Aus dem Institut für Virologie
der Heinrich-Heine-Universität Düsseldorf
Direktor: Univ.-Prof. Dr. med. Jörg Timm

Virusevolution unter Selektionsdruck – Von Sequenzvariationen zu Resistenz, Immunevasion und Genomischer Surveillance

Habilitationsschrift
zur Erlangung der Venia Legendi
für das Fach Virologie
des Fachbereichs Medizin der
Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Dr. rer. nat. Andreas Walker
Düsseldorf, 2025

1 Inhaltsverzeichnis

2 Abkürzungsverzeichnis

bnAbs	breitneutralisierende Antikörper
CD	Cluster of Differentiation
COVID-19	Coronavirus Disease 2019
Ct	cycle-threshold
DAA	Direkt wirkende antivirale Substanz
DNA	Desoxyribonukleinsäure
EMA	Europäische Arzneimittel-Agentur (European Medicines Agency)
FDA	U.S. Food and Drug Administration
IFN	Interferon
IRES	Interne Ribosomen Eintrittsstelle
HAV	Hepatitis A Virus
HBV	Hepatitis B Virus
HBeAg	HBV e Antigen
HBsAg	HBV Oberflächenprotein
HCV	Hepatitis C Virus
HIV	Humanes Immundefizienzvirus
HLA	Humane Leukozyten-Antigene
ID	Identifikationsnummer
mAB	monoklonale Antikörper
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
MHC	Haupthistokompatibilitätskomplex
ml	Milliliter
MSM	Männer, die Sexualverkehr mit Männern haben
NGS	Next-Generation Sequencing
ORF	Offener Leserahmen
PBMC	Mononukleäre Zellen des peripheren Blutes
PCR	Polymerase-Kettenreaktion
PWID	Menschen mit intravenösem Drogenabusus
RBD	Rezeptor-Bindungs-Domäne
RNA	Ribonukleinsäure
RT	Reverse Transpription
SARS-CoV-2	Severe acute respiratory syndrome coronavirus type 2
UTR	Nichttranslatierte Region
VOC	variants of concern
VOI	variants of interest
WHO	Weltgesundheitsorganisation

3 Eigene Veröffentlichungen

Der Inhalt dieser Habilitationsschrift basiert auf den folgenden Publikationen (siehe Anhang):

[1] Walker, A., et al., *Natural prevalence of resistance-associated variants in hepatitis C virus NS5A in genotype 3a-infected people who inject drugs in Germany*. **J Clin Virol**, 2015. 70: p. 43-45.

[2] Walker, A., et al., *A genotype independent, full-genome reverse-transcription protocol for HCV genotyping and resistance testing*. **J Clin Virol**, 2017. 91: p. 42-48.

[3] Walker, A., et al., *A pan-genotypic Hepatitis C Virus NS5A amplification method for reliable genotyping and resistance testing*. **J Clin Virol**, 2019. 113: p. 8-13.

[4] Walker, A., et al., *Detection of a genetic footprint of the sofosbuvir resistance-associated substitution S282T after HCV treatment failure*. **Virol J**, 2017. 14(1): p. 106.

[5] Walker, A., et al., *Distinct Escape Pathway by Hepatitis C Virus Genotype 1a from a Dominant CD8+ T Cell Response by Selection of Altered Epitope Processing*. **J Virol**, 2016. 90(1): p. 33-42.

[6] Walker, A., et al., *Immune escape pathways from the HBV core(18-27) CD8 T cell response are driven by individual HLA class I alleles*. **Front Immunol**, 2022. 13: p. 1045498.

[7] Schwarz, T., et al., *HBV shows different levels of adaptation to HLA class I-associated selection pressure correlating with markers of replication*. **J Hepatol**, 2025. 82(5): p. 805-815.

[8] Walker, A., et al., *Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020*. **Euro Surveill**, 2020. 25(22).

[9] Walker, A., et al., *Characterization of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection Clusters Based on Integrated Genomic Surveillance, Outbreak Analysis and Contact Tracing in an Urban Setting*. **Clin Infect Dis**, 2022. 74(6): p. 1039-1046.

[10] Houwaart, T., et al., *Integrated genomic surveillance enables tracing of person-to-person SARS-CoV-2 transmission chains during community transmission and reveals extensive onward transmission of travel-imported infections, Germany, June to July 2021*. **Euro Surveill**, 2022. 27(43).

4 Zusammenfassung

Die kontinuierliche Evolution viraler Varianten prägt sämtliche Bereiche der Virologie und stellt fortlaufend neue Herausforderungen dar. Dazu zählen mutierte Primer-Bindungsstellen in der molekularen Diagnostik, die Entwicklung antiviraler Resistenzen sowie Fragestellungen im Zusammenhang mit der Wirksamkeit monoklonaler Antikörper oder T-Zelltherapien. Vor diesem Hintergrund gewinnt die Sequenzierung viraler Genome zunehmend an Bedeutung. Während bis vor wenigen Jahren vorwiegend kurze DNA-Fragmente mittels Sanger-Sequenzierung analysiert wurden, hat sich spätestens seit Beginn der SARS-CoV-2-Pandemie die hochauflösende Vollgenomsequenzierung durch Next-Generation-Sequencing (NGS) als Standard etabliert.

Die rasante Entwicklung direkt antiviral wirkender Substanzen (DAAs) gegen HCV brachte zahlreiche neue diagnostische Fragestellungen mit sich. Im ersten Teil dieser Arbeit konnte gezeigt werden, dass NS5A-RAVs bereits in therapienaiven Patienten vorkommen. Die volllängen cDNA-Synthese sowie die genotypunabhängige Amplifikation von NS5A ermöglichen eine simultane Bestimmung von Genotyp und Resistenzprofil. In einer Fallstudie konnte gezeigt werden, dass selbst bei Wirkstoffen mit hoher Resistenzbarriere resistente Virusvarianten in vivo selektioniert werden können. Daher wäre die Sequenzierung aller HCV-Neudiagnosen im Rahmen einer integrierten genomischen HCV-Surveillance wünschenswert.

Auch für das Verständnis virus-spezifischer Immunantworten ist die Virusevolution von zentraler Bedeutung. Die Erkenntnis, dass $CD8^+$-T-Zellen abhängig von der viralen Sequenz entweder in einen „memory-like“ oder einen „exhausted“ Zustand übergehen, hat weitreichende Implikationen für die Entwicklung neuer Therapien. Immuntherapien, die auf die Erkennung viraler Epitope abzielen, verlieren ihre Wirksamkeit, wenn Immunevasionsmutationen die Antigenpräsentation beeinträchtigen. Ebenso wird die Effektivität therapeutischer Impfstoffe eingeschränkt, wenn Impfstoff und virale Epitope nicht optimal übereinstimmen oder lediglich vorhandene *memory-like* T-Zellantworten reaktiviert werden. Die in dieser Arbeit dargestellten Daten leisten einen wesentlichen Beitrag zum aktuellen Verständnis der Immunevasion bei chronischen Virusinfektionen.

Im Kontext der SARS-CoV-2-Pandemie hat sich die Ganzgenomsequenzierung als zentrale Methode für Surveillance und Kontaktverfolgung etabliert. Angesichts der raschen Entstehung neuer Varianten mit unterschiedlichen Transmissions- und Pathogenitätsprofilen war deren schnelle Identifizierung entscheidend. Die hier präsentierten Analysen zur Übertragung von SARS-CoV-2 in der frühen Pandemiephase zeigen, dass die Lockdown-Maßnahmen im Kreis Heinsberg die weitere Virusverbreitung wirksam eindämmten. Nachfolgende Studien belegten, dass eine integrierte genomische Surveillance (IGS) in einer deutschen Metropole realisierbar ist und die Aufklärung von Infektionsketten sowie Ausbrüchen ermöglicht.

Zusammenfassend verdeutlichen die vorgestellten Ergebnisse, dass Analysen zur Virusevolution unter Selektionsdruck essentiell für das Verständnis von Resistenzentwicklung, Immunevasion und epidemiologischer Überwachung sind. Die interdisziplinäre Verknüpfung virologischer Diagnostik mit Bioinformatik und epidemiologischen Daten bildet die Grundlage einer effektiven Integrierten Genomischen Surveillance, die sowohl die klinische Versorgung als auch das öffentliche Gesundheitsmanagement nachhaltig verbessert.

5 Einleitung

5.1 Evolution von Virusvarianten

Obwohl DNA- und RNA-Viren verschiedene Replikationsstrategien und Anpassungsmechanismen an ihren Wirt haben, führen bei beiden Fehler in der Replikation zur Evolution neuer Virusvarianten. Abhängig von der Genomgröße und Komplexität haben Viren unterschiedliche Mutationsraten. Allgemein formuliert mutieren RNA-Viren schneller als DNA-Viren, einzelsträngige Viren mutieren schneller als doppelsträngige Viren, und die Genomgröße korreliert negativ mit der Mutationsrate [11, 12]. Die Mutationsrate ist auch von viralen und zellulären Faktoren abhängig, wie z.B. von Proof-Reading Proteinen bei Coronaviren oder von Wirts-kodierte Cytosin/Adenosin Deaminasen [13]. Daher handelt es sich bei dem Virusisolat eines Infizierten nicht um ein molekular eindeutig definiertes Virus, sondern um eine Population sehr nahe verwandter Virusvarianten, der sogenannten Quasispezies. Die Quasispezies-Theorie geht auf die grundlegenden Arbeiten von Eigen und Schuster zurück [14, 15] und beschreibt Viren als ein organisiertes Spektrum an Mutanten, welche von einer Master-Sequenz dominiert werden. Diese hat die höchste Fitness oder Replikationskapazität, unterliegt aber der ständigen Selektion [16]. Selektionsfaktoren können zum Beispiel das Immunsystem, antivirale Medikamente oder andere Umweltfaktoren sein.

5.2 Das Hepatitis B Virus (HBV)

Das Hepatitis B Virus (HBV) ist der Erreger der akuten und chronischen Hepatitis (Leberentzündung) B beim Menschen. Nach Schätzungen der Weltgesundheitsorganisation (World Health Organisation, WHO) haben weltweit ungefähr 2 Milliarden Menschen eine HBV-Infektion erlitten, die bei ca. 250 Millionen Menschen zu einem chronischen Verlauf führte [17]. Patienten mit chronischer Infektion weisen ein deutlich erhöhtes Risiko für die Entwicklung einer Leberzirrhose und eines primären Leberkarzinoms auf. Jedes Jahr sterben etwa 1,1 Million Menschen an den Folgen einer chronischen HBV-Infektion, und obwohl eine prophylaktische Impfung vorhanden ist, infizieren sich immer noch etwa 1 Million Menschen jährlich mit HBV [17]. Die Übertragung des HBV erfolgt in den hoch endemischen Gebieten Südostasiens und Afrikas meist perinatal, also von der infizierten Mutter auf das Kind während der Geburt oder im Rahmen des Stillens. In Ländern mit niedriger Prävalenz erfolgt die Infektion im jungen Erwachsenenalter über ungeschützten Geschlechtsverkehr, aber auch unsterile medizinische Behandlungen oder das Teilen von Injektionsnadeln beim intravenösen Drogenkonsum sind relevante Infektionsrisiken. Eine Behandlung mit Nukleos(t)id-Analoga, unterdrückt effizient die virale Replikation, führt jedoch nur selten zu einer vollständigen Heilung und muss daher lebenslang eingenommen werden.

5.2.1 Die Entdeckung von HBV

Hepatitis-ähnliche Krankheitsausbrüche wurden bereits in der Antike beschrieben, aber erst mit der systematischen Erfassung von Infektionen im zweiten Weltkrieg wurden zwei Formen, die *„infektiöse Hepatitis (Typ A)“ und die „Serum Hepatitis (Typ B)“* unterschieden [18]. Für beide konnte zunächst kein Erreger gefunden werden. Bei der Analyse von Antigenen im Blut entdeckten Blumberg und Alter 1963 [19] ein neues *low density* Lipoprotein, welches häufig in Seren von australischen Ureinwohnern, aber nur selten in gesunden Amerikanern detektiert werden konnte. Daraufhin wurde es zunächst als

„Australia Antigen“ bezeichnet. In späteren Studien zeigte sich, dass das Australia Antigen gehäuft bei Transfusionspatienten vorkam und eine geographische und familiäre Häufung in bestimmten Regionen hatte. Im Jahre 1968 wurden sphärische Partikel im Blut von Australia Antigen-positiven Patienten gefunden und ein Jahr später konnte Blumberg eine Assoziation mit einer akuten Hepatitis herstellen [20-23]. Zunächst konnte keine Nukleinsäure in diesen Partikeln detektiert werden. Erst durch bessere Separationsmethoden konnte David Dane 1970 DNA in den Viren nachweisen [24].

5.2.2 Genomorganisation und Replikation

Das Hepatitis B Virus (Orthohepadnavirus hominoidei) gehört zur Gattung der Orthohepadnaviren in der Familie der Hepadnaviridae (International Committee on Taxonomy of Viruses; ICTV). Orthohepadnaviren infizieren Säugtiere und haben ein relativ enges Wirtsspektrum. Seit der Entdeckung von HBV sind viele weitere Spezies-spezifische Orthohepadnaviren identifiziert worden; Sowohl in Orang-Utans (GoHBV) [25], Schimpansen (ChHBV) [26], Wollaffen (WMHBV) [27], in Nagetieren (Waldmurmeltier (WHV)) [28], in Hörnchen (GSHV und ASHV) [29] und zuletzt auch in Fledermäusen [30]. Die Avihepadnaviren kommen in verschiedenen Vögeln vor [31] und mit den Parahepadnaviren wurden auch unbehüllte Hepadnaviren in Fischen entdeckt, die sich evolutionär vor über hunderten Millionen Jahren von den Vorfahren der Tetrapoden getrennt haben [32, 33], was auf eine sehr lange Existenz der Hepadnaviren hindeutet.

HBV besitzt ein 3,2 kb großes, zirkuläres, doppelsträngiges DNA-Genom mit einer äußerst kompakten Organisation [34]. Alle Nukleotide haben kodierende Funktion, wobei über die Hälfte davon in zwei verschiedenen Leserastern kodieren (**Abbildung 1**). Das HBV-Genom besitzt vier offene Leserahmen (Open Reading Frames, kurz ORF) die insgesamt für 7 virale Proteine kodieren: Der längste ORF kodiert für die virale Polymerase (Pol) und nimmt fast das komplette Genom ein. Der preCore/Core ORF kodiert für das Nukleokapsidprotein (Core Protein, HBc) und das Prä-Core Protein [35], welches eine N-Terminal verlängerte, sekretierte Variante des Core-Proteins darstellt. Das Prä-Core Protein wird auch als HBeAg bezeichnet und hat eine immunmodulatorische Funktion [36]. Der ORF für die Hüllproteine liegt vollständig im Bereich des Pol-ORF. Die drei Hüllproteine werden als HBsAg bezeichnet und sind modular aufgebaut. Sie werden als S (*Small*), M (*Middle*, PräS2) und L (*Large*, PräS1) bezeichnet [37], wobei das jeweils größere alle Bestandteile des vorherigen enthält. Der X-ORF kodiert für das regulatorische X-Protein (HBx), welches nur in den Orthohepadnaviren vorhanden ist und für die effiziente Transkription des HBV Genoms benötigt wird [38].

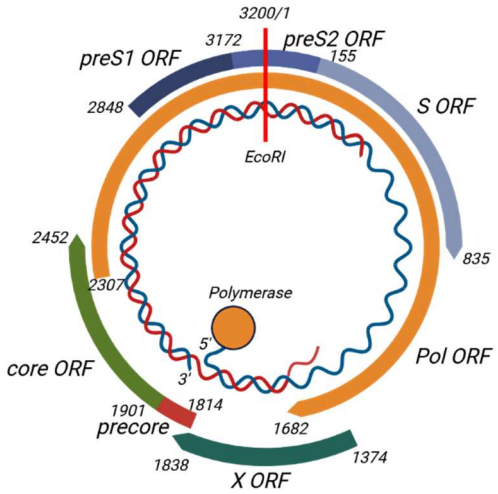


Abbildung 1: Schematische Darstellung des HBV Genoms.
Das HBV Genom in der partiell doppelsträngigen Konformation, kovalent gebunden an die Polymerase, wie es in den Virionen vorliegt. In Rot ist die Prä-Core Region, in Grün der Core ORF, in Orange die Polymerase, in Blau die verschiedenen ORFs der Oberflächenproteine und in dunkelgrün der X- ORF. Der Startpunkt des Genoms bei der EcoRI Schnittstelle ist mit einem roten Strich markiert. Abbildung freundlicher Weise von Dr. Maximilian Damagnez zur Verfügung gestellt.

Das HBV-Virion besteht aus dem icosahedralen Nukleokapsid, in welchem sich die von einer Lipidhülle umschlossene virale DNA befindet. In der Lipidhülle sind die drei Oberflächenproteine im Verhältnis 1:1:4 (*Large, Middle, Small*) eingebettet [39]. Das Genom liegt im Virion als partiell doppelsträngige, relaxierte, zirkuläre DNA (rcDNA) vor, wobei der Minusstrang am 5'-Ende noch ein kurzes Stück RNA besitzt und am 3'-Ende die Polymerase kovalent an die DNA gebunden ist [34].

Bei der Infektion bindet die im L-HBs befindliche PräS1-Domäne an den Gallensäure-Rezeptor NTCP (Natrium-Taurocholate Co-transporting Polypeptide), was zur Internalisierung des Virus führt [40]. Nach Internalisierung und Fusion der Virushülle mit der Endosomenmembran gelangt das Nukleokapsid ins Cytoplasma [41]. Im Cytoplasma schließt die virale Polymerase die Plusstrang-Synthese ab, wodurch das Nukleokapsid zerfällt und das HBV-Genom in den Zellkern transportiert wird. Im Zellkern wird die rcDNA durch zelluläre Enzyme in die kovalent geschlossene zirkuläre cccDNA umgewandelt [42]. Diese cccDNA dient als Matrize für alle viralen mRNAs und die prägenomische RNA (pgRNA), welche als Matrize für die reverse Transkription dient. Die Bindung der viralen Polymerase an die pgRNA führt zur Verpackung durch das Core-Protein. Die reverse Transkription findet dann im Nukleokpasid statt. Als Primer dient dabei die Hydroxyl-Gruppe des Tyrosinrestes 93 in der Polymerase, deshalb ist die Polymerase auch kovalent mit dem Minusstrang verbunden [43]. Nach der Minusstrang-Synthese wird das Nukleokapsid verpackt und über das Endoplasmatische Retikulum und den Golgi-Apparat sezerniert [41, 44]. Die Synthese des Plusstrangs findet erst nach der Verpackung des Nukleokapsids statt. Innerhalb des fertig verpackten Kapsids stehen nur begrenzt Nukleotide zur Verfügung, sodass die Polymerase den viralen Plusstrang nur teilweise vervollständigen kann.

Neben den infektiösen Virionen werden auch subvirale Partikel in etwa 10.000-fachen Überschuss gebildet, die ausschließlich die Oberflächenproteine enthalten [37]. Es wird vermutet, dass diese als immunologische „Attrappen" dienen. Zudem wird in der frühen Phase der Infektion eine sekretierte Version des Core-Proteins produziert. Dieses Prä-Core-Protein ist identisch mit dem Core Protein, besitzt aber am N-Terminus 29 zusätzliche Aminosäuren, die als Signalpeptid für das Endoplasmatische Retikulum dienen und die die Sekretion aus der Zelle veranlassen. Das sekretierte Prä-Core-Protein wird als HBeAg [35] bezeichnet und dient der Unterdrückung einer Immunantwort gegen das Core-Protein [36, 45]. Es wird vermutet, dass das HBeAg eine Rolle bei der Etablierung einer chronischen HBV-Infektion bei perinataler Infektion spielt, da das HBeAg im Gegensatz zu HBV-Virionen plazentagängig ist. Im Verlauf einer chronischen Infektion können sich aber auch Mutationen im HBV-Genom entwickeln, die die Expression des Prä-Core Protein reduzieren oder komplett ausschalten. Diese Serokonversion zu HBeAg-negativ ist oft mit einer signifikanten Reduktion der Viruslast und der Leberentzündung verbunden.

5.2.3 Die HBV-Genotypen

Basierend auf den Unterschieden in der Nukleotidsequenz wird HBV momentan in neun unterschiedliche Genotypen (A bis I; >7,5 % Sequenzunterschied) und über 30 Subtypen (4-7,5 % Sequenzunterschied) eingeteilt [46]. Während die Genotypen A bis H seit langem als eigenständige Genotypen anerkannt sind, wurde Genotyp I erstmals 2008 beschrieben [47]. Er ist eine Rekombination der Genotypen A, C und G. Aufgrund der geringen Nukleotid-Differenz zu Genotyp C wurde er lange nicht als eingeständiger Genotyp akzeptiert [47, 48]. Erst die Isolation weiterer Isolate in Laos, Nordindien und China führte zu Akzeptanz als eigenständiger Genotyp [49, 50]. Ein als Genotyp „J" vorgeschlagenes Isolat stammt von einem Japaner, der auf Borneo lebte [51]. Phylogenetische

Analysen legen nahe, dass es sich dabei um eine Rekombination aus dem Genotyp C und einem Primaten-HBV handelt [52]. Da bisher noch keine weiteren Fälle einer solchen Rekombination bekannt sind, ist er noch nicht als eigner Genotyp anerkannt.

5.2.4 Die HBV-spezifische Immunantwort

Die Rolle des angeborenen Immunsystem bei der Bekämpfung von HBV im Menschen ist noch nicht vollständig aufgeklärt. Es wird davon ausgegangen, dass das HBV eine Art „Stealth Virus“ ist, nicht oder nur sehr schlecht vom angeborenen Immunsystem erkannt wird [53, 54]. Im Vergleich zu anderen Viren interferiert HBV nicht mit der Interferon-vermittelten Signalkaskade, sondern es verhindert die Induktion von Typ I Interferon [53, 55-57]. Trotz der fehlenden Induktion ist HBV aber sehr sensitiv gegenüber Interferon, wie Co-Infektionen mit Hepatitis C (HCV) deutlich zeigen. Obwohl HBV und HCV in der gleichen Hepatozyte replizieren können [58], führt die Co-Infektion mit HCV und die damit verbundene HCV-induzierte Interferon-Antwort [59, 60] zu einer stark reduzierten HBV-Replikation [61]. Infolge dessen, kann es bei HBV/HCV ko-infizierten Personen nach einer erfolgreichen HCV-Therapie zu einer fulminanten Reaktivierung der HBV-Infektion kommen [62].

Die adaptive Immunantwort hingegen ist für die Kontrolle der HBV-Infektion essentiell. Auffällig ist, dass das Einsetzen der adaptiven Immunantwort in der akuten Infektion sehr spät im Vergleich zu anderen Virusinfektionen beginnt [63], was maßgeblich durch die immunsuppressive Umgebung der Leber beeinflusst ist [64]. Antikörper gegen alle Virusbestandteile sind erst nach etwa 8-12 Wochen [37] im Blut messbar, aber nur Antikörper gegen das HBsAg (Anti-HBs) haben neutralisierende Wirkung. Für die Kontrolle der Infektion während der akuten Infektionsphase sind $CD8^+$ T-Zellen von entscheidender Bedeutung [63, 65]. $CD8^+$ T-Zellen erkennen infizierte Hepatozyten über die Interaktion ihres T-Zellrezeptors (TCR) mit Virus-Peptiden (Epitope), die auf der Oberfläche der infizierten Zellen durch sogenannte Humane Leukozyten-Antigene Klasse I (HLA)-Moleküle präsentiert werden. Frühere Studien haben gezeigt, dass Patienten, die eine HBV-Infektion spontan eliminieren, polyklonale und multispezifische $CD8^+$ T-Zell-Immunantworten aufweisen [66-68]. Im Gegensatz dazu sind bei Patienten mit chronischer HBV-Infektion die $CD8^+$ T-Zell-Immunantworten schwach [7, 69] und überwiegend gegen das Core-Protein, jedoch kaum gegen die Polymerase und nur selten gegen das Hüllprotein gerichtet [70, 71]. Dies ist besonders bei HBeAg-positiven Patienten ausgeprägt, weshalb man diese früher als „immuntolerante“ Patienten bezeichnet hatte. In diesen Patienten sind kaum oder nur sehr schwach ausgeprägte $CD8^+$ T-Zellantworten zu finden [6, 72]. Interessanterweise führen Infektionen mit HBeAg-negativen Virusvarianten öfters zu einer fulminanten Hepatitis [73].

In der chronischen Infektion führt die dauerhafte Antigenstimulation zu einer T-Zell-Erschöpfung [74], die mit einer hohen Expression von inhibitorischen Rezeptoren wie Programmed Cell Death Protein 1 (PD-1), Cytotoxic T Lymphocyte Associated Protein 4 (CTLA-4), Lymphocyte Activation Gene 3 Protein (LAG-3), T Cell Immunoglobulin Mucin Domain 3 (TIM-3; HAVCR2), CD244/2B4 sowie CD160 assoziiert ist [75-80]. Zudem weisen diese T-Zellen eine gestörte T-Zell-Homöostase auf, was sich in einer hohen Expression von pro-apoptotischen Proteinen (z.B. BIM) und mitochondrialer Dysfunktion zeigt [81]. Zusammen führen diese Mechanismen zu einer eingeschränkten Funktionsfähigkeit und schließlich zu einem Verlust der Virus-spezifischen $CD8^+$ T-Zellen [74, 82, 83]. Für erfolgreiche Immuntherapien wird daher die Unterdrückung der Virusreplikation in Kombination mit Strategien zur Wiederherstellung der antiviralen $CD8^+$ T-Zellfunktion als unerlässlich angesehen [76, 84, 85]. Zusätzlich werden im Verlauf der chronischen Infektion auch Mutationen in den T-Zell-Epitopen selektiert, die eine

Erkennung des Virus durch die T-Zellen verhindern. Diese Mutationen können sowohl die Bindung der Peptide an das HLA-Klasse I oder die Antigenpräsentation durch den TCR behindern [86].

5.3 Das Hepatitis C Virus

HCV wird über Blut und Blutprodukte übertragen. Obwohl die Zahl der jährlichen Neuinfektionen von 1,5 Million im Jahr 2019 auf 1 Million im Jahr 2024 gesunken ist, stecken sich immer noch viele Menschen hauptsächlich aus Hochrisikogruppen an. Risikogruppen für eine HCV-Infektion sind Menschen, die Drogen injizieren, und Männer, die mit HIV leben und Sex mit Männern haben (MSM). [87]. Eine chronische HCV-250.000 Menschen jedes Jahr an den Folgen einer chronischen HCV-Infektion sterben [17]. Im Rahmen der Agenda 2030 für nachhaltige Entwicklung definierte die WHO im Jahr 2016 HBV, HCV, HIV und anderen sexuell übertragbaren Infektionen (STIs) als Bedrohung für die öffentliche Gesundheit und forderte ihre Eliminierung bis 2030. Das Ziel dabei ist es, 90 % der weltweiten HCV Fälle zu diagnostizieren und 80 % dieser Fälle einer Therapie zuzuführen. Ein positives Beispiel stellt dabei Ägypten dar, das lange Zeit eines der Länder mit der höchsten HCV-Prävalenz weltweit war. Die hohe HCV-Infektionsrate in Ägypten war auf unsterile Injektionspraktiken bei Massenbehandlungen gegen Bilharziose in den 1950er Jahren zurückzuführen und war eines der bedeutendsten Gesundheitsprobleme in Ägypten [88]. Die ägyptische Regierung startete daraufhin 2015 das nationale „*100 Million Healthy Lives*“-Programm, das eine flächendeckende Testung und eine kostenlose Behandlung gegen HCV beinhaltete.4,5 Millionen Ägypter erfolgreich behandelt [89]. Die HCV-Prävalenz in Ägypten sank als Folge der Maßnahmen von über 10 % im Jahre 2016 auf unter 1 % im Jahre 2019 [90]. Dies stellt einen bemerkenswerten Erfolg im globalen Kampf gegen Hepatitis C dar und zeigt, dass bei guter Planung und korrekter Durchführung die WHO Ziele erreichbar sind [91].

5.3.1 Die Entdeckung von HCV

Nach der Identifizierung von HAV und HBV fiel auf, dass zwischen 5-10 % der Posttransfusions-Hepatitiden von keinen der beiden genannten Viren verursacht wurden. Diese als „*NonA-NonB* Hepatitis“-bezeichnete Erkrankung verlief in der Regel mild, persistierte aber in den meisten Patienten. Im Jahr 1978 konnten Harvery Alter und Kollegen zeigen, dass es sich bei der NonA-NonB Hepatitis um eine Infektionskrankheit handelte, die von einem Erreger ausgelöst wurde, der vom Menschen auf den Schimpansen übertragbar war [92]. Im Jahre 1989 gelang es dem Team um Michael Houghton, ein Flavivirus aus dem Blut von infizierten Schimpansen zu identifizieren [93], welches Hepatitis C Virus genannt wurde. 1997 konnte die Arbeitsgruppe von Charly Rice zeigen, dass die Transfektion von Hepatitis C Virus-RNA in die Leber von Schimpansen ausreicht, um eine Hepatitis C-Infektion zu verursachen [94]. Alter, Houghton und Rice wurden dafür 2020 mit dem Nobel Preis in Medizin geehrt. Nach der Identifizierung von HAV und HBV fiel auf, dass zwischen 5-10% der Posttransfusions-Hepatitiden weder von HAV noch HBV verursacht wurden. Diese *NonA-NonB* Hepatitis verlief in der Regel mild, aber persistierte in den meisten Patienten. Im Jahr 1978 konnten Harvery Alter und Kollegen zeigen, dass es sich bei der NonA-NonB Hepatitis um einen infektiösen Erreger handelt, der vom Menschen auf den Schimpansen übertragbar war [92]. Im Jahre 1989 gelang es dem Team um Michael Houghton ein Flavivirus aus dem Blut von infizierten Schimpansen zu identifizieren [93], welches Hepatitis C Virus genannt wurde. 1997 konnte die Arbeitsgruppe von Charly Rice zeigen, dass die Transfektion von Hepatitis C Virus-RNA in die Leber von Schimpansen ausreicht, um eine Hepatitis

C Infektion zu verursachen [94]. Alter, Houghton und Rice wurden dafür 2020 mit dem Nobel Preis in Medizin geehrt.

5.3.2 Die HCV-Genomorganisation und -Replikation

Das Hepatitis C Virus (*Hepacivirus hominis*) gehört zur Familie der Flaviviridae und besitzt ein einzelsträngiges, ca. 9,6 kb großes RNA-Genom mit Plusstrang-Orientierung, welches nur für einen einzigen Open Reading Frame kodiert (**Abbildung 2**). Flankiert wird der ORF an beiden Seiten von nicht-translatierten Regionen (UTR). Die am 5‘-Ende gelegene, hochkonservierte UTR ist ca. 360 Nukleotide lang und enthält die *Internal Ribosom Entry Site* (IRES), die eine Cap-unabhängige Translation des Polyproteins erlaubt [95]. Des Weiteren enthält sie die Bindungsstelle für die leberspezifische microRNA-122, welche für die leberspezifische Translation und Replikation des Virus essentiell ist [96]. Die 3‘-UTR enthält eine ca. 40 Nukleotid lange Poly-Uridin-Wiederholung, sowie einen 98 Nukleotid langen hochkonservierten Bereich (X-tail), der für die Replikation essentiell ist .

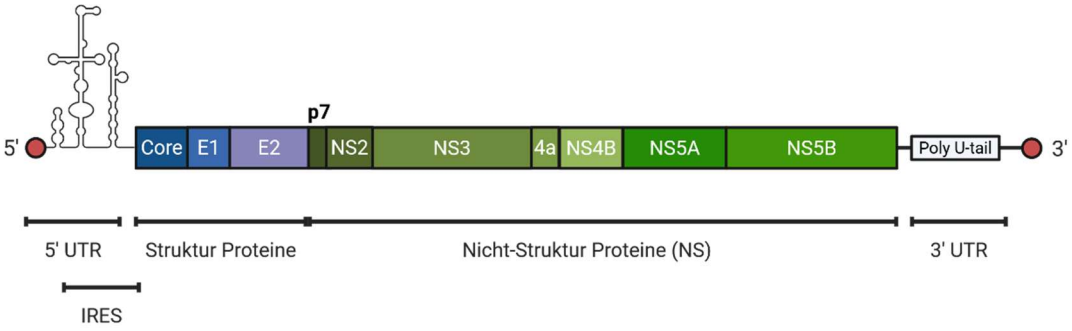


Abbildung 2: Schematische Darstellung des HCV Genoms.
Das HCV Genom wie es in den Virionen vorliegt. Am 5‘Ende befindet sich die Interne Ribosomen Eintrittsstelle (IRES) und am 3‘ Ende die Poly-Uridin Wiederholung. Die Struktur Proteine sind in Blautönen und die Nichtstruktur Proteine in Grüntönen dargestellt. Die Abbildung wurde in BioRender mit Daten aus [95] erstellt.

Während der Translation wird der ORF in ein ca. 3.000 Aminosäuren großes Polyprotein translatiert, welches von zellulären und viralen Proteasen in die zehn viralen Proteine gespalten wird: Core, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B [95]. Die Strukturproteine Core, E1 und E2 bilden das Virion. Die Proteine p7 und das Nicht-Strukturprotein NS2 sind an der Assemblierung beteiligt, werden aber nicht in das Virion eingebaut. Die restlichen Nicht-Strukturproteine („NS-“) bilden die Replikationsmaschinerie und sind alleine in der Lage, die RNA zu replizieren [97]. NS3 ist ein multifunktionelles Protein, das sowohl eine RNA-Helikase, als auch die wichtigste Protease enthält. Zusammen mit dem Cofaktor NS4A spaltet NS3 alle Nicht-Strukturproteine, abgesehen von NS2 aus dem Polyprotein [98]. Des Weiteren ist die NS3/4A-Protease auch für die Spaltung von Immunsensoren wie MAVS beteiligt und unterdrückt so die angeborene Immunantwort [99]. NS4B ist ein sehr hydrophobes Protein, was für die Biogenese der Replikationsorganelle, des sogenannten „Membranous Web“ verantwortlich ist [100]. NS5A ist ein multifunktionelles, multimeres Phosphoprotein, das an der RNA-Replikation, dem Virion-Assembly und der Modulation von zellulären Signalwegen beteiligt ist [95]. NS5B ist die RNA-abhängige RNA-Polymerase und damit für die RNA-Replikation essentiell. Eine Besonderheit bei HCV ist, dass alle Proteine Membranverankert sind und somit der komplette Lebenszyklus des Virus an Membranen stattfindet [95].

Wie alle Positiv-Strang RNA-Viren induziert auch HCV nach einer Infektion eine starke Membranveränderung in den Hepatozyten, welche auch als Replikationsorganellen bezeichnet

werden. Sie bestehen hauptsächlich aus Doppelmembran-Vesikeln, die von der ER-Membran abgeleitet sind, aber einen erhöhten Anteil an Cholesterin und Sphingolipiden aufweisen. Innerhalb der Replikationsorganelle findet die RNA-Replikation geschützt vor Sensoren des angeborenen Immunsystems wie RIG-I statt [100].

Die Formation des Virions ist ein stark regulierter Prozess, an dem verschiedene virale und zelluläre Proteine beteiligt sind. Nach der Translation und Prozessierung aus dem Polyprotein durch NS2 wird das reife Kapsidprotein zu sogenanntem Lipid droplets transportiert. Hier trifft es auf die Oberflächenproteine E1 und E2 wie auch auf die replizierte RNA. Der exakte Ablauf der Assemblierung ist noch nicht abschließend verstanden. Es ist bekannt, dass das Umschalten von RNA-Replikation auf RNA-Verpackung durch den Phosphorylierungsgrad des NS5A-Proteins reguliert wird. Neben NS5A sind auch noch p7, NS2 und NS3 und zelluläre Faktoren an der Assemblierung beteiligt [95]. Nach der Assemblierung liegt das HCV-Virion nicht als freies Partikel vor, sondern als sogenanntes Lipo-Viro-Partikel. Durch die Einlagerung von Cholesterol und anderen Triglyceriden in die Doppelmembran entsteht eine 50-100 nm großer Partikel, der den Very-Low Density Lipoproteine ähnelt. Wie auch bei den *Very-Low Density Lipoproteine* werden in die Virionen verschiedene Apolipoproteine eingebaut [101]. Die Existenz als Lipo-Viro-Partikel vermittelt wahrscheinlich einen Schutz vor der humoralen Immunantwort.

5.3.3 Die HCV-Genotypen

Basierend auf den Unterschieden in der Nukleotidsequenz wird HCV momentan in acht unterschiedliche Genotypen (1-8; > 10 % Sequenzunterschied) und über 60 Subtypen (< 10 % Sequenzunterschied) eingeteilt. Weltweit kommt der Genotyp 1 am häufigsten vor, gefolgt vom Genotyp 3. Während die Genotypen 1-6 seit langem als eigenständige Genotypen anerkannt sind, wurden die Genotypen 7 und 8 bisher erst bei wenigen Patienten gefunden und erst 2015 bzw. 2018 in der Literatur beschrieben [102, 103]. Darüber hinaus treten bei HCV so genannte Rekombinanten auf. Im Gegensatz zu HIV sind Rekombinanten bei HCV seltener, wurden jedoch bereits in der Literatur beschrieben. Ein Beispiel ist die *St. Petersburg*-Variante, eine Rekombinante aus den Genotypen 2k und 1b [104]. Der Rekombinationspunkt befindet sich im NS2 Protein, wobei die Strukturproteine vom Genotyp 2k und die Nicht-Strukturproteine vom Genotyp 1b abstammen. Während bei der St. Petersburg-Varianten der Rekombinationspunkt immer identisch ist, scheint er bei anderen Varianten wie 2b/1a, 2b/1b, oder 2k/1b variabel zu sein, was dafür spricht, das Rekombinationsereignisse häufiger vorkommen können [105]. Die hohen Sequenzunterschiede zwischen den Genotypen haben direkte Auswirkungen auf die Effizienz von antiviralen Medikamenten, insbesondere vor dem Hintergrund, dass die ersten Medikamente für den Genotyp 1 entwickelt wurden.

5.3.4 Die HCV-spezifische Immunantwort

Eine funktionsfähige adaptive Immunantwort ist für die Elimination einer HCV-Infektion essentiell. Die Beobachtung, dass eine HCV-Infektion in Agammaglobulinämie-Patienten ohne Hilfe von B-Zellen ausheilen kann, führte zunächst zu der Annahme, dass B-Zellen keine bedeutende Rolle bei der Elimination der HCV-Infektion inne haben [106]. Diese Vermutung hat sich mittlerweile relativiert. So konnte gezeigt werden, dass die frühe induktion von breitneutralisierende Antikörpern (bnAbs) vor einem chronischen Infektionsverlauf schützen [107]. Auch Antikörper, die durch Impfungen mit löslichen E1-/E2-Proteinen generiert wurden, schützen gegen das homologe Virusisolat [108]. Die

Mehrzahl der neutralisierenden Antikörper ist dabei gegen die hypervariable Region 1 (HVR1) im N-terminalen Bereich des E2-Proteins gerichtet. HVR1-spezifische Antikörper sind meistens Isolat-spezifisch und der Antikörper vermittelte Selektionsdruck ist Grund für die hohe Mutationsrate der HVR1-Region. Durch Immunisierung mit verschiedenen HVR1-Isolaten können in Mäusen bnAbs induziert werden [109] und in verschiedenen Studien schützten bnAbs sowohl vor homologer als auch vor heterologen Infektionen [110-113]. Es bleibt jedoch weiterhin unklar, warum bnAbs nicht in chronisch HCV-infizierten Patienten induziert werden und wie das Virus der Antikörperantwort entkommen kann. Es gibt Vermutungen, dass das Virion durch seine sehr enge Assoziation mit VLDL-Partikel, nicht gut für B-Zellen zugänglich ist oder dass die Unterstützung durch follikuläre $CD4^+$ Helferzellen hierbei fehlt [114].

Für die Viruselimination ist eine funktionelle, multispezifische T-Zellantwort essentiell. Dabei werden sowohl $CD4^+$ als auch $CD8^+$ T-Zellen benötigt. Die Detektion von HCV-spezifischen T-Zellen ist eng mit dem Ausbruch der Lebererkrankung assoziiert [115-117] und die Depletion von $CD4^+$ oder $CD8^+$ T-Zellen in Schimpansen führten zu einer Chronifizierung der HCV-Infektion. Interessanterweise führte eine Depletion von $CD8^+$ Zellen zu einer prolongierten Infektion, die erst mit der Erholung der $CD8^+$ T-Zellzahl ausheilt [118]. Des Weiteren führte die Depletion der $CD4^+$ Zellen zu einer chronischen Infektion, die mit der Entwicklung von viralen Immunevasionsvarianten in den HCV-spezifischen $CD8^+$ T-Zellepitopen einherging [119].

Evidenz dafür, dass HCV durch die Entwicklung von viralen Immunevasionsvarianten in der Lage ist, an den $CD8^+$ T-Zell Immundruck zu adaptieren, stammt aus Infektionsexperimenten in Schimpansen [120, 121] und aus Studien von humanen Infektionsketten, bei denen Donor und Empfänger bekannt waren [122-124]. In beiden Fällen zeigte sich, dass bei der Übertragung in einen neuen Wirt insbesondere in den von $CD8^+$ T-Zellen erkannten Bereichen der Virussequenz Mutationen angereichert werden. Diese Mutationen finden überwiegend während der ersten Monate der Infektion statt [125]. Es ist noch nicht geklärt, warum es bei HCV im Gegensatz zu HIV keine Immunantworten gegen die neuen Epitopvarianten gibt. Eine Erklärung könnte die fehlende CD4-Antwort sein, die im Verlauf der Chronifizierung verloren geht [118, 126].

Funktionell gibt es drei Mechanismen, wie sich HCV an den Wirt anpasst, um der $CD8^+$ T-Zellantwort zu entkommen. Ein Mechanismus ist die Mutation von Ankeraminosäuren in den $CD8^+$ T-Zellepitopen, die eine Bindung des Peptids an das MHC-I bzw. HLA-Klasse I Molekül verhindern [127]. In diesem Falle wird das Epitop nicht mehr auf der Zelloberfläche präsentiert. Eine weitere Möglichkeit stellen Mutationen im Epitop dar, die nicht die Ankerpositionen betreffen, aber die Erkennung durch den T-Zellrezeptor (TCR) verhindern. In diesem Fall wird das Epitop weiter präsentiert, jedoch nicht mehr vom TCR erkannt. T-Zellen mit einem alternativen TCR könnten das Epitop jedoch gegebenenfalls weiterhin erkennen, sodass eine breite oder kreuzreaktive T-Zellantwort das Virus weiterhin effektiv bekämpfen kann [128]. Die dritte Möglichkeit besteht in der Selektion von Mutationen, die die Prozessierung des Epitopes in der Zelle beeinflussen. Die Antigenpräsentation ist ein vielschichtiger Prozess, der den korrekten Abbau viraler Proteine durch das Proteasom, den Transport der Peptide in das ER, die Bindung an das MHC-I Molekül sowie das Trimmen der Peptide auf die korrekte Länge erfordert [129]. Mutationen in der Virussequenz, die einen dieser Schritte inhibieren, können sowohl innerhalb als auch außerhalb des Epitops liegen [124, 130, 131]. Analog zu den Mutationen in der Ankerposition führen sie dazu, dass das Epitop nicht mehr auf der Zelloberfläche präsentiert wird. Welche Immunevasionsmutationen in einem Virus selektiert wird, hängt überwiegend davon ab, wie

stark die $CD8^+$ T-Zellantwort ist und welche Mutationen die geringste Auswirkung auf die virale Fitness hat. Eine erfolgreiche Immunantwort induziert wahrscheinlich breit, kreuzreaktive T-Zellantworten gegen Epitope, deren Mutationen hohe Fitnesskosten haben.

Immunevasionsmutationen beeinflussen den Phänotyp von $CD8^+$ T-Zellen. Bei chronischen Infektionen sind häufig etwa 50 % der potenziellen Epitope der Viren mutiert. Dennoch lassen sich in den Patienten weiterhin HCV-spezifische T-Zellen nachweisen, die diese konservierten Epitope erkennen. Diese Zellen zeigen eine ausgeprägte Hochregulation inhibitorischer Checkpoint-Rezeptoren wie PD-1, TIM-3, CTLA-4, 2B4 und LAG-3 [132-137], was zu einem Verlust ihrer Effektorfunktionen führt. Im Gegensatz dazu besitzen $CD8^+$ T-Zellen, die Epitope mit Immunevasionsmutationen erkennen, ein Profil, das typischerweise nach einer akuten Infektion beobachtet wird [138, 139]. Sie besitzen einen Gedächtnisphänotyp, gekennzeichnet durch eine starke Hochregulation von CD127. Dieser Befund ist kohärent, da das entsprechende Antigen auf Grund der Mutation für die betroffenen T-Zellen nicht länger verfügbar ist, wodurch sie von einem aktivierten zu einem ruhenden, langfristig überlebensfähigen Gedächtniszustand übergehen.

5.3.5 Antivirale Therapie

Mit der Zulassung zahlreicher direkt wirkender antiviraler Substanzen („Directly Acting Antivirals“; DAA) gegen HCV wurde die Behandlung der chronischen Hepatitis C revolutioniert [140, 141]. Die rasante klinische Entwicklung in der Therapie der HCV-Infektion hat 2011 mit der Zulassung der ersten Protease-Inhibitoren Telaprevir und Boceprevir begonnen. Aufgrund der niedrigen Resistenzbarriere mussten diese jedoch noch mit Interferon-α (IFNα) kombiniert werden [142]. Mittlerweile ist das IFNα, das in den ersten 25 Jahren wesentlicher Bestandteil der HCV-Therapie war, bedeutungslos geworden [143]. Die wichtigsten Angriffsziele der antiviralen Therapie der HCV- Infektion sind die NS3/4A-Protease, der virale Replikationsfaktor NS5A und die RNA-abhängige RNA-Polymerase NS5B [144]. Letztere kann durch nukleosidische oder nicht-nukleosidische Inhibitoren gehemmt werden [145]. Bis auf den nukleosidischen NS5B-Inhibitor Sofosbuvir sind mittlerweile alle Wirkstoffe der ersten Generation durch modernere Wirkstoffe mit pan-genotypischer und nebenwirkungsärmerer Wirkung abgelöst worden (Siehe **Tabelle 1** und [140]). Moderne Therapien dauern 8-12 Wochen und bestehen dabei aus Kombinationen von zwei oder drei Wirkstoffklassen mit pan-genotypischen Wirksamkeit und einer einmal täglichen Dosierung. Mit diesen Therapien sind auch unter realen Behandlungs-bedingungen Heilungsraten von über 95 % der Patienten möglich [140].

Tabelle 1: Auflistung der durch die EMA zugelassene HCV HCV-Medikamente (Stand Dezember 2025) [140].

Wirkstoffklasse	Wirkstoff	Genotyp[a]	Hersteller	Handelsname	Zulassung
Protease-Inhibitoren	Telaprevir (TVR)	1	Janssen-Cilag	Incivo®	2011-2016
	Boceprevir (BOC)	1	MSD	Victrelis®	2011-2018
	Simeprevir (SMV)	1 / 4	Janssen-Cilag	Olysio®	2014-2018
	Paritaprevir (PTV)[b]	1 / 4	AbbVie	Viekirax®	2015-2024
	Grazoprevir (GRZ)[h]	1 / 4	MSD	Zepatier®	2016-
	Voxilaprevir (VOX)[g]	1-6	Gilead	Vosevi®	2017-
	Glecaprevir (GLE)[f]	1-6	AbbVie	Maviret ®	2017-
NS5A-Inhibitoren	Daclatasvir (DCV)	1-6	BMS	Daklinza®	2014-2019
	Ledipasvir (LDV)[c]	1, 3, 4, 6	Gilead	Harvoni®	2014-
	Ombitasvir (OMV)[b]	1 / 4	AbbVie	Viekirax®	2015-2024
	Elbasvir(ELB)[h]	1 / 4	MSD	Zepatier®	2016-
	Velpatasvir (VEL)[e]	1-7	Gilead	Epclusa® Vosevi®	2016-
	Pibrentasvir (PIB)[f]	1-6	MSD	Maviret®	2017-
Polymerase-Inhibitoren	*nukleosidische:* Sofosbuvir (SOF)[d]	1-7	Gilead	Sovaldi® Harvoni® Epclusa® Vosevi®	2014-
	nicht-nukleosidische: Dasabuvir (DSV)	1	AbbVie	Exviera®	2015-2024

a) laut Zulassung durch die EMA, b) als fix-dose Kombination von Paritaprevir plus Ritonavir und Ombitasvir , c) als fix-dose Kombination von Ledipasvir und Sofosbuvir (Harvoni®), d) als fix-dose Kombination mit Ledipasvir, Velpatasvir oder allein, e) als fix-dose Kombination von Velpatasvir und Sofosbuvir (Epclusa®), f) als fix-dose Kombination von Glecaprevir und Pibrentasvir (Maviret®/Mavyret®), g) als fix-dose Kombination von Velpatasvir, Voxilaprevir und Sofosbuvir (Vosevi®), h) als fix-dose Kombination von Grazoprevir und Elbasvir (Zepatier®)

5.3.6 Resistenzen gegen HCV-Inhibitoren

Schon während der ersten klinischen Studien mit DAAs in Kombination mit peg-IFNα wurde deutlich, dass bei Versagen der Therapie typischerweise Resistenz-assoziierte Varianten (RAVs) selektiert wurden. Dies lag daran, dass die Resistenzbarriere der DAAs der ersten Generation noch sehr niedrig war und bei schlechtem Ansprechen der IFNα-Therapie die Patienten faktisch mit einer Monotherapie behandelt wurden [146, 147]. Viele der damals selektierten RAVs waren bereits zuvor in subgenomischen HCV-Replikonsystemen *in vitro* als resistent identifiziert worden [148]. Auch gegen die neueren Medikamente wurde die Selektion von RAVs beobachtet. Dies gilt sowohl für Patienten mit einer Durchbruchsinfektion unter Therapie („break through"), als auch für solche mit einem wieder Erscheinen der HCV-RNA nach Abschluss der Therapie („relapse"). In Abhängigkeit von der viralen Fitness persistieren diese RAVs nur wenige Tage (z.B. bei Sofosbuvir) bis hin zu vielen Monaten (z.B. bei den NS5A-Inhibitoren) [4]. Neben der viralen Fitness spielt auch die Anzahl der Mutationen, die für einen Aminosäuren-Austausch notwendig sind, eine entscheidende Rolle hinsichtlich der Resistenzbarriere gegenüber einem antiviralen Medikament. So benötigt der Genotyp 1a nur eine

Substitution, um die Resistenz R155K in der Protease zu realisieren, während beim Genotyp 1b zwei Substitutionen benötigt werden. Dementsprechend sind die Ansprechraten bei DAA-Kombinationen mit Protease-Inhibitoren bei Genotyp 1b höher als bei Genotyp 1a [148]. Die höchste Resistenzbarriere der zugelassenen DAAs zeigt das Nukleosid-Analogon Sofosbuvir. Bisher konnten die RAVs S282T gegen dieses Medikament lediglich für sehr kurze Zeit in wenigen Patienten nach Therapieversagen nachgewiesen werden, da Viren mit der S282T-Mutation eine sehr stark eingeschränkte Replikationsfähigkeit zeigen [149-151].

Im Gegensatz zu HIV wird bei HCV üblicherweise nur dann eine genotypische Resistenzanalyse durchgeführt, wenn ein Therapieversagen vorliegt. Für die Resistenzanalyse werden die therapierelevanten Genregionen amplifiziert, sequenziert und die Resistenzen mittels eines Regel-basierten Algorithmus bestimmt. Regelbasierte Algorithmen wie beispielsweise *geno2pheno HCV* (https://www.https://hcv.geno2pheno.org/) basieren meisten auf *in vitro*- und *in vivo*-Daten aus Zellkulturexperimenten, klinischen Studien, Kasuistiken und Expertenwissen [152].

5.4 Das Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

5.4.1 Die Entdeckung von SARS-CoV-2

Im Dezember 2019 traten in der chinesischen Stadt Wuhan (Provinz Hubei) erstmals Fälle einer bislang unbekannten Atemwegserkrankung auf, die bald mit einem neuartigen Coronavirus in Verbindung gebracht wurde. Das Virus wurde als Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) bezeichnet [153], da es genetisch eng mit dem SARS-CoV verwandt ist, das bereits 2002/2003 in Asien eine regionale Epidemie ausgelöst hatte. Die durch SARS-CoV-2 verursachte Erkrankung erhielt den Namen COVID-19 (Coronavirus Disease 2019). Die WHO stufte SARS-CoV-2 bereits am 30. Januar 2020 als „gesundheitliche Notlage von internationaler Tragweite“ ein und erhöhte diese Einstufung schließlich im März zur weltweiten Pandemie [154]. Die Ansteckung mit SARS-COV-2 erfolgt analog zu anderen respiratorischen Erregern primär durch Tröpfcheninfektion durch bereits infizierte Personen. Infektionen können auch von asymptomatischen Trägern ausgehen, da Infizierte bereits 1-2 Tage vor Symptombeginn infektiös sind [155]. Die durchschnittliche Inkubationszeit dauerte fünf bis sechs Tage bei frühen SARS-CoV-2-Isolaten hat sich aber mit auftreten der Omikron-Varianten auf drei Tage reduziert. Die Infektiosität dauert bei immunkompetenten Personen etwa zehn Tage nach Infektion, kann jedoch bei immunsupprimierten Personen deutlich länger bestehen [156-158]. Zu den häufigsten Symptomen zählen Husten, Fieber, Schnupfen und der Verlust des Geruchs- und Geschmacksinns. In ca. 80 % der Fälle verläuft der Krankheitsverlauf mild bis moderat. In 15-20 % der Fälle können schwere Pneumonien mit Lungenversagen und Tod auftreten [159]. Das Risiko eines schweren Verlaufs ist von verschiedenen Vorerkrankungen wie kardiovaskulären Erkrankung, chronischen Lungen-, Leber- und Nierenerkrankungen, Diabetes mellitus, Adipositas (BMI > 35) und dem Alter abhängig [160].

5.4.2 Genomaufbau und Replikation

SARS-CoV-2 ist ein einzelsträngiges ca. 30 kb großes RNA-Virus mit Plusstrang-Orientierung und gehört zu den Betacoronaviren, zu welchen neben den hochpathogenen SARS-CoV-1 und MERS-CoV (Middle East Respiratory Sysndrom Coronavirus) auch die saisonalen Coronaviren HCoV-HKU1 und HCoV-OC43 gezählt werden. Während die saisonalen Coronaviren schon seit vielen Jahrzehnten weltweit endemisch verbreitet sind und überwiegend milde Erkältungskrankheiten verursachen, sind SARS-CoV, MERS-CoV und SARS-CoV-2 erst vor kürzerer Zeit aus tierischen Reservoirs auf den Menschen übergetreten [161] und können zu schweren Erkrankungen mit tödlichem Verlauf führen.

Phylogenetische Analysen positionieren das Virus in die Gruppe der eng verwandten Fledermaus (Bat)-Coronaviren, insbesondere dem Isolat RaTG13, das aus Rhinolophus-Fledermäusen in Yunnan isoliert wurde und eine 96,2 % Sequenzidentität zu Sars-CoV-2 aufweist [162]. Diese enge Verwandtschaft legt eine zoonotische Herkunft nahe, wobei Zwischenwirte wie das Malaiische Schuppentier (*Manis javanica*) oder andere Farmtiere als mögliche Rekombinationsorte diskutiert werden. Ein Laborunfall als Ursprung für die neue Coronavariante wird immer wieder kontrovers diskutiert. Bisher fehlen zu dieser Hypothese allerdings überzeugende wissenschaftlich nachvollziehbare Daten. Das Spike-Protein des Corona-Virus ist das primäre Oberflächenprotein, das für den Zelleintritt verantwortlich ist. Über die Rezeptor-Bindungs-Domäne (RBD) des Spike-Proteins bindet das Virus an den Angiotensin Converting Enzyme 2 (ACE2)-Rezeptor. Anschließend wird das Spike-Protein durch die Wirts-Proteasen (TMPRSS2 oder Cathepsin L) gespalten, wodurch das Fusionspeptid frei wird und sich in die Membran inserieren kann. Dies führt schließlich zur Fusion von Virus und Plasmamembran [163]. Nach erfolgreichem Eintritt in die Zelle dient das virale Genom als Translationsvorlage.

Das ca. 30 kb große, einzelsträngige RNA-Genom besitzt eine 5'- UTR, gefolgt von einem großen ORF1a/1b (**Abbildung 3**). Dieser wird durch ein gezieltes ribosomales Frameshifting in die Polyproteine pp1a und pp1ab translatiert und kodiert für die Nicht-Struktur-Proteine NSP1-NSP16. Die Polyproteine werden durch NSP3 und NSP5 in die einzelnen NSPs gespalten, welche den viralen Replikations- und Transkriptionskomplex (RTC) bilden. Im RTC finden die RNA-Synthese, das Proofreading (nsp14-Exonuklease-Aktivität) und die Modifikationen der viralen mRNA statt. Dabei synthetisiert der virale RTC zunächst den Negativstrang, welcher dann als Vorlage für das Plusstrang-Genom sowie für subgenomische mRNAs dient. Von den subgenomischen mRNAs werden die Struktur Proteinen Spike (S), Membran (M), Envelope (E) und Nucleocapsid (N) kodiert, denen dann zusätzliche akzessorische Proteine ORF3a, ORF6, ORF7a/b, ORF8, ORF9b, ORF10 folgen, die für die Immunmodulation und die Pathogenese verantwortlich sind [164].

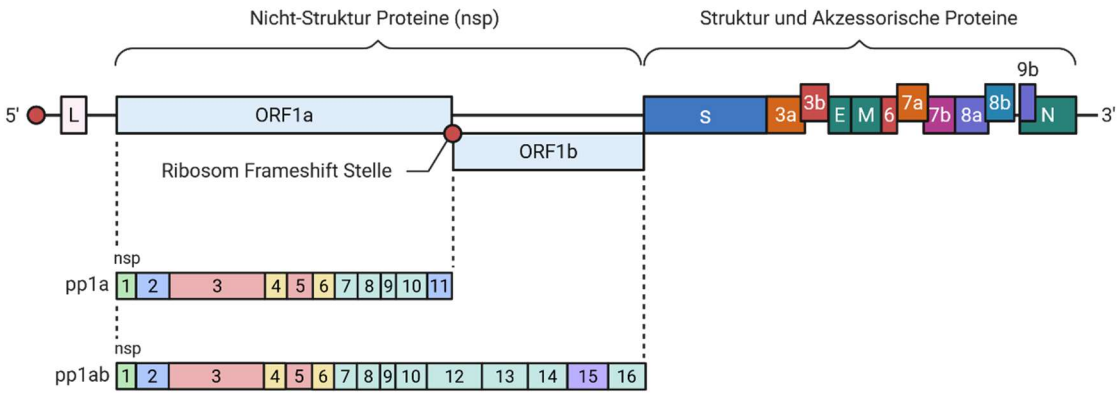


Abbildung 3: Schematische Darstellung des SARS-CoV-2 Genoms.
Das SARS-COV-2 Genom kann in Nicht-Struktur, Struktur und Akezessorische Proteine unterteilt werden. Die nichtstrukturellen Proteine werden in ORF1a und ORF1b kodiert. Die Cap-abhängige Translation beginnt bei ORF1a und produziert pp1a, das nsp1–11 umfasst, oder pp1ab, ein längeres Polypeptid, das nsp12–16 umfasst. Die Produktion beider Polypeptide hängt davon ab, ob das Stopcodon bei ORF1a vom Ribosom erkannt oder durch eine Änderung des Leserahmens durch die Ribosom-Frameshifting-Stelle umgangen wird. Die strukturellen und akzessorischen Proteine werden durch Translation ihrer jeweiligen subgenomischen mRNAs synthetisiert. Die Abbildung basiert auf einer BioRender Template von Jessica M Tucker, Jung-Hee Lee und Britt Glaunsinger basierend auf [163].

5.4.3 SARS-CoV-2-Varianten

Obwohl SARS-CoV-2 das Proofreading Protein nsp14 besitzt, entstehen bei der Replikation Fehler im Genom [165]. Das führte seit Beginn der Pandemie zu einer bemerkenswerten genetischen Diversität, die in einer Vielzahl verschiedener Virus-Isolate resultierte, die je nach potentiellem Risiko als Varianten von Interesse (VOI) bzw. als Varianten von Besorgnis (VOC) bezeichnet wurden. Die VOC B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta) sowie seit Ende 2021 B.1.1.529 (alias BA.1, Omikron) illustrieren dabei unterschiedliche evolutionäre Pfade, bei denen vor allem Mutationen in der RBD des Spike-Proteins selektiert wurden. So zeigte die erstmals in Südafrika detektierte Omikron Variante BA.1 über 30 Mutationen im Spike-Protein. Davon lagen 15 dieser Mutation in der RBD, dem Bereich, gegen die die meisten neutralisierenden Antikörper gerichtet sind. Dies führte schließlich zu einer deutlich reduzierten Empfindlichkeit des Virus gegenüber neutralisierenden Antikörpern [166, 167]. Aufgrund der Immunisierung gegen SARS-CoV-2 und der natürlichen Immunität nach Infektion ist das Spike-Protein einem starken Selektionsdruck ausgesetzt. Aufgrund der hohen strukturellen Plastizität der RBD [168], ist daher auch in Zukunft mit weiteren Immunevasionsvarianten zu rechnen.

5.4.4 Impfstoffentwicklung gegen SARS-CoV-2

Die Impfstoffentwicklung gegen SARS-CoV-2 stellte einen beispiellosen wissenschaftlichen und gesellschaftlichen Kraftakt dar. Bereits wenige Wochen nach der Veröffentlichung der ersten Virussequenzen waren die ersten Impfstoffkandidaten generiert und der prä-klinischen Forschung zugegangen. Besonders durch die Nutzung der neuen mRNA-Technologie konnten Antigene schnell synthetisiert und evaluiert werden. Aber auch klassische Plattformen, wie inaktivierte Viren, Vektor- oder Protein-basierte Impfstoffe wurden von diversen pharmazeutischen Herstellern in sehr kurzer Zeit entwickelt, so dass bereits innerhalb eines Jahres nach Ausbruch der Pandemie die ersten Impfstoffe zur Verfügung standen. Retrospektiv gesehen, schützten fast alle Impfstoffe vor schweren

symptomatischen COVID-19-Verläufen, Hospitalisation und Tod, wobei die Wirksamkeit je nach Impfstofftyp, Virusvariante und Altersgruppe zwischen 70 % und 95 % lag [169-173].

5.4.5 Antivirale Therapie gegen SARS-CoV-2

Neben der Impfung kamen unterschiedliche Virostatika und monoklonale Antikörper (mAbs) gegen SARS-CoV-2 zum Einsatz, deren Effektivität jedoch stark von der jeweils zirkulierenden Variante abhing. Aktuell zugelassene Medikamente sind der Polymerase-Inhibitor Remdesivir (RDV, Veklury®) und der Protease-Inhibitor Nirmatrelvir/Ritonavir (N/r, Paxlovid™). Beide Inhibitoren sind auch gegen die momentan zirkulierende Omikron-Variante wirksam [174, 175]. Der in der Pandemie eingesetzte Polymerase-Inhibitor Molupiravir (MPV, Lagerio®) hat in der EU keine Zulassung erhalten. Die zugelassenen mAbs zeigen seit dem Auftreten der Omikron-Varianten BA.2 keine ausreichende Wirksamkeit mehr und sind seither weitgehend obsolet[176].

5.4.6 Die Kontaktnachverfolgung als Schlüsselinstrument zur Unterbrechung von Transmissionen

Die Kontaktnachverfolgung (Contact Tracing) und Isolierung von Infizierten war gerade zu Beginn der Coronapandemie neben dem Tragen von Masken und der Einhaltung von Abstandsregeln die einzige Möglichkeit, die exponentielle Ausbreitung des Virus zu verhindern [177-179]. Ziel der Maßnahmen war es, potenzielle Infektionsketten frühzeitig zu identifizieren und infizierte bzw. exponierte Personen zu isolieren, um damit die effektive Reproduktionszahl (R_e) des Virus unter 1 zu senken. So konnte der erste Ausbruch von SARS-CoV-2 in Deutschland durch die umfassende Kontaktnachverfolgung und strenge Quarantäne der Infizierten eingedämmt werden [180]. Aufgrund mangelnder Digitalisierung war die Kontaktnachverfolgung zu Beginn der Pandemie allerdings sehr personalaufwändig da z.B. gefaxte Befunde im Gesundheitsamt analog abgearbeitet werden mussten, bevor Kontaktpersonen informiert und unter Quarantäne gestellt werden konnten.

Als Konsequenz wurde digitale Gesundheitsplattformen und Sentinel-Surveillance-Systeme in kürzester Zeit entwickelt, die zuvor aufgrund mangelnder finanzieller wenig priorisiert wurden. Weitere digitale Fortschritte ermöglichten Apps für Smartphones (z.B. Corona-Warn-App), die direkte und indirekte Kontakte über Bluetooth-Daten anonym registrierten; Wurde ein Person später SARS-CoV-2-positiv getestet, konnten alle Kontakte der letzten 48 Stunden darüber anonym informiert werden [181]. Trotz datenschutzrechtlicher Bedenken erwiesen sich diese Technologien als wirksam, insbesondere wenn viele Menschen sie verwendeten und Testergebnisse schnell und digital zur Verfügung standen. Empirische Studien zeigen, dass jede erfolgreich isolierte Kontaktperson die durchschnittliche Anzahl sekundärer Fälle um etwa 0,5 bis 1 Fall reduzierte, was bei hohen Fallzahlen zu einer signifikanten Gesamtreduktion der Transmission führte [182]. Die Effektivität der Kontaktnachverfolgung hing jedoch stark von der Geschwindigkeit der Fallidentifikation, der Testverfügbarkeit und der öffentlichen Compliance ab. Verzögerungen von mehr als 48 Stunden zwischen Symptombeginn und Isolation reduzierten den Nutzen drastisch [182]. Deshalb wurden in vielen Regionen hybride Modelle entwickelt, die manuelle und digitale Ansätze miteinander verknüpften, um sowohl die Reichweite als auch die Genauigkeit der Kontaktnachverfolgung zu maximieren.

5.4.7 Genetische SARS-CoV2-Surveillance

Schon früh in der Pandemie zeigte sind, dass eine Surveillance-Infrastruktur unverzichtbar für das Verständnis der SARS-CoV-2 Dynamik war. So visualisierte Nextstrain (https://nextstrain.org/) alle weltweit verfügbaren Daten in einer interaktiven Weltkarte mit einer phylogenetischen Verteilung der Sequenzen [183]. Diese weltweite Vernetzung von Laboren ermöglichte die Echtzeit-Überwachung von Mutationen, was insbesondere beim Aufkommen der Varianten Alpha, Beta, Gamma, Delta und Omikron von kritischer Bedeutung war. Durch die Integration von Genomdaten in phylogenetische Analysen konnten Transmission-Ketten rekonstruiert, die geografische Verbreitung von Varianten kartiert und die Wirksamkeit bestehender Impfstoffe gegen neue Mutationen bewertet werden [184-186]. Diese Daten bildeten die Grundlage für die Abschätzung von Immunevasionsvarianten, Impfstoffen und Inzidenz-Modellen [187].

6 Ziele der vorgelegten Arbeiten

In den hier zusammengeführten Arbeiten soll gezeigt werden, wie die Evolution von Virusvarianten die Etablierung von chronischen Infektionen oder die Entwicklung von Resistenzen ermöglicht. Neben diesen für den Menschen nachteiligen Entwicklungen, soll aber auch gezeigt werden, wie Virusvarianten gezielt für die genetische Kontaktnachverfolgung verwendet werden können, um damit Infektionen eindämmen zu können.

Eine erfolgreiche HCV-Therapie war, besonders in den Anfängen der DAA-Therapien, entscheidend vom Genotyp, den vorhandenen Substitutionen in den Zielbereichen der Medikamente und den verwendeten Medikamenten abhängig. Die zuverlässige und schnelle Bestimmung von Genotypen und Resistenzen war deshalb ein fester Bestandteil der Therapieentscheidung. Obwohl die Bedeutung der Genotypisierung durch die Zulassung von pangenotypischen DAA-Therapien nachgelassen hat, ist sie weiterhin ein wichtiger Bestandteil der Bewertung von Therapieversagern. Zusätzlich kann in Hochrisikokollektiven eine Reinfektion von einem Relapse nach Therapie oft nur durch Genotypisierung ausgeschlossen werden. Diese Unterscheidung sind für das Behandlungsregime der Betroffenen von therapeutischer Bedeutung. Die Anpassung an einen neuen Wirt ist für chronische Viruserkrankungen ein wichtiger Bestandteil der Pathogenese und führt bei nicht optimalen Therapien zur Etablierung von Resistenzen

Das Verständnis der Chronifizierung bei Hepatitis B und C Viren ist wichtig, um neue Impfstrategien und Immuntherapien zu entwickeln bzw. bestehende Therapieoptionen zu optimieren. Insbesondere das Verständnis von Mutationen, die hoch-konservierte Bereich betreffen, können einen wichtigen Beitrag leisten, um Regionen zu identifizieren, die für eine effiziente Impfung wichtig sind. Obwohl es viele Bestrebungen gibt, eine T-Zell-basierte Impfung bei HBV zu entwickeln, ist immer noch nicht abschließend geklärt, welche Bereiche des Virus für eine T-Zell-Immunisierung am besten geeignet sind. Die hier dargestellten Arbeiten leisten einen substantiellen Beitrag zur Identifizierung solcher geeigneten Bereiche und können damit die Entwicklung zukünftiger Immunisierungsstrategien für zukünftige Immunisierungsstrategien bei chronischen Virusinfektionen prägen.

Darüber hinaus ermöglichen es so genannte „Flaschenhals-Effekte“ bei der Virus-Übertragung, Infektionsketten nachzuvollziehen. Flaschenhals-Effekte treten auf, wenn bei der Übertragung eines Virus von einem Wirt auf einen anderen nur ein kleiner Teil der Viruspopulation den „Engpass“ passiert, wodurch die genetische Vielfalt stark reduziert wird. Durch die genetische Verwandtschaft unterschiedlicher Virusvarianten lassen sich in Verbindung mit der Kontaktnachverfolgung Infektionsketten aufklären bzw. Ausbrüchen eindämmen. Das Potential einer solchen genomischen Infektionskettennachverfolgung zeigte sich insbesondere während der SARS-COV-2 Pandemie, die eine enorme Herausforderung für den öffentlichen Gesundheitsdienst darstellte. In der Hochphase der Pandemie mussten lokale Gesundheitsämter die Kontakte von hunderten Infizierten nachverfolgen, ohne das Wissen darüber, ob das Lagebild aus der Kontaktnachverfolgung dem tatsächlichen Infektionsgeschehen entsprach. Die schnelle Virussequenzierung und Kontaktnachverfolgung anhand von genetischen Varianten stellte dabei ein neues innovatives Tool für die Kontaktnachverfolgung zur Verfügung.

7 Eigene Untersuchungen und Ergebnisse

7.1 Entwicklung von Genotypisierungs- und Resistenztestungs-methoden für das Hepatitis C Virus

Die Therapie der HCV-Infektion ist mittlerweile für alle Patienten empfohlen. Während die heute verwendetet pangenotypischen direkt wirkenden DAAs bei über 90 % der Patienten eine Heilung erzielen, waren die ab 2012 zugelassenen DAA-Präparate oft Genotyp-abhängig und wiesen bei Mutationen in den DAA-Zielbereichen teilweise ein stark reduziertes Ansprechen auf. Aus diesem Grund war die Genotypisierung und anschließende Resistenztestung ein fester Bestandteil der HCV-Diagnostik. Moderne HCV-Therapien zeichnen sich durch eine pangenotypische Wirksamkeit und hohe Resistenzbarrieren aus. Dennoch kann die Genotypisierung zu einer verkürzten und kostenoptimierten Therapie beitragen. Auch das Wissen, ob bereits Resistenzen vor Therapiebeginn sogenannte „Baseline-Resistenzen“ vorliegen kann zu einer optimierten Therapiestrategie verwendet werden.

7.1.1 Analyse der NS5A Baseline-Resistenzen von HCV-Genotyp 3a-Isolaten

Mit der Zulassung des ersten NS5A-Inhibitor Daklinza® (Daclatasvir) im Jahre 2014 war eine völlig neue Substanzklasse zur Therapie der HCV-Infektion verfügbar, wodurch ein zusätzlicher Bedarf an NS5A-Resistenzanalysen entstand. Das Betraf nicht nur Therapieversager, sondern auch therapienaive Patienten, da *in vitro*-Studien einige Resistenzmutationen (Resistance Associated Variant, RAV) bereits in therapienaiven Patientenidentifizierten, die als Baseline-Resistenzen bezeichnet werden [188]. Während Prävalenzdaten für Baseline RAVs in Genotyp 1 schon recht gut beschrieben war, gab es für den Genotyp 3a nur sehr wenige Daten. Da Genotyp 3a infizierte Patienten in Deutschland nahezu 50% der Infizierten ausmachen, wurden zunächst Baseline RAVs im Genotyp 3a- untersucht.

Titel der Original Arbeit:

Natural prevalence of resistance-associated variants in hepatitis C virus NS5A in genotype 3a-infected people who inject drugs in Germany

Andreas Walker, Holger Siemann, Svenja Groten, R. Stefan Ross, Norbert Scherbaum, Jörg Timm

Dazu wurde von 112 Patient, aus einer Kohorte von Genotyp 3a-infizierten Menschen mit intravenösem Drogenabusus (People Who Inject Drugs; PWID), die NS5A-Region mittels selbst entwickelten Primern amplifiziert und sequenziert. Der NS5A-Bereich konnte in 110 von 112 Probanden (98.2 %) erfolgreich amplifiziert werden. Die beiden nicht erfolgreichen amplifizierten Proben hatten eine niedrige Viruslast von 3502 IU/ml bzw. 930 IU/ml. Eine phylogenetische Analyse mit weiteren GT3a-Isolaten aus Europa, zeigte eine gleichmäßige Verteilung unserer Proben im phylogenetischen Baum, was darauf hinweist, dass die Isolate eine repräsentative Stichprobe darstellten (**Abbildung 4**).

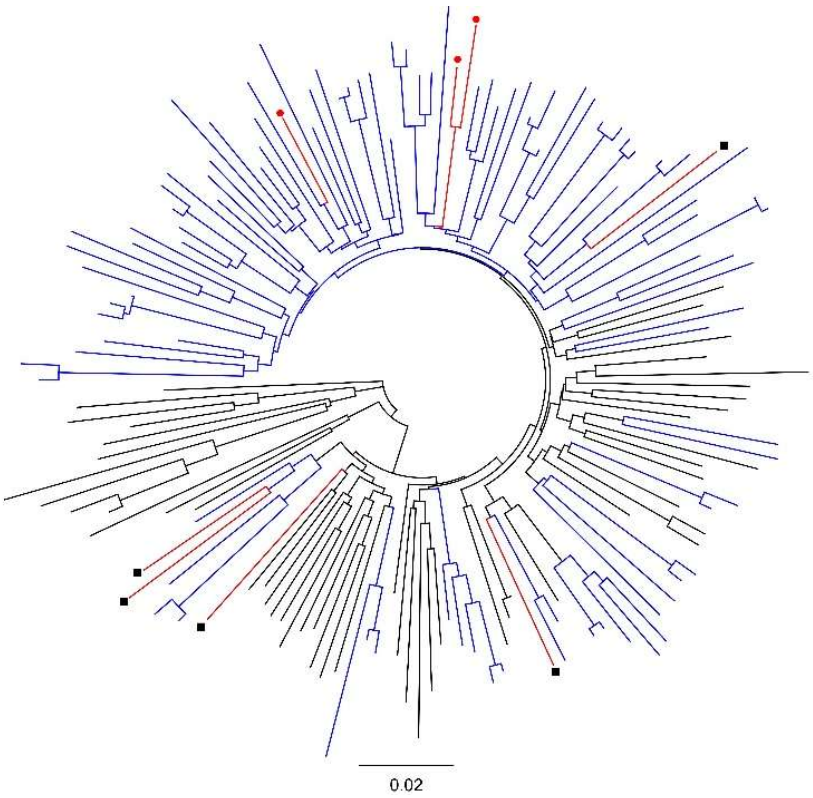


Abbildung 4: Phylogenetischer Analyse der NS5A NS5A-Sequenzen. *Die NS5A NS5A-Sequenzen aus der deutschen PWID PWID-Kohorte (in Bblau und Rrot) und Genotyp 3a 3a-Referenz sSequenzen aus der HCV HCV-Datenbank (in schwarz) wurden aligniert und ein phylogenetischer Baum wurde mit Hilfe der Neighbor-Joining Joining-Methode unter Verwendung des Jukes–Cantor Cantor-Distanzmodels berechnet. Die Äste von Sequenzen mit RAVs sind in rot gefärbt und , Sequenzen mit der Y93H Y93H-Substitution sind durch ein schwarzes Quadrat, Sequenzen mit der A30K Mutation durch einen schwarzen Punkt markiert. Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV.*

In subgenomischen HCV-Replikationsmodellen wurden bereits einige NS5A-RAVs für den Genotyp3a ermittelt. Diese betrafen überwiegen die Substitutionen A30K, L31V/M und Y93H. Weitere mit Resistenz assoziierte Positionen, die im Genotyp 1 gefunden wurden, waren die Positionen 28, 32 und 58. Da unklar war, ob diese Substitutionen auch im Genotyp 3a vorkommen, wurde von allen Sequenzen ein Aminosäuren-Alignment erstellt und anschließend jede bisher beschriebene Positionen auf Substitutionen untersucht (Tabelle 2).

Die Aminosäurerepositionen L31 und P32 waren in unserer Kohorte vollständig konserviert. An der Position M28 gab es bei 2 von 110 Sequenzen Substitutionen zu M28V und M28L. Die Position A30 wies bei 12 von 110 Sequenzen (10,7 %) Substitutionen auf (A30K, A30T, A30V, A30S, A30L). Von diesen Polymorphismen wurde nur die A30K-Substitution mit einer verringerten Empfindlichkeit gegenüber den NS5A-Hemmern Daclatasvir und Ledipasvir in Verbindung gebracht. Diese war bei 5 von 110 (4,5 %) Probanden vorhanden. In NS5A ist die Position Tyrosin 93 am häufigsten mit hochgradigen Resistenzen assoziiert. An dieser Position wiesen 4 von 110 (3,6 %) Probanden eine Substitution auf. Ein Proband hatte die Y93F Mutation und drei weitere Probanden die Substitution Y93H, die bekanntermaßen eine hochgradige Resistenz gegenüber NS5A-Inhibitoren verleiht. Insgesamt konnten bei 18 von 110 (16,4 %) Probanden RAVs nachgewiesen werden. Die hochgradig Resistenz-vermittelte Substitutionen A30K und Y98H konnten bei 5 bzw. 3 PWID detektiert werden.

Tabelle 2: Die Frequenz von RAV in GT3a-infizierten, behandlungsnaiven PWIDs. *Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV.*

Position	28	(n)	30	(n)	31	(n)	32	(n)	93	(n)
Prototype	M	(108)	A	(98)	L	(110)	P	(110)	Y	(106)
Substitutions	V	(1)	**K**	**(5)**[a]					**H**	**(3)**[b]
	L	(1)	T	(3)					F	(1)
			V	(2)						
			L	(1)						
			S	(1)						

[a] 44-fold IC50 change (6).
[b] 2154-fold IC50 change (6).

Obwohl die Zahl von Therapieversagern in klinischen Studien gering ist [189, 190], finden sich unabhängig vom Genotyp in der Mehrzahl dieser Probanden RAVs. In der Zulassungsstudie für die Kombinationstherapie aus Sofosbuvir und Daclatasvir [190] hatten nur 7 von 13 (54 %) Patienten mit der Baseline-RAV Y93H eine erfolgreiche Therapie, wohingegen 124 von 143 (92,5 %) Patienten ohne Baseline-RAV erfolgreich therapiert wurden. Aus diesem Grund kann eine Baseline-Resistenztestung vor Therapie einen wichtigen Beitrag dazu leisten, die Therapiestrategie individuell an den Patienten anzupassen. Die Daten dieser Studie zeigen zudem, dass auch in GT3a-infizierten Probanden in Deutschland die Prävalenz von Baseline-Resistenzen vergleichbar mit anderen Ländern ist und manche Patienten von einer NS5A Resistenztestung vor Therapieeinleitung profitieren könnten.

7.1.2 Etablierung einer voll-längen HCV cDNA Synthese für die Routine-diagnostik

Mit der Einführung der DAA-Therapien stieg der Bedarf an spezifischer Resistenztestung in der Diagnostik stark an. Dabei wurde klassischerweise zuerst der Genotyp mittels Amplifikation der Core-Region am 5'-Ende des Genoms bestimmt, anschließend die Genregionen der Nicht-Strukturproteine NS3, NS5A und NS5B einzeln in cDNA umgeschrieben und schließlich diese Bereiche spezifisch amplifiziert. Eine vollständigen Genotypisierung und Resistenztestung erforderte daher die Synthese und Amplifikation von 4 individuellen cDNAs. Da die cDNA-Synthese den kostenintensivsten Schritt darstellt, wurde untersucht, ob die individuellen cDNA-Synthesen durch eine optimierte Reverse Transkription ersetzt werden könnten.

Titel der Original Arbeit:

A genotype independent, full-genome reverse-transcription protocol for HCV genotyping and resistance testing

Andreas Walker, Matthias Bergmann, Jennifer Camdereli, Rolf Kaiser, Nadine Lübke, Jörg Timm

Historisch bedingt verwenden viele Labore die Core-Region um den Genotyp zu bestimmen(„$core_{GT}$-PCR"). Dieser Bereich ist variabel genug, um Genotypen gut zu unterscheiden, wird jedoch von konservierten Bereichen flankiert, was eine zuverlässige Amplifikation ermöglicht. Der Bereich liegt im 5'-Bereich des HCV-Genoms, während die therapierelevanten Gene im 3'-Bereich des Genoms liegen. In früheren Arbeiten konnte gezeigt werden, dass es unter speziellen Voraussetzungen und Labormethoden möglich ist, dass komplette HCV-Genom in cDNA umzuschreiben [191]. Diese RNA-Extraktionsmethoden waren jedoch nicht routinetauglich, da sie eine manuelle Extraktion mit spezifischen Zwischenschritten erforderte. Aus diesem Grund wurde untersucht, ob eine volllängen cDNA-Synthese speziell für die Routinediagnostik etabliert werden kann. Dazu wurde RNA aus einer Plasmaprobe eines GT1a-infizierten Patienten mittels eines EZ1-Extraktors der Firma Qiagen automatisch extrahiert und mit dem im NS5B-Gen bindenden Primer sv542as reverse transkribiert (**Abbildung 5A**).

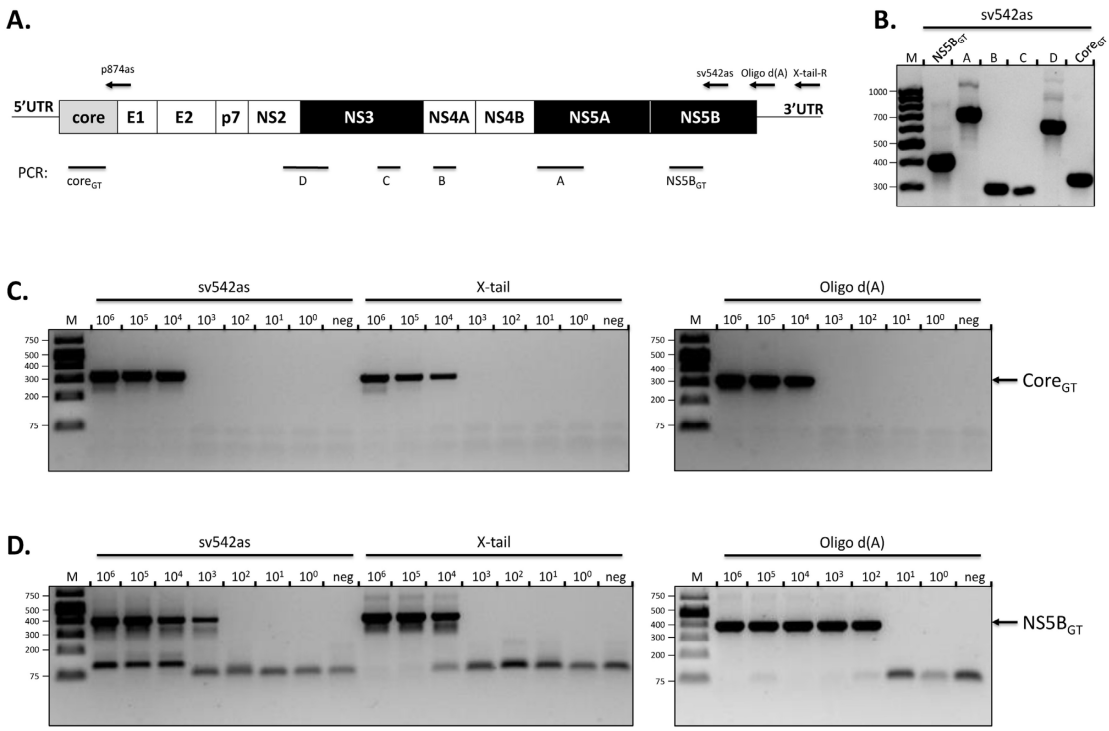


Abbildung 5: Rationale für den Ansatz der reversen Transkription. *(A) Schema des HCV-Genoms basierend auf dem Referenzgenom H77. Die Position der RT-Primers ist oben, die der in der Arbeit verwendeten PCR (NS5BGT, A-D, CoreGT) unterhalb und DAA-Targets sind schwarz dargestellt. (B) Untersuchung der cDNA-Länge mittels Agarose Gelelektrophorese. HCV-RNA wurde mit den Primern sv542as revers transkribiert und die cDNA-Länge mit den verschiedenen PCRs überprüft. (C) Amplifikationseffizienz der Core- und (D) der NS5B-Region mit verschiedenen RT-Primern. Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV.*

Anschließend wurde die cDNA-Länge durch verschiedene, kurze PCRs, die verschiedene Bereiche des Genoms amplifizieren, untersucht (**Abbildung 5B**). Wie in **Abbildung 5B** dargestellt ist, konnten alle PCR-Fragmente erfolgreich amplifiziert werden, inklusive des Core-Fragmentes am äußersten 5'-Ende des Genoms. Nachdem gezeigt werden konnte, dass die volllängen cDNA-Synthese möglich ist, wurde in weiteren Versuchen, Primer-Sequenz für die Reverse Transkription optimiert um eine möglichst lange cDNA inklusive des vollständigen NS5B-Proteins zu erzeugen. Dabei zeigte sich, dass alle verwendeten reverse Primer bis zu einer Viruslast von 10.000 IU/ml eine volllängen cDNA-Synthese ermöglichen (**Abbildung 5C**). Insgesamt erzielte der Oligo-d(A)-Primer im Vergleich die höchste cDNA-Syntheseeffizienz, und ermöglichte eine vollständige Transkription in Proben mit einer Verdünnung von 100 IU/ml (**Abbildung 5D**). Durch die zusätzliche Verwendung des HCV-spezifischen Primers p874as, der in der $Core_{GT}$-PCR als Reverse Primer verwendet wird, konnte auch die Sensitivität der $Core_{GT}$-PCR in Proben mit bis zu 1000 IU/ml verbessert werden. (**Abbildung 6**).

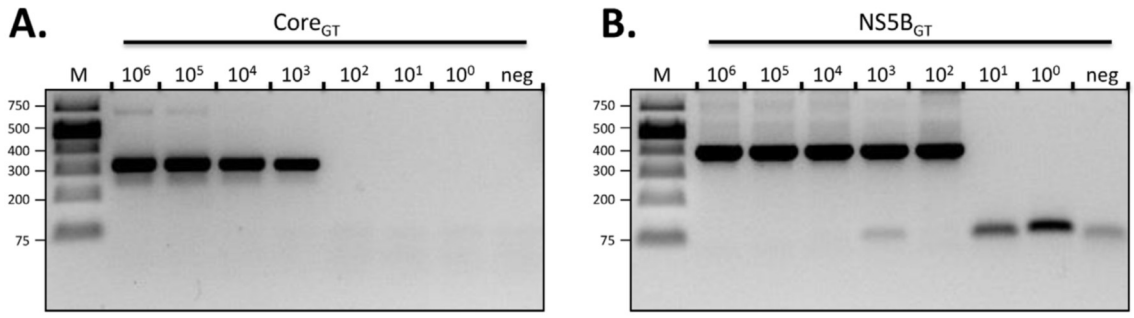


Abbildung 6: Erhöhung der Sensitivität durch den Primers p874as.
RNA-Verdünnungen wurden mit Oligo d(A) und p874as reverse transkribiert. Die Effizienz der cDNA-Synthese wurde mittels der Core$_{GT}$-PCR (A) oder mittels NS5B$_{GT}$-PCR (B) untersucht. Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV.

Nach Optimierung der cDNA-Synthese wurde die Amplifikationseffizienz der DAA- Regionen NS3, NS5A und NS5B aus der erzeugten cDNA untersucht. Dazu wurden Patientenproben mit bekanntem Genotyp und Viruslasten zwischen 8.000 - 12.000.000 IU/ml (log 3,4 - 7,09) aus der Routinediagnostik mit den Primern p874as und Oligo d(A) reverse transkribiert und die DAA-Regionen mit Genotyp-spezifischen PCRs amplifiziert. Dabei zeigte sich, dass bis auf wenige Ausnahmen, alle Genbereiche aus der voll-längen cDNA amplifiziert werden konnte und die Methode auch bei niedrigen Viruslasten mit hoher Zuverlässigkeit funktioniert (**Abbildung 7**).

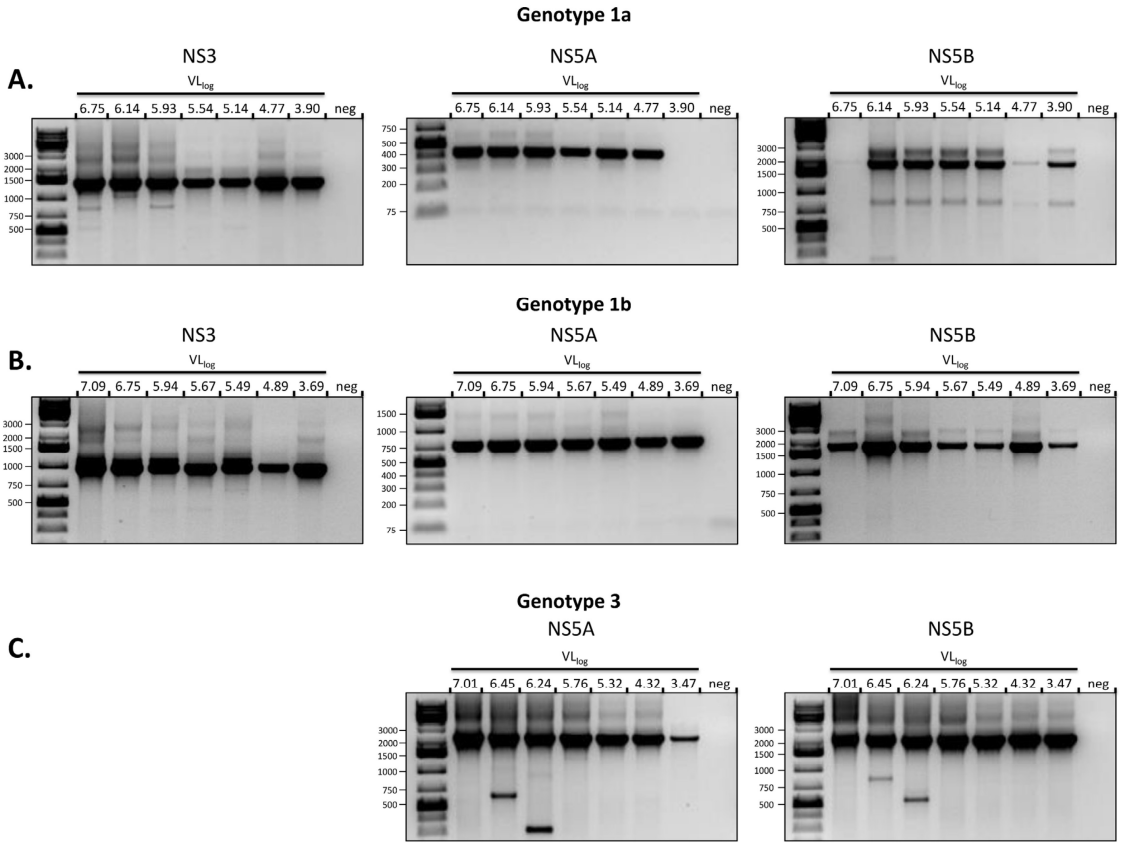


Abbildung 7: Amplifikation von HCV-Resistenzgenen. *RNA von Patientenisolaten wurde automatisiert extrahiert und mittels Primer Oligo d(A) and p874as reverse transkribiert. Nested-PCR für die Amplifikation der relevanten Resistenzgene (A) im Genotyp 1a, (B) im Genotyp 1b, (C) im Genotyp 3a. Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV.*

Diese Ergebnisse zeigten, dass eine voll-längen cDNA-Synthese des HCV-Genoms unter Routinebedingungen möglich ist. Der Vorteil der voll-längen cDNA-Synthese ist, dass ca. 50 % der Kosten gespart und der Arbeitsaufwand und die Analysezeit deutlich reduziert werden können. Zusätzlich erlaubt sie die Amplifikation langer Fragmente, wie man sie für Quasispeziesanalysen benötigt. Die Methode wurde erfolgreich in der Routinediagnostik des Instituts für Virologie an der Universitätsklinik Düsseldorf etabliert.

7.1.3 Etablierung einer simultanen, Genotyp-unabhängigen NS5A-Genotypisierungs- und Resistenztestung

Mit der Zulassung von Harvoni®, Viekirax® und Zepatier® enthielten ab dem Jahr 2015 alle zugelassenen Kombinationstherapien gegen HCV einen NS5A-Inhibitor. Da Baseline-Resistenzen in der NS5A-Region bei allen Wirkstoffkombinationen einen relevanten Einfluss auf das virologische Therapieansprechen zeigten [192, 193], stieg der Bedarf an prätherapeutischen Resistenztestungen deutlich an. Eine kombinierte Resistenzanalyse und Genotypisierung direkt aus der NS5A-Region ist daher von erheblichem diagnostischem Nutzen gewesen. Aufgrund der ausgeprägten genetischen Variabilität dieser Region war eine zuverlässige Amplifikation jedoch bis dahin nicht möglich. Durch Kombination verschiedener degenerierter Primer wurde daher versucht, eine Genotyp-unabhängige Amplifikationsmethode für die NS5A-Region zu entwickelt. Zusätzlich sollte die Zuverlässigkeit der Genotyp-Bestimmung in NS5A mit der etablierten Genotypisierung in Core verglichen werden.

Titel der Original Arbeit:

A pan-genotypic Hepatitis C Virus NS5A amplification method for reliable genotyping and resistance testing

Andreas Walker, Kim Sophie Ennker, Rolf Kaiser, Nadine Lübke, Jörg Timm

Um eine Genotyp-unabhängige Amplifikation zu erreichen, musste zunächst ein konservierter Bereich im Genom identifiziert werden. Mittels Referenzsequenzen aus der Los Alamos HCV-Sequenzdatenbank konnte dabei die Bereiche um die Nukleotide 6069–6146, 6810–6851, 7500-7541 (H77-Nummerierung) identifiziert werden (**Abbildung 8A**). In früheren Arbeiten hatten wir festgestellt, dass degenerierte Primer besser binden, wenn sie länger sind und ein konstantes 5‘ Ende besitzen. Daher wurden die Primer so designed, dass sie am 5´-Ende 18-23 Nukleotide der H77 Referenzsequenz (GT1a) besaßen, gefolgt von einem 23-24 Nukleotiden langen pan-genotypischen Teil, der bis zu 8 Basen mit Ambiguitäten enthielt. Für die Genotypen 2 und 4, die sich in diesen Bereichen deutlich von den anderen Genotypen unterschieden, wurden separate Primer generiert.

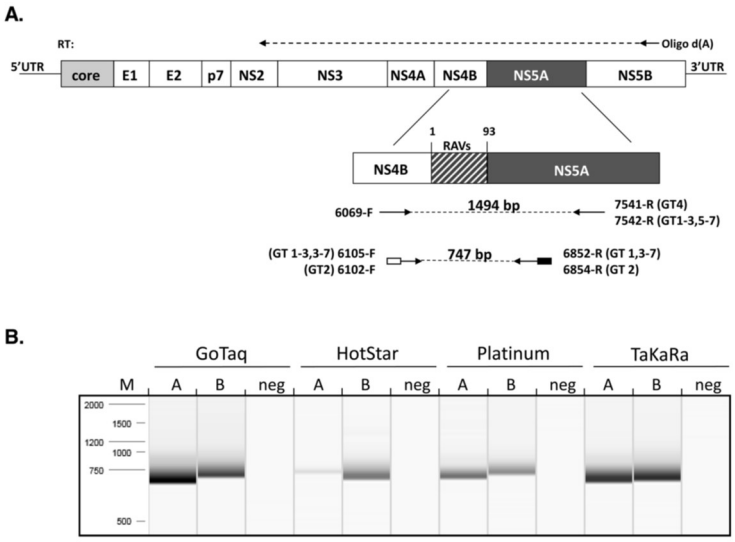


Abbildung 8: Schematische Darstellung des PCR Ansatzes: *(A) Darstellung der Primer-Position und der amplifizierten Region. Die Bindesstelle für die Sequenzierprimer befinden sich am 5'-Ende der PCR-II Primer und sind als weiße bzw. schwarze Box dargestellt. (B) Amplifikationseffizienz kommerziell erhältlicher HotStart-Polymerasen. GoTaq Hot Start (Promega), HotStar Taq (Qiagen), Platinum HotStart (Thermo Scientific) und TaKaRa Ex Taq (TaKaRa). Proben eines (A) GT1a- und eines (B) GT1b-infizierten Patienten wurde amplifiziert und über ein QIAxcel Kapillarelektrophorese-System aufgetrennt. (M) DNA-Leiter 250 bp – 4 kb (Qiagen). Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV.*

Da Primer, die Inosine als degenerierte Base enthalten, nicht von allen Polymerasen akzeptiert werden, musste zunächst überprüft werden, welche der kommerziell erhältlichen Polymerasen in der Lage sind, die generierten Primer zu verwenden. Dazu wurden drei Standard-Polymerasen (GoTaq, HotStar und PlatiniumTaq) und eine Proofreading Polymerase (Ex Taq) untersucht.Dabei erwies sich die Proofreading Polymerase der Firma TaKaRa als am besten geeignet (**Abbildung 8B**).

Im Anschluss an die Validierung der Primer und der Polymerase wurden 262 HCV-Patientenisolate aus Düsseldorf mit dem optimierten Protokoll amplifiziert und der NS5A-Bereich sequenziert. Der Genotyp war durch die Sequenzierung der Core-Region bereits bekannt. Die mediane Viruslast betrug in den untersuchten Proben $1{,}07\times10^6$ IU/ml (zwischen 248 – 21×10^6 IU/ml).

Insgesamt konnte bei 257 von 262 (98,1 %) der Proben das erwartete 747 bp-Fragment amplifiziert und sequenziert werden. Die fünf negativen Proben wiesen alle eine sehr niedrige Viruslast von unter 3.000 IU/ml auf.

Eine Sequenzierung von NS5A erlaubt theoretisch eine simultane Resistenztestung und Genotypisierung. Aus diesem Grund wurde untersucht, wie zuverlässig die Genotypisierung in NS5A im Vergleich zum Goldstandard der Core-Genotypisierung war. Zu diesem Zweck wurde für alle Sequenzen der 257 Isolate eine phylogenetische Analyse der NS5A- und der Core-Sequenzen durchgeführt (**Abbildung 9**). In 250 von 257 (97,3 %) wurde der selbe Genotyp in NS5A wie in Core detektiert. Unter den 7 diskrepanten Isolaten befanden sich drei korrekt bestimmte, 2k/1b-Rekombinanten (core/NS5A 2k/1b) und 4 sehr seltene Subtypen, die in den phylogenetischen Bäumen nicht hinreichend aufgelöst werden konnten (core/NS5A: 1i/1d, 2b/2k, 2q/2f und 2k/2c). Dies zeigte, dass die Zuverlässigkeit der Genotypisierung aus dem NS5A-Bereich vergleichbar ist, wie die Genotypisierung aus dem Core-Bereich.

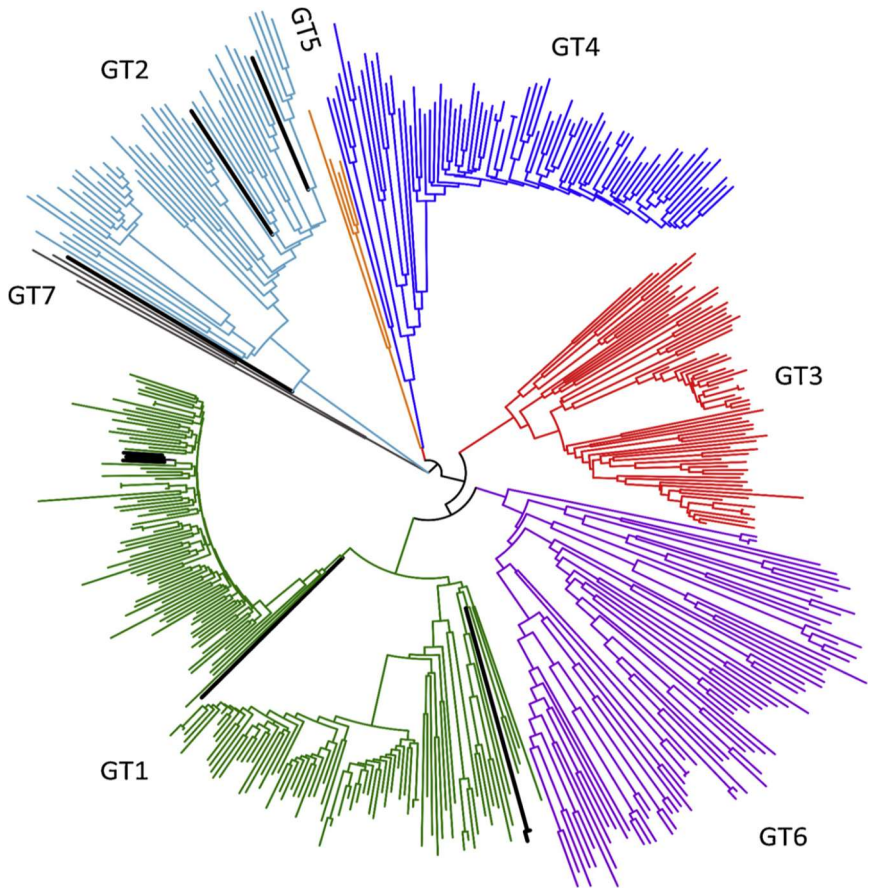


Abbildung 9: Phylogenetische Analyse der NS5A-Sequenzen. *Die NS5A-Sequenzen wurden gegen Referenzsequenzen mit bekanntem Genotyp aligniert und ein phylogenetischer Baum wurde mit dem Tamura-Nei-Distanzmodell und der Neighbor-Joining-Methode berechnet. Proben, bei denen Core und NS5A unterschiedliche Genotypen zeigten, sind in schwarz markiert. Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV.*

Resistenzvorhersage-Algorithmen wie Geno2Pheno [HCV] [152] bestimmen neben den Resistenzen auch immer den zugrundeliegenden Genotyp. Daher wurde die Übereinstimmung zwischen den phylogenetischen Analysen und der Geno2Pheno [HCV] Vorhersage abgeglichen. Im Vergleich zur phylogenetischen Analyse des Core-Fragments sagte Geno2Pheno [HCV] für das NS5A-Fragment in 248 von 257 Fällen (96,5 %) den korrekten Genotyp voraus. Die Abweichungen betrafen erneut die drei rekombinanten 2k/1b-Varianten sowie vier Proben, die bereits in der phylogenetischen NS5A-Analyse diskrepant waren. Zwei Genotyp 1h-Proben wurden von Geno2Pheno [HCV] fälschlicherweise als Genotyp 1a vorhergesagt (Sequenzhomologie zum Referenzstamm (79 % bzw. 80 %). Der Grund für die fehlerhafte Zuordnung war, dass zu diesem Zeitpunkt noch keine GT 1h-Referenzsequenz in der Geno2Pheno [HCV] hinterlegt war.

Zusammenfassend stellt das von uns entwickelte Protokoll eine deutliche Verbesserung zur bisherigen Diagnostik dar. Es ermöglichte erstmals die Genotypisierung und die Vorhersage von Resistenzen aus dem wichtigen NS5A-Bereich. Der pan-genotypische Ansatz reduzierte die praktische Arbeitszeit und die Kosten im Labor um etwa 50 %. Die Genauigkeit der Genotypisierung über NS5A unter Verwendung von Geno2Pheno [HCV] ist vergleichbar mit der phylogenetischen Analyse, bietet aber den Vorteil der simultanen Resistenzbestimmung. Damit erhält man zusätzlich zum Genotyp auch Informationen zu möglichen Baseline-RAV, selbst bei Patienten, die normalerweise keine Resistenztestung bekommen hätten.

7.1.4 Selektion der Sofosbuvir-Resistenzmutation S282T in einer GT3a-infizierten Patientin

Der NS5B-Inhibitor Sofosbuvir (SOF) ist ein Uracil-Analog, das durch eine ProTide-Modifikation pharmakologisch aktiv und oral bioverfügbar gemacht wurde. Sofosbuvir wurde 2007 von Michael J. Sofia bei Pharmasset entwickelt und später von Gilead Sciences übernommen. Im Jahr 2014 erfolgte die Zulassung durch die Europäische Arzneimittel-Agentur (EMA) zur Behandlung chronischer HCV-Infektionen [194]. Sofosbuvir ist wirksam gegen alle bekannten HCV-Genotypen, und selbst in klinischen Monotherapiestudien wurden nur vereinzelt Resistenzmutationen beobachtet [151]. Aus diesen Gründen ist Sofosbuvir Bestandteil sämtlicher von Gilead entwickelter Kombinationspräparate. Die in Zellkultur beschriebene Resistenzmutation S282T vermittelt je nach Genotyp eine 2-18-fach reduzierte Empfindlichkeit gegenüber Sofosbuvir, reduziert jedoch gleichzeitig die Replikationsfähigkeit je nach Genotyp um 89 % – 99 %. Daher konnte sie in den klinischen Zulassungsstudien lediglich bei einem einzigen mit Sofosbuvir monotherapierten Patienten der FISSION-Studie nachgewiesen werden [195]. Angesichts der geringen Zahl dokumentierter Resistenzfälle kommt der Untersuchung individueller klinischer Verläufe besondere Bedeutung zu, um die virale Evolution unter Sofosbuvir-Therapie besser zu verstehen.

Titel der Original Arbeit:

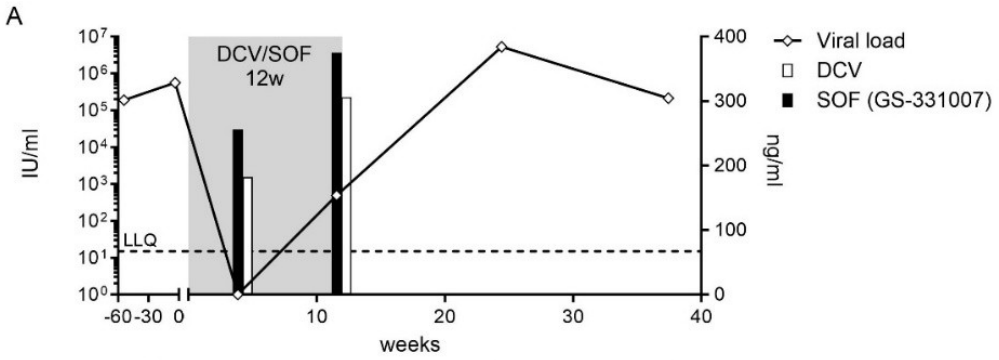
Detection of a genetic footprint of the sofosbuvir resistance-associated substitution S282T after HCV treatment failure

Andreas Walker, Sandra Filke, Nadine Lübke, Martin Obermeier, Rolf Kaiser, Dieter Häussinger, Jörg Timm and Hans H. Bock

Im Folgenden wird ein Fall vorgestellt, bei denen die Resistenzen Y93H in NS5A und S282T in NS5B in einer Patientin unter einer Sofosbuvir (SOF)-/ Daclatasvir (DCV)-Kombinationstherapie selektiert wurde. Bei der Patientin handelte es sich um eine 59-jährige Frau, die mit dem HCV-Genotyp 3a infiziert war und noch keine Leberzirrhose ausgebildet hatte. Die HCV-Viruslast vor Therapie betrug 562.530 IU/ml, andere Virusinfektionen wie HIV und HBV wurden nicht nachgewiesen. Die Patientin wurde Leitliniengerecht über 12 Wochen mit einer täglichen Dosis von 400 mg SOF und 60 mg DCV behandelt. Aufgrund von Schluckbeschwerden verwendete die Patientin den Tablettenüberzug MEDCOAT®. Zunächst konnte ein gutes Therapieansprechen erreicht werden und die Patientin hatte kein detektierbares Virus bei der ersten Kontrolle in Woche 4 (**Abbildung 10**).

Zwei Tage vor Ende der Therapie in Woche 12 war HCV mit einer sehr niedrige Viruslast von 493 IU/ml nachweisbar und stieg in den nächsten 12 Wochen nach Ende der Behandlung auf 5.290 000 IU/ml an (**Abbildung 10A**). Die retrospektive Analyse der Plasmaspiegel deuteten darauf hin, dass die Patientin ihre Medikamente nicht regelmäßig eingenommen hat. Eine Phagophobie konnte von Seiten der Klinik nicht ausgeschlossen werden. Die retrospektive Resistenzanalyse mittels ultratiefer Sequenzierung (Next Generation Sequencing, kurz NGS) zeigte, dass keine RAVs vor Therapiestart in NS5A und NS5B vorhanden waren (Häufigkeit < 0,2 %). Mittels Patienten-spezifischen Primern konnte das Virus auch

aus der Probe mit sehr niedriger Viruslast (Woche 12 nach Therapiestart), also noch unter Therapie, amplifiziert werden. Dabei konnten in der kompletten Quasispezies die Resistenzmutationen Y93H in NS5A und S282T in NS5B nachgewiesen werden (**Abbildung 10B**). Zwölf Wochen nach Ende der Therapie (FU12) war die RAV S282T wieder vollständig verschwunden, während die RAV Y93H in NS5A über einen längeren Zeitraum in der Quasispezies nachweisbar war. Die sehr niedrige Viruslast unter Therapie und das sofortige verschwinden der S282T-Mutation in NS5B weisen auf eine sehr geringe Replikationsfähigkeit der entsprechenden Mutante hin.



B

RAS	BL	week 12	FU12	FU24
NS5A-Y93H	0.17%	99.6%	68.5%	0.20%
NS5B-S282T	0%	99.8%	0.07%	0.02%

C

codon usage pos. 282				
S282 (AGT)	98.2%	0%	31.2%	75.0%
S282 (AGC)	1.1%	0%	0.07%	0.25%
S282 (TCT)	0%	0%	68.0%	24.2%
T282 (ACT)	0%	99.8%	0.07%	0.02%

Abbildung 10: Infektionsverlauf bei einer Genotyp 3a infizierten Patienten unter einer Sofosbuvir/Daclatasvir Therapie. *(A) Gezeigt sind die HCV Viruslast (schwarze Linie), die Dauer der Therapie (graue Schattierung) und die Plasmalevel von Daclatasvir (weiße Balken) und dem Sofosbuvir Metabolit GS-331007 (schwarzer Balken). (B) Häufigkeit der RAV NS5A Y93H und NS5B-S282T zum Zeitpunkt vor Therapie (Woche 0), am Ende der Behandlung (Woche 12) and 24 Wochen nach dem Ende der Therapie. (C) Codon Nutzung in der NS5B Position 282. Die Veröffentlichung erfolgt unter CC BY 4.0.*

Interessanterweise fiel bei der NGS-Analyse auf, dass zwei unterschiedliche evolutionäre Wege für die Rückbildung der S282T-Mutation gewählt wurden. In der FU12-Probe konnte bei 31 % der Quasispezies das ursprüngliche Codon für Serin (AGT) nachgewiesen werden, während 68 % der Quasispezies das alternative TCT Codon für Serin (TCT) verwendeten (**Abbildung 10C**). Diese alternative Codon-Verwendung für Serin an Position 282 war einzigartig und kam in keinem der 593 zuvor analysierten HCV-Genome aus der Los Alamos HCV Sequenzdatenbank vor. Während der Nachbeobachtung in Woche 24 sank die Häufigkeit der TCT-Variante in der Quasispezies auf 24,2 %, was zeigte, dass diese Mutation auch auf HCV-RNA-Ebene einer negativen Selektion unterlag.

Der Case Report zeigt, dass bei Patienten, die ihre Medikamente nicht regelmäßig nehmen, RAVs in sehr kurzer Zeit selektioniert werden können. Während Resistenzen in NS5A nur einen geringen Einfluss auf die Replikationsfähigkeit zu haben scheinen, führen Resistenzen an Position 282 in NS5B nachweislich zu einer deutlichen Beeinträchtigung der Replikation. Aufgrund dieser stark reduzierten Replikationsfähigkeit tritt die S282T-Variante in klinischen Proben nur sehr selten auf. Interessanterweise veröffentlichte Gane et al. zeitgleich zu unserer Studie in einer Metaanalyse die

gesammelten Resistenzdaten zu S282T aus sämtlichen Zulassungsstudien von Gilead Medikamenten in denen Sofosbuvir enthalten war. Dabei zeigte sich, dass 1.025 von 8.598 behandelten Patienten ein Therapieversagen hatten. Von diesen war bei 10 (1 %) Patienten die S282T-Mutation nachweißbar, wovon bei 5 Patienten die alternative Codon-Benutzung nachgewiesen werden konnte. Bei einem weiteren Patienten konnte die ursprüngliche S282T-Mutation nicht mehr nachgewiesen werden, jedoch das alternative TCT-Codon [195]. Das alternative Serin-Codon an Position 282 ist damit eine Art molekulare Signatur für eine vorherige Sofosbuvir-induzierte RAV. Aus diesem Grund ist es entscheidend, nicht nur einzelne Aminosäuresubstitutionen zu betrachten, sondern gegebenenfalls auch charakteristische Resistenzsignaturen zu berücksichtigen, die auf eine spezifische Selektionsdynamik hinweisen.

7.2 Evolution von Virusvarianten unter dem Selektionsdruck durch das adaptive Immunsystem

Neben der Anpassung an Medikamente spielt die natürliche Virusevolution auch eine wichtige Rolle bei der Anpassung an einen neuen Wirt bzw. an das Immunsystem des neuen Wirtes. Nach Schätzungen der WHO sind weltweit rund 250 Millionen Menschen chronisch mit HBV und rund 50 Millionen Menschen chronisch mit HCV infiziert. In beiden Virusinfektionen spielen $CD8^+$ T-Zellen eine wichtige Rolle bei der Kontrolle der Infektion. Im folgenden Abschnitt wurde untersucht, wie Viren sich über verschiedene Wege an diesen $CD8^+$ T-Zell-Selektionsdruck anpassen, um eine chronische Infektion zu etablieren.

7.2.1 Strategien zur Umgehung der $CD8^+$ T-Zell-vermittelten Immunantwort am Beispiel des Hepatitis C-Virus

Antivirale $CD8^+$ T-Zellen sind die zentrale Komponente des adaptiven Immunsystems gegen Virusinfektionen. Sie erkennen infizierte Zellen über kurze 8-10 Aminosäure lange Peptide, sogenannte Epitope, die auf HLA-Molekülen der Klasse I auf der Oberfläche von infizierten Zellen präsentiert werden. Mit über 100 verschiedenen Allelen gehören die HLA-Moleküle der Klasse I zu den variabelsten Genen des Menschen. Je nach HLA-Allel werden unterschiedliche Epitope präsentiert, womit Viren beim Wirtswechsel ihr Genom an die Bereiche anpassen müssen, die vom neuen Wirt präsentiert und damit von $CD8^+$ T-Zellen erkannt werden.

Titel der Original Arbeit:

Distinct Escape Pathway by Hepatitis C Virus Genotype 1a from a Dominant CD8 T Cell Response by Selection of Altered Epitope Processing

Andreas Walker, Kathrin Skibbe, Eike Steinmann, Stephanie Pfaender, Thomas Kuntzen, Dominik A. Megger, Svenja Groten, Barbara Sitek, Georg M. Lauer, Arthur Y. Kim, Thomas Pietschmann, Todd M. Allen, Joerg Timm

Für die Entwicklung von $CD8^+$ T-Zell-basierten prophylaktischen Impfstoffen gelten immundominante Epitope, also solche, die bei nahezu allen Patienten mit dem entsprechenden HLA-Typ eine Immunantwort hervorrufen, als besonders geeignet. Idealerweise sollten diese zwischen den verschiedenen HCV-Genotypen konserviert sein. In einer multizentrischen Studie konnten wir die $CD8^+$ T-Zellantwort gegen das hochkonservierte HLA-B*51-restringierte Epitop $NS3_{1373\text{-}1380}$ (IPFYGKAI) charakterisieren und zeigen, wie HCV der $CD8^+$ T-Zellantwort entkommt.

Um die $CD8^+$ T-Zellantwort gegen dieses Epitop zu charakterisieren, wurden zunächst 43 HLA-B*51-positive PWIDs untersucht, inklusive 15 PWIDs, die eine ausgeheilte HCV-Infektion hatten (sogenannte „Resolver“). Dazu wurden mononukleäre Zellen des peripheren Blutes (PBMCs) aus Blutproben der Probanden isoliert, für 10 Tage mit dem HLA-B*51-restringierten Peptid IPFYGKAI expandiert und die

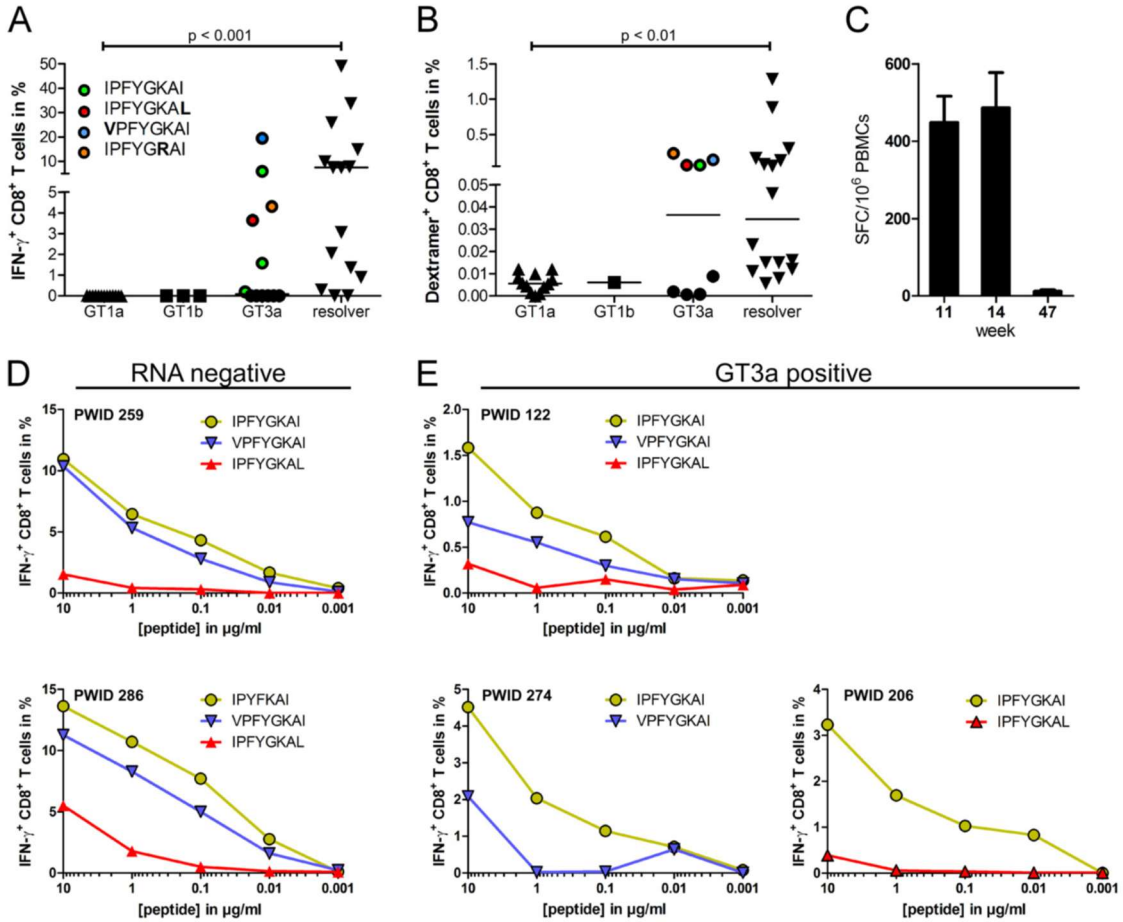


Abbildung 11: Die CD8+ T-Zellantwort gegen das Epitope IPFYGKAI1373-1380. *(A) CD8+ T-Zellen wurden für 10 Tage in vitro expandiert und die IFNγ-Antwort mittels Durchflusszytometrie gemessen. Die autologe Epitopsequenz der Patientenisolate ist angegeben. (B) Die HLA-B*51 1373-1380-spezifischen CD8⁺ T-Zellen wurden direkt ex vivo mit einem HLA-B*51:01 Dextramer gefärbt. (C) Die HLA-B*51 1373-1380-spezifische CD8⁺ T-Zellantwort in einem Patienten mit akuter Hepatitis C- Infektion (D and E) IFNγ-Antwort nach Restimulation mit Peptidverdünnungen der Prototype (grün), I1373V (blau) und I1380L (rot)-Sequenz in Resolvern und GT3a-infizierten Patienten. Die Veröffentlichung erfolgt mit Genehmigung der American Society for Microbiology.*

Frequenz der IFNγ-produzierenden T-Zellen nach Restimulation mit dem Peptid mittels Durchflusszytometrie analysiert. Nach Expansion der IPFYGKAI-spezifischen CD8⁺ T-Zellen konnte bei 12 von 15 (80 %) Resolvern und bei 6 von 12 (50 %) chronisch infizierten Patienten IFNγ-produzierende CD8⁺ T-Zellen detektiert werden. Auffällig war dabei, dass nur Genotyp 3a- aber nicht Genotyp 1-infizierte Patienten eine Antwort zeigten (**Abbildung 11A&B**).

Sequenzanalysen des Epitops zeigten, dass bis auf 5 PWIDs alle Probanden die Prototyp-Sequenz besaßen. Die gefundenen Substitutionen im Epitop waren I1373V, K1377R und I1380L, während ein Patient zusätzlich die Substitution S1368P außerhalb des Epitops aufwies (**Abbildung 11A&B**). Bemerkenswerterweise wurde diese Substitution auch bei dem Patienten mit akuter HCV-Infektion des Genotyps 1a beobachtet, bei dem ein rascher Rückgang der CD8⁺ T-Zell-Reaktion eintrat (**Abbildung 11C**). Um den Einfluss dieser Mutationen im Epitop auf die T-Zellantwort zu untersuchen,

wurden erneut PBMC für 10 Tage mit dem Prototyp-Peptid expandiert und dann mit verschiedenen Varianten in verschiedenen Konzentrationen restimuliert. Dabei zeigte sich, dass die I1373V-Substitution die $CD8^+$ T-Zellantwort kaum beeinflusst, während die I1380L-Substitution die $CD8^+$ T-Zellantwort deutlich reduziert und damit eine klassische Immunevasionsmutation darstellt (**Abbildung 11D&E**).

Als nächste wurde untersucht, ob die Immunevasionsmutationen auch bei Patienten mit HLA-B*51 Allelen gehäuft vorkommen. In einer Analyse von insgesamt 442 Genotyp-1a-Sequenzen aus einer multizentrischen Kohorte (**Abbildung 12A**) fanden wir keinerlei Evidenz für Selektion von Polymorphismen innerhalb des Epitops. Die Substitution I1373V war leicht angereicht, es wurde jedoch kein signifikanter Unterschied zu HLA-B*51-negativen Patienten festgestellt. Ähnliche Ergebnisse wurde in einer multizentrischen GT1b-Kohorte mit 145 Patienten beobachtet (**Abbildung 12B**). Interessanterweise ergab die Analyse eines sehr großen Genotyp 1b-Ausbruchs, der auf kontaminiertes Anti-D Immunoglobulin zurückzuführen war (die Anti-D-Kohorte), eine signifikante Anreicherung der Polymorphismen I1373V und I1380L in HLA-B*51-positiven Patientinnen (**Abbildung 12C**). Die Selektion des Polymorphismus I1373V konnte auch in einer GT3a-Kohorte mit 102 Patienten bestätigt werden (**Abbildung 12D**).

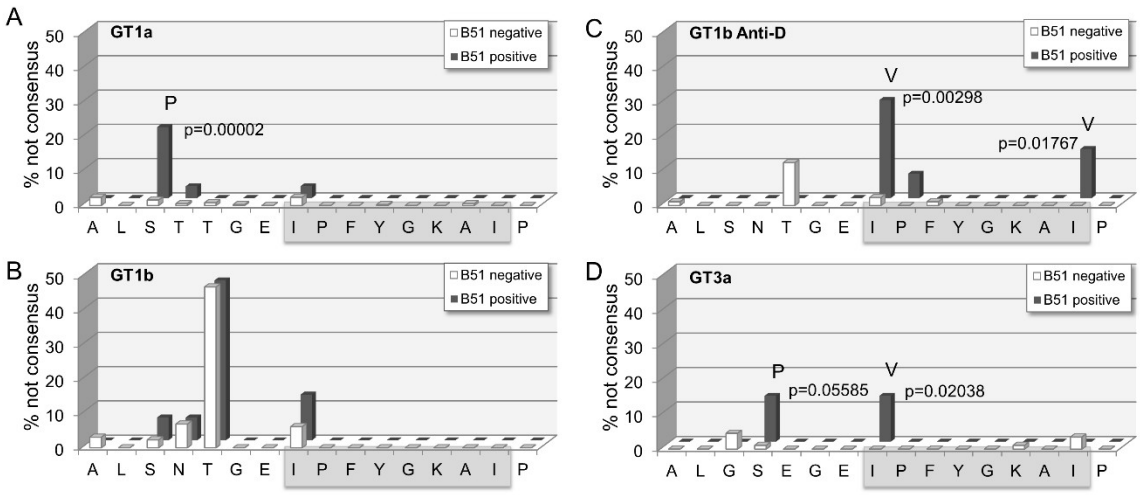


Abbildung 12: Die Frequenz der HLA-B*51-assoziierten Sequenzpolymorphismen in der Epitopregion. *Gezeigt ist die Häufigkeit von Abweichungen von der Prototypsequenz in (A) GT1a, (B) GT1b, (C) der ostdeutschen Anti-D-Kohorte, (D) in GT3a. Patienten mit dem HLA-B*51-Allel sind grau und Patienten ohne das HLA-B*51-Allel sind weiß dargestellt. Positionen mit signifikanten Unterschieden in der Frequenz bei An- verglichen mit Abwesenheit von HLA-B*51 sind markiert und die p-Werte (Fisher-Exakter-Test) sowie die häufigste Aminosäurevariante sind angegeben. Die Veröffentlichung erfolgt mit Genehmigung der American Society for Microbiology.*

Wenngleich bei GT1a keine Polymorphismen innerhalb des Epitope identifiziert wurden, gab es signifikante Anreicherung der Mutation S1368P, also 5 Aminosäuren vor dem Epitop. Diese Mutation war in 20,7 % der HLA-B*51-positiven Patienten vorhanden, aber nur in 0,8 % der HLA-B*51-negativen Patienten (**Abbildung 12A**; p=0.00002). Auch in GT3a wurde dieser Polymorphismus in HLA-B*51-positiven Patienten selektioniert (**Abbildung 12D**).

Die Beobachtung, dass in GT1b und GT3a Substitutionen innerhalb des Epitopes, aber in GT1a nur in den flankierenden Bereichen des Epitopes selektioniert werden, war einzigartig. Da *in silico* Vorhersagen kein weiteres HLA-B*51-restringiertes Epitop in diesem Bereich vorhersagten, wurde daraus geschlossen, dass diese Substitution die Prozessierung des Epitops beeinflusst. Um diese Hypothese zu untersuchen, wurden ein endogener Prozessierungs-Assay entwickelt. Dazu wurden IPFYGKAI-spezifische $CD8^+$ T-Zellen über einen Zeitraum von zehn Tagen *in vitro* expandiert und anschließend mit HLA-B*51-positiven Zielzellen stimuliert. Diese Zielzellen exprimierten das grün fluoreszierende Protein (GFP) sowie ein Fragment des NS3-Proteins, welches das Epitop wie auch die flankierenden Sequenzbereiche enthielt.

In diesen Experimenten zeigte sich, dass die Anwesenheit der S1368P-Mutation die IFNγ-Antwort um 76 % beim Genotyp 1a bzw. um 82 % beim Genotyp 3a reduzierte (**Abbildung 13**). Die Ergebnisse deuten darauf hin, dass die S1368P-Mutation die Antigenprozessierung beeinträchtigt und damit weniger Antigen auf den Zielzellen vorhanden ist.

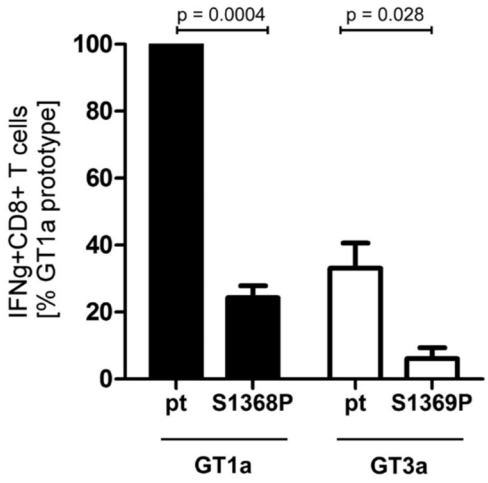


Abbildung 13: Die $CD8^+$ T-Zellantwort gegen das endogen prozessiertes Epitop IPFYGKAI1373–1380. *Expandierte $CD8^+$ Effektor-T-Zellen wurden mit HLA-B*51-positiven Zielzellen, die das Epitop exprimieren, stimuliert. Gezeigt wird die auf Genotyp 1a Prototyp normalisierte IFNγ-Antwort von drei Experimenten. Die Veröffentlichung erfolgt mit Genehmigung der American Society for Microbiology.*

Der zugrundeliegende Mechanismus konnte in weiterführenden *in vitro* Experimenten mittels Proteasomverdau aufgeklärt werden. Beim Verdau von 30 Aminosäuren langen Peptiden, die der Prototyp-Sequenz entsprachen, konnten im Massenspektrometer mehrere 11-13 Aminosäuren lange Vorläufer des Epitops detektiert werden. Beim Verdau des S1368P-Peptids war ausschließlich ein 13 Aminosäure langes Produkt vorhanden, das exakt mit dem Prolin an Position 1368 anfing. Aus der Literatur war bekannt, dass 10-13 Aminosäuren lange Peptide bevorzugt von TAP (Transporter Associated With Prosessing) ins ER transportiert werden und dort auf das HLA-Molekül geladen werden. Für einen erfolgreichen Transport zur Oberfläche müssen die Peptide zuvor jedoch noch von der Aminopeptidase ERAP (Endoplasmic Reticulum Aminopeptidase 1&2) gekürzt werden. Die Aktivität von ERAP wird dabei stark von der N-terminalen Aminosäure beeinflusst. Während Leucin, Lysin und Phenylalanin sehr gut gekürzt werden, werden die Aminosäuren Prolin und Glutaminsäure nur sehr schlecht gekürzt. Das deutet darauf hin, dass in GT1a die Mutation S1368P selektioniert wird, da das beim proteasomalen Verdau entstehende Peptid nicht oder nur in geringen Umsatz von ERAP

gekürzt werden kann. In Folge dessen, wird das Epitop nur in geringen Mengen auf der Oberfläche von infizierten Zellen präsentiert.

Um zu untersuchen, warum die Mutation S1368P nur in GT1a selektioniert werden, wurden infektiöse Viren mit den verschiedenen Substitutionen hergestellt und in Replikationsexperimenten untersucht (**Abbildung 14**). Dabei zeigte sich, dass die Mutation I1380L, die auch die stärksten Immunevasion bei Peptidstimulation zeigte, eine deutlich reduzierte Replikationsfähigkeit hatte. Im Gegensatz dazu war die Replikationsfähigkeit durch die Substitutionen S1368P und I1373V nicht signifikant eingeschränkt.

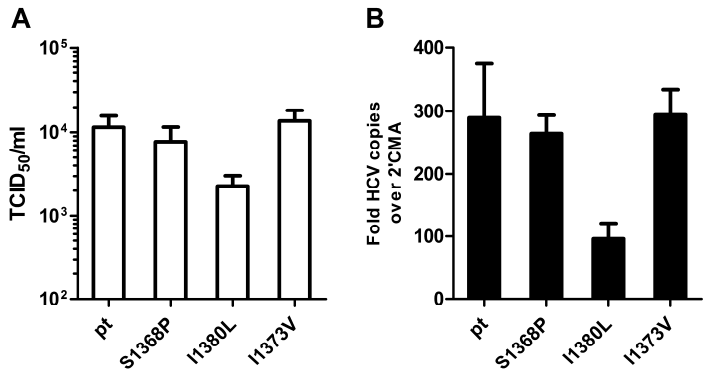


Abbildung 14: Die Infektiosität der Escape-Mutationen. *Huh7.5-Zellen wurden mit infektiöse Prototype bzw. Varianten Viren infiziert. Als Kontrolle dienten Zellen, in denen die Replikation mit 2'methyladenosin (2'CMA) blockiert wurde. (A) Infektiosität nach 72 h ermittelt durch serielle Verdünnungsreihen (TCID50) (B) RNA Konzentration 72 h nach Transfektion ermittelt durch qRT-PCR, normalisiert auf 2'CMA behandelte Zellen. Gezeigt sind die Ergebnisse von vier Experimenten. Die Veröffentlichung erfolgt mit Genehmigung der American Society for Microbiology.*

Zusammenfassend konnten wir in dieser Studie zeigen, dass Immunevasionsmutationen nicht nur innerhalb von Epitopen auftreten können, sondern auch in den flankierenden Bereichen. Obwohl das Epitop in allen Genotypen hochkonserviert ist, scheint es unterschiedliche Wege zum Erreichen der Immunevasion zu geben. Die Variante I1380L ist die effizienteste Immunevasionsvariante, hat aber auch den größten Replikationsnachteil. Die Variante I1373V repliziert gut, hat aber eine hohe Kreuzreaktivität. Die Substitution I1368P scheint damit für das Virus den besten Kompromiss zwischen Immunevasion und Replikation im Genotyp 1a aufzuweisen. Die Selektion von $CD8^+$ Immunevasionsvarianten ist somit immer eine Abwägung zwischen einer reduzierte Erkennung durch das Immunsystem des Wirts und einer Replikationshemmung in Folge der Mutation.

7.2.2 Strategien zur Umgehung der $CD8^+$ T-Zell-vermittelten Immunantwort am Beispiel des Hepatitis B-Virus

In der HBV Forschung gibt es momentan große Bestrebungen, eine vollständige Heilung der HBV Infektion (eine so genannte funktionale Heilung) durch T-Zell-basierte Therapien zu erreichen. Dazu zählen verschiedene Therapieansätze, die in der Onkologie bereits erfolgreich eingesetzt wurden, wie z.B. Checkpoint-Inhibitoren, T-Zell-Vakzine und T-Zellrezeptor-modifizierte Effektorzellen (CAR-T-Zellen). Alle diese Ansätze basieren jedoch auf der Erkennung von HLA-Klasse-I-präsentierten viralen Peptiden auf infizierten Zellen. Das Epitop an der Position 18-27 im HBV Core-Protein ($Core_{18-27}$) ist dabei momentan das Hauptziele für T-Zell-basierte Therapien, da es immundominant ist und von den HLA-Klasse-I Typen HLA-A*02, HLA-B*35 und HLA-B*51 präsentiert werden kann. Es war jedoch bislang unbekannt, ob alle Subtypen der beschrieben HLA- Klasse-I Typen das Epitop präsentieren können und

welche Immunevasionsvarianten bei welchem HLA-Typ selektioniert werden. Aus diesem Grund wurde eine hochauflösende Analyse der $Core_{18-27}$-spezifischen $CD8^+$ T-Zellen, der präsentierenden HLA-Subtypen und der korrespondierenden Immunevasionsvarianten durchgeführt.

Titel der Original Publikation

Immune escape pathways from the HBV core18-27 CD8 T cell response are driven by individual HLA class I alleles

Andreas Walker, Tatjana Schwarz, Janine Brinkmann-Paulukat, Karin Wisskirchen, Christopher Menne, Elahe Salimi Alizei, Helenie Kefalakes, Martin Theissen, Daniel Hoffmann, Julian Schulze zur Wiesch, Mala K. Maini, Markus Cornberg, Anke RM Kraft, Verena Keitel, Hans H. Bock, Peter A. Horn, Robert Thimme, Heiner Wedemeyer, Falko M. Heinemann, Tom Luedde, Christoph Neumann-Haefelin, Ulrike Protzer and Jörg Timm

Das $Core_{18-27}$-Epitop wurde ursprünglich als HLA-A*02-restringiertes, immundominantes Epitop beschrieben. In der Literatur gab es Hinweise, dass das Epitop auch durch andere HLA-Typen präsentiert werden kann. Aus diesem Grund wurde zunächst eine detaillierte Analyse der HLA-Restriktion in 154 Patienten durchgeführt. Dazu wurden PBMCs für 10 Tage mit dem $Core_{18-27}$-Peptid, IL2 und CD3/28-Beads stimuliert und die Frequenz der IFNγ-produzierenden $CD8^+$ T-Zellen analysiert. Dabei fiel auf, dass eine $Core_{18-27}$-spezifische $CD8^+$ T-Zellantwort fast ausschließlich bei HBeAg-negativen Patienten detektierbar war (**Abbildung 15A**). Aus diesem Grund wurden in den weiteren Untersuchungen nur HBeAg-negative Patienten näher analysiert. In Übereinstimmung mit der Literatur konnte sowohl bei Patienten, die HLA-A*02, HLA-B*35 und HLA-B*51-positiv waren, eine $CD8^+$ T-Zellantwort detektiert werden (**Abbildung 15B**). Bei der Analyse der Subtypen zeigten sich jedoch Unterschiede auf. So konnte eine T-Zellantwort bei den Subtypen HLA-B*35:01 und HLA-B*35:03 detektiert werden, aber keine bei HLA-B*35:02 und HLA-B*35:08. Aufgrund der geringen Zahl an nicht-HLA-A*02:01 Subtypen konnten hier keine fundierten Schlussfolgerungen mit der Ausnahm von Typ HLA-A*02:05 gezogen werden, bei dem eine T-Zellantwort detektierbar war. Auch bei HLA-B*51 gab es in unserer Kohorte nur den Subtyp HLA-B*51:01. Interessanterweise zeigte die Gruppe der Patienten, mit mehreren relevanten HLA-Klasse-I-Allelen, im Median eine stärkere $CD8^+$ T-Zellantwort als Patienten, die nur ein relevantes HLA-Klasse-I-Allel besaßen (**Abbildung 15B**). Diese Daten deuten darauf hin, dass das Vorhandensein mehrerer relevanter HLA-Klasse-I-Allele mit einer robusteren Immunantwort assoziiert sein könnte.

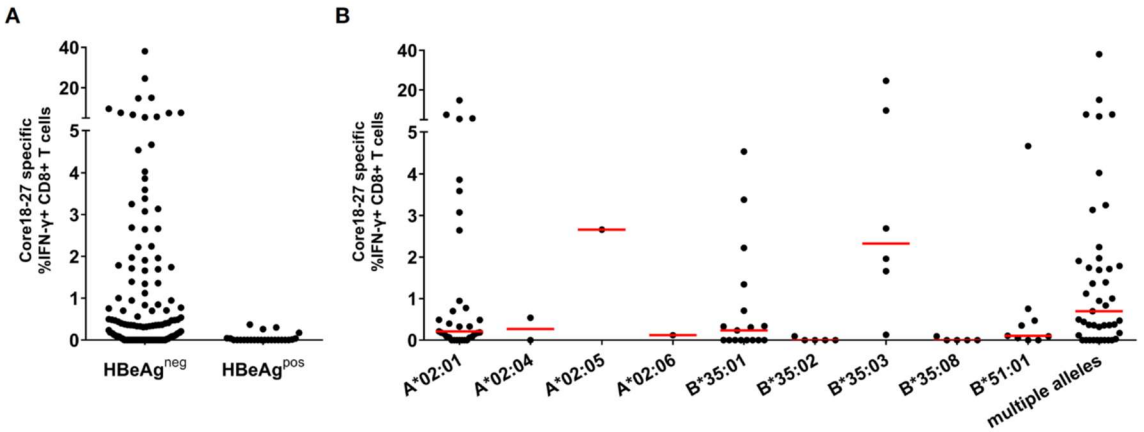


Abbildung 15: Das Core$_{18-27}$ -Epitop kann durch mehrere HLA-Klasse-I Allele präsentiert werden. *PBMCs von 154 chronisch HBV-infizierten Patienten wurden mit dem Core$_{18-27}$ -Peptid für 10 Tage stimuliert und nach Restimulation für 5 h die IFNγ-Produktion mittels Durchflusszytometrie bestimmt. (A) Die IFNγ-Produktion in HBeAg-negativen und -positiven Patienten. (B) Die IFNγ-Produktion der HBeAg-negative Patienten nach HLA-Klasse-I Allelen aufgeteilt. Die Veröffentlichung erfolgt unter CC BY 4.0.*

Nachdem bestätigt werden konnte, dass das Core$_{18-27}$-Epitop von mehreren HLA-Typen nicht aber von allen Subtypen präsentiert werden konnte, wurde als nächstes die Variabilität des Epitops untersucht. In früheren Arbeiten konnten wir und andere bereits zeigen, dass Sequenzvarianten, die in chronisch infizierten HBV-Patienten gefunden werden, Immunevasionsvarianten darstellen [86, 196]. Im Weiteren sollte deshalb untersucht werden, welche Varianten durch die einzelnen HLA-Klasse-I-Moleküle selektioniert werden. Dazu wurde insgesamt 409 Patienten mit einer chronischen HBV-Infektion an verschiedenen Standorten in Europa rekrutiert. Von allen Patienten wurden der HLA-Typ und die Sequenz des Hepatitis B-Virus bestimmt. Beim Vergleich der Aminosäuresequenzen zeigte sich, dass die Core$_{18-27}$-Region hochgradig polymorph ist, wobei die meisten Substitutionen an den Positionen 4,7 und 10 auftraten (**Abbildung 16A**). Auch bei 23 % (26 von 112) Patienten, die keines der relevanten HLA-Moleküle besaßen, traten Substitutionen im Core$_{18-27}$-Epitop auf, wobei die Substitution S4A am häufigsten war. Diese Substitution war auch in Patienten mit HLA*02, HLA-B*35:01/03 oder HLA-B*51:01 nicht signifikant unterschiedlich, was darauf hindeutet, dass sie nicht durch CD8$^+$ T-Zelldruck verursacht wurde.

Bei den anderen Substitutionen gab es jedoch klare Assoziationen mit den jeweiligen HLA-Klasse I Allelen. Beispielsweise wurde bei HLA-B*35:01/03-positiven Patienten die S4T-Substitution reproduzierbar selektioniert, während diese bei HLA-A*02-positiven Patienten nicht oder nur selten selektioniert wurde (**Abbildung 16A&B**). Auch bei HLA-B*51:01-positiven Patienten wurde die Substitution S4T selektioniert, jedoch wurden hier auch Mutationen wie S4G, S4V und andere seltene Substitutionen beobachtet, die bei den anderen HLA-Allelen nicht nachweisbar waren. Im Gegensatz dazu wurde bei HLA-A*02:01-positiven Patienten die F7Y-Substitution allein oder in Kombination mit anderen Substitutionen beobachtet. Die selektive Anreicherung der S4T-Substitution bei HLA-B*35-positiven Patienten und die Anreicherung der F7Y-Substitution bei HLA-A*02-Patienten legt nahe, dass je nach HLA-Klasse-I Typ des Patienten unterschiedliche Immunevasionswege gewählt werden (**Abbildung 16A&B**). Ähnliche Beobachtungen konnten auch auf der Quasispezies-Ebene gemacht werden. Bei der ultratiefen Sequenzanalyse von 96 zufällig ausgewählten Patienten (**Abbildung 16C**)

zeigte sich erneut, dass bei HLA-A*02:01-positiven Patienten die Varianten mit der F7Y-Substitution (Rottöne) verstärkt vertreten waren, während die Variante S4T überhaupt nicht auftrat.

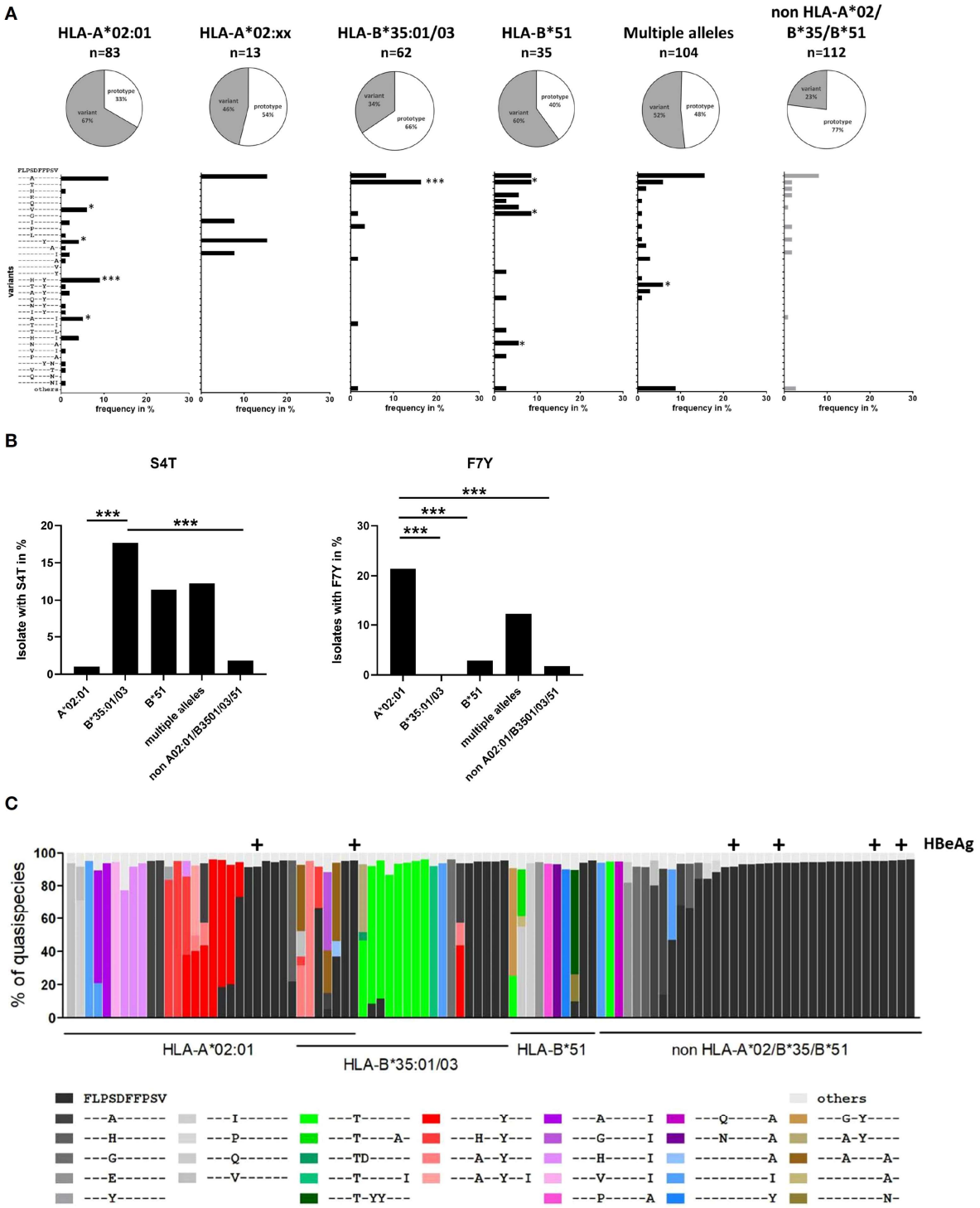


Abbildung 16. Die Core$_{18\text{-}27}$-Epitop ist hoch-variabel in HBeAg-negativen Patienten. *Von 409 Patienten wurde ein 717 bp großes, das komplette Core-Protein abdeckendes Fragment per PCR amplifiziert und sequenziert. (A) Das Kuchendiagramm zeigt die Frequenz der Sequenzen mit einer Abweichung in Relation zum HLA-Klasse-I Typ. Das Balkendiagramm die häufigsten Epitop-Varianten. (B) Die Frequenz der Polymorphismen S4T (links) und F7Y (rechts) unabhängig von anderen Varianten im Epitop. (C) Analyse von HLA-assoziierten Substitutionen in der Quasispezies. Zufällig ausgewählte PCR-Produkte aus A von 96 Patienten wurden mit einer durchschnittlichen Sequenztiefe von 15-20000 x ultratief sequenziert. Jede Säule repräsentiert einen Patienten mit den individuellen Varianten wie in der Legende angegeben. Nur Varianten > 5 % sind gezeigt und HBeAg-positive Patienten sind mit einem (+) gekennzeichnet. Die Veröffentlichung erfolgt unter CC BY 4.0.*

Bei HLA-B*35:01/03-positiven Patienten waren Varianten mit der S4T-Substitution (Grüntöne) häufiger und stellten oft fast 100 % der Quasispezies dar. Der Nachweis von mehreren Epitop-Varianten innerhalb eines Patienten zeigte zusätzlich, dass er zu einer kontinuierlichen Selektion innerhalb der Quasispezies eines Patienten kommt.
Um den Einfluss der Epitop-Variation auf die CD8$^+$ T-Zell-Reaktion zu analysieren, wurden PBMCs in Gegenwart des Prototyp-Epitops expandiert und an Tag 10 mit den häufigsten Epitop-Varianten restimuliert. Dabei zeigte sich, dass die Doppelvarianten wie S4H/F7Y oder S4H/V10I zu einer deutlich reduzierten T-Zellantwort (**Abbildung 17**) in HLA-A*02-positiven Patienten führte. Diese Daten unterstützen damit die Annahmen, dass es sich dabei um Immunevasionsmutation handelt. Interessanterweise zeigte die Variante, die das einzelne F7Y besaß, keine reduzierte CD8$^+$ T-Zellantwort, obwohl sie die einzige Variante war, die in HLA-A*02-positiven Patienten signifikant angereichert vorkam (**Abbildung 16A**). Auch bei der in HLA-B*35:01/03-positiven Patienten selektionierte S4T-Variante gab es keine funktionellen Hinweise auf eine Immunevasion, wenn die Zellen mit exogen hinzugefügtem Peptid stimuliert wurden (**Abbildung 17**, rechts). Dies deutet darauf hin, dass weder die HLA-Klasse-I-Bindung noch die TCR-Bindung an den HLA-Klasse-I/Peptid-Komplex durch diese beiden Varianten beeinträchtigt wird.

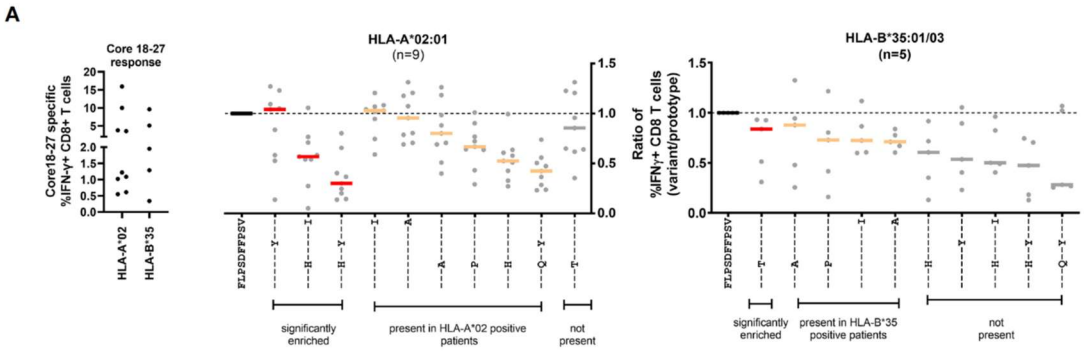


Abbildung 17: Die Kreuzreaktivität der Core$_{18-27}$-spezifischen CD8$^+$ T-Zellen gegenüber natürlich vorkommenden Sequenz Variationen. *PBMCs von HBeAg-negativen Patienten wurden mit dem Core$_{18-27}$ Peptid expandiert und an Tag 10 mit verschiedenen Core$_{18-27}$ Varianten restimuliert. Links: die Core$_{18-27}$-spezifische CD8$^+$ T-Zellantwort, die für die Kreuzreaktivitätsversuche verwendet wurden. Mittig und Rechts davon die auf den Prototyp (FLPSDFFPSV) normalisiert IFNγ-Produktion nach Stimulation mit den Peptidvarianten, die in HLA-A*02:01- oder HLA-B*35:01/03-positiven Patienten beobachtet wurden. Die Veröffentlichung erfolgt unter CC BY 4.0.*

Weitere funktionelle Tests und Bindungsanalysen ergaben, dass die Peptidbindung an HLA-Klasse-I-Moleküle bei exogener Zugabe des synthetischen Peptids nicht beeinträchtigt war. Da diese Assays jedoch keine endogene Prozessierung und Präsentation berücksichtigen, wurde untersucht, ob Sequenzvarianten innerhalb oder in der Nähe des Epitops die Antigenprozessierung beeinflussen.

Dazu wurden HepG2-Zellen, die natürlicherweise HLA-A*02:01 exprimieren, stabil mit unterschiedlichen Varianten des Core-Proteins transduziert und als Target-Zellen für 10 Tage expandierte PBMCs verwendet. Zellen, die exogen mit dem jeweiligen Peptid beladen wurden, dienten dabei als Kontrolle (**Abbildung 18**).

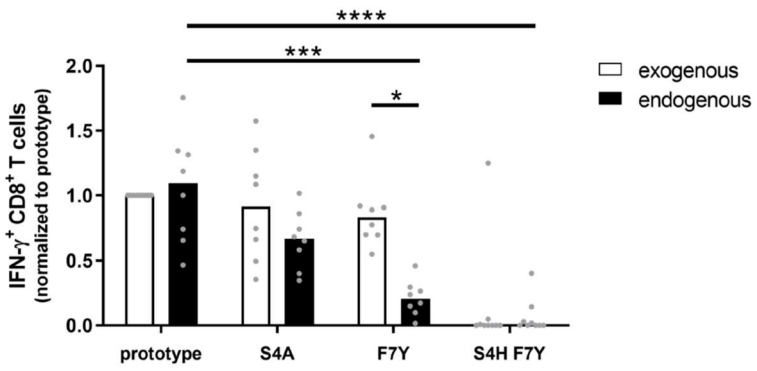


Abbildung 18: Die Core$_{18-27}$-spezifische CD8$^+$ T-Zellantwort gegen endogen prozessierte Epitop-Varianten. *Zehn Tage mit Core$_{18-27}$ expandierte Effektorzellen wurden mit HepG2-Zielzellen, die entweder den Prototyp oder eine Variante des HBV-Core-Protein exprimierten, restimuliert (schwarze Balken). HepG2, die über Nacht mit dem Peptid exogen beladen wurden, dienten als Kontrolle (weiße Balken). Die Veröffentlichung erfolgt unter CC BY 4.0.*

Unabhängig davon, ob das Peptid endogen prozessiert wurde oder exogen auf die Zellen beladen wurde, gab es bei der Prototyp-Sequenz keine signifikanten Unterschiede in der CD8$^+$ T-Zellantwort (**Abbildung 18**). Die S4A-Mutation führte zu einer leicht, aber nicht signifikant verringerten IFNγ-Produktion bei der endogenen Prozessierung. In Einklang mit Daten aus früheren Experimenten führte die Doppelsubstitution S4H-F7Y zu einer deutlich reduzierte CD8$^+$ T-Zellantwort. Interessanterweise zeigte das F7Y-Peptid bei exogener Zugabe wieder eine vergleichbare T-Zellantwort wie die Prototyp-Sequenz. Wenn das Epitop dagegen endogen prozessiert wurde, reduzierte sich die IFNγ-Antwort um 75 % (**Abbildung 18**). Das deutet darauf hin, dass auch bei HBV-Infektionen Varianten selektioniert werden, die durch eine gestörte Antigenprozessierung der CD8$^+$ T-Zellantwort entkommen.

Zusammenfassend zeigte diese Studie, dass das Core$_{18-27}$ -Epitope tatsächlich von mehreren HLA-Klasse-I Allelen präsentiert werden kann. Für T-Zell-basierte Immuntherapien stellt das einen enormen Vorteil dar, da es die Zahl an behandelbaren Patienten deutlich erhöht. Gleichzeitig konnten wir aber auch zeigen, dass nicht alle Subtypen der jeweiligen Klasse-I Allee das Core$_{18-27}$-Epitop präsentiert, was das therapeutische Spektrum wieder eingrenzt und eine hochauflösende HLA-Klasse-I Bestimmung vor Therapie erforderlich macht. Die Sequenzanalysen bestätigten frühere Beobachtungen, dass das Core$_{18-27}$-Epitope unter starkem Selektionsdruck steht. Interessanterweise werden in Abhängigkeit des HLA-Klasse-I Alles unterschiedliche Mutationen selektioniert. Die meisten dieser Mutationen scheinen die Bindung des T-Zellrezeptors an den Peptid-/HLA-Klasse-I-Komplex zu beinträchtigen, womit eine Therapie mit breit-kreuzreaktiven TCR möglich sein könnte. Gleichzeitig konnten wir auch Substitutionen finden, die die Antigen-Prozessierung und die Präsentation von Epitopen auf der Zelloberfläche reduzieren. Viren, die solche Substitutionen enthalten, werden gegenüber CD8$^+$-basierten Immuntherapien weniger anfällig sein. Die Sequenzierung des viralen Genoms vor einer T-Zelltherapie sollte deshalb ein fester Bestandteil zukünftiger T-Zelltherapien sein.

7.2.3 Die Selektion von Immunevasionsvarianten im HBV Core-Protein korreliert mit einer verminderter Viruslast

Aufgrund der HLA-Restriktion bei CD8$^+$ T-Zellen, sind die daraus resultierenden Immunevasionsmutationen stets mit einem bestimmten HLA-Klasse-I-Allel assoziiert. Dies konnte in der oben beschriebenen Publikation exemplarisch anhand der Substitutionen S4T und F7Y gezeigt werden, die nur bei HLA-B*35:01/03- bzw. HLA-A*02:01-positiven Patienten vorkommen. Durch die Analyse zahlreicher Patientensequenzen in Kombination mit den jeweiligen HLA-Klasse-I-Allele lassen sich damit auf Populationsebene neue Epitope identifizieren und grundlegende Aussagen zum CD8$^+$ T-Zell-

bedingten Selektionsdruck treffen. Um das Ausmaß des HLA-Klasse-I-vermittelten Selektionsdrucks auf das komplette HBV-Genom zu intersuchen, wurde die Analyse auf alle HBV-Proteine ausgeweitet.

Titel der Original Arbeit

HBV shows different levels of adaptation to HLA class I-associated selection pressure correlating with markers of replication

Tatjana Schwarz, Johannes Ptok, Maximilian Damagnez, Christopher Menne, Elahe Salimi Alizei, Julia Lang-Meli, Michelle Maas, Daniel Habermann, Daniel Hoffmann, Julian Schulze zur Wiesch, Georg M. Lauer, Helenie Kefalakes, Markus Cornberg, Anke R.M. Kraft, Smaranda Gliga, Hans H. Bock, Peter A. Horn, Mala K. Maini, Robert Thimme, Heiner Wedemeyer, Jacob Nattermann, Falko M. Heinemann, Tom Luedde, Christoph Neumann-Haefelin, Andreas Walker, Jörg Timm

Um dem Selektionsdruck durch das Immunsystem zu untersuchen, wurde von 532 Patienten der genaue HLA-Klasse-I-Genotyp bestimmt und das komplette HBV-Genom sequenziert. HLA-assoziierte Mutationen (sogenannte HAMs) wurden mithilfe eines von uns mitentwickelten Bayesschen-Verfahren, dem sogenannten HAM-Detektor ermittelt [197]. Der HAM-Detektor berechnet dabei für jede Position im HBV-Genom, in der es eine Abweichung von der Konsensussequenz gibt, die A-Posteriori-Wahrscheinlichkeit (HAM-Score), dass es sich dabei um eine HLA-assoziierte Mutation handelt. Als Bayesscher Priori werden neben dem HLA-Typ, die Mutationsverteilung im Alignment, das HLA-Bindungsmotiv, die phylogenetische Verwandtschaft der Sequenzen sowie der Grad der Konservierung berücksichtigt (**Abbildung 19A**).

Der Algorithmus gibt dabei an, wie wahrscheinlich es ist, dass eine Mutation mit einem HLA-Klasse-I-Typ assoziiert ist. Ein Posterior-p-Wert> 0,5 bedeutet, dass die Aminosäure häufiger mit dem spezifischen HLA-Klasse-I Allel anzutreffen ist, während bei einem Posterior-p-Wert < 0,5 die Aminosäure weniger häufig in Patienten vorkommt, die dieses Allel tragen. In **Abbildung 19B** ist exemplarisch die A-Posteriori-Wahrscheinlichkeit der Aminosäuren im $Core_{18-27}$-Epitop für das HLA-A*02:01-Allele dargestellt. Da die meisten HLA-A*02:01-positiven Patienten an der Position 24 im Core-Protein (F7Y in vorheriger Publikation) ein Tyrosin tragen, beträgt der Posterior-p-Wert für Tyrosin 1,0. Gleichzeitig ist in HLA-A*02:01-positiven Patienten die Prototyp-Aminosäure Phenylalanin deutlich seltener, womit sie einen Posterior-p-Wert von 0 besitzt. Der Vorteil des Bayesschen-Verfahren liegt in der quantitativen Bestimmung der Substitution für jedes HLA-Klasse-I Allele in der Kohorte. In Validierungsstudien zum HAM-Detektor zeigte sich, dass die meisten experimentell bestätigten Immunevasionsmutationen einen HAM-Score von über 0,8 haben. Dadurch können mit großer Wahrscheinlichkeit neue Epitope gefunden werden, die durch die hohe Variabilität der Sequenzen in bisherigen HLA-Assoziationsstudien bislang nicht identifiziert werden konnten. So konnte beispielsweise für das potentielle Epitop in der Polymerase (pol_{14-22} CPTVKASKL) eine hohe A-Posteriori-Wahrscheinlichkeit für Phenylalanin an Position 4 des Epitops (V17F) bei HLA-B*35:01-positiven Personen detektiert werden (**Abbildung 19C**). In der konventionellen statistischen Analyse wäre die V17F-Substitution mit einem p-Wert von 0,11 nicht detektiert worden.

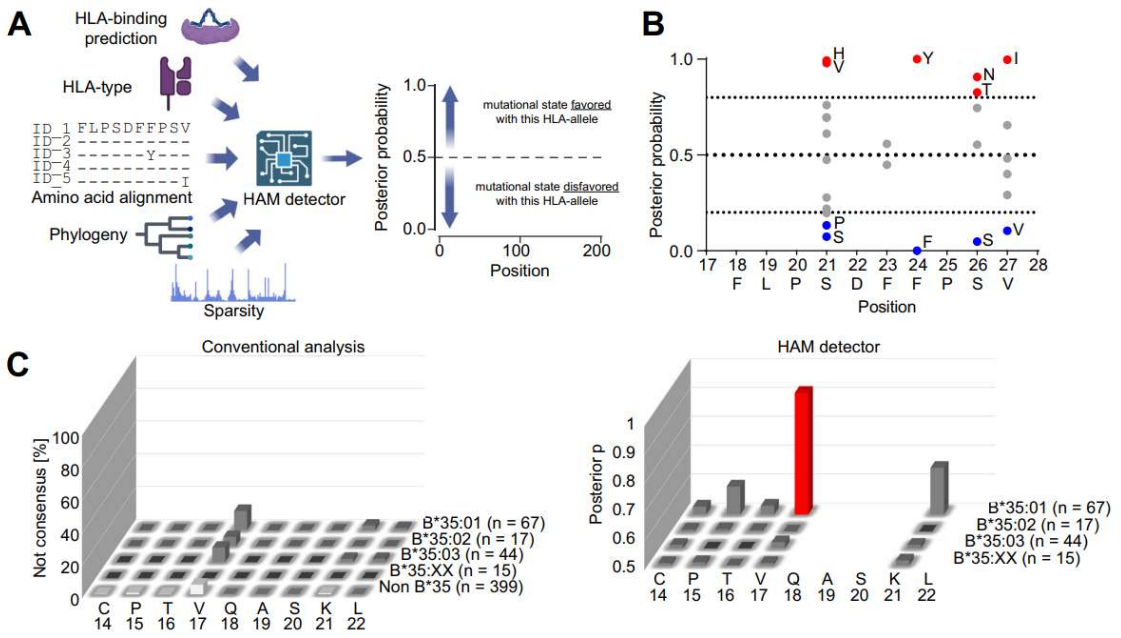


Abbildung 19 Die Vorteile des HAM-Detektor zur Identifizierung von HLA-Klasse-I-assoziierten Substitutionen im Vergleich zu bisherigen statistischen Methoden. *(A) Schematische Darstellung des HAM-Detektors. Das Bayessche-Model integriert neben der Virussequenz und dem HLA-Typ des Patienten auch die Phylogenie, das Bindungsmotiv des HLA-Moleküls sowie den Konservierungsgrad der Position. Substitutionen mit einer hohen A-Posteriori-Wahrscheinlichkeit (HAM-Score) sind häufig mit dem HLA-Typ anzutreffen, Substitutionen mit einem niedrigen HAM-Score dagegen selten. (B) Darstellung des HAM-Scores für das HLA-A*02-restringierte Epitop Core$_{18-27}$. (C) Vergleich zwischen der konventionellen Analyse (links) und der HAM-Detektoranalyse (rechts) für das neu identifizierte HLA-B*35:01-restringierte Epitop in der Polymerase (pol$_{14-22}$ CPTVKASKL). Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV*

Bei der Analyse aller HBV-Proteine zeigten sich deutliche unterschiedliche hinsichtlich der HAM-Verteilungen. Während bei einer randomisierten Kontrolle keine Substitution über 0,8 detektiert wurde (**Abbildung 20A**), konnten im Core-Protein insgesamt 247 Substitutionen mit einer A-Posteriori-Wahrscheinlichkeit über 0,8 detektiert werden (**Abbildung 20B**). Im Gegensatz dazu waren im HBs-Oberflächenprotein und dem X-Protein nur vereinzelt HAMs detektierbar. In der viralen Polymerase wurden HAMs fast ausschließlich im N-terminalen, mit dem Core-Protein überlappenden Bereich detektiert (**Abbildung 20D-E**).

Die mittels HAM-Detektor identifizierten Häufungen von HAMs im Core-Protein erwies sich als unerwartet. Grundsätzlich werden Zwei mögliche Erklärungen für dieses Phänomen diskutiert: Entweder enthält im Core-Protein tatsächlich mehrere Epitope, oder der auf das Core-Protein wirkende Selektionsdruck ist höher als auf den übrigen viralen Proteinen. Um ersteres zu untersuchen, wurden daher alle bisher beschriebenen und bestätigten $CD8^+$ T-Zellepitope aus der Immunepitop-Datenbank näher analysiert. Diese Analyse ergab, dass bisher 54 Epitope in der Polymerase, 42 im Core-Protein, 27 im Envelope-Protein und 10 im X-Protein beschrieben wurden, womit der erste Erklärungsansatz verworfen wurde. Analysiert man den HAM-Score der bereits beschriebenen Epitope, fällt auf, dass Epitope im Core-Protein einen substantiell höheren HAM-Score haben als die Epitope in den anderen HBV-Proteinen (**Abbildung 20F**). Dies ist ein starker Indikator dafür, dass der Selektionsdruck auf dem Core-Protein tatsächlich höher ist als auf andere virale Proteine im HBV-Genom.

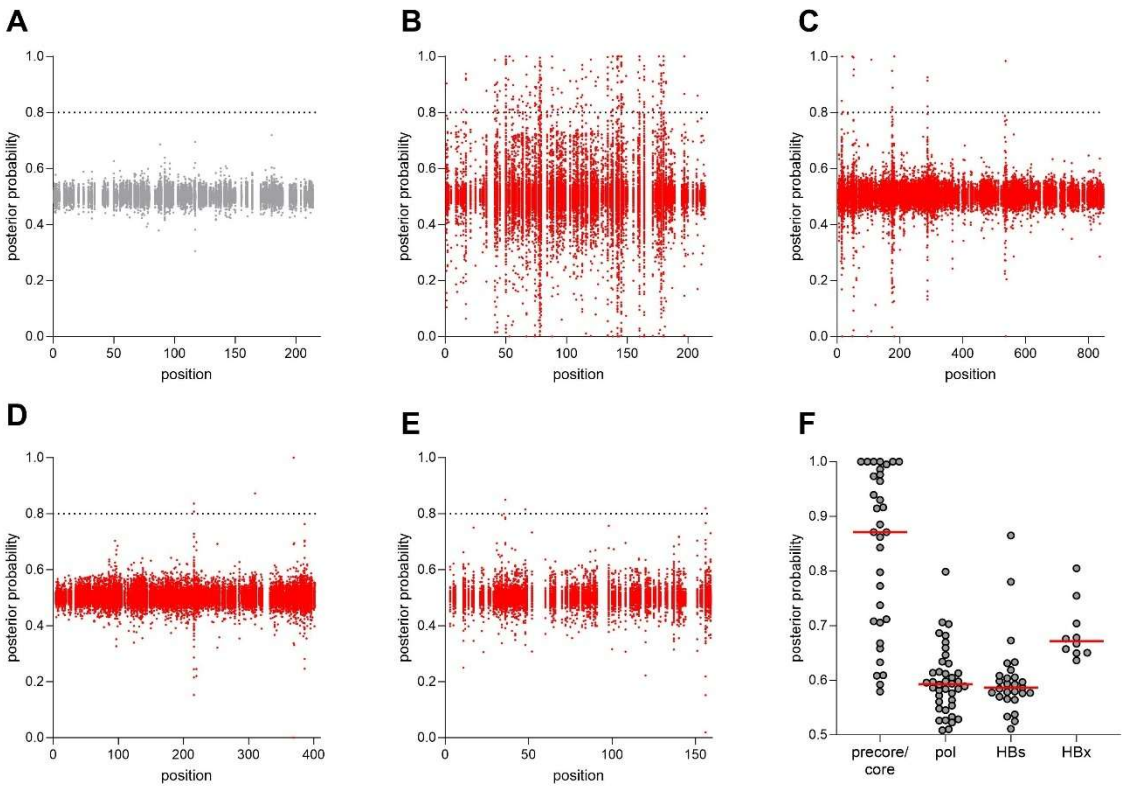


Abbildung 20: HLA-Klasse-I-assoziierte Substitutionen in den HBV-Proteine Precore/Core, Polymerase, HBsAg und dem X-Protein. Die Alignments der Proteine wurden mit dem HAM-Detektor analysiert. (A) HAM-Detektoranalyse für das Precore-/Core-Protein mit randomisierten HLA-Typen. (B) HAM-Detektoranalyse für das Precore-/Core-Protein mit korrekt zugeordnetem HLA-Typ, (C) Polymerase, (D) HBsAg (E) HBx. Die gepunktete Linie bei 0,8 zeigt den empirischen Schwellenwert, ab der in früheren Arbeiten die häufigsten Substitutionen in echten $CD8^+$ T-Zellepitope lagen. (F) Der HAM-Score von bereits bestätigten $CD8^+$ T-Zellepitopen aus der IEDP-Datenbank. Dargestellt ist jeweils der höchste HAM-Score jedes Epitopes. *Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV*

Nachdem gezeigt werden konnte, dass auf dem Core-Protein ein erhöhter Selektionsdruck liegt, stellte sich die Frage, ob ein erhöhter Selektionsdruck eine geringere Viruslasten bedingt. Dazu wurde für jeden Patienten ein individueller Adaptationsgrad berechnet (siehe **Abbildung 21A**). Zu diesem Zweck wurden zunächst alle Mutationen des Patientenisolates im Vergleich zur Konsensussequenz ermittelt. Anschließend wurde anhand des HLA-Typs des Patienten der HAM-Score jeder Mutation erfasst. Die Summe aller HAM-Scores ergibt den Patienten-spezifischen Adaptationsgrad (Adaptation-Score). Ein niedriger Adaptationsgrad steht für wenig Mutationen bzw. für wenig Mutationen in Bereichen, die unter HLA-Klasse-I-Selektionsdruck stehen, währen ein hoher Score für ein stark mutiertes Virus mit vielen HLA-Klasse-I assoziierten Mutationen steht. Beim Vergleich zwischen HBeAg-positiven und HBeAg-negativen Patienten zeigte sich, dass HBeAg-positive Pateinten einen niedrigeren Adaptation-Score haben als HBeAg-negative Patienten (**Abbildung 21B**). Diese Daten stehen in Einklang mit der Lehrmeinung, dass HBeAg-positive Patienten in einer immuntoleranten Phase sind. Interessanterweise zeigte sich, dass Patienten mit einem hohen Adaptation-Score signifikant niedrigere Viruslasten (**Abbildung 21C**) und signifikant weniger HBsAg (**Abbildung 21D**) aufwiesen. Zudem stieg der Adaptation-Score mit dem Alter an (**Abbildung 21E**), was damit erklärt werden kann, dass die ständige Anpassung an den Immundruck in HBeAg-negativen Patienten über die Zeit zu einer Akkumulation von replikationshemmenden Mutationen im HBV-Genom führt.

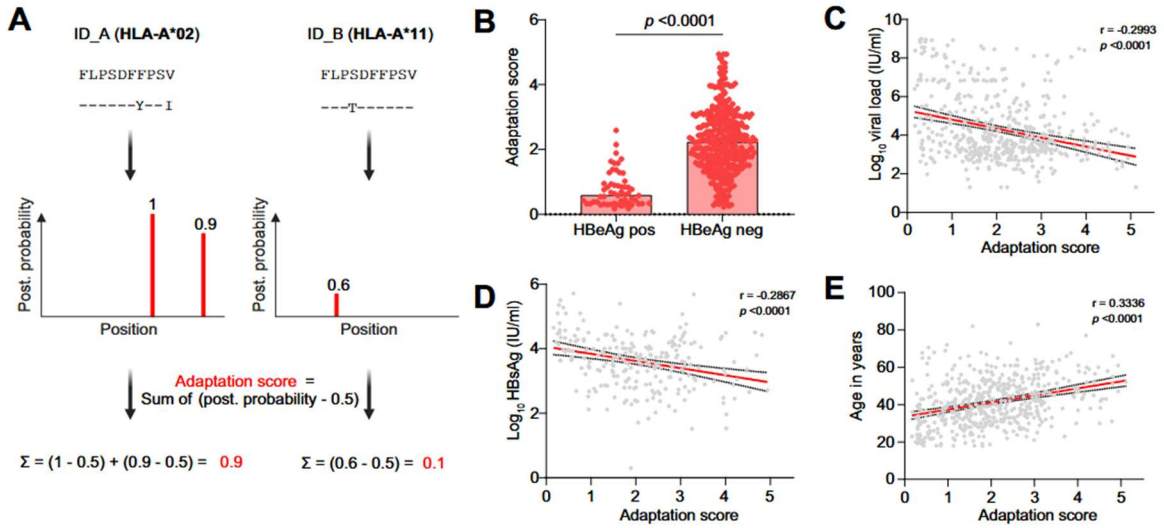


Abbildung 21: Die Anpassung an den HLA-Klasse-I-induzierten Selektionsdruck führt zu verminderter Replikation. *(A) Schematische Darstellung der Berechnung des Adaptationsgrades (Adaptation-Score). (B) Vergleich des Adaptationsgrades in HBeAg-positive und -negativen Patienten (die statistische Signifikanz wurde mit dem Welch's t-Test berechnet). Korrelation des Adaptationsgrades mit (C) der Viruslast, (D) dem Serumlevel des HBsAg, (E) dem Alter des Patienten. Die statistische Signifikanz wurde mit einer einfachen linearen Regression berechnet. Die Veröffentlichung erfolgt mit Genehmigung des Verlags Elsevier BV*

Diese Studie lieferte wichtige Erkenntnisse über das Ausmaß des HLA-Klasse I-assoziierten Selektionsdrucks auf HBV. Insgesamt konnten 295 Aminosäuren unter Selektionsdruck identifiziert werden, von denen ein Großteil in bisher noch nicht beschriebenen $CD8^+$ T-Zellepitopen lag. Darüber hinaus konnte gezeigt werden, dass der höchste Selektionsdruck auf dem Core-Protein liegt und die Sequenzdiversität im Core-Protein durch den $CD8^+$ T-Zell-vermittelten Selektionsdruck entsteht. Die Korrelation von niedrigeren Viruslasten mit hohen Adaptations-Scores kann am ehesten dadurch erklärt werden, dass die Replikation durch die Vielzahl an Mutationen eingeschränkt wird. Dazu würde auch die Beobachtung passen, dass der Grad der Anpassung mit dem Patientenalter zunimmt.

Für die Entwicklung von T-Zell-basierten Therapieansätzen wird der Adaptationsgrad des Virus in Zukunft eine wichtige Rolle spielen, da davon auszugehen ist, dass bereits stark angepasst Viren nicht effizient auf T-Zell-basierten Therapien reagieren werden. Daher ist entscheidend, Epitope auszuwählen, die nur einem geringen Adaptationsgrad unterliegen oder deren Substitutionen mit einer sehr niedrigen Viruslast assoziiert sind, was auf eine eingeschränkte Virusfitness hinweist.

7.3 Die Verwendung der natürlichen SARS-CoV-2 Evolution für die Integrierte Genomische Surveillance

Insbesondere zu Beginn der Corona-Pandemie war die Ermittlung von Kontakten und die Quarantäne potentiell infizierter Personen die einzige Möglichkeit, die Verbreitung des SARS-CoV-2-Virus zu verhindern. Die Idee, Erreger zu sequenzieren und über das Erbgut Infektionsketten nachzuweisen, wurde bei bakteriellen Ausbrüchen schon länger verwendet, war aber zu diesem Zeitpunkt in der Virologie nicht weit verbreitet. Aufgrund der natürlich auftretenden Mutationen und der Entstehung neuer Virusvarianten besitzt jedes Virus einen eigenen genetischen Barcode. Die durchschnittliche Mutationsrate von 2 Mutationen pro Monat bei SARS-CoV-2 ist ausreichend, um unterschiedliche Isolate voneinander abzugrenzen. In diesem Kapitel soll gezeigt werden, wie natürliche Mutationen für die Genomische Surveillance von SARS-CoV-2 verwendet werden können und wie schließlich eine Integrierte Genomische Surveillance unter Einbeziehung der Kontaktnachverfolgungsdaten für Düsseldorf aufgebaut wurde (**Abbildung 22**).

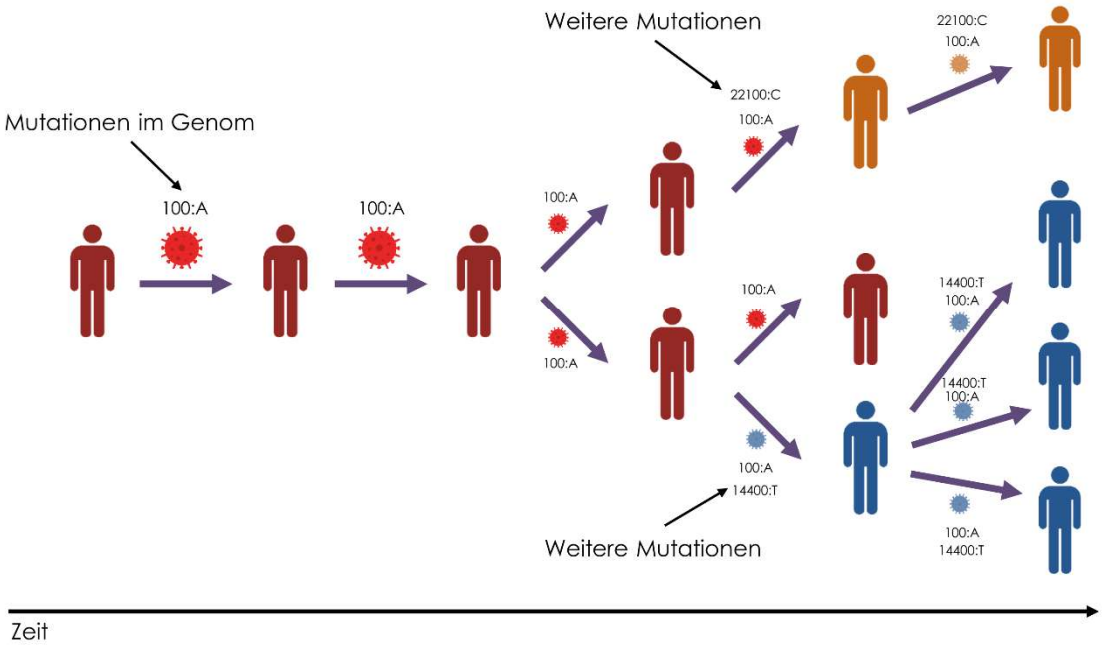


Abbildung 22: Schematische Darstellung der Genomischen Surveillance. *Anhand der Untersuchung spezifischer Mutationsmuster im Virusgenom und mithilfe von weiteren Informationen wie Symptombeginn oder Testdatum können Infektionsketten zuverlässig zurückverfolgt werden. Eigene Abbildung*

7.3.1 Die Aufklärung des zweiten großen deutschen SARS-CoV-2-Ausbruchs in Heinsberg

Zu Beginn der Corona-Pandemie war unklar, wie schnell sich SARS-CoV-2 ausbreiten würde und welche Infektionswege für die Übertragung relevant sind. Die ersten SARS-CoV-2 Fälle in Deutschland in Bayer und Baden-Württemberg wurden in einer Firma durch eine infizierte Reiserückkehrerin aus China ausgelöst. Aufgrund der Nachvollziehbarkeit der Kontakte konnte durch konsequente Kontaktnachverfolgung und Quarantäne-Maßnahmen der Ausbruch eingedämmt werden [180]. Am 24./25. Februar 2020 wurde im Universitätsklinikum Düsseldorf ein Ehepaar aus Heinsberg erstmals positiv auf SARS-CoV-2 getestet, welches keine Reiseanamnese in SARS-CoV-2-Risikogebiete hatte. Drei Tage

später waren bereits 37 Fälle bestätigt und einen Monat später am 22. April war die Fallzahl bereits auf über 1.700 Fälle angestiegen. Retrospektiv konnte die Kappensitzung des örtlichen Karnevalvereins in der Gemeinde Gangelt im Kreis Heinsberg als wahrscheinlichstes Superspreading-Event ermittelt werden. Mit dem Auftreten der ersten SARS-CoV-2 Fälle in Düsseldorf am 3. März 2020 wurde der Verdacht geäußert, dass die Düsseldorfer SARS-CoV-2 Fälle aufgrund der geographischen Nähe aus Heinsberg stammten.

Titel der Original Publikation

Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020

Andreas Walker, Torsten Houwaart, Tobias Wienemann, Malte Kohns Vasconcelos, Daniel Strelow, Tina Senff, Lisanna Hülse, Ortwin Adams, Marcel Andree, Sandra Hauka, Torsten Feldt, Björn-Erik Jensen, Verena Keitel, Detlef Kindgen-Milles, Jörg Timm, Klaus Pfeffer, Alexander T Dilthey

Bereits Anfang März wurde die SARS-CoV-2 Vollgenomsequenzierung basierend auf dem ARTIC-Protokoll am Universitätsklinikum Düsseldorf etabliert. Daher konnten Vollgenomsequenzen der Isolate beider Index-Fälle bereits am 12. März bei GISAID veröffentlicht werden. Um zu untersuchen, ob die Düsseldorf Index-Fälle tatsächlich auf den Ausbruch in Heinsberg zurückzuführen waren, wurden alle Düsseldorfer SARS-CoV-2 Isolate, die zu diesem Zeitpunkt zur Verfügung standen (n=45) sequenziert und zur Bestimmung der genetischen Verwandtschaft ein phylogenetischer Baum in Form eines Minimum-Spanning Trees berechnet. Dabei zeigte sich, dass die zwischen dem 25. und 28. Februar 2020 entnommen Proben, einschließlich der beiden Index-Fälle, von klonalem Ursprung waren (**Abbildung 23**, schwarze Punkte). Insgesamt konnten fünf unterschiedliche Virusvarianten aus Heinsberg identifiziert werden, die zwischen zwei und sechs polymorphen Positionen im Vergleich zur Wuhan-Referenzsequenz hatten (**Abbildung 23**, schwarze Punkte). Alle Heinsberg-Proben wiesen im Vergleich zur SARS-CoV-2 Wuhan-1 Referenzsequenz zwei für den Heinsberg-Cluster charakteristische Mutationen auf. Im Vergleich zu anderen öffentlich verfügbarer SARS-CoV-2-Sequenzen ergab sich kein eindeutiger Ursprung des Heinsberg-Ausbruchs. Auch zeigten die Heinsberg-Isolate keine Verwandtschaft zu den Sequenzen früherer Isolate aus Deutschland oder den Niederlanden.

Im Gegensatz dazu gehörten die Isolate von Düsseldorfer Patienten zu mindestens fünf verschiedenen Clustern. Bemerkenswert ist, dass nur vier Patientenisolate mit dem Heinsberg-Ausbruch verwandt waren (**Abbildung 23**, gelbe Punkte im Heinsberg Cluster); Zwei dieser Isolate stammten von Bewohnern eines an den Kreis Heinsberg angrenzenden Bezirks, die im Universitätsklinikum Düsseldorf behandelt wurden. Die Herkunft der anderen beiden Isolate blieb unklar. In den anderen Clustern waren einige Isolate identisch oder beinah identisch (genetische Distanz 0 oder 1) zu Isolaten aus Australien, dem Vereinigten Königreich, den Vereinigten Staaten und zahlreichen weiteren Ländern, was stark auf mehrere unabhängigen Eintragungen durch Reiserückkehrer hindeutete. Somit fanden sich keine Hinweise auf eine weitverbreitete Eintragung von Heinsberg-abgeleiteten SARS-CoV-2-Stämmen in den Raum Düsseldorf.

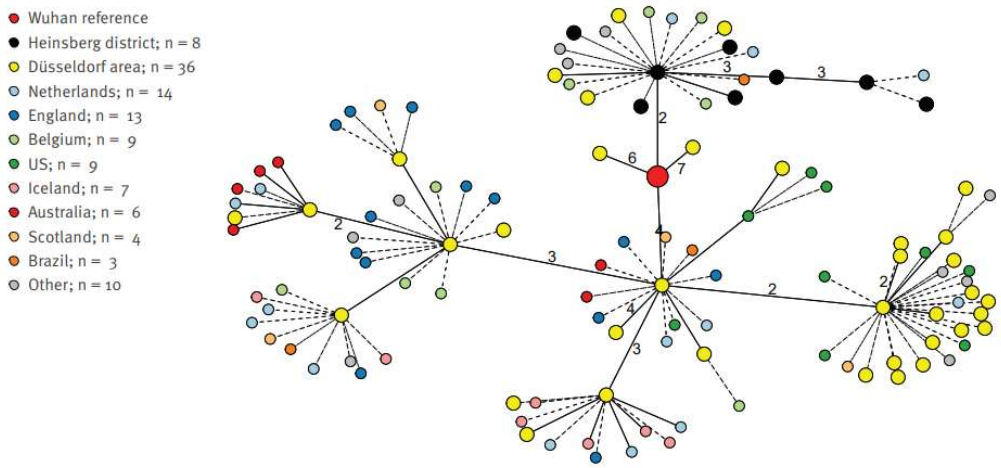


Abbildung 23: Die genetische Verwandtschaft der frühen SARS-CoV-2 Isolate aus Düsseldorf und Heinsberg. *Gezeigt sind vollständig sequenzierte SARS-CoV-2 Isolate. Die Distanzberechnung wurde mit der mst-Funktion der R-Software pegas berechnet und mit sna visualisiert. Jeder Punkt stellt ein Isolat dar und die genetische Distanz zwischen den Isolaten ist 0 bei gestrichelten Linien, 1 bei durchgezogenen Linien oder ansonsten sind die Nukleotidunterschiede als Zahl zwischen den Punkten angegeben. Die Veröffentlichung erfolgt unter CC BY 4.0*

Durch die neu eingeführte SARS-CoV-2-Ganzgenomsequenzierung konnten wir zeigen, dass der Heinsberg-Ausbruch tatsächlich ein klonales Ereignis war, wahrscheinlich ausgelöst durch die Kappensitzung am 15. Februar in Gangelt. Die fehlende genetische Verwandtschaft zwischen den Heinsberg-Proben und anderen frühen deutschen Isolaten weist auf ein unabhängiges Einführungsgeschehen hin, möglicherweise über die Niederlande, China oder ein drittes Land.

Im Gegensatz dazu waren die ersten SARS-CoV-2-Isolate aus Düsseldorf polyklonal und hatten eine enge genetische Verwandtschaft mit internationalen Stämmen, was auf unabhängige Eintragungen hindeutet. Trotz der nahen geographischen Lage von Düsseldorf zu Heinsberg (70 km) waren nur 4 Isolate dem Heinsberg-Cluster zugehörig, was darauf schließen lässt, dass die früh ergriffenen Lockdown-Maßnahmen in Heinsberg eine starke Verbreitung verhindert haben.

7.3.2 Die Etablierung einer Integrierten Genomischen SARS-CoV-2 Surveillance in Düsseldorf

Die Analyse des Heinsberg-Ausbruchs zeigte, dass die Genomische Surveillance die Zuordnung zu Infektionsclustern erlaubt und damit Infektionsketten nachvollziehbar werden. Eine exakte Rekonstruktion des Infektionsgeschehens benötigt aber immer auch weitere Daten aus der Kontaktnachverfolgung. Im Sommer 2020 stellten wir uns deshalb die Frage, ob es möglich ist, in einer Großstadt wie Düsseldorf durch die Integration der Genomischen Surveillance mit den Daten der Kontaktnachverfolgung des zuständigen Gesundheitsamtes das Infektionsgeschehen in der Stadt abbilden zu können. Der Fokus der Tätigkeit lag insbesondere auf der Untersuchung von Superspreading-Events und auf bisher unbekannte Infektionsquellen.

Titel der Original Arbeit

Characterization of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection Clusters Based on Integrated Genomic Surveillance, Outbreak Analysis and Contact Tracing in an Urban Setting

Andreas Walker, Torsten Houwaart, Patrick Finzer, Lutz Ehlkes, Alona Tyshaieva, Maximilian Damagnez, Daniel Strelow, Ashley Duplessis, Jessica Nicolai, Tobias Wienemann, Teresa Tamayo, Malte Kohns Vasconcelos, Lisanna Hülse, Katrin Hoffmann, Nadine Lübke, Sandra Hauka, Marcel Andree, Martin P. Däumer, Alexander Thielen, Susanne Kolbe-Busch, Klaus Göbels, Rainer Zotz, Klaus Pfeffer, Jörg Timm and Alexander T. Dilthey; on behalf of the German COVID-19 OMICS Initiative (DeCOI)

In Zusammenarbeit mit dem Gesundheitsamt und einem diagnostischen Labor aus Düsseldorf wurden zwischen August und Dezember 320 zufällig ausgewählte SARS-CoV-2-positive Patientenproben aus Düsseldorf am Institut für Virologie sequenziert. Im Durchschnitt wurden 19 Proben pro Woche sequenziert, womit im Beobachtungszeitraum ca. 3 % der Düsseldorfer SARS-CoV-2-Proben sequenziert wurden (**Abbildung 24A**).

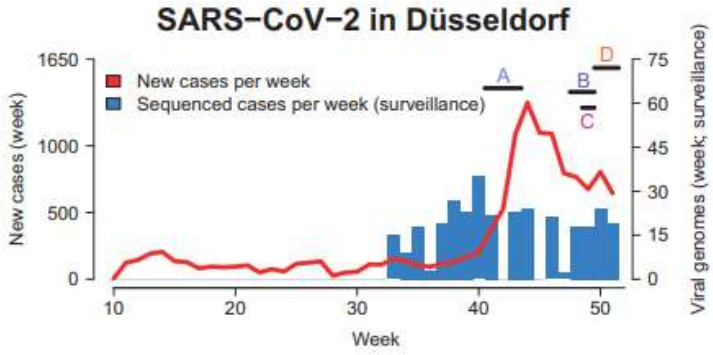


Abbildung 24: Die Entwicklung der SARS-CoV-2- Infektionen von September bis Dezember 2020. *Neudiagnosen sind in rot und sequenzierte Fälle als blaue Balken dargestellt. Die Kalenderwochen sind auf der X-Achse dargestellt. Die horizontalen Striche stellen die Proben-entnahmezeitpunkte der Krankenhaus-ausbrüche auf den Stationen A–D des Düsseldorfer Universitätsklinikums dar. Die Veröffentlichung erfolgt unter CC BY 4.0*

Die Entwicklung der viralen Varianten entsprach im Wesentlichen den in Europa vorkommenden Trends. Die so genannte Klade 20E, die zunächst mit geringer Häufigkeit nachgewiesen wurde, machte am Ende des Untersuchungszeitraumes nahezu die Hälfte der sequenzierten Genome aus (**Abbildung 25A**). Die Varianten „Alpha“, „Beta“ und „Gamma“ (nach Pangolin-Nomenklatur) waren zu diesem Zeitpunkt in Deutschland noch nicht vorhanden. Ein Vergleich der Düsseldorfer Sequenzen mit internationalen SARS-CoV-2-Genomen aus der GISAID-Datenbank zeigte zudem, dass viele der identifizierten Isolate in den globalen Sequenzdatenbanken bislang nicht vertreten waren.

Zunächst wurden SARS-CoV-2-Ausbrüche auf Stationen im Universitätsklinikum Düsseldorf und ihre Verbindungen zur lokalen Bevölkerung untersucht. Bei allen vier nosokomialen Ausbrüchen im Universitätsklinikum Düsseldorf konnte mittels Sequenzierung bestätigt werden, dass es sich um klonale Ausbrüche auf den Stationen handelte (**Abbildung 25B**). Bei den Ausbrüchen auf den Stationen A und C konnte hingegen keine Verbindungen zwischen den untersuchten Ausbrüchen und bereits

sequenzierten Viren aus dem Düsseldorf Raum bzw. der GISAID-Datenbank nachgewiesen werden. Auf Station B gab es ein Isolat (Z279) mit einer genetischen Distanz von 0 zu den Ausbruchsisolaten. Da diese Probe jedoch bereits 6 Wochen vor dem Ausbruch isoliert wurde und es keine Daten aus der Kontaktnachverfolgung zu dem Patienten gegeben hat, konnte keine detailliertere Rückverfolgung stattfinden (**Abbildung 25B&C**).

Der Ausbruch auf Station D wurde Mitte Dezember entdeckt und betraf 16 Patienten und 13 Mitarbeitende. Bei der Analyse der Surveillance-Daten wurden zwei Isolate aus Düsseldorf (Z164 und Z794) gefunden, die genetisch zum Ausbruch auf Station D gehören (**Abbildung 25B&C**). Z164 wurde Anfang Oktober diagnostiziert, war jedoch identisch zu den Ausbruchsisolaten von Station D war.

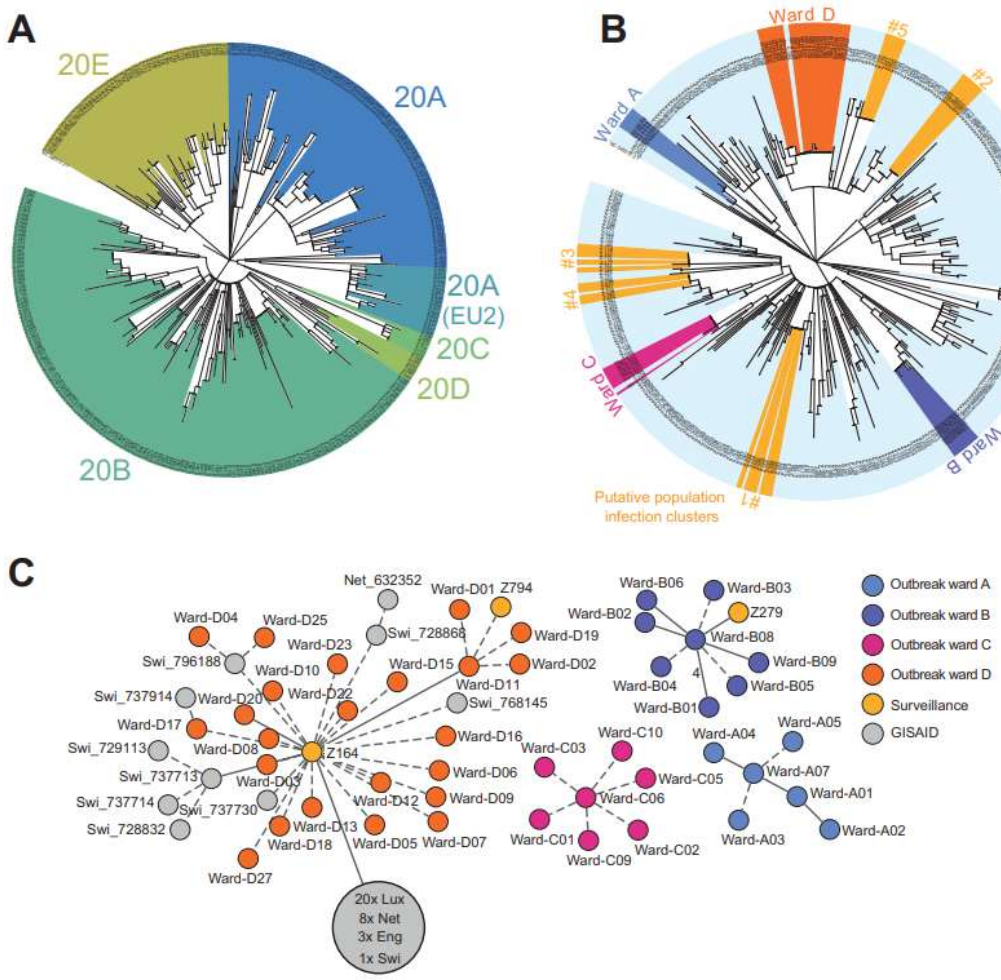


Abbildung 25: Phylogenetische Analyse der Verwandtschaft zwischen den Surveillance-Proben und den hospitalisierten Fällen am Universitätsklinikum Düsseldorf. *Die Farben entsprechen der Varianten-Farbgebung von Nextstrain. (B) Phylogenetischer Baum der Surveillance-Probenaus A und den 44 Proben aus den Krankenhausausbrüchen der Stationen A-D. Potentielle Infektionscluster basierend auf dem Cluster-Algorithmus sind farblich markiert. Beide Bäume wurden mit itol visualisiert. (C) Minimum spanning tree (berechnet mit der Software "networkx version 2.5"; visualisiert mit Cytoscape version 3.8.2 und Inkscape version 0.92). Gezeigt sind alle Krankenhausausbrüche und die identischen oder nahe verwandte Isolate aus der GISAID-Datenbank und den Surveillance-Proben aus Düsseldorf. Gestrichelte Linien entsprechen einem Abstand von 0 Nukleotiden, durchgezogene Linien ohne Nummer entsprechen einem Unterschied von einem Nukleotid. Die Veröffentlichung erfolgt unter CC BY 4.0*

Bei der detaillierten Analyse dieses Falles durch das Gesundheitsamtes stellte sich heraus, dass ein Familienmitglied von Z164 Ende Oktober in einer anderen Station des Uniklinikums behandelt wurde. Da zu diesem Zeitpunkt nur symptomatische Patienten getestet wurden, könnte das eine mögliche Infektionsroute ins Klinikum gewesen sein. Auch für Z794 konnte eine Verbindung ins Klinikum hergestellt werden. Ein naher Verwandter von Z794 wurde auf einer Station behandelt, auf die Patienten von Station D verlegt wurden. Durch intensive Mitarbeiterschulungen, Aufnahmestopps und umfassende Testung aller Patienten konnten die Ausbrüche eingedämmt werden.

Um zu untersuchen, ob mit der ungezielten Sequenzierung SARS-CoV-2-Infektionscluster und Infektionsketten in der Bevölkerung identifiziert werden können, wurde die genetische Verwandtschaft aller sequenzierten Proben mittels eines Cluster-Algorithmus untersucht. Damit konnten 5 Cluster identifiziert werden (**Abbildung 25B**), die einen genetischen Abstand von 0 oder 1 hatten. Von diesen waren 4 dem Gesundheitsamt bereits als Ausbrüche bekannt. Bei Cluster #1 handelte es sich um einen Ausbruch während eines Schulausfluges, bei Cluster #2 um einen Ausbruch in einem Altenheim, sowie der Familie einer Altenpflegerin und einer Schule. Die Cluster #4 und #5 bildeten Infektionsketten innerhalb von größeren Haushalten.

Cluster #3 war dem Gesundheitsamt bisher unbekannt, weshalb alle Personen erneut einer detaillierten telefonischen Nachuntersuchung durch das Gesundheitsamt unterzogen wurden. Dabei stellte sich heraus, dass Person Z132 Kontakt zu einer positiv getesteten Person X hatte, diese in der ersten Befragung durch das Gesundheitsamt aber nicht als Kontakt angegeben hatte. Die Nachbefragung ergab, dass beide Mitglieder in einem Kontaktsport-Verein waren. Der Beginn der Symptome passte dabei sehr gut auf ein gemeinsames Training als potentielle Infektionsquelle. Person X und Z177 wiederum standen in engem persönlichem Kontakt, was diese Infektionen erklärt. Zwei weitere Personen waren Mitglieder desselben Haushalts. Für die anderen Isolate konnte keine Verbindung hergestellt werden.

Zusammenfassend zeigte die Studie, dass bereits bei einer Sequenzierungstiefe von etwa 3 % der Fälle Infektionscluster in der Bevölkerung identifiziert werden können. Die meisten dieser Cluster waren dem Gesundheitsamt bereits im Rahmen der routinemäßigen Kontaktnachverfolgung bekannt. Jedoch konnten durch die Genomische Surveillance auch bislang unbekannte Cluster aufgedeckt werden. Das Potential der Genomischen Surveillance liegt insbesondere darin, etablierte epidemiologische Zusammenhänge zu bestätigen und dort zusätzliche Informationen zu liefert, wo ein Zusammenhang vermutet, aber noch nicht eindeutig belegt ist. In solchen Fällen können die Mitarbeitenden der Kontaktnachverfolgung gezielt nach den entsprechenden Verbindungen suchen. Die routinemäßige Sequenzierung von Krankenhausfällen bietet einen präzisen Überblick über das Infektionsgeschehen im Krankenhaus. Sie ermöglicht auch eine präzise Unterscheidung zwischen tatsächlichen Ausbrüchen und einer zufälligen Häufung gleichzeitig eingelieferter COVID-19-Patienten. Diese Daten schaffen Transparenz und Sicherheit für das Hygienemanagement, die für das Pandemiemanagement von entscheidender Bedeutung sein können.

7.3.3 Echtzeit-Rückverfolgung von SARS-CoV-2-Übertragungsketten in der Düsseldorfer Bevölkerung

Für eine wirksame Eindämmung eines Erregers durch Kontaktnachverfolgung und Quarantäne ist ein umfassendes Verständnis der relevanten Übertragungswege und Infektionsketten entscheidend. In der Praxis sind jedoch weder klassische Kontaktverfolgung noch Genomische Surveillance allein ausreichend, um dieses Ziel zu erreichen. Nachdem wir zeigen konnten, dass die Implementierung einer Genomischen Surveillance auf Ebene einer Großstadt grundsätzlich möglich ist, bestand unser nächstes Anliegen darin, Analysen nicht mehr nur retrospektiv durchzuführen, sondern eine aktive, zeitnahe Überwachung des aktiven Infektionsgeschehens zu ermöglichen. Dafür mussten zwei zentrale Herausforderungen bewältigt werden: Erstens mussten die Ergebnisse der Sequenzierung und deren Interpretation möglichst schnell vorliegen, um eine zeitnahe Entscheidungsfindung zu ermöglichen. Zweitens galt es, eine möglichst umfassende Abdeckung durch Sequenzierung zu erreichen, um relevante Infektionsereignisse nicht zu übersehen. In enger Kooperation mit dem Gesundheitsamt Düsseldorf wurde daher eine SARS-CoV-2 Integrierte Genomische Surveillance (IGS) etabliert, im Rahmen derer möglichst viele SARS-CoV-2-Infektionen in Düsseldorf sequenziert und die Ergebnisse spätestens 48 Stunden nach Vorliegen des positiven Testergebnisses über ein Dashboard bereitgestellt werden sollte.

Titel der Original Arbeit

Integrated genomic surveillance enables tracing of person-to-person SARS-CoV-2 transmission chains during community transmission and reveals extensive onward transmission of travel-imported infections, Germany, June to July 2021

Torsten Houwaart, Samir Belhaj, Emran Tawalbeh, Dirk Nagels, Yara Fröhlich, Patrick Finzer, Pilar Ciruela, Aurora Sabrià, Mercè Herrero, Cristina Andrés, Andrés Antón, Assia Benmoumene, Dounia Asskali, Hussein Haidar, Janina von Dahlen, Jessica Nicolai, Mygg Stiller, Jacqueline Blum, Christian Lange, Carla Adelmann, Britta Schroer, Ute Osmers, Christiane Grice, Phillipp P. Kirfel, Hassan Jomaa, Daniel Strelow, Lisanna Hülse, Moritz Pigulla, Pascal Kreuzer, Alona Tyshaieva, Jonas Weber, Tobias Wienemann, Malte Kohns Vasconcelos, Katrin Hoffmann, Nadine Lübke, Sandra Hauka, Marcel Andree, Claus Jürgen Scholz, Nathalie Jazmati, Klaus Göbels, Rainer Zotz, Klaus Pfeffer, Jörg Timm, Lutz Ehlkes, Andreas Walker, Alexander T. Dilthey, German COVID-19 OMICS Initiative (DeCOI)

Für die Etablierung einer IGS wurden alle SARS-COV-2-positiven Proben aus zwei großen Laboren, die ca. 50 % der SARS-CoV-2 Testung in Düsseldorf durchgeführt haben, täglich von uns gesammelt, amplifiziert und noch am selben Tag mittels Nanopore-Technologie sequenziert. Die Nanopore-Sequenzierung bietet im Vergleich zu anderen Plattformen den Vorteil der kontinuierlichen Datengenerierung, so dass bereits während der Sequenzierung erste Analysen gemacht werden können. Die bioinformatische Arbeitsgruppe von Professor Alexander Dilthey hat dazu einen vollautomatischen Analysealgorithmus geschrieben, der stündlich die Datenmenge und Qualität der Sequenzierung überprüfte und bei ausreichend Daten ein komplettes Virusgenom berechnete. Die genetische Verwandtschaftsverhältnisse zwischen den neuen Virus-Isolaten und bereits existierenden

Isolaten wurden dann sofort in einem Cluster-Algorithmus überprüft und über ein Dashboard angezeigt.

Mitte Juli 2021 wurde in der routinemäßigen Cluster-Analyse zwei neue große Gruppen eng verwandter Virusisolate identifiziert, inklusive mehrerer absolut identischer Varianten aus unterschiedlichen Stadtteilen (**Abbildung 26**). Im Zeitraum zwischen dem 15. Juni und 1. August wurden 518 der 976 (53 %) in Düsseldorf gemeldeten Fälle vollständig sequenziert, wobei 60 Virusisolate zu Cluster 1 und 42 zu Cluster 2 zugeordnet werden konnten. Bemerkenswert war, dass in der initialen Befragung durch das Gesundheitsamt alle frühen Fälle des Cluster 1 angegeben haben, nicht zu wissen, wo sie sich angesteckt hatten. Mit dem Wissen, dass es sich um genetisch identische Isolate handelte, konnte das Gesundheitsamt dann gezielte, strukturierte Fallinterviews durchführen (**Abbildung 27A**).

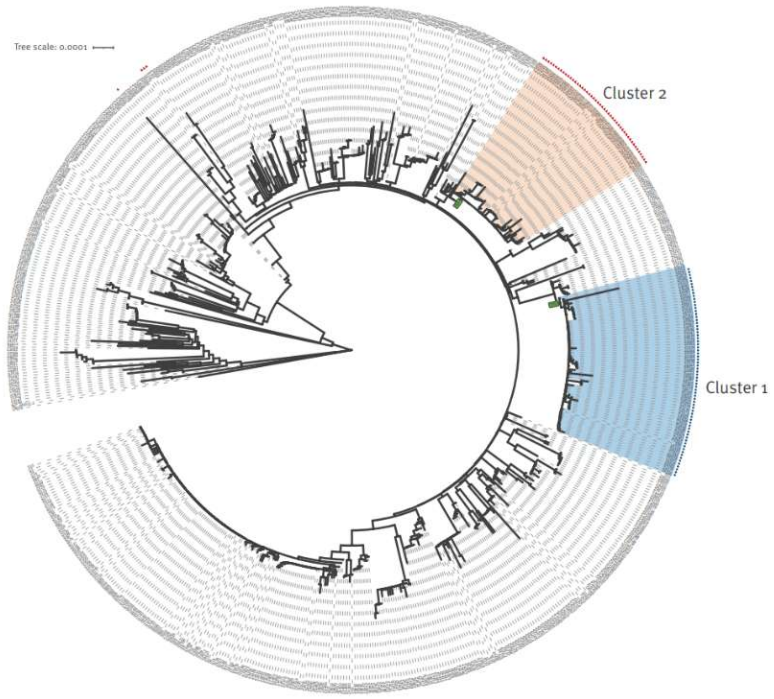


Abbildung 26: Phylogenetischer Baum der SARS-CoV-2 Isolate. *Alle vollständigen sequenzierten SARS-COV-2-Genome vom 15. Juni bis 1. August 2021 wurden mittels MAFFT zusammengefügt und ein phylogenetischer Baum mit dem Tamura-Nei-Distanzmodels und der Neighbour-Joining-Methode berechnet. Der Baum wurde mit iTol coloriert. Cluster 1 (blau) und Cluster 2 (rot) sind farblich markiert. Die Gabelungen I361 und I584, die den Clustern zugrunde liegen, sind mit kleinen Kästchen mit grünem Hintergrund dargestellt. Die Markermutationen T14064C (blaue Punkte; Cluster 1) und C18744T (rote Dreiecke; Cluster 2), die charakteristisch für den jeweiligen Cluster waren, sind außen am Rand des Baumes markiert. Die Veröffentlichung erfolgt unter CC BY 4.0*

Der Ursprung von Cluster 1 konnte dadurch auf mehrere Kneipenbesuche der Indexperson 1 (IP1) zurückgeführt werden (**Abbildung 27B**). IP1 hatte sich wahrscheinlich während eines Urlaubs auf der Insel Mallorca mit SARS-CoV-2 infiziert und dann mindestens acht Personen (KP1-KP8) während zweier Kneipenbesuche am 30. Juni 2021 in der Düsseldorfer Altstadt infiziert.

Weitere Ansteckungen erfolgten dann durch die Kontakte ersten Grades während weiterer Kneipenbesuche in der Altstadt am 2. und 3. Juli (**Abbildung 27B**, rechts). Es wurden aber auch zufällige weitere Personen infiziert, wie KP9 und KP10, die sich am 03. Juli auf einer Kneipentour in der Altstadt befanden. Anschließend breitete sich diese Virusvariante dann in der Düsseldorfer Bevölkerung aus. Die dann einsetzende intensive Kontaktnachverfolgung konnten weitere 15 Fälle identifizieren (P6-P24), die vermutlich über Kontakte des zweiten Grades infiziert wurden (**Abbildung 27B**). Untersuchungen des Düsseldorfer Gesundheitsamtes sowie in sozialen Medien veröffentlichte Videos

zeigten schwere Verstöße gegen die damals geltenden Coronaschutzverordnung. So wurde getanzt, die Maskenpflicht in Innenräumen missachtet und die Kontaktdaten der Gäste nicht erfasst.

Im Falle von Cluster 2 konnte keine Indexperson in Deutschland identifiziert werden. Ein Großteil der Fälle war auf Haushaltskontakte zurückzuführen, mit insgesamt sehr kurzen bereits aufgeklärten Infektionsketten. Auffällig war jedoch, dass 10 von 42 Fällen Reiserückkehrer aus Barcelona/Katalonien und fünf weitere Fälle Reiserückkehrer aus Südfrankreich waren (**Abbildung 28A**).

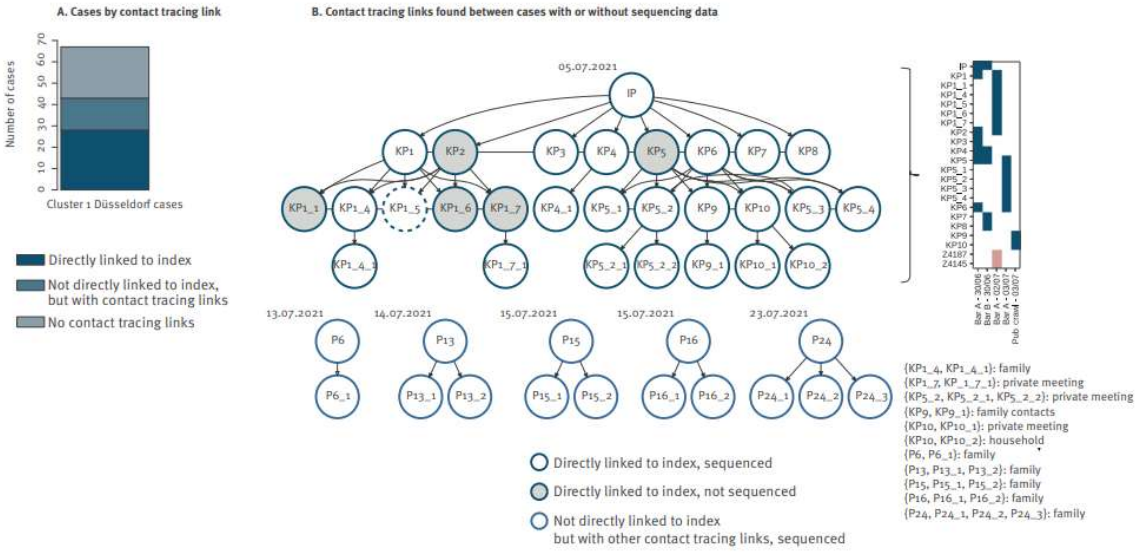


Abbildung 27: Ergebnisse der Kontaktverfolgung von Cluster 1. *(A) Balkendiagramm, das die Anzahl der Fälle in der jeweiligen Kategorie anzeigt. (B) Visualisierung der rekonstruierten epidemiologischen Struktur von Cluster 1. Jeder Kreis im Diagramm der Übertragungskette steht für einen Fall. Grau schattierte Kreise stehen für epidemiologisch verknüpfte, aber nicht sequenzierte Fälle. Nicht schattierte Kreise stehen für epidemiologisch verknüpfte und sequenzierte Fälle. Kreise mit gestrichelter Umrandung sind epidemiologisch verknüpfte aber besitzen einen größeren genetischen Abstand. Die Einfügung rechts neben dem Übertragungsketten-Diagramm zeigt die Besuche der Index-Person und den Kontakten erster und zweiter Ordnung in zwei Bars in der Düsseldorfer Altstadt. Die beiden mit roten Kästchen markierten Fälle besuchten ebenfalls Bar A, sie gehören jedoch aufgrund ihrer typischen Virusvariante zu Cluster 2. Die Veröffentlichung erfolgt unter CC BY 4.0*

In Zusammenarbeit mit dem Gesundheitsamt in Barcelona, konnten Virusisolate identifiziert werden, die sehr eng (genetische Distanz 0 und 1) mit den Düsseldorfer Isolaten verwandt waren und Ende Juni in Barcelona gefunden wurden. Bei Cluster 2 handelte es sich somit wahrscheinlich um mehrere unabhängige Eintragungen derselben Virusvariante, die in Katalonien, Spanien und anderen europäischen Ländern zirkulierte, gefolgt von einer diffusen Übertragung innerhalb der Bevölkerung in Düsseldorf (**Abbildung 28C**). Passend zur diffusen Weiterübertragung dieser Varianten konnte die Kontaktnachverfolgung bei fünf Fällen in Düsseldorf und einer Person in Solingen erneut die Düsseldorfer Altstadt als Infektionsursprung identifizieren. Kurioser Weise waren davon zwei Personen (Z4187 und Z4145) am selben Abend, in derselben Bar, in der auch der Cluster 1-Ausbruch stattfand. Obwohl sie damit epidemiologisch wahrscheinlich zu Cluster 1 gezählt worden wären, zeigen die Sequenzdaten, dass sich definitiv zum unabhängigen Cluster 2 gehören (**Abbildung 28C**).

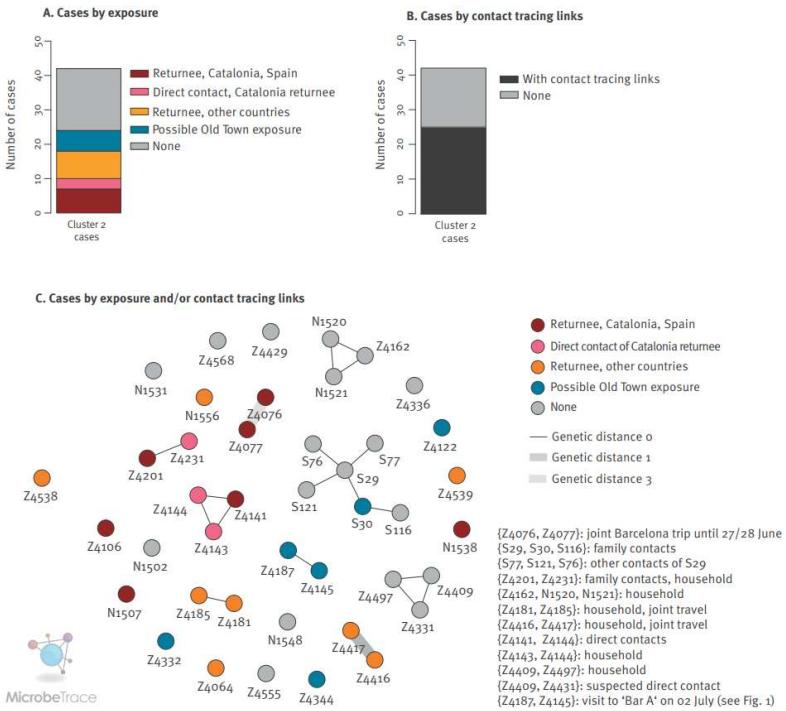


Abbildung 28: Ergebnisse der Kontaktverfolgung von Cluster 2. *(A&B) Balkendiagramm, das die Anzahl der Fälle in der jeweiligen Kategorie anzeigt. (C) Visualisierung der genetischen Distanzen und Kontaktverfolgungsergebnisse für Cluster 2 erstellt mit MicrobeTrace [21]. Jeder Knotenpunkt steht für einen Fall, die Kanten zwischen den Knotenpunkten zeigen die durch die Kontaktverfolgung identifizierten Verbindungen an. Proben-IDs mit „N" und "Z" sind aus Düsseldorf, Proben mit „S" aus Solingen. Die Farben der Kanten geben die genetischen Distanzen zwischen den durch die Kontaktverfolgung verbundenen Proben an. Die genaue Beziehung der Fälle zueinander ist rechts aufgeführt. Die Veröffentlichung erfolgt unter CC BY 4.0*

Gemeinsam betrachtet, waren im Juli 2021 8 % der SARS-CoV-2-Infektionen in Düsseldorf auf Cluster 1 und 4 % auf Cluster 2 zurückzuführen (**Abbildung 29**). In Anbetracht der Tatsache, dass ca. die Hälfte aller gemeldeten Fälle sequenziert wurden ist davon auszugehen, dass diese beiden Infektionsereignisse für rund ein Viertel der SARS-CoV-2-Infektionen in Düsseldorf verantwortlich waren.

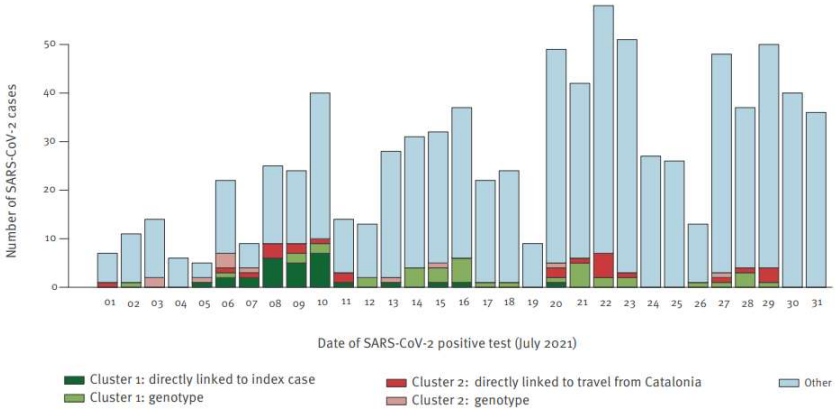


Abbildung 29:Verteilung der SARS-CoV-2 Fälle in Düsseldorf nach Zeit und der genomische Zuordnung. *Die Abbildung zeigt die täglich neu registrierten SARS-CoV-2-Fälle in Düsseldorf im Juli 2021 und wie viele davon mit Cluster 1 oder Cluster 2 in Verbindung standen. Farblich hervorgehoben sind die Fälle, die durch die Kontaktnachverfolgung direkt dem Index-Fall oder Reiserückkehrern aus Katalonien zugeordnet wurden („directly linked"), oder Fälle, die durch die phylogenetischen Analysen („genotype") den Clustern zugeordnet wurden. Die Veröffentlichung erfolgt unter CC BY 4.0*

Zusammenfassend konnten wir in dieser Studie zeigen, dass die Integration von klassischer Kontaktverfolgung und Genomischer Surveillance einen deutlichen Mehrwert besitzt. Während beide identifizierten Cluster in Düsseldorf durch Reiserückkehrer verursacht wurde, zeigten sie doch Unterschiede bezüglich der Transmissionsmuster und der Auswirkungen auf die öffentliche Gesundheit. Cluster 1 wurde vermutlich durch eine Indexperson ausgelöst, deren Verbreitung durch Verstöße gegen die Coronaschutzverordnung stark begünstigt wurde. Im Gegensatz dazu resultierte Cluster 2 aus mehreren unabhängigen Importen einer Virusvariante. Der Hauptübertragungsweg in Düsseldorf bestand dabei aus Haushaltskontakten, aber auch hier kam es zu Übertragungen in der Düsseldorfer Altstadt. Strenge Testanforderungen für Rückkehrreisende oder verpflichtende Antigen-Schnellteste hätten bei beiden Clustern zu einer signifikanten Reduktion der Fälle geführt. Beide Cluster hätten jedoch nicht ohne eine IGS aufgelöst werden können. Viele Fälle konnten erst durch die strukturierten Fallinterviews, die nach der Identifizierung der genetischen Verknüpfung durchgeführt wurden, aufgedeckt werden. Andererseits wären beide Cluster ohne die Genomische Surveillance aufgrund ihrer Gemeinsamkeiten wie Reiserückkehrer und Ansteckungen im Düsseldorfer Nachtleben vermutlich als ein einziger Ausbruch definiert worden.

Somit zeigte diese Studie, dass es möglich ist, SARS-CoV-2 Infektionsketten auch in Zeiten von hohen Inzidenzen mittels einer IGS nachzuverfolgen. Während das hier entwickelte IGS-System auf die Rückverfolgung von SARS-CoV-2 beschränkt war, ist die IGS auch auf andere Krankheitserreger anwendbar.

8 Literaturverzeichnis

1. Walker, A., et al., *Natural prevalence of resistance-associated variants in hepatitis C virus NS5A in genotype 3a-infected people who inject drugs in Germany.* J Clin Virol, 2015. **70**: p. 43-45.
2. Walker, A., et al., *A genotype independent, full-genome reverse-transcription protocol for HCV genotyping and resistance testing.* J Clin Virol, 2017. **91**: p. 42-48.
3. Walker, A., et al., *A pan-genotypic Hepatitis C Virus NS5A amplification method for reliable genotyping and resistance testing.* J Clin Virol, 2019. **113**: p. 8-13.
4. Walker, A., et al., *Detection of a genetic footprint of the sofosbuvir resistance-associated substitution S282T after HCV treatment failure.* Virol J, 2017. **14**(1): p. 106.
5. Walker, A., et al., *Distinct Escape Pathway by Hepatitis C Virus Genotype 1a from a Dominant CD8+ T Cell Response by Selection of Altered Epitope Processing.* J Virol, 2016. **90**(1): p. 33-42.
6. Walker, A., et al., *Immune escape pathways from the HBV core(18-27) CD8 T cell response are driven by individual HLA class I alleles.* Front Immunol, 2022. **13**: p. 1045498.
7. Schwarz, T., et al., *HBV shows different levels of adaptation to HLA class I-associated selection pressure correlating with markers of replication.* J Hepatol, 2025. **82**(5): p. 805-815.
8. Walker, A., et al., *Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020.* Euro Surveill, 2020. **25**(22).
9. Walker, A., et al., *Characterization of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection Clusters Based on Integrated Genomic Surveillance, Outbreak Analysis and Contact Tracing in an Urban Setting.* Clin Infect Dis, 2022. **74**(6): p. 1039-1046.
10. Houwaart, T., et al., *Integrated genomic surveillance enables tracing of person-to-person SARS-CoV-2 transmission chains during community transmission and reveals extensive onward transmission of travel-imported infections, Germany, June to July 2021.* Euro Surveill, 2022. **27**(43).
11. Drake, J.W., *The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes.* Ann N Y Acad Sci, 1999. **870**: p. 100-7.
12. Duffy, S., *Why are RNA virus mutation rates so damn high?* PLoS Biol, 2018. **16**(8): p. e3000003.
13. Smith, E.C. and M.R. Denison, *Coronaviruses as DNA wannabes: a new model for the regulation of RNA virus replication fidelity.* PLoS Pathog, 2013. **9**(12): p. e1003760.
14. Eigen, M., *Selforganization of matter and the evolution of biological macromolecules.* Naturwissenschaften, 1971. **58**(10): p. 465-523.
15. Eigen, M. and P. Schuster, *The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle.* Naturwissenschaften, 1977. **64**(11): p. 541-65.
16. Domingo, E. and C. Perales, *Viral quasispecies.* PLoS Genet, 2019. **15**(10): p. e1008271.
17. WHO, *Global hepatitis report 2024: action for access in low- and middle-income countries*. 2024, Geneva.
18. Feinstone, S.M., *History of the Discovery of Hepatitis A Virus.* Cold Spring Harb Perspect Med, 2019. **9**(5).
19. Blumberg, B.S., H.J. Alter, and S. Visnich, *A "New" Antigen in Leukemia Sera.* JAMA, 1965. **191**: p. 541-6.

20. Bayer, M.E., B.S. Blumberg, and B. Werner, *Particles associated with Australia antigen in the sera of patients with leukaemia, Down's Syndrome and hepatitis.* Nature, 1968. **218**(5146): p. 1057-9.
21. Blumberg, B.S., A.I. Sutnick, and W.T. London, *Hepatitis and leukemia: their relation to Australia antigen.* Bull N Y Acad Med, 1968. **44**(12): p. 1566-86.
22. Levene, C. and B.S. Blumberg, *Additional specificities of Australia antigen and the possible identification of hepatitis carriers.* Nature, 1969. **221**(5176): p. 195-6.
23. London, W.T., A.I. Sutnick, and B.S. Blumberg, *Australia antigen and acute viral hepatitis.* Ann Intern Med, 1969. **70**(1): p. 55-9.
24. Dane, D.S., C.H. Cameron, and M. Briggs, *Virus-like particles in serum of patients with Australia-antigen-associated hepatitis.* Lancet, 1970. **1**(7649): p. 695-8.
25. Warren, K.S., et al., *A new group of hepadnaviruses naturally infecting orangutans (Pongo pygmaeus).* J Virol, 1999. **73**(9): p. 7860-5.
26. Vaudin, M., et al., *The complete nucleotide sequence of the genome of a hepatitis B virus isolated from a naturally infected chimpanzee.* J Gen Virol, 1988. **69 (Pt 6)**: p. 1383-9.
27. Lanford, R.E., et al., *Isolation of a hepadnavirus from the woolly monkey, a New World primate.* Proc Natl Acad Sci U S A, 1998. **95**(10): p. 5757-61.
28. Summers, J., J.M. Smolec, and R. Snyder, *A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks.* Proc Natl Acad Sci U S A, 1978. **75**(9): p. 4533-7.
29. Marion, P.L., et al., *A virus in Beechey ground squirrels that is related to hepatitis B virus of humans.* Proc Natl Acad Sci U S A, 1980. **77**(5): p. 2941-5.
30. Drexler, J.F., et al., *Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes.* Proc Natl Acad Sci U S A, 2013. **110**(40): p. 16151-6.
31. Dallmeier, K., U. Schultz, and M. Nassal, *Heterologous replacement of the supposed host determining region of avihepadnaviruses: high in vivo infectivity despite low infectivity for hepatocytes.* PLoS Pathog, 2008. **4**(12): p. e1000230.
32. Lauber, C., et al., *Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses.* Cell Host Microbe, 2017. **22**(3): p. 387-399 e6.
33. Hahn, C.M., et al., *Characterization of a Novel Hepadnavirus in the White Sucker (Catostomus commersonii) from the Great Lakes Region of the United States.* J Virol, 2015. **89**(23): p. 11801-11.
34. Beck, J. and M. Nassal, *Hepatitis B virus replication.* World J Gastroenterol, 2007. **13**(1): p. 48-64.
35. Magnius, L.O. and A. Espmark, *A new antigen complex co-occurring with Australia antigen.* Acta Pathol Microbiol Scand B Microbiol Immunol, 1972. **80**(2): p. 335-7.
36. Milich, D.R., *Is the function of the HBeAg really unknown?* Hum Vaccin Immunother, 2019. **15**(9): p. 2187-2191.
37. Gerlich, W.H., *Medical virology of hepatitis B: how it began and where we are now.* Virol J, 2013. **10**: p. 239.
38. Decorsiere, A., et al., *Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor.* Nature, 2016. **531**(7594): p. 386-9.
39. Seitz, S., et al., *Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions.* EMBO J, 2007. **26**(18): p. 4160-7.

40. Yan, H., et al., *Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus.* Elife, 2012. **1**: p. e00049.
41. Jiang, B. and E. Hildt, *Intracellular Trafficking of HBV Particles.* Cells, 2020. **9**(9).
42. Nassal, M., *HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B.* Gut, 2015. **64**(12): p. 1972-84.
43. Nassal, M., *Hepatitis B viruses: reverse transcription a different way.* Virus Res, 2008. **134**(1-2): p. 235-49.
44. Bruss, V., et al., *Post-translational alterations in transmembrane topology of the hepatitis B virus large envelope protein.* EMBO J, 1994. **13**(10): p. 2273-9.
45. Milich, D. and T.J. Liang, *Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection.* Hepatology, 2003. **38**(5): p. 1075-86.
46. Velkov, S., et al., *The Global Hepatitis B Virus Genotype Distribution Approximated from Available Genotyping Data.* Genes (Basel), 2018. **9**(10).
47. Tran, T.T., T.N. Trinh, and K. Abe, *New complex recombinant genotype of hepatitis B virus identified in Vietnam.* J Virol, 2008. **82**(11): p. 5657-63.
48. Kurbanov, F., et al., *When should "I" consider a new hepatitis B virus genotype?* J Virol, 2008. **82**(16): p. 8241-2.
49. Arankalle, V.A., et al., *A novel HBV recombinant (genotype I) similar to Vietnam/Laos in a primitive tribe in eastern India.* J Viral Hepat, 2010. **17**(7): p. 501-10.
50. Yu, H., et al., *Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype "I".* PLoS One, 2010. **5**(2): p. e9297.
51. Tatematsu, K., et al., *A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J.* J Virol, 2009. **83**(20): p. 10538-47.
52. Sun, B., et al., *Origin and dispersal history of Hepatitis B virus in Eastern Eurasia.* Nat Commun, 2024. **15**(1): p. 2951.
53. Wieland, S.F. and F.V. Chisari, *Stealth and cunning: hepatitis B and hepatitis C viruses.* J Virol, 2005. **79**(15): p. 9369-80.
54. Chiale, C., A.M. Marchese, and M.D. Robek, *Innate immunity and HBV persistence.* Curr Opin Virol, 2021. **49**: p. 13-20.
55. Cheng, X., et al., *Hepatitis B virus evades innate immunity of hepatocytes but activates cytokine production by macrophages.* Hepatology, 2017. **66**(6): p. 1779-1793.
56. Mutz, P., et al., *HBV Bypasses the Innate Immune Response and Does Not Protect HCV From Antiviral Activity of Interferon.* Gastroenterology, 2018. **154**(6): p. 1791-1804 e22.
57. Suslov, A., et al., *Hepatitis B Virus Does Not Interfere With Innate Immune Responses in the Human Liver.* Gastroenterology, 2018. **154**(6): p. 1778-1790.
58. Bellecave, P., et al., *Hepatitis B and C virus coinfection: a novel model system reveals the absence of direct viral interference.* Hepatology, 2009. **50**(1): p. 46-55.
59. Guidotti, L.G. and F.V. Chisari, *Cytokine-induced viral purging--role in viral pathogenesis.* Curr Opin Microbiol, 1999. **2**(4): p. 388-91.
60. Wieland, S.F., et al., *Interferon prevents formation of replication-competent hepatitis B virus RNA-containing nucleocapsids.* Proc Natl Acad Sci U S A, 2005. **102**(28): p. 9913-7.
61. Chu, C.M., C.T. Yeh, and Y.F. Liaw, *Low-level viremia and intracellular expression of hepatitis B surface antigen (HBsAg) in HBsAg carriers with concurrent hepatitis C virus infection.* J Clin Microbiol, 1998. **36**(7): p. 2084-6.

62. Collins, J.M., et al., *Hepatitis B Virus Reactivation During Successful Treatment of Hepatitis C Virus With Sofosbuvir and Simeprevir.* Clin Infect Dis, 2015. **61**(8): p. 1304-6.
63. Thimme, R., et al., *CD8(+) T cells mediate viral clearance and disease pathogenesis during acute hepatitis B virus infection*, in *J Virol*. 2003. p. 68-76.
64. Protzer, U., M.K. Maini, and P.A. Knolle, *Living in the liver: hepatic infections.* Nat Rev Immunol, 2012. **12**(3): p. 201-13.
65. Maini, M.K., et al., *Direct ex vivo analysis of hepatitis B virus-specific CD8(+) T cells associated with the control of infection.* Gastroenterology, 1999. **117**(6): p. 1386-96.
66. Aoki, J., et al., *Kinetics of peripheral hepatitis B virus-specific CD8+ T cells in patients with onset of viral reactivation.* J Gastroenterol, 2013. **48**(6): p. 728-37.
67. Rehermann, B., et al., *The hepatitis B virus persists for decades after patients' recovery from acute viral hepatitis despite active maintenance of a cytotoxic T-lymphocyte response.* Nat Med, 1996. **2**(10): p. 1104-8.
68. Rehermann, B., et al., *The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis.* J Exp Med, 1995. **181**(3): p. 1047-58.
69. Penna, A., et al., *Cytotoxic T lymphocytes recognize an HLA-A2-restricted epitope within the hepatitis B virus nucleocapsid antigen.* J Exp Med, 1991. **174**(6): p. 1565-70.
70. Hoogeveen, R.C., et al., *Phenotype and function of HBV-specific T cells is determined by the targeted epitope in addition to the stage of infection.* Gut, 2019. **68**(5): p. 893-904.
71. Schuch, A., et al., *Phenotypic and functional differences of HBV core-specific versus HBV polymerase-specific CD8+ T cells in chronically HBV-infected patients with low viral load.* Gut, 2019. **68**(5): p. 905-915.
72. Chen, M.T., et al., *A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14913-8.
73. Liang, T.J., et al., *A hepatitis B virus mutant associated with an epidemic of fulminant hepatitis.* N Engl J Med, 1991. **324**(24): p. 1705-9.
74. Heim, K., et al., *Heterogeneity of HBV-Specific CD8(+) T-Cell Failure: Implications for Immunotherapy.* Front Immunol, 2019. **10**: p. 2240.
75. Boni, C., et al., *Characterization of hepatitis B virus (HBV)-specific T-cell dysfunction in chronic HBV infection.* J Virol, 2007. **81**(8): p. 4215-25.
76. Fisicaro, P., et al., *Antiviral intrahepatic T-cell responses can be restored by blocking programmed death-1 pathway in chronic hepatitis B.* Gastroenterology, 2010. **138**(2): p. 682-93, 693 e1-4.
77. Nebbia, G., et al., *Upregulation of the Tim-3/galectin-9 pathway of T cell exhaustion in chronic hepatitis B virus infection.* PLoS One, 2012. **7**(10): p. e47648.
78. Raziorrouh, B., et al., *The immunoregulatory role of CD244 in chronic hepatitis B infection and its inhibitory potential on virus-specific CD8+ T-cell function.* Hepatology, 2010. **52**(6): p. 1934-47.
79. Schurich, A., et al., *Role of the coinhibitory receptor cytotoxic T lymphocyte antigen-4 on apoptosis-Prone CD8 T cells in persistent hepatitis B virus infection.* Hepatology, 2011. **53**(5): p. 1494-503.
80. Wu, W., et al., *Blockade of Tim-3 signaling restores the virus-specific CD8(+) T-cell response in patients with chronic hepatitis B.* Eur J Immunol, 2012. **42**(5): p. 1180-91.

81. Lopes, A.R., et al., *Bim-mediated deletion of antigen-specific CD8 T cells in patients unable to control HBV infection.* J Clin Invest, 2008. **118**(5): p. 1835-45.
82. Schurich, A., et al., *Distinct Metabolic Requirements of Exhausted and Functional Virus-Specific CD8 T Cells in the Same Host.* Cell Rep, 2016. **16**(5): p. 1243-1252.
83. Bosch, M., et al., *A liver immune rheostat regulates CD8 T cell immunity in chronic HBV infection.* Nature, 2024. **631**(8022): p. 867-875.
84. Bengsch, B., B. Martin, and R. Thimme, *Restoration of HBV-specific CD8+ T cell function by PD-1 blockade in inactive carrier patients is linked to T cell differentiation.* J Hepatol, 2014. **61**(6): p. 1212-9.
85. Fisicaro, P., et al., *Targeting mitochondrial dysfunction can restore antiviral activity of exhausted HBV-specific CD8 T cells in chronic hepatitis B.* Nat Med, 2017. **23**(3): p. 327-336.
86. Kefalakes, H., et al., *Adaptation of the hepatitis B virus core protein to CD8(+) T-cell selection pressure.* Hepatology, 2015. **62**(1): p. 47-56.
87. Polaris Observatory, H.C.V.C., *Global change in hepatitis C virus prevalence and cascade of care between 2015 and 2020: a modelling study.* Lancet Gastroenterol Hepatol, 2022. **7**(5): p. 396-415.
88. Frank, C., et al., *The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt.* Lancet, 2000. **355**(9207): p. 887-91.
89. Elbaz, T., et al., *Leading Role of Sofosbuvir/Daclatasvir in Achieving Hepatitis C Elimination in Egypt.* J Viral Hepat, 2025. **32**(7): p. e70032.
90. Gomaa, A., et al., *Hepatitis C Elimination in Egypt: Story of Success.* Pathogens, 2024. **13**(8).
91. WHO. *Egypt becomes the first country to achieve WHO validation on the path to elimination of hepatitis C*. 2023 [cited 2025 24.11.2025]; Available from: https://www.emro.who.int/media/news/egypt-becomes-the-first-country-to-achieve-who-validation-on-the-path-to-elimination-of-hepatitis-c.html.
92. Alter, H.J., et al., *Transmissible agent in non-A, non-B hepatitis.* Lancet, 1978. **1**(8062): p. 459-63.
93. Choo, Q.L., et al., *Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome.* Science, 1989. **244**(4902): p. 359-62.
94. Kolykhalov, A.A., et al., *Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA.* Science, 1997. **277**(5325): p. 570-4.
95. Tabata, K., C.J. Neufeldt, and R. Bartenschlager, *Hepatitis C Virus Replication.* Cold Spring Harb Perspect Med, 2020. **10**(3).
96. Schult, P., et al., *microRNA-122 amplifies hepatitis C virus translation by shaping the structure of the internal ribosomal entry site.* Nat Commun, 2018. **9**(1): p. 2613.
97. Lohmann, V., et al., *Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line.* Science, 1999. **285**(5424): p. 110-3.
98. Lohmann, V., J.O. Koch, and R. Bartenschlager, *Processing pathways of the hepatitis C virus proteins.* J Hepatol, 1996. **24**(2 Suppl): p. 11-9.
99. Meylan, E., et al., *Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus.* Nature, 2005. **437**(7062): p. 1167-72.
100. Romero-Brey, I., et al., *Three-dimensional architecture and biogenesis of membrane structures associated with hepatitis C virus replication.* PLoS Pathog, 2012. **8**(12): p. e1003056.

101. Cosset, F.L., et al., *HCV Interplay with Lipoproteins: Inside or Outside the Cells?* Viruses, 2020. **12**(4).
102. Borgia, S.M., et al., *Identification of a Novel Hepatitis C Virus Genotype From Punjab, India: Expanding Classification of Hepatitis C Virus Into 8 Genotypes.* J Infect Dis, 2018. **218**(11): p. 1722-1729.
103. Murphy, D.G., et al., *Hepatitis C virus genotype 7, a new genotype originating from central Africa.* J Clin Microbiol, 2015. **53**(3): p. 967-72.
104. Raghwani, J., et al., *Origin and evolution of the unique hepatitis C virus circulating recombinant form 2k/1b.* J Virol, 2012. **86**(4): p. 2212-20.
105. Hostager, R., et al., *Hepatitis C virus genotype 1 and 2 recombinant genomes and the phylogeographic history of the 2k/1b lineage.* Virus Evol, 2019. **5**(2): p. vez041.
106. Adams, G., et al., *Natural recovery from acute hepatitis C virus infection by agammaglobulinemic twin children.* Pediatr Infect Dis J, 1997. **16**(5): p. 533-4.
107. Pestka, J.M., et al., *Rapid induction of virus-neutralizing antibodies and viral clearance in a single-source outbreak of hepatitis C.* Proc Natl Acad Sci U S A, 2007. **104**(14): p. 6025-30.
108. Law, M., *Antibody Responses in Hepatitis C Infection.* Cold Spring Harb Perspect Med, 2021. **11**(3).
109. Lange, M., et al., *Hepatitis C virus hypervariable region 1 variants presented on hepatitis B virus capsid-like particles induce cross-neutralizing antibodies.* PLoS One, 2014. **9**(7): p. e102235.
110. Giang, E., et al., *Human broadly neutralizing antibodies to the envelope glycoprotein complex of hepatitis C virus.* Proc Natl Acad Sci U S A, 2012. **109**(16): p. 6205-10.
111. Law, M., et al., *Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge.* Nat Med, 2008. **14**(1): p. 25-7.
112. Morin, T.J., et al., *Human monoclonal antibody HCV1 effectively prevents and treats HCV infection in chimpanzees.* PLoS Pathog, 2012. **8**(8): p. e1002895.
113. Vanwolleghem, T., et al., *Polyclonal immunoglobulins from a chronic hepatitis C virus patient protect human liver-chimeric mice from infection with a homologous hepatitis C virus strain.* Hepatology, 2008. **47**(6): p. 1846-55.
114. Kemming, J., R. Thimme, and C. Neumann-Haefelin, *Adaptive Immune Response against Hepatitis C Virus.* Int J Mol Sci, 2020. **21**(16).
115. Lechner, F., et al., *Analysis of successful immune responses in persons infected with hepatitis C virus.* J Exp Med, 2000. **191**(9): p. 1499-512.
116. Thimme, R., et al., *Viral and immunological determinants of hepatitis C virus clearance, persistence, and disease.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15661-8.
117. Thimme, R., et al., *Determinants of viral clearance and persistence during acute hepatitis C virus infection.* J Exp Med, 2001. **194**(10): p. 1395-406.
118. Grakoui, A., et al., *HCV persistence and immune evasion in the absence of memory T cell help.* Science, 2003. **302**(5645): p. 659-62.
119. Shoukry, N.H., et al., *Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection.* J Exp Med, 2003. **197**(12): p. 1645-55.
120. Cooper, S., et al., *Analysis of a successful immune response against hepatitis C virus.* Immunity, 1999. **10**(4): p. 439-49.
121. Erickson, A.L., et al., *The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes.* Immunity, 2001. **15**(6): p. 883-95.

122. Cox, A.L., et al., *Cellular immune selection with hepatitis C virus persistence in humans.* J Exp Med, 2005. **201**(11): p. 1741-52.
123. Tester, I., et al., *Immune evasion versus recovery after acute hepatitis C virus infection from a shared source.* J Exp Med, 2005. **201**(11): p. 1725-31.
124. Timm, J., et al., *CD8 epitope escape and reversion in acute HCV infection.* J Exp Med, 2004. **200**(12): p. 1593-604.
125. Callendret, B., et al., *Transmission of clonal hepatitis C virus genomes reveals the dominant but transitory role of CD8(+) T cells in early viral evolution.* J Virol, 2011. **85**(22): p. 11833-45.
126. Schulze Zur Wiesch, J., et al., *Broadly directed virus-specific CD4+ T cell responses are primed during acute hepatitis C infection, but rapidly disappear from human blood with viral persistence.* J Exp Med, 2012. **209**(1): p. 61-75.
127. Timm, J. and C.M. Walker, *Mutational escape of CD8+ T cell epitopes: implications for prevention and therapy of persistent hepatitis virus infections.* Med Microbiol Immunol, 2015. **204**(1): p. 29-38.
128. Meyer-Olson, D., et al., *Limited T cell receptor diversity of HCV-specific T cell responses is associated with CTL escape.* J Exp Med, 2004. **200**(3): p. 307-19.
129. Sijts, E.J. and P.M. Kloetzel, *The role of the proteasome in the generation of MHC class I ligands and immune responses.* Cell Mol Life Sci, 2011. **68**(9): p. 1491-502.
130. Kimura, Y., et al., *Escape mutations alter proteasome processing of major histocompatibility complex class I-restricted epitopes in persistent hepatitis C virus infection.* J Virol, 2005. **79**(8): p. 4870-6.
131. Seifert, U., et al., *Hepatitis C virus mutation affects proteasomal epitope processing.* J Clin Invest, 2004. **114**(2): p. 250-9.
132. Bengsch, B., et al., *Coexpression of PD-1, 2B4, CD160 and KLRG1 on exhausted HCV-specific CD8+ T cells is linked to antigen recognition and T cell differentiation.* PLoS Pathog, 2010. **6**(6): p. e1000947.
133. Golden-Mason, L., et al., *Negative immune regulator Tim-3 is overexpressed on T cells in hepatitis C virus infection and its blockade rescues dysfunctional CD4+ and CD8+ T cells.* J Virol, 2009. **83**(18): p. 9122-30.
134. Nakamoto, N., et al., *Synergistic reversal of intrahepatic HCV-specific CD8 T cell exhaustion by combined PD-1/CTLA-4 blockade.* PLoS Pathog, 2009. **5**(2): p. e1000313.
135. Radziewicz, H., et al., *Liver-infiltrating lymphocytes in chronic human hepatitis C virus infection display an exhausted phenotype with high levels of PD-1 and low levels of CD127 expression.* J Virol, 2007. **81**(6): p. 2545-53.
136. Schlaphoff, V., et al., *Dual function of the NK cell receptor 2B4 (CD244) in the regulation of HCV-specific CD8+ T cells.* PLoS Pathog, 2011. **7**(5): p. e1002045.
137. Urbani, S., et al., *PD-1 expression in acute hepatitis C virus (HCV) infection is associated with HCV-specific CD8 exhaustion.* J Virol, 2006. **80**(22): p. 11398-403.
138. Kasprowicz, V., et al., *Hepatitis C virus (HCV) sequence variation induces an HCV-specific T-cell phenotype analogous to spontaneous resolution.* J Virol, 2010. **84**(3): p. 1656-63.
139. Rutebemberwa, A., et al., *High-programmed death-1 levels on hepatitis C virus-specific T cells during acute infection are associated with viral persistence and require preservation of cognate antigen during chronic infection.* J Immunol, 2008. **181**(12): p. 8215-25.

140. EASL, *EASL recommendations on treatment of hepatitis C: Final update of the series(☆).* J Hepatol, 2020. **73**(5): p. 1170-1218.
141. Papatheodoridis, G.V., et al., *Hepatitis C: The beginning of the end-key elements for successful European and national strategies to eliminate HCV in Europe.* J Viral Hepat, 2018. **25 Suppl 1**: p. 6-17.
142. Pan, Q., et al., *Telaprevir/boceprevir era: from bench to bed and back.* World J Gastroenterol, 2012. **18**(43): p. 6183-8.
143. Pawlotsky, J.M., et al., *From non-A, non-B hepatitis to hepatitis C virus cure.* J Hepatol, 2015. **62**(1 Suppl): p. S87-99.
144. Buhler, S. and R. Bartenschlager, *[Molecular mechanisms of hepatitis C virus (HCV) replication - implications for the development of antiviral drugs].* Z Gastroenterol, 2011. **49**(7): p. 836-44.
145. Bartenschlager, R., V. Lohmann, and F. Penin, *The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection.* Nat Rev Microbiol, 2013. **11**(7): p. 482-96.
146. Sarrazin, C., et al., *Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir.* Gastroenterology, 2007. **132**(5): p. 1767-77.
147. Susser, S., et al., *Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients.* Hepatology, 2009. **50**(6): p. 1709-18.
148. Kieffer, T.L., et al., *Hepatitis C viral evolution in genotype 1 treatment-naive and treatment-experienced patients receiving telaprevir-based therapy in clinical trials.* PLoS One, 2012. **7**(4): p. e34372.
149. Donaldson, E.F., et al., *Clinical evidence and bioinformatics characterization of potential hepatitis C virus resistance pathways for sofosbuvir.* Hepatology, 2015. **61**(1): p. 56-65.
150. Hedskog, C., et al., *Evolution of the HCV viral population from a patient with S282T detected at relapse after sofosbuvir monotherapy.* J Viral Hepat, 2015. **22**(11): p. 871-81.
151. Svarovskaia, E.S., et al., *Infrequent development of resistance in genotype 1-6 hepatitis C virus-infected subjects treated with sofosbuvir in phase 2 and 3 clinical trials.* Clin Infect Dis, 2014. **59**(12): p. 1666-74.
152. Kalaghatgi, P., et al., *Geno2pheno[HCV] - A Web-based Interpretation System to Support Hepatitis C Treatment Decisions in the Era of Direct-Acting Antiviral Agents.* PLoS One, 2016. **11**(5): p. e0155869.
153. Coronaviridae Study Group of the International Committee on Taxonomy of, V., *The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2.* Nat Microbiol, 2020. **5**(4): p. 536-544.
154. WHO, *Director-General's opening remarks at the media briefing on COVID19 -March 2020.* 2020.
155. Kim, M.C., et al., *Duration of Culturable SARS-CoV-2 in Hospitalized Patients with Covid-19.* N Engl J Med, 2021. **384**(7): p. 671-673.
156. Aydillo, T., et al., *Shedding of Viable SARS-CoV-2 after Immunosuppressive Therapy for Cancer.* N Engl J Med, 2020. **383**(26): p. 2586-2588.
157. Koff, A.G., et al., *Prolonged incubation of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in a patient on rituximab therapy.* Infect Control Hosp Epidemiol, 2021. **42**(10): p. 1286-1288.

158. Keitel, V., et al., *Case Report: Convalescent Plasma Achieves SARS-CoV-2 Viral Clearance in a Patient With Persistently High Viral Replication Over 8 Weeks Due to Severe Combined Immunodeficiency (SCID) and Graft Failure.* Front Immunol, 2021. **12**: p. 645989.
159. STAKOB, *Hinweise zu Erkennung, Diagnostik und Therapie von Patienten mit COVID-19.* 2022.
160. RKI, *Epidemiologischer Steckbrief zu SARS-CoV-2 und COVID-19.* 2021.
161. Cui, J., F. Li, and Z.L. Shi, *Origin and evolution of pathogenic coronaviruses.* Nat Rev Microbiol, 2019. **17**(3): p. 181-192.
162. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* Nature, 2020. **579**(7798): p. 270-273.
163. Hoffmann, M., et al., *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p. 271-280 e8.
164. V'Kovski, P., et al., *Coronavirus biology and replication: implications for SARS-CoV-2.* Nat Rev Microbiol, 2021. **19**(3): p. 155-170.
165. Robson, F., et al., *Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting.* Mol Cell, 2020. **79**(5): p. 710-727.
166. Pather, S., et al., *SARS-CoV-2 Omicron variants: burden of disease, impact on vaccine effectiveness and need for variant-adapted vaccines.* Front Immunol, 2023. **14**: p. 1130539.
167. Wilhelm, A., et al., *Limited neutralisation of the SARS-CoV-2 Omicron subvariants BA.1 and BA.2 by convalescent and vaccine serum and monoclonal antibodies.* EBioMedicine, 2022. **82**: p. 104158.
168. Dadonaite, B., et al., *Spike deep mutational scanning helps predict success of SARS-CoV-2 clades.* Nature, 2024. **631**(8021): p. 617-626.
169. Dickerman, B.A., et al., *Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines in U.S. Veterans.* N Engl J Med, 2022. **386**(2): p. 105-115.
170. El Sahly, H.M., et al., *Efficacy of the mRNA-1273 SARS-CoV-2 Vaccine at Completion of Blinded Phase.* N Engl J Med, 2021. **385**(19): p. 1774-1785.
171. Jin, L., et al., *CoronaVac: A review of efficacy, safety, and immunogenicity of the inactivated vaccine against SARS-CoV-2.* Hum Vaccin Immunother, 2022. **18**(6): p. 2096970.
172. Polack, F.P., et al., *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.* N Engl J Med, 2020. **383**(27): p. 2603-2615.
173. Voysey, M., et al., *Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK.* Lancet, 2021. **397**(10269): p. 99-111.
174. Kim, J.M., et al., *Effectiveness of Paxlovid, an Oral Antiviral Drug, Against the Omicron BA.5 Variant in Korea: Severe Progression and Death Between July and November 2022.* J Korean Med Sci, 2023. **38**(27): p. e211.
175. Gilead-Sciences. *Gilead Statement on Veklury® (Remdesivir) and the SARS-CoV-2 Omicron Variant*. 2021 [cited 2025; Available from: https://www.gilead.com/company/company-statements/2021/gilead-statement-on-veklury-remdesivir-and-the-sars-cov-2-omicron-variant.
176. Chen, Z., et al., *Potent monoclonal antibodies neutralize Omicron sublineages and other SARS-CoV-2 variants.* Cell Rep, 2022. **41**(5): p. 111528.

177. Kang, J., et al., *South Korea's responses to stop the COVID-19 pandemic.* Am J Infect Control, 2020. **48**(9): p. 1080-1086.
178. Pung, R., et al., *Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures.* Lancet, 2020. **395**(10229): p. 1039-1046.
179. Miller, J.S., et al., *COVID-19 Case Investigation and Contact Tracing in Central Washington State, June-July 2020.* J Community Health, 2021. **46**(5): p. 918-921.
180. Bohmer, M.M., et al., *Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series.* Lancet Infect Dis, 2020. **20**(8): p. 920-928.
181. Wikipedia. *Covid-19-Apps*. 2025; Available from: https://de.wikipedia.org/wiki/COVID-19-App.
182. Kretzschmar, M.E., et al., *Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study.* Lancet Public Health, 2020. **5**(8): p. e452-e459.
183. Hadfield, J., et al., *Nextstrain: real-time tracking of pathogen evolution.* Bioinformatics, 2018. **34**(23): p. 4121-4123.
184. consortiumcontact@cogconsortium.uk, C.-G.U., *An integrated national scale SARS-CoV-2 genomic surveillance network.* Lancet Microbe, 2020. **1**(3): p. e99-e100.
185. Douglas, J., et al., *Real-Time Genomics for Tracking Severe Acute Respiratory Syndrome Coronavirus 2 Border Incursions after Virus Elimination, New Zealand.* Emerg Infect Dis, 2021. **27**(9): p. 2361-2368.
186. Hjorleifsson, K.E., et al., *Reconstruction of a large-scale outbreak of SARS-CoV-2 infection in Iceland informs vaccination strategies.* Clin Microbiol Infect, 2022. **28**(6): p. 852-858.
187. Smith, M.R., et al., *Rapid incidence estimation from SARS-CoV-2 genomes reveals decreased case detection in Europe during summer 2020.* Nat Commun, 2021. **12**(1): p. 6009.
188. Pawlotsky, J.M., *Hepatitis C Virus Resistance to Direct-Acting Antiviral Drugs in Interferon-Free Regimens.* Gastroenterology, 2016. **151**(1): p. 70-86.
189. Afdhal, N., et al., *Ledipasvir and sofosbuvir for previously treated HCV genotype 1 infection.* N Engl J Med, 2014. **370**(16): p. 1483-93.
190. Nelson, D.R., et al., *All-oral 12-week treatment with daclatasvir plus sofosbuvir in patients with hepatitis C virus genotype 3 infection: ALLY-3 phase III study.* Hepatology, 2015. **61**(4): p. 1127-35.
191. Zhang, E.Z., et al., *Development of a sensitive RT-PCR method for amplifying and sequencing near full-length HCV genotype 1 RNA from patient samples.* Virol J, 2013. **10**: p. 53.
192. Wyles, D., et al., *Post-treatment resistance analysis of hepatitis C virus from phase II and III clinical trials of ledipasvir/sofosbuvir.* J Hepatol, 2017. **66**(4): p. 703-710.
193. Zeuzem, S., et al., *NS5A resistance-associated substitutions in patients with genotype 1 hepatitis C virus: Prevalence and effect on treatment outcome.* J Hepatol, 2017. **66**(5): p. 910-918.
194. *Sofosbuvir*, in *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury*. 2012: Bethesda (MD).
195. Gane, E.J., et al., *The emergence of NS5B resistance associated substitution S282T after sofosbuvir-based treatment.* Hepatol Commun, 2017. **1**(6): p. 538-549.
196. Bertoletti, A., et al., *Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells.* Nature, 1994. **369**(6479): p. 407-10.

197. Habermann, D., et al., *HAMdetector: a Bayesian regression model that integrates information to detect HLA-associated mutations.* Bioinformatics, 2022. **38**(9): p. 2428-2436.

9 Danksagung

Mein herzlicher Dank gilt allen, die mich über die Jahre hinweg bei meiner Forschung unterstützt haben und ohne deren Hilfe diese Arbeit nicht möglich gewesen wäre.

Ein besonderer Dank gilt Prof. Dr. Jörg Timm, der bereits früh an mich glaubte und mich kontinuierlich förderte. Für die Freiheit neue Ideen auszuprobieren, für Deine unermüdliche Unterstützung sowie für die zahlreichen Diskussionen und wertvollen Ratschläge danke ich dir, Jörg.

Ich danke allen Kooperationspartnern, Klinikerinnen und Klinikern sowie den Patientinnen und Patienten, die uns ihre Proben für die Experimente zur Verfügung gestellt haben. Besonders hervorheben möchte ich Prof. Alexander Dilthey, dessen stetiger Optimismus und Tatkraft die SARS-Kontaktnachverfolgung in Düsseldorf erst ermöglicht haben. „Hey man“, Danke!

Mein besonderer Dank gilt meiner Kollegin und Büronachbarin PD Dr. Nadine Lübke. Ohne deine engagierte Mitarbeit, die vielen Diskussionen, Tipps und Tricks wäre diese Habilitationsschrift kaum realisierbar gewesen. Danke, Nadine!

Den aktuellen als auch den ehemaligen Mitgliedern der Arbeitsgruppe, für die zahllosen Experimente, die unzähligen Sequenzierungsläufe und die stets inspirierenden Gespräche. Besonders möchte ich mich bei Dr. Martha Paluschinski für die Unterstützung bei der Arbeit und für die unzähligen Datenschutz- und Ethikanträge bedanken.

Mein Dank richtet sich zudem an Prof. Michael Nassal und Prof. Michael Roggendorf für ihre wertvolle Unterstützung während meiner Promotion und der ersten Post-Doc-Phase.

Abschließend möchte ich mich bei meiner gesamten Familie und meinen Eltern bedanken, die mich immer unterstützt haben. Mein besonderer Dank geht an meine Frau Maraike, die mir immer den Rücken freigehalten hat und mir die Kraft gegeben hat weiterzumachen, selbst in schwierigen Zeiten. Ebenso möchte ich meinen Kindern Malte, Zoe, Talea und Ronja für die gemeinsamen und schönen Momente bedanken.

10 Erklärungen und eidesstattliche Versicherung

Hiermit erkläre ich, dass bei den wissenschaftlichen Arbeiten, die Gegenstand meiner Habilitationsleistung sind, ethische Grundsätze und die jeweils gültigen Empfehlungen zur Sicherung guter wissenschaftlicher Praxis durch mich beachtet wurden.

Düsseldorf, 11.12.2025 _____

Dr. rer. nat. Andreas Walker

Hiermit erkläre ich, dass keine weiteren Habilitationsverfahren eingeleitet oder erfolglos abgeschlossen worden sind.

Düsseldorf, 11.12.2025 _____

Dr. rer. nat. Andreas Walker

Hiermit versichere ich an Eides statt, dass ich die Beiträge zu den Publikationen, die meiner Habilitationsschrift zugrunde liegen, eigenständig geleistet habe.

Düsseldorf, 11.12.2025 _____

Dr. rer. nat. Andreas Walker

11 Zugrunde liegende Originalarbeiten

Der Anhang der Publikationen erfolgt mit freundlicher Genehmigung der Verlage.

Journal of Clinical Virology 70 (2015) 43–45

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

Short communication

Natural prevalence of resistance-associated variants in hepatitis C virus NS5A in genotype 3a-infected people who inject drugs in Germany

Andreas Walker[a], Holger Siemann[b], Svenja Groten[c], R. Stefan Ross[c], Norbert Scherbaum[b], Jörg Timm[a,*]

[a] Institute for Virology, Heinrich-Heine-University, University Hospital, Düsseldorf, Germany
[b] LVR-Hospital Essen, Department of Addictive Behavior and Addiction Medicine, Faculty of Medicine, University of Duisburg-Essen, Germany
[c] Institute of Virology, University of Duisburg-Essen, University Hospital Essen, Essen, Germany

ARTICLE INFO

Article history:
Received 30 April 2015
Received in revised form 30 June 2015
Accepted 6 July 2015

Keywords:
Hepatitis C virus
NS5A
Antiviral treatment
Genotype 3a

ABSTRACT

Background: People who inject drugs (PWID) are the most important risk group for incident Hepatitis C virus (HCV) infection. In PWID in Europe HCV genotype 3a is highly prevalent. Unfortunately, many of the recently developed directly acting antiviral drugs against HCV (DAAs) are suboptimal for treatment of this genotype. Detection of resistance-associated variants (RAV) in genotype 3a may help to optimize treatment decisions, however, robust protocols for amplification and sequencing of HCV NS5A as an important target for treatment of genotype 3a are currently lacking.
Objectives: The aim of this study was to establish a protocol for sequencing of HCV NS5A in genotype 3a and to determine the frequency of RAVs in treatment-naïve PWID living in Germany.
Study design: The full NS5A region was amplified and sequenced from 110 HCV genotype 3a infected PWID using an in-house PCR protocol.
Results: With the established protocol the complete NS5A region was successfully amplified and sequenced from 110 out of 112 (98.2%) genotype 3a infected PWID. Phylogenetic analysis of sequences from PWID together with unrelated genotype 3a sequences from a public database showed a scattered distribution without geographic clustering. Viral polymorphisms A30K and Y93H known to confer resistance in a GT3a replication model were present in 8 subjects (7.2%).
Conclusions: A protocol for amplification of nearly all GT3a samples was successfully established. Substitutions conferring resistance to NS5A inhibitors were detected in a few treatment-naive PWID.

1. Background

During the last few years, several directly acting antivirals (DAAs) became available for the treatment of HCV [1], starting a new era of therapy of chronic hepatitis C. Most novel compounds were optimized for inhibition of viral replication of HCV genotype 1a and 1b and typically show no or much weaker activity against genotype 3a. At the moment, only the nucleoside analogue sofosbuvir and the NS5A inhibitors ledipasvir and daclatasvir are approved by the european medicines agency (EMA) for treatment of patients infected with HCV genotype 3a. If an NS5A inhibitor is used for treatment, a combination with sofosbuvir with or without ribavirin is currently recommended (guidelines of the German medical society) [2]. Although the number of patients that have been treated with this combination is still limited, sustained viral response rates between 63% and 96% have been reported, depending on the fibrosis stage, prior treatment experience and the duration of therapy. Illicit intravenous drug use has become the most important risk factor for incident HCV infection. In 2013 more than 80% of the newly diagnosed HCV infections in Germany were most likely transmitted by intravenous drug use [3]. Notably, HCV genotype 3a is associated with injection drug use [4] and is the most frequent subtype in a single-center cohort of PWID collected in Essen, Germany (data not shown).

2. Objectives

The role of resistance associated variants (RAV) for treatment of patients infected with HCV genotype 3a is largely unclear. For

* Corresponding author. Institute for Virology, Heinrich-Heine-University Düsseldorf, University Hospital Düsseldorf, Building No. 22.21.02.49, Universitätsstr. 1, 40225 Düsseldorf, Germany.
E-mail address: joerg.timm@med.uni-duesseldorf.de (J. Timm).

http://dx.doi.org/10.1016/j.jcv.2015.07.008
1386-6532/

Table 1
Primer used in this study.

Reaction	Primer name	Primer sequence (5′-3′)	Amplicon size
RT	GT3a-7544-R	ATCGCCCGGCTCYCCCTCGA	
PCR-1	GT3a-5145-F	CCRAGYTGGGACGAGAYGTGGA	2403 bp
	GT3a-7544-R	ATCGCCCGGCTCYCCCTCGA	
PCR-2	GT3a-5275-F	TCATGGYATGCATGTCAGCYGA	2175 bp
	GT3a-7444-R	GTGTGTCRACCCCRGAGGATGA	
Sequencing	GT3a-5275-F	TCATGGYATGCATGTCAGCYGA	
	GT3a-5769-F	AACATACTCGGGGGGTGGGTTGC	
	GT3a-7444-R	GTGTGTCRACCCCRGAGGATGA	

Table 2
The frequency of RAVs in GT3a infected, treatment-naïve PWIDs.

Position	28	(n)	30	(n)	31	(n)	32	(n)	93	(n)
Prototype	M	(108)	A	(98)	L	(110)	P	(110)	Y	(106)
Substitutions	V	(1)	**K**	**(5)**[a]					**H**	**(3)**[b]
	L	(1)	T	(3)					F	(1)
			V	(2)						
			L	(1)						
			S	(1)						

[a] 44-fold IC50 change (6).
[b] 2154-fold IC50 change (6).

detection and monitoring of RAVs reliable protocols for amplification and sequencing of NS5A are required. The aim of this study was to establish a set of primers for amplification of genotype 3a and to utilize the developed PCR protocol to determine the frequency of RAVs in HCV genotype 3a in a cohort of treatment-naïve PWID in Germany.

3. Study Design

Blood samples from patients with a history of injection drug use were collected from the ward for inpatient detoxification treatment of drug addicts or the clinic for opioid maintenance treatment (OMT) at the Department of Addictive Behavior and Addiction Medicine, LVR-Hospital Essen, Hospital of the University of Duisburg-Essen. Written informed consent was obtained from all study participants and the study was approved by the ethics committee of the Medical Faculty of the University of Duisburg-Essen in accordance with the Declaration of Helsinki. Viral RNA was extracted from all patients ($n = 112$) infected with genotype 3a (determined by sequencing of the core region [5]) with the QIAamp viral RNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. An alignment of all available GT3a sequences from the HCV sequence database between position nt5000 and nt8000 was used to design primers covering nearly all GT3a variants (Table 1). RNA was transcribed with Superscript III (Invitrogen) with the reverse primer GT3a-7544-R (final concentration 0.5 pmol/μl). NS5A was amplified in a two-step nested PCR using GoTaq Polymerase (Promega) with the primers as outlined in Table 1 and the following PCR conditions: 120 s at 94 °C followed by 35 cycles each 30 s 94 °C, 30 s 55 °C and 160 s 72 °C followed by 10 min at 72 °C. The PCR products were directly sequenced and sequences were aligned with the software Geneious 7.1.5 (Biomatters, Auckland, New Zealand). All sequences were submitted to GenBank (#KR082016–KR082126) Table 2.

4. Results

The PCR protocol was validated with samples of 112 HCV-RNA positive treatment-naïve PWID with a broad range of viral load (median 2.4×10^5 IU/ml, range 615–3.4×10^7 IU/ml). Amplification of the expected 2 kb fragment was successful in 110 of 112 (98.2%) samples. The samples for which the PCR has failed had a viral load of

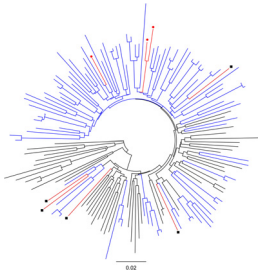


Fig. 1. Phylogenetic tree of NS5A sequences from HCV genotype 3a. NS5A sequences from PWID collected in Germany (blue or red) and from the HCV database (black) were aligned and a phylogenetic tree was calculated with the neighbor-joining method and the Jukes–Cantor distance model. Sequences with RAVs are shown in red and are marked with red dots (Y93H) or black squares (A30K). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

3.502 and 930 IU/ml, respectively. Notably, 10 of 12 samples with a viral load below 10.000 IU/ml were positive in the PCR including the sample with the lowest viral load in the cohort (615 IU/ml). Bulk sequencing of all 110 PCR products was performed with primers listed in Table 1. A neighbor-joining phylogenetic tree of the resulting NS5A nucleotide sequences revealed that the sequences from PWID in Germany are equally distributed among other unrelated GT3a sequences (Fig. 1). Polymorphisms known to confer resistance against NS5A inhibitors in genotype 3a were previously identified in a subgenomic replication model and include the substitutions A30K, L31 V/M and Y93H [6]. Additional resistance associated positions identified in genotype 1 are 28, 32 and 58, however, it is not clear if substitutions at these positions in genotype 3a impact susceptibility to NS5A inhibitors. The residue L31 and P32 were fully conserved in our cohort and substitutions at position M28 were observed in 2 of 110 sequences (M28 V and M28L). Residue A30 was more polymorphic with 12 of 110 sequences (10.7%) harboring substitutions (A30KTVSL). Of these polymorphisms, only the A30K substitution was shown to confer reduced susceptibility to the NS5A inhibitors daclatasvir and ledipasvir and was present in 5 of 110 (4.5%) subjects. The most important position for resistance to NS5A inhibitors is position 93. Variations from the prototype residue tyrosine (Y93HF) were observed in 4 of 110 patients (3.6%) including three patients (2.7%) with the substitution Y93H known to confer high-level resistance to NS5A inhibitors. Taken together,

polymorphisms at positions associated with resistance to NS5A inhibitors were detected in 18 of 110 (16.4%) PWID infected with genotype 3a including two substitutions (A30K and Y93H) known to confer resistance to NS5A inhibitors in genotype 3a that were detected in 5 and 3 PWIDS, respectively.

5. Discussion

The clinical relevance of resistance mutations for treatment of HCV genotype 3a with DAAs is currently largely unclear. Although irrespective of the infecting genotype RAVs were detected in the majority of patients who fail to achieve sustained virological response (SVR) in clinical studies, the frequency of treatment failures with new combinations of DAAs is overall low [7,8]. Nevertheless, there is evidence from HCV genotype 3a infected patients treated with sofosbuvir and daclatasvir that pre-existing RAVs in NS5A were more frequent in the small number of patients experiencing viral relapse. In the ALLY-3 trial six of 13 patients (46%) who carried the Y93H substitution before treatment had a viral relapse whereas only ten of 134 (7.5%) of patients carrying the prototype residue tyrosine in this position failed to achieve SVR [8]. This strongly suggests that in genotype 3a presence of this substitution at baseline negatively impacts the response to combination therapy with an NS5A inhibitor and longer treatment duration may be considered. Importantly, it has been reported that RAVs in NS5A selected during treatment are associated with low fitness costs and can be detected for months after treatment cessation as the predominant variant of the quasispecies [9]. This is in line with the detection of NS5A RAVs in treatment naïve patients. The Y93H substitution was reported in 13 of 147 (8.8%) in the ALLY-3 trial. Similar frequencies were previously reported in a study by Hernandez at al. (8 of 100; 8%) [6]. In PWIDs from Germany we observed the Y93H substitution as the dominant variant of the quasi species in 3 of 110 (2.7%) patients. Future studies specifically aimed to determine if detection of this substitution prior to therapy impacts treatment outcome or influences therapy decisions will be needed. The protocol presented here for amplification and sequencing of the NS5A region in genotype 3a may help addressing this question.

Competing interests

None declared.

Ethical approval

The study was approved by the ethics committee of the Medical Faculty of the University of Duisburg-Essen in accordance with the Declaration of Helsinki.

Acknowledgement

The study was supported by a grant of the German Ministry of Health to the National Reference Centre for Hepatitis C (Essen).

References

[1] S. Zeuzem, G.M. Dusheiko, R. Salupere, A. Mangia, R. Flisiak, R.H. Hyland, A. Illeperuma, E. Svarovskaia, D.M. Brainard, W.T. Symonds, G.M. Subramanian, J.G. McHutchison, O. Weiland, H.W. Reesink, P. Ferenci, C. Hezode, R. Esteban, V. Investigators, Sofosbuvir and ribavirin in HCV genotypes 2 and 3, N. Engl. J. Med. 370 (2014) 1993–2001.

[2] Christoph Sarrazin, Thomas Berg, Peter Buggisch, Matthias Dollinger, Holger Hinrichsen, Harald Hofer, Dietrich Hüppe, Michael Manns, Stefan Mauss, Jörg Petersen, Karl-Georg Simon, Ingo van Thiel, Heiner Wedemeyer, Zeuzem1 S. 2015. Aktuelle Empfehlung zur Therapie der chronischen Hepatitis C. Addendum zur Hepatitis C Leitlinie.

[3] S. Schink, R. Zimmermann, Virushepatitis C im Jahr 2013, Epidemiologisches Bull. 2014 (2014) 275–284.

[4] O.G. Pybus, A. Cochrane, E.C. Holmes, P. Simmonds, The hepatitis C virus epidemic among injecting drug users, Infect. Genet. Evol.: J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 5 (2005) 131–139.

[5] S. Viazov, S.S. Ross, K.K. Kyuregyan, J. Timm, C. Neumann-Haefelin, O.V. Isaeva, O.E. Popova, P.N. Dmitriev, F. El Sharkawi, R. Thimme, M.I. Michailov, M. Roggendorf, Hepatitis C virus recombinants are rare even among intravenous drug users, J. Med. Virol. 82 (2010) 232–238.

[6] D. Hernandez, N. Zhou, J. Ueland, A. Monikowski, F. McPhee, Natural prevalence of NS5A polymorphisms in subjects infected with hepatitis C virus genotype 3 and their effects on the antiviral activity of NS5A inhibitors, J. Clin. Virol. 57 (2013) 13–18.

[7] N. Afdhal, K.R. Reddy, D.R. Nelson, E. Lawitz, S.C. Gordon, E. Schiff, R. Nahass, R. Ghalib, N. Gitlin, R. Herring, J. Lalezari, Z.H. Younes, P.J. Pockros, A.M. Di Bisceglie, S. Arora, G.M. Subramanian, Y. Zhu, H. Dvory-Sobol, J.C. Yang, P.S. Pang, W.T. Symonds, J.G. McHutchison, A.J. Muir, M. Sulkowski, P. Kwo, I.O.N. Investigators, Ledipasvir and sofosbuvir for previously treated HCV genotype 1 infection, N. Engl. J. Med. 370 (2014) 1483–1493.

[8] D.R. Nelson, J.N. Cooper, J.P. Lalezari, E. Lawitz, P.J. Pockros, N. Gitlin, B.F. Freilich, Z.H. Younes, W. Harlan, R. Ghalib, G. Oguchi, P.J. Thuluvath, G. Ortiz-Lasanta, M. Rabinovitz, D. Bernstein, M. Bennett, T. Hawkins, N. Ravendhran, A.M. Sheikh, P. Varunok, K.V. Kowdley, D. Hennicken, F. McPhee, K. Rana, E.A. Hughes, A.-S. Team, All-oral 12-week treatment with daclatasvir plus sofosbuvir in patients with hepatitis C virus genotype 3 infection: ALLY-3 phase III study, Hepatology 61 (2015) 1127–1135.

[9] K. Kosaka, M. Imamura, C.N. Hayes, H. Abe, N. Hiraga, S. Yoshimi, E. Murakami, T. Kawaoka, M. Tsuge, H. Aikata, D. Miki, H. Ochi, H. Matsui, A. Kanai, T. Inaba, K. Chayama, Emergence of resistant variants detected by ultra-deep sequencing after asunaprevir and daclatasvir combination therapy in patients infected with hepatitis C virus genotype 1, J. Viral hepatitis 22 (2015) 158–165.

Journal of Clinical Virology 91 (2017) 42–48

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

A genotype independent, full-genome reverse-transcription protocol for HCV genotyping and resistance testing

Andreas Walker[a], Matthias Bergmann[a], Jennifer Camdereli[a], Rolf Kaiser[b], Nadine Lübke[a], Jörg Timm[a,*]

[a] Institute of Virology, Heinrich-Heine-University, University Hospital, Düsseldorf, Germany
[b] Institute of Virology, University of Cologne, Cologne, Germany

ARTICLE INFO

Keywords:
Hepatitis C virus
Genotyping
Antiviral treatment
NS5A
Reverse transcription
Resistance testing

ABSTRACT

Background: HCV treatment options and cure rates have tremendously increased in the last decade. Although a pan-genotype HCV treatment has recently been approved, most DAA therapies are still genotype specific. Resistance-associated variants (RAVs) can limit the efficacy of DAA therapy and are associated with increased risk for therapy failure. With the approval of DAA regimens that recommend resistance testing prior to therapy, correct assessment of the genotype and testing for viruses with RAVs is clinically relevant. However, genotyping and resistance testing is generally done in costly and laborious separate reactions.

Objective: The aim of the study was to establish a genotype-independent full-genome reverse transcription protocol to generate a template for both genotyping and resistance testing and to implement it into our routine diagnostic setup.

Study design: The complete HCV genome was reverse transcribed with a pan-genotype primer binding at the 3′end of the viral RNA. This cDNA served as template for transcription of the genotyping amplicon in the core region as well as for the resistance testing of NS3, NS5A, and NS5B.

Results: With the established RT-protocol the HCV core region was successfully amplified and genotyped from 124 out of 125 (99.2%) HCV-positive samples. The amplification efficiency of RAV containing regions in NS3, NS5A, NS5B was 96.2%, 96.6% and 94.4%, respectively.

Conclusions: We developed a method for HCV full-genome cDNA synthesis and implemented it into a routine diagnostic setup. This cDNA can be used as template for genotyping amplicons covering the core or NS5B region as well as for resistance testing amplicons in NS3, NS5A and NS5B.

1. Background

HCV treatment options and viral cure rates (sustained virological response, SVR) have tremendously increased in the last decade and SVR rates > 95% are now achieved with combinations of direct acting antivirals (DAA). Although drugs active against all HCV genotypes have been approved, most DAA therapies are still genotype specific. Data from a large multi-center cohort showed that approximately 50% of unsuccessful IFN-free DAA therapies is caused by misclassification of the HCV genotype or subtype [1]. Therefore, correct identification of HCV genotype and subtype is required for treatment decisions and a predictor for the outcome of DAA-based therapy. Although with commercial HCV genotyping assays correct identification of the genotype is typically achieved (93–99%), subtype prediction is less efficient [2–4]. Sequencing of highly discriminating fragments either from the core or the NS5B region is considered as the gold standard for HCV genotyping and also permits identification of recombinants [5,6]. However, the work flow requires multiple steps including generation of amplicons, sequencing and genotype prediction and is therefore time consuming.

Resistance-associated variants (RAVs) can limit the efficacy of DAA therapy and have been associated with increased risk for therapy failure in some treatment combinations [7,8]. NS5A inhibitors are standard components of DAA combination therapies and virtually all patients with treatment failure have selected RAVs in NS5A [9]. These RAVs can persist after therapy and possibly impair SVR rates upon re-therapy [10]. Baseline NS5A RAVs can be found in up to 16% of therapy naive patients [8,11,12]. Hence, resistance testing for the presence of virus with NS5A RAVs before therapy is recommended for selected DAA combinations [13,14] and may help to optimize treatment decisions in

* Corresponding author at: Institute for Virology, Heinrich-Heine-University Düsseldorf, University Hospital Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany.
E-mail address: joerg.timm@med.uni-duesseldorf.de (J. Timm).

http://dx.doi.org/10.1016/j.jcv.2017.04.008
Received 31 January 2017; Received in revised form 31 March 2017; Accepted 6 April 2017

patients with negative predictors for SVR (e.g. patients with cirrhosis, infected with genotype 3 or with prior treatment failure).

2. Objective

The enormous sequence variability of HCV requires subgenotype-specific PCRs for successful amplification of RAV containing regions. Thus, at present, identification of the viral genotype is a precondition for subsequent PCR and sequencing for resistance analysis. Usually, the genotype is predicted by either commercial hybridization tests or by in-house amplicon sequencing. Here, we aimed to optimize the work flow to a single reverse transcription reaction to generate cDNA that allows subsequent amplification of all genomic regions that are relevant for treatment decisions (core, NS3, NS5A and NS5B) in a fast, reliable and cheap routine diagnostic setup.

3. Study design

3.1. Plasma samples

Plasma samples of patients with HCV-infection were obtained in the framework of the German PEPSI study for the surveillance of HCV baseline resistance. Written informed consent was obtained from all study participants and the study was approved by the ethics committee of the Medical Faculty of the University of Dusseldorf (study number #2012048) in accordance with the Declaration of Helsinki. Viral RNA from 400 µl plasma was extracted automatically using the EZ1 Virus Mini Kit v2.0 on an EZ1 Advanced XL robot or manually with the QIAamp Viral RNA Mini Kit (both Qiagen, Hilden, Germany) according to the manufacturer's protocol. RNA was eluted in a volume of 60 µl and stored at −20 °C.

3.2. Reverse transcription

RNA was reverse transcribed in vitro with Superscript III (Invitrogen). To reduce hands-on time all mixes were prepared in batches as following: 3 µl/well "frozen-mix" containing 1 µl reverse primer (Table 1), 1 µl dNTPs (Roche,10 mM each) and 1 µl water were aliquoted in 8-strips with hinged-caps (Eppendorf #951010022) and stored at −20 °C until usage. When the mix contained two reverse primers water was omitted. For reverse transcription an appropriate number of tubes was thawed and 10 µl RNA was added to the "frozen-mix". Secondary RNA-structures were reduced by melting for 5 min at 65 °C. The mixture was cooled down to 25 °C and reverse transcription was started by addition of 7 µl/well freshly prepared reverse transcription mix (4 µl SSIII-Buffer, 1 µl DTT, 1 µl RNase Inhibitor (NEB) and 1 µl SSIII) with the following conditions: 10 min at 25 °C, 60 min at 42 °C, 30 min at 50 °C, 30 min 55 °C, 15 min at 75 °C and 4 °C.

3.3. Nested PCR and genotyping

A two-step nested PCR for HCV genotyping or resistance testing was performed by using genotype specific primer combinations (Table 2) and the GoTaq Polymerase (Promega) according to the manufacturer's protocol. First, "frozen PCR-mixes" corresponding to the required target region were prepared. Per well 46 µl PCRI mixture containing GoTaq Buffer colorless (Promega), 200 µM dNTPs (Roche), 0.5 µM each Primer, 1.25 Units GoTaq Polymerase were aliquoted in 8-strips with hinged-caps (Eppendorf #951010022) and stored at −20 °C until usage. PCRII mixes were identical to PCRI except that the volume was 47 µl and GoTaq Buffer green was used. For the amplification step an appropriate number of tubes was thawed and four microliter of cDNA were used for the 1st round of nested-PCR (for PCR conditions see Suppl. Table 1). Subsequently, three microliter of PCR-product from the 1st round was used for the 2nd round of nested-PCR. Positive PCR-products were purified with the QIAquick PCR Purification Kit (Qiagen, Hilden) and were Sanger sequenced. Sequences were aligned and a consensus sequence was generated with the software Geneious 7.1.5 (Biomatters, Auckland, New Zealand).

The $core_{GT}$ consensus sequence was used for genotype prediction by a HCV blast search (http://hcv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html) and the NCBI Genotyping tool (http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi). In case of discrepancies or low sequence similarity (< 90%) a phylogenetic tree with genotype reference sequences was used to determine the genotype. In cases of ambiguous prediction or evidence for 1b/2k recombinants the $NS5B_{GT}$ fragment was also amplified and sequenced. Resistance analysis was performed by the a prediction algorithm as implemented in $geno2pheno_{[HCV]}$ (http://hcv.bioinf.mpi-inf.mpg.de/index.php) [15].

4. Results

Since the highly predictive region for genotyping in core and the DAA resistance regions are located at opposite ends of the HCV genome an important question was, if a genotype independent generation of a full-length cDNA was possible in a routine diagnostic setup. To address this, RNA of a plasma samples from a GT1a infected patient with high viremia was extracted using the EZ1 automatic extraction device and reverse transcription was initiated with primer sv542as, binding to a conserved region in NS5B [16]. After reverse transcription the length of the synthesized cDNA was verified by nested PCRs covering different shorter regions of the HCV genome. As seen in Fig. 1B, fragments across the complete HCV genome could be successfully amplified, including the core fragment used for genotyping ($core_{GT}$) near the 5′end of the genome [17]. Successful whole genome cDNA synthesis was also confirmed with HCV RNA purified by a manual protocol (QIAamp Viral RNA Mini Kit, data not shown).

Efficient resistance testing requires sequencing of complete NS5B including highly conserved regions (NS5B aa555–560) that usually are used as primer binding sites [18,19]. Therefore, full genome reverse transcription with primers binding to the thymidine tract or the conserved X-tail region was analyzed next. Primer sv543as binding in NS5B served as a control. Serial RNA dilutions were used for cDNA synthesis and the success was probed with the $core_{GT}$ specific PCR. Synthesis of shorter cDNAs was tested by PCR amplification of a fragment in NS5B that can also be used for HCV genotyping ($NS5B_{GT}$) [16]. Amplification of the $core_{GT}$ fragment was possible from all cDNAs synthesized from HCV-RNA dilutions down to 10,000 IU/ml of HCV-RNA (Fig. 1C). In contrast, the amplification efficiency of the $NS5B_{GT}$ fragment was different depending on the primer used for cDNA synthesis. After cDNA synthesis with the primer in the highly structured X-tail region, amplification of the $NS5B_{GT}$ fragment was successful down to HCV-RNA concentrations of 10,000 IU/ml. When cDNA was synthesized with the primer sv542as, amplification of the $NS5B_{GT}$ was

Table 1
Primer used for reverse transcription.

Reverse transcription primer	Sequence 5′ → 3′	Final concentration [µM]	Ref.
sv542as	GGAGGGGCGGAATACCTGGTCATRGCYTCCGTRAA	0.5	[16]
Oligo d(A)	AAAAAAAAAAAAAAAAAAAA	0.5	[19]
X-tail	AGCTAGCCGTGACTAGGGCTAAG	2.5	–

Table 2
Primer combination used for nested PCR.

Genotype	Region	Nested PCR	Forward primer	Sequence 5′ → 3′	Reverse primer	Sequence 5′ → 3′	Amplicon size (bp)
Pan-GT	$Core_{GT}$	PCR1	p417s	GGYGGYGGNCAGATCGTTGG	p874as	ARGAAGATAGARAARGAGCAACC	457
		PCR2	p439s	GAGTWTACBTGYTGCCGCGCAG	p1As	ATRTACCCCATGAGRTCGGC	312
Pan-GT	$NS5B_{GT}$	PCR1	sv360s & sv361s	CTTCTCATATGACACCCGCTGYTTYGA CTTCTCATATGACACCCGWTGCTTYGA	sv542as	GGAGGGGCGGAATACCTGGTCATRGCYTCCGTRAA	397
		PCR2	sv362s & sv363s	CATATGACACCCGCTGYTTYGACTC CATATGACACCCGWTGCTTYGAYTC	sv542as	GGAGGGGCGGAATACCTGGTCATRGCYTCCGTRAA	394
GT1a	NS3	PCR1	HCV1a_3a-F	ATGTGGCCTCTCCTCCTGC	HCV1a_4a-R	TCTCCGGTGGTGGACAGAGC	1719
		PCR2	HCV1a_3b-F	TGGRTTCCCCCCCTCAACG	HCV1a_3e-R	ATCCGTGGAGTGGCACTCG	1370
	NS5A	PCR1	HCV1a_5c-F	TGAGCGGTGAGGTCCCCTCC	HCV1a_5e-R	GAGGGATCAGTGAGCATGG	912
		PCR2	HCV1a_5d-F	TGGGACTGGATATGCGAGG	HCV1a_5d-R	GTGAAAAATTCGGGCGATGG	408
	NS5B	PCR1	HCV1a + b-7527-F	AYCCIGATCTCAGCGACGGRTC	HCV1a_7e-R_int	CCTGCAGCAAGCAGGAGTAGGC	1772
		PCR2	HCV1a + b-7587-F	GAYGTCGTSTGCTGCTCRATGTC	HCV1a_7e-R_int	CCTGCAGCAAGCAGGAGTAGGC	1712
GT1b	NS3	PCR1	HCV1b_3b-F	GCCGCGATGCCATCATCC	HCV1b_4a-R HCV1b_4b-R	CACGACRAGCCGCGCTCC CATTAGAGCGTCTGTTGC	1693
		PCR2	HCV1b_3c-F	TTGCGGTGGCAGHAGAGC	HCV1b_3e-R	CGCCCGTGGTGATGGTCC	940
	NS5A	PCR1	HCV1b_5c-F	ATAGGCCTTGGGAAGGTGC	HCV1b_6a-R	ATGTTYCCGCCCATCTCCTGCCG	1176
		PCR2	HCV1b_5d-F	CCCCACGCACTATGTGCC	HCV1b_5e-R	AGGGGTCGGTRAGCATGG	728
	NS5B	PCR1	HCV1a + b-7527-F	AYCCIGATCTCAGCGACGGRTC	HCV1b_7e-R	GGGGAGCAGGTAGATGCC	1829
		PCR2	HCV1a + b-7587-F	GAYGTCGTSTGCTGCTCRATGTC	GT1b_9325-R	CCYACRGARAGTAGGAGTAGGC	1738
GT2	NS5A	PCR1	GT2-6075-F	GCGGTCCARTGGATGAAYAGRCT	GT2a-7038-R & GT2b-7038-R	GT2a-7038-R: GCATCCACCATGTCCACYTCAT GT2b-7038-R: GCATCCACCATGTCACAYTCAT	963
		PCR2	GT2-6102 -F	GCCTTYGCYTCCAGAGGAAACCA	GT2-6816-R	TCAGGRTCRCARGGRAGCTG	714
	NS5B	PCR1	GT2-7497-F	TCCTCCATGCCYCCCCTYGAG	sv542as	GGAGGGGCGGAATACCTGGTCATRGCYTCCGTRAA	1119
		PCR2	GT2-7599-F	TGCTCYATGTCATACTCCTGGAC	sv542as	GGAGGGGCGGAATACCTGGTCATRGCYTCCGTRAA	1017
GT3	NS5A	PCR1	(+)-GT3a-5145	CCRAGYTGGGACGAGAYGTGGA	(−)-GT3a-7544	ATCGCCCGGCTCYCCCTCGA	2399
		PCR2	(+)-GT3a-5275	TCATGGYATGCATGTCAGCYGA	(−)-GT3a-7444	GTGTGTCRACCCCRGAGGATGA	2169
	NS5B	PCR1	HCV3a-6840-F	GATGTBTCTGTGCTGACCTCGATG	HCV3a-9276-R	CACGYGACACGCTGTGATAAATGTC	2436
		PCR2	GT3a-7283-F	ACCACCAACTGTCCATGG	HCV3a-9171-R	GTACCGCCCAATTRAAGAGRTA	1888
GT4	NS3	PCR1	(+)-HCV4-3264	RTRTTYACRSCYATGGAGAAG	(−)-HCV4-4275	GARTGGCAYTCRTCACARATGATRAT	1011
		PCR2	(+)-HCV4-3364	ATGAGATMYTGCTCGGRCCRGC	(−)-HCV4-4275	GARTGGCAYTCRTCACARATGATRAT	911
	NS5A	PCR1	(+)-GT4-5937-F	TTYAARATCATGAGYGGYGA	(−)-GT4d-6677-R	ACCTGGCAGGGRCACTTGAT	780
		PCR2	(+)-GT4-6069-F	GAGGGRGCRGTYCAGTGGATGAA	(−)-GT4d-6674-R	CAGGGGCACTTGATGTTRTC	645
	NS5B	PCR1	HCV4-7933-F	ATTCCCGCAAGGCCRTTAACCA	GTx-9276-R	CAYGAGACAIGCTGTGATARATGTC	1343
		PCR2	HCV4-7948-F	TTAACCACATCARCTCCGTGTGG	HCV4-9269-R	GCTGTGATAAATGTCYCCCCCG	1321

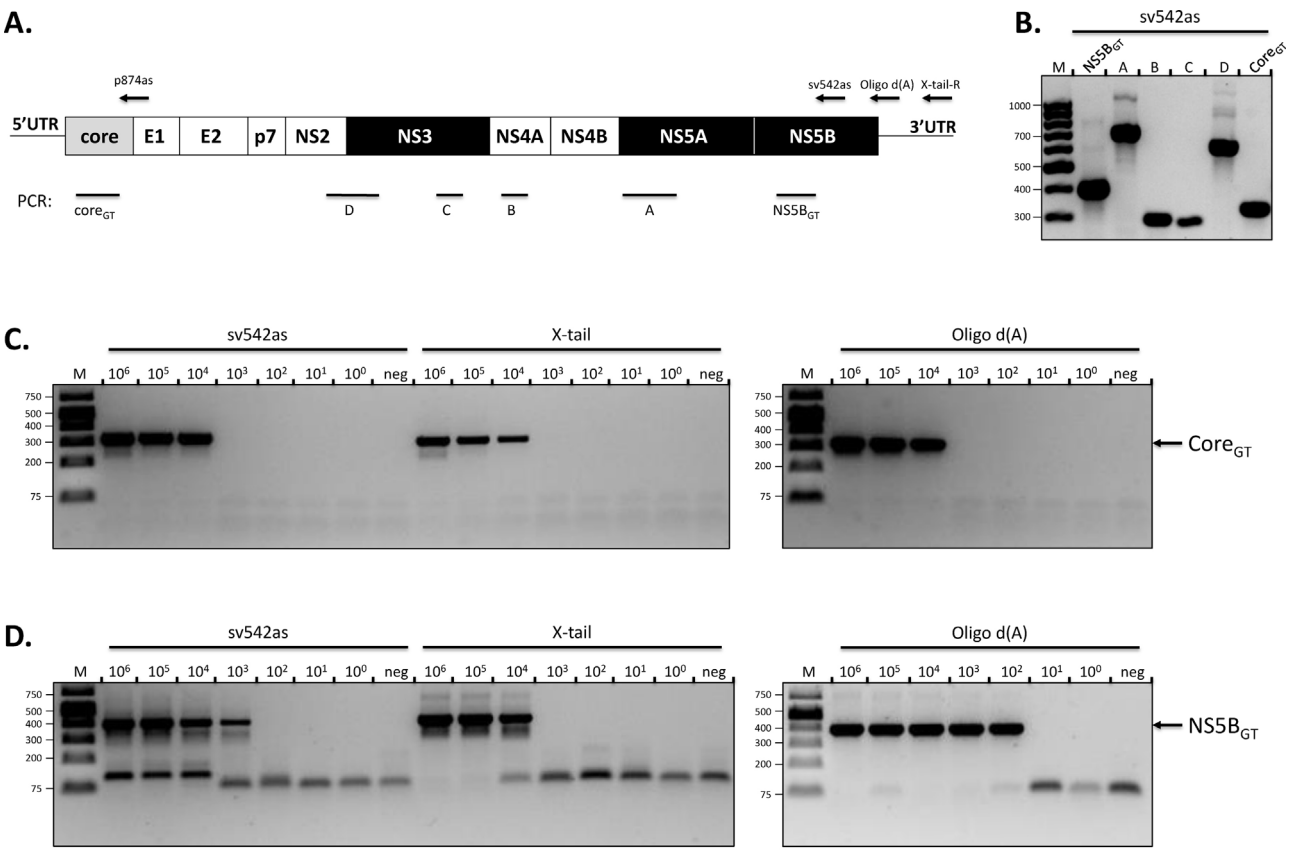


Fig. 1. Rationale for the reverse transcription approach. (A) Scheme of the HCV genome based on the H77 reference genome. The location of reverse transcription primer is depicted above and amplicons for analysis of cDNA synthesis length are depicted below the HCV genome. RAV containing proteins are depicted in black. (B) Probing of cDNA length. RNA from a patient with high viral load (GT1a; 10^7 IU/ml HCV-RNA) was reverse transcribed with primer sv542as and length of the cDNA was probed with PCRs depicted in A. 5 µl aliquots of the different PCRs were separated in a 1% agarose gel and stained with ethidium bromide. M. Marker 100 bp ladder (PeqLab). (C) Full genome cDNA synthesis. Serial RNA dilutions were reverse transcribed with the indicated primers and efficiency of full length cDNA synthesis was probed by nested PCR with core specific primers ($core_{GT}$). M. Marker 1 kb ladder plus (Fermentas). (D) Synthesis efficiency of shorter cDNAs. Synthesis efficiency of shorter cDNAs was analyzed by amplification of the $NS5B_{GT}$ fragment from the reverse transcription shown in C. M. Marker 1 kb ladder plus (Fermentas).

possible down to concentrations of 1000 IU/ml. Interestingly, the highest sensitivity was achieved with the Oligo d(A) primer. Here, amplification of $NS5B_{GT}$ was possible from 100 IU/ml (≈ 20 copies/reaction) (Fig. 1D). Therefore primer Oligo d(A) was chosen for all further experiments.

Oligo d(A) permitted $core_{GT}$ amplification from 10,000 IU/ml HCV-RNA, however, this sensitivity was considered insufficient for routine genotyping. We therefore analyzed if the addition of a core specific primer improves amplification of the $core_{GT}$ amplicon from samples with low viral load. Therefore, serial RNA dilutions were reverse transcribed with the primers p874as and Oligo d(A) and cDNA synthesis was verified with the $core_{GT}$ PCR. As shown in Fig. 1C this combination increased the sensitivity of the $core_{GT}$ specific PCR down to 1000 IU/ml HCV-RNA without influencing the amplification of $NS5B_{GT}$ (Fig. 2A and B). Thus, the primer combination p874as and Oligo d(A) was used for cDNA synthesis in all further experiments.

After optimizing cDNA synthesis we analyzed the amplification efficiency of the DAA targets NS3, NS5A and NS5B from these long transcripts. A collection of patient samples with known genotype from the routine diagnostic was reverse transcribed with p874as and Oligo d(A) and the $core_{GT}$ region was amplified for genotyping. After sequencing and genotype prediction, subtype specific PCRs for the NS3, NS5A and NS5B region of samples with viral loads ranging between 8000 and 6,000,000 IU/ml were performed using the same cDNA. All genotype 1a NS3 fragments could be amplified and for NS5A and NS5B only individual samples with less than 100,000 IU/ml failed (Fig. 3A). Since NS3 as the most distant region from the starting point of cDNA synthesis (NS3) was successfully amplified in all genotype 1a samples,

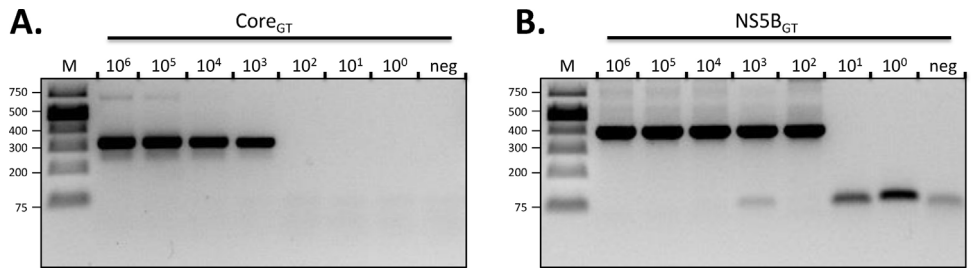


Fig. 2. The sensitivity of genotyping can be enhanced by addition of primer p874as. Serial RNA dilutions were reverse transcribed with primer Oligo d(A) and p874as. (A) cDNA synthesis was probed by nested PCR with core specific primers ($core_{GT}$). (B) Synthesis efficiency of shorter cDNAs was analyzed with the $NS5B_{GT}$ PCR. M. Marker 1 kb ladder plus (Fermentas).

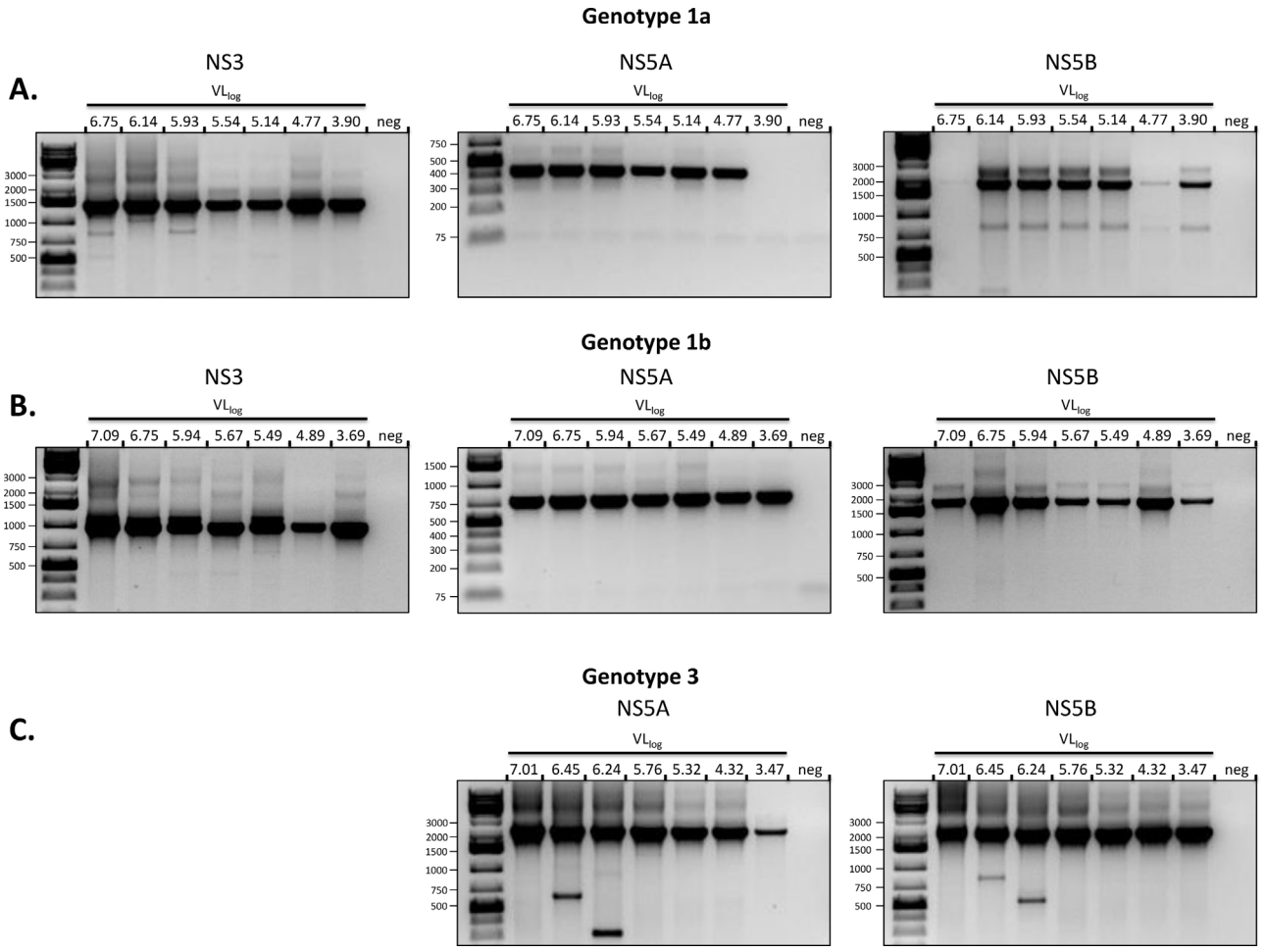


Fig. 3. Amplification of resistant relevant regions. RNA from patients was extracted and reverse transcribed with primer Oligo d(A) and p874as. For genotyping, the $core_{GT}$ fragment was amplified, sequenced and the genotype was predicted using HCV blast search. (A) Nested PCR for amplification of genotype 1a NS3, NS5A and NS5B. Based on the genotype information genotype specific amplicons for NS3, NS5A and NS5B were generated by nested PCRs. (B) Amplification of genotype 1b NS3, NS5A and NS5B. (C) Amplification of genotype 3 NS5A and NS5B. Primer and fragment sizes are given in Table 2. The viral load is given in $log_{(10)}$ and patients are arranged according to their viral load. M. Marker 1 kb ladder plus (Fermentas).

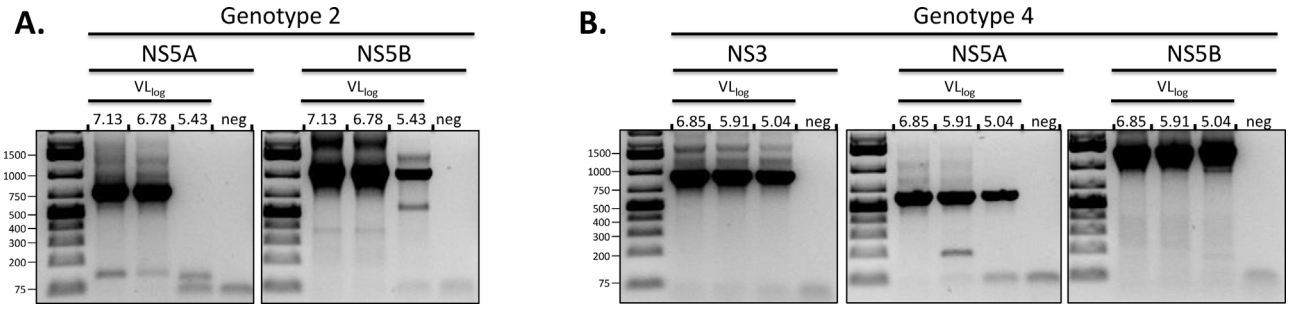


Fig. 4. Amplification of resistant relevant regions from genotype 2 and 4. (A) Amplification of NS5A and NS5B. (B) Nested PCRs for amplification of genotype 4 NS3, NS5A and NS5B. Patient samples were treated as described in Fig. 3 and for details see Fig. 3. M. Marker 1 kb ladder plus (Fermentas).

the reduced amplification efficiency for NS5A and NS5B was most likely due to primer mismatches in the nested PCR. In genotype 1b all samples could be amplified including samples with viral loads less than 10,000 IU/ml (Fig. 3B). This could also be confirmed for genotype 3a for the relevant target regions NS5A and NS5B (Fig. 3C). For genotype 4, again, all fragments were successfully amplified from the full length HCV cDNA (Fig. 4B). In genotype 2 in two out of three patients the NS5A amplicon and in all three patients the NS5B amplicon was successfully amplified.

After successful validation the protocol was implemented in our routine diagnostic. Between August 2016 and March 2017 one-hundred-twenty-four of 125 (99.2%) HCV-positive samples were successfully genotyped with this method (Median VL 937,970 IU/ml; Range 23–15,789,652). The sample for which the $core_{GT}$ PCR has failed had a viral load of 23 IU/ml. Notably, five of six samples with a viral load below 10,000 IU/ml were positive in the PCR including a sample with 938 IU/ml. From the genotyped samples 26 of 27 (96.2%) NS3, 86 of 89 (96.6%) NS5A and 17 of 18 (94.4%) NS5B resistance analyses were successfully performed.

5. Discussion

With the approval of DAA regimens that recommend resistance testing prior to therapy [13,14], testing for viruses with RAVs is

clinically relevant. Even in the era of pan-genotype DAAs the correct assessment of HCV genotypes and subtypes still has important clinical and epidemiological implications. Currently, HCV genotyping and resistance testing are done in separate reactions because most resistance testing protocols use genotype-specific primers for amplicon generation. The approach presented here has several advantages. First, a combined protocol reduces hands on time and cuts cost by approximately 50% because reverse transcription is the most expensive step. Second, the genotype-independent cDNA synthesis enhances the overall success rate, because failure of nested PCR even from samples with high viral load [20] due to mismatch of the reverse transcription primer is excluded.

Current full-length HCV genome protocols report the use of special RNA extraction and handling methods [19,21]. This is in contrast to our observation that full genome reverse transcription is also possible under routine clinical laboratory practice. Complete cDNA synthesis was possible with RNA extracted either by spin columns or by an automatic extraction system and synthesis efficiency solely depended on viral load and the primers used for RT. Theoretically, random primers, widely used for cDNA synthesis, circumvent the need of specific primers, however, typically yield short cDNAs and have reduced priming efficiency compared to specific primer [22]. In our analysis, no differences in full-length cDNA generation irrespective of the primer was observed, however, amplification of the $NS5B_{GT}$ fragment was more efficient with primer sv542as and especially Oligo d(A), indicating that also shorter cDNAs are produced. Primer X-tail-R binds in the conserved but highly structured RNA X-tail region which may interfere with correct initiation of cDNA synthesis; however, why this only affects shorter cDNA and not the full-genome cDNA synthesis is unknown. Of note, previous reports by other groups showing [21] that amplification of target region is possible when RT primers were omitted could not be observed (data not shown). $NS5B_{GT}$ could be amplified from samples with 100 IU/ml HCV-RNA when primer Oligo d(A) was used for cDNA synthesis. This primer binds to the poly-uracil tract of HCV and most likely absence of secondary structures of this region permits more efficient binding and initiation of cDNA synthesis compared to the other regions. Interestingly, cDNA synthesis with Oligo d(A) significantly reduced unspecific amplification with the $NS5B_{GT}$, a major limitation of the $NS5B_{GT}$ primer combination [16]. To permit genotyping of low viral load samples we combined Oligo d (A) with the core-specific primer p874as for efficient $core_{GT}$ amplicon generation. This combination did not influence amplification of NS3, NS5A and NS5B resistance amplicons as shown for selected samples from genotype 1–4.

From the 125 clinical samples analyzed in the routine diagnostic set up, all except one with a viral load of 23 IU/ml could be genotyped including samples < 1000 IU/ml of HCV-RNA. Amplification of requested target regions NS3, NS5A and NS5B was possible in 96.2%, 96.6% and 94.4%. Theoretically, the use of pan-genotypic primers for nested PCRs protocols [23–25] could further simplify genotyping and resistance testing. However, due to the lack of sufficient full-length genomes of rare genotypes genotyping with NS-proteins is difficult and the reported amplification rates are lower (78–90%) [24,25] compared to our genotype-specific protocols.

In this study only the requested regions were analyzed which lead to small sample sizes for NS3 and NS5B and therefore the amplification efficiency can vary when larger series have been tested. The other limitation is that the protocol was only used for genotype 1–4 samples, because we did not have access to genotype 5–7 samples. However, the binding site of the reverse primer Oligo d(A) is conserved in all genotypes, therefore we anticipate that this reverse transcription protocol will also work for other genotypes.

In summary we developed a reverse transcription method for HCV full-genome cDNA synthesis and implemented it into a routine diagnostic setup. This cDNA can be used as template for genotyping amplicons covering the core or NS5B region as well as for resistance testing amplicons covering NS3, NS5A or NS5B. The usage of frozen-mixes is of further convenience and considerably reduces hands-on time.

Conflict of interest

The authors have no conflict of interest.

Author contribution

Acquisition of samples and data (AW, MB, JC, NL, RK), analysis and interpretation of data, drafting of the manuscript (AW, NL, JT), critical revision of the manuscript (AW, NL, JT), study concept, design and supervision (AW, JT).

Acknowledgments

The authors thank Eugen Bäcker and Iris Herrmann for technical help. This work was funded by DFG grant TI 323/4-1. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jcv.2017.04.008.

References

[1] Group HCVIRNS, V.C. Di Maio, V. Cento, I. Lenci, M. Aragri, P. Rossi, S. Barbaliscia, M. Melis, G. Verucchi, C.F. Magni, E. Teti, A. Bertoli, F. Antonucci, M.C. Bellocchi, V. Micheli, C. Masetti, S. Landonio, S. Francioso, F. Santopaolo, A.M. Pellicelli, V. Calvaruso, L. Gianserra, M. Siciliano, D. Romagnoli, R. Cozzolongo, A. Grieco, J. Vecchiet, F. Morisco, M. Merli, G. Brancaccio, A. Di Biagio, E. Loggi, C.M. Mastroianni, V. Pace Palitti, P. Tarquini, M. Puoti, G. Taliani, L. Sarmati, A. Picciotto, V. Vullo, N. Caporaso, M. Paoloni, C. Pasquazzi, G. Rizzardini, G. Parruti, A. Craxi, S. Babudieri, M. Andreoni, M. Angelico, C.F. Perno, F. Ceccherini-Silberstein, Multiclass HCV resistance to direct-acting antiviral failure in real-life patients advocates for tailored second-line therapies, Liver Int. (2017).
[2] N. Chueca, I. Rivadulla, R. Lovatti, G. Reina, A. Blanco, J.A. Fernandez-Caballero, L. Cardenoso, J. Rodriguez-Granjer, M. Fernandez-Alonso, A. Aguilera, M. Alvarez, J.C. Galan, F. Garcia, Using NS5B sequencing for hepatitis C virus genotyping reveals discordances with commercial platforms, PLOS ONE 11 (2016) e0153754.
[3] S. Chevaliez, M. Bouvier-Alias, R. Brillet, J.M. Pawlotsky, Hepatitis C virus (HCV) genotype 1 subtype identification in new HCV drug development and future clinical practice, PLoS ONE 4 (2009) e8209.
[4] S. Larrat, J.D. Poveda, C. Coudret, K. Fusillier, N. Magnat, A. Signori-Schmuck, V. Thibault, P. Morand, Sequencing assays for failed genotyping with the versant hepatitis C virus genotype assay (LiPA), version 2.0, J. Clin. Microbiol. 51 (2013) 2815–2821.
[5] S. De Keukeleire, P. Descheemaeker, M. Reynders, Diagnosis of hepatitis C virus genotype 2k/1b needs NS5B sequencing, Int. J. Infect. Dis.: IJID 41 (2015) 1–2.
[6] V. Kartashev, M. Doring, L. Nieto, E. Coletta, R. Kaiser, S. Sierra, H.C.V.E.S. group, New findings in HCV genotype distribution in selected West European, Russian and Israeli regions, J. Clin. Virol. 81 (2016) 82–89.
[7] J.M. Pawlotsky, Hepatitis C virus resistance to direct-acting antiviral drugs in interferon-free regimens, Gastroenterology (2016).
[8] C. Sarrazin, The importance of resistance to direct antiviral drugs in HCV infection in clinical practice, J. Hepatol. 64 (2016) 486–504.
[9] S. Zeuzem, M. Mizokami, S. Pianko, A. Mangia, K.H. Han, R. Martin, E. Svarovskaia, H. Dvory-Sobol, B. Doehle, C. Hedskog, C. Yun, D.M. Brainard, S. Knox, J.G. McHutchison, M.D. Miller, H. Mo, W.L. Chuang, I. Jacobson, G.J. Dore, M. Sulkowski, NS5A resistance-associated substitutions in patients with genotype 1 hepatitis C virus: prevalence and effect on treatment outcome, J. Hepatol. (2017).
[10] E. Lawitz, S. Flamm, J.C. Yang, P.S. Pang, Y. Zhu, E. Svarovskaia, J.G. McHutchison, D. Wyles, P. Pockros, Retreatment of patients who failed 8 or 12 weeks of ledipasvir/sofosbuvir-based regimens with ledipasvir/sofosbuvir for 24 weeks, J. Hepatol. 62 (2015) S192.
[11] A. Walker, H. Siemann, S. Groten, R.S. Ross, N. Scherbaum, J. Timm, Natural prevalence of resistance-associated variants in hepatitis C virus NS5A in genotype 3a-infected people who inject drugs in Germany, J. Clin. Virol. 70 (2015) 43–45.
[12] D. Hernandez, N. Zhou, J. Ueland, A. Monikowski, F. McPhee, Natural prevalence of NS5A polymorphisms in subjects infected with hepatitis C virus genotype 3 and their effects on the antiviral activity of NS5A inhibitors, J. Clin. Virol. 57 (2013) 13–18.
[13] US Food and Drug Administration, Prescribing Information Zepatier, (2016).
[14] US Food and Drug Administration, Prescribing Information Daklinza, (2016).

[15] P. Kalaghatgi, A.M. Sikorski, E. Knops, D. Rupp, S. Sierra, E. Heger, M. Neumann-Fraune, B. Beggel, A. Walker, J. Timm, H. Walter, M. Obermeier, R. Kaiser, R. Bartenschlager, T. Lengauer, Geno2pheno[HCV] – a web-based interpretation system to support hepatitis C treatment decisions in the era of direct-acting antiviral agents, PLOS ONE 11 (2016) e0155869.
[16] S. Viazov, S.S. Ross, K.K. Kyuregyan, J. Timm, C. Neumann-Haefelin, O.V. Isaeva, O.E. Popova, P.N. Dmitriev, F. El Sharkawi, R. Thimme, M.I. Michailov, M. Roggendorf, Hepatitis C virus recombinants are rare even among intravenous drug users, J. Med. Virol. 82 (2010) 232–238.
[17] S. Viazov, A. Zibert, K. Ramakrishnan, A. Widell, A. Cavicchini, E. Schreier, M. Roggendorf, Typing of hepatitis C virus isolates by DNA enzyme immunoassay, J. Virol. Methods 48 (1994) 81–91.
[18] M. Ruhl, P. Chhatwal, H. Strathmann, T. Kuntzen, D. Bankwitz, K. Skibbe, A. Walker, F.M. Heinemann, P.A. Horn, T.M. Allen, D. Hoffmann, T. Pietschmann, J. Timm, Escape from a dominant HLA-B*15-restricted CD8+ T cell response against hepatitis C virus requires compensatory mutations outside the epitope, J. Virol. 86 (2012) 991–1000.
[19] E.Z. Zhang, D.J. Bartels, J.D. Frantz, S. Seepersaud, J.A. Lippke, B. Shames, Y. Zhou, C. Lin, A. Kwong, T.L. Kieffer, Development of a sensitive RT-PCR method for amplifying and sequencing near full-length HCV genotype 1 RNA from patient samples, Virol. J. 10 (2013) 53.
[20] A. Walker, K. Skibbe, E. Steinmann, S. Pfaender, T. Kuntzen, D.A. Megger, S. Groten, B. Sitek, G.M. Lauer, A.Y. Kim, T. Pietschmann, T.M. Allen, J. Timm, Distinct escape pathway by hepatitis C virus genotype 1a from a dominant CD8+ T cell response by selection of altered epitope processing, J. Virol. 90 (2015) 33–42.
[21] X. Fan, Y. Xu, A.M. Di Bisceglie, Efficient amplification and cloning of near full-length hepatitis C virus genome from clinical samples, Biochem. Biophys. Res. Commun. 346 (2006) 1163–1172.
[22] M. Stangegaard, I.H. Dufva, M. Dufva, Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA, Biotechniques 40 (2006) 649–657.
[23] B. Besse, M. Coste-Burel, N. Bourgeois, C. Feray, B.M. Imbert-Marcille, E. Andre-Garnier, Genotyping and resistance profile of hepatitis C (HCV) genotypes 1–6 by sequencing the NS3 protease region using a single optimized sensitive method, J. Virol. Methods 185 (2012) 94–100.
[24] I. Lindstrom, M. Kjellin, N. Palanisamy, K. Bondeson, L. Wesslen, A. Lannergard, J. Lennerstrand, Prevalence of polymorphisms with significant resistance to NS5A inhibitors in treatment-naive patients with hepatitis C virus genotypes 1a and 3a in Sweden, Infect. Dis. 47 (2015) 555–562.
[25] A.S. Nishiya, C. de Almeida-Neto, S.C. Ferreira, C.S. Alencar, C. Di-Lorenzo-Oliveira, J.E. Levi, N.A. Salles, A. Mendrone Jr., E.C. Sabino, HCV genotypes, characterization of mutations conferring drug resistance to protease inhibitors, and risk factors among blood donors in Sao Paulo, Brazil, PLOS ONE 9 (2014) e86413.

Journal of Clinical Virology 113 (2019) 8–13

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

A pan-genotypic Hepatitis C Virus NS5A amplification method for reliable genotyping and resistance testing

Andreas Walker[a,*], Kim Sophie Ennker[a], Rolf Kaiser[b], Nadine Lübke[a], Jörg Timm[a]

[a] Institute of Virology, Heinrich-Heine-University, University Hospital, Düsseldorf, Germany
[b] Institute of Virology, University of Cologne, Cologne, Germany

ARTICLE INFO

Keywords:
Hepatitis C Virus
Genotype
Antiviral treatment
NS5A
Resistance testing
RAS

ABSTRACT

Background: Chronic infection with the Hepatitis C Virus (HCV) is associated with the risk of progressive liver disease. Although, HCV treatment options and viral cure rates have tremendously increased over the last decade, all currently licensed combination therapies contain inhibitors of the replication complex NS5A. Resistance-associated substitutions (RAS) in NS5A can limit the efficacy of therapy; however, resistance testing is routinely not recommended for all patients. Notably, pan-genotypic combinations have been approved, however the correct identification of the HCV genotype is still required for treatment decisions and is a good predictor for treatment success.

Objective: The aim of this study was the establishment of a pan-genotypic NS5A amplification method for reliable genotyping and simultaneous resistance testing in a fast and cheap routine diagnostic setup.

Study design: Pan-genotypic degenerated nested PCR primer were designed and tested in 262 HCV-patients. The collection included samples from genotypes 1–7 and the median viral load was 1.07×10^6 IU/ml (range 248-21×10^6 IU/ml).

Results: Amplification of the expected 747bp fragment was successful in 257 of 262 (98.1%) samples including samples < 1000 IU/ml. The direct comparison of the genotype information obtained with core sequencing to those obtained by NS5A prediction showed high concordance (97.3%) and discrepancies occurred only for relatively rare subtypes. Resistance analysis using Geno2Pheno$_{[HCV]}$ showed NS5A-RAS in 23 of 257 (8.9%) of samples.

Conclusions: We successfully developed a routine diagnostic method for pan-genotypic amplification of NS5A. This amplicon can be used for simultaneous genotyping and resistance testing for enhancing and improving routine HCV diagnostic.

1. Background

Worldwide about 70 million people are chronically infected with the Hepatitis C Virus (HCV) associated with the risk of progressive liver disease [1,2]. HCV treatment options and viral cure rates (sustained virological response, SVR) have tremendously increased over the last decade, developing from Interferon-based therapies (50% SVR) to Interferon free combinations of direct acting antivirals (DAA) with SVR rates > 95% [3,4]. Current DAA targets are the NS3/4 A protease, the viral NS5A replication-complex and the RNA-dependent RNA-polymerase encoded in NS5B. Meanwhile, pan-genotypic treatment regimens with improved safety profiles and relatively high resistance-barrier are available [5–9]. Notably, inhibitors of the NS5A replication-complex are included in all currently licensed combination therapies.

Resistance-associated substitutions (RAS) can limit the efficacy of DAA therapy and have been associated with increased risk for therapy failure [10–13]. The majority of patients with therapy failure after treatment with NS5A inhibitors have selected RAS in NS5A [14,15]. These RAS persist over prolonged time after therapy and possibly impair SVR rates upon re-therapy [16–18]. In addition to RAS selected during treatment, they are also detected in treatment-naïve patients in relevant frequencies. In studies where viral resistance genotyping was performed, NS5A-RAS were detected in up to 16% at baseline [11,19,20]. Hence, resistance testing for the presence of virus with NS5A-RASbefore therapy is recommended for selected DAA combinations [5,21,22] and may help to optimize treatment decisions in patients with negative predictors for SVR [22] (e.g. patients with cirrhosis or DAA therapy failure or those infected with genotype 3).

* Corresponding author at: Institute for Virology, University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, Universitätsstr. 1, 40225, Düsseldorf, Germany.
E-mail address: Andreas.Walker@med.uni-duesseldorf.de (A. Walker).

https://doi.org/10.1016/j.jcv.2019.01.012
Received 14 October 2018; Received in revised form 15 January 2019; Accepted 29 January 2019

2. Objective

Although treatment combinations with pan-genotypic activity have been approved, the determination of the HCV genotype is still recommended before initiating antiviral therapy [23,24]. Depending on the regimen even correct identification of the viral subtype (genotype 1a or 1b) is still required for treatment decisions and is a predictor for treatment success [24]. Although commercial HCV genotyping assays are available and have a good test performance with a reported specificity up to 93–99% [25–27], sequencing of highly discriminating fragments either from the core or the NS5B region is still considered as the gold standard for HCV genotyping and also permits identification of recombinants [28,29]. Genotyping by sequencing however, requires multiple steps including generation of amplicons, sequencing and genotype prediction and is therefore time consuming and expensive.

As the HCV genotype is required and the NS5A resistance information would be beneficial for treatment decisions we aimed to generate a pan-genotypic NS5A PCR for simultaneous genotyping and resistance testing of NS5A in a fast, reliable and cheap routine diagnostic setup.

3. Study design

3.1. Patient samples

Plasma samples of patients with HCV-infection were obtained in the framework of the German PEPSI study for the surveillance of HCV baseline resistance from Dusseldorf (D) and Cologne (K). Written informed consent was obtained from all study participants and the study was approved by the ethics committee of the Medical Faculties of Dusseldorf (study number 5945R) and Cologne (#2012048). Viral RNA from 400 µl plasma was extracted automatically using the EZ1 Virus Mini Kit v2.0 on an EZ1 Advanced XL robot or manually with the QIAamp Viral RNA Mini Kit (both Qiagen, Hilden, Germany) according to the manufacturer's protocol. RNA was eluted in a volume of 60 µl and stored at −80 °C.

3.2. Reverse transcription

A "primer-mix" containing 1 µl reverse primer Oligo d(A) (50 pmol/µl), 1 µl dNTPs (10 mM each) and 1 µl water were aliquoted in 8-strips with hinged-caps (Eppendorf #951010022) and stored at −20 °C until usage. For reverse transcription, an appropriate number of tubes was thawed and 10 µl RNA was added to the "primer-mix". Secondary RNA-structures were reduced by melting for 5 min at 65 °C before cooling down to 25 °C. RNA was next reverse transcribed in vitro with Superscript III (Invitrogen) as previously described [30] by addition of 7 µl/well reverse transcription mix (4 µl SSIII-Buffer, 1 µl DTT, 1 µl RNase Inhibitor (NEB) and 1 µl SSIII) with the previously described conditions: 10 min at 25 °C, 60 min at 42 °C, 30 min at 50 °C, 30 min 55 °C, 15 min at 75 °C and 4 °C [30,31].

3.3. Nested PCR

Two-step nested PCRs were performed with primer combinations as shown in Table 1 and TaKaRa Ex Taq® DNA Polymerase Hot-Start (TaKaRa) according to the manufacturer's protocol. To reduce hands on time all PCR mixes were prepared in large batches and frozen ("frozen-PCR mixes"). Per well 45 µl PCRI mixture containing 1x TaKaRa Ex Taq polymerase buffer, 200 µM dNTPs (TaKaRa), 0.5 µM each Primer, 1.25 Units Polymerase were aliquoted in 8-strips and stored at −20 °C until usage. PCRII mixes were identical to PCRI except the final volume of 47 µl. For the amplification step an appropriate number of tubes was thawed and five microliter of cDNA were used for the 1 st round of nested-PCR. PCR condition were 180 s at 94 °C followed by 35 cycles each 30 s 95 °C, 30 s 55 °C and 120 s 72 °C followed by 10 min at 72 °C and hold at 10 °C. Subsequently, three microliter of PCR-product from the first round was used for the second round of nested-PCR. PCR conditions were identical to the first round.

3.4. Genotyping and resistance testing

PCR-products from second round of PCR were purified with ExoSAP-IT (Thermo Fisher) or QIAquick PCR Purification Kit (Qiagen, Hilden) and Sanger sequenced (Eurofin Genomics) with sequencing primer NS5A-Seq-F and NS5A-Seq-R (see Table1). Sequences were aligned and a consensus sequence was generated with the software Geneious 10.1.3 (Biomatters, Auckland, New Zealand). The NS5A consensus sequences were then genotyped by phylogenetic analysis with reference sequences suggested by Smith et al. [32]. Moreover, genotype and resistance analysis was also performed using the prediction algorithm geno2pheno$_{[HCV]}$ version 0.92 (http://hcv.bioinf.mpi-inf.mpg.de/index.php) [33]. (32)Genotyping results based on the core region [34] served as a reference for comparisons.

4. Results

4.1. Amplification strategy

To develop an amplification strategy that permits simultaneously genotyping and identification of resistance-associated variants in the NS5A region, degenerate primers covering the N-terminal part of NS5A were designed (Fig. 1A). Therefore, 488 HCV sequences from the Los Alamos Sequence database [35] were aligned and three conserved regions around nt 6069–6146, nt 6810–6851, nt7500-7541 (H77 numbering) were identified. We have previously observed [34] that longer primers consisting of a conserved part without degenerate nucleotides followed by 3′ variable region with degenerate nucleotides have better binding characteristics than shorter primers. Therefore, all primers were designed with a 5′ conserved part consisting of 18–23 nucleotides derived from GT1a followed by a pan-genotypic variable 3′ part with up to 8 degenerate bases (Table 1). The conserved 5′ part enhances primer binding and allows direct sequencing of the amplicon. Due to low amplification efficiency of GT2 (Suppl. Fig. 1) with the original primer set we generated additional PCR-II primers for GT2 (6102-GT2-F and 6854-GT2-R) and included them into the PCR-II mix (Table 1)

4.2. Polymerase selection

Inosine containing primers are not equally accepted by all polymerases [36]. Therefore, we first analyzed the amplification efficiency of different commercially available polymerases. To address this, RNA from GT1a and GT1b infected patients was reverse transcribed with primer Oligo d(A) and the amplification efficiency was analyzed by nested PCRs with the pan-genotypic NS5A primers using four different hot start polymerases; GoTaq Hot Start (Promega), Qiagen HotStar Taq (Qiagen), Platinum HotStart (Thermo Scientific) and TaKaRa Ex Taq (TaKaRa) (Fig. 1B). As seen in Fig. 1B, all polymerases were able to amplify the 747 bp fragment with differences in the performance. Notably, the GoTaq Hot Start and the TaKaRa Ex taq polymerase were the most efficient. Because the TaKaRa Ex Taq polymerase provides the advantage of proofreading activity, this polymerase was used throughout all further experiments.

4.3. Düsseldorf cohort

After validation of the primers and polymerase combinations, 262 HCV-RNA positive patient samples with known genotype from the routine diagnostic were amplified with the pan-genotypic protocol. The genotype was determined by core sequencing, the collection included samples from genotypes 1–7 and the median viral load was 1.07 × 10 [6] IU/ml (range 248-21 × 10 [6] IU/ml). Amplification of the

Table 1
Primer combination used for amplification of NS5A.

Reaction	Primer Name	direction	sequence 5'- > 3'	genotypes
RT	Oligo d(A)	reverse	AAAAAAAAAAAAAAAAAAAA	1-7
PCR I	6069-F	forward	GAGGGGGCAGTGCAATGGATGAAYMGIYTIATIGCITTYGC	1-7
	7542-R	reverse	GAGATCCGGATCCCCAGGYTCICCYTCIAGIGGIGGCATIGA	1,2,3,5,6,7
	7541-R	reverse	AAGTCCGGGTCACCGGGGTCCCCYTCWARYGRAGGCWTTGA	4
PCR II	6105-F	forward	*TTCGCCTCCCGGGGGAAC*CAYGTITCICCIACICAYTAYGT	1,3-7
	6102-GT2-F	forward	GCC*TTCGCCTCCCGGGGGAAC*CAYGTYGCCCCIACICAYTAYGT	2
	6852-R	reverse	*CACGGCTACGTCCGGTTC*IGGITCRCAIGGIAGYTGIGAYCC	1,3-7
	6854-GT2-R	reverse	CAA*CACGGCTACGTCCGGTTC*AGGITCRCARGGRAGYTGIGAICC	2
Sequencing	NS5A-Seq-F	forward	TTCGCCTCCCGGGGGAAC	na
	NS5A-Seq-R	reverse	CACGGCTACGTCCGGTTC	na

Sequencing sites in PCR II primers are indicated in italic.

expected 747 bp fragment was successful in 257 of 262 (98.1%) samples. The five negative samples had all very low viral loads (< 3000IU/ml) and were not enriched for a specific genotype (Table 2).

Notably, a similar pan-genotypic one-step amplification strategy of NS5A was recently published [37], using slightly different primer binding sites. To directly compare our procedure to the published protocol, we first analyzed the amplification efficiency testing titration series of the viral RNA. As seen in Fig. 2, both protocols were able to amplify the NS5A fragment, however, the amplification efficiency and DNA yield was higher with our nested PCR protocol compared to the one-step protocol. For a head-to head comparison on clinical samples a collection of rare genotypes (Suppl. Table 1) and samples < 20.000 IU/ml were amplified side to side. While no differences were observed for the amplification of rare genotypes, amplification and sequencing of samples < 20.000 IU/ml was more effective with our novel protocol compared to the published protocol. With our protocol amplification

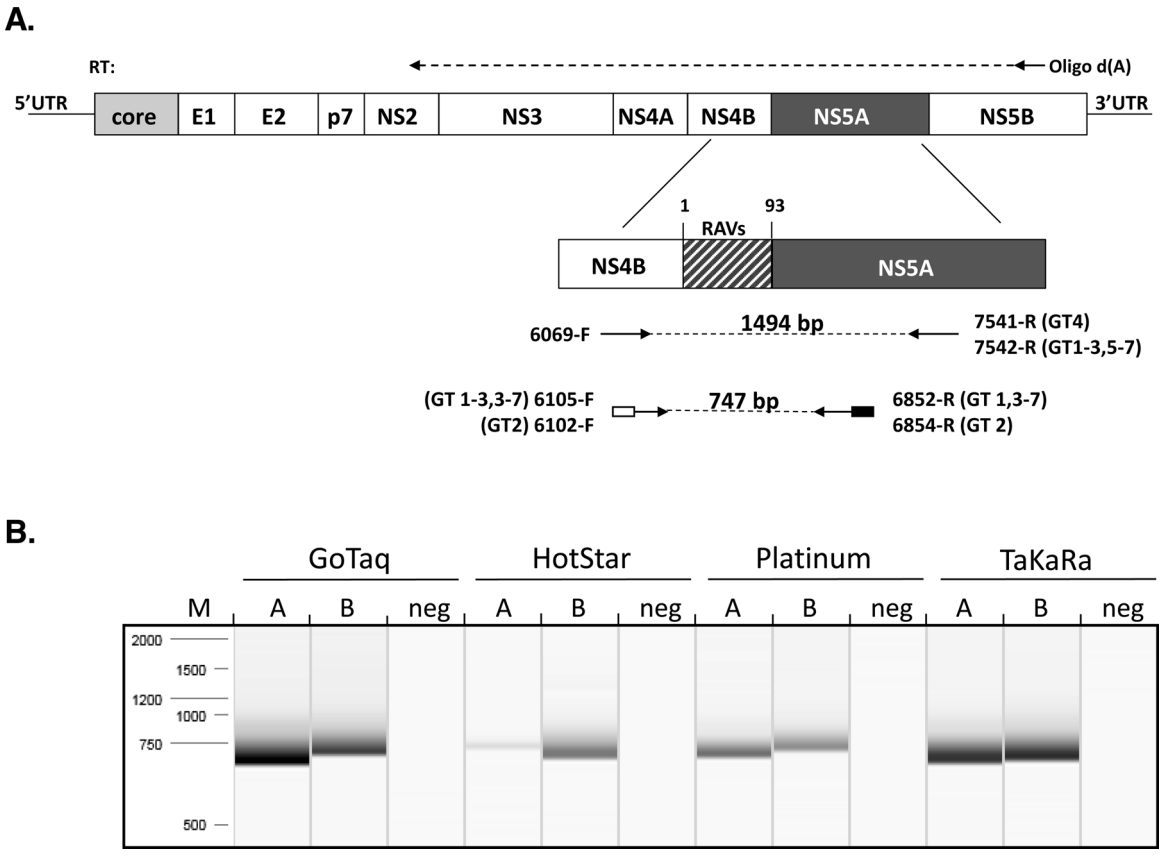


Fig. 1. Rationale for the pan-genotypic NS5A approach.
(A) Scheme of the HCV genome based on the H77 reference genome. The location of reverse transcription primer is depicted above and the nested-PCR primers for genotyping and resistance analysis are depicted below the HCV genome. The NS5A RAV containing region is depicted by horizontal stripes. For enhancement of binding and direct sequencing of the amplicon all primers were designed with a pan-genotypic variable 3′ part with up to 8 degenerate or Inosin bases followed by 18–23 nucleotides derived from GT1a. Amplicons are directly sequenced with primer Seq-F and Seq-R (depicted as white and black box on PCR-II primer, respectively) binding to the GT1a sequence in the pan-genotype primers. (B) Amplification efficiency of commercially available HotStart polymerases. RNA from a GT1a and a GT1b infected patient was extracted using the EZ1 automatic extraction device and reverse transcribed with primer Oligo d(A). cDNA was amplified by nested PCRs with the pan-genotypic NS5A primers using four different Taq polymerases; GoTaq Hot Start (Promega), HotStar Taq (Qiagen), Platinum HotStart (Thermo Scientific) and TaKaRa Ex Taq (TaKaRa). PCR products were separated on a QIAxcel Advanced capillary electrophoresis device using a DNA screening cartridge with a 250bp 4 kb DNA ladder (Qiagen). The expected fragment is 747 bp. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Efficiency of different pan-genotypic NS5A amplification protocols.

Patienten-ID	core Genotyp	viral load	Walker et al.	Andre-Garnier et al.
D16-36079	4a	248	–	–
D17-23546	1b	715	–	–
D17-36276	3a	752	Yes	–
D17-20294	1b	852	–	–
D16-50221	4	948	Yes	Yes
D16-47267	3a	1970	Yes	Yes
K16-15343	4a	2288	–	–
D16-57092	3a	2405	Yes	–
D17-17585	1a	2797	–	–
D16-41860	3a	3477	Yes	Yes
D17-21569	1b	4568	Yes	Yes
D17-24705	1b	6213	Yes	Yes
D17-38540	1a	9978	Yes	–
D17-2015	1b	11254	Yes	Yes
D16-53895	3a	11675	Yes	–
D17-37507	2a	14464	Yes	Yes

was successful in 11 out of 16 (68.75%) samples, while the one-step protocol was only able to amplify 7 out of 16 (43.75%) samples (Table 2).

4.4. Genotyping

In principle, sequencing of NS5A would allow both genotyping and resistance prediction, provided that genotyping by NS5A is as reliable as core sequencing. To analyze this, NS5A amplicons were bulk sequenced with primers listed in Table 1, aligned with the HCV reference dataset [32] and the genotype was resolved by a phylogenetic analysis (Fig. 3). For comparison, the core sequences from the routine diagnostic were also aligned and resolved by a phylogenetic tree. The comparison showed identical genotypes in 250 of 257 (97.3%) (Table 3). Three samples with discrepant results were true 2k/1b variants with 2k in core and 1b in NS5A. The remaining four discrepant samples were assigned with the same genotype but the subtypes were predicted differently (core/NS5A: 1i/1d, 2b/2k, 2q/2f and 2k/2c). All discrepancies occurred for relatively rare subtypes with limited data available in sequence databases. With the exception of correct identification of recombinants, genotyping based on NS5A seemed to be reliable and as good as genotyping based on core.

Since simultaneous genotyping and resistance prediction using a prediction algorithm like Geno2Pheno$_{[HCV]}$ [33] would be faster and more comfortable, the NS5A sequences were also used to predict the viral genotype with Geno2Pheno$_{[HCV]}$. Compared to the phylogenetic analysis of the core fragment, Geno2Pheno$_{[HCV]}$ predicted for the NS5A fragment correct genotypes in 248 of 257 (96.5%). The discrepancies included again the three recombinant 2k/1b variants and the four samples that were already discrepant in the NS5A tree. Two additional samples were 1 h in the phylogenetic analysis of the core and the NS5A fragment but predicted as 1a by Geno2Pheno$_{[HCV]}$. Notably, for both samples a low sequence homology to the reference strain was reported by Geno2Pheno$_{[HCV]}$ (79% and 80%). Taken together, genotype prediction based on NS5A is reliable and prediction by Geno2Pheno is nearly as reliable as phylogenetic analysis. In case of low sequence identity with the reference sequence reported by Geno2Pheno$_{[HCV]}$, correct genotyping needs validation by phylogenetic analysis.

Resistance associated substitutions (RAS) Prediction of RAS was performed with Geno2Pheno$_{[HCV]}$. Overall, the frequency of RAS was low with 23 of 257 (8.9%) patients harboring RAS. RAS were detected in different genotypes including GT1a/b, GT3a, GT4a/d and a GT1 and a GT3 isolate with unassigned subtype. Notably, four of the 23 patients carrying RAS were previously treated with a DAA combination containing an NS5A inhibitor.

5. Discussion

Even in the era of pan-genotypic HCV treatment, the HCV genotype is still an important diagnostic parameter. Genotyping by sequencing is the most reliable genotyping method, especially for rare genotypes and recombinants [28,38]. A major challenge for a NS5A pan-genotypic PCR protocols is the enormous sequence diversity between genotypes and subtypes that hinders identification of ideal primer binding sites for efficient and unbiased amplification of all genotypes. Here, we developed and optimized a nested PCR that allows successful amplification of all genotypes and subtypes even from samples with low viral HCV-RNA concentrations.

Sequencing of the NS5A region produces highly reliable genotype information as seen by the high concordance of our NS5A genotype prediction with the core genotyping. This is in line with previous finding [37,39]. The four samples with different subtypes in core/NS5A all belonged to rare or untyped sub-genotypes and most likely present genotyping inaccuracy rather than real discrepancies. Genotype prediction by Geno2Pheno was also highly specific. Again, discrepancies were only detected for rare subtypes and were marked by low sequence identity with the reference sequence (Table 3). Thus, genotyping by NS5A allows precise genotyping and additionally has the advantage, that recombinant viral variants are correctly genotyped in the relevant DAA target region. Genotyping with Geno2Pheno has the advantage of being fast, comfortable and simultaneously determines RAS.

The clinical relevance of resistance mutations for treatment of HCV genotype with DAAs is still not clear [40]. RAS are frequently detected in treatment naïve patients, however, lower SVR rates of patients with RAS compared to patients without RAS are only detected in GT1a and GT3 infected patients [41,42] whereas baseline RAS in GT1b are not of clinical importance [43,44]. The overall frequency of baseline RAS (8.9%) was lower compared to other studies [11,45,46]; however the frequency of RAS in GT3 was comparable to data from other West-

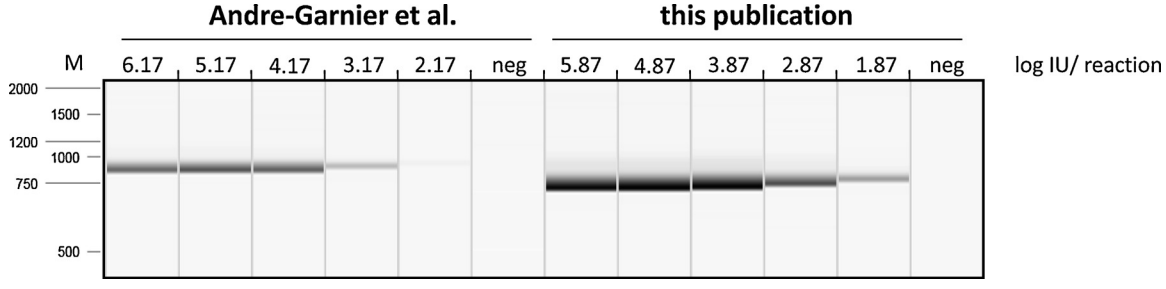


Fig. 2. Comparison of different pan-genotypic approach for NS5A amplification.
Efficiency of NS5A amplification was analyzed by serial dilution of RNA (22 × 10 [7] IU/ml) and amplification with the indicated protocols. PCR-aliquots were separated on a QIAxcel Advanced System. M. QX DNA Size Marker 250 bp – 4 kb (Qiagen). As the protocols use different amounts of input RNA the viral load per reaction is indicated for better estimation of amplification efficiency. Of note, both protocols amplify the lowest dilution (equal to 1000 IU/ml plasma), however with different efficacy.

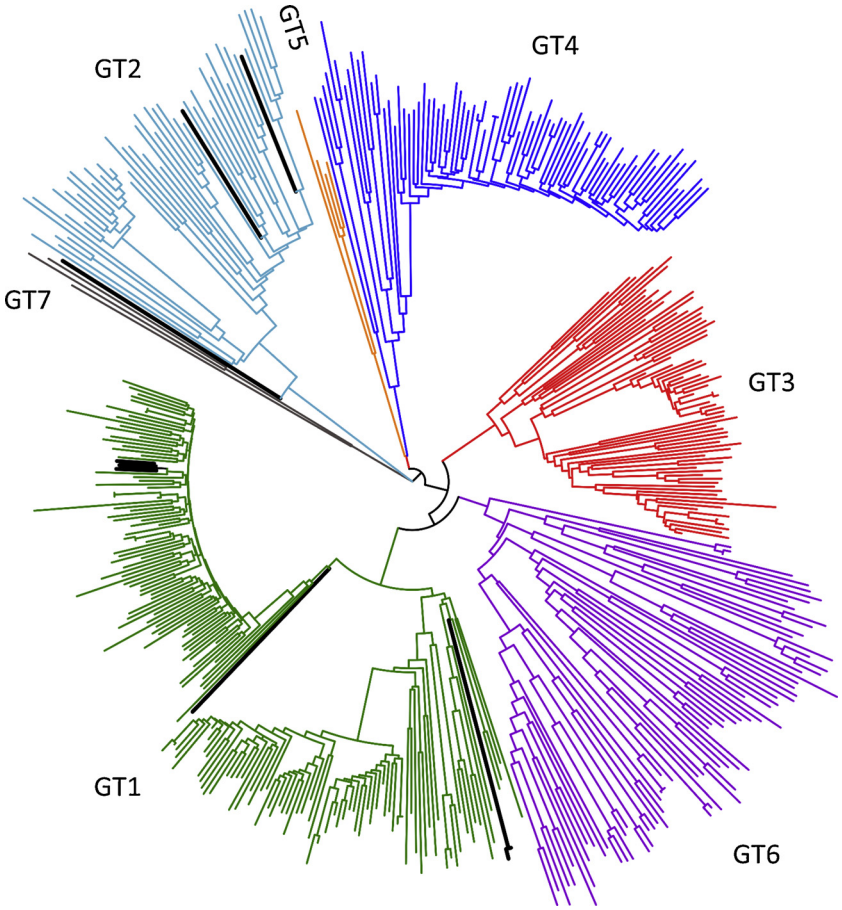


Fig. 3. Phylogenetic tree of NS5A sequences.
The NS5A sequences were aligned with the reference genomes provided by Smith et al [32] and a phylogenetic tree was generated by the Neighbor-Joining method using the Tamura-Nei genetic distance model. The tree was generated using Geneious R.10 and visulized using iTOL3 [47]. The branches of disconcordant sequences are marked in **black**. The tree is color coded according to the genotype.

Table 3
Discrepant samples.

sample number	core tree	NS5A tree	Geno2Pheno
D16-56229	2k	1b	1b (91%)
D17-32430	2k	1b	1b (90%)
D17-44030	2k	1b	1b (91%)
D17-55572	1i	1d	1b (81%)
D17-17662	1h	1h	1a (79%)
D17-51295	1h	1h	1a (80%)
D16-28550	2b	2t	2f (74%)
D17-32482	2q	2f	2f (85 %)
K15-15437	2k	2c	2c (83%)

German cohorts [19]. Most polymorphisms were found in GT1b and here the tyrosine to histidine at position 93 alone or in combination with polymorphisms at position 31 was predominant.

Recently, another pan-genotypic approach for NS5A genotyping [37] with slightly different primer binding sites using a modified one-step PCR protocol was published. Both protocols showed comparable results for samples with high viremia, however the two-step PCR was more efficient in amplification of samples < 10.000 IU/ml. Notably, the one-step protocol by Andre-Garnier requires the expensive T4 gene32 in the RT-reaction for efficient amplification of samples with low virus concentrations. Finally, with medium to low virus concentrations the one-step protocol yielded only low amounts of PCR product that did not permit sequencing.

In this study we only analyzed RAS from amplicons amplified with the pangenotypic protocol. For formally correct analysis, RAS detection should be compared to the detection rate with genotype specific amplification. However, resistance analysis was not the primary focus of this work and from the 10 out of 23 patients with RAS where we have genotype specific amplicons there was no discrepancy. The other limitation is that we had only access to a limited number of GT5, 6 and 7 samples. Since the binding site of the pangenotypic primer are relative conserved between the different sub/genotypes, we anticipate that this protocol will also work for other genotypes.

Taken together this protocol is an improvement to the currently available protocols and permits genotyping and resistance prediction in the most important RAV region NS5A. The pan-genotypic approach reduces hands on time and costs by 50%. Moreover, besides genotype information the clinician also gets information on baseline RAS even in patients that normally would not be screened for RAS. Finally, reverse transcription with the poly d(A) primer yield whole HCV genome cDNA [30] that also allows cheap and easy re-amplification of other HCV regions.

Conflict of interest

The authors have no conflict of interest.

Author contributions

Acquisition of samples and data (AW, KSE, NL, RK), analysis and interpretation of data, drafting of the manuscript (AW, JT) critical revision of the manuscript (AW, NL, JT), study concept, design and supervision (AW, JT).

Acknowledgments

The authors thank Alexandra Graupner and Jennifer Camdereli for technical help. This work was funded by DFG grant TI 323/4-1 and

intramural funding. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jcv.2019.01.012.

References

[1] Polaris Observatory HCVC, Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study, Lancet Gastroenterol. Hepatol. 2 (2017) 161–176.
[2] A.P. Thrift, H.B. El-Serag, F. Kanwal, Global epidemiology and burden of HCV infection and HCV-related disease, Nat. Rev. Gastroenterol. Hepatol. 14 (2017) 122–132.
[3] A. Kohli, A. Shaffer, A. Sherman, S. Kottilil, Treatment of hepatitis C: a systematic review, JAMA 312 (2014) 631–640.
[4] N. Alkhouri, E. Lawitz, F. Poordad, Novel treatments for chronic hepatitis C: closing the remaining gaps, Curr. Opin. Pharmacol. 37 (2017) 107–111.
[5] US Food and Drug Administration, Prescribing Information Vosevi, (2017).
[6] US Food and Drug Administration, Prescribing Information Mavyret, (2017).
[7] Y.N. Lamb, Glecaprevir/Pibrentasvir: first global approval, Drugs 77 (16) (2017) 1797–1804, https://doi.org/10.1007/s40265-017-0817-y.
[8] R. Voaklander, I.M. Jacobson, Sofosbuvir, velpatasvir and voxilaprevir combination for the treatment of hepatitis C, Expert Rev. Gastroenterol. Hepatol. 11 (2017) 789–795.
[9] European Association for the Study of the Liver, EASL recommendations on treatment of hepatitis C 2016, Electronic address eee. J. Hepatol. 66 (2017) 153–194.
[10] J.M. Pawlotsky, Hepatitis C Virus resistance to direct-acting antiviral drugs in interferon-free regimens, Gastroenterology 151 (1) (2016) 70–86, https://doi.org/10.1053/j.gastro.2016.04.003.
[11] C. Sarrazin, The importance of resistance to direct antiviral drugs in HCV infection in clinical practice, J. Hepatol. 64 (2016) 486–504.
[12] S. Zeuzem, M. Mizokami, S. Pianko, A. Mangia, K.H. Han, R. Martin, E. Svarovskaia, H. Dvory-Sobol, B. Doehle, C. Hedskog, C. Yun, D.M. Brainard, S. Knox, J.G. McHutchison, M.D. Miller, H. Mo, W.L. Chuang, I. Jacobson, G.J. Dore, M. Sulkowski, NS5A resistance-associated substitutions in patients with genotype 1 Hepatitis C Virus: prevalence and effect on treatment outcome, J. Hepatol. 66 (5) (2017) 900–918, https://doi.org/10.1016/j.jhep.2017.01.007.
[13] R. Esteban, J.A. Pineda, J.L. Calleja, M. Casado, M. Rodriguez, J. Turnes, L.E. Morano Amado, R.M. Morillas, X. Forns, J.M. Pascasio Acevedo, R.J. Andrade, A. Rivero, J.A. Carrion, S. Lens, M. Riveiro-Barciela, B. McNabb, G. Zhang, G. Camus, L.M. Stamm, D.M. Brainard, G.M. Subramanian, M. Buti, Efficacy of Sofosbuvir and Velpatasvir, with and without ribavirin, in patients with HCV genotype 3 infection and cirrhosis, Gastroenterology 155 (4) (2018) 1120–1127, https://doi.org/10.1053/j.gastro.2018.06.042.
[14] S. Viazov, A. Zibert, K. Ramakrishnan, A. Widell, A. Cavicchini, E. Schreier, M. Roggendorf, Typing of Hepatitis C Virus isolates by DNA enzyme immunoassay, J. Virol. Methods 48 (1994) 81–91.
[15] D. Wyles, H. Dvory-Sobol, E.S. Svarovskaia, B.P. Doehle, R. Martin, N.H. Afdhal, K.V. Kowdley, E. Lawitz, D.M. Brainard, M.D. Miller, H. Mo, E.J. Gane, Post-treatment resistance analysis of Hepatitis C Virus from phase II and III clinical trials of Ledipasvir/Sofosbuvir, J. Hepatol. 66 (2017) 703–710.
[16] E. Lawitz, S. Flamm, J.C. Yang, P.S. Pang, Y. Zhu, E. Svarovskaia, J.G. McHutchison, D. Wyles, P. Pockros, Retreatment of patients who failed 8 or 12 weeks of Ledipasvir/Sofosbuvir- based regimens with Ledipasvir/Sofosbuvir for 24 weeks, J. Hepatol. 62 (2015) S192-S192.
[17] C. Sarrazin, H. Dvory-Sobol, E.S. Svarovskaia, B.P. Doehle, P.S. Pang, S.M. Chuang, J. Ma, X. Ding, N.H. Afdhal, K.V. Kowdley, E.J. Gane, E. Lawitz, D.M. Brainard, J.G. McHutchison, M.D. Miller, H. Mo, Prevalence of resistance-associated substitutions in HCV NS5A, NS5B, or NS3 and outcomes of treatment with Ledipasvir and Sofosbuvir, Gastroenterology 151 (2016) 501–512 e501.
[18] S. Zeuzem, M. Mizokami, S. Pianko, A. Mangia, K.H. Han, R. Martin, E. Svarovskaia, H. Dvory-Sobol, B. Doehle, C. Hedskog, C. Yun, D.M. Brainard, S. Knox, J.G. McHutchison, M.D. Miller, H. Mo, W.L. Chuang, I. Jacobson, G.J. Dore, M. Sulkowski, NS5A resistance-associated substitutions in patients with genotype 1 Hepatitis C Virus: prevalence and effect on treatment outcome, J. Hepatol. 66 (2017) 910–918.
[19] A. Walker, H. Siemann, S. Groten, R.S. Ross, N. Scherbaum, J. Timm, Natural prevalence of resistance-associated variants in Hepatitis C Virus NS5A in genotype 3a-infected people who inject drugs in Germany, J. Clin. Virol. 70 (2015) 43–45.
[20] D. Hernandez, N. Zhou, J. Ueland, A. Monikowski, F. McPhee, Natural prevalence of NS5A polymorphisms in subjects infected with Hepatitis C Virus genotype 3 and their effects on the antiviral activity of NS5A inhibitors, J. Clin. Virol. 57 (2013) 13–18.
[21] US Food and Drug Administration, Prescribing Information Zepatier, (2016).
[22] EASL, EASL recommendations on treatment of hepatitis C 2016, J. Hepatol. 66 (2017) 153–194.
[23] AASLD, Recommendations for Testing, Managing, and Treating Hepatitis C, [last Accessed January 2019] (2019).
[24] EASL, EASL recommendations on treatment of hepatitis C 2018, J. Hepatol. 69 (2) (2018) 461–511, https://doi.org/10.1016/j.jhep.2018.03.026.
[25] S. Chevaliez, M. Bouvier-Alias, R. Brillet, J.M. Pawlotsky, Hepatitis C Virus (HCV) genotype 1 subtype identification in new HCV drug development and future clinical practice, PLoS One 4 (2009) e8209.
[26] N. Chueca, I. Rivadulla, R. Lovatti, G. Reina, A. Blanco, J.A. Fernandez-Caballero, L. Cardenoso, J. Rodriguez-Granjer, M. Fernandez-Alonso, A. Aguilera, M. Alvarez, J.C. Galan, F. Garcia, Using NS5B sequencing for Hepatitis C Virus genotyping reveals discordances with commercial platforms, PLoS One 11 (2016) e0153754.
[27] S. Larrat, J.D. Poveda, C. Coudret, K. Fusillier, N. Magnat, A. Signori-Schmuck, V. Thibault, P. Morand, Sequencing assays for failed genotyping with the versant Hepatitis C Virus genotype assay (LiPA), version 2.0, J. Clin. Microbiol. 51 (2013) 2815–2821.
[28] V. Kartashev, M. Doring, L. Nieto, E. Coletta, R. Kaiser, S. Sierra, group HCVES, New findings in HCV genotype distribution in selected West European, Russian and Israeli regions, J. Clin. Virol. 81 (2016) 82–89.
[29] S. De Keukeleire, P. Descheemaeker, M. Reynders, Diagnosis of Hepatitis C Virus genotype 2k/1b needs NS5B sequencing, Int. J. Infect. Dis. 41 (2015) 1–2.
[30] A. Walker, M. Bergmann, J. Camdereli, R. Kaiser, N. Lubke, J. Timm, A genotype independent, full-genome reverse-transcription protocol for HCV genotyping and resistance testing, J. Clin. Virol. 91 (2017) 42–48.
[31] E.Z. Zhang, D.J. Bartels, J.D. Frantz, S. Seepersaud, J.A. Lippke, B. Shames, Y. Zhou, C. Lin, A. Kwong, T.L. Kieffer, Development of a sensitive RT-PCR method for amplifying and sequencing near full-length HCV genotype 1 RNA from patient samples, Virol. J. 10 (53) (2013).
[32] D.B. Smith, J. Bukh, C. Kuiken, A.S. Muerhoff, C.M. Rice, J.T. Stapleton, P. Simmonds, Expanded classification of Hepatitis C Virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource, Hepatology 59 (2014) 318–327.
[33] P. Kalaghatgi, A.M. Sikorski, E. Knops, D. Rupp, S. Sierra, E. Heger, M. Neumann-Fraune, B. Beggel, A. Walker, J. Timm, H. Walter, M. Obermeier, R. Kaiser, R. Bartenschlager, T. Lengauer, Geno2pheno[HCV] – a web-based interpretation system to support hepatitis C treatment decisions in the era of direct-acting antiviral agents, PLoS One 11 (2016) e0155869.
[34] S. Viazov, S.S. Ross, K.K. Kyuregyan, J. Timm, C. Neumann-Haefelin, O.V. Isaeva, O.E. Popova, P.N. Dmitriev, F. El Sharkawi, R. Thimme, M.I. Michailov, M. Roggendorf, Hepatitis C Virus recombinants are rare even among intravenous drug users, J. Med. Virol. 82 (2010) 232–238.
[35] C. Kuiken, K. Yusim, L. Boykin, R. Richardson, The Los Alamos hepatitis C sequence database, Bioinformatics 21 (2005) 379–384.
[36] T. Knittel, D. Picard, PCR with degenerate primers containing deoxyinosine fails with Pfu DNA polymerase, PCR Methods Appl. 2 (1993) 346–347.
[37] E. Andre-Garnier, B. Besse, A. Rodallec, O. Ribeyrol, V. Ferre, C. Luco, L. Le Guen, N. Bourgeois, J. Gournay, E. Billaud, F. Raffi, M. Coste-Burel, B.M. Imbert-Marcille, An NS5A single optimized method to determine genotype, subtype and resistance profiles of Hepatitis C strains, PLoS One 12 (2017) e0179562.
[38] E. Knops, E. Heger, C. Koenig, U. Moebius, N. Lubke, R. Kaiser, V. Di Cristanziano, L. Husgen, C. Kocyigit, J. Rupp, S. Sierra, Accurate Hepatitis C Virus genotyping and selection of optimal therapy: lessons from a St Petersburg strain infection, Clin. Microbiol. Infect. 24 (2018) 440–441.
[39] S. Mansoor, A. Javed, A. Ali, A. Mansoor, Heterogeneous genomic locations within NS3, NS4A and NS4B identified for genotyping and subtyping of Hepatitis C Virus: a simple genome analysis approach, Infect. Genet. Evol. 44 (2016) 61–68.
[40] A. Walker, R. Kaiser, R.J.T. Bartenschlager, Genotypic resistance testing of HCV – is there a clinical need? GMS Infect. Dis. 2016 (4) (2016) Doc05.
[41] P. Krishnan, G. Schnell, R. Tripathi, J. Beyer, T. Reisch, X. Zhang, C. Setze, L. Rodrigues Jr., M. Burroughs, R. Redman, K. Chayama, H. Kumada, C. Collins, T. Pilot-Matias, Analysis of Hepatitis C Virus genotype 1b resistance variants in japanese patients treated with paritaprevir-ritonavir and ombitasvir, Antimicrob. Agents Chemother. 60 (2016) 1106–1113.
[42] M.P. Curry, J.G. O'Leary, N. Bzowej, A.J. Muir, K.M. Korenblat, J.M. Fenkel, K.R. Reddy, E. Lawitz, S.L. Flamm, T. Schiano, L. Teperman, R. Fontana, E. Schiff, M. Fried, B. Doehle, D. An, J. McNally, A. Osinusi, D.M. Brainard, J.G. McHutchison, R.S. Brown Jr., M. Charlton, A.- Investigators, Sofosbuvir and Velpatasvir for HCV in patients with decompensated cirrhosis, N. Engl. J. Med. 373 (2015) 2618–2628.
[43] J.J. Feld, I.M. Jacobson, C. Hezode, T. Asselah, P.J. Ruane, N. Gruener, A. Abergel, A. Mangia, C.L. Lai, H.L. Chan, F. Mazzotta, C. Moreno, E. Yoshida, S.D. Shafran, W.J. Towner, T.T. Tran, J. McNally, A. Osinusi, E. Svarovskaia, Y. Zhu, D.M. Brainard, J.G. McHutchison, K. Agarwal, S. Zeuzem, Investigators A.-, Sofosbuvir and Velpatasvir for HCV genotype 1, 2, 4, 5, and 6 infection, N. Engl. J. Med. 373 (2015) 2599–2607.
[44] M. Mizokami, O. Yokosuka, T. Takehara, N. Sakamoto, M. Korenaga, H. Mochizuki, K. Nakane, H. Enomoto, F. Ikeda, M. Yanase, H. Toyoda, T. Genda, T. Umemura, H. Yatsuhashi, T. Ide, N. Toda, K. Nirei, Y. Ueno, Y. Nishigaki, J. Betular, B. Gao, A. Ishizaki, M. Omote, H. Mo, K. Garrison, P.S. Pang, S.J. Knox, W.T. Symonds, J.G. McHutchison, N. Izumi, M. Omata, Ledipasvir and Sofosbuvir fixed-dose combination with and without ribavirin for 12 weeks in treatment-naive and previously treated Japanese patients with genotype 1 hepatitis C: an open-label, randomised, phase 3 trial, Lancet Infect. Dis. 15 (2015) 645–653.
[45] N. Palanisamy, P. Kalaghatgi, D. Akaberi, A. Lundkvist, Z.W. Chen, P. Hu, J. Lennerstrand, Worldwide prevalence of baseline resistance-associated polymorphisms and resistance mutations in HCV against current direct-acting antivirals, Antivir. Ther. (23) (2018) 485–493, https://doi.org/10.3851/IMP3237.
[46] Y. Hirotsu, T. Kanda, H. Matsumura, M. Moriyama, O. Yokosuka, M. Omata, HCV NS5A resistance-associated variants in a group of real-world Japanese patients chronically infected with HCV genotype 1b, Hepatol. Int. 9 (2015) 424–430.
[47] I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, Nucleic Acids Res. 44 (2016) W242–245.

Walker et al. Virology Journal (2017) 14:106
DOI 10.1186/s12985-017-0779-4

Virology Journal

CASE REPORT Open Access

Detection of a genetic footprint of the sofosbuvir resistance-associated substitution S282T after HCV treatment failure

Andreas Walker[1], Sandra Filke[2], Nadine Lübke[1], Martin Obermeier[3], Rolf Kaiser[4], Dieter Häussinger[2], Jörg Timm[1*] and Hans H. Bock[2]

Abstract

Background: The major resistance-associated substitution for sofosbuvir (S282T) in HCV NS5B causes severe viral fitness costs and rapidly reverts back to prototype in the absence of selection pressure. Accordingly, resistance against sofosbuvir is rarely detected even in patients after treatment failure.

Case presentation: We report a case of a GT3a infected patient with viral breakthrough under SOF/DCV therapy. At the time of breakthrough the RAS S282T was predominant in NS5B and then rapidly disappeared during follow-up by week 12 after treatment. Interestingly, despite only serine was encoded in position 282 during follow-up, two distinct genetic pathways for reversion were detectable. In 31% of the quasispecies the original codon for serine was present whereas in the majority of the quasispecies an alternative codon was selected. This alternative codon usage was unique for all GT3a isolates from the HCV database and remained detectable as a genetic footprint for prior resistance selection at the RNA level for at least 6 months.

Conclusions: Comparative analyses of viral sequences at the codon level before and after DAA treatment may help to elucidate the patient's history of resistance selection, which is particularly valuable for highly unfit substitutions that are detectable only for a short period of time. If such codon changes increase the risk of re-selection of resistance upon a second exposure to SOF remains to be addressed.

Keywords: Sofosbuvir, Daclatasvir, Resistance, Genotype 3a, Breakthrough, S282T, MEDCOAT

Background

During the last few years, several directly acting antivirals (DAAs) became available for treatment of chronic hepatitis C, starting a new era of therapy of hepatitis C virus (HCV) infection. DAAs were typically optimized for inhibition of viral replication of HCV genotype 1a and 1b and are in most cases less active against genotype 3a. Currently, only a combination of the nucleoside analogue sofosbuvir (SOF) and one of the NS5A inhibitors Velpatasvir (VEL) or daclatasvir (DCV) are recommended for treatment of patients infected with HCV genotype 3a [1]. Depending on the fibrosis stage, prior treatment experience and the duration of therapy, sustained viral response rates (SVR) between 63 and 96% have been reported [2, 3]. In clinical studies resistance-associated substitutions (RAS) were detected in the majority of patients who failed to achieve SVR (reviewed in [4, 5]). In Genotype 3a the most common NS5A-RAS are Y93H, A30K and L31I [6]. In the ALLY-3 trial NS5A-Y93H RAS was present in 15 of 16 patients with relapse, whereas NS5B RASs associated with resistance to SOF (aa159, 282, or 321) were not detected [2]. Up to date the major SOF resistance associated substitution S282T was only found in few patients with viral relapse [7–9].

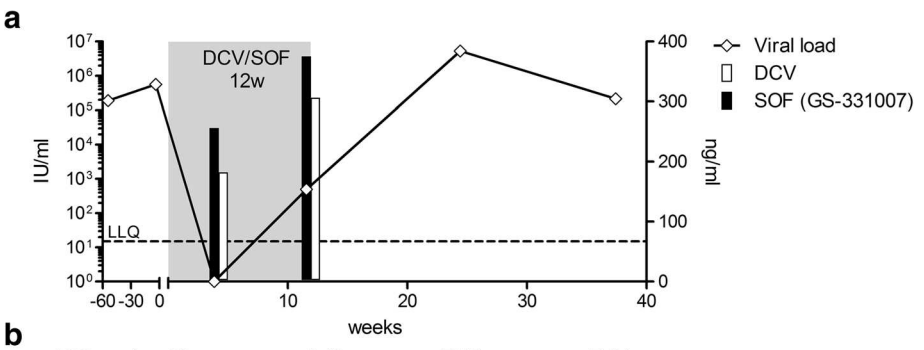
* Correspondence: joerg.timm@med.uni-duesseldorf.de
[1]Institute for Virology, Heinrich-Heine-University, University Hospital, Universitätsstr. 1, 40225 Düsseldorf, Germany
Full list of author information is available at the end of the article

Case presentation

We report here a case of a GT3a infected treatment naïve patient with viral breakthrough under SOF/DCV therapy. The patient was a 59 year old non-cirrhotic, HIV-negative woman treated at the University Hospital Düsseldorf, Germany. In accordance with the German and international guidelines treatment was initiated with 400 mg Sofosbuvir and 60 mg Daclatasvir once daily. Viral load at baseline was 562.530 IU/ml and dropped below the detection limit (<15 IU/ml) at week 4 (Fig. 1). Two days prior to the end of treatment at week 12 HCV-RNA was detectable again at low levels (493 IU/ml) consistent with a viral breakthrough. During follow-up 12 weeks after the end of treatment viral load further increased to high levels (5.290.000 IU/ml). Notably, due to difficulties with swallowing of tablets the patient had used MEDCOAT® tablet coating for assistance. Plasma drug levels at week 4 and at breakthrough were 182 ng/ml and 306 ng/ml for DCV and 256 ng/ml and 375 ng/ml for the SOF metabolite GS-331007 (Fig. 1). Amplicon sequencing [10] and resistance phenotyping using Geno2Pheno [HCV] (http://hcv.bioinf.mpi-inf.mpg.de/index.php; [11]) showed no RAS in NS5A and NS5B at baseline. Resistance testing at week 12 (day 82), when the patient was still under treatment, revealed the RAS S282T in NS5B and Y93H in NS5A (Fig. 1). Y93H persisted in 68.5% of the quasispecies during follow-up at week 12 and continued to revert back to nearly undetectable levels at follow-up at week 24. The substitution S282T had already fully reverted back to prototype at follow-up at week 12 consistent with severe fitness costs associated with this RAS. Interestingly, two distinct pathways were selected for reversion to the prototype residue. Ultra-deep sequencing revealed that 31% of the quasispecies reverted back to the original codon for serine (AGT) whereas in 68% of the quasispecies an alternative codon for serine (TCT) was selected. This alternative codon usage was unique and was absent from 114 GT3a isolates analyzed in our laboratory and from all 489 reference sequences from the HCV database. In line with negative selection of alternative codon usage in this position, the frequency of variants in the quasispecies utilizing the alternative codon further decreased to 24.2% during follow up at week 24.

Discussion and conclusions

We report here, to our knowledge, the first case of selection of the RAS S282T in a patient infected with genotype GT3a. Selection of the RAS was associated with a viral breakthrough detected at week 12, when the plasma drug levels were in a low therapeutic range. If the low plasma drug levels are a consequence of the MEDCOAT® coating of the tablets is unknown since no peer reviewed data on the pharmacokinetics are available. However, it has been described that co-medication with proton pump inhibitors (PI) during HCV-treatment is a risk factor for treatment failure [12]. Accordingly, it seems plausible that the usage of MEDCOAT® might be a risk factor for treatment failure as well.



b

RAS	BL	week 12	FU12	FU24
NS5A-Y93H	0.17%	99.6%	68.5%	0.20%
NS5B-S282T	0%	99.8%	0.07%	0.02%

c

codon usage pos. 282				
S282 (AGT)	98.2%	0%	31.2%	75.0%
S282 (AGC)	1.1%	0%	0.07%	0.25%
S282 (TCT)	0%	0%	68.0%	24.2%
T282 (ACT)	0%	99.8%	0.07%	0.02%

Fig. 1 Therapy failure of a patient with HCV GT3a infection treated with SOF/DCV for 12 weeks. **a** HCV viral load (black line), time of treatment (gray box) and plasma drug levels for DCV (white bar) and the SOF metabolite GS-331007 (black bar) are shown. **b** Frequency of the RAS NS5A-Y93H and NS5B-S282T at baseline (BL), end of treatment (week 12) and follow-up (FU). **c** Codon usage at NS5B position 282 as determined by ultra-deep sequencing

Although reversion of NS5A RAS has been described, in the majority (85–94%) of genotype 1 infected patients NS5A RAS persist in the viral population over years after treatment [13–15]. For genotype 3a no systematic analysis of the destiny of NS5A RAS after treatment failure is available. In line with previous reports the S282T substitution rapidly reverted back to prototype by week 12 after treatment [5]. Interestingly, despite only serine was encoded in position 282 at that time, the codon usage had changed leaving a detectable genetic footprint of prior resistance selection that was unique for GT3a isolates. This genetic footprint of prior resistance selection remained detectable at the RNA level for at least 6 months, consistent with lower fitness costs compared to the original RAS. We conclude that comparative analyses of viral sequences at the codon level before and after DAA treatment may help to elucidate the patient's history of resistance selection, which is particularly valuable for highly unfit RAS that are detectable only for a short period of time. If such codon changes increase the risk of re-selection of resistance upon a second exposure to SOF is unknown. The patient reported here was not re-treated until now, as no advanced fibrosis was yet detectable.

Abbreviations
DCV: Daclatasvir; HCV: Hepatitis C Virus; RAS: resistance-associate substitution; SOF: Sofosbuvir

Acknowledgements
We thank Iris Herrmann and Jennifer Camdereli for excellent technical help.

Funding
This work was funded by DFG grant TI 323/4–1

Availability of data and materials
The datasets used and analyzed in the current study are available from the corresponding author on reasonable request.

Authors' contributions
Acquisition of data (AW, SF, NL, MO, RK, JT), study concept and design (DH, JT, HB), analysis and interpretation of data (AW, NL, JT, HB), drafting of the manuscript (AW, JT, HB) critical revision of the manuscript (AW, SF, NL, MO, RK, DH, JT,HB), study supervision (JT,DH,HB). All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Obtained from the patient.

Ethics approval and consent to participate
Informed consent was obtained from the patient, and the study protocol was approved by the ethics committee of the Medical Faculty of the University of Düsseldorf (Ethic #2012048) in accordance with the guidelines of the Declaration of Helsinki.

Publisher's Note

Author details
[1]Institute for Virology, Heinrich-Heine-University, University Hospital, Universitätsstr. 1, 40225 Düsseldorf, Germany. [2]Department of Gastroenterology, Hepatology and Infectious Diseases, Heinrich-Heine-University, University Hospital, Düsseldorf, Germany. [3]Medical Center for Infectious Diseases (MIB), Berlin, Germany. [4]Institute of Virology, University of Cologne, Cologne, Germany.

Received: 1 February 2017 Accepted: 6 June 2017
Published online: 08 June 2017

References
1. European Association for the Study of the Liver. Electronic address, e.e.e., EASL Recommendations on Treatment of Hepatitis C 2016. J Hepatol. 2017; 66(1):153–94. PMID: 27667367.
2. Nelson DR, et al. All-oral 12-week treatment with daclatasvir plus sofosbuvir in patients with hepatitis C virus genotype 3 infection: ALLY-3 phase III study. Hepatology. 2015;61(4):1127–35.
3. Ferenci P. Treatment of hepatitis C in difficult-to-treat patients. Nat Rev Gastroenterol Hepatol. 2015;12(5):284–92.
4. Pawlotsky JM. Hepatitis C virus resistance to direct-acting antiviral drugs in interferon-free regimens. Gastroenterology. 2016;151(1):70-86. PMID: 27080301.
5. Sarrazin C. The importance of resistance to direct antiviral drugs in HCV infection in clinical practice. J Hepatol. 2016;64(2):486–504.
6. Hernandez D, et al. Natural prevalence of NS5A polymorphisms in subjects infected with hepatitis C virus genotype 3 and their effects on the antiviral activity of NS5A inhibitors. J Clin Virol. 2013;57(1):13–8.
7. Donaldson EF, et al. Clinical evidence and bioinformatics characterization of potential hepatitis C virus resistance pathways for sofosbuvir. Hepatology. 2015;61(1):56–65.
8. Hedskog C, et al. Evolution of the HCV viral population from a patient with S282T detected at relapse after sofosbuvir monotherapy. J Viral Hepat. 2015; 22(11):871–81.
9. Svarovskaia ES, et al. Infrequent development of resistance in genotype 1-6 hepatitis C virus-infected subjects treated with sofosbuvir in phase 2 and 3 clinical trials. Clin Infect Dis. 2014;59(12):1666–74.
10. Walker A, et al. Natural prevalence of resistance-associated variants in hepatitis C virus NS5A in genotype 3a-infected people who inject drugs in Germany. J Clin Virol. 2015;70:43–5.
11. Kalaghatgi P, et al. Geno2pheno[HCV] - a web-based interpretation system to support hepatitis C treatment decisions in the era of direct-acting antiviral agents. PLoS One. 2016;11(5):e0155869.
12. Tapper EB, et al. Evaluation of proton pump inhibitor use on treatment outcomes with ledipasvir and sofosbuvir in a real-world cohort study. Hepatology. 2016;64(6):1893–9.
13. Dvory-Sobol H, et al. Long-term persistence of HCV NS5A variants after treatment with NS5A inhibitor ledipasvir. J Hepatol. 2015;62(Suppl 2):S221.
14. Krishnan P, et al. O057 : Long-term follow-up of treatment-emergent resistance-associated variants in NS3, NS5A and NS5B with paritaprevir/r-, ombitasvir- and dasabuvir-based regimens. J Hepatol. 2015;62(Suppl 2):S220.
15. McPhee F, et al. Resistance analysis of hepatitis C virus genotype 1 prior treatment null responders receiving daclatasvir and asunaprevir. Hepatology. 2013;58(3):902–11.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Distinct Escape Pathway by Hepatitis C Virus Genotype 1a from a Dominant CD8$^+$ T Cell Response by Selection of Altered Epitope Processing

Andreas Walker,[a] Kathrin Skibbe,[b] Eike Steinmann,[c] Stephanie Pfaender,[c] Thomas Kuntzen,[d] Dominik A. Megger,[e] Svenja Groten,[b] Barbara Sitek,[e] Georg M. Lauer,[f] Arthur Y. Kim,[g] Thomas Pietschmann,[c] Todd M. Allen,[d] Joerg Timm[a]

Institute for Virology, Heinrich-Heine-University, University Hospital, Düsseldorf, Germany[a]; Institute of Virology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany[b]; Institute of Experimental Virology, Twincore Centre for Experimental and Clinical Infection Research, a joint venture between the Medical School Hanover and the Helmholtz Centre for Infection Research, Hanover, Germany[c]; Ragon Institute of MGH, MIT, and Harvard, Charlestown, Massachusetts, USA[d]; Medizinisches Proteom-Center, Ruhr-Universität Bochum, Bochum, Germany[e]; Gastrointestinal Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA[f]; Division of Infectious Diseases, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA[g]

ABSTRACT

Antiviral CD8$^+$ T cells are a key component of the adaptive immune response against HCV, but their impact on viral control is influenced by preexisting viral variants in important target epitopes and the development of viral escape mutations. Immunodominant epitopes highly conserved across genotypes therefore are attractive for T cell based prophylactic vaccines. Here, we characterized the CD8$^+$ T cell response against the highly conserved HLA-B*51-restricted epitope IPFYGKAI$_{1373–1380}$ located in the helicase domain of NS3 in people who inject drugs (PWID) exposed predominantly to HCV genotypes 1a and 3a. Despite this epitope being conserved in both genotypes, the corresponding CD8$^+$ T cell response was detected only in PWID infected with genotype 3a and HCV-RNA negative PWID, but not in PWID infected with genotype 1a. In genotype 3a, the detection of strong CD8$^+$ T cell responses was associated with epitope variants in the autologous virus consistent with immune escape. Analysis of viral sequences from multiple cohorts confirmed HLA-B*51-associated escape mutations inside the epitope in genotype 3a, but not in genotype 1a. Here, a distinct substitution in the N-terminal flanking region located 5 residues upstream of the epitope (S1368P; $P = 0.00002$) was selected in HLA-B*51-positive individuals. Functional assays revealed that the S1368P substitution impaired recognition of target cells presenting the endogenously processed epitope. The results highlight that, despite an epitope being highly conserved between two genotypes, there are major differences in the selected viral escape pathways and the corresponding T cell responses.

IMPORTANCE

HCV is able to evolutionary adapt to CD8$^+$ T cell immune pressure in multiple ways. Beyond selection of mutations inside targeted epitopes, this study demonstrates that HCV inhibits epitope processing by modification of the epitope flanking region under T cell immune pressure. Selection of a substitution five amino acids upstream of the epitope underlines that efficient antigen presentation strongly depends on its larger sequence context and that blocking of the multistep process of antigen processing by mutation is exploited also by HCV. The pathways to mutational escape of HCV are to some extent predictable but are distinct in different genotypes. Importantly, the selected escape pathway of HCV may have consequences for the destiny of antigen-specific CD8$^+$ T cells.

Based on phylogenetic analysis, hepatitis C virus (HCV) can be classified into at least seven genotypes and multiple subtypes that differ up to 20% at the amino acid level (1). The HCV genotypes have distinct epidemiological characteristics, they are associated with different transmission risk factors, and their frequencies in a population are regionally different. In Europe and North America the HCV genotypes 1 and 3 are most common (2, 3). Since routine screening via nucleic acid amplification for HCV eliminated the risk for infection through blood products, the most important risk group for incident HCV infection are people who inject drugs (PWID). The high prevalence of HCV infection with seroprevalence rates up to 80% paired with frequent risk practices for HCV transmissions in PWID results in incidence rates between 8 and 25% per year in young adult injectors (4, 5), and there is strong evidence that multiple exposures are common in this risk group (6, 7). The degree of sequence diversity between genotypes and subtypes precludes broad protection against reinfection. Accordingly, multiple infections of the same individual with different viruses have been reported (7).

Even at the subtype level HCV isolates typically differ between hosts. Lack of a proof reading function of the virus encoded RNA-

Received 7 August 2015 **Accepted** 1 October 2015

Accepted manuscript posted online 7 October 2015

Citation Walker A, Skibbe K, Steinmann E, Pfaender S, Kuntzen T, Megger DA, Groten S, Sitek B, Lauer GM, Kim AY, Pietschmann T, Allen TM, Timm J. 2016. Distinct escape pathway by hepatitis C virus genotype 1a from a dominant CD8$^+$ T cell response by selection of altered epitope processing. J Virol 90:33–42. doi:10.1128/JVI.01993-15.

Editor: M. S. Diamond

Address correspondence to Joerg Timm, joerg.timm@med.uni-duesseldorf.de.

A.W. and K.S. contributed equally to this article.

dependent RNA polymerase results in a high error rate during RNA replication. As a consequence, HCV exists in chronically infected patients as a quasispecies of closely related but genetically distinct viral variants. There is now strong evidence that viral genetic variation between hosts is the product of continuous selection of mutations by host immune pressure (8–14). Collectively, this inherent sequence diversity of HCV at the genotype, subtype, and quasispecies level is a major obstacle to vaccine design (15). Strategies that aim to develop prophylactic vaccines against HCV have to cope with this genetic heterogeneity either by inducing immune responses with a high degree of cross-reactivity (16, 17), by inducing multiple responses against different sequence variants (18), or by focusing the immune response on highly conserved regions of the virus.

Dominant $CD8^+$ T cell epitopes that are conserved across different HCV genotypes are rare (19, 20). Here, we characterized a highly conserved dominant HLA-B*51-restricted $CD8^+$ T cell epitope ($IPFYGKAI_{1373–1380}$) in HCV NS3 in PWID predominantly exposed to genotype 1a and 3a. Vigorous responses were detected in PWID with spontaneous immune control of HCV and in PWID with genotype 3a infection, but not in PWID infected with genotype 1a. Although selection of mutations inside the epitope was overall rare, there was evidence for mutational escape in HCV genotype 1b and 3a by population sequence analysis. Interestingly, HCV genotype 1a followed a distinct escape pathway by selecting a substitution (S1368P) located five amino acids upstream of the epitope. Further analysis of the functional relevance revealed that the S1368P substitution altered epitope processing. The results demonstrate that beyond selection of mutations inside $CD8^+$ epitopes HCV adapts to immune pressure by selecting mutations in the epitope flanking region. The evolutionary escape pathways differ between HCV genotypes, indicating distinct genetic plasticity.

MATERIALS AND METHODS

Patients. Blood samples from patients with a history of injection drug use were collected from the ward for inpatient detoxification treatment of drug addicts or the clinic for opioid maintenance treatment at the Department of Addictive Behavior and Addiction Medicine, Rhine State Hospital Essen, Hospital of the University of Duisburg-Essen. Written informed consent was obtained from all study participants and the study was approved by the ethics committee of the Medical Faculty of the University of Duisburg-Essen in accordance with the Declaration of Helsinki. In total 43 HLA-B*51-positive treatment-naive subjects who were HCV antibody positive (by anti-HCV chemiluminescent microparticle immunoassay from Abbott) were analyzed, including 15 HCV-RNA-negative patients and 28 subjects with detectable HCV-RNA. Peripheral blood mononuclear cells (PBMCs) were isolated from blood via Ficoll gradient centrifugation and subsequently cryopreserved.

Analysis and alignment of HCV sequences. The frequency of HLA-B*51-associated sequence polymorphisms was analyzed in 405 HCV genotype 1a sequences and 145 HCV genotype 1b sequences from a multicenter cohort (21) and in a large HCV genotype 1b outbreak (22). A region covering the epitope $IPFYGKAI_{1373–1380}$ was amplified and sequenced from additional 37 HCV genotype 1a isolates, as well as 102 HCV genotype 3a isolates collected in Germany. All sequences have been submitted to GenBank (accession no. FJ864775 to FJ864816, KJ130249 to KJ130317, and KJ668232 to KJ668268).

Analysis of the $CD8^+$ T cell response. Antigen-specific T cells were expanded from cryopreserved HLA-B*51-positive PBMCs utilizing synthetic peptides (>70% purity) purchased from EMC, Tübingen, Germany. After thawing, the PBMCs were cultured in RPMI 1640 medium containing 10% fetal calf serum, 100 U of penicillin/ml, 100 μg of streptomycin/ml, 10 mM HEPES buffer, and 25 U of recombinant interleukin-2 (IL-2)/ml and then stimulated with HLA-B*51 peptide 1373 (1 μg/ml) and 0.1 μg of anti-CD28 and anti-CD49d/ml. After 7 days medium containing IL-2 was added. On day 10 the cells were restimulated with the same peptide (10 μg/ml) in the presence of brefeldin A (100 ng/ml) for 4 h and then analyzed for their $CD4^+$, $CD8^+$ and gamma interferon (IFN-γ) expression via flow cytometry. To determine the degree of cross-reactivity between different B*51-1373 variants, PBMCs were cultivated in the presence of the HLA-B*51-1373 prototype or the variant peptide. After 10 days, both cultures were restimulated with the prototype and the variant peptide at different concentrations before intracellular IFN-γ staining. For the detection of HCV-specific cells *ex vivo* and after peptide-specific *in vitro* expansion, thawed PBMCs were stained with phycoerythrin-labeled IPFYGKAI-specific HLA-B*5101 dextramer (Immudex, Denmark), followed by surface staining with $CD8^+$ PerCP-Cy5.5 (eBioscience). All samples were acquired using a FACSCanto (BD), and the data were analyzed by using FlowJo software (Tree Star, Inc.).

Expression plasmids for HCV-GFP fusion proteins. All constructs were based on the parental plasmids pEGFP-N1 (Clontech, Germany) featuring a cytomegalovirus promoter-controlled enhanced green fluorescent protein for N-terminal fusion. Fragments of NS3 were generated by amplifying NS3 1329-1433 by nested PCR with the primers 4b-F (CC TACGGCAAGTTCCTTGC) and 4b-R_new (GCAGTCTATCACCGAG TCG) and the primers B51-for-EcoRI (CCAAGGGAATTCTTGGCT TCGTCTTCACCCTCGGCATCGGCACYGTCCTTGACCAAG) and B51-Rev-SalI (CGGATACCGTCGACCCGCGGTARTAMGCCACGGC). Subsequently, the fragment was cloned via EcoRI/SalI into pEGFP-N1. The substitution S1368P was introduced via site-directed mutagenesis (Stratagene) using the primer B51-S1368P-for (GGAGGTTGCTCTGCC CACCACCGGAGA) and B51-S1368P-rev (TCTCCGGTGGTGGGCAG AGCAACCTCC) according to the manufacturer's instructions. All constructs were verified by DNA sequencing.

Analysis of the $CD8^+$ T cell response against endogenously processed antigens. HLA-B*51 PBMCs from healthy donors were obtained from buffy coats (Department of Transfusion Medicine, University Hospital Essen, Essen, Germany). For analysis of endogenously processed antigens, HLA-B*51-positive cells were electroporated with plasmids pac-NS3-S1368S-GFP or pac-NS3-S1368P-GFP encoding an NS3-GFP fusion protein utilizing the Amaxa T Cell Nucleofector kit (vpa-1002; Lonza). In brief, after thawing PBMCs were cultured in RPMI medium. After 24 h, 7×10^6 cells were resuspended in 100 μl of Nucleofector solution, mixed with 7.5 μg of DNA and pulsed with the optimized protocol for unstimulated human T cells (Program v024; Amaxa2b) with an Amaxa apparatus (Lonza). After electroporation, the PBMCs were cultured in RPMI medium for 24 h cells before GFP-positive cells were sorted using a FACSAria II (BD). Subsequently, 10^5 B*51-1373-specific cells from a 10-day culture were restimulated with GFP-positive cells in a 1:1 ratio for 4 h before intracellular IFN-γ staining.

***In vitro* proteasome digestion.** The synthetic peptides EVALSTTGEI PFYGKAIPLEAIKGG and EVALPTTGEIPFYGKAIPLEAIKGG (≥95% purity) were purchased from EMC, Tübingen, Germany, and digested using a 20S proteasome assay kit complemented with human constitutive or immunoproteasome (Boston Biochem). In brief, 20 μg of peptide was digested with 4 μg of sodium dodecyl sulfate-activated proteasome in a 500-μl reaction. The reaction was stopped on indicated time points by adding 3 volumes of ice-cold acetone. For precipitation of proteasome, the samples were subsequently frozen at −20°C for 30 min and then centrifuged for 30 min at 4°C with 15,000 rpm. The supernatants were collected, evaporated to dryness, and dissolved in 50 μl of 0.1% trifluoroacetic acid. The peptide concentration of the resulting solution was determined by amino acid analysis as previously described (23). For the subsequent liquid chromatography-tandem mass spectrometry experiment, the sample was diluted to a concentration of 0.33 pmol/μl, and 5 pmol was analyzed using an UltiMate 3000 RSLCnano system online coupled to a

Velos Pro linear ion trap mass spectrometer (both from Thermo Scientific, Bremen, Germany) as described earlier (23). The acquired raw files were further analyzed with Proteome Discoverer software (Thermo Scientific, v1.3.0.339) and searched with Sequest (24) against a self-written database containing the investigated prototype sequence (pt) and S1368P peptides. Precursor and fragment ion mass tolerance was set to 0.4 Da, and the confidence level of peptide identification was set to a false discovery rate of 1%. Relative peptide quantification was carried out via spectral counting. Then, a normalized spectral index was calculated for each of the identified peptides by dividing the number of acquired peptide spectrum matches (PSMs) of a particular peptide by the number of PSMs acquired in the whole sample.

Generation and analysis of TNcc viruses. Plasmid pTNcc encoding for the GT1a virus TNcc was kindly provided by Jens Bukh (25). Mutations S1368P, I1373V, and I1380L were introduced by site-directed mutagenesis, and all constructs were verified by DNA sequencing. *In vitro* transcription and electroporation of Huh-7.5 cells was performed as described before (26). At 4 h posttransfection, the medium was changed with fresh Dulbecco modified Eagle medium (2 mM L-glutamine, nonessential amino acids, 100 U of penicillin/ml, 100 μg of streptomycin/ml, 10% fetal calf serum) supplemented with or without 10 mM 2′CMA (kindly provided by T. Tellinghuisen [27]). At 72 h posttransfection, cell-free supernatant was filtered through 0.45-μm-pore-size filters and 10× concentrated through Amicon centrifugal filters (Millipore). Virus titers were determined as described elsewhere (28) with slight modifications. In brief, Huh-7.5 cells were seeded in 96-well plates at a density of 10^4 cells per well 24 h prior to inoculation with dilutions of filtered cell culture supernatant (at least six wells were used per dilution). After 3 days, cells were washed with phosphate-buffered saline (PBS), fixed for 20 min with ice-cold methanol at −20°C, and washed three times with PBS. HCV-infected cells were detected with anti-core C7.50 (1:300) (29) and anti-NS5A 9E10 (1:1,000) (28) antibody in PBS for 45 min at room temperature. The cells were washed as described above, and bound antibodies were detected by incubation with horseradish peroxidase-conjugated antibodies specific to murine IgG (Sigma-Aldrich, Steinheim, Germany) diluted 1:200 in PBS. After 1 h of incubation at room temperature, the cells were washed as specified above. The peroxidase activity was detected by using carbazole substrate (0.32% [wt/vol] of 3-amino-9-ethylcarazole [Sigma] in *N*,*N*-dimethyl-formamide was diluted at a ratio of 1:3.3 with 15 mM acetic acid, 35 mM sodium acetate, [pH 8.0], and 0.4% H_2O_2). Virus titers (50% tissue culture infective doses [$TCID_{50}$]/ml) were calculated based on the method of Spearman and Kärber. HCV RNA was quantified by X-tail reverse transcriptase-PCR as described previously (22). For the detection of HCV core protein, virus-containing supernatant was inactivated by the addition of Triton X-100 to a final concentration of 1% (vol/vol) and the amount of released core protein was determined by a commercially available core enzyme-linked immunosorbent assay (Architect HCV Core AG test; Abbott, Wiesbaden, Germany).

Statistical analysis. All statistical tests were performed using GraphPad Prism 5.0 software (GraphPad Software, San Diego, CA).

RESULTS

$CD8^+$ T cells directed against the epitope $IPFYGKAI_{1373–1380}$ are detected after spontaneous resolution and in chronic genotype 3a infection but not in chronic genotype 1a infection. We previously reported an HLA-B*51-restricted epitope in NS3 ($IPFYGKAI_{1373–1380}$) that is >90% conserved across the HCV genotypes 1a, 1b, and 3a (19). The $CD8^+$ T cell immune response directed against the epitope $IPFYGKAI_{1373–1380}$ was analyzed in a cohort of anti-HCV-positive PWID carrying the HLA-B*51 allele. A total of 43 subjects were analyzed, including 15 anti-HCV-positive PWID with undetectable HCV-RNA, 13 PWID infected with GT1a, 3 PWID infected with GT1b, and 12 PWID infected with GT3a. After 10 days of *in vitro* expansion IPFYGKAI-specific $CD8^+$ T cells were detectable in 12 of 15 (80%) individuals with undetectable viremia and in 6 of 12 (50%) patients chronically infected with genotype 3a (Fig. 1A). In contrast, $CD8^+$ T cells directed against this epitope were not detectable after antigen-specific expansion in patients with genotype 1a or 1b infection (Fig. 1A). Similar results were obtained when $CD8^+$ T cell were directly analyzed *ex vivo* with HLA class I/peptide dextramers. IPFYGKAI-specific $CD8^+$ T cells were detectable in most HCV-RNA-negative PWID and in some PWID with a genotype 3a infection. Notably, the four genotype 3a-infected patients with high frequencies were identical to those where antigen-specific $CD8^+$ T cells were efficiently expanded *in vitro* (Fig. 1A). In turn, the analysis with HLA class I/peptide dextramers confirmed lack of detection of IPFYGKAI-specific $CD8^+$ T cells in patients infected with genotype 1 (Fig. 1B). The $CD8^+$ T cell response against the epitope $IPFYGKAI_{1373–1380}$ was also studied in a patient with acute HCV genotype 1a infection. The patient was previously reported in a study on the impact of HLA-B*57 on HCV infection outcome (30) and was also HLA-B*51 positive. Interestingly, here, the response was detectable by week 11 after infection but became completely undetectable in an IFN-γ enzyme-linked immunospot (ELISpot) assay by week 47 (Fig. 1C).

The autologous virus of 22 PWID from the total of 26 with detectable HCV-RNA was sequenced. The sequences of the epitope and the flanking region are shown in Table 1 and are indicated in Fig. 1A for PWID with genotype 3a infection with a detectable $CD8^+$ T cell response (colored in Fig. 1A). In 3 of 10 HCV genotype 3a sequences, the virus harbored substitutions in the epitope region (I1373V, K1377R, or I1380L). Of note, all three patients mounted robust responses against the prototype sequence (Fig. 1A). In functional assays the I1380L variant impaired recognition by antigen-specific $CD8^+$ T cells, whereas the I1373V variant was cross-reactive in HCV-RNA-negative PWID and to a lesser extent in GT3a-infected PWID (Fig. 1D and E). Notably, the three remaining GT3a-infected patients with detectable $CD8^+$ T cell responses and no evidence for escape mutations had a history of injection drug use of less than 6 months, consistent with more recent exposure and infection with HCV, possibly at a stage prior to mutational escape. One of nine GT1a sequences and one of three GT1b sequences harbored the I1373V substitution in position 1 of the epitope. In both patients no $CD8^+$ T cell response was detectable. One PWID with genotype 1a infection harbored a S1368P substitution. Notably, this substitution was also selected in the patient with acute HCV genotype 1a infection and with the rapid decline of the $CD8^+$ T cell response (Fig. 1C and data not shown).

The residue under selection pressure in HLA-B*51-positive individuals depends on the viral genotype. Given the relatively high reproducibility of IPFYGKAI-specific $CD8^+$ T cell responses in HCV-RNA negative PWID, we aimed to address whether there is evidence for immune escape in patients with persistent HCV infection at a population level. We therefore analyzed the impact of HLA-B*51 expression on the frequency of sequence polymorphisms in this epitope in different cohorts. In an analysis of 442 genotype 1a sequences from a multicenter cohort (Fig. 2A), there was no evidence for mutational escape inside the epitope. Although an I1373V polymorphism was observed, this substitution was not enriched in HLA-B*51-positive patients. In contrast, the I1373V polymorphism was slightly enriched in HLA-B*51-posi-

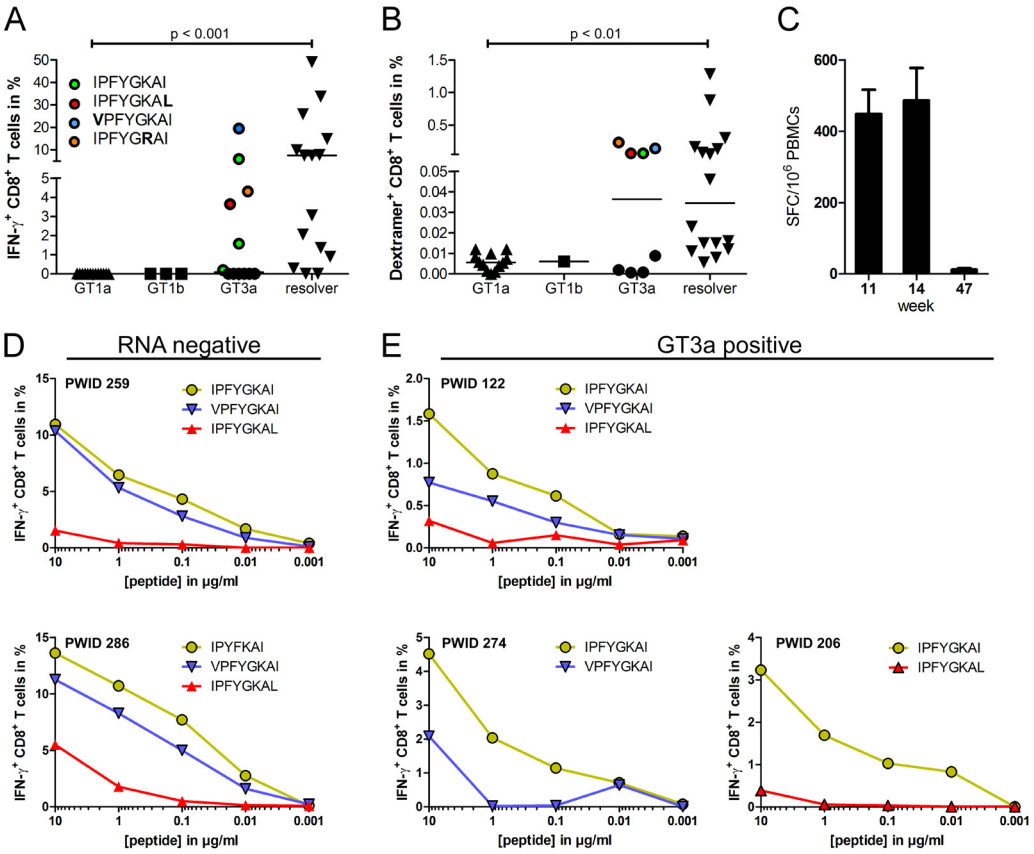


FIG 1 CD8$^+$ T cell response against the epitope IPFYGKAI$_{1373–1380}$ in PWID exposed to HCV. (A) T cells were expanded for 10 days from PBMCs in the presence of the peptide IPFYGKAI. After *in vitro* expansion, the cells were restimulated with the peptide before intracellular IFN-γ staining. The autologous viral epitope sequences from GT3a-infected PWID with detectable responses are indicated. (B) HLA-B*51$_{1373–1380}$-specific T cells were detected ex *vivo* via IPFYGKAI-specific HLA-B*5101 dextramer staining. (C) The HLA-B*51$_{1373–1380}$ specific T cell response was determined by ELISpot assay at the indicated time points in a patient with acute HCV infection. (D and E) Serial peptide dilutions of the prototype (green), I1373V (blue), or I1380L (red) sequence of the B51-1373 peptide were tested in RNA-negative (D) or GT3a-infected (E) patients. Statistical comparisons between groups in panels A and B were done with a Kruskal-Wallis test, and significant *P* values are indicated.

tive patients from a multicenter cohort of 145 genotype 1b-infected patients (Fig. 2B), although this difference was not statistically significant. However, in a large genotype 1b single source outbreak there was statistical evidence for selection of polymorphisms in position 1 (I1373V) and 8 (I1380L) of the epitope in the presence of HLA-B*51 (Fig. 2C). Finally, we also analyzed a local cohort of 102 patients infected with genotype 3a (Fig. 2D). Here, there was statistical evidence for selection of the I1373V polymorphism inside the epitope. Importantly, even though in genotype 1a there was no evidence for mutational escape inside the epitope, there was strong statistical evidence for HLA-B*51-associated selection pressure on a residue located five amino acids upstream of the epitope (S1368) (Fig. 2A). In genotype 1a, 20.7% of the HLA-B*51-positive patients carried a S1368P substitution; in contrast, only 0.8% of the HLA-B*51-negative patients carried any substitutions in this position ($P = 0.00002$). Notably, in genotype 3a an S1369P substitution was also enriched in HLA-B*51-positive individuals; however, the difference was not statistically significant (Fig. 2D). Taken together, by analyses of different cohorts we found statistical evidence for mutational escape inside the HLA-B*51-restricted epitope (IPFYGKAI) in genotype 1b and 3a and evidence for the selection of a distinct substitution in the epitope flanking region in genotype 1a.

The S1368P substitution impairs targeting of the endogenously processed HLA-B*51-restricted epitope IPFYGKAI$_{1373–1380}$. HCV genotype 1a was unique because evidence for HLA-B*51-associated selection pressure was only observed in the epitope flanking region. Of note, this region does not contain a second HLA-B*51 binding motif by analysis with prediction algorithms for HLA class I binding (www.immuneepitope.org). We therefore hypothesized that the S1368P may impair processing and presentation of the epitope. To address this experimentally, an assay that allows analysis of endogenously processed antigens was established. Effector CD8$^+$ T cells directed against the HLA-B*51-restricted epitope IPFYGKAI were obtained by 10 days of antigen-specific expansion from different PWID with spontaneous immune control of HCV. To generate target cells that present the endogenously processed epitope HLA-B*51-positive buffy coat cells were transfected with expression plasmids encoding a fusion protein of a short fragment of NS3 containing the epitope fused to GFP. At 24 h after transfection GFP-positive cells were sorted and used as targets for specific CD8$^+$ T. As a positive-control native target cells were pulsed with the synthetic peptide overnight. Figure 3A shows a representative result of one experiment. Upon stimulation with peptide-pulsed targets 6.3% IFN-γ$^+$ CD8$^+$ T cells were detectable. When targets were transfected with a genotype 1a prototype

TABLE 1 Patient characteristics

Patient ID	Genotype	Viral load (IU/ml)	Sequence[a]														
			A	L	S	T	T	G	E	I	P	F	Y	G	K	A	I	P
138	1a	2,148,000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
154	1a	777,200	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
175	1a	2,200,000	.	.	P	.	.	.	.	.	.	.	.	.	.	.	.	.	.
251	1a	61,790	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
278	1a	3,015,000	.	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	.
283	1a	2,422,000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
299	1a	275,700	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
324	1a	2,469,000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
348	1a	83,080	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
393	1a	5,282,000	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
408	1a	1,562,000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
423	1a	320,900	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
581	1a	3,942,000	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
084	1b	61,550	.	.	.	N	I	.	.	.	.	.	.	.	.	.	.	.	.
117	1b	1,033,000	.	.	.	N	I	.	.	V	.	.	.	.	.	.	.	.	.
332	1b	220,000	.	.	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.
096	3a	247,700	.	.	G	S	E	.	.	.	.	.	.	.	.	.	.	.	.
113	3a	2258,000	.	.	G	S	E	.	.	.	.	.	.	.	.	.	.	.	.
122	3a	87,400	.	.	G	S	E	.	.	.	.	.	.	.	.	.	.	.	.
137	3a	1,046,000	.	.	G	S	E	.	.	.	.	.	.	.	.	.	.	.	.
176	3a	924,700	.	.	G	S	E	.	.	.	.	.	.	.	.	.	.	.	.
206	3a	368,000	.	.	G	S	E	.	.	.	.	.	.	.	.	.	.	L	.
240	3a	154,500	.	.	G	S	E	.	.	.	.	.	.	.	.	.	.	.	.
257	3a	3,642	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
274	3a	146,300	.	.	G	S	E	.	.	V	.	.	.	.	.	.	.	.	.
292	3a	80,690	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
466	3a	34,010,000	.	.	G	S	E	.	.	.	.	.	.	.	.	.	.	.	.
533	3a	511,500	.	.	G	S	D	.	.	.	.	.	.	.	R	.	.	.	.
042	RNA neg[b]																		
062	RNA neg																		
110	RNA neg																		
161	RNA neg																		
196	RNA neg																		
242	RNA neg																		
264	RNA neg																		
286	RNA neg																		
344	RNA neg																		
365	RNA neg																		
417	RNA neg																		
418	RNA neg																		
474	RNA neg																		
574	RNA neg																		
587	RNA neg																		

[a] The H77 prototype sequence is indicated in the column subheadings. Differences from the prototype sequence are specified by the appropriate letter in the table. A period indicates no difference from the prototype sequence. –, Not done.
[b] RNA neg, RNA negative.

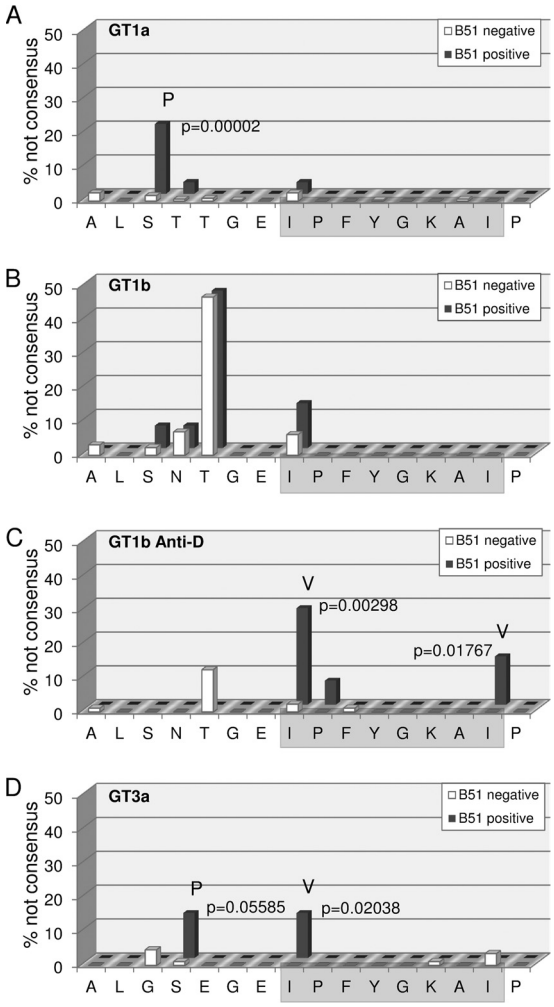


FIG 2 Frequency of HLA-B*51-associated viral polymorphisms in the epitope region. The frequency of variations from the reported prototype sequence of the epitope region in GT1a (A), GT1b (B), the East-German Anti-D cohort (C), and GT3a (D) are shown for patients carrying the HLA-B*51 allele (red) and patients not carrying the HLA-B*51 allele (green). Positions with significant differences in polymorphism frequencies in the absence or presence of HLA-B*51 are marked, and the *P* values (Fisher exact test) and the most frequent variant amino acid are indicated.

sequence 7.3% of $CD8^+$ T cells secreted IFN-γ. In contrast, when targets were transfected with the plasmid harboring the S1368P substitution the number of IFN-γ^+CD8^+ T cells was reduced to 1.3%. In seven independent experiments the $CD8^+$ T cell response against targets transfected with the S1368P variant was reproducibly reduced to levels of 24% compared to prototype 1a (Fig. 3B). This suggests that less antigen was presented on S1368P-transfected target cells consistent with impaired endogenous processing associated with this substitution. To compare the processing efficiency between genotype 1a and genotype 3a, the same fusion proteins were constructed with the 3a prototype sequence and the genotype 3a S1369P substitution. In three independent experiments the $CD8^+$ T cell response against targets transfected with prototype 3a were reproducibly weaker (33%) compared to prototype 1a. The S1369P substitution further reduced the response to 6% compared to prototype 1a. This suggests that the epitope is less efficiently processed in a genotype 3a context compared to a genotype 1a context. Moreover, both substitutions in the epitope flanking region (S1368P in genotype 1a and S1369P in genotype 3a) further reduce the antigen processing efficiency.

Differential proteasomal cleavage of peptides with the S1368P substitution. It was next addressed whether the S1368P substitution selected in genotype 1a has an impact on proteasomal cleavage consistent with altered processing. Therefore, synthetic peptides 25 amino acids in length either with the prototype sequence (pt) or harboring the S1368P substitution (S1368P) were digested with constitutive or immune proteasome, and the cleavage products were analyzed by mass spectrometry. In Fig. 4 the normalized spectral indices for cleavage products containing the full epitope sequence are shown at different time points after digestion with the constitutive proteasome (Fig. 4A) or the immune proteasome (Fig. 4B). There was no significant difference in the relative frequency of epitope containing peptides upon digestion of the prototype or the S1368P variant at any time. In fact, there

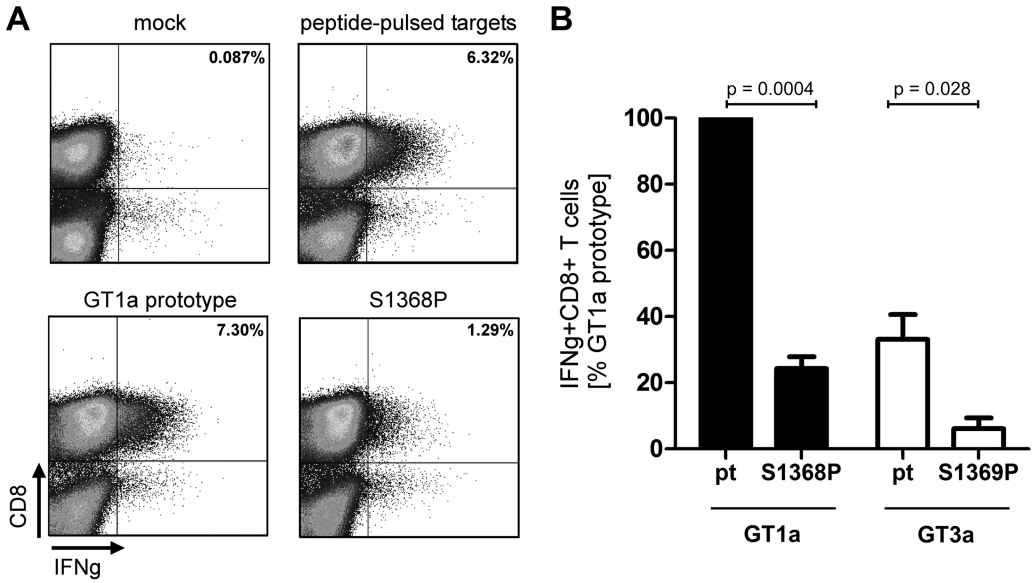


FIG 3 CD8$^+$ T cell response against the endogenously processed epitope IPFYGKAI$_{1373–1380}$. Effector T cells were expanded for 10 days from PBMCs in the presence of the peptide IPFYGKAI. HLA-B*51-positive target cells were generated by electroporation with a NS3(aa1330-2420)-GFP fusion protein. GFP-positive cells were sorted and used as targets for restimulation of IPFYGKAI-specific effector CD8$^+$ T cells in an effector/target ratio of 1:1 for 4 h, followed by an intracellular cytokine staining (ICS). Mock transfected targets and peptide-pulsed targets served as negative or positive controls, respectively. (A) Representative fluorescence-activated cell sorting results of one ICS. (B) The IFN-γ response against targets transfected with GT1a prototype NS3 was normalized to 100% and compared to the response against other targets as indicated. The data represent results from at least three independent experiments. The *P* values were calculated using a one-sample *t* test with a hypothetical value of 100 (S1368P versus pt1a) or an unpaired *t* test (pt GT3a versus S1369P).

was a minor trend toward higher frequencies of epitope containing peptides after 24 h when the S1368P variant was digested (Fig. 4A and B). Since carboxypeptidase activity is basically absent from the endoplasmic reticulum (ER), antigen presentation requires peptides with correct C-terminal ends after proteasomal cleavage (31). We therefore focused the analysis on peptide products ending with the epitope sequence IPFYGKAI. Figure 4C and D show the normalized spectral indices of individual peptides after digestion by the constitutive proteasome (Fig. 4C) or the immune proteasome (Fig. 4D) at different time points. After digestion of the prototype peptide the most prevalent cleavage product was STTGEIPFYGKAI carrying the residue that was under selection pressure in HLA-B*51-positive patients in genotype 1a at its N-terminal end (boxed in Fig. 4C and D). Also highly prevalent were N-extended peptides with three (TGEIPFYGKAI) or four (TTGEIPF YGKAI) additional residues. When the S1368P variant was digested the most prevalent cleavage product (PTTGEIPFYGKAI) carried the substitution selected in HLA-B*51-positive patients in genotype 1a at its N-terminal end. Importantly, shorter N-extended cleavage products with three or four additional residues were nearly absent after digestion of the S1368P variant. Similar results were obtained after digestion by the constitutive proteasome (Fig. 4C) or the immune proteasome (Fig. 4D).

The S1368P substitution does not impair viral fitness in HCV genotype 1a. We hypothesized that fitness constraints caused preferred selection of the S1368P substitution in genotype 1a. To address this experimentally, the recently described genotype 1a full genome virus TNcc (25) was utilized to study the replication fitness and infectivity associated with the substitutions S1368P, I1373V or I1380L. At 72 h after the electroporation of virus RNA into Huh-7.5 cells, the supernatants were used for infection of naive Huh-7.5 cells (Fig. 5A). Upon transfection with the parental HCV TN genome the $TCID_{50}$ was 1.1×10^4 infectious particles per ml. Approximately the same amount of infectious particles was released upon transfection with the S1368P variant (0.8×10^4 $TCID_{50}$/ml) or the I1373V variant (1.3×10^4 $TCID_{50}$/ml). In contrast, the I1380L substitution was associated with a decreased $TCID_{50}$ of 0.2×10^4 per ml. The same hierarchy was observed at the replication level, with only a slight reduction of replication for the S1368P and I1373V substitution and a 3-fold reduced replication for the I1380L substitution. This indicates that only the substitution I1380L impairs viral fitness in genotype 1a, whereas the substitutions S1368P and I1380L do not.

DISCUSSION

Here, we characterized a highly conserved HLA-B*51-restricted CD8$^+$ T cell epitope located in the helicase domain of HCV NS3. The high frequency of CD8$^+$ T cells specific for this epitope in PWID who achieved spontaneous immune control of HCV infection suggests that this CD8$^+$ T cell response is reproducibly mounted in HLA-B*51-positive patients. Given the high degree of conservation of the epitope, it was therefore surprising that in chronic infection CD8$^+$ T cell responses against this epitope were detectable only in PWID infected with genotype 3a and were completely absent in patients infected with genotype 1a. Although viral sequence polymorphisms were overall rare, there was selection pressure on the epitope containing region, as evidenced by HLA-B*51-associated viral sequence polymorphisms in all studied genotypes. Notably, in genotype 1b there was statistical support for the selection of escape mutations inside the epitope in the single-source outbreak, whereas there was only a nonsignificant trend in the genotype 1b multicenter cohort. It is possible that escape patterns inside this epitope differ between cohorts. Alternatively, because the natural sequence variability in this epitope in the geno-

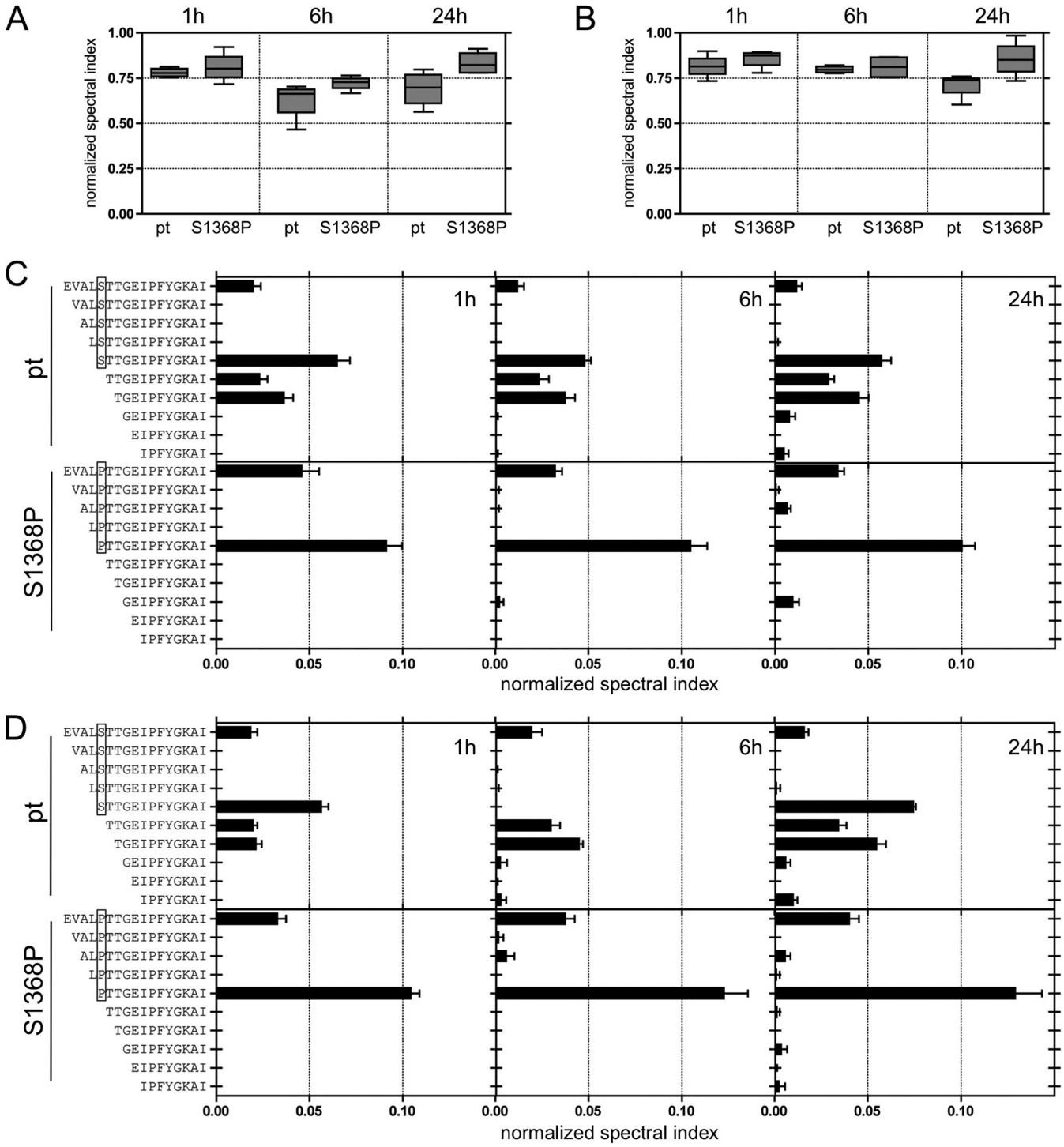


FIG 4 Impact of the S1368P substitution on proteasomal degradation. (A and B) The relative abundance of epitope-containing cleavage products after digestion of 25mer peptides with the prototype sequence (pt) or carrying the S1368P substitution (S1368P) with constitutive proteasome (A) or immune proteasome (B) is shown by normalized spectral indices. (C and D) The relative abundance of cleavage products with the correct C-terminal epitope end is shown by normalized spectral indices of individual peptides after digestion with constitutive proteasome (C) or immune proteasome (D).

type 1b multicenter cohort is higher than in the single-source outbreak, the size of the multicenter cohort may be too small to unmask an existing effect of HLA-B*51-associated selection pressure.

The predominant escape mechanisms of substitutions selected by $CD8^+$ T cell immune pressure in HCV epitopes are impaired binding of the variant epitope to the HLA class I-molecule or impaired binding of the T cell receptor to the HLA class I/peptide-

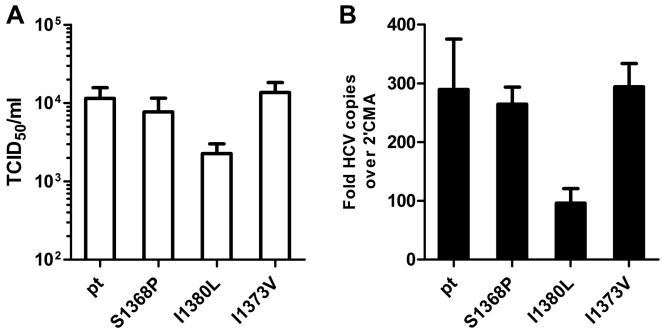


FIG 5 Infectivity and replication of wild-type and mutant TNcc strains. RNA transcripts of the parental HCV TN genome and TN mutants were transfected into Huh-7.5 cells. (A) Seventy-two hours later, cell-free supernatant was used for inoculation of naive Huh-7.5 cells. The $TCID_{50}$ of the variants was determined by a limiting-dilution assay and staining with core and NS5A-specific antibodies. (B) RNA levels at 72 h posttransfection were measured by quantitative real-time reverse transcription-PCR. The RNA levels in the presence of 2′methyladenosin (2′CMA) was used for normalization. The ratio of measured HCV RNA to measured RNA levels in the presence of 2′CMA was calculated for each construct. The data from four independent experiments are shown.

complex. As an alternative mechanism, processing of the variant epitope can be affected by substitutions in the epitope region. We previously reported $CD8^+$ T cell selection of a substitution inside the HLA-B*08-restricted epitope NS3 $HSKKKCDEL_{1395–1403}$ located in position 9 of the epitope (9). Functional analyses revealed that the selected leucine to valine or phenylalanine substitution impaired presentation of the endogenously processed epitope. Similarly, Kimura et al. (32) described substitutions selected inside targeted $CD8^+$ T cell epitopes that impaired proteasomal processing in the chimpanzee model. In contrast, $CD8^+$ T cell selection of substitutions in the epitope flanking region has not been described yet in HCV. Seifert et al. (33) demonstrated that a tyrosine to phenylalanine polymorphism located in the C-terminal position +1 of the HLA-A*02-restricted epitope NS3 CVNG $VCWTV_{1073–1081}$ influenced carboxy-terminal cleavage of the epitope by the proteasome; however, there was no conclusive evidence for immune selection pressure on this residue (22, 33). In HIV larger population studies of HLA class I-associated viral adaptation suggested that selection of processing mutations in the epitope flanking region contributes to mutational immune escape (34). This included substitutions in the C-terminal and N-terminal epitope flanking regions. Only a few studies directly addressed the functional consequences of such putative processing mutations (35, 36). The example of selection of altered processing in HLA-B*51-positive patients infected with HCV genotype 1a supports that similar immune escape strategies are utilized by HCV.

Processing and presentation of viral epitopes is a multistep process starting with production of peptide precursors by proteasomal degradation of viral proteins (reviewed in reference 37). These peptide precursors undergo N-terminal trimming by different endopeptidases and aminopeptidases in the cytosol and by the aminopeptidases ERAP1 and ERAP2 after the peptides have been transported by TAP to the ER (38, 39). Importantly, since carboxypeptidase activity is absent from the ER, production of peptide precursors with correct C-terminal ends by the proteasome is required for subsequent HLA class I presentation (31). It has been highlighted that the trimming efficiency of N-extended epitopes strongly depends on the amino acid composition of the epitope flanking region. Some amino acids, such as leucine, lysine, phenylalanine, and methionine, are cleaved with high activity, whereas amino acids such as proline and glutamic acid are only poorly cleaved (40). Accordingly, these latter amino acids are underrepresented in flanking regions of known $CD8^+$ T cell epitopes (41). In HCV genotype 1a there was HLA-B*51-associated selection pressure on position −5 of the epitope with a substitution from serine to proline. Digestion of the N- and C-extended variant peptide by the constitutive and the immune proteasome yielded predominantly precursor peptides starting with proline, whereas digestion of the prototype yielded also shorter N-extended precursor peptides. The shorter peptides are advantageous for the transport into the ER (42), and more importantly, N-terminal trimming of the peptide precursor starting with proline is inhibited (31). This suggests that the serine to proline substitution altered proteasomal production of epitope precursors with the consequence of a very low abundance of cleavable N-extended epitopes in the ER, although the latter was not formally tested.

Selection of $CD8^+$ T cell escape mutations is a trade-off between functional escape and viral constraints. Indeed, HCV may explore its replication space with different substitutions in targeted epitopes before some mutations reach fixation (9, 43). The example presented here underlines that even despite an epitope being conserved across HCV genotypes the selected escape pathways and the $CD8^+$ T cell response can substantially differ. Although not fully conclusive, our data suggest that selection of the S1368P substitution represents the optimal trade-off between functional immune escape and viral fitness in genotype 1a. The I1380L substitution may be disadvantageous due to its fitness costs despite efficient immune escape. In turn, the degree of cross-reactivity of the I1373V variant may prevent fixation in genotype 1a despite its low fitness costs. Interestingly, we observed interindividual differences in the level of cross-reactivity of the I1373V substitution with high cross-reactivity in HCV-RNA-negative PWID and lower levels of cross-reactivity in PWID infected with genotype 3a (Fig. 1D and E), suggesting that the degree of cross-reactivity may influence the outcome of infection as previously reported (17). Absence of the $CD8^+$ T cell response in PWID with chronic genotype 1a infection is unlikely caused by lack of priming. The rapid decline of the response in the patient with acute HCV genotype 1a infection rather suggests secondary failure. This patient harbored a virus with the S1368P substitution. Although the $CD8^+$ T cell response was clearly impaired against endogenously processed antigens carrying the S1368P substitution, there was still some degree of T cell reactivity in our assays. This residual epitope presentation may cause continuous $CD8^+$ T cell stimulation associated with progressive dysfunction or exhaustion of antigen-specific $CD8^+$ T cells (44–46) and possibly ultimate extinction of these cells (46).

The degree of $CD8^+$ T cell reactivity against the S1368P substitution in the context of genotype 1a was comparable to the reactivity against the endogenously processed genotype 3a antigen, suggesting that the epitope is less efficiently processed in the context of genotype 3a. Notably, in genotype 3a the substitution S1369P also impaired recognition of the endogenously processed antigen; however, here the I1373V was preferentially selected. One possibility is that in combination with the lower processing efficiency of the epitope in genotype 3a the I1373V substitution is sufficient to escape from the immune response. A second plausible

explanation is that the I1373V is associated with lower fitness costs compared to S1369P and I1380L in genotype 3a. However, this cannot be addressed until robust genotype 3a infectious viruses are publicly available. In either case, at least in some individuals the corresponding $CD8^+$ T cell response was preserved in our cohort of PWID infected with genotype 3a, a finding consistent with a memory T cell phenotype after mutational escape of the targeted antigen as previously reported (44, 45).

Taken together, HCV is able to evolutionarily adapt to $CD8^+$ T cell immune pressure in multiple ways. The pathways to mutational escape are predictable but are distinct in different genotypes. Beyond selection of mutations inside targeted epitopes that impair HLA class I binding of the variant peptide ligand or TCR binding to the variant HLA class I/peptide complex HCV also inhibits epitope processing by modification of the epitope flanking region under T cell immune pressure. Importantly, the selected escape pathway of HCV may have consequences for the destiny of antigen-specific $CD8^+$ T cells.

ACKNOWLEDGMENTS

We thank Jens Bukh for the TNcc strain, Darius Moradpour and Charles Rice for monoclonal antibodies, and Michael Engelmann and Lejla Timmer for technical assistance.

FUNDING INFORMATION

Deutsche Forschungsgemeinschaft (DFG) provided funding to Joerg Timm under grant numbers TRR60 and RTG1045. The Helmholtz Association provided funding to Thomas Pietschmann under grant number SO-024.

REFERENCES

1. **Pybus OG, Cochrane A, Holmes EC, Simmonds P.** 2005. The hepatitis C virus epidemic among injecting drug users. Infect Genet Evol **5:**131–139. http://dx.doi.org/10.1016/j.meegid.2004.08.001.
2. **Esteban JI, Sauleda S, Quer J.** 2008. The changing epidemiology of hepatitis C virus infection in Europe. J Hepatol **48:**148–162. http://dx.doi.org/10.1016/j.jhep.2007.07.033.
3. **Rustgi VK.** 2007. The epidemiology of hepatitis C infection in the United States. J Gastroenterol **42:**513–521. http://dx.doi.org/10.1007/s00535-007-2064-6.
4. **Page K, Hahn JA, Evans J, Shiboski S, Lum P, Delwart E, Tobler L, Andrews W, Avanesyan L, Cooper S, Busch MP.** 2009. Acute hepatitis C virus infection in young adult injection drug users: a prospective study of incident infection, resolution, and reinfection. J Infect Dis **200:**1216–1226. http://dx.doi.org/10.1086/605947.
5. **Mehta SH, Astemborski J, Kirk GD, Strathdee SA, Nelson KE, Vlahov D, Thomas DL.** 2011. Changes in blood-borne infection risk among injection drug users. J Infect Dis **203:**587–594. http://dx.doi.org/10.1093/infdis/jiq112.
6. **Micallef JM, Macdonald V, Jauncey M, Amin J, Rawlinson W, van Beek I, Kaldor JM, White PA, Dore GJ.** 2007. High incidence of hepatitis C virus reinfection within a cohort of injecting drug users. J Viral Hepat **14:**413–418. http://dx.doi.org/10.1111/j.1365-2893.2006.00812.x.
7. **Aitken CK, Lewis J, Tracy SL, Spelman T, Bowden DS, Bharadwaj M, Drummer H, Hellard M.** 2008. High incidence of hepatitis C virus reinfection in a cohort of injecting drug users. Hepatology **48:**1746–1752. http://dx.doi.org/10.1002/hep.22534.
8. **von Hahn T, Yoon JC, Alter H, Rice CM, Rehermann B, Balfe P, McKeating JA.** 2007. Hepatitis C virus continuously escapes from neutralizing antibody and T-cell responses during chronic infection in vivo. Gastroenterology **132:**667–678. http://dx.doi.org/10.1053/j.gastro.2006.12.008.
9. **Timm J, Lauer GM, Kavanagh DG, Sheridan I, Kim AY, Lucas M, Pillay T, Ouchi K, Reyor LL, Schulze zur Wiesch J, Gandhi RT, Chung RT, Bhardwaj N, Klenerman P, Walker BD, Allen TM.** 2004. CD8 epitope escape and reversion in acute HCV infection. J Exp Med **200:**1593–1604. http://dx.doi.org/10.1084/jem.20041006.
10. **Ruhl M, Knuschke T, Schewior K, Glavinic L, Neumann-Haefelin C, Chang DI, Klein M, Heinemann FM, Tenckhoff H, Wiese M, Horn PA, Viazov S, Spengler U, Roggendorf M, Scherbaum N, Nattermann J, Hoffmann D, Timm J.** 2011. $CD8^+$ T-cell response promotes evolution of hepatitis C virus nonstructural proteins. Gastroenterology **140:**2064–2073. http://dx.doi.org/10.1053/j.gastro.2011.02.060.
11. **Erickson AL, Kimura Y, Igarashi S, Eichelberger J, Houghton M, Sidney J, McKinney D, Sette A, Hughes AL, Walker CM.** 2001. The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes. Immunity **15:**883–895. http://dx.doi.org/10.1016/S1074-7613(01)00245-X.
12. **Cox AL, Mosbruger T, Mao Q, Liu Z, Wang XH, Yang HC, Sidney J, Sette A, Pardoll D, Thomas DL, Ray SC.** 2005. Cellular immune selection with hepatitis C virus persistence in humans. J Exp Med **201:**1741–1752. http://dx.doi.org/10.1084/jem.20050121.
13. **Ray SC, Fanning L, Wang XH, Netski DM, Kenny-Walsh E, Thomas DL.** 2005. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. J Exp Med **201:**1753–1759. http://dx.doi.org/10.1084/jem.20050122.
14. **Farci P, Alter HJ, Wong DC, Miller RH, Govindarajan S, Engle R, Shapiro M, Purcell RH.** 1994. Prevention of hepatitis C virus infection in chimpanzees after antibody-mediated in vitro neutralization. Proc Natl Acad Sci U S A **91:**7792–7796. http://dx.doi.org/10.1073/pnas.91.16.7792.
15. **Fauvelle C, Lepiller Q, Felmlee DJ, Fofana I, Habersetzer F, Stoll-Keller F, Baumert TF, Fafi-Kremer S.** 2013. Hepatitis C virus vaccines: progress and perspectives. Microb Pathog **58:**66–72. http://dx.doi.org/10.1016/j.micpath.2013.02.005.
16. **Ziegler S, Skibbe K, Walker A, Ke X, Heinemann FM, Heinold A, Mok JY, van Esch WJ, Yang D, Wolfl M, Timm J.** 2014. Impact of sequence variation in a dominant HLA-A*02-restricted epitope in hepatitis C virus on priming and cross-reactivity of $CD8^+$ T cells. J Virol **88:**11080–11090. http://dx.doi.org/10.1128/JVI.01590-14.
17. **Yerly D, Heckerman D, Allen TM, Chisholm JV, III, Faircloth K, Linde CH, Frahm N, Timm J, Pichler WJ, Cerny A, Brander C.** 2008. Increased cytotoxic T-lymphocyte epitope variant cross-recognition and functional avidity are associated with hepatitis C virus clearance. J Virol **82:**3147–3153. http://dx.doi.org/10.1128/JVI.02252-07.
18. **Yusim K, Dilan R, Borducchi E, Stanley K, Giorgi E, Fischer W, Theiler J, Marcotrigiano J, Korber B, Barouch DH.** 2013. Hepatitis C genotype 1 mosaic vaccines are immunogenic in mice and induce stronger T-cell responses than natural strains. Clin Vaccine Immunol **20:**302–305. http://dx.doi.org/10.1128/CVI.00605-12.
19. **Giugliano S, Oezkan F, Bedrejowski M, Kudla M, Reiser M, Viazov S, Scherbaum N, Roggendorf M, Timm J.** 2009. Degree of cross-genotype reactivity of hepatitis C virus-specific $CD8^+$ T cells directed against NS3. Hepatology **50:**707–716. http://dx.doi.org/10.1002/hep.23096.
20. **von Delft A, Humphreys IS, Brown A, Pfafferott K, Lucas M, Klenerman P, Lauer GM, Cox AL, Gaudieri S, Barnes E.** 2015. The broad assessment of HCV genotypes 1 and 3 antigenic targets reveals limited cross-reactivity with implications for vaccine design. Gut http://dx.doi.org/10.1136/gutjnl-2014-308724.
21. **Kuntzen T, Timm J, Berical A, Lennon N, Berlin AM, Young SK, Lee B, Heckerman D, Carlson J, Reyor LL, Kleyman M, McMahon CM, Birch C, Schulze Zur Wiesch J, Ledlie T, Koehrsen M, Kodira C, Roberts AD, Lauer GM, Rosen HR, Bihl F, Cerny A, Spengler U, Liu Z, Kim AY, Xing Y, Schneidewind A, Madey MA, Fleckenstein JF, Park VM, Galagan JE, Nusbaum C, Walker BD, Lake-Bakaar GV, Daar ES, Jacobson IM, Gomperts ED, Edlin BR, Donfield SM, Chung RT, Talal AH, Marion T, Birren BW, Henn MR, Allen TM.** 2008. Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients. Hepatology **48:**1769–1778. http://dx.doi.org/10.1002/hep.22549.
22. **Ruhl M, Chhatwal P, Strathmann H, Kuntzen T, Bankwitz D, Skibbe K, Walker A, Heinemann FM, Horn PA, Allen TM, Hoffmann D, Pietschmann T, Timm J.** 2012. Escape from a dominant HLA-B*15-restricted $CD8^+$ T cell response against hepatitis C virus requires compensatory mutations outside the epitope. J Virol **86:**991–1000. http://dx.doi.org/10.1128/JVI.05603-11.
23. **Megger DA, Bracht T, Kohl M, Ahrens M, Naboulsi W, Weber F, Hoffmann AC, Stephan C, Kuhlmann K, Eisenacher M, Schlaak JF, Baba HA, Meyer HE, Sitek B.** 2013. Proteomic differences between hepatocellular carcinoma and nontumorous liver tissue investigated by a combined gel-based and label-free quantitative proteomics study. Mol

Cell Proteomics **12:**2006–2020. http://dx.doi.org/10.1074/mcp.M113.028027.

24. **Borisenko IA, Viazovichenko Iu E, Gudkov VI.** 1994. An improvement in information support in the interests of the epidemiological health welfare of the troops. Voenno-Meditsinskii Zhurnal **37–42:**80. (In Russian.)
25. **Li YP, Ramirez S, Jensen SB, Purcell RH, Gottwein JM, Bukh J.** 2012. Highly efficient full-length hepatitis C virus genotype 1 (strain TN) infectious culture system. Proc Natl Acad Sci U S A **109:**19757–19762. http://dx.doi.org/10.1073/pnas.1218260109.
26. **Steinmann E, Brohm C, Kallis S, Bartenschlager R, Pietschmann T.** 2008. Efficient trans-encapsidation of hepatitis C virus RNAs into infectious virus-like particles. J Virol **82:**7034–7046. http://dx.doi.org/10.1128/JVI.00118-08.
27. **Marukian S, Jones CT, Andrus L, Evans MJ, Ritola KD, Charles ED, Rice CM, Dustin LB.** 2008. Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. Hepatology **48:**1843–1850. http://dx.doi.org/10.1002/hep.22550.
28. **Lindenbach BD, Evans MJ, Syder AJ, Wolk B, Tellinghuisen TL, Liu CC, Maruyama T, Hynes RO, Burton DR, McKeating JA, Rice CM.** 2005. Complete replication of hepatitis C virus in cell culture. Science **309:**623–626. http://dx.doi.org/10.1126/science.1114016.
29. **Moradpour D, Wakita T, Tokushige K, Carlson RI, Krawczynski K, Wands JR.** 1996. Characterization of three novel monoclonal antibodies against hepatitis C virus core protein. J Med Virol **48:**234–241. http://dx.doi.org/10.1002/(SICI)1096-9071(199603)48:3<234::AID-JMV4>3.0.CO;2-9.
30. **Kim AY, Kuntzen T, Timm J, Nolan BE, Baca MA, Reyor LL, Berical AC, Feller AJ, Johnson KL, Schulze zur Wiesch J, Robbins GK, Chung RT, Walker BD, Carrington M, Allen TM, Lauer GM.** 2011. Spontaneous control of HCV is associated with expression of HLA-B 57 and preservation of targeted epitopes. Gastroenterology **140:**686–696 e681. http://dx.doi.org/10.1053/j.gastro.2010.09.042.
31. **van Endert P.** 2011. Post-proteasomal and proteasome-independent generation of MHC class I ligands. Cell Mol Life Sci **68:**1553–1567. http://dx.doi.org/10.1007/s00018-011-0662-1.
32. **Kimura Y, Gushima T, Rawale S, Kaumaya P, Walker CM.** 2005. Escape mutations alter proteasome processing of major histocompatibility complex class I-restricted epitopes in persistent hepatitis C virus infection. J Virol **79:**4870–4876. http://dx.doi.org/10.1128/JVI.79.8.4870-4876.2005.
33. **Seifert U, Liermann H, Racanelli V, Halenius A, Wiese M, Wedemeyer H, Ruppert T, Rispeter K, Henklein P, Sijts A, Hengel H, Kloetzel PM, Rehermann B.** 2004. Hepatitis C virus mutation affects proteasomal epitope processing. J Clin Invest **114:**250–259. http://dx.doi.org/10.1172/JCI200420985.
34. **Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, Mullins JI, Walker BD, Harrigan PR, Goulder PJ, Heckerman D.** 2008. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol **4:**e1000225. http://dx.doi.org/10.1371/journal.pcbi.1000225.
35. **Allen TM, Altfeld M, Yu XG, O'Sullivan KM, Lichterfeld M, Le Gall S, John M, Mothe BR, Lee PK, Kalife ET, Cohen DE, Freedberg KA, Strick DA, Johnston MN, Sette A, Rosenberg ES, Mallal SA, Goulder PJ, Brander C, Walker BD.** 2004. Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. J Virol **78:**7069–7078. http://dx.doi.org/10.1128/JVI.78.13.7069-7078.2004.
36. **Milicic A, Price DA, Zimbwa P, Booth BL, Brown HL, Easterbrook PJ, Olsen K, Robinson N, Gileadi U, Sewell AK, Cerundolo V, Phillips RE.** 2005. $CD8^+$ T cell epitope-flanking mutations disrupt proteasomal processing of HIV-1 Nef. J Immunol **175:**4618–4626. http://dx.doi.org/10.4049/jimmunol.175.7.4618.
37. **Sijts EJ, Kloetzel PM.** 2011. The role of the proteasome in the generation of MHC class I ligands and immune responses. Cell Mol Life Sci **68:**1491–1502. http://dx.doi.org/10.1007/s00018-011-0657-y.
38. **Hammer GE, Kanaseki T, Shastri N.** 2007. The final touches make perfect the peptide-MHC class I repertoire. Immunity **26:**397–406. http://dx.doi.org/10.1016/j.immuni.2007.04.003.
39. **Serwold T, Gonzalez F, Kim J, Jacob R, Shastri N.** 2002. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. Nature **419:**480–483. http://dx.doi.org/10.1038/nature01074.
40. **Zhang SC, Martin E, Shimada M, Godfrey SB, Fricke J, Locastro S, Lai NY, Liebesny P, Carlson JM, Brumme CJ, Ogbechie OA, Chen H, Walker BD, Brumme ZL, Kavanagh DG, Le Gall S.** 2012. Aminopeptidase substrate preference affects HIV epitope presentation and predicts immune escape patterns in HIV-infected individuals. J Immunol **188:**5924–5934. http://dx.doi.org/10.4049/jimmunol.1200219.
41. **Schatz MM, Peters B, Akkad N, Ullrich N, Martinez AN, Carroll O, Bulik S, Rammensee HG, van Endert P, Holzhutter HG, Tenzer S, Schild H.** 2008. Characterizing the N-terminal processing motif of MHC class I ligands. J Immunol **180:**3210–3217. http://dx.doi.org/10.4049/jimmunol.180.5.3210.
42. **van Endert PM, Tampe R, Meyer TH, Tisch R, Bach JF, McDevitt HO.** 1994. A sequential model for peptide binding and transport by the transporters associated with antigen processing. Immunity **1:**491–500. http://dx.doi.org/10.1016/1074-7613(94)90091-4.
43. **Uebelhoer L, Han JH, Callendret B, Mateu G, Shoukry NH, Hanson HL, Rice CM, Walker CM, Grakoui A.** 2008. Stable cytotoxic T cell escape mutation in hepatitis C virus is linked to maintenance of viral fitness. PLoS Pathog **4:**e1000143. http://dx.doi.org/10.1371/journal.ppat.1000143.
44. **Kasprowicz V, Kang YH, Lucas M, Schulze zur Wiesch J, Kuntzen T, Fleming V, Nolan BE, Longworth S, Berical A, Bengsch B, Thimme R, Lewis-Ximenez L, Allen TM, Kim AY, Klenerman P, Lauer GM.** 2010. Hepatitis C virus (HCV) sequence variation induces an HCV-specific T-cell phenotype analogous to spontaneous resolution. J Virol **84:**1656–1663. http://dx.doi.org/10.1128/JVI.01499-09.
45. **Bengsch B, Seigel B, Ruhl M, Timm J, Kuntz M, Blum HE, Pircher H, Thimme R.** 2010. Coexpression of PD-1, 2B4, CD160 and KLRG1 on exhausted HCV-specific $CD8^+$ T cells is linked to antigen recognition and T cell differentiation. PLoS Pathog **6:**e1000947. http://dx.doi.org/10.1371/journal.ppat.1000947.
46. **Wherry EJ.** 2011. T cell exhaustion. Nat Immunol **12:**492–499.

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 10 November 2022
DOI 10.3389/fimmu.2022.1045498

OPEN ACCESS

EDITED BY
Michael H. Lehmann,
Ludwig-Maximilians-Universität München, Germany

REVIEWED BY
Antonio Bertoletti,
Duke-NUS Medical School, Singapore
Vincenzo Barnaba,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Jörg Timm
timm@hhu.de

SPECIALTY SECTION
This article was submitted to Viral Immunology, a section of the journal Frontiers in Immunology

RECEIVED 15 September 2022
ACCEPTED 19 October 2022
PUBLISHED 10 November 2022

CITATION
Walker A, Schwarz T, Brinkmann-Paulukat J, Wisskirchen K, Menne C, Alizei ES, Kefalakes H, Theissen M, Hoffmann D, Schulze zur Wiesch J, Maini MK, Cornberg M, Kraft ARM, Keitel V, Bock HH, Horn PA, Thimme R, Wedemeyer H, Heinemann FM, Luedde T, Neumann-Haefelin C, Protzer U and Timm J (2022) Immune escape pathways from the HBV core$_{18-27}$ CD8 T cell response are driven by individual HLA class I alleles. *Front. Immunol.* 13:1045498. doi: 10.3389/fimmu.2022.1045498

Immune escape pathways from the HBV core$_{18-27}$ CD8 T cell response are driven by individual HLA class I alleles

Andreas Walker[1], Tatjana Schwarz[1], Janine Brinkmann-Paulukat[1], Karin Wisskirchen[2,3], Christopher Menne[1], Elahe Salimi Alizei[4], Helenie Kefalakes[5], Martin Theissen[6], Daniel Hoffmann[6], Julian Schulze zur Wiesch[7,8], Mala K. Maini[9], Markus Cornberg[10,11], Anke RM Kraft[10,11], Verena Keitel[12], Hans H. Bock[12], Peter A. Horn[13], Robert Thimme[4], Heiner Wedemeyer[10,11], Falko M. Heinemann[13], Tom Luedde[12], Christoph Neumann-Haefelin[4], Ulrike Protzer[2,3] and Jörg Timm[1]*

[1]Institute of Virology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [2]Institute of Virology, School of Medicine, Technical University of Munich, Helmholtz Zentrum München, Munich, Germany, [3]German Center for Infection Research (DZIF), Site Munich, Munich, Germany, [4]Department of Medicine II, University Hospital Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany, [5]Institute of Virology, University of Duisburg-Essen, University Hospital Essen, Essen, Germany, [6]Research Group Bioinformatics, Faculty of Biology, University of Duisburg-Essen, Essen, Germany, [7]Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [8]German Center for Infection Research (DZIF), Site Hamburg, Hamburg, Germany, [9]Division of Infection and Immunity, Institute of Immunity and Transplantation, University College London, London, United Kingdom, [10]Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany, [11]German Center for Infection Research (DZIF), Site Hannover, Hannover, Germany, [12]Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [13]Institute for Transfusion Medicine, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

Background and aims: There is growing interest in T cell-based immune therapies for a functional cure of chronic HBV infection including check-point inhibition, T cell-targeted vaccines or TCR-grafted effector cells. All these approaches depend on recognition of HLA class I-presented viral peptides. The HBV core region 18-27 is an immunodominant target of CD8+ T cells and represents the prime target for T cell-based therapies. Here, a high-resolution analysis of the core$_{18-27}$ specific CD8+ T cell and the selected escape pathways was performed.

Methods: HLA class I typing and viral sequence analyses were performed for 464 patients with chronic HBV infection. HBV-specific CD8+ T-cell responses against the prototype and epitope variants were characterized by flow cytometry.

Results: Consistent with promiscuous presentation of the core$_{18-27}$ epitope, antigen-specific T cells were detected in patients carrying HLA-A*02:01, HLA-B*35:01, HLA-B*35:03 or HLA-B*51:01. Sequence analysis confirmed

reproducible selection pressure on the core$_{18-27}$ epitope in the context of these alleles. Interestingly, the selected immune escape pathways depend on the presenting HLA-class I-molecule. Although cross-reactive T cells were observed, some epitope variants achieved functional escape by impaired TCR-interaction or disturbed antigen processing. Of note, selection of epitope variants was exclusively observed in HBeAg negative HBV infection and here, detection of variants associated with significantly greater magnitude of the CD8 T cell response compared to absence of variants.

Conclusion: The core$_{18-27}$ epitope is highly variable and under heavy selection pressure in the context of different HLA class I-molecules. Some epitope variants showed evidence for impaired antigen processing and reduced presentation. Viruses carrying such escape substitutions will be less susceptible to CD8+ T cell responses and should be considered for T cell-based therapy strategies.

KEYWORDS

hepatitis B virus, CD8 T cell response, TCR-engineered T cells, immune escape, chronic infection

Introduction

Although, a prophylactic vaccine against hepatitis B virus (HBV) is available, persistent infections associated with chronic liver disease are still a global health problem. Chronic HBV-infections can be treated with IFNα or nucleot(s)ide analogues. However, a functional cure with persistently undetectable HBV-DNA and absence of liver inflammation is only rarely achieved (1, 2). Accordingly, there is growing interest in novel strategies leading to clearance of chronic HBV-infection. There is strong evidence that the immune response by CD8$^+$ T cells contributes to sustained immune control of HBV infection (3–5). In turn, in chronic infection HBV-specific CD8$^+$ T cells exhibit an exhausted phenotype with upregulation of inhibitory receptors and progressive dysfunction (4, 6). One of the proposed strategies for novel treatment interventions is therefore a combination of antiviral treatment with immune therapies including immune checkpoint inhibitors, T-cell inducing vaccines or transfer of TCR-grafted effector cells (4, 6, 7). These T cell-based therapies ultimately rely on presentation of viral epitopes by HLA class I molecules on infected hepatocytes. Given the sequence diversity of HBV, there is concern that viral variants in important epitopes may impact the efficacy of such novel treatment strategies.

We and others have shown previously, that viral variants in targeted CD8$^+$ T cell epitopes impair the immune response (8, 9). Indeed, enrichment of certain viral sequence polymorphisms in epitopes presented by particular HLA class I molecules strongly suggest that epitope variants are selected as an immune evasion mechanism to escape from CD8$^+$ T cell selection pressure. The region core$_{18-27}$ represents an immunodominant HLA-A*02-restricted epitope that is well described in the literature (10–18). Interestingly, this epitope is a promiscuous binder and is also presented by HLA-B*35 and HLA-B*51 (9, 12). Given the high reproducibility of T cell responses against this epitope, it is considered an attractive target for T cell-based immune therapies. Here, we provide a high-resolution analysis of the HLA class I subtypes restricting the CD8$^+$ T cell immune response to this epitope region, the selected HLA class I subtype-dependent pathways to immune escape and the responsible immune escape mechanisms.

Material and methods

Patients

In a multi-center effort for analysis of HLA class I-associated selection pressure 464 HBV patients were recruited at the Hepatology Units in Düsseldorf, Essen, Hamburg, Hannover, Freiburg (all Germany) and London (UK); (Supplemental Table 1). Only HBV genotype A and D infected patients were

included. Informed consent was obtained from each patient, and the study protocol was approved by the local ethics committee of the Medical Faculty of Düsseldorf in accordance with the guidelines of the Declaration of Helsinki. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll density gradient centrifugation (Biocoll; Biochrom) (9). DNA for HLA-typing was extracted from PBMCs using spin columns (Qiagen). High resolution (second field) HLA-A and HLA-B typing was performed by use of sequence-specific oligonucleotides (LABType^{TM}) methodology (19), provided by One Lambda (Thermo Fisher Inc.) at department of transfusion medicine of the University Hospital Essen.

Amplification and sequence analysis of the HBV core region

Two-step nested PCRs were performed with GoTaq HotStart-Polymerase (Promega) Polymerase according to the manufacturer's protocol and the following primer combinations for PCR-I: TS-1585_F_int (TTCGCTTCACCTCTGCACGT); TS-2419_R (GCGACGCGGNGATTGAGAYCT) and PCR-II: HBVCoreF (TGTCAACGACCGACCTTGAGG); TS-2397_R_int (CGTCTGCGAGGYGAGGGAGTTC). Per reaction 95 µl PCRI mixture containing 1x GoTaq polymerase buffer, 200 µM dNTPs (Bio-Budget), 0.5 µM each Primer and 1.25 units polymerase were mixed with 5 µl HBV-DNA. PCRII mixes were identical to PCRI except the final volume of 97 µl. PCR condition were 180 s at 94°C followed by 35 cycles each 30 s 95°C, 30 s 55°C and 120 s 72°C followed by 10 min at 72°C and hold at 10°C. Subsequently, three microliter of the first round PCR-product was used for the second round of nested-PCR with the same PCR conditions. PCR products were purified with the QIAquick PCR-Purification Kit (Qiagen, Hilden) and Sanger sequenced with sequencing primer HBVCoreF and TS-2397_R_int. For high-throughput sequencing Core-PCRs were sent to an external provider (SeqIt Kaiserlautern, Germany) and were sequenced on a MiSeq2 (Illumina) with a mean coverage of 15-20.000x. All obtained sequences were aligned with the software Geneious 10.2.6 (RRID : SCR_010519). Sequences were submitted to Genbank and are available under accession numbers (MZ043025-MZ043097; MZ097624-MZ097884).

Analysis of HBV-specific CD8$^+$ T cells

HBV-specific CD8$^+$T cells were detected after antigen-specific expansion as previously described (20). Briefly, PBMCs were resuspended in RPMI medium containing 10% fetal calf serum and stimulated with individual peptides (1 µg/mL), anti-CD28/CD49d (0.5 µg/mL; BD Biosciences) and recombinant interleukin-2 (20 U/mL; Hoffmann-La Roche). On day 10 the cells were restimulated with prototype or variant peptides and secretion of interferon-γ (IFNγ) was analyzed by intracellular cytokine staining (ICS) and subsequent analysis on a FACS Canto (Becton Dickinson).

Analysis of the HLA-restriction of antigen-specific CD8$^+$ T cells

Partially HLA-matched PBMCs from healthy donors expressing either HLA-A*02:01, HLA-B*35:01 or HLA-B*51:01 were used as targets for antigen-specific CD8 T cells as effector cells, that have been expanded from PBMCs of HBV infected patients in the presence of the core 18-27 prototype peptide. The target cells were pulsed with the peptide (10µg/ml) overnight at 37°C and carefully washed five 5 times in PBS the next day. After washing 200.000 target cells were co-cultured with 200.000 effector cells in a 24 well plate for 5h followed by a standard ICS. Exogenously added peptide in the absence of target cells served as a positive control, target cells cultured overnight in the absence of peptide served as a negative control. The frequency of IFNγ+ CD8+ T cells in the negative control was considered as background and was subtracted from the results in the presence of peptide.

Generation of HepG2 cells stably expressing the HBV-core-mCherry fusion protein

For generation of stable cell lines the GFP sequence in the bicistronic pWPI vector (kindly provided by D. Trono, EPFL) was replaced by a Blasticidin-S deaminase and the synthetic sequence for an HBV core$_{aa1-183}$-mCherry fusion protein (Eurofins genomics) was inserted by Gibson assembly (NEB). The different variants in the core$_{18-27}$ sequence were introduced by site-directed mutagenesis. Generation of lentiviral pseudoparticles was described previously (21). In brief, 2.5×10^6 293T cells were co-transfected with pWPI-core-mCherry-BSD, pCMVR8.74 and pMD2.G using Mirus TransIT$^®$-LT1 (Mirus) transfection reagent and pseudoparticles were harvested after 48h and 72h. For transduction 2x 10^6 HepG2 cells were spinoculated (MOI 1-5) for 30min at 37°C and 700g and seeded in DMEM. Twenty-four hours after infection cells were

10.3389/fimmu.2022.1045498

selected with 25 µg/ml Blasticidin (*In vivo*gen). HEK293T and HepG2 cells were cultured in Dulbecco's modified Eagle Medium (2mML-glutamine, 10 mM HEPES, nonessential amino acids, 100 U of penicillin/ml, 100 µg of streptomycin/ml, 10% fetal calf serum).

Analysis of the CD8$^+$ T cell response against endogenously processed antigens

Stable HBV-core-mCherry expressing HepG2 cells were seeded at a density of 2x10^5 cells/well in a 6-well Plate in DMEM without blasticidin. The next day, 2x10^{5}HBV core-specific CD8$^+$ T cells were co-cultured for 4h with the HepG2 cells before IFN-γ was analyzed by ICS. As control HepG2 cells were incubated with 10 mg/ml prototype or variant peptide overnight, washed extensively and were then used as targets as above described.

Results

Influence of the HLA class I subtype on the CD8$^+$ T cell response against the core$_{18\text{-}27}$ region

The HLA-A*02 restricted epitope in HBV core$_{18\text{-}27}$ is frequently targeted by patients and is a prime target for T cell therapy (22). Since there is evidence, that the core$_{18\text{-}27}$ region is also immunogenic in the context of HLA-A*02, HLA-B*35 and HLA-B*51 alleles (11, 12, 20), the HLA-restriction was analyzed in more detail. Therefore, the CD8$^+$ T cell immune response against HBV core$_{18\text{-}27}$ was determined in 154 chronic HBV infected patients. PBMCs were cultured for 10 days in the presence of the core$_{18\text{-}27}$ peptide and the frequency of IFNγ-producing cells upon re-stimulation with the peptide was determined by flow cytometry. Notably, in this analysis only HBeAg negative patients showed detectable CD8$^+$ T cell responses directed against this epitope (Figures 1A; Supplemental Figure S2). Accordingly, in a subsequent analysis

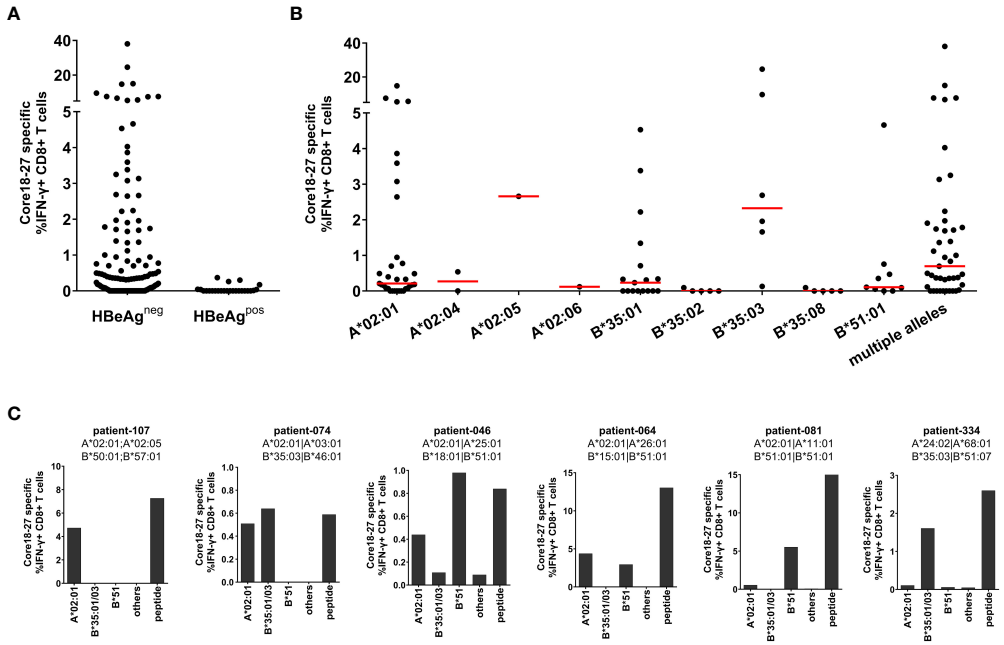


FIGURE 1
The core$_{18\text{-}27}$ epitope is restricted by multiple HLA class I molecules. PBMCs from 154 patients with chronic HBV infection were stimulated for 10 days in the presence of the core$_{18\text{-}27}$ peptide and re-stimulated for 5h before IFNγ production was determined by ICS. Depicted are IFNγ producing HBV core 18-27 specific CD8 T cells, the IFNγ response from non-stimulated cell was subtracted **(A)** IFNγ responses according to the HBeAg status. **(B)** IFNγ responses according to the HLA-type of the HBeAg negative patients from **(A)**. Patients depicted with single HLA-types had no other relevant HLA-type (e.g. A*02:01 are patients negative for HLA-B*35:01/03 or HLA-B*51 and vice versa), patients with "multiple alleles" had at least two relevant HLA-alleles (e.g. A*02:01 and B*51). **(C)** The HLA-restriction of antigen-specific CD8$^+$ T cells was determined with partially HLA-matched peptide-pulsed target cells expressing only HLA-A*02:01 (not B*35 or B*51), B35:01/03 (not A*02:01 or B*51), HLA-B*51 (not A*02:01 or B35), others (not A*02:01, B*35 or B*51). As control core18-27 peptide was directly added to the culture (peptide). The frequency of IFNγ -producing CD8$^+$ T cells was determined by ICS and a representative example is shown in Supplemental Figure S3.

of the relevant HLA class I-alleles for the $CD8^+$ T cell response, only HBeAg negative patients were included. Consistent with promiscuous binding and presentation of the peptide, antigen-specific production of IFNγ by $CD8^+$ T cells was detected in patients carrying HLA-A*02, HLA-B*35 and HLA-B*51 (Figure 1B).

Importantly, there were differences in the $CD8^+$ T cell response between different HLA class I-subtypes. In the context of HLA-B*35 only patients carrying the subtype HLA-B*35:01 and B*35:03 mounted an immune response against the epitope whereas patients with HLA-B*35:02 and HLA-B*35:08 did not. The most robust and reproducible $CD8^+$ T cell response was detected in HLA-B*35:03 positive patients. In the group of HLA-A*02 positive patients the most frequent subtype was HLA-A*02:01. Although the number of patients with non-HLA-A*02:01 subtypes was too low to draw solid conclusions, at least one patient with HLA-A*02:05 also mounted a robust immune response against the epitope. In the group of HLA-B*51 positive patients, all analysed patients had the subtype HLA-B*51:01. Interestingly, the group of patients carrying multiple relevant HLA class I-alleles showed a higher median of the $CD8^+$ T cell response (0.69%) than HLA-A*02:01 (median 0.21%), HLA-B*35:01 (median 0.23%) and HLA-B*51:01 (median 0.10%) positive patients (Figure 1B), suggesting that presence of multiple relevant HLA class I-alleles might be associated with a more robust immune response.

The restricting HLA-molecule was further analysed in six patients carrying one or two relevant HLA class I alleles, with heterologous partially HLA-matched peptide-pulsed target cells. These analyses revealed, that in some patients two distinct $CD8^+$ T cell immune responses were detectable, each restricted by different HLA class I molecules (Figure 1C, patients 074, 046 and 064). In other cases, only one HLA class I molecule dominated the response despite presence of a second relevant allele (Figure 1C, patient 081). Taken together, the $core_{18-27}$ region is presented by multiple HLA class I subtypes with the strongest evidence for $CD8^+$ T cell immune responses in the context of HLA-A*02:01, HLA-B*35:01, HLA-B*35:03 and HLA-B*51:01.

Sequence analysis of the $core_{18-27}$ region

Earlier studies suggested that naturally occurring sequence variants in $core_{18-27}$ were functionally associated with immune escape (8) and were a product of selection pressure in HLA-A*02-positive individuals (9). To decipher the individual effect of each HLA-molecule on selection pressure, we analyzed core sequences from 409 HBV-patients, infected with genotype A or D. The cohort included 83 patients positive for HLA-A*02:01 (not other HLA-A*02-subtypes, B*35 or B*51), 13 patients positive for HLA-A*02:xx (not HLA-A*02:01, B*35 or B*51), 62 patients positive for B*35:01 or B*35:03 (not A*02 or B*51), 35 patients positive for HLA-B*51:01 (not A*02 or B*35), 104 patients with combinations of these four HLA class I alleles and 112 patients negative for all four HLA alleles.

Overall, the epitope region is highly polymorphic with frequent substitutions predominantly in positions 4, 7 and 10 of the epitope (Figure 2A). The frequency of any variation from the prototype sequence in the absence of the relevant HLA class I alleles was 23% (26 of 112) with the S4A substitution being the most frequent (8%). Notably, the frequency of the S4A substitution was not significantly different in the presence of HLA*02, HLA-B*35:01/03 or HLA-B*51:01 suggesting that this substitution was not driven by CD8 T cell pressure in the context of these alleles. However, the frequency of sequences with other substitutions in the epitope region was significantly enriched in patients with these relevant HLA class I alleles. For example, in HLA-B*35:01/03 positive patients an S4T substitution was reproducibly selected. Here, ten of 62 HLA-B*35:01/03 positive patients (16.1%) had the S4T substitution. This was in contrast to HLA-A*02 positive patients and patients lacking any of the relevant alleles, where this substitution was not or only rarely observed ($p<0.0001$) (Figures 2A, B). Of note, also in HLA-B*51:01 positive patients substitutions were preferentially selected in position 4 of the epitope, but in addition to S4T, other mutations such as S4G, S4V and other rare substitutions were observed that were nearly undetectable in the absence of any of the relevant alleles. In contrast, in HLA-A*02:01 positive patients the F7Y substitution alone or in combination with substitutions in position 4 (e.g. S4H) was significantly enriched compared to HLA-B*51 or HLA-B*35:01/03 positive patients or patients lacking the relevant HLA alleles ($p<0.05$ and $p<0.0001$; Figures 2A, B). The selective enrichment of the S4T substitution in HLA-B*35:01/03 positive patients and of the F7Y substitution in HLA-A*02:01 positive patients suggests that the pathways to $CD8^+$ T cell escape differ between the presenting HLA class I molecules. Phylogenetic analysis showed no founder effect or genotype association of the substitutions (Supplemental Figure S1) supporting the conclusion that the variants were selected by individual immune pressure.

To analyze the epitope diversity at the quasispecies level, the core-region was amplified from 96 randomly-chosen patients and sequenced on a MiSeq Illumina platform with a median coverage of 12,358-fold (range 1,028 - 52,336-fold). The frequency of individual epitope variants within the quasispecies is shown in Figure 2C. The prototype sequence is shown in black and each variant is color-coded as indicated. HBeAg positive patients are marked above the columns. For clarity, variants including the S4T substitution are coded in shades of green. Variants including the F7Y substitution are coded in shades of red. Variants with two substitutions in position 4 and 10 are coded in shades of pink. Interestingly,

10.3389/fimmu.2022.1045498

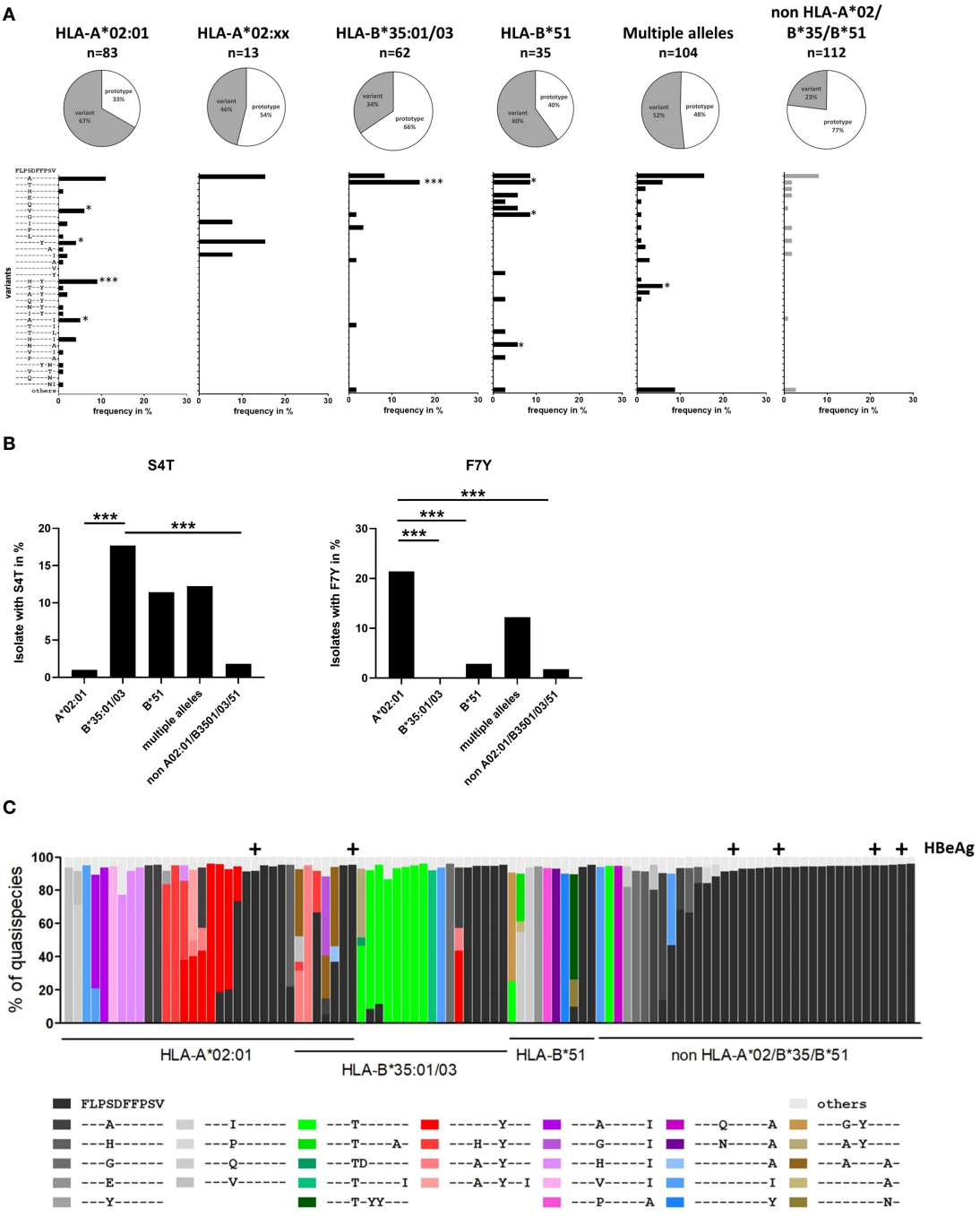


FIGURE 2
Frequency of sequence polymorphisms in the CD8$^+$ T-cell epitope HBV core$_{18\text{-}27}$. **(A)** 717 bp amplicon from the core region was amplified and Sanger sequenced from 409 patients with known HLA-type and chronic HBV infection (genotype A or D). The pie charts show the frequencies of core$_{18\text{-}27}$ variants in patients with the depicted HLA-type. The lower panels show the frequencies of individual epitope variants for isolates from patients carrying only HLA-A*02:01, HLA-A*02 subtypes others than A*02:01 (A*02:xx), only HLA-B*35:01/03, only HLA-B*51:01, multiple alleles or none of the relevant alleles. Epitope variants that are significantly more frequent compared to patients lacking any of the relevant HLA class alleles by Fisher's exact test are indicated (*$p < 0.05$; ***$p < 0.001$). **(B)** Frequency of sequence polymorphisms in the core$_{18\text{-}27}$ epitope. Cumulative frequencies of sequence polymorphism were calculate regardless of other sequence polymorphism in the core 18-27 epitope for S4T (left) and F7Y (Right). **(C)** Analysis of HLA-associated frequencies of individual variants in the core$_{18\text{-}27}$ region within the quasispecies. Amplicons of 96-randomly chosen patients from **(A)** were submitted to high-throughput sequencing with a mean coverage of 15-20.000x. Reads were aligned to the patient-specific reference sequence generated by Sanger sequencing. Each bar represents the core$_{18\text{-}27}$ region of one patient with the individual epitope variants being color coded as indicated. Only variants with frequencies >5% of the quasispecies are shown. Samples from HBeAg-positive patients are marked above the column with (+).

detection of multiple epitope variants within the quasispecies was common in patients carrying any of the relevant HLA class I alleles, which would be consistent with continuous selection. In line with the population sequence data, variants were enriched in patients carrying any of the relevant HLA class I alleles and there was a clear difference between the HLA-types. In HLA-A*02:01 positive patients, variants with the F7Y substitution (shades of red) were enriched, while the variant S4T did not occur at all, not even in minor frequencies. In HLA-B*35:01/03 positive patients variants with the S4T substitution (shades of green) were more frequent and often the major variant.

Absence or lower frequencies of substitutions in patients lacking the relevant HLA-molecules suggests negative selection of epitope variants in the core18-27 region in the absence of immune pressure. We therefore tested, if the substitutions in the epitope region were associated with lower HBV-DNA concentrations in patients. Therefore, HBV-DNA concentrations and the core18-27 sequence were analyzed in treatment-naïve patients (n=280). The patients were classified according to their serological HBeAg status. As expected, the viral load was significantly higher in HBeAg-positive patients than in HBeAg-negative patients (Figure 3A). Interestingly, in the HBeAg positive group all isolates carried the prototype sequence consistent with absence of selection pressure by CD8 T cells. In HBeAg negative patients, there was no significant difference in viral load between prototype and the different variants. It was also addressed, if epitope variants associated with a distinct magnitude of the CD8 T cell response after 10 days of antigen-specific expansion. As no CD8 T cell responses and no epitope variants were detected in HBeAg positive patients (Figure 1A), HBeAg positive individuals were excluded from the analysis to avoid a bias based on the HBeAg status (Figure 3B). Notably, despite exclusion of HBeAg positive patients, infection with a variant was associated with a significantly stronger T-cell response compared to infection with the prototype (Figure 3B; p=0.0006), consistent with selection pressure in these patients.

Cross-reactivity of CD8$^+$ T cells with variants of the core$_{18\text{-}27}$ epitope

To analyze the impact of epitope variation on the CD8$^+$ T cell response, PBMCs were expanded in the presence of the prototype epitope and restimulated at day 10 with different epitope variants. The peptides included significantly enriched variants in the presence of HLA-A*02:01 or HLA-B*35:01/03 and additional variants that were detect in the presence of these alleles without statistical support for selection as well as variants that were not detected in the presence of these allele. The frequency of IFNγ-producing cells was determined by flow cytometry (Figure 4A, left panel and Supplemental Figure S4) and the results were normalized to the prototype response (Figure 4A, middle and right panel). In line with functional immune escape from HLA-A*02:01-restricted CD8$^+$ T cells the double variant S4H/F7Y and to a lesser extent the S4H/V10I variant showed substantial reduction in the IFNγ response (Figure 4A). Interestingly, the variant carrying the single F7Y

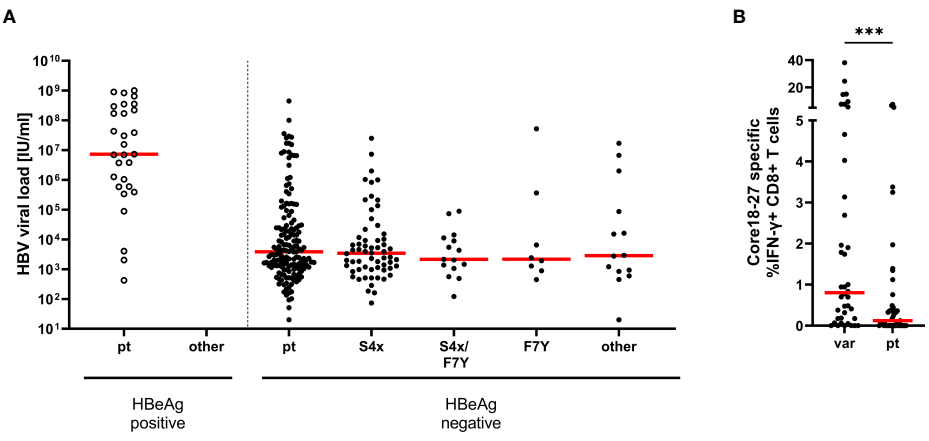


FIGURE 3
Association between HBV-DNA and variants in the epitope core$_{18\text{-}27}$. **(A)** The HBV-DNA concentration is shown for 280 treatment-naive patients (genotype A and D) according to the absence and presence of substitutions in position 4 and 7 of the epitope. Patients were grouped according their epitope sequence and their serological HBeAg status as indicated. pt = prototype, S4X = any variant at position 4. **(B)** core$_{18\text{-}27}$-specific CD8$^+$ T cells responses from patient samples shown in Figure 1A, in HBeAg negative patients according to their epitope sequence (Supplemental Table 1); var = any variation in the core 18-27 epitope; pt = prototype. Statistical significance was calculated with the Mann-Whitney-test (p = 0.0001). *** = p<0.001.

10.3389/fimmu.2022.1045498

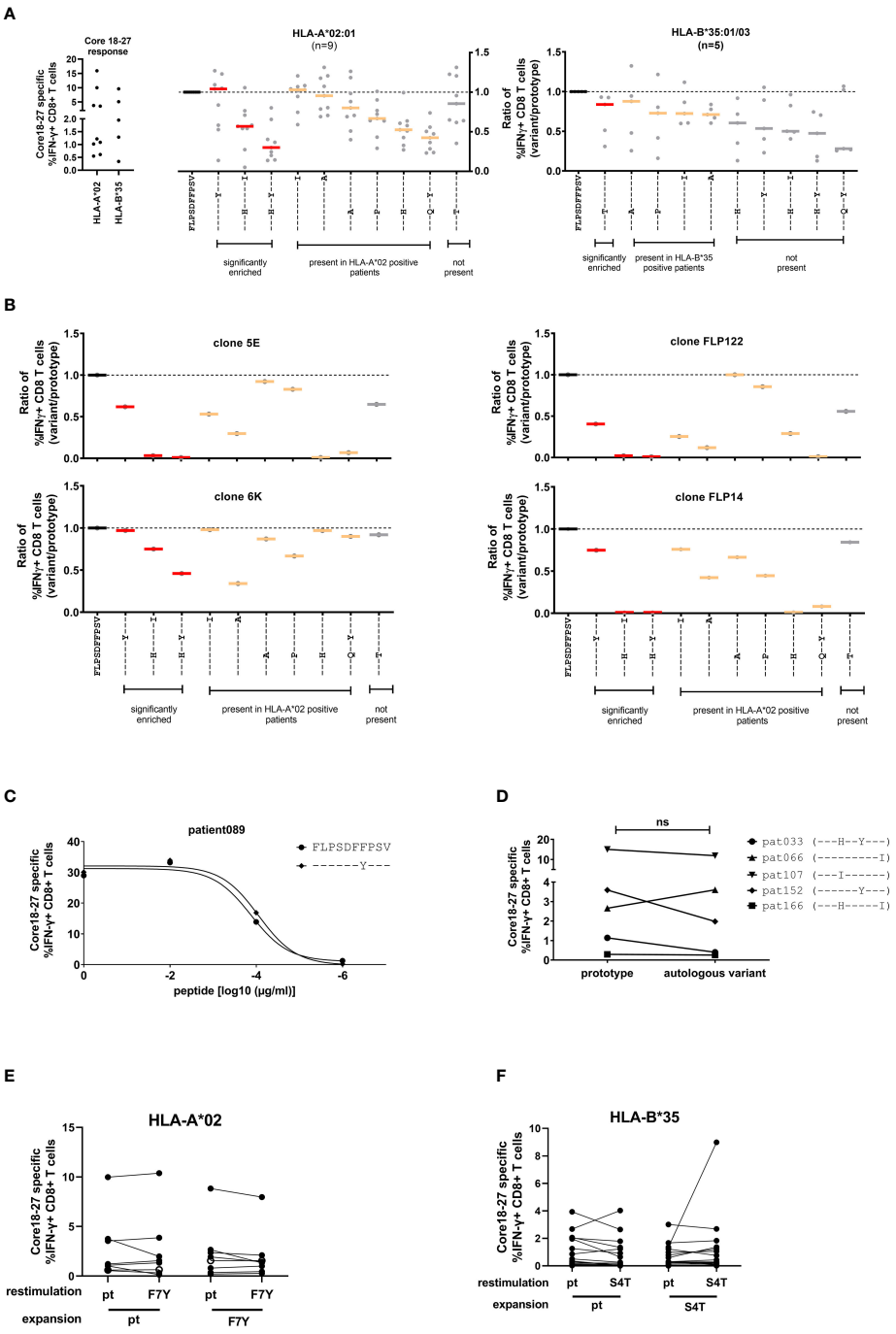


FIGURE 4
Cross-reactivity of core$_{18\text{-}27}$-specific CD8^{CD8+} T cells with naturally occurring sequence variants. After antigen-specific expansion for 10 days, core$_{18\text{-}27}$-specific CD8^{+} T cells were restimulated with the most common variants in naturally occurring isolates from Figure 2 and IFNγ-producing CD8^{+} T cells were determined by ICS. For comparison the responses were normalized to the prototype (FLPSDFFPSV) response. **(A)** *Left:* the core 18-27 specific CD8 T cell response of the patients used for the cross-reactivity analysis before normalization. Middle/Right: Cross-reactivity of core$_{18\text{-}27}$-specific CD8^{+} T cells with peptide variants in HLA-A*02:01 or HLA-B*35:01/03 positive patients. The response against variants with statistical support for selection on the sequence level (Figure 2) are marked in red. **(B)** Cross-reactivity of HLA-A*02 restricted grafted T effector cells. **(C)** The CD8^{+} T-cell response against different concentrations of the prototype peptide or the F7Y variant determined by ICS. **(D)** The core 18-27 CD8^{+}-T cell response against the prototype or the autologous viral sequence present in the corresponding patient. **(E)** Expansion of core 18-27 CD8^{+}-T cells with the prototype or F7Y-variant peptide and restimulation with the indicated peptide in HLA-A*02 positive patients (n=8) **(F)** Expansion of core 18-27 CD8^{+}-T cells with the prototype or S4T-variant peptide and restimulation with the indicated peptide in HLA-B*35 positive patients (n=20). ns = not significant p >0.05.

substitution, that was reproducibly selected in HLA-A*02:01 positive patients, showed no clear tendency towards functional immune escape in this assay. Notably, the same degree of cross-reactivity of the F7Y variant was observed, when serial dilutions of the peptide were tested, suggesting that the functional avidity was not impaired by the variant peptide (Figure 4C). Also in HLA-B*35:01/03 positive individuals, the selected S4T substitution was highly cross-reactive (Figure 4A). Importantly, there was substantial variation between individuals regarding the degree of cross-reactivity with epitope variants. Therefore, the cross-reactivity of individual T cell receptors was determined by utilizing $CD8^+$ T cells transduced with defined TCRs from donors with an HLA-A*02:01-restricted T cell response against this epitope (16, 23) (Figure 4B; Supplemental Figure S5). Indeed, the cross-reactivity profiles differed between individual TCRs. The TCR 6K showed substantial cross-reactivity with the selected variants including the variants carrying double substitutions. In contrast, the other TCRs were more sensitive to substitutions in the epitope. Interestingly, even here, the F7Y variant was reactive for all four TCRs, suggesting that it does not represent an optimal TCR-escape mutation.

Given the individual differences in the cross-reactivity patterns between patients and TCRs, we hypothesized, that the selected viral variant in a given host is the result of a highly individualized process. We therefore tested in patients with chronic HBV-infection the cross-reactivity with the individual's autologous virus (Figure 4D). Again, in most cases we did not find clear evidence for functional immune escape with the variants. We also tested the ability of the variant peptides F7Y and S4T to expand antigen-specific CD8 T cells in HLA-A*02:01 or HLA-B*35:01 positive individuals (Figures 4E, F; Supplemental Figure S6). Again, there was no clear reduction of the IFNγ response by the variant. The frequency of antigen-specific CD8 T cells expanded with the variant peptide was at the same level compared to prototype and showed the same degree of cross-reactivity (Figures 4E, F). Of note, in line with previous results (24, 25), there were also no differences regarding the maturation state and expression of inhibitory receptors on antigen-specific $CD8^+$ T cells when compared between patients with or without escape mutations (data not shown).

Taken together, variants carrying two substitutions selected in the context of HLA-A*02:01 showed functional evidence for immune escape. Interestingly, despite strong statistical evidence for selection of the F7Y variant in HLA-A*02:01-positive patients and the S4T variant in HLA-B*35:01/03 positive patients, there was no functional evidence for immune escape from assays with exogenously added peptide. This suggests that neither HLA class I-binding nor TCR binding to the HLA class I/peptide complex was impaired by these two variants.

Influence of $core_{18-27}$ epitope variants on processing and presentation by HLA-A*02:01

Given the statistical support for selection of the F7Y and S4T substitution in the presence of HLA-A*02:01 or HLA-B*35:01/03, the absence of a clear functional impact on $CD8^+$ T cell cross-reactivity was unexpected. Our functional assays as well as the predicted binding affinities suggested that binding to the HLA class I molecule is not impaired when the synthetic peptide is exogenously added to the assay. This assay, however, does not include endogenous processing and presentation of the peptide from the viral protein and it has been previously shown that sequence variants within the epitope as well as in the epitope flanking region can associate with altered epitope processing (26–28). To address the influence of sequence variants on epitope processing, HepG2 cells were stably transduced with different variants of the core protein. HepG2 naturally express HLA-A*02:01 on their surface and the epitope is processed from endogenous protein repertoire. The complete core protein (genotype D) was fused to mCherry to compare expression levels and the frequency of mCherry positive cells and the mean fluorescence intensity (MFI) were analyzed before each experiment (Supplemental Figure S7). These cell lines were used as targets for $core_{18-27}$-specific HLA-A*02:01-restricted $CD8^+$ T cells and the IFNγ response was determined by flow cytometry. Figure 5 shows the results of nine independent experiments. An example and the gating strategy is outlined in Supplemental Figure S8. The same frequency of IFNγ-positive cells was detected when target cells were transfected with the prototype core protein or when they were pulsed with the synthetic prototype peptide, indicating that the prototype epitope can be processed from the fusion protein. The frequent S4A substitution, showed a slightly but not significantly reduced IFNγ production when endogenously processed. In line with our previous results, the variant with the double S4H-F7Y substitution showed a substantially impaired immune response, when the synthetic peptide was exogenously added to the assay or endogenously processed. In contrast, the exogenously added F7Y variant peptide did not impair the immune response, however, here, the frequency of IFNγ-producing $CD8^+$ T cells was reduced 4-fold when the variant epitope was endogenously processed, suggesting that the F7Y mutation interferes with correct antigen processing.

Discussion

The core region 18-27 is one of the immunodominant targets of the $CD8^+$ T cell response and is considered a prime target for future T cell based immune therapies. Here, we performed an in-depth analysis of the presenting HLA class I

10.3389/fimmu.2022.1045498

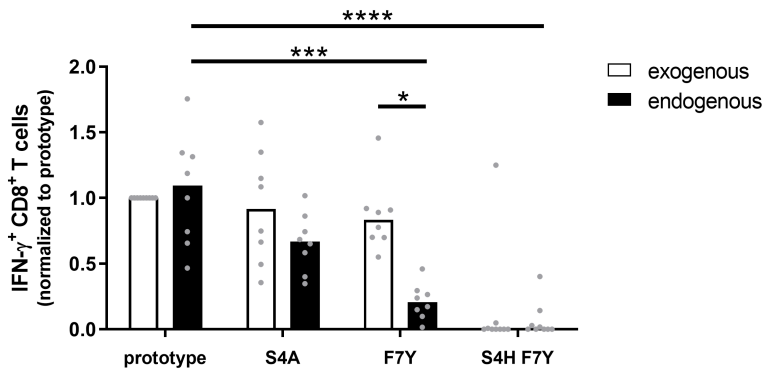


FIGURE 5
$CD8^+$ T cell response against the endogenously processed epitope $core_{18-27}$. Target cells stably expressing the HBV core protein were generated by transduction of HepG2 cells with prototype or variant HBV core (aa1-183) fused to mCherry. Effector T cells from different HLA-A*02:01-positive donors were expanded for 10 days in the presence of the $core_{18-27}$ peptide FLPSDFFPSV. HBV core expressing HepG2 cells (endogenous) or peptide-pulsed HepG2 cells (exogenous) were used as targets for re-stimulation of $core_{18-27}$-specific effector $CD8^+$ T cells in an effector:target ratio of 1:1 followed by an ICS. The IFNγ response was normalized to the response against HepG2 cells with prototype peptide. P-values were calculated by One-way Anova with a Bonferroni's Multiple Comparison Test. *p < 0.05; ***p < 0.001; ****p < 0.0001.

molecules, the sequence diversity of the epitope between and within individuals, the HLA-dependent pattern of selection pressure and propose a mechanisms of immune escape.

There are different reports suggesting promiscuous binding and presentation of the $core_{18-27}$ epitope by different HLA class I molecules. In MHC-class-I binding assays, binding to a variety of HLA-class I-molecules including HLA-A*02, B*07, B*35, B*44, B*51 (13) has been described. Moreover, beyond $CD8^+$ T-cell responses against $core_{18-27}$ in the context of HLA-A*02, earlier studies demonstrated also $CD8^+$ T cell responses restricted by HLA-B*35 (11, 13) and HLA-B*51 (11, 12). Notably, despite reported binding to HLA-B*07 (11, 13) we did not detect any robust responses in HLA-B*07-positive patients with chronic infection (data not shown). In the context of T-cell based immune therapies promiscuous binding and presentation of epitopes by different HLA class I molecules may have the advantage of broader HLA coverage in the target population. In our cohort, 62.2% of the patients carried at least one relevant HLA class I allele consistent with a high frequency of individuals mounting a response against this particular epitope.

Interestingly, despite promiscuous binding of the $core_{18-27}$ epitope to different HLA class I types, we observed an impact of the HLA class I-subtype on the ability to mount a $CD8^+$ T cell response. This was most evident in the context of HLA-B*35, as here $CD8^+$ T cell responses were only detected when the epitope was presented by HLA-B*35:01 or HLA-B*35:03 but not when presented by HLA-B*35:02 or HLA-B*35:08. The exact subtype-specific restriction was less clear in the context of HLA-A*02. Most HLA-A*02-subtypes, other than HLA-A*02:01, were in the "multiple alleles" group and it is therefore difficult to draw solid conclusions for these less frequent alleles. Notably, it was previously described, that the $core_{18-27}$ epitope is immunogenic in HLA-A*02:01 but not in HLA-A*02:03 (14), suggesting that also here, subtype-specific differences play a role for the immune response. Collectively, these results highlight that high-resolution HLA class I-typing is required for optimal results of T cell-based therapies.

In line with selection pressure by $CD8^+$ T cells in the context of HLA-A*02:01, HLA-B*35:01/03 and HLA-B*51:01 we observed increased frequencies of substitutions in the epitope region in patients carrying any of the relevant HLA class I-alleles compared to patients lacking these alleles. In a previous study, we could show, that viral substitutions in the $core_{18-27}$ epitope are significantly enriched in HLA-A*02 positive patients (9). Here, we extended the analysis and provide evidence, that the $core_{18-27}$ region is also under selection pressure when presented by other HLA class I-molecules. Notably, variants were predominantly detected in HBeAg-negative patients and not in HBeAg-negative patients. This is in line with the concept of a tolerogenic function attributed to the HBeAg [reviewed in (29)]. In transgenic mouse model, mice carrying a core-specific TCR lacked an immune response against the HBcAg when the HBeAg was expressed at high levels (30). In turn, when the HBeAg was expressed from a promotor with low activity, the mice mounted a weak core-specific immune response and when no HBeAg was present, a strong and robust T cell response was observed. Although the mechanism of the putative immunomodulatory HBeAg function is not understood, a consequence may be lack of immune selection of CD8 T cell epitope variants in the presence of HBeAg as observed in our study. Notably, epitope variants are also absent during acute resolving HBV infection. In a previous study the core region from HLA-A*02-positive patients with acute HBV infection was sequenced (24). Here,

eight patients infected with genotype A or D were longitudinally followed and the earliest and last viremic sample was sequenced. All patients carried the prototype sequence in the $core_{18-27}$ epitope at the earliest time point. For five patients, a second time point between 2-4 weeks later was sequenced and in none of them a variant was selected before they continued to spontaneously resolve the infection.

The role of viral escape from the $CD8^+$ T cell response has been debated in the context of hepatitis B. Early studies by Rehermann et al. concluded that mutational escape does not contribute to immune evasion of HBV, based on the observation that no epitope variants were observed in patients with HBV infection (17). However, the studied cohort was dominated by HBeAg positive infections, where we also do not see any evidence immune selection and the T cell response against epitope variants was not analyzed. Notably, the observed V10I substitution represents the prototype residue in HBV genotypes B/C, which are predominant in China. In line with a potential immune evasive effect, a prior study reported a reduced T cell response against the $core_{18-27}$ epitope in a Chinese cohort (14). We also found evidence for selection of the V10I substitution in genotype D however, in our assays a reduction in the CD8 T cell response was only detectable in combination with an additional substitution at position 4. Other studies specifically analyzing the impact of sequence variants in the $core_{18-27}$ region on the $CD8^+$ T cell response by Bertoletti et al. were consistent with immune escape by common variants (8, 10, 18). In line with our data, the authors found that the F7Y substitution did not impair HLA-A*02:01 binding and observed a similar pattern of reactivity (18). Previous work showed that epitope core 10-27 can be also presented by HLA-class II (18), its therefore possible that this region is also under HLA-class II selection pressure, which was not addressed in our study.

Different mechanisms for functional impairment of the $CD8^+$ T cell response by selected substitutions in targeted epitopes have been described (31). Here, the degree of cross-reactivity of $core_{18-27}$-specific CD8 T cells with epitope variants suggests that binding to the HLA class I-molecule is not impaired in most cases. In line with earlier studies, our results rather suggest that the interaction of the TCR with some of the variant peptides in complex with the HLA molecule is altered, which can present as an antagonistic effect of the variant peptide on T cell function (10). Importantly, impairment of TCR binding was seemingly only achieved when two substitutions within the epitope were selected. In contrast, the immunological assays were less conclusive for the single F7Y substitution in the context of HLA-A*02 and S4T substitution in the context of HLA-B*35:01/03, despite there was strong statistical support for selection in the presence of these alleles. This was in line with previous studies by Bertoletti et al. showing the ability of $core_{18-27}$ specific CD8 T cells to tolerate substitutions at epitope positions 4-7 (18). Lack of a functional impact on the CD8 T cell response by putative escape mutations in the targeted epitope was previously described in HCV in the chimpanzee model and in the human system (26, 32). In these cases, evidence for functional immune escape was obtained from experiments, in which the endogenously processed epitope was studied. Similar to escape mutations in HCV (28), the F7Y substitution did not show evidence for immune escape in assays when the synthetic peptide was exogenously added, however, the immune response to the endogenously processed epitope was impaired by the substitution. Most likely, this is explained by reduced amounts of variant peptide antigen on the surface of cells as a consequence of altered antigen processing. Although we do not know at what extent a reduction of the T cell response is biologically relevant, we can speculate that cells infected with viral variants have less antigen on their surface and are therefore less susceptible to T cells. Unfortunately, HepG2 cells do not express HLA-B*35:01/03 and therefore the impact of the S4T substitution could not be addressed with the same assay.

The viral sequence data suggest that different pathways lead the virus to escape from the $CD8^+$ T cell response. In HIV-1, of all the HLA-associated polymorphisms that occur within or near optimally described CTL epitopes, roughly 20% occur at anchor positions for HLA binding (33). The remaining substitutions are selected in other positions of the epitope and either impair antigen processing or TCR binding. Notably, there are also examples for selection of distinct escape variants within an identical epitope depending on the presenting HLA class I subtype (34, 35). The overall selection process is likely influenced by viral factors such as the impact of individual substitutions on the replication capacity as well as host factors such as the presenting HLA class molecule. Indeed, our data also suggest that different escape pathways are selected in the context of HLA-B*35:01/03 compared to the context of HLA-A*02:01. In HLA-B*35:01/03 positive patients a highly reproducible S4T substitution is selected whereas in HLA-A*02 positive patients this substitutions is rarely selected and even undetectable at the quasispecies level. In HLA-A*02, the escape pathways seem more complex and include substitutions in different positions. Multiple sites under selection pressure may suggest either continuous evolution if no optimal escape mutation can be selected or different solutions for optimal immune escape. In turn, the barrier to escape may be lower for HLA-B35:01/03. Lower expression levels of HLA-B*35 compared to other HLA class I-molecules have been reported (36) which may be consistent with a lower threshold to full escape. Distinct mutational escape pathways depending on the presenting HLA-class I-molecule may contribute to differential viral evolution in HLA-diverse populations (37, 38).

Multiple sites under selection pressure in the context of the same HLA class I-subtype may also reflect the highly individualized pathways to immune escape. Given the enormous diversity of the TCR it seems plausible that the selection process also reflects the predominant TCRs recruited by the individual's $CD8^+$ T cell response. Although we could not find clear evidence

for functional immune escape even for some of the selected autologous variants in patients with chronic infection, we believe that the individual TCR plays an important role (39). Indeed, the analysis of the cross-reactivity profiles of TCR-clones indicated differences between the responses of individual TCRs to variations of the epitope. In case different epitope variants are equally well processed and presented on the cell surface, individual broadly cross-reactive TCRs will likely be advantageous compared to TCRs with a narrowly focused cross-reactivity profile. In the setting of infections with highly variable pathogens such as HBV, TCRs with broad cross-reactivity - such as clone 6K in our analysis – will therefore be quite valuable for $CD8^+$ T cell-based therapies.

It is well established that HBeAg positive HBV infection associates with higher viral loads compared to HBeAg negative infection (40). Strikingly, we did not observe any variation from the prototype sequence in HBeAg positive patients, which is in line with absence of detectable core-specific CD8 T cell responses (30). In HBeAg negative patients the frequency of variants was enriched and the presence of variants was associated with stronger CD8 T cell responses compared to patients with prototype sequence. Whether this truly reflects a greater magnitude of the response or a better capacity of specific CD8 T cells to expand, needs to be studied. In HCV infection differentiation towards a memory phenotype after mutational escape has been described (41), however, this has not been observed in the context of HBV (24). In line with the idea that the immune response is an important factor for the viral replication level, we could not identify differences in the viral load associated with individual substitutions in HBeAg negative patients. However, we cannot exclude that lower viral loads are also caused by fitness costs. To directly analyze the influence of individual substitutions on viral fitness, a robust replication model for HBV is required. It has yet to be determined whether the recently developed *in vitro* models for HBV replication are sufficiently robust to address this in the future (16, 42)

Taken together, our data show that the dominant epitope region $core_{18\text{-}27}$ is highly variable and under heavy selection pressure in the context of different HLA class I-molecules. The selection process is rather complex and may be driven by host factors such as the presenting HLA-molecule and the TCR recruited for $CD8^+$ T cell response. Mechanistically, the escape pathways seem to impair binding of the TCR to the variant peptide/HLA class I-complex, which suggests that broadly cross-reactive TCRs may be beneficial in this context. However, we also find evidence for impaired antigen processing and reduced presentation of epitope variants. Viruses carrying such escape substitutions will be less susceptible to $CD8^+$ T cell responses and viral genome sequencing should be considered when strategies for T cell therapies are further developed.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MZ043025-MZ043097, MZ097624-MZ097884.

Ethics statement

The studies involving human participants were reviewed and approved by Ethikkommission an der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf. The patients/participants provided their written informed consent to participate in this study.

Author contributions

The project was conceived by AW, JB, TS, and JT. Experiments were performed by AW, TS, JB, KW, CM, EA, HK, AK, and FH. Data were analyzed by all authors. The manuscript was written by AW and JT with input from all authors. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by grants from the DFG (TI 323/4–1), the Stiftung zur Erforschung infektiös-immunologischer Erkrankungen (AW, 10-16-72), the Jürgen Manchot Foundation and the European Union's Horizon 2020 research and innovation program (no. 848223; TherVacB consortium). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgments

The authors are very grateful to the study participants for taking part in the study. The authors thank Didier Trono, EPFL Lausanne for plasmids and protocols and Alexandra Graupner, Anja Voges and Eugen Bäcker for technical help.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2022.1045498/full#supplementary-material

References

1. Hu J, Protzer U, Siddiqui A. Revisiting hepatitis b virus: Challenges of curative therapies. *J Virol* (2019) 93(20): e01032–19. doi: 10.1128/JVI.01032-19

2. Revill PA, Chisari FV, Block JM, Dandri M, Gehring AJ, Guo H, et al. A global scientific strategy to cure hepatitis b. *Lancet Gastroenterol Hepatol* (2019) 4 (7):545–58. doi: 10.1016/S2468-1253(19)30119-0

3. Maini MK, Burton AR. Restoring, releasing or replacing adaptive immunity in chronic hepatitis b. *Nat Rev Gastroenterol Hepatol* (2019) 16(11):662–75. doi: 10.1038/s41575-019-0196-9

4. Gehring AJ, Protzer U. Targeting innate and adaptive immune responses to cure chronic HBV infection. *Gastroenterology* (2019) 156(2):325–37. doi: 10.1053/j.gastro.2018.10.032

5. Thimme R, Wieland S, Steiger C, Ghrayeb J, Reimann KA, Purcell RH, et al. CD8(+) T cells mediate viral clearance and disease pathogenesis during acute hepatitis b virus infection. *J virology* (2003) 77(1):68–76. doi: 10.1128/JVI.77.1.68-76.2003

6. Maini MK, Pallett LJ. Defective T-cell immunity in hepatitis b virus infection: why therapeutic vaccination needs a helping hand. *Lancet Gastroenterol Hepatol* (2018) 3(3):192–202. doi: 10.1016/S2468-1253(18)30007-4

7. Bertoletti A, Le Bert N. Immunotherapy for chronic hepatitis b virus infection. *Gut Liver* (2018) 12(5):497–507. doi: 10.5009/gnl17233

8. Bertoletti A, Costanzo A, Chisari FV, Levrero M, Artini M, Sette A, et al. Cytotoxic T lymphocyte response to a wild type hepatitis b virus epitope in patients chronically infected by variant viruses carrying substitutions within the epitope. *J Exp Med* (1994) 180(3):933–43. doi: 10.1084/jem.180.3.933

9. Kefalakes H, Budeus B, Walker A, Jochum C, Hilgard G, Heinold A, et al. Adaptation of the hepatitis b virus core protein to CD8 T-cell selection pressure. *Hepatology* (2015) 62(1):47–56. doi: 10.1002/hep.27771

10. Bertoletti A, Sette A, Chisari FV, Penna A, Levrero M, De Carli M, et al. Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells. *Nature* (1994) 369(6479):407–10. doi: 10.1038/369407a0

11. Bertoni R, Sidney J, Fowler P, Chesnut RW, Chisari FV, Sette A. Human histocompatibility leukocyte antigen-binding supermotifs predict broadly cross-reactive cytotoxic T lymphocyte responses in patients with acute hepatitis. *J Clin Invest* (1997) 100(3):503–13. doi: 10.1172/JCI119559

12. Thimme R, Chang KM, Pemberton J, Sette A, Chisari FV. Degenerate immunogenicity of an HLA-A2-restricted hepatitis b virus nucleocapsid cytotoxic T-lymphocyte epitope that is also presented by HLA-B51. *J virology* (2001) 75 (8):3984–7. doi: 10.1128/JVI.75.8.3984-3987.2001

13. Depla E, van der Aa A, Livingston BD, Crimi C, Allosery K, De Brabandere V, et al. Rational design of a multiepitope vaccine encoding T-lymphocyte epitopes for treatment of chronic hepatitis b virus infections. *J virology* (2008) 82(1):435–50. doi: 10.1128/JVI.01505-07

14. Tan AT, Loggi E, Boni C, Chia A, Gehring AJ, Sastry KS, et al. Host ethnicity and virus genotype shape the hepatitis b virus-specific T-cell repertoire. *J virology* (2008) 82(22):10986–97. doi: 10.1128/JVI.01124-08

15. Maini MK, Boni C, Ogg GS, King AS, Reignat S, Lee CK, et al. Direct ex vivo analysis of hepatitis b virus-specific CD8(+) T cells associated with the control of infection. *Gastroenterology* (1999) 117(6):1386–96. doi: 10.1016/S0016-5085(99)70289-1

16. Wisskirchen K, Kah J, Malo A, Asen T, Volz T, Allweiss L, et al. T Cell receptor grafting allows virological control of hepatitis b virus infection. *J Clin Invest* (2019) 129(7):2932–45. doi: 10.1172/JCI120228

17. Rehermann B, Pasquinelli C, Mosier SM, Chisari FV. Hepatitis b virus (HBV) sequence variation of cytotoxic T lymphocyte epitopes is not common in patients with chronic HBV infection. *J Clin Invest* (1995) 96(3):1527–34. doi: 10.1172/JCI118191

18. Bertoletti A, Southwood S, Chesnut R, Sette A, Falco M, Ferrara GB, et al. Molecular features of the hepatitis b virus nucleocapsid T-cell epitope 18-27: interaction with HLA and T-cell receptor. *Hepatology* (1997) 26(4):1027–34. doi: 10.1002/hep.510260435

19. Oudshoorn M, Horn PA, Tilanus M, Yu N. Typing of potential and selected donors for transplant: methodology and resolution. *Tissue Antigens* (2007) 69 Suppl 1:10–2. doi: 10.1111/j.1399-0039.2006.758_5.x

20. Kefalakes H, Jochum C, Hilgard G, Kahraman A, Bohrer AM, El Hindy N, et al. Decades after recovery from hepatitis b and HBsAg clearance the CD8+ T cell response against HBV core is nearly undetectable. *J Hepatol* (2015) 63(1):13–9. doi: 10.1016/j.jhep.2015.01.030

21. Lange M, Fiedler M, Bankwitz D, Osburn W, Viazov S, Brovko O, et al. Hepatitis c virus hypervariable region 1 variants presented on hepatitis b virus capsid-like particles induce cross-neutralizing antibodies. *PloS One* (2014) 9(7): e102235. doi: 10.1371/journal.pone.0102235

22. Gehring AJ, Xue SA, Ho ZZ, Teoh D, Ruedl C, Chia A, et al. Engineering virus-specific T cells that target HBV infected hepatocytes and hepatocellular carcinoma cell lines. *J Hepatol* (2011) 55(1):103–10. doi: 10.1016/j.jhep.2010.10.025

23. Wisskirchen K, Metzger K, Schreiber S, Asen T, Weigand L, Dargel C, et al. Isolation and functional characterization of hepatitis b virus-specific T-cell receptors as new tools for experimental and clinical use. *PloS One* (2017) 12(8): e0182936. doi: 10.1371/journal.pone.0182936

24. Hoogeveen RC, Robidoux MP, Schwarz T, Heydmann L, Cheney JA, Kvistad D, et al. Phenotype and function of HBV-specific T cells is determined by the targeted epitope in addition to the stage of infection. *Gut* (2019) 68(5):893–904. doi: 10.1136/gutjnl-2018-316644

25. Schuch A, Salimi Alizei E, Heim K, Wieland D, Kiraithe MM, Kemming J, et al. Phenotypic and functional differences of HBV core-specific versus HBV polymerase-specific CD8+ T cells in chronically HBV-infected patients with low viral load. *Gut* (2019) 68(5):905–15. doi: 10.1136/gutjnl-2018-316641

26. Kimura Y, Gushima T, Rawale S, Kaumaya P, Walker CM. Escape mutations alter proteasome processing of major histocompatibility complex class I-restricted epitopes in persistent hepatitis c virus infection. *J virology* (2005) 79 (8):4870–6. doi: 10.1128/JVI.79.8.4870-4876.2005

27. Seifert U, Liermann H, Racanelli V, Halenius A, Wiese M, Wedemeyer H, et al. Hepatitis c virus mutation affects proteasomal epitope processing. *J Clin Invest* (2004) 114(2):250–9. doi: 10.1172/JCI200420985

28. Walker A, Skibbe K, Steinmann E, Pfaender S, Kuntzen T, Megger DA, et al. Distinct escape pathway by hepatitis c virus genotype 1a from a dominant CD8+ T cell response by selection of altered epitope processing. *J virology* (2015) 90(1):33–42. doi: 10.1128/JVI.01993-15

29. Kramvis A, Kostaki EG, Hatzakis A, Paraskevis D. Immunomodulatory function of HBeAg related to short-sighted evolution, transmissibility, and clinical manifestation of hepatitis b virus. *Front Microbiol* (2018) 9:2521. doi: 10.3389/fmicb.2018.02521

30. Chen MT, Billaud JN, Sallberg M, Guidotti LG, Chisari FV, Jones J, et al. A function of the hepatitis b virus precore protein is to regulate the immune response to the core antigen. *Proc Natl Acad Sci U S A.* (2004) 101(41):14913–8. doi: 10.1073/pnas.0406282101

31. Timm J, Walker CM. Mutational escape of CD8+ T cell epitopes: implications for prevention and therapy of persistent hepatitis virus infections. *Med Microbiol Immunol* (2015) 204(1):29–38. doi: 10.1007/s00430-014-0372-z

32. Timm J, Lauer GM, Kavanagh DG, Sheridan I, Kim AY, Lucas M, et al. CD8 epitope escape and reversion in acute HCV infection. *J Exp Med* (2004) 200 (12):1593–604. doi: 10.1084/jem.20041006

33. Carlson JM, Le AQ, Shahid A, Brumme ZL. HIV-1 adaptation to HLA: a window into virus-host immune interactions. *Trends Microbiol* (2015) 23(4):212–24. doi: 10.1016/j.tim.2014.12.008

34. Carlson JM, Listgarten J, Pfeifer N, Tan V, Kadie C, Walker BD, et al. Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. *J Virol* (2012) 86(9):5230–43. doi: 10.1128/JVI.06728-11

35. Yagita Y, Kuse N, Kuroki K, Gatanaga H, Carlson JM, Chikata T, et al. Distinct HIV-1 escape patterns selected by cytotoxic T cells with identical epitope specificity. *J virology* (2013) 87(4):2253–63. doi: 10.1128/JVI.02572-12

36. Yarzabek B, Zaitouna AJ, Olson E, Silva GN, Geng J, Geretz A, et al. Variations in HLA-b cell surface expression, half-life and extracellular antigen receptivity. *Elife* (2018) 7:e34961. doi: 10.7554/eLife.34961

37. Xia Y, Pan W, Ke X, Skibbe K, Walker A, Hoffmann D, et al. Differential escape of HCV from CD8(+) T cell selection pressure between China and Germany depends on the presenting HLA class I molecule. *J Viral Hepat* (2019) 26(1):73–82. doi: 10.1111/jvh.13011

38. Kloverpris HN, Leslie A, Goulder P. Role of HLA adaptation in HIV evolution. *Front Immunol* (2015) 6:665. doi: 10.3389/fimmu.2015.00665

39. Ziegler S, Skibbe K, Walker A, Ke X, Heinemann FM, Heinold A, et al. Impact of sequence variation in a dominant HLA-A*02-restricted epitope in hepatitis c virus on priming and cross-reactivity of CD8+ T cells. *J virology* (2014) 88(19):11080–90. doi: 10.1128/JVI.01590-14

40. EASL EASL 2017 clinical practice guidelines on the management of hepatitis b virus infection *J Hepatol* (2017) 67(2):370–98. doi: 10.1016/j.jhep.2017.03.021

41. Bengsch B, Seigel B, Ruhl M, Timm J, Kuntz M, Blum HE, et al. Coexpression of PD-1, 2B4, CD160 and KLRG1 on exhausted HCV-specific CD8+ T cells is linked to antigen recognition and T cell differentiation. *PloS Pathog* (2010) 6(6):e1000947. doi: 10.1371/journal.ppat.1000947

42. Hu J, Lin YY, Chen PJ, Watashi K, Wakita T. Cell and animal models for studying hepatitis b virus infection and drug development. *Gastroenterology* (2019) 156(2):338–54. doi: 10.1053/j.gastro.2018.06.093

Research Article
Viral Hepatitis

JOURNAL
OF HEPATOLOGY

HBV shows different levels of adaptation to HLA class I-associated selection pressure correlating with markers of replication

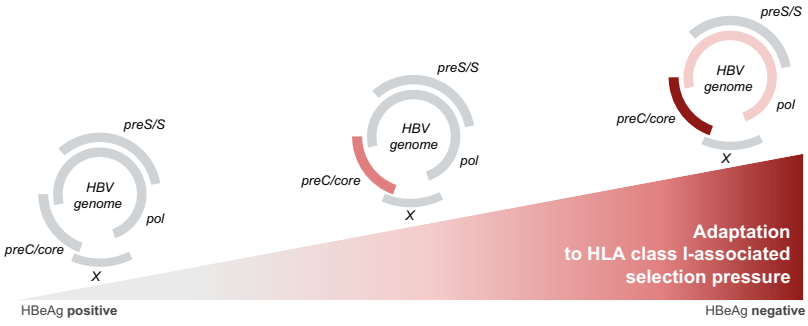
Authors

Tatjana Schwarz, Johannes Ptok, Maximilian Damagnez, …, Christoph Neumann-Haefelin, Andreas Walker, Jörg Timm

Correspondence

joerg.timm@uni-duesseldorf.de (J. Timm).

Graphical abstract



Highlights

- 532 complete viral genomes from people with chronic HBV infection were sequenced.
- HLA-associated mutational states (HAMs) indicative for CD8 T cell pressure were identified.
- HAMs are more frequent in HBV core compared to other HBV proteins.
- HBV genomes show different levels of adaptation between patients and viral proteins.
- The level of adaptation of HBV to CD8 T cell pressure correlates with markers of replication.

Impact and implications

The immune response mediated by CD8 T cells plays a critical role in controlling HBV infection and shows promise for therapeutic strategies aimed at achieving a functional cure. This study demonstrates that mutational escape within CD8 T-cell epitopes is common in HBV and represents a key factor in the failure of immune control. Notably, the HBV core protein emerges as the primary target of CD8 T-cell selection pressure. Additionally, the observed correlation between HBV adaptation levels and viral replication markers indicates that CD8 T-cell immunity may influence transitions between phases of chronic HBV infection.

https://doi.org/10.1016/j.jhep.2024.10.047
 J. Hepatol. 2025, 82, 805–815

Research Article
Viral Hepatitis

JOURNAL
OF HEPATOLOGY

HBV shows different levels of adaptation to HLA class I-associated selection pressure correlating with markers of replication

Tatjana Schwarz[1,11,†], **Johannes Ptok**[1,†], Maximilian Damagnez[1], Christopher Menne[1], Elahe Salimi Alizei[2], Julia Lang-Meli[2], Michelle Maas[2], Daniel Habermann[3], Daniel Hoffmann[3], Julian Schulze zur Wiesch[4], Georg M. Lauer[5], Helenie Kefalakes[6], Markus Cornberg[6], Anke R.M. Kraft[6], Smaranda Gliga[7], Hans H. Bock[7], Peter A. Horn[8], Mala K. Maini[9], Robert Thimme[2], Heiner Wedemeyer[6], Jacob Nattermann[10], Falko M. Heinemann[8], Tom Luedde[7], Christoph Neumann-Haefelin[2,12], Andreas Walker[1,‡], Jörg Timm[1,*,‡]

Journal of Hepatology **2025**. vol. 82 | 805–815

Background & Aims: Immune responses by CD8 T cells are essential for control of HBV replication. Although selection of escape mutations in CD8 T-cell epitopes has previously been described in HBV infection, its overall influence on HBV sequence diversity and correlation with markers of HBV replication remain unclear.

Methods: Whole-genome sequencing was applied to HBV isolates from 532 patients with chronic HBV infection and high-resolution HLA class I genotyping. Using a Bayesian model (HAMdetector) for identification of HLA-associated mutational states (HAMs), the frequency and location of residues under CD8 T-cell selection pressure were determined and the levels of adaptation of individual isolates were quantified.

Results: Using previously published thresholds for the identification of HAMs, a total of 295 residues showed evidence of CD8 T-cell escape, the majority of which were located in previously unidentified epitopes. Interestingly, HAMs were highly enriched in the HBV core protein compared to all other proteins. When individual HBV isolates were compared, different levels of adaptation to HLA class I immune pressure were noted. The level of adaptation increased with patient age and correlated with markers of replication, with low levels of adaptation in HBeAg-positive infection. Furthermore, the levels of adaptation negatively correlated with HBV viral load and HBsAg levels, consistent with high levels of HLA class I-associated selection pressure in patients with low replication levels.

Conclusions: HBV sequence diversity is shaped by HLA class I-associated selection pressure with the HBV core protein being a predominant target of selection. Importantly, different levels of adaptation to immune pressure were observed between HBV infection stages, which need to be considered in the context of T-cell-based therapies.

Introduction

HBV infection remains a significant global health concern, with over 250 million individuals estimated to be chronically infected worldwide.[1] HBV is a small, enveloped DNA virus belonging to the Hepadnaviridae family that primarily targets the liver, leading to a wide spectrum of clinical outcomes, ranging from asymptomatic carriers to severe liver disease, including cirrhosis and hepatocellular carcinoma.[2] The hepatitis B e antigen (HBeAg) is an important marker for classification of the different clinical stages of HBV infection. Although not required for replication, there is evidence that HBeAg is involved in the establishment of chronic infection and functions as an immune tolerogen (for a review, see[3]). Loss of HBeAg is typically associated with a flare of hepatitis, activation of CD8 T cells[4] and reduced levels of replication. However, in some patients, high-level replication of HBeAg-negative viral variants persists.[2]

Immune responses by HBV-specific CD8 T cells play a central role in determining the outcome of infection.[5] CD8 T cells contribute to controlling HBV infection but may also contribute to liver pathology during chronic hepatitis B.[6] CD8 T cells target HBV-infected hepatocytes via interactions of their T-cell receptor (TCR) with specific viral epitopes presented on the surface of infected cells in the context of the HLA class I complex. Failure of the CD8 T-cell response against HBV in the context of chronic

* Corresponding author. Address: Institute of Virology, University Hospital Düsseldorf, Medical Faculty, Heinrich-Heine University, Moorenstr. 5, 40225 Düsseldorf, Germany; Tel.: +49(0) 211-81, fax: +49(0) 211-81.
E-mail address: joerg.timm@uni-duesseldorf.de (J. Timm).
† Equal contribution
‡ Equal contribution
https://doi.org/10.1016/j.jhep.2024.10047

infection has been described and includes a progressive dysfunction of HBV-specific CD8 T cells upon continuous antigen stimulation in the liver, also termed T-cell exhaustion.[7,8] In addition, selection of viral epitope variants associated with immune escape also contributes to CD8 T-cell failure in HBV infection.[9] Selected mutations in HBV epitopes can impair binding to the HLA class I molecule or alter the interaction of the TCR with the variant epitope in the HLA-complex.

Understanding the extent and the mechanisms underlying the selection of escape mutations within CD8 T-cell epitopes is of paramount importance in the context of both natural HBV infection and therapeutic interventions, such as immunotherapies. In HBV, the existing data on CD8 T-cell escape is correlative, showing that substitutions are enriched in targeted epitopes during chronic infection and functionally impair the CD8 T-cell response.[10–14] In prior HLA class I-association studies, these substitutions have been observed to be statistically more frequent in patients carrying the relevant HLA class I alleles consistent with their specific selection.[13] Notably, data on CD8 T-cell escape in HBV has only been obtained for individual immunodominant epitopes or substitution patterns in the HBV core protein.[10–13,15] The full extent of HLA class I-associated mutations across all HBV proteins and the different levels of adaptation of individual HBV isolates is therefore unknown.

Herein, we combined full HBV genome sequence data from 532 patients with chronic infection with high-resolution HLA class I genotyping and applied novel tools for detection of HLA-associated mutational states (HAMs) in the viral genome to get a comprehensive map of the frequency and distribution of substitutions selected by HLA class I-restricted CD8 T cells. The results show that selection pressure is predominantly exerted on epitopes in the HBV core protein. Moreover, different frequencies of HAMs were noted for individual viral genomes correlating with markers of viral replication, such as the HBeAg status and HBV DNA concentrations in serum. Collectively, the data demonstrate extensive but also different levels of adaptation of viral genomes to CD8 T-cell pressure in chronic hepatitis B. Different levels of adaptation may influence the perspective for functional cure of hepatitis B, especially when T-cell-based therapeutic concepts are applied.

Patients and methods

Patients

In a multicenter effort, samples from 532 patients with HBV were collected at the Universities of Bonn, Düsseldorf, Essen, Freiburg, Hamburg, Hannover (all Germany) and London (UK) (Table 1 and Table S1). Informed consent was obtained from each patient, and the study protocol was approved by the local ethics committee of the Medical Faculty of Düsseldorf in accordance with the guidelines of the Declaration of Helsinki. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll density gradient centrifugation (Biocoll; Biochrom).[13] DNA for HLA-typing was extracted from PBMCs using spin columns (Qiagen). High-resolution (second field) HLA-A and HLA-B typing was performed by use of sequence-specific oligonucleotides (LABType™) methodology,[16] provided by One Lambda (Thermo Fisher Inc.).

Table 1. Patient characteristics.

	n (%)
Total, N	532 (100)
Sex	
Male	292 (55)
Female	224 (42)
No data	28 (5)
Age (years)	
Median	40
Range	18-85
Genotype	
A	101 (19)
B	27 (5)
C	13 (2)
D	372 (70)
E	20 (4)
Viral load (IU/ml)	
Median	4,494
IQR	1,294-94,846
HBsAg (IU/ml)	
Median	6,304
IQR	1,316-11,031
HBeAg status	
Positive	58
Negative	350
No data	124

HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen.

Amplification and sequence analysis of the HBV genome

Viral nucleic acid from 400 μl plasma was extracted automatically using the EZ1 Virus Mini Kit v2.0 on an EZ1 Advanced XL robot or manually (both Qiagen) and the complete HBV genome was amplified in two fragments as previously described.[17,18] In brief, two-step nested PCRs were performed for the core region (nucleotide [nt] 1,683-nt 2,399; 717 base pairs [bp] according to the reference genome NC_003977.2) and the polymerase region (nt 2,299-nt 1,798; 2,682 bp according to the reference genome NC_003977.2) with primers listed in Table S2. Amplicons were Sanger sequenced with primers listed in Table S2. Sequences were aligned with the software Geneious 10.2.6 (RRID:SCR_010519). Sequences were submitted to GenBank and are available under accession numbers (MZ043025-MZ043097; MZ097624-MZ097884).

Phylogenetic analysis of viral sequences and HAMdetector analysis

To analyze the phylogeny and to provide the input files for the HAMdetector tool, all obtained sequences were aligned with the software Geneious 10.2.6 (RRID:SCR_010519) using MAFFT.[19] For phylogenetic analysis, a tree based on the complete HBV sequence, with references from Genebank, was calculated in Geneious 10.2.8 with the Mr. Bayes plugin.[20] For visualization, the output was exported as Newick file with support values and visualized with iTol.[21] HAMdetector[22] is implemented as a julia package for identifying HLA-associated substitutions based on aligned viral sequences paired to host HLA class I data. It integrates information from epitope prediction via MHCflurry 2.0 and phylogeny (based on RAxML-NG). The model is fit using Stan and the complete source code and documentation is available at https://github.com/HAMdetector/Escape.jl. For prediction, the open-reading frames for precore, polymerase (pol), L-HBsAg (HBs) and the

x protein (HBx) were inferred from the HBV alignment used above and were translated into an amino acid sequence. No adjustments were made to sequences where the amber-stop codon at position 27 in precore was the majority or sequences with insertion or deletions in the pol/preS1 region. Therefore, the numbering of amino acids in the HAM detector output is not consistent with IEDB (Immune Epitope Database) or GenBank numbering. The final amino acid alignment, including numbering, can be found in the supplementary material. For HAMdetector phylogeny, the MAFFT nucleotide alignment used for phylogenetic analysis was used. The complete output data and sequence alignments are available as described in the data availability statement.

Analysis of HAMs in previously published epitopes

Previously published HBV epitopes were retrieved from the IEDB. Parameters for the database search were as follows: linear peptides, organism hepatitis B virus (ID 10407), host human, assay T-cell with positive result, MHC restriction class I. A list of 310 entries was generated, which was further manually edited to remove epitopes with an undefined restricting HLA class I type, insufficiently mapped epitopes (length ≥13 amino acids) and duplicate entries of different variants of the same epitope. The final list contained 141 entries (42 in precore/core, 54 in pol, 34 in HBsAg and 11 in the HBx protein). To provide evidence for HLA class I-associated selection pressure on residues within the epitopes, the maximum posterior probabilities for the relevant HLA class I types were determined using HAMdetector for all positions in the epitope.

Quantification of HLA class I adaptation of individual HBV isolates

The level of adaptation of the HBV polyprotein to HLA class I-associated selection pressure was quantified. For this purpose, a majority consensus sequence was generated for each HBV genotype and all isolates of the corresponding genotype were aligned to the consensus sequence. Only positions that differed from the consensus sequence were included in the subsequent calculation of an adaptation score. For these variant amino acids, the maximum posterior probabilities for all relevant HLA class I alleles of the individual patient were determined and subtracted by 0.5, as only values >0.5 are indicative of HAMs. The adaptation score was then calculated as the sum of the maximum posterior probabilities for all variant positions after subtraction of 0.5.

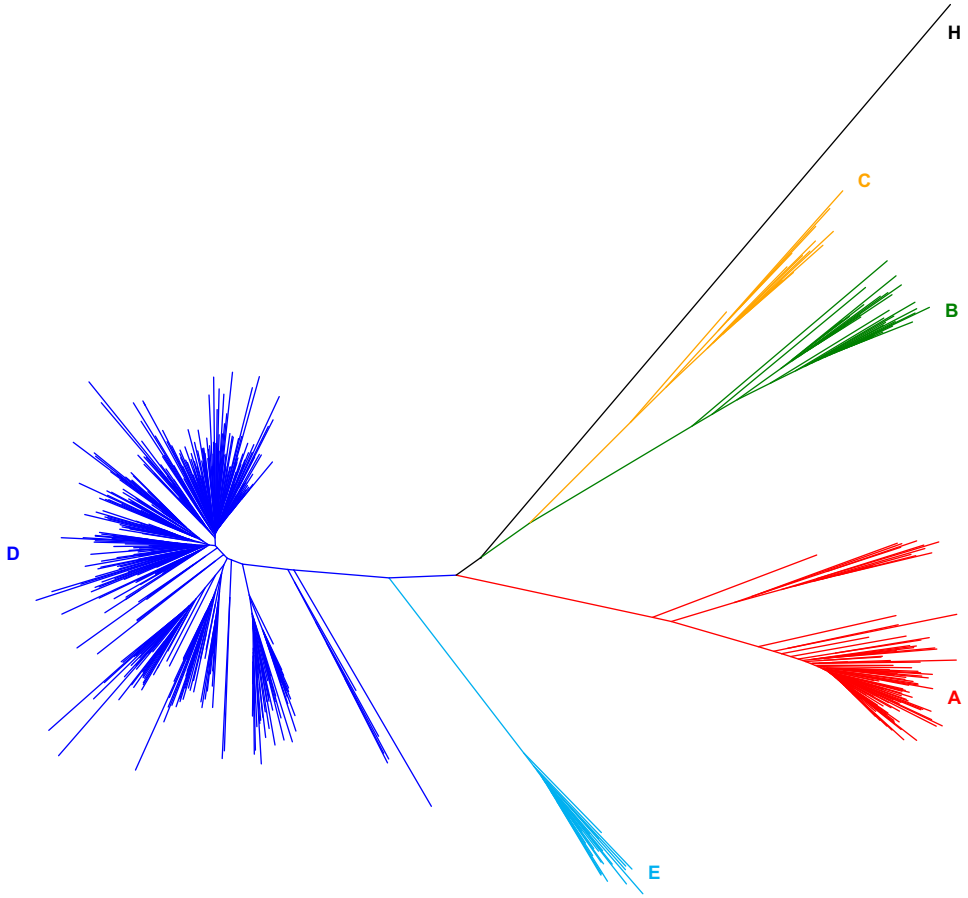


Fig. 1. Phylogenetic tree of all HBV sequences of the multicenter cohort. A total of 532 complete HBV sequences were aligned with genotype reference sequences [from ICTV] using MAFFT. A phylogenetic tree was calculated with the Mr. Bayes Plugin[20] in the software *Geneious 10.2.8* using the GTR genetic distance model and genotype H as outgroup. Genotypes are color coded (A-H) as indicated.

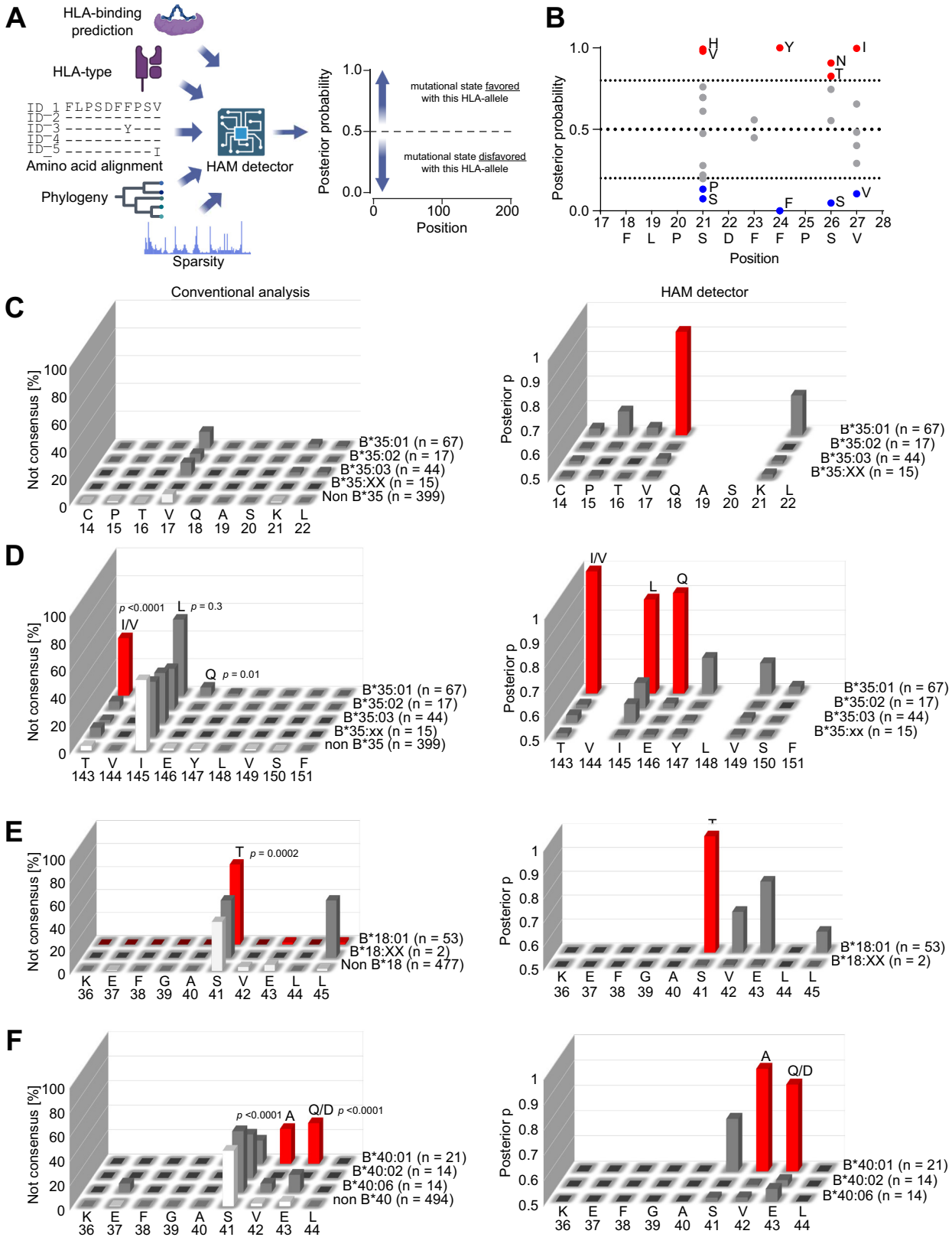


Fig. 2. HAMdetector and conventional statistical analysis for identification of HLA class I-associated viral sequence polymorphisms. (A) Schematic illustration of the input and output information of HAMdetector. The model integrates the alignment of viral sequences, the HLA alleles of patients, sparsity of HLA substitution associations, the phylogeny of viral sequences and potential HLA class I-binding motifs and assigns a posterior probability to all amino acids in the alignment. High posterior probabilities indicate that the amino acid is favored in individuals with the HLA allele, low posterior probabilities indicate that the amino acid is disfavored in

Results

Identification of residues under selection pressure with the HAMdetector tool

The full viral genome from 532 patients with chronic HBV infection was amplified in two fragments and sequenced. For genotyping, a phylogenetic tree with all viral sequences from the study as well as reference sequences from the NCBI was constructed (Fig. 1). The majority of the viral sequences from the cohort were genotype D (n = 372; 69.9%) and genotype A (n = 100; 18.8%). As expected for a European cohort, other genotypes such as genotype B (n = 27; 5.0%), genotype C (n = 13; 2.4%) and genotype E (n = 20; 3.7%) were less frequent (Table 1). The HLA class I-genotype was determined for all patients, allowing for association studies with viral sequence features. The complete patient characteristics are shown in Table S1.

For detection of residues under HLA class I-associated selection pressure (HAMs) in HBV, we utilized the HAMdetector tool.[22] The tool comprises a unified Bayesian regression model that combines multiple sources of information. These integrated factors include the alignment of viral sequences, patients' HLA alleles, sparsity of HLA substitution associations, the phylogeny of viral sequences and potential HLA class I-binding motifs (Fig. 2A). The outcome of this integration is a set of posterior probabilities for HLA substitution associations, which range from 0 to 1. The majority of posterior probabilities are close to 0.5, indicating that the amino acid is neither favored nor disfavored in the presence of the HLA class I allele. High posterior probabilities with values close to 1 indicate that the amino acid is favored in individuals with this HLA class I allele consistent with CD8 T-cell escape. Low posterior probabilities indicate that the amino acid is disfavored in the presence of this allele, suggesting that this amino acid is not a CD8 T-cell escape mutation. In Fig. 2B the results are illustrated for the well-described epitope $core_{18-27}$. For example, in this epitope, the F24Y substitution is a described immune escape variant.[17] Phenylalanine (F) at position 24 has a posterior probability of 1 for being a favored mutational state in HLA-A*02-positive patients. Conversely, the amino acid tyrosine (Y) at the same position has a posterior probability of 0, indicating it is a highly disfavored mutational state in patients with HLA-A*02.

Fig. 2C–F presents conventional analyses of substitution frequencies for three candidate epitope regions, compared to results obtained from HAMdetector. Traditional statistical methods for identifying "HLA footprints" involve creating a 2x2 table that counts the number of viral sequences with and without a substitution and with or without a specific HLA allele, followed by a statistical test, such as Fisher's exact test. The results of Fisher's exact test for associations identified by HAMdetector are shown in Fig. 2. Many HAMs identified by HAMdetector were also recognized as statistically significant in conventional analyses, although there are exceptions. For instance, for the candidate epitope (pol_{14-22} CPTVKASKL), there was a high posterior probability for phenylalanine at position 4 of the epitope (V17F) as the favored amino acid in HLA-B*35:01-positive individuals. In conventional analysis, a p value of 0.11 would not support immune selection of the V17F substitution. The advantage of HAMdetector lies in its inclusion of additional relevant information. In the model, a substitution that is "unexpected" at a certain position based on phylogeny is more likely to result from selection pressure. Conversely, polymorphisms typical of distinct genotypes or phylogenetic clades are less likely to be the result of selection pressure. Another crucial factor is sparsity. Comprehensive epitope maps show that HLA epitopes often overlap. Statistically, this means an association between a substitution and one HLA allele suggests that associations with other HLA alleles are also more likely. Additionally, relevant HLA class I motif information is included. HAMdetector considers substitutions in regions matching the binding motif of the relevant HLA class I type more likely to be HAMs. Based on this additional information, the algorithm assigns a high posterior probability to the V17F substitution as a HAM. Notably, the candidate epitope pol_{14-22} CPTVKASKL was experimentally confirmed in immunological assays (a representative result is shown in Fig. S1).

In a second HLA-B*35:01-restricted candidate epitope identified in the core region ($core_{143-151}$ TVIEYLVSF; Fig. 2D), HAMs were detected in three positions. Conventional analysis, however, indicated CD8 T-cell escape at only one position. Notably, HAMdetector identified a putative escape residue at the highly polymorphic position three of the epitope, with leucine being favored at this position in HLA-B35:01-positive individuals. We were able to experimentally confirm this novel candidate epitope in HLA-B*35-positive individuals and also tested peptide variants with the putative escape mutations (Fig. S1). The results further support that the substitutions functionally act as immune escape mutations.

Finally, in a third candidate epitope in the core ($core_{36-44}$ KEFGASVEL), HAMdetector identified HAMs in the context of HLA-B*18:01 (Fig. 2E) and distinct HAMs in the same epitope region in the context of HLA-B*40:01 (Fig. 2F). In this case, the HAMs would also have been identified by conventional statistical analysis. The candidate epitope was experimentally confirmed in the context of HLA-B*18:01 (Fig. S1). In conclusion, the results demonstrate that the HAMdetector is a powerful tool for the identification of HLA class I-associated viral sequence polymorphisms and allows for the identification of novel CD8+ T-cell epitopes.

The majority of HAMs are located in the HBV precore/core protein

Multiple sequence alignments were created for precore/core, pol, large HBsAg and HBx for identification of HAMs in all HBV

individuals with the HLA allele. A posterior probability of 0.5 indicates that the amino acid is neither favored nor disfavored. (B-D) Comparison of conventional statistical analysis (left column) with HAMdetector results (right column) for different candidate epitope regions and HLA class I genotypes. (B) Illustration of HAMdetector scores for the HLA-A*02-restricted epitope $core_{18-27}$. (C) HLA-B*35:01 pol_{14-22} CPTVKASKL (Statistical significance was calculated by Fisher's exact test), (D) HLA-B*35:01 $core_{143-151}$ TVIEYLVSF (Statistical significance was calculated by Fisher's exact test), (E) HLA-B*18:01 $core_{36-44}$ KEFGASVEL (Statistical significance was calculated by Fisher's exact test) and (F) HLA-B*40:01 $core_{36-44}$ KEFGASVEL (Statistical significance was calculated by Fisher's exact test). Positions with p <0.05 by Fisher's exact test or posterior probability ≥0.8 in HAMdetector are colored in red. The results of the Fisher's exact test for all HAMs and the favored amino acid are shown. HAM, HLA-associated mutational state.

proteins. All posterior probabilities are shown for the different proteins in Fig. 3. When precore/core sequences were analyzed, posterior probabilities >0.8 were calculated for a total of 295 residues, consistent with strong evidence for HLA class I selection (Fig. 3B). In contrast, there were no posterior probabilities >0.8 when HLA class I genotypes were randomly assigned to the sequences (Fig. 3A), confirming that HLA class I genotypes strongly contribute to HAMdetector results. In the majority of cases with high posterior probabilities (212 of 295; 71.9%), the amino acids were variations from the majority consensus sequence. However, there were also instances (83 of 295) where high posterior probabilities were observed for the majority consensus amino acid. Similar to what has been observed in HIV and HCV, escape mutations may have accumulated at the population level and become the majority consensus residue.[23–29] Notably, evidence for selection of the consensus amino acid was more frequently observed for high-frequency HLA alleles in our cohort. For example, selection of consensus was linked to the high-frequency allele HLA-A*02 in 18.1% of cases, whereas selection of variations from consensus was driven by HLA-A*02 in only 6.6% of cases ($p = 0.0046$).

Interestingly, there were strong differences between the different HBV proteins. In contrast to the precore/core protein, only a few positions had posterior probabilities ≥0.8 in the viral envelope protein (large HBsAg; Fig. 3C) and the HBx protein (Fig. 3D), suggesting that HLA class I selection is less reproducible here. In HBV pol, posterior probabilities ≥0.8 were mainly observed in the N-terminal region of the protein that overlaps with the core region on the HBV genome (Fig. 3E). The complete list with all posterior probabilities for all HBV proteins are provided as described in the data availability statement.

Although HAMdetector integrates information on HLA class I-binding motifs into the model, it does not incorporate existing information on HLA class I-restricted epitopes in HBV. There are several possible mechanisms for increased evidence of HLA class I-associated selection pressure on the precore/core protein. These include a higher frequency of epitopes in the precore/core region or stronger selection pressure on epitopes in the precore/core region. To address this, all previously described and fully mapped HLA class I-restricted epitopes from the immune epitope database were analyzed in more detail. The complete list of epitopes with the respective posterior probabilities are available as described in the data availability statement. Although the number of described and experimentally matched epitopes is higher for the polymerase (54 epitopes) than the core protein (42 epitopes), the maximum posterior probabilities were higher for residues within the core epitope region (Fig. 3E). The previously described epitopes in HBsAg and HBx also showed lower maximum posterior probabilities compared to epitopes in the core protein. Taken together, the results are consistent with a higher degree of HLA class I-associated selection pressure on epitopes in the HBV precore/core protein compared to the remaining HBV proteins.

Viral adaptation to HLA class I-associated selection pressure correlates with markers of viral replication

The posterior probability of a given residue is a quantitative measure of the probability that the residue is a HAM. We next sought to establish a score as a quantitative measure for the level of adaptation to HLA class I expression in a complete protein. To calculate such an "*adaptation score*", we created a majority consensus reference sequence for each genotype and compared it to the individual sequences of the patient. The adaption score of an individual sequence was then calculated as the sum of all maximum posterior probabilities, for all relevant HLA class I alleles of the individual patient, in positions that differed from the consensus sequence (Fig. 4A). Accordingly, viral proteins with high sequence homology to the consensus sequence tend to have low adaptation scores. Conversely, viral proteins with multiple differences from the consensus sequence tend to have a high adaptation score if the differences are likely to be HAMs according to the HAMdetector results.

We then investigated whether these adaptation scores correlated with markers of viral replication. An important marker for the clinical classification of chronic HBV infection is HBeAg serostatus.[2] Therefore, we compared the HLA class I adaption scores between 58 HBeAg-positive patients and the 350 HBeAg-negative patients for whom the HBeAg serostatus was available. Consistent with a higher degree of CD8 T-cell selection pressure in HBeAg-negative hepatitis B, HLA class I adaptation scores were significantly higher in HBeAg-negative patients compared to HBeAg-positive patients (Fig. 4B; $p < 0.0001$). There was a negative correlation between the HLA class I adaptation scores and the serum HBV DNA concentrations (Fig. 4C; $p < 0.0001$; $r = -0.2993$) and the HBsAg level (Fig. 4D; $p < 0.0001$; $r = -0.2867$). Notably, the adaptation scores also positively correlated with age (Fig. 4E; $p < 0.0001$; $r = 0.3336$), suggesting that adaptation to HLA class I-associated selection pressure increases over time.

Taken together, our analysis of all HBV proteins suggested different levels of HLA class I-associated selection pressure, with particularly strong selection pressure on the HBV precore/core protein. Furthermore, quantification of the levels of HLA class I adaptation for individual isolates revealed differences, which correlate with markers of replication and age. HBeAg-positive HBV infection is associated with low levels of HLA class I adaptation. Moreover, high levels of HLA class I adaptation were observed in patients with low viral load and low HBsAg levels.

Discussion

Selection of escape mutations in targeted epitopes of the CD8 T-cell response has been well described in chronic viral hepatitis.[9,30] In HBV infection, there is also strong evidence for selection of escape mutations in the epitope core$_{18\text{-}27}$, which is supported by functional experiments showing the impact of epitope variants on the CD8 T-cell response.[10,11,13,17,31] Reproducible immune selection of virus mutations in hosts sharing the same HLA class I allele can be detected as statistical associations between viral sequence polymorphisms and HLA class I alleles at the population level.[13,32–34] In a previous analysis of viral sequences from the HBV precore/core we found strong evidence for HLA class I-associated selection pressure and were able to use a viral sequencing approach for identification of novel CD8 T-cell epitopes.[13] Here, we extended the analysis to all HBV open-reading frames, including the proteins precore/core, pol, HBsAg and HBx. In this analysis we utilized HAMdetector, which was specifically

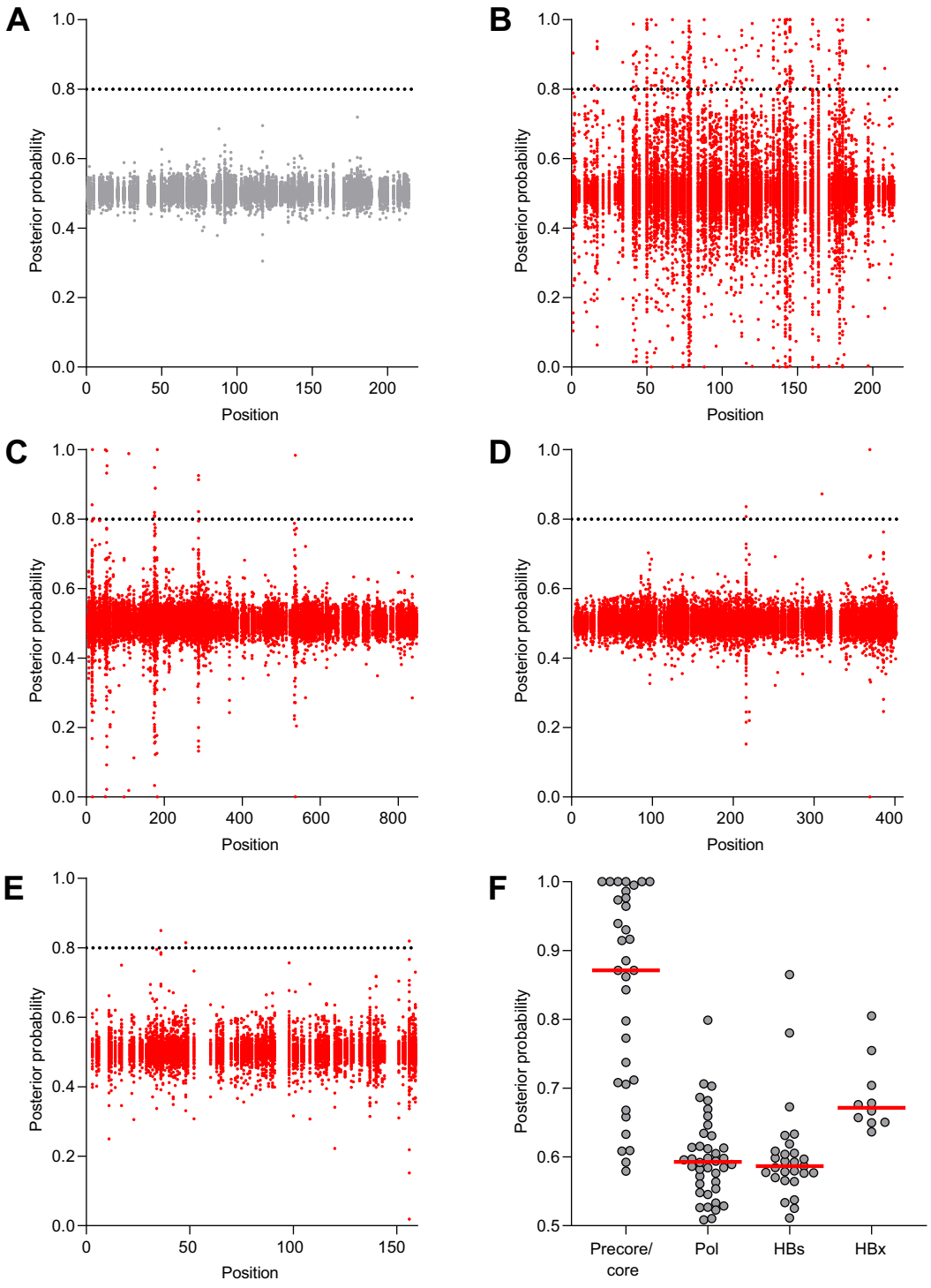


Fig. 3. HLA-associated mutational states in HBV proteins. Alignments of the four proteins precore/core, pol, HBsAg and HBx were analyzed with HAMdetector. (A) HAMdetector results for the precore/core protein with randomly assigned HLA class I genotypes (Posterior probability was calculated using the Bayesian approach, as implemented in the HAMdetector). (B) HAMdetector results for the precore/core protein with correctly assigned HLA class I genotypes (Posterior probability was calculated using the Bayesian approach, as implemented in the HAMdetector). (C-E) HAMdetector results for (C) pol, (D) HBsAg and (E) HBx (Posterior probability was calculated using the Bayesian approach, as implemented in the HAMdetector). The dotted line represents a 0.8 posterior probability threshold. In a previous study[22] posterior probabilities ≥0.8 were strongly indicative of true CD8 T-cell epitopes. (F) Previously described and fully mapped HLA class I-restricted epitopes were retrieved from a public database (IEDB). For each epitope, the maximal posterior probability for all amino acid position of the epitope was calculated for the relevant restricting HLA class I type and is shown for epitopes located in precore/core, pol, HBsAg and HBx (Posterior probability was calculated using the Bayesian approach, as implemented in the HAMdetector). HAM, HLA-associated mutational state; HBsAg, hepatitis B surface antigen; HBx, hepatitis B x protein.

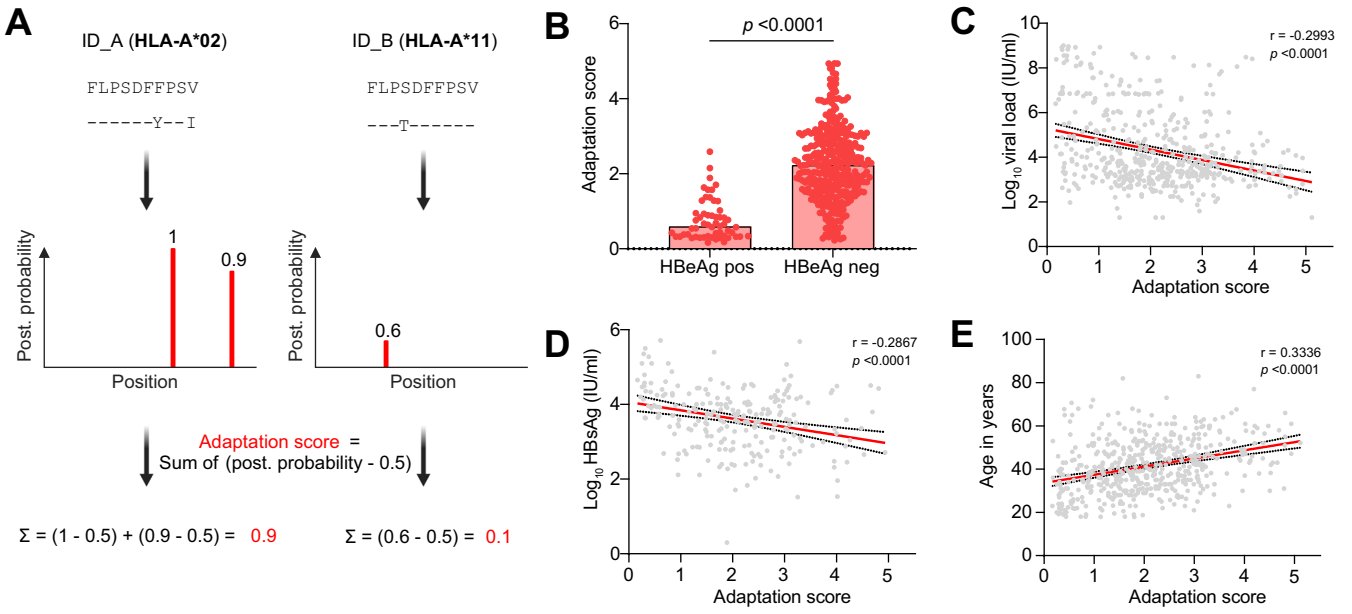


Fig. 4. Adaptation of HBV to HLA class I-associated selection pressure correlates with markers of HBV replication. (A) Schematic illustration of the adaptation score calculation. The adaptation score represents the number of amino acid substitutions in each sequence compared to a consensus sequence and uses maximum posterior probabilities from HAMdetector to weigh the substitutions based on the likelihood that they were selected by HLA. (B) Comparison of adaptation scores for isolates from HBeAg-positive and HBeAg-negative HBV infection as determined by the HBeAg serostatus (Statistical significance was calculated by Welch's t-test). (C) Correlation of adaptation scores of viral sequences with HBV DNA concentrations in serum (Statistical significance was calculated by Simple linear regression). (D) Correlation of adaptation scores of viral sequences with HBsAg levels in serum (Statistical significance was calculated by Simple linear regression). (E) Correlation of adaptation scores with age (Statistical significance was calculated by Simple linear regression). The red line represents the linear regression analysis and the dotted line the margins of the 95% confidence intervals. HAM, HLA-associated mutational state; HBeAg, hepatitis B e antigen.

designed for detection of HAMs in viral genomes.[22] Compared to conventional statistical approaches utilizing HLA class I genotypes and mutation frequencies in the viral genome, HAMdetector has the advantage that additional information such as HLA class I motifs and the phylogeny of the viral sequences are integrated into one model. This allows for more accurate detection of HAMs by increasing the sensitivity and decreasing the rate of false positive results.[22] HAMdetector has been applied to different datasets of viral sequences including HIV, HBV and HDV.[18,22] The in-depth analysis of residues under selection pressure provided here may promote future studies on epitope identification and mapping that will be required when exploring the functional differences of CD8 T cells directed against different HBV proteins.

In most cases, the consensus amino acid will represent the viral state in the absence of immune pressure. However, this does not hold true for all situations when viral infections are studied at the population level. There are multiple instances where the consensus sequence of circulating viral isolates within a population is modulated by immune selection. For example, in HIV and HCV, the accumulation of escape mutations at the population level and their fixation in the consensus sequence has been well-documented.[23–29] Several mechanisms contribute to this accumulation, including the transmission of escape mutations to new hosts,[35,36] lack of reversion in the absence of immune pressure when the fitness cost of the escape variant is low,[28,35,37] and selection by highly frequent HLA alleles in the population.[24,27,38] Together, these mechanisms influence substitution frequencies in circulating isolates, potentially leading to the replacement of the original consensus residue with an escape residue. Indeed, evidence from HIV and HCV suggests that consensus sequences may differ between populations due to variations in HLA allele frequencies.[24,27,38] This is consistent with our dataset, where 28.1% of residues with evidence for immune selection pressure correspond to the majority consensus amino acid. This suggests that HBV may already have adapted to some extent to high-frequency HLA alleles at the population level.

When HAMdetector was applied to alignments of the different HBV proteins, the results suggested that the frequency of HAMs was substantially higher in the precore/core protein compared to the other proteins. Importantly, systematic studies of the breadth of the CD8 T-cell response against different HBV proteins are lacking. It is therefore unclear if the overall epitope density in the core protein is higher or if the individual CD8 T-cell responses against epitopes in core exert more reproducible selection pressure. Although not definitive, our analysis of described CD8 T-cell epitopes in HBV proteins from a public database suggests that reproducible selection of escape mutations seems to be a characteristic of core epitopes. This would suggest either host differences in the quantity or quality of the CD8 T-cell response between viral proteins or differences between the targeted proteins in their ability to accommodate escape mutations. In line with differences between the magnitude of the CD8 T-cell response against different HBV proteins, more robust CD8 T-cell responses against the core protein than the envelope protein have been reported in chronic HBeAg-negative infection, but this difference was not fully consistent and depended on the disease stage.[39–41] Interestingly, the phenotype of CD8 T cells directed against the core protein also differed from that of CD8 T cells directed against HBV pol,[42,43] suggesting that functional

differences may contribute to the higher degree of selection pressure on the core protein. That said, we cannot exclude that the ability of the virus to accommodate escape mutations differs between viral proteins. Of note, as an essential part of the virus particle, the core protein is subject to functional constraints that limit its sequence diversity. When we compared the entropy at the amino acid level between the HBV proteins, we did not find any differences for amino acid positions with strong support for selection in HAMdetector (Fig. S2).

Importantly, based on the secreted form of the precore protein, different stages of chronic HBV infection are distinguished.[2] During HBeAg-positive HBV infection, the precore protein is translated and further processed and secreted as the HBeAg, which can be detected in serological assays. The exact functional role of the secreted HBeAg is not fully understood, but it has been associated with a state of "tolerance" characterized by high HBV DNA concentrations and no or only mild inflammation in the liver.[3] In our analysis, HBeAg-positive HBV infection was associated with lower levels of HLA class I adaptation compared to HBeAg-negative HBV infection. This is in line with prior studies of the precore/core protein where substitutions were enriched during the HBeAg-negative phase of infection.[13,17,44–46] Collectively, the data strongly support that the sequence diversity in HBV core is strongly influenced by CD8 T-cell pressure in HBeAg-negative stages but that there is less selection pressure in HBeAg-positive infection.

Different levels of adaptation to HLA class I-associated selection pressure also correlated with HBV DNA and HBsAg concentrations in serum. High levels of adaptation to HLA class I-associated selection pressure were observed in isolates from patients with low viral load and low HBsAg levels. The correlations also suggest that viral adaptation to HLA class I pressure may cause fitness costs by impairing viral replication. Notably, the negative correlation was mainly driven by HLA adaptation of the precore/core protein, for which impaired virion production was previously observed in HBeAg-negative hepatitis.[47] Analysis of the mechanisms by which substitutions in the precore/core protein may cause impairment of replication would require further studies.[48] Importantly, it is also possible that viral fitness is not impaired and low HBV DNA concentrations are the consequence of the immune response. In this case, high HLA class I adaptation scores may simply be a marker of a functional CD8 T-cell response that is responsible for inhibition of viral replication.

The cohort studied here came from different institutions in Central Europe and represents the patients in the hepatology outpatient clinics from this area, with a predominance of genotypes D and A. Accordingly, although the analysis tool controls for the underlying phylogeny of the sequences, there may be genotype-specific differences that were not detected in this study. The viral sequence was obtained prior to antiviral treatment; however, this may lead to a selection bias in our cohort when patients with low viral load and normal transaminase levels were preferentially included. We specifically tried to include HBeAg-positive patients; however, untreated HBeAg-positive HBV infection is uncommon in such cohorts and was present in only 14.2% of our cohort. Nevertheless, the strong difference in the level of HLA class I adaptation between HBeAg-positive and HBeAg-negative patients is striking. Unfortunately, we were only able to perform cross-sectional studies between these groups. Thus, it is unclear if HAMs are only selected upon transition from HBeAg-positive to HBeAg-negative HBV infection or if this is a continuous process during persistent HBeAg-negative HBV infection. Selection of variants and increased substitution rates in the HBV quasispecies have been described during the HBeAg seroconversion phase.[44] However, the level of adaptation also correlated with age. Using patient age as a proxy for duration of infection, this strongly suggests that viral genomes continuously accumulate HAMs over time. More detailed longitudinal analyses of viral sequences combined with CD8 T-cell studies would be required to address this, which is difficult to perform as most patients with relevant HBV replication receive antiviral treatment to prevent potential liver disease.

In summary, our study provides important insights into the extent of HLA class I-associated selection pressure on HBV and highlights that CD8 T-cell pressure strongly contributes to sequence diversity in the HBV core protein. Moreover, different levels of adaptation to CD8 T-cell pressure were observed, which correlate with markers of viral replication. These different levels of adaptation are relevant for the further development of T cell-based therapeutic strategies, as they may represent different levels of susceptibility to immune therapy. We would hypothesize that less adapted isolates would be more susceptible to T-cell therapy, which needs to be further studied.

Affiliations

[1]*Institute of Virology, University of Düsseldorf, Faculty of Medicine, Düsseldorf, Germany;* [2]*Clinic for Internal Medicine II, Freiburg University Medical Center, Faculty of Medicine, University of Freiburg, Freiburg im Breisgau, Germany;* [3]*Bioinformatics and Computational Biophysics, Center for Medical Biotechnology (ZMB), University of Duisburg-Essen, Essen, Germany;* [4]*Department of Internal Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany;* [5]*Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts, USA;* [6]*Department of Gastroenterology, Hepatology, Infectious Diseases and Endocrinology, Hannover Medical School, Hannover, Germany;* [7]*Department of Gastroenterology, Hepatology and Infectious Diseases, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University, Düsseldorf, Germany;* [8]*Institute for Transfusion Medicine, University Hospital Essen, University Duisburg-Essen, Essen, Germany;* [9]*Division of Infection and Immunity, Institute of Immunity and Transplantation, University College London, London, United Kingdom;* [10]*Department of Internal Medicine I, University Hospital of Bonn, Bonn, Germany;* [11]*Institute of Virology, Charité-Universitätsmedizin Berlin, Corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany;* [12]*Department of Gastroenterology and Hepatology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany*

Abbreviations

HAM, HLA-associated mutational state; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBx, hepatitis B x protein; PBMC, peripheral blood mononuclear cells; TCR, T-cell receptor.

Financial support

The study was partially funded by the German Research Foundation (DFG; TI 323/4–1 to J.T. and TRR 179, project number 272983813, to C.N.H.), the Stiftung zur Erforschung infektiös-immunologischer Erkrankungen (AW, 10-16-72) and the Jürgen Manchot Foundation, Germany. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Authors' contributions

The project was conceived by TS, AW and JT. Experiments were performed by TS, JP, MD, CM, EA, JLM, MM, HK, AK, SG, FH and AW. EA, JSW, GL, HK, MC, AK, SG, HHB, PAH, MKM, RT, HW, JN, FMH, TL and CNH contributed samples

and clinical data. All authors contributed to the interpretation of the results. The original draft was written by AW and JT and finalized with input from all authors. All authors contributed to the article and approved the submitted version.

Conflict of interest

TS, JP, MD, SG, DH, FMH, JLM, TL, CM, AK, ESA, CNH, DH, MM, RT, PH declare no conflict of interest. JT reports honoraria for lectures from AbbVie and GSK. AW has received grants for HCV genomic surveillance from Gilead Sciences. HB reports honoraria for lectures or presentations from AbbVie and Gilead Sciences, travel support from AbbVie and Gilead Sciences, and participates on advisory boards for AbbVie, Gilead Sciences, and Ipsen. MKM reports funding from Gilead Sciences for HBV immunology research and received consulting fees from Astrivax and Moderna. MC reports consulting fees from Gilead Sciences, AbbVie, MSD Sharp & Dohme, Falk Foundation, and is the medical CEO of the German Liver Foundation. JN reports grants from the German Research Council, the German Center for Infection Research, the Hector Foundation, and the German Cancer Aid. JSzW reports grants from EU Horizon 20/20, DZIF, DFG CRC1328. HW reports grants from Abbott, Biotest, consulting fees from Abbott, Bristol-Myers-Squibb, F. Hoffmann-La Roche Ltd., Gilead Science, GlaxoSmithKline, Janssen, Roche Diagnostics International Ltd., Vir Biotechnology Inc. and honoraria from Biotest Ag and Gilead Science. HK reports grants from the German Center for Infection Research, EU Horizon Health, German Research Foundation, and internal Funds from Hannover Medical School and support for attending the International Delta-Cure Meeting and EASL/AASLD Masterclass. GL reports funding from NIH.

Please refer to the accompanying ICMJE disclosure forms for further details.

Data availability statement

The authors openly provide the data presented in this manuscript through the OSF Home service (https://osf.io/e7spw/?view_only=0e72873cc5fa413a9a3c8b0d2b9c9e37). Sequence data are available at GenBank (MZ043025-MZ043097; MZ097624-MZ097884).

Acknowledgements

The authors are very grateful to the study participants for taking part in the study. The authors thank Alexandra Graupner, Anja Voges and Eugen Bäcker for technical help.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2024.10.047.

References

Author names in bold designate shared co-first authorship

[1] World Health Organization (WHO). Fact sheet hepatitis B, Update 2022. 2022 [cited; Available from: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b; 2022.
[2] European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol 2017;67:370–398.
[3] Milich DR. Is the function of the HBeAg really unknown? Hum Vaccin Immunother 2019;15:2187–2191.
[4] Tsai SL, Chen PJ, Lai MY, et al. Acute exacerbations of chronic type B hepatitis are accompanied by increased T cell responses to hepatitis B core and e antigens. Implications for hepatitis B e antigen seroconversion. J Clin Invest 1992;89:87–96.
[5] Maini MK, Burton AR. Restoring, releasing or replacing adaptive immunity in chronic hepatitis B. Nat Rev Gastroenterol Hepatol 2019;16:662–675.
[6] Bertoletti A, Ferrari C. Adaptive immunity in HBV infection. J Hepatol 2016;64:S71–S83.
[7] Lang-Meli J, Neumann-Haefelin C, Thimme R. Immunotherapy and therapeutic vaccines for chronic HBV infection. Curr Opin Virol 2021;51:149–157.
[8] Maini MK, Pallett LJ. Defective T-cell immunity in hepatitis B virus infection: why therapeutic vaccination needs a helping hand. Lancet Gastroenterol Hepatol 2018;3:192–202.
[9] Salimi Alizei E, Hofmann M, Thimme R, et al. Mutational escape from cellular immunity in viral hepatitis: variations on a theme. Curr Opin Virol 2021;50:110–118.
[10] Bertoletti A, Costanzo A, Chisari FV, et al. Cytotoxic T lymphocyte response to a wild type hepatitis B virus epitope in patients chronically infected by variant viruses carrying substitutions within the epitope. J Exp Med 1994;180:933–943.
[11] Bertoletti A, Sette A, Chisari FV, et al. Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells. Nature 1994;369:407–410.
[12] Ferrari C, Penna A, Bertoletti A, et al. Cellular immune response to hepatitis B virus-encoded antigens in acute and chronic hepatitis B virus infection. J Immunol 1990;145:3442–3449.
[13] Kefalakes H, Budeus B, Walker A, et al. Adaptation of the hepatitis B virus core protein to CD8 T-cell selection pressure. Hepatology 2015;62:47–56.
[14] Rehermann B, Fowler P, Sidney J, et al. The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis. J Exp Med 1995;181:1047–1058.
[15] Khakoo SI, Ling R, Scott I, et al. Cytotoxic T lymphocyte responses and CTL epitope escape mutation in HBsAg, anti-HBe positive individuals. Gut 2000;47:137–143.
[16] Oudshoorn M, Horn PA, Tilanus M, et al. Typing of potential and selected donors for transplant: methodology and resolution. Tissue Antigens 2007;69(Suppl 1):10–12.
[17] Walker A, Schwarz T, Brinkmann-Paulukat J, et al. Immune escape pathways from the HBV core(18-27) CD8 T cell response are driven by individual HLA class I alleles. Front Immunol 2022;13:1045498.
[18] **Magvan B**, **Kloeble AA**, Ptok J, et al. Sequence diversity of hepatitis D virus in Mongolia. Front Med (Lausanne) 2023;10:1108543.
[19] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772–780.
[20] Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 2001;17:754–755.
[21] Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 2016;44:W242–W245.
[22] Habermann D, Kharimzadeh H, Walker A, et al. HAMdetector: a Bayesian regression model that integrates information to detect HLA-associated mutations. Bioinformatics 2022;38:2428–2436.
[23] Timm J, Li B, Daniels MG, et al. Human leukocyte antigen-associated sequence polymorphisms in hepatitis C virus reveal reproducible immune responses and constraints on viral evolution. Hepatology 2007;46:339–349.
[24] Neumann-Haefelin C, Frick DN, Wang JJ, et al. Analysis of the evolutionary forces in an immunodominant CD8 epitope in hepatitis C virus at a population level. J Virol 2008;82:3438–3451.
[25] Payne R, Muenchhoff M, Mann J, et al. Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. Proc Natl Acad Sci U S A 2014;111:E5393–E5400.
[26] Leslie A, Kavanagh D, Honeyborne I, et al. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. J Exp Med 2005;201:891–902.
[27] Carlson JM, Brumme CJ, Martin E, et al. Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. J Virol 2012;86:13202–13216.
[28] Kawashima Y, Pfafferott K, Frater J, et al. Adaptation of HIV-1 to human leukocyte antigen class I. Nature 2009;458:641–645.
[29] Kloverpris HN, Leslie A, Goulder P. Role of HLA adaptation in HIV evolution. Front Immunol 2015;6:665.
[30] Timm J, Walker CM. Mutational escape of CD8+ T cell epitopes: implications for prevention and therapy of persistent hepatitis virus infections. Med Microbiol Immunol 2015;204:29–38.
[31] Bertoletti A, Southwood S, Chesnut R, et al. Molecular features of the hepatitis B virus nucleocapsid T-cell epitope 18-27: interaction with HLA and T-cell receptor. Hepatology 1997;26:1027–1034.
[32] Ruhl M, Knuschke T, Schewior K, et al. CD8+ T-cell response promotes evolution of hepatitis C virus nonstructural proteins. Gastroenterology 2011;140:2064–2073.
[33] Moore CB, John M, James IR, et al. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science 2002;296:1439–1443.
[34] **Karimzadeh H**, **Kiraithe MM**, **Oberhardt V**, et al. Mutations in hepatitis D virus allow it to escape detection by CD8(+) T cells and evolve at the population level. Gastroenterology 2019;156:1820–1833.
[35] Schneidewind A, Brumme ZL, Brumme CJ, et al. Transmission and long-term stability of compensated CD8 escape mutations. J Virol 2009;83:3993–3997.
[36] Carlson JM, Du VY, Pfeifer N, et al. Impact of pre-adapted HIV transmission. Nat Med 2016;22:606–613.
[37] Davenport MP, Loh L, Petravic J, et al. Rates of HIV immune escape and reversion: implications for vaccination. Trends Microbiol 2008;16:561–566.
[38] Xia Y, Pan W, Ke X, et al. Differential escape of HCV from CD8(+) T cell selection pressure between China and Germany depends on the presenting HLA class I molecule. J Viral Hepat 2019;26:73–82.

[39] Aliabadi E, Urbanek-Quaing M, Maasoumy B, et al. Impact of HBsAg and HBcrAg levels on phenotype and function of HBV-specific T cells in patients with chronic hepatitis B virus infection. Gut 2022;71:2300–2312.
[40] **Le Bert N**, **Gill US**, **Hong M**, et al. Effects of hepatitis B surface antigen on virus-specific and global T cells in patients with chronic hepatitis B virus infection. Gastroenterology 2020;159:652–664.
[41] Park JJ, Wong DK, Wahed AS, et al. Hepatitis B virus–specific and global T-cell dysfunction in chronic hepatitis B. Gastroenterology 2016;150:684–695.e685.
[42] Schuch A, Salimi Alizei E, Heim K, et al. Phenotypic and functional differences of HBV core-specific versus HBV polymerase-specific CD8+ T cells in chronically HBV-infected patients with low viral load. Gut 2019;68:905–915.
[43] Hoogeveen RC, Robidoux MP, Schwarz T, et al. Phenotype and function of HBV-specific T cells is determined by the targeted epitope in addition to the stage of infection. Gut 2019;68:893–904.
[44] Lim SG, Cheng Y, Guindon S, et al. Viral quasi-species evolution during hepatitis Be antigen seroconversion. Gastroenterology 2007;133:951–958.
[45] Kramvis A, Kostaki EG, Hatzakis A, et al. Immunomodulatory function of HBeAg related to short-sighted evolution, transmissibility, and clinical manifestation of hepatitis B virus. Front Microbiol 2018;9:2521.
[46] Rehermann B, Pasquinelli C, Mosier SM, et al. Hepatitis B virus (HBV) sequence variation of cytotoxic T lymphocyte epitopes is not common in patients with chronic HBV infection. J Clin Invest 1995;96:1527–1534.
[47] **Volz T**, **Lutgehetmann M**, Wachtler P, et al. Impaired intrahepatic hepatitis B virus productivity contributes to low viremia in most HBeAg-negative patients. Gastroenterology 2007;133:843–852.
[48] Blondot ML, Bruss V, Kann M. Intracellular transport and egress of hepatitis B virus. J Hepatol 2016;64:S49–S59.

Keywords: Hepatitis B Virus; CD8 T-cell pressure; sequence adaptation; HLA-associated mutational states (HAM); viral load; HBeAg; whole-genome sequencing.
Received 1 March 2024; received in revised form 8 October 2024; accepted 28 October 2024; available online 12 November 2024

Rapid communication

Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020

Andreas Walker[1,2] , Torsten Houwaart[2,3] , Tobias Wienemann[3] , Malte Kohns Vasconcelos[3] , Daniel Strelow[3] , Tina Senff[1] , Lisanna Hülse[3] , Ortwin Adams[1] , Marcel Andree[1] , Sandra Hauka[1] , Torsten Feldt[4] , Björn-Erik Jensen[4] , Verena Keitel[4] , Detlef Kindgen-Milles[5] , Jörg Timm1 , Klaus Pfeffer[3] , Alexander T Dilthey[3]

1. Institute of Virology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
2. These authors contributed equally
3. Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
4. Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
5. Department of Anaesthesiology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Correspondence: Alexander T Dilthey (alexander.dilthey@med.uni-duesseldorf.de)

Citation style for this article:
Walker Andreas , Houwaart Torsten , Wienemann Tobias , Vasconcelos Malte Kohns , Strelow Daniel , Senff Tina , Hülse Lisanna , Adams Ortwin , Andree Marcel , Hauka Sandra , Feldt Torsten , Jensen Björn-Erik , Keitel Verena , Kindgen-Milles Detlef , Timm Jörg , Pfeffer Klaus , Dilthey Alexander T . Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020. Euro Surveill. 2020;25(22):pii=2000746. https://doi.org/10.2807/1560-7917.ES.2020.25.22.2000746

Article submitted on 25 Apr 2020 / accepted on 04 Jun 2020 / published on 04 Jun 2020

We whole-genome sequenced 55 SARS-CoV-2 isolates from Germany to investigate SARS-CoV-2 outbreaks in 2020 in the Heinsberg district and Düsseldorf. While the genetic structure of the Heinsberg outbreak indicates a clonal origin, reflecting superspreading dynamics from mid-February during the carnival season, distinct viral strains were circulating in Düsseldorf in March, reflecting the city's international links. Limited detection of Heinsberg strains in the Düsseldorf area despite geographical proximity may reflect efficient containment and contact-tracing efforts.

We report on the genetic structure of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in North-Rhine Westphalia, Germany's most populous state (18 million inhabitants). Our analysis includes the 'Heinsberg outbreak' [1], which started in the second half of February 2020 – comprising a superspreading event at a carnival session in Gangelt, a small municipality of ca 12,000 inhabitants on the border between Germany and the Netherlands – and subsequent outbreak dynamics in March, in the state capital Düsseldorf, located 70 km from Gangelt and an international economic and air travel hub of ca 600,000 inhabitants.

Severe acute respiratory syndrome coronavirus 2 genome sequencing

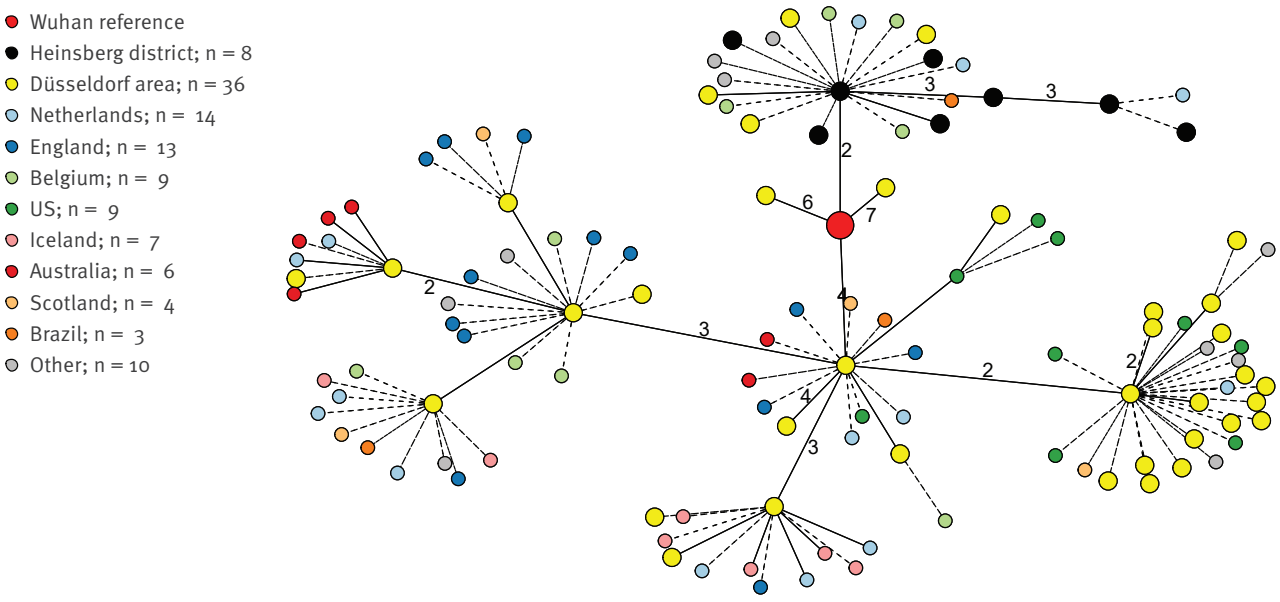
The institute of virology at Düsseldorf University Hospital was one of the first laboratories to offer SARS-CoV-2 diagnostics in North-Rhine Westphalia. A total of 55 SARS-CoV-2 isolate samples were acquired from diagnostic swabs sent to this institute in February and March 2020. Of these, 10 were directly linked to the Heinsberg outbreak (obtained from medical practices in the Heinsberg district or from patients treated at Düsseldorf University Hospital who were Heinsberg district residents) and 45 originated from the city of Düsseldorf and surrounding districts.

RNA extraction and reverse transcription were carried out as previously described [2]. DNA amplification and sequencing on the Oxford Nanopore platform were carried out according to the Artic protocol [3,4] (Supplementary Text), yielding between 31 and 582 Mb of raw sequencing data per sample (Supplementary Table S1). Bioinformatic analysis was based on the Artic pipelines and additional manual curation was carried out (Supplementary Text), yielding completely resolved genomes with 2–13 polymorphic positions (Supplementary Table S2) relative to the SARS-CoV-2 reference genome [5].

Of note, we observed evidence for ambiguities at polymorphic positions in 11 of 55 samples (Supplementary Table S2); for one such sample (NRW-39; 13 positions called as multi-allelic), PCR was repeated and a separate sequencing run was carried out, confirming the detected ambiguities (Supplementary Text). Further work is necessary to investigate whether ambiguities represent within-patient viral quasispecies.

Figure 1

Minimum spanning tree of severe acute respiratory syndrome coronavirus 2 sequences, showing 44 unambiguously[a] resolved genomes from the Heinsberg district (n = 8) and the Düsseldorf area (n = 36), Germany, February–March 2020



GISAID: Global Initiative on Sharing All Influenza Data; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; US: United States.

[a] Eleven isolate genomes from 55 in our study with ambiguities are not considered 'unambiguously resolved' for the purposes of this analysis and are thus omitted.

The original Wuhan SARS-CoV-2 reference genome; and a sample of closely related publicly available SARS-CoV-2 genomes from GISAID (distance to any of the Heinsberg/Düsseldorf genomes of 0 or 1; see Supplementary Text for details) are included in the analysis.

Dashed and solid edges without adjacent numbers indicate distances of 0 and 1, respectively; all other distances are shown explicitly.

In a proof-of-concept experiment, we also successfully sequenced reverse-transcribed viral cDNA from patient material without an intermediate PCR-based amplification step (Supplementary Text), potentially enabling simplified sample preparation and increased read lengths for some samples in the future.

Ethical statement

Our study was Institutional Review Board (IRB)-approved by the ethics committee of the Heinrich Heine University Düsseldorf (#2020–839).

The Heinsberg outbreak

The first cases of SARS-CoV-2 infection in Germany were detected in late January 2020 and could be linked to recent travel to Northern Italy and China [1]. On 24 and 25 February 2020, however, two members of the same household from the Heinsberg district with no known travel history to SARS-CoV-2 risk areas were diagnosed with SARS-CoV-2; by 28 February 2020, the number of confirmed infections in the Heinsberg district had grown to 37; by 22 April 2020, to›1,700 [6]. Contact tracing later showed that many of the early SARS-CoV-2 cases could be linked to a carnival session attended by the two index cases. The carnival event was held on 15 February 2020 in the municipality of Gangelt, which is part of the Heinsberg district [1]. Epidemiological investigation revealed that the index cases had travelled to the Netherlands, not considered a risk area at the time, 7 days prior to attending the Gangelt carnival session. The 'Heinsberg outbreak' represented one of the first large-scale SARS-CoV-2 outbreaks in Germany, seeded by community transmission and amplified by superspreading-type dynamics.

Genomic analysis of 10 SARS-CoV-2 isolates from the Heinsberg outbreak, sampled between 25 and 28 February and including those from the index cases, demonstrated the clonal origin of the outbreak (Figure); all Heinsberg samples shared the same two mutations compared to the SARS-CoV-2 reference genome (Supplementary Table S2). Viral diversity in the Heinsberg samples varied between two and six polymorphic positions relative to the SARS-CoV-2 reference genome, and five distinct viral variants (i.e. haplotypes) could be identified (Supplementary Table S2).

An analysis (Supplementary Text) of other publicly available SARS-CoV-2 sequences did not reveal an obvious origin of the Heinsberg outbreak (Supplementary Table S3); the Heinsberg isolates are not related to early sequences from other German outbreak areas

(Bavaria, Baden-Wuerttemberg), and, despite intense Dutch viral sampling (585 available viral genomes from the Netherlands at the time of analysis), our analysis identified only two closely related isolates from the Netherlands (one collected on 21 March, the other with undefined collection date). The role of the index cases' short vacation in the Netherlands 7 days before the Gangelt carnival session [7], while suggestive in terms of reported SARS-CoV-2 incubation periods thus remains ambiguous [8]. Moreover, large numbers of closely related isolates are circulating in many countries, for example England, Wales, and Iceland (Supplementary Table S3). The small number of polymorphisms shared by all samples in the Heinsberg outbreak (n = 2), compared with a maximum number of six per-isolate polymorphic positions in the same samples, likely acquired over a period of a few weeks, is compatible with a relatively recent introduction from China.

Düsseldorf outbreak dynamics

The first SARS-CoV-2 cases in Düsseldorf, 70 km from Gangelt, were diagnosed in early March 2020 [9]; as at 21 April 2020, the outbreak had grown to more than 900 confirmed cases [10]. The set of 55 whole-genome-sequenced isolates included 45 samples from Düsseldorf and nearby districts, collected between 3 and 23 March. A minimum spanning tree analysis of 44 unambiguously resolved viral sequences (Figure) showed that there were at least five clusters of viral stains circulating in the Düsseldorf area; the number of polymorphic positions relative to the SARS-CoV-2 reference genome in the Düsseldorf samples varied between 2 and 13 (Supplementary Table S2). Closely related strains (distance 0 or 1) were found in Australia, the United Kingdom, the United States and many other countries (Supplementary Table S3), strongly suggesting multiple independent introduction events. Of note, four 'Düsseldorf area' isolates clustered with the Heinsberg outbreak (Figure 1); of these, two were collected from residents of a district next to Heinsberg, who had been treated at the Düsseldorf University Hospital, and two remained of unclear origin (patient data not available). Thus, there was no evidence for widespread community circulation of Heinsberg-derived SARS-CoV-2 strains in the Düsseldorf area.

Illumina validation

To verify the accuracy of Nanopore-based viral assembly, additional Illumina sequencing was carried out for the first 11 samples, according to date of collection, of our cohort (Supplementary Table S1; Supplementary Text); data analysis was carried out with iVar [11]. For 41 of 45 polymorphic positions identified by either Nanopore or Illumina across the 11 samples, the called alleles agreed; manual inspection of the discordant positions revealed low coverage for two discordant positions and one missed multi-allelic call for each sequencing technology (Supplementary Table S4).

Discussion

Since its emergence in the Chinese city of Wuhan in late 2019, SARS-CoV-2 has infected more than 6 million individuals and led to more than 370,000 deaths worldwide as at 03 June 2020 [12]. As SARS-CoV-2 case numbers and the social and economic consequences of social distancing and lock-down measures continue to rise, many countries are facing difficult trade-offs. Improved methods to characterise the dynamics of viral transmission are urgently needed.

More than 10,000 globally sourced SARS-CoV-2 genomes are publicly available, and powerful data sharing and analysis platforms like the Global Initiative on Sharing All Influenza Data (GISAID) EpiCoV database [13] and Nextstrain [14] enable the collaborative analysis of viral population structure on a global level. Additional insights into transmission dynamics can be gained from focused investigations of individual outbreaks and by integrating genomic data with classical epidemiology.

Here we have investigated the genetic structure of two SARS-CoV-2 outbreaks, which occurred at two nearby locations in North-Rhine Westphalia using Nanopore sequencing, which has additional applications in many fields such as human genetics [15] and microbial metagenomics [16]. We have demonstrated the clonal origin of the Heinsberg outbreak. This is consistent with available epidemiological data pointing to a carnival session in Gangelt as the epicentre of the outbreak [1]. The lack of association between the Heinsberg samples and other early German outbreak isolates is suggestive of a separate introduction event, possibly via the Netherlands, China, or a third country. By contrast, SARS-CoV-2 isolates circulating in Düsseldorf were highly polyclonal and could be grouped into at least five clusters of viral haplotypes.

Despite the geographical proximity between Heinsberg and Düsseldorf, only four of 36 unambiguously resolved samples from the Düsseldorf area clustered with the Heinsberg outbreak, and two of these were derived from residents of a district neighbouring Heinsberg. Limited detection of Heinsberg strains in the Düsseldorf area may reflect the effectiveness of the contact-tracing efforts conducted by the German public health authorities; of note, 'lockdown'-type restrictions with limits on public gatherings in Germany were only imposed on 23 March 2020 [17], i.e. on the day on which the last sample of our study was collected.

More extensive sampling of SARS-CoV-2 isolates from North-Rhine Westphalia will be required to investigate the effect of various containment measures on transmission chains at a genomic level. Consistent with reports from Iceland [18], New York [19], and data on Nextstrain, our study has demonstrated the simultaneous circulation of distinct viral variants (i.e. haplotypes) in a metropolitan region. In the Heinsberg outbreak, we could identify five distinct variants. As

SARS-CoV-2 genomes continue to diverge as part of ongoing viral evolution, the application of genomic epidemiology [20,21] for the identification and targeted interruption of viral transmission chains will become increasingly feasible.

Data availability

All generated viral genome assemblies have been submitted to GISAID; all generated assemblies and the raw sequencing data are also available on NCBI (BioProject PRJNA627229).

Acknowledgements

This work was supported by the Jürgen Manchot Foundation and by funding from the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung; Award number 031L0184B).

We gratefully acknowledge the Authors, the Originating and Submitting Laboratories for their sequence and metadata shared through GISAID. All submitters of data may be contacted directly via GISAID. The Acknowledgments Table for GISAID is part of the Supplement (Supplementary Table S5).

We would also like to thank Nicholas Loman and Josh Quick for advice and discussions.

Conflict of interest

None declared.

Authors' contributions

AW, TH, TW, MKV, JT, KP and ATD conceptualised and designed the study. AW, DS, TS, LH, and OA designed and implemented wet-lab protocols. TH and ATD developed and implemented sequencing data analysis approaches. OA, MA, SH, TF, BJ, VK, and DKM provided clinical samples and epidemiological data and gave input into the study design. All authors have commented on the draft and approved the final version.

References

1. Robert Koch Institute (RKI). Coronavirus Disease 2019 (COVID-19) Daily Situation Report 05/03/2020. Berlin: RKI; Mar 2020. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-03-05-en.pdf?__blob=publicationFile
2. Walker A, Ennker KS, Kaiser R, Lübke N, Timm J. A pan-genotypic Hepatitis C Virus NS5A amplification method for reliable genotyping and resistance testing. J Clin Virol. 2019;113:8-13. https://doi.org/10.1016/j.jcv.2019.01.012 PMID: 30771598
3. Quick, J. ARTIC amplicon sequencing protocol for MinION for nCoV-2019.2020. https://doi.org/http://dx.doi.org/10.17504/protocols.io.bdp7i5rn
4. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12(6):1261-76. https://doi.org/10.1038/nprot.2017.066 PMID: 28538739
5. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265-9. https://doi.org/10.1038/s41586-020-2008-3 PMID: 32015508
6. Kreis Heinsberg. Coronavirus im Kreis Heinsberg. [Coronavirus in the district of Heinsberg]. Heinsberg: Kreisverwaltung Heinsberg; 2020. Available from: https://www.kreis-heinsberg.de/aktuelles/aktuelles/?pid=5149
7. Netherlands National Institute for Public Health and the Environment (RIVM). Duitse coronapatiënt niet ziek tijdens verblijf in Limburg. [German corona patient not ill during stay in Limburg]; 26 Feb 2020. Bilthoven: RIVM. Available from: https://www.rivm.nl/nieuws/duitse-coronapatient-niet-ziek-tijdens-verblijf-in-limburg
8. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020;172(9):577-82. https://doi.org/10.7326/M20-0504 PMID: 32150748
9. Pressedienst Landeshauptstadt Düsseldorf. Zwei Düsseldorfer mit Coronavirus infiziert. [Two Düsseldorfers infected with coronavirus]. Düsseldorf: Pressedienst Landeshauptstadt Düsseldorf; 2020. [Accessed 04 Jun 2020]. Available from: https://www.duesseldorf.de/medienportal/pressedienst-einzelansicht/pld/zwei-duesseldorfer-mit-coronavirus-infiziert.html
10. Pressedienst Landeshauptstadt Düsseldorf. Die Coronazahlen vom 21. April. [Coranavirus case numbers 21 April]. Düsseldorf: Pressedienst Landeshauptstadt Düsseldorf; 2020. https://www.duesseldorf.de/medienportal/pressedienst-einzelansicht/pld/die-coronazahlen-vom-21-april.html
11. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20(1):8. https://doi.org/10.1186/s13059-018-1618-7 PMID: 30621750
12. World Health Organization (WHO). Coronavirus disease 2019 (COVID-19) Situation Report – 135. Geneva: WHO; 03 Jun 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200603-covid-19-sitrep-135.pdf?sfvrsn=39972feb_2
13. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017;22(13):30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494 PMID: 28382917
14. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121-3. https://doi.org/10.1093/bioinformatics/bty407 PMID: 29790939
15. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018;36(4):338-45. https://doi.org/10.1038/nbt.4060 PMID: 29431738
16. Dilthey AT, Jain C, Koren S, Phillippy AM. Strain-level metagenomic assignment and compositional estimation for long reads with MetaMaps. Nat Commun. 2019;10(1):3066. https://doi.org/10.1038/s41467-019-10934-2 PMID: 31296857
17. Robert Koch Institute (RKI). Coronavirus Disease 2019 (COVID-19) Daily Situation Report 23/03/2020. Berlin: RKI; Mar 2020. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-03-23-en.pdf?__blob=publicationFile
18. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. N Engl J Med. 2020;NEJMoa2006100. https://doi.org/10.1056/NEJMoa2006100 PMID: 32289214
19. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary A, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. medRxiv, 2020.04.08.20056929 (preprint). https://doi.org/10.1126/science.abc1917
20. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, et al. Tracking virus outbreaks in the twenty-first century. Nat Microbiol. 2019;4(1):10-9. https://doi.org/10.1038/s41564-018-0296-2 PMID: 30546099
21. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet. 2018;19(1):9-20. https://doi.org/10.1038/nrg.2017.88 PMID: 29129921

License, supplementary material and copyright

Any supplementary material referenced in the article can be found in the online version.

Clinical Infectious Diseases

MAJOR ARTICLE

Characterization of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection Clusters Based on Integrated Genomic Surveillance, Outbreak Analysis and Contact Tracing in an Urban Setting

Andreas Walker,[1,a] Torsten Houwaart,[2,a] Patrick Finzer,[2,3,a] Lutz Ehlkes,[4,a] Alona Tyshaieva,[2] Maximilian Damagnez,[1] Daniel Strelow,[2] Ashley Duplessis,[1] Jessica Nicolai,[2] Tobias Wienemann,[2] Teresa Tamayo,[2] Malte Kohns Vasconcelos,[2] Lisanna Hülse,[2] Katrin Hoffmann,[3] Nadine Lübke,[1] Sandra Hauka,[1] Marcel Andree,[1] Martin P. Däumer,[5] Alexander Thielen,[5] Susanne Kolbe-Busch,[2] Klaus Göbels,[4] Rainer Zotz,[3] Klaus Pfeffer,[2] Jörg Timm,[1] and Alexander T. Dilthey[2,6,7]; on behalf of the German COVID-19 OMICS Initiative (DeCOI)

[1]Institute of Virology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany; [2]Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University Düsseldorf, Düsseldorf, Germany; [3]Zotz | Klimas, Düsseldorf, Germany; [4]Düsseldorf Health Department (Gesundheitsamt Düsseldorf), Düsseldorf, Germany; [5]SeqIT GmbH, Pfaffplatz 10, 67655 Kaiserslautern, Germany; [6]Institute of Medical Statistics and Computational Biology, University of Cologne, Cologne, Germany; and [7]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany

Background. Tracing of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission chains is still a major challenge for public health authorities, when incidental contacts are not recalled or are not perceived as potential risk contacts. Viral sequencing can address key questions about SARS-CoV-2 evolution and may support reconstruction of viral transmission networks by integration of molecular epidemiology into classical contact tracing.

Methods. In collaboration with local public health authorities, we set up an integrated system of genomic surveillance in an urban setting, combining a) viral surveillance sequencing, b) genetically based identification of infection clusters in the population, c) integration of public health authority contact tracing data, and d) a user-friendly dashboard application as a central data analysis platform.

Results. Application of the integrated system from August to December 2020 enabled a characterization of viral population structure, analysis of 4 outbreaks at a maximum care hospital, and genetically based identification of 5 putative population infection clusters, all of which were confirmed by contact tracing. The system contributed to the development of improved hospital infection control and prevention measures and enabled the identification of previously unrecognized transmission chains, involving a martial arts gym and establishing a link between the hospital to the local population.

Conclusions. Integrated systems of genomic surveillance could contribute to the monitoring and, potentially, improved management of SARS-CoV-2 transmission in the population.

Keywords. genomic epidemiology; infection chain; community transmission; rapid sequencing; Nanopore sequencing.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a pandemic coronavirus first detected in late 2019 [1, 2], has infected >135 million individuals and led to >2.9 million associated deaths [3]. Until the wide availability of vaccines, non-pharmaceutical interventions to limit SARS-CoV-2 transmission will continue to play an important role in pandemic management. The specific source of SARS-CoV-2 infections during community transmission, however, often remains unknown even in public health systems that operate effective contact tracing regimes (eg, for around 40% of cases in the city of Düsseldorf; Düsseldorf Health Department internal data).

Genomic epidemiology [4, 5], that is, the application of modern genomic technologies to characterize viral transmission chains [6–10], can crucially contribute to the design and evaluation of viral containment strategies. Its possible applications include the targeted investigation of putative outbreaks, for example, in hospitals [11] and care homes, as well as untargeted "surveillance sequencing" to monitor transmission dynamics and viral evolution in the population at large [12–16]. In an integrated genomic epidemiology approach, the joint analysis of surveillance, outbreak, and contact tracing data can enable the improved analysis of infection chains in the population and healthcare settings [17].

Received 19 April 2021; editorial decision 22 June 2021; published online 28 June 2021.

[a]A. W., T. H., P. F., and L. E. contributed equally to this work.

Correspondence: A. Dilthey, Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University Düsseldorf, Universitätsstr. 2, 40225 Düsseldorf, Germany (dilthey@hhu.de).

Clinical Infectious Diseases® 2022;74(6):1039–46

https://doi.org/10.1093/cid/ciab588

In summer 2020, we established and tested a fully integrated SARS-CoV-2 genomic epidemiology system in Düsseldorf, the capital of North Rhine Westphalia, a city of about 600 000 inhabitants in Germany's largest metropolitan area. Our approach combined untargeted longitudinal surveillance sequencing, implemented in collaboration with a large commercial diagnostic laboratory, analysis of putative SARS-CoV-2 outbreaks from the city's largest hospital, the integration of local public health authorities, and the development of a user-friendly dashboard to facilitate data analysis and exchange by all participating stakeholders (Figure 1).

METHODS

Surveillance Sample Collection

Surveillance sample collection was implemented in collaboration with the local diagnostic laboratory Zotz | Klimas, the largest commercial SARS-CoV-2 testing laboratory in Düsseldorf. A convenience sampling approach, arbitrarily targeting 20 – 30 samples per week with Ct value <32, was implemented; sample selection was typically carried out on a single day and no metadata were used to determine sampling choices. Selected samples were shipped to the Institute of Virology at the Heinrich Heine University for amplification.

Outbreak Sample Collection

Samples from outbreaks at Düsseldorf University Hospital were collected by local clinical staff and the employee health department and sent to the Institute of Virology at the University Hospital of Düsseldorf, which is responsible for diagnostic testing of patients and staff. All 4 putative SARS-CoV-2 outbreaks identified by the hospital's hygiene staff between September and December 2020 are included in this article. Outbreak samples were sequenced locally on the Oxford Nanopore platform or externally in collaboration with SeqIT GmbH (Kaiserslautern, Germany); see Supplementary Table 2.

Sequencing and Assembly

Full sequencing and assembly protocols for Nanopore and Illumina are specified in Supplementary Text 3. Nanopore sequencing was based on the Artic protocol [18–20].

Quality Control and Isolate Assembly Inclusion Criteria

All consensus sequences with >3000 undefined ("N") characters were classified as low quality and excluded from all further analyses, leading to the exclusion of 21 surveillance isolate assemblies (pre-filtering: 341 surveillance assemblies; post-filtering: 320 surveillance assemblies). For the remaining isolates, higher Ct values are associated with increased numbers of "N" characters (Supplementary Figure 5).

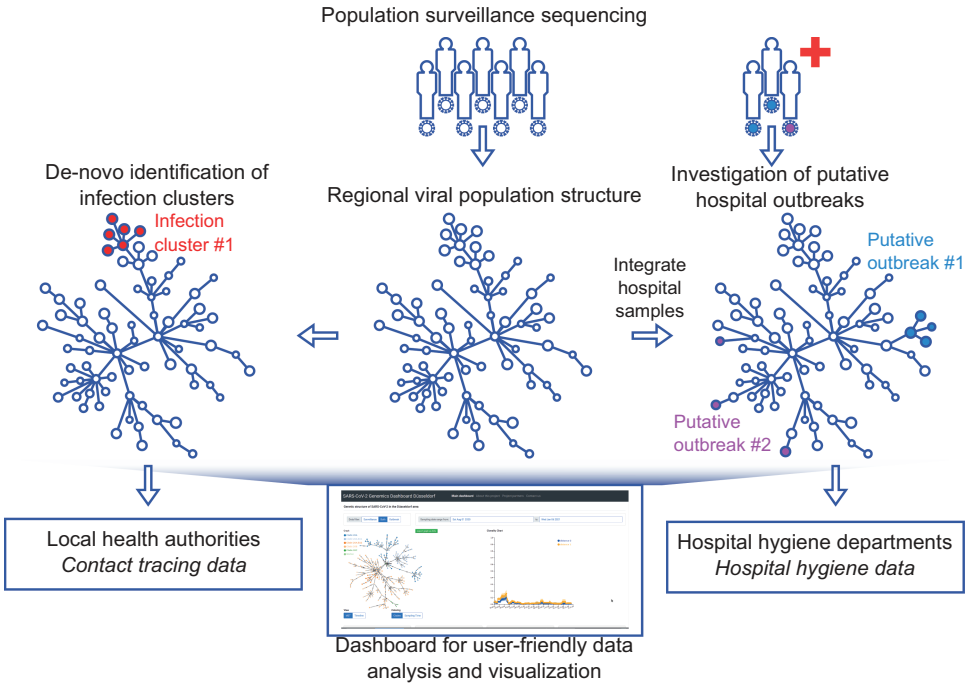


Figure 1. Integrated genomic surveillance in the Düsseldorf area. Population surveillance sequencing enables the characterization of local SARS-CoV-2 population structure, facilitating the discrimination between clonal hospital outbreaks (here: putative outbreak 1) or simultaneously detected but unrelated SARS-CoV-2 hospital ward cases (here: putative outbreak 2). Viral population surveillance data can also enable the *de novo* identification of infection clusters in the population based on the genetic data. Added value of genomic surveillance is maximized when genetic data are integrated with complementary epidemiological data or approaches, such as contact tracing or hospital outbreak data. Utilization of viral genetic data by diverse stakeholders is facilitated by providing a user-friendly real-time web application ("dashboard") for analysis and visualization of the generated viral genomes. Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

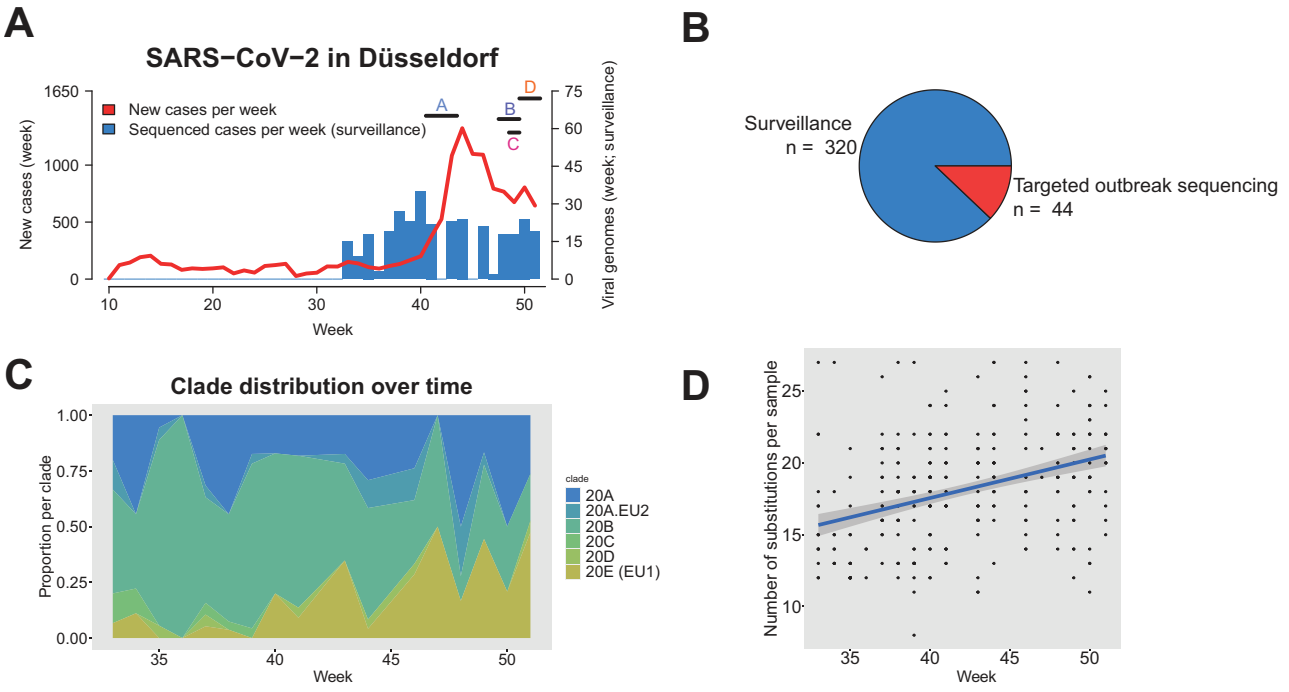


Figure 2. Local development of SARS-CoV-2 from September to December 2020. *A*, Newly diagnosed (red line) and sequenced (blue bars; by sample collection week) cases of SARS-CoV-2 by calendar week of 2020 in Düsseldorf. Horizontal bars indicate sample collection times for 4 hospital outbreaks on different wards (*A*–*D*) of Düsseldorf University Hospital. *B*, Sequenced samples by sample origin. *C*, Clade composition of surveillance samples by sample collection week, using the NextStrain [21] color scheme. *D*, Substitutions per sequenced surveillance sample and sample collection line; each dot represents one viral genome, blue line: linear fit. Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Clade Assignment and Variant Calling

Sample clades, variants and other summary statistics were computed with the NextClade tool of NextStrain [21]. Pangolin (https://github.com/cov-lineages/pangolin) was used to screen for the presence of B.1.1.7 and B.1.351.

SARS-CoV-2 Dashboard Application

The implementation of the SARS-CoV-2 dashboard web application is described in Supplementary Text 4 and visualized in Supplementary Figure 6.

Phylogenetic and Minimum Spanning Tree Analyses

Details of phylogenetic and minimum spanning tree analyses are described in Supplementary Text 3.

Identification of Putative Population Infection Clusters

Putative infection clusters in the surveillance sequencing data were identified by greedily clustering all isolate genomes with edit distance 0, using the dashboard distance matrix (see above), and filtering for clusters with ≥4 members. All candidate clusters were manually inspected.

Integration of Contact Tracing Data

We integrated contact tracing and case information data available at and collected by Düsseldorf Health Department (Gesundheitsamt Düsseldorf). All personally identifiable information remained at Düsseldorf Health Department.

RESULTS

Genomic Surveillance in the Düsseldorf Region

In collaboration with a local diagnostic lab and employing a convenience sampling approach, we obtained 320 high-quality SARS-CoV-2 viral isolate genomes from samples collected in Düsseldorf between August and December 2020 (median: 19 samples/week). The collected genomes represented 3.1% of 10 276 newly diagnosed polymerase chain reaction (PCR)-confirmed cases during the sampling period; the proportion of sequenced cases on a weekly basis varied between 0% and 20% (Supplementary Table 1), and complete or near-complete dropout due to challenges with sampling logistics during periods of high incidence was observed for 3 weeks in Fall 2020 (Supplementary Table 1). During the last 4 weeks of the sampling period, the proportion of sequenced cases stabilized between 2% and 3%. Sequencing data, sample metadata, and assembly quality are summarized in Supplementary Tables 2 and 3. By sample genome inclusion criteria (see Methods), all included isolate genomes were of high quality (<3000Ns). In total, 80/320 isolate genome consensus sequences contained at least

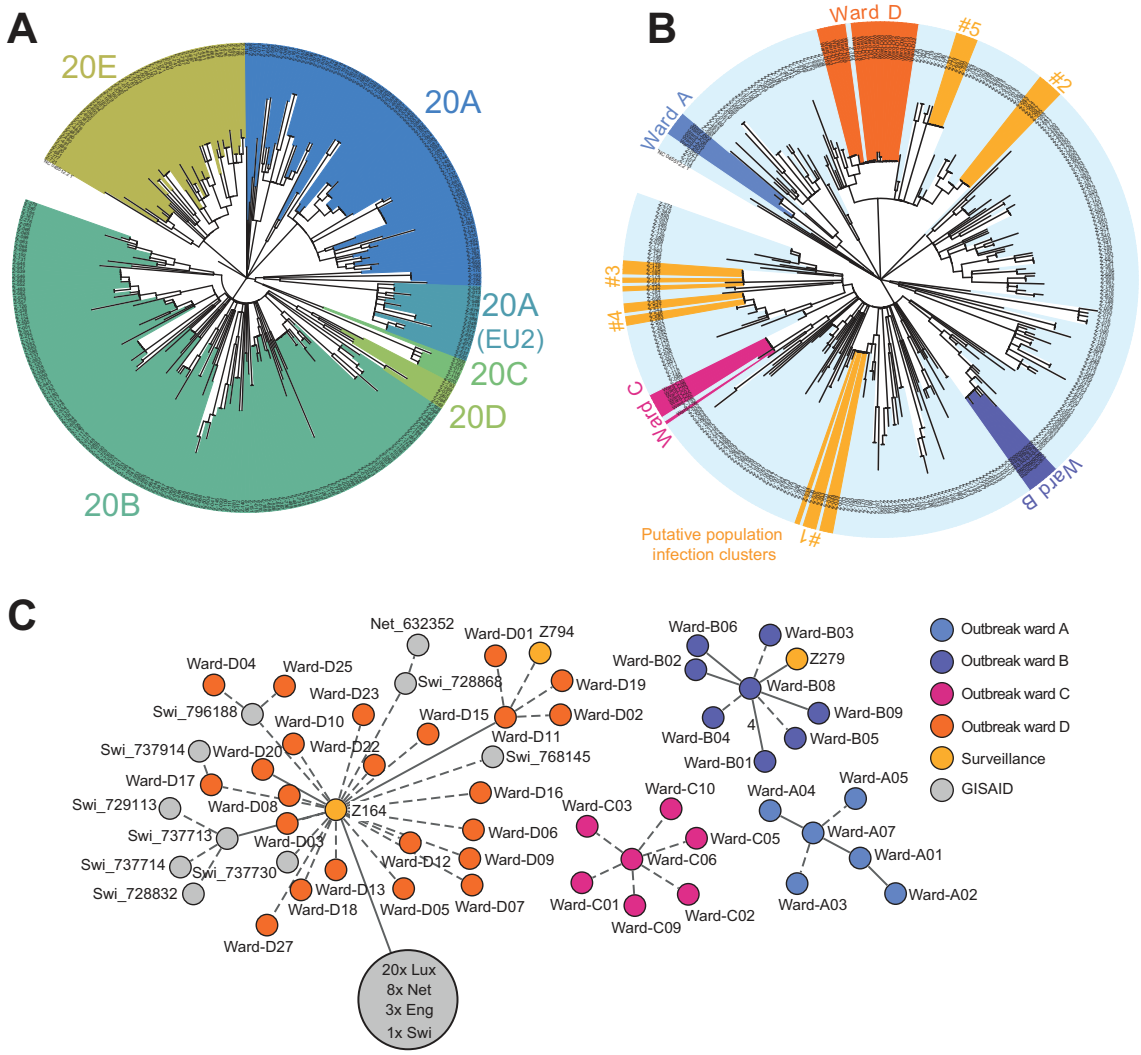


Figure 3. *A*, Phylogenetic tree of the 320 surveillance samples collected during this study; colours are assigned according to the NextStrain [21] clade system. *B*, Joint phylogenetic tree of 44 samples from 4 hospital outbreaks (Ward *A*–*D*) and 320 surveillance samples. For a description of the outbreaks, see main text. Putative population infection clusters are highlighted in yellow (1–5). Gaps in the corresponding shaded areas correspond to related samples not identified by the greedy clustering algorithm (see Methods). Tree visualization based on iTol [22]. *C*, Minimum spanning tree (calculated with the Python library networkx version 2.5; visualized with Cytoscape version 3.8.2 and Inkscape version 0.92) visualization of the 4 hospital outbreaks, including all identical or near-identical (distance = 0 or distance = 1) from GISAID and the surveillance sequencing cohort. Samples from GISAID are labelled with their country of origin (Lux = Luxemburg; Net = Netherlands; Swi = Switzerland; Eng = England). The large gray circle represents a cluster of identical and near-identical GISAID samples. Solid lines without number indicate distance = 1 and dashed lines indicate distance = 0 between samples. Abbreviation: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

one ambiguous character (average: 0.48 ambiguous characters / genome), indicating potential intra-patient strain variability.

The development of viral population structure was largely consistent with the developments in Europe at large; for example, although clade 20E was initially found at low frequencies, it accounted for nearly half of the sequenced genomes towards the end of the sampling period (Figure 2C, Figure 3A). One notable exception was the absence of clade 20I/501Y.V1 (equivalent to B.1.1.7/"Alpha" in Pangolin nomenclature; https://github.com/cov-lineages/pangolin), which had already reached significant frequencies in some European countries by the end of December 2020 (eg, 48% in the UK or 15% in the Netherlands); the viral variants B.1.351/"Beta," first identified in South Africa, and P.1/"Gamma" were also not detected. A comparison between the locally generated data and 385 109 non-Düsseldorf samples from GISAID [23] showed that many of the detected isolate genomes were not yet represented in global databases (see Supplementary Table 4 and Supplementary Text 1 for details).

SARS-CoV-2 Dashboard

To enable the independent interrogation of viral sequencing data by all involved stakeholders, we developed a user-friendly web application ("Dashboard"; see Methods; https://covgen.

hhu.de and https://covgen.hhu.de/paper for the version used here). The dashboard was continuously updated with sequenced population and outbreak isolate genomes and enabled the targeted development of hypotheses about the genetic structure of outbreaks or transmission chains by hospital hygiene staff or local health authorities. Inter-sample genetic relatedness was visualized using a minimum spanning tree (MST), and the interface supported various filtering and visualization options, for example, enabling the targeted display of outbreaks and genetically related samples from the local viral population (Supplementary Figure 2). During the sampling period, the dashboard was a key tool to assist the interrogation of the genetic structure of the sequenced viral genomes.

Hospital Outbreak Analysis

We used the developed integrated system to characterize the genetic structure of 4 SARS-CoV-2 hospital outbreaks at Düsseldorf University Hospital and to search for putative transmission chains connecting the local population to the hospital. For 3 of 4 characterized outbreaks (Wards A, B, and C), sequencing (Supplementary Tables 2 and 3) confirmed the clonal structure of the outbreaks (Figure 3B, Figure 3C) within 15 to 32 days (Supplementary Table 3, Supplementary Text 2), but no links between the outbreaks and the local viral population were identified. Infection control and prevention measures that were put into place included staff re-training, improved room ventilation, and upgrades to patient protective equipment; a detailed description of the outbreaks and improved hospital infection control and prevention measures is given in Supplementary Text 2.

For 1 outbreak (Ward D) that affected 16 patients and 13 healthcare workers in mid-December, sequencing confirmed the clonal nature of the outbreak and identified 2 potential links between the outbreak and the local population (Figure 3C); first, analyzed sequencing data were available within 26 days after the detection of the outbreak. For a detailed analysis of the outbreak samples in the context of the Düsseldorf surveillance data, we split the outbreak samples into 2 subgroups. Subgroup-01 represented the majority viral type of the outbreak, and all samples within Subgroup-01 were genetically identical; Subgroup-02 represented samples with a distance (see Supplementary Text 1) of 1 single-nucleotide polymorphism (SNP) to Subgroup-01. An analysis of the surveillance data showed that the viral type of Subgroup-01 was present in the local population as early as 1 October (sample Z164; distance to majority of outbreak samples = 0) and that the viral type of Subgroup-02 was detected again in the surveillance data, although the outbreak was ongoing (sample Z794, sampled on 16 December). Contact tracing data collected by Düsseldorf public health authorities showed that a family member of the individual that sample Z164 was taken from was treated at another ward of the clinical department of Düsseldorf University Hospital in October 2020, establishing a possible link between the surveillance samples and the hospital outbreak on ward "D"; a SARS-CoV-2 antibody test of this family member was weakly positive for immunoglobulin A (IgA). Contact tracing data also showed that a close relative of the individual who provided sample Z794 was treated at a coronavirus disease 2019 (COVID-19) ward of Düsseldorf University Hospital, to which SARS-CoV-2-positive cases from ward "D" had been transferred to. Interestingly, 5 highly related isolates, identical (distance = 0) to samples part of the ward "D" outbreak, were also identified in Switzerland (GISAID IDs 796188, 728868, 737730 768145, 737914; sampled between 9 November and 28 December). Near-identical samples (distance = 1) were also found in the Netherlands and the United Kingdom (Supplementary Table 4). Implemented infection control and prevention measures included comprehensive screening, a temporary stop of admissions to the ward, and re-education of staff in nonpharmaceutical interventions. Further details on the Ward D outbreak are presented in Supplementary Text 2.

Genetically Based Identification of Population Infection Clusters

To evaluate whether untargeted surveillance sequencing data could enable the genetically based identification of infection clusters and transmission chains in the population during community transmission, we applied a simple greedy clustering algorithm (Methods) to the viral sequence surveillance data, identifying groups of genetically identical samples. This analysis identified 5 putative infection clusters within the Düsseldorf samples (Figure 3B; Supplementary Table 5). Routine public health authority contact tracing data were subsequently integrated and showed that the identified clusters reflected epidemiologically relevant associations.

PopClust#1 consisted of 7 patient samples collected in late August; contact tracing data revealed that this cluster corresponded to a known transmission event during a school excursion in August 2020.

PopClust#2 consisted of 6 patient samples collected in mid-/late September; of these, 3 were linked via a care home, 2 were collected from members of the same family, and the remaining sample was linked to an otherwise unrelated primary school outbreak.

PopClust#4 and PopClust#5, consisting of 4 and 5 patient samples collected in November and December, respectively, contained samples that were linked via joint household membership, as well as samples without any obvious connections to the identified households.

Finally, investigation of PopClust#3, representing five samples collected in early October, enabled the discovery of a previously unrecognized population transmission chain. For samples Z132 and Z177, a reanalysis of the originally collected

contact tracing data pointed to a connection between the cases involving another positively tested individual X (not part of the study): (Z132, X) were connected via active membership of the same martial arts gym, and (X, Z177) had a close personal relationship. Z132 and X had not identified each other as direct contacts; the identified putative link between Z132 and X was further supported by the timing of the infections and highlights the potential for fomite and/or aerosol transmission in martial arts contexts. Two additional samples of PopClust#3 were members of the same household, and for the remaining sample in PopClust#3, no obvious connection to the other samples was identified.

Sample ID and contact tracing information for the identified clusters are summarized in Supplementary Table 5.

Rapid Nanopore Sequencing Experiment

To investigate whether a rapid viral surveillance sequencing workflow could be implemented, we measured total turnaround time for 24 samples with different cycle threshold (Ct) values (range 17–31) from the surveillance cohort from sample receipt to bioinformatic analysis when using streamlined workflows and processes (Supplementary Text 3). After a total time of 28 hours (11 hours for sample and library preparation, 15 hours for sequencing, and ≤2 hours for bioinformatic analysis; Supplementary Figure 3), 19 of 24 genomes were resolved to high quality (<3000 N positions; Supplementary Figure 4*A*). Increasing the sequencing time by at least 2 hours increased the number of high-quality resolved genomes to 20; the number of resolved bases across all samples saturated after 37 hours of sequencing (Supplementary Figure 4*B*).

DISCUSSION

An improved understanding of SARS-CoV-2 transmission chains is key to effective viral containment. We have shown that an integrated local SARS-CoV-2 genomic epidemiology system implemented in the state capital city of Düsseldorf could enable the retrospective detection of SARS-CoV-2 infection chains through hospitals and the local population during ongoing community transmission. We could confirm the clonal nature of 4 outbreaks in a regional maximum care hospital and contribute to the design and implementation of refined infection control intervention measures, minimizing the risk of nosocomial SARS-CoV-2 transmission (see Supplementary Text 3). We also developed a simple algorithmic approach to identify putative infection clusters in the local population based on genetic data alone and found 5 such clusters in the generated surveillance data from August to December 2020. Integration with contact tracing data showed that the untargeted sequencing data captured epidemiologically relevant viral transmissions in settings of societal importance, such as care homes, schools, and recreational physical activities. Intriguingly, we found 2 potential links between the local population and a hospital outbreak and identified a previously unrecognized population transmission chain in a martial arts gym, confirming the potential utility of untargeted sequencing for identification of transmission chains in the population during ongoing community transmission. Key features of our approach include the joint analysis of population and outbreak sequencing data, the integration of genetic data with contact tracing data, and the availability of a user-friendly dashboard as a central data analysis platform for all participating stakeholders.

When applied at scale, fully integrated local genomic epidemiology systems could contribute to a more effective management of the SARS-CoV-2 pandemic on multiple levels. First, significant uncertainties remain with respect to the relative importance of viral transmission in various settings of societal importance, such as restaurants, public transport, childcare facilities, or schools. Routine large-scale untargeted surveillance sequencing could contribute to a more quantitative understanding of transmission risks in such settings and thus enable the design of improved infection prevention measures. Second, as shown here, genomic epidemiology can enable the identification of large infection clusters and superspreading events in the population, which play an important role in driving the pandemic [24]. Third, genetically characterizing a large proportion of local cases may enable locally adapted strategies for containing the spread of variants of concern [25]. Fourth, genomic epidemiology could increase accountability for the prevention and management of infection chains at the individual or organizational level, as sequencing can provide supporting evidence as well as evidence against individual transmissions or transmission contexts suggested by classical epidemiological approaches. A simple model calculation (see Supplementary Table 6 for an interactive Excel sheet) shows that a strategy of sequencing all positive cases would increase the total cost of the testing system by only 30–40%. Although not insignificant, these additional costs could potentially be offset by the economic benefits of improved pandemic management; as the current costs of lockdown-like measures in many countries are significant (eg, estimated at EUR 25–75 billion per week for Germany; see Florian et al [26]), even a small improvement could translate into overall cost-effectiveness.

Important remaining challenges include the logistics of achieving turnaround times compatible with public health authority decision making [27], as well as the downstream integration of viral sequencing data. As we and others have demonstrated, the Nanopore technology can enable turnaround times as low as 28 hours for SARS-CoV-2 [17, 28, 29]. Further improvements may be possible in the fields of sample logistics, real-time control of sequencing runs (eg, with RAMPART; https://artic.network/rampart), and streaming data analysis. The developed SARS-CoV-2 dashboard with automatic detection of infection clusters could be a first step toward a more routine

integration of viral sequencing into the processes of local public health authorities; further work, however, is required, for example, for integration with classical contact tracing software. In addition, the simple greedy algorithm used here to identify candidate population infection clusters could be improved by population genetics modeling [30, 31], as well as by the incorporation of additional dimensions such as sampling time.

Limitations of this study include the utilized convenience sampling scheme, potentially associated with a biased selection of samples; the relatively low proportion of sequenced positive cases over the sampling period, potentially limiting generalizability; the retrospective nature of the study, limiting the potential actionability of the generated surveillance data from the perspective of public health; and the fact that intervention based on genetic data was limited to the investigated hospital outbreaks.

Based on the methods and results developed here, these limitations are currently addressed in a follow-up study started in Spring 2021. Representing a natural extension of the work presented here, the follow-up study will enable a further investigation of the full potential of "real-time" genomic surveillance for supporting public health decision making, based on an improved characterization of pathogen transmission chains in the population at large. Preliminary results (n = 4260) of the study show that case sequencing rates of over 50% can be achieved even at 7-day newly diagnosed case incidence rates >150 per 100 000 population; that "swab-to-sequence" times of <72 hours can be achieved in routine sequencing setups; and that real-time integration of these sequencing data can enable the routine detection of transmission chains in settings such as care homes, primary schools and families, consistent with the results presented here. Although a full analysis will be presented after the conclusion of the follow-up study in late 2021, a continuously updated view of local viral population structure and putative infection clusters is provided via the publicly available dashboard web application presented here (https://covgen.hhu.de).

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Author Contributions. A. W., T. H., P. F., and L. E. contributed to the design of the study and contributed to the data collection, data analysis, data interpretation, literature review, and contributed to the writing of the manuscript. A. T. and J. N. contributed to the data collection, data analysis, data interpretation, figures, and contributed to the writing of the manuscript. M. D., D. S., A. D., T. W., T. T., D. K. M., L. H., K. H., N. L., S. H., M. A., M. P. D., A. T., and S. K.-B. contributed to the data collection, data analysis, and data interpretation. J. T., K. P., R. Z., and K. G. contributed to the design and supervision of the study and to the writing of the manuscript. A. T. D. contributed to the design and the supervision of the study, to data collection, data analysis, data interpretation, literature review, and wrote the first draft of the manuscript. All authors reviewed and approved the final manuscript.

Deutsche COVID-19 Omics Initiative (DeCOI)

Janine Altmüller, Angel Angelov, Anna C. Aschenbrenner, Robert Bals, Alexander Bartholomäus, Anke Becker, Daniela Bezdan, Michael Bitzer, Helmut Blum, Ezio Bonifacio, Peer Bork, Nicolas Casadei, Thomas Clavel, Maria Colome-Tatche, Inti Alberto De La Rosa Velázquez, Andreas Diefenbach, Alexander Dilthey, Nicole Fischer, Konrad Förstner, Sören Franzenburg, Julia-Stefanie Frick, Gisela Gabernet, Julien Gagneur, Tina Ganzenmüller, Marie Gauder, Alexander Goesmann, Siri Göpel, Adam Grundhoff, Hajo Grundmann, Torsten Hain, André Heimbach, Michael Hummel, Thomas Iftner, Angelika Iftner, Stefan Janssen, Jörn Kalinowski, René Kallies, Birte Kehr, Andreas Keller, Oliver Keppler, Sarah Kim-Hellmuth, Christoph Klein, Michael Knop, Oliver Kohlbacher, Karl Köhrer, Jan Korbel, Peter G. Kremsner, Denise Kühnert, Ingo Kurth, Markus Landthaler, Yang Li, Kerstin Ludwig, Oliwia Makarewicz, Manja Marz, Alice McHardy, Christian Mertes, Maximilian Münchhoff, Sven Nahnsen, Markus Nöthen, Francine Ntoumi, Peter Nürnberg, Uwe Ohler, Stephan Ossowski, Jörg Overmann, Silke Peter, Klaus Pfeffer, Anna R. Poetsch, Ulrike Protzer, Alfred Pühler, Nikolaus Rajewsky, Markus Ralser, Olaf Rieß, Stephan Ripke, Ulisses Rocha, Philip Rosenstiel, Emmanuel Saliba, Leif Erik Sander, Birgit Sawitzki, Simone Scheithauer, Philipp Schiffer, Jonathan Schmid-Burgk, Wulf Schneider, Eva-Christina Schulte, Joachim Schultze , Alexander Sczyrba, Mariam L. Sharaf, Yogesh Singh , Michael Sonnabend, Oliver Stegle, Jens Stoye, Fabian Theis, Janne Vehreschild, Thirumalaisamy P. Velavan, Jörg Vogel, Max von Kleist, Andreas Walker, Jörn Walter, Dagmar Wieczorek, Sylke Winkler, John Ziebuhr.

Acknowledgments. The authors acknowledge Jeff Barrett for discussions and for introducing the notion of clonality / heterozygosity as a useful summary statistic of surveillance sequencing data. They would like to thank the hospital hygiene staff and the employee health department medical officers for their support.

The authors gratefully acknowledge the authors, the originating and submitting laboratories for their sequence and metadata shared through GISAID. All submitters of data may be contacted directly via GISAID. The Acknowledgments Table for GISAID is part of Supplementary Table 7. Computational support and infrastructure were provided by the "Centre for Information and Media Technology" (ZIM) at the University of Düsseldorf (Germany). Sequencing and the design of sequencing workflows were supported by the Biologisch-Medizinisches Forschungszentrum der Heinrich Heine University Düsseldorf (BMFZ). Viral genome sequences are available on GISAID. Accessions are listed in Supplementary Table 3. All source code for the dashboard is available under the MIT license and can be downloaded from https://github.com/DiltheyLab/SARS-CoV2-Dashboard-Releases. The version of the dashboard and the data for the analyses presented here are archived under https://covgen.hhu.de/paper. This study was approved by the ethics committee of the Medical Faculty of Heinrich Heine University Düsseldorf (#2020–839).

Financial support. Jürgen Manchot Foundation German Research Foundation award number 428994620 German Federal Ministry of Education and Research award numbers 031L0184B and B-FAST (grant number 01KX2021).

Potential conflicts of interests. The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References

1. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature **2020**; 579:265–9.
2. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature **2020**; 579:270–3.
3. World Health Organization. WHO COVID-19 Weekly Epidemiological Update (11 April 2021). Available at: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---13-april-2021. Accessed 26 January 2021.

4. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet **2018**; 19:9–20.
5. Grubaugh ND, Ladner JT, Lemey P, et al. Tracking virus outbreaks in the twenty-first century. Nat Microbiol **2019**; 4:10–9.
6. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature **2016**; 530:228–32.
7. Grubaugh ND, Ladner JT, Kraemer MUG, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature **2017**; 546:401–5.
8. Gonzalez-Reiche AS, Hernandez MM, Sullivan M, et al. Introductions and early spread of SARS-CoV-2 in the New York city area. medRxiv **2020**: 2020.04.08.20056929.
9. Lu J, du Plessis L, Liu Z, et al. Genomic epidemiology of SARS-CoV-2 in Guangdong province, China. Cell **2020**; 181:997–1003 e9.
10. Popa A, Genger JW, Nicholson MD, et al. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. Sci Transl Med **2020**; 12:eabe2555.
11. MacFadden DR, McGeer A, Athey T, et al. Use of genome sequencing to define institutional influenza outbreaks, Toronto, Ontario, Canada, 2014–15. Emerg Infect Dis **2018**; 24:492–7.
12. COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe **2020**; 1:e99–e100.
13. Geoghegan JL, Ren X, Storey M, et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. Nat Commun **2020**; 11:6351.
14. Pattabiraman C, Habib F, P K H, et al. Genomic epidemiology reveals multiple introductions and spread of SARS-CoV-2 in the Indian state of Karnataka. PLoS One **2020**; 15:e0243412.
15. Oude Munnink BB, Nieuwenhuijse DF, Stein M, et al; Dutch-Covid-19 response team. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat Med **2020**; 26:1405–10.
16. Gudbjartsson DF, Stefansson K. Early spread of SARS-CoV-2 in the Icelandic population. Reply. N Engl J Med **2020**; 383:2184–5.
17. Meredith LW, Hamilton WL, Warne B, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. Lancet Infect Dis **2020**; 20:1263–71.
18. Quick J. ARTIC amplicon sequencing protocol for MinION for nCoV-2019. Available at: https://dx.doi.org/10.17504/protocols.io.bbmuik6w. Accessed 20 June 2021.
19. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc **2017**; 12:1261–76.
20. Tyson JR, James P, Stoddart D, et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. bioRxiv **2020**. doi:10.1101/2020.09.04.283077
21. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics **2018**; 34:4121–3.
22. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res **2016**; 44:W242–5.
23. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data: from vision to reality. Euro Surveill **2017**; 22:30494.
24. Laxminarayan R, Wahl B, Dudala SR, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. Science **2020**; 370:691–7.
25. Rambaut A, Loman N, Pybus O, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Available at: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563. Accessed 20 June 2021.
26. Florian D, Clemens F, Marcell G, et al. Die volkswirtschaftlichen Kosten des corona-shutdown für Deutschland: eine szenarienrechnung. ifo Schnelldienst **2020**; 73:29–35.
27. Köser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med **2012**; 366:2267–75.
28. Greninger AL, Naccache SN, Federman S, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med **2015**; 7:99.
29. Quick J, Ashton P, Calus S, et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. Genome Biol **2015**; 16:114.
30. Bouckaert R, Heled J, Kühnert D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol **2014**; 10:e1003537.
31. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol **2005**; 22:1185–92.

Integrated genomic surveillance enables tracing of person-to-person SARS-CoV-2 transmission chains during community transmission and reveals extensive onward transmission of travel-imported infections, Germany, June to July 2021

Torsten Houwaart[1,*], Samir Belhaj[2,*], Emran Tawalbeh[2,*], Dirk Nagels[2], Yara Fröhlich[3], Patrick Finzer[1,4], Pilar Ciruela[5,6], Aurora Sabrià[5], Mercè Herrero[5], Cristina Andrés[7], Andrés Antón[7], Assia Benmoumene[3], Dounia Asskali[3], Hussein Haidar[3], Janina von Dahlen[3], Jessica Nicolai[1], Mygg Stiller[8], Jacqueline Blum[8], Christian Lange[8], Carla Adelmann[9], Britta Schroer[9], Ute Osmers[10], Christiane Grice10, Phillipp P. Kirfel[10], Hassan Jomaa[10], Daniel Strelow[1], Lisanna Hülse[1], Moritz Pigulla[2], Pascal Kreuzer[2], Alona Tyshaieva[1], Jonas Weber[1], Tobias Wienemann[1], Malte Kohns Vasconcelos[1], Katrin Hoffmann[4], Nadine Lübke[3], Sandra Hauka[3], Marcel Andree[3], Claus Jürgen Scholz[11], Nathalie Jazmati[11], Klaus Göbels[2], Rainer Zotz[4,12], Klaus Pfeffer[1], Jörg Timm[3], Lutz Ehlkes[2,], Andreas Walker[3,**], Alexander T. Dilthey[1,**], German COVID-19 OMICS Initiative (DeCOI)[13]**

1. Institute of Medical Microbiology and Hospital Hygiene, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
2. Düsseldorf Health Authority (Gesundheitsamt Düsseldorf), Düsseldorf, Germany
3. Institute of Virology, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
4. Zotz | Klimas, Düsseldorf, Germany
5. Sub-Directorate General of Surveillance and Response to Public Health Emergencies, Public Health Agency of Catalonia, Barcelona, Spain
6. CIBER Epidemiologia y Salud Pública (CIBERESP), Instituto Salud Carlos III, Madrid, Spain
7. Microbiology Unit, Vall d'Hebron University Hospital, Barcelona, Spain
8. Medizinische Laboratorien Düsseldorf, Düsseldorf, Germany
9. Solingen Health Authority (Gesundheitsamt Solingen), Solingen, Germany
10. MVZ SYNLAB Leverkusen GmbH, Leverkusen, Germany
11. Labor Dr. Wisplinghoff, Cologne, Germany
12. Department of Hemostasis and Transfusion Medicine, Heinrich Heine University Medical Center, Düsseldorf, Germany
13. Members of German COVID-19 OMICS Initiative (DeCOI) are listed under Collaborators.

* These authors both contributed equally and share first authorship.
** These authors contributed equally and share last authorship.

Correspondence: Alexander T. Dilthey (dilthey@hhu.de)

Collaborators: The collaborators are listed at the end of the article.

Citation style for this article:
Houwaart Torsten, Belhaj Samir, Tawalbeh Emran, Nagels Dirk, Fröhlich Yara, Finzer Patrick, Ciruela Pilar, Sabrià Aurora, Herrero Mercè, Andrés Cristina, Antón Andrés, Benmoumene Assia, Asskali Dounia, Haidar Hussein, von Dahlen Janina, Nicolai Jessica, Stiller Mygg, Blum Jacqueline, Lange Christian, Adelmann Carla, Schroer Britta, Osmers Ute, Grice Christiane, Kirfel Phillipp P., Jomaa Hassan, Strelow Daniel, Hülse Lisanna, Pigulla Moritz, Kreuzer Pascal, Tyshaieva Alona, Weber Jonas, Wienemann Tobias, Kohns Vasconcelos Malte, Hoffmann Katrin, Lübke Nadine, Hauka Sandra, Andree Marcel, Scholz Claus Jürgen, Jazmati Nathalie, Göbels Klaus, Zotz Rainer, Pfeffer Klaus, Timm Jörg, Ehlkes Lutz, Walker Andreas, Dilthey Alexander T., German COVID-19 OMICS Initiative (DeCOI). Integrated genomic surveillance enables tracing of person-to-person SARS-CoV-2 transmission chains during community transmission and reveals extensive onward transmission of travel-imported infections, Germany, June to July 2021. Euro Surveill. 2022;27(43):pii=2101089. https://doi.org/10.2807/1560-7917.ES.2022.27.43.2101089

Article submitted on 19 Nov 2021 / accepted on 16 Sept 2022 / published on 27 Oct 2022

Background: Tracking person-to-person SARS-CoV-2 transmission in the population is important to understand the epidemiology of community transmission and may contribute to the containment of SARS-CoV-2. Neither contact tracing nor genomic surveillance alone, however, are typically sufficient to achieve this objective. **Aim:** We demonstrate the successful application of the integrated genomic surveillance (IGS) system of the German city of Düsseldorf for tracing SARS-CoV-2 transmission chains in the population as well as detecting and investigating travel-associated SARS-CoV-2 infection clusters. **Methods:** Genomic surveillance, phylogenetic analysis, and structured case interviews were integrated to elucidate two genetically defined clusters of SARS-CoV-2 isolates detected by IGS in Düsseldorf in July 2021. **Results:** Cluster 1 (n = 67 Düsseldorf cases) and Cluster 2 (n = 36) were detected in a surveillance dataset of 518 high-quality SARS-CoV-2 genomes from Düsseldorf (53% of total cases, sampled mid-June to July 2021). Cluster 1 could be traced back to a complex pattern of transmission in nightlife venues following a putative importation by a SARS-CoV-2-infected return traveller (IP) in late June; 28 SARS-CoV-2 cases could be epidemiologically directly linked to IP. Supported by viral genome data from Spain, Cluster 2 was shown to represent multiple independent introduction events of a viral strain circulating in Catalonia and other European countries, followed by diffuse community transmission in Düsseldorf.

Conclusion: IGS enabled high-resolution tracing of SARS-CoV-2 transmission in an internationally connected city during community transmission and provided infection chain-level evidence of the downstream propagation of travel-imported SARS-CoV-2 cases.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a worldwide pandemic with › 593 million cases and › 6.4 million associated deaths up to August 2022 [1]. SARS-CoV-2 vaccines have greatly contributed to reductions in coronavirus disease (COVID-19)-associated morbidity and mortality in many countries; however, non-pharmaceutical interventions (NPIs) to limit viral spread and reduce the healthcare burden of SARS-CoV-2 remain important in many contexts. Such contexts include instances of low vaccine availability or high rates of vaccine hesitancy in some countries, the potential for vaccine breakthrough infections and, more generally, the emergence of novel viral variants.

As the aim of NPIs is to interrupt or prevent pathogen transmission chains, a comprehensive understanding of these transmission chains in the population – who infected whom, and in which epidemiological context – could be greatly beneficial. Contact tracing regimes, which typically employ structured case interviews and which are operated by many countries, are an important data source on pathogen transmission in the population. Contact tracing is generally recognised as an important element of SARS-CoV-2 mitigation strategies [2-5]. However, the ability of classical contact tracing regimes to reliably track transmission chains in the population is limited and a substantial number of infections typically remain unexplained. For example, in the German city of Düsseldorf, an international economic and air travel hub of ca 600,000 inhabitants, ca 45% of SARS-CoV-2 infections remained unexplained in 2021 (Düsseldorf Health Department internal data), despite the operation of a well-staffed and comprehensive contact tracing effort. Similar numbers have been reported from other localities [6]. Genomic surveillance, another potential data source on the structure of population transmission chains, has also emerged as an important element of SARS-CoV-2 mitigation strategies [7,8]. However, due to the relatively low mutation rate of SARS-CoV-2 [9] and the fact that many genomic surveillance systems only sample a limited proportion of total cases, genomic surveillance by itself is typically not sufficient to enable reconstruction of transmission chains at the person-to-person level in the population outside of confined outbreak scenarios.

Integrated genomic surveillance (IGS) is an emerging approach that refers to the integrated analysis of genetic and complementary epidemiological data. As we and others have shown [10-13], IGS can contribute to identification of otherwise unrecognised SARS-CoV-2 transmission chains in the general population even under conditions of high-incidence community transmission and thus provide important complementary information for the design and implementation of NPIs.

Here we use the IGS system of Düsseldorf (IGSD) to investigate person-to-person transmission chains in this city in late June and July 2021.

Methods

Integrated genomic surveillance in Düsseldorf

The IGSD has been described elsewhere [10]. Briefly, when fully operational, the system operates as follows. First, a large proportion of SARS-CoV-2 from local cases is rapidly sequenced. Viral genomes (Z* samples) are primarily generated by the Centre for Medical Microbiology, Hospital Hygiene, and Virology of Heinrich Heine University Düsseldorf as part of a dedicated local sequencing effort. Viral genome sequences of local cases generated under the national German SARS-CoV-2 surveillance programme by a collaborating large diagnostic laboratory (N* samples) are also integrated. In 2021, the achieved sequencing rate typically varied between 40 and 60% of new cases on a weekly basis; the 'routine Düsseldorf surveillance dataset' described below consists of sequence data obtained as described here (Z* and N* samples).

Second, putative infection clusters are identified with a search algorithm for groups of pairwise-identical samples ('cliques').

Third, the generated sequencing data and identified putative infection clusters are displayed in a visual form ('dashboard'; available at https://covgen.hhu.de); this visualisation is continuously-updated. This system is used as the main information exchange mechanism with the Düsseldorf Health Authority.

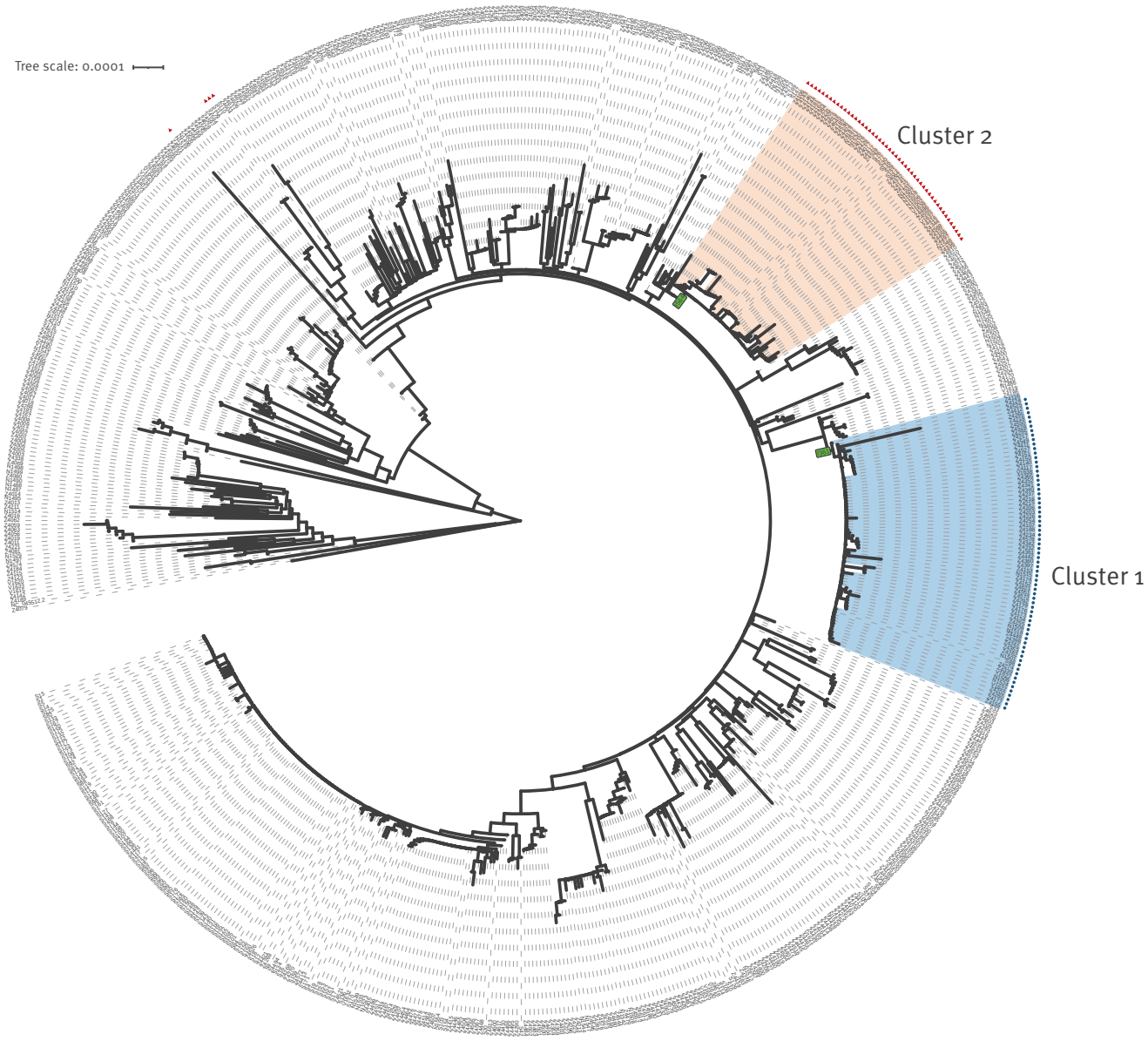
Fourth, the identified putative infection clusters are investigated at the Düsseldorf Health Authority. The investigation combines (i) routine data collected as part of Düsseldorf Health Authority's contact tracing activities, including on symptom onset, travel history and contact persons and (ii) information obtained from structured case interviews ('deep backward contact tracing'; see below) to elucidate potential case connections not captured by standard contact tracing.

Of note, the IGSD does not comprise the routine collection of clinical metadata, and case severity is not used as a sample selection criterion.

To investigate the applicability of the developed system beyond Düsseldorf, a trial run of the IGS system was carried out in the nearby smaller city of Solingen in July and August 2021; the Solingen data were processed and analysed separately from the Düsseldorf data and only integrated with the Düsseldorf data during the phylogenetic cluster refinement analysis (see below).

Figure 1

Phylogenetic analysis of the extended dataset of sequences of SARS-CoV-2 registered cases, Düsseldorf and Solingen, Germany, 15 June–01 August 2021 (n = 699 sequences)[a]



SARS-CoV-2: severe acute respiratory coronavirus 2.

[a] The phylogenetic tree, generated with Geneious (see Methods) and visualised with iTol [19], shows 699 sequences. Nine sequences were found to exhibit low-quality alignments in the multiple sequence alignment of the 708 input sequences by manual inspection and were removed from the alignment before construction of the tree.

Cluster 1 and Cluster 2 are highlighted (areas of the tree shaded in blue and red, respectively). The nodes serving as the two clusters' root nodes, I361 and I584, are displayed as little boxes with green background. The presence of T14064C and C18744T, mutations used in the process of defining Cluster 1 and Cluster 2, is indicated by blue circles and red triangles, respectively.

Inter-sample distance metric

Inter-sample genetic distances were calculated as defined previously [10], with one modification (point (iv) below). Briefly, a multiple sequence alignment (MSA) of all sequences was built with MAFFT [14], using GISAID [15] instructions. The distance d(x, y) d(x, y) between two samples *x* and *y* was defined as the number of differences between the MSA entries of xx and yy , (i) ignoring leading or trailing gap characters, (ii) counting matches and mismatches according to International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes, (iii) counting subsequent non-matching gaps columns as a difference of 1, (iv) ignoring deletions aligned to 'N' regions in the other genome, and (v) ignoring any mismatches in the MSA regions between the beginning of the MSA and the 20th ACGT character of either sequence and the end of the MSA and the 20 last ACGT characters of either sequence.

Structured case interviews

A specialised team of interviewers within Düsseldorf Health Authority conducted structured case interviews. These covered (i) occupation and place of work; (ii) utilisation of public transport; (iii) social, household and family contacts; (iv) utilisation of medical services; (v) supermarket and retailer visits; (vi) gastronomy and nightlife; (vii) travel history. Before a case was classified as unavailable, a minimum of three contact attempts were carried out using available landline or mobile phone numbers; participation in the structured case interviews was voluntary.

Phylogenetic cluster refinement analysis

To refine the definition of two large groups of genetically-related-SARS-CoV-2-infected cases, which are referred to as Cluster 1 and Cluster 2, phylogenetic analysis of an 'extended' dataset (see Results section) was carried out using the neighbour-joining method with the Tamura–Nei genetic distance model as implemented in Geneious version 10.2.6 (Figure 1). The samples previously flagged by the IGSD routine cluster analysis algorithms were located in the phylogenetic tree. Once the presence of two large clusters of genetically-related isolates was confirmed, an analysis of the mutational patterns observed downstream of the putative cluster-associated root nodes was carried out. FASTA files and the phylogenetic tree are publicly available (see Data availability).

Cluster 1 strain-of-origin analysis background dataset

To investigate potential origins of the viral strain of Cluster 1, a background dataset was assembled by combining (i) a random sample of non-Cluster 1 Düsseldorf sequences (n = 30); (ii) the set of all SARS-CoV-2 sequences from the Balearic Islands sampled between 15 June and 01 July 2021 available on GISAID (n = 173); (iii) the sequence of EPI_ISL_2710175, a viral genome from the Balearic Islands sampled on 14 June 2021 that was identified using the GISAID Audacity Instant Search [15]; (iv) 61 GISAID sequences related to Cluster 1. This GISAID set was assembled by carrying out a tree neighbourhood search in the GISAID 'Global Phylogeny' tree from August 2021 (GISAID-hCoV-19-phylogeny-2021–08–16; representing 624,052 sequences). Specifically, two Cluster 1 sequences (N1501, N1506) with genetic distance 0 to an individual SARS-CoV-2-infected traveller returning to Düsseldorf from the island of Mallorca (IP), who was retrospectively identified as a likely Cluster 1 index case in the city, were located in the tree. The identities of all leaves with a tree distance (defined as the cumulative length of the edges along the shortest path between two nodes) of ≤ 3/29,903 to either of the two Cluster 1 sequences and sampling date ≤ 15 July 2021 were extracted.

The corresponding viral genome sequences were obtained from the GISAID MSA, and details and acknowledgements are provided in Supplementary Table 1.

Results

Detection and refinement of two large clusters in July 2021

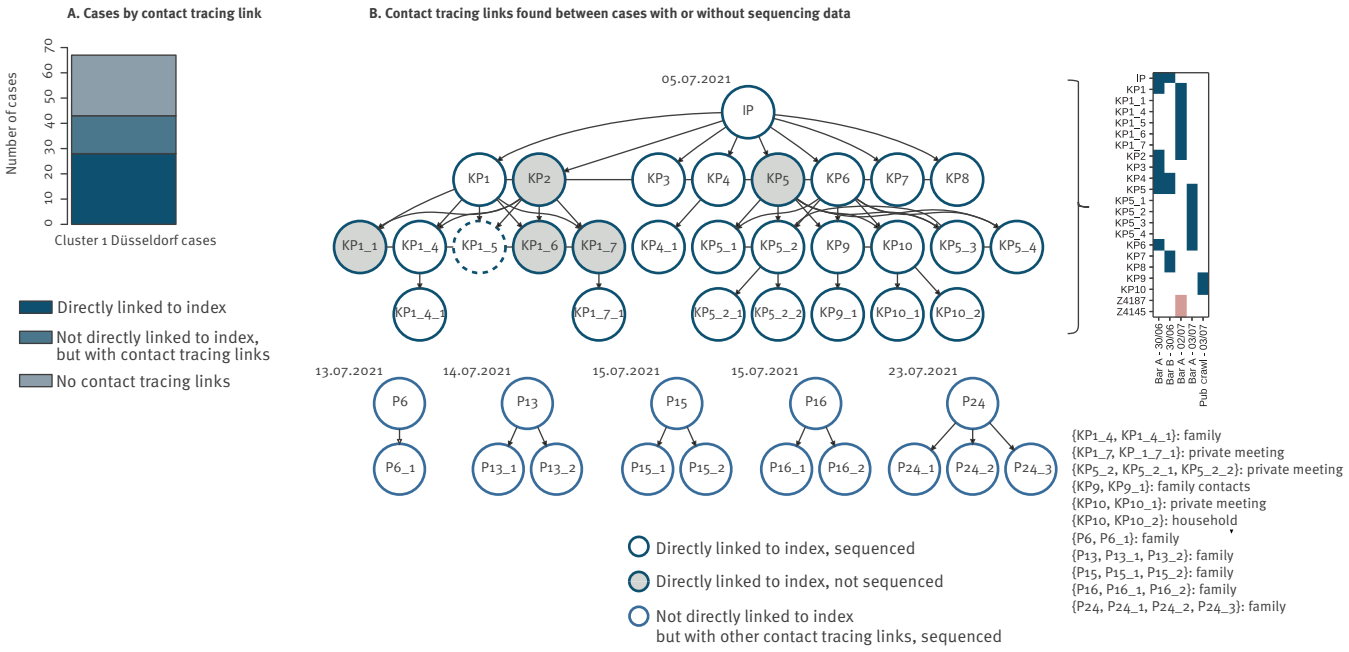
Between 15 June and 01 August 2021, 541 SARS-CoV-2 surveillance genome sequences from Düsseldorf were registered within the IGSD, of which 518 were high-quality sequences (< 5,000 undefined nt; Supplementary Table 2); this set is referred to as the 'routine Düsseldorf surveillance dataset'. Over the same period, 976 new SARS-CoV-2 cases were registered in Düsseldorf, i.e. a high-quality viral genome sequence was available for ca 50% of cases.

In mid-July 2021, the emergence of multiple overlapping putative infection clusters ('cliques' in the pairwise isolate distance matrix, i.e. groups of multiple pairwise-identical viral isolates) was detected by the system's routine cluster analysis algorithms in the routine Düsseldorf surveillance dataset and indicated the presence of two novel large groups of closely related viral isolates (Delta variant; Phylogenetic Assignment of Named Global Outbreak (Pango) lineage designation: B.1.617.2 [16]).

Phylogenetic analysis (see Methods section; Figure 1) was carried out based on an expanded dataset, referred to as the 'extended dataset', with 708 viral genome sequences sampled between 15 June and 01 August 2021 (Supplementary Table 2) that comprised the original Düsseldorf routine surveillance dataset (n = 518); lower-quality Düsseldorf surveillance sequences (n = 23); Düsseldorf University Hospital patient sequences from the same period (n = 27); and available sequences from the nearby city of Solingen, where a trial run of the IGS system took place in July and August 2021 (n = 140). In this analysis, quality thresholds were applied after the construction of the phylogenetic tree; the rationale for including

Figure 2

Cluster 1 contact tracing results, Düsseldorf and Solingen, Germany, 15 June–01 August 2021 (n = 68 cases)



IP: infected return traveller, which is considered as the putative index case.

A. Bar plot showing (i) the number cases who had a direct contact-tracing link to IP (i.e. first-, second- and third-order contacts), (ii) the number of cases with no contact-tracing link to IP but with an epidemiological link to other Cluster-1 cases, who were also not related to IP and (iii) the number of cases with no contact-tracing links to any Cluster 1 cases.

B. Visualisation of the reconstructed epidemiological structure of Cluster 1. Each node in the transmission chain graph represents one case. Nodes shaded in grey represent epidemiologically linked, but non-sequenced cases (see text). Unshaded nodes represent epidemiologically linked and sequenced cases. Of note, the sequence of IP was not used for Figure 1, as this case and respective sequence data were found later in the investigation. While KP1_5 and KP5_2_1 cases had sequence data, these were initially not included in the phylogenetics-based definition of Cluster 1: KP1_5 exhibited an increased genetic distance to the other samples in the cluster (and is therefore shown with a dashed border; see Supplementary Note for a discussion) and KP5_2_1 had›5,000 undefined nt. Upon contact tracing findings, however, these two cases were included as part of Cluster 1. Test dates of assumed index cases, who were epidemiologically unconnected to IP, but for whom contact-tracing suggested that they further transmitted the Cluster-1 strain to other IP-unconnected cases, are respectively shown above the respective nodes of these assumed index cases. The inset to the right of the transmission-chain graph shows the complex patterns of visits of IP and IP's first- and second-order contacts to two bars in the Old Town District of Düsseldorf. The two cases shown with red boxes also visited Bar A, but their sequenced viral isolates group with Cluster 2 in phylogenetic analysis. Cases KP9 and KP10 participated in a pub crawl in the Düsseldorf Old Town area around Bar A. The text inset details the precise nature of the identified case relationships. Transmission chain graphs were plotted with Graphviz [20].

sequences from Solingen was to investigate potential transmission beyond Düsseldorf.

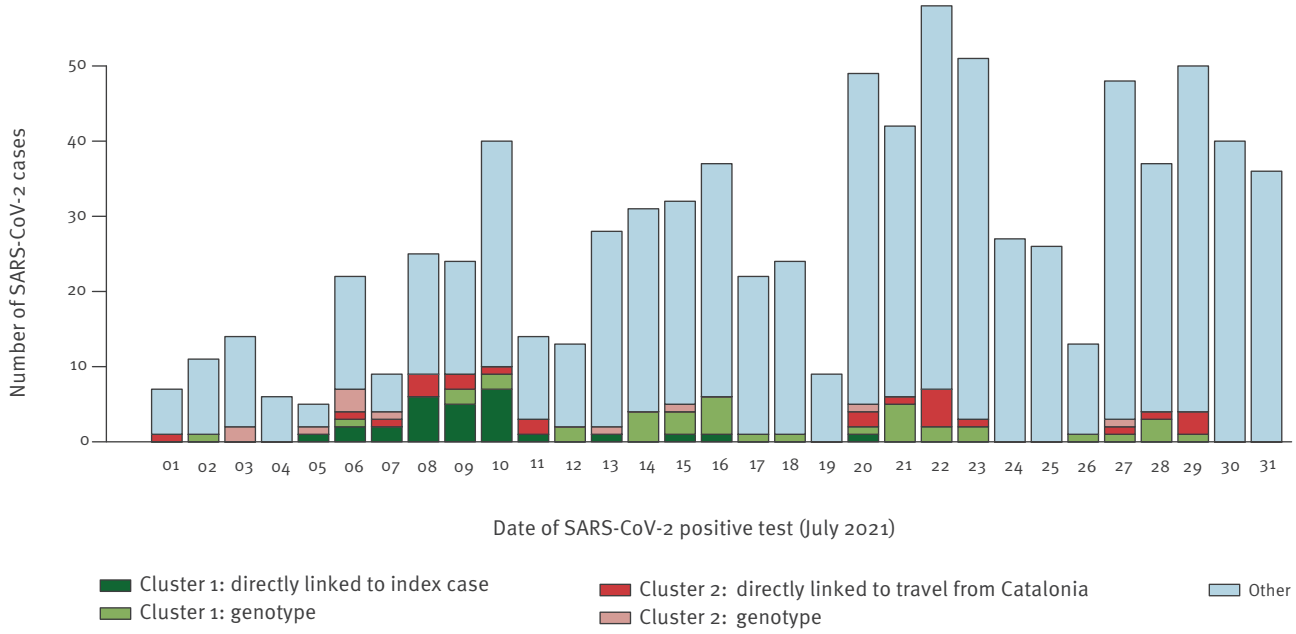
The phylogenetic analysis identified the mutation T14064C (blue circles in Figure 1) as associated with Cluster 1, and C18744T (red triangles) as associated with Cluster 2. The internal nodes I361 and I584 of the constructed phylogenetic tree (see Figure 1 and 'Data availability') were chosen as the root nodes for Cluster 1 and Cluster 2 respectively. The isolate clusters Cluster 1 and Cluster 2 were provisionally defined as the sets of leaf-level descendants of these nodes, including Z4116, a sample carrying an isolated undefined genotype ('N') at position 14064 with distance 0 to other Cluster 1 samples (e.g. IP).

Subsequent to the creation of the phylogenetic tree (Figure 1) with the extended dataset, phylogenetic outliers, repeat samples from the same individual, sequences with›5,000 undefined nt, and non-surveillance Düsseldorf University Hospital sequences sampled after 09 July were removed (see Supplementary Table 3 for a full list of included and removed samples). For Cluster 1 (n = 71 leaf-level descendants of node I361 in the tree), this meant removing 11 sequences (two outlying – including a sequence from a case named KP1_5; four redundant; two with›5,000 undefined nt – including one from a case named KP5_2_1; and three non-surveillance). For Cluster 2 (n = 49 leaf-level descendants of node I548 in the tree), this meant taking away seven sequences (one with›5,000 undefined nt; four redundant; and two non-surveillance).

The phylogenetics-based definition of Cluster 1 comprised 60 viral sequences with an average pairwise genetic distance of 0.91 (59 from Düsseldorf,

Figure 3

Distribution of SARS-CoV-2 cases according to time, their cluster, and epidemiological or genomic basis for inclusion in a cluster, Düsseldorf, Germany, 01 July–31 July 2021 (n = 850 cases)



IP: infected return traveller, which is considered as the index case; SARS-CoV-2: severe acute respiratory coronavirus syndrome coronavirus 2.

The figure shows daily new registered SARS-CoV-2 cases in Düsseldorf in July 2021, and how many of these were associated with Cluster 1 or Cluster 2. For Cluster 1, 'directly linked to index case' refers to uninterrupted, contact tracing-supported putative transmission chains between the linked cases and IP; 'genotype' refers to respective case samples that were identified as belonging to Cluster 1 in the phylogenetic tree analysis (see text). For Cluster 2, 'directly linked to travel from Catalonia' refers to cases who either recently returned from Catalonia and to cases who were directly linked to Catalonia returnees in a manner supported by contact tracing; 'genotype' refers to respective case samples that were identified as belonging to Cluster 2 in the phylogenetic tree analysis (see text).

one from Solingen), sampled from 05 July onwards (Supplementary Table 4); of Cluster 2, 42 viral sequences with an average pairwise genetic distance of 1.89 (36 from Düsseldorf, six from Solingen), sampled from 30 June onwards (Supplementary Table 5).

The FASTA file used for the phylogenetic analysis comprising all analysed sequences as well as the constructed tree in Newick format are publicly available (see 'Data availability'). To further investigate Cluster 1 and Cluster 2, integration of routine contact-tracing data and structured case interviews were carried out (see Methods' section).

Cluster 1 was associated with nightlife spreading events following a putative travel-associated importation

The emergence of Cluster 1 in Düsseldorf could be traced back to multiple nightlife spreading events following putative importation of the Cluster 1-associated strain by IP, an individual SARS-CoV-2-infected traveller returning to Düsseldorf from the island of Mallorca on 28 June (the sequence of IP was not included to compile the phylogenetic tree in Figure 1 as this case was identified during the epidemiological investigation). The identified epidemiological links between Cluster 1 cases are visualised in Figure 2. Transmission of the imported viral strain (Delta variant) in Düsseldorf was likely initiated during encounters between IP and eight first-order contacts (KP1–KP8) in two bars ('Bar A', 'Bar B') in the Old Town District of Düsseldorf, a popular area for nightlife activities with narrow streets and more than 200 bars, on 30 June. Additional transmissions took place in a complex pattern of additional visits of the first-order contacts to Bar A on 02 July (KP2 and KP1 were present in the bar in the same time as KP1_1–KP1_7) and 03 July (Figure 2B Inset) during a likely encounter between the first-order contacts and KP9 and KP10, who were on a pub crawl in the area around Bar A on 03 July (where KP5 and KP6 were present, as well as KP5_1–KP5_4); and from the second-order contacts into the local population via private meetings, family and household contacts (Figure 2B). Contact tracing and structured interviews also uncovered links between an additional 15 cases without direct links to IP (Figure 2B); these likely represented ongoing community transmission of the introduced viral strain or secondary introduction events (see below). Apart from IP, the other cases had no recorded travel histories.

Of note, looking into a potential link between IP and Cluster 1 begun after it emerged during routine contact

tracing that IP had frequented Düsseldorf Old Town nightlife venues on 30 June; when this link started to be explored, the investigation of Cluster 1 was already under way and had identified the Old Town and 30 June as focal points for Cluster 1-related viral transmission. The positive PCR test of IP was carried out in a laboratory not located in Düsseldorf and therefore not covered by the IGSD; the viral genome of IP (available on GISAID under EPI_ISL_3044996), however, was sequenced under Germany's SARS-CoV-2 national genomic surveillance programme and could be requested by Düsseldorf Health Authority after identification of IP. Analysis of the viral genome sequence of IP confirmed that it was highly related to Cluster 1, carrying the T14064C mutation and exhibiting a genetic distance of 0 to 34 of the 60 Cluster 1 sequences phylogenetically defined (Supplementary Table 4).

The assignment of IP as the likely Cluster 1 index case was based on the reconstructed pattern of likely infection events as well as on the dates of symptom onset (Supplementary Table 4) of IP (01 July) and KP1–8 (04–07 July for all cases but KP1, who reported symptom onset on 02 July). IP's symptom onset on 01 July rendered an infection on 30 June unlikely and favoured Mallorca or the return flight to Düsseldorf as infection contexts. In addition, apart from IP, only KP4 and KP5 were present in both Bar A and Bar B (Bar B was where KP7 and KP8 were likely infected) on 30 June, and KP4 and KP5 reported symptom onset on 04 July, consistent with an infection transmitted by IP on 30 June.

To further investigate potential origins of the viral strain of Cluster 1, we analysed the sequences of Cluster 1 against a background dataset of other contemporaneous sequences from Düsseldorf, the Balearic Islands, and GISAID samples related to Cluster 1 (see Methods section). Consistent with an assumed infection of IP on Mallorca, phylogenetic analysis (Supplementary Figure 1) showed that the sequences of Cluster 1 and a small number of isolates from the Balearic Islands and GISAID formed a distinct cluster. Furthermore, an analysis of genetic distances (Supplementary Table 6) showed that the Cluster 1-related sequences from the Balearic Islands were as closely related (genetic distance 1) to the sequence of IP as any of the GISAID sequences up to a sampling date of 07 July, approximately 1 week after the initiation of Cluster 1 transmission in Düsseldorf. IP-identical viral isolates started appearing in the GISAID dataset with sampling dates from 07 July onwards; the 'originating laboratory' record of the earliest three such isolates, however, indicated a likely sampling location in the area around Düsseldorf and thus a likely connection to Cluster 1. The first IP-identical isolates in the GISAID dataset from another German state were collected from 13 July onwards; these, as well as earlier Cluster 1-related sequences from June and July with genetic distance 1, may reflect wider circulation of Cluster 1-related strains in Europe and highlight the possibility of independent introduction events, as well as, in particular for the GISAID samples collected from mid-July onwards, potential export of the Cluster 1 viral strain from the Düsseldorf area.

Including IP; two individuals (KP2, KP5) who were in the company of KP1, KP3, KP4, and KP6 when they were likely infected by IP; three individuals (KP1_1, KP1_6, KP1_7) who were with KP1_4 and KP1_5 when they were likely infected by KP1 or KP2; KP5_2_1 (the viral genome of whom had more than 5,000 undefined nt and who was therefore removed from the initial results of the phylogenetic analysis); and KP1_5 (the viral genome of whom exhibited an increased genetic distance to the other samples in the cluster; see Supplementary Note), 28 SARS-CoV-2 cases in Düsseldorf could be directly linked to IP (defined as the identification of an uninterrupted, contact tracing-supported putative transmission chain between the linked cases and IP; Figure 2A), with a median serial interval of 3 days. With these cases included, Cluster 1 comprised 67 Düsseldorf cases, or 8% of new SARS-CoV-2 cases registered in Düsseldorf in July (Figure 3), and one Solingen case. Of note, two cases belonging to Cluster 2, Z4187 and Z4145, also visited Bar A on 02 July (without recorded direct contacts to other Cluster 1 cases); despite this potential epidemiological link, the genetic data clearly showed that these belonged to a different infection cluster. For 24 Düsseldorf cases and S88, the only Solingen sample in Cluster 1, no links to other Cluster 1 samples or other putative infection sources were identified.

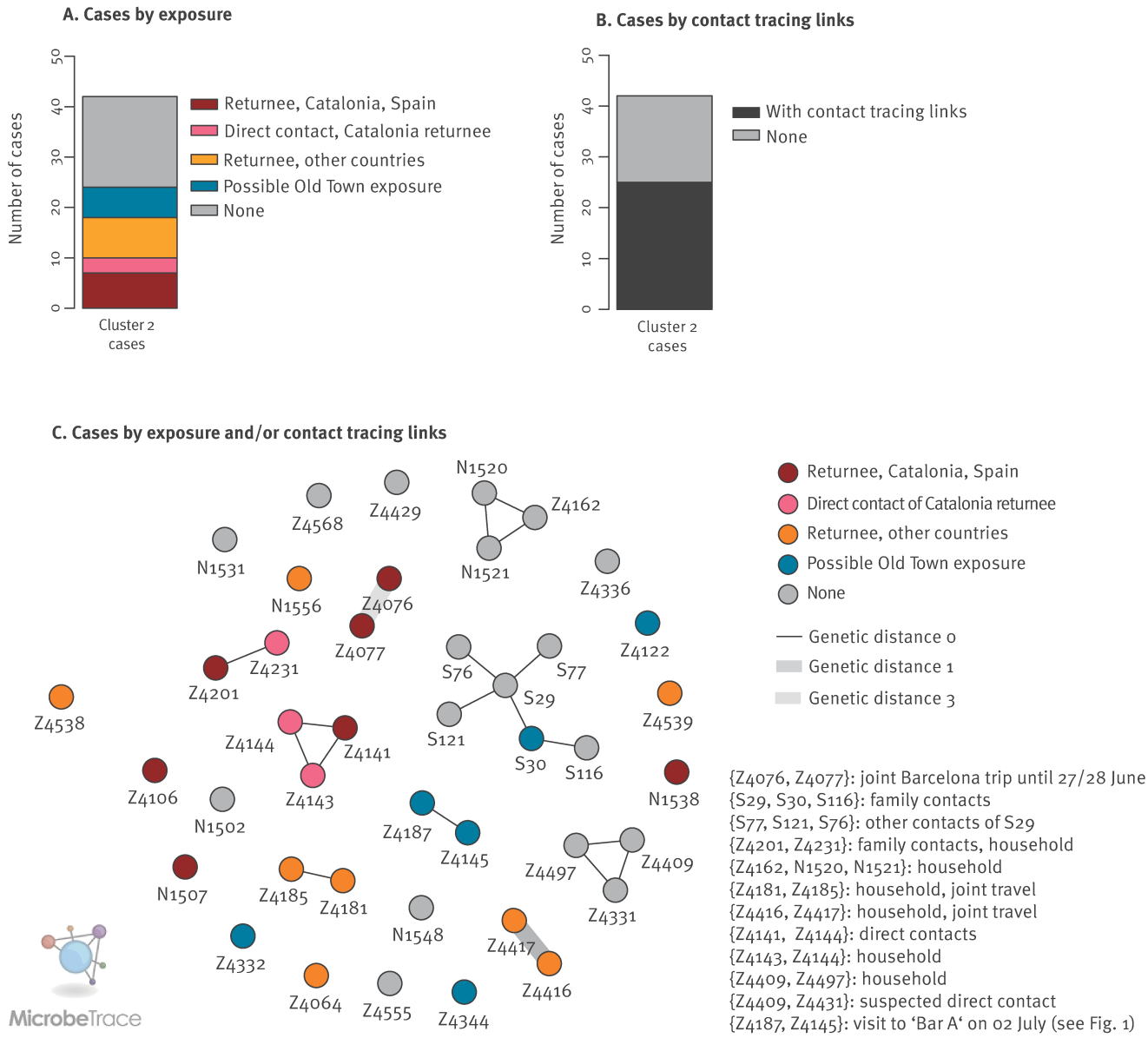
Investigations by the Düsseldorf Health Authority and the discovery of a video posted to social media channels showed limited adherence to mandatory SARS-CoV-2 infection prevention measures in Bars A and B in force at the time, including dancing and non-compliance with indoor masking rules. In addition, the investigations demonstrated insufficient tracking of customer contact details, also in violation of mandatory German pandemic regulations in place at the time.

Cluster 2 represents multiple independent importation events linked to return travel

For Cluster 2, detected from 30 June onwards and comprising 42 cases, no clear index case could be identified. While the integration of contact tracing data and structured case interviews enabled delineating relationships for 25 cases (e.g. household contacts), the overall size of the identified transmission chains was limited compared with Cluster 1 (Figure 4A). Examination of the travel history of the cases, however, showed that almost a quarter of the 42 Cluster 2 cases could be linked to return travel from Catalonia (seven returnees from Catalonia and three associated downstream infections in Düsseldorf); an additional five cases had been travelling to France before testing positive for SARS-CoV-2 (Figure 4B, Supplementary Table 5). Analysis of symptom onset (Supplementary Table 5) in relation to return travel dates suggested that eight of 15 return travellers in Cluster 2 were likely infected during their

Figure 4

Cluster 2 genetic structure and contact tracing results, Düsseldorf and Solingen, Germany, 15 June–01 August 2021 (n = 42 cases)



{Z4076, Z4077}: joint Barcelona trip until 27/28 June
{S29, S30, S116}: family contacts
{S77, S121, S76}: other contacts of S29
{Z4201, Z4231}: family contacts, household
{Z4162, N1520, N1521}: household
{Z4181, Z4185}: household, joint travel
{Z4416, Z4417}: household, joint travel
{Z4141, Z4144}: direct contacts
{Z4143, Z4144}: household
{Z4409, Z4497}: household
{Z4409, Z4431}: suspected direct contact
{Z4187, Z4145}: visit to 'Bar A' on 02 July (see Fig. 1)

A. Bar plot showing the number of cases according to their status with regard to travel, contacts with returning travellers and exposure to the Old Town of Düsseldorf in Cluster 2.

B. Bar plot showing the number of cases in Cluster 2 with contact-tracing links found to other cases in Cluster 2 and the number of cases with no links identified by contact tracing.

C. Visualisation of genetic distances and contact tracing results for Cluster 2, created with MicrobeTrace [21]. Each node represents one case; edges between nodes indicate links identified by contact tracing; edge colours indicate genetic distances between the samples linked by contact tracing. The text inset details the precise nature of the identified case relationships.

stay abroad because their symptoms had started either when abroad or within the first 24 hours after return. For some of these cases, exposure in Düsseldorf could be ruled out with certainty (e.g. Z4106, with symptom onset on the same day as the return flight). The viral sequences of Z4077 and Z4076, which belonged to some of the earliest Cluster 2 cases in Düsseldorf, exhibited a genetic distance of 3. The cases corresponding to these two sequences were likely infected during a joint trip to Barcelona (symptom onset 1 and 3 days after return to Düsseldorf, respectively), and likely represented two independent infection events in Barcelona. Cluster 2 thus likely reflected multiple independent introduction events of a viral strain also circulating in Catalonia, Spain and other European countries, followed by diffuse community transmission in Düsseldorf. Consistent with this, isolates closely related to Z4076 (EPI_ISL_2982413; genetic distance 1; sampling date 30 June) and Z4077 (EPI_ISL_3306421, EPI_ISL_3009886, EPI_ISL_3009934; genetic distance 0; sampling dates from 06 July onwards) were identified by genomic surveillance in Barcelona, in a collaborative effort to investigate this cluster. Of note, an additional direct contact of Z4077 and Z4076 in Barcelona also tested positive for SARS-CoV-2 after return to Düsseldorf and reported many contacts at work and in social gatherings in Düsseldorf and other cities, possibly contributing to community transmission; the viral sequence of this case could be obtained through a commercial diagnostic laboratory in Cologne and was found to also cluster with Cluster 2 (data not shown). Consistent with diffuse community transmission in and around Düsseldorf, for five additional cases from Düsseldorf, and one case from Solingen at the beginning of a six-person transmission chain, case interviews suggested a possible exposure to the virus in the Old Town district of Düsseldorf (Figure 4). Furthermore, consistent with circulation of the Cluster 2 viral strain in other European regions, sequence-identical samples from multiple European countries (e.g. Belgium France, Germany) were identified via a GISAID Audacity Instant search for the sequence of Z4077, with sample collection dates of the identified sequences beginning on 24 June, i.e. approximately 1 week before sample collection of the first Cluster 2 cases in Düsseldorf (30 June). Over the course of July, Cluster 2 accounted for 4% of total SARS-CoV-2 infections in Düsseldorf.

Discussion

Increased understanding of SARS-CoV-2 transmission chains in the population is important to support improved containment strategies. Here, we used IGS, an emerging approach integrating genetic and classical epidemiological data, to investigate two large clusters of genetically-related SARS-CoV-2 isolates occurring in Düsseldorf. Taken together, isolates from these two clusters accounted for more than 10% of SARS-CoV-2 cases in the city during the considered period (Figure 3). We show that IGS allowed to trace complex SARS-CoV-2 transmission chains in both clusters, each involving cases who had journeyed abroad.

While both identified clusters were related to travel-imported SARS-CoV-2 infections, they exhibited different patterns with respect to transmission and potential public health implications. A large proportion of detected cases in Cluster 1 could be linked to a putative index case, who was a returning traveller with an initially undetected SARS-CoV-2 infection; non-adherence to infection prevention rules at nightlife venues contributed to an environment conducive to subsequent spread of infections. Cluster 1-associated case load could likely have been reduced at multiple points: by detecting the infection of IP upon return to Germany e.g. through a PCR test at the border; by mandatory quarantine regardless of infection status for return travellers; or by measures to reduce infection risks in nightlife settings, such as mandatory PCR or rapid antigen testing of patrons or strict adherence to physical distancing and indoor masking rules. What is more, compliance with mandatory customer contact data collection rules in place would have benefited post-hoc cluster investigation efforts, and it may also have contributed to further infection containment efforts. Of note, Spain, including the Balearic Islands, was only declared a region of high COVID-19 risk by the German health authorities on 27 July [17], i.e. almost 4 weeks after the first transmission in the reported clusters occurred. Cluster 2, by contrast, was driven by multiple independent introductions of a viral strain also circulating in Catalonia and other parts of Europe, and a larger number (n = 6) of Cluster 2 cases were found in another city, Solingen, which is part of the wider metropolitan area around Düsseldorf. Strict testing requirements for return travellers would likely also have contributed to a significant reduction of Cluster 2 cases.

IGS was a necessary approach for the investigation of the two clusters; in many instances, links between cases were only uncovered by the structured case interviews carried out after genetic links had been identified. On the other hand, Clusters 1 and 2 may have been considered as connected without the additional information gathered through genomic surveillance. Indeed, two cases from Cluster 2 (Z4187 and Z4145) were present in one of the bars where Cluster 1-related transmissions took place. Furthermore, the genomic data collected during the investigation of Cluster 2 could be analysed together with genomic data from the city of Barcelona to trace infection chains beyond Germany. The joint effort between researchers in Düsseldorf and Barcelona, which contributed to understanding the spread of the virus in Cluster 2, also demonstrates the potential of pan-European collaboration.

This study has multiple potential limitations. First, relevant cases may be missing from the analysis of the two clusters. Reasons for this may include undetected infections in asymptomatic individuals or failure to identify relevant cases during contact tracing or based on genetic data. Second, high-quality viral genomes were only available for ca 50% of SARS-CoV-2 cases in Düsseldorf during the considered period, so relevant

cases in the genetic analysis may have been missed (see previous point). There remains also uncertainty with respect to cases for whom no sequencing data were available, such as the epidemiologically linked additional KP cases of Cluster 1. In addition, the assignment of KP1_5 to Cluster 1 remained ambiguous even with genetic data. Third, while the assignment of IP as the index case of Cluster 1 was supported by dates of symptom onset, the reconstructed pattern of putative transmission events in different bars, and the detection of related viral isolates from the Balearic Islands, a degree of uncertainty remained, as there was no way to rule out the presence of undetected cases or independent importation events at the beginning of the detected transmission chains. The detection of related sequences in other regions highlights the possibility of additional independent introduction events, in particular for Cluster 1 cases from mid-July onwards for which no link to IP could be identified, as well as the possibility of export of viral strains from the Düsseldorf area, as inferring the directionality of transmission from genetic data alone is generally not possible. In addition, it is possible that the structured case interviews failed to uncover relevant case travel histories. Fourth, the possibility of multiple exposures to SARS-CoV-2 during high-incidence periods, likely observed e.g. for two Cluster 2 cases in this study, represents a general challenge for the accurate tracing of transmission chains; inferred links that are not supported by both genomic and epidemiological evidence should be interpreted with caution. Fifth, the current study was carried out on a timescale of weeks and retrospective in nature; in the future, the sequencing speeds achievable with modern single-molecule sequencing technologies (from 'swab to sequence' in<72 hours [10]) may enable implementations of IGS that support 'real-time' containment efforts. Sixth, the IGSD does not comprise the routine collection of clinical metadata, and case severity is not used as a sample selection criterion; inclusion of clinical metadata may further increase the utility of IGS.

Due to the emerging nature of IGS, there are many remaining open questions with respect to how to best design and implement an IGS system. For example, it is unclear at which level – national, regional, or at the level of a city – the integration between genetic and contact tracing data should be carried out, and how structured case interviews for backward contact tracing should best be conducted. The two clusters presented here clearly demonstrate the benefits of integrating local knowledge acquired 'on-the-ground' with data gathered by surveillance systems in different cities or countries; in the future, integration of local systems with larger national or European networks may enable the improved characterisation of introduction events and viral strain flow across cities and states.

Our work demonstrates the feasibility of tracing SARS-CoV-2 infection chains through a locally implemented system and during the later phases of the pandemic with high-incidence community transmission in an internationally connected city. This study complements existing studies from earlier phases of the pandemic [12,18] or from national or state-level genomic surveillance systems [7,11,13]. While the developed IGS system is currently limited to the tracing of SARS-CoV-2, its future potential applications include other emerging pathogens or multi-resistant bacterial pathogens.

Collaborators

German COVID-19 OMICs Initiative (DeCOI): Janine Altmüller, Angel Angelov, Anna C. Aschenbrenner, Robert Bals, Alexander Bartholomäus, Anke Becker, Matthias Becker, Daniela Bezdan, Michael Bitzer, Conny Blumert, Ezio Bonifacio, Peer Bork, Bunk Boyke, Helmut Blum, Nicolas Casadei, Thomas Clavel, Maria Colome-Tatche, Markus Cornberg, Inti Alberto De La Rosa Velázquez, Andreas Diefenbach, Alexander Dilthey, Nicole Fischer, Konrad Förstner, Sören Franzenburg, Julia-Stefanie Frick, Gisela Gabernet, Julien Gagneur, Tina Ganzenmueller, Marie Gauder, Janina Geißert, Alexander Goesmann, Siri Göpel, Adam Grundhoff, Hajo Grundmann, Torsten Hain, Frank Hanses, Ute Hehr, André Heimbach, Marius Hoeper, Friedemann Horn, Daniel Hübschmann, Michael Hummel, Thomas Iftner, Angelika Iftner, Thomas Illig, Stefan Janssen, Jörn Kalinowski, René Kallies, Birte Kehr, Andreas Keller, Oliver T. Keppler, Sarah Kim-Hellmuth, Christoph Klein, Michael Knop, Oliver Kohlbacher, Karl Köhrer, Jan Korbel, Peter G. Kremsner, Denise Kühnert, Ingo Kurth, Markus Landthaler, Yang Li, Kerstin U. Ludwig, Oliwia Makarewicz, Manja Marz, Alice C. McHardy, Christian Mertes, Maximilian Münchhoff, Sven Nahnsen, Markus Nöthen, Francine Ntoumi, Peter Nürnberg, Stephan Ossowski, Jörg Overmann, Silke Peter, Klaus Pfeffer, Isabell Pink, Anna R. Poetsch, Ulrike Protzer, Alfred Pühler, Nikolaus Rajewsky, Markus Ralser, Kristin Reiche, Olaf Rieß, Stephan Ripke, Ulisses Nunes da Rocha, Philip Rosenstiel, Antoine-Emmanuel Saliba, Leif Erik Sander, Birgit Sawitzki, Simone Scheithauer, Philipp Schiffer, Jonathan Schmid-Burgk, Wulf Schneider, Eva-Christina Schulte, Joachim L. Schultze, Alexander Sczyrba, Mariam L. Sharaf, Yogesh Singh, Michael Sonnabend, Oliver Stegle, Jens Stoye, Fabian Theis, Thomas Ulas, Janne Vehreschild, Thirumalaisamy P. Velavan, Jörg Vogel, Sonja Volland, Max von Kleist, Andreas Walker, Jörn Walter, Dagmar Wieczorek, Sylke Winkler, John Ziebuhr

Ethical statement

This study was approved by the ethics committee of the Medical Faculty of Heinrich Heine University Düsseldorf (#2020–839).

Data availability

FASTA files and the phylogenetic tree of all Düsseldorf and Solingen samples have been made available via OSF (DOI 10.17605/OSF.IO/B7TU3). All viral genome assemblies from Düsseldorf surveillance samples that meet GISAID quality criteria have also been submitted to GISAID. Accessions are listed in Supplementary Table 2.

Acknowledgements

This work was supported by the Ministry for Work, Health and Social Affairs of the State of North Rhine-Westphalia ("24.04.01 Gen. u. IKA" and CPS-1-1A); the Jürgen Manchot

Foundation; the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung; award numbers 031L0184B and 01KX2021); the German Research Foundation (award 428994620).

We gratefully acknowledge the Authors, the Originating and Submitting Laboratories for their sequence and metadata shared through GISAID. All submitters of data may be contacted directly via GISAID. The Acknowledgments Table for GISAID is part of the Supplement (Supplementary Table 1). Computational support and infrastructure were provided by the "Centre for Information and Media Technology" (ZIM) at the University of Düsseldorf (Germany). Sequencing and the design of sequencing workflows were supported by the Biologisch-Medizinisches Forschungszentrum der Heinrich Heine University Düsseldorf (BMFZ).

Rainer Zotz sadly passed away in September 2022 prior to the publication of this work. This work would not have been possible without his contributions and his support, and he will be deeply missed.

Conflict of interest

None declared

Authors' contributions

TH, KP, JT, LE, AW, and ATD conceptualised and designed the study. SB, ET, DN, PF, JN, CA, BS, TW, MKV, PK, MP, AT, JW, NL, SH, MA, KG, RZ gave input into the study design and carried out data analyses and integration. YF, PCN, ASS, MHG, CAV, AAP, AB, DA, HH, JD, LS, JB, CL, UO, CG, PPK, HJ, DS, LH, KH, CJS, and NJ analysed clinical samples, designed and implemented sequencing data analysis approaches, and contributed sequencing data. All authors have commented on the draft and approved the final version. German COVID-19 OMICS Initiative (DeCOI) contributed to the design of wet lab protocols and bioinformatics workflows.

References

1. World Health Organization (WHO). COVID-19 Weekly Epidemiological Update, Edition 106, published 24 August 2022. Geneva: WHO; 2022.
2. Pung R, Chiew CJ, Young BE, Chin S, Chen MI, Clapham HE, et al. , Singapore 2019 Novel Coronavirus Outbreak Research Team. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. Lancet. 2020;395(10229):1039-46. https://doi.org/10.1016/S0140-6736(20)30528-6 PMID: 32192580
3. Kang J, Jang YY, Kim J, Han SH, Lee KR, Kim M, et al. South Korea's responses to stop the COVID-19 pandemic. Am J Infect Control. 2020;48(9):1080-6. https://doi.org/10.1016/j.ajic.2020.06.003 PMID: 32522606
4. Peak CM, Kahn R, Grad YH, Childs LM, Li R, Lipsitch M, et al. Individual quarantine versus active monitoring of contacts for the mitigation of COVID-19: a modelling study. Lancet Infect Dis. 2020;20(9):1025-33. https://doi.org/10.1016/S1473-3099(20)30361-3 PMID: 32445710
5. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health. 2020;5(8):e452-9. https://doi.org/10.1016/S2468-2667(20)30157-2 PMID: 32682487
6. Miller JS, Bonacci RA, Lash RR, Moonan PK, Houck P, Van Meter JJ, et al. COVID-19 Case Investigation and Contact Tracing in Central Washington State, June-July 2020. J Community Health. 2021;46(5):918-21. https://doi.org/10.1007/s10900-021-00974-5 PMID: 33689116
7. Lane CR, Sherry NL, Porter AF, Duchene S, Horan K, Andersson P, et al. Genomics-informed responses in the elimination of COVID-19 in Victoria, Australia: an observational, genomic epidemiological study. Lancet Public Health. 2021;6(8):e547-56. https://doi.org/10.1016/S2468-2667(21)00133-X PMID: 34252365
8. COVID-19 Genomics UK (COG-UK) consortiumcontact@cogconsortium.uk. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe. 2020;1(3):e99-100. https://doi.org/10.1016/S2666-5247(20)30054-9 PMID: 32835336
9. Callaway E. The coronavirus is mutating - does it matter? Nature. 2020;585(7824):174-7. https://doi.org/10.1038/d41586-020-02544-6 PMID: 32901123
10. Walker A, Houwaart T, Finzer P, Ehlkes L, Tyshaieva A, Damagnez M, et al. , German COVID-19 OMICS Initiative (DeCOI). Characterization of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection Clusters Based on Integrated Genomic Surveillance, Outbreak Analysis and Contact Tracing in an Urban Setting. Clin Infect Dis. 2022;74(6):1039-46. PMID: 34181711
11. Hjorleifsson KE, Rognvaldsson S, Jonsson H, Agustsdottir AB, Andresdottir M, Birgisdottir K, et al. Reconstruction of a large-scale outbreak of SARS-CoV-2 infection in Iceland informs vaccination strategies. Clin Microbiol Infect. 2022;28(6):852-8. https://doi.org/10.1016/j.cmi.2022.02.012 PMID: 35182757
12. Popa A, Genger JW, Nicholson MD, Penz T, Schmid D, Aberle SW, et al. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. Sci Transl Med. 2020;12(573):eabe2555. https://doi.org/10.1126/scitranslmed.abe2555 PMID: 33229462
13. Douglas J, Geoghegan JL, Hadfield J, Bouckaert R, Storey M, Ren X, et al. Real-Time Genomics for Tracking Severe Acute Respiratory Syndrome Coronavirus 2 Border Incursions after Virus Elimination, New Zealand. Emerg Infect Dis. 2021;27(9):2361-8. https://doi.org/10.3201/eid2709.211097 PMID: 34424164
14. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. https://doi.org/10.1093/molbev/mst010 PMID: 23329690
15. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017;22(13):30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494 PMID: 28382917
16. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5(11):1403-7. https://doi.org/10.1038/s41564-020-0770-5 PMID: 32669681
17. Robert Koch Institute (RKI). Informationen zur Ausweisung internationaler Risikogebiete, 23. Juli 2021. [Information on the Risk Classification of Regions outside Germany, 23 July 2021]. Berlin: RKI; 2021. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Transport/Archiv_Risikogebiete/Risikogebiete_2021-07-23.pdf?__blob=publicationFile
18. Walker A, Houwaart T, Wienemann T, Vasconcelos MK, Strelow D, Senff T, et al. Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020. Euro Surveill. 2020;25(22). https://doi.org/10.2807/1560-7917.ES.2020.25.22.2000746 PMID: 32524946
19. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44(W1):W242-5. https://doi.org/10.1093/nar/gkw290 PMID: 27095192
20. Gansner ER, North SC. An open graph visualization system and its applications to software engineering. Softw Pract Exper. 2000;30(11):1203-33. https://doi.org/10.1002/1097-024X(200009)30:11<1203::AID-SPE338>3.0.CO;2-N
21. Campbell EM, Boyles A, Shankar A, Kim J, Knyazev S, Cintron R, et al. MicrobeTrace: Retooling molecular epidemiology for rapid public health response. PLOS Comput Biol. 2021;17(9):e1009300. https://doi.org/10.1371/journal.pcbi.1009300 PMID: 34492010

License, supplementary material and copyright

Any supplementary material referenced in the article can be found in the online version.