Aus dem Institut für Virologie der Heinrich-Heine-Universität Düsseldorf

Direktor: Univ.- Prof. Dr. med. Jörg Timm

Entwicklung einer Methode zur Amplifikation und Sequenzierung des kompletten Genoms des Hepatitis C Virus

Dissertation

Zur Erlangung des Grades eines Doktors der Medizin

der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Camilla Sophia Neuroth

2025

Angaben der Gutachter/innen

Als Inauguraldissertation gedruckt mit der Genehmigung der

Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez:

Dekan/in: Prof. Dr. med. Nikolaj Klöcker

Erstgutachter/in: Prof. Dr. med. Jörg Timm

Zweitgutachter/in: PD Dr. rer. nat. Doreen M. Floß

Zusammenfassung

Nach Angaben der Weltgesundheitsorganisation sind weltweit circa 50 Millionen Menschen chronisch mit dem Hepatitis C Virus infiziert. Diese Erkrankung kann unbehandelt zu schwerwiegenden Leberschäden und zum Tod führen. Eine frühzeitige Diagnose und eine effektive Therapie sorgen für die Unterbrechung von Infektionsketten und sind daher Ziele in der Bekämpfung der Virushepatitis. Seit 2015 können durch *direct-acting antivirals* Virusheilungsraten von > 95 % erreicht werden. Dennoch stagniert der zuvor beobachtete Rückgang der gemeldeten Infektionen mit dem Hepatitis C Virus. Fragen, wo und durch wen sich Menschen mit dem Virus anstecken, können durch Kontaktnachverfolgung und phylogenetische Analyse besser beantwortet werden, wie diese zum Beispiel (z.B.) bereits in der Ebola-Epidemie 2013-2016 oder der Corona-Pandemie 2020-2022 erfolgreich durchgeführt wurden. Für eine optimal aufgelöste phylogenetische Analyse ist die Aufschlüsselung des gesamten Virusgenoms notwendig. Ziel dieser Dissertationsarbeit ist daher insbesondere die Entwicklung einer Ganzgenom- *polymerase chain reaction* (PCR).

In der Entwicklung einer Ganzgenom-PCR des Hepatitis C Virus stellen unter anderem die hohe Variabilität als auch die acht unterschiedlichen Genotypen eine Herausforderung dar. In dieser Dissertationsarbeit wurde der Ansatz einer Genotyp-spezifischen *Tiling*-PCR verfolgt, bei der das Genom in vielen kleinen, überlappenden Abschnitten amplifiziert, anschließend sequenziert und bioinformatisch wieder zusammengesetzt wird. Nach der PCR erfolgte eine MinION-Sequenzierung mittels Nanopore sowie eine bioinformatische Datenauswertung. Dabei wurde sich auf die in der Düsseldorfer Kohorte vorherrschenden Genotypen (GT) 1a, 1b, 3a, 4a und 4d konzentriert.

Mit einer Ganzgenom-PCR soll eine möglichst hohe Genomabdeckung erzeugt werden. Mit der Tiling-PCR wurde eine Genomabdeckung > 90 % bei 64 % der Proben des GT 1a, bei 77 % der Proben des GT 1b, bei 98 % der Proben des GT 3a, bei 60 % der Proben des GT 4a und bei 77 % der Proben des GT 4d erreicht. Eine Genomabdeckung > 95 % wurde bei 25 % der Proben des GT 1a, bei 47 % der Proben des GT 1b, bei 90 % der Proben des GT 3a, bei 60 % der Proben des GT 4a und bei 62 % der Proben des GT 4d beobachtet. Für alle fünf Genotypen gemeinsam betrachtet wurde im Mittel eine Genomabdeckung von 91 % ($\pm$ 15 %) erreicht. Limitierende Faktoren der PCR stellen z.B. die Viruslast und die Probenqualität dar. Weiterführende Arbeiten haben als Ziel die Optimierung und Validierung der Tiling-PCR, sodass eine neue Standarddiagnostikmethode im Institut für Virologie der Uniklinik Düsseldorf etabliert werden kann.

Summary

According to the World Health Organization, around 50 million people worldwide are chronically infected with the hepatitis C virus. If left untreated, this disease can lead to serious liver changes and death. Early diagnosis and effective treatment ensure that chains of infection are interrupted and are therefore goals in the fight against viral hepatitis. Since 2015, direct-acting antivirals have been able to achieve viral cure rates of > 95 %. Nevertheless, the previous drop in new infections with the hepatitis C virus has stagnated since 2015. Questions about where and through whom people become infected with the virus can be answered by contact tracing and phylogenetic analysis, as has already been successfully carried out in the Ebola epidemic 2013-2016 or the Corona pandemic 2020-2022, for example. For a high-resolution phylogenetic analysis, it is necessary to break down the entire virus genome. The aim of this dissertation is therefore especially to develop a whole genome polymerase chain reaction (PCR).

In the development of a whole genome PCR of the hepatitis C virus, the high variability as well as the eight different genotypes pose a challenge. In this dissertation, the approach of a genotype-specific tiling PCR was pursued, in which the genome is amplified in many small, overlapping sections, then sequenced and bioinformatically reassembled. The PCR was followed by MinION sequencing using a nanopore and bioinformatic data analysis. The focus was on the predominant genotypes (GT) 1a, 1b, 3a, 4a and 4d in the Düsseldorf cohort.

The aim of whole genome PCR is to achieve the highest possible genomic coverage. With tiling PCR, genomic coverage > 90 % was achieved in 64 % of the samples of the GT 1a, 77 % of the GT 1b samples, 98 % of the GT 3a samples, 60 % of the GT 4a samples and 77 % of the GT 4d samples. Genomic coverage > 95 % was observed in 25 % of GT 1a samples, in 47 % of GT 1b samples, in 90 % of GT 3a samples, in 60 % of GT 4a samples and in 62 % of GT 4d samples. Considering all five genotypes together, an average genomic coverage of 91 % (± 15 %) was achieved. Limiting factors of the PCR are for example viral load and sample quality. The aim of further work is to optimize and validate the tiling PCR so that a new standard diagnostic method can be established at the Institute of Virology at Düsseldorf University Hospital.

Abkürzungsverzeichnis

A	Adenin
C	Cytosin
ca.	circa
cDNA	Engl. *complementary deoxyribonucleic acid*
DAA	Engl. *direct antiviral agents*
DNA	Engl. *deoxyribonucleic acid*
dNTP	Desoxy-Nukleosidtriphosphat
DTT	Dithiothreitol
E	Engl. *envelope protein*
G	Guanin
GT	Genotyp
HAV	Hepatitis A Virus
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HIV	humanes Immundefizienz Virus
HVR	Hypervariable Region
IDU	Engl. *Injection-drug-use*
IUPAC	Engl. *International Union of Pure and Applied Chemistry*
i.v.	intravenös
Min	Minuten
NANBH	Nicht-A, Nicht-B-Hepatitis
NGS	Engl. *Next Generation Sequencing*
Nr.	Nummer
NS	Engl. *Non-structual protein*
nt	Nukleotidposition
NTR	nicht-translatierte Region
PCR	Engl. *polymerase chain reaction*

PEPSI	Engl. *Screening Project of Protease Inhibitor*
RAS	Resistenz-assoziierte Substitutionen
RKI	Robert Koch-Institut
RNA	Engl. *ribonucleic acid*
RT	Reverse Transkription
SARS-CoV-2	Engl. *severe acute respiratory syndrome coronavirus 2*
Sek	Sekunde(n)
SFB	Engl. *Short Fragment Buffer*
sog.	sogenannt
T	Thymin
TBE	TRIS-Borat-EDTA
u.a.	unter anderem
UTR	Untranslatierte Region
Vgl.	Vergleiche
VI-Nr.	Virologie-Nummer
WHO	Weltgesundheitsorganisation
z.B.	zum Beispiel
#	Nummer

Inhaltsverzeichnis

1 Einleitung 1

1.1 Hepatitis C Virus (HCV) Epidemiologie 1

1.2 HCV Infektion 1

1.3 HCV Struktur 2

1.4 HCV Genotypen 3

1.5 HCV Transmission 5

1.6 HCV Diagnostik 6

1.7 HCV Therapie 7

1.8 Problematiken in der HCV-Therapie 8

1.9 Multiplex PCR 9

1.10 HCV Ganzgenom-PCR 11

2 Ziele dieser Arbeit 14

3 Material und Methoden 15

3.1 Material 15

3.1.1 Geräte 15

3.1.2 Verbrauchsmaterialien 16

3.1.3 Chemikalien und Enzyme 17

3.1.4 Kits 19

3.1.5 Oligonukleotide 19

3.1.6 Referenzsequenzen 20

3.1.7 Verwendete Software 20

3.2 Patientenproben 21

3.3 Methoden 22

3.3.1 Probenaufbereitung 22

3.3.2 Reverse Transkription von HCV RNA 22

3.3.3 Tiling - PCR 23

3.3.4 Agarose-Gelelektrophorese 24

3.3.5 DNA-Quantifizierung 25

3.3.6 MinION - Sequenzierung 25

3.3.7 Illumina - Sequenzierung 28

3.3.8 Bioinformatische Auswertung der Rohdaten der Sequenzierung 29

3.3.9 Resistenztestung 30

3.3.10 Statistische Auswertung 30

4 Ergebnisse und Validierung 32

4.1 Konzeptentwicklung der Tiling PCR und Sequenzierung 32

4.2 Optimierung der Reversen Transkription 32

4.3 Entwicklung der Tiling-PCR Primer-Sets 34

4.3.1 Etablierung Genotyp 1a 34

4.3.2 Etablierung Genotyp 1b 41

4.3.3 Etablierung Genotyp 3a 46

4.3.4 Etablierung Genotyp 4 51

4.4 Überprüfung der Methode mit höherer Fallzahl 54

4.4.1 Genotyp 1a 54

4.4.2 Genotyp 1b 60

4.4.3 Genotyp 3a 62

4.4.4 Genotyp 4 65

4.4.5 Zusammenfassende Gegenüberstellung sowie Mittelwerte der Genotypen 68

4.5 Beschreibung der Düsseldorfer Kohorte 70

4.5.1 Phylogenetische Datenanalyse 70

4.5.2 Resistenztestung 72

4.6 Validierung der Methode im Vergleich zu anderen Verfahren 73

4.6.1 Vergleich zwischen der Tiling-PCR mit Nanopore-Sequenzierung und der konventionellen Sanger-Sequenzierung 73

4.6.2 Vergleich zwischen der Tiling-PCR mit Illumina- und mit Nanopore-Sequenzierung 74

4.6.3 Vergleich der Nanopore- und Illumina- Sequenzierung zur Sanger-Sequenzierung 77

5 Diskussion 78

5.1 Etablierung der Tiling-PCR 78

5.1.1 Reverse Transkription 79

5.1.2 Workflow und Einflussgrößen für den Sequenziererfolg 79

5.1.3 Korrektur der Primer-Sets 81

5.1.3.1 Genotyp 1a und 3a 81

5.1.3.2 Genotyp 1b 82

5.1.3.3 Genotyp 4 83

5.1.3.4 Zusammenfassung 83

5.2 Überprüfung der Methode mit höherer Fallzahl 83

5.2.1 Relative Häufigkeiten des Erfolgs 83

5.2.2 Korrelationen 84

5.2.3 Resistenztestung 85

5.3 Vergleich der Düsseldorfer IDU-Kohorte 86

5.4 Vergleich unterschiedlicher Sequenziermethoden 87

5.5 Vergleich zu anderen Ganzgenom-PCR Methoden 89

6 Schlussfolgerungen / Ausblick 92

7 Abbildungsverzeichnis 94

8 Tabellenverzeichnis 95

9 Literaturverzeichnis 96

10 Anhang 101

11 Danksagung 118

1 Einleitung

1.1 Hepatitis C Virus (HCV) Epidemiologie

Die Hepatitis C – Infektion hat weltweit eine große Bedeutung. Nach aktuellen Schätzungen der Weltgesundheitsorganisation (WHO) gibt es jährlich circa (ca.) 1,0 Millionen Neuinfektionen und ca. 50 Millionen Menschen leiden an einer chronischen Hepatitis C – Infektion. Die Krankheitslast in Europa hat daran einen Anteil von 18 %, was ca. 9 Millionen chronisch infizierter Menschen bedeutet. [1]

Die chronische HCV-Infektion ist für 27 % der Fälle mit Leberzirrhose sowie 25 % der Fälle mit hepatozellulärem Karzinom verantwortlich und macht gemeinsam mit der Hepatitis B – Infektion die mehrheitliche Ursache des Karzinoms aus. [2]

1.2 HCV Infektion

Im Jahr 1975 untersuchte Harvey Alter Transfusions-bedingte Infektionen der Leber. Hierbei konnte er sowohl eine Hepatitis A Virus (HAV) -Infektion ausschließen, da es für diese bereits eine Nachweismethode gab, als auch eine Hepatitis B Virus (HBV) -Infektion, da kein S-Antigen im Blut gefunden wurde. H. Alter definierte daher die Nicht-A, Nicht-B-Hepatitis (NANBH). [3] Der Nachweis eines NANBH-Antigens erfolgte dann im Jahr 1989 durch eine Arbeitsgruppe von M. Houghton. Mit Hilfe einer aus Blut eines infizierten Schimpansen gebildeten komplementären Desoxyribonukleinsäure- (cDNA-) Bibliothek konnten Antikörper gegen Virusproteine im Blut eines NANBH-infizierten Patienten entdeckt werden. So wurde eine eigenständige dritte Gruppe von Hepatitis-Viren nachgewiesen [4] und gleichzeitig der Grundstein zur Entwicklung von Antikörpertests zur Diagnostik von HCV gelegt. [4, 5] Der Kausalzusammenhang, dass das neu gefundene Virus eine Hepatitis auslösen kann, erfolgte 1997 durch Charles Rice. Er stellte einen genetischen Klon des Virus her und injizierte diesen in die Leber eines Schimpansen, welcher nachfolgend an Hepatitis erkrankte. [6]

Nach der Entdeckung des Hepatitis C Virus zeigte sich schnell, dass es sich bei dieser Virusinfektion um eine weltweit verbreitete Erkrankung handelt. Trotzdem ist sie nur wenig im öffentlichen Bewusstsein vertreten, weshalb eine Infektion mit HCV nicht immer oder meist spät diagnostiziert wird. Dies führt zu einer Vielzahl an Todesfällen, die nach Schätzungen der WHO vergleichbar sind zu Infektionen durch Malaria, humanes Immundefizienz-Virus (HIV) oder Tuberkulose. [7]

Der Verlauf der HCV-Infektion unterteilt sich in eine akute und eine chronische Phase. Nach einer Inkubationszeit, die zwischen zwei Wochen bis sechs Monaten betragen kann, verläuft die akute Erstinfektion bei ca. 80 % der Patienten symptomfrei. Mögliche Symptome sind zum Beispiel (z.B.) Fieber, Übelkeit, Bauchschmerzen oder ein Ikterus. Die akute Infektion ist in der Regel nicht lebensbedrohlich und 15-45 % heilen spontan innerhalb von 6 Monaten aus, ohne dass die Patienten eine Therapie bekommen haben. Die Mehrheit der Erkrankten, genauer 55-85 %, entwickeln eine chronische HCV – Infektion, die definitionsgemäß länger als 6 Monate besteht. Innerhalb von 20 Jahren entwickelt sich als Folge der chronischen Phase in 15-30 % der Fälle eine Leberzirrhose. [1]

Der einzige natürliche Wirt des Hepatitis C Virus ist der Mensch [8], weshalb theoretisch eine Elimination des Virus durch eine Impfung möglich wäre.

1.3 HCV Struktur

Das Hepatitis C Virus gehört zur Familie der *Flaviviridae* und bildet aufgrund seiner physikochemischen Eigenschaften sowie seiner Genomstruktur und -organisation eine eigenständige Gattung. [8]

Es handelt sich um ein behülltes, einsträngiges, lineares Ribonukleinsäure- (RNA-) Virus mit positiver Polarität und einer Länge von ca. 9600 Nukleotiden. Im Genom befindet sich ein großer offener Leserahmen, der für ein einziges Polyprotein von ca. 3000 Aminosäuren Länge codiert. Dieses Polyprotein wird von viralen und zellulären Proteasen in mindestens 10 Proteine aufgespalten, welchen unterschiedliche Funktionen zugeordnet werden können. Zu den Strukturproteinen werden das Kapsidprotein (Core) sowie zwei Hüllproteine (engl. *envelope*, E) E1 und E2 gezählt, die Bestandteile des Viruspartikels sind. Die Nicht-Strukturproteine (NS) 2 bis 5 sowie das hydrophobe Protein (p7) haben wichtige Funktionen bei der RNA-Replikation sowie der Virusmontage (engl. *assembly*) und -freisetzung. [9-11]

Der offene Leserahmen wird von nicht-translatierten Regionen (NTR) umgeben, die als hochgradige Strukturen am 5‘- und 3‘-Ende der HCV-RNA zu finden sind. Vor dem Startcodon befindet sich die 5‘-NTR, die sich durch eine komplexe Faltung kennzeichnet. Sie dient der RNA-Replikation und gleichzeitig als interne Ribosomeneintrittsstelle, an der das virale Polyprotein translatiert wird. Des Weiteren finden sich in der 5‘-NTR zwei Bindungsstellen für die leberspezifische micro-RNA 22. Besonders ist, dass die Bindung an das virale RNA-Genom zu einer Genexpressions-Steigerung führt. [9, 12, 13]

Am Ende des Genoms befindet sich die 3‘-NTR, welche eine essenzielle Funktion für die Vermehrung des RNA-Genoms hat. Sie ist strukturell dreigeteilt und besteht aus einer 40 Nukleotid langen, variablen Region, einer Polyuridin-reichen Region verschiedener Länge sowie einem hochkonservierten Bereich von 98 Nukleotiden Länge, der als X-tail bezeichnet wird. [14]

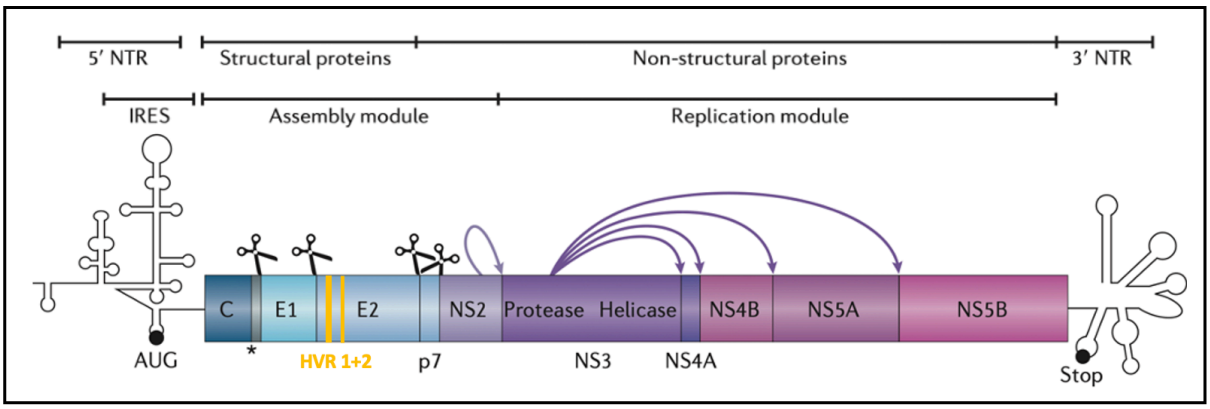

Abbildung 1: Strukturübersicht des Genoms des Hepatitis C Virus
Schematische Darstellung der viralen Genomstruktur des Hepatitis C Virus. An den Genomenden mit 5‘- und 3’-Nicht-translatierten Regionen sind Sekundärstrukturen abgebildet sowie das Startcodon (AUG) und das Stopcodon (Stop) des Polyproteins. Die Polyproteinspaltstellen durch diverse Proteasen sind als Scheren und Pfeile symbolisiert. In der Kapsidregion (C) stellt ein Sternchen die Spaltstelle einer Signalpeptid-Peptidase dar. An der internen Ribosomeneintrittsstelle (IRES) wird das virale Polyprotein translatiert. Mit gelben Strichen sind die hypervariablen Regionen (HVR) 1 und 2 innerhalb des Proteins E2 markiert. Abbildung modifiziert nach [15] und entnommen von Bartenschlager et al. [16]

Im Jahr 1991 konnten zwei hypervariable Regionen im Hüllprotein E2 nachgewiesen werden, die gemeinsam den variabelsten Abschnitt des Genoms darstellen. [17] Es zeigte sich, dass die hypervariable Region (HVR) 1 ein immundominantes Motiv und eine kritische Neutralisationsdomäne ist und sich aus diesem Grund bei Kontakt mit dem menschlichen Immunsystem neutralisierende Antikörper bilden. [18] Die Variabilität dieser Regionen stellt für das Immunsystem, die Entwicklung von Impfstoffen als auch für die Entwicklung einer Ganzgenom-PCR und -Sequenzierung eine große Herausforderung dar.

1.4 HCV Genotypen

Das Hepatitis C Virus weist eine große genetische Heterogenität auf. Diese ist bedingt durch eine hohe Replikationsrate sowie dem Mangel einer Korrekturlesefunktion der RNA-abhängigen RNA-Polymerase, die für das virale Nichtstrukturprotein 5b (NS5B) kodiert. [19] Phylogenetische Datenanalysen und Analogien zu anderen Viren lassen vermuten, dass die Divergenz der einzelnen Genotypen mindestens 500 bis 2000 Jahre zurück liegt. [20] In den 90er Jahren wurde von Simmonds [21, 22] auf Basis der phylogenetischen Sequenzanalyse

von NS5B ein einheitliches Klassifikationskonzept von HCV vorgeschlagen. Bis in die 2000er erfolgte die Einteilung in Genotyp, Subtyp sowie den Isolat eines Subtyps anhand der phylogenetischen Analyse sowie dem paarweisen Sequenzvergleich der Regionen Core, E1 und NS5. Ein Genotyp, Subtyp und Isolat eines Subtyps werden definiert durch eine Sequenzähnlichkeit von 60-70 %, 75-85 % und 90-98 %. Mit diesem Klassifikationssystem zeigten sich sechs verschiedene Genotypen und mehr als 90 Subtypen. [19]

Das Konzept musste erweitert werden, als ab 2005 eine Reihe von vollständigen HCV-Genomsequenzen veröffentlicht wurden. Ein Subtypname wurde dann als bestätigt oder vorläufig definiert, je nachdem ob eine vollständige oder partielle Nukleotidsequenz vorliegt. Außerdem erfolgt eine Bestätigung nur, wenn mindestens drei Beispiele des neuen Subtyps beschrieben wurden. [19, 23] Diese Kriterien definieren zum Zeitpunkt des Schreibens der Dissertation acht Genotypen und 90 bestätigten Subtypen. [24, 25]

Die Fehleranfälligkeit der RNA-abhängigen RNA-Polymerase ist durch das Fehlen einer Korrekturlesefähigkeit begründet [26] und führt zu einer Fehlerrate von einer Mutation pro 10^{-3} bis 10^{-5} Basen in einem Replikationszyklus. [27, 28] So kann sich innerhalb eines infizierten Individuums eine Mischung eng verwandter Sequenzvarianten bilden, die sich nur um ein Nukleotid von der in der Population verbreiteten Sequenz unterscheiden. Die Gesamtheit dieser Sequenzvarianten wird als Quasispezies bezeichnet. [28] Die hohe Variabilität innerhalb der Quasispezies führt zu einem Anpassungsvorteil an Umweltbedingungen und stellt eine Theorie dar, die chronische Persistenz im Wirt zu begründen. [29]

Die HCV-Infektion zeigt eine hohe globale Verbreitung, wobei die globale Verteilung einzelner Genotypen durch historische Einflüsse, wie Kolonialisierung, Sklavenhandel und Migration, erklärbar ist. Die verschiedenen Genotypen zeigen eine deutliche geographische Verteilung. Außerdem ist ein Zusammenhang mit Geschlecht, Alter und dem Übertragungsweg beobachtet worden. [30, 31]

Die weltweit höchste Prävalenz wurde mit 46,2 % der HCV-Fälle bei Genotyp (GT) 1 beobachtet, wovon ca. ein Drittel im ostasiatischen Raum vorkommt. In 99 % der Fälle treten dabei die Subtypen 1a (31 %) und 1b (68 %) auf. [31]

Phylogenetische Analysen konnten zeigen, dass sich der GT 1 von Westafrika ausgehend weltweit verbreitet hat. Dabei geschah die Ausbreitung des Subtyps 1b einige Jahre früher und steht im Zusammenhang mit der internationalen Verbreitung von kontaminierten Blutprodukten, während die Ausbreitung des Subtyps 1a erst nach 1960 erfolgte und durch die Zunahme des intravenösen (i.v.) -Drogenkonsums erklärbar ist. [32]

Der zweithäufigste HCV-Genotyp ist der GT 3. Er kommt in 30,1 % der weltweiten Fälle vor, wovon etwa drei Viertel in Südasien zu finden sind. In Westeuropa sind besonders junge Menschen mit Bezug zu i.v.-Drogenkonsum oder Tätowierungen mit dem Subtyp 3a infiziert. [33, 34]

Für die Mehrzahl der verbleibenden Fälle ist der GT 2 mit 9,1 %, der GT 4 mit 8,3 % und der GT 6 mit 5,4 % verantwortlich. Dabei treten die GT 2 und GT 6 gehäuft im ostasiatischen Raum auf und der GT 4 ist vermehrt zwischen Zentralafrika und dem Nahen Osten beobachtet worden. Die GT 5 und GT 7 machen nur < 1 % der weltweiten Fälle aus und hatten ihren Ursprung überwiegend auf dem afrikanischen Kontinent. [31]

Im Jahr 2018 wurden in Indien erstmals vier Patienten mit einer Genomsequenz des Virus entdeckt, die sich zu > 30 % von den bereits beschrieben Genotypen unterschied. Diese neue Sequenz wurde als Genotyp 8 bezeichnet. [35]

Im Rahmen dieser Dissertationsarbeit wurde sich zur Entwicklung einer Ganzgenom-Polymerasekettenreaktion (PCR) auf die in Deutschland dominierenden Geno- / Subtypen 1a, 1b, 3a, 4a und 4d beschränkt.

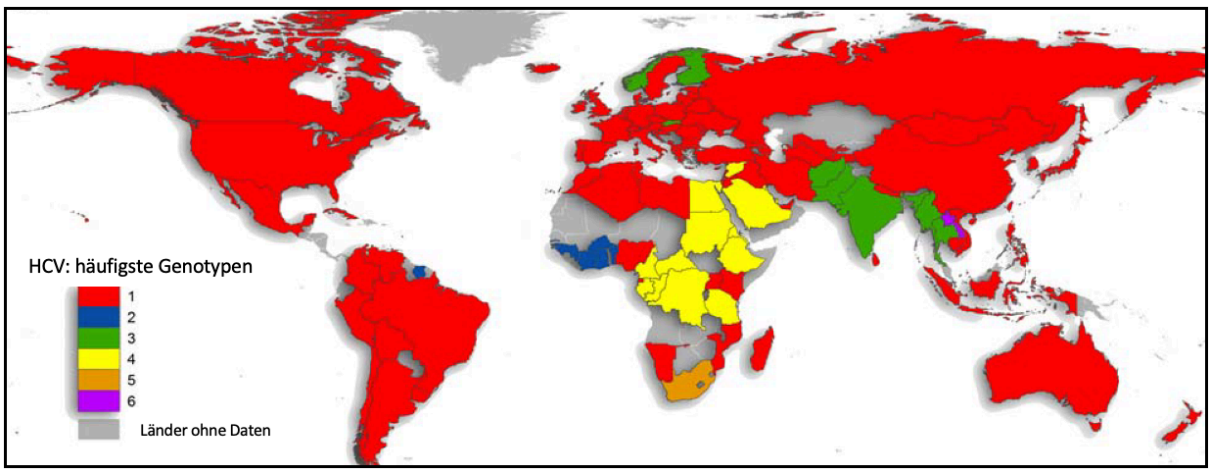

Abbildung 2: Übersicht der globalen Genotyp-Verteilung vom Hepatitis C Virus (HCV)
Die abgebildete Weltkarte stellt farblich die globale Genotyp-Verteilung von HCV dar. Jede Farbe stellt den mehrheitlich in diesem Land aufgetretenen Genotypen dar. Weltweit dominiert Genotyp 1 (rot), gefolgt von Genotyp 4 (gelb) und Genotyp 3 (grün). Nur in wenigen Ländern führend sind Genotyp 2 (blau), Genotyp 5 (orange) sowie Genotyp 6 (lila). Die Abbildung wurde entnommen und übersetzt aus den Studien von Messina. [31]

1.5 HCV Transmission

Das Hepatitis C Virus ist ein über Blut übertragenes Virus, dessen einziger bekannter Wirt der Mensch ist. Die weltweit häufigsten Übertragungsmöglichkeiten erfolgen durch Wiederverwendung oder unzureichende Sterilisation von kontaminierten Medizinprodukten,

Transfusion von ungetestetem Blut oder Blutprodukten sowie der gemeinsamen Verwendung von Injektionsutensilien im Rahmen von injizierendem Drogenkonsum. Letzterer stellt in Deutschland den häufigsten Übertragungsweg dar. [1, 8]

Des Weiteren ist eine vertikale Transmission von einer Mutter auf ihr Kind möglich sowie eine Übertragung bei sexuellen Praktiken, die zur Exposition gegenüber Blut führen. Es ist keine Verbreitung des Virus über Muttermilch, Nahrung, Wasser oder gelegentlichen Kontakt wie Umarmen oder Küssen bekannt. [1]

Ein beruflich bedingtes Infektionsrisiko bei medizinischem Personal beträgt bei europäischen Patienten ca. 0,42 % und ist beispielsweise durch Stichverletzung mit HCV-kontaminierten Kanülen möglich. [36]

Als Risikofaktor für eine HCV-Infektion wurde auch eine unsachgemäß durchgeführte Tätowierung, z.B. in Haftanstalten, identifiziert. [8, 37]

1.6 HCV Diagnostik

Da eine Neuinfektion mit Hepatitis C häufig asymptomatisch verläuft, wird die Erkrankung vieler Patienten erst Jahre nach der Infektion diagnostiziert, wenn diese sich bereits chronifiziert und häufig sekundäre Leberschäden verursacht hat. [1]

Die Diagnosestellung erfolgt in zwei Schritten. Zuerst wird mit einem indirekten, serologischen Test auf Anti-HCV Antikörper untersucht, ob der Patient bereits Kontakt zu dem Hepatitis C Virus hatte. Dabei ist eine serologische Lücke von 7-8 Wochen zu beachten. Anschließend erfolgt der direkte Nachweis von HCV-RNA mittels PCR, um zwischen einer ausgeheilten und infektiösen HCV-Erkrankung zu unterscheiden. [1, 8]

Nach der Sicherung einer aktiven HCV-Infektion sollte zur Prüfung der Therapieoptionen der HCV-Genotyp sowie die Konzentration der HCV-RNA, daher die Viruslast, bestimmt werden. [8] An der Uniklinik Düsseldorf erfolgt die Viruslastbestimmung durch eine quantitative PCR mit Hilfe des Vollautomats Cobas® 6800/8800 der Firma Roche und die Genotypisierung durch eine Sequenzierung des Genomabschnitts NS5A. [38] Weitere Resistenz-relevante Regionen können bei Bedarf mit Hilfe von Genotyp-abhängigen Protokollen sequenziert werden.

Bei einer chronischen HCV-Infektion sollte des Weiteren der Grad der Leberschädigung, z.B. eine Fibrose oder Zirrhose, festgestellt werden. Dies ist z.B. durch eine Leberbiopsie möglich. [1]

Trotz dieser diagnostischen Mittel kannten im Jahr 2022 weltweit nur ca. 36 % der 50 Millionen Personen mit HCV-Infektion ihre Diagnose. [1]

1.7 HCV Therapie

Die akute HCV-Infektion muss in der Regel nicht therapiert werden, da die Immunreaktion die Infektion bei einigen Menschen beseitigt. Wird die Infektion jedoch chronisch, ist eine Behandlung erforderlich, mit dem Ziel, die Erkrankung zu heilen. [1]

Die ersten Therapieansätze mit Interferon-α und Ribavirin waren mit Heilungsraten um 50 % wenig erfolgreich und zeigten starke Nebenwirkungen. [39, 40]

Anfang der 2000er wurde die Entwicklung der direkt antiviral wirksamen Substanzen (DAA) möglich, weil ein lang gesuchtes Zellkulturmodell zur molekularen Klonierung des viralen Genoms entdeckt wurde. [41] Bei dem sogenannten (sog.) „Replikon-System“ handelt es sich um gentechnisch miniaturisierte HCV-Genome, die sich in humanen Hepatomzellen vermehren. Dazu sind selbst-replizierende subgenomische RNAs, die Replikons, notwendig, welche die für die Genomreplikation verantwortlichen Nichtstrukturproteine (NS3-5) beinhalten. [42] Durch Nutzung der Replikons konnten diverse DAAs effizient getestet werden. [43] Dies ermöglichte die Entwicklung und Zulassung dreier Substanzklassen, welche die Virusvermehrung an unterschiedlichen Stellen blockieren. Die NS3/4A-Proteasehemmer inhibieren die proteolytische Spaltung des viralen Polyproteins, die NS5B-Polymerasehemmer behindern die Synthese der viralen RNA und die NS5A-Inhibitoren hemmen die Regulation von Virusreplikation und -partikelbildung. [42]

Heute empfiehlt die WHO eine Interferon-freie Therapie mit pan-genotypischen DAAs für Erwachsene und Kinder ab 3 Jahren mit chronischer HCV-Infektion. [1] Dabei erfolgt eine Kombination aus zwei oder mehr Substanzklassen, zumeist als eine Tablette, wie es bereits aus der HIV-Kombinationstherapie bekannt ist. Die Vorteile dieser HCV-Therapie sind eine orale Verfügbarkeit, eine kurze Behandlungsdauer von 12 bis 24 Wochen je nach vorliegender Zirrhose, kaum Nebenwirkungen sowie Heilungsraten von > 95 %. Nachteilig sind die hohen Kosten der Therapie, die sich in Deutschland auf 25.000 bis 60.000 Euro belaufen. Die Preise konnten durch die Einführung von Generika bereits drastisch gesenkt werden, dennoch wurden bis Ende 2022 nur ca. 20 % der mit chronischer HCV-Infektion diagnostizierten Patienten mit DAA behandelt. [1, 42]

Zur Überprüfung des Therapieansprechen wird mindestens 12 Wochen nach Ende der antiviralen Therapie eine Messung der HCV-RNA durchgeführt. Von einem dauerhaften virologischen Therapieansprechen ist bei einem fehlenden HCV-RNA-Nachweis auszugehen, der die Eradikation der HCV-Infektion anzeigt. [8]

1.8 Problematiken in der HCV-Therapie

Obwohl sich der Zugang zur HCV-Therapie bereits verbessert hat, ist er weiterhin begrenzt. Nur 12,5 der 50 Millionen mit HCV infizierten Menschen wurden im Jahr 2022 mit DAAs behandelt. [1] Ursächlich sind hierfür unter anderem (u.a.) fehlende Screening-Verfahren der HCV-Infektion, die zu einer späten oder fehlenden Diagnosestellung und unwissentlicher Verbreitung des Virus führen. [44] So kannten im Jahr 2022 nach Schätzungen der WHO nur 36 % (18 Millionen) der 50 Millionen infizierten Menschen ihre Diagnose. [1] In Deutschland wurde deshalb seit dem 1. Oktober 2021 ein einmaliges Hepatitis B und C -Screening für Versicherte ab 35 Jahren eingeführt. [45]

Weitere Hindernisse der HCV-Therapie stellen Re-Infektionen oder seltener Rückfälle (engl. *relaps*) dar. Re-Infektionen sind möglich, da aufgrund der Variabilität in der HVR1 eine erfolgreiche Therapie keinen immunologischen Schutz gegen eine erneute Infektion bietet. Sie korrelieren dabei häufig mit bestimmten Risikoverhalten, wie z.B. intravenösem Drogenabusus. [42, 44] Trotz der hohen virologischen Heilungsraten der DAA-Therapie wird die Infektion je nach Patientengruppe und Behandlungsschema bei 1-15 % der Betroffenen nicht beseitigt. Als Einflussfaktoren wurden u.a. die Verstoffwechselung der DAA-Wirkstoffe durch den Patienten, das Vorliegen einer ausgeprägten Leberfibrose oder -zirrhose, die Therapieadhärenz, genetische Faktoren sowie Resistenzen von HCV gegen die direkt wirkenden Virostatika beobachtet. [46]

Wie bereits beschrieben zeigt das Hepatitis C Virus eine hohe Replikationsrate und genetische Variabilität, die die Entstehung von Mutationen begünstigt und deshalb eine Quasispezies-Verteilung aufzeigt (Vergleiche (Vgl.) 1.4). In einem Patienten findet sich eine komplexe Mischung von genetisch unterschiedlichen, aber eng verwandten Viruspopulationen unterschiedlicher Größe. Die Replikationsfähigkeit (definiert als Fitness) einer Population bedingt ihren jeweiligen Anteil. Die HCV-Populationen koexistieren, aber jede Veränderung der Umweltbedingungen führt zu einer Neueinstellung des Gleichgewichts und damit zu einer Umverteilung in der Abundanz der Quasispezies-Populationen. [46, 47] Das Virus hat so die Möglichkeit sich schnell an veränderte Bedingungen anzupassen. Unter DAA-Therapie entsteht eine positive Selektion der Viruspopulation, die eine reduzierte Suszeptibilität für das Medikament aufweist. Gleichzeitig erweitert sich der Replikationsraum für diese Viruspopulation durch die vollständige Hemmung von DAA-sensitiven Wildtyp-Viren. Diese „sekundäre Aminosäuresubstitution“ oder „fitnessassoziierte Substitution“ erhöht die

schnelle Ausbreitung der resistenten Varianten während der Behandlung (Durchbruch) oder nach der Behandlung (Rückfall) und beeinflusst ihre Persistenz. [46] In der klinischen Praxis ist es daher sinnvoll, bei Patienten mit Reinfektion oder Relaps ein Resistenzprofil zu erstellen. [42, 48, 49]

Ein individuelles Therapieregime ist außerdem sinnvoll bei Patienten mit Begleit-Erkrankungen und Co-Infektionen, wie z.B. Leberzirrhose, hepatozellulärem Karzinom, eingeschränkter Nierenfunktion, HBV- oder HIV-Infektion. Häufig müssen in diesen Fällen auch Medikamenteninteraktionen berücksichtigt werden. [50]

Die Möglichkeit einer Immunisierung durch einen Impfstoff ist aufgrund der hohen Variabilität des Virus nicht gegeben. [1]

1.9 Multiplex PCR

Bei einer Multiplex PCR werden mehrere Desoxyribonukleinsäure (DNA) -Sequenzen (engl. *amplicon*) gleichzeitig in einer Reaktion amplifiziert. Erstmals gelang dies 1988, als mehrere, weit voneinander entfernte Sequenzen des humanen Muskeldystrophie-Gens zeitgleich amplifiziert werden konnten. [51, 52]

Eine Multiplex PCR kann zwei verschiedenen Ansätzen folgen. Entweder werden multiple Sequenzabschnitte mit einem einzelnen Primer oder einem Primer-Set amplifiziert oder es werden mehrere Primerpaare hinzugefügt, die auf jeweils unterschiedliche Sequenzabschnitte abzielen. Im folgenden Abschnitt wird sich auf die Entwicklung letzterer Methode bezogen. [52]

Seit der Entdeckung der PCR Anfang der 1980er Jahre durch K. Mullis [53] hat sich die Molekulardiagnostik zu einem Schlüsselverfahren der virologischen Diagnostik entwickelt. Eine Verwendungs-Limitierung der PCR lag für die Laboratorien zunächst in den hohen Kosten, bedingt durch den großen zeitlichen und personellen Aufwand für eine Probe. Eine Multiplex-Methode sollte deshalb diese Unzulänglichkeit überwinden und eine höhere diagnostische Kapazität erzeugen. [54]

Es zeigte sich, dass das nützlichste Testdesign eine empirische Auswahl von Oligonukleotid-Primern sowie die Verwendung von *Hot-Start*-basierten PCR-Methoden enthält. Diese PCR-Methoden konnten die Sensitivität und Spezifität verbessern, die Automatisierung erleichtern und die Kosten der PCR senken, wodurch die Multiplex-PCR besonders für die Diagnose von Infektionserregern mit viralen Nukleinsäuren relevant wurde. [54] Im Jahr 1996 wurde z.B.

bereits eine Dual-Target-PCR für die Charakterisierung von Masern [55] und 1998 für die Identifikation von Influenza B eingesetzt. [56]

Die Entwicklung einer Multiplex PCR gliedert sich in 6 Schritte: [52]

1. Auswahl der Amplicon-Positionen
2. Positionierung der Primer (in Abhängigkeit von der gewünschten Amplicon-Größe)
3. Primerdesign (mit ähnlicher Reaktionskinetik)
4. Entwicklung der PCR Konditionen für jedes Primer-Set
5. Zugabe der Primer-Sets zur Reaktion
6. Anpassen der Reaktionskomponenten (Magnesium, Desoxy-Nukleosidtriphosphat (dNTP), Polymeraseanforderung) sowie Zyklusbedingungen der PCR

Ein weiterer Vorteil der Multiplex PCR ist die verbesserte Detektion von falsch-negativen Ergebnissen, da jedes Amplicon eine interne Kontrolle für die anderen amplifizierten Fragmente darstellt. Wenn eine Deletion nur einen Teil der Sequenz betrifft, welche die Multiplex PCR vervielfältigt, werden einige Fragmente anzeigen, dass die Reaktion trotzdem erfolgreich war und die vorhandenen Lücken auf die Sequenz der Probe zurückzuführen sind. Gleichzeitig sind große Deletionen in der Regel zusammenhängend [57], weshalb nicht zusammenhängende Deletionen auf Artefakte durch Versagen einiger Fragmente der PCR schließen lassen. Zusätzlich kann durch Zugabe eines Kontroll-Amplicons außerhalb der Zielsequenz ein informatives Nicht-Amplifikationsergebnis von einem PCR-Fehler unterschieden werden. [52]

In den letzten 20 Jahren zeigte sich die Multiplex PCR als valide Methode zur schnellen und präzisen Identifizierung von Atemwegsviren, wie z.B. Influenza A + B, Respiratorisches Synzytial-Virus, humanes Metapneumo-Virus oder dem Adenovirus. [58-60] Außerdem hat sich die Technik als leistungsfähiges und kostengünstiges Instrument zur (Sub-)Typisierung von Virusstämmen in verschiedenen epidemiologischen Studien erwiesen. [54]

Die Arbeitsgruppe um J. Quick etablierte im Jahr 2017 eine Multiplex PCR-Methode zur MinION- und Illumina-Sequenzierung des Zika-Virus, die aber auch auf andere Viren übertragbar sei. Besonders ist an diesem Protokoll, dass das gesamte Genom des Zika-Virus sequenziert wird. Die Methode wurde u.a. für den Einsatz in Feldstudien und Diagnostik in Regionen mit limitierter Infrastruktur konzipiert, weshalb beispielsweise eine Analyse ohne Internetzugang möglich ist und die virale Consensus Sequenz einer klinischen Probe innerhalb von 1-2 Tagen mit einem einfachen Laborarbeitsablauf erstellt werden kann. Dabei wird das Virusmaterial direkt von der klinischen Probe ohne vorherige Isolation genutzt. Das Protokoll

ist ausgerichtet auf virale Genome mit wenig Genomkopien, wie es beim Zika-Virus häufig vorkommt und was eine Genomsequenzierung grundsätzlich erschwert. [61]

Wie u.a. in der Ebola-Epidemie in den Jahren 2013-2016 gesehen wurde, ist es nützlich, die Bewältigung von Virusausbrüchen durch eine genomische Echtzeit-Überwachung zu unterstützen sowie die Transmissions-, Ausbreitungs- und Anpassungsmechanismen zu verstehen. [62, 63] Eine Ganzgenom-Sequenzierung von Ebola direkt aus klinischen Proben war dabei aufgrund der extrem hohen Viruslast während einer akuten Infektion möglich. [64] Ganzgenomische Ansätze zur Sequenzierung des Zika-Virus zeigten sich zunächst aufgrund der geringeren Anzahl an Genomkopien als problematisch und ergaben eine geringe Genomabdeckung. Deshalb soll eine Amplifizierung durch eine PCR eine Anreicherung der Nukleinsäure ergeben. Dazu wurde ein sog. *Tiling*-Amplicon-Schema mit kurzen Amplicons verwendet, da die Wahrscheinlichkeit, dass lange Fragmente in der Probe vorhanden sind, mit geringerer Virusdichte sinkt. Die Verwendung kürzerer Amplicons erzeugt eine größere Produktanzahl, über deren bioinformatische Zusammenfügung ein Zielgenom erzeugt werden kann. Um hohe Kosten für Zeit und Verbrauchsmaterialien sowie das Risiko von Kreuzkontaminationen durch einzelne Reaktionen zu verringern, wurde eine Multiplex PCR entwickelt. [61]

Das Protokoll von J. Quick wurde ab dem Frühjahr 2020 auch mit abgewandelten Primern in der Covid-19-Pandemie zur Diagnostik des *severe acute respiratory syndrome coronavirus 2* (SARS-CoV2) verwendet. Hierbei zeigte sich eine Ganzgenom-Sequenzierung als nützlich zur Entdeckung neuer Varianten sowie epidemiologischen Analyse und Kontaktnachverfolgung. [65]

1.10 HCV Ganzgenom-PCR

Eine Ganzgenom-PCR und -Sequenzierung ermöglicht die Aufschlüsselung der gesamten Nukleotidabfolge des Virus. Das bringt zum einen Vorteile bei der Identifikation von Mutationen, die eine Resistenz des Virus bedingen können, zum anderen können mit Hilfe von phylogenetischer Datenanalyse Übertragungswege besser verstanden werden.

Geringe Viruslasten, verminderte Probenqualität oder eine Vorbehandlung der Patienten stellen Herausforderungen für die Sequenzierung von RNA-Viren dar. [66] Für das Hepatitis C Virus erschwert die hohe Variabilität der Virussequenz zusätzlich die Entwicklung einer solchen Methode, weshalb in der Literatur bereits verschiedene Lösungsansätze beschrieben wurden.

Die Arbeitsgruppe von F. Manso etablierte eine Genotyp-unabhängige Ganzgenomsequenzierung, die mit ca. 120 bp langen Hybridisierungssonden die Sequenz erkennt und dabei bis zu 20 % Nichtübereinstimmungen toleriert. [67] Mit diesem Test (engl. *assay*) wurde der Genotyp sowie die antivirale Resistenz von 72 klinischen Proben verschiedener Genotypen bestimmt. Nach Bindung von HCV-spezifischen biotinylierten Oligonukleotid-Sonden an die HCV-DNA durch Hybridisierung und Inkubation, wurde eine Sequenzierung mit einem MiSeq™ -Gerät von Illumina durchgeführt. Im anschließenden bioinformatischen Prozess wurden die im Text basierten FASTq-Format gespeicherten Daten von Inhalten des menschlichen Genoms befreit und gegen ein HCV-Referenzgenom kartiert. Des Weiteren wurde der Algorithmus VICUNA sowie ein hausinternes C++ Programm (QuasiBAM) zur Bildung einer Consensus Sequenz genutzt. Zur Interpretation des Genotyps und der Resistenzen wurde Geno2Pheno verwendet.

Eine 2022 veröffentlichte Übersichtsarbeit von C. Munyuza stellt auf Anreicherung durch Hybridisierungssonden basierende Methoden der Sequenzierung von HIV und HCV dar. [68] Als ein Vorteil dieser Methoden wird die erfolgreiche Sequenzierung von Proben mit niedrigen Viruslasten oder durch Therapie oder Lagerung beeinträchtigten Proben genannt.

Ein anderer Ansatz für eine nahezu vollständige Ganzgenomsequenzierung wurde 2016 von der Arbeitsgruppe von R.A. Bull beschrieben. [69] Die Herangehensweise kennzeichnet sich durch eine verschachtelte (engl. *nested*) PCR, die ein sehr langes und ein kurzes PCR-Produkt erzeugt. Dabei werden zwei PCR-Reaktionen hintereinander durchgeführt, wobei das Primerpaar der zweiten Reaktion innerhalb eines PCR-Produktes der ersten Reaktion bindet. Dies erhöht die Spezifität der PCR. Im Rahmen der Studie wurden 122 HCV-Proben des GT 1-6 untersucht. Nach der RNA-Extraktion und reverser Transkription erfolgt die nested PCR. Dabei wird ein 9206 bp langes, nahezu Ganzgenomfragment und ein kürzeres, 4kb langes Fragment am 5‘-Ende des Genoms gebildet. Von beiden PCR-Schritten sind die Vorwärts-Primer Genotyp-unabhängig und die Rückwärts-Primer unterschiedlich für jeden Genotyp. Die Sequenzierung erfolgt nach Aufreinigung des PCR-Produktes mit zwei verschiedenen Methoden. Eine Methode stellt die MiSeq-Illumina-Sequenzierung dar, mit der 2x 300 bp lange Sequenzierstücke gebildet wurden. Diese wurden mit dem Algorithmus Bowtie 2 der Software Geneious kartiert und daraus eine Consensus Sequenz gebildet. Gemischte Infektionen wurde durch das Auftreten von mehreren überlappenden DNA-Stücken derselben genetischen Quelle (engl. *contigs*) mit zwei oder mehr Referenzgenomen identifiziert. In diesem Fall wurde eine Consensus Sequenz für jeden der contigs erstellt. Die andere

Sequenziermethode stellte das PacBio SMRT sequencing dar, die ausgewählt wurde, weil sie > 9 kb große Sequenzierstücke bildet. Die bioinformatische Auswertung erfolgte nach Korrektur durch eine Haplotyp-Rekonstruktionsanalyse mit SMRT Analysis v2.3 sowie dem Algorithmus VICUNA, sodass auch eine Consensus Sequenz gebildet werden konnte.

Eine Arbeitsgruppe von H. R. Lapointe beschreibt 2021 ebenfalls eine nested-PCR-Methode zur Genotyp-unabhängigen nahezu Ganzgenom-Sequenzierung. [70] Für die Studie wurden Proben von Therapie-naiven als auch mit DAA therapierten HCV-Infizierten der GT 1-6 untersucht. Nach der RNA-Extraktion und Reverse-Transkriptase-PCR (RT-PCR) erfolgte die nested PCR. Das lange „WG amplicon“ deckt dabei die Abschnitte Core bis zum Kodon 336 des NS5B ab. Überlappend befindet sich dahinter auf dem Genom das „MiDi amplicon“, welches sich vom Kodon 228 bis zum 3‘-Ende von NS5B erstreckt. Es wurden einzelne Sequenzier-Bibliotheken (engl. *libraries*) für die Fragmente hergestellt und dann in einem 7:1 Verhältnis nach der geschätzten Amplicon-Länge kombiniert. Die Sequenzierung erfolgt mit einem MiSeq™-Gerät und die Datenauswertung mit Hilfe eines hausinternen Arbeitsablaufs „MiCall“, in dem die Daten auf Referenzgenome kartiert und eine Consensus Sequenz gebildet wurde.

2 Ziele dieser Arbeit

Zwischen 2004 und 2011 zeigt sich ein Rückgang der jährlichen, an das Robert Koch-Institut (RKI) übermittelten HCV-Erstdiagnosen in Deutschland, der anschließend allerdings stagnierte. [71] Dies steht im Widerspruch zu den sehr hohen Heilungsraten der aktuellen Therapie, wodurch sich die Frage stellt, wo, wie und durch wen sich die Menschen weiterhin mit HCV infizieren. Für eine Elimination der Hepatitis C Virus -Infektion, die sich die WHO bis zum Jahr 2030 zum Ziel gesetzt hat [1], ist es notwendig diese Frage zu klären, um präventive und therapeutische Maßnahmen präzise einsetzen zu können.

Wie schon von der Bekämpfung des Ebola-, Zika- oder SARS-CoV2-Virus bekannt, kann eine Kontaktnachverfolgung anhand phylogenetischer Analysen Aufschluss über mögliche Risikogruppen und das Infektionsgeschehen geben. Zu diesem Zweck soll eine Ganzgenom-PCR entwickelt werden, die das bisherige Spektrum an Diagnostikmethoden von Hepatitis C des Instituts für Virologie am Universitätsklinik Düsseldorf ergänzt.

Zusätzlich zur epidemiologischen Anwendung kann eine Ganzgenom-PCR zur Resistenztestung der DAAs eingesetzt werden. Obwohl es aufgrund der hohen Heilungsraten der DAA nur bei wenigen der optimal behandelten Patienten (2-5 %) zum Therapieversagen kommt, ist die absolute Zahl von Patienten, die weltweit eine erneute Behandlung benötigen, dennoch erheblich. [72] Das schnelle Auftreten Resistenz-assoziierter Substitutionen (RAS), z.B. im Zusammenhang mit unvollständigem Medikamentendruck aufgrund Inadhärenz des Patienten, hat Auswirkungen auf das Therapieansprechen sowie -versagen. [73] RAS in NS3, NS5A sowie NS5B wurden mit einer geringeren Anfälligkeit für DAAs in Verbindung gebracht, trotzdem gibt es aufgrund der hohen genetischen Variabilität von HCV bisher keine standardisierte Methode zum Nachweis von DAA-RAS. [74] Diese soll mit einer Ganzgenom-PCR etabliert werden.

Im Gegensatz zu den in 1.10 bereits beschriebenen Ganzgenom-PCR-Methoden soll im Rahmen dieser Dissertation der Ansatz einer sog. *Tiling-PCR* verfolgt werden. Dabei wird das Genom in vielen kleinen Abschnitten amplifiziert sowie sequenziert und anschließend bioinformatisch wieder zusammengesetzt werden.

Durch Entwicklung dieser Methode im Rahmen dieser Dissertation soll eine Grundlage für die Einrichtung einer integrierten genomischen Surveillance von HCV geschaffen werden, womit ein besseres Verständnis der Hepatitis-C-Epidemiologie erreicht werden kann.

3 Material und Methoden

3.1 Material

3.1.1 Geräte

Automatisiertes Nukleinsäure-Reinigungssystem	EZ1 Advanced XL, Qiagen, Deutschland EZ1 Advanced Virus Card v2.0, Qiagen, Deutschland
Gefrierschrank (-80 °C)	Ultra- Low Temperature Freezer, New Brunswick, Eppendorf, Deutschland
Gefrierschrank (-20 °C)	Bosch, Deutschland & Liebherr, Deutschland
Gel-Dokumentationsgerät	Intas Gel iX Imager, Intas Science Imaging Instruments GmbH, Deutschland
Heizblock	AkkuBlock™ Digital Dry Bath, Labnet International, Inc., USA Thermostat Plus, Eppendorf, Deutschland
Hula-Mixer Sample Mixer	Invitrogen by Thermo Fisher Scientific, USA
Kühlschrank (4 °C)	Bosch, Deutschland
Luminometer	Tristar 2 LB 942 Multimode Reader, Berthold Technologies, Deutschland
Magnetständer	DynaMag™-2, Invitrogen by Thermo Fisher Scientific, USA
Magnetplatte	96-Well Microtiter Plate Magnetic Separation Rack, New England Biolabs® Inc., USA
Mikrowelle	Küppersbusch, Österreich
MinION	Oxford Nanopore Technologies, England
Präzisionswaage	Mettler Toledo, Deutschland
Spannungsquelle	Biometra Power Pack P25 / P25T, Analytik Jena, Deutschland
Thermocycler	Biometra Trio, Analytik Jena, Deutschland C1000 Touch™, Bio-Rad Laboratories, USA

Vortex-Schüttler	Ministar, VWR, USA
	Press to mix 34524, Snijders, Niederlande
	VV3, VWR, USA
Zentrifugen	Centrifuge 5415 D, Eppendorf, Deutschland
	Centrifuge 5427 R, Eppendorf, Deutschland
	Color Sprout Mini-Zentrifuge, Biozym Scientific GmbH, Deutschland
	PCV-2400, Grant Instruments LTD, Großbritannien

3.1.2 Verbrauchsmaterialien

Desinfektionsmittel	Softa-man acute, B. Braun Melsungen AG, Deutschland
	Bode Sterilium, Bode Chemie Hamburg
Desinfektionstücher	Schülke wipes Spendereimer, Schülke & Mayr GmbH, Deutschland
Elektrophoresekammer	Biometra Compact Multi-Wide, Analytik Jena, Deutschland
Flow Cell (R9.4.1) FLO-MIN 106D	Oxford Nanopore Technologies, England
MinION	Oxford Nanopore Technologies, England
Pipetten	Eppendorf® Research® plus Pipette / Reference® 2 Pipette, Eppendorf, Deutschland
Pipettenspitzen	Starlab, Deutschland
Präzisionstücher	Kimtech Science, Kimberly-Clark Professional Tapira Plus, Großverbraucherspezialisten eG, Deutschland
Reagenzkartuschen	EZ1®, Virus Mini Kit v2.0, Qiagen, Deutschland
Reaktionsgefäße	96 Well Black Plate, Thermo Scientific™ by Thermo Fisher Scientific, USA
	Corning™ Costar™ sterile Einweg-Reagenz Reservoir (50 ml), Corning, USA
	DNA Low Binding 1,5 ml Reagiergefäß, Sarstedt AG&Co.KG, Deutschland

	Hard-Shell PCR-Plate-96, Perkin Elmer, Deutschland
	SafeSeal Reagiergefäß 1,5 ml, Sarstedt AG&Co.KG, Deutschland
	PCR Tube Stripe 0,2 ml, Eppendorf, Deutschland
Verschlussfolie für Platten	Microseal®'B'seal, Bio-Rad Laboratorien, UK

3.1.3 Chemikalien und Enzyme

0,1 M Dithiothreitol (DTT)	Invitrogen by Thermo Fisher Scientific, USA
5x SSIV first strand Buffer	Invitrogen by Thermo Fisher Scientific, USA
20 % Glycerol-Lösung	Carl Roth GmbH + Co. KG, Deutschland
Adapter Mix II	Oxford Nanopore Technologies, England
Agarose	Biozym Scientific GmbH, Deutschland
Blunt/TA Ligase Master Mix	New England Biolabs® Inc., USA
Bromphenol-Blau	Merck KGaA, Deutschland
dNTP Mix, PCR Grade (10 mM)	Qiagen, Deutschland
Elution Buffer	Oxford Nanopore Technologies, England
Ethanol	Merck KGaA, Deutschland
Ethidiumbromid 0,025 %	Carl Roth GmbH + Co. KG, Deutschland
Flush Buffer	Oxford Nanopore Technologies, England
Flush Tether	Oxford Nanopore Technologies, England
Größenmarker Gelelektrophorese	Gene Ruler DNA Ladder Mix, Thermo Fisher Scientific, USA
H_2O (Nuclease-free)	Applied biosystems by Thermo Fisher Scientific, USA
	UltraPure™ Distilled Water, Invitrogen by Thermo Fisher Scientific, USA
Loading Beads	Oxford Nanopore Technologies, England
Native Barcoding Expansion 96 EXP-NBD 196	Oxford Nanopore Technologies, England
NEBNext® Quick Ligation Reaction Buffer	New England Biolabs® Inc., USA

NEBNext® Quick T4 DNA Ligase	New England Biolabs® Inc., USA
NEBNext® Sample Purification Beads	New England Biolabs® Inc., USA
NEBNext® Ultra II End Prep Enzyme Mix	New England Biolabs® Inc., USA
NEBNext® Ultra II End Prep Reaction Buffer	New England Biolabs® Inc., USA
RNase Inhibitor, Murine / Human Placenta	New England Biolabs® Inc., USA
Sequencing Buffer	Oxford Nanopore Technologies, England
SuperScript™ IV Reverse Transkriptase (200 U μl)	Invitrogen by Thermo Fisher Scientific, USA
Q5® Hot Start High-Fidelity DNA-Polymerase (2000 U/ml)	New England Biolabs® Inc., USA
Q5® Reaction Buffer	New England Biolabs® Inc., USA
Quant-iT™ 1x ds DNA Hs Working Solution	Invitrogen by Thermo Fisher Scientific, USA
S Fragment Buffer	Oxford Nanopore Technologies, England
TRIS-Borat-EDTA (TBE) - Puffer	Sigma-Aldrich, USA

3.1.4 Kits

Flow Cell Priming Kit EXP-FLP002	Oxford Nanopore Technologies, England
Ligation Sequencing Kit SQK-LSK109	Oxford Nanopore Technologies, England
Native Barcoding Expansion Kit EXP-NBD196	Oxford Nanopore Technologies, England
NEBNExt® ARTIC SARS-CoV-2 Companion Kit (Oxford Nanopore Technologies®)	New England Biolabs® Inc., USA
Q5® Hot Start High-Fidelity DNA-Polymerase (2.000 U/ml)	New England Biolabs® Inc., USA
Sequencing Auxiliary Vials EXP-AUX001	Oxford Nanopore Technologies, England
SFB Expansion Kit EXP-SFB001	Oxford Nanopore Technologies, England
SuperScript™ IV Reverse Transcriptase (200 U/μl)	Invitrogen by Thermo Fisher Scientific, USA

3.1.5 Oligonukleotide

Für die Reverse Transkription wurden Random Hexamere sowie der Oligo-d(A)-Primer verwendet. Die Random Hexamere wurden von invitrogen by Thermo Fisher Scientific erworben, in einer Konzentration von 50 μM geliefert und bestehen aus Abschnitten mit je sechs zufälligen Nukleotiden. Der Primer Oligo-d(A) wird ebenfalls in einer Konzentration von 50 μM verwendet, aber von der Firma Eurofins Scientific bezogen. Es handelt sich um einen Rückwärts-Primer aus Adenin (A) mit der 5'→3'-Sequenz AAAAAAAAAAAAAAAAAA, der an den hochkonservierten Polyuridin-Schwanz am 3'-Ende der HCV-RNA anknüpft. Er ist in der Literatur für die Genotypen 1-7 beschrieben. [75]

Alle weiteren Oligonukleotide der Tiling-PCR wurden in einer Sequenz von 5'→ 3' von der Firma Eurofins Scientific bezogen. Zur erleichterten Übersicht wurden die Genotyp-abhängigen Primersequenzen im Anhang angefügt (Vgl. Anhang 1.1, 1.2, 1.3, 1.4).

3.1.6 Referenzsequenzen

Zur Bildung des Primersets sowie zur Analyse für mögliche Primerkorrekturen wurden für jeden Genotyp unterschiedliche Referenzisolate verwendet. Die entsprechenden *GenBank Accession Numbers* sind im Anhang angefügt (Vgl. Anhang 2.1, 2.2, 2.3, 2.4).
Zum Kartieren der Sequenzierdaten in Geneious Prime® wurde ebenfalls eine Referenzsequenz benötigt. Deren *GenBank Accession Numbers* sind im Folgenden aufgelistet.

Genotyp 1a	M67463, AF009606 (HCV-H77)
Genotyp 1b	EU781827
Genotyp 3a	X76918
Genotyp 4a	Y11604
Genotyp 4d	DQ418786

3.1.7 Verwendete Software

Geneious Prime® 10.2.6 und 11.0.12 +7	Biomatters ltd., Neuseeland
GraphPad Prism Version 8.0.2	GraphPad Software, Inc., California
IBM® SPSS® Statistics Version 28.0.1.0	International Business Machines Corporation, USA
Intas GDS	Intas Science Imaging Instruments GmbH, Deutschland
MinKNOW Version 21.10.4	Oxford Nanopore Technologies, England
MicroWin 2000, Version 4.41	Mikrotek Laborsysteme GmbH, Overath, Deutschland

3.2 Patientenproben

Zur Etablierung der Methode wurden rückgestellte Proben der Studie „Die Bedeutung von Resistenzmutationen gegenüber Protease Inhibitoren bei HCV-positiven Patienten“ (PEPSI) verwendet. Im Rahmen dieser Studie wurde von HCV-infizierten Patienten 10 ml Blut während der routinemäßigen Blutabnahme gewonnen und anschließend das Plasma abgetrennt, welches die benötigte, nicht aufgereinigte Virus-RNA enthielt.

Der Verarbeitung des Patientenmaterials für wissenschaftliche Zwecke haben alle Studienteilnehmer*innen schriftlich zugestimmt. Gemäß der Deklaration von Helsinki wurde der PEPSI-Studie durch den Ethikausschuss der Medizinisches Fakultät der Heinrich-Heine-Universität Düsseldorf eine Zulassung zur Probennutzung mit der Studiennummer AKN 2012048 erteilt.

Des Weiteren wurden Rückstellproben von zuvor in Essen gelagerten pseudonymisierten Proben der *injection-drug-use*- (IDU) Kohorte Düsseldorf verwendet. Dazu liegt ebenfalls eine gültige Ethikgenehmigung mit dem Aktenzeichen 13-4543-BO vor.

3.3 Methoden

3.3.1 Probenaufbereitung

Die Aufbereitung der Patientenproben erfolgt durch Nukleinsäureextraktion aus freien Viren mit Hilfe des Automaten „EZ1 Advanced XL“ von Qiagen®. Die Methode kombiniert eine silikabasierte Nukleinsäureextraktion mit der Anwendung von magnetischen Partikeln.
Das Protokoll entspricht den Angaben des Herstellers Qiagen®. Dabei sollten die aufzubereitenden Proben innerhalb von 24h nach der Blutentnahme verarbeitet werden und die Probenmenge mindestens 200 μl betragen, bei einem geringeren Volumen sollte mit Negativ-Plasma aufgefüllt werden. Zur vereinfachten Durchführung der Aufreinigungsprozedur wurden vorprogrammierte Protokolle von der „EZ1 Advanced Virus Card v2.0“ Qiagen® genutzt. Ebenfalls von Qiagen® beinhaltet das „EZ1 Virus Mini Kit v2.0“ Reagenzkartuschen mit den benötigten Komponenten, wie z.B. Lysepuffer, Waschpuffer und RNAse-freie Elutionspuffer. Der Einsatz von Proteinkinase a sorgte für eine Lyse viraler Partikel aus Körperflüssigkeiten sowie für eine Inaktivierung von RNAsen. Eine höhere Ausbeute an RNA wurde durch die Verwendung von carrier-RNA’s erzeugt. Eine carrier-RNA besteht aus ca. 20-30 Thymin-Nukleinbasen, die durch eigene Bindung an die Filtersäule RNAsen von dem Abbau viraler RNA ablenken.
Der Automat „EZ1 Advanced XL“ ermöglicht die Extraktion viraler Nukleinsäuren in verschiedenen Volumina. Die im Rahmen dieser Dissertationsarbeit verwendeten Proben wurden zu einem Volumen von 60 μl extrahiert und anschließend bei – 20°C eingelagert.

3.3.2 Reverse Transkription von HCV RNA

Die Methode zur Umwandlung viraler RNA in cDNA erfolgte mittels der Reversen Transkriptase SuperScript™ IV von Invitrogen. Das Vorgehen entspricht einer Weiterentwicklung des Protokolls zur Genotyp-unabhängigen, Voll-Genom Reversen Transkription (RT) von Walker et al. [76] Der Prozess erfolgte in zwei Schritten, die im Folgenden „RT-1“ und „RT-2-long“ benannt werden. Für jeden Schritt wurde ein Reagenz-Mix angesetzt, dessen Volumen je nach Anzahl der durchzuführenden Proben variierte.
Der „RT-Mix-1“ besteht aus Random Hexameren (50 μM), Oligo-d(A)-Primer (50 μM) sowie dNTP’s (10 mM each). Nach Zugabe von 10 μl RNA-Eluat wurde das Gemisch 5 Minuten (Min) bei 65 °C im Thermocycler erhitzt und anschließend wieder auf 25 °C abgekühlt. Dieser Schritt „RT-1“ dient der Aufschmelzung von Sekundärstrukturen.

Im Anschluss wurden zügig 7 μl des „RT-Mix-2“ zugefügt. Dieser besteht aus 5x SSIV first strand Buffer, DTT (0,1 M), RNAse Inhibitor sowie dem Superscript IV RT Enzym. Im anschließenden Schritt „RT-2-long“ erfolgte die Reverse Transkription durch Erwärmung für 10 Min auf 25 °C, für 60 Min auf 42 °C, für 30 Min auf 50 °C und für 30 Min auf 55 °C im Thermocycler. Zuletzt wurde die Lösung zur Denaturierung der übrigen RNA sowie dem Superscript™ IV Reverse Transkriptase Enzym auf 75 °C erhitzt.

Bei einer anschließenden Abkühlung auf 4 °C konnte die cDNA kurzfristig stabil gehalten werden. Sollte eine längere Aufbewahrung erfolgen, wurde das RT-Produkt bei – 20 °C tiefgefroren.

Tabelle 1: Protokoll der Reversen Transkription
Tabellarische Zusammenfassung der Reagenzien und Thermocycler-Programme der Reserven Transkription.

RT Mix 1	**RT 1**	**RT Mix 2**	**RT 2 long**
1 μl Random Hexamere (50 μM)	Volumen: 13 μl	4 μl 5x SSIV first strand Buffer	Volumen: 20 μl
1 μl Oligo-d(A) (50 μM)	5 Min bei 65 °C	1 μl DTT (0,1 M each)	10 Min bei 25 °C
1 μl dNTP‘s (10 μM each)	2 Min bei 25 °C	1 μl RNAse-Inhibitor	60 Min bei 42 °C
10 μl RNA		1 μl Superscript™ IV RT Enzym (200 U/ μl)	30 Min bei 50 °C
		13 μl RT Mix 1	30 Min bei 55 °C
			15 Min bei 75 °C
			∞ bei 4 °C

3.3.3 Tiling - PCR

Das Protokoll ist angelehnt an die Multiplex PCR – Methode für die MinION- und Illumina-Sequenzierung von J. Quick. [61] Mit Hilfe eines Primer-Design-Tools namens Primal Scheme (https://primalscheme.com) wurden Primerpaare mit einer Amplicon-Länge von 400-600 bp erstellt, deren Produkte sich überlappen. Die kurze Amplicon-Länge erhöht die Wahrscheinlichkeit ein kontinuierliches PCR- und Sequenzierprodukt zu erhalten, welches bioinformatisch zusammengesetzt wurde. Um zu verhindern, dass sich durch die Überlappung unerwünschte Primerpaare zwischen den beieinanderliegenden Forward- und Reverse-Primern bilden, wurden die Primer in zwei Pools (Genotyp 3a, 4a und 4d) bzw. drei Pools (Genotyp 1a und 1b) aufgeteilt. Durch solche unerwünschten Primerpaare würde ein kürzeres,

überamplifiziertes Amplicon entstehen, was die Amplifikation des gesamten Genoms behindert. Die einzelnen Primer-Pools wurden als 100 μM Stock-Solution zusammen pipettiert und anschließend auf 10 μM Arbeits-Solution verdünnt.

Als Ausgangsmaterial für die Tiling-PCR dienten pro Probe 2,5 μl der synthetisierten cDNA aus der Reversen Transkription. Diese wurden in einen neuen Ansatz von 22,5 μl überführt, der das Q5® Hot Start High-Fidelity DNA Polymerase Kit von New England Biolabs® nutzt. Die beschriebene Polymerase zeichnet sich laut dem Hersteller durch besonders hohe Genauigkeit mit höchster Zuverlässigkeit aus und ist ca. 280x genauer als die Taq DNA Polymerase. [77] Der verwendete PCR-Ansatz setzt sich aus den folgenden Reagenzien zusammen: 5 μl 5x Q5® Reaction Buffer, 0,25 μl Q5® Hot Start High-Fidelity DNA Polymerase, 0,5 μl dNTPs (10 mM), Arbeits-Solution eines Primer-Pools mit einer Ziel-Konzentration von 0,015 μM sowie Nuklease-freiem Wasser, dessen Volumen je nach Primeranzahl im jeweiligen Pool variiert.

Das PCR-Programm kennzeichnet sich durch die Kombination des Annealing- und Elongationsschritts und wird unter folgenden Zyklusbedingungen durchgeführt:

Tabelle 2: Programm der Tiling-Polymerasekettenreaktion (PCR)
Darstellung der Temperaturen und Zyklen einer Tiling-PCR zur Anwendung im Thermocycler.

Temperatur [°C]	Zeit [Min]	Zyklus	Funktion
98.0	00:30	1 x	DNA-Denaturierung, Polymerase-Aktivierung
98.0	00:15	40x	DNA-Denaturierung
63.0	00:10		Primer Hybridisierung, Verlängerung
4.0	∞	-	Kühlung + Lagerung

3.3.4 Agarose-Gelelektrophorese

Die Agarose-Gelelektrophorese ist eine Methode zur Auftrennung von Nukleinsäuren nach der Größe. Eine alternative Kapillarelektrophorese kann zur Detektion der PCR-Produkte einer Tiling-PCR nicht angewandt werden, da es sich um zu kleine Fragmente handelt, die nicht adäquat aufgetrennt werden können.

Zur Herstellung eines 1 %-igen Agarosegels wurden 1,6 g Agarosepulver in 160 ml 1x TBE-Puffer in der Mikrowelle zum Sieden gebracht. Durch diesen Prozess löst sich die Agarose und

es bildet sich eine klare Flüssigkeit. Nach Abkühlung der Lösung auf Handwärme wurden 9 Tropfen Ethidiumbromid zugefügt und das Gemisch in eine Gelelektrophoresewanne gegossen, die fixierte Gelkämme zum Erhalt von Geltaschen enthielt. Die Aushärtungsphase beträgt in der Kühlkammer bei -4 °C ca. 15 Min. Im Anschluss konnte das erstarrte Gel mit 1x TBE-Puffer vollständig begossen und die Kämme zur Bildung der Gel-Taschen entfernt werden. Die erste Geltasche wurde mit 10 μl Größenmarker „Gene Ruler Ladder Mix“ gefüllt, die folgenden Geltaschen erhielten 10 μl PCR-Produkt, welches zuvor mit Blaupuffer vermischt wurde. Der Blaupuffer wurde aus 20 %iger Glycerol-Lösung, 10x TBE-Puffer sowie einer Spatelspitze Bromphenolblau hergestellt.

Die DNA wurde anschließend bei 400 mA und 120 V für 35 Min aufgetrennt und in einer Gel-Dokumentationskammer (von Intas Science Imaging Instruments GmbH) mittels UV-Strahlung sichtbar gemacht. Mit Hilfe des Intas Geldokumentationssystems (GDS) konnten die Ergebnisse bildlich dargestellt werden. Eine Beurteilung der vorliegenden Nukleinsäuren ist durch den eingesetzten Größenmarker möglich.

3.3.5 DNA-Quantifizierung

Die DNA-Quantifizierung von PCR-Produkten erfolgte mit Hilfe des Centro LB 963 Lumineszenz Mikroplatten Readers von Berthold. Ein Luminometer misst schwache Emissionen von sichtbarem Licht aus einer Probe mit Hilfe eines Photomultipliers. Nach Angaben des Herstellers emittieren die Proben dabei Licht aufgrund einer chemischen Reaktion und benötigen keine Anregungslichtquelle. [78] Zur Messung wurde jeweils 1 μl PCR-Produkt mit 99 μl Quant-iT 1x ds DNA Hs Working Solution auf einer lichtunempfindlichen Platte vermischt und dies mit einer Standard-Reihe verglichen, die aus 8 x 10μl Standard in unterschiedlichen Konzentrationen mit jeweils 90μl Quant-iT 1x ds DNA Hs Working Solution besteht.

Nach Einsetzen der Platte in den Tristar 2 Multimode Reader B942 wurde sie mit dem Programm „Quant-it ds DNA Meassurement v4“ der Software MikroWin für 30 Sekunden (Sek) geschüttelt, für 2 Min inkubiert und anschließend gemessen. Die DNA-Konzentration konnte dann mit Hilfe der Standardreihe berechnet werden.

3.3.6 MinION - Sequenzierung

Die MinION-Sequenzierung ist eine Methode zur DNA-Sequenzierung. Sie wurde von *Oxford Nanopore Technologies* entwickelt und stellt eine Form des „*Next Generation Sequencing*“ (NGS) dar. Das Prinzip beruht auf der Messung einer Spannungsveränderung beim Durchlaufen der Nukleinsäure durch eine Pore in Nanometergröße, wodurch eine

DNA-Sequenz-Bestimmung in Echtzeit möglich wird. Es können dabei Fragmente von bis zu 4 Mb am Stück sequenziert werden und die Methode arbeitet mit 400bp/sek deutlich schneller als andere NGS-Methoden. [79] Nachteilig ist eine relativ hohe Fehlerquote [80], die aber, wie bei anderen Hochdurchsatz-Sequenziermethoden, durch die große Anzahl sequenzierter Moleküle ausgeglichen werden kann. [81, 82]

Die Sequenzierung erfolgt in Fließzellen (engl. *Flow Cells*), in denen die Nanoporen untergebracht sind. Die hochstabilen heptameren Protein-Nanoporen haben einen Innendurchmesser von 1 nm und sind in eine elektrisch-resistente Polymermembran eingesetzt. Sog. „*Microscaffolds*“, daher kleine Gerüstverbindungen, halten die zahlreichen Nanoporen innerhalb einer Flow Cell stabil. Jede Microscaffold ist mit einer Elektrode zum Sensor-Array-Chip verbunden. Durch anwendungsspezifische integrierte Schaltungen können mehrere Nanopore-Experimente parallel innerhalb einer Flow Cell durchgeführt werden.

Der MinION ist ein tragbarer Sequenzer in Taschenformat, wodurch eine Analyse überall möglich ist und nah am Patienten stattfinden kann. Der MinION kann einen theoretischen Output von 50 GB an Daten für die Sequenzierung von DNA, cDNA oder nativer RNA erzeugen. Die prozessive Bewegung der Basen durch die Pore führt zu einer kontinuierlichen Veränderung des Stroms, die als „*Squiggle*“ bezeichnet wird. Die MinKNOW^{TM}-Software wandelt das Squiggle in Echtzeit in Sequenzierstücke (engl. *reads*) um, wobei jeder read einem einzelnen DNA-/RNA-Strang entspricht und die Rohdaten in Dateien des FAST5-Formats übersetzt werden. Mit Hilfe von Basecalling-Algorithmen wird das gespeicherte Signal in ein FASTq-Format decodiert, sodass die Basensequenz ersichtlich wird.

Die Durchführung der MinION-Sequenzierung richtet sich mit wenigen Abwandlungen nach dem NEWNext ARTIC SARS-CoV-2 Companion Kit E7660 Instruction manual [83] sowie dem Amplicon barcoding with Native Barcoding Expansion 96 (EXP-NBD196) – Protokoll von Oxford Nanopore [84] und gliedert sich in acht Schritte.

Zur Aufreinigung der Proben wurde ein Volumen von mind. 50 μl benötigt, sollte das PCR-Produkt weniger Volumen betragen, wurde dies mit Nuklease-freiem Wasser aufgefüllt. Zum PCR-Produkt wurden im Verhältnis 1 : 1,4, daher 70 μl, NEBNext® Sample Purification Beads hinzugegeben und bei 22 °C für 5 Min im Hula-Mixer vermischt. Die Beads haben die Funktion, die DNA zu binden, sodass sich der Probenverlust während der Aufreinigung verringert. Anschließend wurden die Beads mit Hilfe einer Magnetplatte am Boden gesammelt und die übrige Flüssigkeit konnte abpipettiert werden. Es folgten zwei Waschschritte mit 200 μl 70 %igen Ethanol nach dem gleichen Prinzip mit Hilfe der Magnetplatte. Eine anschließende

Eluierung mit 25 μl Nuklease-freiem Wasser für 2 Min bei Raumtemperatur sorgte für ein Lösen der DNA von den Beads. Der Überstand ohne Beads konnte nun für den nächsten Schritt verwendet werden. Dieser erste Schritt kann an der Heinrich-Heine-Universität Düsseldorf auch durch einen Pipettierroboter des Biologisch-Medizinischen Forschungszentrum durchgeführt werden.

Im zweiten Schritt folgte nun der sog. *End Prep*, in dem aus den DNA-Fragmenten reparierte DNA mit 5‘ phosphorylierten und 3‘ dA-tailierten Enden entsteht. Hierzu wurden 12,5 μl des aufgereinigten PCR-Produktes mit 1,75 μl NEBNext® Ultra II End Prep Reaction Buffer und 0,75 μl NEBNext® Ultra II End Prep Enzyme Mix versetzt. Anschließend erfolgte eine Inkubation im Thermocycler für 5 Min bei 20 °C und für 5 Min bei 65 °C.

Der dritte Schritt der Barcode Ligation dient der Verhinderung einer Probenverwechslung bei der gleichzeitigen Analyse vieler Proben. Hierzu wurde pro Well 1,5 μl Nativ Barcoding Expansion 96, 10 μl Blunt/TA Ligase Master Mix und 3 μl Nuklease-freies Wasser zu 1,5 μl End-prepped DNA gegeben. Anschließend erfolgte wieder eine Inkubation im Thermocycler, diesmal für 20 Min bei 20 °C und für 10 Min bei 65 °C.

Im vierten Schritt erfolgt das sog. *Pooling*. Das Zielvolumen für jeden Pool betrug 480 μl und es wurden zwei Pools für eine 96-Well Platte angefertigt. Bei einer vollen Platte entsprach dies 10 μl gebarcodeter DNA, die jedem der beiden Pools zugeführt wurde.

Der fünfte Schritt bestand für jeden Pool aus mehreren Waschschritten. Wie bereits bei Schritt Eins erfolgte dies mit Hilfe von Beads, die an die DNA binden. Pro Pool wurden 192 μl NEBNext® Sample Purification Beads hinzugefügt und eine Durchmischung durch Drehungen auf dem Hula Mixer für 5 Min bei Raumtemperatur erzeugt. Anschließend wurden die Beads auf einem Magnetständer am Boden des Probengefäßes gesammelt und die übrige Flüssigkeit konnte abpipettiert werden. Der Ablauf wiederholte sich bei einer zweimaligen Waschung mit 200 μl Short Fragment Buffer (SFB). Der SFB hat die Funktion kleine Fragmente < 3 kbp aus dem PCR-Produkt zu lösen. Danach erfolgte ein Waschschritt mit 200 μl 70 %igen Ethanol sowie eine Elution in 35 μl Nukleasen-freien Wassers für 2 Min bei Raumtemperatur, durch die sich die DNA erneut von den Beads löste.

Der konzentriertere Überstand konnte für den sechsten Schritt, die Adapter Ligation, verwendet werden. In diesem Schritt wird mit Hilfe einer Ligase ein Adapter an die DNA gebunden, sodass die Nanopore sie erkennt und aufnimmt. Hierzu wurden 5 μl Adapter Mix II, 10 μl NEBNext® Quick Ligation Reaction Buffer und 5 μl NEBNext® Quick T4 DNA Ligase

zu den 35 μl gepoolten, gebarcodeten Proben-Pools gegeben. Anschließend wurden beide Pools für 10 Min bei Raumtemperatur inkubiert.

Der siebte Schritt besteht erneut aus mehreren Waschschritten, die nach dem gleichen Prinzip mit Hilfe der Beads und des Magnetständers durchgeführt werden. Nach Zugabe von 20 μl Beads wurde zweimal mit 125 μl SFB gewaschen. Anschließend wurden pro Pool 15 μl Elution Buffer zugegeben und für 5 Min bei Raumtemperatur inkubiert, wodurch sich die DNA von den Beads lösen konnte und der Überstand überführt wurde.

Den letzten Schritt stellt die sog. *Library Prep* dar. Zuvor sollte die Flow Cell mit Hilfe eines Flow Cell Checks mit der Software MinKNOW vorbereitet werden. Außerdem sollte die Flow Cell ca. 15-20 Min vor Zugabe des Proben-Pools mit Puffer befüllt werden, der aus 30 μl Flush Tether und aus 1170 μl Flush Buffer besteht. Von diesem Mix wurden ca. 800 μl in den Puffer-Port der Flow Cell gefüllt, wobei zu beachten war, dass keine Luft in die Flow Cell gelangen durfte. Für die Library Prep wurden pro Pool 37,5 μl Sequencing Buffer und 25,5 μl Loading Beads zu den 15 μl DNA gegeben. Die gesamten ca. 80 μl des Pools konnten nun langsam tropfend auf den Proben-Port einer Flow Cell aufgetragen werden, während beide Ports geöffnet sein sollten.

Im MinKNOW Betriebsfeld konnte anschließend der Sequenzierlauf gestartet werden, sodass die Messung begann.

3.3.7 Illumina - Sequenzierung

Die Illumina-Sequenzierung ist ein weiteres Beispiel für NGS, welche die parallele Sequenzierung von Millionen DNA-Abschnitten in einem Durchlauf ermöglicht.

Der entsprechende Vertragspartner der Heinrich-Heine-Universität ist der Molekularbiologische Service SEQ-IT GmbH & Co KG in Kaiserslautern. Zur Analyse werden 40 μl PCR-Produkt verdünnt mit 5 μl Nuklease-freiem Wasser in PCR-Tubes per Post verschickt und die Ergebnisse innerhalb von 1-2 Wochen zurückerhalten.

Alle Sequenzierungen werden bei SEQ-IT- GmbH & Co KG auf Illumina Sequenzierern bearbeitet. Nach eigenen Angaben der Firma werden bei dabei bis zu 800 Millionen Reads bei einer maximalen Readlänge von 2x150 bp erzeugt. Dies entspricht einer Gesamtleistung von bis zu 120 GB pro Sequenzierlauf. [85]

Die Illumina -Sequenzierung basiert auf dem „*Sequencing by synthesis*“-Prinzip. Dabei wird die Synthese eines komplementären Stranges bei jedem einzelnen Nukleotid verfolgt und durch die Verwendung von fluoreszierenden dNTPs in Echtzeit detektiert.

Zunächst wird in einer Library Preparation die DNA in einheitlich große Stücke fragmentiert, die Enden der Fragmente repariert und anschließend werden durch Ligation oder Tagmentation Adapter an die DNA-Enden angehangen. Diese Adapter stellen zum einen eine Verbindung der Sequenz mit der Flow Cell dar und sind gleichzeitig komplementär zur Primer-Sequenz, um die Sequenzierung zu ermöglich. Zum anderen enthalten sie einen Barcode zur Identifikation der Probe.

Anschließend wird eine sog. Brücken-PCR durchgeführt. Nach Gabe der Probe auf die Fließzelle bindet der eine Adapter am Ende der DNA an die Flow Cell. Der freie Adapter bindet an einen komplementären Primer, der auf der Flow Cell gebunden ist, sodass sich der DNA-Strang als Brücke darstellt. Der Primer dient nun als gebundener Startpunkt für die DNA-Polymerase, die einen Rückwärtsstrang bildet. Nach vielen Zyklen ist das Ergebnis ein Cluster aus DNA-Vorwärts- und -Rückwärts-Strängen des identischen DNA-Fragments. Die Vervielfältigung dient der Signalverstärkung der Sequenzanalyse, da ein einzelnes Signal eines DNA-Moleküls zu schwach wäre, um von einer Kamera detektiert zu werden.

Für die Sequenzierung baut die Polymerase nun spezielle dNTPs ein, die ein fluoreszierendes Label mit unterschiedlichen Wellenlängen enthalten. Dieses Label blockiert die Polymerase, sodass nach dem Einbau eines Nukleotids die Synthese so lange abbricht, bis ein Enzym das Label gespalten hat. In diesem Zeitraum regt ein Laser die fluoreszenten Labels an, sodass auf der Platte das im Cluster eingebaute Nukleotid aufleuchtet. Dies wird durch eine Kamera festgehalten und durch eine Software detektiert. Der Vorgang wird für jedes Nukleotid wiederholt, bis das gesamte Fragment synthetisiert ist. [86]

Ein Vorteil der Illumina-Sequenzierung ist die hohe Genauigkeit der Rohdaten, die vom Hersteller mit 99 % angegeben wird. Diese kann dadurch erreicht werden, dass jedes einzelne Nukleotid des DNA-Stücks analysiert wird. Nachteilig ist, dass es durch diese sehr genaue Methode länger dauert, Ergebnisse zu erhalten und höhere Kosten anfallen.

3.3.8 Bioinformatische Auswertung der Rohdaten der Sequenzierung

Die als FAST5-Datei gespeicherten Rohdaten der Sequenzierung werden mit Hilfe von Basecalling-Algorithmen in eine FASTq-Datei decodiert. Die erhaltenen FASTq-Dateien wurden in die Software Geneious Prime® importiert. Jede Probe wurde zu einer entsprechenden Referenz des gleichen Genotyps kartiert (engl. „*Map to reference*“), sodass die Fragmente in einer gebildeten Zusammenstellung (engl. *assembly*) in die richtige Reihenfolge gebracht wurden. Eine anschließend gebildete Consensus Sequenz zeigte das

Nukleotid an, welches zu mind. 60 % an dieser Position auftrat, sodass die mehrheitliche (≥ 60 %) Nukleotidabfolge dargestellt wird. Des Weiteren wurde definiert, dass die Consensus Sequenz ein N enthält, wenn an dieser Nukleotidposition ≤ 30 Reads vorlagen. Der sequenzierte Genomanteil zeigt anschließend die Anzahl der Ns in der gesamten Consensus Sequenz an und dient der Qualitätskontrolle. Ein Kartieren der Sequenzier-reads zu einer Referenz mit den verwendeten Primern bildete ein erneutes Assembly. Davon ließ sich die Anzahl der Reads pro Nukleotidposition graphisch darstellen. Diese Häufigkeit, mit der ein Nukleotidabschnitt sequenziert wurde, wird im Folgenden als Abdeckung des Genoms (engl. *Coverage*) bezeichnet und wurde für die Analyse des Erfolgs der Methode angewendet.

Zum Vergleich der Sequenzierergebnisse der Nanopore zur Sanger-Sequenzierung wurde aus den Ausrichtungen (engl. *alignements*) ein phylogenetischer Baum gebildet. Dies erfolgte ebenfalls in der Software Geneious Prime® mit den Einstellungen des Gendistanzmodells Tamura-Nei und der Baumbildungsmethode Neighbor-Joining. Anschließend wurden beispielhaft die zusammen gehörigen Proben hinsichtlich ihrer Unterschiede in der Nukleotidabfolge miteinander verglichen. Genauso wurde bei dem Vergleich zur Illumina-Sequenzierung verfahren.

Zur Untersuchung der phylogenetischen Zusammenhänge innerhalb der Düsseldorfer IDU-Kohorte wurde mit den gleichen Einstellungen mit Geneious Prime® ein weiterer phylogenetischer Baum erstellt. Diese Daten wurden zur Visualisierung mit dem Online Tool iTOL3 (https://itol.embl.de) in einer Grafik verarbeitet.

3.3.9 Resistenztestung

Die Ermittlung von Resistenzen innerhalb des HCV-Genoms erfolgte mit Hilfe eines online-basierten Interpretationssystems namens Geno2Pheno$_{[HCV]}$ der Gesellschaft für nachhaltige Forschung e.V. aus Deutschland (https://hcv.geno2pheno.org/). Die Software identifiziert anhand der eingefügten Sequenzdaten die genomischen Regionen, generiert Prognosen für den HCV-Geno- und Subtyp und erstellt einen Arzneimittel-Vorhersagebericht mit Resistenzen zu den einzelnen DAAs. [87]

3.3.10 Statistische Auswertung

Die statistische Auswertung der Daten erfolgt mit Hilfe der Software SPSS Statistics. Für jeden Genotyp wurde eine eigenständige Datenbank angelegt, worin zunächst die metrischen Variablen Viruslast, Anzahl Ns in der Consensus Sequenz als Maß für den sequenzierten Genomanteil, Anzahl Ns in der Consensus Sequenz in % und der Anzahl der Reads der

Sequenzierung (Readmenge) erfasst wurden. Es erfolgte eine Prüfung auf eine Normalverteilung der Variablen mittels Kolmogorov-Smirnov- sowie Shapiro-Wilk-Test, worin sich zeigte, dass mindestens eine Variable nicht normal verteilt war. Aus diesem Grund wurden anschließend paarweise Korrelationen als Spearman-Rangkorrelation berechnet und graphisch dargestellt. Des Weiteren wurde durch die Berechnung 100 - Anzahl Ns in der Consensus Sequenz in % eine neue Variable sequenzierter Genomanteil in % definiert, von der im Anschluss ein Mittelwert mit Standardabweichung berechnet wurde.

Zur weiteren Analyse wurde die nominalen Variablen Erfolg90/95 als Anzahl Ns in % $\leq$ 10 bzw. 5 % definiert, um analysieren zu können, wie häufig eine erfolgreiche PCR (definiert als Coverage, daher Reads pro Nukleotidposition, > 30) bei über 90 bzw. 95 % des Genoms erreicht wurde. Diese relativen Häufigkeiten wurden weiter aufgetrennt nach der Viruslast sowie der Readmenge. Dazu wurden die nominalen Variablen vlü100.000 als Viruslast > 100.000 IU/ml, vlü1.000.000 als Viruslast > 1.000.000 IU/ml sowie readü100.000 als Readmenge > 100.000 IU/ml definiert.

Zum Vergleich eines zusätzlichen 4. Pools des GT 1a wurde die nominale Variable Pool 4 definiert, wodurch die Häufigkeiten der Variable Erfolg90/95 unterschieden werden konnten. Beim Vergleich der unterschiedlichen Primersets des GT 1b wurde die nominale Variable Primerset (alt/neu) definiert und ebenfalls danach unterschieden. Des Weiteren wurde nach diesem Prinzip für den Vergleich zwischen der Nanopore- und Illumina-Sequenzierung des GT 3a vorgegangen, wofür die nominale Variable Methode definiert wurde.

Zur graphischen Darstellung der Coverage einzelner Proben zu ihren Nukleotidpositionen wurde die Software GraphPad Prism verwendet. Außerdem wurden Balkendiagramme zur Gegenüberstellung des Erfolgs der Genotypten gebildet. Die zugehörigen Häufigkeiten der Genomabdeckung > 90 % bzw. > 95 % entstammten der statistischen Auswertung aus SPSS.

4 Ergebnisse und Validierung

4.1 Konzeptentwicklung der Tiling PCR und Sequenzierung

Die Entwicklung der Ganzgenom-PCR und -Sequenzierung des Hepatitis C Virus teilte sich in zwei Schritte. In der Etablierung wurden Genotyp-spezifische Primersets entworfen und bestmöglich an Referenz-Isolate angepasst. Anschließend wurde die Methode mit höherer Fallzahl überprüft und statistisch ausgewertet. Es wurden Primersets für die in Deutschland am häufigsten vorkommenden Genotypen 1a, 1b, 3a, 4a und 4d entwickelt. Zur Etablierung wurden Proben von HCV-infizierten Patienten der PEPSI-Studie verwendet. Dabei wurden ausschließlich Proben mit einer Viruslast über 100.000 IU/ml verwendet, um ein Fehlschlagen durch zu wenig Virusmaterial auszuschließen. Um die entwickelten Primersets an einer höheren Fallzahl zu überprüfen, wurden Rückstellproben der Düsseldorfer IDU-Kohorte verwendet.

Die Tiling-PCR ist angelehnt an das Protokoll der Multiplex PCR für MinION und Illumina-Sequenzierung von J. Quick. [61] Dabei wurde eine spezifische Primerpoolentwicklung und -anpassung für jeden Genotyp durchgeführt, die im Folgenden genauer beschrieben ist.

4.2 Optimierung der Reversen Transkription

Zur Amplifikation des gesamten Genoms ist zuvor die Bildung eines cDNA-Strangs durch eine Reverse Transkription notwendig. Im Labor des Instituts für Virologie der Uniklinik Düsseldorf wird für die Amplifikation der Core- und NS5A-Region eine RT mit dem Primer p874a und Oligo-d(A)-Primer durchgeführt. Der p874a-Primer bindet in der Core-Region (an Referenzsequenz H77 an Nukleotidposition (nt) 852-874) während der Oligo-d(A)-Primer an einem konservierten Poly-T Stretch am 3’-Ende des HCV-Genoms (nt 9418-9437) bindet.

Es sollte zunächst untersucht werden, ob diese Genotyp-unabhängige Reverse Transkription [76] auch für die Ganzgenom-Amplifikation verwendet werden kann. Da für den Genotyp 1b im Institut bereits ein Primer-Set zur Tiling-PCR entworfen worden war, wurden zunächst Proben des Genotyps 1b mit einer Viruslast von über 1.000.000 IU/ml amplifiziert und sequenziert. Anschließend wurden die Sequenzstücke auf das Referenzgenom kartiert und die Coverage analysiert (Vgl. 3.3.2, 3.3.3, 3.3.6). Mit Ausnahme von kleinen Lücken konnte bei den meisten Proben über das gesamte Genom eine Coverage über 30 erreicht werden. Das bedeutet, für die Genomabschnitte liegen Daten von mindestens 30 Sequenzierungen vor und

die Sequenzdaten haben damit ein Mindestmaß an Verlässlichkeit erreicht. Allerdings wurde bei einigen Proben eine unvollständige Sequenzabdeckung beobachtet, was am Beispiel der Probe mit Virologie-Nummer (VI-Nr.) 41790 gezeigt ist. In Teil A der Abbildung 3 ist die Coverage dieser Probe dargestellt und es ist ersichtlich, dass erst ab Position 4800 eine Coverage über 30 erreicht wurde. Als Minimum für eine erfolgreiche Sequenzierung wurde ein Wert der Coverage über 30 definiert. Eine gute Coverage sollte Werte > 100 erreichen, das angestrebte Ziel der Coverage stellen Werte > 10.000 dar.

Als Ursache wurde die Bildung einer unvollständigen cDNA nach der RT bei einer geringen Viruslast vermutet. Der verwendete Oligo-d(A)-Primer bindet am 3‘-Ende des Genoms und scheint die RT nicht effizient bis zum 5‘-Ende durchführen zu können. Der p874a-Primer bindet an Position 874 im Genom und ermöglicht nur eine sehr kurze cDNA-Bildung. Zur Lösung des Problems wurde die RT in einem nächsten Versuch mit Random Hexameren an Stelle des p874a-Primers durchgeführt. Random Hexamere bestehen aus sechs Nukleotiden und können zufällig verteilt im Genom binden und daher auch eine vollständige cDNA am Anfang des HCV-Genoms gewährleisten. Im Unterschied zum p874a- und Oligo-d(A)-Primer werden durch Random Hexamere viele kleine Stücke produziert, was für eine Tiling-PCR aber keinen Nachteil darstellt. Zu beachten ist allerdings die Bildung von unspezifischen PCR-Fragmenten, da Random Hexamere auch jede Nicht-HCV-RNA binden und replizieren können. Das neue RT-Protokoll wurde zuerst an Proben des Genotyp 1b untersucht. Dabei zeigte sich eine gleichmäßige Amplifikation sowie eine erfolgreiche Sequenzierung des vollständigen Genoms, was in Teil B der Abbildung 3 am Beispiel der Probe mit VI-Nr. 41790 gezeigt ist.

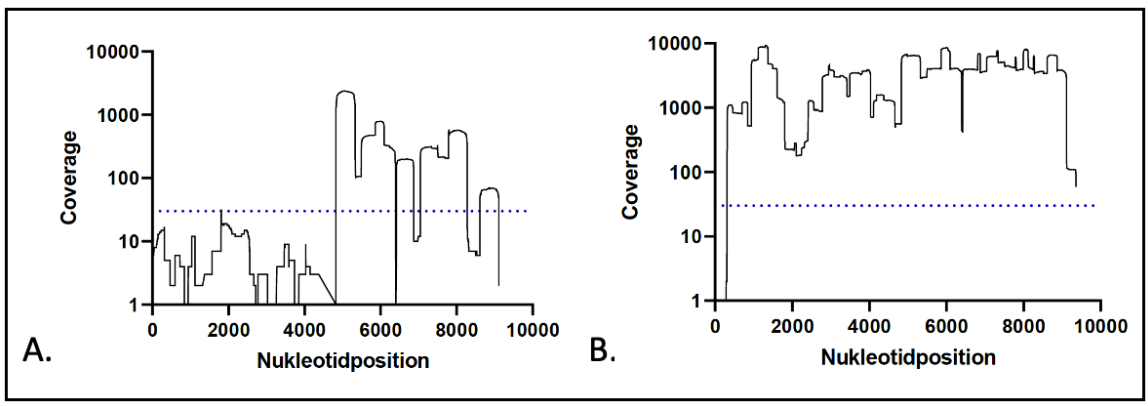

Abbildung 3: Coverage der Probe #41790 des Genotyps 1b im Versuch 1 und 2
Dargestellt ist die Anzahl der Reads pro Nukleotidposition (engl. *Coverage*) im Verhältnis zur Nukleotidposition auf dem Genom einer Tiling-Polymerasekettenreaktion einer Hepatitis C Probe #41790 des Genotyps 1b. Eine Coverage < 30 wurde als nicht ausreichend definiert und ist mit der blau gepunkteten Linie markiert. In Teil **A.** ist der erste Versuch mit einer Reversen Transkription (RT) mit p874a- sowie Oligo-d(A)-Primer dargestellt, wobei erst ab einer Position von 4800 Nukleotiden eine hinreichende Coverage erreicht wurde. In Teil **B.** ist der zweite Versuch mit einer RT mit Random Hexameren anstelle des p874a-Primers abgebildet, wobei eine ausreichende Coverage über nahezu das gesamte Genom erreicht wurde.

4.3 Entwicklung der Tiling-PCR Primer-Sets

4.3.1 Etablierung Genotyp 1a

Um eine Ganzgenom-Sequenzierung des Hepatitis C Virus zu entwickeln, wurden Primer für eine Tiling-PCR entworfen. Da der Genotyp 1a in Deutschland häufig vorkommt und er sich von dem ebenfalls häufigen Genotyp 1b unterscheidet, wurde ein spezifisches Primerset für den GT 1a entwickelt. Zur Erstellung konnte auf bereits vorhandene Primer-Sets des Instituts zurückgegriffen werde, die in drei Primer-Pools aufgeteilt waren. Pro Pool sollten dadurch 12 (Pool 2 +3) oder 13 (Pool 1) jeweils ca. 400 bp lange PCR-Fragmente entstehen, die nach anschließender Sequenzierung bioinformatisch durch Kartieren auf eine Referenzsequenz zusammengesetzt werden sollten. Die Primer sind in Tabelle 1 des Anhangs aufgeführt.

In einem ersten Versuch wurden sieben Genotyp 1a-Proben mit einer Viruslast > 800.000 IU/ml mittels Oligo-d(A)-Primer und Random Hexameren in cDNA umgeschrieben und mit den bereits vorhandenen Primern in drei Pools amplifiziert. In einer anschließenden Gel-Elektrophorese (Vgl. 3.3.4) konnte die Bildung von PCR-Produkten nachgewiesen werden, was in Abbildung 4 dargestellt ist. Die intensivsten Banden sind bei ca. 400 bp sichtbar und stellen die gebildeten PCR-Fragmente dar. Des Weiteren sind einige größere Banden erkennbar, die durch PCR-Fragmente zwischen weiter entfernten Primerpaaren entstanden sind.

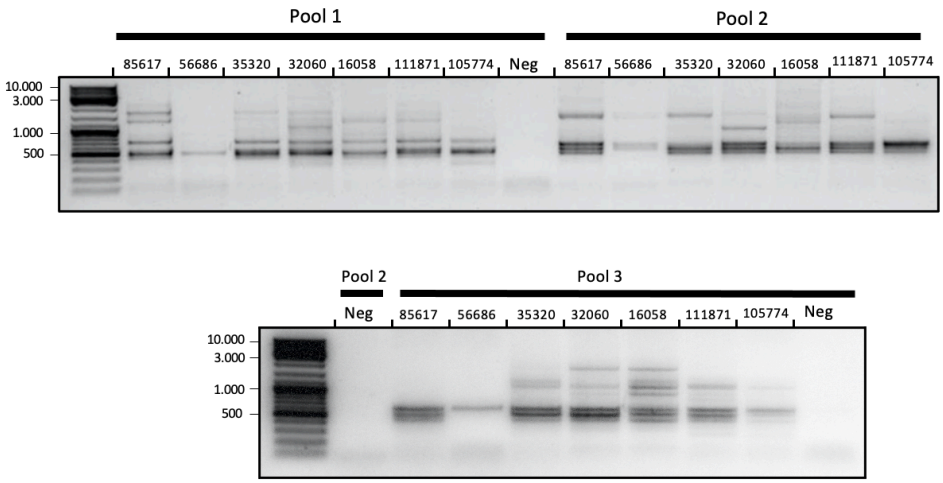

Abbildung 4: Gel-Elektrophorese von sieben Proben des Genotyps 1a im Versuch 1
Dargestellt ist das Ergebnis einer Gel-Elektrophorese einer Tiling-Polymerasekettenreaktion (PCR) von sieben Proben des Hepatitis C Virus des Genotyps 1a. Ziel war der Nachweis der erwarteten PCR-Produkte, welche ca. 400 Basenpaar lange Fragmente enthalten. Ebenfalls zeigten sich einige größere Fragmente, die zufällig zwischen zwei nicht zusammengehörigen Primern entstanden sind.

Nach der Sequenzierung des PCR-Produkts mittels Nanopore sowie der bioinformatischen Datenverarbeitung (Vgl. 3.3.6, 3.3.8) konnte eine Analyse der Coverage sowie der Primerbindestellen erfolgen. Dazu wurde für die Coverage ein Cut-Off von 30 als Minimum definiert, welcher so u.a. auch in der Humangenetik üblich ist. Des Weiteren wurde eine Coverage $\geq$ 100 als gute und $\geq$ 10.000 als gewünschte Coverage festgelegt, da eine Coverage $\geq$ 10.000 für eine Resistenztestung notwendig ist. In Regionen mit einer Coverage < 30 wurden die zugehörigen Primer mit 144 Genotyp 1a Referenzisolaten (Vgl. Abschnitt 2.1 des Anhangs) verglichen und nach Verbesserungsmöglichkeiten untersucht. Dieser Prozess soll im Folgenden an einigen Beispielen erläutert werden.

Eine Region des HCV-Genoms, die sich durch große Vielfalt sowie Quasispezies auszeichnet und daher für eine Sequenzierung besonders schwer abzudecken ist, wird als hypervariable Region bezeichnet. Dabei handelt es sich um einen Bereich des viralen Hüllproteins, der sich unter dem Selektionsdruck von Antikörpern kontinuierlich verändert. Durch diese Region kann ein bioinformatisches Problem entstehen, in dem unterschiedliche, aber durch die PCR korrekt amplifizierte Fragmente nicht ganz auf die Referenzsequenz passen und deshalb vom Algorithmus als schlechte Sequenz mit einer niedrigen Coverage erkannt werden. Deshalb wurde für die HVR zusätzlich das in der bioinformatischen Datenverarbeitung gebildete Assembly analysiert, welches bei sechs der sieben Proben des GT 1a an Position 1500-1620 Inhomogenitäten zeigte. Die nachfolgend gebildete Consensus Sequenz und deren Coverage war für die Position der HVR dennoch ausreichend abgedeckt, was in Teil A. der Abbildung 5 am Beispiel der Probe mit VI-Nr. 32060 dargestellt und mit dem orangenen Pfeil markiert ist.

Um diesen komplexen und inhomogenen Sequenzabschnitt trotzdem noch besser abzubilden und das bioinformatische Problem zu reduzieren, wurden die zugehörigen Primer 1e, 2a und 2b mit Referenzisolaten verglichen und eine Optimierung der Primer durchgeführt. Um eine höhere Variabilität und mehr Quasispezies mit der PCR amplifizieren zu können, wurden Mehrdeutigkeitssymbole (wie z.B. Y, R, K, auch *International Union of Pure and Applied Chemistry* (*IUPAC)-Codes* genannt) innerhalb der Primer eingesetzt. Wenn z.B. der IUPAC-Code Y im Primer geschrieben wird, werden in der Produktion dieses Primers 50 % der Isolate an dieser Stelle ein Cytosin (C) und 50 % ein Thymin (T) aufweisen, wodurch der Primer an mehr HCV-Genome passt. Für das Primer-Set des Genotyps 1a wurde im Primer 1e-R an Position 19 ein T zu einem K verändert, weil in den komplementären Referenzsequenzen zu 60 % A und zu 40 % C vorkommen. Es wurde darauf geachtet, nicht mehr als zwei Mehrdeutigkeitssymbole zu verwenden, da dies die Bindung des Primers an das PCR-Produkt

destabilisiert. Aus diesem Grund wurde im Primer 1e-R an Position 13 ein K zu einem T verändert, da an dieser Stelle in den komplementären Referenzisolaten zu 97,9 % ein A zu finden war. Die Primer 1e-F und 2a-R wurden als Ganzes entlang des Genoms verschoben, um eine bessere Passgenauigkeit zum Referenzgenom zu ermöglichen.

Es wurden weitere Abschnitte des Genoms und deren Primer untersucht, bei denen eine reduzierte Coverage vorlag. Z.B. wurde im Primer 3c-F an Position 1 ein T zu einem C korrigiert, da in den Referenzisolaten zu 100 % ein C vorkam. Der Primer 3e-R wurde um ein Nukleotid gekürzt, da an der ersten Position zu 17 % ein T und zu 83 % ein C in den Referenzisolaten vorkommen und eine Länge von 19 Nukleotiden bereits für eine stabile Bindung des Primers ausreichend ist. Die Primer 4b-F und 4b-R wurden wie die Primer 1e-F und 2a-R als Ganzes entlang des Genoms verschoben. Die Veränderung der Primersequenzen sind in folgender Tabelle 3 dargestellt.

Tabelle 3: Neue Primer Genotyp 1a
Dargestellt sind die veränderten Primer und deren Sequenz für eine Tiling-Polymerasekettenreaktion zur Sequenzierung des Hepatitis C Virus des Genotyps 1a.

Primer	Alte Sequenz	Neue Sequenz
1e-F / 1e-F_new	GGTGTGTCCCTTGCGTTCG	CCGGGGTGTGTCCCTTGCGT
1e-R / 1e-R_new	CTATTGAKGTGCCAACTGCC	CKATTGATGTGCCAACTGCC
2a-R / 2a-R_new	AGGTYTTGGRGGGTAGTGCC	TAGTGCCAGCAGTAGGGGCG
3c-F / 3c-F_new	TCCRAATGGAGACCAAGC	CCCRAATGGAGACCAAGC
3e-R / 3e-R_short	ATCCGTGGAGTGGCACTCG	ATCCGTGGAGTGGCACTC
4b-F / 4b-F_new	CCTACGGCAAGTTCCTTGC	GTAAGAGCACCAAGGTCCC
4b-R / 4b-R_new	AGCGTGRTTGTCTCAATGG	GCAGTCTATCACCGAGTCG

Die neuen Primer wurden dem vorhandenen Primerset hinzugefügt und anschließend eine PCR mit den gleichen sieben Proben erneut durchgeführt. Da die einzelnen Fragmente nicht in der Gel-Elektrophorese dargestellt sind und zuvor experimentell bestätigt wurde, dass die Tiling-PCR funktioniert, wurde im weiteren Verlauf auf eine Gel-Elektrophorese verzichtet und der Erfolg der PCR nur noch durch eine DNA-Quantifizierung mittels Berthold-Luminometer nachgewiesen (Vgl. 3.3.5). Eine Auswertung dieser DNA-Quantifizierung ist beispielhaft im Anhang (Vgl. Abschnitt 3 und Tabelle 6 des Anhangs) dargestellt. Im Anschluss erfolgte erneut eine Sequenzierung mittels MinION sowie Analyse der Coverage und Primerbindestellen.

Auch nach Optimierung der Primer der HVR zeigte sich weiterhin eine Inhomogenität im Assembly und nur wenig Veränderung der Coverage, was in Teil B. der Abbildung 5 am Beispiel der Probe mit VI-Nr. 32060 gezeigt und mit dem orangenen Pfeil markiert ist. An der Nukleotidposition 1865-1900 war bei dieser Probe im ersten Versuch (Teil A. der Abbildung 5) eine verminderte Coverage beobachtet worden. Die zugehörigen Primer 2b-F/R passten aber genau auf die Sequenz und wurden daher im Primerset belassen. Im zweiten Versuch konnte an dieser Nukleotidposition eine ausreichende Coverage erreicht werden, was in Teil B. der Abbildung 5 dargestellt und mit dem roten Pfeil markiert ist.

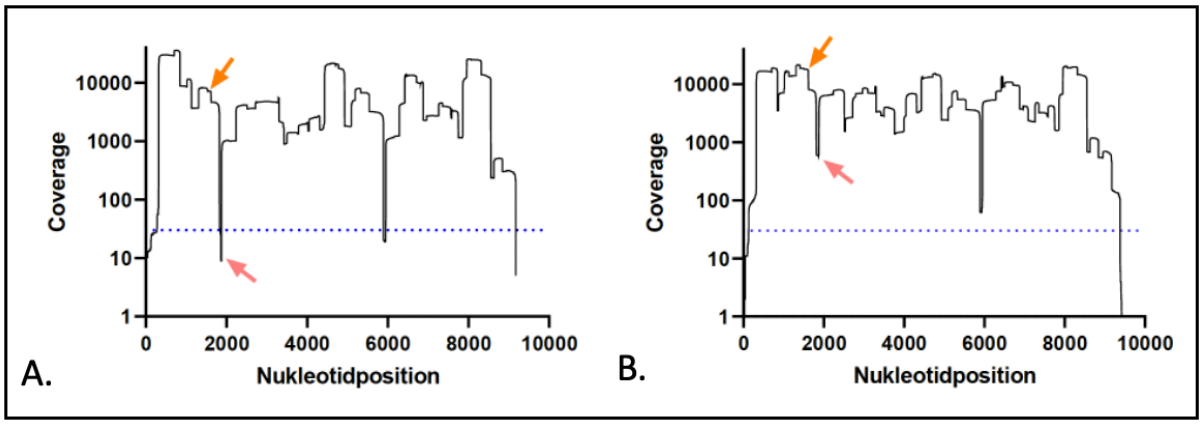

Abbildung 5: Coverage der Probe #32060 des Genotyps 1a im Versuch 1 und 2
Analog zu Abbildung 3 ist die Anzahl der Reads pro Nukleotidposition (engl. *Coverage*) im Verhältnis zur Nukleotidposition auf dem Genom einer Tiling-Polymerasekettenreaktion einer Hepatitis C Probe #32060 des Genotyps 1a dargestellt. In Teil **A.** ist der erste Versuch mit dem ursprünglichen Primerset dargestellt, wo an Nukleotidposition 1865-1900 die Coverage das Minimum von 30 unterschreitet, was mit dem roten Pfeil markiert ist. Die gleiche Position ist in Teil **B.**, welcher den zweiten Versuch mit neuen Primern abbildet, ebenfalls mit dem roten Pfeil markiert und zeigt eine ausreichende Coverage. Mit einem orangenen Pfeil markiert wurde in Teil **A. und B.** die hypervariable Region, die in beiden Versuchen eine Coverage weit über 30 erreicht hat. Ebenfalls in beiden Abbildungsteilen zeigt sich eine verminderte Coverage circa an Nukleotidpositionen 5800-5900.

Bei weiterer Coverage-Analyse fiel bei vier der sieben Proben eine Minderung im Bereich der Primer 5b und 5c auf (ca. Position: 5800-5900). Dies war bereits im ersten Versuch bei drei der sieben Proben beobachtet worden, es konnte aber weder im Vorwärts (engl. *Forward*) noch im Rückwärts (engl. *Reverse*) -Primer eine verbesserte Primersequenz erzeugt werden, da die vorhandenen Primer bereits exakt auf die Referenzgenome passten. Die verminderte Coverage ist in Teil A. der Abbildung 6 am Beispiel der Probe mit VI-Nr. 35320 gezeigt.

Da sowohl Forward- und Reverse-Primer 5b als auch 5c zum Referenzgenom passend erschienen, wurde in einem 3. Versuch die Konzentration dieser Primer im Pool verdoppelt, um eine besser Genomabdeckung zu erreichen. Dadurch ließ sich an entsprechender Position eine deutliche Steigerung der Coverage beobachten, was graphisch in Teil B. der Abbildung 6 am Beispiel von Probe mit VI-Nr. 35320 dargestellt ist.

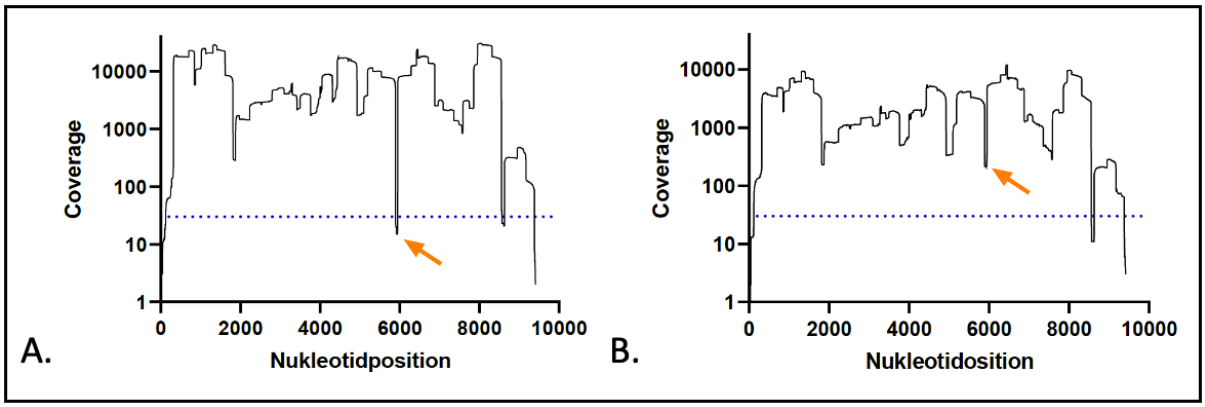

Abbildung 6: Coverage der Probe #35320 des Genotyps 1a im Versuch 2 und 3
Analog zu Abbildung 3 ist die Genomabdeckung (engl. *Coverage*) im Verhältnis zur Nukleotidposition auf dem Genom einer Tiling-Polymerasekettenreaktion einer Hepatitis C Probe des Genotyps 1a #35320 dargestellt. In Teil **A.** ist der Versuch 2 mit einfacher Konzentration der Primers 5b und 5c dargestellt, wogegen Teil **B.** den Versuch 3 mit doppelter Konzentration dieser Primer zeigt. Der orangene Pfeil markiert die den beiden Primern zugehörige Nukleotidposition 5800-5900, die in Teil **A.** eine verminderte Coverage und in Teil **B.** eine deutliche Erhöhung der Coverage darstellt. In beiden Abbildungsteilen fällt eine reduzierte Coverage zu Genombeginn und -ende sowie an der Nukleotidposition 9080-9160 auf.

Im Vergleich der drei durchgeführten Versuche fiel auf, dass trotz der Optimierung der RT mit Random Hexameren bei fast allen Proben die Anfangssequenz bis zum 300. Nukleotid, am häufigsten bis zum 130. Nukleotid, fehlte. Dies war verwunderlich, da der zugehörige Primer 1a-F (am Referenzgenom H77 an nt 36-52) in der 5'- Untranslatierten Region (UTR) bindet und genau wie der Reverse-Primer 1a-R (am Referenzgenom H77 an nt 435-454) exakt auf die Referenzsequenzen passte.

Im Primerset des Genotyps 3a liegt der Primer für das erste Fragment an einer anderen Position (am Referenzgenom H77 an nt 93-113) und in Etablierungs-Versuchen (Vgl. 4.3.3) wurde das erste Fragment in jedem der elf untersuchten Isolate amplifiziert. Da der Primer HCV-3a_1L ebenfalls auf die Sequenz des GT 1a-Referenzisolats passte, wurde in einem weiteren Versuch für drei Proben zusätzlich dieser erste Primer des GT 3a zum 1. Pool des GT 1a hinzugefügt. Dabei zeigte sich in allen drei Proben eine Besserung durch einen früheren Sequenzbeginn, in dem das erste Fragmente mindestens teilweise amplifiziert werden konnte. Bildlich dargestellt ist dies beispielhaft an der Probe mit VI-Nr. 35320 in Teil A. und B. der Abbildung 7.

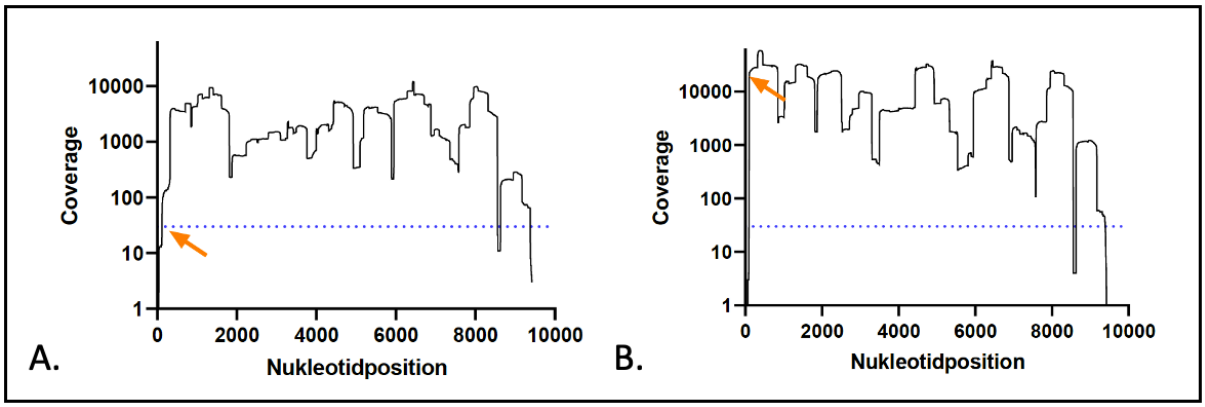

Abbildung 7: Coverage der Probe #35320 des Genotyps 1a im Versuch 3 und 4
Analog zu Abbildung 3 ist die Anzahl der Reads pro Nukleotidposition (engl. *Coverage*) im Verhältnis zur Nukleotidposition auf dem Genom einer Tiling-Polymerasekettenreaktion einer Hepatitis C Probe des Genotyps 1a #35320 abgebildet. In Teil **A.** ist der Versuch 3 mit dem ursprünglichen Primerset dargestellt, wogegen Teil **B.** den Versuch 4 mit einem zusätzlichen Primer HCV-3a_1L zeigt. Der orangene Pfeil markiert die Anfangssequenz des Genoms, an dem die Coverage mit dem ursprünglichen Primerset (Teil **A.**) vermindert war und mit dem zusätzlichen Primer HCV-3a_1L (Teil **B.**) deutlich erhöht ist. In beiden Abbildungsteilen fällt eine reduzierte Coverage zu Genomende sowie an der Nukleotidposition 9080-9160 auf.

Das optimierte Genotyp 1a -Primerset aus drei Pools ist in Tabelle 4 dargestellt und wurde anschließend zur Überprüfung der Methode mit einer höheren Fallzahl verwendet.

Tabelle 4: Primerset Genotyp 1a
Abgebildet ist das Primerset einer Tiling-Polymerasekettenreaktion für eine Sequenzierung des Hepatitis C Virus des Genotyp 1a. Es teilt sich in drei Pools auf.

Pool 1		Pool 2		Pool 3	
1a-F	1a-R	1b-F	1a-R	1c-F	1c-R
1d-F	1d-R	1e-F, 1e-F_new	1e-R 1e-R_new	2a-F	2a-R 2a-R_new
2b-F	2b-R	2c-F	2c-R	2c-F	2d-R
2e-F	2e-R	3a-F	3a-R	3b-F	3b-R
3c-F 3c-F_new	3c-R	3d-F	3d-R	3e-F	3e-R 3e-R_short
4a-F	4a-R	4b-F 4b-F_new	4b-R 4b-R_new	4b-F_new	4b-R_new
4c-F	4c-R	4d-F	4d-R	4e-F	4e-R
5a-F	5a-R	5b-F	5b-R	5c-F	5c-R
5d-F	5d-R	5e-F	5e-R	6a-F	6a-R
6b-F	6b-R	6c-F	6c-R	6d-F	6d-R
6e-F	6e-R	7a-F	7a-R	7b-F	7b-R
7c-F	7c-R	7d-F	7d-R	7e-F	7e-R
7e-R_int	7e-R_ext				

4.3.2 Etablierung Genotyp 1b

Da der Genotyp 1b des Hepatitis C Virus in Deutschland vermehrt bei Patienten mit einer HCV-Infektion vorkommt, sollte wie für den Genotyp 1a ein spezifisches Primerset zur Etablierung einer Ganzgenom-Sequenzierung entwickelt werden. Ebenfalls konnte dabei auf bereits vorhandene Primer-Sets des Instituts zurückgegriffen werden, die sich in drei Pools aufteilten und elf (Pool 3) oder zwölf (Pool 1+2) ca. 400 bp lange Fragmente bildeten. Die Primer sind in Tabelle 3 des Anhangs aufgeführt.

In einem ersten Versuch wurde für sieben Proben mit einer Viruslast über 1.000.000 IU/ml eine Reverse Transkription mit dem Oligo-d(A)-Primer und eine anschließende Tiling-PCR durchgeführt (Vgl. 3.3.2, 3.3.3). Die Bildung der PCR-Produkte wurde mittels Gel-Elektrophorese nachgewiesen, was in Abbildung 8 beispielhaft an vier Proben gezeigt ist. Genau wie für den Genotyp 1a sind die intensivsten Banden bei ca. 400 bp ersichtlich, die das gebildete PCR-Produkt repräsentieren und anzeigen, dass die Tiling-PCR funktioniert hat. Weitere längere Banden sind durch Fragmente zwischen weiter entfernten Primern entstanden.

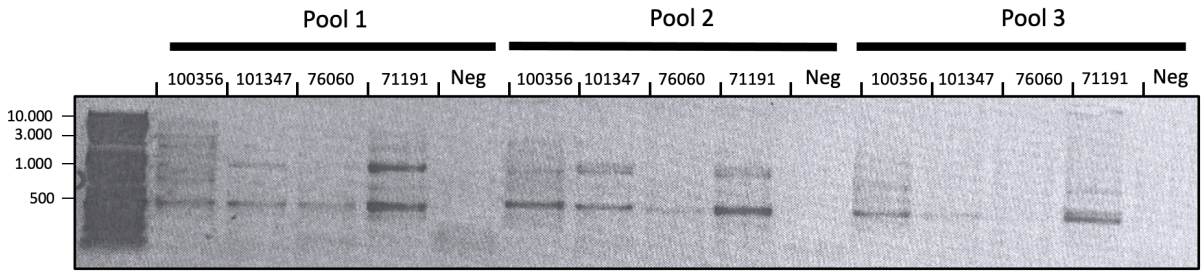

Abbildung 8: Gel-Elektrophorese von vier Proben des Genotyps 1b im Versuch 1
Dargestellt ist das Ergebnis einer Gel-Elektrophorese einer Tiling-Polymerasekettenreaktion von vier Proben des Hepatitis C Virus des Genotyps 1b. Die erwarteten PCR-Produkte, welche aus ca. 400 Basenpaar langen Fragmenten bestehen, konnten nachgewiesen werden. Ebenfalls zeigten sich einige größere Fragmente, die zufällig zwischen zwei nicht zusammengehörigen Primern entstanden sind.

Wie bei Genotyp 1a erfolgte anschließend eine Sequenzierung mittels Nanopore, eine bioinformatische Auswertung sowie eine Analyse der Genomabdeckung mit Hilfe von 229 Referenzgenomen (Vgl. Abschnitt 2.2 des Anhangs). Dabei fiel eine fehlende Coverage zu Beginn des Genoms auf, was durch die Verwendung von Random Hexameren bei der Reversen Transkription verbessert werden konnte. (Vgl. 4.2)

In einem zweiten Versuch wurden elf Proben des Genotyps 1b mit dem gleichen Primer-Set, aber der modifizierten RT mit Random Hexameren und Oligo-d(A)-Primer angesetzt (Vgl. 4.2) und nach dem gleichen Prinzip verfahren. Dabei konnte ein früherer Sequenzbeginn bei einigen Proben beobachtet werden, was für eine Verbesserung der Reversen Transkription spricht und deshalb in den nächsten Versuchen mit Random Hexameren und Oligo-d(A)-Primer durchgeführt wurde.

Wie für den Genotyp 1a beschrieben (Vgl. 4.3.1), stellt die Sequenzierung der HVR eine besondere Herausforderung dar. Auch für den Genotyp 1b war an Genomposition 1490-1550 bei allen elf Proben eine Inhomogenität im Assembly während der bioinformatischen Datenverarbeitung erkennbar. Um die Amplifikation zu verbessern (siehe Teil A. der Abbildung 9), wurden die zugehörigen Primer analysiert und optimiert. Z.B. wurde an Primer HCV-1b_1e-R an Position 6 ein Guanin (G) zu einem R ausgetauscht, da in der komplementären Sequenz der Referenzgenome zu 82 % ein C und zu 17 % ein T vorkamen. Außerdem wurde am Primer HCV-1b_2a-R an Position 17 ein R gegen ein A ausgetauscht, da in den komplementären Referenzgenomen zu 75 % ein C und zu 21 % ein T vorkamen. Zusätzlich zur Nukleotidanalyse der Primer wurde überprüft, dass die sog. *Annealing*-Temperatur der Primer mind. 65 °C beträgt. Dies wurde mittels des Tm Calculators von New England Biolabs® (https://tmcalculator.neb.com/#!/main) überprüft. Wurde eine entsprechende Temperatur nicht erreicht, musste der Primer um einige Nukleotide verlängert werden. Dies war z.B. beim Primer HCV-2a-F der Fall, der um drei Nukleotide ergänzt wurde.

Nach dem gleichen Prinzip wurden Primer optimiert, deren zugehöriger Nukleotidbereich eine verminderte Coverage zeigte. So wurde darauf geachtet, den Primer an eine Position zu setzen, an der die meisten der Referenzsequenzen ein einheitliches Nukleotid haben oder ein Mehrdeutigkeitssymbol einsetzbar war. Ein Mehrdeutigkeitssymbol sollte, wie bei Genotyp 1a beschrieben, nicht häufiger als zweimal pro Primer verwendet werden.

In Tabelle 5 sind alle durchgeführten Veränderungen an den Primersequenzen dargestellt.

Tabelle 5: Neue Primer Genotyp 1b
Dargestellt sind die veränderten Primer und deren Sequenz für eine Tiling-Polymerasekettenreaktion zur Sequenzierung des Hepatitis C Virus des Genotyps 1b.

Primer	Alte Sequenz	Neue Sequenz
HCV-1b_1e-R / _NEW	TTGATGTGCCAGCTGCC	TTGATGTGCCARCTGCC
HCV-1b_2a-F / _NEW	GGGATATGATGATGAACTGG	TGGGATATGATGATGAAYTGGTC
HCV-1b_2a-R / _NEW	GAGGYGCGTAGTGCCAGC	GRGGYGCGTAGTGCCAGC
HCV-1b_2e-R / _NEW	CCRTTTGGACATAATGACC	CCATTTGGACATAATGRCCCCC
HCV-1b_3b-R / _NEW	ATGGGCGCRAGGAGTCGC	ATRGGCGCRAGGAGTCGC
HCV-1b_3c-F / _NEW	TTGCGGTGGCAGHAGAGC	TTGCGGTGGCAGTWGAGC
HCV-1b_4c-F / _NEW	CTATGGCAAAGCCATCCC	CCCTGGTGTATTTAGGTAAGCCC
HCV-1b_4e-R / _NEW	TCTGCTTGAAYTGCTCGG	TCTGYTTGAAYTGCTCGGCGA
HCV-1b_5d-R / _NEW	GGGCAYTTTACGTTGTCRGTGG	GGGCAYTTYACGTTGTCAGTGG
HCV-1b_7e-F / _NEW	CYCACTTCTTCTCCATCC	TCTGATGACYCAYTTCTTCTCCATCCT

Die zehn veränderten Primer wurden anstelle ihrer Vorgänger in drei neue Pools pipettiert. Davon wurde nach einer RT mit Random Hexameren und Oligo-d(A)-Primer erneut eine Tiling-PCR für die gleichen elf Proben angesetzt. Wie bei Genotyp 1a erfolgte der Nachweis des PCR-Produkts in den folgenden Versuchen durch eine DNA-Quantifizierung mittels eines Lumineszenz Mikroplatten Reader von Berthold (Vgl. 3.3.5). Anschließend erfolgte ebenfalls eine Sequenzierung mittels Nanopore, eine bioinformatische Auswertung sowie eine Coverage-Analyse.

Im Bereich der HVR wurde trotz der optimierten Primer im Assembly eine Inhomogenität beobachtet, dennoch war die Coverage ausreichend, was den Beobachtungen für den Genotyp 1a entspricht. Die Coverage-Analyse aller Proben zeigte, dass bei sechs der elf Proben häufiger Verringerungen bzw. Lücken in der Coverage im Vergleich zum Versuch mit den alten Primern festgestellt werden konnten. Zum Beispiel traten bei der Probe mit VI-Nr. 11981 zwei neue Lücken bei Position 1830-2000 sowie bei Position 2600-2675 auf, was in folgender Abbildung 9 erkennbar und mit den orangenen Pfeilen markiert ist.

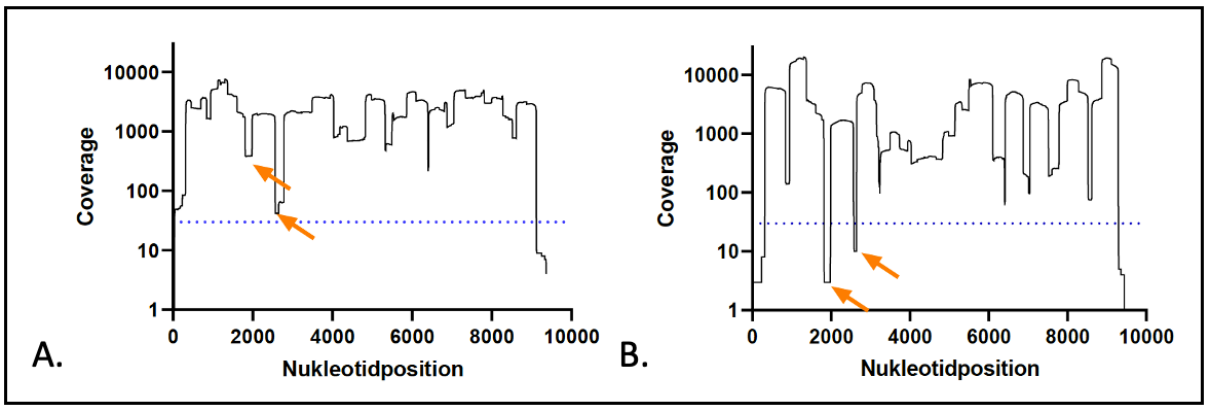

Abbildung 9: Coverage der Probe #11981 des Genotyps 1b im Versuch 2 und 3
Analog zu Abbildung 3 ist die Anzahl der Reads pro Nukleotidposition (engl. *Coverage*) im Verhältnis zur Nukleotidposition auf dem Genom einer Tiling-Polymerasekettenreaktion einer Hepatitis C Probe #11981 des Genotyps 1b abgebildet. Teil **A.** der Abbildung stellt den Versuch 2 mit dem alten Primer-Set dar, wohingegen Teil **B.** den Versuch 3 mit dem neuen Primer-Set zeigt. Die orangenen Pfeile markieren die Nukleotidpositionen 1830-2000 sowie 2600-2675, an denen in Teil **B.** die Coverage das Minimum von 30 unterschreitet. An gleicher Position war in Teil **A.** mit dem alten Primer-Set eine ausreichende Coverage vorhanden.

Da die zugehörigen Primer HCV-1b_2b-F/-R und HCV-1b_2d-F/-R exakt mit den Referenzgenomen übereinstimmen, erscheint es unklar, wieso die Bindung bei dieser Probe nicht funktioniert hat. Mögliche Ursachen können eine seltene Sequenzabfolge in der Probe sein, die der häufigen Variante der Referenzgenome widerspricht, oder eine geringere Rohdatenmengen in der Sequenzierung. Letzteres ist in diesem Fall aber eher auszuschließen, da im Versuch mit den neuen Primern eine größere Datenmenge erzeugt wurde als beim Versuch mit den alten Primern.

Zur Verifizierung, welches Primer-Set effektiver auf die Sequenzen deutscher HCV GT 1b-Proben passt, wurde die Tiling-PCR und Sequenzierung jeweils mit beiden Primer-Sets mit einer größeren Fallzahl von 13 GT 1b-Proben der IDU-Kohorte Düsseldorf durchgeführt. Im Durchschnitt erreichte der Versuch mit den alten Primern eine Coverage von 93,3 % (± 4 %) und der Versuch mit den neuen Primern eine Coverage von 93,6 % (± 3 %). Eine Coverage von > 95 % wurde dabei zu 38,5 % (5 von 13 Proben) im Versuch mit den neuen Primern und zu 47,1 % (8 von 17 Proben) im Versuch mit den alten Primern erreicht. Eine Coverage von > 90 % bei 84,6 % (11 von 13 Proben) wurde im Versuch mit den neuen Primern und bei 76,5 % (13 von 17 Proben) mit den alten Primern beobachtet.

Anschließend wurden die Mittelwerte der Coverage in besonders relevanten Regionen des HCV-Genoms (Core, E1, NS3, NS5A und NS5B) bestimmt, um zu überprüfen, welches Primerset diese besser abdeckt. Dieser Vergleich ist in Tabelle 6 dargestellt und zeigt, dass die Coverage im Versuch mit dem alten Primer-Set in allen untersuchten Abschnitten deutlich höher war.

Tabelle 6: Vergleich der Mittelwerte der Coverage des Genotyps 1b
Dargestellt sind die Mittelwerte der Coverage von Proben der injection drug-use (IDU) – Kohorte Düsseldorf des Genotyps 1b von einer Tiling-Polymerasekettenreaktion mit einem alten und einem überarbeiteten, neuen Primer-Set. Die Nukleotidpositionen der relevanten Proteine von Hepatitis C wurden mit Hilfe der Referenz AF009606 (H77) bestimmt. Die Coverage zeigt sich im Versuch mit dem alten Primer-Set deutlich erhöht.

Protein	Nukleotidposition	Coverage-Mittelwert Versuch mit alten Primern	Coverage-Mittelwert Versuch mit neuen Primern
Core	342-914	3886	2035
E1	915-1490	16249	11664
NS3	3420-5312	8896	4559
NS5A	6258-7601	20552	7406
NS5B	7602-9374	21789	8125

Anhand dessen wurde entschieden, das alte Primerset für folgende Untersuchung und statistische Verifizierungen zu verwenden. Die Verteilung des Primersets in seine drei verschiedenen Pools ist in Tabelle 7 dargestellt.

Tabelle 7 Primerset Genotyp 1b
Abgebildet ist das Primerset einer Tiling-Polymerasekettenreaktion für eine Sequenzierung des Hepatitis C Virus des Genotyps 1a. Es teilt sich in drei Pools auf.

Pool 1		Pool 2		Pool 3	
HCV1b_1a-F	HCV1b_1a-R	HCV1b_1b-F	HCV1b_1b-R	HCV1b_1c-F	HCV1b_1c-R
HCV1b_1d-F	HCV1b_1d-R	HCV1b_1e-F	HCV1b_1e-R	HCV1b_2a-F	HCV1b_2a-R
HCV1b_2b-F	HCV1b_2b-R	HCV1b_2c-F	HCV1b_2c-R	HCV1b_2d-F	HCV1b_2d-R
HCV1b_2e-F	HCV1b_2e-R	HCV1b_3a-F	HCV1b_3a-R	HCV1b_3b-F	HCV1b_3b-R
HCV1b_3c-F	HCV1b_3c-R	HCV1b_3d-F	HCV1b_3d-R	HCV1b_3e-F	HCV1b_3e-R
HCV1b_4a-F	HCV1b_4a-R	HCV1b_4b-F	HCV1b_4b-R	HCV1b_4c-F	HCV1b_4c-R
HCV1b_4d-F	HCV1b_4d-R	HCV1b_4e-F	HCV1b_4e-R	HCV1b_5a-F	HCV1b_5a-R
HCV1b_5b-F	HCV1b_5b-R	HCV1b_5c-F	HCV1b_5c-R	HCV1b_5d-F	HCV1b_5d-F
HCV1b_5e-F	HCV1b_5e-R	HCV1b_6a-F	HCV1b_6a-R	HCV1b_6b-F	HCV1b_6b-R
HCV1b_6c-F	HCV1b_6c-R	HCV1b_6d-F	HCV1b_6d-R	HCV1b_6e-F	HCV1b_6e-R
HCV1b_7a-F	HCV1b_7a-R	HCV1b_7b-F	HCV1b_7b-R	HCV1b_7c-F	HCV1b_7c-R
HCV1b_7d-F	HCV1b_7d-R	HCV1b_7e-F	HCV1b_7e-R		

4.3.3 Etablierung Genotyp 3a

Zur Etablierung einer Ganzgenom-Sequenzierung des in Deutschland häufigen Genotyps 3a des Hepatitis C Virus wurde ebenfalls ein Genotyp-spezifisches Primerset entworfen. Da im Institut bisher kein Primerset zur Verfügung stand, wurde im Rahmen dieser Dissertation ein neues Primerset für den Genotyp 3a erstellt. Hierzu wurde ein Primer-Design-Tool namens Primal Scheme (https://primalscheme.com) verwendet. Da die vorgeschlagenen Primer nicht optimal passten, wurden sie überarbeitet, indem jeder Primer mit 44 Referenzisolaten (Vgl. Abschnitt 2.3 des Anhangs) verglichen wurde. Insbesondere die ersten drei Positionen des 3‘-Endes sollten an eine Stelle im Genom passen, an der die meisten der Referenzsequenzen das gleiche Nukleotid haben, da sie besonders entscheidend für eine Bindung zwischen Primer und dem viralen Genom sind. An Positionen, an denen zwei unterschiedliche Nukleotide häufig vorkommen, (z.B. 55 % A und 45 % G) wurden Mehrdeutigkeitssymbole verwendet, damit der Primer beide Genomvarianten binden kann. Allerdings sollten solche IUPAC-Codes, wie bereits bei Genotyp 1a beschrieben (Vgl. 4.3.1), nicht öfter als zweimal pro Primer verwendet werden, um die Stabilität des Primers nicht zu gefährden. Teilweise war es nötig, den Primer als Ganzes entlang des Genoms zu verschieben, wenn eine zu hohe Variabilität an der vom Primer-Design-Tool vorgeschlagenen Position bestand.

Das gebildete Primerset (Vgl. Tabelle 4 des Anhangs) bestand aus zwei Pools mit jeweils 17 Forward- und Reverse-Primern, durch die ca. 400 bp lange Fragmente gebildet werden sollten. Mit diesem Primerset wurde nach einer RT mit dem Oligo-d(A)-Primer) und Random Hexameren ein erster Versuch mit vier Genotyp 3a-Proben mit einer Viruslast über 1.000.000 IU/ml durchgeführt. Anschließend wurden die entstandenen Produkte durch eine Gel-Elektrophorese nachgewiesen. In Abbildung 10 sind die intensivsten Banden bei ca. 400 bp zu sehen, die den gebildeten Fragmenten entsprechen. Größere Banden, welche durch weiter entfernte Primerpaare entstanden, sind ebenfalls zu sehen.

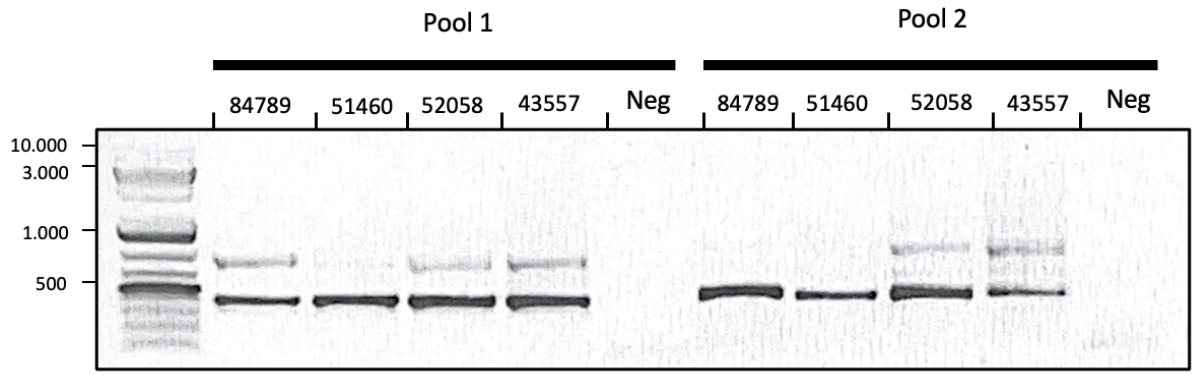

Abbildung 10: Gel-Elektrophorese von vier Proben des Genotyps 3a im Versuch 1
Dargestellt ist das Ergebnis einer Gel-Elektrophorese einer Tiling-Polymerasekettenreaktion (PCR) von vier Proben des Hepatitis C Virus des Genotyps 3a. Es dient als Nachweis der erwarteten PCR-Produkte, welches aus ca. 400 Basenpaar langen Fragmenten bestehen. Ebenfalls zeigten sich einige größere Fragmente, die zufällig zwischen zwei nicht zusammengehörigen Primern entstanden sind.

Nach der Sequenzierung mittels Nanopore erfolgte die bioinformatische Auswertung sowie eine Analyse der Coverage. Dabei zeigte sich bei allen vier Proben eine Coverage deutlich über 30, welche als Minimum für eine erfolgreiche Sequenzierung definiert wurde. Dies ist beispielhaft an der Probe mit VI-Nr. 52058 in Abbildung 11 dargestellt.

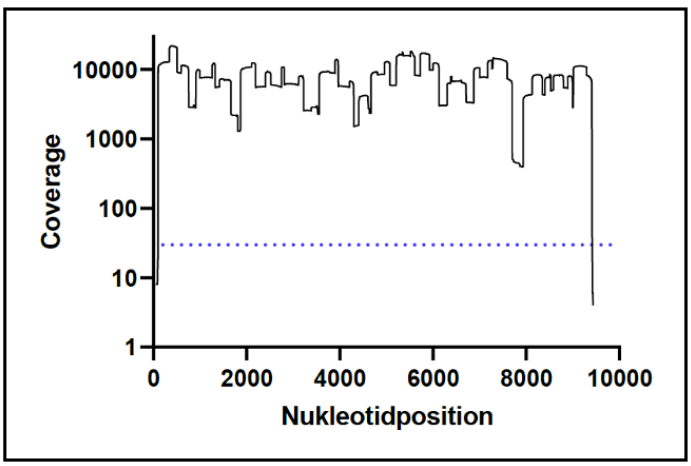

Abbildung 11: Coverage der Probe #52058 des Genotyps 3a im Versuch 1
Analog zu Abbildung 3 ist die Anzahl der Reads pro Nukleotidposition (engl. *Coverage*) im Verhältnis zur Nukleotidposition auf dem Genom einer Tiling-Polymerasekettenreaktion und Sequenzierung einer Hepatitis C Probe #52058 des Genotyps 3a abgebildet. Über nahezu dem gesamten Genom zeigt sich eine ausreichend hohe Coverage.

Zur weiteren Überprüfung des Primer-Sets wurde in einem zweiten Versuch für sieben Genotyp 3a-Proben mit einer Viruslast über 1.000.000 IU/ml eine RT mit Oligo-d(A)-Primer und Random Hexameren, eine Tiling-PCR und eine Sequenzierung mittels Nanopore durchgeführt.

Eine Analyse der Coverage sowie der Primerbindestellen und deren Abgleich mit Referenzisolaten führten zu Korrekturen an sechs Primern, die in Tabelle 8 abgebildet sind von denen einige im Folgenden beispielhaft erläutert werden.

Tabelle 8: Neue Primer Genotyp 3a
Dargestellt sind die veränderten Primer und deren Sequenz für eine Tiling-Polymerasekettenreaktion zur Sequenzierung des Hepatitis C Virus des Genotyps 3a.

Primer	Alte Sequenz	Neue Sequenz
HCV-3a_5L / -alternativ1/2	GCTCRCTGTACCCAGGCCA	CGTCGCCATCAAACGGTCCAGA TTYCTYGTGGGACAAGCCTTCAC
HCV-3a_5R / _new	CCAGCTATGAACCCGGTGTT	CCCAGCTATGAACCCKGTGT
HCV-3a_6L / _new	GGCCTAGCCTATTACTCCATGC	GGCCTAGCCTATTACTCCATGCA
HCV-3a_6R / _new	TAGTGCCAGCAGTACGGTCT	ACRACCACTGGCGATGGTGT
HCV-3a_7L / _new	TACTGCTTCACACCATCGCC	GACAAACCGTACTGCTGGCA
HCV-3a_28L / _new	AGAGCGTGGTCTGCTGCTCT	AGAGYGTRGTCTGCTGCTCT

Wie bereits für den Genotyp 1a und 1b beschrieben (Vgl. 4.3.1, 4.3.2), zeigten sich auch für den Genotyp 3a Inhomogenitäten im Assembly in der Region der HVR, die im Bereich der Primer 5 und 6 (ca. nt 1400-1700) liegt. Um die Coverage dieser Region, wie in Abbildung 12 dargestellt, noch weiter zu erhöhen, wurde die zugehörigen Primer 5 und 6 nochmals optimiert. So wurden für den Primer HCV-3a-5L aufgrund einiger Unterschiede in den Referenzisolaten zwei verschiedene Primervarianten gebildet, um möglichst viel Genvarietät durch das Primerset abzudecken. Außerdem wurde der Primer HCV-3a-5R um eine Position verschoben und an der ursprünglichen Position 5 ein G in ein K verändert, da in den komplementären Referenzisolaten zu 55 % ein C und zu 45 % ein A vorkam, und der Primer HCV-3a-6L um ein Nukleotid verlängert, um eine höhere Annealing-Temperatur zu erhalten.
Des Weiteren trat bei vier der sieben Proben eine Verminderung der Coverage ca. an Nukleotidposition 1770-1830 auf, was beispielhaft an der Probe mit VI-Nr. 43557 in Teil A. der Abbildung 12 dargestellt und mit einem orangefarbigen Pfeil markiert ist. Bei der Primeranalyse dieser Genomposition fiel auf, dass eine Lücke im Primerset zwischen den Primer HCV-3a-6R und -7L bestand. Die Primer wurden deshalb als Ganzes neu im Genom platziert. Im Beispiel der Probe mit VI-Nr. 43557 ist des Weiteren eine Verminderung der Coverage bei Position 7650-7920 zu beobachten (Vgl. Teil A. der Abbildung 12, mit einem orangenfarbenen Pfeil markiert), die ebenfalls bei einer weiteren Probe auftrat. Im zugehörigen Primer HCV-3a-28L wurde deshalb an Position 5 ein C in ein Y verändert, da in

den Referenzisolaten zu 15 % ein T und zu 84 % ein C vorkommen und in beiden eigenen Proben ein T an dieser Position gefunden wurde.

Die korrigierten sieben Primer wurden an Stelle ihrer Vorgänger in ein neues Primerset eingesetzt. Damit wurde in einem weiteren Versuch eine RT mit Oligo-d(A)-Primer und Random Hexameren sowie eine Tiling-PCR mit den gleichen sieben Proben durchgeführt. Die PCR-Produkte wurden durch eine DNA-Quantifizierung mittels eines Lumineszenz Mikroplatten Reader von Berthold nachgewiesen (Vgl. 3.3.5) Nach anschließender Sequenzierung sowie bioinformatischer Auswertung konnte erneut eine Analyse der Coverage und Primerbindestellen erfolgen. Im Beispiel der Probe mit VI-Nr. 43557 konnte eine Verbesserung der Coverage im Bereich der veränderten Primer HCV-3a-6R, -7L sowie -28L beobachtet werden, was in Teil B. der Abbildung 12 gezeigt ist, da die Coverage über das gesamte Genom einen Wert > 1000 erreicht hat.

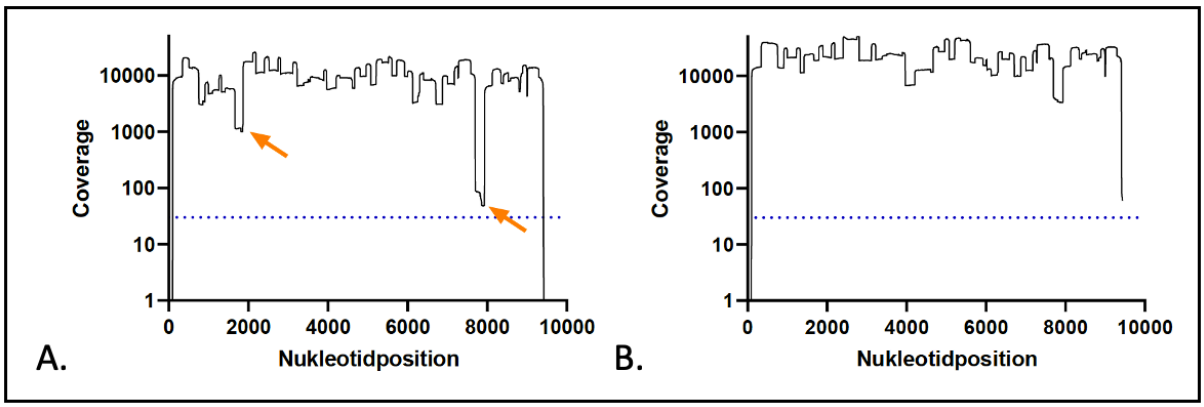

Abbildung 12: Coverage der Probe #43557 des Genotyps 3a im Versuch 2 und 3
Analog zu Abbildung 3 ist die Anzahl der Reads pro Nukleotidposition (engl. *Coverage*) im Verhältnis zur Nukleotidposition auf dem Genom einer Tiling-Polymerasekettenreaktion und Sequenzierung einer Hepatitis C Probe #43557 des Genotyps 3a abgebildet. In Teil **A.** wurde die PCR im Versuch 2 mit alten Primern durchgeführt. An Genomposition 1770-1830 und 7650-7920 zeigen sich verminderte Coverages, die mit orangefarbigen Pfeilen markiert sind. In Teil **B.** wurde die PCR im Versuch 3 mit neuen Primern durchgeführt und hier zeigt sich eine durchgängig hohe Coverage > 1000.

Wie bereits bei Genotyp 1a und 1b beobachtet, führten Primerveränderungen, die den Bereich der HVR abdecken, zu keiner nennenswerten Veränderung des Assembly oder der Coverage (Vgl. 4.3.1, 4.3.2).

Das optimierte Primerset ist in folgender Tabelle 9 in seinen zwei Pools beschrieben worden und wurde anschließend verwendet, um die Tiling-PCR an einem größeren Patientenkollektiv des Genotyps 3a zu verifizieren.

Tabelle 9: Primerset Genotyp 3a
Abgebildet ist das Primerset einer Tiling-Polymerasekettenreaktion für eine Sequenzierung des Hepatitis C Virus des Genotyps 3a. Es teilt sich in zwei Pools auf.

Pool A		Pool B	
HCV-3a_1L	HCV-3a_1R	HCV-3a_2L	HCV-3a_2R
HCV-3a_3L	HCV-3a_3R	HCV-3a_4L	HCV-3a_4R
HCV-3a_5L-alternative1 / 2	HCV-3a_5R_new	HCV-3a_6L_new	HCV-3a_6R_new
HCV-3a_7L_new	HCV-3a_7R	HCV-3a_8L	HCV-3a_8R
HCV-3a_9L	HCV-3a_9R	HCV-3a_10L	HCV-3a_10R
HCV-3a_11L	HCV-3a_11R	HCV-3a_12L	HCV-3a_12R
HCV-3a_13L	HCV-3a_13R	HCV-3a_14L	HCV-3a_14R
HCV-3a_15L	HCV-3a_15R	HCV-3a_16L	HCV-3a_16R
HCV-3a_17L	HCV-3a_17R	HCV-3a_18L	HCV-3a_18R
HCV-3a_19L	HCV-3a_19R	HCV-3a_20L	HCV-3a_20R
HCV-3a_21L	HCV-3a_21R	HCV-3a_22L	HCV-3a_22R
HCV-3a_23L	HCV-3a_23R	HCV-3a_24L	HCV-3a_24R
HCV-3a_25L	HCV-3a_25R	HCV-3a_26L	HCV-3a_26R
HCV-3a_27L	HCV-3a_27R	HCV-3a_28L_new	HCV-3a_28R
HCV-3a_29L	HCV-3a_29R	HCV-3a_30L	HCV-3a_30R
HCV-3a_31L	HCV-3a_31R	HCV-3a_32L	HCV-3a_32R
HCV-3a_33L	HCV-3a_33R	HCV-3a_34L	HCV-3a_34R

4.3.4 Etablierung Genotyp 4

Weltweit werden nur ca. 8,3 % aller HCV-Infektionen durch den Genotyp 4 bedingt, jedoch ist er in manchen Hochrisikogruppen, wie z.B. Männern, die Sex mit Männern haben, stärker verbreitet. Deshalb wurde auch für den GT 4 ein spezifisches Primerset für eine Tiling-PCR entwickelt.

Dazu wurde wie für den Genotyp 3a das Primer-Design-Tool Primal Scheme (https://primalscheme.com) verwendet. Das Primerset wurde mit 62 Referenzisolaten (Vgl. Abschnitt 2.4 des Anhangs) des Genotyps 4 verglichen, wobei besonders Wert auf die häufigen Nukleotide der Subtypen 4a und 4d gelegt wurde, da alle anderen GT 4-Subtypen eher selten in Deutschland auftreten. Wenn sich die Genotypen stark voneinander unterschieden, wurde jeweils eine eigene Primerversion für den Genotyp 4a und 4d erstellt. Auf diesem Wege wurde ein Primerset (Vgl. Tabelle 5 des Anhangs) aus zwei Pools mit jeweils 30 Forward- und Reverse-Primern gebildet, welches anschließend nach den gleichen Kriterien wie für den Genotyp 3a überarbeitet wurde. (Vgl. 4.3.3)

In zwei aufeinanderfolgenden Versuchen wurde eine RT mit Oligo-d(A)-Primer sowie Random Hexameren und eine Tiling-PCR mit insgesamt zehn GT 4a- und GT 4d-Proben mit einer Viruslast > 100.000 IU/ml durchgeführt. Die Bildung der PCR-Fragmente wurde für jeweils zwei Proben jedes Genotyps in einem ersten Versuch in einer Gel-Elektrophorese überprüft. In Abbildung 13 sind intensivste Banden bei ca. 400 bp zu sehen, die den gebildeten Fragmenten entsprechen. Ebenfalls sind größere Banden zu sehen, die sich durch Fragmente zwischen weiter entfernten Primerpaaren gebildet haben.

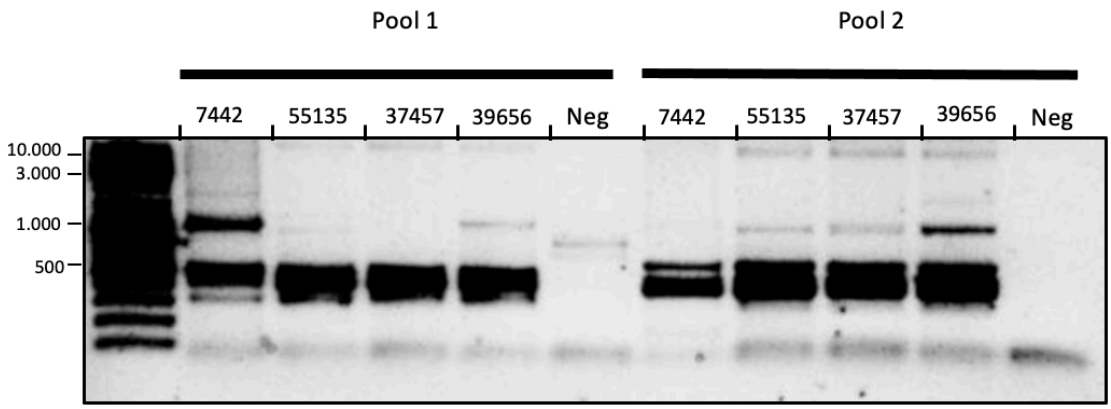

Abbildung 13: Gel-Elektrophorese von vier Proben des Genotyps 4 im Versuch 1
Dargestellt ist das Ergebnis einer Gel-Elektrophorese einer Tiling-Polymerasekettenreaktion (PCR) von vier Proben des Hepatitis C Virus des Genotyps 4, wovon die ersten beiden dem Subtyp 4a und die letzten beiden 4d zugehörig sind. Die Gel-Elektrophorese dient als Nachweis der erwarteten PCR-Produkte, welche aus ca. 400 Basenpaar langen Fragmenten bestehen. Ebenfalls zeigten sich einige größere Fragmente, die zufällig zwischen zwei nicht zusammengehörigen Primern entstanden sind.

In einem zweiten Versuch wurde das PCR-Produkt von jeweils drei weiteren Proben des GT 4a und GT 4d durch eine DNA-Quantifizierung mittels Lumineszenz Mikroplatten Reader von Berthold nachgewiesen (Vgl. 3.3.5). Anschließend erfolgte entsprechend der anderen Genotypen bei beiden Versuchen eine Sequenzierung mittels Nanopore, eine bioinformatische Auswertung sowie eine Analyse der Genomabdeckung.

Die Sequenz der fünf Proben des Genotyps 4d sowie vier Proben des Genotyps 4a konnte dabei fast vollständig mit einer hohe Coverage abgebildet werden, was am Beispiel der Proben mit IDU-Nr. 743 und Nr. 775 in Abbildung 14 gezeigt ist.

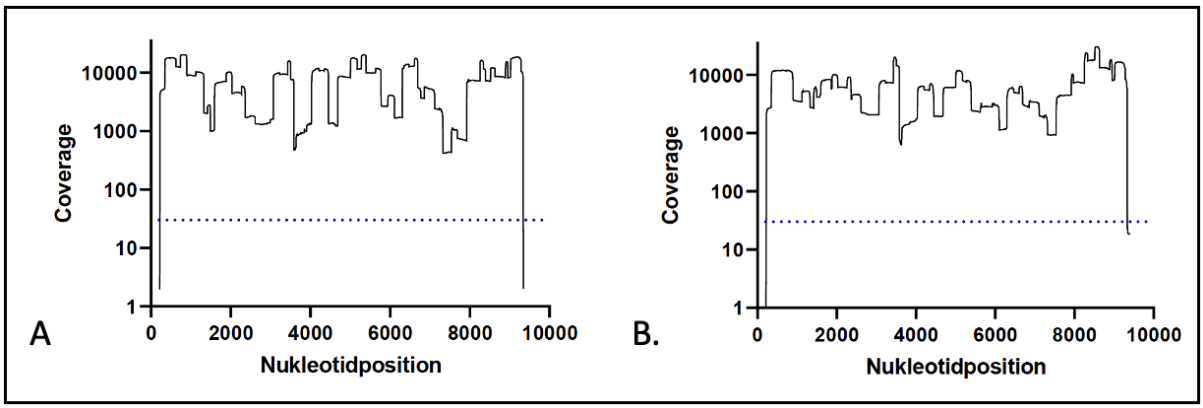

Abbildung 14: Coverage der Probe #743 des Genotyps 4d und #775 des Genotyps 4a
Analog zu Abbildung 3 ist die Anzahl der Reads pro Nukleotidposition (engl. *Coverage*) im Verhältnis zur Nukleotidposition auf dem Genom einer Tiling-Polymerasekettenreaktion und Sequenzierung einer Hepatitis C Probe #743 des Genotyps 4d (Teil **A**.) und der Probe #775 des Genotyps 4a (Teil **B.**) dargestellt. Eine Coverage < 30 wurde nur zu Beginn und Ende des Genoms unterschritten, dazwischen zeigt sich durchgängig eine hohe Coverage.

Da die Amplifikation der GT 4a- und GT 4d-Proben erfolgreich war, wurde entschieden, das Primerset für den GT 4 ohne weitere Korrekturen für die Überprüfung mit höherer Fallzahl zu behalten. In folgender Tabelle 10 ist das Primerset aufgeteilt in seine zwei Pools für den GT 4 dargestellt.

Tabelle 10: Primerset Genotyp 4
Abgebildet ist das Primerset einer Tiling-Polymerasekettenreaktion für eine Sequenzierung des Hepatitis C Virus des Genotyps 4. Es teilt sich in zwei Pools auf und es gibt eine Version für den Genotyp 4a und 4d, wobei sich der Primer HCV-4_5_LEFT unterscheidet.

Pool A		Pool B	
HCV-4_1_LEFT	HCV-4_1_RIGHT	HCV-4_2_LEFT	HCV-4_2_RIGHT
HCV-4_3_LEFT	HCV-4_3_RIGHT	HCV-4_4_LEFT	HCV-4_4_RIGHT
HCV-4_5_LEFT_4a HCV-4_5_LEFT_4d	HCV_4_5_RIGHT	HCV-4_6_LEFT	HCV-4_6_RIGHT
HCV-4_7_LEFT	HCV-4_7_RIGHT	HCV-4_8_LEFT	HCV-4_8_RIGHT
HCV-4_9_LEFT	HCV-4_9_RIGHT	HCV-4_10_LEFT	HCV-4_10_RIGHT
HCV-4_11_LEFT	HCV-4_11_RIGHT	HCV-4_12_LEFT	HCV-4_12_RIGHT
HCV-4_13_LEFT	HCV-4_13_RIGHT	HCV-4_14_LEFT	HCV-4_14_RIGHT
HCV-4_15_LEFT	HCV-4_15_RIGHT	HCV-4_16_LEFT	HCV-4_16_RIGHT
HCV-4_17_LEFT	HCV-4_17_RIGHT	HCV-4_18_LEFT	HCV-4_18_RIGHT
HCV-4_19_LEFT	HCV-4_19_RIGHT	HCV-4_20_LEFT	HCV-4_20_RIGHT
HCV-4_21_LEFT	HCV-4_21_RIGHT	HCV-4_22_LEFT	HCV-4_22_RIGHT
HCV-4_23_LEFT	HCV-4_23_RIGHT	HCV-4_24_LEFT	HCV-4_24_RIGHT
HCV-4_25_LEFT	HCV-4_25_RIGHT	HCV-4_26_LEFT	HCV-4_26_RIGHT
HCV-4_27_LEFT	HCV-4_27_RIGHT	HCV-4_28_LEFT	HCV-4_28_RIGHT
HCV-4_29_LEFT	HCV-4_29_RIGHT	HCV-4_30_LEFT	HCV-4_30_RIGHT

4.4 Überprüfung der Methode mit höherer Fallzahl

Nach erfolgreicher Etablierung der Primer-Sets wurden diese zur weiteren Überprüfung in einer größeren Kohorte validiert. Dazu wurden Proben der IDU-Kohorte aus Düsseldorf genutzt, deren Genotyp bereits bekannt war. Die IDU-Kohorte enthielt 95 Proben des Genotyps 1a, 17 Proben des Genotyps 1b, 86 Proben des Genotyps 3a und acht Proben des Genotyps 4.

Zur Analyse des Erfolgs der PCR und Sequenzierung wurde der sequenzierte Genomanteil als Messgröße ausgewählt. Der sequenzierte Genomanteil wird definiert durch die Anzahl vorkommender Ns in der Consensus Sequenz, in der ein N gebildet wird, sobald die Coverage, daher die Anzahl der Reads pro Nukleotidposition, < 30 liegt. Für die Ganzgenom-PCR wurde ein sequenzierter Genomanteil > 90 % als erfolgreich definiert. Dies ist angelehnt an die Qualitätsvorgabe des RKIs zu einer SARS-CoV2-Sequenzierung nach der Corona-Surveillance-Verordnung. [88]

Ein sequenzierter Genomanteil > 90 % bedeutet, dass an > 90 % der Nukleotidpositionen im HCV-Genom eine Coverage von mindestens 30 Reads pro Nukleotidposition vorhanden sind müssen. Zusätzlich wurden die Ergebnisse auf einen sequenzierten Genomanteil > 95 % hin untersucht.

4.4.1 Genotyp 1a

Zur Validierung des Primer-Sets für den Genotyp 1a wurde eine Tiling-PCR mit 95 Proben der Düsseldorfer IDU-Kohorte durchgeführt. Einen sequenzierten Genomanteil > 90 % erreichten 61 von 95 Proben (64,2 %) und einen sequenzierten Genomanteil > 95 % wurde von 24 von 95 Proben (25,3 %) erreicht. Als Mittelwert von allen 95 Proben wurde ein sequenzierter Genomanteil von 84 % ($\pm$ 19 %) bestimmt.

Bei einer Auftrennung dieser Werte nach der Viruslast zeigt sich, dass bei Proben mit einer Viruslast über 1.000.000 IU/ml 45 von 51 Proben (88,2 %) einen sequenzierten Genomanteil > 90 % und 20 von 51 Proben (39,2 %) einen sequenzierten Genomanteil > 95 % erreichten. Bei einer Viruslast zwischen 100.000 und 1.000.000 IU/ml wurde bei 15 von 32 Proben (46,9 %) ein sequenzierter Genomanteil > 90 % und bei vier von 32 Proben (12,5 %) ein sequenzierter Genomanteil > 95 % beobachtet. Proben mit einer Viruslast unter 100.000 IU/ml konnten nur in einem von zwölf Fällen (8,3 %) einen sequenzierten Genomanteil > 90 % erreichen und in keinem Fall > 95 %.

Dieser Zusammenhang zeigt sich ebenfalls in der statistisch signifikanten Spearman-Rangkorrelation ($r = 0{,}658$, $p < 0{,}001$ (2-seitig)) zwischen dem sequenzierten Genomanteil und der Viruslast, der in Abbildung 15 unter Teil A. dargestellt ist.

Eine weitere mögliche Einflussgröße auf den sequenzierten Genomanteil ist die Anzahl der Reads der Sequenzierung, daher die Readmenge. Bei einer Readmenge von > 100.000 / Sequenzierung konnten 60 von 83 Proben (72 %) einen sequenzierten Genomanteil von > 90 % erreichen, wohingegen dies nur eine von zwölf Proben bei einer Readmenge < 100.000 /Sequenzierung erreichen konnte. Bei Proben mit einer Readmenge > 100.000 / Sequenzierung konnte in 24 von 83 Proben (29 %) ein sequenzierter Genomanteil von > 95 % beobachtet werden, während dies von keiner Probe mit einer Readmenge < 100.000 / Sequenzierung erreicht wurde.

Die Spearman-Rangkorrelation zeigt einen statistisch-signifikanten Zusammenhang ($r = 0{,}686$, $p < 0{,}001$ (2-seitig)) zwischen dem sequenzierten Genomanteil und der Readmenge und ist in Abbildung 15 unter Teil B. gezeigt. Auch eine Korrelation zwischen der Readmenge und der Viruslast zeigte sich als statistisch signifikant ($r = 0{,}436$, $p < 0{,}01$ (2-seitig)) und ist in Abbildung 15 unter Teil C. wiedergegeben. Bei Vergleich der Zusammenhänge fällt auf, dass die Daten zwischen dem sequenzierten Genomanteil und der Viruslast sowie der Readmenge stärker korrelieren als die zwischen der Readmenge und der Viruslast.

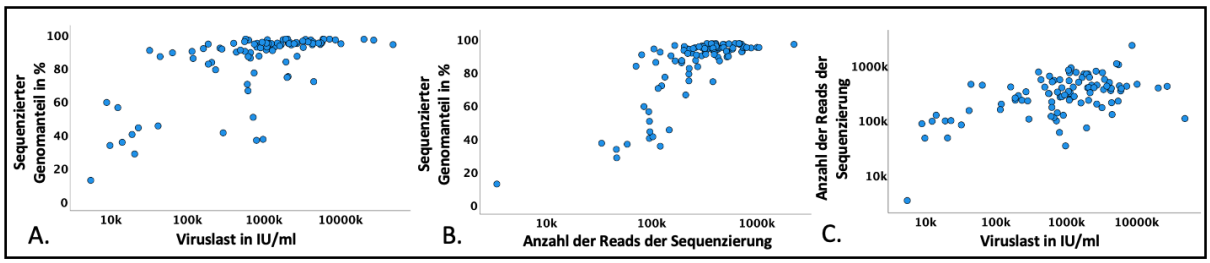

Abbildung 15: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 1a

Dargestellt sind Korrelationen zwischen dem sequenzierten Genomanteil in % und der Viruslast in IU/ml (**A.**), dem sequenzierten Genomanteil in % und der Readmenge (**B**.) sowie der Readmenge und der Viruslast in IU/ml (**C.**) von Proben der injection-drug-use (IDU) – Kohorte Düsseldorf des Genotyps 1a.

In der statistischen Auswertung zeigte sich, dass die Abdeckung in einigen Proben, besonders denen mit geringer Viruslast, niedrig ist, weshalb zur Erhöhung des sequenzierten Genomanteils ein zusätzlicher 4. Primer-Pool gebildet werden sollte. Dieser Pool sollte die Regionen zur Genotypisierung sowie Resistenztestung abdecken, weshalb er aus Primern für die Regionen E1, NS3, NS5A, NS5B und Core bestand, welche in Tabelle 2 des Anhangs aufgeführt sind. Diese Primer sind im Institut bereits zur Bestimmung des Genotyps bzw. zur Resistenztestung etabliert gewesen. Für 14 Proben wurde eine Tiling-PCR mit vier Pools und gleichzeitig zur Kontrolle eine Tiling-PCR mit den drei bekannten Pools durchgeführt. Anschließend erfolgte eine Sequenzierung mittels Nanopore sowie die bioinformatische Auswertung.

Dabei zeigte sich, dass ein sequenzierter Genomanteil von > 90 % von elf der 14 (78,6 %) Proben in dem Versuch mit dem zusätzlichen Pool 4 und von nur einer der 14 Proben im Versuch ohne den vierten Pool erreicht wurde. Einen sequenzierten Genomanteil von > 95 % erreichten mit dem 4. Pool sieben von 14 Proben (50 %), während keine von diesen Proben einen solchen sequenzierten Genomanteil ohne den 4. Pool erzielt hatte. Im Durchschnitt erreichten die 14 Proben im Versuch mit dem zusätzlichen Pool 4 einen sequenzierten Genomanteil von 91 % (± 11 %) und im Versuch ohne den Pool 4 einen sequenzierten Genomanteil von 59 % (± 22 %).

Wie bereits für die Versuche mit dem Primerset aus drei Pools beschrieben, hatte die Viruslast auch Einfluss auf den Sequenziererfolg, wenn der 4. Primer-Pool zusätzlich eingesetzt wurde. So wurde ein sequenzierter Genomanteil von > 95 % bei fünf von fünf Proben (100 %) bei einer Viruslast > 1.000.000 IU/ml erreicht, von zwei von drei Proben (40 %) mit einer Viruslast zwischen 100.000 und 1.000.000 IU/ml und von keiner von vier Proben (0 %) mit einer Viruslast < 100.000 IU/ml. Trotz dieses Zusammenhangs war durch den vierten Pool eine erfolgreiche Sequenzierung bei niedrigerer Viruslast beobachtbar. So hatte die einzige erfolgreiche Probe ohne den vierten Pool eine Viruslast von > 1.000.000 IU/ml, genauer 2.382.595 IU/ml. Im Vergleich konnte ein sequenzierter Genomanteil von > 90 % bei einer Probe mit dem vierten Pool mit der niedrigsten Viruslast von 42.625 IU/ml bestimmt werden. Auch die Anzahl der Reads der Sequenzierung beeinflusste die Höhe des sequenzierten Genomanteils. Alle Proben mit oder ohne den 4. Pool, die einen sequenzierten Genomanteil > 90 % oder > 95 % erreichten, hatten eine Readmenge von > 100.000 / Sequenzierung.

Für die Tiling-PCR mit dem 4. Pool konnte, wie für die ursprüngliche PCR mit drei Pools, eine statistische signifikante, positive Korrelation zwischen dem sequenzierten Genomanteil und der Viruslast in IU/ml beobachtet werden (mit Pool 4: r = 0,675, p = 0,008 (2-seitig); ohne Pool 4: r = 0,864, p < 0,01 (2-seitig)), wie in Teil A. der Abbildung 16 ersichtlich ist. Ebenfalls wurde eine signifikante, positive Korrelation zwischen dem sequenzierten Genomanteil und der Anzahl der Reads beobachtet (mit Pool 4: r= 0,670, p = 0,009 (2-seitig), ohne Pool 4: r = 0,807, p < 0,001 (2-seitig)). Die Korrelation zwischen der Readmenge und der Viruslast zeigte sich nur in der Versuchsreihe mit dem 4. Pool als signifikant (mit Pool 4: r = 0,727, p = 0,003 (2-seitig), ohne Pool 4: r = 0,433, p = 0,122 (2-seitig)). Dies ist in folgender Abbildung 16 unter Teil B. und C. dargestellt.

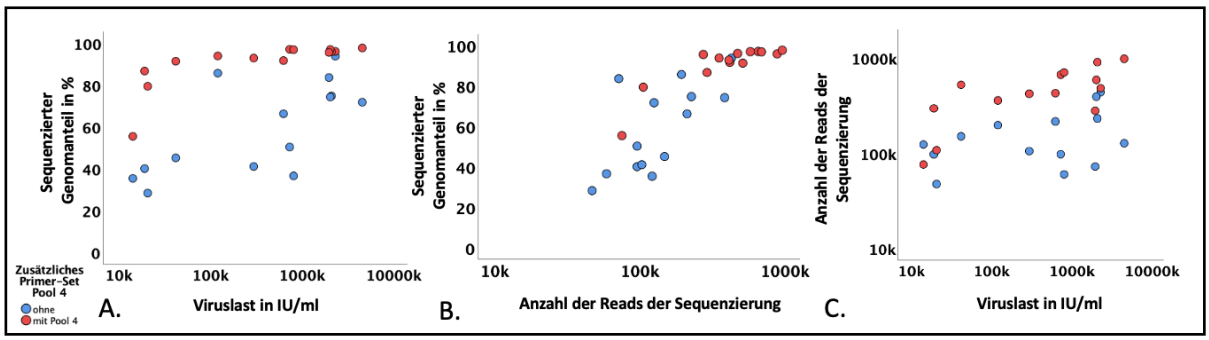

Abbildung 16: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 1a mit und ohne vierten Primer-Pool
Dargestellt ist der Vergleich einer Sequenzierung mit (rot) und ohne (blau) eines zusätzlichen vierten Primer-Pools des Genotyps 1a. Dazu sind Zusammenhänge zwischen dem sequenzierten Genomanteil in % und der Viruslast in IU/ml (**A.**), dem sequenzierten Genomanteil in % und der Readmenge (**B.**) sowie der Readmenge und der Viruslast in IU/ml (**C.**) von Proben der injection-drug-use (IDU) – Kohorte Düsseldorf des Genotyps 1a abgebildet. Mit dem 4. Primer-Pool wurden höhere Ergebnisse in dem sequenzierten Genomanteil und der Readmenge erreicht.

Zur Untersuchung, ob der vierte Pool in den relevanten Regionen des HCV-Genoms einen positiven Einfluss hat, wurden Mittelwerte der Coverage, daher der Anzahl der Reads pro Nukleotidposition, in den Regionen Core, E1, NS3, NS5A und NS5B bestimmt. Hierzu wurden die Nukleotid-Abschnitte mit Hilfe der Referenz AF009606 (HCV-H77) festgelegt. Eine minimale Coverage für eine erfolgreiche Sequenzierung wurde als 30 definiert (Vgl. 4.2) und konnte in jeder Region mit und ohne dem 4.Pool erreicht werden. Als gewünschte Coverage wurde eine Cut-Off von 10.000 definiert, bei dessen Wert auch eine Resistenztestung möglich wäre. Dies konnte für jede der fünf Regionen nur mit dem 4. Pool erreicht werden. In allen fünf Regionen zeigt sich im Mittel eine höhere Coverage, wenn der 4. Primer-Pool Teil der PCR war, was in folgender Tabelle 11 dargestellt ist.

Tabelle 11: Vergleich der Mittelwerte der Coverage von Genotyp 1a mit und ohne Pool 4
Dargestellt sind die Mittelwerte der Coverage von Proben der injection drug-use (IDU) – Kohorte Düsseldorf des Genotyps 1a von einer Tiling-Polymerasekettenreaktion mit und ohne eines vierten Primer-Pools, der spezifische Primer für diese Regionen enthielt. Die Nukleotidpositionen der relevanten Proteine von Hepatitis C wurden mit Hilfe der Referenz AF009606 (H77) bestimmt. Die Coverage zeigt sich mit dem vierten Pool erhöht.

Protein	Nukleotidposition	Coverage-Mittelwert ohne Pool 4	Coverage-Mittelwert mit Pool 4
Core	342-914	11915	19969
E1	915-1490	20883	32274
NS3	3420-5312	7955	11605
NS5A	6258-7601	11887	19332
NS5B	7602-9374	12689	18340

Mit Hilfe des Programms Geno2Pheno (Vgl. 3.3.9) wurde anhand der gebildeten Sequenzen eine Resistenztestung durchgeführt und die Anzahl der Resistenzen für die fünf zugelassenen Substanzen Sofosbuvir, Velparasvir, Pibrentasvir, Voxilaprevir und Glecaprevir in Tabelle 12 dargestellt. Im Vergleich fällt auf, dass ohne den 4. Pool bei den Substanzen Voxila- und Glecaprevir eine Resistenz beobachtet wurde, die im Versuch mit dem 4. Pool nicht vorhanden war. Beide Resistenzen traten innerhalb einer Probe (VI-Nr. #1024) auf, in der auch noch einige weitere Resistenzen gegen andere Substanzen auftraten. Die vorliegenden Mutationen (122R, 155K, 156V, 168L, 170L für Voxilaprevir und 166V, 168L für Glecaprevir) lagen allesamt im Nukleotidabschnitt des Zielproteins NS3. In den Versuchen ohne den 4. Pool wurde dieser Abschnitt im Durchschnitt am schlechtesten abgedeckt, wogegen mit dem 4. Pool eine mittlere Coverage von > 10.000 erreicht wurde (Vgl. Tabelle 11). Der zu geringe sequenzierte Genomanteil ohne den 4. Pool führte daher vermutlich zu einer Fehlmessung einzelner Basen, die zu einem falsch-positiven Ergebnis der Resistenztestung führten. Mit dem 4. Pool können dagegen durch den hohen sequenzierten Genomanteil auch kleine Sequenzunterschiede sicher bestimmt werden.

Tabelle 12: Vergleich Resistenztestung mit und ohne Pool 4
Dargestellt sind die Anzahl der beobachteten Resistenzen von fünf zugelassenen Substanzen bei einer Resistenztestung von jeweils 10 Proben des Genotyps 1a. Die Sequenz ist mit Hilfe einer Tiling-PCR sowie Sequenzierung gebildet und die Resistenztestung anschließend mittels Geno2Pheno erstellt worden. Bei der Tiling-PCR erfolgten zwei verschiedene Ansätze, ohne und mit einem 4. Pool, der zusätzliche Primer für Resistenz-relevante Regionen enthielt.

Substanz (Zielprotein)	**Sofosbuvir (NS5B)**	**Velpatasvir (NS5A)**	**Pibrentasvir (NS5A)**	**Voxilaprevir (NS3)**	**Glecaprevir (NS3)**
Ohne Pool 4	0	0	0	1	1
Mit Pool 4	0	0	0	0	0

Zusammenfassend konnte für den Genotyp 1a durch eine Tiling-PCR und Sequenzierung bei 61 von 95 Proben (64,2 %) das Genom zu > 90 % abgedeckt werden. Positiven Einfluss auf den sequenzierten Genomanteil hatten dabei eine hohe Viruslast der Proben sowie eine hohe Readmenge pro Sequenzierung. Ein höherer sequenzierter Genomanteil wurde außerdem durch die Einführung eines vierten Pools mit Primern der Regionen E1, NS3, NS5A, NS5B und Core erreicht. Im Versuch mit dem zusätzlichen 4. Primer-Pool konnte in elf von 14 Proben (78,6 %) ein sequenzierter Genomanteil von > 90 % erreicht werden.

4.4.2 Genotyp 1b

In der Etablierung des Primer-Sets für eine Tiling-PCR des Genotyps 1b wurde sich nach einem Vergleich für das alte Primer-Set entschieden (Vgl. 4.3.2), welches an 17 Proben der IDU-Kohorte überprüft werden sollte. Dabei wurde ein sequenzierter Genomanteil von > 90 % bei 13 von 17 Proben (76,5 %) und ein sequenzierter Genomanteil von > 95 % bei acht von 17 Proben (47,1 %) erreicht. Im Durchschnitt wurde ein sequenzierter Genomanteil von 93 % (± 4 %) gemessen.

Wie bei Genotyp 1a wurde der Einfluss der Viruslast sowie der Readmenge auf den sequenzierten Genomanteil untersucht, dabei ist zu berücksichtigen, dass alle 17 Proben eine Viruslast > 100.000 IU/ml aufwiesen. Bei Proben mit einer Viruslast von > 1.000.000 IU/ml wurde in sechs von neun Fällen (66,7 %) ein sequenzierter Genomanteil von > 90 % und in vier von neun Fällen (44,4 %) ein sequenzierter Genomanteil von > 95 % beobachtet. Bei Proben mit einer Viruslast zwischen 100.000 und 1.000.000 IU/ml erreichten sieben von acht Fälle (87,5 %) einen sequenzierten Genomanteil von > 90 % und vier von acht Fälle (50 %) einen sequenzierten Genomanteil von > 95 %. Anders als für den Genotyp 1a besteht für die 17 Proben des Genotyps 1b eine nicht-signifikanten Spearman-Rangkorrelation (r = 0,131, p = 0,620 (2-seitig)) zwischen der Viruslast und dem sequenzierten Genomanteil. In Teil A. der Abbildung 17 ist dieser Zusammenhang graphisch darstellt, in dem dennoch ein positiver Trend erkennbar ist.

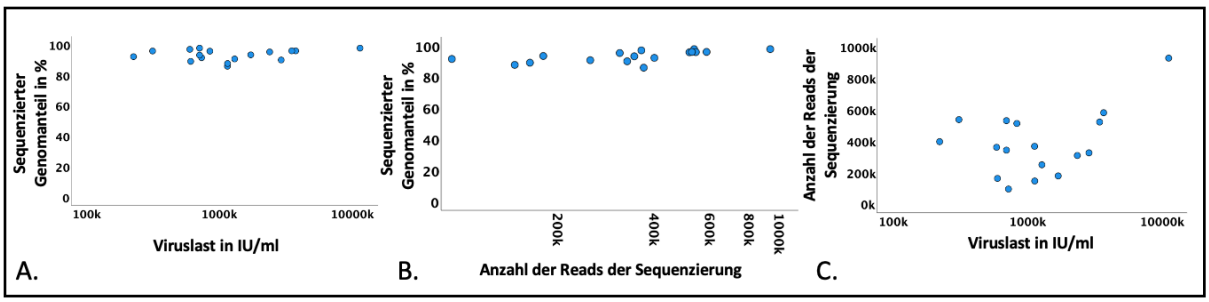

Abbildung 17: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 1b
Dargestellt sind Zusammenhänge zwischen dem sequenzierten Genomanteil in % und der Viruslast in IU/ml (**A.**), dem sequenzierten Genomanteil in % und der Readmenge (**B**.) sowie der Readmenge und der Viruslast in IU/ml (**C.**) von Proben der injection-drug-use (IDU) – Kohorte Düsseldorf des Genotyps 1b.

Eine Readmenge von < 100.000 / Sequenzierung trat nur bei einer der 17 Proben auf und zeigte dabei weder einen sequenzierten Genomanteil von > 90 % noch von > 95 %. Eine Korrelation zwischen dem sequenzierten Genomanteil und der Anzahl der Reads der Sequenzierung zeigte sich als positiv und signifikant (r= 0,717, p = 0,001 (2-seitig)) und ist in Teil B. der Abbildung 17 gezeigt. Zwischen der Readmenge und der Viruslast konnte kein

statistisch signifikanter Zusammenhang nachgewiesen werden (r= 0,137, p = 0,599 (2-seitig)), dennoch ist graphisch ein positiver Trend in Teil C. der Abbildung 17 erkennbar.

Zur spezifischen Analyse des sequenzierten Genomanteil des Sequenzierprodukts wurde ein Mittelwert der Coverage der 17 Proben in besonders relevanten Regionen bestimmt, was in Tabelle 13 dargestellt ist. Zur Definition der Position der Nukleotidabschnitte wurde sich dabei auf das Referenzgenom AF009606 (HCV-H77) bezogen. Der minimale Cut-Off einer Coverage von 30 für eine erfolgreiche Sequenzierung wurde in allen untersuchten Regionen erreicht. Eine Coverage von 100.000, wie sie für eine Resistenztestung nötig wäre, wurde in den Bereichen E1, NS5A und NS5B gemessen.

Tabelle 13: Mittelwerte der Coverage des Genotyps 1b
Dargestellt sind die Mittelwerte der Coverage von Proben der injection drug-use (IDU) – Kohorte Düsseldorf des Genotyps 1b von einer Tiling-Polymerasekettenreaktion und Sequenzierung. Die Nukleotidpositionen der relevanten Proteine von Hepatitis C wurden mit Hilfe der Referenz AF009606 (H77) bestimmt.

Protein	Nukleotidposition	Mittelwert der Coverage des GT 1b
Core	342-914	3886
E1	915-1490	16249
NS3	3420-5312	8896
NS5A	6258-7601	20552
NS5B	7602-9374	21789

Zusammenfassend konnte für den Genotyp 1b eine Tiling-PCR und Sequenzierung bei 13 von 17 Proben (76,5 %) das Genom zu > 90 % abdecken. Dabei zeigte sich ein positiver Einfluss der Readmenge pro Sequenzierung auf den sequenzierten Genomanteil. In Gegensatz dazu konnte keine statistisch signifikante Korrelation zwischen der Viruslast und dem sequenzierten Genomanteil der Proben nachgewiesen werden, dennoch zeigte sich graphisch ein positiver Trend. Für die Genomabschnitte E1, NS5A und NS5B konnte im Durchschnitt eine besonders hohe Coverage erreicht werden, mit der eine Resistenztestung möglich wäre.

4.4.3 Genotyp 3a

Zur Validierung der Tiling-PCR mit dem erfolgreich etablierten Primer-Set wurde eine PCR mit 86 Genotyp 3a-Proben durchgeführt. Dabei konnte bei 84 von 86 Proben (97,7 %) ein sequenzierter Genomanteil von > 90 % und bei 77 von 86 Proben (89,5 %) ein sequenzierter Genomanteil von > 95 % beobachtet werden. Im Durchschnitt wurde ein sequenzierter Genomanteil von 98 % (± 3 %) erreicht. Wie bereits für Genotyp 1a und 1b gezeigt, kann die Viruslast sowie die Anzahl der Reads der Sequenzierung Einfluss auf den sequenzierten Genomanteil haben und wurde deshalb statistisch genauer untersucht. Bei drei Proben war die Viruslast nicht bekannt, weshalb sie nicht in die Viruslast-abhängige Auswertung einbezogen werden konnten.

Eine Analyse, getrennt nach der Viruslast zeigt, dass ein sequenzierter Genomanteil von > 90 % in 38 von 38 Fällen (100 %) mit einer Viruslast von > 1.000.000 IU/ml, in 24 von 25 Fällen (96 %) mit einer Viruslast zwischen 100.000 und 1.000.000 IU/ml und in 19 von 20 Proben (95 %) mit einer Viruslast < 100.000 IU/ml erreicht werden konnte. Ein sequenzierter Genomanteil von > 95 % wurde in 38 von 38 Fällen (100 %) mit einer Viruslast > 1.000.000 IU/ml, in 22 von 25 Fällen (88 %) mit einer Viruslast zwischen 100.000 und 1.000.000 IU/ml und in 15 von 20 Fällen (75 %) mit einer Viruslast < 100.000 IU/ml beobachtet. Eine Spearman-Rangkorrelation zeigte einen positiven, statistisch signifikanten Zusammenhang ($r = 0{,}499$, $p < 0{,}001$ (2-seitig)) zwischen dem sequenzierten Genomanteil in % zur Viruslast in IU/ml, welchen der Teil A. der Abbildung 18 wiedergibt.

Eine Aufteilung nach der Readmenge zeigt, dass ein sequenzierter Genomanteil von > 90 % bei 80 von 82 Proben (97,6 %) mit einer Readmenge > 100.000 / Sequenzierung und bei vier von vier Proben (100 %) mit einer Readmenge < 100.000 / Sequenzierung beobachtet wurde. Ein sequenzierter Genomanteil von > 95 % wurde von 74 von 82 Proben (90,2 %) mit einer Readmenge > 100.000 / Sequenzierung und in drei von vier Proben (75 %) mit einer Readmenge < 100.000 / Sequenzierung erreicht. Ebenfalls konnte ein statistisch signifikanter Zusammenhang ($r = 0{,}322$, $p = 0{,}002$ (2-seitig)) zwischen dem sequenzierten Genomanteil und der Anzahl der Sequenzierreads dargestellt werden. Diesen Zusammenhang zeigt Teil B. der Abbildung 18.

Die Korrelation der Readmenge zur Viruslast stellt sich für den Genotyp 3a als nicht signifikant ($r = 0{,}184$, $p = 0{,}095$(2-seitig)) dar, graphisch zeigt sich aber ein positiver Trend, der in folgender Abbildung 18 unter Teil C. dargestellt ist.

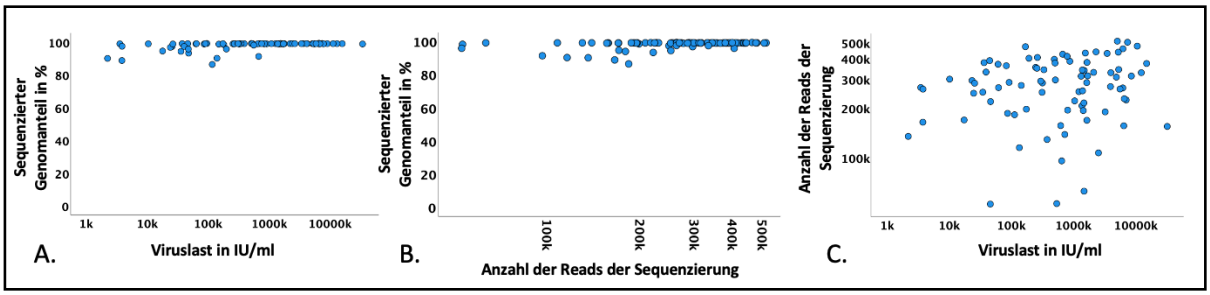

Abbildung 18: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 3a
Dargestellt sind Zusammenhänge zwischen dem sequenzierten Genomanteil in % und der Viruslast in IU/ml (**A.**), dem sequenzierten Genomanteil in % und der Readmenge (**B.**) sowie der Readmenge und der Viruslast in IU/ml (**C.**) von Proben der injection-drug-use (IDU) – Kohorte Düsseldorf des Genotyps 3a abgebildet.

Zur Überprüfung, wie effektiv die Tiling-PCR diagnostisch relevante Abschnitte des HCV-Genoms abdeckt, wurde ein Mittelwert der Coverage der Regionen Core, E1, NS3, NS5A und NS5B gebildet. Zur Berechnung dieses Mittelwerts wurden die Nukleotidpositionen der Regionen mit Hilfe der Referenzsequenz X76918 definiert. Die mittlere Coverage der fünf Regionen überschritten den minimalen Cut-Off von 30 für eine erfolgreiche Sequenzierung deutlich. Die Region Core und NS3 stellen sich dabei als stärker abgebildet dar, aber nur die Region Core zeigt durchschnittlich eine Coverage > 10.000, die für eine Resistenztestung ausreichend wäre.

Tabelle 14: Mittelwerte der Coverage des Genotyps 3a
Dargestellt sind die Mittelwerte der Coverage von Proben der injection drug-use (IDU) – Kohorte Düsseldorf des Genotyps 3a der Tiling-Polymerasekettenreaktion. Die Nukleotidpositionen der relevanten Proteine von Hepatitis C wurden mit Hilfe der Referenz X76918 bestimmt.

Protein	Nukleotidposition	Mittelwert der Coverage des GT 3a
Core	299-871	10504
E1	872-1447	7735
NS3	3395-5287	9570
NS5A	6233-7588	6314
NS5B	7589-9361	8441

Zur weiteren Überprüfung der Methode wurden zehn Proben mit dem geringsten sequenzierten Genomanteil ausgewählt und wiederholt. Dabei konnte eine Veränderung des sequenzierten Genomanteils beobachtet werden. Ein sequenzierter Genomanteil von > 90 % wurde bei acht von zehn Proben (80 %) im ersten Versuch und bei sieben von zehn (70 %) im zweiten Versuch erreicht. Ein sequenzierter Genomanteil von > 95 % wurde im ersten Versuch nur bei einer von zehn Proben (10 %) erreicht, wogegen dies im zweiten Versuch zu bei vier von zehn Proben (40 %) beobachtet werden konnte. Zum Vergleich des Verfahrens gegen eine andere Sequenziermethode wurde der gleiche PCR-Ansatz dieser zehn Proben anschließend zur Illumina-Sequenzierung verschickt. (Vgl. 4.6.2)

Zusammengefasst zeigte die Tiling-PCR des Genotyps 3a in 84 von 86 Proben (97,9 %) einen sequenzierten Genomanteil von > 90 % und funktioniert daher sehr erfolgreich. Positive Einflussfaktoren stellen eine hohe Viruslast der Probe als auch eine hohe Anzahl der Reads der Sequenzierung dar. Für die Genomabschnitte Core konnte im Durchschnitt eine besonders hohe Coverage erreicht werden, mit der eine Resistenztestung möglich wäre.

4.4.4 Genotyp 4

Nach erfolgreicher Etablierung der Primersets sollte zur Prüfung der Methode eine Tiling-PCR mit 18 Genotyp 4 -Proben durchgeführt werden. Da der Genotyp 4 in der Düsseldorfer IDU-Kohorte selten vorkommt, wurden für den Genotyp 4 acht unbekannte und fünf aus der Etablierung des Primersets bekannte Proben verwendet. Für den Genotyp 4a konnten ausschließlich fünf bekannte Proben einbezogen werden, da es keine weiteren Proben dieses Genotyps in der Düsseldorfer IDU-Kohorte gab.

Im gesamten Genotyp 4 wurde bei 13 von 18 Proben (72,2 %) ein sequenzierter Genomanteil von > 90 % und bei elf von 18 Proben (61,1 %) ein sequenzierter Genomanteil von > 95 % beobachtet. Im Durchschnitt wurde dabei ein sequenzierter Genomanteil von 91 % (± 9 %) erzeugt. Die einzelne Betrachtung beider Subtypen 4a und 4d zeigte ähnliche Ergebnisse, wobei der GT 4a einen etwas niedrigeren sequenzierten Genomanteil zeigte. So erreichte der GT 4a jeweils in drei von fünf Proben (60 %) einen sequenzierten Genomanteil von über 90 und 95 % und im Durchschnitt einen sequenzierten Genomanteil von 89 % (± 13 %). Der GT 4d zeigte bei zehn von 13 (76,9 %) Proben einen sequenzierten Genomanteil von > 90 %, bei acht von 13 Proben (61,5 %) einen sequenzierten Genomanteil von > 95 % und im Durchschnitt einen sequenzierten Genomanteil von 92 % (± 7 %).

Eine getrennte Analyse nach der Viruslast zeigt für den GT 4a, dass ein sequenzierter Genomanteil von > 90 % bei einer von zwei Proben (50 %) mit einer Viruslast > 1.000.000 IU/ml und bei zwei von drei Proben (66,7 %) mit einer Viruslast zwischen 100.000 und 1.000.000 IU/ml erreicht wurde.

Für den GT 4d wurde ein sequenzierter Genomanteil von > 90 % bei einer Viruslast > 1.000.000 IU/ml in neun von elf Proben (81,8 %) und bei einer Viruslast zwischen 100.000 IU/ml und 1.000.000 IU/ml in einem von zwei Fällen (50 %) erreicht. Ein sequenzierter Genomanteil von > 95 % wurde für den GT 4a in gleichem Maße wie der sequenzierte Genomanteil von > 90 % erreicht. Für den GT 4d wurde ein sequenzierter Genomanteil von > 95 % bei sieben von elf Proben (63,6 %) mit einer Viruslast > 1.000.000 IU/ml und bei einer von zwei Proben (50 %) mit einer Viruslast zwischen 100.000 IU/ml und 1.000.000 IU/ml erreicht.

Bei weiteren gemeinsamen, statistischen Analysen des Genotyps 4 zeigten sich die Spearman-Rangkorrelationen zwischen dem sequenzierten Genomanteil und der Viruslast, (r = -0,266, p = 0,285 (2-seitig)), dem sequenzierten Genomanteil und der Readmenge (r = 0,207, p = 0,410 (2-seitig)) sowie der Readmenge und der Viruslast (r = 0,399, p = 0,101 (2-seitig)) als nicht

signifikant. Dabei wies eine graphische Tendenz bezüglich des Zusammenhangs zwischen dem sequenzierten Genomanteil und der Viruslast auf einen negativen Verlauf hin und zeigte sich damit anders als bei den Graphen der weiteren beobachteten Genotypen. Dies ist in Teil A. der Abbildung 19 dargestellt.

Eine Readmenge < 100.000 / Sequenzierung wurde nur bei einer Probe für den GT 4d beobachtet, die einen sequenzierten Genomanteil von > 95 % erreichte. Alle anderen Proben hatten eine Readmenge > 100.000 / Sequenzierung und zeigten dabei teilweise einen geringeren sequenzierten Genomanteil. Die graphischen Zusammenhänge zwischen dem sequenzierten Genomanteil und der Readmenge sowie der Readmenge und der Viruslast gaben Hinweise auf einen ähnlichen, positiven Trend zum Verlauf der Graphen, die bei den anderen Genotypen beobachtet worden sind, dies ist in folgender Abbildung 19 unter Teil B. und C. ersichtlich.

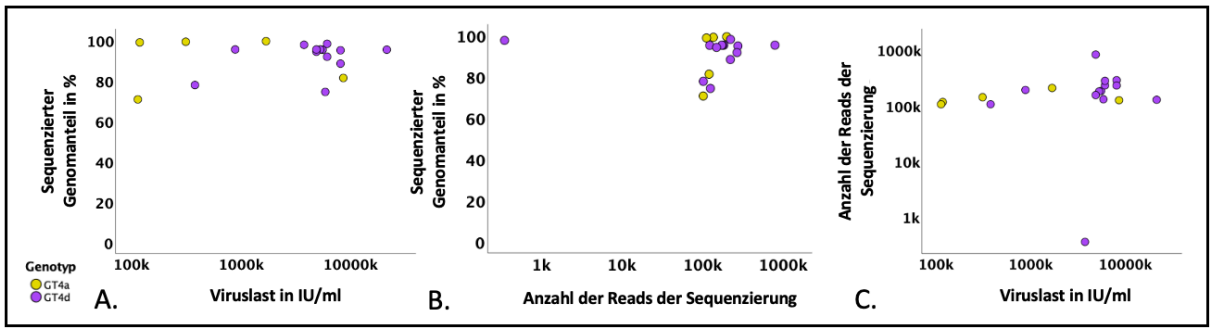

Abbildung 19: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 4

Dargestellt sind Ergebnisse einer Sequenzierung von Proben der injection-drug-use (IDU) – Kohorte Düsseldorf des Genotyps 4a (gelb) und 4d (lila). Es sind Zusammenhänge zwischen dem sequenzierten Genomanteil in % und der Viruslast in IU/ml (**A.**), dem sequenzierten Genomanteil in % und der Readmenge (**B.**) sowie der Readmenge und der Viruslast in IU/ml (**C.**) abgebildet.

Um zu untersuchen, wie effektiv diagnostisch relevante Abschnitte des HCV-Genoms durch die Tiling-PCR abdeckt werden, wurde ein Mittelwert der Coverage der Regionen Core, E1, NS3, NS5A und NS5B gebildet. Zur Definition der Nukleotidpositionen dieser Regionen wurde die Referenzsequenz JX227964 verwendet. Tabelle 15 stellt die mittleren Messwerte der Coverage der verschiedenen Regionen gegenüber. In allen fünf Regionen wurde eine Coverage weit > 30 erreicht, die als Minimum für eine erfolgreiche Sequenzierung definiert wurde. In den Regionen Core und NS5B konnte außerdem eine mittlere Coverage > 10.000 erreicht werden, die z.B. für eine Resistenztestung ausreichend wäre.

Tabelle 15: Mittelwerte der Coverage des Genotyps 4
Dargestellt sind die Mittelwerte der Coverage von Proben der injection drug-use (IDU) – Kohorte Düsseldorf des Genotyps 4 der Tiling-Polymerasekettenreaktion. Die Nukleotidpositionen der relevanten Proteine von Hepatitis C wurden mit Hilfe der Referenz JX227964 bestimmt.

Protein	Nukleotidposition	Mittelwert der Coverage des Genotyps 4
Core	275-847	13623
E1	848-1423	6858
NS3	3353-5245	7043
NS5A	6191-7534	5176
NS5B	7535-9307	11089

Zusammengefasst konnte der Genotyp 4 durch eine Tiling-PCR bei 13 von 18 Proben (72,2 %) eine Abdeckung des gesamten Genoms von > 90 % erreichen. Zusammenhänge zwischen dem sequenzierten Genomanteil, der Viruslast und der Readmenge waren nicht statistisch-signifikant, was durch die geringe Fallzahl der Genotyps 4 bedingt sein könnte. Für die Genomabschnitte Core und NS5B konnte im Durchschnitt eine besonders hohe Coverage erreicht werden, mit der eine Resistenztestung möglich wäre.

4.4.5 Zusammenfassende Gegenüberstellung sowie Mittelwerte der Genotypen

Um den Sequenziererfolg der fünf untersuchten Genotypen GT 1a, GT 1b, GT 3a, GT 4a und GT 4d gemeinsam darzustellen und zu vergleichen, wurden aus den vorangegangen Ergebnissen der Überprüfung mit einer höheren Fallzahl (Vgl. 4.4.1, 4.4.2, 4.4.3, 4.4.4) relative Häufigkeiten des sequenzierten Genomanteils von > 90 % und von > 95 % errechnet. In folgender Abbildung 20 wurden diese einander gegenübergestellt und zusätzlich die Fallzahl der untersuchten Proben angegeben. Dabei fällt auf, dass der Genotyp 3a am häufigsten eine erfolgreiche Ganzgenom-Sequenzierung aufweist. Ebenfalls zeigen der Genotyp 4a und 4d häufiger einen höheren sequenzierten Genomanteil von > 95 % als der Genotyp 1a und 1b. Ein sequenzierter Genomanteil von > 90 % wird dagegen durch den Genotyp 1a und 1b häufiger erreicht als durch den Genotyp 4a, da für den Genotyp 4a gleich häufig ein sequenzierter Genomanteil von > 90 % und von > 95 % beobachtet wurde.

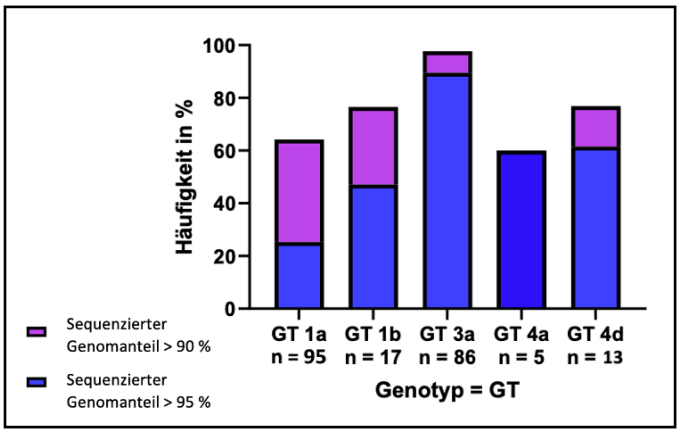

Abbildung 20: Sequenziererfolg Tiling-PCR Gesamt
Dargestellt sind relative Häufigkeiten in % des Erfolgs einer Genotyp-spezifischen Tiling-Polymerasekettenreaktion (PCR) des Hepatitis C Virus der Genotypen (GT) 1a, 1b, 3a, 4a und 4d. Als Erfolg wurde der sequenzierte Genomanteil definiert, bei dem in > 90 % (lila) oder > 95 % (blau) des gesamten Genoms eine ausreichende Coverage (mind. > 30 Reads pro Nukleotidposition) vorlag. Des Weiteren sind die Fallzahlen n der erfolgten Experimente angegeben.

Eine gemeinsame Auswertung der Ergebnisse aller Proben der fünf Genotypen zeigt, dass durchschnittlich ein sequenzierter Genomanteil von 91 % (± 15 %) erreicht wurde. Ein sequenzierter Genomanteil von > 90 % wurde dabei in 171 von 216 Proben (79 %) erreicht und ein sequenzierter Genomanteil von > 95 % in 117 von 216 Proben (54 %).

Im Rahmen der statistischen Auswertung wurden Korrelationen zwischen dem sequenzierten Genomanteil, der Viruslast und der Readmenge untersucht. Bei den Genotypen mit hoher Fallzahl, daher Genotyp 1a und 3a, konnten jeweils positive, statistisch signifikante Korrelationen zwischen diesen Variablen beobachtet werden. Für den Genotyp 1b, 4a und 4d zeigt sich dieser Zusammenhang als nicht-statistisch signifikant, aber mit einem graphisch

positiven Trend, weshalb vermutlich die zu geringe Fallzahl eine Signifikanz verhindert. Daraus lässt sich schließen, dass eine Einschränkung des Probenkollektivs hinsichtlich ihrer Viruslast sinnvoll sein könnte, um bessere Sequenziererfolge zu ermöglichen. In folgender Abbildung 21 sind daher relative Häufigkeiten des Sequenziererfolgs aufgetrennt nach der Viruslast der Proben dargestellt. In Teil A. finden sich die Ergebnisse von Proben mit einer Viruslast zwischen 100.000 IU/ml und 1.000.000 IU/ml, in Teil B. die der Proben mit einer Viruslast > 1.000.000 IU/ml. Es fällt auf, dass für die Genotypen 1a, 1b, 3a und 4d häufiger ein sequenzierter Genomanteil von > 90 % und von > 95 % erreicht wurde, wenn eine Viruslast > 1.000.000 IU/ml bestand. Lediglich für den Genotyp 4a zeigte sich dieser Zusammenhang nicht, was durch die geringe Fallzahl begründet sein könnte.

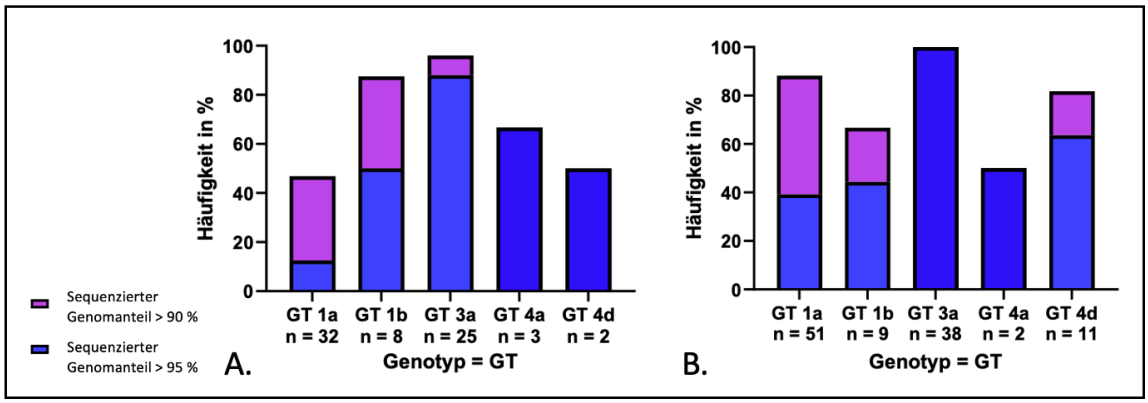

Abbildung 21: Sequenziererfolg Tiling-PCR unterteilt nach Viruslast
Dargestellt sind relative Häufigkeiten in % des Erfolgs einer Genotyp-spezifischen Tiling-Polymerasekettenreaktion (PCR) des Hepatitis C Virus der Genotypen (GT) 1a, 1b, 3a, 4a und 4d. Als Erfolg wurde der sequenzierte Genomanteil definiert, bei dem in > 90 % (lila) oder > 95 % (blau) des gesamten Genoms eine ausreichende Coverage (mind. > 30 Reads pro Nukleotidposition) vorlag. In Teil **A.** sind die Ergebnisse von Proben mit einer Viruslast zwischen 100.000 IU/ml und 1.000.000 IU/ml dargestellt und in Teil **B.** die von Proben mit einer Viruslast > 1.000.000 IU/ml. Des Weiteren sind die Fallzahlen n der erfolgten Experimente angegeben.

Werden die Ergebnisse aller fünf Genotypen gemeinsam betrachtet, wurde im Durchschnitt bei einer Viruslast < 100.000 IU/ml ein sequenzierter Genomanteil von 80 % (± 27 %) erreicht. Von den insgesamt 32 Proben erreichten 63 % einen sequenzierten Genomanteil von > 90 % und 47 % einen sequenzierten Genomanteil von > 95 %. Für Proben mit einer Viruslast zwischen 100.000 IU/ml und 1.000.000 IU/ml wurde im Durchschnitt ein sequenzierter Genomanteil von 90 % (± 13 %) ermittelt. Von den 70 Proben erreichten 70 % einen sequenzierten Genomanteil von > 90 % und 47 % einen sequenzierten Genomanteil von > 95 %. Proben mit einer Viruslast > 1.000.000 IU/ml zeigten im Durchschnitt einen sequenzierten Genomanteil von 95 % (± 8 %). Von den 111 Proben konnten 89 % einen sequenzierten Genomanteil von > 90 % und 60 % einen sequenzierten Genomanteil von > 95 % erreichen.

4.5 Beschreibung der Düsseldorfer Kohorte

4.5.1 Phylogenetische Datenanalyse

Zur Darstellung des Infektionsnetzwerkes und möglicher Cluster-Bildung innerhalb der Düsseldorfer IDU-Kohorte wurde aus den Daten der Sequenzierung mit Hilfe von Geneious Prime® und dem Design-Tool iTOL3 (Vgl. 3.3.8) ein phylogenetischer Baum gebildet, der in Abbildung 22 dargestellt ist.

In der Düsseldorfer Kohorte wurde am häufigsten der Genotyp 1 (zu 52 %), gefolgt von dem Genotyp 3 (zu 40 %) beobachtet. Innerhalb des Genotyps 1 trat zu 85 % der Subtyp 1a und zu 15 % der Subtyp 1b auf und der Genotyp 3 wurde ausschließlich durch den Subtyp 3a gebildet. Selten wurde der Genotyp 4 (zu 8 %) beobachtet, innerhalb dessen ausschließlich die Subtypen 4a (zu 28 %) und 4d (zu 72 %) auftraten.

In allen beschriebenen Genotypen treten Isolate auf, die sehr geringe genetische Abstände haben und beispielhaft in Abbildung 22 mit den orangefarbigen Kreisen markiert sind. Innerhalb des Genotyps 3a sind z.B. zwei Proben von unterschiedlichen Individuen mit einem genetischen Abstand von 0 erkennbar, weshalb hier von einer kürzlichen Übertragung ausgegangen werden kann. Auch innerhalb der Genotypen 1a und 1b sind mehrmals Cluster von mehreren Proben mit genetisch sehr geringem Unterschied zu sehen. Diese Beispiele zeigen eindrücklich, dass es in der Kohorte weiterhin zu Übertragungen kommt und die Ganzgenomsequenzierung diese Übertragungen sichtbar machen kann.

Durch Kenntnis solch enger genetischer Verwandtschaften wäre es möglich, die zugehörigen Patienten auf mögliche gemeinsame Ansteckungswege oder zu risikoreichen Verhaltensweisen zu befragen. Dies könnte hilfreich sein, um Zusammenhänge zu erkennen und das Infektionsgeschehen von Hepatitis C besser zu verstehen.

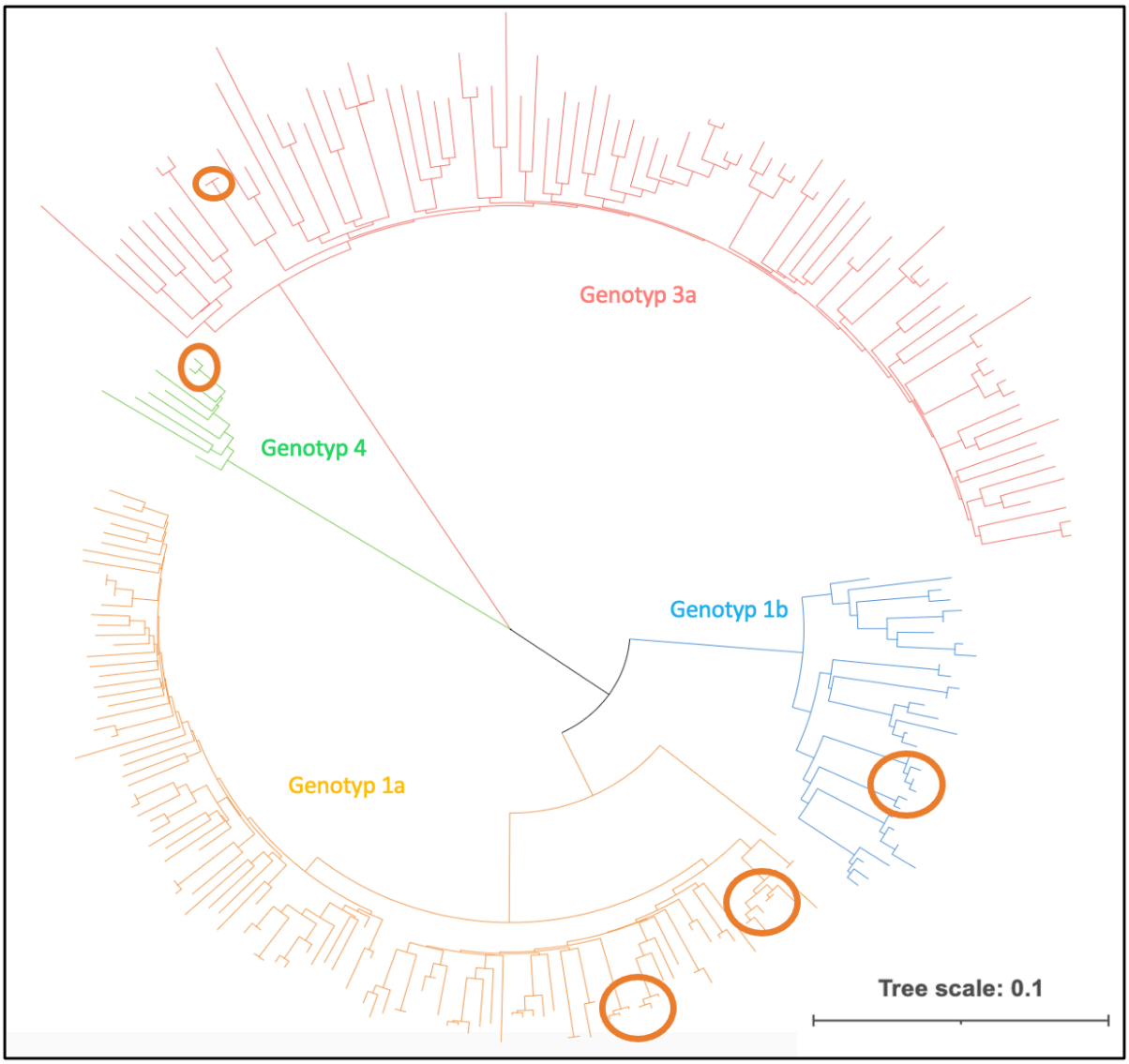

Abbildung 22: Phylogenetischer Baum der mit Hepatitis C Virus infizierten Kohorte Düsseldorf
Dargestellt sind die phylogenetischen Zusammenhänge des sequenzierten Hepatitis C Virus aus 216 Proben der Düsseldorfer injection drug use (IDU) – Kohorte Düsseldorf. Der Baum wurde durch die Nachbar-Verbindungsmethode mit dem Tamura-Nei-genetischen Distanz-Modell hergestellt, wozu die Software Geneious Prime 11.0.12 verwendet wurde. Zur Visualisierung wurde das Online-Tool iTOL3 verwendet. Erkennbar sind die vier unterschiedlichen Subgenotypen 1a, 1b, 3a und 4. Mit orangefarbigen Kreisen sind beispielhaft für jeden Genotyp Zusammenhänge mit sehr geringem genetischem Abstand markiert.

4.5.2 Resistenztestung

Die Proben der IDU-Kohorte eignen sich sehr gut für die Bestimmung von Baseline-Resistenzen, da alle Patienten Therapie-naiv sind. Aus diesem Grund wurden alle 206 Sequenzen mit Hilfe eines online-basierten Interpretationssystems Geno2Pheno$_{[HCV]}$ auf das Vorliegen von Resistenzen gegen die fünf zugelassenen Substanzen Sofosbuvir, Velpatasvir, Pibrentasvir, Voxilaprevir und Glecaprevir untersucht (Vgl. 3.3.9). In folgender Tabelle 16 ist die Anzahl der aufgetretenen Resistenzen nach Genotyp sowie im Gesamten für jedes der fünf Substanzen gegenübergestellt. Es fällt auf, dass bei Proben des Genotyps 1a die meisten Resistenzen beobachtet wurden, darunter gegen die Substanzen Velpatasvir, Voxilaprevir und Glecaprevir. Für den Genotyp 3a wurden ebenfalls Resistenzen gegen Velpatasvir beobachtet. Keine Resistenzen wurden für Proben der Genotypen 1b und 4 als auch gegen die Substanzen Sofosbuvir und Pibrentasvir gefunden. Baseline-Resistenzen wurden im Rahmen dieser Testung nicht gefunden und scheinen daher keine nennenswerte Rolle zu spielen.

Tabelle 16: Resistenztestung der Düsseldorfer IDU-Kohorte
Beschrieben sind die Anzahl der beobachteten Resistenzen gegen fünf zugelassene Substanzen bei einer Resistenztestung von 206 HCV-Proben der Düsseldorfer IDU-Kohorte. Die Sequenzen wurden mit Hilfe einer Tiling-PCR sowie einer Nanopore-Sequenzierung gebildet und die Resistenztestung anschließend mittels Geno2Pheno erstellt worden. Die Anzahl der Resistenzen wurde für die einzelnen untersuchten Genotypen (GT) als auch im Gesamten beschrieben.

Substanz (Zielprotein)	**Sofosbuvir (NS5B)**	**Velpatasvir (NS5A)**	**Pibrentasvir (NS5A)**	**Voxilaprevir (NS3)**	**Glecaprevir (NS3)**
GT 1a	0	4	0	8	1
GT 1b	0	0	0	0	0
GT 3a	0	4	0	0	0
GT 4	0	0	0	0	0
Gesamt	0	8	0	8	1

4.6 Validierung der Methode im Vergleich zu anderen Verfahren

4.6.1 Vergleich zwischen der Tiling-PCR mit Nanopore-Sequenzierung und der konventionellen Sanger-Sequenzierung

Zur Verifizierung der entwickelten Methode sollten die gebildeten Sequenzen mit einer anderen publizierten PCR- und Sequenziermethode verglichen werden. Hierzu wurde die Sanger-Sequenzierung ausgewählt, da diese bereits seit Jahrzehnten etabliert ist. [89]
Im Rahmen einer Dissertation von Marie-Annett Bernard wurden alle Proben des Genotyps 3a der Düsseldorfer IDU-Kohorte im ca. 750 Nukleotid-langen NS5A-Abschnitt amplifiziert und nach der Methode von Sanger sequenziert. Diese Daten konnten zum Vergleich des NS5A-Bereichs beider Methoden vom Institut zur Verfügung gestellt werden. Nach der Erstellung eines phylogenetischen Baums konnten gleiche Proben einander zugeordnet und die Unterschiede analysiert werden.
Stichprobenartig wurden zehn Proben mit unterschiedlichen Unterscheidungsgrad (ein bis zehn Nukleotide Differenz) zwischen den Methoden herausgesucht und die Sequenz auf Nukleotid-Ebene verglichen. Bei den meisten unterschiedlichen Nukleotiden handelte es sich um sogenannte Ambiguitäten. So wurde in 25 der 35 unterschiedlichen Positionen mit der Methode der Nanopore eine Mischung aus zwei Basen erkannt, während mit der Sanger-Methode nur eine Base erkannt wurde. Jedoch entsprach die Base der Sanger-Sequenzierung dem Basengemisch der Nanopore. An neun der 35 Positionen wurde mit der Nanopore ein N und in der Sanger-Sequenzierung ein G gemessen. Bei Analyse der Rohdaten der Nanopore zeigte sich jedoch, dass zu 60 % ein G vorlag. Diese Positionen wurden scheinbar durch den Algorithmus nicht richtig erkannt. In vier der 35 unterschiedlichen Nukleotide kam es in der Sanger-Sequenzierung vor, dass zwei unterschiedliche Basen an einer Position der Sequenzabfolge auftraten, während in der Nanopore ein einzelnes Nukleotid erkannt wurde. Ein Basenunterschied mit eindeutiger Zuordnung der sequenzierten Base ohne Ambiguität konnte an sechs Positionen beobachtet werden.

4.6.2 Vergleich zwischen der Tiling-PCR mit Illumina- und mit Nanopore-Sequenzierung

Zur weiteren Überprüfung der Tiling-PCR wurden zehn Proben des Genotyps 3a ausgewählt, die eine überwiegend hohe Readmenge, aber einen geringen sequenzierten Genomanteil aufwiesen. Nach einem doppelten Ansatz wurde das PCR-Produkt zum einen erneut mit der Nanopore sequenziert und zum anderen zur Illumina-Sequenzierung vorbereitet und zum Vertragspartner verschickt. (Vgl. 3.3.7) Mit diesem Vergleich sollte überprüft werden, ob eine Illumina-Sequenzierung für Proben mit verminderten sequenzierten Genomanteil eine effiziente, alternative Option darstellt.

Ein sequenzierter Genomanteil von > 90 % wurde von sieben der zehn Proben bei der erneuten Sequenzierung mittels Nanopore und bei allen zehn Proben bei der Illumina-Sequenzierung erreicht. Ein sequenzierter Genomanteil von > 95 % konnte bei vier Proben mittels Nanopore- und bei acht Proben mittels Illumina-Sequenzierung beobachtet werden. Im Durchschnitt wurde bei der Sequenzierung mittels Nanopore ein sequenzierter Genomanteil von 90 % ($\pm$ 11 %) erreicht und bei der Sequenzierung mittels Illumina ein sequenzierter Genomanteil von 98 % ($\pm$ 3 %).

Um den Einfluss der Viruslast und der Readmenge zu untersuchen, wurden die Ergebnisse nach diesen Variablen unterteilt statistisch ausgewertet. Mit einer Viruslast < 100.000 IU/ml zeigten eine von fünf Proben (20 %) mittels Nanopore und vier von fünf Proben (80 %) mittels Illumina einen sequenzierten Genomanteil von > 95 %, und drei von fünf Proben (60 %) mittels Nanopore sowie fünf von fünf Proben (100 %) mittels Illumina einen sequenzierten Genomanteil von > 90 %. Bei einer Viruslast zwischen 100.000 und 1.000.000 IU/ml wurde ein sequenzierter Genomanteil von > 95 % mittels Nanopore sowie mittels Illumina bei jeweils zwei von drei Proben (66.7 %) beobachtet, ein sequenzierter Genomanteil von > 90 % wurde durch beide Methoden bei drei von drei Proben (100 %) erreicht. Mit einer Viruslast > 1.000.000 IU/ml konnte eine Probe durch beide Methoden einen sequenzierten Genomanteil von > 95 % erreichen.

Eine Spearman-Rangkorrelation zwischen dem sequenzierten Genomanteil in % und der Viruslast in IU/ml zeigte unterschiedliche Ergebnisse bei ähnlicher Tendenz. Bei der Nanopore-Sequenzierung wurde eine positive, statistisch signifikante Korrelation ($r = 0{,}783$, $p = 0{,}013$ (2-seitig)) beobachtet, wogegen bei der Illumina-Sequenzierung keine signifikante Korrelation ($r = 0{,}25$, $p = 949$ (2-seitig)) gemessen werden konnte. Graphisch deutet sich aber ebenfalls ein leicht positiver Zusammenhang an, der in Teil A. der Abbildung 23 dargestellt ist.

Eine Readmenge < 100.000 lag bei einzig einer der Proben vor. Diese zeigte mittels Nanopore einen sequenzierten Genomanteil zwischen 90-95 %, wogegen mittels Illumina durchgeführt ein sequenzierter Genomanteil von > 95 % beobachtet werden konnte. Eine Korrelation zwischen dem sequenzierten Genomanteil und der Anzahl der Reads der Sequenzierung ist sowohl für die Sequenzierung mittels Nanopore als auch mittels Illumina nicht signifikant (Nanopore: r = 0,207, p = 0,410 (2-seitig); Illumina: r = 0,266, p = 0,285 (2-seitig)). Graphisch dargestellt zeigt sich ein ähnlicher, positiver Zusammenhang, was in folgender Abbildung 23 unter Teil B. erkennbar ist. Außerdem stellt sich dar, dass mit der Illumina-Sequenzierung deutlich mehr Reads pro Sequenzierung gebildet wurden.

Eine Korrelation zwischen der Readmenge zur Viruslast weist für beide Sequenziermethoden ebenfalls einen nicht-signifikanten Zusammenhang (Nanopore: r = 0,317, p = 0,406 (2-seitig); Illumina: r = 0,250, p = 0,516 (2-seitig)) auf und ist graphisch in Teil C. in folgender Abbildung 23 beschrieben.

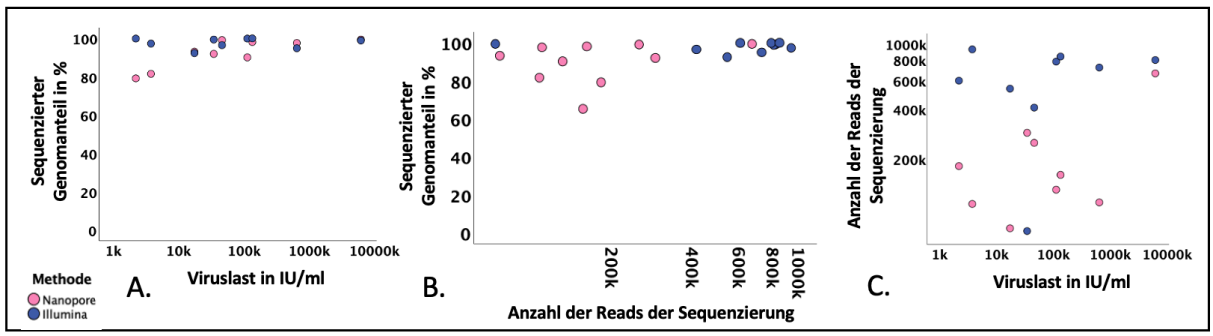

Abbildung 23: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Readmenge und der Viruslast in IU/ml des Genotyps 3a mit Nanopore und Illumina-Sequenzierung
Dargestellt ist der Vergleich einer Sequenzierung von Proben der injection-drug-use (IDU) – Kohorte Düsseldorf des Genotyps 3a, durchgeführt mittels Nanopore- (rosa) oder Illumina- (blau) Sequenzierung. Es sind Zusammenhänge zwischen dem sequenzierten Genomanteil in % und der Viruslast in IU/ml (**A.**), dem sequenzierten Genomanteil in % und der Readmenge (**B.**) sowie der Readmenge und der Viruslast in IU/ml (**C.**) abgebildet.

Bei Betrachtung der Abdeckung von für die Diagnostik relevanten Proteinregionen Core, E1, NS3, NS5A und NS5B des HCV-Genoms wurden die Mittelwerte der Coverage, daher die Anzahl der Reads pro Nukleotiposition, dieser Regionen gebildet. Zuvor waren Nukleotidabschnitte der Proteinregionen mit Hilfe der Referenzsequenz X76918 definiert worden. In folgender Tabelle 17 ist erkennbar, dass sowohl mit der Illumina- als auch der Nanopore-Methode eine als Minimum für erfolgreiche Sequenzierung definierte Coverage von mind. 30 in allen fünf Regionen erreicht wurde. Im Vergleich konnte die Sequenzierung mittels Illumina jeweils eine höhere Coverage erreichen, besonders groß ist der Unterschied

in den Regionen E1 und NS3. Dort konnte mittels Illumina-Sequenzierung eine Coverage > 10.000 erreicht werden, welche ein relevanter Cut-Off für die Resistenztestung darstellt.

Tabelle 17: Vergleich der Mittelwerte der Coverage zwischen Illumina- und Nanopore-Sequenzierung
Dargestellt sind die Mittelwerte der Coverage von Proben der injection drug-use (IDU) – Kohorte Düsseldorf des Genotyps 3a der Tiling-Polymerasekettenreaktion, sequenziert mittels Illumina- oder Nanopore-Methode. Die Nukleotidpositionen der relevanten Proteine von Hepatitis C wurden mit Hilfe der Referenz X76918 bestimmt. Es zeigt sich eine deutliche erhöhte Coverage mittels Illumina-Sequenzierung in allen untersuchten Abschnitten.

Protein	Nukleotidposition	Mittelwert der Coverage, Methode: Illumina	Mittelwert der Coverage; Methode: Nanopore
Core	299-871	8315	8201
E1	872-1447	13279	7000
NS3	3395-5287	14757	8891
NS5A	6233-7588	8247	6425
NS5B	7589-9361	12218	8031

Mit Hilfe von Geno2Pheno (Vgl.3.3.9) wurde anhand der gebildeten Sequenzen eine Resistenztestung durchgeführt und die Anzahl der Resistenzen für die fünf zugelassenen Substanzen Sofosbuvir, Velparasvir, Pibrentasvir, Voxilaprevir und Glecaprevir in folgender Tabelle 18 dargestellt. Dabei wurde kein Unterschied in der Anzahl der Resistenzen zwischen der Sequenzierung mittels Nanopore oder Illumina gefunden. Mit beiden Methoden wurde eine Resistenz für die Substanz Velpatasvir bestimmt.

Tabelle 18: Vergleich Resistenztestung Nanopore vs. Illumina
Dargestellt sind die Anzahlen der beobachteten Resistenzen von fünf zugelassenen Substanzen bei einer Resistenztestung von jeweils 10 Proben des Genotyps 3a. Die Sequenz ist mit Hilfe einer Tiling-PCR amplifiziert und danach entweder mit der Nanopore- oder Illumina-Technik sequenziert worden. Anschließend erfolgte eine Resistenztestung mittels Geno2Pheno.

Substanz (Zielprotein)	Sofosbuvir (NS5B)	Velpatasvir (NS5A)	Pibrentasvir (NS5A)	Voxilaprevir (NS3)	Glecaprevir (NS3)
Nanopore	0	1	0	0	0
Illumina	0	1	0	0	0

4.6.3 Vergleich der Nanopore- und Illumina- Sequenzierung zur Sanger-Sequenzierung

Um eine weitere Vergleichsmöglichkeit zwischen der Nanopore- und Illumina-Sequenzierung zu erhalten, wurden die Ergebnisse der Illumina-Sequenzierung ebenfalls mit den bekannten NS5A-Sequenzen der Sanger-Sequenzierung in einem phylogenetischen Baum zusammengeführt und dessen Unterschiede von gleichen Proben analysiert. Die Sequenz der Probe mit IDU-Nr. 602 und IDU-Nr. 983 mussten für diesen Vergleich ausgeschlossen werden, da keine NS5A-Sequenz der Sanger-Sequenzierung vorlag.

Es zeigten sich sowohl im Vergleich mit der Nanopore- als auch der Illumina- Sequenzierung häufig Nukleotidunterschiede zur Sanger- Sequenzierung, deren Anzahl pro Probe in Tabelle 19 aufgeführt sind. Die Nukleotidunterschiede könnten u.a. mit der geringeren Viruslast und Readmenge zu erklären sein, weshalb diese zehn Proben des GT 3a ursprünglich ausgewählt wurden. Überwiegend ähneln sich die Anzahl der Unterschiede der Nanopore- und Illumina-Methode bis auf einige Ausnahmen. Die 271 Unterschiede der Probe mit IDU-Nr. 444, die 156 Unterschiede der Probe mit IDU-Nr. 156 sowie die 258 Unterschiede der Probe mit IDU-Nr. 983 durch die Nanopore sind fast ausschließlich durch Ns zu Stande gekommen, was auf keine falschen Nukleotide, sondern eine Unterbrechung der Sequenz an dieser Stelle hindeutet. Auch 99 der 199 Unterschiede der Probe mit IDU-Nr. 890 haben die gleiche Ursache. Weitere Unterschiede der Nanopore- zur Sanger-Sequenzierung zeigen, wie bereits beschrieben, häufig einen Zusammenhang zu Ambiguitäten. Im Vergleich wird sichtbar, dass die Illumina-Sequenzierung mehrfach weniger Unterschiede zur Sanger-Sequenzierung zeigt. Nur bei der Probe mit IDU-Nr. 814 konnte, wie bei der Nanopore-Sequenzierung, ein langer Abschnitt aus Ns als Ursache des Unterschieds erkannt werden. Auch bei der Illumina-Sequenzierung sind Unterschiede durch Ambiguitäten häufig, wenn auch mit 15 von 24 Unterschieden etwas seltener als bei der Nanopore-Sequenzierung.

Tabelle 19: Vergleich Nanopore-/Illumina- zur Sanger-Sequenzierung
Dargestellt sind die Anzahl der Nukleotidunterschiede der Sequenzen von jeweilig zusammengehörigen Proben der injection-drug-use (IDU)-Kohorte Düsseldorf. Die NS5A-Region des Hepatitis C Virus wurde mittels Sanger- und mittels Nanopore-/Illumina-Sequenzierung untersucht und die gebildeten Sequenzen miteinander auf ihre Unterschiede hin analysiert.

IDU-Nr	**440**	**798**	**444**	**495**	**1075**	**1069**	**890**	**814**
Nanopore	0	2	271	3	3	8	119	156
Illumina	0	0	0	2	6	8	5	131

5 Diskussion

Ziel der vorliegenden Arbeit stellte die Entwicklung einer Ganzgenom-Sequenzierung des Hepatitis C Virus dar. Trotz hoher virologischer Heilungsraten wurden im Jahr 2021 in Deutschland knapp 5000 Neuinfektionen mit Hepatitis C diagnostiziert. [90] Eine genomische Surveillance des Virus wäre durch Abbildung des gesamten Genoms möglich und könnte helfen, um diesen Widerspruch besser zu verstehen. Des Weiteren könnte die Abbildung des gesamten Genoms eine Kontaktnachverfolgung ermöglichen, die Aufschluss über Infektionsketten und Identifizierung von Ausbrüchen der Erkrankung bietet.

In der vorliegenden Dissertation wurde eine Genotyp-spezifische Tiling-PCR entwickelt, deren methodischer Ansatz darauf beruht, das gesamte Genom in vielen kleinen Abschnitten zu amplifizieren, zu sequenzieren und anschließend bioinformatisch wieder zusammen zu setzen. Hierzu wurden Primer-Sets mit Primern von 400-600 bp Länge entwickelt, die sich gegenseitig überlappen und deshalb in zwei bzw. drei Pools aufgeteilt wurden. Während der Etablierung wurden diese Primersets mit Hilfe von europäischen Referenzisolaten des jeweiligen Genotyps überarbeitet und angepasst. Anschließend wurde die Methode mit höherer Fallzahl mittels Proben der Düsseldorfer IDU-Kohorte überprüft und die Ergebnisse statistisch ausgewertet.

Zur Validierung wurden die durch verschiedene Methoden gebildeten Sequenzen des NS5A-Abschnitts der Proben der IDU-Kohorte miteinander verglichen. Die Methode der Sanger-Sequenzierung [89] wurde hierfür als Kontrollgruppe verwendet und mit der Tiling-PCR mit Nanopore- oder mit Illumina-Sequenzierung verglichen.

5.1 Etablierung der Tiling-PCR

Zur Etablierung der Genotyp-spezifischen Ganzgenom-PCR wurden unterschiedliche Primersets neu designt (Genotyp 3a, 4a und 4d) oder es konnten auf im Institut bereits vorhandene Primersets zurückgegriffen werden (Genotyp 1a und 1b). Eine Überarbeitung der Primer erfolgte durch den Vergleich der Sequenz von europäischen HCV-Referenzisolaten. Insgesamt wurden fünf unterschiedliche Primersets für die Genotypen 1a, 1b, 3a, 4a, und 4d etabliert.

5.1.1 Reverse Transkription

In Vorarbeiten am Institut der Virologie Düsseldorf konnte bereits gezeigt werden, dass das komplette HCV-RNA-Genom mittels eines Primers am 3'Ende in einer cDNA umgeschrieben werden kann. [76] Dabei wird ein Oligo-d(A) -Primer verwendet, der an den Poly (T) -Bereich zwischen NS5B und der 3'-NRT bindet. Zur besseren cDNA-Synthese des 5'-UTRs wird zusätzlich noch der Primer p874a, der in der Core-Region bindet, eingesetzt.

In Experimenten dieser Dissertation zeigte sich jedoch, dass viele Sequenzen im 5'-UTRs eine niedrige Genomabdeckung bzw. große Lücken aufwiesen. Der Primer p874a, der an die Genomposition 874 bindet, ermöglicht zwar eine cDNA-Synthese der ersten 800 Basenpaare, jedoch fehlte bei einigen Sequenzen der Bereich zwischen nt 800 und 3600. Dieses Problem konnte durch den Einsatz von Random Hexameren gelöst werden. Random Hexamere sind dafür bekannt, nur kleine, zufällige Bereiche in cDNA umzuschreiben. [38] Dies kann bei größeren PCR-Fragmenten über 1000 bp ein Problem darstellen. Bei kleineren PCR-Fragmenten, wie den in dieser Dissertation gebildeten 400 bp langen Fragmenten, scheint das kein Nachteil zu sein. Durch den Einsatz von Random Hexameren und Oligo-d(A)-Primer im RT-Protokoll konnte bei vielen Proben eine deutlich gleichmäßigere Amplifizierung und Sequenzierung des 5'-Bereiches (Vgl. 4.2) erreicht werden, weshalb das Protokoll für alle untersuchten Genotypen übernommen wurde.

Nachteilig ist bei der Verwendung von Random Hexameren die mögliche Bildung von unspezifischen PCR-Fragmenten, da diese auch Nicht-HCV-RNA binden können, wodurch eine erhöhte Gefahr der Bildung von unspezifischen, z.B. humanen Reads besteht. Außerdem entstehen überwiegend kurze cDNA-Fragmente, was aber für die Tiling-PCR keinen Nachteil darstellt, da hier nur kurze Fragmente nötig sind. Eine Alternative zu Random Hexameren stellt die Verwendung von Random Pentadecameren dar. Damit kann eine höhere cDNA-Ausbeute und -Qualität erzeugt werden, was für Proben mit einer niedrigen Viruslast von Vorteil sein könnte. [91]

5.1.2 Workflow und Einflussgrößen für den Sequenziererfolg

Nach der erfolgreichen Reversen Transkription wurde die Tiling-PCR in zwei bzw. drei Pools durchgeführt. Das gebildete PCR-Produkt wurde in den ersten Versuchen mittels Gel-Elektrophorese und anschließend mittels DNA-Quantifizierung durch ein Luminometer nachgewiesen. Die Sequenzierung erfolgte mittels Oxford Nanopore Sequenzierung. Nach der Sequenzierung des PCR-Produkts mittels Nanopore erfolgte die bioinformatische

Datenverarbeitung der FastQ-Daten in Geneious Prime®. Die Bildung eines Assembly erzeugt die richtige Reihenfolge der Sequenzierstücke durch Orientierung an einem Referenzgenom und durch die Bildung einer Consensus Sequenz wurde die mehrheitliche Nukleotidabfolge angezeigt (Vgl. 3.3.8). Die anschließend bestimmte Coverage, daher die Anzahl der Reads pro Nukleotidposition, dient als Qualitätskontrolle der Sequenzierung. Es wurde ein Cut-Off von 30 als Minimum für eine erfolgreiche Sequenzierung definiert (Vgl. 4.2). Erstrebenswert wäre eine Coverage über 10.000, da sich die gebildete Sequenz dann für eine Resistenztestung eignet. Des Weiteren wurde als Erfolgsmaßstab der sequenzierte Genomanteil durch die Anzahl vorkommender Ns in der Consensus Sequenz definiert. Ein N wurde gebildet, sobald die Coverage an dieser Position < 30 lag und daher nicht ausreichend war.

Eine reduzierte Genomabdeckung ist meistens auf ein fehlendes PCR-Fragment zurückzuführen, weil z.B. die zugehörigen Primer nicht gebunden haben. Eine fehlende Bindung kann z.B. durch Fehlpaarungen oder Sekundärstrukturen entstehen, wenn z.B. die Nukleotidabfolge nicht zur cDNA passt oder sich ein zu hoher G/C-Gehalt am 3‘-Ende oder ein Poly-A-Abschnitt im Primer befindet. [92] Weitere Ursachen für eine schlecht funktionierende PCR können vielfältig sein. Beispiele sind eine unvollständige cDNA durch die Reverse Transkription, ein frühzeitiger Abbruch oder fehlerhaftes Arbeiten der Polymerase, wobei letzteres z.B. durch Korrekturlesefunktionen der Polymerase verhindert werden soll. [93]

Die Bindung des Primers an die cDNA ist ebenfalls durch unterschiedliche Einflüsse bedingt. Die ersten drei Nukleotide des Primers sollten exakt zur Sequenz der cDNA passen und auch alle weiteren Nukleotide sollten überwiegende Ähnlichkeit aufweisen. Um dies zu gewährleisten, wurden die zugehörigen Primer aller Regionen mit reduzierter Coverage mit Referenzisolaten des jeweiligen Genotyps verglichen und korrigiert (Vgl. 4.3.1, 4.3.2, 4.3.3). Im Falle von erheblichen Unterschieden zwischen den Referenzisolaten wurden bis zu zwei Mehrdeutigkeitssymbole (IUPAC-Codes) in einen Primer eingebracht, um möglichst viele Quasispezies durch die PCR abdecken zu können.

Des Weiteren ist die Primer-Schmelztemperatur T_m relevant, welche angibt, zu welcher Temperatur 50 % des Primers an die cDNA gebunden ist. Sie ist abhängig von der Nukleotidlänge des Primers sowie dem Gehalt der Nukleotide C und G. Die Schmelztemperatur sollte maximal 3-5°C unterhalb der Annealing-Temperatur des PCR-Zyklus liegen, um eine ausreichende Bindung des Primers an die cDNA zu erreichen. Bei zu niedriger Schmelztemperatur würde keine effiziente Bindung des Primers an die cDNA erfolgen, bei zu hoher Temperatur könnte die Bindung zu unspezifisch erfolgen, weshalb Artefakte oder eine

nicht-funktionierende PCR die Folge wären. Nach dem verwendeten Protokoll von J. Quick [61] entspricht die Annealing-Temperatur 63°C. Aus diesem Grund wurden die Primer der Tiling-PCR auf eine Schmelztemperatur zwischen 60-68°C überprüft und entsprechend in ihrer Nukleotidlänge angepasst.

Im verwendeten Protokoll stellt der Annealing-Schritt eine Kombination mit dem Elongationsschritt dar. Dies ist sinnvoll, um einem Ungleichgewicht der Produkte unterschiedlicher Primer vorzubeugen, da durch die Potenzierung eines Ungleichgewichts mit jedem PCR-Durchlauf auch Bindungsunterschiede der Primer von wenigen Prozent relevant wären. Mit Verlängerung der Annealing-Phase bzw. Kombination mit der Elongationsphase können auch schlechter bindende Primer an ihr Target binden, sodass alle PCR-Produkte in gleichmäßiger Menge gebildet werden.

Des Weiteren kann auch ein bioinformatisches Problem die Ursache für eine scheinbar reduzierte Genomabdeckung sein. Müssen z.B. unterschiedliche PCR-Produkte aufgrund verschiedener Quasispezies, wie sie häufig in der HVR vorkommen, gebildet werden, passen diese PCR-Produkte bei der Bildung des Assembly nicht zum Referenzgenom und der Algorithmus kann sie nicht richtig zuordnen.

5.1.3 Korrektur der Primer-Sets

5.1.3.1 Genotyp 1a und 3a

An Positionen im Genom, wo eine reduzierte Coverage beobachtet wurden, wurden die zugehörigen Primer mit Referenzkollektiven auf Nukleotidebene verglichen und anschließend optimiert. Für den Genotyp 3a wurde so nach Korrektur der drei Primer HCV-3a-6R, -7L und -28L eine deutliche Steigerung der Coverage erreicht (Vgl. 4.3.3). Für den Genotyp 1a konnte eine Optimierung z.B. durch Verdopplung der Konzentration der Primer 5b und 5c beobachtet werden (Vgl. 4.3.1).

Nach erfolgreichen ersten Sequenzierungen stellte sich heraus, dass bei einigen GT 1a Isolaten die ersten Nukleotide (bis zum 300. Nukleotid) nicht ausreichend sequenziert wurden und sich dies durch die veränderte Reverse Transkription reduzierten, aber nicht beheben ließ. Eine mögliche Ursache, wieso das Genom in der 5‘-UTR schwerer zu amplifizieren ist, könnte im Aufbau des Virusgenoms liegen. Innerhalb der 5‘-UTR befindet sich eine komplexe RNA-Struktur, die eine interne Ribosomen-Eingangsstelle darstellt und essentiell zur Vermehrung des RNA-Genoms ist. [9] Eine solche RNA-Struktur kann z.B. eine starke Sekundärstruktur ausbilden, die schwerer zu amplifizieren ist.

Da für den Genotyp 3a die Anfangssequenz gut abgebildet werden konnte und die Sequenz der 5‘-UTR zwischen GT 1a und GT 3a an der Primerbindungsstelle identisch ist, wurde für drei GT 1a -Proben untersucht, ob durch die zusätzliche Verwendung des Primers HCV-3a_1L des Genotyps 3a eine bessere Genomabdeckung möglich ist. (Vgl. 4.4.1) Bei allen drei Proben wurde eine Verbesserung beobachtet, aber dennoch begann die Sequenz z.B. erst ab dem 95. Nukleotid. Dies konnte ebenfalls bei dem Genotyp 1b beobachtet werden. Wie bei der HVR ist es fraglich, ob die Sequenz nicht gebildet wird oder ob es sich um ein bioinformatisches Artefakt, z.B. aufgrund eines minderwertigen Assemblys, handelt.

5.1.3.2 Genotyp 1b

Die Korrektur des Primer-Sets des Genotyps 1b zeigte uneindeutige Ergebnisse. Im Versuch mit den optimierten Primern wurden mehrmals reduzierte Coverages an Positionen beobachtet, wo zuvor eine ausreichende Coverage erreicht wurde (Vgl. Abbildung 9), obwohl deren korrigierte Primer exakter zu den Referenzisolaten passten. Zur weiteren Verifizierung, welches Primerset sich als geeigneter darstellt, wurden PCRs mit beiden Primersets in höherer Fallzahl durchgeführt (Vgl. 4.3.2) Dabei konnte eine geringfügig höhere Coverage im Versuch mit den alten Primern beobachtet werden, die sich ebenfalls in den Mittelwerten der Coverage wichtiger Nukleotidregionen zeigte (Vgl. Tabelle 6).

Die veränderten Primer enthielten häufiger Ambiguitäten, was möglicherweise zu mehr Instabilität in der Primerbindung geführt haben könnte. Dennoch wurden, wie auch für die anderen Genotypen beschrieben, nicht mehr als zwei Ambiguitäten pro Primer verwendet. Des Weiteren wurden einige Primer verlängert, um eine Annealing-Temperatur > 65°C zu erreichen. Möglicherweise könnte es dadurch zu vermehrter Bildung von Sekundärstrukturen, wie z.B. einer Primer-Dimer-Bildung gekommen sein, die zu einer schlechteren Bindung an die cDNA führen könnte. Ein anderer Grund für diese Beobachtungen könnte z.B. auch eine schlechtere Synthese der Primer durch den Hersteller sein, was durch eine Neubestellung und Wiederholung der Versuche verifizierbar wäre.

Anders als bei Genotyp 1a waren für den Genotyp 1b neue Primer-Pools gebildet worden und nicht die veränderten Primer zusätzlich zum Pool hinzugefügt worden. Für den Genotyp 1a sind dadurch für die Regionen der veränderten Primer höhere Konzentrationen durch die doppelten, aber unterschiedlichen Primer vorhanden, was eine Verbesserung der Coverage erzeugen könnte. Entgegen dieser Theorie spricht, dass für den Genotyp 3a wie für den Genotyp 1b neue Pools gebildet wurden und dabei die Coverage erhöht werden konnte.

5.1.3.3 Genotyp 4

Die Tiling-PCR des Genotyps 4 konnte bei neun von zehn Proben ein fast vollständiges Genom abdecken und funktionierte daher sehr erfolgreich. Lediglich bei einer Probe #7442 zeigten sich größere Lücken in der Sequenz (Vgl. 4.3.4), die sich auch in einer Wiederholung der PCR und Sequenzierung erneut darstellten. Die Primer der entsprechenden Regionen passten zu den Referenzisolaten und wiesen keinen Korrekturbedarf auf. Mögliche Ursachen für die unvollständige Abdeckung könnte die niedrige Viruslast von 116.642 IU/ml dieser Probe darstellen oder eine schlechte Probenqualität, z.B. durch schlechte Lagerungsqualität oder zu häufiges Auf- und Abtauen.

5.1.3.4 Zusammenfassung

Im Rahmen der Etablierung wurden Primersets für eine Tiling-PCR zur Diagnostik von HCV der Genotypen 1a, 1b, 3a, 4a und 4d gebildet. Durch Korrektur der Primer auf Nukleotidebene, Veränderung der Länge zur Beeinflussung der Primerschmelztemperatur oder Erhöhung der Konzentration einzelner Primer konnte eine optimale Anpassung erreicht werden, die eine hohe Coverage über das gesamte Genom erzeugt. In Abbildung 6 und Abbildung 12 ist beispielhaft eine solche Optimierung der Coverage dargestellt.

5.2 Überprüfung der Methode mit höherer Fallzahl

Nach Überarbeitung der Genotyp-spezifischen Primersets wurde die Tiling-PCR mit einer größeren Fallzahl durchgeführt. Dazu konnten Rückstellproben der Düsseldorf IDU-Kohorte sequenziert werden. Je nach Genotyp ist eine sehr unterschiedlich große Fallzahl vorhanden, was bei der statistischen Auswertung berücksichtigt werden sollte.

5.2.1 Relative Häufigkeiten des Erfolgs

Als Erfolgsmerkmal der Sequenzierung wurde der sequenzierte Genomanteil definiert, der über 90 % bzw. 95 % liegen sollte und dessen relative Häufigkeiten aller Genotypen in Abbildung 20 dargestellt sind. Bei der Gegenüberstellung fällt auf, dass der Genotyp 3a deutlich häufiger einen erfolgreichen sequenzierten Genomanteil aufweist. Ebenfalls zeigen der Genotyp 4a und 4d häufiger einen sequenzierten Genomanteil von > 95 % als der Genotyp 1a und 1b. Eine mögliche Ursache könnte z.B. die Aufteilung der Pools sein, die für die Genotypen 3a, 4a und 4d nur aus zwei Pools bestehen und für die Genotypen 1a und 1b aus drei Pools. Zwei Pools sind im labortechnischen Handling weniger komplex und könnten deshalb weniger fehleranfällig sein. Ein weiterer Vorteil von einer PCR mit zwei Primer-Pools

ist die geringere Nutzung von Probenmaterial pro PCR. Für weiterführende Arbeiten könnte es deshalb sinnvoll sein, neue Primer für den Genotyp 1a und 1b in jeweils zwei Pools zu generieren.

5.2.2 Korrelationen

Um Störfaktoren (engl. *Confounder*) auf den Erfolg der PCR zu identifizieren, wurden im Rahmen der statistischen Auswertung Korrelationen zwischen dem sequenzierten Genomanteil, der Viruslast und der Anzahl der Reads der Sequenzierung (Readmenge) untersucht. Bei den Genotypen mit hoher Fallzahl, daher Genotyp 1a und 3a, wurden dabei jeweils positive, statistisch hoch-signifikante Korrelationen zwischen dem sequenzierten Genomanteil und der Viruslast sowie dem sequenzierten Genomanteil und der Readmenge beobachtet. Für die Genotypen 1b, 4a und 4d zeigte sich graphisch ein positiver Trend (Vgl. Abbildung 17 und Abbildung 19), aber durch die geringe Fallzahl (GT 1b: n = 17, GT 4a: n = 5, GT 4d n = 13) zeigte sich der Zusammenhang statistisch als nicht-signifikant. Es ist daher wahrscheinlich, dass ein positiver Zusammenhang zwischen dem sequenzierten Genomanteil und der Viruslast sowie der Readmenge besteht. Daraus lässt sich schließen, dass die Tiling-PCR besser funktioniert, wenn eine höhere Anzahl an Reads gebildet wurde und wenn eine Probe eine höhere Viruslast aufweist. Letzteren Zusammenhang bildet auch die Abbildung 21 ab. Darin ist der Sequenziererfolg aufgetrennt nach der Viruslast dargestellt und für die Genotypen 1a, 1b, 3a und 4d wird ein häufigeres Erreichen eines sequenzierten Genomanteils von > 90 % oder > 95 % angezeigt, wenn die Viruslast > 1.000.000 IU/ml beträgt.

Der Zusammenhang zwischen der Viruslast und der Anzahl der Reads der Sequenzierung zeigte sich uneindeutiger. Eine statistisch signifikante Korrelation wurde für den Genotyp 1a beobachtet, wobei die Daten geringer korrelierten als zwischen dem sequenzierten Genomanteil und der Viruslast sowie der Readmenge (Vgl. 4.4.1). Für die Genotypen 1b, 3a, 4a und 4d wurde keine signifikante Korrelation zwischen der Viruslast und der Readmenge beobachtet, allerdings zeigte sich auch bei diesen Genotypen graphisch ein positiver Trend (Vgl. Abbildung 17, Abbildung 18, Abbildung 19). Der Zusammenhang, dass eine höhere Viruslast zu einer höheren Anzahl sequenzierter Reads führt, ist daher möglich, kann aber nicht sicher nachgewiesen werden.

Eine Vergrößerung der Fallzahl, insbesondere die der GT 1b, GT 4a und GT 4d, könnte weiter Aufschluss über die Zusammenhänge zwischen dem sequenzierten Genomanteil, der Viruslast und der Anzahl der gebildeten Reads geben.

Ein weiterer Einflussfaktor auf den sequenzierten Genomanteil als auch die Readmenge stellt die Qualität eines einzelnen Sequenzierlaufs dar, die anhand der gebildeten Datenmenge beurteilbar ist. Ein Sequenzierergebnis mit wenig Reads und einem niedrigen sequenzierten Genomanteil trotz viel viralem Ausgangsmaterial kann zwei Gründe haben. Zum einen könnte zu wenig amplifiziert worden sein, daher habe die PCR schlecht funktioniert, oder zum anderen könnte der Sequenzierlauf in der Nanopore nicht ausreichend gewesen sein. Um den Einfluss der Sequenzierung stabil zu halten, wurden, wie in der Coronasequenzierung, Sequenzierläufe mit weniger als zehn Gigabasen wiederholt.

5.2.3 Resistenztestung

Die Resistenz-relevanten Regionen des HCV-Genoms (E1, NS3, NS5A, NS5B und Core) sind für die Diagnostik besonders wichtig und sollten bestenfalls durch die Tiling-PCR einen hohen sequenzierten Genomanteil erreichen. Um dies zu gewährleisten, wurden für den Genotyp 1a ein vierter Primer-Pool aus bereits etablierten Primern der Diagnostik des Instituts für Virologie der Uniklinik Düsseldorf erstellt und eine Tiling-PCR durchgeführt. Ein Vergleich der Mittelwerte der Coverage zwischen einem Versuch mit und ohne dem vierten Pool zeigte eindeutig eine höhere Coverage in jeder der untersuchten Regionen (Vgl. Tabelle 11). In einer Probe wurden ohne den vierten Pool zwei falsch-positive Resistenzen bestimmt, weshalb sich bestätigte, dass für eine sichere Resistenztestung eine Coverage > 10.000 notwendig ist und diese mit dem vierten Pool auch erreicht werden kann. Des Weiteren zeigte sich mit dem vierten Pool häufiger eine Coverage > 90 oder > 95 % sowie eine höhere durchschnittliche Coverage (Vgl. 4.4.1). Eine Erstellung eines vierten Primer-Pools mit Primern der Regionen E1, NS3, NS5A, NS5B und Core für die Genotypen 1b, 3a, 4a und 4d könnte daher sinnvoll als Ergänzung zu den schon bestehenden Primer-Pools sein, da durch die doppelte Abdeckung die für Resistenzen relevanten Regionen des Genoms mit höherer Sicherheit erfolgreich sequenziert werden können.

Von den insgesamt 216 Proben der Düsseldorfer IDU-Kohorte wurden mit der Tiling-PCR Sequenzen ermittelt, die mit Hilfe des Tools Geno2Pheno$_{[HCV]}$ auf das Vorliegen von Resistenzen gegen die fünf Medikamente Sofosbuvir, Velpatasvir, Pibrentasvir, Voxilaprevir und Glecaprevir untersucht wurden (Vgl. Tabelle 16). Es fiel auf, dass die meisten Resistenzen beim Genotyp 1a und gegen die Medikamenten Velpatasvir und Voxilaprevir auftraten. Diese Beobachtung deckt sich mit Daten aus einer Übersichtsarbeit von Z. Lui [94] zur weltweiten Prävalenz von Resistenzen gegen DAAs.

5.3 Vergleich der Düsseldorfer IDU-Kohorte

Zum Abschätzen der Inzidenz sowie dem besseren Verständnis des Infektionsgeschehens sollte nach einem ähnlichen Vorgehen, wie für SARS-CoV-2 durchgeführt [95], eine Untersuchung der genetischen Verwandtschaft zwischen den Patienten der Düsseldorfer IDU-Kohorte durchgeführt werden. Dazu wurden die Sequenzen mit Hilfe von Geneious Prime® und dem Design-Tool iTOL3 als phylogenetischer Baum dargestellt (Vgl. Abbildung 22).
Dabei zeigte sich für jeden Genotyp, dass einige Proben sehr geringe genetische Unterschiede zueinander aufweisen, was auf eine kürzliche Übertragung hindeutet. Durch die Ganzgenom-PCR können daher Übertragungen innerhalb der Kohorte sichtbar gemacht werden, was für zukünftige Kontaktnachverfolgung genutzt werden könnte.
Der phylogenetische Baum zeigte zusätzlich auf, in welchem Verhältnis die unterschiedlichen HCV-Genotypen innerhalb der Düsseldorfer Kohorte auftraten (Vgl. 4.5.1). Zur Einordnung, wie repräsentativ die Düsseldorfer IDU-Kohorte zum HCV-Infektionsgeschehen Deutschlands ist, wurden diese Daten mit denen der Studie zu Drogen und chronischen Infektionskrankheiten (sog. DRUCK-Studie) des RKIs [96] verglichen. Im Rahmen dieser Studie wurden zwischen 2011 und 2015 in acht deutschen Städten Daten erhoben, die das Verhalten sowie den Serostatus von HCV, HBV und HIV von Menschen, die i.v.-Drogen konsumierten, enthielten. Außerdem wurde bei Proben aus sechs dieser Städte eine HCV-Genotypbestimmung durchgeführt, wobei der Genotyp 1 (zu 47 %) und der Genotyp 3 (zu 46 %) am häufigsten beobachtet wurden. Innerhalb der Genotyp 1 -Infektionen lag dabei zu 80 % eine Infektion des Subtyps 1a, zu 16 % des Subtyps 1b und zu 3 % eine Mischinfektion vor. Innerhalb der Genotyp 3 -Infektionen handelte es sich in allen Fällen um den Subtyp 3a. Innerhalb der seltenen Genotypen 2 (zu 3,5 %) und 4 (zu 2,9 %) wurde in den meisten Fällen der Subtyp 2b und 4a bestimmt. Für den Genotyp 1 und 3 entspricht dies einer sehr ähnlichen Verteilung wie in der Düsseldorfer IDU-Kohorte. Lediglich in den seltenen Genotypen 2 und 4 lassen sich Unterschiede in der Verteilung beobachten. In der Düsseldorf IDU-Kohorte trat Genotyp 2 nur sehr selten auf und wurde deshalb nicht amplifiziert, außerdem war der Subtyp 4d dem Subtyp 4a überlegen.
Damit stellt die Düsseldorfer IDU-Kohorte für die häufigen Genotypen eine ähnliche Verteilung wie in anderen deutschen Großstädten dar und ist repräsentativ für Menschen in Deutschland, die i.v.-Drogen konsumieren, anzunehmen.

5.4 Vergleich unterschiedlicher Sequenziermethoden

Zur Verifizierung der Tiling-PCR und MinION-Sequenzierung sollte ein Vergleich zur etablierten Sequenziermethode nach Sanger [89] erfolgen. Dazu wurden die Sequenzen beider Methoden des 750 Nukleotid-langen NS5A-Abschnitts des Genotyps 3a einander mit Hilfe eines phylogenetischen Baums gegenübergestellt. Es zeigte sich, dass alle untersuchten Isolate eine hohe genetische Verwandtschaft hatten und auf demselben Ast im phylogenetischen Baum lagen, was bedeutet, dass zwischen den Sequenzen beider Methoden nur geringe Unterschiede bestanden.

Stichprobenartig wurden die Sequenzen von zehn Proben auf Nukleotid-Ebene miteinander verglichen (Vgl.4.6.1). Die Unterschiede betrugen 1-10 nt und waren fast ausschließlich auf Positionen mit Ambiguitäten, daher unterschiedlichen Virusvarianten, die sich an einer Stelle unterschieden, zurückzuführen. Werden zwei unterschiedliche Basen an der gleichen Stelle, z.B. aufgrund von Quasispezies, gemessen, wird das als Ambiguität oder Mehrdeutigkeit bezeichnet. Häufiger wurden diese Ambiguitäten nach der Messung mit der Nanopore angezeigt und in den meisten Fällen wies die Sanger-Sequenzierung an dieser Stelle eines der beiden Nukleotide auf. In der Sanger-Sequenzierung werden die Balken des Chromatogramms angezeigt und der Algorithmus bildet die Sequenz nach der Mehrheit der Nukleotid-Peakhöhe. Aus diesen technischen Gründen können deshalb Ambiguitäten < 25 % nicht richtig erkannt werden. In der Nanopore-Sequenzierung wird nicht nur ein Chromatogramm-Peak, sondern jedes PCR-Produkt einzeln betrachtet und anschließend die Gesamtheit statistisch ausgewertet. Zur Bildung einer Consensus-Sequenz von hoher Qualität wird ein Cut-Off von 60 % gewählt, was bedeutet, dass 60 % der häufigsten Reads berücksichtigt werden. Dies können auch zwei oder mehr unterschiedliche Nukleotide sein, die dann als Ambiguität interpretiert werden. Die Nanopore-Sequenzierung zeigt dementsprechend mehr Variabilität aufgrund einer tieferen Auflösung. Da der Vergleich zwischen der Tiling-PCR mit Nanopore-Sequenzierung zur Sanger-Sequenzierung nur stichprobenartig und ausschließlich für den Genotyp 3a durchgeführt wurde, besteht weiterer Validierungsbedarf.

Eine genauere, tiefere, aber auch teurere Sequenzierung stellt die Illumina-Sequenzierung dar. [81] Ob dies eine Alternative zur Nanopore-Sequenzierung sein könnte, sollte am Beispiel von zehn Genotyp 3a-Proben untersucht werden, die in der Tiling-PCR mit Nanopore-Sequenzierung einen verminderten sequenzierten Genomanteil gezeigt hatten (Vgl. 4.6.2).

Die Tiling-PCR mit Illumina-Sequenzierung erreichte dabei sowohl einen höheren sequenzierten Genomanteil über das gesamte Genom als auch höhere Coverage-Mittelwerte in den Resistenz-relevanten Regionen (Vgl. Tabelle 17) im Vergleich zur Nanopore-Sequenzierung.

Eine statistische Auswertung mit Korrelationen zwischen dem sequenzierten Genomanteil, der Viruslast und der Readmenge ist aufgrund der geringen Fallzahl von zehn Proben nicht abschließend möglich, aber es zeigte sich ein positiver Trend zwischen diesen Zusammenhängen. Vergleicht man die durchschnittliche Readmenge der zehn Proben miteinander, fällt auf, dass mittels Illumina-Sequenzierung mehr als dreimal so viele Sequenzierstücke als mittels Nanopore-Sequenzierung produziert werden, was den höheren sequenzierten Genomanteil erklären könnte. In Zukunft könnte dies durch die Verwendung von FlowCells mit mehr Poren, wie z.B. einer PromethION FlowCell [97] gelöst werden. Mit mehr Poren können in der gleichen Zeit mehr Nukleotide abgelesen und somit mehr Reads produziert werden.

Ein Vergleich der Sequenz beider Methoden zur NS5A-Sequenz der Sanger-Sequenzierung sollte Aufschluss über die Fehleranfälligkeit bzw. Genauigkeit der Methoden geben. Dabei zeigte sich jeweils im Vergleich zur Sanger-Methode, dass eine Tiling-PCR gefolgt von Illumina Sequenzierung weniger Nukleotidunterschiede aufwies, als wenn dieselbe Probe mittels Nanopore sequenziert wurde (Vgl. Tabelle 19).

Sowohl in der Höhe des sequenzierten Genomanteils als auch der Genauigkeit zeigt somit die Illumina- gegenüber der Nanopore-Sequenzierung Vorteile. Nachteilig sind hingegen höhere Kosten (ca. 120 Euro Sequenzierkosten pro Probe) als auch ein größerer Zeitaufwand (ca. zwei Wochen) durch einen längeren Sequenzierzeitraum als auch das Verschicken der Proben zu einem Vertragspartner. Durch eine geringe Fallzahl ist die Aussagekraft des Vergleiches zwischen den Sequenziermethoden eingeschränkt und wurde im Rahmen der vorliegenden Dissertation nur beispielhaft erarbeitet. Folgeuntersuchungen mit höherer Fallzahl und weiteren Genotypen wären daher sinnvoll. Dennoch decken sich die Beobachtungen mit bereits vorhandenen Validierungen aus der Literatur. Dabei zeigte die Sequenzierung mittels Nanopore eine höhere Fehlerrate im Vergleich zur Illuminasequenzierung, welche aber bei einer Ganzgenomsequenzierung durch überlappende Reads und der Bildung einer Consensus Sequenz teilweise kompensiert werden kann. [80, 98] Bei der in dieser Dissertation verwendeten Fließzelle R9.4 wurde nachgewiesen, dass eine 100fache Coverage ausreicht, um entstehende Fehler zu kompensieren. [99] Mit Verwendung der kürzlich veröffentlichten

Fließzelle R10 kann dies sogar auf eine 20fache Coverage reduziert werden [100], weshalb die Verwendung dieser für Folgearbeiten empfohlen wird.

Zusammenfassend zeigt die Nanopore-Sequenzierung mehr Vorteile, wenn eine schnelle, lokale und kostengünstige Diagnostik erwünscht ist [101] und die Illumina-Sequenzierung, wenn eine detailliertere Analyse mit hoher Qualität, z.B. zur Identifizierung von Minderheitsvarianten innerhalb von Patienten das Ziel ist. [98]

5.5 Vergleich zu anderen Ganzgenom-PCR Methoden

In der Literatur sind bereits mehrere Ansätze zur Etablierung einer Ganzgenom-PCR beschrieben. Im Folgenden sollen deren Ergebnisse mit denen der in dieser Arbeit verwendeten Tiling-PCR-Methode verglichen werden. Das genaue Vorgehen der einzelnen Ansätze wurde in 1.10 bereits beschrieben.

Im Gegensatz zu einer Tiling-PCR mit kurzen Fragmenten entwickelte A. Bull [69] im Jahr 2016 eine Ganzgenom-Sequenzierung, welche fast das gesamte Genom mittels einer langen und einer kurzen PCR amplifiziert. Für Proben der Genotypen 1-6 mit Viruslasten > 14.800 IU/ml konnten dabei Erfolgsraten von 90,5 % erreicht werden. Der Amplifikationserfolg wurde dabei mit phylogenetischer Analyse zu einem Referenzgenom als auch durch Densitometrie der ausreichenden Länge des PCR-Produkts mittels ImageJ überprüft. Bei geringen Viruslasten funktionierte die Bildung des langen PCR-Produkts nicht, auch eine vorherige Ultrazentrifugation konnte nur eine geringe Verbesserung von 30 % erreichen.

Eine ähnliche Strategie verfolgte H. R. Lapointe [70] im Jahr 2021 mit einer Genotyp-unabhängigen nested-PCR mit einem langen und einem kurzen Fragment. In den Versuchen zeigte sich wie für die Tiling-PCR ein Zusammenhang zur Viruslast und eine unterschiedliche Erfolgsquote der einzelnen Genotypen. Bei einer Viruslast von > 316.000 IU/ml wurde eine Sequenzerfolgsrate (definiert als > 100fache Abdeckung des Genoms) von durchschnittlich annähernd 90 % erreicht werden.

Ein direkter Vergleich zu der in dieser Dissertation untersuchten Tiling-PCR ist schwierig, da der Sequenziererfolg anders definiert wurde. Durch den Algorithmus wurde in der Consensus Sequenz ein N gebildet, wenn eine < 30-fache Genomabdeckung vorlag. Daraus wurde ein sequenzierter Genomanteil in % erzeugt, der angab, wie viel Prozent des Genoms Nukleotidbuchstaben und daher kein N enthält. Der Durchschnitt aller Proben der fünf untersuchten Genotypen zeigte einen sequenzierten Genomanteil von 91 % ($\pm$ 15 %) an.

Davon hatten 53 Proben eine Viruslast < 300.000 IU/ml, deren sequenzierter Genomanteil im Durchschnitt 84,4 % (± 22 %) betrug. Die Probe mit der niedrigsten Viruslast stellt z.B. eine GT 3a-Probe mit 2.242 IU/ml dar, welche mit einem sequenzierten Genomanteil von 90 % erfolgreich abgedeckt werden konnte. Damit ist die Tiling-PCR auch für Proben mit geringerer Viruslast effektiver anwendbar und kann mindestens einen Großteil des Genoms abdecken. Der entscheidende Vorteil der Tiling-PCR gegenüber einer PCR mit langen PCR-Fragmenten ist die Tatsache, dass es nicht ein Alles-oder-nichts-Ergebnis gibt, sondern dass sich mit der Tiling-PCR immer mindestens einige Bereiche erfolgreich amplifizieren und sequenzieren lassen. Somit kann auch bei Proben mit niedriger Viruslast oder schlechter RNA-Qualität noch eine Sequenzierung durchgeführt werden.

Ein anderer Ansatz einer Ganzgenom-PCR des Hepatitis C Virus ist die Genotyp-unabhängige Ganzgenomsequenzierung mittels 120 bp langen Hybridisierungssonden. F. Manso [67] beschrieb im Jahr 2021 einen solchen Ansatz und etablierte ihn als Teil des klinischen Behandlungsprogramms des National Health Service in Großbritannien. Mit dieser Methode wurde die Bestimmung von antiviralen Resistenzen sowie die Bestimmung des Genotyps untersucht. Bei Proben mit Viruslasten > 398.100 IU/ml wurde eine Sensitivität von > 90 % gemessen, bei Proben mit Viruslasten > 31.600 IU/ml von > 50 %. Dies beschreibt eine ähnliche Abhängigkeit zur Viruslast, wie auch für die Tiling-PCR beobachtet wurde.
Eine Anreicherung durch Hybridisierungssonden kann höhere Genomabdeckungen bei Proben mit geringerer Viruslast oder bei durch Therapie beeinträchtigten Proben erreichen, schlussfolgert die Arbeitsgruppe um C. Munyuza [68] im Rahmen ihrer Übersichtsarbeit 2022. Durch die spezifische Bindung eines Barcodes und Adapters an die DNA der Zielregion kann diese anschließend angereichert und tiefensequenziert werden. Durch diese Anreicherung können virale Sequenzen hundert- oder tausendfach sensitiver nachgewiesen werden. [102] Für den klinischen Alltag stellt diese Methode aber bisher keine Alternative dar, da sie teuer und mit langen und komplexen Protokollen zu aufwendig ist. [68] Dennoch könnte dieses Verfahren in Zukunft eine Alternative für Proben mit niedrigen Viruslasten werden.
Auch die Verwendung von Random Pentadecameren während der Reversen Transkription könnte eine Möglichkeit darstellen, um eine höhere cDNA-Ausbeute und -Qualität zu erzeugen, was für Proben mit einer niedrigen Viruslast von Vorteil sein könnte. [91]

Zusammenfassend zeigt sich die Tiling-PCR im Vergleich zu anderen Ganzgenom-PCR-Methoden des Hepatitis C Virus als kostengünstige Methode mit einem unkomplizierten Workflow. Auch wenn die Ansätze aufgrund unterschiedlicher Erfolgsdefinitionen nicht gänzlich vergleichbar sind, kann die Tiling-PCR eine ähnlich hohe oder höhere Genomabdeckung erreichen. Bei niedrigen Viruslasten, z.B. < 300.000 IU/ml, oder einem inkompletten viralen Genom kann durch die Tiling-PCR immer noch ein Großteil des Genoms abgedeckt werden, während Methoden, die ein langes PCR-Fragment bilden, schon nicht mehr funktionieren. Dennoch arbeitet auch die Tiling-PCR bei Proben mit niedriger Viruslast schlechter und ist dadurch limitiert. Dies könnte z.B. durch Kombination einer Anreicherung mit Hybridisierungssonden und einer anschließenden Tiling-PCR verbessert werden.

6 Schlussfolgerungen / Ausblick

Die WHO hat sich zum Ziel gesetzt, dass bis zum Jahr 2030 die Virushepatitis keine Bedrohung der öffentlichen Gesundheit mehr darstellt. [1] Um dies zu erreichen, ist die Diagnose der Infektion unerlässlich. Durch eine Sequenzierung des gesamten HCV-Genoms kann sowohl eine frühzeitige Anpassung der Medikamente bei Resistenzen erfolgen als auch durch phylogenetische Untersuchungen die Infektionswege nachverfolgt werden. Dadurch könnten prophylaktische Maßnahmen optimiert werden, um Neuinfektionen zu vermeiden.

Hierfür wurde im Rahmen dieser Dissertation eine Ganzgenom-PCR für die Genotypen 1a, 1b, 3a, 4a und 4d entwickelt. Mögliche Ziele weiterführender Arbeiten könnten die Optimierung, Validierung und Etablierung dieser Methode in die klinische Diagnostik des Instituts für Virologie der Uniklinik Düsseldorf darstellen.

Während der Entwicklung der Ganzgenom-PCR wurden die Primer mit Hilfe von Referenzsequenzen optimal angepasst, ggf. ihre Länge zur Anpassung der Annealing-Temperatur variiert oder auch die Konzentration einzelner Primer erhöht, damit eine hohe Genomabdeckung über das gesamte Genom erreicht werden konnte. Dabei zeigte sich die Einteilung der Primer in zwei Pools, wie für den Genotyp 3a und 4 entwickelt, als vorteilhafter als die Einteilung in drei Pools, weshalb dies in Zukunft auch für die Primersets der Genotypen 1a und 1b anzustreben wäre.

Für eine Resistenzbestimmung ist eine besonders hohe Coverage von > 10.000 nötig. Um dies zu erreichen, wurde beispielhaft für den Genotyp 1a ein vierter Pool mit etablierten Primern der Genomabschnitte der Protein-Regionen E1, NS3, NS5A, NS5B und Core gebildet und zu den bestehenden Pools der PCR hinzugefügt (Vgl. 4.4.1). Die Entwicklung eines solchen weiteren Primer-Pools könnte auch für die anderen Genotypen vorteilhaft sein und in Zukunft etabliert werden.

Stichprobenartig wurde für den Genotyp 3a ein weiteres Sequenzierungsverfahren für Proben getestet, die zuvor einen vermindertem Sequenziererfolg hatten (Vgl.4.6.2 und 4.6.3.). Anstelle der Nanopore-Sequenzierung wurde das Illumina-Verfahren genutzt, welches tiefer und genauer sequenziert. Dabei wurden sowohl höhere Readmengen als auch ein höherer sequenzierter Genomanteil und damit ein höherer Sequenziererfolg beobachtet. Eine Illumina-Sequenzierungen ist allerdings erheblich kosten- und zeitaufwendiger, weshalb sie eher für spezifische Fragestellungen oder für Proben, die mit der Nanopore-Sequenzierung nicht zufriedenstellend bearbeitet werden konnten, eine Alternative darstellt. Eine Nanopore-

Sequenzierung könnte sich hingehen besser als kostengünstige, zeitsparende und vor Ort durchführbare Methode für die Routinediagnostik eignen.

Im Vergleich zu anderen Ganzgenom-PCR-Methoden lassen die dargestellten Ergebnisse darauf schließen, dass das HCV-Genom mit der entwickelten Tiling-PCR mindestens vergleichbar oder sogar erfolgreicher sequenziert werden konnte (Vgl. 5.5). Limitierend sind dabei weiterhin geringe Viruslasten oder verminderte Probenqualitäten. Methoden, die mit einer Anreicherung durch Sondenköder arbeiten, zeigen in dieser Hinsicht Vorteile und könnten daher als Alternative für solche Proben in Frage kommen.

Das beschriebene Protokoll der Tiling-PCR ist Genotyp-spezifisch, weshalb eine vorherige Bestimmung des Genotyps nötig ist. Daher könnte in weiterführenden Arbeiten ein Genotyp-unabhängiges Primerset erarbeitet werden, da dies eine Zeitersparnis im diagnostischen Ablauf ermöglichen würde. Außerdem könnte eine Erweiterung der Tiling-PCR auf alle acht Genotypen des Hepatitis C Virus angestrebt werden.

Mit der entwickelten Tiling-PCR kann das gesamte Genom des Hepatitis C Virus zeit- und kostengünstig sequenziert werden. Durch die Bildung kleiner Fragmente kann sogar bei einem nicht kompletten viralen Genom dennoch der Großteil der Sequenz abgebildet werden. Aus den genannten Gründen sollte sich die entwickelte Methode besonders in der klinischen Routinediagnostik sowie zur integrierten genomischen Surveillance eignen. Hierfür wird sie seit dem Frühjahr 2022 am Institut für Virologie der Uniklinik Düsseldorf eingesetzt.

7 Abbildungsverzeichnis

Abbildung 1: Strukturübersicht des Genoms des Hepatitis C Virus 3
Abbildung 2: Übersicht der globalen Genotyp-Verteilung vom Hepatitis C Virus (HCV)..... 5
Abbildung 3: Coverage der Probe #41790 des Genotyps 1b im Versuch 1 und 2 33
Abbildung 4: Gel-Elektrophorese von sieben Proben des Genotyps 1a im Versuch 1 34
Abbildung 5: Coverage der Probe #32060 des Genotyps 1a im Versuch 1 und 2 37
Abbildung 6: Coverage der Probe #35320 des Genotyps 1a im Versuch 2 und 3 38
Abbildung 7: Coverage der Probe #35320 des Genotyps 1a im Versuch 3 und 4 39
Abbildung 8: Gel-Elektrophorese von vier Proben des Genotyps 1b im Versuch 1 41
Abbildung 9: Coverage der Probe #11981 des Genotyps 1b im Versuch 2 und 3 44
Abbildung 10: Gel-Elektrophorese von vier Proben des Genotyps 3a im Versuch 1 47
Abbildung 11: Coverage der Probe #52058 des Genotyps 3a im Versuch 1 47
Abbildung 12: Coverage der Probe #43557 des Genotyps 3a im Versuch 2 und 3 49
Abbildung 13: Gel-Elektrophorese von vier Proben des Genotyps 4 im Versuch 1 51
Abbildung 14: Coverage der Probe #743 des Genotyps 4d und #775 des Genotyps 4a 52
Abbildung 15: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 1a..... 55
Abbildung 16: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 1a mit und ohne vierten Primer-Pool 57
Abbildung 17: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 1b 60
Abbildung 18: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 3a..... 63
Abbildung 19: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Anzahl der Reads der Sequenzierung (Readmenge) und der Viruslast in IU/ml des Genotyps 4 66
Abbildung 20: Sequenziererfolg Tiling-PCR Gesamt..... 68
Abbildung 21: Sequenziererfolg Tiling-PCR unterteilt nach Viruslast 69
Abbildung 22: Phylogenetischer Baum der mit Hepatitis C Virus infizierten Kohorte Düsseldorf 71
Abbildung 23: Korrelationen zwischen dem sequenzierten Genomanteil in %, der Readmenge und der Viruslast in IU/ml des Genotyps 3a mit Nanopore und Illumina-Sequenzierung 75

8 Tabellenverzeichnis

Tabelle 1: Protokoll der Reversen Transkription 23
Tabelle 2: Programm der Tiling-Polymerasekettenreaktion (PCR) 24
Tabelle 3: Neue Primer Genotyp 1a 36
Tabelle 4: Primerset Genotyp 1a 40
Tabelle 5: Neue Primer Genotyp 1b 43
Tabelle 6: Vergleich der Mittelwerte der Coverage des Genotyps 1b 45
Tabelle 7 Primerset Genotyp 1b 45
Tabelle 8: Neue Primer Genotyp 3a 48
Tabelle 9: Primerset Genotyp 3a 50
Tabelle 10: Primerset Genotyp 4 53
Tabelle 11: Vergleich der Mittelwerte der Coverage von Genotyp 1a mit und ohne Pool 4 .. 58
Tabelle 12: Vergleich Resistenztestung mit und ohne Pool 4 59
Tabelle 13: Mittelwerte der Coverage des Genotyps 1b 61
Tabelle 14: Mittelwerte der Coverage des Genotyps 3a 63
Tabelle 15: Mittelwerte der Coverage des Genotyps 4 67
Tabelle 16: Resistenztestung der Düsseldorfer IDU-Kohorte 72
Tabelle 17: Vergleich der Mittelwerte der Coverage zwischen Illumina- und Nanopore-Sequenzierung 76
Tabelle 18: Vergleich Resistenztestung Nanopore vs. Illumina 76
Tabelle 19: Vergleich Nanopore-/Illumina- zur Sanger-Sequenzierung 77

9 Literaturverzeichnis

1. (WHO), W.H.O. *Hepatitis C Key facts*. 2024 [cited 2025 30.01.]; 09.04.2024:[Available from: https://www.who.int/news-room/fact-sheets/detail/hepatitis-c.
2. Perz, J.F., et al., *The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide.* J Hepatol, 2006. 45(4): p. 529-38.
3. Alter, H., et al., *CLINICAL AND SEROLOGICAL ANALYSIS OF TRANSFUSION-ASSOCIATED HEPATITIS.* The Lancet, 1975. **306**(7940): p. 838-841.
4. 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.
5. Kuo, G., et al., *An Assay for Circulating Antibodies to a Major Etiologic Virus of Human Non-A, Non-B Hepatitis.* Science, 1989. **244**(4902): p. 362-364.
6. Kolykhalov, A.A., et al., *Transmission of Hepatitis C by Intrahepatic Inoculation with Transcribed RNA.* Science, 1997. **277**(5325): p. 570-574.
7. Bartenschlager, R., S. Urban, and U. Protzer, *Towards curative therapy of chronic viral hepatitis.* Z Gastroenterol, 2019. **57**(1): p. 61-73.
8. Institut, R.K. *Hepatitis C RKI Ratgeber*. 2018 [cited 2025 30.01.]; 31.01.2018:[Available from: https://www.rki.de/DE/Aktuelles/Publikationen/RKI-Ratgeber/Ratgeber/Ratgeber_HepatitisC.html?nn=16911148.
9. Bühler, 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.
10. Lohmann, V. and R. Bartenschlager, *On the history of hepatitis C virus cell culture systems.* J Med Chem, 2014. **57**(5): p. 1627-42.
11. Moradpour, D. and F. Penin, *Hepatitis C virus proteins: from structure to function.* Curr Top Microbiol Immunol, 2013. **369**: p. 113-42.
12. Jopling, C.L., et al., *Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA.* Science, 2005. **309**(5740): p. 1577-81.
13. Henke, J.I., et al., *microRNA-122 stimulates translation of hepatitis C virus RNA.* Embo j, 2008. **27**(24): p. 3300-10.
14. Friebe, P. and R. Bartenschlager, *Genetic analysis of sequences in the 3' nontranslated region of hepatitis C virus that are important for RNA replication.* J Virol, 2002. **76**(11): p. 5326-38.
15. Tsukiyama-Kohara, K. and M. Kohara, *Hepatitis C Virus: Viral Quasispecies and Genotypes.* Int J Mol Sci, 2017. **19**(1).
16. 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.
17. Hijikata, M., et al., *Hypervariable regions in the putative glycoprotein of hepatitis C virus.* Biochem Biophys Res Commun, 1991. **175**(1): p. 220-8.
18. Farci, P., et al., *Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein.* Proc Natl Acad Sci U S A, 1996. **93**(26): p. 15394-9.
19. Nakano, T., et al., *An updated analysis of hepatitis C virus genotypes and subtypes based on the complete coding region.* Liver Int, 2012. **32**(2): p. 339-45.
20. Smith, D.B., et al., *The origin of hepatitis C virus genotypes.* J Gen Virol, 1997. **78 (Pt 2)**: p. 321-8.

21. Simmonds, P., et al., *Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region.* J Gen Virol, 1993. **74 (Pt 11)**: p. 2391-9.
22. Simmonds, P., *Variability of hepatitis C virus.* Hepatology, 1995. **21**(2): p. 570-83.
23. Simmonds, P., et al., *Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes.* Hepatology, 2005. **42**(4): p. 962-73.
24. Smith, D.B., et al., *Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource.* Hepatology, 2014. **59**(1): p. 318-27.
25. Peter Simmons, I., *Confirmed HCV genotypes/subtypes.* 2019.
26. Steinhauer, D.A., E. Domingo, and J.J. Holland, *Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase.* Gene, 1992. **122**(2): p. 281-8.
27. Drake, J.W. and J.J. Holland, *Mutation rates among RNA viruses.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13910-3.
28. Fishman, S.L. and A.D. Branch, *The quasispecies nature and biological implications of the hepatitis C virus.* Infect Genet Evol, 2009. **9**(6): p. 1158-67.
29. Martell, M., et al., *Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution.* J Virol, 1992. **66**(5): p. 3225-9.
30. Gower, E., et al., *Global epidemiology and genotype distribution of the hepatitis C virus infection.* J Hepatol, 2014. **61**(1 Suppl): p. S45-57.
31. Messina, J.P., et al., *Global distribution and prevalence of hepatitis C virus genotypes.* Hepatology, 2015. **61**(1): p. 77-87.
32. Magiorkinis, G., et al., *The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis.* PLoS Med, 2009. **6**(12): p. e1000198.
33. 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.
34. Kartashev, V., et al., *New findings in HCV genotype distribution in selected West European, Russian and Israeli regions.* J Clin Virol, 2016. **81**: p. 82-9.
35. 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.
36. Kubitschke, A., et al., *[Injuries from needles contaminated with hepatitis C virus: how high is the risk of seroconversion for medical personnel really?].* Internist (Berl), 2007. **48**(10): p. 1165-72.
37. Jafari, S., et al., *Tattooing and the risk of transmission of hepatitis C: a systematic review and meta-analysis.* Int J Infect Dis, 2010. **14**(11): p. e928-40.
38. 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.
39. Rice, C.M. and M. Saeed, *Hepatitis C: Treatment triumphs.* Nature, 2014. **510**(7503): p. 43-4.
40. Poynard, T., et al., *Viral hepatitis C.* Lancet, 2003. **362**(9401): p. 2095-100.
41. Lohmann, V., et al., *Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line.* Science, 1999. **285**(5424): p. 110-3.
42. Boettler, T., V. Lohmann, and R. Bartenschlager, *[Hepatitis C: From Individual Cure to Worldwide Elimination?].* Dtsch Med Wochenschr, 2019. **144**(8): p. 535-542.
43. Schaefer, E.A. and R.T. Chung, *Anti-hepatitis C virus drugs in development.* Gastroenterology, 2012. **142**(6): p. 1340-1350.e1.

44. Chung, R.T. and T.F. Baumert, *Curing chronic hepatitis C--the arc of a medical triumph.* N Engl J Med, 2014. **370**(17): p. 1576-8.
45. (KBV), K.B. *Gesundheitscheck-up - Screening auf Hepatitis B und C*. 2022 15.07.2022 [cited 2025 30.01.]; Available from: https://www.kbv.de/html/5540.php.
46. Pawlotsky, J.M., *Hepatitis C Virus Resistance to Direct-Acting Antiviral Drugs in Interferon-Free Regimens.* Gastroenterology, 2016. **151**(1): p. 70-86.
47. Pawlotsky, J.M., *Hepatitis C virus population dynamics during infection.* Curr Top Microbiol Immunol, 2006. **299**: p. 261-84.
48. DGVS. *S3-Leitlinie „Prophylaxe, Diagnostik und Therapie der Hepatitis-C-Virus (HCV) - Infektion“*. AWMF Register-No: 021/012 2018 [cited 2025 30.01.]; 28.03.2018:[Available from: https://register.awmf.org/de/leitlinien/detail/021-012.
49. Walker, A., et al., *Genotypic resistance testing of HCV – is there a clinical need?* GMS Infectious Diseases, 2016. **4**.
50. Sandmann, L., et al., *Treatment of Chronic Hepatitis C: Efficacy, Side Effects and Complications.* Visc Med, 2019. **35**(3): p. 161-170.
51. Chamberlain, J.S., et al., *Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification.* Nucleic acids research, 1988. **16**(23): p. 11141-11156.
52. Edwards, M.C. and R.A. Gibbs, *Multiplex PCR: advantages, development, and applications.* PCR Methods Appl, 1994. **3**(4): p. S65-75.
53. Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.* Cold Spring Harb Symp Quant Biol, 1986. **51 Pt 1**: p. 263-73.
54. Elnifro, E.M., et al., *Multiplex PCR: optimization and application in diagnostic virology.* Clin Microbiol Rev, 2000. **13**(4): p. 559-70.
55. Jin, L., A. Richards, and D.W. Brown, *Development of a dual target-PCR for detection and characterization of measles virus in clinical specimens.* Mol Cell Probes, 1996. **10**(3): p. 191-200.
56. Zou, S., C. Stansfield, and J. Bridge, *Identification of new influenza B virus variants by multiplex reverse transcription-PCR and the heteroduplex mobility assay.* J Clin Microbiol, 1998. **36**(6): p. 1544-8.
57. Chamberlain, J.S., et al., *Diagnosis of Duchenne and Becker muscular dystrophies by polymerase chain reaction. A multicenter study.* Jama, 1992. **267**(19): p. 2609-15.
58. Huang, H.S., et al., *Multiplex PCR system for the rapid diagnosis of respiratory virus infection: systematic review and meta-analysis.* Clin Microbiol Infect, 2018. **24**(10): p. 1055-1063.
59. Reijans, M., et al., *RespiFinder: a new multiparameter test to differentially identify fifteen respiratory viruses.* J Clin Microbiol, 2008. **46**(4): p. 1232-40.
60. Li, H., et al., *Simultaneous detection and high-throughput identification of a panel of RNA viruses causing respiratory tract infections.* J Clin Microbiol, 2007. **45**(7): p. 2105-9.
61. Quick, J., 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): p. 1261-1276.
62. Gardy, J., N.J. Loman, and A. Rambaut, *Real-time digital pathogen surveillance - the time is now.* Genome Biol, 2015. **16**(1): p. 155.
63. Dudas, G., et al., *Virus genomes reveal factors that spread and sustained the Ebola epidemic.* Nature, 2017. **544**(7650): p. 309-315.
64. Gire, S.K., et al., *Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak.* Science, 2014. **345**(6202): p. 1369-72.

65. Walker, A., et al., *Characterization of SARS-CoV-2 infection clusters based on integrated genomic surveillance, outbreak analysis and contact tracing in an urban setting.* Clin Infect Dis, 2021.
66. Fitzpatrick, A.H., et al., *High Throughput Sequencing for the Detection and Characterization of RNA Viruses.* Front Microbiol, 2021. **12**: p. 621719.
67. Manso, C.F., et al., *Technical Validation of a Hepatitis C Virus Whole Genome Sequencing Assay for Detection of Genotype and Antiviral Resistance in the Clinical Pathway.* Front Microbiol, 2020. **11**: p. 576572.
68. Munyuza, C., H. Ji, and E.R. Lee, *Probe Capture Enrichment Methods for HIV and HCV Genome Sequencing and Drug Resistance Genotyping.* Pathogens, 2022. **11**(6).
69. Bull, R.A., et al., *A method for near full-length amplification and sequencing for six hepatitis C virus genotypes.* BMC Genomics, 2016. **17**: p. 247.
70. Lapointe, H.R., et al., *Validation of a Genotype-Independent Hepatitis C Virus Near-Whole Genome Sequencing Assay.* Viruses, 2021. **13**(9).
71. Koch-Institut, R. *GBE-Themenheft Hepatitis C*. 2016 [cited 2025 30.01.]; 25.07.2016:[Available from: https://www.rki.de/DE/Themen/Nichtuebertragbare-Krankheiten/Downloads/Themenhefte/hepatitis_c_2016_inhalt.html?nn=16911148.
72. Pawlotsky, J.M., *Retreatment of Hepatitis C Virus-Infected Patients with Direct-Acting Antiviral Failures.* Semin Liver Dis, 2019. **39**(3): p. 354-368.
73. Malandris, K., et al., *The Role of RASs /RVs in the Current Management of HCV.* Viruses, 2021. **13**(10).
74. Bartlett, S.R., et al., *Sequencing of hepatitis C virus for detection of resistance to direct-acting antiviral therapy: A systematic review.* Hepatol Commun, 2017. **1**(5): p. 379-390.
75. 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.* Virology journal, 2013. **10**: p. 53-53.
76. 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.
77. GmbH, N.E.B. *Q5 High Fidelity DNA Polymerase*. 2021 [cited 2025 30.01.]; 01.01.2021:[Available from: https://www.neb-online.de/pcr-dna-amplifikation/high-fidelity-pcr/q5-high-fidelity-dna-polymerase/.
78. Co.KG, B.T.G. *Luminometer*. [cited 2025 30.01.]; Available from: https://www.berthold.com/de/bioanalytik/wissen/glossar-bioanalytik/luminometer/.
79. Technologies, O.N. *How nanopore sequencing works*. [cited 2025 30.01.]; Available from: https://nanoporetech.com/platform/technology.
80. Magi, A., B. Giusti, and L. Tattini, *Characterization of MinION nanopore data for resequencing analyses.* Brief Bioinform, 2017. **18**(6): p. 940-953.
81. Slatko, B.E., A.F. Gardner, and F.M. Ausubel, *Overview of Next-Generation Sequencing Technologies.* Curr Protoc Mol Biol, 2018. **122**(1): p. e59.
82. Greninger, A.L., et al., *Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis.* Genome Med, 2015. **7**: p. 99.
83. Inc., N.E.B. *NEBNext ARTIC SARS-CoV-2 Companion Kit Instruction Manual Version 6.0_2/22*. 2022 [cited 2025 30.01.]; Available from: https://international.neb.com/-/media/nebus/files/manuals/manuale7660-with-varskip-short-v2-sars-cov-2-primers-v6.pdf?rev=0f8fe49e8fe8447c9015857bf3705a1d&hash=6C621E952E48D482F3CB2E3096B71E15.
84. Technologies, O.N. *Nanopore Protocol Amplicon barcoding with Native Barcoding Expansion 96*. 2023 10.03.2023 [cited 2025 30.01.]; Available from:

https://community.nanoporetech.com/docs/prepare/library_prep_protocols/amplicon-barcoding-with-native-barcoding-expansion-96-exp-nbd196-and-sqk/v/nba_9102_v109_revj_09jul2020.

85. Co.KG, S.-I.G. *„Next-Generation-Sequencing“*. [cited 2025 30.01.]; Available from: https://seq-it.de/sequenzierung/.
86. Inc., I. *Illumina Sequencing Technology*. 2010 [cited 2025 30.01.]; Available from: https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf.
87. 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.
88. (RKI), R.K.I., *Qualitätskriterien Rohdaten DESH-BI-QC-v2*. 2021.
89. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* J Mol Biol, 1975. **94**(3): p. 441-8.
90. (RKI), R.K.I. *Epidemiologisches Bulletin 2022*. 2022 [cited 2025 30.01.]; Available from: https://www.rki.de/DE/Aktuelles/Publikationen/Epidemiologisches-Bulletin/2022/38_22.pdf?__blob=publicationFile&v=1.
91. Stangegaard, M., I.H. Dufva, and M. Dufva, *Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA.* Biotechniques, 2006. **40**(5): p. 649-57.
92. Koch, D., *PCR: Polymerase-Kettenreaktion*, in *Biochemie des Menschen*, F. Horn, Editor. 2018, Georg Thieme Verlag.
93. Steitz, T.A., *DNA polymerases: structural diversity and common mechanisms.* J Biol Chem, 1999. **274**(25): p. 17395-8.
94. Liu, Z., et al., *World-wide Prevalence of Substitutions in HCV Genome Associated With Resistance to Direct-Acting Antiviral Agents.* Clin Gastroenterol Hepatol, 2021. **19**(9): p. 1906-1914.e25.
95. Smith, M.R., et al., *Rapid incidence estimation from SARS-CoV-2 genomes reveals decreased case detection in Europe during summer 2020.* Nature Communications, 2021. **12**(1): p. 6009.
96. Zimmermann, R., *Abschlussbericht der Studie „Drogen und chronischen Infektionskrankheiten in Deutschland“ (DRUCK-Studie)*. 2016, Robert Koch-Institut: Berlin.
97. Technologys, O.N. *Flow Cells*. [cited 2025 30.01.]; Available from: https://store.nanoporetech.com/eu/flow-cells.html.
98. Nieuwenhuijse, D.F., et al., *Towards reliable whole genome sequencing for outbreak preparedness and response.* BMC Genomics, 2022. **23**(1): p. 569.
99. Oude Munnink, B.B., et al., *Towards high quality real-time whole genome sequencing during outbreaks using Usutu virus as example.* Infect Genet Evol, 2019. **73**: p. 49-54.
100. Oude Munnink, B.B., et al., *Validating Whole Genome Nanopore Sequencing, using Usutu Virus as an Example.* J Vis Exp, 2020(157).
101. Alleweldt, F., et al., *Economic evaluation of whole genome sequencing for pathogen identification and surveillance - results of case studies in Europe and the Americas 2016 to 2019.* Euro Surveill, 2021. **26**(9).
102. Miyazato, P., et al., *Application of targeted enrichment to next-generation sequencing of retroviruses integrated into the host human genome.* Sci Rep, 2016. **6**: p. 28324.

10 Anhang

1. Oligonukleotide

Alle Oligonukleotide der Tiling-PCR wurden in einer Sequenz von 5‘→ 3‘ der Firma Eurofins Scientific in einem Volumenverhältnis von 100 pM/μl bezogen.

1.1 Genotyp 1a

Olionukleotide mit dem Suffix „_new“ wurden nachträglich als Verbesserungsoption gebildet und als neue Primer dem Primerset hinzugefügt. Ein zusätzlicher Pool 4 wurde für 15 Proben des GT 1a verwendet, dabei sind alle relevanten Regionen zur Erkennung von Resistenzen im HCV-Genom abgedeckt.

Tabelle Anhang 1: Oligonukleotide des Genotyps 1a
Dargestellt sind Oligonukleotide in einer Sequenz von 5‘ → 3‘ des Genotyps 1a.

Name	Sequenz
1a-F	ATCACTCCCCTGTGAGG
1a-R	GGCAACAAGTAAACTCCACC
1b-F	CGGGAGGTCTCGTAGACC
1b-R	CCTGTTGCATAGTTCACGC
1c-F	GGTAAGGTCATCGATACC
1c-R	GATRTGACGTCGAAGCTGCG
1d-F	TCTTCCTTCTGGCCCTGC
1d-R	TCCCCAGTGRGCACCAGCG
1e-F	GGTGTGTCCCTTGCGTTCG
1e-F_new	CCGGGGTGTGTCCCTTGCGT
1e-R	CTATTGAKGTGCCAACTGCC
1e-R_new	CKATTGATGTGCCAACTGCC
2a-F	TATGATGATGAACTGGTCCC
2a-R	AGGTYTTGGRGGGTAGTGCC
2a-R_new	TAGTGCCAGCAGTAGGGGCG

2b-F	TCCATGGTGGGGAACTGGG
2b-R	CCAGTTGAGTTCATCCAGG
2c-F	TGGTGGTGGGAACGACCG
2c-R	CAGGAGAACGACGTACTCCC
2d-F	ATGTACGTGGGAGGGGTCG
2d-R	ACAACRCCGCCACACGACG
2e-F	CGTCTGCTCCTGCTTGTGG
2e-R	TGRACGCGCACRAAGTAGGG
3a-F	ATGTGGCCTCTCCTCCTGC
3a-R	CCCACGTGATGAGCTTGG
3b-F	TGGRTTCCCCCCCTCAACG
3b-R	ACCCCTTGGAGACCATTCC
3c-F	TCCRAATGGAGACCAAGC
3c-F_new	CCCRAATGGAGACCAAGC
3c-R	AATGACATCGGCGTGCCTCG
3d-F	GACAAAAACCAAGTGGAGGG
3d-R	GCCACCTGGAAGCTCTGGG
3e-F	ATAGCAGGGGYAGCCTGC
3e-R	ATCCGTGGAGTGGCACTCG
3e-R_short	ATCCGTGGAGTGGCACTC
4a-F	AGTGCCCCAGAGCTTCCAGG
4a-R	TCTCCGGTGGTGGACAGAGC
4b-F	CCTACGGCAAGTTCCTTGC
4b-F_new	GTAAGAGCACCAAGGTCCC
4b-R	AGCGTGRTTGTCTCAATGG
4b-R_new	GCAGTCTATCACCGAGTCG
4c-F	CGAGGAGGTTGCTCTGTCC
4c-R	AGCACAGCCYGCGTCATAGC

4d-F	TGTGTCACYCAGACAGTCG
4d-R	ACCCAGGTGCTCGTGACG
4e-F	ACCTGGTAGCGTACCAAGCC
4e-R	GGGCCCTTCTGCTTGAACTGC
5a-F	ACCCTCCATGGGCCAACACC
5a-R	GAGGACCTTCCCCAGTCC
5b-F	GCTCGCYGAGCAGTTCAAGC
5b-R	GGCTATYAGCCGGTTCATCC
5c-F	TGAGCGGTGAGGTCCCCTCC
5c-R	CATGTCCAGTGATCTCAGC
5d-F	TGGGACTGGATATGCGAGG
5d-R	GTGAAAAATTCGGGCGATGG
5e-F	ACACTCGCTGCCACTGTGG
5e-R	GAGGGATCAGTGAGCATGG
6a-F	AGGAACATGTGGAGTGGG
6a-R	GTTYTCTGACTCAACCCTGG
6b-F	CGCAATTACCTTGCGAGCC
6b-R	GATTCRGTGAGGACCACC
6c-F	TCCGCTCCATCTCTCAAGGC
6c-R	TYGACCATGACCCGTCGC
6d-F	AACCACCTGTGGTCCATGGC
6d-R	TGCCTTTGGCAAGCACTGCG
6e-F	GGARGAYGTCGTGTGCTGC
6e-R	TTACGACCCCCCTTCTCAGG
7a-F	GTAGAGGAAGCTTGCAGCC
7a-R	ATTGCCTCCTCCGTACGG
7b-F	CCACATCAACTCCGTGTGG
7b-R	CCACACACGAGCATGGTGC

7c-F	GGATTCCAATACTCACCAGG
7c-R	TTATGTTGCCTAGCCAGG
7d-F	GCTATGACCAGGTACTCCG
7d-R	TTGCCACATATGGCAGCC
7e-F	GCCTGCTACTCCATAGAACC
_7e-R	TTCATCGGTTGGGGAGGAGG
7e-R_int	CCTGCAGCAAGCAGGAGTAGGC
7e-R_ext	AGGCCGGAGTGTTTACC

Tabelle Anhang 2: Pool 4 für Resistenz-relevante Regionen des HCV-Genoms
Dargestellt sind Oligonukleotide in einer Sequenz von 5‘ → 3‘ eines 4. Pools für Resistenz-relevante Regionen des Hepatitis C Virus (HCV) – Genom des Genotyps 1a.

Region / Herkunft	Name	Sequenz
Core	p417s	GGYGGYGGNCAGATCGTTGG
	p874a	ARGAAGATAGARAARGAGCAACC
NS3 (GT 1a)	HCV1a_3a_F	ATGTGGCCTCTCCTCCTGC
	HCV1a_4a_R	TCTCCGGTGGTGGACAGAGC
NS5A (GT 1a)	HCV1a_5c_F	TGAGCGGTGAGGTCCCCTCC
	HCV1a_5e_R	GAGGGATCAGTGAGCATGG
NS5B (GT 1a)	HCV1a+b_7527-F	AYCCIGATCTCAGCGACGGRTC
	HCV1a_7e-R_int	CCTGCAGCAAGCAGGAGTAGGC
E1	493s_H77	GCAACAGGGAACCTTCCTGGTTGCTC
	987R_H77	CGTAGGGGACCAGTTCATCATCAT
NS5A_Genotypisierung	NS5A-6069-F	GAGGGGGCAGTGCAATGGATGAAYMGIYTIATIGCITTYGC
	NS5A-7541-R	AAGTCCGGGTCACCGGGGTCCCCYTCWARYGRAGGCWTTGA
	NS5A-7542-R	GAGATCCGGATCCCCAGGYTCICCYTCIAGIGGIGGCATIGA

1.2 Genotyp 1b

Oligonukleotide mit dem Suffix „_NEW“ wurden, wie bei Genotyp 1a, nachträglich bestellt. Bei diesem Primerset wurden die Pools neu zusammen pipettiert und die alten Primer durch die neuen Primer ersetzt.

Tabelle Anhang 3: Oligonukleotide des Genotyps 1b
Dargestellt sind Oligonukleotide in einer Sequenz von 5‘ → 3‘ des Genotyps 1b.

Name	Sequenz
HCV-1b_1a-F	ATCACTCCCCTGTGAGG
HCV-1b_1a-R	GCGCGGCAACAGGTAAACTCC
HCV-1b_1b-F	CGGGAGGTCTCGTAGACC
HCV-1b_1b-R	CCTGTTGCATAGTTCACGC
HCV-1b_1c-F	GGTAAGGTCATCGATACC
HCV-1b_1c-R	AGATCCCCCACGTACATRGC
HCV-1b_1d-F	CATGTCACGAACGACTGC
HCV-1b_1d-R	ACRGCTTGTGGGATCCGG
HCV-1b_1e-F	CAYGTCGAYTTGCTCGTTGG
HCV-1b_1e-R	TTGATGTGCCAGCTGCC
HCV-1b_1e-R_NEW	TTGATGTGCCARCTGCC
HCV-1b_2a-F	GGGATATGATGATGAACTGG
HCV-1b_2a-F_NEW	TGGGATATGATGATGAAYTGGTC
HCV-1b_2a-R	GAGGYGCGTAGTGCCAGC
HCV-1b_2a-R_NEW	GRGGYGCGTAGTGCCAGC
HCV-1b_2b-F	AACACCAAYGGCAGYTGG
HCV-1b_2b-R	TGCTTCCGGAAGCARTCCG
HCV-1b_2c-F	TGGTTYGGCTGTACATGG
HCV-1b_2c-R	GCTATCAGCAGCATCATCC
HCV-1b_2d-F	CTCCAYCAGAACATCGTGG
HCV-1b_2d-R	GGAGGATGATGGCATCGC
HCV-1b_2e-F	TGTTCTTCTGTGCYGCCTGG

HCV-1b_2e-R	CCRTTTGGACATAATGACC
HCV-1b_2e-R_NEW	CCATTTGGACATAATGRCCCCC
HCV-1b_3a-F	ATGGCTGCATCGTGCGGAGG
HCV-1b_3a-R	GAGCGCGYACRAAGTACGG
HCV-1b_3b-F	GCCGCGATGCCATCATCC
HCV-1b_3b-R	ATGGGCGCRAGGAGTCGC
HCV-1b_3b-R_NEW	ATRGGCGCRAGGAGTCGC
HCV-1b_3c-F	TTGCGGTGGCAGHAGAGC
HCV-1b_3c-F_NEW	TTGCGGTGGCAGTWGAGC
HCV-1b_3c-R	CCAARTAAAGGTCCGAGCTGCC
HCV-1b_3d-F	ACAAGAACCAGGTCGAGGG
HCV-1b_3d-R	GGAGCGTGTAGATGGGCC
HCV-1b_3e-F	CCTACYTGAAGGGCTCYTCGGG
HCV-1b_3e-R	CGCCCGTGGTGATGGTCC
HCV-1b_4a-F	ATGGAAACTACYATGCGG
HCV-1b_4a-R	CACGACRAGCCGCGCTCC
HCV-1b_4b-F	AAGGACCATCACCACGGG
HCV-1b_4b-R	CATTAGAGCGTCTGTTGC
HCV-1b_4c-F	CTATGGCAAAGCCATCCC
HCV-1b_4c-R_NEW	CCCTGGTGTATTTAGGTAAGCCC
HCV-1b_4c-R	GGTGTATTTAGGTAAGCCCG
HCV-1b_4d-F	GCATMTACAGGTTTGTGACTCC
HCV-1b_4d-R	CCAGGTGCTVGTGACGACC
HCV-1b_4e-F	CATCGTGGGAYCARATGTGG
HCV-1b_4e-R	TCTGCTTGAAYTGCTCGG
HCV-1b_4e-R_NEW	TCTGYTTGAAYTGCTCGGCGA
HCV-1b_5a-F	GTGGAAGTGYCTCAYACG
HCV-1b_5a-R	TCCCGCTGATGAARTTCC

HCV-1b_5b-F	CGARCAGGGAATGCAGCTCGC
HCV-1b_5b-R	AGCCGGTTCATCCACTGC
HCV-1b_5c-F	ATAGGCCTTGGGAAGGTGC
HCV-1b_5c-R	GCCAGACTCCCTTGTACCC
HCV-1b_5d-F	CCCCACGCACTATGTGCC
HCV-1b_5d-R	GGGCAYTTTACGTTGTCRGTGG
HCV-1b_5d-R_NEW	GGGCAYTTYACGTTGTCAGTGG
HCV-1b_5e-F	CACCTGCCCATGTGGAGC
HCV-1b_5e-R	AGGGGTCGGTRAGCATGG
HCV-1b_6a-F	CAACGCRTACACCACGGG
HCV-1b_6a-R	ATGTTYCCGCCCATCTCCTGCCG
HCV-1b_6b-F	TCACAGCTCCCATGYGAGCC
HCV-1b_6b-R	ACCGTCCTCTTYCTCCGTGG
HCV-1b_6c-F	ATGGGCGGRAACATCACCCG
HCV-1b_6c-R	TCAAGGGGGGGCATGGAGG
HCV-1b_6d-F	CCAATACCACCTCCACGG
HCV-1b_6d-R	CATCTCCTTGAGCACGTCCC
HCV-1b_6e-F	GTCCTACACATGGACAGG
HCV-1b_6e-R	TARAGGGCCATYTTCTCGC
HCV-1b_7a-F	ACCGGGACGTGCTBAAGG
HCV-1b_7a-R	CCGTTGAGTCAAARCAGCG
HCV-1b_7b-F	TTGAYACCACCATCATGGC
HCV-1b_7b-R	CTTYGCAGCTCGACAGGC
HCV-1b_7c-F	CGCTGYTTTGACTCAACGG
HCV-1b_7c-R	CATCAGAATCATCCTTGCCC
HCV-1b_7d-F	CACGGAGGCTATGACTAGG
HCV-1b_7d-R	CTTCTGGCCCGATGTCTCC
HCV-1b_7e-F	CYCACTTCTTCTCCATCC

HCV-1b_7e-F_NEW	TCTGATGACYCAYTTCTTCTCCATCCT
HCV-1b_7e-R	GGGGAGCAGGTAGATGCC

1.3 Genotyp 3a

Oligonukleotide mit dem Suffix „_new“ oder „alternative1/2“ wurden ebenfalls nachträglich bestellt. Wie bei Genotyp 1b wurden die Pools des Primersets neu zusammenpipettiert und die alten Primer durch die neuen Primer ersetzt.

Tabelle Anhang 4: Oligonukleotide des Genotyps 3a
Dargestellt sind Oligonukleotide in einer Sequenz von 5‘ → 3‘ des Genotyps 3a.

Name	Sequenz
HCV-3a_1L	AGTATGAGTGTCGTGCAGCCT
HCV-3a_1R	CGTTCAGAAGTTTTACGCGTCG
HCV-3a_2L	AACATGAGCACACTTCCTAAACC
HCV-3a_2R	GTACCCCATGAGGTCGGC
HCV-3a_3L	GCCCCTCTATGGTAACGAGG
HCV-3a_3R	CGGCCTCATACACAATACTGCT
HCV-3a_4L	CAGCTAGTYTAGAGTGGCGGAA
HCV-3a_4R	CAATTCATCATCATRTCCCAAGCCA
HCV-3a_5L	GCTCRCTGTACCCAGGCCA
HCV-3a-5L-alternative1	CGTCGCCATCAAACGGTCCAGA
HCV-3a-5l-alternative2	TTYCTYGTGGGACAAGCCTTCAC
HCV-3a_5R	CCAGCTATGAACCCGGTGTT
HCV-3a_5R_new	CCCAGCTATGAACCCKGTGT
HCV-3a_6L	GGCCTAGCCTATTACTCCATGC
HCV-3a_6L_new	GGCCTAGCCTATTACTCCATGCA
HCV-3a_6R	TAGTGCCAGCAGTACGGTCT
HCV3a-6R-new	ACRACCACTGGCGATGGTGT
HCV-3a_7L	TACTGCTTCACACCATCGCC
HCV3a-7L-new	GACAAACCGTACTGCTGGCA
HCV-3a_7R	CCATGCATCGAGGTGTCAACC

HCV-3a_8L	ACCGACTGCTTCAGGAAACATC
HCV-3a_8R	AACTCCCATTTCARCGCCCA
HCV-3a_9L	CCCATGCCTGCATTGTCAAC
HCV-3a_9R	CTGTCTTCACCCGACCAAGC
HCV-3a_10L	GTCCCTAGCATTGCTYGTCCT
HCV-3a_10R	AGGTGGTCRTATAGGTAGGTGTT
HCV-3a_11L	TGGTTCGCCTTTGCATGCT
HCV-3a_11R	TTCTTRTCCCTGCCAGTCAAGC
HCV-3a_12L	TGAGGTGTTGTTGGGRCCTG
HCV-3a_12R	ATCRGCWTCGCGGGTAACCA
HCV-3a_13L	CCCAGACCTTCCTAGGTACAA
HCV-3a_13R	AGGGTTTCCACTGGTATGAACTGTA
HCV-3a_14L	ATCTTTAGGGCYGCTGTGTG
HCV-3a_14R	ACATCATATGCYCCCCCGGA
HCV-3a_15L	AACCGCACCGTYACAACTGG
HCV-3a_15R	GCTACAGCRTTGAGCCCCA
HCV-3a_16L	AACTCCCCCRGGCAGCAT
HCV-3a_16R	CGCRTCTTGGGGAGCAGT
HCV-3a_17L	CGCTACTGACGCYCTCATGA
HCV-3a_17R	GGGCATCTATGTGAGTYAGTCC
HCV-3a_18L	CTGCTGAGACCACAGTCAGACT
HCV-3a_18R	CCAAGCAACACCCAGGTACT
HCV-3a_19L	AACACTACATGGACCCACGC
HCV-3a_19R	GGGCTCAATGACAGCTTGTTG
HCV-3a_20L	CAAGCTCAGGYAATAGCCCACC
HCV-3a_20R	CTGCCAGGATGTCAAGCAAGAC
HCV-3a_21L	TGGCGTCTCTTATGGCGTTC
HCV-3a_21R	CGAATGCGATGAGCCTGTTCAT

HCV-3a_22L	TCAACCTRTTGCCCGCCATA
HCV-3a_22R	CCTTTTGACAGGAAATRAAGGGCA
HCV-3a_23L	ACCATCTGGGACTGGGTTTG
HCV-3a_23R	ACTTGGCACGGACACTTGAG
HCV-3a_24L	AACTACACTCGCGCACTATGG
HCV-3a_24R	GGCCTTCARCGACGGAGC
HCV-3a_25L	TGACCTCGATGTTGAGAGACCC
HCV-3a_25R	TTTCCAGCGRTCCAACAGTG
HCV-3a_26L	GCTGCAGAGTGTTTCAAGAAACC
HCV-3a_26R	AGTCGCAACTCAAGTCYGGATC
HCV-3a_27L	ACCACCAACTGTCCATGGATG
HCV-3a_27R	AGTTTCTCCTCCTCAGCACTACA
HCV-3a_28L	AGAGCGTGGTCTGCTGCTCT
HCV-3a_28L_new	AGAGYGTRGTCTGCTGCTCT
HCV-3a_28R	TTCCAGCAAGTCCTCCCAGA
HCV-3a_29L	AGTTCGGGTATAGTGCGAAGGA
HCV-3a_29R	TCTTCCACCCTGATGTCCTGT
HCV-3a_30L	GGGTGCGTGTCTGTGAGAA
HCV-3a_30R	CCTTGATGTAACAAGTGATTGTRTTGCC
HCV-3a_31L	ACGGAGCGGCTTTACTGC
HCV-3a_31R	GTGGCATCACGGGTGAGGTA
HCV-3a_32L	GATGATCTRGTCGTGGTGGC
HCV-3a_32R	AGAGTAAGTGGCCCCGTACAT
HCV-3a_33L	GGGTGCGCATGGTRATGATGA
HCV-3a_33R	CCAACCGTAAACCAGCTGGA
HCV-3a_34L	ACCAGCAATCATTGAAAGACTCCA
HCV-3a_34R	AGCTGGCAGGAGAAAGATGC

1.4 Genotyp 4

Die Oligonukleotide für den Genotyp 4 sind insbesondere auf die Subtypen 4a und 4d zugeschnitten. Für den Primer HCV-4_5_LEFT gibt es zwei unterschiedliche Versionen für jeden der beiden Subtypen. Beide Primer wurden im gleichen Primerpool eingesetzt, sodass eine Tiling-PCR für die Subtypen 4a und 4d gemeinsam durchgeführt werden konnte.

Tabelle Anhang 5: Oligonukleotide des Genotyps 4
Dargestellt sind Oligonukleotide in einer Sequenz von 5‘ → 3‘ des Genotyps 4. Der Primer HCV-4_5_LEFT liegt in einer unterschiedlichen Version für den Genotyp 4a und 4d vor.

Name	Sequenz
HCV-4_1_LEFT	CGGAAATTTGGGCGTGCC
HCV-4_1_RIGHT	YGCCCACCCRCARCCCTC
HCV-4_2_LEFT	AGCACGAATCCTAAACCTCAAAGA
HCV-4_2_RIGHT	CACGAGAGAAGTGCCARGAGG
HCV-4_3_LEFT	GCCGACCTCATGGGATACATC
HCV-4_3_RIGHT	GCTCCARTTCATCATCATGTCCCA
HCV-4_4_LEFT	GTCATGTGGATYTGATGGTGGGYGCYGC
HCV-4_4_RIGHT	ACCCCWGCAAAGAGGAAYAGGACCA
HCV-4_5_LEFT_4a	GGAGTGGCYTAYTTCAGCATGCA
HCV-4_5_LEFT_4d	GGCRTWGCGTACTTCAGCATGCA
HCV-4_5_RIGHT	CAGTACACGGGGCCRCACAC
HCV-4_6_LEFT	GGRAGCTGGCAYATCAACAGRAC
HCV-4_6_RIGHT	GAACCCRGTGCYGTTCATCCA
HCV-4_7_LEFT	CCTGTCGTGGTCGGRACMAC
HCV-4_7_RIGHT	TGCCACKGYGTGGTGGAAAGGAGCA
HCV-4_8_LEFT	GGASGCRGCATGCAACTGGAC
HCV-4_8_RIGHT	CAGCATTGATGKTRATCAGGTTGG
HCV-4_9_LEFT	TCTGTGCGTGCCTNTGGATGAT
HCV-4_9_RIGHT	ATGCRTAAGCCCTYTCGGGCA
HCV-4_10_LEFT	CGCNTGCGGGATGTGGCC
HCV-4_10_RIGHT	AGGCCCTCRGCRGCCCA

HCV-4_11_LEFT	CGTCCCTTACTTYGTGAGGGC
HCV-4_11_RIGHT	GGTGTCYCTGCCRGTGAGGCT
HCV-4_12_LEFT	CACYGCRTACGCGCAGCAGAC
HCV-4_12_RIGHT	CATGGGGCACAGCAGCGGMCC
HCV-4_13_LEFT	CARATGTACACCAATGTYGACCAAGA
HCV-4_13_RIGHT	CCYGTGGTGATGGTCCTGAC
HCV-4_14_LEFT	ACRGGAAGYGGCAAGAGCAC
HCV-4_14_RIGHT	GTCGGCARGGCGACYTCCTC
HCV-4_15_LEFT	CCAAGCGGAGACYGCTGGAG
HCV-4_15_RIGHT	GTNGTCTCTATGGAGAAGGTGGGGTC
HCV-4_16_LEFT	GACTTTGACTCAGTGATAGACTGCA
HCV-4_16_RIGHT	GCAYTTCCACATGGTGTCCCA
HCV-4_17_LEFT	CTKGARTTCTGGGAGAGCGTCTT
HCV-4_17_RIGHT	CCGACRATCACCACGCTGCC
HCV-4_18_LEFT	CATCATGGCYTGCATGTCMGC
HCV-4_18_RIGHT	AAWAGGAGGGTTTGTTGRGTGGT
HCV-4_19_LEFT	GCGAARCAYATGTGGAACTTCATCAG
HCV-4_19_RIGHT	GCAATSAGRCGGTTCATCCACTG
HCV-4_20_LEFT	GTCACYTTYAAGATCATGAGCGGCGA
HCV-4_20_RIGHT	GAGCCRTTYTTGAYGTGGCC
HCV-4_21_LEFT	GTRCTGAGTGACTTCAARACGTGGCT
HCV-4_21_RIGHT	ACYTGGCAGGGRCACTTGATGTT
HCV-4_22_LEFT	GAGGARTACGTGGAGGTTCGCAG
HCV-4_22_RIGHT	TCAGCCACMAGTGGCTCRAAAGAGTC
HCV-4_23_LEFT	ACATCYATGCTGACAGAYCCATC
HCV-4_23_RIGHT	TCCTRGGRGGAGGAACGGGGG
HCV-4_24_LEFT	ACGTGGAAGCARCMGGACTAC
HCV-4_24_RIGHT	AGTGARTTGCTCAGGGG

HCV-4_25_LEFT	GACTTGACATCAGAYTCTTGGTCCAC
HCV-4_25_RIGHT	GCAAGTCCTYCCACACGGAG
HCV-4_26_LEFT	CSGAGCCATTCCCGCAAGGCC
HCV-4_26_RIGHT	CGGGCTCYARGTCACAACACTG
HCV-4_27_LEFT	ACCCGCTGCTTTGACTCCAC
HCV-4_27_RIGHT	CTCGTCATAGCCTCCGTGAA
HCV-4_28_LEFT	CTRAARGACTGCACCATGCTGGT
HCV-4_28_RIGHT	TADGTGACTCCGTACATRTCGAAGTC
HCV-4_29_LEFT	CCRAAGCCAGGARGCCCTTGA
HCV-4_29_RIGHT	GTGATAAATGTCYCCCCCGCC
HCV-4_30_LEFT	CAYGGATACTCTCCACACGAACTCAA
HCV-4_30_RIGHT	AAGTAGGAGTAGGCACAGGAGTAA

2. Referenzsequenzen

Zur Bildung des Primersets und für mögliche Primerkorrekturen wurden für jeden Genotyp unterschiedliche Referenzisolate verwendet, deren *GenBank Accession Numbers* im Folgenden aufgeführt sind.

2.1 Genotyp 1a

NC004102, AF009606, AF271632, AF290978, AF511948, AF511949, AF511950, AJ278830, AX100563, AX663428, D10749, EF032886, EF407411, EF407412, EF407413, EF407414, EF407415, EF407417, EF407418, EF407419, EF407421, EF407422, EF407423, EF407425, EF407426, EF407427, EF407428, EF407431, EF407432, EF407433, EF407434, EF407435, EF407436, EF407437, EF407438, EF407439, EF407440, EF407441, EF407442, EF407443, EF407444, EF407445, EF407446, EF407447, EF407449, EF407450, EF407451, EF407452, EF407453, EF407454, EF407455, EF407456, EF407457, EF621489, EU155213, EU155214, EU155215, EU155216, EU155233, EU155236, EU155237, EU155238, EU155239, EU155240, EU155241, EU155242, EU155243, EU155244, EU155245, EU155246, EU155247, EU155248, EU155249, EU155250, EU155251, EU155252, EU155265, EU155266, EU155267, EU155268, EU155269, EU155270, EU155271, EU155272, EU155273, EU155274, EU155275, EU155276, EU155277, EU155278, EU155282, EU155283, EU155284, EU155285, EU155286, EU155287, EU155288, EU155289, EU155290, EU155291, EU155292, EU155293, EU155294, EU155295, EU155296, EU155297, EU155298, EU155299, EU155309, EU155310, EU155311, EU155312, EU155313, EU155314, EU155319, EU155320, EU155321, EU155322, EU155323, EU155338, EU155339, EU155340, EU155341, EU155342, EU155343, EU155344, EU155345, EU155346, EU155347, EU155348, EU155349, EU155350, EU155351, EU155352, EU155353, EU155354, EU155355, EU155378, EU155379, EU155380, EU234064, EU250017, M62321, M67463

2.2 Genotyp 1b

AJ000009, EU155356, EU155357, EU155358, EU155359, EU155360, EU155361, EU155362, EU155363, EU155364, EU155365, EU155366, EU155367, EU155368, EU155369, EU155370, EU155371, EU155372, EU155373, EU155374, EU155375, EU155376, EU155377, AY587016, D10934, AY460204, L02836, EU155381, EU155382, AJ238799, AJ132997, U45476, AJ238800, M58335, D90208, D50480, D50481, D50484, AF139594, D63857, AB191333, D14484, X61596,

D11355, AF165045, AF165046, AF165064, AF165048, AF165050, AF165052, AF165054, AF165056, AF165058, AF165060, AF165062, D89872, U01214, AF176573, AY587844, AF483269, U89019, M84754, EU155324, EU155325, EU155326, EU155327, EU234061, EU155328, EU155329, EU155330, EU155331, EU155332, EU155333, EU155334, EU234062, EU155335, EU155336, EU155337, EU155217, EU155218, EU155219, EU155220, EU155221, EU155222, EU155223, EU155224, EU155225, EU155226, EU155227, EU155228, EU155229, EU155230, EU155231, EU155232, EU155300, EU155301, EU155302, EU155303, EU155304, EU155305, EU155306, EU155307, EU155308, EU155253, EU155254, EU155255, EU155256, EU155257, EU155258, EU155259, EU155260, EU155261, EU155262, EU155263, EU155264, EU155315, EU155316, EU155317, EU155318, EU155279, EU155280, EU155234, EU155235, EU155281, EF032892, EF407459, EF407481, EF407467, EF407495, EF407502, EF407491, EF407475, EF407476, EF407493, EF407478, EF407480, EF407468, EF407498, EF407469, EF407486, EF407472, EF407479, EF407483, EF407470, EF407497, EF407461, EF407471, EF407463, EF407484, EF407503, EF407465, EF407462, EF407482, EF407487, EF407490, EF407499, EF407492, EF407485, EF407501, EF407494, EF407460, EF407488, EF407477, EF407464, EF407504, AB016785, AX739971, CQ819761, D89815, DD462968, DD495793, AF054250, AY045702, AF333324, D85516, AF356827, AB049087, AB049088, AB049089, AB049090, AB049091, AB049092, AB049093, AB049094, AB049095, AB049096, AB049097, AB049098, AB049099, AB049100, AB049101, D45172, D30613, M96362, E03766, E04420, E05027, E06261, E06457, E07579, E08399, E08461, E09631, E10035, AB080299, AF207752, AF207753, AF207754, AF207755, AF207756, AF207757, AF207758, AF207759, AF207760, AF207761, AF207762, AF207763, AF207764, AF207765, AF207766, AF207767, AF207768, AF207769, AF207770, AF207771, AF207772, AF207773, AF207774, AF208024

2.3 Genotyp 3a

X76918, AF046866, AY956467, D17763, D28917, DQ430819, DQ430820, DQ437509, GQ275355, GQ356200, GQ356201, GQ356202, GQ356203, GQ356204, GQ356205, GQ356206, GQ356207, GQ356208, GQ356209, GQ356210, GQ356211, GQ356212, GQ356213, GQ356214, GQ356215, GQ356217, GU814263, HQ639941, HQ639942, HQ738645, HQ912953, JF509175, JF509176, JF509177, JN714194, JQ717254, JQ717255, JQ717256, JQ717257, JQ717258, JQ717259, JQ717260, NC_009824, GU294484

2.4 Genotyp 4

JX227964, GU814265, NC_009825, Y11604, DQ516084, AB795432, DQ418782, DQ418783, DQ418784, DQ418788, DQ418787, DQ418789, FJ462435, FJ025854, FJ025855, FJ025856, FJ462436, EU392172, FJ462437, DQ516083, DQ418786, EU392169, EU392170, EU392174, EU392175, FJ462432, JX227963, JX227971, EU392171, EU392173, FJ462438, FJ839870, JX227957, JX227958, FJ462433, JX227961, JX227972, FJ462441, JX227970, FJ462440, JX227977, JX227978, JX227979, FJ462431, FJ462434, FJ462439, JX227962, JX227976, FJ839869, JX227959, JX227960, DQ418789, DQ418788, DQ418787, DQ418786, DQ418784, DQ418783, DQ418782, NC_009825, JX227964, DQ516084, DQ516083

3. DNA-Quantifizierung mittels Berthold-Luminometer

Da einzelne Fragmente der Tiling-PCR in der Gel-Elektrophorese nicht nachgewiesen werden können und ein Funktionieren der PCR bereits experimentell bestätigt wurde, wurde im weiteren Verlauf der Erfolg des PCR-Versuchs durch eine DNA-Quantifizierung nachgewiesen. Hierzu wurde eine DNA-Quantifizierung mittels Lumineszenz Mikroplatten Reader von Berthold durchgeführt (Vgl. 3.3.5), was beispielhaft an einem Versuch mit sieben Genotyp 1a Proben gezeigt werden soll. In der Tiling-PCR wurden für jede Probe drei Pools gebildet, deren DNA-Gehalt gleichzeitig mit einer Standardreihe durch das Luminometer gemessen wurde. Dabei wurden Rohdaten bestimmt, die im oberen Abschnitt der folgenden Tabelle Anhang 6 dargestellt sind. Nach Umrechnung der Rohdaten mit Hilfe der Standardkurve konnten folgende DNA-Konzentrationen in den einzelnen Pools gemessen werden, die im unteren Abschnitt der Tabelle Anhang 6 dargestellt sind. Die PCR wurde als erfolgreich angesehen, wenn mehr als 20 ng DNA in einem Pool enthalten war. In der letzten Zeile sind die Negativ-Kontrollen der drei Pools verzeichnet, die erwartungsgemäß kaum DNA enthielten.

Tabelle Anhang 6: Rohdaten, Standardreihe und Ergebnis einer DNA-Quantifizierung
Dargestellt ist die Auswertung einer DNA-Quantifizierung mittels Berthold-Luminometer. Im oberen Abschnitt sind die gemessenen Rohdaten aufgelistet, die mit Hilfe der Standardkurve in DNA-Konzentrationen umgerechnet werden konnten. Diese Ergebnisse sind im unteren Abschnitt der Tabelle dargestellt.

Standard	Pool 1	Pool 2	Pool 3	Standard Kurve		
2441	129333	160396	90018	ng	Value	calculated ng
19540	69748	81705	76916	0	244	0,074
37628	160165	135537	125541	0,5	1954	0,5904
67518	145558	145104	121572	1	3763	1,137
149443	168125	153121	143640	2	6752	2,040
204863	137955	123969	112508	4	14944	4,5151
295815	137402	122392	86907	6	20486	6,1896
297066	18042	10524	14245	8	29582	8,9375
				10	29707	8,9753
				Steigung:	3309,8	

Standard	Pool 1	Pool 2	Pool 3
	39,1	48,5	27,2
	21,1	24,7	23,2
	48,4	41,0	37,9
	44,0	43,8	36,7
	50,8	46,3	43,4
	41,7	37,5	34,0
	41,5	37,0	26,3
Neg. Kontrolle	5,5	3,2	4,3

11 Danksagung

Die vorliegende Dissertation wurde vom Oktober 2021 bis zum Frühjahr 2025 im Institut der Virologie der Uniklinik Düsseldorf sowie der Heinrich-Heine-Universität unter der Leitung von Herrn Dr. Jörg Timm angefertigt. Mein besonderer Dank gilt den folgenden Personen, die mich auf meinem Weg zur Verwirklichung dieser Dissertation unterstützt haben:

Meinem Doktorvater Herrn **Prof. Dr. med. Jörg Timm**, der mir die Möglichkeit gegeben hat, diese Dissertationsarbeit in seinem Institut zu schreiben und mein Thema mit Inspiration und Anregungen förderte. Die Nutzung seiner Vorarbeiten sowie seine fachliche Expertise und das konstruktive Feedback haben mir sehr geholfen, in meiner Arbeit erfolgreich voranzukommen.

Meinem Betreuer **Dr. Andreas Walker** für die hervorragende Betreuung sowie die Ausrichtung und thematische Eingrenzung meiner Arbeit. Durch seine engagierte Erklär-, Diskussions- und Hilfsbereitschaft habe ich mich von Beginn an sehr wohlgefühlt, was ausschlaggebend für den Erfolg dieser Dissertationsarbeit war. Auch herzlichen Dank für die Anleitung im Labor, die geduldigen Erläuterungen, insbesondere der bioinformatischen Themen, und die mühevolle Arbeit des Korrekturlesens.

Meiner Co-Betreuerin **PD Dr. Doreen M. Floß** für ihre zur Verfügung gestellte Zeit und ihre konstruktiven Vorschläge in der Begutachtung meiner Dissertation.

Den Lehrenden des Instituts für Virologie, darunter insbesondere Frau **Dr. Nadine Lübke**, Herr **Prof. Dr. Ortwin Adams**, Frau **Dr. Lisa Müller**, Herr **Dr. Marcel Andrée** und Herr **Prof. Dr. Ingo Drexler**, die mich im Themenblock „Infektion und Abwehr“ für die Inhalte der Virologie begeisterten und mich dadurch zu einer Dissertation in diesem Fachbereich inspirierten. Außerdem stand mir immer die Türe für einen fachlichen Austausch offen, insbesondere während des Wahlfachs zu Viruserkrankungen im Klinikalltag, welches sehr hilfreich für mein virologisches Grundverständnis war.

Danksagung

Der gesamten „AG Timm“ für die freundliche Arbeitsatmosphäre, die vielen wertvollen Anregungen und eine stete Hilfsbereitschaft, die wesentlich zum Gelingen dieser Arbeit beigetragen hat. Insbesondere Frau **Anja Voges** und Frau **Jennifer Camdereli**, die mich durch die Einarbeitung und Unterstützung in Labortätigkeiten entscheidend in meiner Arbeit unterstützt haben.

Frau **Yara Fröhlich**, Frau **Assia Benmoumene** und Frau **Dounia Asskali** und allen weiteren studentischen Hilfskräften für die tatkräftige Unterstützung in der Nanopore-Sequenzierung sowie den anregenden kollegialen Austausch in den Mittagspausen.

Herr **Prof. Dr. Norbert Scherbaum** der Uniklinik Essen für die Nutzungsmöglichkeit der in Essen gelagerten pseudonymisierten Proben der Düsseldorfer IDU-Kohorte.

Meinen Eltern **Sonja Christine Neuroth** und **Dr. Rudolf Neuroth** für die uneingeschränkte und vielseitige Unterstützung während meines Studiums sowie die liebevolle Motivation, die wesentlich zum Gelingen dieser Arbeit beigetragen hat.

Meiner Kindergartenfreundin **Valerie Schramm** für ihre sehr wertvolle Hilfe beim Korrekturlesen und den sprachlichen Feinheiten.

Meinem Verlobten **Patrick Alexander Bahlau**, der mir zu jeder noch so späten Stunde mit Rat und Tat zur Seite stand und mir damit die nötige Zuversicht und Sicherheit für die Erstellung dieser Arbeit gab.

Allen Mitarbeiter*innen des Instituts für Virologie am Universitätsklinikum Düsseldorf für die harmonische Zusammenarbeit und schöne Erfahrung im Team.

Curriculum Vitae

Persönliche Daten

Name	Camilla Sophia Neuroth
Geburtsdatum	15.01.1996
Geburtsort	Frankfurt am Main

Schulausbildung

2001 - 2005	Wilhelm Busch Grundschule, Ratingen-Hösel
2005 -2013	Gymnasium Essen-Werden, Essen

Berufs- und Hochschulausbildung

2013 - 2016	Ausbildung zur Gesundheits- und Krankenpflegerin, St. Elisabeth Akademie, Düsseldorf, St. Marien Krankenhaus, Ratingen
Oktober 2016 - Februar 2017	Gesundheits- und Krankenpflegerin, St. Marien Krankenhaus, Ratingen
Sommersemester 2017	Bachelor of applied science (Chemie, Physik), Bergische Universität Wuppertal
Wintersemester 2017/18	Humanmedizin, Medizinische Fakultät der Universität zu Lübeck
Sommersemester 2018 - Wintersemester 2024/25	Humanmedizin, Medizinische Fakultät der Heinrich-Heine-Universität Düsseldorf
Mai 2019 - September 2023	Studentische Hilfskraft in der Mikrobiologie der Medizinischen Laboratorien Düsseldorf
29.03.2021	1. Abschnitt der Ärztlichen Prüfung
12.10.2023	2. Abschnitt der Ärztlichen Prüfung
13.12.2024	3. Abschnitt der Ärztlichen Prüfung
2021 - 2025	Promotion im Institut für Virologie der Uniklinik Düsseldorf