissen in diesem Fachgebiet beitragen und Grundlage für weiterführende Experimente sein werden.

#### **4.5 Implikationen für Praxis und zukünftige Forschung**

Die Identifizierung von Virusvarianten bei chronischen Infektionen soll dazu dienen, Mechanismen der Chronifizierung und den damit einhergehenden Adaptationsprozess des HEV und dessen Evolution besser zu verstehen. Insertionen scheinen in diesem Prozess gehäuft aufzutreten und spielen möglicherweise eine entscheidende Rolle. Die Identifizierung und Überwachung dieser Veränderungen im viralen Genom sind daher wichtig, um besonders pathogene Varianten sicher zu identifizieren. Weitere *in vitro* Untersuchungen in Zellkultur sind notwendig, um die Relevanz der Insertionen zu evaluie-

ren. Auch der Prozess des Einbaus und der Rekombination von viralen und humanen Genfragmenten sowie die Relevanz von Deletionen erfordert weitere Forschungsarbeiten.

Die große Vielfalt innerhalb der viralen Quasispezies von RNA-Viren konnte durch diese Arbeit unterstrichen werden. Diese Diversität erfordert daher auch detaillierte Sequenzierungsmethoden, welche sich dieser Vielfalt bestmöglich annähern. Diese Arbeit zeigt somit auch die Bedeutung geeigneter, besonders genauer und innovativer Untersuchungsmethoden auf.

Die Analyse des HEV-Genoms könnte in Zukunft dazu dienen, schwere Verläufe oder Therapieversagen vorherzusagen. Dazu ist eine umfassendere Erfassung der Veränderungen im HEV-Genom wünschenswert. Diese Sequenzierungsdaten könnten zukünftig dazu genutzt werden, individuelle Anpassungen der Therapie vorzunehmen und Resistenzen frühzeitig zu identifizieren.

## 5. Schlussfolgerungen

In dieser Arbeit wurden Varianten des Hepatitis-E-Virus aus Blutproben von Patientinnen und Patienten untersucht und deren klinische Relevanz für den chronischen Verlauf einer Hepatitis E beurteilt.

Diese Untersuchung konnte bestätigen, dass Insertionen in der HVR bei chronisch HEV-infizierten Personen gehäuft vorkommen. Das erhöhte Vorkommen von Insertionen bei diesen Personen deutet auf deren Relevanz für den Verlauf der HEV-Infektion hin. Eine Insertion konnte auch bei einer akut HEV-infizierten Person nachgewiesen werden. In der Zusammenschau mit den derzeit bekannten Forschungsergebnissen scheint es sich bei dieser Beobachtung vermutlich um Einzelfälle zu handeln.

Die Sequenzen mit Insertionen weisen bestimmte Eigenschaften auf. Es wurde ein erhöhter Anteil an Lysin innerhalb der Aminosäuresequenz sowie eine erhöhte Anzahl von posttranslationalen Modifikationsstellen festgestellt. Insgesamt deutet dies auf eine funktionelle Relevanz der Insertionen hin.

Die Analyse von Virusvarianten mit Hilfe von sensitiven Sequenzierungsmethoden könnte in der klinischen Praxis künftig dazu dienen, Risikopatientinnen und -patienten auszumachen und individuelle Therapieanpassungen vorzunehmen. Weitere Forschung ist notwendig, um den Mechanismus hinter dem Einbau der Insertionen zu verstehen, die funktionelle Relevanz zu verifizieren und weitere Angriffspunkte für mögliche Therapien zu identifizieren.

## Literaturverzeichnis

1. Cao G, Jing W, Liu J, Liu M. The global trends and regional differences in incidence and mortality of hepatitis A from 1990 to 2019 and implications for its prevention. *Hepatol Int.* 2021;15(5):1068-82.
2. Hsu YC, Huang DQ, Nguyen MH. Global burden of hepatitis B virus: current status, missed opportunities and a call for action. *Nat Rev Gastroenterol Hepatol.* 2023;20(8):524-37.
3. Roudot-Thoraval F. Epidemiology of hepatitis C virus infection. *Clin Res Hepatol Gastroenterol.* 2021;45(3):101596.
4. Bender D, Glitscher M, Hildt E. [Viral hepatitis A to E: prevalence, pathogen characteristics, and pathogenesis]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2022;65(2):139-48.
5. Stockdale AJ, Kreuels B, Henrion MYR, Giorgi E, Kyomuhangi I, de Martel C, Hutin Y, Geretti AM. The global prevalence of hepatitis D virus infection: Systematic review and meta-analysis. *J Hepatol.* 2020;73(3):523-32.
6. Khuroo MS. Study of an epidemic of non-A, non-B hepatitis. Possibility of another human hepatitis virus distinct from post-transfusion non-A, non-B type. *Am J Med.* 1980;68(6):818-24.
7. Balayan MS, Andjaparidze AG, Savinskaya SS, Ketiladze ES, Braginsky DM, Savinov AP, Poleschuk VF. Evidence for a virus in non-A, non-B hepatitis transmitted via the fecal-oral route. *Intervirology.* 1983;20(1):23-31.
8. Songtanin B, Molehin AJ, Brittan K, Manatsathit W, Nugent K. Hepatitis E Virus Infections: Epidemiology, Genetic Diversity, and Clinical Considerations. *Viruses.* 2023;15(6).
9. Li P, Liu J, Li Y, Su J, Ma Z, Bramer WM, Cao W, de Man RA, Peppelenbosch MP, Pan Q. The global epidemiology of hepatitis E virus infection: A systematic review and meta-analysis. *Liver Int.* 2020;40(7):1516-28.
10. Robert Koch-Institut: Hepatitis E - RKI Ratgeber. 2023 [https://www.rki.de/DE/Content/Infekt/EpidBull/Merkblaetter/Ratgeber\\_HepatitisE.html](https://www.rki.de/DE/Content/Infekt/EpidBull/Merkblaetter/Ratgeber_HepatitisE.html). Abfragedatum: 04.08.2024 [Internet].
11. Robert Koch-Institut: SurvStat@RKI 2.0, <https://survstat.rki.de>, Abfragedatum: 04.08.2024 [Internet].
12. Faber MS, Wenzel JJ, Jilg W, Thamm M, Höhle M, Stark K. Hepatitis E virus seroprevalence among adults, Germany. *Emerg Infect Dis.* 2012;18(10):1654-7.
13. Capai L, Falchi A, Charrel R. Meta-Analysis of Human IgG anti-HEV Seroprevalence in Industrialized Countries and a Review of Literature. *Viruses.* 2019;11(1).
14. Rein DB, Stevens GA, Theaker J, Wittenborn JS, Wiersma ST. The global burden of hepatitis E virus genotypes 1 and 2 in 2005. *Hepatology.* 2012;55(4):988-97.
15. Schemmerer M, Wenzel JJ, Stark K, Faber M. Molecular epidemiology and genotype-specific disease severity of hepatitis E virus infections in Germany, 2010-2019. *Emerg Microbes Infect.* 2022;11(1):1754-63.
16. Dalton HR, Hunter JG, Bendall RP. Hepatitis E. *Curr Opin Infect Dis.* 2013;26(5):471-8.
17. Purdy MA, Drexler JF, Meng XJ, Norder H, Okamoto H, Van der Poel WHM, Reuter G, de Souza WM, Ulrich RG, Smith DB. ICTV Virus Taxonomy Profile: Hepeviridae 2022. *J Gen Virol.* 2022;103(9).

18. Yang YL, Nan YC. Open reading frame 3 protein of hepatitis E virus: Multi-function protein with endless potential. *World J Gastroenterol*. 2021;27(20):2458-73.
19. Domingo E, Sheldon J, Perales C. Viral Quasispecies Evolution. *Microbiology and Molecular Biology Reviews*. 2012;76(2):159-216.
20. Pischke S, Behrendt P, Bock CT, Jilg W, Manns MP, Wedemeyer H. Hepatitis E in Germany--an under-reported infectious disease. *Dtsch Arztebl Int*. 2014;111(35-36):577-83.
21. Abravanel F, Pique J, Couturier E, Nicot F, Dimeglio C, Lhomme S, Chiabrando J, Saune K, Péron JM, Kamar N, Evrard S, de Valk H, Cintas P, Izopet J. Acute hepatitis E in French patients and neurological manifestations. *J Infect*. 2018;77(3):220-6.
22. Purcell RH, Emerson SU. Hepatitis E: an emerging awareness of an old disease. *J Hepatol*. 2008;48(3):494-503.
23. Larrue H, Abravanel F, Péron JM. Hepatitis E, what's the real issue? *Liver Int*. 2020;40 Suppl 1:43-7.
24. Wedemeyer H, Pischke S, Manns MP. Pathogenesis and treatment of hepatitis e virus infection. *Gastroenterology*. 2012;142(6):1388-97.e1.
25. Blasco-Perrin H, Madden RG, Stanley A, Crossan C, Hunter JG, Vine L, Lane K, Devooght-Johnson N, McLaughlin C, Petrik J, Stableforth B, Hussaini H, Phillips M, Mansuy JM, Forrest E, Izopet J, Blatchford O, Scobie L, Peron JM, Dalton HR. Hepatitis E virus in patients with decompensated chronic liver disease: a prospective UK/French study. *Aliment Pharmacol Ther*. 2015;42(5):574-81.
26. Khuroo MS, Kamili S. Aetiology, clinical course and outcome of sporadic acute viral hepatitis in pregnancy. *J Viral Hepat*. 2003;10(1):61-9.
27. Wu C, Wu X, Xia J. Hepatitis E virus infection during pregnancy. *Virol J*. 2020;17(1):73.
28. Kamar N, Izopet J, Pavio N, Aggarwal R, Labrique A, Wedemeyer H, Dalton HR. Hepatitis E virus infection. *Nat Rev Dis Primers*. 2017;3:17086.
29. Treagus S, Wright C, Baker-Austin C, Longdon B, Lowther J. The Foodborne Transmission of Hepatitis E Virus to Humans. *Food Environ Virol*. 2021;13(2):127-45.
30. Ankcorn MJ, Tedder RS. Hepatitis E: the current state of play. *Transfus Med*. 2017;27(2):84-95.
31. Schlosser B, Stein A, Neuhaus R, Pahl S, Ramez B, Krüger DH, Berg T, Hofmann J. Liver transplant from a donor with occult HEV infection induced chronic hepatitis and cirrhosis in the recipient. *J Hepatol*. 2012;56(2):500-2.
32. Kamar N, Selves J, Mansuy J-M, Ouezzani L, Péron J-M, Guitard J, Cointault O, Esposito L, Abravanel F, Danjoux M, Durand D, Vinel J-P, Izopet J, Rostaing L. Hepatitis E virus and chronic hepatitis in organ-transplant recipients. *The New England Journal of Medicine*. 2008;358(8):811-7.
33. EASL Clinical Practice Guidelines on hepatitis E virus infection. *Journal of Hepatology*. 2018;68(6):1256-71.
34. Ma Z, de Man RA, Kamar N, Pan Q. Chronic hepatitis E: Advancing research and patient care. *J Hepatol*. 2022;77(4):1109-23.
35. Kamar N, Garrouste C, Haagsma EB, Garrigue V, Pischke S, Chauvet C, Dumortier J, Cannesson A, Cassuto-Viguié E, Thervet E, Conti F, Lebray P, Dalton HR, Santella R, Kanaan N, Essig M, Mousson C, Radenne S, Roque-Afonso AM, Izopet J, Rostaing L. Factors associated with chronic hepatitis in patients with hepatitis E virus infection who have received solid organ transplants. *Gastroenterology*. 2011;140(5):1481-9.

36. le Coutre P, Meisel H, Hofmann J, Röcken C, Vuong GL, Neuburger S, Hemmati PG, Dörken B, Arnold R. Reactivation of hepatitis E infection in a patient with acute lymphoblastic leukaemia after allogeneic stem cell transplantation. *Gut*. 2009;58(5):699-702.
37. Beer A, Holzmann H, Pischke S, Behrendt P, Wrba F, Schlue J, Drebber U, Neudert B, Halilbasic E, Kreipe H, Lohse A, Sterneck M, Wedemeyer H, Manns M, Dienes HP. Chronic Hepatitis E is associated with cholangitis. *Liver Int*. 2019;39(10):1876-83.
38. Dalton HR, Kamar N, van Eijk JJ, McLean BN, Cintas P, Bendall RP, Jacobs BC. Hepatitis E virus and neurological injury. *Nat Rev Neurol*. 2016;12(2):77-85.
39. Ripellino P, Norton B, van Eijk J, Dalton HR. Non-traumatic neurological injury and hepatitis E infection. *Expert Rev Anti Infect Ther*. 2018;16(4):255-7.
40. Kamar N, Abravanel F, Selves J, Garrouste C, Esposito L, Lavayssière L, Cointault O, Ribes D, Cardeau I, Nogier MB, Mansuy JM, Muscari F, Peron JM, Izopet J, Rostaing L. Influence of immunosuppressive therapy on the natural history of genotype 3 hepatitis-E virus infection after organ transplantation. *Transplantation*. 2010;89(3):353-60.
41. Kamar N, Izopet J, Tripon S, Bismuth M, Hillaire S, Dumortier J, Radenne S, Coilly A, Garrigue V, D'Alteroche L, Buchler M, Couzi L, Lebray P, Dharancy S, Minello A, Hourmant M, Roque-Afonso A-M, Abravanel F, Pol S, Rostaing L, Mallet V. Ribavirin for Chronic Hepatitis E Virus Infection in Transplant Recipients. *New England Journal of Medicine*. 2014;370(12):1111-20.
42. Gorris M, van der Lecq BM, van Erpecum KJ, de Bruijne J. Treatment for chronic hepatitis E virus infection: A systematic review and meta-analysis. *J Viral Hepat*. 2021;28(3):454-63.
43. Lhomme S, Abravanel F, Dubois M, Sandres-Saune K, Rostaing L, Kamar N, Izopet J. Hepatitis E virus quasispecies and the outcome of acute hepatitis E in solid-organ transplant patients. *Journal of virology*. 2012;86(18):10006-14.
44. Narayanan S, Abutaleb A, Sherman KE, Kottilil S. Clinical features and determinants of chronicity in hepatitis E virus infection. *Journal of Viral Hepatitis*. 2019;0(0).
45. Suneetha PV, Pischke S, Schlaphoff V, Grabowski J, Fyttili P, Gronert A, Bremer B, Markova A, Jaroszewicz J, Bara C, Manns MP, Cornberg M, Wedemeyer H. Hepatitis E virus (HEV)-specific T-cell responses are associated with control of HEV infection. *Hepatology*. 2012;55(3):695-708.
46. Kemming J, Gundlach S, Panning M, Huzly D, Huang J, Lütgehetmann M, Pischke S, Schulze Zur Wiesch J, Emmerich F, Llewellyn-Lacey S, Price DA, Tanriver Y, Warnatz K, Boettler T, Thimme R, Hofmann M, Fischer N, Neumann-Haefelin C. Mechanisms of CD8+ T-cell failure in chronic hepatitis E virus infection. *J Hepatol*. 2022;77(4):978-90.
47. Debing Y, Ramière C, Dallmeier K, Piorkowski G, Trabaud M-A, Lebossé F, Scholtès C, Roche M, Legras-Lachuer C, de Lamballerie X, André P, Neyts J. Hepatitis E virus mutations associated with ribavirin treatment failure result in altered viral fitness and ribavirin sensitivity. *Journal of Hepatology*. 2016;65(3):499-508.
48. Todt D, Gisa A, Radonic A, Nitsche A, Behrendt P, Suneetha PV, Pischke S, Bremer B, Brown RJP, Manns MP, Cornberg M, Bock CT, Steinmann E, Wedemeyer H. In vivo evidence for ribavirin-induced mutagenesis of the hepatitis E virus genome. *Gut*. 2016;65(10):1733-43.
49. van Tong H, Hoan NX, Wang B, Wedemeyer H, Bock CT, Velavan TP. Hepatitis E Virus Mutations: Functional and Clinical Relevance. *EBioMedicine*. 2016;11:31-42.

50. Chen Z, Zhou X, Lunney JK, Lawson S, Sun Z, Brown E, Christopher-Hennings J, Knudsen D, Nelson E, Fang Y. Immunodominant epitopes in nsp2 of porcine reproductive and respiratory syndrome virus are dispensable for replication, but play an important role in modulation of the host immune response. *J Gen Virol*. 2010;91(Pt 4):1047-57.
51. Lhomme S, Garrouste C, Kamar N, Saune K, Abravanel F, Mansuy JM, Dubois M, Rostaing L, Izopet J. Influence of polyproline region and macro domain genetic heterogeneity on HEV persistence in immunocompromised patients. *J Infect Dis*. 2014;209(2):300-3.
52. Koonin EV, Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. *Proceedings of the National Academy of Sciences*. 1992;89(17):8259.
53. Purdy MA, Lara J, Khudyakov YE. The Hepatitis E Virus Polyproline Region Is Involved in Viral Adaptation. *PloS one*. 2012;7(4):e35974.
54. Shukla P, Nguyen HT, Torian U, Engle RE, Faulk K, Dalton HR, Bendall RP, Keane FE, Purcell RH, Emerson SU. Cross-species infections of cultured cells by hepatitis E virus and discovery of an infectious virus–host recombinant. *Proceedings of the National Academy of Sciences*. 2011;108(6):2438.
55. Johne R, Reetz J, Ulrich RG, Machnowska P, Sachsenröder J, Nickel P, Hofmann J. An ORF1-rearranged hepatitis E virus derived from a chronically infected patient efficiently replicates in cell culture. *Journal of Viral Hepatitis*. 2014;21(6):447-56.
56. Nguyen HT, Torian U, Faulk K, Mather K, Engle RE, Thompson E, Bonkovsky HL, Emerson SU. A naturally occurring human/hepatitis E recombinant virus predominates in serum but not in faeces of a chronic hepatitis E patient and has a growth advantage in cell culture. *Journal of General Virology*. 2012;93(3):526-30.
57. Lhomme S, Abravanel F, Dubois M, Sandres-Saune K, Mansuy J-M, Rostaing L, Kamar N, Izopet J. Characterization of the polyproline region of the hepatitis E virus in immunocompromised patients. *Journal of virology*. 2014;88(20):12017-25.
58. Lhomme S, Nicot F, Jeanne N, Dimeglio C, Roulet A, Lefebvre C, Carcenac R, Manno M, Dubois M, Peron J-M, Alric L, Kamar N, Abravanel F, Izopet J. Insertions and Duplications in the Polyproline Region of the Hepatitis E Virus. *Frontiers in microbiology*. 2020;11:1-.
59. Muñoz-Chimeno M, Cenalmor A, Garcia-Lugo MA, Hernandez M, Rodriguez-Lazaro D, Avellon A. Proline-Rich Hypervariable Region of Hepatitis E Virus: Arranging the Disorder. *Microorganisms*. 2020;8(9):1417.
60. Schulz M, Biedermann P, Bock C-T, Hofmann J, Choi M, Tacke F, Hanitsch LG, Mueller T. Rituximab-Containing Treatment Regimens May Imply a Long-Term Risk for Difficult-To-Treat Chronic Hepatitis E. *International Journal of Environmental Research and Public Health*. 2020;17(1):341.
61. Papp CP, Biedermann P, Harms D, Wang B, Kebelmann M, Choi M, Helmuth J, Corman VM, Thürmer A, Altmann B, Klink P, Hofmann J, Bock CT. Advanced sequencing approaches detected insertions of viral and human origin in the viral genome of chronic hepatitis E virus patients. *Sci Rep*. 2022;12(1):1720.
62. Jothikumar N, Cromeans TL, Robertson BH, Meng XJ, Hill VR. A broadly reactive one-step real-time RT-PCR assay for rapid and sensitive detection of hepatitis E virus. *Journal of Virological Methods*. 2006;131(1):65-71.
63. Wang B, Harms D, Papp CP, Niendorf S, Jacobsen S, Lütgehetmann M, Pischke S, Wedermeyer H, Hofmann J, Bock CT. Comprehensive Molecular Approach for

- Characterization of Hepatitis E Virus Genotype 3 Variants. *Journal of clinical microbiology*. 2018;56(5):e01686-17.
64. Smith DB, Izopet J, Nicot F, Simmonds P, Jameel S, Meng X-J, Norder H, Okamoto H, van der Poel WHM, Reuter G, Purdy MA. Update: proposed reference sequences for subtypes of hepatitis E virus (species Orthohepevirus A). *The Journal of General Virology*. 2020;101(7):692-8.
65. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*. 2013;10(10):996-8.
66. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*. 2014;30(15):2114-20.
67. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*. 2010;26(5):589-95.
68. Machnowska P, Hauser A, Meixenberger K, Altmann B, Bannert N, Rempis E, Schnack A, Decker S, Braun V, Busingye P, Rubaihayo J, Harms G, Theuring S. Decreased emergence of HIV-1 drug resistance mutations in a cohort of Ugandan women initiating option B+ for PMTCT. *PloS One*. 2017;12(5):e0178297.
69. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011;27(21):2957-63.
70. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2010;26(5):589-95.
71. Pagès H, Aboyou P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings2020 2020.
72. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*. 2017;59(1):65-98.
73. Deng W, Wang C, Zhang Y, Xu Y, Zhang S, Liu Z, Xue Y. GPS-PAIL: prediction of lysine acetyltransferase-specific modification sites from protein sequences. *Scientific Reports*. 2016;6:39787.
74. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology*. 1999;294(5):1351-62.
75. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004;4(6):1633-49.
76. Qiu W-R, Xiao X, Lin W-Z, Chou K-C. iUbiq-Lys: prediction of lysine ubiquitination sites in proteins by extracting sequence evolution information via a gray system model. *Journal of Biomolecular Structure & Dynamics*. 2015;33(8):1731-42.
77. Wang P, Xiao X, Chou K-C. NR-2L: a two-level predictor for identifying nuclear receptor subfamilies based on sequence-derived features. *PloS One*. 2011;6(8):e23505.
78. Madeira F, Park Ym, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*. 2019;47(W1):W636-W41.
79. Biedermann P, Klink P, Nocke MK, Papp CP, Harms D, Kebelmann M, Thürmer A, Choi M, Altmann B, Todt D, Hofmann J, Bock CT. Insertions and deletions in the hypervariable region of the hepatitis E virus genome in individuals with acute and chronic infection. *Liver Int*. 2023;43(4):794-804.
80. Mannervik B, Awasthi YC, Board PG, Hayes JD, Di Ilio C, Ketterer B, Listowsky I, Morgenstern R, Muramatsu M, Pearson WR. Nomenclature for human glutathione transferases. *The Biochemical Journal*. 1992;282 ( Pt 1):305-6.

81. Villordo SM, Carballeda JM, Filomatori CV, Gamarnik AV. RNA Structure Duplications and Flavivirus Host Adaptation. *Trends Microbiol.* 2016;24(4):270-83.
82. Eshaghi A, Duvvuri VR, Lai R, Nadarajah JT, Li A, Patel SN, Low DE, Gubbay JB. Genetic variability of human respiratory syncytial virus A strains circulating in Ontario: a novel genotype with a 72 nucleotide G gene duplication. *PLoS One.* 2012;7(3):e32807.
83. Trento A, Galiano M, Videla C, Carballal G, García-Barreno B, Melero JA, Palomo C. Major changes in the G protein of human respiratory syncytial virus isolates introduced by a duplication of 60 nucleotides. *J Gen Virol.* 2003;84(Pt 11):3115-20.
84. Hotard AL, Laikhter E, Brooks K, Hartert TV, Moore ML. Functional Analysis of the 60-Nucleotide Duplication in the Respiratory Syncytial Virus Buenos Aires Strain Attachment Glycoprotein. *J Virol.* 2015;89(16):8258-66.
85. Kenney SP, Meng X-J. The Lysine Residues within the Human Ribosomal Protein S17 Sequence Naturally Inserted into the Viral Nonstructural Protein of a Unique Strain of Hepatitis E Virus Are Important for Enhanced Virus Replication. *Journal of virology.* 2015;89(7):3793.
86. Lhomme S, Nicot F, Jeanne N, Dimeglio C, Roulet A, Lefebvre C, Carcenac R, Manno M, Dubois M, Peron J-M, Alric L, Kamar N, Abravanel F, Izopet J. Insertions and Duplications in the Polyproline Region of the Hepatitis E Virus. *Frontiers in microbiology.* 2020;11:1.

## Eidesstattliche Versicherung

„Ich, Paula Biedermann, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: Charakterisierung und klinische Relevanz von Virusvarianten für den chronischen Verlauf einer Hepatitis E (Characterization and clinical relevance of virus variants for a chronic infection with hepatitis E virus) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; [www.icmje.org](http://www.icmje.org)) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

---

## Anteilserklärung an den erfolgten Publikationen

Paula Biedermann hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: **Biedermann P**, Klink P, Nocke MK, Papp CP, Harms D, Kebelmann M, Thürmer A, Choi M, Altmann B, Todt D, Hofmann J, Bock CT. **Insertions and deletions in the hypervariable region of the hepatitis E virus genome in individuals with acute and chronic infection**. Liver Int. 2023 Apr;43(4):794-804. doi: 10.1111/liv.15517. Epub 2023 Jan 13. PMID: 36617681.

Beitrag im Einzelnen:

- Maßgeblicher Anteil bei Konzeption der Forschungsarbeit
- Federführung bei der Literaturrecherche
- Auswahl der Proben aus der Labordatenbank
- Alleinige Durchführung der Laborarbeiten (RNA-Extraktion, qPCR, cDNA-Synthese, PCR, Vorbereitung der Proben für Sangersequenzierung und NGS) nach technischer Beratung durch Labormitarbeiter
- Mitentwicklung der Methode zur Datenauswertung mittels bioinformatischer Pipeline
- Hauptsächliche Datenauswertung (Auswertung der Patientendaten, Clusterbildung mittels UP-ARSE Algorithmus; Anwendung bioinformatischer Pipeline, Analyse der Insertionen und Deletionen, Analyse der Aminosäurezusammensetzung, Analyse der Eigenschaften der Sequenzen); dargestellt in Tabelle 1, Tabelle 2, Abbildung 2,3 und 4 sowie Tabelle S1 des „Supplement“
- Statistische Auswertung nach Beratung mit den Koautoren und Koautorinnen in Rstudio der Programmiersprache R; Bestandteil der Abbildungen 3 und 4 sowie der Tabelle S2 des „Supplement“
- Erstellung der Publikation und Umsetzung des Überarbeitungsprozess unter Berücksichtigung der Kommentare der Koautoren und Koautorinnen sowie der Reviewer bis zur Annahme der Publikation

---

Unterschrift, Datum und Stempel des/der erstbetreuenden Hochschullehrers/in

---

Unterschrift des Doktoranden/der Doktorandin

# Druckexemplar der Publikation

Received: 16 September 2022 | Revised: 30 November 2022 | Accepted: 4 January 2023  
DOI: 10.1111/liv.15517

ORIGINAL ARTICLE



## Insertions and deletions in the hypervariable region of the hepatitis E virus genome in individuals with acute and chronic infection

Paula Biedermann<sup>1,2</sup> | Patrycja Klink<sup>1</sup> | Maximilian K. Nocke<sup>3</sup> | Christian-Patrick Papp<sup>1,2</sup> | Dominik Harms<sup>1</sup> | Marianne Kebelmann<sup>1</sup> | Andrea Thürmer<sup>4</sup> | Mira Choi<sup>5</sup> | Britta Altmann<sup>1</sup> | Daniel Todt<sup>3,6</sup> | Jörg Hofmann<sup>2,7</sup> | Claus-Thomas Bock<sup>1,8</sup>

<sup>1</sup>Division of Viral Gastroenteritis and Hepatitis Pathogens and Enteroviruses, Department of Infectious Diseases, Robert Koch Institute, Berlin, Germany

<sup>2</sup>German Centre for Infection Research, Institute of Virology, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin Institute of Health, Berlin, Germany

<sup>3</sup>Department of Molecular and Medical Virology, Ruhr University Bochum, Bochum, Germany

<sup>4</sup>Genome Sequencing, Methodology and Research Infrastructure, Robert Koch Institute, Berlin, Germany

<sup>5</sup>Department of Nephrology and Intensive Medical Care, Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

<sup>6</sup>European Virus Bioinformatics Center (EVBC), Jena, Germany

<sup>7</sup>Labor Berlin, Charité-Vivantes GmbH, Berlin, Germany

<sup>8</sup>Institute of Tropical Medicine, University of Tuebingen, Tuebingen, Germany

### Correspondence

Claus-Thomas Bock, Department of Infectious Diseases, Robert Koch Institute, 13353 Berlin, Germany.  
Email: [bockc@rki.de](mailto:bockc@rki.de)

### Funding information

German Federal Ministry of Health (BMG), Grant/Award Number: ZMVI1-2518FSB705; German Bundestag by the Federal Government; German Research Association DFG, Grant/Award Number: 448974291; German Federal Ministry of Education and Research (BMBF), Grant/Award Number: 01KI2106; Claussen-Simon-Stiftung; Claussen-Simon Foundation; CSF; Fazit-Stiftung 'Promotions Stipendium'; Investitionsbank Berlin, Germany, Grant/Award Number: 10169028; EFRE, Grant/Award Number: 10169096

Handling Editor: Alessio Aghemo

### Abstract

**Background and Aims:** Hepatitis E virus is a major cause of acute hepatitis worldwide and can progress to chronicity in immunocompromised individuals. Various virus–host recombination events have been reported in the hypervariable region of the hepatitis E virus genome, but the patterns of assembly and selection remain unclear.

**Methods:** To gain further insight into viral evolution, we assessed the presence of low abundance variants in 16 samples from individuals with acute or chronic infection using a targeted next-generation sequencing approach.

**Results:** In seven samples, different variants with insertions and/or deletions were identified. Among them, eight insertions originating either from human genes or from the hepatitis E virus genome. Five different deletions could be identified. The amino acid composition of sequences with insertions showed a higher frequency of lysine and a lower abundance of proline, and additionally acetylation and ubiquitination sites were more frequent than in hepatitis E virus wild-type sequences.

**Abbreviations:** ALAT, alanine aminotransferase; ASAT, aspartate aminotransferase;  $\gamma$ -GT,  $\gamma$ -glutamyltransferase; cDNA, complementary DNA; DNA, deoxyribonucleic acid; GSTA1, glutathione S-transferase alpha 1; HEV, hepatitis E virus; HVR, hypervariable region; Indel, insertion and/or deletion; KTR, kidney transplant recipient; NGS, next-generation sequencing; NHL, non-Hodgkin lymphoma; ORF, open reading frame; PCR, polymerase chain reaction; qPCR, quantitative PCR; RdRp, RNA-dependent RNA Polymerase; RNA, ribonucleic acid; RPL18, ribosomal protein L18; RT, reverse transcription; RTX, rituximab.

Daniel Todt, Jörg Hofmann, Claus-Thomas Bock contributed equally to this work and thus shared last authorship.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Liver International* published by John Wiley & Sons Ltd.

**Conclusions:** These findings suggest that the nucleotide composition of insertions and sites for post-translational modification may contribute to recombination events. Although the impact of low-level hepatitis E virus variants is uncertain, our results highlight the importance of a highly sensitive next-generation sequencing approach to capture the full diversity of hypervariable region.

**KEYWORDS**

gene insertion, hepatitis E virus, hypervariable region, next-generation sequencing, recombination, sequence deletion

## 1 | INTRODUCTION

Hepatitis E virus (HEV) is one of the leading causes of viral hepatitis worldwide and responsible for large outbreaks and epidemics with an acute course of disease in resource-poor countries due to waterborne/faecal-oral transmission of the virus.<sup>1</sup> In industrialized countries, sporadic cases with acute and chronic cases occur mainly due to zoonotic transmission.<sup>2</sup> The single-stranded positive sense RNA-virus belongs to the genus *Paslahepevirus*, species *Paslahepevirus balayani* in the family of Hepeviridae and is divided into 8 genotypes of which 4 (HEV-1–HEV-4) are known to be pathogenic in humans.<sup>3,4</sup> In addition, one case was reported in which a person became chronically infected with HEV-7 after consuming camel meat and milk.<sup>5</sup> HEV genotype 3 (HEV-3) is the predominant genotype causing HEV infection in industrialized countries.<sup>6</sup> Over the past decade, an increasing number of autochthonous cases with HEV-3 have been identified, which can lead to a chronic course of infection.<sup>7</sup> Chronic infections primarily occur in immunocompromised individuals, such as organ transplant recipients, patients with haematological disorders, and patients with HIV/AIDS.<sup>8</sup> In immunocompromised individuals, treatment options are scarce with an increased risk of progression to liver cirrhosis.<sup>9</sup> Therefore, chronic HEV infections are an increasingly important clinically and public health relevant issue.

The exact mechanisms leading to chronic infections are still unknown. Host factors resulting in alteration of the immune system may play an important role.<sup>8</sup> The T-cell response has been identified as a key variable in the body's capability to control the virus.<sup>10–12</sup> In addition, viral characteristics may contribute to chronification. The hypervariable region (HVR) shows a highly divergent nucleotide composition compared with other regions of the HEV genome and is known to influence virus adaption *in vitro* and *in vivo*.<sup>13,14</sup> In recent studies, insertions in the HEV HVR have been shown to enhance replication ability *in vitro*.<sup>15–19</sup> Furthermore, virus-host recombinants have been detected in samples of chronically infected individuals, suggesting a potential role of HVR diversity in the development of persistence.<sup>14–20</sup> Recently, insertions have also been detected in an acutely infected individual.<sup>21</sup> However, the impact of these recombination events in the HVR on the chronic course of the infection remains unclear.

### Key Points

Hepatitis E Virus infections lead to acute or chronic courses. In chronically infected patients, fragments of human mRNA or of the own viral genome inserted into the hypervariable region have previously been reported. In this study, we found previously unknown genetic rearrangements and investigated their impact on protein sequences and their characteristics.

The aim of this study was to analyse and characterize the HVR variants of individuals with acute and chronic HEV infection to provide detailed evidence of the involvement of HVR variants in the chronification of HEV infection. The application of a targeted NGS approach allows the detection of low-frequency viral variants that may carry insertions that would not be detectable with Sanger sequencing. In addition, the insertions detected in the sequences of the analysed samples were characterized in terms of amino acid composition and post-translational modifications to provide further evidence on the potential role of HVR in the chronification of HEV infection.

## 2 | MATERIALS AND METHODS

### 2.1 | Patient samples

Sixteen blood samples including 15 EDTA-plasma samples and one serum sample from HEV-infected individuals were analysed. The selection was based on availability, availability of required volume, and viral load needed for PCR and sequencing. Baseline characteristics of the study cohort and individual patient characteristics were summarized in Tables 1 and 2. The samples were obtained from routine diagnostics and were pseudonymized before further investigation. Sample 19-0195 was a follow-up sample of patient 1 described recently by Papp et al.<sup>22</sup> Nine samples were from acute and seven from chronically HEV-infected individuals (Table 2). Chronic HEV infection is defined as the persistence of HEV RNA for more than 3 months.<sup>23</sup>

TABLE 1 Baseline characteristics of the study cohort

	Reference range	Total (n = 16)	Acute HEV (n = 9)	Chronic HEV (n = 7)
Age (years)		57.5 (18–90)	59 (18–90)	56 (24–74)
Sex (ratio m:f)		1:8	9:0	1:2.5
ASAT (U/L)	<50 U/L	56 (27–814)	55 (27–814)	57 (45–152)
ALAT (U/L)	<41 U/L	103.5 (30–1201)	119 (52–1201)	74 (30–134)
Bilirubin (mg/dl)	<1 mg/dl	0.535 (0.09–4.02)	0.58 (0.3–4.02)	0.35 (0.09–3.84)
γ-GT (U/L)	8–61 U/L	111 (60–1401)	109 (60–811)	113 (76–1401)
Viral load (IU/ml)		1.19E+06 (3.24E+04–3.46E+07)	1.58E+06 (3.24E+04–2.92E+07)	6.39E+05 (3.68E+04– 3.46E+07)

Note: Data in median and range.

Abbreviations: m, male; f, female; ASAT, aspartate aminotransferase; ALAT, alanine aminotransferase; γ-GT, γ-glutamyltransferase; HEV, hepatitis E viral infection.

## 2.2 | Ethics statement

This study was approved by the local ethics committee (approval number EA1/367/16) and written informed consent was obtained from all participating individuals. Patient samples were de-identified for this study. All experiments were performed in accordance with relevant guidelines and regulations.

## 2.3 | RNA extraction and quantification

RNA was extracted from 140 μl EDTA-plasma or serum using the QIAamp Viral RNA mini Kit (Qiagen) and the QIAcube (Qiagen) according to the manufacturer's instructions. For quantification, a real-time RT-PCR was performed as described previously.<sup>24</sup>

## 2.4 | Genotyping and phylogenetic analysis

For genotyping and subsequent phylogenetic analysis, a nested RT-PCR assay modified from the protocol described by Wang et al<sup>24</sup> targeting a 307 nt fragment in the ORF 1 was performed. In detail, first-round RT-PCR was performed using the QIAGEN OneStep RT-PCR kit (Qiagen) and a modified primer pair (38\_forward: GAGGC YATGGTSGAGAARG and 39\_reverse: GCCATGTTCCAGACRGT RTTCC). The PCR was performed under the following conditions: 15 min at 95°C followed by 40 cycles consisting of 10 s at 94°C, 30 s at 50°C and extension at 72°C for 50 s and a subsequent final elongation at 72°C for 2 min. For nested PCR, a modified primer pair (37\_forward: GGTCCGYGCTATTGARAARG and 27\_reverse: TCRCCAGAGTGYTTCTTCC) and HotStarTaq Master Mix (Qiagen, Hilden, Germany) were used under the following conditions: 15 min at 95°C, followed by 30 cycles at 94°C for 10 s, 49°C for 30 s, 72°C for 20 s and a final elongation step of 2 min at 72°C. Due to the short fragment length and lack of further NGS sequencing, no proof-reading polymerase was deemed necessary for this amplification step. The fragment was primer and quality trimmed to 256 nt. The HEV geno- and subtypes were assigned by phylogenetic

analysis using genotype 3 reference sequences according to the latest proposal of the ICTV for Hepeviridae and the most closely related sequence of each sample from GenBank according to the Max Score value.<sup>25</sup> To generate a cladogram, all genotype 3 reference sequences described earlier and genotype 3 sequences from the HEV-GLUE Sequence Database were downloaded.<sup>25,26</sup> A multi-sequence alignment (MSA) of all sequences was built using Clustal-Omega version 1.2.4.<sup>27</sup> Based on this MSA, all loaded sequences were cut to the same genomic section as the patient sequences. A second MSA was created by QIAGEN CLC Genomics Workbench 22.0 (<https://digitalinsights.qiagen.com/>) providing the basis for the cladogram, which was computed and visualized using the neighbour-joining construction method from the same tool.

## 2.5 | Amplification and sequencing of the HVR

RNA was reverse transcribed using the SuperScript IV Reverse Transcriptase Synthesis System (Invitrogen) according to the manufacturer's recommendations with a prolonged incubation step of 20 min at 50°C and 15 min at 55°C using oligo(dT)s and random hexamers. The nested PCR and amplicon preparation was performed according to Papp et al<sup>22</sup> with the following modifications: The Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific) was used and the first PCR was performed under the following conditions: 30 s at 98°C and 35 cycles of 5 s at 98°C, 30 s at 62°C, 20 s at 72°C, followed by 2 min at 72°C. Conditions of the second PCR were: 30 s at 98°C, 13 cycles of 5 s at 98°C, 30 s at 62°C and 10 s at 72°C, 15 cycles of 5 s at 98°C and 30 s at 72°C, followed by 2 min at 72°C.

## 2.6 | Next-generation sequencing of HVR

Next-Generation sequencing was performed on the MiSeq Sequencer (Illumina, Inc.) with paired-end technology and 300 bp read length as previously described.<sup>22</sup> In addition, sequences were confirmed by Sanger sequencing performed as described recently.<sup>22</sup>

TABLE 2 Patient characteristics

Sample ID	Sex	Age <sup>a</sup>	AST <sup>b</sup> (U/L)	ALT <sup>b</sup> (U/L)	Bilirubin <sup>b</sup> (mg/dl)	γ-GT <sup>b</sup> (U/L)	Days since first positive PCR	Status of infection	Reason for impaired immune response	Viral load (IU/ml) (sample type)
19-0142	M	33	46	109	0.98	60	27	Acute	KTR	1.58E+06 (plasma)
19-0143	M	41	77	160	0.55	811	0	Acute		1.47E+06 (plasma)
19-0144	W	56	49	44	0.52	102	1679	Chronic	KTR	5.72E+05 (plasma)
19-0150	M	63	57	106	0.35	118	187	Chronic	KTR	3.68E+04 (plasma)
19-0151	M	66	45	46	0.26	113	359	Chronic	KTR	9.07E+05 (plasma)
19-0157	M	54	48	30	0.46	167	265	Chronic	KTR	1.48E+07 (plasma)
19-0162	M	18	27	52	0.55	61	6	Acute	RTX-patient	8.50E+06 (plasma)
19-0164	M	56	814	1201	0.61	495	0	Acute		2.92E+07 (plasma)
19-0166	M	90	n.a.	n.a.	n.a.	n.a.	0	Acute		3.18E+05 (plasma)
19-0171	M	74	60	134	0.09	76	686	Chronic	KTR	5.31E+04 (plasma)
19-0174	M	62	54	119	0.58	109	69	Acute	KTR	2.14E+06 (serum)
19-0194	M	90	208	265	4.02	152	0	Acute		3.24E+04 (plasma)
19-0195	M	24	152	74	3.84	1401	404	Chronic	RTX-patient, Bruton's disease <sup>c</sup>	3.46E+07 (plasma)
19-0196	W	50	68	101	0.35	93	489	Chronic	IgG-deficiency, T-lymphopaenia	6.39E+05 (plasma)
19-0201	M	59	n.a.	n.a.	n.a.	n.a.	8	Acute	NHL	3.13E+06 (plasma)
19-0204	M	64	55	63	0.3	63	0	Acute	Immuno-suppression	4.09E+05 (plasma)

Abbreviations: KTR, kidney transplant recipient; RTX, rituximab; NHL, non-Hodgkin lymphoma; n.a., not available.

<sup>a</sup>Age - age in years at sample date.

<sup>b</sup>For reference ranges see Table 1.

<sup>c</sup>Agammaglobulinemia.

## 2.7 | Bioinformatics

The different amplicon sequences with their different indel-patterns, each representing at least one viral haplospecies, can be handled as molecular 'operational taxonomic units' (OTU). Using the tool UPARSE, amplicon sequences were clustered into OTU's based on similarity.<sup>28</sup> Similar OTUs were grouped by length and their indel-pattern and within each group, the OTU with the largest cluster size was picked as group representative. Subsequently, we offered all representative OTUs as mapping reference in a python-based pipeline, which originally was designed for the molecular surveillance of HIV and includes a merging step using FLASH (version 1.2.11), trimming with Trimmomatic (version 0.36) and using the BWA mapper (version 0.7.15).<sup>29-32</sup> Briefly, paired sequences were merged using FLASH with a minimum overlap of 15 and a maximum overlap of 300bp and a max-mismatch density of 0.25. Merged reads as well as unmerged reads were than quality trimmed by Trimmomatic with the options sliding window (8:20) and minimum sequence length of 50 base pairs. Subsequently, all survived reads were mapped against a reference file containing all representative OTUs of one sample plus the wild-type HVR sequence, using the mem-algorithm of BWA with a mismatch-penalty of 2. The resulting alignments were analysed by their codon frequencies and a consensus sequence for each offered reference was generated based on the most frequent codon-variants. Samtools (version 1.11) was used to analyse the distribution of reads between

the offered reference sequences.<sup>33</sup> Sequences with at least 100 reads and a minimum share of 0.5% reads were included. For the following analyses, the sequences were assigned into three different groups: sequences without indels (insertions and/or deletions), with deletions, and with insertions. Sequences containing both an insertion and a deletion were assigned to the group 'sequences with insertions'. The programming language Julia version 1.7.3 was utilized for data processing and pipeline organization, while R version 4.2.0 was used for visualization.<sup>34,35</sup> For visualization, the HEV-3c reference strain wbGER27 (Acc. No. FJ705359.1) as proposed by Smith et al<sup>25</sup> was used.

## 2.8 | Analysis of regulation sites and of amino acid characteristics

To analyse the amino acid composition and for the creation of a graph, the biostrings package from the 'bioconductor software project' based on the programming language R was used.<sup>35,36</sup> The following tools were used for the in silico determination of putative post-translational modifications: prediction of acetylation on internal lysines: GPS-PAIL with medium threshold and all putative acetylation (<http://pail.biocuckoo.org/>)<sup>37</sup>; prediction of ubiquitination: BDM-PUB with a balanced cut-off (<http://bdmpub.biocuckoo.org/>); prediction of phosphorylation: (<http://www.cbs.dtu.dk/services/NetPhos/>)<sup>38</sup>; prediction of glycosylation: (<http://www.cbs.dtu.dk/services/NetPhos/>)<sup>38</sup>.

dtu.dk/services/NetNGlyc/)<sup>39</sup>; prediction of methylation: (<http://www.jci-bioinfo.cn/iMethyl-PseAAC>).<sup>40,41</sup> For the analysis of amino acid characteristics, the "Emboss Pepstats Tool" was used ([https://www.ebi.ac.uk/Tools/seqstats/emboss\\_pepstats/](https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/)).<sup>42</sup>

## 2.9 | Statistical analysis

Quantitative variables were compared by employing the Wilcoxon rank sum test. The analysis was performed using the programming language R.<sup>35</sup>  $p < .05$  were considered statistically significant.

## 2.10 | Nucleotide sequence accession numbers

Sequences generated using NGS have been deposited in the NCBI GenBank under the accession numbers OL469119–OL469148.

## 3 | RESULTS

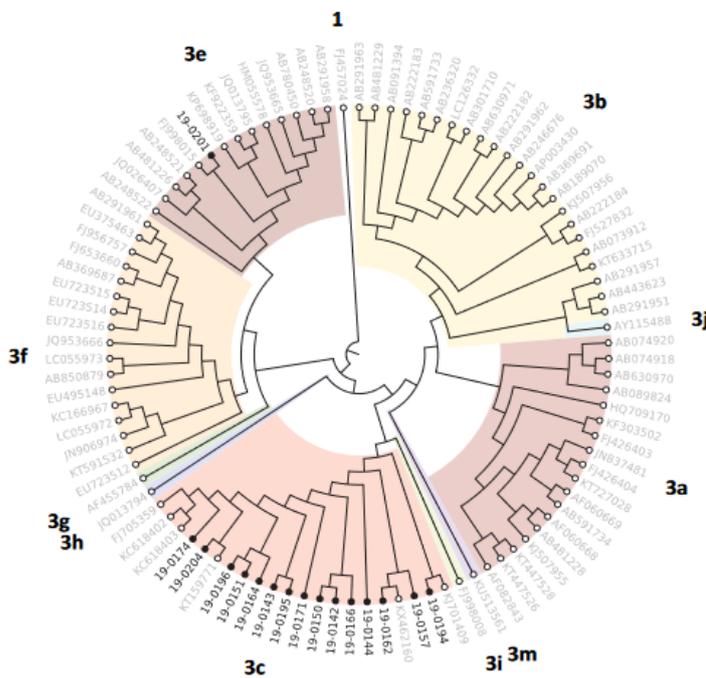
### 3.1 | Characteristics of the study cohort

Samples of 16 HEV-positive individuals were analysed, of which seven were chronically infected and nine had an acute infection. Fourteen individuals were male and two were female. The median age was 56 and 59 in the chronically and acutely infected group with an age range

between 24 and 74 years and 18–90 years, respectively (Table 1). The enzymes aspartate aminotransferase (AST), alanine aminotransferase (ALT),  $\gamma$ -glutamyltransferase ( $\gamma$ -GT) were moderately elevated with no significant differences between acutely or chronically infected individuals. Furthermore, there was no significant difference regarding the viral load between the group of acutely infected and the group of chronically infected patients (Wilcoxon rank sum test; data not shown) (Table 1). The duration of infection among the chronically infected patients ranged from 6 to 55 months (Table 2). Out of seven chronically infected individuals, five were kidney transplant recipients and two suffered from immunodeficiencies (Table 2). Two patients from the group of acutely infected patients developed a chronic infection (19-0174 and 19-0204). All sequences were assigned to HEV genotype 3. Fifteen sequences clustered within HEV subtype 3c and one sequence assigned to HEV subtype 3e (Figure 1).

### 3.2 | NGS amplicon sequencing of HVR

Sequences without indels were present in all samples. In addition, insertions and deletions were detected in the HVR domain of seven samples (Table S1). All indel-patterns found were in frame. In total, eight different insertions were identified in the HVR region of four samples, of which three were derived from individuals in a chronic state of infection (19-0144, 19-0195, 19-0196) and one from an acutely infected patient (19-0174) (Figure 2). All quasispecies with an indel contained more than one indel variant (Figure 2 and Table S1). The insertions



**FIGURE 1** Cladogram of the patient-derived HEV sequences based on partial ORF1 sequences (256 nt). Subgenotyping results were confirmed with HEVnet (<https://www.rivm.nl/mpf/typingtool/hev/>). The HEV genotype 1 sequence FJ457024 was set as an outgroup. HEV sequences from this study are marked in black font (e.g. 19-0194) and were compared with HEV reference sequences representing HEV-3 (depiction: HEV subtype, accession number, e.g., 3e FJ998015).

ranged from 96 to 162 nt in size and were present with frequencies between 0.8% and 95.9% (Table S1). Among these eight insertions, five were derived from the HEV genome itself, either from the HVR or the RNA-dependent RNA polymerase (RdRp) region (19-0144, 19-0174, 19-0195, 19-0196; Figure 2 and Table S1). The remaining three insertions were derived from human genes (19-0195, 19-0196; Figure 2 and Table S1). Of these, one insertion showed the highest similarity with the human gene for the nucleoprotein Ahnak (19-0195). Ahnak is located on chromosome 11 and possibly plays a role in tumour metastasis and calcium regulation in cardiomyocytes and acts as a scaffold protein.<sup>43</sup> The insertion consisted of two Ahnak fragments (Figure 2 and Table S1). The second insertion in the HVR sequence of the sample 19-0195 was derived from the ribosomal protein L18 (RPL18). The remaining human-derived insertion in the HVR sequence of sample 19-0196 showed the highest identity to Glutathione S-transferase A1 (GSTA1), a gene that is abundantly expressed within the liver (Figure 2 and Table S1).<sup>44</sup> All insertions deriving from the HEV genome itself were made up of at least two fragments with one of them being a duplication of the adjacent region (Figure 2). Three insertions consisted of one HVR and one RdRp fragment. Two insertions were a combination of two different HVR fragments.

Deletions were observed in the HEV viral quasispecies of five samples (19-0142, 19-0151, 19-0157, 19-0174, 19-0196; Figure 2 and Table S1). Two samples within this group were collected from acutely infected patients (19-0142, 19-0174) and three from chronically infected patients (19-0151, 19-0157, 19-0196). The deletions ranged from 6 to 45 nt in size and were present in frequencies between 0.5% and 87.6% (Table S1). In the HVR sequence of sample 19-0174, a combination of two deletions consisting of a 7 and an 8 nt deletions was detected (Figure 2 and Table S1). Two variants were detected which featured both a deletion and an insertion in one strain (19-0174, 19-0196; Figure 2 and Table S1).

In addition, one 90 nt insertion not meeting our inclusion criteria (frequency: 0.4%) was identified in the HEV viral quasispecies of sample 19-0174 which was derived from the human gene NCKAP1 and was in frame (data not shown). The NCKAP1 Gene is located on chromosome 2 and is part of the WAVE complex that regulates lamellipodia formation.<sup>45</sup>

### 3.3 | Position and fragment composition of insertions

In total, the position of the identified insertions varied, but all were located between nucleotide position (nt) nt 2227 and nt 2400

(numbering according to the HEV reference strain FJ705359.1). No pattern in regard to the nucleotides and amino acid composition before and after the insertion was identified.

### 3.4 | Amino acid composition of the HVR

For the amino acid analysis, the sequences were divided into three different groups: (1) sequences without indels, (2) sequences with insertions, and (3) sequences with deletions. Furthermore, the ratio of specific amino acids within each group (group (1)–(3)) was analysed (Figures 3 and 4). All the analysed sequences were rich in alanine (A), proline (P), and serine (S). The sequences with insertions contained significantly more lysine (K) (5.5% vs. 2.2%;  $p < .0001$ ) and glycine (G) (4% vs. 2.5%;  $p$ -value: .013) compared with the groups without insertions (groups 1 and 3). On the other hand, these sequences (group 2) contained significantly less proline (P) (23.3% vs 28.5%;  $p$ -value: .00064), glutamate (E) (4.2% vs. 4.7%;  $p$ -value: .040) and valine (V) (5.6% vs. 6.8%;  $p$ -value: .032).

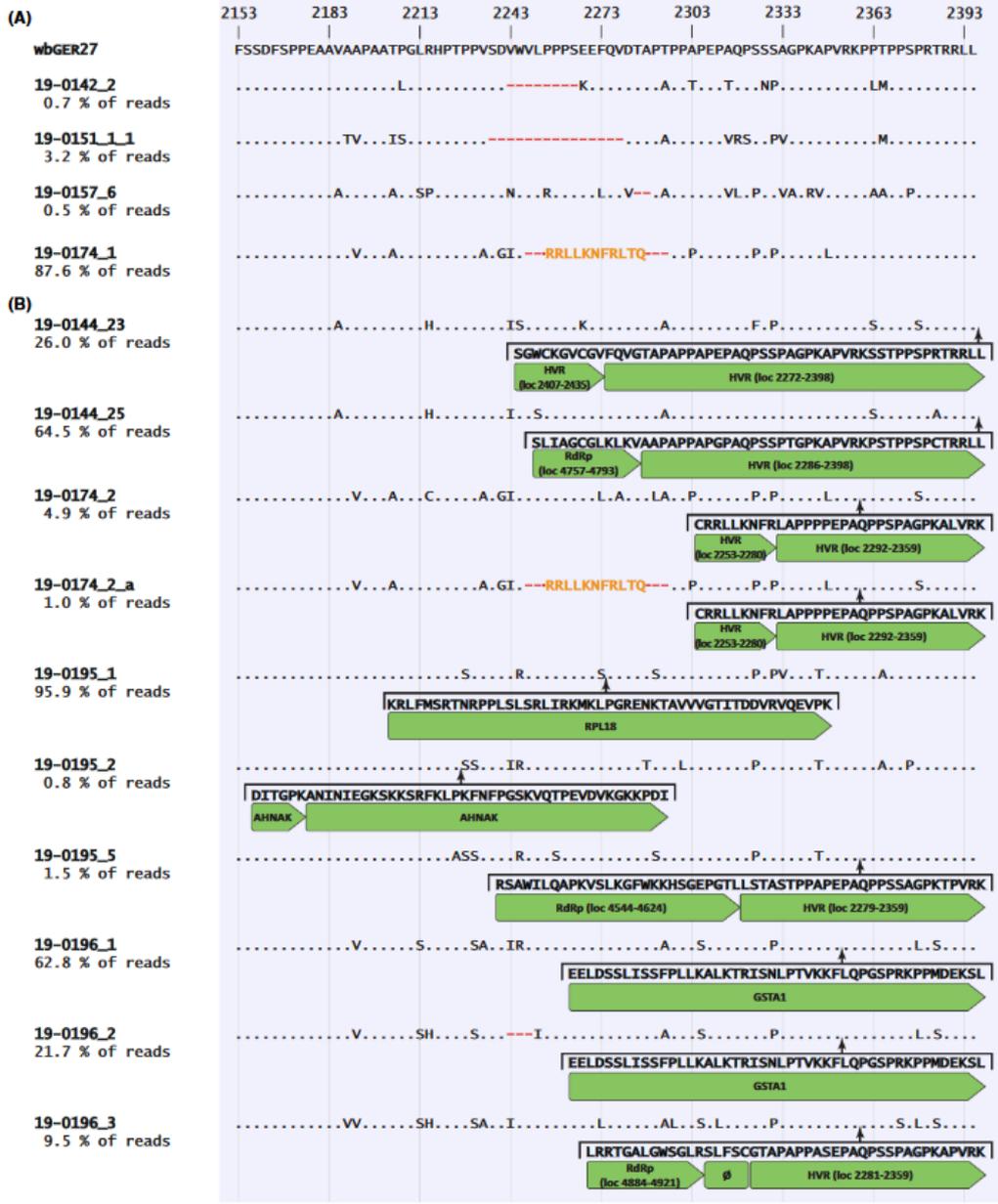
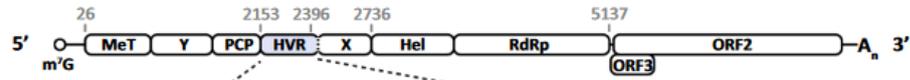
Sequences with deletions consisted of significantly more arginine (R) (8.4% vs. 6.2%;  $p$ -value: .0062), asparagine (N) (1.0% vs. 0.3%;  $p$ -value: .017), histidine (H) (1.3% vs. 1.2%;  $p$ -value: .012), and lysine (K) (3.0% vs. 2.2%;  $p$ -value: .048) compared with sequences without indels (group 1).

Sequences with insertions displayed significantly more polar ( $p$ -value: .029) and basic amino acids ( $p < .0001$ ) and less small ( $p$ -value: .0023) and non-polar amino acids ( $p$ -value: .029) compared with sequences without indels (Figure 4). Sequences with deletions consisted of more basic ( $p$ -value: .014) and less acidic amino acids ( $p$ -value: .0065) (Figure 4). Further details are given in Table S2.

### 3.5 | Characteristics of sequences

Sequences with insertions provided significantly more putative sites for acetylation ( $p$ -value: .010), methylation (lysine) ( $p$ -value: .0018), and ubiquitination ( $p$ -value: .010) compared with sequences without indels (Figure 4). In detail, sequences with insertions provided between 3.2 and 8.6 putative sites for ubiquitination and acetylation per 100 amino acids in comparison to 1.2 and 2.4 putative sites in sequences without indels. The sequences without indels carried between one and two putative post-translational modification sites for acetylation and ubiquitination. The number of putative post-translational modifications was high in the sequences with

**FIGURE 2** Amino acid (aa) sequence alignment of HEV HVR genomic regions. Depiction of insertions and deletions in the HEV HVR region of samples from HEV-infected individuals. The pseudonym of the patient is indicated above each aa sequence alignment (e.g., 19-0144). The upper sequence depicts the HEV reference strain wbGER27 (FJ705359.1) and numbering is according to this HEV reference strain. A hyphen indicates a deletion reference strain. The green bars indicate the human or HEV sequence insertion in the HEV HVR region. (A) displays sequences with deletions only, while (B) shows all sequences with insertions. Below the HVR variant name, the percentage of reads is displayed. For more detailed information, view Table S1. HEV sequence data of the patient samples are available at NCBI GenBank database (Acc. No. OL469119–OL469148).



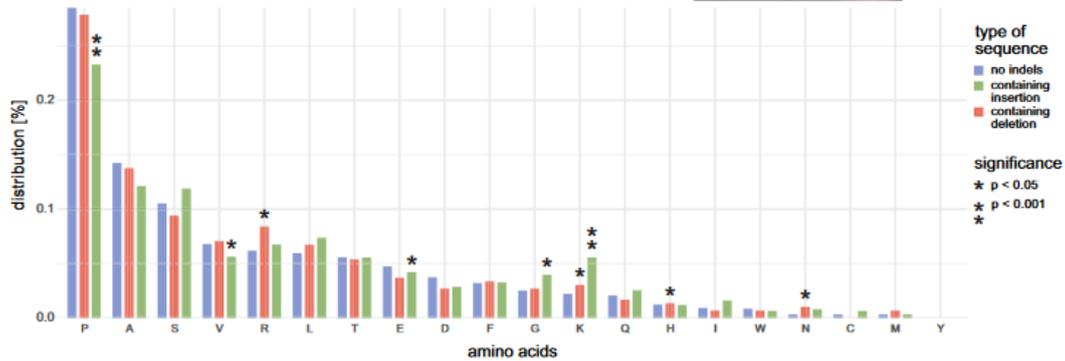


FIGURE 3 Amino acid composition of the HVR sequences of the three groups: Sequences without indels (in blue), with insertions (in green), and with deletions (in red). The mean percentage composition of amino acids is indicated on the left (y-axis). The single letter code was used to represent the amino acids (x-axis). The amino acids are ranked from most common to least common according to the group of sequences without indels. Asterisks indicate significant statistical differences between sequences without indels compared with sequences with insertions or sequences with deletions.  $p < .05^*$  and  $< .001^{**}$  indicate statistical significance.

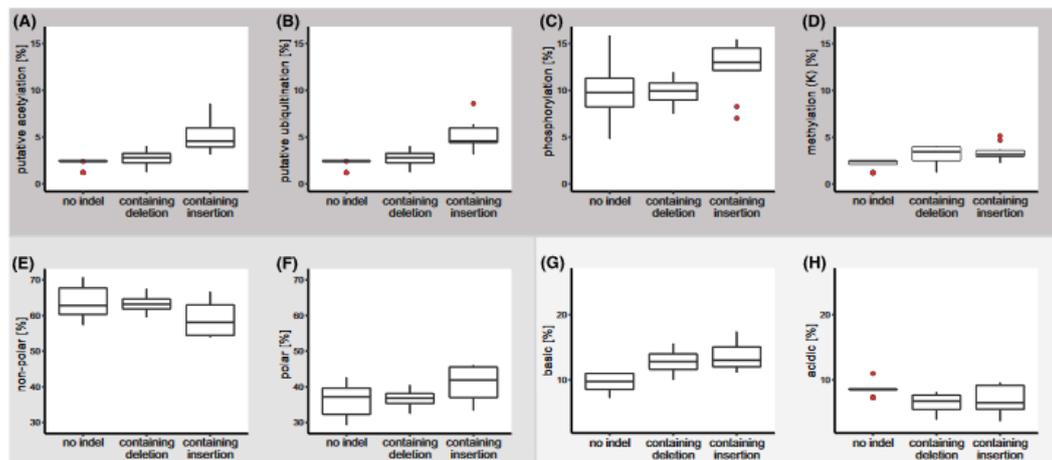


FIGURE 4 Comparison of potential post-translational modifications (dark grey background, A–D) and amino acid characteristics, that is, polarity (medium grey background, E, F) as well as ratio of acidic and basic amino acids (light background, G, H) in sequences without indels, sequences with deletions and sequences with insertions. Boxplots range from minimum to maximum, excluding outliers, which are represented by red dots. Second and third quartiles are displayed as a box divided by the median. The data shown are a subset of Table S2 (Supplement), which points out differences between the three sequence groups.

the Ahnak insertion (19-0195\_2) which carried the most sites with 11 putative acetylation and ubiquitination sites. Furthermore, the sequences with the GSTA1 insertion (19-0196\_1, 19-0196\_2) provided the second most ubiquitination and acetylation sites with eight putative post-translational modification sites. Sequences with deletions provided significantly more sites for putative acetylation ( $p$ -value: .048), ubiquitination ( $p$ -value: .048) and methylation (lysine) ( $p$ -value: .037) post-translational modifications with absolute numbers of putative post-translational modification sites ranging between one and three per sequence (Figure 4).

## 4 | DISCUSSION

Over the past decade, the awareness and assessment of the risk potential of the HEV infection has changed. This is partly due to the recent observation that autochthonous HEV infections in industrialized countries can cause severe chronic hepatitis in immunocompromised individuals, and partly due to new sensitive and specific molecular HEV detection methods that allow certain detection of the virus. However, the knowledge of the viral life cycle, molecular virology, and pathogenesis of HEV infection is still incomplete.

Recent studies revealed that recombination events in the HEV HVR can enhance replication ability *in vitro* and can influence virus adaptation *in vitro* and *in vivo*.<sup>13–19</sup>

To gain further insights into HEV HVR recombination events, we analysed the sequences of the HVR obtained from 16 samples of individuals with acute or chronic HEV infection using a targeted NGS approach. This amplicon-based NGS approach also allows the detection of low-frequency variants in the viral population that would go unnoticed with Sanger sequencing.<sup>22</sup> In this study cohort, all but two of the individuals were men, and the mean age was 57.5 years. This observation is not unexpected, as patients infected with HEV are known to be predominantly middle aged to elderly men.<sup>46</sup> In contrast to other findings where higher elevations of ALT were observed in acutely HEV-infected patients compared with chronically infected patients, we detected no significant difference (Tables 1 and 2).<sup>1</sup> In the phylogenetic analysis, all samples but one were assigned to HEV subgenotype 3c (Figure 1), the most common HEV subgenotype in Germany.<sup>47</sup>

By applying the targeted NGS strategy, insertions, deletions, and a mixture of both were identified in the HVR quaspecies of seven samples (Figure 2). Eight different insertions in the sequences of four patient samples were identified, all of which showing more than one recombinant variant in their viral population (Figure 2 and Table S1). The insertions derived either from the HVR and/or the RdRp region of the HEV genome or from human genes such as Anhak, GSTA1, and RPL18 (Table S1). An insertion deriving from GSTA1 gene has not been described before; however, Anhak and RPL18 sequences have already been reported recently<sup>22</sup>; the RPL18 fragment identified in our study was 3 nt shorter. Viral-host recombinants in the HVR region including duplications of the HVR and insertions of the RdRp have already been detected in chronically HEV-infected individuals and recently in one acutely infected individual.<sup>14–17,19–21</sup> In this study, three of the samples with insertions in the HVR sequences were obtained from chronically HEV-infected individuals (19-0144, 19-0195, 19-0196) and one from an acutely HEV-infected individual (19-0174). The sample from the acutely infected individual was drawn 69 days after the first positive PCR and the patient became chronically infected afterwards. Further investigations are necessary to assess whether screening for insertions using a sensitive NGS approach could become a powerful tool in a personalized medicine approach to identify individuals at risk for developing a chronic infection.

Analysis of the amino acid composition of the sequences identified in this study revealed significant differences between HEV HVR variants with and without insertions as previously reported (Figure 3).<sup>14,21</sup> The sequences with insertions consisted of significantly more lysine and glycine and less proline and valine which is consistent with the results reported by Lhomme et al.<sup>21</sup> However, especially the decrease in proline and the increase in lysine are remarkable. Proline is the key characteristic of this genome region.<sup>48</sup> Lysine, on the other hand, offers broad possibilities for post-translational modifications, and thus increased lysine residues indicate the possible significance in regulatory processes. In our observation,

sequences with insertions also contained significantly more positively charged amino acids (Figure 4). In Lhomme et al.<sup>21</sup> human insertions increased the proportion of positively charged amino acids and HEV duplications decreased the proportion of negatively charged amino acids. In Munoz et al.,<sup>20</sup> human insertions equally increased the proportion of positive amino acids, but HEV duplications increased the proportion of negative amino acids. However, we did not subdivide into human insertions and HEV duplications for the calculations in this study. Overall, these observations are of particular importance as Scholz et al.<sup>49</sup> showed in reverse genetics experiments that the amino acid sequences not the nucleotide sequences or the length of the inserted fragment seemed to be important for the deregulation of viral replication.

The analysis of post-translational modifications revealed that HVR sequences with insertions provided significantly more sites for post-translational acetylation than sequences without indels. Furthermore, an increase in the putative sites for methylation (lysine) was observed in this study which has not been described before. Comparably with other reports, the increase in putative acetylation and ubiquitination sites in HVR sequences with insertions could be confirmed, but an increase in phosphorylation or in glycosylation sites was not observed, contrary to previous reports.<sup>20,21</sup> The significance of the increased number of post-translational modification sites remains still unclear, however, as no further functional analyses were performed in our study.

In the analysed sequences, deletions in the HVR sequences of five patient samples could be identified. In our examination, sequences with deletions also provided significantly more putative sites for acetylation, ubiquitination, and methylation compared with sequences without indels. Deletions in the HEV genome have only been described by Nyugen et al so far and most recently by Munoz et al.<sup>19,20</sup> While the deletions detected by Munoz et al were only up to 30 nt in length, Nguyen et al observed large deletions in the HVR of a sample from a chronically HEV-infected patient. With a range from 42 nt to more than 700 nt, the size of the detected deletions was much larger compared with the deletions observed in this study. However, using our NGS approach, deletions of this dimension (up to 700 nt) could not have been detected. Equally larger insertions could possibly go unnoticed due to the limited amplicon size with a 2×300 bp sequencing approach. Another limitation of this study is the limited sample size. To gain a deeper insight into the prevalence and structure of indels in the HVR region larger cohorts should be studied.

## 5 | CONCLUSIONS

The data presented here supports the hypothesis that the amino acid composition with its increased proportion of lysine residues and putative post-translational modification sites within the HVR region is significant for the sustainability and breakthrough of indels in the persistence of HEV infection. Further studies are needed to evaluate the significance of an insertion in the acute stage of disease and later progression to chronic infection and whether this could be a

screening parameter to identify individuals at risk for a prolonged and potentially more severe course of HEV infection as part of a personalized medicine approach in the future.

#### ACKNOWLEDGEMENTS

We are grateful to Steffen Zander (RKI) and the staff members of the laboratory of the FG15 and MF2 at RKI for their excellent technical assistance. Open Access funding enabled and organized by Projekt DEAL.

#### FUNDING INFORMATION

This research was funded by grants from the German Federal Ministry of Health (BMG) with regard to a decision of the German Bundestag by the Federal Government (CHED-project grant No: ZMVI1-2518FSB705). D.T. was supported by a grant from the German Research Association DFG (Project number 448974291) and by the German Federal Ministry of Education and Research (BMBF) [project: VirBio, grant number: 01KI2106]. D.H. is supported by the Claussen-Simon-Stiftung (Claussen-Simon Foundation; CSF) 'Dissertation Plus' program, Germany, and the Fazit-Stiftung 'Promotions Stipendium'. B.A. is supported by ProFIT grant of the Investitionsbank Berlin, Germany (IBB, ProFIT No. 10169028 co-funded by EFRE no. 10169096). The funders BMG, DFG, BMBF, CSF, Fazit, and ProFit had no role in the design of the study, in the collection, analyses or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ETHICS APPROVAL AND PATIENT CONSENT STATEMENT

This study was approved by the local ethics committee (approval number EA1/367/16) and written informed consent was obtained from all participating individuals. Patient samples were de-identified for this study. All experiments were performed in accordance with relevant guidelines and regulations.

#### ORCID

Maximilian K. Nocke  <https://orcid.org/0000-0003-4632-525X>

Daniel Todt  <https://orcid.org/0000-0002-3564-1014>

Claus-Thomas Bock  <https://orcid.org/0000-0002-2773-486X>

#### REFERENCES

- Kamar N, Bendall R, Legrand-Abravanel F, et al. Hepatitis E. *Lancet*. 2012;379(9835):2477-2488.
- Webb GW, Dalton HR. Hepatitis E: an underestimated emerging threat. *Ther Adv Infect Dis*. 2019;6:2049936119837162.
- Purdy MA, Drexler JF, Meng XJ, et al. ICTV virus taxonomy profile: hepeviridae 2022. *J Gen Virol*. 2022;103(9):1-2. doi:10.1099/jgv.0.001778
- Walker PJ, Siddell SG, Lefkowitz EJ, et al. Recent changes to virus taxonomy ratified by the international committee on taxonomy of viruses (2022). *Arch Virol*. 2022;167:2429-2440.
- Lee G-H, Tan B-H, Teo EC-Y, et al. Chronic infection with camelid hepatitis E virus in a liver transplant recipient who regularly consumes camel meat and Milk. *Gastroenterology*. 2016;150(2):355-357.e3.
- Izopet J, Tremeaux P, Marion O, et al. Hepatitis E virus infections in Europe. *J Clin Virol*. 2019;120:20-26.
- Kamar N, Selves J, Mansuy J-M, et al. Hepatitis E virus and chronic hepatitis in organ-transplant recipients. *N Engl J Med*. 2008;358(8):811-817.
- Kamar N, Izopet J, Dalton HR. Chronic hepatitis e virus infection and treatment. *J Clin Exp Hepatol*. 2013;3(2):134-140.
- Kamar N, Garrouste C, Haagsma EB, et al. Factors associated with chronic hepatitis in patients with hepatitis E virus infection who have received solid organ transplants. *Gastroenterology*. 2011;140(5):1481-1489.
- Narayanan S, Abutaleb A, Sherman KE, Kottitil S. Clinical features and determinants of chronicity in hepatitis E virus infection. *J Viral Hep*. 2019;26(4):414-421.
- Suneetha PV, Pischke S, Schlaphoff V, et al. Hepatitis E virus (HEV)-specific T-cell responses are associated with control of HEV infection. *Hepatology*. 2012;55(3):695-708.
- Lhomme S, Abravanel F, Dubois M, et al. Hepatitis E virus quasispecies and the outcome of acute hepatitis E in solid-organ transplant patients. *J Virol*. 2012;86(18):10006-10014.
- Purdy MA, Lara J, Khudyakov YE. The hepatitis E virus Polyproline region is involved in viral adaptation. *PLoS One*. 2012;7(4):e35974.
- Lhomme S, Abravanel F, Dubois M, et al. Characterization of the polyproline region of the hepatitis E virus in immunocompromised patients. *J Virol*. 2014;88(20):12017-12025.
- Shukla P, Nguyen HT, Faulk K, et al. Adaptation of a genotype 3 hepatitis E virus to efficient growth in cell culture depends on an inserted human gene segment acquired by recombination. *J Virol*. 2012;86(10):5697-5707.
- John R, Reetz J, Ulrich RG, et al. An ORF1-rearranged hepatitis E virus derived from a chronically infected patient efficiently replicates in cell culture. *J Viral Hep*. 2014;21(6):447-456.
- Shukla P, Nguyen HT, Torian U, et al. Cross-species infections of cultured cells by hepatitis E virus and discovery of an infectious virus-host recombinant. *Proc Natl Acad Sci USA*. 2011;108(6):2438-2443.
- Kenney SP, Meng X-J. The lysine residues within the human ribosomal protein S17 sequence naturally inserted into the viral non-structural protein of a unique strain of hepatitis E virus are important for enhanced virus replication. *J Virol*. 2015;89(7):3793-3803.
- Nguyen HT, Torian U, Faulk K, et al. A naturally occurring human/hepatitis E recombinant virus predominates in serum but not in faeces of a chronic hepatitis E patient and has a growth advantage in cell culture. *J Gen Virol*. 2012;93(3):526-530.
- Muñoz-Chimeno M, Cenalmor A, Garcia-Lugo MA, Hernandez M, Rodriguez-Lazaro D, Avellon A. Proline-rich hypervariable region of hepatitis E virus: arranging the disorder. *Microorganisms*. 2020;8(9):1417.
- Lhomme S, Nicot F, Jeanne N, et al. Insertions and duplications in the polyproline region of the hepatitis E virus. *Front Microbiol*. 2020;11:1.
- Papp CP, Biedermann P, Harms D, et al. Advanced sequencing approaches detected insertions of viral and human origin in the viral genome of chronic hepatitis E virus patients. *Sci Rep*. 2022;12(1):1720.
- European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu; European Association for the Study of the Liver. EASL clinical practice guidelines on hepatitis E virus infection. *J Hepatol*. 2018;68(6):1256-1271.
- Wang B, Harms D, Papp CP, et al. Comprehensive molecular approach for characterization of hepatitis E virus genotype 3 variants. *J Clin Microbiol*. 2018;56(5):e01686-e01617.
- Smith DB, Izopet J, Nicot F, et al. Update: proposed reference sequences for subtypes of hepatitis E virus (species Orthohepevirus A). *J Gen Virol*. 2020;101(7):692-698.

26. HEV-GLUE 2019. Accessed May 12, 2022. <http://hev.glue.cvr.ac.uk/#/home>.
27. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.
28. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10(10):996-998.
29. Machnowska P, Hauser A, Meixnerberger K, et al. Decreased emergence of HIV-1 drug resistance mutations in a cohort of Ugandan women initiating option B+ for PMTCT. *PLoS One*. 2017;12(5):e0178297.
30. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011;27(21):2957-2963.
31. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
32. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. 2010;26(5):589-595.
33. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993.
34. Bezares J, Edelman A, Karpinski S, Shah VB. Julia: a fresh approach to numerical computing. *SIAM Rev*. 2017;59(1):65-98.
35. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021.
36. Pagès H, Aboyou P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. R package version 2.62.0. <https://bioconductor.org/packages/Biostrings2021>
37. Deng W, Wang C, Zhang Y, et al. GPS-PAIL: prediction of lysine acetyltransferase-specific modification sites from protein sequences. *Sci Rep*. 2016;6:39787.
38. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*. 1999;294(5):1351-1362.
39. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004;4(6):1633-1649.
40. Qiu W-R, Xiao X, Lin W-Z, Chou K-C. iUbiq-Lys: prediction of lysine ubiquitination sites in proteins by extracting sequence evolution information via a gray system model. *J Biomol Struct Dyn*. 2015;33(8):1731-1742.
41. Wang P, Xiao X, Chou K-C. NR-2L: a two-level predictor for identifying nuclear receptor subfamilies based on sequence-derived features. *PLoS One*. 2011;6(8):e23505.
42. Madeira F, Ym P, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47(W1):W636-W641.
43. Davis TA, Loos B, Engelbrecht A-M. AHNAK: the giant jack of all trades. *Cell Signal*. 2014;26(12):2683-2693.
44. Sheehan D, Meade G, Foley VM, Dowd CA. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem J*. 2001;360(Pt 1):1-16.
45. Bladt F, Aippersbach E, Gelkop S, et al. The murine Nck SH2/SH3 adaptors are important for the development of mesoderm-derived embryonic structures and for regulating the cellular Actin network. *Mol Cell Biol*. 2003;23(13):4586-4597.
46. Dalton HR, Bendall RP, Rashid M, et al. Host risk factors and autochthonous hepatitis E infection. *Eur J Gastroenterol Hepatol*. 2011;23(12):1200-1205.
47. Adlhoc C, Avellon A, Baylis SA, et al. Hepatitis E virus: assessment of the epidemiological situation in humans in Europe, 2014/15. *J Clin Virol*. 2016;82:9-16.
48. Koonin EV, Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. *Proc Natl Acad Sci USA*. 1992;89(17):8259-8263.
49. Scholz J, Falkenhagen A, John R. The translated amino acid sequence of an insertion in the hepatitis E virus strain 47832c genome, but not the RNA sequence, is essential for efficient cell culture replication. *Viruses*. 2021;13(5):762.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Biedermann P, Klink P, Nocke MK, et al. Insertions and deletions in the hypervariable region of the hepatitis E virus genome in individuals with acute and chronic infection. *Liver Int*. 2023;43:794-804. doi:[10.1111/liv.15517](https://doi.org/10.1111/liv.15517)

## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

## Komplette Publikationsliste

- **Biedermann P**, Klink P, Nocke MK, Papp CP, Harms D, Kebelmann M, Thürmer A, Choi M, Altmann B, Todt D, Hofmann J, Bock CT. Insertions and deletions in the hypervariable region of the hepatitis E virus genome in individuals with acute and chronic infection. *Liver Int.* 2023 Apr;43(4):794-804. doi: 10.1111/liv.15517. Epub 2023 Jan 13. PMID: 36617681.
- Papp CP, **Biedermann P**, Harms D, Wang B, Kebelmann M, Choi M, Helmuth J, Corman VM, Thürmer A, Altmann B, Klink P, Hofmann J, Bock CT. Advanced sequencing approaches detected insertions of viral and human origin in the viral genome of chronic hepatitis E virus patients. *Sci Rep.* 2022 Feb 2;12(1):1720. doi: 10.1038/s41598-022-05706-w. PMID: 35110582; PMCID: PMC8811047.
- Schulz M\*, **Biedermann P\***, Bock CT, Hofmann J, Choi M, Tacke F, Hanitsch LG, Mueller T. Rituximab-Containing Treatment Regimens May Imply a Long-Term Risk for Difficult-To-Treat Chronic Hepatitis E. *Int J Environ Res Public Health.* 2020 Jan 3;17(1):341. doi: 10.3390/ijerph17010341. PMID: 31947836; PMCID: PMC6982013. (\*geteilte Erstautorenschaft)

## Danksagung

Ich möchte mich ausdrücklich bei meinen Betreuern, Prof. Dr. Jörg Hofmann und Prof. Dr. Claus-Thomas Bock, für die Betreuung dieser Promotionsarbeit und die produktive Zusammenarbeit bedanken.

Weiterhin gilt mein Dank den Mitarbeitenden der FG15 „Virale Gastroenteritis- und Hepatitisserreger und Enteroviren“ am Robert Koch-Institut und insbesondere Steffen Zander für die Unterstützung dieser Arbeit während der Labortätigkeiten. Insbesondere möchte ich mich bei Dr. Patrycja Klink und Britta Altmann (FG 15) für die gute Zusammenarbeit während der langwierigen bioinformatischen Auswertung bedanken. Zuletzt möchte ich mich bei Dr. Daniel Todt und Maximilian Nocke der „Abteilung für Molekulare und Medizinische Virologie“ der Ruhr-Universität Bochum für die wichtigen Impulse in der Schlussphase der Publikation bedanken.