

Nanopore sequencing in laboratory medicine

-

Novel molecular diagnostic prospects for Familial Mediterranean fever and SARS-CoV-2 infections

Von der Fakultät für Umwelt und Naturwissenschaften
der Brandenburgischen Technischen Universität Cottbus–Senftenberg
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

vorgelegt von

Master of Science (M.Sc.)

Jonas Paul Christian Schmidt

aus Titisee-Neustadt

Gutachter: Honorarprof. Dr. rer. nat. Dirk Roggenbuck

Gutachter: Prof. Dr. med. Meike Burger

Tag der mündlichen Prüfung: 04.05.2023

Abstract

Nanopore sequencing, a third-generation sequencing technique that applies nanometre sized pores to transduce the physical and chemical properties of specific nucleobases into measurable electrical signals, shows attractive advantages over conventional next-generation sequencing techniques. However, primarily due to high sequencing error rates this technique has rarely been used so far in clinical laboratory diagnostics.

In this cumulative dissertation Nanopore sequencing was established and validated in clinical diagnostics using the example of the molecular diagnosis of Familial Mediterranean fever (FMF) and SARS coronavirus-2 (SARS-CoV-2) infections. First, a novel data analysis pipeline for accurate single nucleotide polymorphism (SNP) genotyping using Nanopore sequencing data was developed and validated with the corresponding sequencing protocol against conventional Sanger sequencing using 47 samples of patients with clinical suspicion of FMF. This method comparison showed a perfect agreement between both methods rendering current Nanopore sequencing in principle suitable for SNP genotyping in human genetics.

The bioinformatic analysis of sequencing data is one of the most challenging parts in Nanopore sequencing experiments and complicates the application in a clinical diagnostic setting. Therefore, six different bioinformatic tools for sequence alignment were evaluated regarding their applicability to Nanopore sequencing data. This evaluation revealed a good suitability of all except one of these tools although differences in quality and performance exist. Since Nanopore sequencing showed a robust performance in SNP genotyping, a SARS-CoV-2 whole genome sequencing (WGS) protocol was established to enable onsite viral WGS in a clinical laboratory. This was especially important for viral molecular biological surveillance during the pandemic as shown by analysing viral genetic data over the course of one year. Applying this approach in a clinical research project to investigate host-virus interaction by aggregating for the first time viral genetic data, serological data and clinical data, showed diverse humoral immune responses to SARS-CoV-2, that appear to be influenced by age, obesity and disease severity. Further, even small viral genetic changes may influence the clinical presentation of the associated disease COVID-19.

Additionally, a novel reverse transcriptase (RT)- loop mediated isothermal amplification (LAMP) assay for the detection of SARS-CoV-2 was developed and validated for diagnostic use by method comparison with conventional RT-polymerase chain reaction (PCR).

In summary, by presenting advancements of sequencing and bioinformatic workflows with the focus on an application in clinical diagnostics, the results of this thesis may pave the way for a broader application of Nanopore sequencing in laboratory medicine in the near future.

Zusammenfassung

Die Nanopore-Sequenzierung, eine Nukleinsäure-Sequenzieretechnologie der dritten Generation, bei der Poren im Nanometerbereich verwendet werden, um die physikalischen und chemischen Eigenschaften einzelner Nukleobasen in elektrisch messbare Signale umzuwandeln, weist einige Vorteile im Vergleich zu konventionellen Next-Generation Sequenzierverfahren auf. Dennoch wird die Technik im Wesentlichen aufgrund einer hohen Fehlerrate bisher kaum in der klinischen Diagnostik verwendet.

In der vorliegenden kumulativen Dissertation wurde die Nanopore-Sequenzierung am Beispiel der molekularen Diagnostik von familiärem Mittelmeerfieber (FMF) und SARS Coronavirus-2 (SARS-CoV-2) Infektionen für die klinische Diagnostik etabliert und validiert. Hierzu wurde im ersten Schritt eine neue Datenanalyse Pipeline für die zuverlässige Bestimmung von Einzelbasenaustauschen (SNPs) unter Verwendung von Nanopore-Sequenzierdaten entwickelt und mit dem dazugehörigen Sequenzierprotokoll durch einen Methodenvergleich mit konventioneller Sanger-Sequenzierung unter Verwendung von 47 klinischen Proben von FMF-Patienten validiert. Dabei konnte eine perfekte Übereinstimmung zwischen beiden Methoden gezeigt werden, was die Anwendbarkeit der Nanopore-Sequenzieretechnologie für SNP-Genotypisierung in der Humangenetik demonstriert.

Die bioinformatische Auswertung der Daten stellt eine wesentliche Herausforderung bei der Durchführung von Nanopore-Sequenzierexperimenten dar und erschwert damit eine Anwendung in der klinischen Diagnostik. Um dies zukünftig zu vereinfachen, wurden sechs verschiedene Sequenz-Alignment Programme hinsichtlich Ihrer Anwendbarkeit auf Nanopore-Sequenzierdaten überprüft. Dabei konnte gezeigt werden, dass trotz deutlicher Unterschiede hinsichtlich Qualität und Leistungsfähigkeit alle bis auf eines dieser Programme für die Analyse entsprechender Datensätze verwendet werden können.

Da sich die Nanopore-Sequenzierung als gut geeignet für SNP-Genotypisierung erwies, wurde zusätzlich ein SARS-CoV-2 Vollgenom-Sequenzierprotokoll etabliert, um die Vollgenom-Sequenzierung im klinischen Diagnostiklabor zu ermöglichen. Dies hat sich insbesondere als wichtig für die molekularbiologische Surveillance im Rahmen der Pandemie erwiesen, wie die Auswertung der genetischen Daten über den Zeitraum von einem Jahr zeigte. Zusätzlich wurde das etablierte Vollgenom-Sequenzierprotokoll in einem klinischen Forschungsprojekt verwendet, um Wirt-Virus Interaktionen zu untersuchen, indem zum ersten Mal virale genetische Daten, serologische Daten und klinische Daten aggregiert ausgewertet wurden. Dabei konnte eine breit gefächerte humorale Immunantwort auf das Virus beobachtet werden, die im Wesentlichen von den Faktoren Alter, Übergewicht und dem Schweregrad der Erkrankung beeinflusst

wurde. Außerdem konnte dargestellt werden, dass vermutlich schon kleine Veränderungen im viralen Genom das klinische Bild der assoziierten Erkrankung COVID-19 beeinflussen.

Abschließend wurde im Rahmen der Dissertation eine neue reverse Transkriptase (RT)- Loop Mediated Isothermal Amplification (LAMP) Methode zum Nachweis von SARS-CoV-2 entwickelt und für die diagnostische Anwendung durch einen umfassenden Methodenvergleich mit konventioneller RT-Polymerasekettenreaktion (PCR) validiert.

Zusammenfassend können die Ergebnisse dieser Arbeit durch Weiterentwicklung und Validierung von Sequenzierprotokollen und bioinformatischen Ansätzen zur Datenauswertung mit Fokus auf eine Anwendung in der klinischen Diagnostik den Grundstein für einen zukünftig breiteren Einsatz der Nanopore-Sequenzierertechnologie im Bereich der Labormedizin legen.

Content

List of Abbreviations.....	VI
List of Figures	VIII
1. Introduction	1
1.1 Nanopore sequencing	1
1.1.1 Biotechnological aspects of Nanopore sequencing	1
1.1.2 Bioinformatic data analysis	4
1.1.3 Application examples in clinical diagnostics and research	6
1.2 Loop-mediated isothermal amplification (LAMP).....	6
1.2.1 Application of LAMP in clinical diagnostics.....	7
1.2.2 Special features of LAMP	7
1.3 Familial Mediterranean fever	8
1.3.1 Genetic background.....	8
1.3.2 Human genetic diagnostics.....	8
1.4 SARS-CoV-2 and Coronavirus disease 2019 (COVID-19).....	9
1.4.1 Clinical presentation and reaction of the immune system.....	10
1.4.2 Laboratory diagnostics	10
2. Objective	12
3. Results	13
3.1 Establishment and evaluation of Nanopore sequencing for application in clinical laboratory diagnostics	13
3.1.1 Genotyping of familial Mediterranean fever gene (MEFV)- single nucleotide polymorphism - Comparison of Nanopore with conventional Sanger sequencing (Schmidt et al. 2022).....	13
3.1.2 Comparative analysis of alignment tools for application on Nanopore sequencing data (Becht, Schmidt et al. 2021)	16
3.2 Application of Nanopore sequencing for SARS-CoV-2 whole genome sequencing.....	18
3.2.1 Serological and viral genetic features of patients with COVID-19 in a selected German patient cohort—correlation with disease characteristics (Schmidt et al. 2021) .	19
3.2.2 Epidemiological surveillance of SARS-CoV-2 by whole genome sequencing	24
3.3 Development of a RT-LAMP assay for SARS-CoV-2 detection	25
3.3.1 A semi-automated, isolation-free, high-throughput SARS-CoV-2 reverse transcriptase (RT) loop-mediated isothermal amplification (LAMP) test (Schmidt et al. 2021).....	25
4. Discussion	30
4.1 Establishment and evaluation of Nanopore sequencing for application in clinical laboratory diagnostics	30
4.2 Application of Nanopore sequencing for SARS-CoV-2 whole genome sequencing.....	33
4.3 Development of a RT-LAMP assay for SARS-CoV-2 detection	37

Content

5. Summary	39
References	41
Documentation / List of Publications	48
Acknowledgements	51
Eidesstattliche Erklärung.....	52
Appendix	53

List of Abbreviations

ACE2	Angiotensin Converting Enzyme 2
BAM	binary alignment map
BAL	bronchoalveolar lavage
BMI	body mass index
CAPS	Cryopyrin-associated periodic syndrome
cDNA	complementary DNA
COVID-19	Coronavirus disease 2019
CPU	central processing unit
CT	computer tomography
Ct	cycle threshold
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double stranded DNA
ELISA	enzyme-linked immunosorbent assay
FMF	Familial Mediterranean fever
FN	false negative
FP	false positive
GLM	generalized linear model
IL-1	Interleukin-1
LAMP	loop-mediated isothermal amplification
LoD	limit of detection
MERS-CoV	Middle East Respiratory Syndrome coronavirus
MKD	Mevalonate kinase deficiency
NAT	nucleic acid amplification technique
NGS	next generation sequencing
ONT	Oxford Nanopore Technologies
ORF	open reading frame
PCR	polymerase chain reaction
POC	point-of-care
PON	point-of-need
RAM	random-access memory
RBD	receptor binding domain
RNA	ribonucleic acid
RSV	Respiratory Syncytial Virus
RT-LAMP	reverse transcription loop-mediated isothermal amplification
RT-PCR	reverse transcription polymerase chain reaction
SARS-CoV	Severe acute respiratory syndrome-coronavirus
SARS-CoV-2	Severe acute respiratory syndrome-coronavirus type 2
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
TP	true positive
TRAPS	Tumour necrosis factor receptor-associated periodic syndrome

List of Abbreviations

Tt	threshold time
UTR	untranslated region
VOC	variant of concern
WGS	whole genome sequencing
WHO	World Health Organisation

List of Figures

Figure 1: Schematic representation of the basic principle of Nanopore sequencing.	2
Figure 2: Electrical raw signal of a single read which was generated using a MinION sequencing device.	5
Figure 3: Overview of the gene structure of <i>MEFV</i> and genetic variants which were identified in this gene in 47 clinical samples.	14
Figure 4: Data analysis pipeline which was established specifically to analyze the Nanopore sequencing data and identify SNPs within <i>MEFV</i>	15
Figure 5: Comparison of six different sequence alignment tools regarding speed and memory consumption.	17
Figure 6: Distribution of A) match rate, B) mismatch rate and C) error rate which were calculated for the application of the different alignment tools on each data set.	18
Figure 7: Viral genetic variants which were identified by Nanopore sequencing of 55 RNA isolates.	20
Figure 8: Phylogenetic analysis of 55 SARS-CoV-2 consensus sequences.	21
Figure 9: Anti-SARS-CoV-2 antibody levels in 49 patients with COVID-19.	22
Figure 10: Distribution of SARS-CoV-2 variants of concern (VOC) in the area of the clinical diagnostic laboratory MVZ Laborärzte Singen.	24
Figure 11: Amplification curves of the SARS-CoV-2 RT-LAMP assay.	26
Figure 12: Conventional agarose gel (1.5%) electrophoresis of SARS-CoV-2 RT-LAMP reaction products.	27
Figure 13: Correlation between SARS-CoV-2 RT-LAMP Tt values and RT-PCR Ct values.	28

1. Introduction

1.1 Nanopore sequencing

Over the last few years Nanopore sequencing, which is considered a third-generation sequencing technology, has become increasingly important for sequencing of nucleic acids in research and diagnostic laboratories^{1,2}. So far, the technique has been only commercialized by the British company Oxford Nanopore Technologies (ONT) and was introduced into the market in 2014³. Until today there have been many technological improvements and the technique is still under active development to continuously increase sequencing performance regarding throughput and accuracy². Compared to conventional next generation sequencing (NGS) techniques, like Illumina sequencing or Ion Torrent sequencing, important advantages of the third-generation include a sequencing process and data acquisition/analysis in real-time as well as the possibility of sequencing native desoxyribonucleic acids (DNA) or ribonucleic acids (RNA) without the need for amplification by polymerase chain reaction (PCR)^{1,2,4}. A comprehensive review by Schmidt et al. (2020) summarises the technical aspects of current Nanopore sequencing and provides an overview of possible applications in clinical research and diagnostics².

1.1.1 Biotechnological aspects of Nanopore sequencing

As outlined by Schmidt et al. Nanopore sequencing is based on nanometre-sized pores which are used to decipher an unknown nucleotide sequence by transducing the physical and chemical properties of specific nucleobases into measurable ionic currents^{2,5,6}. These pores are embedded into a membrane which separates a reservoir containing an electrolyte solution into two compartments (cis and trans chamber), each containing one electrode^{2,5,6}. By applying a voltage bias to the electrodes, the electrolytes in solution are electrophoretically driven through the pore which generates a measurable ionic current signal^{2,5}. Further, any negatively-charged DNA or RNA molecules that are present in the cis chamber also pass through the pore^{2,5}. While passing through, the single stranded molecules partially block the pore which results in a measurable interruption of the current signal (Figure 1)^{2,5}. By using complex classification algorithms, the amplitude and duration of the transient current blockades can be used to identify the unknown nucleotide sequence since the signal shape is directly related to a specific group of nucleobases^{2,5-7}.

Since it is not yet possible to sense single nucleobases, currently available Nanopore sequencing platforms detect specific groups of six nucleobases for DNA sequencing and groups of five for RNA sequencing^{2,6}. In order to achieve controlled and standardized signal acquisition, in current Nanopore sequencing devices, the translocation speed of the single stranded target molecules is controlled by applying helicase motor proteins (Figure 1)^{2,6}. These proteins work like a ratchet so that each nucleobase is held in the pore for a few milliseconds^{2,6}. For DNA sequencing, combined with a special sequencing adapter, the motor proteins are ligated during library preparation to the terminal nucleotides of the double stranded DNA (dsDNA) molecules which are to be sequenced^{2,6}. During the sequencing process, the adapter is translocated through the pore in front of the dsDNA to which it is attached^{2,6}. This brings the helicase into contact with the pore, where it is activated and unzips the dsDNA into two single strands of which one is translocated through the pore in a stepwise fashion^{2,6}. By using this mechanism, the translocation speed is regulated at a continuous rate of approximately 450 bases per second^{2,6,8}.

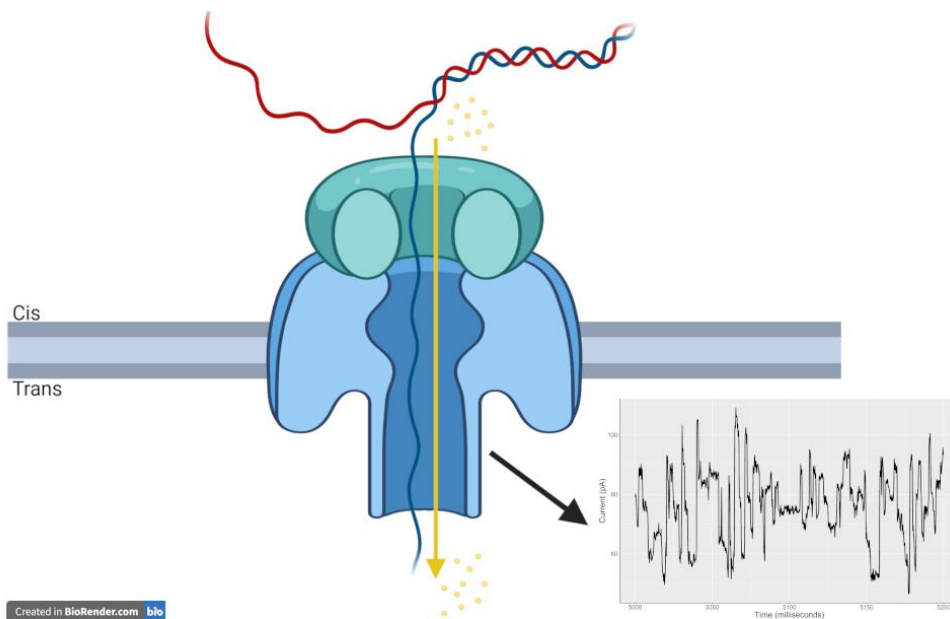


Figure 1: Schematic representation of the basic principle of Nanopore sequencing. A nanopore (blue), embedded into a membrane, is partially blocked by a DNA strand passing through with the support of a helicase motor protein (green). The resulting interruption of the ion current (yellow) generates a measurable electrical signal. Since the signal shape is directly related to specific groups of nucleobases, it can be used to identify the unknown nucleotide sequence by using complex classification algorithms (Created with BioRender)².

Commercial available Nanopore sequencing devices apply sensor arrays which contain thousands of nanopores^{2,6}. It is therefore possible to sequence many different DNA strands in parallel which greatly increases sequencing throughput and the amount of data generated^{2,6}.

In general, as summarised in the review by Schmidt et al., two different types of nanopores have been used so far for Nanopore sequencing^{2,5}. They include biological as well as synthetic solid-state pores^{2,5,6,9}. Up to now, sequencing of nucleic acids has mainly been performed by using biological pores which are transmembrane protein channels inserted into a suitable cellular or artificial substrate^{2,5,6}. These proteins can be isolated from bacteria or bacteriophages and well-studied examples include α -hemolysin (*Staphylococcus aureus*), MspA (*Mycobacterium smegmatis*) and a pore which is isolated from the bacteriophage phi29^{2,5,6,9-11}. Although biological pores are characterised by a highly-reproducible size and structure as well as the potential to be easily modified using biotechnological engineering techniques, they require special environmental conditions regarding temperature, electrolyte concentration and pH which makes them more fragile^{2,5,6}. Therefore, research effort has been directed towards developing different kinds of more stable synthetic nanopores by using semiconductor materials including silicon nitride, silicon dioxide, aluminium oxide, boron nitride and graphene^{2,9}. However, since the field of solid-state nanopore based sequencing is still in an early phase, this approach is currently not widely used for sequencing of nucleic acids^{2,9}. This also applies to the idea of combining biological pores with solid-state components to form hybrid pores and use the advantages of both techniques^{2,5,9}.

Among other relevant advantages (Table 1) compared to conventional NGS techniques, an important feature of Nanopore sequencing is its ability to create long reads up to multiple megabases which makes it a so called “long-read” sequencing technique^{2,6}. However, this requires gentle nucleic acid isolation techniques as well as a comprehensive quality control of the isolate to ensure sufficient sample quality^{2,6}. Basic quality measures include fragment size analysis, absorbance spectrometry as well as fluorometric quantification^{2,6}. Finally, there are different library preparation kits available from ONT which can be used to prepare the purified samples for sequencing^{2,6}. All of them have in common the fact that the sequencing adapter, which contains the motor protein, is attached to the target molecules during the preparation procedure. Further, some of them offer the possibility of labelling different samples during the library preparation procedure by using unique molecular identifiers^{2,6}. These identifiers are normally short oligonucleotides, which can be attached to the samples during the library preparation procedure and do not appear naturally in the sample. Such multiplexing approaches can be used to greatly increase efficiency by sequencing multiple samples in parallel in a single run^{2,6}. After

the sequencing run containing a pool of labelled samples is finished, based on the molecular labelling it is possible to assign the individual reads back to the corresponding sample².

A major disadvantage of Nanopore sequencing is its low single read accuracy². Although this quality measure is influenced by different parameters like sequencing instrument, sequencing protocol and sample type, Nanopore sequencing shows a remarkably higher error rate (~6%) compared to other sequencing techniques including PacBio sequencing (~1.5%), Illumina sequencing (~0.5%) and conventional Sanger sequencing (~0.001%)¹²⁻¹⁶.

As outlined by Schmidt et al., together with the large amount of data which is generated during a single sequencing run, the complex error profile makes bioinformatic data analysis challenging and requires a lot of computational power which is another major disadvantage of Nanopore sequencing².

Table 1: Advantages and disadvantages of current Nanopore sequencing. These points are relevant with regard to an application in a clinical diagnostic laboratory².

Advantage	Disadvantage
Long read sequencing ^a	Low single read accuracy
Real-time sequencing	Complex data analysis
Simple library preparation with the possibility of multiplexing	High requirements on computational infrastructure
Native sequencing without amplification	Limited experience in clinical diagnostic application
Low capital costs	
Highly specialized laboratory infrastructure not required	

^a The maximum achievable read length is limited by the sample preparation technique and not by the sequencing principle itself. With the current technical capabilities, it is possible to reach a maximum read length up to 2 megabases (Mb).

1.1.2 Bioinformatic data analysis

Bioinformatic data analysis is a central task for interpreting the data generated during a Nanopore sequencing experiment. Due to the large amount of data and the complex error profile of the generated reads, it is challenging and time consuming². Up to now a large number of freely available algorithms and tools has been developed for the main applications including base calling, demultiplexing, quality control, data handling, read mapping, De novo assembly, and variant discovery^{2,17}. However, since there are hardly any standardized data analysis pipelines available, it is up to the researcher to apply the suitable tools for a specific research question and develop their own pipelines². As outlined in a bioinformatic research article by Becht, Schmidt et al. (2021), comparing different available data analysis tools regarding their

performance is a mandatory step during development of such pipelines in order to get reliable results¹⁸.

A major step, which is included in every data analysis pipeline, is base calling. During this process the raw read signal, which describes the changes in the ionic current signal over time as the target molecules are passed through the pore, is used to identify the unknown DNA or RNA sequence (Figure 2)^{2,6,17}. For this purpose modern machine learning approaches are applied, which is one reason why Nanopore sequencing data analysis requires a lot of computational power^{2,17}. Besides the nanopores and the sequencing chemistry itself, the base calling process is a key factor for raw read accuracy of the resulting nucleic acid sequences. Optimization of the tools, that are used for base calling, has led to a remarkable increase of sequencing accuracy over the last few years.

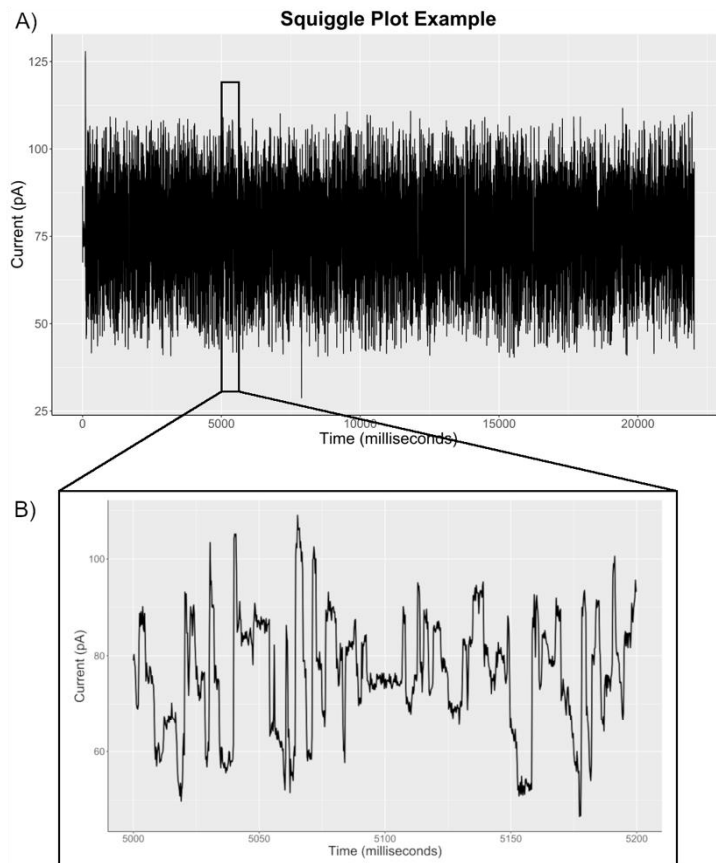


Figure 2: Electrical raw signal of a single read which was generated using a MinION sequencing device. A) Full length signal which is represented in a so called “squiggle plot”. The total read length after base calling was 8685 bases. B) Time interval from 5000 milliseconds to 5200 milliseconds showing the fluctuation of the current due to the DNA strand passing through the pore².

1.1.3 Application examples in clinical diagnostics and research

As summarised by Schmidt et al., since Nanopore sequencing has not been available on the market for long and comprehensive method validation data is not available, it is rarely applied in clinical diagnostics so far². A possible application in clinical diagnostics is further complicated by the complex error profile of Nanopore sequencing data as well as the challenging data analysis². However, the technique shows promising results in a variety of research applications².

Especially in the field of diagnostics of infectious diseases it is used to sequence bacterial and viral genomes^{2,19-21}. Other important applications include the field of cancer research². Here, Nanopore sequencing can be used to identify mutated genes and provide critical diagnostic information^{2,22,23}. Last but not least, there are some research applications in human genetics where Nanopore sequencing is used to sequence the human genome and diagnose heritable diseases^{2,24,25}.

1.2 Loop-mediated isothermal amplification (LAMP)

The amplification of nucleic acids is a fundamental technique in all fields of molecular biological science²⁶. Besides PCR, which is the most widely used approach, many other alternative techniques including loop-mediated isothermal amplification (LAMP) have been developed²⁶. Similar to PCR, a LAMP reaction mix contains deoxynucleotide triphosphates (dNTPs), a polymerase, a fluorescent dye, target specific primers and the DNA template²⁶. In contrast to conventional PCR, LAMP primers consist of a set of four or six different primers which bind to six or eight different regions on the target sequence²⁶. These primer sets contain two outer primers, two inner primers and two loop primers²⁶. The loop primers are optional but they can accelerate the LAMP reaction by factor of two²⁷. Combined with a *Bst* DNA polymerase, which is originally isolated from *Geobacillus stearothermophilus* and has a high strand displacement activity, this primer design can be used to induce stem-loop DNA structures which are the starting point for a self-priming exponential amplification²⁶⁻²⁸. The reaction can be performed at isothermal conditions with an optimum between 60-65 °C and generates long amplicons with cauliflower-like structures²⁷. Detection of the reaction products can be done by multiple different target sequence-independent methods including turbidity measurement, endpoint detection with the naked eye, gel electrophoresis, colorimetric reactions, melting and annealing curve analysis, intercalating fluorescent dyes, bioluminescence and electrochemiluminescence²⁷. Further,

sequence-specific detection is possible through the use of target-specific probes, modified primers as biorecognition elements or sequencing²⁷.

The isothermal reaction condition is an important advantage of LAMP compared to PCR, because it obviates the need for complex thermal cycling²⁷. This makes LAMP combined with simple detection methods especially interesting for applications like point-of-care (POC) or point-of-need (PON) testing, where specialized laboratory equipment is not available²⁷.

1.2.1 Application of LAMP in clinical diagnostics

LAMP has been used in various different studies for the detection of human pathogens including bacteria (e.g. *Mycobacterium tuberculosis*, *Streptococcus pneumoniae*, *Bordetella pertussis*, *Klebsiella pneumoniae*, *Salmonella typhi*, *Campylobacter jejuni*, *Helicobacter pylori*, *Neisseria meningitidis* and *Listeria monocytogenes*), viruses (e.g. adenovirus, varicella zoster, cytomegalovirus), protozoan parasites (e.g. *Plasmodium* spp., *Loa loa* and *Leishmania*) and fungi (e.g. *Candida albicans*, *Cryptococcus neoformans* and *Mucor racemosus*)²⁶.

By combining LAMP with reverse transcriptase enzymes, it is possible to detect RNA in a one-step nucleic acid amplification reaction referred to as reverse transcription loop-mediated isothermal amplification (RT-LAMP)²⁶. The reverse transcriptase is used to make complementary DNA (cDNA) from RNA, which is then further amplified by the *Bst* DNA polymerase in the same reaction mix²⁶. This technique is especially important because it allows very effective detection of RNA viruses²⁶. Published examples include the detection of Dengue virus, influenza viruses, Hepatitis C virus, Ebola virus, respiratory syncytial virus (RSV), Zika virus and Middle East Respiratory Syndrome (MERS) coronavirus²⁶.

1.2.2 Special features of LAMP

Although LAMP is an effective diagnostic tool with high sensitivity and specificity, a robust reaction principle and a simple reaction setup, a major disadvantage for diagnostic use is its high risk of carryover contamination²⁶. Since the LAMP reactions shows a very high amplification efficiency and the reaction products are stable and do not degrade easily, unintended carryover contamination can lead to false positive test results²⁶. To reduce this risk in a diagnostic setting, it is important to take appropriate measures including spatial separation of different laboratory workplaces and immediate discarding of the amplification product without opening reaction tubes²⁶.

Further, compared to conventional PCR the design of LAMP assays is more elaborate due to the complex primer schemes²⁶.

1.3 Familial Mediterranean fever

Familial Mediterranean fever (FMF) is a common autoinflammatory disease which shows a high prevalence among Turkish, Armenian, Jewish and Arabic communities from the eastern Mediterranean region²⁹⁻³¹. In these ethnic groups the prevalence varies between 1:150 and 1:10000³⁰. The disease is clinically diagnosed and is mainly characterised by recurrent fever and serositis (e.g. peritonitis, pleuritis, synovitis) symptoms³⁰⁻³³. In untreated individuals, amyloidosis is a severe complication which can progress to end-stage renal failure³⁰⁻³³. After clinical diagnosis the disease is commonly treated with colchicine^{29,31,34}. In colchicine refractory cases interleukin-1 (IL-1) blockade is suggested^{29,31,34}.

1.3.1 Genetic background

It is considered that FMF is inherited autosomal recessive and is mainly associated with point mutations (single substitutions) in the Mediterranean Fever (*MEFV*) gene³⁰⁻³². The gene is located on the short arm of chromosome 16 in minus strand orientation and is composed of 10 exons^{30,31}. It encodes a protein (pyrin) containing 781 amino acids and is mainly expressed in neutrophils, eosinophils, dendritic cells and fibroblasts²⁹⁻³². Pyrin plays a key role in cellular apoptosis and inflammatory pathways²⁹⁻³². Therefore, it is thought that mutated pyrin causes FMF by inducing an excessive inflammatory response through uncontrolled IL-1 secretion^{29,31,34}.

1.3.2 Human genetic diagnostics

Since FMF is diagnosed based on clinical criteria, genetic testing is performed to support the clinical diagnosis of FMF and screen relatives at risk^{31,32}. According to expert consensus guidelines this can be done by either targeted mutation analysis which only detects the most common *MEFV* mutations or by sequencing of selected exons^{31,35}. However, a minimum diagnostic screen should include clearly pathogenic variants which are frequently identified in patients (Table 2)^{31,35}.

Besides DNA sequencing, which is used in most laboratories for variant identification, targeted approaches can be performed by using PCR based or reverse-hybridisation based assays^{31,35}. These targeted approaches as well as conventional Sanger sequencing suffer from the technological limitation that it is not possible to cover all clinically relevant genetic regions within a single run³¹. Therefore, to overcome this limitation, diagnostic NGS protocols have been developed which allow sequencing of gene panels including not only *MEFV* for FMF diagnosis but also genes which are associated with other periodic fever syndromes encompassing mevalonate kinase deficiency (MKD, gene *MVK*), tumour necrosis factor receptor-associated periodic syndrome (TRAPS, gene *TNFRSF1A*) and cryopyrin-associated periodic syndrome (CAPS, gene *NLRP3*)^{31,35,36}.

Table 2: Recommendations for the screening of sequence variants for the genetic diagnosis of FMF³⁵.

Disease	Gene	Exon	Variants ^a
FMF	<i>MEFV</i>	2	p.E148Q, p.E167D, p.T267I, p.R202Q
FMF	<i>MEFV</i>	3	p.P396S, p.R408Q
FMF	<i>MEFV</i>	5	p.F479L
FMF	<i>MEFV</i>	9	p.I591T
FMF	<i>MEFV</i>	10	p.M680I, p.M694V, p.M694I, p.V726A, p.A744S, p.R761H, p.I692del, p.K695R

^a Amino acid changes with regard to reference sequence NM_000243.2 are shown.

1.4 SARS-CoV-2 and Coronavirus disease 2019 (COVID-19)

Severe acute respiratory syndrome-coronavirus type 2 (SARS-CoV-2) is an enveloped, positive-sense single-stranded RNA virus which has spread pandemically worldwide since its first description at the end of 2019^{37,38}. Unlike endemic “common cold” coronaviruses, SARS-CoV-2 is classified as highly pathogenic and shows similar characteristics to SARS-CoV and MERS-CoV^{37,39}.

The viral genome with a size of ~30 kilobases (kb) consists of at least 14 open-reading frames (ORF)^{37,39-41}. They encode 16 non-structural proteins which form the foundation for virus replication within the host cell as well as nine accessory and four structural proteins, including spike (S), envelop, membrane and nucleocapsid (N) proteins^{37,39,40}. During contact with the

host cell, the S protein is cleaved into two subunits (S1/S2) by proteases^{37,40}. Both subunits are essential for viral entry into the host cell and define tissue tropism as well as viral host range^{37,40,42}.

1.4.1 Clinical presentation and reaction of the immune system

The incubation period after an infection with SARS-CoV-2 is approximately 4-12 days^{37,40,42,43}. Overall the clinical features of the resulting Coronavirus disease 2019 (COVID-19) are diverse and vary in onset and severity^{37,40}. The main symptoms include fever, cough, gastrointestinal illnesses, anosmia and dyspnoea but in severe cases, initially mild symptoms may later exacerbate to a life-threatening systemic inflammation with a cytokine storm syndrome^{37,38,40}. This causes an acute respiratory distress syndrome followed by respiratory failure which are considered the leading causes of death in patients with severe COVID-19^{37,38,40}. In addition to acute symptoms, COVID-19 may also be associated with long-term effects, such as myocardial inflammation^{37,40}.

As outlined in a review article by Schmidt et al. (2021) an infection with SARS-CoV-2 triggers both humoral and cellular immune responses⁴⁴. The viral S and N proteins are most immunogenic and show distinct IgG, IgM, and IgA responses^{37,45}. Besides the humoral response, a cellular immunity mediated by T lymphocytes is also induced⁴⁴. Based on this, various different vaccines applying innovative vaccination techniques were developed in record time during the course of the pandemic⁴⁴. However, so far it is unclear how long the immunity lasts, both after wild virus infection and vaccination⁴⁴.

1.4.2 Laboratory diagnostics

Diagnosis of SARS-CoV-2 infections is mainly performed by using nucleic acid amplification techniques (NAT) due to their high sensitivity and specificity⁴⁶. Besides RT-PCR, which is considered as the gold standard method, isothermal amplification techniques are also applied⁴⁶. Common sample specimens for SARS-CoV-2 testing include nasopharyngeal swabs and oropharyngeal swabs⁴⁶. Also deep respiratory specimens, including bronchoalveolar lavage (BAL) and endotracheal aspirate, are very well suited for diagnostics⁴⁶. Due to the limited capacity of NAT, rapid antigen testing has become widely available and has developed into a diagnostic tool which is used to test symptomatic and asymptomatic individuals in the general population.

However, due to the reduced sensitivity of such testing regimens it should only be used jointly with RT-PCR⁴⁶. Antibody testing should not be used for the diagnosis of an acute infection but to detect a prior infection, proof of vaccination success and for retrospective assessment of the extent of outbreaks⁴⁶.

Besides direct or indirect detection of SARS-CoV-2 by means of laboratory analysis, chest computer tomography (CT) scans are also frequently used to support the clinical diagnosis of COVID-19⁴⁶.

During the course of the pandemic molecular genetic surveillance has been shown to be a valuable tool for public health authorities to track the viral spread as well as the appearance of new viral variants over time. To reduce further viral spread individuals who tested positive are quarantined and contact tracing is performed. Further, representative proportions of SARS-CoV-2 positive samples are genotyped at regular time intervals to identify viral variants by using variant specific PCR assays or whole genome sequencing.

2. Objective

Because of the high demands regarding diagnostic accuracy and performance, it is mandatory to comprehensively validate new technologies before they are applied in routine laboratory diagnostics. Since Nanopore sequencing is a revolutionary new sequencing technique, which offers various different advantages over conventional techniques that make it attractive for clinical diagnostics, the main objective of this work is to establish and validate the technique itself as well as to optimize the bioinformatic data analysis for an application in a clinical diagnostic laboratory.

Due to the comparatively high sequencing error of current Nanopore sequencing, first of all, this requires the development of a novel data analysis pipeline that applies bioinformatic procedures for error correction and produces high quality consensus sequences from raw sequencing data which can be used for reliable identification of variants. Further, to meet the special needs of a clinical diagnostic laboratory, the pipeline should work without requiring extensive computational resources and the results should be available in a reasonable time frame.

The analytical performance of the whole workflow including sequencing and data analysis is assessed by sequencing selected regions of *MEFV* for the genetic diagnosis of FMF in a selected patient cohort with Nanopore sequencing and comparing the results to conventional Sanger sequencing.

Since Nanopore sequencing is also a promising technique for viral whole genome sequencing which is besides SNP genotyping in human genetics another important analytical issue in modern laboratory medicine, a second objective of this work is to support SARS-CoV-2 epidemiological surveillance during the COVID-19 pandemic by enabling on-side viral whole genome sequencing in a clinical diagnostic laboratory without a specialized NGS department. In addition, to study host-virus interactions after SARS-CoV-2 infection in a selected German patient cohort, viral genetic features and the ensued humoral immune response are jointly analysed for the first time using clinical samples.

Finally, to overcome the temporary shortages of laboratory supplies and PCR reagents caused by the pandemic situation a third subsidiary objective is to establish and validate a novel diagnostic SARS-CoV-2 assay that is based on LAMP and can be processed semi-automatically without prior RNA isolation.

3. Results

3.1 Establishment and evaluation of Nanopore sequencing for application in clinical laboratory diagnostics

Independent of its intended use, an analytical method should offer a robust performance with great reproducibility. However, the requirements vary slightly between clinical diagnostic laboratories and research laboratories². Besides analytical performance, sample throughput, possible degree of automation, time to result, ease of use as well as capital and consumable costs are important parameters for an application in clinical diagnostics². In addition, a comprehensive validation including method comparison experiments with gold standard methods is of the utmost importance prior to an application in a diagnostic setting.

Since Nanopore sequencing is a third generation sequencing technique which is still under active development and has only been available for a comparatively short time, the number of studies comparing this technique with conventional Sanger sequencing as the established gold standard method for human genetic diagnostics has been limited³¹.

3.1.1 Genotyping of familial Mediterranean fever gene (*MEFV*)- single nucleotide polymorphism - Comparison of Nanopore with conventional Sanger sequencing (Schmidt et al. 2022)

This method orientated publication evaluated the clinical performance of current Nanopore sequencing. The sequencing technique was used in combination with a novel dedicated data analysis pipeline for single nucleotide polymorphism (SNP) genotyping of selected regions of *MEFV* in 47 patients with clinical suspicion of FMF³¹. The results were validated against diagnostic Sanger sequencing as the gold standard method³¹.

The routine diagnostic workflow for FMF includes DNA isolation, PCR amplification of selected targets covering relevant regions of *MEFV* and Sanger sequencing of the resulting eight amplicons (Figure 3B)³¹. After this routine workflow has been completed, the amplicons obtained from the amplification step were pooled per sample and Nanopore sequencing was performed³¹.

3. Results

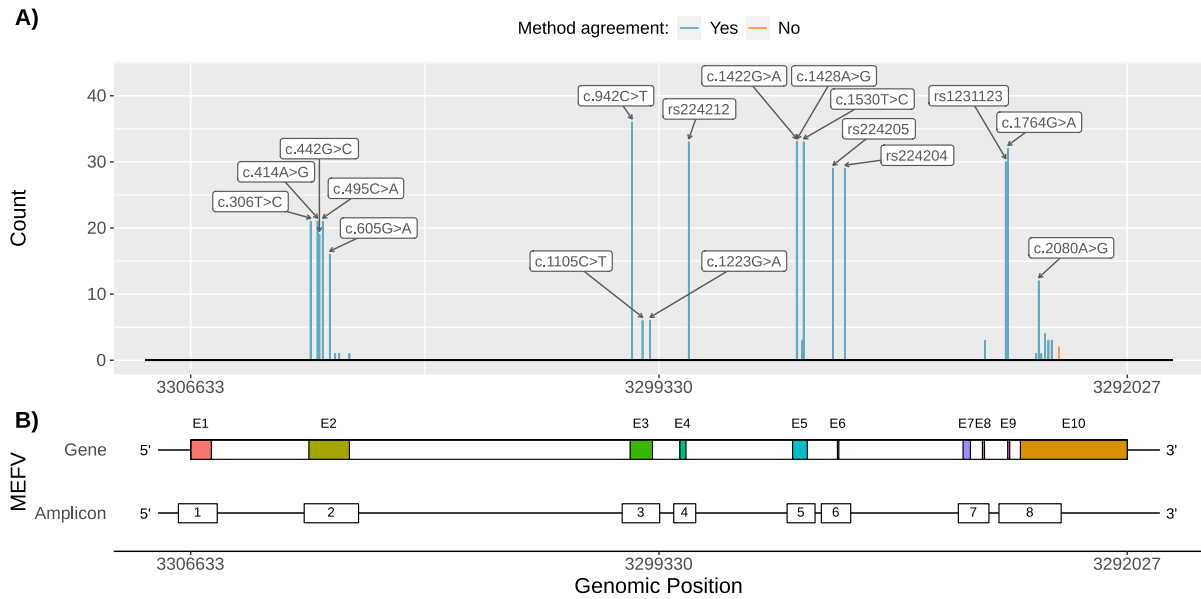


Figure 3: Overview of the gene structure of *MEFV* and genetic variants which were identified in this gene in 47 clinical samples. A) Frequency of the single nucleotide polymorphisms identified by conventional Sanger sequencing and Nanopore sequencing. Variants with a complete agreement between both methods are coloured in blue and differing variants are coloured in orange. cDNA or dbSNP labels are given for the most common variants. B) Structure of *MEFV* and localization of the amplicons which were generated to sequence selected gene regions. Genomic positions on the hg19 reference genome (NC_000016.9) are shown in minus strand orientation³¹. E; exon

Bioinformatic analysis of the Nanopore sequencing data was performed by using a dedicated data analysis pipeline which was established specifically for this purpose and implemented into a shell script for automation purposes (available from GitHub: <https://github.com/j4yo/MEFV-SNP-Genotyping-Pipeline>). The basic sequential steps include base calling and demultiplexing of the raw data, quality control, read filtering, sequence alignment to the reference genome, variant calling, variant filtering and variant annotation (Figure 4)³¹. All results were inspected manually after the automated data analysis pipeline finished³¹.

To perform the method comparison, Nanopore sequencing variant calls were compared to the Sanger sequencing reference for genomic position, nucleotide change, zygosity, amino acid position, and amino acid change³¹. Variants without a complete match and variants which were missed by Nanopore sequencing were classified as false negative (FN) and variants solely identified by Nanopore sequencing as false positive (FP)³¹. Based on this, comparative performance measures including Precision (True positive (TP)/(TP + FP)), Recall (TP/(FN + TP)) and F1-Score ($2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$) were calculated³¹.

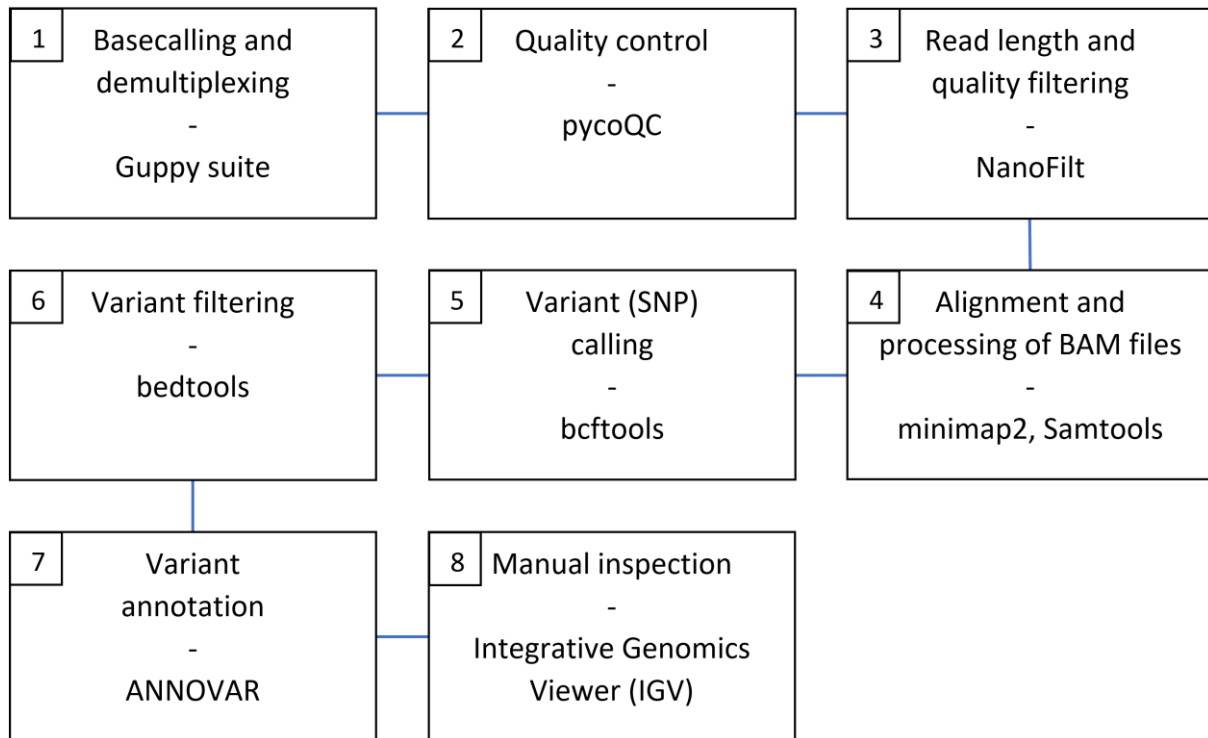


Figure 4: Data analysis pipeline which was established specifically to analyze the Nanopore sequencing data and identify SNPs within *MEFV*. The bioinformatic tools which are used for the individual tasks are shown. Step 1 to 7 were implemented in a shell script for automation purposes³¹. SNP; single nucleotide polymorphism, BAM; binary alignment map

Conventional Sanger sequencing identified 284 heterozygous and 149 homozygous SNPs in the investigated sample collective³¹. These comprised 28 unique variants including the common non-synonymous SNPs p.E148Q, p.R202Q, p.M694V, p.P369S, and p.R408Q (see Table 1 Schmidt et al. 2022)³¹. By using Nanopore sequencing in combination with the dedicated data analysis pipeline it was possible to correctly identify all 433 SNPs which had been confirmed in the sample collective by diagnostic Sanger sequencing before (Figure 3A)³¹. Complete matches regarding genomic position, nucleotide change, zygosity, amino acid position, and amino acid change were achieved³¹. In addition, a transversion from guanine (G) to thymine (T) was identified by Nanopore sequencing in the 3' untranslated region (UTR) in two related patients³¹. The variant was located in a region of the corresponding amplicon which could not be sufficiently resolved by initial Sanger sequencing³¹. By sequencing an additional amplicon which was designed to cover this region of interest of the 3' UTR, it was possible to confirm the transversion in both patients also by Sanger sequencing³¹. Nevertheless, a comprehensive data base research, including ClinVar and dbSNP, did not reveal any further information on the clinical significance of this SNP³¹.

By including the results of the additional Sanger sequencing experiments, a perfect agreement between Nanopore sequencing and Sanger sequencing was shown³¹. However, the initially diverging variant was treated as false positive for the calculation of the performance measures

since it was not identified by initial diagnostic Sanger sequencing³¹. With this assumption in mind, the Nanopore sequencing method achieved a Precision of 0.995, a Recall of 1 and a F1-Score of 0.998³¹.

3.1.2 Comparative analysis of alignment tools for application on Nanopore sequencing data (Becht, Schmidt et al. 2021)

As mentioned before, the bioinformatic data analysis of Nanopore sequencing data is a challenging task due to the complex error profile and the large amount of data which is generated during a single sequencing run. Since bioinformaticians are rarely available in clinical diagnostic laboratories, this can present an important bottleneck for the application of Nanopore sequencing in clinical laboratory diagnostics. So far, a large variety of different tools which can be used for Nanopore sequencing data analysis has been developed and made freely available by the scientific community. A crucial component of many data analysis pipelines is sequence alignment which is the process of referring the generated reads back to a reference sequence¹⁸. This procedure is one example of a bioinformatic task that is aggravated by the complex error profile of Nanopore sequencing reads. Therefore, the publication by Becht, Schmidt et al. 2021 evaluates the performance of different alignment tools applied to Nanopore sequencing data¹⁸.

For this purpose, three different data sets were processed using six different state-of-the-art alignment tools (BLASR, BWA MEM, GraphMap, LAST, minimap2, NGMLR)¹⁸. The three data sets were produced by I) lambda phage whole genome sequencing, II) amplicon sequencing of *MEFV* and III) SARS-CoV-2 whole genome sequencing and encompassed reads of different quality, size and read length distribution (see Table 1 Becht, Schmidt et al. 2021)¹⁸. While the data sets were processed by using the individual tools, performance measures including computational time (central processing unit (CPU) time) and memory (random-access memory (RAM)) usage were recorded for an application in single- and multi-thread mode¹⁸. After the data analysis was finished, quality measures of the alignments including match rate, mismatch rate and error rate were calculated from the resulting alignment files¹⁸. This was done by extracting the frequency of matches (identical nucleotides between query and reference), mismatches (differing nucleotides between query and reference), deletions (missing nucleotides in query compared to reference) and insertions (additional nucleotides in query compared to reference) for each alignment from the alignment output files of the different tools as described by Becht, Schmidt et al.¹⁸.

3. Results

The comparison of the performance measures showed clear differences between the different tools for all three data sets in single thread mode as well as in multi thread mode (Figure 5)¹⁸. Multithreading decreased the computational time while increasing the peak memory consumption¹⁸. Overall, the tool minimap2 showed a superior performance in speed with moderate memory consumption¹⁸.

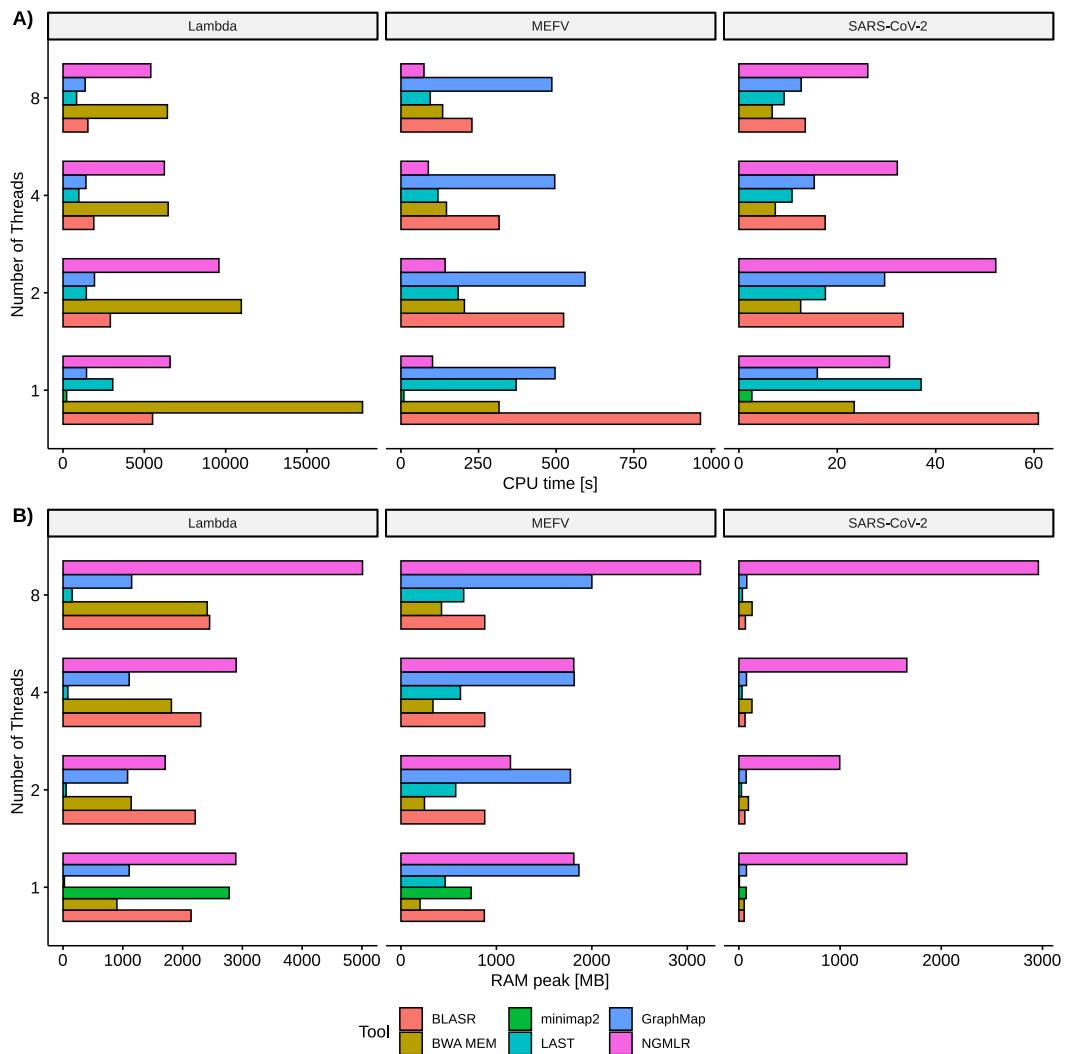


Figure 5: Comparison of six different sequence alignment tools regarding speed and memory consumption. A) Computational time and B) peak memory which are required by each tool for processing the three test data sets using 1, 2, 4 or 8 threads. minimap2 was only used in single-thread mode since this tool does not support multithreading. CPU; Central processing unit, RAM; Random-access memory

Alignment quality assessment by calculating match rate, mismatch rate and error rate as quality measures revealed a similar distribution for all tools except LAST (Figure 6)¹⁸. A generalized linear model (GLM) assuming a Poisson distribution was applied on the underlying count data to investigate the influence of the tool choice on the different rates¹⁸. The alignment length was used as an offset variable¹⁸. For the Lambda phage data set, a highly significant positive

3. Results

influence of the tool LAST was observed on the median mismatch rate (Estimate [E] = 1.145, Standard Error [SE] = 0.155, $P < 0.001$) as well as on the median error rate (E = 0.405, SE = 0.100, $P < 0.001$)¹⁸. The other two test data sets presented no significant influence of the tool choice on any of the quality measures¹⁸.

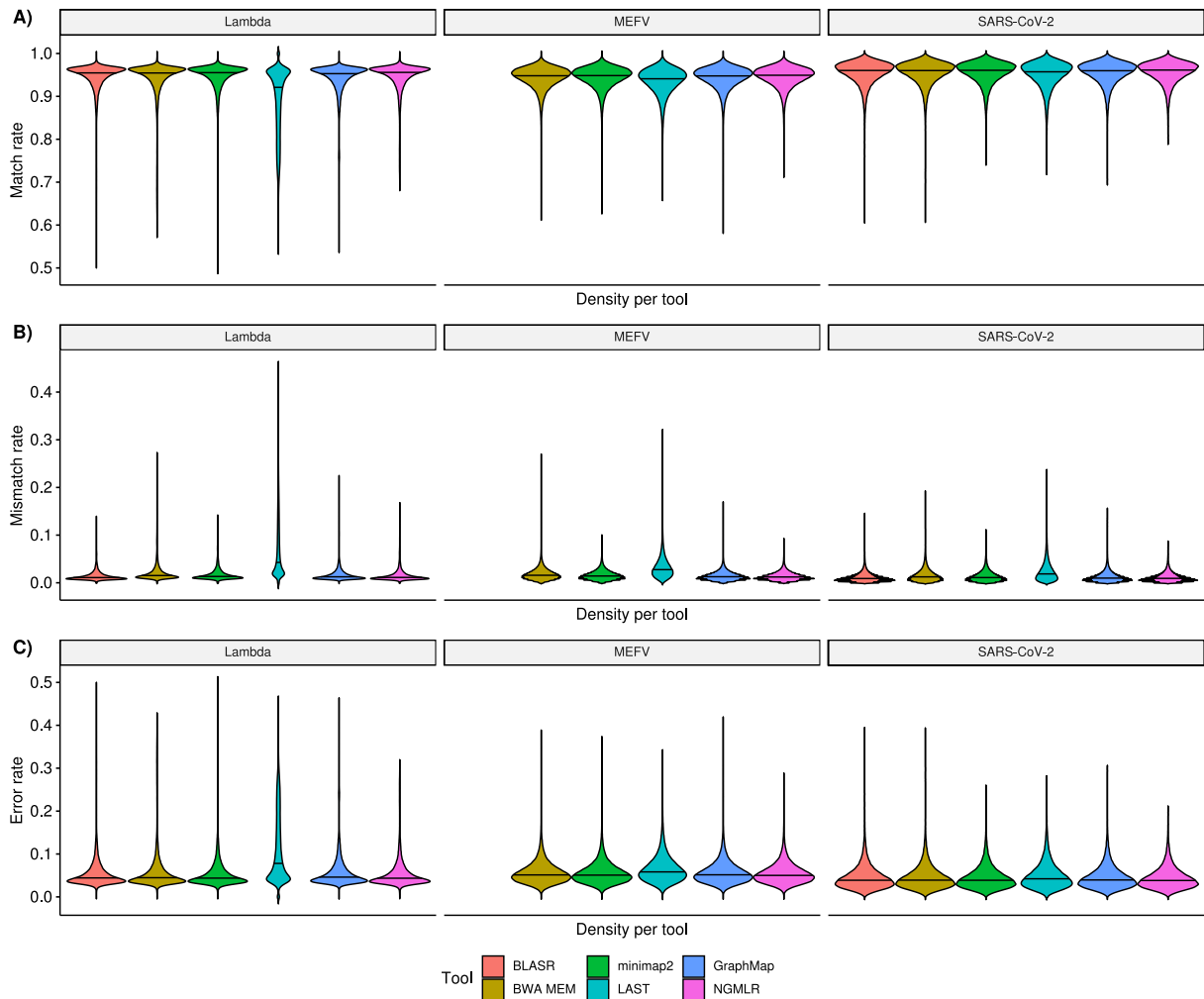


Figure 6: Distribution of A) match rate, B) mismatch rate and C) error rate which were calculated for the application of the different alignment tools on each data set. The tool BLASR was excluded from the calculations for the *MEFV* data set due to insufficient alignment position output, which made an unbiased comparative analysis of the amplicon sequencing data impossible for this tool.

3.2 Application of Nanopore sequencing for SARS-CoV-2 whole genome sequencing

Since Nanopore sequencing offers high throughput and the possibility of sequencing multiple samples within a single run, it is well suited for viral whole genome sequencing (WGS). This procedure can be applied in a research setting to resolve the viral genetic features but also in a clinical diagnostic setting to identify chains of transmission and gain additional epidemiological

information. Further, WGS is the only diagnostic tool available which can be used to accurately monitor the dynamic viral evolution over time and identify new viral strains.

3.2.1 Serological and viral genetic features of patients with COVID-19 in a selected German patient cohort—correlation with disease characteristics (Schmidt et al. 2021)

To study host-virus interaction after SARS-CoV-2 infection, Schmidt et al. investigated the genetic virus characteristics and the ensuing humoral immune response³⁷.

In this study, a total of 55 patients diagnosed with COVID-19 from the State of Baden-Württemberg in Germany were included (see Figure 1 Schmidt et al. 2021)³⁷. Inclusion criteria encompassed a positive SARS-CoV-2 PCR test and a sample of the viral RNA present in the long-term sample archive³⁷.

Clinical data including patient data, risk factors, symptoms and duration of the disease, long-term effects, treatment and epidemiological data were collected using a comprehensive questionnaire (see Table 1 Schmidt et al. 2021)³⁷.

Serum samples for serological testing were collected by venepuncture, on average, 83 days (mean 83.3 days, SD 14.3 days) after a positive PCR result and tested for anti-SARS-CoV-2 IgG and IgM levels to a mixture of S and N proteins (anti-S/N) by using two commercial ELISA kits³⁷. Additionally, anti-SARS-CoV-2 IgG was differentiated in IgG to S1 (anti-S1), S2 (anti-S2) and N proteins (anti-N) by using a third immunoassay as recommended by the manufacturer³⁷.

SARS-CoV-2 WGS of the 55 archived RNA samples was performed on a MinION nanopore sequencing device applying the ARTIC nCoV-2019 sequencing protocol³⁷.

A complete follow-up was not possible in six of the 55 cases because the individual was deceased (5 cases) or not available for sample collection (1 case)³⁷. The questionnaire was returned by 48 out of 49 available patients from which serum samples were collected³⁷.

All 55 viral genomic consensus sequences which were generated during WGS were included in further downstream analysis since the coverage was above 85% (min 88.9%, max 99.6%)³⁷. Variants which were called against the MN908947.3 reference showed a clear distribution over the whole SARS-CoV-2 genome (Figure 7A)³⁷. Overall, 90 different unique variants including 34 synonymous single nucleotide variants (SNVs), 48 non-synonymous SNVs, 2 non-frameshift insertions, 1 frameshift insertion and 5 unclassified variants were identified³⁷. Four variants were identified in all 55 samples (c.C2772T (ORF1ab F924F), c.C14144T (ORF1ab

P4715L), c.A1841G (S D614G), C241T (5' UTR))³⁷. The only invariant region in the investigated sample collective was ORF10 (Figure 7B)³⁷. Median variant count per sample was eight and the highest number of variants was found in ORF1ab, followed by S, 5' UTR, and ORF3a (Figure 7B)³⁷.

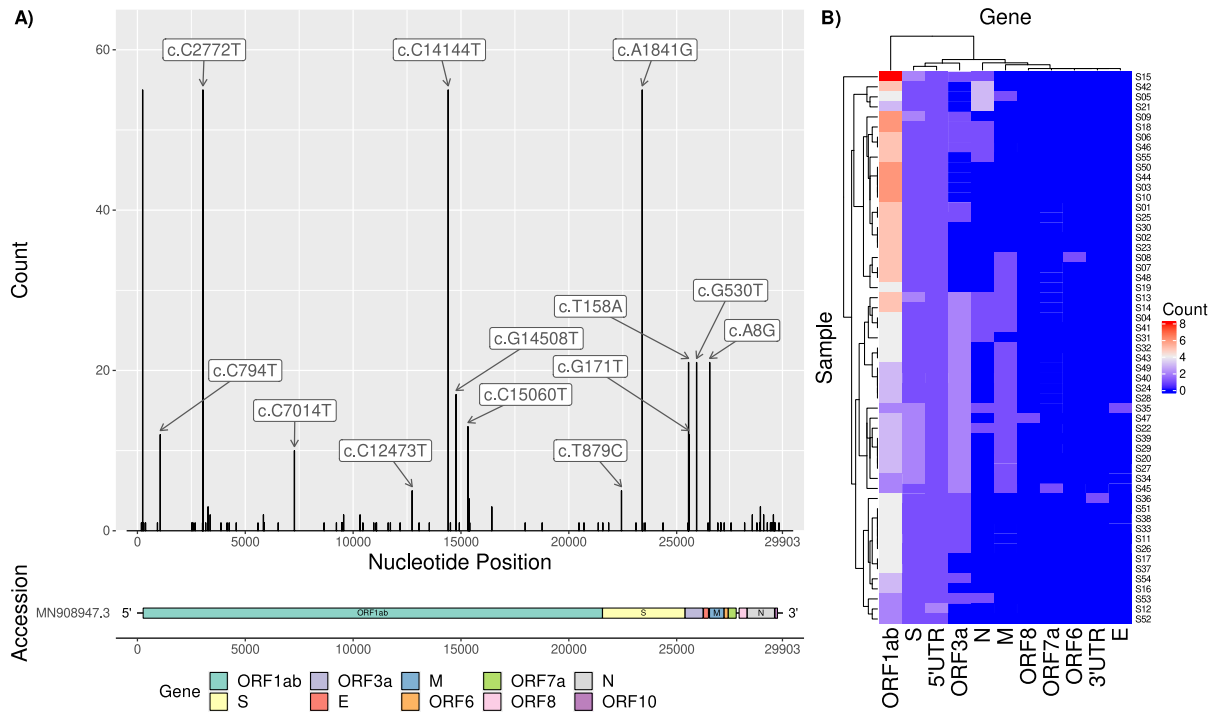


Figure 7: Viral genetic variants which were identified by Nanopore sequencing of 55 RNA isolates. A) Genome-wide distribution and frequency of genetic variants. B) Individual variant count per gene in each study sample. All variants were called against the MN908947.3 reference genome³⁷.

A generalized linear model to assess the variation rate of the individual genes relative to their length revealed a highly positive influence of the N gene on the normalized variation rate ($P=0.0096$, estimate 0.876, standard error [SE] 0.338) as well as a significant negative influence of ORF1ab ($P=0.04$, estimate: -0.528, SE: 0.258)³⁷. This means that, normalized to the respective gene length, N shows a significant larger and ORF1ab a significant lower number of unique variants compared to the other genes³⁷.

To supplement the genetic variation analysis, a phylogenetic analysis of the data was performed (Figure 8)³⁷. For this purpose, multiple sequence alignment was applied and the alignments were analysed in a phylogenetic framework using Maximum-Likelihood Phylogenies³⁷.

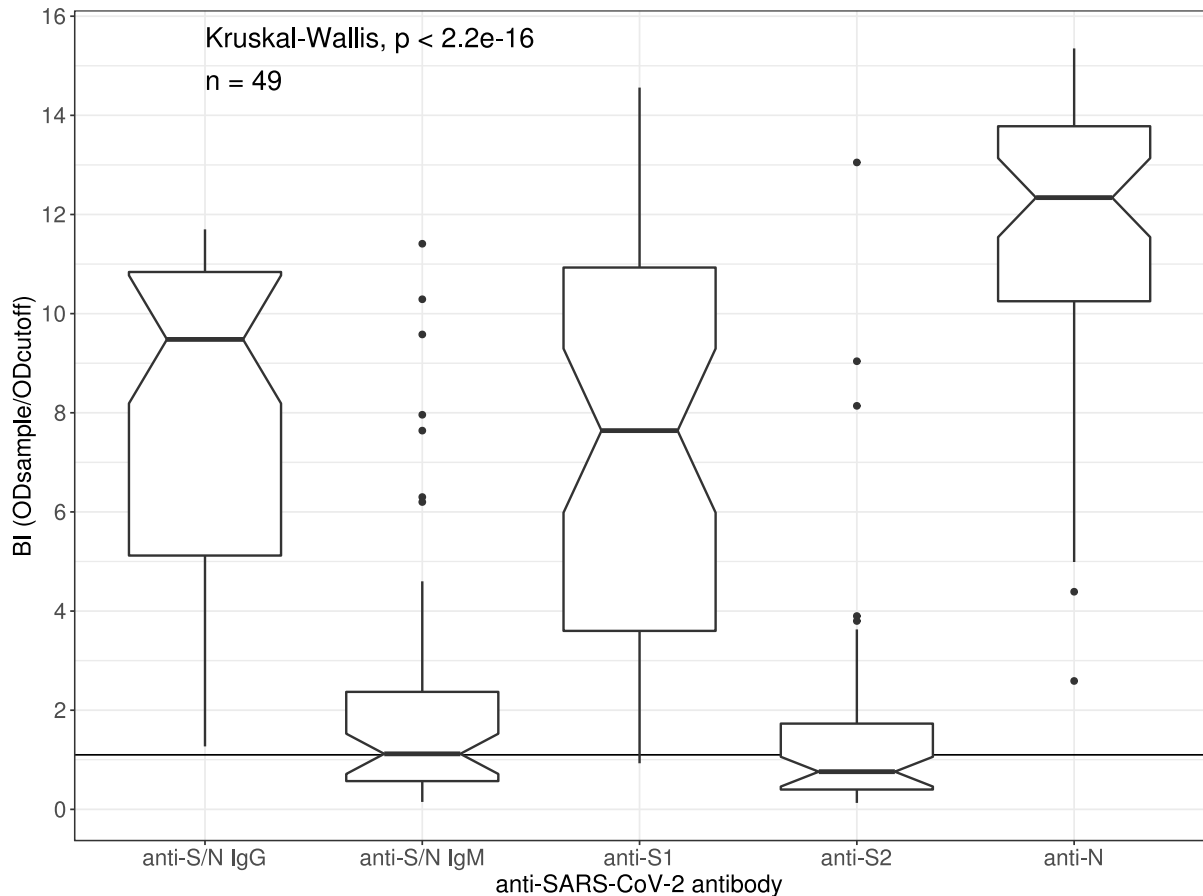


Figure 9: Anti-SARS-CoV-2 antibody levels in 49 patients with COVID-19. Anti-SARS-CoV-2 IgG and IgM against a mixture of spike glycoprotein and nucleocapsid (anti-S/N) were assessed. In addition, IgG antibodies were further differentiated into reactivity against spike glycoprotein domain 1 (anti-S1), domain 2 (anti-S2) and the nucleocapsid protein (anti-N). The positive cut-off of the assays is located at a binding index (BI) of 1.1³⁷.

Anti-SARS-CoV-2 antibody levels were further differentiated by age groups, body mass index (BMI) groups and COVID-19 severity (see Figure 5 Schmidt et al. 2021)³⁷.

Differentiation by age showed significantly higher anti-S/N, anti-S1 and anti-N IgG levels in patients older than 65 compared to younger age groups ($P < 0.05$, respectively)³⁷. Further, significant higher anti-S/N IgM levels were identified in patients older than 65 compared to patients aged 30-65 ($P = 0.012$) (see Figure 5A Schmidt et al. 2021)³⁷.

Correlation of the BMI with the various anti-SARS-CoV-2 antibodies showed significantly higher antibody levels in overweight patients (BMI 25-35) and severely overweight patients (BMI > 35) compared to the normal weight group (BMI < 25) for all tested antibodies except anti-S/N IgM and anti-S2 IgG ($P < 0.05$, respectively)³⁷. Anti-S/N IgM levels were only significantly higher in patients of the overweight group compared to patients in the normal weight group ($P = 0.013$) (see Figure 5B Schmidt et al. 2021)³⁷.

Hospitalized COVID-19 patients showed significant higher levels of anti-S/N IgM and anti-S1 IgG compared to the non-hospitalized group ($P < 0.05$, respectively)³⁷. No significant difference

regarding the need for hospitalization was identified for the other antibodies tested (see Figure 5C Schmidt et al. 2021)³⁷.

To investigate a possible influence of other patient and viral genetic characteristics on antibody levels, univariate followed by multivariate regression analysis was performed (see Table 3 Schmidt et al. 2021)³⁷. In this way, age was established as an independent predictor for higher anti-S/N, anti-S1 and anti-N IgG levels³⁷. Being overweight (BMI>25) was an additional independent predictor for higher anti-S/N and anti-S1 IgG levels³⁷. Further, the absence of the viral genetic variant NSP3 D218E was an independent predictor for higher anti-S1 IgG levels while the absence of chronic liver diseases predicted higher anti-S/N IgG levels³⁷. The presence of tumour diseases was the only independent predictor for higher anti-S/N IgM levels³⁷. Higher anti-S2 IgG levels were only predicted by the presence of the viral genetic variant NSP3 D218E³⁷.

A possible association between the clinical outcome and various independent predictor variables including patient characteristics, antibody levels and viral genetic features was further evaluated by applying univariate followed by multivariate regression analyses³⁷. In the multivariate logistic regression to account for confounding variables, higher anti-S/N IgG and/or IgM levels were found to significantly predict COVID-19 characteristics including appetite loss, night sweat, oxygen need, pneumonia and the need for hospitalization ($P<0.05$, respectively) (see Table 4 Schmidt et al. 2021)³⁷. The occurrence of pneumonia was only predicted by the anti-S/N IgM level (odds ratio [OR] 1.363, $P=0.0317$)³⁷. Further, the prediction of the need for oxygen and hospitalization by higher anti-S/N IgM levels was mainly confounded by cardiovascular diseases ($P<0.05$, respectively)³⁷. Beside blood group A+, cough was additionally predicted by the SNV ORF3a S177I ($P=0.0061$)³⁷. Taste and smell disorders were independently predicted by the SNV NSP12 Q444H (OR 5.444, $P=0.0426$)³⁷. Together with the presence of tumour and chronic lung diseases, higher anti-S/N IgM levels were significantly associated with longer hospitalization (multiple regression analysis, $P<0.05$, respectively)³⁷. The viral SNV N E253A and chronic lung disease were significantly associated with symptom duration (multiple regression analysis, $P<0.05$, respectively)³⁷.

3.2.2 Epidemiological surveillance of SARS-CoV-2 by whole genome sequencing

The protocol described above for SARS-CoV-2 WGS on a Nanopore sequencing device was applied in the molecular biology department of the clinical diagnostic laboratory MVZ Laborärzte Singen to support the epidemiological surveillance during the SARS-CoV-2 pandemic. For this purpose, weekly sequencing runs of selected RNA isolates from nasopharyngeal swabs that tested positive for SARS-CoV-2 by reverse transcription (RT)-PCR in the previous week were performed. To achieve maximum coverage of the viral genome only samples with a cycle threshold (Ct) value below 25 were included. The generated whole genome sequences were classified into viral lineages using the Pango Nomenclature⁴⁷. These results were forwarded to local health authorities to support the identification of chains of infection and appropriate quarantine measures. Special attention was paid to variants of concern (VOC) as designated by the World Health Organisation (WHO) in order to prioritise global monitoring and research⁴⁸.

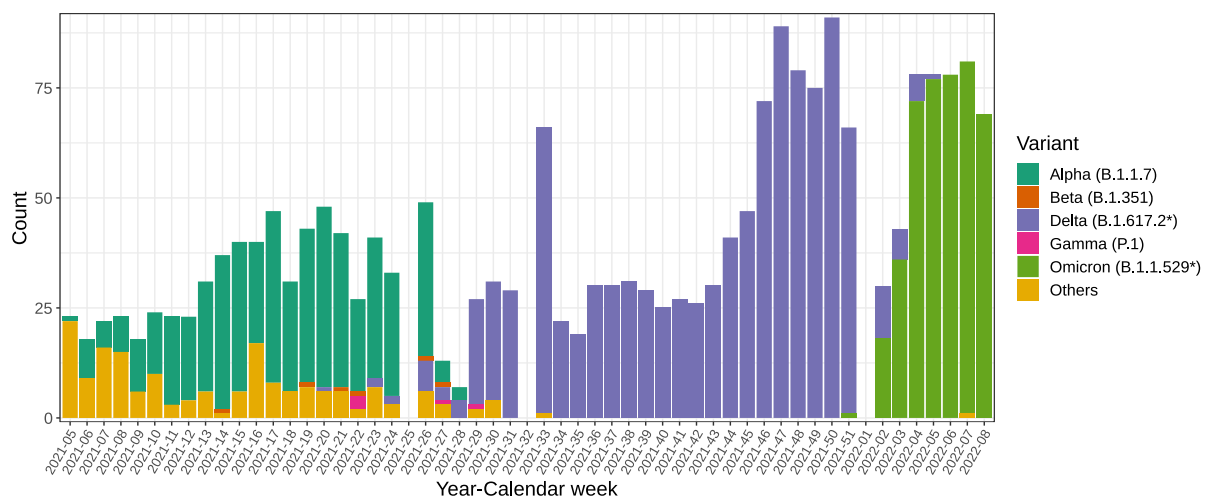


Figure 10: Distribution of SARS-CoV-2 variants of concern (VOC) in the area of the clinical diagnostic laboratory MVZ Laborärzte Singen. The results show weekly SARS-CoV-2 whole genome sequencing runs including selected samples which have been tested positive for SARS-CoV-2 by RT-PCR in the previous week. Sequencing was performed on a MinION Nanopore sequencing device. Variants are labeled according to the WHO and Pango nomenclature. *including sublineages

By applying this molecular biological surveillance concept during the course of 2021, major changes in the distribution of SARS-CoV-2 VOC were observed (Figure 10). The Alpha variant (B.1.1.7) dominated the first six months and was then rapidly displaced by the Delta variant (B.1.617.2). At the end of 2021, Delta was again rapidly displaced by an exponential increase in the Omicron variant (B.1.1.529).

3.3 Development of a RT-LAMP assay for SARS-CoV-2 detection

Fast and reliable identification of infected individuals is a major objective for clinical laboratory diagnostics of SARS-CoV-2. This is essential to support public health initiatives which mainly consist of quarantine measures to reduce the viral spread and decrease the burden on the health care system. As mentioned before, the direct detection of SARS-CoV-2 in nasopharyngeal swabs by RT-PCR is considered the gold-standard method for this purpose. However, the extremely high demand for RT-PCR assays during the pandemic led to a temporary shortage of PCR reagents, plastic consumables and reagents needed for RNA isolation from clinical samples. Therefore, it is important to develop and establish detection methods for SARS-CoV-2 which can be used in addition to conventional RT-PCR. Such methods should ideally show a comparable performance to the gold-standard method, rely on different reagents and consumables, and allow high throughput by automation capability. In addition, it is preferable to establish assays that do not necessarily require RNA isolation prior to analysis since this significantly increases efficiency of the entire workflow where speed and costs are concerned.

3.3.1 A semi-automated, isolation-free, high-throughput SARS-CoV-2 reverse transcriptase (RT) loop-mediated isothermal amplification (LAMP) test (Schmidt et al. 2021)

To support SARS-CoV-2 diagnostics, in this research paper by Schmidt et al. (2021), a robust SARS-CoV-2 RT-LAMP was established to target the E and N gene with high-throughput capabilities and short turnaround times in a clinical laboratory setting⁴⁹. The protocol can be used in combination with standard RNA isolation from nasopharyngeal swabs, but also with a simple enzymatic digestion for sample preparation⁴⁹. Semi-automated high throughput processing is possible by using a liquid handling station⁴⁹. In addition, the method was validated by a comprehensive method comparison with two conventional RT-PCR protocols using 323 clinical samples of individuals with suspected SARS-CoV-2 infection (see Table 1 Schmidt et al. 2021)⁴⁹.

Via fluorescence detection, the established SARS-CoV-2 RT-LAMP protocol enabled a clear identification of SARS-CoV-2 positive samples confirmed by prior standard RT-PCR tests⁴⁹. Compared to negative samples a sigmoid increase of the fluorescence intensity over the isothermal incubation time was observed for positive samples (Figure 11)⁴⁹. By using conventional gel electrophoresis these samples additionally showed a LAMP characteristic banding pattern (Figure 12)⁴⁹. Negative samples, as well as the no-template control, did not show an increase of fluorescence over time and were below a predefined, operator adjustable threshold⁴⁹. Similar

to the calculation of Ct values in RT-PCR experiments, this threshold can be used to calculate the threshold time (Tt) of positive samples⁴⁹.

After establishing the method, the RT-LAMP protocol was compared with an in-house RT-PCR targeting the E gene and with a commercial PCR kit targeting the E, N and ORF1ab genes, using 70 isolated RNA samples⁴⁹. This method comparison showed an almost perfect agreement (Cohen's kappa [κ] > 0.8) with no systematic difference (McNemar's test, $P > 0.05$) between the RT-LAMP and both reference methods⁴⁹. The sensitivity was 94.4% (95% confidence interval [CI] 81.3-99.3%) compared to the in-house RT-PCR, and 89.5% (CI 75.2-97.1%) compared to the commercial RT-PCR kit⁴⁹. In both cases, no false positive results were observed resulting in 100% specificity⁴⁹. A significant positive correlation (Rho [ϕ] > 0.8, $P < 0.001$) was observed between the Tt values of the RT-LAMP assay and the RT-PCR Ct values for all different targets (Figure 13A-D)⁴⁹.

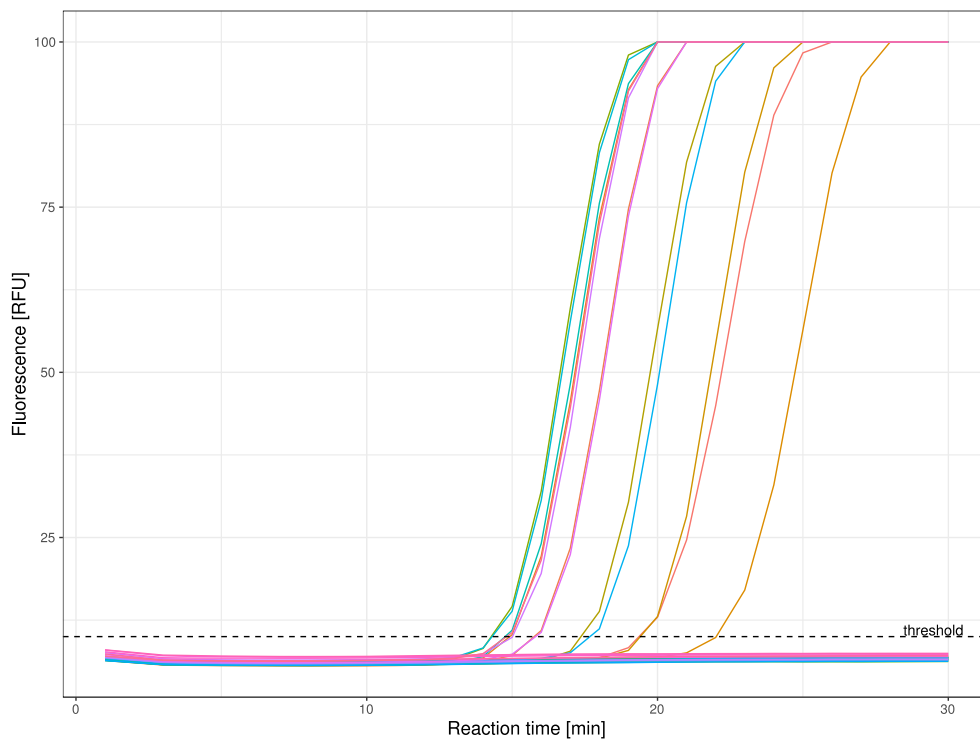


Figure 11: Amplification curves of the SARS-CoV-2 RT-LAMP assay. An intercalating dye was used to detect the amplification products via fluorescence detection. The fluorescence signal is read at minute intervals during the 30 min isothermal incubation at 65 °C on the FAM channel of a conventional RT-PCR cycler. A user defined threshold is applied to identify positive samples and calculate the threshold time (Tt) value⁴⁹.

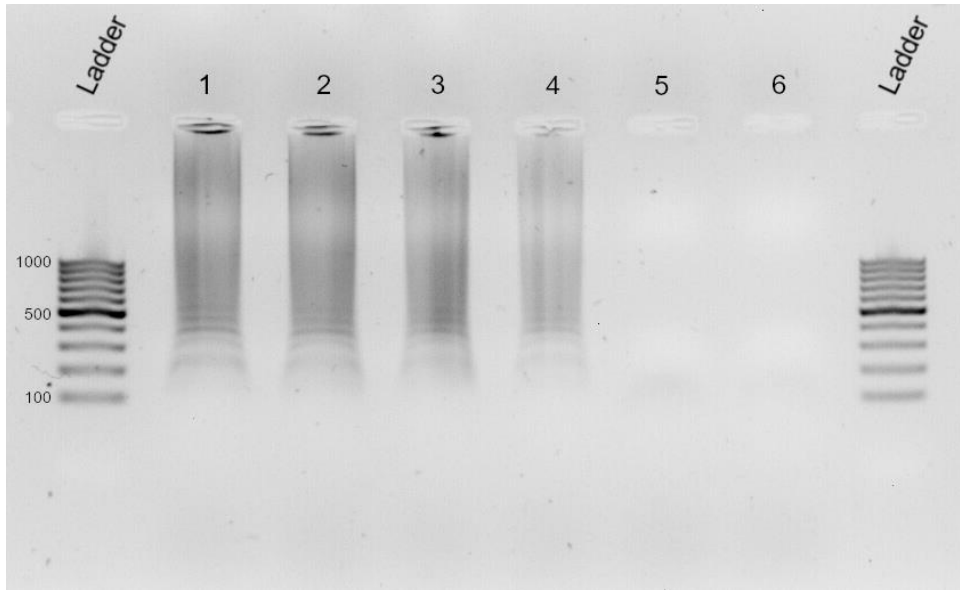


Figure 12: Conventional agarose gel (1.5%) electrophoresis of SARS-CoV-2 RT-LAMP reaction products. The gel was stained with ethidium bromide and a 100 bp DNA Ladder was used as reference. Colour was inverted to improve readability. Reference samples containing 10^7 (1,2) and 10^6 (3,4) SARS-CoV-2 genome copies/ml show a LAMP characteristic banding pattern. This is not present in the no-template controls (5,6)⁴⁹.

Samples classified as false-negative by the RT-LAMP compared to the in-house RT-PCR (2/70) and the commercial RT-PCR (4/70) showed Ct values in the range of 20 to >35 ⁴⁹.

Since high-throughput and short hands-on times are important parameters for the application of an assay in clinical routine diagnostics, the SARS-CoV-2 RT-LAMP protocol was adapted for an application to a liquid handling system⁴⁹. A comparison of this semi-automated approach to the in-house RT-PCR using 188 isolated RNA samples showed a perfect agreement between both methods ($\kappa = 1$)⁴⁹. In this way, sensitivity of 100% (CI 84.6-100%) and a specificity of 100% (CI 97.8-100%) were observed⁴⁹.

To further decrease the sample processing time, a RNA isolation-free sample preparation step by proteinase K digestion was added to the RT-LAMP protocol for nasopharyngeal swabs⁴⁹. Again, for performance assessment a method comparison with the in-house PCR protocol was carried out using 65 nasopharyngeal swabs⁴⁹. To achieve comparable results between both methods, these swabs were initially eluted in NaCl solution (0.9%)⁴⁹.

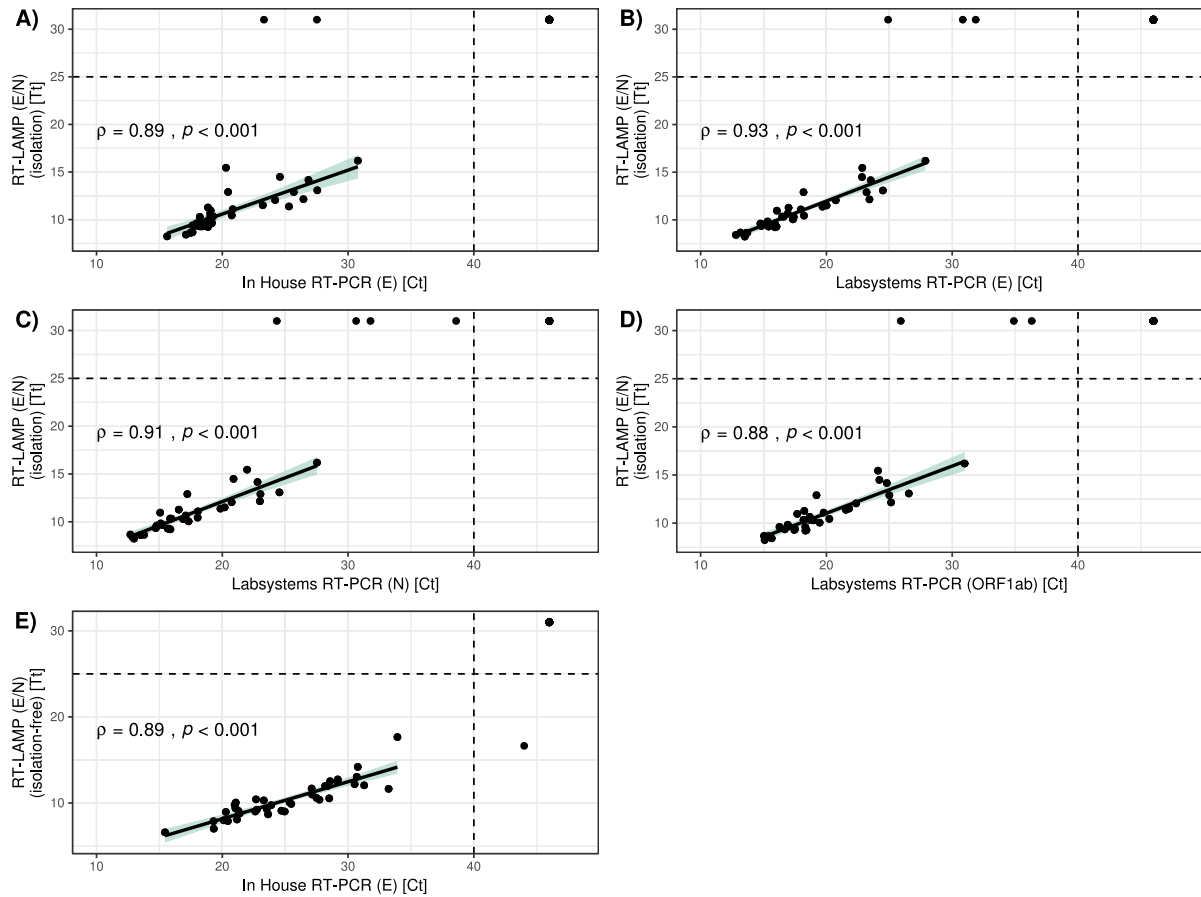


Figure 13: Correlation between SARS-CoV-2 RT-LAMP Tt values and RT-PCR Ct values. The RT-LAMP test with RNA isolation was compared to an in-house RT-PCR assay targeting the E gene (A), as well as a commercial RT-PCR kit targeting the E (B), N (C) and ORF1ab gene (D). Additionally, the isolation free RT-LAMP protocol was compared to the in-house RT-PCR (E). Negative cut-offs are represented by dashed lines. Artificial Ct and Tt values above these cut-offs were assigned to negative samples for data visualization purposes. Only true positive results were included in the correlation analysis and Spearman correlation results are shown.

The eluates were divided into two aliquots, one of which underwent standard RNA isolation followed by RT-PCR and the other was analyzed using the isolation-free RT-LAMP protocol⁴⁹. Again, a near perfect agreement ($\kappa=0.97$) with no systematic difference (McNemar's test, $P=0.79$) was observed between both methods⁴⁹. One sample was classified false-positive by RT-LAMP with a Tt value above 15⁴⁹. Compared to the RT-PCR's Ct values the RT-LAMP Tt values showed a significant positive correlation ($\phi=0.89, P<0.001$) (Figure 13E)⁴⁹. The sensitivity and specificity of the isolation-free RT-LAMP were 100% (CI 91.0-100%) and 96.2% (CI 80.4-99.9%), respectively⁴⁹.

Additional performance characteristics of the isolation-free SARS-CoV-2 RT-LAMP, including the limit of detection (LoD) and the intra-/inter-assay reproducibility, were assessed using reference material with a known viral load as well as positive and negative sample pools⁴⁹. By serial dilution a LoD of 100000 copies/ml was identified, which is equal to 5000 copies in the eluate from a nasopharyngeal swap, or 95 copies in the RT-LAMP reaction itself⁴⁹.

Reproducibility was assessed by measuring both sample pools in five replicates on three consecutive days⁴⁹. The results demonstrated 100% intra- and inter-run reproducibility with low intra-run Tt variability (standard deviation [SD] 0.042, coefficient of variation [CV] 0.4%) and inter-run Tt variability (SD 0.222, CV 2.1%) (see Table 2 Schmidt et al. 2021)⁴⁹.

4. Discussion

4.1 Establishment and evaluation of Nanopore sequencing for application in clinical laboratory diagnostics

To initially establish current Nanopore sequencing in clinical laboratory diagnostics, a sequencing protocol and the corresponding data analysis pipeline to sequence selected regions of *MEFV* and support the diagnosis of FMF were developed. The analytical performance was assessed by performing a comprehensive method comparison with the gold standard Sanger sequencing using 47 clinical samples.

The molecular genetic diagnosis of FMF was selected for comparing both sequencing methods because its diagnosis by conventional Sanger sequencing is very well established which makes it a good reference. Furthermore, FMF is a more common genetic disease which makes it possible to collect a sufficient number of samples for method comparison within a reasonable time frame.

Diagnostic Sanger sequencing revealed the presence of various SNPs including the common non-synonymous variants p.E148Q, p.R202Q, p.M694V, p.P369S and p.R408Q which have been previously described in FMF patients (see Table 1 Schmidt et al. 2022)³¹.

By using Nanopore sequencing on a MinION sequencing device, in combination with a dedicated data analysis pipeline, all relevant regions of the *MEFV* exons were sequenced with a very high read depth and all variants previously identified by Sanger sequencing were detected accurately³¹. Additionally, the Nanopore sequencing results revealed one SNP in two related patients which had not been identified by Sanger sequencing before³¹. It was located in the 3' UTR at the edge of the amplicon covering this region³¹. Poor sequence quality at the beginning and end of an individual read due to primer binding and insufficient base resolution is a very common problem in Sanger sequencing since this technique is based on PCR amplification in combination with capillary electrophoresis³¹. Therefore, by sequencing an additional amplicon, which was designed to span the region of the 3' UTR in which the diverging SNP is located, it was possible to confirm this transversion in both patients also by Sanger sequencing³¹. Taking these additional results into account, overall a perfect agreement between Nanopore sequencing and Sanger sequencing was observed³¹. This is underlined by the Precision, Recall, and F1-Score >0.99, which were calculated treating the initially diverging SNP as a false positive call³¹. A similar high method agreement between Nanopore and Sanger sequencing was also reported by other studies, especially in the fields of microbiology and cancer genomics^{31,50–53}. However,

the relatively small sample size, which is a major limitation of the study by Schmidt et al., makes it mandatory to confirm these results in a larger sample population.

Due to the high read depth in this study it was possible to use Bayesian statistics for accurate variant calling in order to determine the most likely genotype³¹. As the results show, this seems to be sufficient for amplicon sequencing experiments carried out on a Nanopore sequencing device. However, for other modern diagnostic NGS applications including gene panel sequencing, whole exome sequencing, and whole genome sequencing this approach might not be sufficient³¹. In such scenarios, normally the median read depth is much lower compared to amplicon sequencing due to the obviously larger target space³¹. Therefore, for these applications it is necessary to use more advanced tools which can compensate for the Nanopore specific sequencing error even at low read depth³¹. While this study mainly focused on SNP genotyping, other specialized tools are necessary for structural variant calling from Nanopore sequencing data including deletions, inversions, tandem duplications, insertions, transpositions, and translocations³¹.

As well as the analytical performance, the economic aspects of Nanopore sequencing, which are also important for the establishment of an innovative method in clinical laboratory medicine, were evaluated by Schmidt et al. (see Table 2 Schmidt et al. 2022)³¹. The direct comparison to Sanger sequencing shows that Nanopore sequencing, while having a similar analytical performance, requires lower capital costs and has a lower price per sample due to multiplexing³¹. The time to result including DNA isolation, PCR amplification, sequencing and data analysis is comparable between both methods³¹. Overall, this makes Nanopore sequencing attractive for an application in clinical routine diagnostics also from a financial point of view. Compared to other modern sequencing technologies like Illumina sequencing, Ion Torrent sequencing and PacBio sequencing, the choice of Nanopore sequencing for clinical routine diagnostics is especially attractive due to its fast processing time and lower costs combined with the ability to generate long reads^{31,54,55}.

As mentioned before, a major drawback so far of Nanopore sequencing is the complex data analysis. Due to the large amount of data generated during a single sequencing run and the complex error profile developing and maintaining a suitable data analysis pipeline can be a challenging task. Various different tools are currently freely available and these can be used in different applications during the analysis of Nanopore sequencing data. This makes it essential

to compare the performance of different tools for an individual bioinformatic task to identify the tool best suited to the analysis of a specific data set or research question. Therefore, Becht, Schmidt et al. compared six different alignment tools for the application to Nanopore sequencing data¹⁸. The task of sequence alignment was chosen because it is a general-purpose procedure which is part of many different bioinformatic data analysis pipelines.

The comparison was performed using three different data sets with varying data structure including read length distribution, read count, and overall read quality¹⁸. Computational performance of the six tools was assessed and compared by recording the CPU time and RAM usage during the alignment process of the test data set to the respective reference genomes¹⁸. Although the experiments were performed on a standard notebook using only conventional hardware none of the alignment tools failed to process all three data sets¹⁸. One big advantage of Nanopore sequencing over other technologies is the portability of some sequencers¹⁸. This enables researchers to perform sequencing experiments in areas where dedicated high-performance workstations for data analysis might not be available¹⁸. It is encouraging that all the different alignment tools worked also on standard hardware which would support a portable application of the sequencer¹⁸.

The comparison of the computational performance parameters revealed major differences between the different tools¹⁸. The aligner minimap2 was the fastest tool on all three data sets with an intermediate peak RAM consumption¹⁸. This shows the importance of comparing different tools prior to an application in a data analysis pipeline since they are all based on different algorithmic approaches which can result in remarkable performance differences¹⁸. As expected, multithreading reduced the run time at the expense of an increased peak RAM consumption¹⁸. However, this principle was not supported by all tools¹⁸.

The quality of the alignments produced by the different tools was compared by calculating quality measures including match rate, mismatch rate and error rate¹⁸. By doing so, high match rates and low mismatch and error rates for all data sets were observed when applying BLASR, BWA MEM, minimap2, GraphMap and NGMLR¹⁸. This underlines that all of these tools are suitable for application to Nanopore sequencing data¹⁸. Compared to the other tools, LAST showed a significant higher mismatch and error rate for the Lambda data set¹⁸. This has also been observed by others and might be the result of the algorithmic design of this tool^{18,56}.

In addition, the comparison revealed that BLASR, although showing acceptable performance and quality parameters, can hardly be implemented in a Nanopore sequencing data analysis pipeline since binary alignment map (BAM) output is not supported for input data generated by a Nanopore sequencing device¹⁸.

To summarise, this work by Becht, Schmidt et al. shows that a comparison of different available tools prior to the integration in a data analysis pipeline can substantially affect the overall results and performance¹⁸. Since differences among the tools can be explained by different algorithmic approaches and different data set characteristics, the selection of the tool for a specific approach should be done while considering the whole experimental design and further downstream analysis¹⁸.

4.2 Application of Nanopore sequencing for SARS-CoV-2 whole genome sequencing

Besides SNP genotyping in human genetics, where Nanopore sequencing has shown a robust analytical performance, another innovative application in the field of modern laboratory medicine is viral whole genome sequencing.

In the study by Schmidt et al., a group of SARS-CoV-2 PCR-positive patients of a south-western German region was examined to investigate virus-host interactions³⁷. SARS-CoV-2 whole genome sequencing on a MinION sequencing device as well as serological testing was performed and the patients answered a questionnaire on personal and COVID-19 characteristics³⁷.

The sequencing results of the viral genome of 55 COVID-19 patient samples showed genetic alterations mainly as SNVs³⁷. Half of them resulted in changes of the amino acid sequence³⁷. The absolute variant count per gene and sample was highest within ORF1ab representing the largest SARS-CoV-2 ORF³⁷. However, normalized to the gene length, the N gene was the only gene with a significant higher variation rate, while ORF1ab showed a significant lower rate compared to other genes³⁷. The high variation rate of the N gene was also reported elsewhere^{37,57,58}. The only gene without variants was ORF10 which was also demonstrated by others^{37,58}. In general, RNA viruses tend to accumulate variants over time during their replication cycle because RNA copying enzymes are prone to error^{37,59,60}.

Remarkably, four variants were present in all samples investigated (ORF1ab F924F, ORF1ab P4715L, S D614G, and 5'UTR 241C>T), representing signature variants of the SARS-CoV-2 type VI strain, which was most dominant at the time of sample collection^{37,61}. The D614G

exchange in the S protein, especially, has been extensively studied and is assumed to provide a selection advantage by increasing viral infectivity^{37,62–64}.

During phylogenetic analysis of the sequencing data all samples were assigned to the root lineage B based on the nomenclature proposed by Rambaut et al.^{37,47}. The highest level was B.1, which is the lineage that caused the major Italian outbreak in early 2020 and subsequently spread across Europe^{37,47}. The other lineages were sub-lineages of B.1, which match the geographic and chronologic origin of the samples³⁷.

The results of the serological testing revealed detectable anti-S/N and anti-N IgG levels in all patients while only one out of the examined 49 patients did not show anti-S1 IgG³⁷. Since a primary immune response would induce stronger anti-SARS-CoV-2 IgM responses, the higher prevalence of anti-S/N IgG compared to IgM probably results from the effect of an immunological memory likely induced by previous infections with endemic coronaviruses³⁷.

By using rank correlation and multiple regression analyses for the prediction of anti-SARS-CoV-2 antibody levels from genetic SARS-CoV-2 variants and patient characteristics, an association of older (>65 years) and overweight (BMI>25) patients with higher anti-S/N and anti-S1 IgG levels was observed³⁷. Higher anti-N IgG levels were only associated with older patients³⁷. The correlation with older age reflects a stronger humoral inflammatory response reported in aged COVID-19 patients and may hint at an impaired innate or cellular adaptive immune response^{37,38,65}.

Being overweight has been identified as an additional risk factor for severe COVID-19 progression since it is linked with functional impairment of immune cells and decreased immunity as a result of chronic inflammation and hypercytokinemia^{37,66,67}. Therefore, higher anti-S/N and anti-S1 IgG levels in obese patients may result from a unique predisposition to an impaired cellular anti-SARS-CoV-2 response³⁷. As described elsewhere, patients with metabolic syndrome comorbidities also show significant higher anti-SARS-CoV-2 IgG levels^{37,68}.

Overall, old age and obesity were identified as risk factors for severe COVID-19 in a systematic review and meta-analysis^{37,69}.

High anti-S/N IgM and anti-S1 IgG levels were associated with moderate COVID-19 requiring hospitalization of the patients³⁷. Further, both antibodies were positively correlated with the duration of hospitalization³⁷. By multivariate regression analysis only higher anti-S/N IgM levels were identified as a predictor for the need for hospitalization with the presence of cardiovascular diseases as a confounder³⁷. Based on this observation anti-S/N IgM can be employed

as a marker of at least moderate COVID-19, in particular, in patients with cardiovascular diseases³⁷.

In agreement with previous reports, anti-S1 IgG levels were higher than anti-S2 IgG levels^{37,70}. By combining serological data and sequencing data, in the study by Schmidt et al., a positive association of higher anti-S2 IgG levels with the SNV NSP3 D218E was shown for the first time³⁷. Interestingly, the same SNV is negatively associated with higher anti-S1 IgG levels in the investigated patient cohort which may indicate a possible influence of SARS-CoV-2 non-structural protein 3 (NSP3) on antibody formation³⁷. This multi-domain protein is the largest SARS-CoV-2 protein and an essential component of the replication-transcription complex modifying host proteins and interfering with innate immune responses by de-ubiquitination^{37,71}.

Univariate followed by multivariate regression analysis was performed to identify possible associations between patient characteristics, serological markers, genetic changes of SARS-CoV-2 and the clinical expression of COVID-19³⁷.

Higher anti-S/N IgM and IgG levels were established as independent predictors of COVID-19 characteristics including appetite loss, night sweat, oxygen need and pneumonia³⁷. Pneumonia was only associated with higher anti-S/N IgM levels without confounders, which supports published data and the correlation of the IgM response with the need for hospitalization⁷².

The absence of the non-synonymous SNV ORF3a S177I was identified as a confounder for the appearance of cough³⁷. Taste and smell disorders were predicted by the non-synonymous SNV NSP12 Q444H (OR 5.4) without confounders³⁷. NSP12 is a SARS-CoV-2 protein with 932 amino acid residues catalysing replication and transcription of the viral genome^{37,73}. Patients with chronic lung diseases infected with SARS-CoV-2 bearing the non-synonymous SNV N E253A showed a longer symptom duration³⁷. The N protein, which presented a high level of genetic alteration in the study, has multiple functions including complex formation with genomic RNA, interaction with the viral membrane protein during virion assembly and enhancement of the efficiency of virus transcription and assembly^{37,74}.

In summary, the results of this study by Schmidt et al. show the excellent applicability of Nanopore sequencing for SARS-CoV-2 WGS in laboratory medicine. It was the first study combining SARS-CoV-2 WGS with comprehensive SARS-CoV-2 serological testing including IgM and IgG reactivities³⁷. The results show diverse humoral immune responses to SARS-CoV-

2, that appear to be influenced by disease severity, age and obesity³⁷. Furthermore, it was observed that even small viral genetic changes may influence the clinical presentation of COVID-19³⁷.

SARS-CoV-2 WGS using Nanopore sequencing was not only applied to research applications but also used to perform SARS-CoV-2 molecular biological surveillance in clinical routine diagnostics during the course of the pandemic. Special attention was paid to the surveillance of WHO variants of concern (VOC). These variants are defined by amino acid substitutions which significantly change the viral properties⁴⁸. Relevant changes include increased transmissibility, a detrimental change in COVID-19 epidemiology, increased virulence or changes in the clinical presentation of COVID-19⁴⁸. In addition, a VOC can lead to a decrease in the effectiveness of public health and social measures or available diagnostics, vaccines and therapeutics⁴⁸.

At the beginning of 2021 the Alpha variant (B.1.1.7) showed a gradually increasing spread which displaced wildtype variants. B.1.1.7 was first detected in September 2020 in southeast England, which gave this variant the name “British” variant⁷⁵. It contains eight mutations in the spike protein, which are thought to increase transmissibility and reduce susceptibility to some therapeutic monoclonal antibodies⁷⁵. Starting from calendar week 28, a strong spread of the Delta variant (B.1.617.2) was observed, so that infections subsequently were composed exclusively of these variant and sub-lineages. The variant B.1.617.2 was first described during a large outbreak in India, rendering it the “Indian” variant⁷⁶. Compared to B.1.1.7 this variant shows a higher transmissibility which quickly made it the dominant strain worldwide⁷⁶. Remarkably, although current vaccines remained highly effective at preventing admission to hospital and death from B.1.617.2 infections, the overall vaccination efficiency is reduced for this lineage⁷⁶. It therefore caused a large number of cases even in countries with a high vaccination rate⁷⁶. At the turn of 2021/22, B.1.617.2 was displaced by an exponential increase of the Omicron variant (B.1.1.529). This VOC was first described in Botswana in November 2021⁷⁷. It contains up to 59 mutations, most of which are located in the spike protein⁷⁷. As mentioned before, this protein is essential for host cell entry and is therefore the main target of neutralizing antibodies against SARS-CoV-2⁷⁷. Mutations in the receptor binding domain (RBD), which is part of the spike protein, can facilitate escape from vaccine-induced neutralizing antibodies and enhance infectivity by increasing the affinity for the Angiotensin Converting Enzyme 2 (ACE2) receptor⁷⁷. It has been shown that B.1.1.529 is able to evade vaccine-induced neutralizing immunity under standard vaccination regimens and shows a higher infectivity than previous variants⁷⁷.

However, a third “booster” dose of mRNA vaccines seems to induce neutralization against Omicron based on cross-reactivity which may overcome evasion of humoral immunity⁷⁷.

To summarise, the variant pattern, which was observed over the course of nearly a year by SARS-CoV-2 WGS of clinical samples on a Nanopore sequencing device, is a good representation of the highly dynamic viral evolution over time. The utilization of Nanopore sequencing for this application in a clinical diagnostic laboratory gives an impression of the great potential of this technique for implementation in laboratory medicine. Historically, sequencing experiments were performed at larger centres or university research laboratories. However, since Nanopore sequencing devices require low capital costs and are relatively cost efficient and easy to use, this might change in the future.

4.3 Development of a RT-LAMP assay for SARS-CoV-2 detection

The fast and reliable identification of SARS-CoV-2 positive clinical samples by NAT, prior to viral whole genome sequencing, is a major task of diagnostic laboratories to support public health measures and the treatment of infected patients during the pandemic.

Bottlenecks in the supply of PCR reagents, laboratory equipment, and reagents needed for RNA isolation from nasopharyngeal swabs during the SARS-CoV-2 pandemic have affected the availability of standard RT-PCR assays for the detection of this virus. To overcome this, a RT-LAMP protocol was established, which can be used as an alternative to conventional RT-PCR⁴⁹.

Using fluorescence detection, this protocol can be applied on a conventional RT-PCR cycler, it is adaptable to automated processing and it can allow a quantitative analysis if needed⁴⁹. Throughput was further increased and hands-on time decreased by adapting the SARS-CoV-2 RT-LAMP protocol on a liquid handling station⁴⁹. Finally, to further simplify the protocol and increase the robustness of the RT-LAMP assay, it was combined with a simple RNA isolation-free sample preparation using proteinase K digestion⁴⁹. Reducing the complexity of sample preparation decreases processing time and the costs per reaction, which are both important parameters for applying an assay to SARS-CoV-2 mass screening in a clinical diagnostic laboratory⁴⁹. By using the established isolation-free RT-LAMP protocol, it is possible to process samples in less than 90 min, which is significantly faster than conventional RT-PCR protocols^{49,78}.

To evaluate the performance of the RT-LAMP protocol, various method comparison experiments with an in-house as well as a commercial RT-PCR assay were performed using clinical samples⁴⁹. The qualitative results of the RT-LAMP showed near perfect agreement with the RT-PCR assays under all conditions tested⁴⁹. No indication of a systematic difference was observed ($\kappa > 0.8$, McNemar's test, $P > 0.05$)⁴⁹. Sensitivity ranged from 89.5% to 100% and specificity from 96.2% to 100%, which is consistent with the general observation that RT-LAMP assays show a lower sensitivity compared to RT-PCR^{49,78–80}.

A significant positive correlation between RT-LAMP Tt values and RT-PCR Ct values was observed⁴⁹. This supports the use of RT-LAMP in addition to RT-PCR for clinical diagnostics, because both methods are based on the amplification of viral genetic material in the samples and apply an identical procedure for Ct/Tt value calculation⁴⁹.

For the isolation-free SARS-CoV-2 RT-LAMP assay a LoD of 100000 copies/ml or 95 copies per reaction was observed which is consistent with the data reported by others^{49,79,81}.

As shown by Larremore et al. using the dynamic pattern of viral load kinetics, effective SARS-CoV-2 population screening depends primarily on the frequency of testing and the speed of reporting^{49,78}. According to their study test sensitivity is secondary, which supports RT-LAMP assays as a useful alternative or addition to RT-PCR despite the higher LoD⁴⁹.

The isolation-free RT-LAMP protocol developed here showed a high reproducibility with no false classification over three runs with five replicates of a positive and negative sample pool⁴⁹. Little variability in terms of both intra- and inter-run variability was observed for the Tt values of positive samples⁴⁹.

In summary, the RT-LAMP has proved to be a fast and efficient alternative for SARS-CoV-2 screening in clinical diagnostic laboratories⁴⁹. Minor limitations include a reduced analytical and diagnostic sensitivity⁴⁹. In addition, RT-LAMP assays carry a high risk of carry-over contamination due to a highly efficient reaction resulting in large amounts of amplification products, which can lead to false positive results^{26,49}. However, by applying good and careful laboratory practice this risk can be minimized⁴⁹.

5. Summary

Since the first Nanopore sequencing devices entered the market in 2014, this third-generation sequencing technique has developed into a powerful tool for molecular biological research. The unique concept of nucleic acid sequencing by using biological nanopores to convert the physical and chemical properties of specific nucleobases into measurable electrical signals has not only the advantage of real-time sequencing and the generation of long reads but it is also a fast and cost-efficient way of sequencing. Another unique feature of Nanopore sequencing is the portability of some sequencing devices which enables an application in remote areas where a dedicated laboratory infrastructure is not available. Sequencing accuracy, which was a major drawback in the early days of Nanopore sequencing, has improved a lot over time due to continuous innovations of the underlying nanopores, the sequencing chemistries and the bioinformatic data analysis tools. However, compared to other modern NGS techniques the sequencing error is still high which has limited its application to laboratory medicine so far. Nevertheless, by controlling this error, Nanopore sequencing would become attractive for application to diagnostic laboratory medicine due to the above-mentioned advantages.

The aim of this work was to establish and evaluate current Nanopore sequencing as a modern sequencing technique in clinical laboratory medicine especially in the field of human genetics and infectious disease epidemiology.

The genetic diagnosis of Familial Mediterranean fever by single nucleotide polymorphism genotyping was selected as a reference, since this is a well-established diagnostic procedure in human genetics with sufficient samples available. Establishing Nanopore sequencing for this diagnostic issue required the development and establishment of a novel bioinformatic data analysis pipeline and a sequencing protocol to sequence selected regions of the human *MEFV* gene. A comprehensive method comparison to the gold-standard method, Sanger sequencing, using clinical samples revealed a perfect agreement between both methods. This renders current Nanopore sequencing in principle suitable for at least SNP genotyping in human genetics.

The bioinformatic analysis of the sequencing data is one of the most challenging parts in a Nanopore sequencing experiment, especially for an application in clinical laboratory medicine where dedicated bioinformaticians are not regularly available. Therefore, further research was performed to support an application of Nanopore sequencing in clinical laboratory medicine by evaluating different bioinformatic tools for sequence alignment using real data sets.

After Nanopore sequencing showed a robust performance in SNP genotyping a SARS-CoV-2 whole genome sequencing protocol was established on a MinION sequencing device to support molecular biological diagnostics of SARS-CoV-2 during the global pandemic. This approach was further applied in a research project to investigate host-virus interaction by combining viral genetic data, serological data and clinical data for the first time. The results show diverse humoral immune responses to SARS-CoV-2, that appear to be influenced by disease severity, age and obesity. In addition, it was observed that even small viral genetic changes may influence the clinical presentation of COVID-19.

Overall, Nanopore sequencing turned out to be well suited to viral whole genome sequencing. As well as the identification of viral genetic alterations, the generated sequencing data were useful in supporting epidemiological research by phylogenetic interpretation. In addition to this research application, the established SARS-CoV-2 whole genome sequencing protocol was further used for SARS-CoV-2 molecular biological surveillance in clinical laboratory diagnostics. In this way, it was possible to observe and document the viral genetic development over the course of one year.

Due to shortages of PCR reagents and lab supplies during the SARS-CoV-2 pandemic, a novel RT-LAMP assay for the detection of this viral pathogen was established to supplement RT-PCR. A comprehensive method comparison to conventional RT-PCR assays showed that RT-LAMP is a fast and efficient alternative for SARS-CoV-2 screening in clinical diagnostic laboratories.

The development and establishment of modern and innovative techniques, including Nanopore sequencing and LAMP, in clinical diagnostic laboratories are required to expand the capabilities of modern laboratory medicine and to face new diagnostic challenges which arise from continuous progress in medical research. As outlined in this work, the development of standardized workflows and the comprehensive validation of an assay are essential steps prior to an application in clinical diagnostics. Furthermore, by presenting advancements of sequencing and bioinformatic workflows with the focus on an application in clinical diagnostics the results of this thesis may pave the way for a broader application of Nanopore sequencing in laboratory medicine in the near future.

References

1. Liu, L. *et al.* Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 251364 (2012).
2. Schmidt, J., Blessing, F., Fimpler, L. & Wenzel, F. Nanopore Sequencing in a Clinical Routine Laboratory: Challenges and Opportunities. *Clin. Lab.* **66**, (2020).
3. Loose, M. W. The potential impact of nanopore sequencing on human genetics. *Hum. Mol. Genet.* **26**, 202–207 (2017).
4. Kono, N. & Arakawa, K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev. Growth Differ.* **61**, 316–326 (2019).
5. Feng, Y., Zhang, Y., Ying, C., Wang, D. & Du, C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* **13**, 4–16 (2015).
6. *Nanopore sequencing: An introduction.* (World Scientific, 2018).
7. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* **14**, 265–279 (2016).
8. Rang, F. J., Kloosterman, W. P. & Ridder, J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **19**, 90 (2018).
9. Liu, Z., Wang, Y., Deng, T. & Chen, Q. Solid-State Nanopore-Based DNA Sequencing Technology. *J. Nanomater.* **2016**, 1–13 (2016).
10. Haque, F. *et al.* Single pore translocation of folded, double-stranded, and tetra-stranded DNA through channel of bacteriophage phi29 DNA packaging motor. *Biomaterials* **53**, 744–752 (2015).
11. Manara, R. M. A., Wallace, E. J. & Khalid, S. DNA sequencing with MspA: Molecular Dynamics simulations reveal free-energy differences between sequencing and non-sequencing mutants. *Sci. Rep.* **5**, 12783 (2015).
12. Delahaye, C. & Nicolas, J. Sequencing DNA with nanopores: Troubles and biases. *PLoS One* **16**, 0257521 (2021).

13. Jiao, X. *et al.* A Benchmark Study on Error Assessment and Quality Control of CCS Reads Derived from the PacBio RS. *J. Data Mining Genomics Proteomics* **4**, (2013).
14. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR genom. bioinform.* **3**, 019 (2021).
15. Wang, X. V., Blades, N., Ding, J., Sultana, R. & Parmigiani, G. Estimation of sequencing error rates in short reads. *BMC Bioinform.* **13**, 185 (2012).
16. Weirather, J. L. *et al.* Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000research* **6**, (2017).
17. Magi, A., Semeraro, R., Mingrino, A., Giusti, B. & D'Aurizio, R. Nanopore sequencing data analysis: state of the art, applications and challenges. *Brief. Bioinformatics* **19**, 1256–1272 (2018).
18. Becht, C., Schmidt, J., Blessing, F. & Wenzel, F. Comparative analysis of alignment tools for application on Nanopore sequencing data. *Curr. Dir. Biomed. Eng.* **7**, 831–834 (2021).
19. Tyler, A. D. *et al.* Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci. Rep.* **8**, 10931 (2018).
20. Hoenen, T. *et al.* Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerging Infect. Dis.* **22**, 331–4 (2016).
21. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
22. Patel, A. *et al.* MinION rapid sequencing: Review of potential applications in neurosurgery. *Surg. Neurol. Int.* **9**, 157 (2018).
23. Orsini, P. *et al.* Design and MinION testing of a nanopore targeted gene sequencing panel for chronic lymphocytic leukemia. *Sci. Rep.* **8**, 11798 (2018).
24. Bowden, R. *et al.* Sequencing of human genomes with nanopore technology. *Nat. Commun.* **10**, 1869 (2019).

25. Leija-Salazar, M. *et al.* Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. *Mol. Genet. Genomic Med.* **7**, 564 (2019).
26. Wong, Y.-P., Othman, S., Lau, Y.-L., Radu, S. & Chee, H.-Y. Loop-mediated isothermal amplification (LAMP): a versatile technique for detection of micro-organisms. *J. Appl. Microbiol.* **124**, 626–643 (2018).
27. Becherer, L. *et al.* Loop-mediated isothermal amplification (LAMP) – review and classification of methods for sequence-specific detection. *Anal. Methods* **12**, 717–746 (2020).
28. Paik, I. *et al.* Improved Bst DNA Polymerase Variants Derived via a Machine Learning Approach. *Biochemistry* (2021) doi:10.1021/acs.biochem.1c00451.
29. Özen, S., Batu, E. D. & Demir, S. Familial Mediterranean Fever: Recent Developments in Pathogenesis and New Recommendations for Management. *Front. Immunol.* **8**, 253 (2017).
30. Sari, I., Birlik, M. & Kasifoglu, T. Familial Mediterranean fever: An updated review. *Eur. J. Rheumatol.* **1**, 21–33 (2014).
31. Schmidt, J. *et al.* Genotyping of familial Mediterranean fever gene (MEFV)-Single nucleotide polymorphism-Comparison of Nanopore with conventional Sanger sequencing. *PLoS One* **17**, e0265622 (2022).
32. Shohat, M. & Halpern, G. J. Familial Mediterranean fever--a review. *Genet. Med.* **13**, 487–98 (2011).
33. Ozdogan, H. & Ugurlu, S. Familial Mediterranean Fever. *Presse Medicale* **48**, 61–76 (2019).
34. Grandemange, S., Aksentijevich, I., Jeru, I., Gul, A. & Touitou, I. The regulation of MEFV expression and its role in health and familial Mediterranean fever. *Genes Immun.* **12**, 497–503 (2011).
35. Shinar, Y. *et al.* Guidelines for the genetic diagnosis of hereditary recurrent fevers. *Ann. Rheum. Dis.* **71**, (2012).
36. Demir, F. *et al.* Genetic panel screening in patients with clinically unclassified systemic autoinflammatory diseases. *Clin. Rheumatol.* **39**, (2020).

37. Schmidt, J. *et al.* Serological and viral genetic features of patients with COVID-19 in a selected German patient cohort-correlation with disease characteristics. *GeroScience* **43**, 2249–2264 (2021).
38. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
39. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170 (2021).
40. Harrison, A. G., Lin, T. & Wang, P. Mechanisms of SARS-CoV-2 transmission and pathogenesis. *Trends Immunol.* **41**, 1100–1115 (2020).
41. Cao, C. *et al.* The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.* **12**, 3917 (2021).
42. Haque, S. M., Ashwaq, O., Sarief, A. & Azad John Mohamed, Abdul Kalam. A comprehensive review about SARS-CoV-2. *Future Virol.* **15**, 625–648 (2020).
43. Petersen, E. *et al.* Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* **20**, 238–244 (2020).
44. Schmidt, J., Blessing, F. & Gürtler, L. [SARS-CoV-2 vaccines and reaction of the immune system. Can the epidemic spread of the virus be prevented by vaccination?]. *Dtsch. Med. Wochenschr.* **146**, 1085–1090 (2021).
45. Poland, G. A., Ovsyannikova, I. G. & Kennedy, R. B. SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. *Lancet* **396**, 1595–1606 (2020).
46. Gao, J. & Quan, L. Current Status of Diagnostic Testing for SARS-CoV-2 Infection and Future Developments: A Review. *Med. Sci. Monit.* **26**, (2020).
47. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
48. Tracking SARS-CoV-2 variants. <https://www.who.int/health-topics/typhoid/tracking-sars-cov-2-variants> (2022).

49. Schmidt, J. *et al.* A semi-automated, isolation-free, high-throughput SARS-CoV-2 reverse transcriptase (RT) loop-mediated isothermal amplification (LAMP) test. *Sci. Rep.* **11**, 21385 (2021).
50. Neuenschwander, S. M. *et al.* A Sample-to-Report Solution for Taxonomic Identification of Cultured Bacteria in the Clinical Setting Based on Nanopore Sequencing. *J. Clin. Microbiol.* **58**, (2020).
51. Crossley, B. M. *et al.* Nanopore sequencing as a rapid tool for identification and pathotyping of avian influenza A viruses. *J. Vet. Diagn. Invest.* **33**, (2021).
52. Watson, C. M. *et al.* Assessing the utility of long-read nanopore sequencing for rapid and efficient characterization of mobile element insertions. *Lab. Investig.* **101**, (2021).
53. Watson, C. M. *et al.* Long-read nanopore sequencing enables accurate confirmation of a recurrent PMS2 insertion-deletion variant located in a region of complex genomic architecture. *Cancer Genet.* **256–257**, (2021).
54. Thirunavukarasu, D. *et al.* Oncogene Concatenated Enriched Amplicon Nanopore Sequencing for rapid, accurate, and affordable somatic mutation detection. *Genome Biol.* **22**, 1–17 (2021).
55. Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* **13**, 341 (2012).
56. Sović, I. *et al.* Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat. Commun.* **7**, 11307 (2016).
57. Mishra, A. *et al.* Mutation landscape of SARS-CoV-2 reveals five mutually exclusive clusters of leading and trailing single nucleotide substitutions. *bioRxiv* (2020) doi:10.1101/2020.05.07.082768.
58. Kaushal, N. *et al.* Mutational Frequencies of SARS-CoV-2 Genome during the Beginning Months of the Outbreak in USA. *Pathogens* **9**, (2020).
59. Callaway, E. The coronavirus is mutating — does it matter? *Nature* **585**, 174–177 (2020).
60. Yin, C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics* **112**, 3588–3596 (2020).

61. Yang, H.-C. *et al.* Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30679–30686 (2020).
62. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
63. Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y. & Kiyotani, K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J. Hum. Genet.* **65**, 1075–1082 (2020).
64. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812–827 (2020).
65. Klein, S. L. *et al.* Sex, age, and hospitalization drive antibody responses in a COVID-19 convalescent plasma donor population. *J. Clin. Investig.* **130**, 6141–6150 (2020).
66. Frasca, D. *et al.* Effects of obesity on serum levels of SARS-CoV-2-specific antibodies in COVID-19 patients. *medRxiv* (2020) doi:10.1101/2020.12.18.20248483.
67. Korakas, E. *et al.* Obesity and COVID-19: immune and metabolic derangement as a possible link to adverse clinical outcomes. *Am. J. Physiol. Endocrinol. Metab.* **319**, 105–109 (2020).
68. Racine-Brzostek, S. E. *et al.* Postconvalescent SARS-CoV-2 IgG and Neutralizing Antibodies are Elevated in Individuals with Poor Metabolic Health. *J. Clin. Endocrinol. Metab.* **106**, 2025–2034 (2021).
69. Booth, A. *et al.* Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLoS One* **16**, 0247461 (2021).
70. Herroelen, P. H., Martens, G. A., Smet, D., Swaerts, K. & Decavele, A.-S. Humoral Immune Response to SARS-CoV-2. *Am. J. Clin. Pathol.* **154**, 610–619 (2020).
71. Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res.* **149**, 58–74 (2018).
72. Wang, Y. *et al.* Kinetics of viral load and antibody response in relation to COVID-19 severity. *J. Clin. Investig.* **130**, 5235–5244 (2020).

73. Xu, X. *et al.* Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res.* **31**, 7117–30 (2003).
74. McBride, R., van Zyl, M. & Fielding, B. C. The Coronavirus Nucleocapsid Is a Multifunctional Protein. *Viruses* **6**, 2991–3018 (2014).
75. Wang, P. *et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).
76. Singanayagam, A. *et al.* Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. *Lancet Infect. Dis.* **22**, 183–195 (2022).
77. Garcia-Beltran, W. F. *et al.* mRNA-based COVID-19 vaccine boosters induce neutralizing immunity against SARS-CoV-2 Omicron variant. *Cell* **185**, 457-466.e4 (2022).
78. Larremore, D. B. *et al.* Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci. Adv.* **7**, (2021).
79. Dao Thi, V. L. *et al.* A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Sci. Transl. Med.* **12**, (2020).
80. Rödel, J. *et al.* Use of the variplex™ SARS-CoV-2 RT-LAMP as a rapid molecular assay to complement RT-PCR for COVID-19 diagnosis. *J. Clin. Virol.* **132**, 104616 (2020).
81. Yang, Q. *et al.* Saliva TwoStep for rapid detection of asymptomatic SARS-CoV-2 carriers. *medRxiv* (2021) doi:10.1101/2020.07.16.20150250.

Documentation / List of Publications

The listed publications are part of this cumulative dissertation. As a first author I was responsible for the development, establishment and evaluation of Nanopore sequencing and RT-LAMP at the Department of Molecular Biology of the clinical diagnostic laboratory MVZ Laborärzte Singen. This included experimental and study design, literature search, care of the study participants, execution of the experiments, development of bioinformatic data analysis pipelines, data acquisition and analysis, as well as manuscript writing and correction in the review process of the individual journals.

The co-authorship in the publication Becht, Schmidt et al. (2021) is based on the generation and preparation of the data sets, data analysis, manuscript writing and correction in the review process as well as overall supervision of this project.

- **Schmidt, J.**, Blessing, F., Fimpler, L. & Wenzel, F. Nanopore Sequencing in a Clinical Routine Laboratory: Challenges and Opportunities. *Clin. Lab.* 66, (2020). DOI: 10.7754/Clin.Lab.2019.191114
© 2020. Clinical Laboratory Publications GmbH, reprinted with permission
- **Schmidt, J.** et al. Serological and viral genetic features of patients with COVID-19 in a selected German patient cohort-correlation with disease characteristics. *GeroScience* 43, 2249–2264 (2021). DOI: 10.1007/s11357-021-00443-w
published under a CC-BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>)
- **Schmidt, J.**, Blessing, F. & Gürtler, L. SARS-CoV-2 vaccines and reaction of the immune system. Can the epidemic spread of the virus be prevented by vaccination? *Dtsch. Med. Wochenschr.* 146, 1085–1090 (2021). DOI: 10.1055/a-1550-0001
© 2021. Thieme, reprinted with permission
- **Schmidt, J.** et al. A semi-automated, isolation-free, high-throughput SARS-CoV-2 reverse transcriptase (RT) loop-mediated isothermal amplification (LAMP) test. *Sci. Rep.* 11, 21385 (2021). DOI: 10.1038/s41598-021-00827-0
published under a CC-BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>)

- **Schmidt, J.** et al. Genotyping of familial Mediterranean fever gene (MEFV)-Single nucleotide polymorphism-Comparison of Nanopore with conventional Sanger sequencing. *PLoS One* 17, e0265622 (2022). DOI: 10.1371/journal.pone.0265622
published under a CC-BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>)
- Becht, C., **Schmidt, J.**, Blessing, F. & Wenzel, F. Comparative analysis of alignment tools for application on Nanopore sequencing data. *Curr. Dir. Biomed. Eng.* 7, 831–834 (2021). DOI: 10.1515/cdbme-2021-2212
published under a CC-BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>)

Other contribution

- 26.09.2019 Poster at the 16th Annual Congress of the Deutsche Gesellschaft für Klinische Chemie und Laboratoriumsmedizin in Magdeburg “Development of a laboratory protocol for the detection of MEFV mutations by amplicon sequencing on a MinION sequencer.”
- 22.11.2019 Congress presentation at the 38th Annual Congress of the Deutsche Gesellschaft für klinische Mikrozirkulation und Hämorheologie (DGKMH) in Braunschweig “Identifikation von Mutationen im MEFV-Gen mittels Amplicon-Sequencing auf dem MinION Sequencer”
- 06.11.2020 Congress presentation at the 39th Annual Congress of the DGKMH in Hannover (online event) “Tiling Amplicon Sequencing of SARS-CoV-2 on a Nanopore Sequencing Device: Applicability in a routine lab?”
- 05.10.2021 Poster at the 55th Annual Conference of the German Society for Biomedical Engineering (Co-authorship with C. Becht) “Comparative analysis of alignment tools for application on Nanopore sequencing data”
- 07.05.2022 Congress presentation at the Warter Symposium for Pediatrics 2022 “Molekularbiologische und serologische Diagnostik von SARS-CoV-2 – Aktuelle Möglichkeiten der modernen Labormedizin”

- 20.05.2022 Congress presentation at the 40th Annual Congress of the DGKMH in Cottbus-Senftenberg “SARS-CoV-2 Detektion mittels Loop-mediated isothermal amplification (LAMP)”
- 23.06.2022 Congress presentation at the 2nd Micro Med Tec Symposium in Villingen-Schwenningen “Nanopore-Sequencing im klinischen Routine-Labor”
- 16.07.2022 Congress presentation at the 21st annual Symposium of the MVZ Laborärzte Singen in Überlingen “Differenzierte Immunantwort nach Wildvirusinfektion bei SARS-CoV-2 und Virusvariantendiagnostik mittels Vollgenomsequenzierung”

Acknowledgements

This work would not have been possible without the support and encouragement of many people, whom I would like to thank wholeheartedly below.

I would like to express my deepest appreciation to Prof. Dr. Dirk Roggenbuck for agreeing to supervise my doctorate, for helpful scientific advice and discussions as well as for his extremely fast assistance with any question.

In the same way I would like to thank Prof. Dr. Frithjof Blessing for the immense professional and personal support and the great confidence in my abilities. He kept me very motivated all the time and without him my dissertation would have been impossible.

Thanks should also go to Prof. Dr. Meike Burger for her willingness and effort to assess my dissertation.

Sincere thanks belong to Dr. Stefan Rödiger and Prof. Dr. Peter Schierack from the research group “Multiparameter Diagnostics” for their critical comments and support with manuscript writing.

I am also thankful to Prof. Dr. Folker Wenzel for critical discussions and for his guidance during the initial phase of the doctorate.

Correspondingly, I would like to thank Dr. Dagmar Stephan from the Dean’s Office of Faculty 2 for her great support concerning organisational and formal questions.

Special thanks should go to Prof. Dr. Josef Blessing for nightly discussions about the important issues in medical science and for supplying me with healthy snacks during long working days.

I would like to extend my sincere thanks to all my fellow lab members from the MVZ Laborärzte Singen who supported me during the whole process of the doctorate and provided a calm and productive working environment.

I am also grateful to Heather Robertson, PhD and Alastair Robertson, PhD for linguistic proof-reading this work.

Finally, I deeply thank my parents, my brother and my friends for all their support.

Thank you all for your help!

Eidesstattliche Erklärung

Hiermit erkläre ich, Jonas Paul Christian Schmidt, an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Jonas Paul Christian Schmidt

Singen, den 29.07.2022

Appendix

Copies of the publications are attached to this work.

ORIGINAL ARTICLE

Nanopore Sequencing in a Clinical Routine Laboratory: Challenges and Opportunities

Jonas Schmidt^{1,2}, Frithjof Blessing^{1,2}, Lea Fimpler², Folker Wenzel^{1,2}

¹Institute for Laboratory Medicine, Singen, Germany

²Furtwangen University, Faculty of Medical and Life Sciences, Villingen-Schwenningen, Germany

SUMMARY

Background: About forty-five years ago the advent of Sanger sequencing (Sanger and Coulson 1975) was revolutionary as it allowed deciphering of complete genome sequences. A second revolution came when next-generation sequencing (NGS) technologies accelerated and cheapened genome sequencing. Recently, third generation/long-read sequencing methods have appeared, which can directly detect epigenetic modifications on native DNA and allow whole-transcript sequencing without the need for assembly. Nanopore sequencing is one of these third-generation approaches, enabling a single molecule of DNA or RNA to be sequenced in real-time without the need for PCR amplification or chemical labelling of the sample. It works by monitoring changes to an electrical current as nucleic acids are passed through protein or synthetic nanopores.

Methods: A literature search was performed in order to collect and summarize current information about the methodological aspects of nanopore sequencing as well as some application examples.

Results: The review describes concisely and comprehensibly the technical aspects of nanopore sequencing and stresses the advantages and disadvantages of this technique thereby also giving examples of their potential applications in the clinical routine laboratory as are rapid identification of viral pathogens, monitoring Ebola, environmental and food safety monitoring, human and plant genome sequencing, monitoring of antibiotic resistance, and other applications.

Conclusions: It is a useful incitation for such ones being permanently in search of upgrading their laboratory. (Clin. Lab. 2020;66:1097-1104. DOI: 10.7754/Clin.Lab.2019.191114)

Correspondence:

Jonas Schmidt
Institute for Laboratory Medicine
Virchowstraße 10c
78224 Singen (Hohentwiel)
Germany
Phone: +49 07731/99560
Fax: +49 07731/9826831
Email: jonas.schmidt@labor-blessing.de

KEY WORDS

nanopore sequencing, clinical routine laboratory, operating principle, sample preparation, data analysis, research examples, advantages, disadvantages

INTRODUCTION

In the last few years, besides Pacific Bioscience's single-molecule real-time (SMRT) sequencing, nanopore sequencing has become an important representative of the third-generation sequencing technologies [1]. The technique was commercialized by Oxford Nanopore Technologies and entered the market in 2014 [2]. Until today many technical improvements have been made but the technique is still under strong development with a continuous increase of sequencing throughput and ac-

curacy [2,3].

In general, compared to conventional next generation sequencing (NGS) techniques the third-generation shows some important differences [1]. The most important difference is that a polymerase chain reaction (PCR) is not mandatory prior to nanopore sequencing [1]. This speeds up the sample preparation and reduces the possible bias which can be induced by using PCR [1]. A second important difference is that the signal produced during the sequencing process is captured in real-time [1,3].

The rapid technical progress in the field of NGS techniques had an impact on diagnostic procedures in clinical laboratories as well [4]. These days, besides conventional Sanger sequencing, NGS techniques are also frequently used for the diagnosis of human genetic disorders [4]. Additionally, other fields of application are the diagnostics of infectious diseases and immune system related disorders as well as non-invasive prenatal screening [4]. Furthermore, NGS is also used to investigate the clinically relevant properties of somatic cancers related to therapeutic decision making [4].

Unlike research institutions, clinical routine diagnostic laboratories have slightly different demands regarding NGS techniques. First, the applied techniques should be robust with a great reproducibility. Regarding high sample throughput, it is also important that the sample preparation is easy to perform, and the sequencing technique itself is considerably fast. Also, regarding the increasing degree of automation in clinical routine laboratories, adaption to a certain degree of automation should be possible. Finally, the acquisition costs of the sequencers as well as the running costs for consumables and reagents are important points to take into consideration. Like any other sequencing techniques, nanopore sequencing has advantages and disadvantages. They must be considered when thinking about the possible application of this technique for a certain analytical question. Therefore, in the following section we would like to give a brief overview of the technology and show some aspects which need to be taken into account, especially when experimenting with nanopore sequencing in a clinical routine laboratory.

Methodological aspects

Basic principle

The core component in a nanopore sequencing experiment is a nanometer-sized pore which is embedded into a membrane [5]. This membrane separates a reservoir containing an electrolyte solution (e.g., KCl) into two compartments (cis and trans chamber) [5,6]. Both compartments contain one electrode each [5]. If a voltage bias (e.g., ~120 - 180 mV) is applied to the electrodes, the electrolytes in solution are driven electrophoretically through the pore which generates an ionic current signal [5]. Any negatively-charged DNA or RNA molecules which are present in the cis chamber will also pass through the pore [5]. Due to the small diameter of the pore single stranded molecules will partially block the

pore while passing through it which leads to a measurable interruption of the current signal (Figure 1) [5]. Finally, by using complex computational tools the amplitude and duration of the transient current blockades can be used to identify the nucleotide sequence because the shape of the signal is directly related to a specific group of nucleobases [5-7]. Hence, the main function of a nanopore in a nanopore sequencing experiment is to transduce the physical and chemical properties of specific nucleobases into measurable ionic currents, which can be used to identify the nucleotide sequence [6].

With the currently available technology, it is not possible to sense single nucleobases and therefore the signal is detected for a specific group of nucleobases [6]. On currently available Nanopore sequencing platforms, this is normally done in groups of six nucleobases for DNA sequencing and groups of five for RNA sequencing [6]. Another important problem which has to be addressed when talking about nanopore sequencing experiments is the translocation speed of the target molecules through the pore [6]. When the polynucleotides are driven through the pore solely by the voltage bias, one nucleobase takes less than 10 μ s to pass through the pore [6]. This time interval is far too short to allow a suitable detection and differentiation between A, T, C, and G because the signal is masked by the nanopore's electrical background noise [6]. To solve this problem motor enzymes like polymerases or helicases are used in order to slow down the movement of the polynucleotide strand through the pore [6]. Additionally, they work like a ratchet so that each nucleobase is held in the pore for a few milliseconds [6]. Currently, commercially available sequencing platforms mainly use helicase motor proteins [6].

Combined with a special adapter they are ligated during library preparation to the terminal nucleotide of the dsDNA strand which should be sequenced [6]. When starting the sequencing process, the adapter is translocated through the pore together with the dsDNA to which it is attached [6]. This brings the helicase close to the pore where it stays during the sequencing process [6]. Upon contact with the pore, the helicase is activated and unzips the dsDNA into two single strands of which one is translocated through the pore in a stepwise fashion while it is separated from its complementary strand [6]. With the help of the helicase, the translocation speed can be slowed down to a continuous stream of approximately 450 bases per second [6,8].

Nanopore sequencing was mainly commercialized by the British company Oxford Nanopore Technologies (ONT), which announced its first sequencing device in early 2012 [2]. Since that time, the product line has increased significantly with different platforms available. In short there are three core platforms. First, the MinION which is a hand-held sequencing platform [2]. Second, the GridION which is a larger system and third, the PromethION which is the largest currently available nanopore sequencer [2,6]. All of these platforms are based on the same sequencing chemistry but differ in through-

put and data output [2,6]. In general, ONT uses a nanopore sensor array which consists of a silicon-based tri-block co-polymer membrane in which many pores are embedded [6]. This sensor array is placed together with other mandatory parts on a so called flow cell which can be connected to the sequencing device [6]. The sensor array of a MinION flow cell, for example, is designed to hold a maximum of 2,048 individual nanopores [2,6]. However, the actual number of working nanopores on the sensor array is a little bit lower and differs from flow cell to flow cell. The large number of pores on the array makes it possible to sequence many different DNA strands in parallel which significantly increases the sequencing throughput and the amount of generated data [6].

With regard to the application of nanopore sequencing in a clinical routine laboratory, it is important that the MinION sequencer as well as a starter kit containing most reagents can be purchased with low initial capital costs. Also, due to its small size, no special laboratory infrastructure is required and so in most cases it is not necessary to modify the existing infrastructure. With these advantages in mind, the MinION is a good option to become familiar with the technique and to perform smaller sequencing experiments in laboratories that are not specialized exclusively in genetic diagnostics. However, if larger sequencing experiments are planned it might be necessary to scale up to a GridION or even a PromethION.

Different types of nanopores

Basically, there are two different categories of nanopores which can be used for nanopore sequencing [5]. These categories include biological as well as solid-state nanopores [5,6,9]. For DNA sequencing up to now biological pores have been more commonly used [5,6]. They consist of transmembrane protein channels which are inserted into a suitable cellular or artificial substrate [5,6]. In most cases, these transmembrane proteins can be isolated from bacteria or even bacteriophages [5]. Examples for well-studied biological nanopores are α -hemolysin which is an exotoxin secreted by the bacterium *Staphylococcus aureus*, *Mycobacterium smegmatis* porin A (MspA) as well as a pore which is isolated from the bacteriophage phi29 [5,6,9-11]. Important advantages of biological pores involve their highly-reproducible size and structure as well as the possibility to modify them easily by using modern molecular biological techniques [5]. For example, mutating the nucleotide sequence in order to change the amino acid residue at a specific site can be used to modify the structure of a biological pore which can also change the physical and chemical characteristics of the pore [5].

As a practical example, the current MinION sequencer can be used in combination with ONTs R9 pore generation [6]. This pore is a mutant of the CsgG lipoprotein which is isolated from *E. coli* [6]. On the molecular level, the CsgG pore is composed of nine identical subunits that form a beta-barrel pore [6]. *In vivo*, the protein

serves as a transporter for transport polypeptides across the bacterial membrane [6]. To apply it for DNA sequencing, the pore was engineered to allow the translocation of DNA [6].

Due to their origin, the biological pores also show some important disadvantages [9]. They include special environmental demands regarding temperature, electrolyte concentration and pH as well as a reduced stability which makes them easier to break down [5,9]. To overcome such problems different kinds of synthetic nanopores have been developed [5]. They are fabricated in, for example, silicon nitride, silicon dioxide, aluminum oxide, boron nitride or graphene by using advanced fabrication processes for working with such materials [5,9]. The solid-state pores are characterized by an increased robustness and durability, superior mechanical, chemical, and thermal characteristics as well as an easier fabrication process regarding shape and size [9]. However, the field of solid-state nanopore based DNA sequencing is still in an early phase and there might be further developments in the next few years [9]. This is also true for the idea of combining biological and solid-state pores to form hybrid pores and use the advantages of both sides [5,9].

Sample preparation

As with any other analytical technique, sample preparation prior to sequencing also plays a key role. The first challenge is the DNA extraction. When looking for a suitable DNA extraction method it is important to keep in mind that one of the biggest advantages of Nanopore sequencing compared to conventional NGS techniques (e.g., Illumina sequencing) is its ability to generate ultra-long reads (up to megabase pairs) [6]. The maximum read length which can be achieved during a run mainly depends on the DNA itself as well as on the library preparation technique [6]. In general, there are multiple different strategies available which all have their advantages and disadvantages [6]. Examples are spin column based methods, gravity-flow column based methods, magnetic beads as well as phenol-chloroform extraction and dialysis [6]. When it comes to the choice of a suitable extraction method it is important to think about the minimum read length which is required to achieve the goal of the sequencing experiment [6]. Further, factors like the type of sample, sample input, and cost of the extraction method should be taken into consideration [6].

The easiest way to begin with might be to use a commercial DNA extraction kit [6]. Although they are slightly more expensive than a self-developed, manual method, kits offer useful results in combination with a simple workflow and no developmental efforts [6]. Therefore, with regard to getting started with nanopore sequencing in a clinical routine laboratory, depending on the analytical question, DNA extraction kits might be the easiest option. If it turns out that the performance of such kits is not sufficient, more effort must be put into developing an in-house DNA extraction method that

Table 1. Summary of the advantages and disadvantages of nanopore sequencing with special regard to a possible application under routine laboratory conditions.

Advantages	Disadvantages
Long reads possible (up to 2 Mb)	Reduced single read accuracy
Sequencing in real-time	Complex and time-consuming data analysis
Simple library preparation	High demands on IT infrastructure
PCR not mandatory	So far, limited experience with the method under routine conditions
Low acquisition costs	
No highly specialized laboratory infrastructure required	

fits to the specific application [6].

After DNA extraction, the next important point is to assess the DNA quality since high quality DNA is needed to perform a successful sequencing experiment [6]. DNA quality measures which can be recorded are fragment size, absorbance spectrometry values as well as fluorometric quantification [6]. However, this means that additional laboratory equipment is necessary. Size assessment can be done on specialized gel electrophoresis instruments or by conventional gel electrophoresis [6]. To obtain some information about DNA purity by absorbance ratio measurements, a nanodrop photometer or an equivalent spectral photometer is necessary [6]. Last but not least, a fluorescence spectrophotometer is needed for the reliable quantification of the DNA [6]. Even though no highly specialized laboratory infrastructure is required to get started with nanopore sequencing the above-mentioned points represent some mandatory equipment. This is also important for clinical routine laboratories which want to try nanopore sequencing, because, in addition to the sequencing equipment, the purchase of additional instruments for quality assessment might also be necessary.

Finally, it is important to store the extracted DNA in a suitable way [6]. To keep the long fragments and avoid additional shearing the best way is to store the DNA in TE-buffer at 5°C because freezing will cause physical shearing [6]. Also, while handling the DNA, it is important to introduce as little as possible shearing forces to preserve long fragments [6].

After DNA extraction, a sequencing library which can be loaded on the sequencer must be prepared. ONT offers many different kits for this purpose which are based on slightly different principles [6]. Probably the most common principle is the ligation based library preparation which can be done in approximately 60 minutes [6]. In the first step of this procedure the ends of the

DNA fragments are turned into blunt ends by using a combination of polymerases and exonucleases [6]. Afterwards, a “dA-tail” is created by attaching a single deoxy-adenosine to the 3’ ends of the DNA fragments [6]. Finally, by using a T4 DNA ligase, sequencing adapters with complementary 3’ dT-tails are ligated to the fragments [6]. Besides other key components, these sequencing adapters also contain the helicase motor proteins, which are essential for the sequencing process [6]. After the library preparation procedure, the sample is ready to be loaded on a flow cell to start sequencing. Again, for researchers who want to apply nanopore sequencing under routine laboratory conditions, it is important that all currently available library preparation methods are designed to be as simple and fast as possible [6]. Therefore, library preparation can be performed without relying on expensive laboratory equipment or consumables [6]. This makes it easier to establish suitable methods in a clinical routine laboratory.

Last but not least, there are also different barcoding kits available for pooling multiple libraries and sequencing them in one run on a single flow cell [6]. This so-called multiplexing can be applied to use flow cells in a more efficient way when one sample is not enough to take advantage of the full capacity of a single flow cell.

Data analysis

Besides the sequencing procedure itself, the second fundamental part in every nanopore sequencing experiment is the data analysis. So far, there is a large number of algorithms and tools available which can be used for applications like base calling, quality control, data handling, read mapping, de novo assembly, and variant discovery [12].

In general, the ONT sequencing devices are controlled by a software named MinKNOW [12]. This software is essential for the selection of run parameters, data acquisition, real-time signal segmentation, and feed-back for experimental progression [12]. The raw read data is stored by MinKNOW for every read in FAST5 binary files [12]. During the base calling process, these raw signal files which describe the changes in the ionic current signal are used to identify the DNA sequence [6, 12]. This is done by comparing the ionic currents which are sampled at a frequency of 4 kHz with a lookup table containing known current variation patterns [6]. For this purpose machine learning approaches are applied [12]. Figure 2 shows the raw signal of a single read which is visualized as a so called “Squiggle” Plot. One example for a modern base caller is Guppy which is provided by ONT. Guppy has the special feature that the base calling process can be significantly accelerated by using a graphics processing unit (GPU). With this in mind, it is important that computing should be done on a workstation which is equipped with a suitable GPU that is also supported by the base caller.

After base calling, the reads can be aligned against a reference genome in order to identify differences [12]. Due to the long read length of nanopore reads this is a

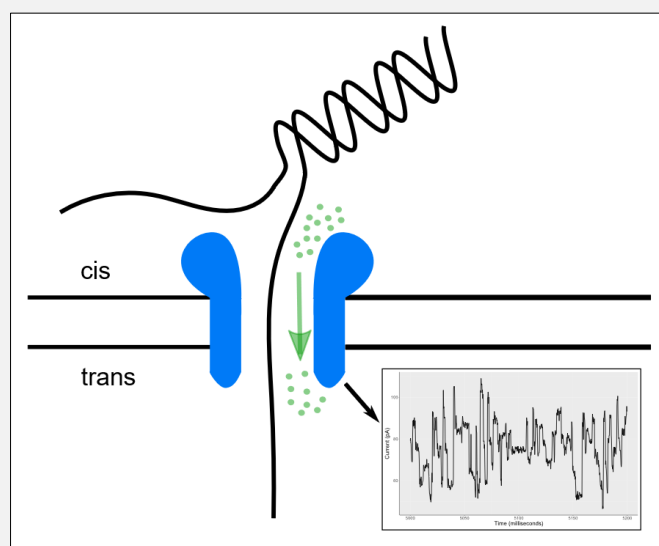


Figure 1. Schematic representation of the basic principle of nanopore sequencing. The nanopore (blue), which is embedded into a membrane, is partially blocked by a DNA single strand passing through it. This leads to an interruption of the ion current (green) and generates a measurable signal. From this signal the sequence can be identified by using complex computational tools.

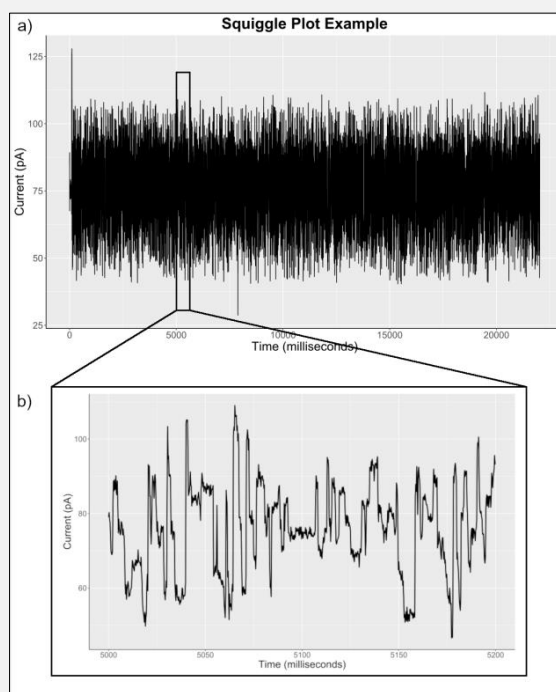


Figure 2. a) Raw signal of a single read which was generated using a MinION sequencer. The total read length after base calling was 8,685 bases. b) Time interval from 5,000 milliseconds to 5,200 milliseconds. The fluctuation of the current due to the DNA strand passing through the pore is clearly visible.

challenging task from a computational point of view [12]. Examples for modern alignment tools are LAST, BLASR, BWA-MEM, GraphMap, and NGMLR [12-17].

Variant discovery includes the identification of single-nucleotide variants (SNVs), small InDels as well as large deletions, duplications, inversions, and translocations [12]. This can be done by searching for differences between the aligned reads and the reference genome [12]. The identification of small variants by using Nanopore sequencing data is still a challenging task due to the sequencing error profile of nanopore reads [12]. However, examples of tools that can be used for this application are marginCaller, Nanopolish and Sniffles [12, 13, 18, 19].

With regard to the application of nanopore sequencing in a clinical routine laboratory, the bioinformatic analysis of the generated data might be the most challenging part because in most cases a dedicated bioinformatician will not be available. Therefore, the first task is to select the tools best suited for a specific research application from the variety of different tools available and combine them into a data analysis pipeline. Here, a search of literature might help in choosing the most suitable tools. Also, it is important to be familiar with Linux since most of the available tools run exclusively in a Linux environment.

Further, another important point to consider are the hardware requirements to perform such computations. Since even the smallest nanopore sequencing device can produce a significant amount of complex data in a 48 hours sequencing run, suitable workstations for data analysis are essential. This must also be considered when talking about the laboratory infrastructure necessary to perform nanopore sequencing experiments.

For smaller applications, a desktop workstation might be sufficient. For example, in our laboratory we use a workstation equipped with an Intel I7-7700K CPU, a NVIDIA GTX-1060 GPU, 32 GB of RAM, and a 1 TB SSD. Long-term data storage is outsourced to external hard drives. For larger sequencing experiments it might be necessary to use even more powerful systems or a computer cluster, if available.

Application examples

Since nanopore sequencing is a very new technique which has not been available on the market for a long time, it is currently not used for routine applications. However, there are a variety of clinical research applications where nanopore sequencing delivers promising results.

Tyler et al. evaluated the MinION sequencer for bacterial genomics and metagenomics especially with regard to the quality, yield, and accuracy of the generated data [20]. For this purpose a set of microbes was sequenced in replicates by two independent laboratories [20]. The results indicate that the technology is suitable in the context of genomic sequencing of microbial isolates important to public health [20]. With regard to the diag-

nostics of infectious diseases another important application example was published by Hoenen et al. They applied the MinION sequencer to sequence the genome of the Ebola virus in the field during the outbreak in Liberia [21]. The generated data was used to investigate the genetic development of the virus during the outbreak [21]. Furthermore, Quick et al. published a protocol for MinION sequencing of Zika and other virus genomes directly from clinical samples [22]. In this protocol, prior to sequencing on the MinION sequencer, an additional multiplex PCR is performed in order to enrich the target genome in the sample and obtain the highest possible coverage of the whole genome [22].

Besides the diagnostics of infectious diseases, there are also some applications in the field of cancer research. In a literature review, Patel et al. mentioned that the MinION sequencer is capable of providing critical diagnostic information for central nervous system (CNS) tumors within a single day [23]. Further, Orsini et al. successfully designed a customized gene panel assay to analyze five frequently mutated genes in chronic lymphocytic leukemia (CLL) by applying nanopore sequencing on the MinION sequencer [24]. In general, their concept offers an easy and affordable workflow for the analysis of prognostically relevant genes in CLL [24]. However, they also mention that further advances are required to improve the accuracy and enable the use in the clinical field [24].

Last but not least, there are studies where nanopore sequencing is applied to sequence the human genome and diagnose heritable diseases. Bowden et al. evaluated the potential of the MinION sequencer for routine whole genome sequencing (WGS) [25]. They sequenced two human genomes in multiple runs on the MinION device [25]. Although the consensus SNV-calling error rates from the nanopore sequencing experiment remained higher than those from short read methods, they could demonstrate substantial benefits of using nanopore sequencing for this application [25]. However, in the same study it was shown that the ongoing improvements to base-calling and SNV-calling processes need to continue in order for nanopore sequencing to be established as a primary method for clinical WGS [25]. Leija-Salazar et al. used nanopore sequencing on the MinION sequencer to evaluate the detection of missense mutations and other variants in the GBA gene [26]. Homozygous mutations in this gene cause Gaucher disease and the presence of heterozygous mutations is an important risk factor for the development of Parkinson's disease [26]. By using nanopore sequencing it was possible to detect missense mutations and an exonic deletion in this difficult gene [26].

Regarding the possible use of Nanopore sequencing products from ONT in clinical routine applications it must be mentioned that they are labeled as "for research use only". However, since the MinION sequencer is CE marked, after a comprehensive in-house validation, it should be possible to use such products for routine applications.

Although there are many published research applications with very promising results, up to now it is still not clear whether nanopore sequencing is suitable for an application in clinical routine laboratories. It is therefore important to perform method comparison studies where the technology is compared to standard methods (e.g., Sanger sequencing) under routine conditions. However, due to the fast technological progress, this is a challenging task because it is difficult to define a suitable starting point and results might quickly be outdated. If there are any major technical improvements during the duration of the method comparison, the whole study must be repeated in order to obtain reliable results.

CONCLUSION

Nanopore sequencing is a modern sequencing technique that can be used in a wide variety of different applications. Important advantages involve the possibility of generating long reads, real-time sequencing, easy and fast library preparation protocols without the need of PCR and comparably low acquisition costs. Disadvantages, on the other hand, are a reduced accuracy in comparison to other NGS techniques as well as the need for complex algorithms and large computational capacity for data analysis (see Table 1).

Due to low requirements regarding laboratory infrastructure, nanopore sequencing, especially on the MinION sequencer, is also promising for application in clinical routine laboratories which are not just specialized in molecular biological diagnostics. However, when thinking about conducting nanopore sequencing experiments in this type of laboratory, there are some additional points to take into consideration. First, it is important to select a suitable isolation technique to isolate the DNA from the samples. DNA quantity and purity have significant influence on the quality of the sequencing results. Therefore, it is crucial that techniques for DNA quantification and determination of the purity are available. The consequence is that in addition to the sequencing platform itself it might be necessary to purchase some additional instruments like a fluorometer and a spectral photometer in order to perform an accurate quality control of the samples prior and during library preparation. The second important point to think about is the data analysis. Depending on the type of experiment, nanopore sequencers will produce large amounts of complex data. Hence, it is necessary to have a computing platform available which can handle these data.

From an organizational point of view, another question is whether genetic analyses should be centralized to a few centers in the future or whether smaller, clinical laboratories will need to offer such analyses. If they are done by highly specialized centers, it might not be lucrative for smaller laboratories to establish modern sequencing technologies. However, this is unlikely, as there are several other applications of modern sequencing techniques in addition to human genetics. Especially

in the field of microbiology, which is an important part of most clinical routine laboratories, molecular biological techniques involving NGS are becoming more and more important [20].

So far, Nanopore sequencing is not used in clinical laboratories to investigate routine diagnostic issues. However, there are numerous published clinical research applications which show the flexibility of nanopore sequencing and evaluate the performance for different analytical issues. Up to now, the biggest disadvantage is the reduced accuracy compared to conventional NGS techniques. Nevertheless, there is considerable active development occurring to increase accuracy and performance in general. If this trend continues, one can assume that nanopore sequencing might become a standard method and establish itself even in clinical routine laboratory.

Acknowledgment:

We thank the anonymous reviewer whose comments helped to improve and clarify this manuscript.

Declaration of Interest:

There are no conflicts of interest.


References:

1. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012;2012:251364 (PMID: 22829749).
2. Loose MW. The potential impact of nanopore sequencing on human genetics. *Hum Mol Genet* 2017;26(R2):R202-R207 (PMID: 28977449).
3. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ* 2019;61:316-26 (PMID: 31037722).
4. Di Resta C, Galbiati S, Carrera P, Ferrari M. Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. *EJIFCC* 2018;29(1):4-14 (PMID: 29765282).
5. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* 2015;13(1):4-16 (PMID: 25743089).
6. Branton D, Deamer DW, editors. *Nanopore sequencing: An introduction*. World Scientific: New Jersey; 2018. <https://lccn.loc.gov/2018049399>
7. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* 2016;14(5):265-79 (PMID: 27646134).
8. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to base-pair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol* 2018;19(1):90 (PMID: 30005597).

9. Liu Z, Wang Y, Deng T, Chen Q. Solid-State Nanopore-Based DNA Sequencing Technology. *Journal of Nanomaterials* 2016; 2016:1-13. <https://doi.org/10.1155/2016/5284786>
10. Haque F, Wang S, Stites C, Chen L, Wang C, Guo P. Single pore translocation of folded, double-stranded, and tetra-stranded DNA through channel of bacteriophage phi29 DNA packaging motor. *Biomaterials* 2015;53:744-52 (PMID: 25890769).
11. Manara RMA, Wallace EJ, Khalid S. DNA sequencing with MspA: Molecular Dynamics simulations reveal free-energy differences between sequencing and non-sequencing mutants. *Sci Rep*;5:12783 (PMID: 26255609).
12. Magi A, Semeraro R, Mingrino A, Giusti B, D'Aurizio R. Nanopore sequencing data analysis: state of the art, applications and challenges. *Brief Bioinform* 2018;19(6):1256-72 (PMID: 28637243).
13. Sedlazeck FJ, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single molecule sequencing. *Nat Methods* 2018;15(6):461-8 (PMID: 29713083).
14. Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res* 2011; 21(3):487-93 (PMID: 21209072).
15. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 2012; 13(1):238 (PMID: 22988817).
16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14):1754-60 (PMID: 19451168).
17. Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat Commun* 2016;7:11307 (PMID: 27079541).
18. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat Methods* 2015;12(8):733-5 (PMID: 26076426).
19. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. *Nat Methods* 2015;12(4):351-6 (PMID: 25686389).
20. Tyler AD, Mataseje L, Urfano CJ, et al. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci Rep* 2018;8(1):10931 (PMID: 30026559).
21. Hoenen T, Groseth A, Rosenke K, et al. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerging Infect Dis* 2016;22(2):331-4 (PMID: 26812583).
22. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017;12(6): 1261-76 (PMID: 28538739).
23. Patel A, Belykh E, Miller EJ, et al. MinION rapid sequencing: Review of potential applications in neurosurgery. *Surg Neurol Int* 2018;9:157 (PMID: 30159201).
24. Orsini P, Minervini CF, Cumbo C, et al. Design and MinION testing of a nanopore targeted gene sequencing panel for chronic lymphocytic leukemia. *Sci Rep* 2018;8(1):11798 (PMID: 30087429).
25. Bowden R, Davies RW, Heger A, et al. Sequencing of human genomes with nanopore technology. *Nat Commun* 2019;10(1):1869 (PMID: 31015479).
26. Leija-Salazar M, Sedlazeck FJ, Toffoli M, et al. Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. *Mol Genet Genomic Med* 2019:e564 (PMID: 30637984).



Serological and viral genetic features of patients with COVID-19 in a selected German patient cohort—correlation with disease characteristics

Jonas Schmidt · Sandro Berghaus · Frithjof Blessing · Folker Wenzel · Holger Herbeck · Josef Blessing · Peter Schierack · Stefan Rödiger · Dirk Roggenbuck 

Received: 19 May 2021 / Accepted: 12 August 2021 / Published online: 1 September 2021
© The Author(s) 2021

Abstract To study host-virus interactions after SARS coronavirus-2 (SARS-CoV-2) infection, genetic virus characteristics and the ensued humoral immune response were investigated for the first time. Fifty-five SARS-CoV-2-infected patients from the early pandemic phase were followed up including serological testing and whole genome sequencing. Anti-spike and nucleocapsid protein (S/N) IgG and IgM levels were determined by screening ELISA and IgG was further characterized by reactivity to

S-subunit 1 (anti-S1), S-subunit 2 (anti-S2) and anti-N. In 55 patients, 90 genetic SARS-CoV-2 changes including 48 non-synonymous single nucleotide variants were identified. Phylogenetic analysis of the sequencing data showed a cluster representing a local outbreak and various family clusters. Anti-S/N and anti-N IgG were detected in 49 patients at an average of 83 days after blood collection. Anti-S/N IgM occurred significantly less frequently than IgG whereas anti-S2 was the least prevalent IgG reactivity ($P < 0.05$, respectively). Age and overweight were significantly associated with higher anti-S/N and anti-S1 IgG levels while age only with anti-N IgG (multiple regression, $P < 0.05$, respectively). Anti-S/N IgG/IgM levels, blood group A+, cardiovascular and tumour disease, NSP12 Q444H and ORF3a S177I were independent predictors of clinical characteristics with anti-S/N IgM being associated with the need for hospitalization (multivariate regression, $P < 0.05$, respectively). Anti-SARS-CoV-2 antibody generation was mainly affected by higher age and overweight in the present cohort. COVID-19 traits were associated with genetic SARS-CoV-2 variants, anti-S/N IgG/IgM levels, blood group A+ and concomitant disease. Anti-S/N IgM was the only antibody associated with the need for hospitalization.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11357-021-00443-w>.

J. Schmidt · S. Berghaus · F. Blessing · H. Herbeck · J. Blessing
Institute for Laboratory Medicine, Singen, Germany

J. Schmidt · F. Blessing · F. Wenzel
Faculty of Medical and Life Sciences, Furtwangen University, Villingen-Schwenningen, Germany

J. Schmidt · P. Schierack · S. Rödiger · D. Roggenbuck (✉)
Faculty Environment and Natural Sciences, Institute of Biotechnology, Brandenburg University of Technology Cottbus-Senftenberg, Großhainer Str. 57, 01968 Senftenberg, Germany
e-mail: dirk.roggenbuck@b-tu.de

P. Schierack · S. Rödiger · D. Roggenbuck
Faculty of Health Sciences Brandenburg, Brandenburg University of Technology Cottbus – Senftenberg, Senftenberg, Germany

Keywords COVID-19 disease characteristics · Serology · Viral genetics · Correlation

Introduction

Severe acute respiratory syndrome-coronavirus type 2 (SARS-CoV-2), an enveloped, positive-sense single-stranded RNA virus, causing coronavirus disease 2019 (COVID-19) has spread rapidly worldwide, with strong economic and social impacts [1, 2]. In contrast to endemic coronaviruses, SARS-CoV-2 is classified as highly pathogenic, with similar characteristics to SARS-CoV and Middle East respiratory syndrome (MERS)-CoV [3].

Its genome consists of 14 open-reading frames (ORF) [3, 4]. They encode 16 non-structural proteins (NSP) which are essential for virus replication within the host cell through the formation of a replicase complex [3, 4]. Additionally, the ORFs encode nine accessory and four structural proteins, which include spike (S), envelope, membrane and nucleocapsid (N) proteins [4]. Upon contact with the host cell, the S protein is cleaved into two subunits (S1/S2) by proteases [4]. Both of them are essential for viral entry and define tissue tropism as well as viral host range [4, 5].

After infection, the incubation period is approximately 4–12 days [4–6]. The clinical features of COVID-19 are diverse and vary in onset and severity [4]. Main symptoms are fever, cough, gastrointestinal illnesses, anosmia and dyspnoea [4]. In addition to these acute symptoms, COVID-19 may also be associated with long-term effects, such as myocardial inflammation [4]. In severe cases, initially mild symptoms may later progress to life-threatening systemic inflammation with a cytokine storm syndrome [1, 4]. This will result in acute respiratory distress syndrome and respiratory failure which are considered leading causes of death in patients with COVID-19 [1, 4].

Infection with SARS-CoV-2 triggers both humoral and cellular immune responses. However, the underlying molecular mechanisms are not fully understood [7]. The S and N proteins are most immunogenic, with distinct IgM, IgG and IgA responses noted in COVID-19 patients [7].

To study host-virus interactions, we combined clinical data of COVID-19 patients of a south-western region of Germany with comprehensive serological data and SARS-CoV-2 whole genome sequencing (WGS) results for the first time to our knowledge.

Material and methods

Study population

Fifty-five patients with COVID-19 diagnosed in accordance with the World Health Organization criteria from the State of Baden-Württemberg in Germany were included in this study. [8] Inclusion criteria were a positive SARS-CoV-2 PCR test and a sample of the viral RNA present in the long-term sample archive (Fig. 1).

In total, 169 individuals were tested positive for SARS-CoV-2 in April 2020 at the beginning of the pandemic. They were contacted at least 2 months later and were invited to participate in serological testing and clinical data collection from June to August. In six cases, a complete follow-up was not possible because the individual was deceased or not available for sample collection. The data collection by using a questionnaire included common patient data, risk factors, symptoms and duration of the disease, long-term effects, therapy and epidemiological questions (Table 1). Of the 49 patients, who underwent anti-SARS-CoV-2 antibody testing, 48 returned the questionnaire. Clinical progression was determined from the responses applying the proposed WHO clinical progression scale [9]. Need for hospitalization was reported by the study participants in the questionnaire.

Serological testing

Serum samples for serological testing were collected by venipuncture and stored at $-20\text{ }^{\circ}\text{C}$ until further analysis. The anti-SARS-CoV-2 IgG and IgM levels to a mixture of S and N proteins (anti-S/N), respectively, were determined according to the manufacturer's manual by two commercial ELISA kits (GA CoV-2 IgG, GA CoV-2 IgM, GA Generic Assays GmbH, Dahlewitz, Germany) on an automated ELISA analyser (Institut Viron-Serion GmbH, Würzburg, Germany). Briefly, a binding index (BI) is calculated by the ratio of optical density (OD) values of samples to a cut-off OD value. Results with a BI ranging from 0.9 to 1.1 were considered borderline [10].

Additionally, an anti-SARS-CoV-2 IgG ELISA, recommended for confirmatory anti-SARS-CoV-2

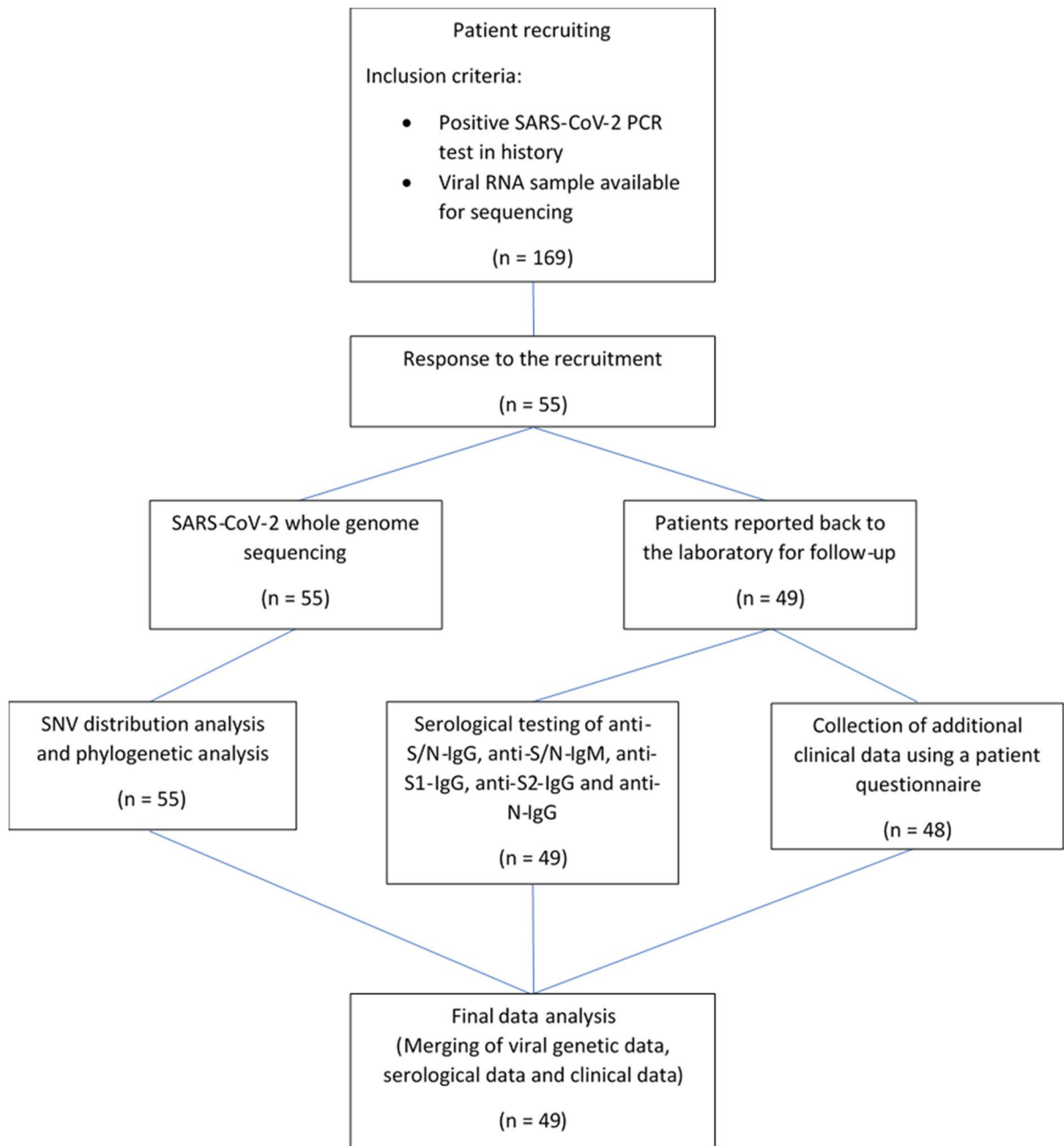


Fig. 1 Recruitment of patients. Inclusion criteria for the study were a positive SARS-CoV-2 PCR test and a sample of the viral RNA present in the long-term sample archive. All patients were tested positive for SARS-CoV-2 in April 2020 at the beginning of the pandemic. anti-S/N, antibodies against a

mixture of the spike glycoprotein with the nucleocapsid; anti-S1 IgG, IgG antibodies to spike glycoprotein domain 1; anti-S2 IgG, IgG antibodies to spike glycoprotein domain 2; anti-N IgG, IgG antibodies to nucleocapsid; SNV, single nucleotide variation

Table 1 Patients' and corresponding COVID-19 characteristics. In total, 48 returned a questionnaire, encompassing patient characteristics and clinical manifestations of the infection. One of the patients reported being completely symptom-free. The symptoms of the remaining 47 patients persisted for a median time of 10 days with an interquartile range of 7 days. Hospitalization due to moderate disease was reported in 6 cases with a mean hospitalization time of 7 days (standard deviation 5 days)

	Number/positive cases	Percentage [%]
Patient characteristics		
SARS-CoV-2 whole genome sequencing	55	100
Anti-SARS-CoV-2 antibody testing	49	89
Questionnaire complete	48	87
Death	5	9
Age		
<30 years	5	9
30–65 years	36	65.4
>65 years	14	25.5
BMI		
<25 ^[a]	17	35.4
25–35 ^[a]	26	54.2
>35 ^[a]	5	10.4
Blood group^[b]		
Type A+	16	33.4
Type A–	3	6.3
Type AB+	3	6.3
Type B+	3	6.3
Type O+	16	33.4
Type O–	2	4.2
Gender		
Female	29	52.7
Male	26	47.3
Clinical characteristics^[a]		
Cardiovascular disease	12	25
Chronic liver disease	2	4.2
Chronic lung disease	8	16.7
Diabetes	6	12.5
Tumour disease	3	6.3
Vitamin D supplementation	6	12.5
COVID-19 characteristics^[a]		
Appetite loss	29	60.4
Breathing difficulties	14	29.2
Bronchial secretions	12	25
Cough	26	54.2
Fatigue	43	89.5
Fever	27	56.3
Hospitalization	6	12.5
Without oxygen need	1	2.1
Oxygen need	5	10.4
Long-term COVID-19 effects	18	37.5
Night sweat	18	37.5
Pneumonia	4	8.3
Shortness of breath	9	18.8
Sore throat	17	35.4
Taste and smell disorders	32	66.7

^[a]Only patients with a complete questionnaire are included ($n=48$)

^[b]Blood groups were only available from 43 patients

IgG testing, was performed according to the manufacturer's protocol (GA CoV-2 IgG+, GA Generic Assays). The assay differentiates IgG to S1 (anti-S1), S2 (anti-S2) and N proteins (anti-N).

All antibody assays showed sensitivities of $\geq 98\%$ after 14 days of SARS-CoV-2 confirmation by PCR. To assess specificity, 1000 blood donor samples collected before and after the COVID-19 outbreak were tested. The anti-S/N IgG and IgM assays showed a specificity of $> 98\%$, respectively. False-positive results may be a consequence of the previous contact with other members of the coronavirus family. No cross-reactions were found by antibodies to the following common infective agents: PIV1-3, Influenza viruses A and B, Haemophilus influenzae, hCoV-229E, hCoV-OC43, hCoV-HKU1, hCoV-NL63, rhinovirus, RSV, adenovirus, M. pneumoniae, C. pneumoniae, CMV, EBV, HSV1 and 2, Toxoplasma, Rubella virus, Coxsackie virus, Parvovirus B19, HCV and HIV. The detected false-positive antibodies were mainly reactive with the N protein. These antibodies were probably generated during previous infections by endemic coronaviruses. Using samples first tested negative for IgG on the GA CoV-2 IgG ELISA, the GA CoV-2 IgG+ reached a specificity of almost 100%.

PCR testing

Viral RNA was isolated from nasopharyngeal swabs using PrepitoViral DNA/RNA300 isolation kits (PerkinElmer, Waltham, USA). PCR testing was performed by using the QuantiTect Probe RT-PCR Kit (Qiagen, Hilden, Germany) with primers and a hydrolysis probe (Biomers, Ulm, Germany) targeting the E gene (Suppl. Material 1). Detection was done on the FAM channel of a LightCycler 96 instrument (Roche, Basel, Switzerland).

SARS-CoV-2 next-generation sequencing

SARS-CoV-2 WGS was performed on a MinION sequencing platform (Oxford Nanopore Technologies, Oxford, UK) using the ARTIC nCoV-2019 sequencing protocol (Suppl. Mat. 2) [11–13]. All 55 samples were divided into three sequencing runs, each including a no-template control and an internal sequencing control. Lambda DNA (Oxford Nanopore Technologies, Oxford, UK) was used as an internal control.

Sequencing data analysis

Rampart was used to monitor the sequencing runs in real time. Oxford Nanopores own basecaller Guppy was employed to rebasecall the produced FAST5 files with a high accuracy model and for demultiplexing. Detailed analysis of sequence data is outlined in Supplemental Material 2. The resulting phylogenetic tree was visualized using R (v4.0.2) (R Foundation for Statistical Computing, Vienna, Austria) and the ggtree package (Suppl. Tab. 1). All consensus sequences from this study are available from GISAID (Suppl. Material 2).

Statistical analysis

Statistical testing was performed using R and ggplot2 package as well as MedCalc (v13.3.00) (MedCalc Software Ltd., Ostend, Belgium). Normality of data was assessed by Shapiro–Wilk test. In the case data was not normally distributed, differences between patient groups were compared using Kruskal–Wallis tests followed by post hoc analysis according to Conover. To compare the variation rate of different genes in the SARS-CoV-2 genome relative to their length, a generalized linear model (GLM) assuming a Poisson distribution was applied. Rank correlation was performed to identify the degree of association between antibody levels and patient characteristics. Logistic regression and multiple regression analyses were performed to predict an association between clinical outcome, serological data and genetic SARS-CoV-2 characteristics.

Results

Clinical presentation of COVID-19

To gain a deeper understanding of SARS-CoV-2 host-virus interactions, a follow-up of 55 COVID-19 patients from April 2020 was performed encompassing (i) SARS-CoV-2 WGS and (ii) serological testing for anti-S/N IgG and IgM as well as IgG to S1, S2 and N. Of 55 COVID-19 patients with PCR-confirmed SARS-CoV-2 infection and viral WGS analysis, 49 patients reported back to the laboratory for antibody testing (Fig. 1). In five of the 6 cases without follow-up, the patient was deceased. Of these

49 patients with a mean age of 52.2 years (standard deviation [SD] 16.2 years), 48 returned a questionnaire, encompassing patient characteristics and clinical manifestations of the infection (Table 1). One of the patients reported being completely symptom-free. The symptoms of the remaining 47 patients persisted for a median time of 10 days (interquartile range [IQR] 7 days). Hospitalization due to moderate disease was reported in 6 cases with a mean hospitalization time of 7 days (SD 5.0 days). Long-term effects of COVID-19 were stated by 18 patients (37.5%), including primarily fatigue and persisting loss of taste and smell.

SARS-CoV-2 whole genome sequencing

Whole genome sequencing of 55 SARS-CoV-2 RNA samples of the recruited COVID-19 patients was performed whereas all obtained sequences could be included in further downstream analysis as the coverage was above 85% (min 88.9%; max 99.6%). Variants to the reference genome MN908947.3 were clearly distributed over the whole SARS-CoV-2

genome (Fig. 2A). In total, 90 different unique variants including 34 synonymous single nucleotide variations (SNVs), 48 non-synonymous SNVs, 2 non-frameshift insertions, 1 frameshift insertion and 5 unclassified variants were identified within the study population (Suppl. Tab. 2). Median variant count per sample was eight and 99.7% of the genomic sites in the total population were without variations. The variants c.C2772T (ORF1ab F924F), c.C14144T (ORF1ab P4715L), c.A1841G (S D614G), and a transition from C to T in the 5' UTR at position 241 were identified in all 55 samples (Fig. 2A). A heat map of the variant count per gene and sample demonstrated that ORF10 was the only invariant region (Fig. 2B). In all samples, the highest numbers of variants were found in ORF1ab, followed by S, 5' UTR and ORF3a.

The variation rate of the individual genes relative to their length was assessed by a general linearized model (Suppl. Figure 1). Here, a highly significant positive influence of the N gene on the normalized variation rate was identified ($P=0.0096$, estimate: 0.876, standard error [SE]: 0.338), which means that this gene shows a significantly larger number of

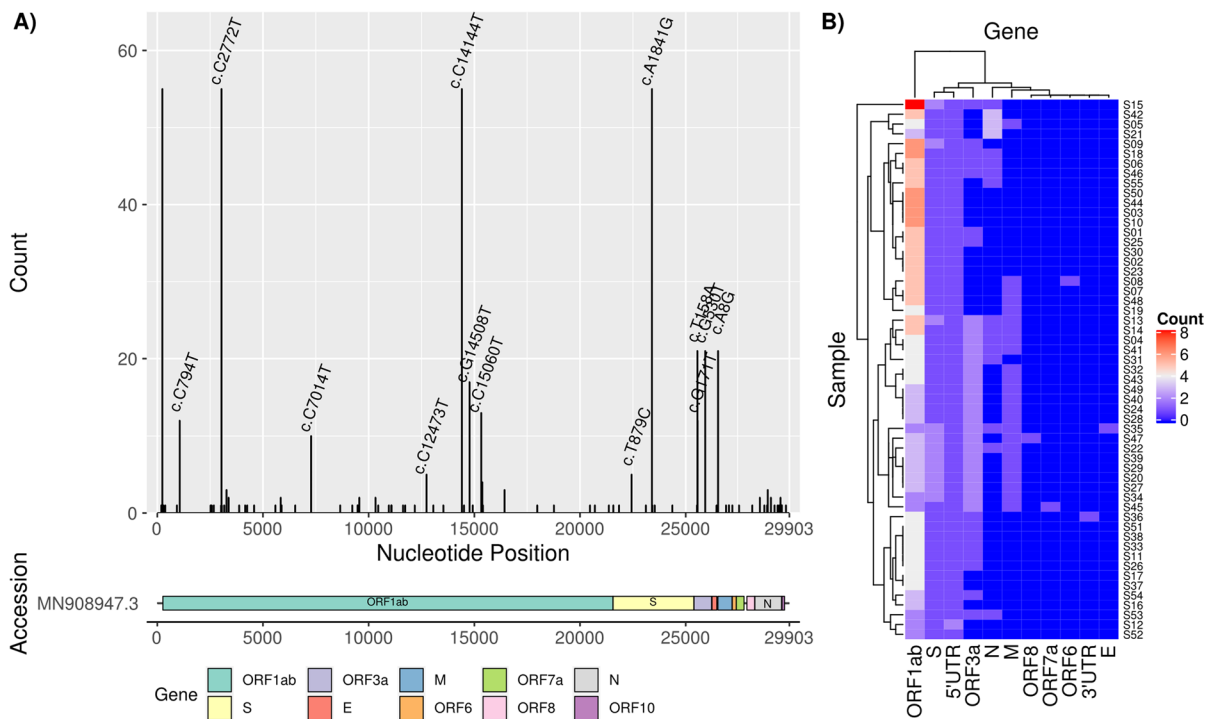


Fig. 2 **A** Overall distribution of SARS-CoV-2 variants identified by whole genome sequencing. The most common variants in coding regions are labelled. **B** Individual variant count per gene in each study sample

unique variants compared to the other regions of the SARS-CoV-2 genome. Further to this, a significant negative influence of ORFlab on the normalized variation rate was observed by applying the model ($P=0.04$, estimate: -0.528 , SE: 0.258).

To analyse the sequencing data from an epidemiological perspective, a phylogenetic analysis was performed (Fig. 3). Six different SARS-CoV-2 lineages, namely B.1, B.1.1, B.1.5, B.1.126, B.1.322 and B.1.353 were identified (Suppl. Tab. 3). The phylogenetic tree showed clear regional clusters in the area of Tuttlingen and Sigmaringen. Deeper analysis of patients' meta-data from the questionnaire revealed that the local cluster in the area of Sigmaringen originated from a local outbreak in a rehabilitation clinic. This was also confirmed by local health authorities. Besides local clustering, distinct clusters were observed within family

members all of whom had an identical SARS-CoV-2 genotype.

Serological testing

Blood drawing was performed on average 83 days (mean 83.3 days, SD 14.3 days) after a positive PCR result. Serological testing encompassed the semi-quantitative detection of anti-S/N IgG and IgM levels. Additionally, IgG levels were differentiated into anti-S1, anti-S2 and anti-N IgG (Fig. 4).

Anti-S/N IgG and anti-N IgG were detected in all 49 patients. Anti-S/N IgM was less frequently detected than anti-S/N IgG (27/49 vs. 49/49, $P<0.0001$). Among the three IgG reactivities investigated, anti-S2 IgG occurred significantly less frequently than anti-S1 and anti-N IgG (19/49 vs. 48/49 and 49/49, $P<0.0001$, respectively).

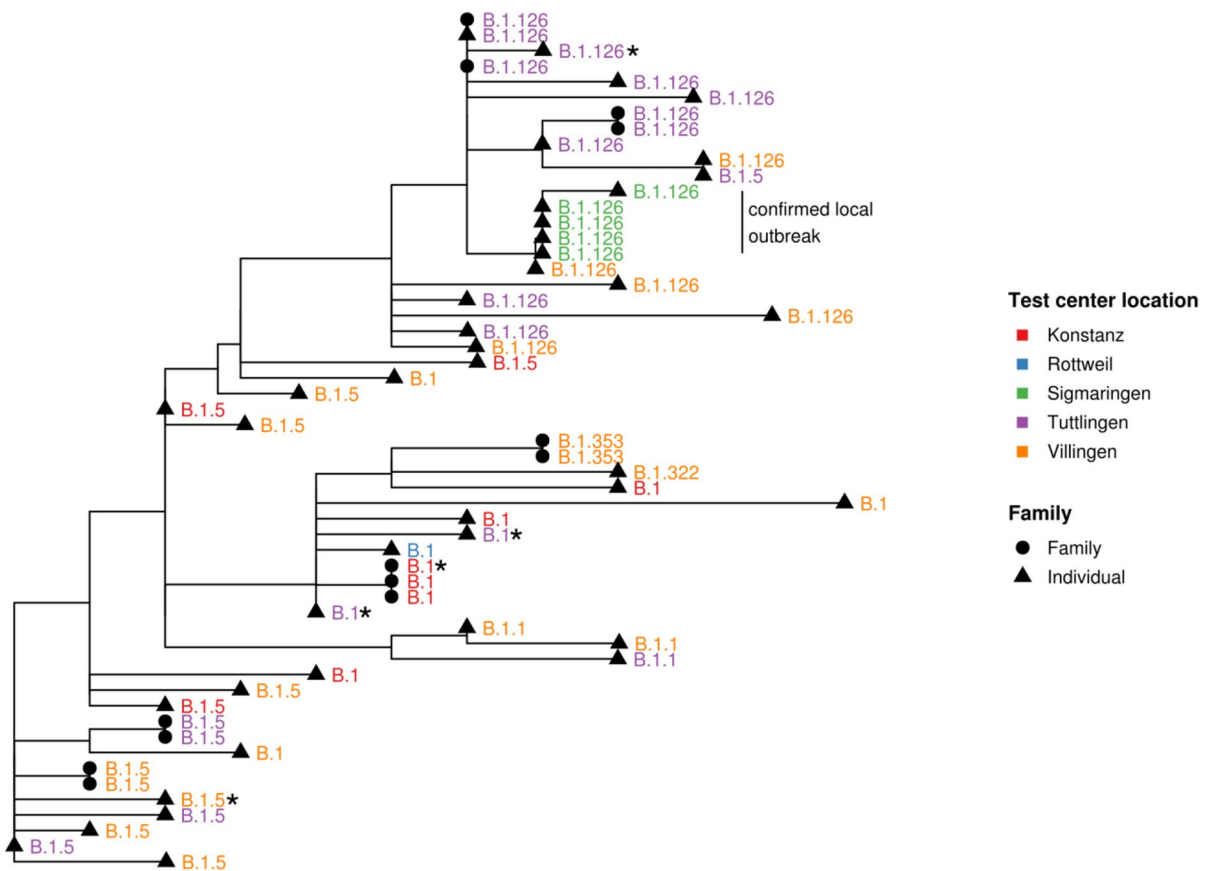
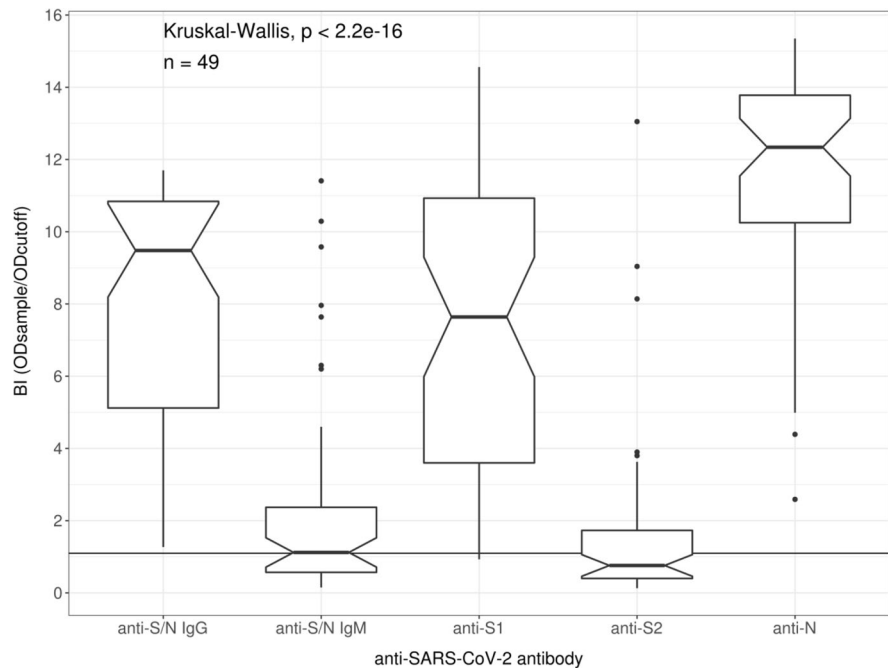


Fig. 3 Phylogenetic analysis of the different SARS-CoV-2 consensus sequences. The tips of the tree are labelled with the identified lineage (*=deceased patients). Samples from

patients, which belong to the same family, present clear clusters. Additionally, a local outbreak in the region of Sigmaringen with the lineage B.1.126 was identified

Fig. 4 Anti-SARS-CoV-2 antibody levels in 49 patients with COVID-19. In total, 49 patient samples were tested for anti-SARS-CoV-2 IgG and IgM against a mixture of the spike glycoprotein with the nucleocapsid (anti-S/N), respectively. Furthermore, IgG against the spike glycoprotein domain 1 (anti-S1), domain 2 (anti-S2) and the nucleocapsid protein (anti-N) were detected. The positive cut-off is located at 1.1 BI (binding index)



Patients demonstrated antibody patterns with varying frequencies (Table 2). The three most prevalent patterns (anti-S/N, anti-S1 and anti-N IgG; anti-S/N IgG and IgM, anti-S1 and anti-N IgG; anti-S/N IgG and IgM, anti-S1, anti-S2 and anti-N IgG) did not show a significantly different prevalence ($P > 0.05$, respectively).

The obtained IgG and IgM levels did not correlate within the examined period of 83 days on average after SARS-CoV-2 PCR testing ($P > 0.05$, respectively).

Anti-SARS-CoV-2 antibody levels in age groups

Rank correlation analysis revealed significant associations of all anti-SARS-CoV-2 antibodies with age (anti-S/N IgG, Spearman's ρ [ϕ]=0.497, $P=0.0003$; anti-S/N IgM, $\phi=0.312$, $P=0.0289$; anti-N IgG, $\phi=0.485$, $P=0.0004$; anti-S1 IgG, $\phi=0.521$, $P=0.0001$; anti-S2 IgG, $\phi=0.288$, $P=0.0451$).

To further investigate the occurrence of anti-SARS-CoV-2 antibodies in relation to age, patients were stratified into three groups: (i) younger than 30 years ($n=5$), (ii) between 30 and 65 years ($n=34$), and (iii) older than 65 years ($n=10$).

Table 2 Patterns of anti-SARS-CoV-2 antibody positivity by ELISA. In total 49 patient samples were tested for anti-SARS-CoV-2 IgG and IgM against a mixture of the spike glycoprotein with the nucleocapsid (anti-S/N), respectively. Furthermore, IgG against the spike glycoprotein domain 1 (anti-S1),

domain 2 (anti-S2) and the nucleocapsid protein (anti-N) were detected. Five different anti-SARS-CoV-2 antibody patterns were identified by ELISA testing. Patterns I, II and III were significantly more prevalent than patterns IV and V (Fisher's exact test, $P < 0.05$, respectively)

Pattern	Anti-S/N IgG	Anti-S/N IgM	Anti-S1 IgG	Anti-S2 IgG	Anti-N IgG	Number	Percentage [%]
I	+	–	+	–	+	15	30.6
II	+	+	+	–	+	15	30.6
III	+	+	+	+	+	12	24.5
IV	+	–	+	+	+	6	12.2
V	+	–	–	+	+	1	2.0

+ = positive, – = negative

Patients older than 65 years showed significantly higher anti-S/N, anti-S1 and anti-N IgG levels in contrast to patients in the two groups with younger age ($P < 0.05$ respectively) (Fig. 5A). Anti-S/N IgM levels were significantly higher only in patients older than 65 years compared to patients aged 30–65 years ($P = 0.012$), but not compared to the age group below 30 years ($P > 0.05$). For anti-S2

IgG, no significant differences between the age groups were observed.

Anti-SARS-CoV-2 antibody levels in groups with different BMI

The body mass index (BMI) was calculated and correlated with the various anti-SARS-CoV-2 antibodies.

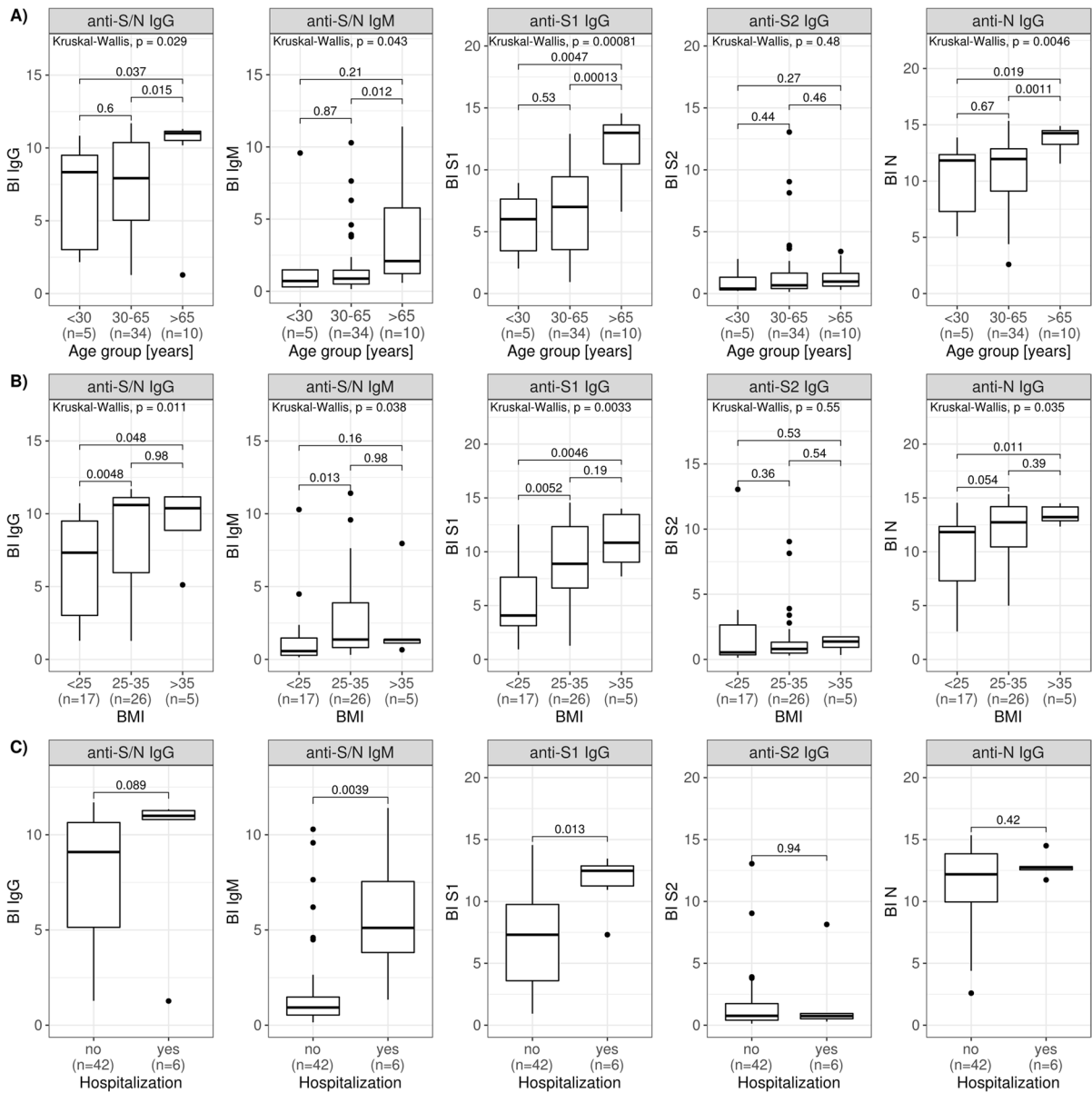


Fig. 5 Anti-SARS-CoV-2 antibody levels differentiated by age, BMI and severity. **A** Anti-SARS-CoV-2 antibody levels in different age groups (n = 49). **B** Anti-SARS-CoV-2 antibody

levels in different body mass index (BMI) groups (n = 48). **C** Anti-SARS-CoV-2 antibody levels in relation to the need for hospitalization (n = 48)

A significant association was established for anti-S/N IgG ($\phi=0.404$, $P=0.0045$), anti-S/N IgM ($\phi=0.355$, $P=0.0133$) and anti-S1 IgG ($\phi=0.451$, $P=0.0013$).

Furthermore, patients were stratified into three different groups: (i) normal weight (BMI < 25; $n=17$), (ii) overweight (BMI 25–35; $n=26$), (iii) severe overweight (BMI > 35, $n=5$). Patients with overweight and severe overweight showed significantly higher antibody levels compared to the normal weight group for all tested antibodies except anti-S/N IgM and anti-S2 IgG ($P<0.05$, respectively) (Fig. 5B). Anti-S/N IgM levels were only significantly higher in patients of the overweight group compared with the ones of the normal weight group ($P=0.013$).

Anti-SARS-CoV-2 antibody levels in relation to the need for hospitalization

Furthermore, antibody levels were compared with regard to the need for hospitalization indicating moderate COVID19 with scores ranging from 4 to 5 (Fig. 5C). Here, significantly higher levels of anti-S/N IgM and anti-S1 IgG were observed in hospitalized patients ($n=6$, $P<0.05$, respectively). All other antibodies tested demonstrated no significant difference regarding the need for hospitalization ($P>0.05$, respectively).

A possible association of anti-SARS-CoV-2 antibodies with hospitalization duration was investigated by rank correlation. Again, a significant association was observed for anti-S/N IgM ($\phi=0.428$, $P=0.0024$) and anti-S1 IgG ($\phi=0.355$, $P=0.0133$).

Association of anti-SARS-CoV-2 antibody levels with genetic SARS-CoV-2 variants and patient characteristics

Given the positive correlation of anti-SARS-CoV-2 antibody levels with age and overweight, univariate followed by multivariate regression analysis was performed to investigate an influence of other patient characteristics and genetic SARS-CoV-2 variants on antibody generation (Table 3). Age was established as an independent predictor for higher anti-S/N, anti-S1 and anti-N IgG levels whereas the latter had no further predictors. In contrast, overweight (BMI > 25, $n=31$) was identified as an additional independent predictor for higher anti-S/N and anti-S1 IgG levels. The absence of the genetic SARS-CoV-2 variant

Table 3 Multiple regression analyses of anti-SARS-CoV-2 antibody levels of 48 patients as dependent variables and independent parameters encompassing patient characteristics listed in Table 1 and SARS-CoV-2 genetic features as predictors. *anti-S/N IgG*, IgG antibodies against a mixture of the spike glycoprotein with the nucleocapsid; *anti-S1 IgG*, IgG antibodies to spike glycoprotein domain 1; *anti-N IgG*, IgG antibodies to nucleocapsid

	Coefficient	Std. Error	P value
Anti-S/N IgG			
Age	0.069	0.026	0.0104
Chronic liver disease	-5.225	2.020	0.0131
Overweight ^[a]	2.163	0.876	0.0174
Anti-S/N IgM			
Tumour disease	5.064	1.567	0.0023
Anti-S1 IgG			
Age	0.088	0.030	0.0049
Overweight ^[a]	2.828	1.005	0.0073
NSP3 D218E	-5.708	2.313	0.0175
Anti-S2 IgG			
NSP3 D218E	4.837	1.662	0.0055
Anti-N IgG			
Age	0.081	0.026	0.0033

^[a]Overweight was characterized by BMI > 25

NSP3 D218E was an additional independent predictor for higher anti-S1 IgG levels whereas the absence of chronic liver disease was one for higher anti-S/N IgG levels.

The only independent predictor for higher anti-S/N IgM levels was the presence of tumour disease with no predictive effect of genetic SARS-CoV-2 variants or other patient characteristics such as age and overweight. For higher anti-S2 IgG levels, the presence of NSP3 D218E was revealed as the only independent predictor, which is in strong contrast to anti-S1 IgG.

Association between clinical outcome, genetic SARS-CoV-2 variability, humoral immune response and patient characteristics

In light of the correlation of anti-SARS-CoV-2 antibody levels with the need for hospitalization and its duration, univariate followed by multivariate regression analyses were performed to evaluate a possible association between the clinical outcome and various independent predictor variables (patient characteristics, antibody levels, viral genetic features).

Univariate analysis revealed a number of clinical characteristics as the dependent variable, which had higher SARS-CoV-2-antibody levels other than anti-S2 IgG levels as independent predictors (Suppl. Tab. 4). A total of five SNVs were found to be independent predictors of COVID-19 traits. All of them were non-synonymous, resulting in amino acid changes in various viral proteins.

In subsequent multivariate logistic regression analysis to account for confounding variables, only higher anti-S/N IgG and/or IgM levels were found to significantly predict COVID-19 characteristics such as appetite loss, night sweat, oxygen need, pneumonia and the need for hospitalization ($P < 0.05$, respectively) (Table 4). Interestingly, anti-S/N IgM was the only variable studied that predicted the occurrence of pneumonia (odds ratio [OR] 1.363, $P = 0.0317$).

Furthermore, the main confounder for higher anti-S/N IgM levels was cardiovascular disease in the prediction of the need for oxygen and hospitalization ($P < 0.05$, respectively). The blood group A+ was identified as an independent predictor for bronchial secretions and cough whereas the latter demonstrated the SNV ORF3a S177I as an additional independent predictor ($P < 0.05$, respectively). The only other SNV identified as independent was NSP12 Q444H for taste and smell disorders (OR 5.444, $P = 0.0426$).

Along with the presence of tumour and chronic lung diseases, a higher anti-S/N IgM level was significantly associated with longer hospitalization (multiple regression analysis, $P < 0.05$, respectively).

Chronic lung disease and the SNV N E253A were significantly associated with symptom duration (multiple regression analysis, $P < 0.05$, respectively).

Table 4 Multivariate regression analyses of COVID-19 characteristics of 48 patients. (A) Multivariate regression analyses of binary COVID-19 patient characteristics by logistic regression analysis. The relationship of dichotomous COVID-19 patient characteristics as dependent variables and independent parameters encompassing patient characteristics listed in Table 1, SARS-CoV-2 genetic features and anti-SARS-CoV-2 antibodies as predictors was analysed. (B) Multiple regression analyses of quantitative COVID-19 patient characteristics as dependent variables and independent parameters encompassing patient characteristics listed in Table 1, SARS-CoV-2 genetic features and anti-SARS-CoV-2 antibodies as predictors

(A) Logistic regression	Coefficient	Std. error	Odds ratio	95% CI	<i>P</i> value
Appetite loss					
Anti-S/N IgG	0.367	0.114	1.443	1.155–1.802	0.0012
Bronchial secretions					
Blood type A+	1.749	0.737	5.750	1.356–24.389	0.0177
Cough					
Blood type A+	2.765	1.144	15.882	1.687–149.490	0.0156
ORF3a S177I	− 3.041	1.108	0.048	0.054–0.419	0.0061
Night sweat					
Anti-S/N IgG	0.404	0.153	1.498	1.109–2.023	0.0084
Anti-S/N IgM	0.300	0.148	1.350	1.011–1.804	0.0419
Oxygen need					
Anti-S/N IgM	0.413	0.188	1.511	1.045–2.185	0.0282
Cardiovascular disease	3.075	1.432	21.647	1.306–358.738	0.0318
Pneumonia					
Anti-S/N IgM	0.310	0.144	1.363	1.027–1.808	0.0317
Hospitalization					
Anti-S/N IgM	0.441	0.201	1.554	1.992–2.306	0.0284
Cardiovascular disease	3.708	1.540	40.773	1.992–834.549	0.0161
Taste and smell disorders					
NSP12 Q444H	1.695	0.836	5.444	1.058–28.011	0.0426
(B) Multiple regression					
Hospitalization duration					
Anti-S/N IgM	0.305		0.109		0.0075
Tumour disease	5.980		1.300		< 0.0001
Chronical lung disease	2.224		3.334		0.0017
Symptom duration					
Chronical lung disease	19.250		7.456		0.0053
N E253A	12.571		4.261		0.0137

Discussion

More than a year after its identification, SARS-CoV-2 has shown a high degree of genome alteration [14]. To investigate virus-host interactions, we examined PCR-positive patients of a south-western German region who were referred to a local reference laboratory and answered a questionnaire on personal and COVID-19 characteristics.

Thus, WGS of the viral genome of 55 enrolled COVID-19 patient samples revealed genetic alterations mainly as SNVs, with about half of these resulting in changes of the amino acid sequence. When looking at the absolute variant count per gene and patient, most variants were located within ORF1ab representing the largest SARS-CoV-2 ORF. Nevertheless, ORF1ab showed a significantly lower variation rate normalized on the gene length compared to the other genes, while the N gene was the only gene with a significantly higher normalized variation rate. Overall, RNA viruses are known to accumulate variants rapidly during their replication cycle because RNA copying enzymes are prone to error [15, 16]. A high variation rate of the N gene was reported elsewhere [17, 18].

ORF10 was the only gene without variants in our study which was also demonstrated elsewhere [18]. Furthermore, our study corroborated published data on the S gene stability [19].

We observed four variants present in all samples (ORF1ab F924F, ORF1ab P4715L, S D614G and 5'UTR 241C>T), representing signature variants of the most dominant SARS-CoV-2 type VI strain [20]. In particular, the D614G exchange in the S protein has been extensively studied and is postulated to provide a selection advantage through increased viral infectivity [21–23].

All samples were assigned to the root lineage B based on Rambaut's nomenclature [24]. The highest level lineage was B.1, encompassing the major Italian outbreak in early 2020 and then spreading across Europe [24]. The other identified lineages were sub-lineages of B.1, which match the geographical origin of the samples. Remarkably, the earliest description dates of the lineages in the Pango strain database coincided with our sample collection date (2020–04–07 to 2020–05–07). At the time of writing this manuscript, the lineages B.1.322, B.1.353 and B.1.5 have already been reassigned as more and more

SARS-CoV-2 whole genomes have been sequenced over time and lineage formation and extinction continue to progress [24].

Given the high genetic variability of SARS-CoV-2, we sought to investigate the emergence of the humoral immune response by determining specific IgM and IgG against the most immunogenic S and N proteins in average 83 days after PCR testing [25, 26]. As expected, all patients revealed detectable anti-S/N and anti-N IgG while only one patient out of the examined 49 did not show anti-S1 IgG. The higher anti-S/N IgG prevalence in contrast to IgM probably indicates the effect of an immunological memory likely induced by previous infections with endemic coronaviruses, as primary immune responses would induce stronger anti-SARS-CoV-2 IgM responses. For all antibodies tested, there was no correlation between time from SARS-CoV-2 PCR testing and antibody levels within the examined period of 83 days on average after SARS-CoV-2 PCR testing. However, it cannot be ruled out that anti-S/N IgM levels, in particular, may have decreased to negative values in the period leading up to blood collection for antibody determination.

Rank correlation and multiple regression analyses using genetic SARS-CoV-2 variants and patient characteristics as independent variables for the prediction of anti-SARS-CoV-2 antibody levels revealed an association of older age (>65 years) and overweight (BMI>25) with higher anti-S/N and anti-S1 IgG levels. In contrast, higher anti-N IgG levels were only associated with older age. The average age of enrolled patients was 52.2 years which is in agreement with the reported age of around 50 years for COVID-19 patients [1, 27]. A systematic review and meta-analysis found old age and obesity as a risk for a severe COVID-19 course [28].

Remarkably, despite a positive correlation of age and BMI with anti-S/N IgM, higher levels of the latter were only associated with the concurrence of tumour disease by multiple regression analysis. On the contrary, the absence of concomitant chronic liver disease was a confounder for the association of older age and overweight with higher IgG levels. The found correlation with older age reflects the stronger humoral inflammatory response reported in aged COVID-19 patients, which may hint at an impaired innate or cellular adaptive immune response [1, 29]. Apart from older age, overweight has been described

as an additional risk factor for severe COVID-19 progression usually linked with functional impairment of immune cells and decreased immunity as a result of chronic inflammation and hypercytokinemia [30, 31]. Therefore, the observed positive association with higher anti-S/N and anti-S1 IgG levels may also be due to a unique predisposition of obese individuals to an impaired cellular anti-SARS-CoV-2 response and requires further investigation. Significantly higher SARS-CoV-2 IgG levels were also previously described in patients with metabolic syndrome comorbidities [32].

In line with previous reports, higher anti-S1 IgG levels were determined in contrast to anti-S2 IgG levels [26]. For the first time, we showed the positive association of higher anti-S2 IgG levels with the SNV NSP3 D218E. This is interesting as the same SNV is negatively associated with higher anti-S1 IgG levels in our patient cohort and may indicate a possible influence of SARS-CoV-2 non-structural protein 3 (NSP3) on antibody formation. The multi-domain Nsp3 is the largest SARS-CoV-2 protein and an essential component of the replication-transcription complex modifying host proteins and interfering with innate immune responses by de-ubiquitination [33].

There was an association of higher anti-S/N IgM and anti-S1 IgG levels with moderate COVID-19 requiring hospitalization of patients. Both anti-SARS-CoV-2 antibodies were also positively correlated with hospitalization duration. Multivariate regression analysis identified only higher anti-S/N IgM levels as predictors for the need for hospitalization with concomitant cardiovascular disease as confounder. This could entail that anti-S/N IgM can be employed as a marker of at least moderate COVID-19 in particular for patients with cardiovascular disease. Cardiovascular disease is an accepted risk factor for severe COVID-19 courses [34, 35].

In light of the diverse clinical expression of COVID-19 in our study cohort, the varying predisposition of patients and the genetic changes of SARS-CoV-2, we performed univariate followed by multivariate regression analysis to identify possible associations. COVID-19 symptoms observed in our study cohort were consistent with other studies [1, 27].

Interestingly, higher anti-S/N IgM and IgG levels were established as independent predictors of COVID-19 traits such as appetite loss, night sweat,

oxygen need and pneumonia. The latter was associated only with higher anti-S/N IgM levels without confounders, supporting published data and the above correlation of the IgM response with the need for hospitalization [36]. In addition to the presence of tumour and chronic lung disease, hospitalization duration was also associated with higher anti-S/N IgM levels.

Another interesting association was the prediction of clinical symptoms such as cough and bronchial secretions by blood type A+. This is consistent with other studies demonstrating a higher risk of individuals with this blood type to develop COVID-19 symptoms after infection [37–39]. While the occurrence of bronchial secretions was only associated with blood type A, the absence of the non-synonymous SNV ORF3a S177I was a confounder for the appearance of cough. The prediction of taste and smell disorders by the non-synonymous SNV NSP12 Q444H (OR 5.4) without confounders is another example in this study that genetic changes may influence the clinical presentation of COVID-19 [22, 40–44]. NSP12 is a large SARS-CoV-2 protein with 932 amino acid residues catalysing replication and transcription of the viral genome [45]. Furthermore, patients with chronic lung disease infected with SARS-CoV-2 bearing the non-synonymous SNV N E253A appear to have a longer symptom duration. This N protein SNV was the only genetic change in structural proteins associated with clinical characteristics in this study. The N protein demonstrating a high level of genetic alteration in the study has multiple functions including complex formation with genomic RNA, interaction with the viral membrane protein during virion assembly and enhancement of the efficiency of virus transcription and assembly [46]. However, it is not part of the replication-transcription complex which is the core component during viral replication [4, 5].

Approximately one-third of patients ($n=18$) in our study population reported having long-term symptoms, particularly persistent anosmia and fatigue after recovery from COVID-19. We could not find statistically significant associations with the persistency of symptoms.

A limitation of our study is the relatively small sample size. In addition, data may be biased by preferential inclusion of patients with symptoms. There was only one patient that did not report COVID-19 symptoms. Therefore, confirmation of the findings

in a larger study population is warranted. Additionally, the associations identified between certain viral and patient characteristics and the clinical outcome of COVID-19 are only descriptive. However, this is the first study combining SARS-CoV-2 WGS with comprehensive anti-SARS-CoV-2 antibody testing encompassing IgM and IgG reactivities.

Conclusion

Our results show diverse humoral immune responses to SARS-CoV-2, which appear to be influenced by disease severity, age and obesity. The serologic profile is more like that of a secondary humoral immune response than a primary one. The non-synonymous SARS-CoV-2 SNV NSP3 D218E is inversely associated with the humoral response to S subunits 1 and 2.

Clinical COVID-19 characteristics are correlated with genetic changes of SARS-CoV-2, anti-S/N IgG and IgM levels as well as patient characteristics such as blood type A+. Anti-S/N IgM is correlated with pneumonia and the need for hospitalization and oxygen. We identified the N gene to be the most variable part of the SARS-CoV-2 genome.

Acknowledgements We thank the reviewers whose comments helped to improve and clarify this manuscript.

Author contribution All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data sets are available in aggregated form on request from the authors.

Code availability Custom R scripts are available on request from the authors.

Declarations

Ethics approval The research related to human use has complied with all the relevant national regulations and institutional policies and is in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee (ethics committee of the Brandenburg University of Technology Cottbus-Senftenberg, EK2020-16).

Informed consent Informed consent was obtained from all individuals included in this study.

Competing interests DR holds an executive position and is a shareholder in Medipan and GA Generic Assays, which are diagnostic manufacturers.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol.* 2021;19(3):141–54. <https://doi.org/10.1038/s41579-020-00459-7>.
2. Plebani M. Laboratory medicine in the COVID-19 era: six lessons for the future. *Clin Chem Lab Med.* 2021;59(6):1035–45. <https://doi.org/10.1515/cclm-2021-0367>.
3. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol.* 2021;19(3):155–70. <https://doi.org/10.1038/s41579-020-00468-6>.
4. Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 transmission and pathogenesis. *Trends Immunol.* 2020;41(12):1100–15. <https://doi.org/10.1016/j.it.2020.10.004>.
5. Haque SM, Ashwaq O, Sarief A, Mohamed Azad John, Abdul Kalam. A comprehensive review about SARS-CoV-2. *Future Virology.* 2020;15(9):625–48. <https://doi.org/10.2217/fvl-2020-0124>.
6. Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis.* 2020;20(9):e238–44. [https://doi.org/10.1016/S1473-3099\(20\)30484-9](https://doi.org/10.1016/S1473-3099(20)30484-9).
7. Poland GA, Ovsyannikova IG, Kennedy RB. SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. *The Lancet.* 2020;396(10262):1595–606. [https://doi.org/10.1016/S0140-6736\(20\)32137-1](https://doi.org/10.1016/S0140-6736(20)32137-1).
8. World Health Organization. WHO COVID-19 Case definition: World Health Organization; 2020 [cited 2021 July 5] Available from: URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2020.2.
9. Marshall JC, Murthy S, Diaz J, Adhikari NK, Angus DC, Arabi YM, et al. A minimal common outcome measure

- set for COVID-19 clinical research. *Lancet Infect Dis.* 2020;20(8):e192–7. [https://doi.org/10.1016/S1473-3099\(20\)30483-7](https://doi.org/10.1016/S1473-3099(20)30483-7).
10. Speletas M, Kyritsi MA, Vontas A, Theodoridou A, Chrysanthidis T, Hatzianastasiou S, et al. Evaluation of two chemiluminescent and three ELISA immunoassays for the detection of SARS-CoV-2 IgG antibodies: implications for disease diagnosis and patients' management. *Front Immunol.* 2020;11:609242. <https://doi.org/10.3389/fimmu.2020.609242>.
 11. Quick J. nCoV-2019 sequencing protocol v2 (GunIt) V.2: protocols.io; 2020.
 12. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc.* 2017;12(6):1261–76. <https://doi.org/10.1038/nprot.2017.066>.
 13. Pattabiraman C, H PK, Habib F, Rasheed R, Prasad P, Reddy V, et al. Genomic epidemiology reveals multiple introductions and spread of SARS-CoV-2 in the Indian state of Karnataka. *PLoS ONE.* 2020;15(12):e0243412. <https://doi.org/10.1371/journal.pone.0243412>.
 14. Banfi G, Lippi G. COVID-19: which lessons have we learned? *Clin Chem Lab Med.* 2021;59(6):1009–11. <https://doi.org/10.1515/cclm-2021-0384>.
 15. Callaway E. The coronavirus is mutating — does it matter? *Nature.* 2020;585:174–7. <https://doi.org/10.1038/d41586-020-02544-6>.
 16. Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics.* 2020;112(5):3588–96. <https://doi.org/10.1016/j.ygeno.2020.04.016>.
 17. Mishra A, Pandey AK, Gupta P, Pradhan P, Dhamija S, Gomes J, et al. Mutation landscape of SARS-CoV-2 reveals five mutually exclusive clusters of leading and trailing single nucleotide substitutions. *bioRxiv.* 2020. <https://doi.org/10.1101/2020.05.07.082768>
 18. Kaushal N, Gupta Y, Goyal M, Khaiboullina SF, Baranwal M, Verma SC. Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA. *Pathogens.* 2020;9(7):565. <https://doi.org/10.3390/pathogens9070565>.
 19. Jia Y, Shen G, Nguyen S, Zhang Y, Huang K-S, Ho H-Y, et al. Analysis of the mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. *bioRxiv* 2020 <https://doi.org/10.1101/2020.04.09.034942>
 20. Yang H-C, Chen C-H, Wang J-H, Liao H-C, Yang C-T, Chen C-W, et al. Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. *Proc Natl Acad Sci U S A.* 2020;117(48):30679–86. <https://doi.org/10.1073/pnas.2007840117>.
 21. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature.* 2021;592(7852):116–21. <https://doi.org/10.1038/s41586-020-2895-3>.
 22. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet.* 2020;65(12):1075–82. <https://doi.org/10.1038/s10038-020-0808-j>.
 23. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell.* 2020;182(4):812–827.e19. <https://doi.org/10.1016/j.cell.2020.06.043>.
 24. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* 2020;5(11):1403–7. <https://doi.org/10.1038/s41564-020-0770-5>.
 25. Siracusano G, Pastori C, Lopalco L. Humoral immune responses in COVID-19 patients: a window on the state of the art. *Front. Immunol.* 2020;11(1049). <https://doi.org/10.3389/fimmu.2020>.
 26. Herroelen PH, Martens GA, de Smet D, Swaerts K, Decavele A-S. Humoral immune response to SARS-CoV-2. *Am J Clin Pathol.* 2020;154(5):610–9. <https://doi.org/10.1093/ajcp/aqaa140>.
 27. Lovato A, de Filippis C. Clinical presentation of COVID-19: a systematic review focusing on upper airway symptoms. *Ear Nose Throat J.* 2020;99(9):569–76. <https://doi.org/10.1177/0145561320920762>.
 28. Booth A, Reed AB, Ponzio S, Yassaee A, Aral M, Plans D, et al. Population risk factors for severe disease and mortality in COVID-19: a global systematic review and meta-analysis. *PLoS ONE.* 2021;16(3): e0247461. <https://doi.org/10.1371/journal.pone.0247461>.
 29. Klein SL, Pekosz A, Park H-S, Ursin RL, Shapiro JR, Benner SE, et al. Sex, age, and hospitalization drive antibody responses in a COVID-19 convalescent plasma donor population. *J Clin Invest.* 2020;130(11):6141–50. <https://doi.org/10.1172/JCI142004>.
 30. Frasca D, Reidy L, Cray C, Diaz A, Romero M, Kahl K, et al. Effects of obesity on serum levels of SARS-CoV-2-specific antibodies in COVID-19 patients. *medRxiv.* 2020;16(3):e0245424. <https://doi.org/10.1101/2020.12.18.20248483>.
 31. Korakas E, Ikonomidis I, Kousathana F, Balampanis K, Kountouri A, Raptis A, et al. Obesity and COVID-19: immune and metabolic derangement as a possible link to adverse clinical outcomes. *Am J Physiol Endocrinol Metab.* 2020;319(1):E105–9. <https://doi.org/10.1152/ajpendo.00198.2020>.
 32. Racine-Brzostek SE, Yang HS, Jack GA, Chen Z, Chadburn A, Ketas TJ, et al. Postconvalescent SARS-CoV-2 IgG and neutralizing antibodies are elevated in individuals with poor metabolic health. *J Clin Endocrinol Metab.* 2021;106(5):e2025–34. <https://doi.org/10.1210/clinem/dgab004>.
 33. Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antiviral Res.* 2018;149:58–74. <https://doi.org/10.1016/j.antiviral.2017.11.001>.
 34. Bae S, Kim SR, Kim M-N, Shim WJ, Park S-M. Impact of cardiovascular disease and risk factors on fatal outcomes in patients with COVID-19 according to age: a systematic review and meta-analysis. *Heart.* 2021;107(5):373–80. <https://doi.org/10.1136/heartjnl-2020-317901>.

35. Bansal M. Cardiovascular disease and COVID-19. *Diabetes Metab Syndr*. 2020;14(3):247–50. <https://doi.org/10.1016/j.dsx.2020.03.013>.
36. Wang Y, Zhang L, Sang L, Ye F, Ruan S, Zhong B, et al. Kinetics of viral load and antibody response in relation to COVID-19 severity. *J Clin Invest*. 2020;130(10):5235–44. <https://doi.org/10.1172/JCI138759>.
37. Zhao J, Yang Y, Huang H, Li D, Gu D, Lu X, et al. Relationship between the ABO blood group and the COVID-19 susceptibility. *Clin Infect Dis*. 2020;73(2):328–31. <https://doi.org/10.1101/2020.03.11.20031096>.
38. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide association study of severe COVID-19 with respiratory failure. *N Engl J Med*. 2020;383(16):1522–34. <https://doi.org/10.1056/NEJMoa2020283>.
39. Zietz M, Zucker J, Tatonetti NP. Associations between blood type and COVID-19 infection, intubation, and death. *Nat Commun*. 2020;11(1):5761. <https://doi.org/10.1038/s41467-020-19623-x>.
40. Young BE, Fong S-W, Chan Y-H, Mak T-M, Ang LW, Anderson DE, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet*. 2020;396(10251):603–11. [https://doi.org/10.1016/S0140-6736\(20\)31757-8](https://doi.org/10.1016/S0140-6736(20)31757-8).
41. Nagy Á, Pongor S, Györfy B. Different mutations in SARS-CoV-2 associate with severe and mild outcome. *Int J Antimicrob Agents*. 2021;57(2): 106272. <https://doi.org/10.1016/j.ijantimicag.2020.106272>.
42. Majumdar P, Niyogi S. ORF3a mutation associated with higher mortality rate in SARS-CoV-2 infection. *Epidemiol Infect*. 2020;148:e262. <https://doi.org/10.1017/S0950268820002599>.
43. Biswas SK, Mudi SR. Genetic variation in SARS-CoV-2 may explain variable severity of COVID-19. *Med Hypotheses*. 2020;143:109877. <https://doi.org/10.1016/j.mehy.2020.109877>.
44. Aiewsakun P, Wongtrakoongate P, Thawornwattana Y, Hongeng S, Thitithanyanont A. SARS-CoV-2 genetic variations associated with COVID-19 severity. *medRxiv* 2020 <https://doi.org/10.1101/2020.05.27.20114546>
45. Xu X, Liu Y, Weiss S, Arnold E, Sarafianos SG, Ding J. Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res*. 2003;31(24):7117–30. <https://doi.org/10.1093/nar/gkg916>.
46. McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*. 2014;6(8):2991–3018. <https://doi.org/10.3390/v6082991>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1 Schmidt, J., Blessing, F. & Gürtler, L. SARS-CoV-2 vaccines and reaction of the immune system. Can the epidemic spread of the virus
2 be prevented by vaccination? Dtsch. Med. Wochenschr. 146, 1085–1090 (2021). DOI: 10.1055/a-1550-0001 © 2021. Thieme,
3 reprinted with permission
4

5 **DMW Standpunkt**

6

7 **SARS-CoV-2 Impfstoffe und Reaktion des Immunsystems.**

8 **Kann die epidemische Ausbreitung des Virus durch Impfung verhindert werden?**

9

10 Jonas Schmidt¹⁾, Frithjof Blessing¹⁾, Lutz Gürtler²⁾

11 1) Labor Blessing, Singen

12 2) Max von Pettenkofer Institut der Ludwig Maximilians Universität, München

13

14

15

16 Korrespondenzadresse:

17 Prof. Dr. med. em. Lutz Gürtler

18 Max-von-Pettenkofer-Institut, LMU München

19 Pettenkofer Str. 9A

20 80336 München

21 Deutschland

22 Phone:

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37 Folgende weitere Autoren haben zur Erstellung des Manuskriptes beigetragen:

38 Philipp Schnee, Sandro Berghaus, Lothar Hehmann, Kerstin Blessing, Josef Blessing¹⁾

39 Thomas Löscher³⁾

40 1) Labor Blessing, Singen

41 3) MVZ Lehelmed, München

42

43 Interessenkonflikt:

44 Die Autorinnen/Autoren geben an, dass kein Interessenkonflikt besteht.

45

46 **Zusammenfassung**

47 Ende 2019 wurde in Wuhan, China ein neues Coronavirus, SARS-CoV-2, identifiziert, welches sich
48 weltweit verbreitet hat und teils eine hohe Letalität bei Erkrankung erzeugte. Neben den
49 hygienischen Maßnahmen zur Eindämmung der Verbreitung wurden Impfstoffe entwickelt, die
50 teils auf den Erfahrungen mit dem Ebolavirus basierten, wie mit Spikeprotein kloniertem
51 Adenovirus von Mensch oder Schimpansen, welches die ACE2 Rezeptor bindende Domäne (RBD)
52 enthält. Weitere Vakzine enthalten mRNA, die für das Spikeprotein codiert und in Lipidnanovesikel
53 verpackt sind und nach Zelleintritt das Spikeprotein synthetisieren. Beide Vakzine lösen eine starke
54 Immunantwort aus, die einige Monate möglicherweise über die T-Zell-Immunität wenige Jahre
55 anhält und bisher gegen alle Varianten (VOC – variant of concern) schützt, wenn auch nicht so
56 effizient wie gegen das Wuhan-Wildvirus. Die für Viren typisch häufigen Mutationen im Genom
57 werden sich fortsetzen und es werden sich weitere Varianten selektionieren, die sich bei
58 mangelnder Immunität verbreiten; bisher waren die Geimpften gegen alle Varianten geschützt,
59 wenn auch manchmal eingeschränkt.

60 Offen ist zurzeit wie lange der Impfschutz anhält, ob nur für mehrere Monate oder länger, ob eine
61 dritte Impfung verabreicht werden soll, um Geimpfte vor einer tödlichen Infektion mit neu
62 aufgetretenen Varianten zu schützen und wieweit eine wiederholte Impfung mit einem Adenovirus
63 basierten Impfstoff verträglich ist.

64

65 Schlüsselwörter: SARS-CoV-2, COVID-19, klonierte Adenovirus-Impfung, mRNA-Impfung,
66 Mutation, Coronavirus-Varianten

67

68 **SARS-CoV-2 vaccines and reaction of the immune system. Can the epidemic spread of the virus
69 be prevented by vaccination?**

70 **Abstract**

71 Since the end of 2019 a new coronavirus, SARS-CoV-2, first identified in Wuhan, China, is spreading
72 around the world partially associated with a high death toll. Besides hygienic measurements to
73 reduce the spread of the virus vaccines have been confected, partially based on the experiences
74 with Ebola virus vaccine, based on recombinant human or chimpanzee adenovirus carrying the
75 spike protein and its ACE2 receptor binding domain (RBD). Further vaccines are constructed by
76 spike protein coding mRNA incorporated in lipid nano vesicles that after entry in human cells
77 produce spike protein. Both vaccine types induce a strong immune response that lasts for months
78 possibly for T-cell immunity a few years. Due to mutations in the coronavirus genome in several
79 parts of the world variants selected, that were partially more pathogenic and partially easier
80 transmissible – variants of concern (VOC). Until now vaccinees are protected against the VOC, even
81 when protection might be reduced compared to the Wuhan wild virus.

82 An open field is still how long the vaccine induced immunity will be sufficient to prevent infection
83 and/or disease; and how long the time period will last until revaccination will be required for life
84 saving protection, whether a third vaccination is needed, and whether revaccination with an
85 adenovirus-based vaccine will be tolerated.

86

87 Key words: SARS-CoV-2, COVID-19, adenovirus-based vaccination, mRNA vaccination, coronavirus
88 variants

89

90

91

92

93

94

95

96 **Epidemiologie**

97 SARS-CoV-2 ist das dritte neu aufgetretene pathogene Coronavirus seit 2002, neben SARS-CoV
 98 (Severe Acute Respiratory Syndrome) und MERS-CoV (Middle East Respiratory Syndrome). Die
 99 Letalität des SARS-CoV-2 aus Wuhan war mit 0,1-1%, gegenüber dem SARS-CoV mit ca. 10% und
 100 dem MERS-CoV mit ca. 30% eher gering. Vier weitere Coronaviren zirkulieren in der deutschen
 101 Bevölkerung seit Jahrzehnten und lösen in der Kindheit typische Erkältungskrankheiten, teils auch
 102 Diarrhoen aus und induzieren nach wiederholter Exposition eine jahrzehntelang persistierende
 103 Immunität [1]. Eine solche erwerbbar Immunität könnte sich bei SARS-CoV-2 und seinen
 104 Varianten möglicherweise ebenfalls entwickeln. Die nach der jetzigen Einteilung bedrohlichen
 105 Virusvarianten (variants of concern, VOC) fasst **Tabelle 1** zusammen.

107 Tabelle 1: Bedrohliche Varianten von SARS-CoV-2 (Variants of concern, VOC). Nach Daten der WHO
 108 2021 [39] sowie des ECDC [40].

WHO Bezeichnung	Pango Stamm	GISAID Stamm	Next strain Stamm	zusätzlich mutierte Aminosäuren	Erstvorkommen
alpha	B.1.1.7	GRY GR/501Y.V1	20I (V1)	S 484K S 452R	United Kingdom Sep 2020
beta	B.1.351 B.1.351.2 B1.351.3	GH/501Y.V2	20H (V2)		Südafrika, Mai 2020
gamma	P.1 P.1. P.1.2	GR/501Y.V3	20J (V3)	S681H	Brasilien Nov 2020
delta	B.1.617.2 AY.1 AY.2	G/478K.V1	21A		Indien Okt 2020

109 Weitere Mutationen im Spikeprotein sind beschrieben. Variants of interest (VOI) sind derzeit von
 110 epsilon, in den USA prävalent, bis lambda eingeteilt worden; bedeutend unter diesen ist
 111 wahrscheinlich der Lambda-Stamm, bezeichnet auch als C.17; GR7452Q.V1; 20D, der im Dezember
 112 2020 in Peru identifiziert wurde.

114
 115 **Inkubationszeit, pathophysiologische Aspekte und Krankheitsverläufe**

116 Die Inkubationszeit von SARS-CoV-2 beträgt 2-20 Tage, im Mittel 3-9 Tage. Die Infektion verläuft
 117 zu etwa 40% asymptomatisch, besonders bei Kindern und Jugendlichen [2]. Sie kann aber auch
 118 schwere Krankheitsbilder auslösen, wie z.B. beatmungspflichtige Pneumonien, und tödlich
 119 verlaufen. Da auch Endothelzellen der Blutgefäße infiziert werden, kann jedes Organ betroffen
 120 sein, also neben der Lunge auch Leber, Niere, Nervensystem u.a. und über die Endothelzell-
 121 Schädigung ist die Krankheit COVID-19 auch mit Thrombosen assoziiert [3]. Thrombosen sind sehr
 122 selten auch als Nebenreaktion der Impfung vorgekommen [4]. Die Virusübertragung erfolgt durch
 123 Tröpfchen und Aerosol aus dem Atemtrakt. Speichel, Sputum, Tracheal- und Bronchialsekrete,
 124 sowie Stuhl und Urin sind ebenfalls infektiös [5]. Nach Übertragung gelangt das Virus über die
 125 ACE2-Rezeptoren in seine Wirtszellen im Nasen-Rachen-Raum, Bronchien, Lungenalveolen und
 126 Darmepithelien [6, 7]. ACE2-Rezeptoren befinden sich auch auf Endothelzellen der Gefäße. SARS-
 127 CoV-2 kann von symptomatisch Erkrankten, aber auch von asymptomatischen Virusträgern
 128 freigesetzt werden [2].

129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178

Reaktionen des humoralen und zellulären Immunsystems

Nach Infektion mit SARS-CoV-2 synthetisiert das Immunsystem außer bindende auch neutralisierende Antikörper gegen die Rezeptor-bindende Domäne (RBD) des Spikeproteins, das sich auf der Virusoberfläche befindet. Neutralisierende Antikörper (NT-AK) inaktivieren das Virus. Sie sind bei jungen Menschen deutlich wirksamer, d.h. neutralisieren den Erreger effektiver als die NT-AK von älteren Erwachsenen [8]. In einer anderen Studie wurde nur eine Korrelation mit der Schwere der Krankheit gefunden, nicht mit dem Alter [9]. Antikörper nach Erstinfektion persistieren meist nur für einige Monate [1,10]: ca. 3 Monate bei Kindern, und ca. 6 Monate bei älteren Patienten [11]. IgG-Antikörpertiter bei alten Menschen sind oft höher als bei jungen, sie binden aber die Viren weniger effektiv. Die überstandene Erstinfektion schützt nur begrenzt vor einer Zweitinfektion [12]. Die Zweitinfektion verursacht häufig nur ein mildes Krankheitsbild, kann jedoch auch schwer und tödlich verlaufen [13-16].

Bisher wurde nicht untersucht, welchen schützenden Effekt die im Blut gemessenen IgG-Antikörper auch in der Mukosa, neben dem exprimierten IgA, haben [17]. Die Impfung führt nicht zu einem Antibody Enhancement, d.h. dass die durch Impfung erzeugten Antikörper bei Infektion mit dem Wildvirus die immunologische Reaktion so stimulieren, dass Gefäßschäden und Schock ausgelöst werden, wie z.B. nach Impfung mit einem Dengue-Fieber-Virus-Impfstoff beschrieben [18].

Neben der humoralen Immunreaktion bildet das Immunsystem eine T-Lymphozyten-Immunität aus: nach Ergebnissen von Analysen an Patienten nach einer SARS-CoV und MERS-CoV-Infektion kann eine T-Zell-Immunität für 1-3, aber auch bis zu 10 Jahre bestehen bleiben [13, 19]. Nach Analyse der zweiten Epidemie in Manaus, Brasilien, war die Immunität, die durch die erste Epidemie ca. 9 Monate früher entstanden ist, nicht ausreichend, um ca. 40% der Bevölkerung vor Infektion mit der neuen Gamma-VOC (P.1) zu schützen, die eine höhere Bindungsaffinität an ACE2 zeigt und 2-fach besser übertragbar ist [16]. Es scheint gegen SARS-CoV-2 eine Immunantwort induziert zu werden, die der gegen die saisonalen Coronaviren ähnlicher ist als die, die gegen SARS-CoV und MERS-CoV erzeugt wird – mit der Einschränkung, dass die Ergebnisse von Langzeitstudien bisher nicht verfügbar sind [20].

Impfstoffe und ihre Wirkungsweise

Aus der Reaktion des Immunsystems auf eine SARS-CoV-2 Infektion folgt, dass auch bei Impfung eine schützende Immunität verzögert aufgebaut wird. In **Tabelle 2** sind die derzeit in Deutschland verfügbaren Impfstoffe dargestellt.

Die Angaben der Hersteller zu den Impfintervallen reichen von einer Einmalgabe (z.B. Johnson & Johnson) bis zu Mehrfachgaben, wobei die 2-malige Gabe bisher überwiegt. Bei den erhobenen Untersuchungen zur Induktion von neutralisierenden Antikörpern und einer T-Zell-Immunität ist davon auszugehen, dass wiederholt geimpft werden muss, um einen ausreichenden Immunschutz zu erreichen [1]. Die bewährten Impfschemata, wie sie z.B. von den FSME- und Hepatitis-B-Virus-Impfungen bekannt sind, beginnend mit einer Grundimmunisierung von 2 Impfungen im Abstand von 4 – 6 Wochen und einer Folgeimpfung nach 6 – 12 Monaten, könnten hierbei Vorbild sein [21].

179 Tabelle 2: Zurzeit verfügbare Vakzine gegen SARS-CoV-2 (Stand Juli 2021). Übersichten: [21–23].

Inhalt Wirksubstanz	Herstellungszeit	Benötigte Impfdosen	Wirksamkeit nach Herstellerangaben (%)	Hersteller
mRNA	schnell	2	ca. 95%	BNT162b2; Pfizer-BioNTech, GER-USA
		2	ca. 94%	mRNA-1273; Moderna, USA
DNA	schnell	2	Ca. 95%	INO-4800; Inovia, USA
Adenovirus Vektor plus S-Protein, nicht replizierend	mittel	1?	ca. 67%	Adv26. COV2.3; Johnson & Johnson, Jansen, BE-USA
		2	ca. 65%- 70% [7]	Chimp Adv5-CoV AZA1222*; AstraZeneca, Oxford, SWE-UK
		2	ca. 98%	Adv5, Adv26; Gamaleya, RUS Adv5; CanSino, CN
S1-Protein auf Nanopartikeln	mittel	2	ca. 89%	NVX-CoV2373; Novavax, USA
inaktiviertes Virus	aufwendig	2-3	ca. 50%	Sinovac; Biotech, Brasilien- Indonesien
		2-3	ca. 98%	Covaxin; Bharat Biotech, IN
		2	ca. 79%	Sinopharm, Wuhan, CN

180 BE = Belgien; CN = China; GER = Deutschland; IN = Indien; RUS = Russland; SWE = Schweden; UK =
181 United Kingdom.

182 * Aufgrund seltener aber schwerer thromboembolischer Nebenwirkungen empfahl die STIKO
183 zeitweise eine Impfstoffgabe für den AstraZeneca-Impfstoff vorzugsweise für Personen im Alter ab
184 60 Jahren, da zerebrale Venenthrombosen 4-16 Tage nach Impfung bei mehr als 170 Personen -
185 meist bei Frauen im Alter unter 60 Jahren - auftraten [24]. Für die Verabreichung der zweiten
186 Impfung wird ein Zeitabstand von 8-12 Wochen empfohlen, teils nach einem DNA-Impfstoff ein
187 mRNA-Impfstoff, bezeichnet als heterologe Impfung. Da sich nach Kenntnis von Verfügbarkeit und
188 Nebenwirkungen die STIKO-Empfehlungen ändern, wird empfohlen, die weiteren Mitteilungen der
189 STIKO zu beachten. <https://www.rki.de/DE/Content/Infekt/Impfen/ImpfungenAZ/COVID-19/COVID-19.html>
190

191
192 Die derzeit verfügbaren und zum Teil noch für die Europäische Union im Zulassungsverfahren
193 befindlichen Impfstoffe können in folgender Weise gegliedert werden:

194
195 **Gruppe der genbasierten Impfstoffe**

196 **1.) Messenger-RNA (m-RNA) Vakzine**

- 197 • BioNTech-Pfizer-Impfstoff (Comirnaty):
198 Etwa 30 µg der in Lipidnanopartikeln eingebetteten mRNA (BNT162b2), kodierend für das
199 Spike-Protein mit Rezeptor-bindender Domäne (RBD), wird intramuskulär verabreicht, von
200 Zellen wie Makrophagen, Fibrozyten und Myozyten aufgenommen und im Zytoplasma in
201 Spike-Protein umgeschrieben. Die freigesetzten Proteine induzieren eine neutralisierende
202 Immunantwort in allen geprüften Altersklassen und führen zu einer CD4+- und CD8+-T-
203 Zell-Antwort. Der Impfstoff muss längerfristig bei minus 60-90°C gelagert werden und ist
204 bis zu 2 Wochen bei minus 25-15°C haltbar.
205
206 • Moderna-mRNA-1273-Impfstoff (Spikevax):

207 Eine Dosis enthält 100 µg mRNA in SM-102 Lipidnanopartikeln eingebettet und wird
208 intramuskulär verabreicht. Das durch die mRNA kodierte und synthetisierte Spike-Protein
209 induziert eine B-Zell-Immunantwort mit der Bildung neutralisierender und bindender
210 Antikörper sowie eine T-Zell-Immunantwort. Der Impfstoff muss bei minus 20°C dauerhaft
211 gelagert werden.

212

213 • Curevac-mRNA-Impfstoff, Tübingen:

214 Derzeit weiter im beschleunigten Zulassungsverfahren während der noch laufenden
215 klinischen Prüfung. Er erreichte bisher eine Wirksamkeit von < 50 % und ist damit zur
216 Impfung nicht geeignet (Stand Juli 2021).

217

218 **2.) Adenovirus-Vektor-Vakzine**

219 • Johnson & Johnson-Janssen-Cilag-Impfstoff (COVID-19 vaccine Janssen):

220 Aus Zellkultur gereinigtes, nicht vermehrungsfähiges Schimpansen-Adenovirus 26, das
221 rekombinantes S1-Glykoprotein enthält. Der intramuskulär verabreichte Impfstoff führt
222 zu einer B- und T-Zell-Immunreaktion – gegen beide Komponenten, Virus und S1-Protein.

223

224 • AstraZeneca-Impfstoff (Vaxzevira):

225 Im wesentlichen S1-RBD-Glykoprotein, das mit rekombinantem Schimpansen-Adenovirus
226 in Zellkultur vermehrt und gereinigt wird. Die Menge der RBD-Komponente pro Impfdosis,
227 die intramuskulär verabreicht wird, entspricht der von etwa 10⁸ Viruspartikeln. Ein Teil
228 der Nebenwirkungen wird durch zelluläre Proteine wie heat shock protein (HSP 90)
229 verursacht, die sich an das Adenovirus anheften [25].

230

231 • Gamaleya-Impfstoff (Sputnik V):

232 Aus Zellkultur gereinigtes S1-Glykoprotein aus rekombinantem menschlichem Adenovirus
233 5/Adenovirus-26-Hybrid-Virus.

234

235 **Gruppe der S1- tragenden Nanopartikel**

236 • NVX-CoV2373 (Novavax USA):

237 Der Impfstoff enthält das S1-Glykoprotein an Nano-Partikel gekoppelt und induziert eine
238 starke Immunreaktion. Das Protein ist in Polysorbat 80 emulgiert und enthält das
239 Membrane-Protein als Adjuvans.

240

241 **Gruppe der inaktivierten Virusvakzine**

242 • Sinovac Biotech: In Brasilien/Indonesien hergestellt und dort in Prüfung.

243 • Sinopharm, Wuhan, China: Zur Wirksamkeit sind kaum Daten publiziert bzw. zugänglich.

244 • Covaxin, Bharat Biotech, Indien: ist wirksam, mit hoher Effizienz; weit über 10 Millionen
245 Impfdosen wurden verabreicht.

246

247 Bisher ist keines der in China hergestellten Vakzine in Deutschland zugelassen, nur eins von der

248 WHO. Die verfügbaren Daten sind in der Publikation von Mallapaty [26] zusammengestellt.

249

250 **Wirksamkeit der derzeit verfügbaren Impfstoffe gegen die bedrohlichen Varianten (VOC)**

251 In kurzer Zeit sind aufgrund des häufigen Auftretens von Mutationen bei der Replikation von SARS-
252 CoV-2 eine Reihe von Mutanten entstanden, von denen die englische Alpha-Variante (B.1.1.7) im
253 europäischen Raum bisher am weitesten verbreitet ist. Die südafrikanische Beta-Variante
254 (B.1.35.1), die portugiesisch-brasilianische Gamma-Variante (P.1) und die indische Variante Delta
255 (B.1.677) sind in Deutschland nachgewiesen, und bezüglich von Kontagiosität und schnellerer

256 Ausbreitung hat die (indische) Delta-Variante mit etwa 50-75% Prävalenz die (englische) Alpha-
257 Variante verdrängt. Die dauerhafte Virulenz dieser Virusmutanten ist bisher nicht sicher
258 beurteilbar, die portugiesisch-brasilianische Variante Gamma ist virulenter als das ursprüngliche
259 Wuhan-Wildvirus [16].

260

261 Folgende Aminosäure-Mutationen führen zu einer Steigerung der Infektiosität und
262 Übertragbarkeit von SARS-CoV-2 [27–29]. Der 1-Buchstaben-Code der Aminosäuren wurde
263 verwendet.

- 264 • D614G, N439K, Y453F, D936Y, S939F, S943T und zusätzlich weitere Mutationen wie in der
- 265 – Alpha-Linie (B.1.1.7): 17 Mutationen, wie N501Y, P681H, Deletion von Histidin 69
266 und Valin 70
- 267 – Beta-Linie (B.1.3.51): 21 Mutationen, wie N501Y, E484K, K417N, und eine
268 Deletion in ORF1b
- 269 – Gamma-Linie (P.1): 17 Mutationen, wie N501Y, E484K, K417N und eine weitere
270 Deletion im ORF1b
- 271 – Delta-Linie (B.1.167.2): 13 Mutationen, wie E484Q, L452R in der RBD

272

273 Zur Wirksamkeit der derzeit verfügbaren Impfstoffe gegen die neuen Varianten liegen bisher nur
274 begrenzte Daten vor. Der BioNTech-Pfizer-Impfstoff und die meisten anderen Impfstoffe schützen
275 vor Erkrankungen durch die Alpha- und Gamma-Variante. Der Nanopartikelimpfstoff „Novavax“
276 soll eine besonders ausgeprägte Wirksamkeit haben. Der Schutz der Vektor-Impfstoffe von
277 AstraZeneca und Johnson & Johnson gegen die Beta-Variante scheint jedoch reduziert [30]. Die in
278 den USA zirkulierende Epsilon-Variante (B.1.427 und B.1.429) scheint schwerer neutralisierbar zu
279 sein [31].

280 Messenger-RNA-Impfstoffe haben den grundsätzlichen Vorteil, dass durch Nukleotidadaptation
281 der RNA an die geänderte Gensequenz der Virusmutanten innerhalb von wenigen Monaten
282 wirksame Vakzine gegen die neuen Varianten hergestellt werden können.

283

284 **Bisher beobachtete mögliche Nebenwirkungen**

285 Als lokale Nebenwirkungen traten auf: Schmerzen, Rötung und Schwellung an der Einstichstelle.

286 Allgemeine Nebenwirkungen sind Fieber, Kopf- und Gliederschmerzen, Abgeschlagenheit,
287 allergische Reaktionen (Exantheme)

288 In England wurden bei 2 als schwere Allergiker bekannten Personen eine anaphylaktische Reaktion
289 beobachtet [32]. Außerdem sind bei mit Adenovirus-Vektorimpfstoff Geimpften in Europa
290 Thrombosen aufgetreten, die die zerebralen Sinusvenen, das Splanchnikusgebiet und arterielle
291 Gefäße betrafen und die bei > 100 Geimpften tödlich verliefen [24, 33]. Bei der Mehrzahl der
292 Betroffenen waren Antikörper gegen Plättchenfaktor 4 nachweisbar, wie sie auch bei der Heparin-
293 induzierten Thrombopenie vom Typ 2 (HIT2) vorkommen und zu Thrombopenie und Ausbildung
294 arterieller oder venöser Thromben führen können [33, 34]. Auch wenn ein Kausalzusammenhang
295 wegen der geringen Zahl bei über 10 Millionen Geimpften und der Mehrzahl von weiblichen
296 Geimpften in der betroffenen Altersgruppe bisher nicht eindeutig beweisbar ist, so handelt es sich
297 bei diesen sehr seltenen, aber schwerwiegenden Ereignissen doch um ein wichtiges Signal,
298 aufgrund dessen die Vektorimpfstoffe in einigen Ländern derzeit nur mit Einschränkungen
299 empfohlen werden. Eine Myokarditis nach Impfung mit dem mRNA-Impfstoff von BioNTech-Pfizer
300 wurde bei >25 männlichen Geimpften beobachtet [35]. Eine weitere, seltene Nebenwirkung nach
301 Gabe des Adenovirus-klonierten Spike-Proteins (Vaxzevria und Janssen-COVID-19 Vakzine) ist das
302 Capillary-leak-Syndrome (CLS), das zur Hypovolämie führt [36, 37]. Wenn dieses Krankheitsbild
303 anamnestisch einmal aufgetreten ist, darf der Impfstoff nicht verabreicht werden.

304 Inwieweit Adjuvanzen, die zur Immunogenitätssteigerung eingesetzt werden müssen, die
305 Nebenwirkungsrate erhöhen, ist bisher nicht untersucht. In die Diskussion geratene mögliche
306 Nebenwirkungen von Impfstoffen, wie z.B. Störung der Blutgerinnung, können nur auf der Basis
307 einer großen Zahl fundierter wissenschaftlicher Daten beurteilt werden [33, 34]. Alle bisher von
308 der Europäischen Zulassungsbehörde für Impfstoffe (EMA) zugelassenen Anti-SARS-CoV-2-Vakzine
309 wurden an mehreren Zehntausend Probanden erprobt, sodass der Sicherheitsaspekt besonders
310 überprüft wurde. Zu Langzeitnebenwirkungen sind derzeit keine begründeten Aussagen möglich,
311 da aufgrund der begrenzten Zeitspanne seit Einführung der Impfung keine Langzeitbeobachtungen
312 vorliegen.

313

314 **Ist eine Unterbrechung der Pandemie möglich?**

315 Es ist absolut wünschenswert, mittels einer nebenwirkungsarmen Impfung die weitere
316 Ausbreitung des Virus und die assoziierte Todesrate einzuschränken bzw. zu beenden. Dass die
317 epidemische Ausbreitung von SARS-CoV-2 und seiner Varianten durch Impfung vermindert bzw.
318 verhindert werden kann, hat sich an den Ergebnissen der Impfung in Israel gezeigt, wo derzeit
319 bereits $\geq 60\%$ der Bevölkerung durchgeimpft sind. Es zeichnet sich in Israel bereits ein guter Erfolg
320 hinsichtlich einer Entschleunigung der Virusausbreitung ab.

321 Bis zum Wirksamwerden einer weitgehenden Durchimpfung sind in Deutschland und vielen
322 anderen Ländern weiterhin Maßnahmen zur Kontaktbeschränkung notwendig - vor allem, um eine
323 Überlastung des Gesundheitssystems zu vermeiden. Insbesondere muss ein Szenario verhindert
324 werden, bei dem eine adäquate intensivmedizinische Versorgung sowohl von COVID-19- wie auch
325 von anderen Patienten nicht mehr gewährleistet ist.

326

327 Eine Abnahme der Infektiosität des Virus ist in Deutschland im Sommerhalbjahr durch höhere
328 Temperaturen und vermehrte UV-Strahlung denkbar [38]. Die Reduktion der Übertragungen in
329 Deutschland im Frühjahr 2020 wird neben dem vermehrten Aufenthalt im Freien und anderen
330 Faktoren darauf zurückgeführt, aber dieser Effekt kann nicht besonders ausgeprägt sein, da in
331 Ländern mit erheblich höheren Sommertemperaturen und UV-Strahlung wie Italien, Spanien und
332 Israel die Virusverbreitung in Nichtgeimpften nicht reduziert ist.

333

334 **Kernaussagen**

- 335 • Durch sofortige Nukleinsäure-Sequenzierung des SARS-CoV-2 im Jahr 2019 sind innerhalb von
336 wenigen Monaten Antikörper- und Virusnachweis-Diagnostik und wirksame Impfstoffe
337 entwickelt worden.
- 338 • Unter dem Druck, die Bevölkerung vor dem SARS-CoV-2 zu schützen, wurde Adenovirus-
339 Vektor-Impfstoff-Konzepte und mRNA-Impfstoffe verfeinert und neu entwickelt und die
340 Voraussetzungen geschaffen, dass für Impfwillige genügend Impfstoff zur Verfügung steht.
- 341 • Die epidemiologische Überwachung der SARS-CoV-2 Ausbreitung hat zur zügigen
342 Identifizierung von Varianten geführt, die als bedrohliche (variant of concern, VOC) und als zu
343 beachtende (variant of interest, VOI) Varianten bezeichnet werden.
- 344 • Durch 2-malige Impfung mit Spike-Protein von SARS-CoV-2 kann eine ausreichende B-
345 Lymphozyten (neutralisierende Antikörper) und T-Lymphozyten (zytotoxische T-Zellen) -
346 Reaktion erzeugt werden, um auch die meisten Varianten zügig zu eliminieren. Die
347 Notwendigkeit und der Nutzen einer dritten Impfung zur Langzeitimmunität bleibt abzuklären.
- 348 • Eine erfolgreiche SARS-CoV-2 Impfung schützt nicht vollständig vor Reinfektion oder
349 Neuinfektion, aber vor schwerem und tödlichem Krankheitsverlauf. Deswegen sind die
350 simplen Hygienemaßnahmen wie Lüftung, Abstand, Händewaschen, Raumhygiene und
351 Masken zur Prävention weiter erforderlich.

352

353 **Literatur**

- 354 [1] Edridge AWD, Kaczorowska J, Hoste ACR, Bakker M, Klein M, Loens K et al. Seasonal coronavirus
355 protective immunity is short lasting: *Nat Med* 2020; 26: 1691-1693
356
- 357 [2] Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection – a narrative review. *Ann*
358 *Int Med* 2021; doi:10.7326/20-3012
359
- 360 [3] Siddiqi HK, Libby P, Ridke PM. COVID-19 – a vascular disease. *Trends Cardiovascul Med* 2021,
361 31: 1-5
362
- 363 [4] Long B, Bridwell R, Gottlieb M. Thrombosis with thrombocytopenia syndrome associated with
364 COVID-19 vaccines. *Am J Emerg Med* 2021; 49: 58-61
365
- 366 [5] Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 transmission and pathogenesis. *Trends*
367 *Immunol* 2020; 41: 1100-1115
368
- 369 [6] Li Y, Zhou W, Yang L, You R. Physiological and pathological regulation of ACE2, the SARS-CoV-2
370 receptor. *Pharmacol Res* 2020; 104833
371
- 372 [7] Koliaraki V, Prados A, Armaka M, Kollias G. The mesenchymal context in inflammation,
373 immunity and cancer. *Nat Immunol* 2020; 2: 974-982
374
- 375 [8] Pierce CA, Preston-Hurlburt P, Dai Y, Aschner CB, Cheshenko N, Galen B et al. Immune response
376 to SARS-CoV-2 infection in hospitalized pediatric and adult patients. *Sci Transl Med* 2020; 12:
377 eabd5487
378
- 379 [9] Trinité B, Tarrés-Freixas F, Rodon J, Pradevas E, Urrea V, Marfil S et al. SARS-CoV-2 infection
380 elicits a rapid neutralization antibody response. *Sci Rep* 2021; 11: 2608
381
- 382 [10] Wölfel R, Corman VM, Guggemoos W, Sellmaier M, Zange S, Müller MA et al. Virological
383 assessment of hospitalized patients with COVID-2019. *Nature* 2020; 581: 465-469
384
- 385 [11] Iyer AS, Jones FK, Nodoushani A, Kelly M, Becker M, Slater D et al. Persistence and decay of
386 human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-
387 19 patients. *Sci Immunol* 2020; doi: 10.1126/sciimmunol.abe0367
388
- 389 [12] Hansen CH, Michlmayr D, Gubbels SM, Mølbak K, Ethelberg S. Assessment of protection
390 against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a
391 population-level observational study. *Lancet* 2021; 397:1204-1212
392
- 393 [13] Iwasaki A. What reinfections mean for COVID-19. *Lancet Infect Dis* 2021; 21: 3-5
394
- 395 [14] Vitale J, Mumoli N, Clerici P, De Paschale M, Evangelista I, Cei M, Mazzone A. Assessment of
396 SARS-CoV-2 reinfection 1 year after primary infection in a population in Lombardy, Italy. *JAMA Int*
397 *Med* 2021; 2212959
398
- 399 [15] Stokel-Walker G. What we know about covid-19 reinfection so far. *BMJ (Br Med J)* 2021; 372:
400 n99
401
- 402 [16] Faria NR, Mellan TK, Whittaker C, Claro IM, Candido DDS, Mishra S et al. Genomics and
403 epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* 2021; 372: 815-821
404

405 [17] Xue M, Zhang T, Hu H, Huang Z, Zhen Y, Liang Y et al. Predictive effects of IgA and IgG
406 combination to assess pulmonary exudation progression in COVID-19 patients. *J Med Virol* 2021;
407 93: 1443-1448
408

409 [18] Halstead SB, Katzelnick L. COVID-19 vaccines: should we fear ADE? *J Infect Dis* 2020; 222: 1946-
410 1950
411

412 [19] Liu W, Zhao M, Lin K, Wong F, Tan W, Gao GF. T-cell immunity of SARS-CoV: implications for
413 vaccine development against MERS-CoV. *Antivir Res* 2017; 137: 82-92
414

415 [20] Rodda LB, Netland J, Shehata L, Pruner KB, Moraski PA, Thouvenel CD et al. Functional SARS-
416 CoV-2-specific immune memory persists after mild COVID-19. *Cell* 2021; 184: 169-183
417

418 [21] Poland GA, Ovsyannikova IG, Kennedy RB. SARS-CoV-2 immunity: review and application to
419 phase 3 vaccine candidates. *Lancet* 2020; 398; 1595-1606
420

421 [22] Castells MC, Phillips EJ. Maintaining safety with SARS-Co-2 vaccines. *N Engl J Med* 2021: 384:
422 643-649
423

424 [23] Chakraborty S, Mallajosyula V, Tato CH, Tan GS, Wang TT. SARS-CoV-2 vaccines in advanced
425 clinical trials: where do we stand. *Adv Drug Delivery R*
426

427 [24] Dakay K, Cooper J, Bloomfield J, Overby P, Mayer SA, Nuoman R et al. Cerebral venous
428 thrombosis as a presentation of COVID-19. *J Stroke Cerebrovascul Dis* 2021; 30: 105434
429

430 [25] Dalidowska I, Gazi O, Prybylski M, Bieganski P. Heat shock protein 90 chaperones E1A early
431 protein of adenovirus 5 and is essential for replication of the virus. *Int J Mol Sci* 2021; 22: 2020
432

433 [26] Mallapaty S. China's COVID-19 vaccines are going global – but questions remain. *Nature* 2021;
434 593: 178-179
435

436 [27] Li Q, Wu J, Nie J, Zhang L et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity
437 and antigenicity. *Cell* 2020; 182: 1284-1294
438

439 [28] Leung K, Shum MHH, Leung GM, Lam TTY, Wu T. Early transmissibility assessment of the N501Y
440 mutant strain of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveill*
441 2021; 26: pii=2002106
442

443 [29] Weber S, Ramirez C, Doerfler W. Signal hotspot mutations in SARS-CoV-2 genome evolve as
444 the virus spreads and actively replicates in different parts of the world. *Virus Res* 2020; 289: 198170
445

446 [30] Emary KRW, Golubchik T, Aley PK, Ariani CV, Angus B, Bibi S et al. Efficacy of ChAdOx1 nCoV-
447 19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory
448 analysis of a randomised controlled trial. *Lancet* 2021; 397: 1351-1362
449

450 [31] McCallum M, Bassl J, De Marco A, Chen A, Walls AC, Julio J et al. SARS-CoV-2 immune evasion
451 by the B.1.427/B.1.429 variant of concern. *Science* 2021; eabl7994, doi: 10.1126/science.abi7994
452

453 [32] GOV UK. Medicines and Healthcare Product, Regulatory Agency. Information for UK recipients
454 on COVID-19 vaccine Astra-Zeneca (Regulation 174): <https://www.gov.uk/publications>.
455

456 [33] McGonagle D, De Marco G, Bridgewood C. Mechanisms of immunothrombosis in vaccine-
457 induced thrombotic thrombocytopenia (VITT) compared to natural SARS-CoV-2 infection. *J*
458 *Autoimmunity* 2021; 121: 102662

459
460 [34] Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic
461 Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. *N Engl J Med.* 2021; 384: 2092-2101
462
463 [35] Montgomery J, Ryan M, Engler R, Hoffman D, McClenathon B, Collins L et al. Myocarditis
464 following immunization with mRNA COVID-19 vaccines in members of the US military. *JAMA*
465 *Cardiol* 2021; doi: 10.1001/jamacardio.2021.2833
466
467 [36] Bahloul M, Ketala W, Lahyemi D, Mayoufi H, Kotti A, Smaoui F. Pulmonary capillary leak
468 syndrome following COVID-19 virus infection. *J Med Virol* 2021; 93: 94-96
469
470 [37] Case R, Ramaniuk A, Martin P, Simpson PJ, Harden C, Ataya A. Systemic Capillary Leak
471 Syndrome secondary to coronavirus disease 2019. *Chest* 2020; 158: e267-e268
472
473 [38] Heilingloh CS, Aufderhorst UW, Schipper L, Dittmer U et al. Susceptibility of SARS-CoV-2 to UV
474 irradiation. *Am J Infect Control* 2020; 48: 1273-1275
475
476 [39] World Health Organization. WHO announces simple, easy-to-say labels for SARS-CoV-2. Im
477 Internet (letzter Zugriff: 13. Juli 2021, [https://www.who.int/news/item/31-05-2021-who-an](https://www.who.int/news/item/31-05-2021-who-announces-simple-easy-to-say-labels-for-sars-cov-2-variants-of-interest-and-concern)
478 [nounces-simple-easy-to-say-labels-for-sars-cov-2-variants-ofinterest-](https://www.who.int/news/item/31-05-2021-who-announces-simple-easy-to-say-labels-for-sars-cov-2-variants-of-interest-and-concern)
479 [and-concern](https://www.who.int/news/item/31-05-2021-who-announces-simple-easy-to-say-labels-for-sars-cov-2-variants-of-interest-and-concern)
480
481 [40] European Centre for Disease Prevention and Control. SARS-CoV-2 variants of concern as of 8
482 July 2021. Im Internet (letzter Zugriff: 13. Juli 2021, [https://www.ecdc.europa.eu/](https://www.ecdc.europa.eu/en/covid-19/variants-concern)
483 [en/covid-19/variants-concern](https://www.ecdc.europa.eu/en/covid-19/variants-concern)
484
485
486



OPEN

A semi-automated, isolation-free, high-throughput SARS-CoV-2 reverse transcriptase (RT) loop-mediated isothermal amplification (LAMP) test

Jonas Schmidt^{1,2,3}, Sandro Berghaus¹, Frithjof Blessing^{1,2}, Folker Wenzel², Holger Herbeck¹, Josef Blessing¹, Peter Schierack^{3,4}, Stefan Rödiger^{3,4} & Dirk Roggenbuck^{3,4}✉

Shortages of reverse transcriptase (RT)-polymerase chain reaction (PCR) reagents and related equipment during the COVID-19 pandemic have demonstrated the need for alternative, high-throughput methods for severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)-mass screening in clinical diagnostic laboratories. A robust, SARS-CoV-2 RT-loop-mediated isothermal amplification (RT-LAMP) assay with high-throughput and short turnaround times in a clinical laboratory setting was established and compared to two conventional RT-PCR protocols using 323 samples of individuals with suspected SARS-CoV-2 infection. Limit of detection (LoD) and reproducibility of the isolation-free SARS-CoV-2 RT-LAMP test were determined. An almost perfect agreement (Cohen's kappa > 0.8) between the novel test and two classical RT-PCR protocols with no systematic difference (McNemar's test, $P > 0.05$) was observed. Sensitivity and specificity were in the range of 89.5 to 100% and 96.2 to 100% dependent on the reaction condition and the RT-PCR method used as reference. The isolation-free RT-LAMP assay showed high reproducibility (Tt intra-run coefficient of variation [CV] = 0.4%, Tt inter-run CV = 2.1%) with a LoD of 95 SARS-CoV-2 genome copies per reaction. The established SARS-CoV-2 RT-LAMP assay is a flexible and efficient alternative to conventional RT-PCR protocols, suitable for SARS-CoV-2 mass screening using existing laboratory infrastructure in clinical diagnostic laboratories.

Severe acute respiratory syndrome corona virus type 2 (SARS-CoV-2), an RNA virus that gives rise to coronavirus disease 2019 (COVID-19), has caused a major pandemic since it was first described in late 2019¹. To better control and monitor the spread of COVID-19, the combined deployment of comprehensive surveillance, diagnostics, research, clinical treatment and vaccine development are required².

Since SARS-CoV-2 is highly contagious, the main goal of laboratory diagnostics should be to identify infected individuals as quickly as possible. To accomplish this, amplification of viral nucleic acid plays a fundamental role in assay strategies that have been established in many clinical diagnostic laboratories^{2,3}. In general, the SARS-CoV-2 genome consists of 14 open reading frames (ORF) which involve possible targets for diagnostic nucleic acid-amplification assays^{4,5}.

Besides reverse transcription (RT) polymerase chain reaction (PCR) (which is the gold standard method), isothermal amplification reactions are alternative techniques for nucleic acid detection^{6,7}. Loop-mediated isothermal amplification (LAMP) has frequently been applied for SARS-CoV-2 detection^{2,8,9}. With this technique, using four or six different target specific primers, and a *Bst* DNA polymerase with a high strand displacement activity, it is possible to detect viral nucleic acids with high sensitivity and specificity¹⁰. The main advantage of LAMP assays is the short isothermal reaction time (typically 10–25 min), which makes it faster and easier to

¹Institute for Laboratory Medicine, Singen, Germany. ²Furtwangen University, Faculty of Medical and Life Sciences, Villingen-Schwenningen, Germany. ³Institute of Biotechnology, Faculty Environment and Natural Sciences, Brandenburg University of Technology Cottbus-Senftenberg, Senftenberg, Germany. ⁴Faculty of Health Sciences Brandenburg, Brandenburg University of Technology, Cottbus-Senftenberg, Universitätsplatz 1, 01968 Senftenberg, Germany. ✉email: dirk.roggenbuck@b-tu.de

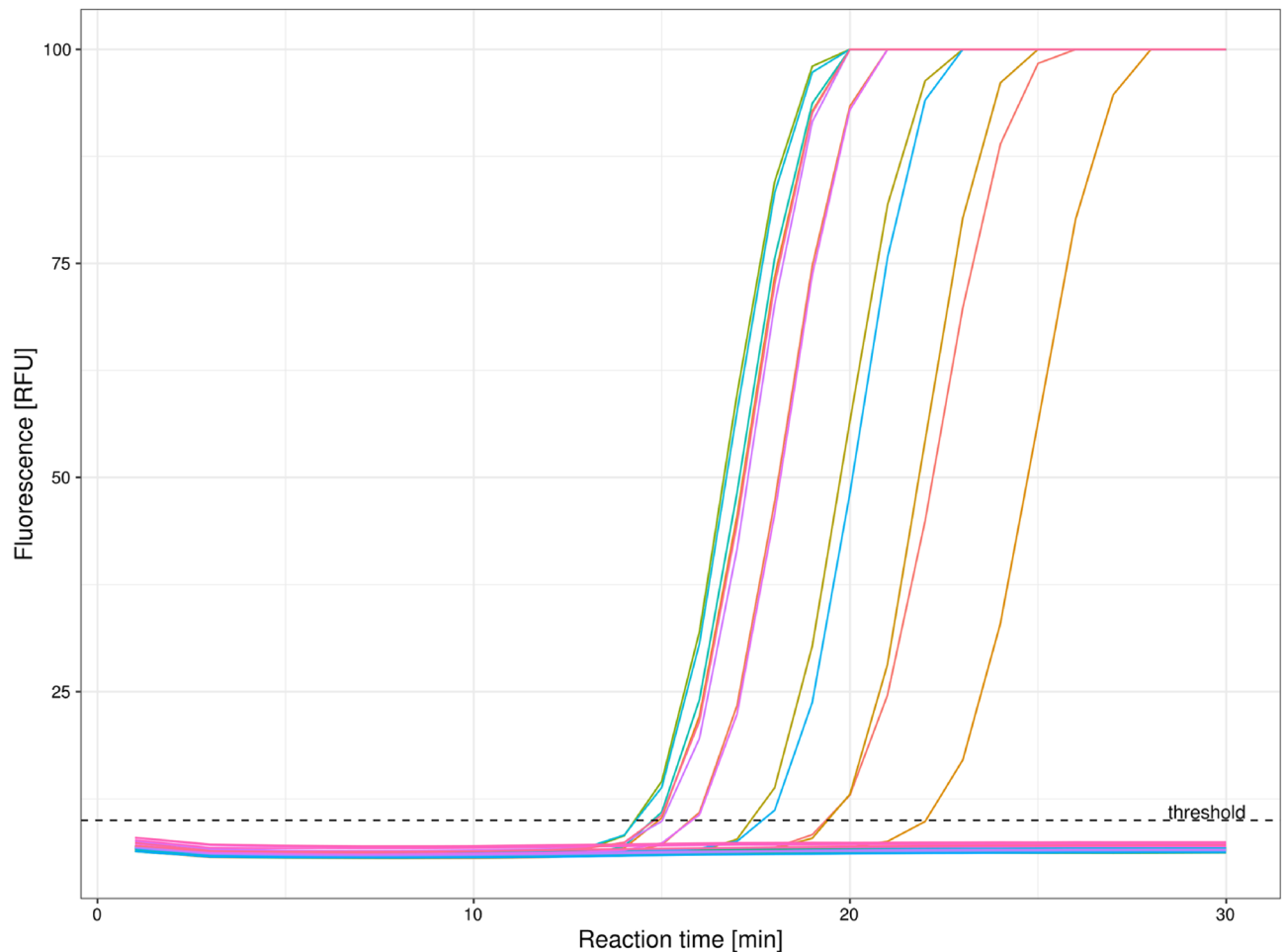


Figure 1. Amplification curves of SARS-CoV-2 RT-LAMP runs using an intercalating fluorescent dye for detection. During the 30 min isothermal incubation (65 °C), the fluorescence signal is read minutely on the FAM channel. A threshold is applied to identify positive samples and calculate the threshold time (T_t) value.

perform compared to conventional RT-PCR¹⁰. There are many different read-out methods available including (real-time) fluorescence detection, colorimetric detection, turbidity, gel electrophoresis, and sequencing^{10–12}. Such assays can therefore be applied under conditions where only basic laboratory equipment is available¹³.

To detect and estimate the amount of DNA in quantitative PCRs, the number of cycles is determined from which the threshold is significantly exceeded¹⁴. In contrast, isothermal amplification reactions are monocyclic. The amount of DNA is not determined by the number of cycles, instead it is determined by how much time is needed to exceed the background signal¹⁵. Since the duration of the cycle is generally constant, virtually any real-time PCR system can be used for the measurement. The specification of PCR-typical quantification points, such as the threshold cycle (Ct) or threshold time (T_t) value have been established in the literature.

Regarding SARS-CoV-2 diagnostics, most of the currently available LAMP protocols focus on qualitative colorimetric detection since interpretation is very easy by eye^{9,16–19}. However, the temporary shortage of PCR supplies during the COVID-19 pandemic has demonstrated the need for alternatives to RT-PCR protocols, even in diagnostic laboratories with high throughput requirements and sophisticated laboratory equipment.

We have established a flexible and robust high-throughput SARS-CoV-2 RT-LAMP protocol, which can not only be used in combination with RNA isolation from nasopharyngeal swabs, but also with simple enzymatic digestion for sample preparation. To evaluate our protocol's performance, we have compared it to two conventional RT-PCR protocols using clinical samples. To enable semi-automated high-throughput processing, we further established the protocol on a liquid handling station.

Results

Applying the established SARS-CoV-2 RT-LAMP protocol, it was possible to distinguish positive and negative samples confirmed by standard RT-PCR test via fluorescence detection (Fig. 1). Positive samples showed a sigmoid increase of the fluorescence intensity over the isothermal incubation period. A banding pattern characteristic for LAMP reactions was observed by using conventional gel electrophoresis (Supplementary Fig. S1). The negative samples, as well as the no-template control, were below a predefined, operator adjustable threshold. The later can be used to calculate the T_t values of positive samples similar to calculation of Ct values in

	RT-LAMP (isolation)—in-house RT-PCR	RT-LAMP (isolation)—Labsystems RT-PCR	RT-LAMP (isolation)—in-house RT-PCR (liquid handler)	RT-LAMP (isolation-free)—in-house RT-PCR
Sample size	70	70	188	65
True positive	34	34	22	39
True negative	34	32	166	25
False positive	0	0	0	1
False negative	2	4	0	0
Cohen's Kappa (κ)	0.94	0.89	1	0.97
McNemar's test (P)	0.55	0.22	NA ^a	0.79
Sensitivity (%) (95% CI) ^b	94.4 (81.3–99.3)	89.5 (75.2–97.1)	100 (84.6–100)	100 (91.0–100)
Specificity (%) (95% CI) ^b	100 (89.7–100)	100 (89.1–100)	100 (97.8–100)	96.2 (80.4–99.9)

Table 1. Method comparison of the SARS-CoV-2 RT-LAMP with RNA isolation with (I) the in-house RT-PCR protocol, (II) the commercial Labsystems RT-PCR kit and (III) the in-house RT-PCR using a liquid handler. Further, the method comparison between the isolation-free protocol of the RT-LAMP and the in-house RT-PCR is shown. ^aNot available. McNemar's test cannot be calculated for perfect agreement. ^bConfidence interval.

conventional RT-PCR. The external control targeting the human beta actin gene (ACTB) also showed a sigmoid increase in fluorescence intensity over time for all investigated clinical samples (Supplementary Fig. S2). The following method comparison experiments were performed without the additional external control reaction mix to increase throughput and efficiency.

After initial establishment, various method comparison experiments were performed using samples from individuals with suspected COVID-19 to compare the RT-LAMP assay with conventional RT-PCR protocols as standard methods for the detection of SARS-CoV-2 (Table 1).

Using 70 isolated RNA samples, the RT-LAMP protocol was compared with our in-house standard RT-PCR protocol targeting the E gene, and with a commercial PCR kit targeting the E, N and ORF1ab genes. An almost perfect agreement (Cohen's kappa [κ] > 0.8) between the RT-PCR protocols and the RT-LAMP protocol was observed with no systematic difference (McNemar's test, $P > 0.05$). Sensitivity was 94.4% (95% confidence interval [CI] 81.3–99.3%) when using the in-house RT-PCR as a reference, and 89.5% (CI 75.2–97.1%) for the commercial RT-PCR kit as a reference. Specificity was 100% in both cases as no false positive results were observed.

The Tt values of the RT-LAMP assay showed a significant positive correlation with the RT-PCR data for all different targets (Rho [ϕ] > 0.8, $P < 0.001$) (Fig. 2A–D). Samples classified as false-negative (2/70) by the RT-LAMP test had Ct values in the range of 20–30 in the in-house RT-PCR. The Ct values of the samples classified as false-negative (4/70) in comparison with the commercial RT-PCR kit were in the range of 20–30 in one case, and above 30 in two cases. The remaining false-negative sample showed only a signal for the N gene target with a Ct value above 35 in the commercial RT-PCR.

In order to increase sample throughput, and decrease manual hands-on time, we adapted the SARS-CoV-2 RT-LAMP workflow on a liquid handling system. To evaluate the performance a method comparison to the in-house RT-PCR was performed using 188 isolated RNA samples. A perfect agreement ($\kappa = 1$) between the two methods was observed resulting in 100% (CI 84.6–100%) sensitivity and 100% (CI 97.8–100%) specificity of the SARS-CoV-2 RT-LAMP (Table 1).

To reduce the processing time further, an RNA isolation-free sample preparation step for nasopharyngeal swabs using proteinase K digestion was added to the RT-LAMP protocol. A method comparison to the in-house RT-PCR was carried out using 65 nasopharyngeal swabs. Near perfect agreement ($\kappa = 0.97$) was observed between both methods with no systematic difference (McNemar's test, $P = 0.79$) (Table 1). Of the 65, a single sample was classified false-positive by the RT-LAMP assay with a Tt value above 15. A significant positive correlation of the RT-LAMP Tt values and the RT-PCR Ct values was observed ($\phi = 0.89$, $P < 0.001$) (Fig. 2E). Sensitivity and specificity of the isolation-free RT-LAMP protocol compared to the in-house RT-PCR protocol were 100% (CI 91.0–100%) and 96.2% (CI 80.4–99.9%), respectively.

Using reference material with a known amount of SARS-CoV-2 genome copies and pools of positive and negative samples, performance characteristics of the isolation-free SARS-CoV-2 RT-LAMP protocol, including LoD and intra-/inter-assay reproducibility, were assessed. The LoD was determined at a concentration of 100,000 copies/ml. This equals 5000 copies in the eluate from a nasopharyngeal swab, or 95 copies in the RT-LAMP reaction (Supplementary Table S1). During the reproducibility runs, all positive and negative pool samples were correctly qualitatively assigned resulting in 100% intra-run and inter-run reproducibility. The intra-run median Tt value (10.81) of the positive sample pool was comparable to the inter-run median Tt value (10.71). As expected, the inter-run Tt variability was slightly higher (standard deviation [SD] 0.222, coefficient of variation [CV] 2.1%) compared to the intra-run Tt variability (SD 0.042, CV 0.4%) (Table 2).

Discussion

Due to supply bottlenecks of PCR reagents, laboratory equipment, and reagents needed for RNA isolation from nasopharyngeal swabs, there have been major difficulties in SARS-CoV-2 routine diagnostic workflows during the recent SARS-CoV-2 pandemic. Therefore, we have established an RT-LAMP protocol for detection of SARS-CoV-2 which can be used as an alternative to conventional RT-PCR protocols. Using fluorescence as read-out,

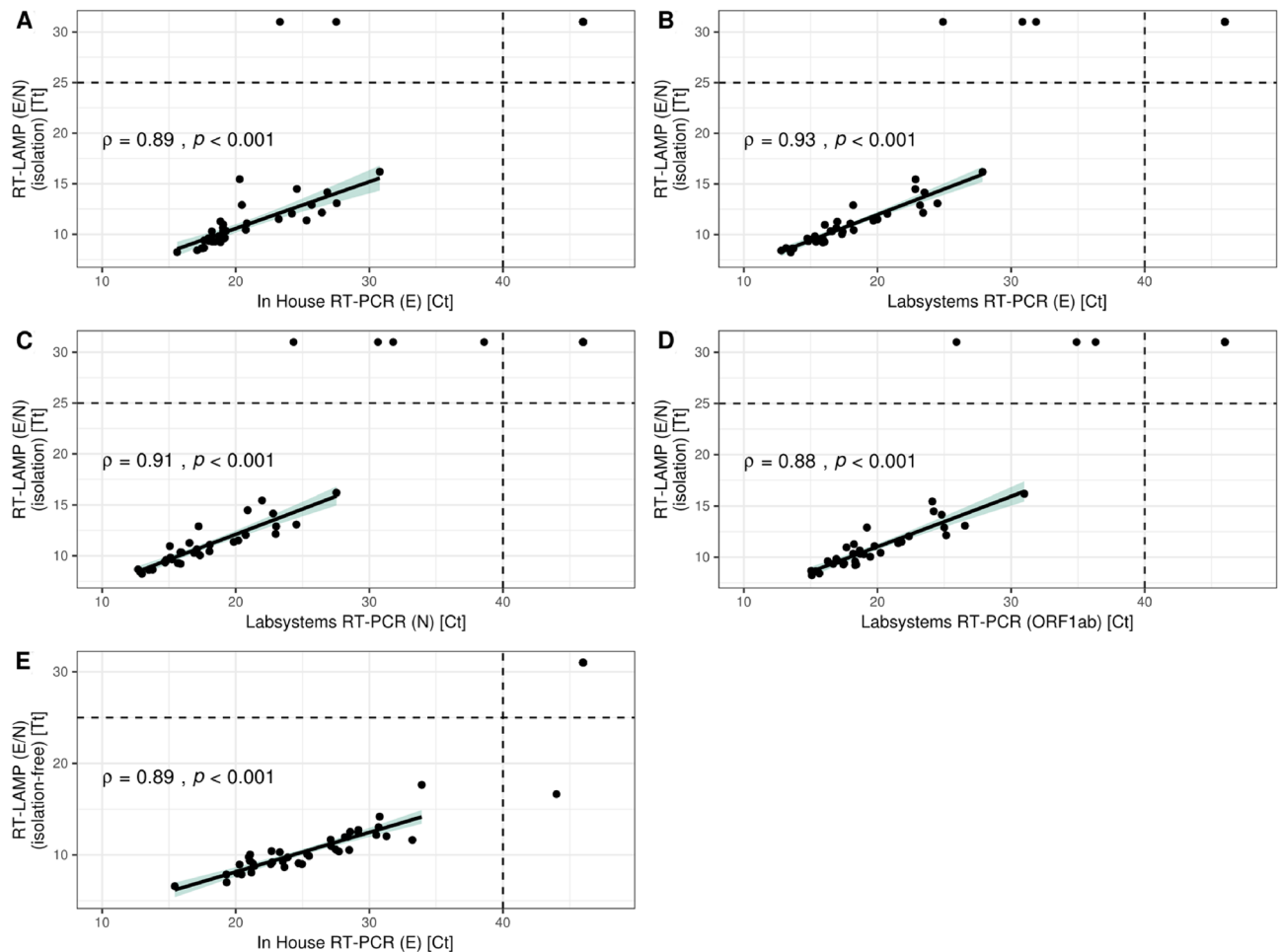


Figure 2. Comparison of the SARS-CoV-2 RT-LAMP Tt values and the RT-PCR Ct values. The dashed lines represent the negative cut-offs. Artificial Ct and Tt values above these cut-offs were assigned to negative samples for data visualization purpose. Spearman correlation results are shown. Only true positive samples were included into the correlation analysis. **(A)** SARS-CoV-2 RT-LAMP with RNA isolation compared to the in-house RT-PCR. **(B)** SARS-CoV-2 RT-LAMP with RNA isolation compared to the Labsystems Dx RT-PCR (E gene target). **(C)** SARS-CoV-2 RT-LAMP with RNA isolation compared to the Labsystems Dx RT-PCR (N gene target). **(D)** SARS-CoV-2 RT-LAMP with RNA isolation compared to the Labsystems Dx RT-PCR (ORF1ab gene target). **(E)** SARS-CoV-2 RT-LAMP without RNA isolation compared to the in-house RT-PCR.

	Intra-run reproducibility	Inter-run reproducibility
Qualitative agreement (%)	100	100
Median Tt value ^a	10.81	10.71
Standard deviation (SD) ^a	0.04	0.22
Coefficient of variation (CV) (%) ^a	0.4	2.1

Table 2. Intra- and inter-run reproducibility of the isolation-free SARS-CoV-2 RT-LAMP protocol. To determine the reproducibility a positive pool sample and a negative pool sample was measured in five replicates on three consecutive days. ^aMedian Tt value, standard deviation and the coefficient of variation are calculated based on the results from the positive pool.

the protocol can be applied on conventional RT-PCR cyclers which are available in most clinical diagnostic laboratories. In addition, this detection method can be easily adapted to automated processing with standard workflows and has the potential to allow a quantitative analysis if needed. To further increase throughput and decrease hands-on time, we successfully adapted the SARS-CoV-2 RT-LAMP protocol on a liquid handling station. This has not only decreased sample processing time, but also reduces the risk of error due to manual processing. To further increase the robustness of the SARS-CoV-2 RT-LAMP assay, we combined the RT-LAMP protocol with a simple RNA isolation-free sample preparation using proteinase K digestion. A similar approach

has been described for the preparation of saliva samples by Vogels et al.²⁰. They demonstrated that saliva is a valid alternative to nasopharyngeal swabs with respect to SARS-CoV-2 testing and their protocol has received Emergency Use Authorization (EUA) from the U.S. Food and Drug Administration (FDA)²⁰.

Since the beginning of the pandemic, the RT-LAMP technology has emerged as an additional tool for laboratory diagnosis of SARS-CoV-2 and COVID-19, in parallel with conventional RT-PCR². Commercial test kits are already offered by different companies including Color Genomics, Lucira Health, and New England Biolabs. There is a high number of published RT-LAMP protocols available describing the detection of SARS-CoV-2 in various clinical specimens including nasopharyngeal swab extractions and saliva. In a point-of-care-test (POCT) setting, detection methods utilising visual inspection of the results are frequently used^{21–26}. For application in a dedicated diagnostic laboratory environment there are different protocols available using fluorescence detection, or absorbance readings as quantitative measurement^{27–33}. These advanced detection methods are also possible in POCT diagnostics and several protocols are available^{34–36}.

By reducing the complexity of the sample preparation, processing time and therefore cost per reaction is reduced. Both aspects are important to consider when adapting assays for SARS-CoV-2 mass screening in clinical diagnostic laboratories. With the isolation-free RT-LAMP protocol, it is possible to process samples in less than 90 min, which is significantly faster than conventional RT-PCR protocols⁸.

During the initial phase of method establishment, an additional reaction mix per sample targeting human ACTB was included as an external control to account for sample quality. Since the results were positive for all investigated clinical samples, we discontinued using the external control during the following method comparison experiments to increase throughput and efficiency. For future experiments, to increase confidence in negative results while retaining high throughput and efficiency, it would be favourable to amplify and detect a human target as internal control in addition to the SARS-CoV-2 targets in a one-tube reaction using target specific probes^{10,11}. This would increase the difficulty of the experimental design and execution but would also expand the overall analytical quality¹⁰.

To compare our RT-LAMP protocols with conventional RT-PCR protocols, we performed various method comparison experiments with clinical samples using an in-house RT-PCR assay as well as a commercial assay as standard. Under all conditions tested, the qualitative results of the RT-LAMP showed near perfect agreement with the RT-PCR assays, with no indication of a systematic difference ($\kappa > 0.8$, McNemar's test, $P > 0.05$). Sensitivity ranged from 89.5% to 100% and specificity from 96.2 to 100%. This is consistent with the general observation that RT-LAMP assays show a reduced sensitivity compared to RT-PCR^{8,16,37}. In our study, RT-PCR Ct values of the samples classified as false-negative by RT-LAMP were in a high range above 30 in most cases. This suggests that it is largely samples with a low viral load that are not detected by RT-LAMP. This could be due to the higher LoD compared to RT-PCR.

The isolation-free SARS-CoV-2 RT-LAMP test showed a LoD of 100,000 copies/ml or 95 copies per reaction which is consistent with data reported elsewhere^{16,38}. Using positive and negative sample pools, our isolation-free RT-LAMP protocol was highly reproducible with no false classification over three runs with five replicates of each pool. The Tt values of the positive sample pool showed little variability in terms of both intra- and inter-run variability.

Using the dynamic pattern of viral load kinetics, Larremore et al. demonstrated that effective SARS-CoV-2 population screening depends primarily on the frequency of testing and speed of reporting⁸. Their study showed that test sensitivity was secondary, supporting RT-LAMP as a useful alternative or addition to RT-PCR despite the higher LoD.

Compared to the use of isolated RNA samples for SARS-CoV-2 detection, the isolation-free RT-LAMP assay showed a slightly reduced specificity of 96.2%. The sample, which was classified false positive by RT-LAMP, demonstrated a high Tt value above 15. Since there is no RNA purification step in the isolation-free protocol, the input sample matrix is more complex in the RT-LAMP reaction under these conditions. This could cause a slightly reduced specificity compared to an assay with RNA isolation. However, laboratory findings should always be interpreted alongside patient clinical history, thus a slightly reduced specificity should be acceptable.

Overall, the RT-LAMP Tt values showed a significant positive correlation with the RT-PCR Ct values. This is a valuable finding since both methods are based on the amplification of viral genetic material in the samples and apply an identical procedure for Ct value/Tt value calculation. As outlined by Engelmann et al., care must be taken when interpreting Ct or Tt values in a clinical setting with regard to viral infectivity since they could be biased by preanalytical issues and there is only little analytical test harmonization available³⁹.

Although RT-LAMP has proven to be a fast and efficient alternative for SARS-CoV-2 screening in clinical diagnostic laboratories, there are some limitations. In addition to reduced analytical and diagnostic sensitivity, RT-LAMP assays carry a high risk for carry-over contamination due to an extremely high efficient reaction, which can lead to false positive results¹⁰. To reduce this risk it is important to take several precautions. Regarding assay design, we added dUTP/UDG to the reaction mixture to reduce the risk of cross contamination from previous runs. Further risk reduction can be achieved by spatial separation of sample preparation, master mix set up, RT-LAMP reaction set up and incubation/detection¹⁰. It is also fundamentally important that the reaction tubes are not opened after the LAMP reaction is complete¹⁰. The tubes should be tightly sealed and discarded immediately. Prior to the use of a RT-LAMP protocol in a clinical diagnostic laboratory, it is important to perform a comprehensive method validation in order to identify pitfalls and take appropriated measures.

SARS-CoV-2 screening is only one possible application in a clinical diagnostic laboratory. As RT-LAMP protocols are quite flexible this technique could be used more widely after the SARS-CoV-2 pandemic. Appropriate protocols have already been established, for Influenza virus, Zika virus, Dengue virus, Hepatitis C, respiratory syncytial virus, and *Streptococcus pneumoniae*^{10,40–43}.

Reagent	Vendor	Volume per reaction (µl)
WarmStart LAMP Master Mix (2X)	New England Biolabs	12.5
Fluorescent dye (50X)	New England Biolabs	0.5
SARS LAMP Primer Mix (×10)/ACTB LAMP Primer Mix (10X)	Integrated DNA Technologies	2.5
H ₂ O	VWR	7.625
dUTP	New England Biolabs	0.175
UDG	New England Biolabs	0.2
WarmStart RTX Reverse Transcriptase	New England Biolabs	0.5

Table 3. Composition of the SARS-CoV-2 RT-LAMP reaction mix. Deoxyuridine triphosphate (dUTP) and Antarctic thermolabile Uracil DNA glycosylase (UDG) were applied to prevent cross-over contamination. Additional WarmStart reverse transcriptase was added to increase reverse transcription efficiency. During the initial method establishment, a second reactions mix was set up per sample using a LAMP primer mix targeting the human beta actin gene (ACTB) instead of SARS-CoV-2 as an external control to account for sample quality.

The measurement method is not bound to a technology, but can be used with different systems. The same applies to the evaluation of the data. In this study we used the software of the manufacturer. However, this can also be done with alternative open source software, as described by us^{6,15,44}.

To summarise, we have established and evaluated a flexible SARS-CoV-2 RT-LAMP protocol which shows acceptable analytical performance and could be used as an alternative to RT-PCR in clinical diagnostic laboratories. Most currently available RT-LAMP protocols are focused on POCT approaches. Our goal was to optimise throughput and turnaround time based on already existing laboratory infrastructure in order to provide an assay which can be easily adapted to different conditions, and is suitable for SARS-CoV-2 mass screening.

Materials and methods

Clinical samples. A total of 323 nasopharyngeal swabs from individuals with suspected COVID-19 sent to the laboratory of the Institute for Laboratory Medicine (Singen, Germany) for SARS-CoV-2 screening were used for method comparison. All samples used were pseudo-anonymized surplus material from routine diagnostics and were retrieved from the laboratory's sample storage only after initial testing was finished. No personal or medical patient data were recorded or analysed. All individuals gave their informed consent. The study has complied with all the relevant national regulations and institutional policies and has been approved by the ethics committee of the Brandenburg University of Technology Cottbus-Senftenberg (EK2020-16).

Sample preparation and SARS-CoV-2 RT-LAMP protocol. During method development, the SARS-CoV-2 RT-LAMP assay was run with and without RNA isolation. RNA isolation from nasopharyngeal swabs was performed on chemagic 360 instruments (PerkinElmer, Waltham, USA) using PrepitoViral DNA/RNA300 isolation kits (PerkinElmer, Waltham, USA).

The isolation free workflow was done using proteinase K digestion²⁰. All sample handling steps were performed under a biosafety cabinet class II. The nasopharyngeal swabs were flushed with 500 µl NaCl -solution (0.9%) by vortexing. 2.5 µl proteinase K solution (50 mg/ml) (ThermoFisher Scientific, Waltham, USA) was prepared in PCR reaction tubes and 50 µl of the sample eluate was added. After vortexing for 1 min, samples were incubated in a thermocycler (ThermoFisher Scientific, Waltham, USA) at 57 °C for 5 min followed by 95 °C for 5 min. After proteinase K digestion the samples were directly analysed by RT-LAMP.

A LAMP primer mix in nuclease free water targeting the N and E gene of SARS-CoV-2 (Supplementary Table S2) was used in combination with a WarmStart RT-LAMP kit including a fluorescent intercalating marker (New England Biolabs, Ipswich, USA)⁴⁵. Individual primers were purchased from IDT (Integrated DNA Technologies, Coralville, USA). Deoxyuridine triphosphate (dUTP) and uracil-DNA glycosylase (UDG) (New England Biolabs, Ipswich, USA) were added to the reaction mix to prevent cross contamination from previous runs. WarmStart RTx Reverse Transcriptase (New England Biolabs, Ipswich, USA) was added to increase RT efficiency. To account for sample quality during the initial method establishment, a separate second reaction mix using a LAMP primer set targeting the human beta actin gene (ACTB) (Supplementary Table S3) was set up per sample as external control⁴⁵.

A master mix was prepared and added to PCR reaction strips (Table 3). After addition of 1 µl RNA template, the strips were tightly sealed, gently mixed, and incubated for 5 min at 25 °C, and then for 30 min at 65 °C in a Rotor-Gene Q device (Qiagen, Hilden, Germany) or a LightCycler 96 (Roche, Basel, Switzerland). Fluorescence readings were taken every minute during this incubation period on the FAM channel as Bst polymerase is mainly active at 65 °C. A no-template control and a positive control (INSTAND, Düsseldorf, Germany) were added for all runs.

RT-PCR protocols. An in-house SARS-CoV-2 RT-PCR protocol and a commercial RT-PCR kit were used for method comparison. Since both assays work only with purified RNA, the above-mentioned RNA isolation procedure was mandatory prior to PCR testing.

The in-house PCR protocol was based on the QuantiTect Probe RT-PCR Kit (Qiagen, Hilden, Germany) with primers and a hydrolysis probe (Biomers, Ulm, Germany) targeting the E gene. Detection was done on the FAM channel of a LightCycler 96 instrument (Roche, Basel, Switzerland). Primer sequences and the temperature profile are shown in the appendix (Supplementary Information 1).

For comparison of the SARS-CoV-2 RT-LAMP protocol with a commercial kit, the LABSYSTEMS COVID-19 Real Time Multiplex RT-PCR Kit (Labsystems Diagnostics OY, Helsinki, Finland) was used. This kit is designed to detect three genes (ORF1ab, E, N) of the SARS-CoV-2 genome simultaneously. The RT-PCR reactions were set up according to the manufacturer's protocol and were analysed on a Rotor-Gene Q instrument.

Calculation of cycle threshold (Ct) and threshold time (Tt) values. The cycle threshold (Ct) values of the RT-PCR assays were calculated from the background corrected amplification curves using the Roto-Gene Q Software [V2.3.5] (Qiagen, Hilden, Germany) or the LightCycler 96 Software [V.1.1.0.1320] (Roche, Basel, Switzerland). Amplification curves from the RT-LAMP were analysed using an identical procedure. As LAMP is an isothermal amplification technique, the reaction time when the fluorescence signal exceeds the threshold is referred to as threshold time (Tt) value.

Semi-automation of the assays. Semi-automated pipetting of the SARS-CoV-2 RT-LAMP protocol and the in-house RT-PCR were run on a Brand Liquid Handling station (Brand, Wertheim, Germany). The master mixes were prepared manually, and the liquid handler was applied to dispense the master mix into the PCR strips. The isolated RNA samples were collected in 96 well plates and the system was set to transfer an appropriate amount of sample into the prepared PCR strips.

RT-LAMP performance evaluation. The performance of the SARS-CoV-2 RT-LAMP protocol was assessed by method comparison. To compare the assay with different RT-PCR protocols, 70 RNA isolates from nasopharyngeal swabs which were initially detected as positive (n = 36) and negative (n = 34) for SARS-CoV-2 by in-house RT-PCR were analysed by all three assays and the qualitative results, as well as Ct values of the RT-PCR assays, and the Tt values of the RT-LAMP were compared.

In order to test the semi-automated procedure on the liquid handling station, another 188 RNA isolates were analysed using the in-house RT-PCR protocol and the RT-LAMP assay.

The isolation-free SARS-CoV-2 RT-LAMP test was compared to the in-house RT-PCR using 65 nasopharyngeal swabs. To obtain comparable results, the eluates from these samples were directly aliquoted after elution with NaCl solution (0.9%). One aliquot was analysed with the isolation-free RT-LAMP and a second underwent standard RNA isolation followed by RT-PCR testing using the in-house protocol.

Performance characteristics including limit of detection (LoD), and intra- and inter-run reproducibility of the isolation-free SARS-CoV-2 RT-LAMP protocol were assessed. LoD was determined by serial dilution of a SARS-CoV-2 reference material containing 10^7 genome copies (INSTAND, Düsseldorf, Germany). Each dilution was analysed in triplicate. To investigate reproducibility of the assay, a negative and a positive sample pool were made from stored SARS-CoV-2 eluates and both were measured in five replicates on three consecutive days.

Statistical analysis. Positive/negative classification of the Tt values and Ct values from different methods was done by using standardised negative cut-offs. For the RT-PCR assays, a cut-off of Ct 40 was applied and for the RT-LAMP test a cut-off of Tt 25. For data visualization, Ct values and Tt values above these cut-offs were artificially assigned to negative samples (Ct 46 for the RT-PCRs and Tt 31 for the RT-LAMP). Qualitative results were analysed by McNemar's test and Cohen's kappa (interrater agreement statistics) for method comparison⁴⁶. Ct values of the different PCR methods were compared with RT-LAMP Tt values by Spearman correlation after testing for normal distribution by applying the Shapiro–Wilk normality test. Statistical testing and data visualization was performed using R [v3.6.3] (R Foundation for Statistical Computing, Vienna, Austria).

Data availability

The raw datasets generated during the current study are available from the corresponding author on reasonable request in aggregated form. All data analysed during this study are included in this published article and its Supplementary Information files.

Received: 23 June 2021; Accepted: 12 October 2021

Published online: 01 November 2021

References

- Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
- Younes, N. *et al.* Challenges in laboratory diagnosis of the novel coronavirus SARS-CoV-2. *Viruses* **12**, 582 (2020).
- Mathuria, J. P. & Yadav, R. Laboratory diagnosis of SARS-CoV-2—A review of current methods. *J. Infect. Public Health* **13**, 901–905 (2020).
- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170 (2021).
- Harrison, A. G., Lin, T. & Wang, P. Mechanisms of SARS-CoV-2 transmission and pathogenesis. *Trends Immunol.* **41**, 1100–1115 (2020).
- Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K. & Weinhäusel, A. A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomol. Detect. Quantif.* **1**, 23–33 (2014).
- Don Wai Luu, L., Payne, M., Zhang, X., Luo, L. & Lan, R. Development and comparison of novel multiple cross displacement amplification (MCDA) assays with other nucleic acid amplification methods for SARS-CoV-2 detection. *Sci. Rep.* **11**, 1–7 (2021).

8. Larremore, D. B. *et al.* Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci. Adv.* **7**, eabd5393 (2021).
9. Aoki, M. N. *et al.* Colorimetric RT-LAMP SARS-CoV-2 diagnostic sensitivity relies on color interpretation and viral load. *Sci. Rep.* **11**, 9026 (2021).
10. Wong, Y.-P., Othman, S., Lau, Y.-L., Radu, S. & Chee, H.-Y. Loop-mediated isothermal amplification (LAMP): A versatile technique for detection of micro-organisms. *J. Appl. Microbiol.* **124**, 626–643 (2018).
11. Becherer, L. *et al.* Loop-mediated isothermal amplification (LAMP)—Review and classification of methods for sequence-specific detection. *Anal. Methods* **12**, 717–746 (2020).
12. Peto, L. *et al.* Diagnosis of SARS-CoV-2 infection with LamPORE, a high-throughput platform combining loop-mediated isothermal amplification and nanopore sequencing. *J. Clin. Microbiol.* **59**, e03271-20 (2021).
13. Craw, P. & Balachandran, W. Isothermal nucleic acid amplification technologies for point-of-care diagnostics: A critical review. *Lab Chip* **12**, 2469–2486 (2012).
14. Ruiz-Villalba, A., Ruijter, J. M. & van den Hoff, M. J. B. Use and misuse of C_q in qPCR data analysis and reporting. *Life* **11**, 496 (2021).
15. Rödiger, S., Burdukiewicz, M., Blagodatskikh, K., Jahn, M. & Schierack, P. R as an environment for reproducible analysis of DNA amplification experiments. *R J.* **7**, 127–150 (2015).
16. Dao Thi, V. L. *et al.* A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Sci. Transl. Med.* **12**, eabc7075 (2020).
17. Baek, Y. H. *et al.* Development of a reverse transcription-loop-mediated isothermal amplification as a rapid early-detection method for novel SARS-CoV-2. *Emerg. Microbes Infect.* **9**, 998–1007 (2020).
18. Huang, W. E. *et al.* RT-LAMP for rapid diagnosis of coronavirus SARS-CoV-2. *Microb. Biotechnol.* **13**, 950–961 (2020).
19. Lalli, M. A. *et al.* Rapid and extraction-free detection of SARS-CoV-2 from saliva with colorimetric LAMP. *medRxiv*. <https://doi.org/10.1101/2020.05.07.20093542> (2020).
20. Vogels, C. B. F. *et al.* SalivaDirect: A simplified and flexible platform to enhance SARS-CoV-2 testing capacity. *Sci. Adv.* **2**, 263–280. e6 (2021).
21. Hu, X. *et al.* Development and clinical application of a rapid and sensitive loop-mediated isothermal amplification test for SARS-CoV-2 infection. *mSphere* **5**, e0080820 (2020).
22. Nawattanapaiboon, K. *et al.* Colorimetric reverse transcription loop-mediated isothermal amplification (RT-LAMP) as a visual diagnostic platform for the detection of the emerging coronavirus SARS-CoV-2. *Analyst* **146**, 471–477 (2021).
23. Lim, B. *et al.* Clinical validation of optimised RT-LAMP for the diagnosis of SARS-CoV-2 infection. *Sci. Rep.* **11**, 16193 (2021).
24. Amaral, C. *et al.* A molecular test based on RT-LAMP for rapid, sensitive and inexpensive colorimetric detection of SARS-CoV-2 in clinical samples. *Sci. Rep.* **11**, 16430 (2021).
25. Janíková, M., Hodosy, J., Boor, P., Klempa, B. & Celec, P. Loop-mediated isothermal amplification for the detection of SARS-CoV-2 in saliva. *Microb. Biotechnol.* **14**, 307–316 (2021).
26. Toppings, N. B. *et al.* A rapid near-patient detection system for SARS-CoV-2 using saliva. *Sci. Rep.* **11**, 13378 (2021).
27. Alekseenko, A. *et al.* Direct detection of SARS-CoV-2 using non-commercial RT-LAMP reagents on heat-inactivated samples. *Sci. Rep.* **11**, 1820 (2021).
28. Lee, J. Y. H. *et al.* Validation of a single-step, single-tube reverse transcription loop-mediated isothermal amplification assay for rapid detection of SARS-CoV-2 RNA. *J. Med. Microbiol.* **69**, 1169–1178 (2020).
29. Sherrill-Mix, S. *et al.* Detection of SARS-CoV-2 RNA using RT-LAMP and molecular beacons. *Genome Biol.* **22**, 169 (2021).
30. Jiang, M. *et al.* Development and validation of a rapid, single-step reverse transcriptase loop-mediated isothermal amplification (RT-LAMP) system potentially to be used for reliable and high-throughput screening of COVID-19. *Front. Cell. Infect. Microbiol.* **10**, 331 (2020).
31. Jang, W. S. *et al.* Development of a multiplex Loop-Mediated Isothermal Amplification (LAMP) assay for on-site diagnosis of SARS-CoV-2. *PLoS One* **16**, e0248042 (2021).
32. Oscorbin, I. P. *et al.* Detection of SARS-CoV-2 RNA by a multiplex reverse-transcription loop-mediated isothermal amplification coupled with melting curves analysis. *IJMS* **22**, 5743 (2021).
33. Dong, Y. *et al.* Comparative evaluation of 19 reverse transcription loop-mediated isothermal amplification assays for detection of SARS-CoV-2. *Sci. Rep.* **11**, 2936 (2021).
34. Bektaş, A. *et al.* Accessible LAMP-enabled rapid test (ALERT) for detecting SARS-CoV-2. *Viruses* **13**, 742 (2021).
35. Garneret, P. *et al.* Performing point-of-care molecular testing for SARS-CoV-2 with RNA extraction and isothermal amplification. *PLoS One* **16**, e0243712 (2021).
36. Mautner, L. *et al.* Rapid point-of-care detection of SARS-CoV-2 using reverse transcription loop-mediated isothermal amplification (RT-LAMP). *Virology* **17**, 160 (2020).
37. Rödel, J. *et al.* Use of the variplex™ SARS-CoV-2 RT-LAMP as a rapid molecular assay to complement RT-PCR for COVID-19 diagnosis. *J. Clin. Virol.* **132**, 104616 (2020).
38. Yang, Q. *et al.* Saliva TwoStep for rapid detection of asymptomatic SARS-CoV-2 carriers. *medRxiv*. <https://doi.org/10.1101/2020.07.16.20150250> (2021).
39. Engelmann, I. *et al.* Preanalytical issues and cycle threshold values in SARS-CoV-2 Real-Time RT-PCR testing: should test results include these?. *ACS Omega* **6**, 6528–6536 (2021).
40. Ahn, S. J. *et al.* Rapid and simple colorimetric detection of multiple influenza viruses infecting humans using a reverse transcriptional loop-mediated isothermal amplification (RT-LAMP) diagnostic platform. *BMC Infect. Dis.* **19**, 676 (2019).
41. Calvert, A. E., Biggerstaff, B. J., Tanner, N. A., Lauterbach, M. & Lanciotti, R. S. Rapid colorimetric detection of Zika virus from serum and urine specimens by reverse transcription loop-mediated isothermal amplification (RT-LAMP). *PLoS One* **12**, e0185340 (2017).
42. Lamb, L. E. *et al.* Rapid detection of Zika virus in urine samples and infected mosquitos by reverse transcription-loop-mediated isothermal amplification. *Sci. Rep.* **8**, 3803 (2018).
43. Takano, C. *et al.* Molecular serotype-specific identification of *Streptococcus pneumoniae* using loop-mediated isothermal amplification. *Sci. Rep.* **9**, 19823 (2019).
44. Rödiger, S., Burdukiewicz, M. & Schierack, P. chipPCR: An R package to pre-process raw data of amplification curves. *Bioinformatics* **31**, 2900–2902 (2015).
45. Zhang, Y. *et al.* Enhancing colorimetric loop-mediated isothermal amplification speed and sensitivity with guanidine chloride. *Biotechniques* **69**, 178–185 (2020).
46. Watson, P. F. & Petrie, A. Method agreement analysis: A review of correct methodology. *Theriology* **73**, 1167–1179 (2010).

Acknowledgements

The authors would like to thank Karina Furie for constructive criticism of the manuscript.

Author contributions

J.S., S.B., F.B. and F.W. designed the work. J.S. carried out the experiments and analysed the data with support from S.R. and D.R. J.S., H.H., J.B., P.S., S.R. and D.R. discussed and interpreted the results. J.S., P.S., S.R. and D.R. wrote the final manuscript. The final version was approved by all authors. F.B. and D.R. supervised the project.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

DR holds an executive position and is a shareholder in Medipan and GA Generic Assays, which are diagnostic manufacturers. FB is the managing director of the Institute for Laboratory Medicine. JS, SB and HH are employees of this institution. The remaining authors declare no potential conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-00827-0>.

Correspondence and requests for materials should be addressed to D.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.


© The Author(s) 2021

RESEARCH ARTICLE

Genotyping of familial Mediterranean fever gene (*MEFV*)—Single nucleotide polymorphism—Comparison of Nanopore with conventional Sanger sequencing

Jonas Schmidt^{1,2,3}, Sandro Berghaus¹, Frithjof Blessing^{1,2}, Holger Herbeck¹, Josef Blessing¹, Peter Schierack^{3,4}, Stefan Rödiger^{3,4}, Dirk Roggenbuck^{3,4} *, Folker Wenzel² 

1 Institute for Laboratory Medicine, Singen, Germany, **2** Faculty of Medical and Life Sciences, Furtwangen University, Villingen-Schwenningen, Germany, **3** Faculty Environment and Natural Sciences, Institute of Biotechnology, Brandenburg University of Technology Cottbus-Senftenberg, Senftenberg, Germany, **4** Faculty of Health Sciences Brandenburg, Brandenburg University of Technology Cottbus–Senftenberg, Senftenberg, Germany

 These authors contributed equally to this work.

* dirk.roggenbuck@b-tu.de



OPEN ACCESS

Citation: Schmidt J, Berghaus S, Blessing F, Herbeck H, Blessing J, Schierack P, et al. (2022) Genotyping of familial Mediterranean fever gene (*MEFV*)—Single nucleotide polymorphism—Comparison of Nanopore with conventional Sanger sequencing. PLoS ONE 17(3): e0265622. <https://doi.org/10.1371/journal.pone.0265622>

Editor: J Francis Borgio, Imam Abdulrahman Bin Faisal University, SAUDI ARABIA

Received: December 3, 2021

Accepted: March 4, 2022

Published: March 17, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0265622>

Copyright: © 2022 Schmidt et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The Nanopore sequencing data underlying the results presented in the study are available from the European

Abstract

Background

Through continuous innovation and improvement, Nanopore sequencing has become a powerful technology. Because of its fast processing time, low cost, and ability to generate long reads, this sequencing technique would be particularly suitable for clinical diagnostics. However, its raw data accuracy is inferior in contrast to other sequencing technologies. This constraint still results in limited use of Nanopore sequencing in the field of clinical diagnostics and requires further validation and IVD certification.

Methods

We evaluated the performance of latest Nanopore sequencing in combination with a dedicated data-analysis pipeline for single nucleotide polymorphism (SNP) genotyping of the familial Mediterranean fever gene (*MEFV*) by amplicon sequencing of 47 clinical samples. Mutations in *MEFV* are associated with Mediterranean fever, a hereditary periodic fever syndrome. Conventional Sanger sequencing, which is commonly applied in clinical genetic diagnostics, was used as a reference method.

Results

Nanopore sequencing enabled the sequencing of 10 target regions within *MEFV* with high read depth (median read depth 7565x) in all samples and identified a total of 435 SNPs in the whole sample collective, of which 29 were unique. Comparison of both sequencing workflows showed a near perfect agreement with no false negative calls. Precision, Recall, and F1-Score of the Nanopore sequencing workflow were > 0.99, respectively.

Nucleotide Archive (Accession: PRJEB49157). The data analysis pipeline as well as the Sanger sequencing results are available from Github (github.com/j4yo/MEFV-SNP-Genotyping-Pipeline).

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Conclusions

These results demonstrated the great potential of current Nanopore sequencing for application in clinical diagnostics, at least for SNP genotyping by amplicon sequencing. Other more complex applications, especially structural variant identification, require further in-depth clinical validation.

1. Introduction

Since its first description in 1996, nanopore-based deoxyribonucleic acid (DNA) sequencing has developed to one of the most powerful sequencing technologies thanks to continuous innovation and improvements [1,2]. Nowadays, different sequencing devices and protocols are commercially available rendering this technique attractive for various areas of molecular biological research and diagnostics, including metagenomics, bacterial and viral infectiology, human genomics, and cancer research [3–11]. The core components of current Nanopore sequencing devices are protein nanopores contained in a membrane [12,13]. As single DNA molecules are passed through these pores, the resulting changes in an ionic current across the membrane are used to infer the sequence of nucleic acids [11–13]. This sequencing approach offers the advantages of real-time sequencing, ultra-long read length (average read length up to 10 kb), high throughput and the possibility of base modification detection as well as native ribonucleic acid (RNA) sequencing [1,13,14]. However, a major drawback compared to other next-generation sequencing (NGS) techniques has been the comparatively high error rate [13]. Although this is a heterogeneous measure, which is influenced by different parameters including sequencing instrument, sequencing protocol and sample type, Nanopore sequencing shows a distinct higher error rate (~6%) compared to PacBio sequencing (~1.5%), Illumina sequencing (~0.5%) and conventional Sanger sequencing (~0.001%) [15–19]. This is especially critical for medical applications such as single nucleotide polymorphism (SNP) genotyping, which require high sequencing accuracy to achieve reliable results [13]. Although the accuracy of Nanopore sequencing has improved considerably by optimization of the underlying sequencing chemistry and bioinformatic analysis tools, it is important to validate the technique against established gold standard methods such as Sanger sequencing to assess a possible application in medical diagnostics [13,20].

A common monogenetic autoinflammatory disease is Familial Mediterranean fever (FMF) which shows a high prevalence among Turkish, Armenian, Jewish and Arabic communities from the eastern Mediterranean region [21,22]. The disease is a clinical diagnosis and mainly characterized by recurrent fever and serositis, with amyloidosis being a severe complication in untreated individuals [22–24]. FMF is considered to be inherited autosomal recessive and is associated with point mutations (single substitutions) in the Mediterranean Fever (*MEFV*) gene [22,24]. This gene consists of 10 exons and is located on the short arm of chromosome 16 in minus strand orientation [22]. It encodes a 781 amino acids containing protein called pyrin, which plays a key role in apoptosis and inflammatory pathways. It is mainly expressed in neutrophils, eosinophils, dendritic cells and fibroblasts [21–23]. Mutated pyrin is thought to cause an excessive inflammatory response through uncontrolled interleukin-1 (IL-1) secretion [21,25]. After clinical diagnosis, the disease is generally treated with colchicine, and IL-1 blockade is suggested in refractory cases [21]. Genetic testing is employed to aid in the clinical diagnosis of FMF and to screen relatives at risk [23]. This can be done either by testing for the most common mutations (targeted mutation analysis) or by sequencing of selected exons [23]. According to expert consensus guidelines for the genetic diagnosis of hereditary recurrent fevers a minimum diagnostic screen should include clearly pathogenic variants which are

frequently identified in patients [26]. For FMF this incorporates the exons 2, 3, 5 and 10 of *MEFV* or a set of nine variants [26]. While DNA sequencing is used in most laboratories for variant analysis, targeted approaches can also be applied by using PCR based or reverse-hybridization based assays [26]. However, these targeted approaches as well as conventional Sanger sequencing suffer from the technological limitation that only a comparably small genetic target range can be covered within a single run. To overcome this limitation, NGS can be applied to sequence gene panels including not only *MEFV* for the diagnosis of FMF but also genes which are associated with other periodic fever syndromes like mevalonate kinase deficiency (MKD, gene *MVK*), tumor necrosis factor receptor-associated periodic syndrome (TRAPS, gene *TNFRSF1A*) and cryopyrin-associated periodic syndrome (CAPS, gene *NLRP3*) [26,27].

In this study, to evaluate the clinical performance of current Nanopore sequencing, we applied this sequencing technique in combination with a dedicated data analysis pipeline for SNP genotyping of selected regions of *MEFV* in 47 patients and validated the results against diagnostic Sanger sequencing as the gold standard method.

2. Material and methods

2.1 Clinical samples

Samples from 25 female and 22 male patients that were drawn for routine *MEFV* assessment were included into this study after routine testing by Sanger sequencing was performed. Median age was 12.1 years (interquartile range [IQR] 12.9). Primary blood samples were collected in EDTA collection tubes by venipuncture and stored at 4°C until further processing. The routine diagnostic workflow includes DNA isolation, polymerase chain reaction (PCR) amplification of selected targets within *MEFV* and Sanger sequencing as described below. Subsequent to routine Sanger sequencing, the amplicons obtained from the amplification step were pooled per sample and Nanopore sequencing was performed.

All included individuals gave their written informed consent. For minor patients, written informed consent was obtained from the parents. The study followed all relevant national regulations and institutional policies, has been approved by the ethics committee of the Landeszärztekammer Baden-Württemberg (F-2018-089) and complies with the World Medical Association Declaration of Helsinki regarding ethical conduct of research involving human subjects and/or animals.

2.2 DNA isolation and PCR amplification

DNA isolation from EDTA whole blood samples was performed on chemagic Prepito-D instruments (PerkinElmer, Waltham, USA) using Prepito NA Body Fluid kits (PerkinElmer) (expected yield: ~2.5 µg).

PCR amplification of the *MEFV* target regions was performed stepwise in eight different PCR reactions using target specific primers (Biomers, Ulm, Germany), Q-Solution (Qiagen, Hilden, Germany), and the AmpliTaq Gold 360 Master Mix (ThermoFisher Scientific, Waltham, USA). The amplicons were designed to span *MEFV* exon 1, exon 2, exon 3, exon 4, exon 5, exon 6, exon 7/8, and exon 9/10 (S1 Table). PCR reactions were performed on an Applied Biosystems Veriti thermal cycler (ThermoFisher Scientific) (S2 and S3 Tables). Nuclease free water was included in all runs as a no template control.

2.3 Sanger sequencing

Prior to sequencing, a clean-up of the amplicons was performed by using ExoSAP-IT clean-up kits (ThermoFisher Scientific). Briefly, 7 µL PCR product were mixed with 1 µL clean-up

reagent by pipetting. This reaction mix was incubated for 15 min at 37°C followed by 15 min at 80°C.

Sanger sequencing of the purified amplicons was performed using the BigDye Terminator Version 3.1 kit (ThermoFisher Scientific) on an Applied Biosystems 3500 Dx Series Genetic Analyzer (ThermoFisher Scientific) according to the manufacturer's protocol. Briefly, sequencing reactions were set up using target specific sequencing primers (Biomers) (S4 Table). After incubation on a thermal cycler, the reaction mix was cleaned by precipitation with ethanol/EDTA/sodium acetate and loaded on the instrument for capillary electrophoresis after resuspending in injection buffer. Sequencing was performed using POP-6 Polymer (ThermoFisher Scientific).

2.4 Nanopore sequencing

Prior to Nanopore sequencing, equal volumes (10 µL) of the amplicons from the target amplification step were pooled for each individual sample. DNA concentration of the pooled samples was measured on a Qubit 4 fluorometer (ThermoFisher Scientific) using the 1x dsDNA HS assay (ThermoFisher Scientific) (S5 Table). Afterwards, a 1.8x AMPure XP bead clean-up was performed according to the manufacturer's protocol (Beckman Coulter, Brea, USA). Sequencing libraries were prepared according to the manufacturer's protocol using native barcoding kits (EXP-NBD104, EXP-NBD114) in combination with ligation sequencing kits (SQK-LSK109) (Oxford Nanopore Technologies (ONT), Oxford, UK). The libraries were prepared with a total of 12 samples per library for each run to ensure a sufficient read count per sample and that the relative proportion of a single sample is comparable (S5 Table). DNA input per sample was 200fmol and 12.5fmol of each barcoded sample were pooled prior to sequencing. Sequencing was performed on a MinION sequencing device (ONT) for 6h using R9.4.1 flow cells (ONT). All samples were sequenced in four different runs using two flow cells. Prior to reuse, the flow cells were purged according to the manufacturer's protocol using flow cell wash kits (EXP-WSH003) (ONT).

2.5 Sequencing data analysis

Sanger sequencing data was analyzed using SEQUENCE Pilot Software [v 3.4.2] (JSI medical systems GmbH, Ettenheim, Germany). Variants were called against the *MEFV* reference (ENSEMBL gene: ENSG00000103313; transcript: ENST00000219596). Identified variants were manually inspected and exported to a comma separated-values (csv) file for comparison with the Nanopore sequencing results.

To analyze the Nanopore sequencing data, a dedicated data analysis pipeline was established by us and implemented into a bash shell script for automation purpose (Fig 1). Raw data in FAST5 file format was basecalled and demultiplexed using the Guppy Basecalling Software [v 5.0.11+2b6dbffa5] (ONT). Basecalling was performed using the "super-accurate" basecalling model (dna_r9.4.1_450bps_sup.cfg). Basic run quality control was performed by applying pycoQC [v 2.5.2] (github.com/leopardi/pycoQC). To remove chimeric and low-quality reads, read filtering was done with NanoFilt [v 2.7.1] (github.com/wdecoster/nanofilt). The filter was set to keep only reads with a read length between 250 and 1200 bases and a quality score equal or larger 15. After filtering, the reads were aligned to chromosome 16 of the hg19 reference genome (NC_000016.9) using minimap2 [v 2.20-r1061] (github.com/lh3/minimap2). The resulting Sequence Alignment Map (SAM) files were sorted and indexed with Samtools [v 1.7] (github.com/samtools/samtools). Afterwards, bcftools [v 1.13] (github.com/samtools/bcftools) was used for variant calling. The tool was set to include only SNPs and skip insertions and deletions. Variant filtering was performed by applying bedtools [v 2.30.0] (github.com/arq5x/bedtools2). Only

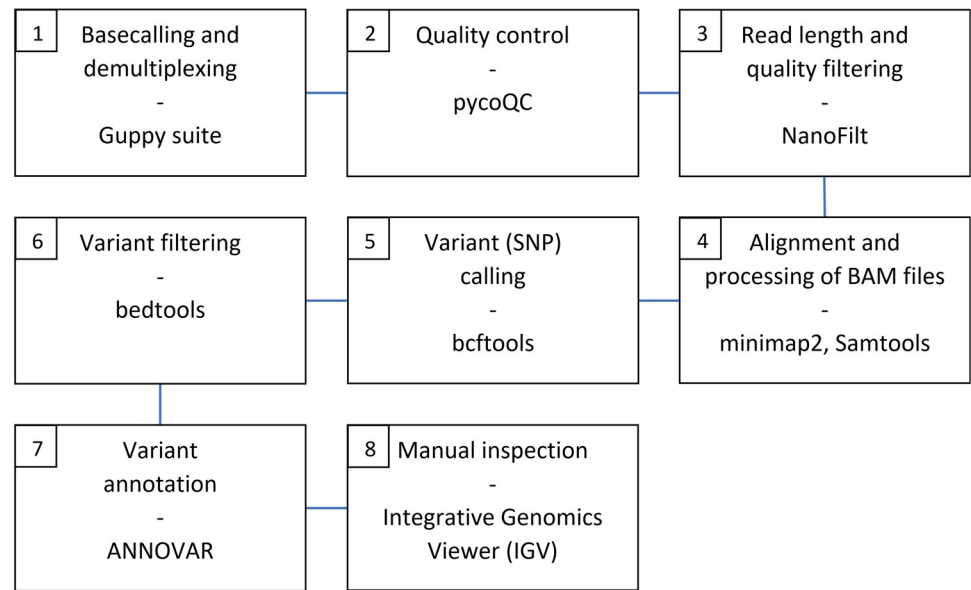


Fig 1. Data analysis pipeline applied for the assessment of the Nanopore sequencing data. Tools used for the different tasks are shown. Step 1 to 7 were implemented in a bash shell script for automation purpose. SNP; single nucleotide polymorphism.

<https://doi.org/10.1371/journal.pone.0265622.g001>

calls in *MEFV* regions covered by the amplicons were included into the final data set. Finally, the identified variants were annotated using ANNOVAR [v 2018-04-16] [28].

Once the automated data analysis pipeline was complete, the results for each individual sample were manually reviewed using the Integrative Genomics Viewer [v2.10.3] (github.com/igvteam/igv).

2.6 Results comparison

Method comparison was done in R [v 3.6.3] (R Foundation for Statistical Computing, Vienna, Austria) [29]. After importing the data sets, Nanopore sequencing variant calls were compared to the Sanger sequencing reference for genomic position, nucleotide change, zygosity, amino acid position, and amino acid change. Nanopore sequencing calls were only classified as true positive (TP) if all five criteria matched to the corresponding Sanger sequencing reference. Variants without a complete match as well as variants which were missed by Nanopore sequencing were classified as false negative (FN) and variants, which were solely identified by Nanopore sequencing as false positive (FP). Based on these classifications, comparative measures including Precision ($TP/(TP + FP)$), Recall ($TP/(FN + TP)$) and F1-Score ($2 * (Precision * Recall)/(Precision + Recall)$) were calculated [30].

Data visualization was performed in R as well using the packages ggVennDiagram, ggplot2, gggenes, and ggpubr. Sequencing depth information was extracted from the SAM files prior to visualization using Samtools.

3. Results

To evaluate the performance of Nanopore sequencing for SNP genotyping, we performed amplicon sequencing of selected *MEFV* regions in 47 clinical samples using a MinION sequencing device and compared the results to conventional Sanger sequencing.

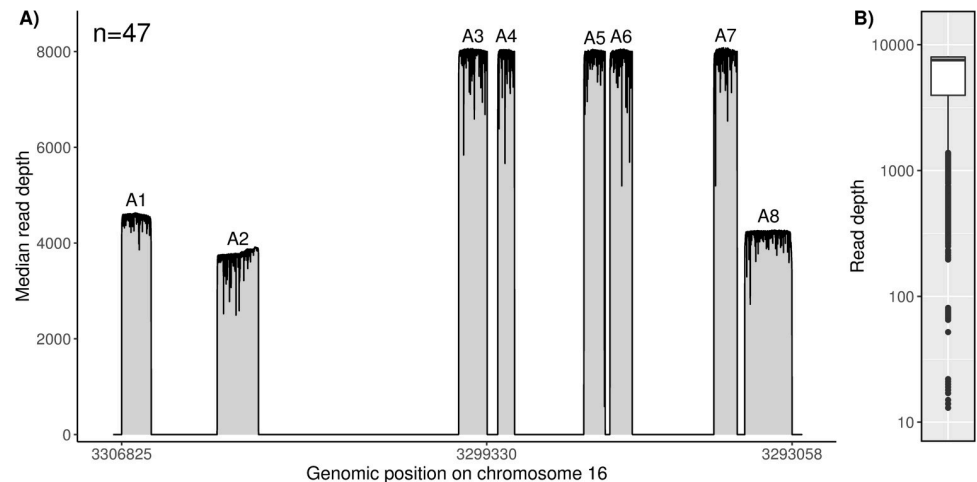


Fig 2. Visualization of the read depth distribution achieved by Nanopore sequencing. (A) Median read depth achieved by amplicon sequencing of selected regions in the *MEFV* gene in 47 clinical samples using a MinION sequencing device. The target regions cover the relevant regions of all 10 exons of this gene. (B) Read depth distribution in the target regions over all 47 samples. A median read depth of 7565x (IQR 4025) was achieved. Outliers with a reduced sequencing depth were observed at the edges of individual amplicons.

<https://doi.org/10.1371/journal.pone.0265622.g002>

By using Nanopore sequencing in combination with a dedicated data analysis pipeline, it was possible to sequence the eight amplicons covering the relevant *MEFV* regions of all 10 exons with a median read depth of 7565x (IQR 4025) over all 47 samples (Fig 2B). A reduced read depth was observed at the edges of individual amplicons (minimum 13x). Furthermore, differences in the median read depth between different amplicons were observed (Fig 2A). Overall, amplicon 1, 2, and 8 showed a lower median read depth compared to the remaining amplicons.

In total, 433 SNPs were identified in the investigated sample collective by Sanger sequencing (284 heterozygous and 149 homozygous). They include 28 unique variants of which 13 are non-synonymous (Table 1). The most common non-synonymous variants include p.E148Q (40.4%), p.R202Q (34.0%), p.M694V (25.5%), p.P369S (12.8%) and p.R408Q (12.8%). In addition, the most common synonymous variants were p.R314R (76.6%), p.E474E (70.2%), p.Q476Q (70.2%), p.D510D (70.2%), and p.P588P (68.1%).

All 433 SNPs confirmed by Sanger sequencing in the sample collective were also identified by Nanopore sequencing with matching genomic position, nucleotide change, zygosity, amino acid position, and amino acid change (Fig 3). Additionally, the Nanopore sequencing results showed a transversion from guanine (G) to thymine (T) in the 3' untranslated region (UTR) at genomic position 3293090 in two patients which has not been identified by initial Sanger sequencing (Figs 3 and S1). Read depth at this genomic position was >7000x in both cases. A data base research, including ClinVar and dbSNP, did not reveal any further information on this SNP. Remarkably, both individuals in whom this SNP was identified were related. By sequencing an additional amplicon, spanning this region, it was possible to confirm the transversion in both samples also by Sanger sequencing (S2 Fig).

For further method comparison, performance parameters such as Precision, Recall, and F1-Score were calculated from the results. The SNP which was only identified by Nanopore sequencing was treated as false positive, since it was not identified during the initial diagnostic Sanger sequencing runs. Based on this assumption, the Nanopore sequencing method in comparison to Sanger sequencing showed a Precision of 0.995, a Recall of 1 and a F1-Score of 0.998.

Table 1. Unique MEFV variants identified in 47 patients. Variant frequency in the sample collective under investigation is shown. One variant in two patients was only identified by Nanopore sequencing and could not be confirmed by initial Sanger sequencing.

Genomic position ^a	cDNA ^b	Protein ^c	Region	Exon ^c	Count (%)	Function ^d	Agreement ^e
3299749	c.942C>T	p.R314R	exonic	3	36 (76.6)	S	yes
3298865	rs224212	-	intronic	-	33 (70.2)	-	yes
3297181	c.1422G>A	p.E474E	exonic	5	33 (70.2)	S	yes
3297175	c.1428A>G	p.Q476Q	exonic	5	33 (70.2)	S	yes
3297073	c.1530T>C	p.D510D	exonic	5	33 (70.2)	S	yes
3293888	c.1764G>A	p.P588P	exonic	9	32 (68.1)	S	yes
3293922	rs1231123	-	intronic	-	30 (63.8)	-	yes
3296616	rs224205	-	intronic	-	29 (61.7)	-	yes
3296429	rs224204	-	intronic	-	29 (61.7)	-	yes
3304762	c.306T>C	p.D102D	exonic	2	21 (44.7)	S	yes
3304654	c.414A>G	p.G138G	exonic	2	21 (44.7)	S	yes
3304573	c.495C>A	p.A165A	exonic	2	21 (44.7)	S	yes
3304626	c.442G>C	p.E148Q	exonic	2	19 (40.4)	NS	yes
3304463	c.605G>A	p.R202Q	exonic	2	16 (34.0)	NS	yes
3293407	c.2080A>G	p.M694V	exonic	10	12 (25.5)	NS	yes
3299586	c.1105C>T	p.P369S	exonic	3	6 (12.8)	NS	yes
3299468	c.1223G>A	p.R408Q	exonic	3	6 (12.8)	NS	yes
3293310	c.2177T>C	p.V726A	exonic	10	4 (8.5)	NS	yes
3297100	c.1503C>T	p.R501R	exonic	5	3 (6.4)	S	yes
3294246	rs77380520	-	intronic	-	3 (6.4)	-	yes
3293257	c.2230G>T	p.A744S	exonic	10	3 (6.4)	NS	yes
3293205	c.2282G>A	p.R761H	exonic	10	3 (6.4)	NS	yes
3293403	c.2084A>G	p.K695R	exonic	10	2 (4.3)	NS	yes
3293090	-	-	UTR3	-	2 (4.3)	-	no
3304380	c.688G>A	p.E230K	exonic	2	1 (2.1)	NS	yes
3304317	c.751G>A	p.E251K	exonic	2	1 (2.1)	NS	yes
3304158	c.910G>A	p.G304R	exonic	2	1 (2.1)	NS	yes
3293447	c.2040G>C	p.M680I	exonic	10	1 (2.1)	NS	yes
3293369	c.2118G>A	p.P706P	exonic	10	1 (2.1)	S	yes

^aGenomic position on the hg19 reference genome (NC_000016.9).

^bdbSNP identifiers are shown for variants in non-coding regions.

^cAmino acid information and exon number are only shown for variants in exonic regions.

^dS = synonymous; NS = non-synonymous.

^eAgreement between Nanopore sequencing and initial Sanger sequencing results.

<https://doi.org/10.1371/journal.pone.0265622.t001>

4. Discussion

To evaluate the performance of Nanopore sequencing for SNP genotyping by amplicon sequencing, we performed a comprehensive method comparison with conventional Sanger sequencing using 47 clinical samples from patients with suspicion of FMF. The number of studies comparing Nanopore and Sanger sequencing in diagnostics has been limited [31–33]. Routine diagnostics using Sanger sequencing, the current gold standard for point-mutation detection so far, revealed the presence of various SNPs, including the non-synonymous variants p.E148Q, p.R202Q, p.M694V, p.P369S and p.R408Q in this sample collective [34]. All of these mutations have been previously described in FMF patients [22,35].

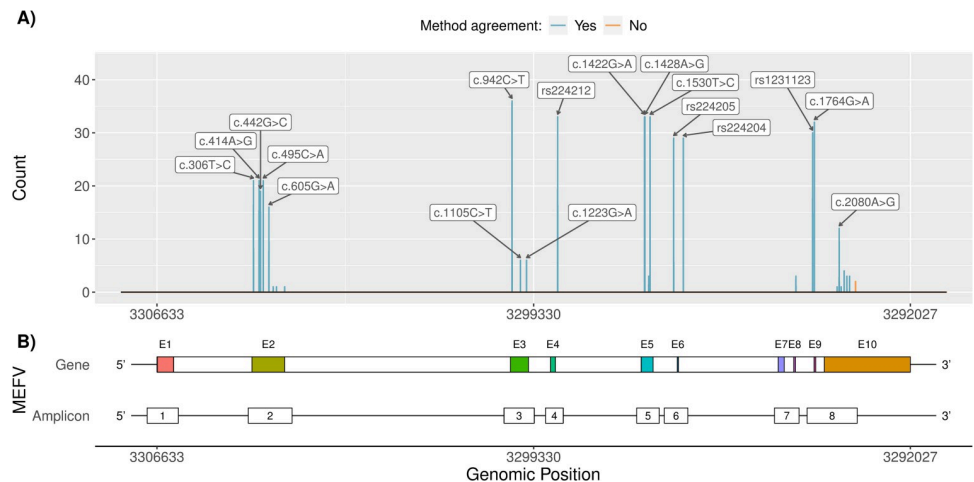


Fig 3. Genetic variants which were identified in selected regions of *MEFV*. (A) Frequency of single nucleotide polymorphisms (SNPs) identified in 47 clinical samples by Sanger and Nanopore sequencing. cDNA labels or dbSNP references are given for the most common variants. Variants with a complete agreement between Sanger and Nanopore sequencing in all 47 clinical samples are coloured in blue and differing variants are coloured in orange. (B) Gene map of *MEFV* and the amplicons used to sequence selected regions of this gene (S1 Table). Genomic positions on the hg19 reference genome (NC_000016.9) are shown in minus strand orientation.

<https://doi.org/10.1371/journal.pone.0265622.g003>

By performing Nanopore sequencing on a MinION sequencing device in combination with a dedicated data analysis pipeline, it was possible to sequence the relevant regions of all *MEFV* exons with a very high read depth. All variants previously identified by diagnostic Sanger sequencing were also accurately detected. Furthermore, Nanopore sequencing revealed only one SNP in two related patients, which had not been identified during initial Sanger sequencing. This SNP was located in the 3' UTR at the edge of the amplicon covering this region. Since current Sanger sequencing is based on PCR amplification and capillary electrophoresis, poor sequence quality due to primer binding and insufficient base resolution is a very common problem at the beginning and end of an individual read [36]. Therefore, low-quality regions are trimmed prior to data analysis. For this reason, the diverging SNP is located in a region of amplicon 8, which cannot be properly sequenced by Sanger sequencing on either the forward or reverse strand. In Nanopore sequencing, a similar problem does not occur since the sequencing adapters are ligated to the ends of the PCR products during library preparation [37]. By sequencing an additional amplicon, spanning the relevant region of the 3' UTR, we were able to confirm the transversion in both patients also by Sanger sequencing. Taking these additional results into account, our data show a complete agreement between Nanopore and Sanger sequencing. Nevertheless, a comprehensive data-base research did not reveal any information about the clinical relevance of this transversion. Since the initial diagnostic Sanger sequencing runs did not identify this variant, the corresponding variant calls were treated as false positive in the calculation of performance measures.

The obtained Precision, Recall, and F1-Score of > 0.99 each demonstrate the excellent agreement between Nanopore and Sanger sequencing for SNP genotyping in our study [38]. This is consistent with other studies that also reported a high degree of agreement for various applications, especially in microbiology and cancer genomics [31,39–41].

The limitations of our study were the small sample size and the focus on targeted SNP genotyping alone. By using targeted amplicon sequencing on the MinION, we were able to sequence the relevant regions of the *MEFV* exons at a high read depth (median read depth 7565x). However, there is a substantial amount of variation in read depth between different

Table 2. Comparison of Nanopore and Sanger sequencing based on various aspects relevant for use in clinical diagnostics.

Aspect	Sanger sequencing	Nanopore sequencing
Capital costs (Instrument, Computing unit, Software) ^a	High (~200000 €)	Low (~3500 €)
Price per <i>MEFV</i> sample [€] ^b	160	75
Time to result [workdays] ^c	3	3
Multiplexing	No	Yes
Data analysis	Simple	Complex
Application in clinical genetics	Reference method	Validation needed

^aBased on current list prices.

^bApproximate price per sample. To archive highest diagnostic accuracy, 11 sequencing reactions must be performed to sequence all target regions with Sanger sequencing, since amplicon 2 and 8 are sequenced in three and two sequencing reactions, respectively. For Nanopore sequencing, the price decreases with increasing degree of multiplexing. ^cIncludes DNA isolation, PCR amplification, sequencing and data analysis.

<https://doi.org/10.1371/journal.pone.0265622.t002>

amplicons within one sample and different samples. This was based on the varying DNA input and varying efficacy of the eight PCR reactions used to amplify the *MEFV* target regions. A more homogeneous read depth distribution could be achieved by determining the concentration of the individual amplicons prior to pooling and subsequent pooling of equimolar amounts. Although this would increase the complexity of the protocol, it would contribute to more homogenous results and probably facilitate a higher degree of multiplexing. Multiplexing of different clinical samples is a key factor in diagnostic NGS as it significantly improves cost efficiency (Table 2) [31]. According to Leija-Salazar et al. a read depth of >100x could be sufficient for accurate variant identification by Nanopore sequencing [10]. Such a threshold would remarkably increase the possible degree of multiplexing in our experimental design. However, due to the inhomogeneous read depth distribution between different amplicons we were not able to evaluate this accurately by subsampling of the data.

Due to the high read depth achieved by amplicon sequencing, we were able to use bcftools for accurate variant calling. This tool employs Bayesian statistics to determine the most likely genotype [38,42]. However, modern diagnostic NGS applications mainly involve gene panel sequencing, whole exome sequencing, and whole genome sequencing [38]. Due to the obviously larger target space, the median read depth in such applications is normally much lower than in amplicon sequencing. Therefore, under these circumstances, it may be necessary to apply more modern tools for accurate variant calling, such as Nanopolish and Medaka (github.com/nanoporetech/medaka), that can handle the unique Nanopore sequencing error profile even at low read depth [43]. Further, structural variant calling including deletions, inversions, tandem duplications, insertions, transpositions, and translocations from Nanopore sequencing data requires also specialised tools [44].

Another important limitation of our study is that we did not utilize the full potential of Nanopore sequencing regarding long read sequencing. By using long reads and tiling amplicon sequencing, it should be possible to sequence the whole gene without the need of amplifying individual exons. While providing the same diagnostic information, this approach would simplify the protocol and reduce the variability in read depth distribution.

Further, prior to clinical application a standardized workflow for sample processing is required.

In the future, in addition to modern bioinformatic data analysis tools, recently announced innovations in nanopores and sequencing chemistry (R10.4 flow cells and Q20+ sequencing chemistry), that increase raw read accuracy, may further improve the performance of Nanopore sequencing for variant identification [45]. Furthermore, they may enable competitive use compared to other

NGS technologies. As mentioned earlier, Nanopore sequencing is especially attractive compared to other technologies like Illumina sequencing, Ion Torrent sequencing or PacBio sequencing due to its fast processing time, lower costs, and ability to generate long reads [45,46].

Summarized, the results of our study show that state-of-the-art Nanopore sequencing in combination with a dedicated data analysis pipeline has a comparable performance to conventional Sanger sequencing for diagnostic SNP genotyping by amplicon sequencing in a clinical setting. Due to continuous technological improvements, after further in-depth clinical validation, this sequencing technique could be applied in clinical genomics and simplify diagnostic workflows in the future.

Supporting information

S1 Fig. Screenshot from IGV showing the MEFV region in which the SNP was solely identified by Nanopore sequencing (red box) in two clinical samples. As Sanger sequencing shows a poor sequence quality at the start and end of a read, this region cannot be sequenced properly by using the routine Sanger sequencing workflow (MF-9-3 = forward sequencing primer Exon 9/10; MF-10-6 = reverse sequencing primer Exon 9/10). By sequencing an additional amplicon, which spans the region containing the variant, it was possible to confirm the transversion in both samples also by Sanger sequencing (MF-10-2 = forward sequencing primer 3' UTR). (TIF)

S2 Fig. Electropherograms from the Sanger sequencing runs of an amplicon spanning the relevant region of the 3' UTR. The transversion from guanine to thymine at genomic position 3293090 is clearly visible in sample 25 (A) and sample 26 (B). (TIFF)

S1 Table. Specific primers used for the amplification of the targets within the MEFV gene. (DOCX)

S2 Table. PCR reaction mixes used for amplification of the targets within MEFV. (DOCX)

S3 Table. PCR reaction programs used for the amplification of the targets within the MEFV gene. (DOCX)

S4 Table. Sequencing primers which were used to sequence the individual amplicons by Sanger sequencing. The final concentration in the reaction mix was 5 μ M. (DOCX)

S5 Table. Overview of the barcode assignment which was used to sequence the clinical samples on a MinION sequencing device. 47 samples were sequenced in four individual runs applying two R9.4.1 flow cells. (DOCX)

Acknowledgments

The authors would like to thank Oliver Woll for technical support with Sanger sequencing.

Author Contributions

Conceptualization: Jonas Schmidt, Sandro Berghaus, Frithjof Blessing, Josef Blessing, Folker Wenzel.

Data curation: Jonas Schmidt, Sandro Berghaus, Holger Herbeck.

Formal analysis: Jonas Schmidt.

Investigation: Jonas Schmidt.

Methodology: Jonas Schmidt.

Project administration: Sandro Berghaus, Josef Blessing, Dirk Roggenbuck.

Resources: Holger Herbeck.

Software: Holger Herbeck.

Supervision: Frithjof Blessing, Dirk Roggenbuck, Folker Wenzel.

Visualization: Jonas Schmidt.

Writing – original draft: Jonas Schmidt.

Writing – review & editing: Jonas Schmidt, Frithjof Blessing, Josef Blessing, Peter Schierack, Stefan Rödiger, Dirk Roggenbuck, Folker Wenzel.

References

1. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics*. 2015; 13:4–16. <https://doi.org/10.1016/j.gpb.2015.01.009> PMID: 25743089
2. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences*. 1996; 93:13770–3. <https://doi.org/10.1073/pnas.93.24.13770> PMID: 8943010
3. Schmidt J, Blessing F, Fimpler L, Wenzel F. Nanopore Sequencing in a Clinical Routine Laboratory: Challenges and Opportunities. *Clin Lab*. 2020; 66. <https://doi.org/10.7754/Clin.Lab.2019.191114> PMID: 32538066
4. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS, Mulvey MR, et al. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci Rep*. 2018; 8:10931. <https://doi.org/10.1038/s41598-018-29334-5> PMID: 30026559
5. Theuns S, Vanmechelen B, Bernaert Q, Deboutte W, Vandenhole M, Beller L, et al. Nanopore sequencing as a revolutionary diagnostic tool for porcine viral enteric disease complexes identifies porcine kobu-virus as an important enteric virus. *Sci Rep*. 2018; 8. <https://doi.org/10.1038/s41598-018-28180-9> PMID: 29959349
6. Smith C, Halse TA, Shea J, Modestil H, Fowler RC, Musser KA, et al. Assessing Nanopore Sequencing for Clinical Diagnostics: a Comparison of Next-Generation Sequencing (NGS) Methods for *Mycobacterium tuberculosis*. *J Clin Microbiol*. 2020; 59. <https://doi.org/10.1128/JCM.00583-20> PMID: 33055186
7. Sakamoto Y, Sereewattanawoot S, Suzuki A. A new era of long-read sequencing for cancer genomics. *J Hum Genet*. 2020; 65:3–10. <https://doi.org/10.1038/s10038-019-0658-5> PMID: 31474751
8. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017; 12:1261–76. <https://doi.org/10.1038/nprot.2017.066> PMID: 28538739
9. Minervini CF, Cumbo C, Orsini P, Anelli L, Zagaria A, Specchia G, et al. Nanopore Sequencing in Blood Diseases: A Wide Range of Opportunities. *Front Genet*. 2020; 11. <https://doi.org/10.3389/fgene.2020.00076> PMID: 32140171
10. Leija-Salazar M, Sedlazeck FJ, Toffoli M, Mullin S, Mokretar K, Athanasopoulou M, et al. Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. *Mol Genet Genomic Med*. 2019:e564. <https://doi.org/10.1002/mgg3.564> PMID: 30637984
11. Bowden R, Davies RW, Heger A, Pagnamenta AT, Cesare Md, Oikkonen, et al. Sequencing of human genomes with nanopore technology. *Nat Commun*.; 10:1869. <https://doi.org/10.1038/s41467-019-09637-5> PMID: 31015479
12. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol*. 2008; 26:1146–53. <https://doi.org/10.1038/nbt.1495> PMID: 18846088

13. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 2016; 17. <https://doi.org/10.1186/s13059-016-1103-0> PMID: 27887629
14. Sonesson C, Yao Y, Bratus-Neuenschwander A, Patrignani A, Robinson MD, Hussain S. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat Commun.* 2019; 10. <https://doi.org/10.1038/s41467-019-11272-z> PMID: 31366910
15. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE.* 2021; 16: e0257521. Epub 2021/10/01. <https://doi.org/10.1371/journal.pone.0257521> PMID: 34597327
16. Jiao X, Zheng X, Ma L, Kutty G, Gogineni E, Sun Q, et al. A Benchmark Study on Error Assessment and Quality Control of CCS Reads Derived from the PacBio RS. *J Data Mining Genomics Proteomics.* 2013; 4. <https://doi.org/10.4172/2153-0602.1000136> PMID: 24179701
17. Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom Bioinform.* 2021; 3:lqab019. Epub 2021/03/27. <https://doi.org/10.1093/nargab/lqab019> PMID: 33817639
18. Wang XV, Blades N, Ding J, Sultana R, Parmigiani G. Estimation of sequencing error rates in short reads. *BMC Bioinformatics.* 2012; 13:185. Epub 2012/07/30. <https://doi.org/10.1186/1471-2105-13-185> PMID: 22846331
19. Weirather JL, Cesare Md, Wang Y, Piazza P, Sebastiano V, Wang X-J, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research.* 2017; 6. <https://doi.org/10.12688/f1000research.10571.2> PMID: 28868132
20. Arteche-López A, Ávila-Fernández A, Romero R, Riveiro-Álvarez R, López-Martínez MA, Giménez-Pardo A, et al. Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. *Sci Rep.* 2021; 11. <https://doi.org/10.1038/s41598-021-85182-w> PMID: 33707547
21. Özen S, Batu ED, Demir S. Familial Mediterranean Fever: Recent Developments in Pathogenesis and New Recommendations for Management. *Front Immunol.* 2017; 8:253. <https://doi.org/10.3389/fimmu.2017.00253> PMID: 28386255
22. Sari I, Birlik M, Kasifoglu T. Familial Mediterranean fever: An updated review. *Eur J Rheumatol.* 2014; 1:21–33. <https://doi.org/10.5152/eurjrheum.2014.006> PMID: 27708867
23. Shohat M, Halpern GJ. Familial Mediterranean fever—a review. *Genet Med.* 2011; 13:487–98. <https://doi.org/10.1097/GIM.0b013e3182060456> PMID: 21358337
24. Ozdogan H, Ugurlu S. Familial Mediterranean Fever. *Presse Med.* 2019; 48:e61–e76. Epub 2019/01/25. <https://doi.org/10.1016/j.jpm.2018.08.014> PMID: 30686512
25. Grandemange S, Aksentijevich I, Jeru I, Gul A, Touitou I. The regulation of MEFV expression and its role in health and familial Mediterranean fever. *Genes Immun.* 2011; 12:497 EP -. <https://doi.org/10.1038/gene.2011.53> PMID: 21776013
26. Shinar Y, Obici L, Aksentijevich I, Bennetts B, Austrup F, Ceccherini I, et al. Guidelines for the genetic diagnosis of hereditary recurrent fevers. *Annals of the rheumatic diseases.* 2012; 71. <https://doi.org/10.1136/annrheumdis-2011-201271> PMID: 22661645
27. Demir F, Doğan ÖA, Demirkol YK, Tekkuş KE, Canbek S, Karadağ ŞG, et al. Genetic panel screening in patients with clinically unclassified systemic autoinflammatory diseases. *Clinical rheumatology.* 2020; 39. <https://doi.org/10.1007/s10067-020-05108-1> PMID: 32458238
28. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38:e164. <https://doi.org/10.1093/nar/gkq603> PMID: 20601685
29. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. Available from: <https://www.R-project.org/>.
30. Zhao S, Agaonov O, Azab A, Stokowy T, Hovig E. Accuracy and efficiency of germline variant calling pipelines for human genome data. *Sci Rep.* 2020; 10. <https://doi.org/10.1038/s41598-020-77218-4> PMID: 33214604
31. Neuenschwander SM, Terrazos Miani MA, Amlang H, Perroulaz C, Bittel P, Casanova C, et al. A Sample-to-Report Solution for Taxonomic Identification of Cultured Bacteria in the Clinical Setting Based on Nanopore Sequencing. *J Clin Microbiol.* 2020; 58. <https://doi.org/10.1128/JCM.00060-20> PMID: 32229603
32. Cumbo C, Minervini CF, Orsini P, Anelli L, Zagaria A, Minervini A, et al. Nanopore Targeted Sequencing for Rapid Gene Mutations Detection in Acute Myeloid Leukemia. *Genes.* 2019; 10:1026. <https://doi.org/10.3390/genes10121026> PMID: 31835432

33. Moon J, Kim N, Lee HS, Shin HR, Lee ST, Jung KH, et al. *Campylobacter fetus* meningitis confirmed by a 16S rRNA gene analysis using the MinION nanopore sequencer, South Korea, 2016. *Emerg Microbes Infect.* 2017; 6. <https://doi.org/10.1038/emi.2017.81> PMID: 29089590
34. Cario R de, Kura ASuraci S, Magi A, Volta A, Marcucci R, et al. Sanger Validation of High-Throughput Sequencing in Genetic Diagnosis: Still the Best Practice. *Front Genet.* 2020; 11. <https://doi.org/10.3389/fgene.2020.592588> PMID: 33343633
35. Comak E, Akman S, Koyun M, Dogan CS, Gokceoglu AU, Arikan Y, et al. Clinical evaluation of R202Q alteration of MEFV genes in Turkish children. *Clin Rheumatol.* 2014; 33:1765–71. <https://doi.org/10.1007/s10067-014-2602-6> PMID: 24718488
36. Crossley BM, Bai J, Glaser A, Maes R, Porter E, Killian ML, et al. Guidelines for Sanger sequencing and molecular assay monitoring. *J Vet Diagn Invest.* 2020; 32:767–75. <https://doi.org/10.1177/1040638720905833> PMID: 32070230
37. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ.* 2019; 61:316–26. <https://doi.org/10.1111/dgd.12608> PMID: 31037722
38. Koboldt DC. Best practices for variant calling in clinical sequencing. *Genome Med.* 2020; 12. <https://doi.org/10.1186/s13073-020-00791-w> PMID: 33106175
39. Crossley BM, Rejmanek D, Baroch J, Stanton JB, Young KT, Killian ML, et al. Nanopore sequencing as a rapid tool for identification and pathotyping of avian influenza A viruses. *J Vet Diagn Invest.* 2021; 33. <https://doi.org/10.1177/1040638720984114> PMID: 33550926
40. Watson CM, La Crinnion, Lindsay H, Mitchell R, Camm N, Robinson R, et al. Assessing the utility of long-read nanopore sequencing for rapid and efficient characterization of mobile element insertions. *Lab Invest.* 2021; 101. <https://doi.org/10.1038/s41374-020-00489-y> PMID: 32989232
41. Watson CM, La Crinnion, Simmonds J, Camm N, Adlard J, Bonthron DT. Long-read nanopore sequencing enables accurate confirmation of a recurrent PMS2 insertion-deletion variant located in a region of complex genomic architecture. *Cancer Genet.* 2021; 256–257. <https://doi.org/10.1016/j.cancergen.2021.05.012> PMID: 34116445
42. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011; 27:2987–93. <https://doi.org/10.1093/bioinformatics/btr509> PMID: 21903627
43. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat Methods.* 2015; 12:733–5. <https://doi.org/10.1038/nmeth.3444> PMID: 26076426
44. Tham CY, Tirado-Magallanes R, Goh Y, Fullwood MJ, Koh BTH, Wang W, et al. NanoVar: accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. *Genome Biol.* 2020; 21. <https://doi.org/10.1186/s13059-020-01968-7> PMID: 32127024
45. Thirunavukarasu D, Cheng LY, Song P, Chen SX, Borad MJ, Kwong L, et al. Oncogene Concatenated Enriched Amplicon Nanopore Sequencing for rapid, accurate, and affordable somatic mutation detection. *Genome Biol.* 2021; 22:1–17. <https://doi.org/10.1186/s13059-020-02207-9> PMID: 33397451
46. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 2012; 13:341. Epub 2012/07/24. <https://doi.org/10.1186/1471-2164-13-341> PMID: 22827831

Chiara Becht, Jonas Schmidt, Frithjof Blessing and Folker Wenzel*

Comparative analysis of alignment tools for application on Nanopore sequencing data

Abstract:

INTRODUCTION: Long-read sequencing techniques such as Oxford Nanopore sequencing, are representing a promising novel approach in molecular-biological methodology, enabling potential facilitation in mapping and de novo assembly. In comparison to conventional sequencing methods, novel alignment tools are mandated to compensate differing data structures (especially high error rate) to achieve acceptably accurate analysis results. **METHODS:** In this study, benchmarking for long read aligners BLASR, GraphMap, LAST, minimap2, NGMLR and the short-read aligner BWA MEM on three experimental datasets was conducted. Obtained alignment results were compared for various quality and performance criteria, such as match rate, mismatch rate, error rate, working memory usage and computational time. **RESULTS:** The comparison yielded differences in alignment quality and performance of tools under test. Tool LAST showed the largest differences among all tools. Minimap2 achieved constant quality with good performance. BLASR, GraphMap, BWA MEM and NGMLR showed slight differences only. **CONCLUSION:** Differences among the tools could be reasoned with dataset characteristics and algorithm approaches of individual tools. All tools except BLASR seem applicable for Nanopore sequencing data. Therefore, selection of the tool should be done under consideration of the experimental design and the further downstream analysis.

Keywords: Alignment tools, Nanopore sequencing, Benchmarking, comparative analysis

<https://doi.org/10.1515/cdbme-2021-2212>

*Corresponding author: **Folker Wenzel:** Faculty of Medical and Life Sciences, Hochschule Furtwangen, Schwenningen, Germany, wfo@hs-furtwangen.de

Chiara Becht: Faculty of Medical and Life Sciences, Hochschule Furtwangen, Schwenningen, Germany

Jonas Schmidt, Frithjof Blessing: MVZ Laborärzte Singen, Singen, Germany

1 Introduction

Sequencing of nucleic acids has advanced to a powerful instrument in molecular biological research and medical diagnostics over time. Developments over three generations evolved in various sequencing techniques, all with the goal of optimizing the core-parameters including sequencing speed and parallelisation, costs, maximum read length and overall error rate [1]. One of the latest technologies is Nanopore sequencing which was commercialized by Oxford Nanopore Technologies (ONT). By applying biological nanopores to translate the sequence information into specific electrical current patterns, this approach offers ultra-long reads accompanied by high sequencing speed and low cost [1]. So far, the most important disadvantage of this technology is the increased error rate compared to other sequencing technologies. However, due to continuous development and improvements of the technology until today the difference has become comparably small [1]. The unique error profile makes the analysis of Nanopore sequencing data challenging and requires specialized algorithms. Handling of errors during different data analysis steps is crucial for the reliability of the result and the final interpretation [2]. One fundamental step in many data analysis pipelines is sequence alignment, which summarizes the process of referring the reads back to a reference sequence [3]. Regarding Nanopore sequencing data, there is a wide variety of specialised alignment tools freely available, which can be used for sequence alignment. Therefore, it is important to compare their performance prior to use and select the best suitable tool for a specific project. This has already been done by others using artificially created dataset [4]. In this work, we focus on the comparison of six state of the art alignment tools with regard to their performance and alignment quality using three different datasets from biological Nanopore sequencing experiments. Further, to account for the flexible use of Nanopore sequencers the comparison was performed using only conventional hardware to test the applicability under standard laboratory conditions.

2 Material and Methods

2.1 Datasets

The test datasets used for comparison of the different alignment tools originate from three different sequencing experiments including I) lambda phage whole genome sequencing, II) amplicon sequencing of the human pyrin innate immunity regulator (MEFV) gene and III) severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) whole genome sequencing (Table 1). All of them were produced during sequencing experiments on a MinION sequencing device using R9.4 flow cells in combination with ligation sequencing following the manufacturers protocol (ONT, Oxford, England). They mainly differ in quality, size and read length distribution.

Prior to the use in this study all datasets were completely anonymized to remove any traceability to a specific sample.

The datasets were prepared for the comparison of the alignment tools in FASTQ file format after base calling raw sequencing data using a high accuracy base calling model of the base caller Guppy [v3.1.5] (ONT). In order to remove chimeric reads, prior to further analysis length filtering based on the expected read length was performed for the datasets originating from amplicon sequencing experiments.

Reference sequences for the alignments were downloaded from the NCBI reference sequence database (SARS-CoV-2 genome: NC_045512.2, human chromosome 16: NC_000016.10, lambda phage genome: NC_001416.1).

Table 1: Characteristics of the three experimental datasets used for the comparison of the alignment tools. All datasets originate from Nanopore sequencing experiments performed on a MinION sequencing device.

Dataset	Target	DNA input	Median Q Score	Total read count	Median read length [bases]
Lambda	Whole Genome	gDNA	13.1	134534	9123
MEFV	MEFV Gene	Amplicons	11.4	154664	463
SARS-CoV-2	Whole Genome	Amplicons	13.2	86418	381

2.2 Computational setup

The comparison of the different alignment tools was conducted on a conventional notebook (Intel Core i7-8565U,

16 GB Random-Access Memory (RAM), 512 GB Solid State Drive (SSD)) running Ubuntu (Ubuntu 18.04.5 LTS) and Microsoft Windows (Windows 10pro version 2004). Base calling was performed on a separate high-performance workstation (Intel Core I7 7700K, 64 GB RAM, Nvidia Geforce RTX2080 TI, 1 TB SSD).

2.3 Comparative analysis

Six different alignment tools were incorporated into the comparison including the dedicated long read alignment tools BLASR [v5.3.3] (github.com/PacificBiosciences/blasr), minimap2 [v2.17-r941] (github.com/lh3/minimap2), LAST [v1060] (github.com/mcfrith/last-genome-alignments), GraphMap [v0.5.2] (github.com/isovic/graphmap) and NGMLR [v0.2.7] (github.com/philres/ngmlr) as well as the state-of-the art short-read alignment tool BWA MEM [v0.7.17-r1188] (github.com/lh3/bwa). For the comparative analysis the tools were used with the default configuration for Nanopore sequencing data as stated in the respective usage manual. In order to compare the performance of the different tools all three datasets were aligned to their respective reference sequence by using each alignment tool and the output was stored in Binary Alignment Map (BAM) files or if not supported by the tool in tab delimited files.

Performance measures including computational time (measured as Central Processing Unit (CPU) time) and working memory consumption (measured as peak RAM usage) of each individual run were recorded by using the time command in Ubuntu. The measurements were performed in single-thread mode and multi-thread mode, when supported by the alignment tool.

The frequencies of matches (identical nucleotides between query and reference), mismatches (differing nucleotides between query and reference), deletions (missing nucleotides in query compared to reference) and insertions (additional nucleotides in query compared to reference) were calculated for each alignment from the alignment output files by using custom R scripts. They were used to calculate quality indicators including match rate (see eq 1), mismatch rate (see eq 2) and error rate (see eq 3).

Median match, error and mismatch rates were statistically compared by applying a generalized linear model (GLM) assuming a Poisson distribution on the underlying count data and using the alignment length as an offset variable.

All data analysis and visualization were done using R [v3.6.3] (R Foundation for Statistical Computing, Vienna, Austria).

$$\text{Match rate} = \frac{\sum \text{matches}}{\sum \text{matches} + \text{mismatches} + \text{deletions} + \text{insertions}} \quad (1)$$

$$\text{Mismatch rate} = \frac{\sum \text{mismatches}}{\sum \text{matches} + \sum \text{mismatches} + \sum \text{deletions} + \sum \text{insertions}} \quad (2)$$

$$\text{Error rate} = \frac{\sum \text{mismatches} + \sum \text{insertions} + \sum \text{deletions}}{\sum \text{matches} + \sum \text{mismatches} + \sum \text{deletions} + \sum \text{insertions}} \quad (3)$$

3 Results

Computational performance of the different alignment tools was assessed by recording CPU time and peak RAM usage (Figure 1). Summarized, all three datasets were successfully processed by each tool on a conventional notebook. The Lambda dataset required the longest median processing time (1913.49 seconds (Lambda) compared to 184.34 seconds (MEFV) and 15.27 seconds (SARS-CoV-2)) and showed the highest median peak RAM usage compared to the other datasets (1760.97 MB (Lambda) compared to 877.68 MB (MEFV) and 75.10 MB (SARS-CoV-2)). Multithreading decreased the CPU time with an increase of the peak RAM usage. The comparison of the different tools showed a superior performance of minimap2 regarding speed and memory consumption on all three datasets.

The performance of the other alignment tools varied between the different datasets. Overall, BWA MEM and LAST showed intermediate CPU time and peak RAM usage. BLASR, GraphMap and NGMLR showed the highest CPU times with varying peak RAM consumption. Especially NGMLR had high peak RAM values.

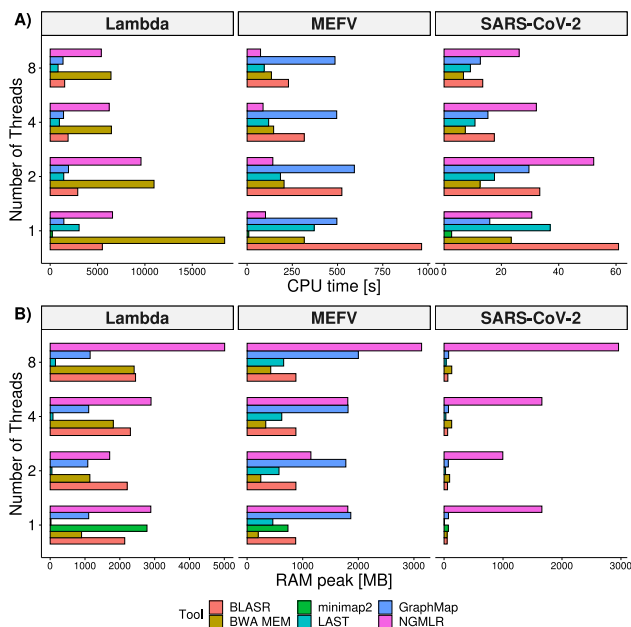


Figure 1: A) CPU time of the different alignment tools running on 1, 2, 4 and 8 threads. B) RAM usage of the different alignment tools when running on 1, 2, 4 and 8 threads. Multithreading was not supported by minimap2.

To compare the alignments produced by the different tools for the three datasets quality measures including match rate, mismatch rate and error rate of the alignments were extracted from the output files.

BLASR was excluded from the calculations of the MEFV dataset due to insufficient position information. Therefore, an analysis of the MEFV amplicon sequencing data under the

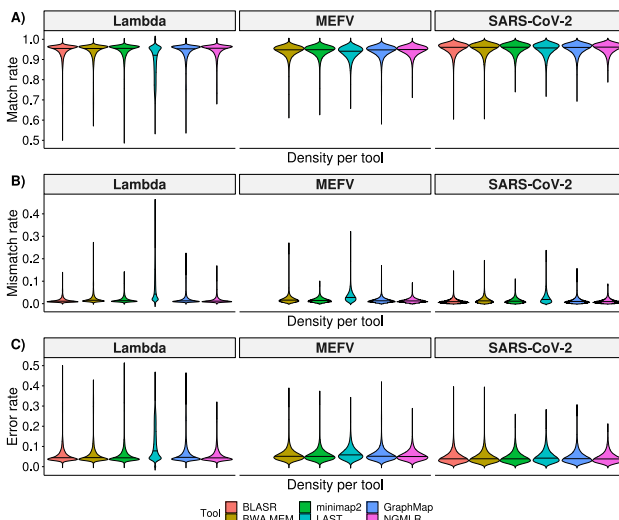


Figure 2: Distribution of A) match rate, B) mismatch rate and C) error rate for the application of the tested alignment tools on all three datasets. BLASR was excluded from the calculations of the MEFV dataset due to insufficient position information, which made the comparative analysis of MEFV exon amplicon sequencing data impossible for this tool.

same conditions as for the other tools was not possible.

All tools except LAST showed a similar distribution of the three rates over all alignments generated per dataset (Figure 2). The achieved match rates ranged mainly between 85 % and 100 % (Figure 2A). The GLM showed no significant influence of the tool choice on the median match rate for all of the datasets. Mismatch rates were mainly observed in the range of 0 % to 5 % for all three datasets (Figure 2B). By applying the GLM a highly significant positive influence (Estimate [E] = 1.145, Standard Error [SE] = 0.155, P < 0.001) of the tool LAST on the median mismatch rate was observed for the Lambda dataset. Additionally, a significant positive influence (E = 0.353, SE = 0.130, P = 0.007) of the tool BWA MEM was observed. For the other two datasets the GLM showed no significant influence of the tool on the median mismatch rate. The error rate which includes mismatches, insertions and deletions ranged mainly between 0 % and 12 % (Figure 2C). Again, tool LAST showed a highly significant positive influence (E = 0.405, SE = 0.100, P < 0.001) on the median error rate for the Lambda dataset. For the other two datasets no significant influences of the tools were observed.

4 Discussion

To evaluate the performance and quality of modern alignment tools for Nanopore sequencing data we compared six long-read alignment tools by using three experimental datasets. The three different datasets were selected to account for varying data structure regarding read length distribution, read count and overall read quality. Computational performance of the different tools was assessed by recording CPU time and peak RAM usage on a conventional notebook during the alignment of the test datasets to the respective reference sequence. Remarkably, by using only standard hardware none of the tools failed to process the alignments. This is a promising result since one big advantage of Nanopore sequencing is the portability of some sequencing devices. To support this feature, it is important that the tools used for data analysis can be applied on standard hardware without the need of high-performance workstations. As expected, the use of multithreading led to reduced runtime with an increased peak RAM consumption. This is important since further process parallelization can seriously speed up the whole data analysis workflow. Major differences in CPU time and peak RAM usage were observed between the different tools. Even without the option of multithreading minimap2 was the fastest tool on all three datasets with an intermediate peak RAM consumption. BLASR, GraphMap and NGMLR showed comparably long run time. This let them appear to be less efficient for the application on long read datasets. In addition, NGMLR showed a high peak RAM usage which might be a bottleneck for larger datasets on less powerful systems. In general, differences in performance between the different tools can be explained by different algorithmic approaches. However, since all tools proved to be functional even with standard hardware, performance should play only a minor role in the selection of a specific alignment tool for a certain experiment. In order to compare the alignments produced by the different tools quality measures including match rate, mismatch rate and error rate were calculated from the output files. Overall, the analysis revealed high match rates and low mismatch and errors rates for all datasets when applying BLASR, BWA MEM, minimap2, GraphMap and NGMLR. This indicates that these tools can be applied for the analysis of Nanopore sequencing data. LAST showed a significantly higher mismatch rate and error rate for the Lambda dataset compared to the other tools. Similar observations were also made by others [5, 6]. Since the Lambda genome is comparably small, long reads should be easy to map. Considering the high match rates and low mismatch/error rates achieved by the other tools, the difference observed for LAST seems to originate from the tool itself [6]. Possible

explanations are the use of scoring algorithm and the formation of local alignments [6]. Although BLASR showed an acceptable performance for the Lambda and the SARS-CoV-2 dataset, this tool can only hardly be implemented in data analysis pipelines for Nanopore sequencing data since output in BAM format is not supported for this kind of data. Benchmarking of alignment tools is frequently done by using synthetic datasets [4]. In this study, we used experimental datasets derived from MinION sequencing runs to evaluate the tools under real experimental conditions. Although this approach helps to evaluate and compare the characteristics of an alignment tool, it has the limitation that true known reference values are missing. Therefore, it is not possible to assess quality criteria like precision and recall of the alignment tools [4]. For extended comparison studies further quality measures including clipping rate, read splitting and the mapping position relative to the reference genome can be included. Summarized, in this work by using experimental datasets we show that all evaluated alignment tools except BLASR are suitable for the analysis of Nanopore sequencing data and the integration in a data analysis pipeline. Especially minimap2 showed a superior performance. Differences among the tools could be reasoned with dataset characteristics and algorithm approaches of individual tools. Therefore, selection of the tool should be done under consideration of the experimental design and further downstream analysis.

Author Statement

Research funding: The author state no funding involved. Conflict of interest: None declared. Informed consent: No individuals were included in this study. Ethical approval: No research related to human use was conducted. The datasets were completely anonymised prior to the use in this study.

References

- [1] Schmidt J, Blessing F, Fimpler L, Wenzel F. Nanopore Sequencing in a Clinical Routine Laboratory: Challenges and Opportunities. *Clin Lab* 2020;66(6).
- [2] Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol.* 2016;34:518–24.
- [3] Pevsner J, Bioinformatics and functional genomics. 3rd ed. Chichester: Wiley Blackwell; 2015.
- [4] Pavlovikj N, Moriyama EN, Deogun JS. Comparative analysis of alignment tools for nanopore reads. *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017.* 2017:169–74.
- [5] Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat Commun* 2016;7:1–11.
- [6] Li H. A few comments on GraphMap. 2015. <https://lh3.github.io/2015/07/30/a-few-comments-on-graphmap>. Accessed 21 Jan 2021.