



#### **Prof. Dr. rer. medic. Barbe Rentsch**

Professorin für Angewandte Biotechnologie sowie Dozentin für biologische, biotechnologische und biomedizinische Module im Studiengang Labor- und Verfahrenstechnik an der Dualen Hochschule Sachsen, Standort Riesa.

Arbeits- und Forschungsschwerpunkte: Molekularbiologie, Zell- und Gewebekultivierung, Bioanalytik, Biomaterialien, Tissue Engineering und Regenerative Medizin.

**Kontakt:** barbe.rentsch@dhsn.de



#### **B. Sc. Sara Alsulaiman Alnaser**

Absolventin des Studiengangs Labor- und Verfahrenstechnik (Studienrichtung Biotechnologie) an der Dualen Hochschule Sachsen, Standort Riesa (Abschlussjahr 2024). Praktische Ausbildung bei der GVG Diagnostics GmbH in Leipzig.

**Kontakt:** saraalnasser40@gmail.com



#### **B. Sc. Leonie Engelmann**

Absolventin des Studiengangs Labor- und Verfahrenstechnik (Studienrichtung Biotechnologie) an der Dualen Hochschule Sachsen, Standort Riesa (Abschlussjahr 2024). Praktische Ausbildung am Institut für Bioprocess- und Analysenmesstechnik e.V. in Heilbad Heiligenstadt.

**Kontakt:** engelmann17@gmx.de



#### **B. Sc. Lucca Heinrich**

Absolvent des Studiengangs Labor- und Verfahrenstechnik (Studienrichtung Biotechnologie) an der Dualen Hochschule Sachsen, Standort Riesa (Abschlussjahr 2024). Praktische Ausbildung bei der FILK Freiberg Institute gGmbH.

**Kontakt:** lucca.heinrich@gmail.com



#### **B. Sc. Melanie Thielemann**

Studium zum Bachelor of Science im Studiengang Labor- und Verfahrenstechnik mit der Studienrichtung Biotechnologie an der Staatlichen Studienakademie Riesa (2009–2012). Seit 2012 tätig als Laboringenieurin an der Dualen Hochschule Sachsen, Standort Riesa, im Studiengang Labor- und Verfahrenstechnik.

Arbeits- und Forschungsschwerpunkte: Organische Chemie, analytische Trennmethoden, instrumentelle Analytik, Mikrobiologie und Umweltmikrobiologie, Biochemie, Molekularbiologie, Zellkulturtechnik und Fermentation.

**Kontakt:** melanie.thielemann@dhsn.de

# Einführung und Evaluation der Nanopore-Sequenzierung: Vergleich mit der Sanger-Methode und Anwendungsbeispiele im Labor

Sara Alsulaiman Alnaser, Leonie Engelmann und Lucca Heinrich,  
Melanie Thielemann, Barbe Rentsch

## Abstract

Die Sequenzierung biologischer Moleküle, insbesondere von DNA und RNA, stellt eine Schlüsseltechnologie in der Molekularbiologie dar. Von der klassischen Sanger-Sequenzierung bis hin zu modernen Next- und Fourth-Generation-Sequencing-Technologien wurden erhebliche Fortschritte erzielt, die sowohl die Geschwindigkeit als auch die Genauigkeit der Datenanalyse revolutionierten. Insbesondere die Oxford Nanopore-Technologie (ONT) bietet durch Echtzeit-Sequenzierung und Portabilität neue Möglichkeiten in der Genomforschung.

Im Rahmen des Moduls Angewandte Biotechnologie wurde die ONT-MinION-Technologie eingeführt und für die Sequenzierung von PCR-Produkten getestet. Ein Protokoll zur DNA-Bibliothekspräparation, Flusszellen-Beladung und Sequenzierung wurde entwickelt und die Qualität der erzeugten Sequenzdaten bewertet. Trotz innovativer Ansätze und einer hohen Sequenzierkapazität zeigt die Nanopore-Technologie im Vergleich zur Sanger-Sequenzierung und Illumina-Plattform noch eine höhere Fehlerrate. Die Analyse der Nanopore-Daten ergab eine durchschnittliche Genauigkeit von 76 % mit typischen Abweichungen wie Insertionen, Deletionen und Substitutionen. Ursachen dafür wurden in PCR-Artefakten und technologischen Limitationen identifiziert.

Trotz dieser Einschränkungen bietet die Nanopore-Technologie das Potenzial für eine kostengünstige und schnelle Analyse großer genomischer Regionen. Die Studie unterstreicht die Notwendigkeit weiterer Optimierungen der Basecalling-Algorithmen und Fehlerkorrekturmethode, um die Genauigkeit der Nanopore-Sequenzierung zu verbessern.

The sequencing of biological molecules, particularly DNA and RNA, represents a key technology in molecular biology. From classical Sanger sequencing to modern next- and fourth-generation sequencing technologies, significant advancements have been made, revolutionizing both the speed and accuracy of data analysis. Notably, Oxford Nanopore Technology (ONT) introduces new possibilities in genome research through real-time sequencing and portability.

As part of the Applied Biotechnology module, the ONT MinION technology was introduced and tested for the sequencing of PCR products. A protocol for DNA library preparation, flow cell loading, and sequencing was developed, and the quality of the resulting sequence data was evaluated. Despite its innovative approach and high sequencing capacity, Nanopore technology still exhibits a higher error rate compared to Sanger sequencing and Illumina platforms. Analysis of the Nanopore data revealed an average accuracy of 76 %, with typical errors such as insertions, deletions, and substitutions. These inaccuracies were attributed to PCR artifacts and technological limitations.

Despite these challenges, Nanopore technology offers the potential for cost-effective and rapid analysis of large genomic regions. The study highlights the need for further optimization of base-calling algorithms and error correction methods to improve the accuracy of Nanopore sequencing.

## Abkürzungen

4G-Seq	engl. Fourth-Generation Sequencing	KB	Kilobasen
AT	engl. dA-Tailing	M	Millionen
bp	Basenpaare	Mb	Megabasenpaar (1.000.000)
ddNTPs	Didesoxyribonukleosid-Triphosphate	NGS	engl. Next-Generation Sequencing
DNA	engl. deoxyribonucleic acid	ONT	Oxford Nanopore-Technologie
EBI	engl. European Bioinformatics Institute	PCR	engl. polymerase chain reaction
ER	engl. End Repair	Q-Score	engl. quality score
FCS	engl. fetal calf serum	RNA	engl. ribonucleic acid
Gb	Giga-Basenpaar (1.000.000.000)	SMRT- Sequenzierung	engl. single-molecule real-time sequencing
IGV	Integrative Genomics Viewer	TGS	engl. Third-Generation Sequencing

## Schlagworte

DNA-Sequenzierung, Echtzeit-Sequenzierung, Oxford Nanopore-Technologie, Nanopore-Sequenzierung, MinION, Sanger-Sequenzierung, PCR-Produkt-Sequenzierung, Myogenin

## Einleitung

Sequenzierung bezeichnet die Bestimmung der genauen Abfolge von Bausteinen in biologischen Molekülen, insbesondere DNA oder RNA. Die Gensequenzierung fokussiert sich auf einzelne Gene oder kurze DNA-Abschnitte, beispielsweise PCR-Produkte (polymerase chain reaction), um genetische Varianten zu identifizieren. Hingegen analysiert die Genomsequenzierung das gesamte Erbgut eines Organismus und wird häufig bei der

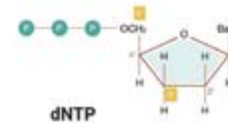
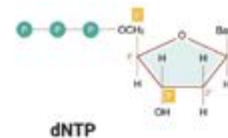
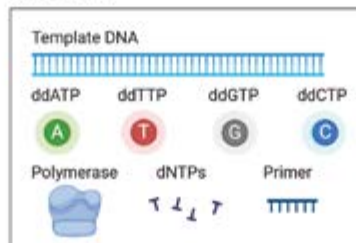
Untersuchung seltener genetischer Erkrankungen eingesetzt. Die Exomsequenzierung beschränkt sich auf die kodierenden Bereiche des Genoms. Die RNA-Sequenzierung dient zur Analyse aller RNA-Moleküle in einer Zelle und liefert wertvolle Einblicke in die Genexpression. [1, 2]

Die Sanger-Sequenzierung, benannt nach ihrem Entwickler Frederick Sanger [3], revolutionierte die Molekularbiologie, indem

### Die drei grundlegenden Schritte

- 1 PCR mit zusätzlichen Kettenabbruchnucleotiden
- 2 Trennung der PCR-Produkte der Größe nach mittels Kapillargelelektrophorese
- 3 Fluoreszenzdetektion, Sequenzanalyse und Rekonstruktion

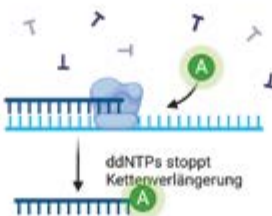
### Reagenzien



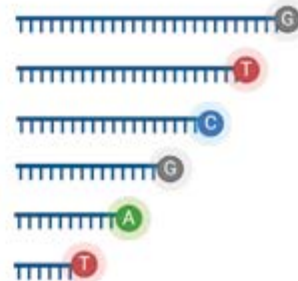
### 1 Primeranlagerung und Kettenverlängerung



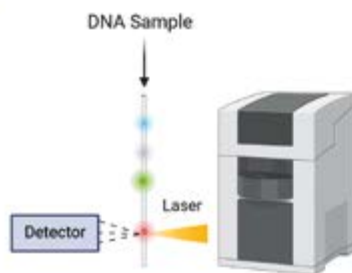
### ddNTP-Bindung und Kettenabbruch



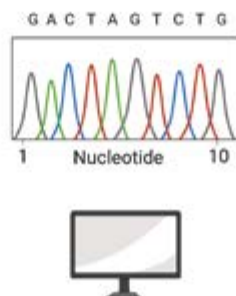
### Fluoreszenzmarkierte DNA-Probe



### 2 Kapillargelelektrophorese und Fluoreszenzdetektion



### 3 Sequenzanalyse und Rekonstruktion



## Abbildung 1: Sequenzieren mithilfe der Kettenabbruchmethode nach Sanger

Bei der Sanger-Methode synthetisiert die DNA-Polymerase, ähnlich einer PCR, ausgehend von einem DNA-Template mit Hilfe eines Primers viele verschiedene DNA-Stränge. Das Enzym bricht nach verschiedenen Nucleotiden (zufälliger Einbau von ddNTPs) ab, sodass Stränge aller möglichen Längen entstehen. Diese werden durch Kapillargelelektrophorese aufgetrennt. Die kleinsten Fragmente bestehen aus dem Primer und dem ersten Nucleotid der DNA-Matrize. Längere Fragmente bestehen dagegen aus dem Primer und längeren Nucleotidsträngen, die komplementär zur DNA-Matrize sind. Ein Laser regt die Fluoreszenzmarkierungen in jedem Fragment an, und ein Detektor erkennt das daraus resultierende emittierte Licht. Da jedes der vier ddNTPs mit einer anderen Fluoreszenzmarkierung versehen ist, kann das emittierte Licht direkt mit der Identität des terminalen ddNTP in Verbindung gebracht werden. Über eine Software werden diese Daten in ein Chromatogramm umgewandelt, in dem die Fluoreszenz-Peaks der einzelnen Nucleotide entlang der DNA-Template aufgezeigt werden. Created in BioRender.com

Quelle in Anlehnung an: [4, 5]

sie eine schnelle und zuverlässige Methode zur Bestimmung der Basensequenz von DNA-Molekülen bereitstellte. Sie war entscheidend für die Entschlüsselung des menschlichen Genoms und legte den Grundstein für zahlreiche weitere Genomprojekte.

Das Prinzip der Sanger-Sequenzierung beruht auf der In-vitro-DNA-Replikation, bei der die DNA-Polymerase komplementäre DNA-Stränge synthetisiert (Abbildung 1). Durch den gezielten Einbau von Didesoxynukleotiden (ddNTPs), denen die 3'-Hydroxylgruppe fehlt, wird die Kettenverlängerung terminiert. Die resultierenden DNA-Fragmente unterschiedlicher Länge werden anschließend mittels Kapillargelelektrophorese aufgetrennt und anhand ihrer Fluoreszenz detektiert. Obwohl die Sanger-Sequenzierung durch modernere Hochdurchsatz-Sequenzierungsverfahren abgelöst wurde, findet sie immer noch Anwendung bei speziellen Aufgaben, wie der Überprüfung von Sequenzierungsdaten oder der Analyse kleiner Genomregionen. [4, 102ff]

Seit der Entwicklung der Sanger-Sequenzierung hat die DNA-Sequenzierung revolutionäre Fortschritte gemacht. Next-Generation Sequencing (NGS)-Technologien wie die Illumina-Plattform ermöglichen die massive parallele Sequenzierung von Millionen von DNA-Fragmenten gleichzeitig. Durch Verfahren wie sequencing by

synthesis, bei dem fluoreszenzmarkierte Nukleotide inkorporiert und detektiert werden, können ganze Genome in kurzer Zeit sequenziert werden. Dies hat zu einer drastischen Reduzierung der Kosten und einer Steigerung des Durchsatzes geführt. Third-Generation Sequencing (TGS)-Technologien wie SMRT-Sequenzierung (single-molecule real-time sequencing) bieten alternative Ansätze. Sie ermöglichen die direkte Sequenzierung einzelner Moleküle ohne vorherige PCR-Amplifikation und liefern oft längere Leselängen. Dadurch können komplexe Genome mit großen Wiederholungsregionen oder epigenetischen Modifikationen besser analysiert werden. [6]

Fourth-Generation Sequencing (4G-Seq)-Technologien wie die Oxford Nanopore-Technologie (ONT) revolutionieren die DNA-Sequenzierung durch ihre Fähigkeit zur Echtzeit-Sequenzierung einzelner Moleküle ohne vorherige PCR-Amplifikation. Bei der Nanopore-Sequenzierung wird ein einzelnes DNA-Molekül durch eine winzige, proteinbasierte oder künstlich hergestellte Pore getrieben. Jedes Nukleotid, das die Pore passiert, verursacht eine charakteristische Veränderung des elektrischen Stroms, die von der Größe und der chemischen Zusammensetzung des Nukleotids abhängt. Diese Veränderungen werden in Echtzeit aufgezeichnet und durch Algorithmen in eine DNA-Sequenz übersetzt (Abbildung 2). [7, 8]

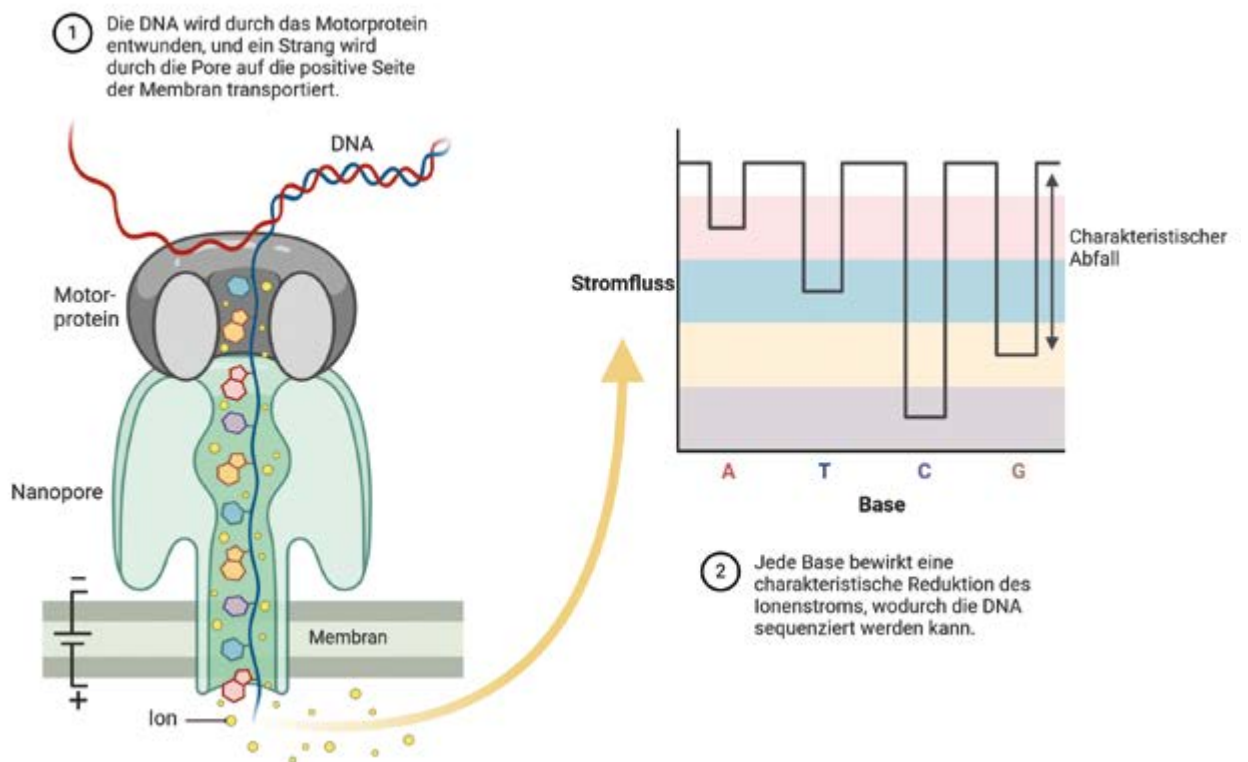


Abbildung 2: Das Prinzip der Nanoporen-Sequenzierung.

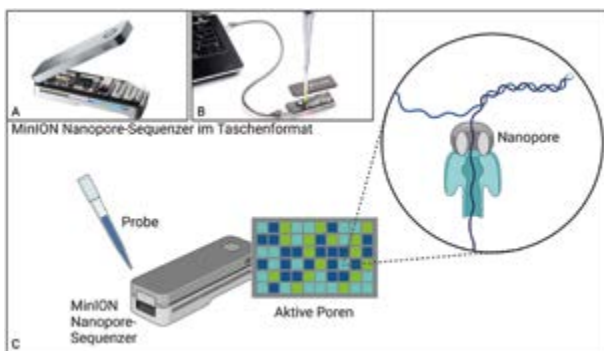
Die Technologie basiert auf einer nanoskaligen Proteinpore oder „Nanopore“, die als Biosensor dient und in eine elektrisch resistente Polymermembran eingebettet ist. In einer elektrolytischen Lösung wird eine konstante Spannung angelegt, um einen Ionenstrom durch die Nanopore zu erzeugen, sodass negativ geladene einzelsträngige DNA- oder RNA-Moleküle durch die Nanopore von der negativ geladenen „cis“-Seite zur positiv geladenen „trans“-Seite getrieben werden. Created in BioRender.com

Quelle in Anlehnung an: [9]

Ein Motor-Enzym zieht das DNA-Molekül schrittweise durch die Nanopore. Dabei trennt es die DNA-Doppelstränge auf, sodass die einzelnen Basen nacheinander die Pore passieren. Die Nanopore-Sequenzierung ermöglicht die direkte Detektion von DNA-Modifikationen wie Methylierungen und bietet somit neue Einblicke in epigenetische Regulationsprozesse. Obwohl die Technologie noch relativ jung ist und eine höhere Fehlerquote im Vergleich zu anderen Sequenziermethoden aufweist, werden durch kontinuierliche Weiterentwicklung und die Kombination mit anderen Technologien immer genauere Ergebnisse erzielt. [9]

Die Nanopore-Sequenzierung findet Anwendung in einer Vielzahl von Bereichen, darunter die De-novo-Assemblierung großer und komplexer Genome, die Analyse von Metagenomen, die Erforschung von Epigenomen und die Echtzeit-Überwachung von Virusmutationen. Aufgrund ihrer Portabilität und der Möglichkeit zur direkten Sequenzierung von langen DNA-Molekülen eröffnet die Nanopore-Sequenzierung neue Perspektiven für die genomische Forschung und die Entwicklung von diagnostischen Tools. [9]

Das MinION von ONT ist ein handliches, tragbares Sequenziergerät, das die Nanopore-Sequenzierung ermöglicht (Abbildung 3). Das Gerät besteht aus einer Durchflusszelle, die Tausende von winzigen Poren enthält, und einem integrierten Computerchip zur Datenerfassung und -verarbeitung. Die Durchflusszelle dient als Träger für die Nanoporen und ermöglicht den kontinuierlichen Fluss der Probenlösung. (Abbildung 3). [8, 10, 11]



**Abbildung 3: MinION-Gerät – Aussehen und Aufbau.**

Das MinION-Gerät ist vollständig tragbar, wiegt 100 g und wird an einen Laptop angeschlossen (A, B). Innerhalb des Gerätes findet sich die Durchflusszelle, auf der die Probe aufgetragen wird und sich die Nanoporen finden, sowie ein Sensor-Chip (ASIC) mit 512 aktiven Kanälen für Datenerfassung und Steuerung (C). Created in BioRender.com  
Quelle in Anlehnung an: [8, 10, 11]

Eine MinION-Durchflusszelle enthält 2048 parallel arbeitende Nanoporen, die in eine Polymermembran eingebettet sind. Jede Pore wird individuell gesteuert und gemessen. Ein Motorprotein zieht einzelsträngige DNA- oder RNA-Moleküle durch die Pore, wobei die charakteristischen Stromänderungen während des Durchtritts jedes Nukleotids zur Bestimmung der Sequenz genutzt werden. ONT hat das System kontinuierlich weiterentwickelt, um die Sequenziergeschwindigkeit und -genauigkeit zu erhöhen.

Aktuelle Versionen wie R10.3 ermöglichen eine Sequenziergeschwindigkeit von bis zu 450 Basen pro Sekunde. [9]

### Zielstellung

Innerhalb des im Studiengang Labor- und Verfahrenstechnik, Studienrichtung Biotechnologie stattfindenden Moduls ‚Angewandte Biotechnologie‘ sollte durch 3 Studierende des 6. Semesters die ONT in Form der MinION-Durchflusszelle im Studiengang eingeführt werden und ein Protokoll zur Sequenzierung von PCR-Produkten mittels der MinION-Technologie entwickelt werden. Dabei sollten folgende Schwerpunkte betrachtet werden:

- Bewertung der Qualität und der Leistungsfähigkeit der MinION-Sequenzierung für die Analyse von DNA-Fragmenten im Vergleich zur etablierten Methode (Sequenzierung nach Sanger),
- Identifizierung optimaler Parameter für die Bibliothekspräparation und Sequenzierung von PCR-Produkten,
- Vergleich der MinION-Daten mit Referenzsequenzen zur Validierung der Methode.

### Material und Methoden

#### Bereitstellung der zu sequenzierenden cDNA

Adhärente C2C12-Maus-Myoblasten (DSMZ: ACC 565, Passage 1) wurden mit einer initialen Zellzahl von 1000 Zellen/cm<sup>2</sup> in Proliferationsmedium (RPMI-Medium, 20 % FCS und 1 % Penicillin/Streptomycin) ausgesät und für 2 Tage (5 % CO<sub>2</sub>, 37 °C) bis zu einer Konfluenz von ca. 60-70 % kultiviert. Die myogene Differenzierung erfolgte in Differenzierungsmedium (reduzierte Serumbedingungen, 2 % FCS) für 4 Tage. [12]

Die Gesamt-RNA der Zellen wurde nach Herstellerangaben isoliert (Nucleospin® RNA, MACHEREY-NAGEL GmbH & Co. KG). Die Quantität sowie die Qualität der RNA wurde mittels eines Nanodrop (Thermo Fisher Scientific) bestimmt. Die cDNA-Synthese wurde in einer 20 µl Reverse Transkriptase (RT)-Reaktion unter Nutzung eines Oligo(dT)<sub>18</sub> Primers (Thermo Fisher Scientific) mit 2 µg RNA nach Herstellerangaben durchgeführt (Omniskript® RT Kit, QIAGEN GmbH). [12]

Das zur Sequenzierung vorgesehene PCR-Produkt (Myogenin, 404 bp) wurde mittels Endpunkt-PCR (Initialisierung 95 °C, 180 s; Denaturierung 95 °C, 10 s; Annealing 60 °C, 30 s; Elongation 72 °C, 60 s; 40 Zyklen) aus der cDNA amplifiziert (Forward-Primer: AAGGTGTG-TAAGAGGAAGTCTGTGTCGG; Reverse-Primer: ACGATGGACGTAAGG-GAGTGACAGATTGT) und im Anschluss per Agarose-Gelelektrophorese getrennt und visualisiert [13, 14].

Die Produkte der Sequenzreaktion wurden im Anschluss nach Herstellerangaben gereinigt (GeneJET PCR Purification Kit, K0701, Thermo Fisher Scientific).

### Nanopore-Sequenzierung

#### Präparation der Enden und Adapterligation

Vor der Nanopore-Sequenzierung müssen die DNA-Fragmente, die häufig unregelmäßige Enden (z. B. 5'- oder 3'-Überhänge) aufweisen, präpariert werden (Abbildung 4). In einem mehrstufigen Prozess wurden diese Enden zunächst geglättet und phosphoryliert (End Repair, ER). Anschließend wurde an das 3'-Ende jedes Fragments ein einzelnes Adeninnukleotid angefügt (dA-Tailing, AT). Dieser dA-Tail diente als Ankerpunkt für die Anbindung von Adaptern, die über einen komplementären Thymin-Überhang verfügen. Die Adapter enthalten spezifische Sequenzen, die für die Bindung an das Motorprotein während der Sequenzierung notwendig sind. [15]

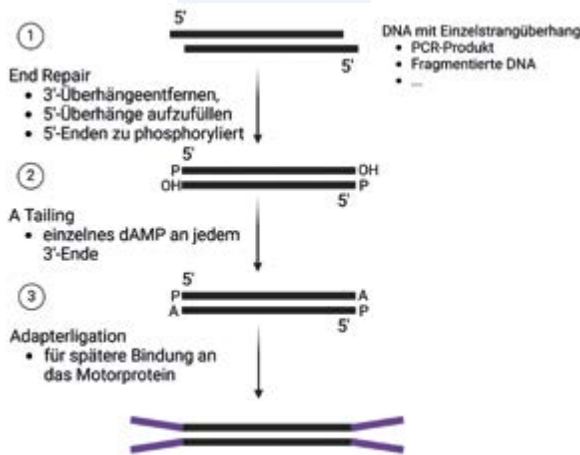


Abbildung 4: Schritt-für-Schritt-Darstellung der Präparation der Enden und Adapterligation

Die DNA wird am Ende repariert, um ein stumpfes Ende zu erzeugen. Ein zusätzliches dAMP wird in das 3'-Ende eines DNA-Fragments eingebaut. Der Adapter wird an beide Enden der DNA ligiert. Created in BioRender.com  
Quelle in Anlehnung an: [15]

Die Präparation der Enden sowie die Adapterligation erfolgten mittels des Ligation Sequencing Kit V14 nach Herstellerangaben (Ligation Sequencing Kit V14, ONT). Im Anschluss an die Adapterligierung wurde die zu sequenzierende DNA unter Nutzung von AMPure XP Beads aufgereinigt.

#### Vorbereiten und Beladen der MinION-Flusszelle sowie die Sequenzierung mittels MinION

Das Vorbereiten (Priming) und Beladen einer MinION-Flusszelle ist ein entscheidender Schritt bei der Nanopore-Sequenzierung. Dabei wurde die Flusszelle gemäß den Herstellerangaben (Ligation Sequencing Kit V14, ONT) für die Sequenzierung vorbereitet, indem die erforderlichen Reagenzien sowie die zu sequenzierende DNA eingeführt wurden (Abbildung 5 A).

Sobald die Flusszelle beladen ist, kann der Sequenzierelauf über die Software MinKNOW gestartet werden (Abbildung 5 B). MinKNOW ist die Betriebssoftware, die zur Steuerung von ONT-Geräten verwendet wird, indem Sequenzierungsparameter eingestellt und Proben verfolgt werden. MinKNOW verwaltet die Datenerfassung und die Echtzeitanalyse, führt lokale Base-Calls durch und gibt die Binärdateien im POD5-Format aus, um sowohl Metadaten, einschließlich des Rohsignals, als auch Leseinformationen zu speichern. Zusätzlich wurden die Daten im FASTQ-Format gespeichert. FASTQ-Dateien sind Textdateien, die Sequenzdaten für jeden Read und die zugehörigen Qualitätsbewertungen pro Base enthalten. [17]

Nach Abschluss der Sequenzierung wurde die Flusszelle mittels des Flow Cell Wash Kits (ONT) gereinigt. Dabei wird die DNA verdaut, ausgespült und das System für das Laden einer neuen Bibliothek vorbereitet. Dieses Verfahren ermöglicht eine mehrfache Verwendung derselben Flusszelle und maximiert so die verfügbare Laufzeit. Alternativ kann ein Lagerungspuffer in die Flusszelle gegeben werden, um sie bei 4-8 °C zu lagern.

#### Analyse der Daten

Aus den Rohdaten der Nanopore-Sequenzierung wurde in MinKNOW ein Basecalling in Echtzeit durchgeführt (Abbildung 6), gefolgt von einem Readcalling. Nach dem Basecalling wurden die Qualitätsmetriken der FASTQ-Dateien überprüft und eine durchschnittliche Read-Länge ermittelt.

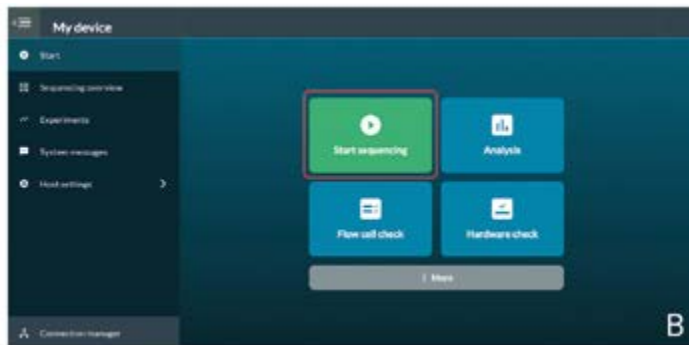
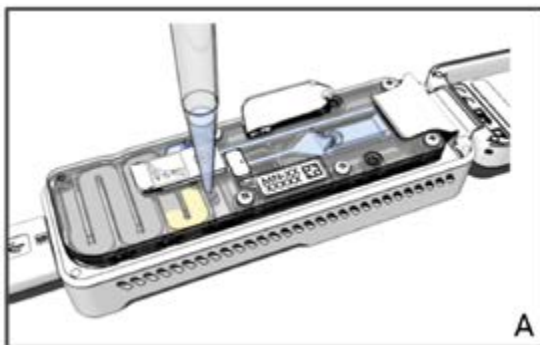


Abbildung 5: Vorbereiten und Beladen der MiniON-Flusszelle sowie die Sequenzierung mittels der MinKNOW-Software

Quelle in Anlehnung an: [16]

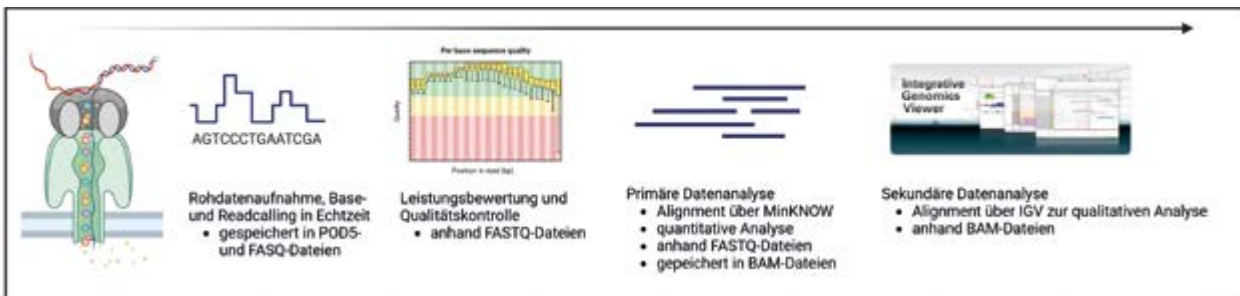


Abbildung 6: Darstellung des analytischen Workflows zur Erzeugung und Auswertung der Daten.

Der Ablauf der Datenanalyse für Oxford Nanopore-Sequenzierungen setzt sich aus der Rohdatenerfassung sowie Basecalling und Readcalling in Echtzeit, der Leistungsbewertung und Qualitätskontrolle, der primäre Datenanalyse, in der die Reads gegen eine Referenzsequenz ausgerichtet werden sowie der sekundären Datenanalyse zusammen. Created in BioRender.com

Im Anschluss an das Base- und Readcalling erfolgte ein primäres Alignment der FASTQ-Daten in MinKNOW (MinKNOW post-run alignment, Tabelle 1). Dieser Workflow ermöglicht ein einfaches Alignment der Oxford Nanopore-Reads gegen eine Referenz und liefert Statistiken zur Qualität des Alignments. Das primäre Alignment in MinKNOW erzeugt BAM-Dateien (Binary Alignment/Map), die eine effiziente Speicherung der Sequenzalignments gegen

Tabelle 1: Darstellung der genutzten Referenzsequenzen und des Workflows in MinKNOW

Referenzsequenz	Accession-Nummer	Workflow
Genesequenz Myogenin	NC_000067.7:134217742-134220286 Mus musculus strain C57BL/6J chromosome 1, GRCm39	MinKNOW Post-run alignment
mRNA Myogenin	NM_031189.2_cds_NP_112466.1_1 [gene=Myog]	MinKNOW Post-run alignment

eine Referenz ermöglichen. [18]

Weiterhin wurden die Alignments unter Nutzung der BAM-Dateien mittels des Integrative Genomics Viewer (IGV) visuell dargestellt [19].

## Ergebnisse und Diskussion

### Rohdatenanalyse sowie Base- und Readcalling

Basecalling ist der Prozess, bei dem die Rohdaten, die während der Sequenzierung (z. B. durch ein Nanopore-Gerät) generiert werden, in die tatsächliche DNA-Sequenz übersetzt werden. Ein Read ist das Resultat des Basecalling-Prozesses. Es handelt sich dabei um die erhaltene Sequenz von Nukleotiden, die aus den Rohdaten der Sequenzierung extrahiert wird. Ein Read kann die vollständige Sequenz einer DNA-Region oder ein Fragment davon darstellen. Als "passed" werden Basen bzw. Reads bezeichnet, wenn diese einen vordefinierten Qualitätsschwellenwert erreichen oder überschreiten, damit wird die Zuverlässigkeit der Daten bestätigt. Im Gegensatz dazu werden Basen bzw. Reads, die unter den Schwellenwert für die Qualitätsbewertung fallen, als

"failed" bezeichnet. [20]

Anhand der Rohdaten wurde zunächst eine Abschätzung vorgenommen und die voraussichtliche Gesamtzahl an Basen (1,82 Gb) berechnet (Abbildung 7, A). Nach Abschluss des Basecalling konnten 1,1 Gb den "passed" Basen und 441,85 Mb den "failed" Basen zugeordnet werden. Insgesamt konnten 3,45 Millionen Reads generiert werden und davon 2,28 Millionen Reads zugeordnet werden (Abbildung 7 B). Der Anteil der "failed" Basen bzw. Reads, die unterhalb des Qualitätsschwellenwertes liegen, werden durch verschiedene Ursachen beeinflusst, die zu einer niedrigeren Sequenzierqualität führen können.

Im Allgemeinen gibt es zwei Quellen, aus denen Sequenzierfehler in Nanopore-Sequenzierungsdaten entstehen können: erstens die Rohsignale, die vom Nanopore-Sequenzierer erzeugt werden, und zweitens der Basecalling-Algorithmus, der die Signaldaten in Nukleotidsequenzen übersetzt. Sequenzierfehler, die durch technische Limitationen, Reagenzienqualität, Signalstärke oder Kontaminationen verursacht werden, führen zu einer Abnahme des Rohsignals und damit zu einer ungenauen Basenidentifikation und beeinflussen die Zuverlässigkeit der gesamten Sequenzanalyse. Zusätzlich zu Verbesserungen der ONT-Geräte und der Chemie für die Nanopore-DNA-Sequenzierung gibt es eine kontinuierliche Entwicklung neuer Basecaller, die sowohl von ONT als auch von der breiteren Forschungsgemeinschaft veröffentlicht werden. [21]

Der N50-Wert, als mittlere Länge eines Sequenzsatzes [22] wird mit 437 Basen angegeben und entspricht damit der erwarteten Länge der sequenzierten DNA (404 Basen, [14]). Der etwas größere Wert kann durch die ligierten Adapter erklärt werden.

Insgesamt wurde über einen Zeitraum von 22 Stunden sequenziert, und es wurden 24,02 GB an Sequenzdaten erzeugt, welche in 864 POD5 bzw. FASTQ-Dateien gespeichert wurden. Davon wurden 575 POD5- bzw. FASTQ-Dateien als "passed" eingestuft (pod5\_pass/ fastq\_pass). Jede POD5- bzw. FASTQ-Datei enthält wiederum ca. 4000 Reads. Dies entspricht in der Summe ca. 2,3

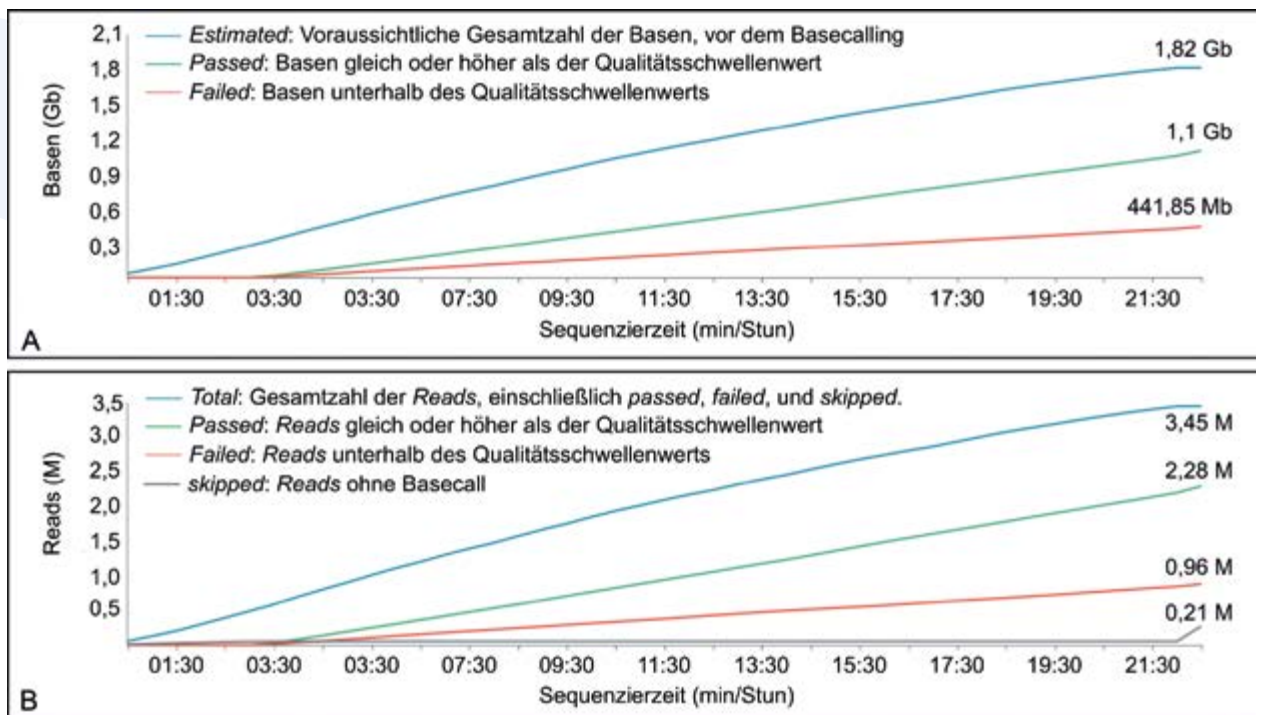


Abbildung 7: kumulatives Ergebnisdiagramm (cumulative output)

Bei der Nanopore-Sequenzierung bieten cumulative output-Metriken einen Überblick über die gesamte Sequenzierungsleistung und Qualitätskontrollmaßnahmen auf der Grundlage der während des Sequenzierens generierten Daten

Millionen Reads. Im Gegensatz dazu erzeugt die traditionelle Sanger-Sequenzierung, die im Studiengang durchgeführt wird, in der Regel einen einzigen Read pro Sequenzreaktion. Ein solcher Lauf dauert etwa eine Stunde. Bei einer kontinuierlichen Nutzung des Sequenzers könnten somit in 24 Stunden rund 24 Reads generiert werden. (Tabelle 3)

**Bewertung der Qualität der Sequenzierung (quality score)**

Der quality score (Q-Score) ist ein Wert, der die Wahrscheinlichkeit bestimmt, dass während des Laufs eine falsche Base aufge-

Q-Score	Wahrscheinlichkeit falscher Basecalls	Genauigkeit
10	1 zu 10	90 %
20	1 zu 100	99 %
30	1 zu 1.000	99,9 %
40	1 zu 10.000	99,99 %

Tabelle 2: Qualitätsergebnisse und Genauigkeit des Basecalling

Quelle: [23]

rufen wurde (Tabelle 2). [21, 23] Die typische Lesegenauigkeit reicht von ~90 % für traditionelle lange Reads bis zu >99 % für kurze Reads. [24]

Ein Q-Score über 30 gilt allgemein als gute Qualität für die meisten Sequenzierungsexperimente. In dieser Arbeit wurde jedoch

ein durchschnittlicher Q-Score von etwa 10,5 erreicht, was einer Fehlerrate von rund 10 % (1 Fehler pro 10 Basen) und einer Genauigkeit von etwa 90 % entspricht.

Die Nanopore-Basecalling-Technologie weist im Vergleich anderen Sequenzierung derzeit eine höhere Fehlerquote von etwa 5 % bis 15 % auf, während die Sanger-Sequenzierung und die Illumina HiSeq-Plattform eine deutlich geringere Fehlerrate von etwa 0,01 % bis 0,1 % haben, mit den meisten Reads über einem Q-Score von 30. [23, 25, 26]

Im Studiengang erreichen Studierende bei der traditionell durchgeführten Sanger-Sequenzierung des Myogenin-PCR-Produkts normalerweise einen Q-Score von etwa 20, was einer Genauigkeit von 99 % und einer Fehlerrate von 1 % entspricht. Niedrige Q-Scores können zu einer erhöhten Anzahl falsch-positiver Calls führen, was ungenaue Schlussfolgerungen und höhere Kosten für Experimente nach sich ziehen kann [23].

Um eine hohe Qualität der Sequenzierungsdaten zu gewährleisten, wurden und werden zahlreiche Verbesserungen in den Bereichen Ausgangsmaterial, Sequenzierungsstrategie und Datenverarbeitung entwickelt. So kommen kontinuierlich optimierte Kits für die Bibliotheksvorbereitung auf den Markt, und die Algorithmen zur Datenanalyse werden fortlaufend weiterentwickelt. [27]

**Primäre und sekundäre Sequenzanalyse**

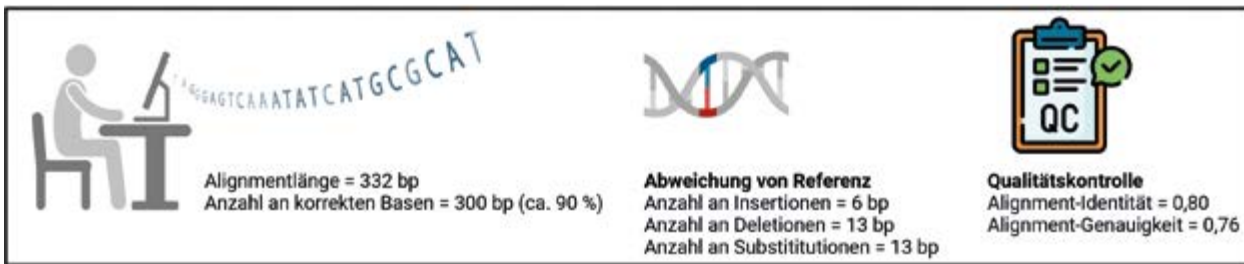


Abbildung 8: Darstellung der Ergebnisse aus dem MinKNOW post-run alignment

Created in BioRender.com

Für ein primäres Sequenzalignment wurden die 575 FASTQ-Dateien (fastq\_pass) einem MinKNOW post-run alignment unterzogen [18]. Als Referenz diente die mRNA von Myogenin (NM\_031189.2). Es wurden 575 BAM-Dateien erzeugt.

Insgesamt wurden 878.275 Einzelsequenzen mit einer durchschnittlichen Länge von ca. 332 bp gegen die Referenzsequenz verglichen (Abbildung 8). Dabei konnten im Durchschnitt 300 bp korrekt ermittelt werden. Abweichungen in Form von Insertionen (6 bp), Deletionen (13 bp) sowie Substitutionen (13 bp) summierten sich durchschnittlich auf 32 bp. Daraus ergibt sich eine Alignment-Identität von 0,80 (80%) und eine Alignment-Genauigkeit von 0,76 (76%).

Die Alignment-Identität, auch als Fehlerrate bezeichnet, gibt den Prozentsatz der identischen Basen zwischen der Abfragesequenz und der Referenzsequenz in einem Alignment an. Die Alignment-Genauigkeit berücksichtigt hingegen sowohl die Identität als auch andere Alignment-Merkmale wie Lücken und Fehlanpassungen [28]

Ähnlich wie beim Q-Score zeigt sich, dass die Genauigkeit von Alignments mit Sequenzen aus der Nanopore-Sequenzierung derzeit noch niedriger ist als bei anderen Sequenzierplattformen. Die Fehlerrate nach Alignment liegt hier zwischen 1 % und

10 % [29, 30]. Im Vergleich dazu erreichen NGS-Technologien wie die Illumina-Plattform Fehlerraten von 0,28 % bis 1,8 %. Die geringste Fehlerrate weist weiterhin die traditionelle Sanger-Sequenzierung auf (0,001 %). [27]

Anhand exemplarischer BAM-Dateien aus dem primären Alignment in MinKNOW wurde Alignments gegen die mRNA von Myogenin (NM\_031189.2) mithilfe des IGV durchgeführt [19]. Diese dienten zur qualitativen Überprüfung der Alignments (Abbildung 9).

Die qualitative Analyse der Alignments ergab eine signifikante Fehlerrate, die mit den Ergebnissen der vorangegangenen quantitativen Analysen übereinstimmt. Wie in Abbildung 9 deutlich wird, treten in den Alignments mehrfach Abweichungen von der Referenzsequenz auf, darunter Insertionen, Deletionen und Substitutionen.

Die Ursachen für diese Fehlerraten sind vielfältig. Eine wesentliche Rolle spielen dabei die PCR-Schritte, die vor der eigentlichen Sequenzierung durchgeführt werden. Während der Amplifikation können Polymerase-Fehler auftreten, bei denen falsche Basen eingebaut werden. Diese Fehler können sich durch die zyklische Verstärkung in der PCR vervielfachen. Der Einsatz von High-Fidelity-Enzymen, die durch ihre hohe Genauigkeit und Zuverlässigkeit während der Amp-

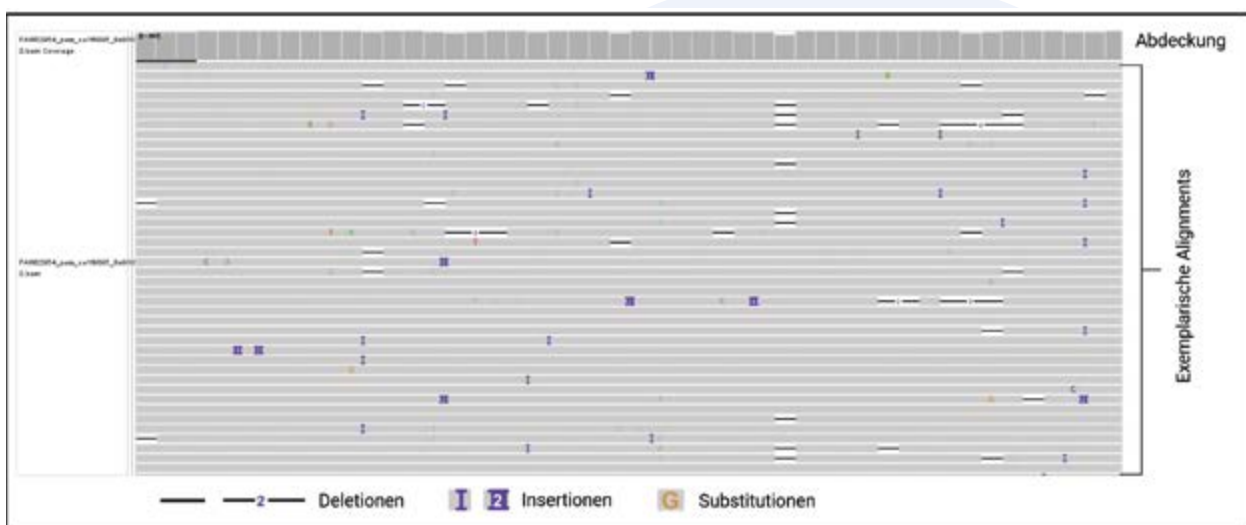


Abbildung 9: Exemplarische Darstellung von Alignments in IGV

Dargestellt ist ein Auszug aus einem Alignment. Alle übereinstimmenden Bereiche zur Referenz sind in Grau dargestellt, Deletionen durch einen Querstrich, Insertionen in Lila und Substitutionen durch die farbige Base auf grauem Grund.

lifikation hervorstechen, bietet eine effektive Möglichkeit, den Einbau fehlerhafter Basen zu minimieren [27]. Darüber hinaus beeinflussen auch die Softwarelösungen für das Base- und Readcalling sowie die verwendeten Algorithmen der Alignmentsoftware die Fehlerrate maßgeblich. Optimierungen in diesen Bereichen, wie die Anwendung fortschrittlicher Fehlermodellierungsansätze oder algorithmischer Verbesserungen, könnten dazu beitragen, die Genauigkeit der Analyse zu erhöhen und die Gesamtfehlerrate zu verringern. [27, 31]

**Zusammenfassung und Fazit**

Innerhalb des von den Studierenden durchgeführten Projekts konnte die Nanopore-Sequenzierung mithilfe des MinION-Geräts von Oxford Nanopore Technologies (ONT) erfolgreich zur Sequenzierung von PCR-Produkten etabliert werden. Die erhobenen Daten zur Qualität und Leistungsfähigkeit der Methode sowie die anschließende Analyse der resultierenden Sequenzen ermöglichten einen Vergleich mit der im Studiengang Labor- und Verfahrenstechnik als Standard etablier-

Merkmal	Nanopore-Sequenzierung	Sanger-Sequenzierung
Vorbereitungszeit (Sequenzreaktion)*	Ca. 60 min	Ca. 90 min
Laufzeit	Ca. 24 h	Ca. 1 h pro Read
Datenmenge	24,02 GB	1 Read pro Sequenzreaktion (ca. 150 KB)
Reads/24 h	4000	24
Read-Länge	Ca. 400	Ca. 340
Q-Score	Ca. 10	Ca. 20
Alignment-Fehlerrate	Ca. 20 %	Ca. 0,001 %
Kosten Gerät**	1000 €	250.000 €
Kosten Sequenzierung pro MB**	2-10 €	500 €
Kosten Sequenzierung pro Lauf**	500-1000 €	5 €

**Tabelle 3. Sequenzierung von Myogenin-PCR-Produkten aus Zellkulturproben – Vergleich Nanopore-Sequenzierung versus Sanger-Sequenzierung**

\*Ohne RNA-Isolierung, RT-Reaktion und Myogenin-PCR; \*\* Die in der Tabelle angegebenen Kosten sind Schätzwerte und können je nach Anbieter und Projekt variieren. Quellen: [32, 33]

ten Sanger-Sequenzierung. Tabelle 3 zeigt einen direkten Vergleich zwischen der Nanopore- und der Sanger-Sequenzierung, wobei der Schwerpunkt auf der Analyse von Myogenin-PCR-Produkten aus Zellkulturproben liegt.

Die Nanopore-Sequenzierung generiert deutlich mehr Daten in kürzerer Zeit. Sie ist somit ideal für Hochdurchsatz-Anwendungen. Die Anschaffungskosten für ein Nanopore-Sequenziergerät sind deutlich niedriger. Die Kosten pro Sequenzierung sind ebenfalls günstiger, insbesondere bei großen Datenmengen. Die Readlänge

ist bei beiden Methoden vergleichbar. Die Sanger-Sequenzierung bietet in der Regel eine höhere Genauigkeit (höherer Q-Score, geringerer Fehlerrate). Die Sanger-Sequenzierung ist gut geeignet für kleinere Projekte und zur Überprüfung einzelner Sequenzen. Die Nanopore-Sequenzierung eignet sich eher für größere Projekte, wie Metagenomik, Transkriptomik und Echtzeit-Sequenzierungen.

**Literaturverzeichnis**

[1] CD Genomics, Targeted Region Sequencing. [Online]. Verfügbar unter: [https://www.cd-genomics.com/targeted-region-sequencing.html?gclid=CjwKCAiAqNSsBhAvEiwAn\\_tmxcUcGDj0\\_9D29SQ3SSo1RzWeo5D79P96T7ofwA9dw1Qud\\_BcPfqUJxoCXekQAvD\\_BwE](https://www.cd-genomics.com/targeted-region-sequencing.html?gclid=CjwKCAiAqNSsBhAvEiwAn_tmxcUcGDj0_9D29SQ3SSo1RzWeo5D79P96T7ofwA9dw1Qud_BcPfqUJxoCXekQAvD_BwE) (Zugriff am: 3. Januar 2024).

[2] D. Qin, "Next-generation sequencing and its clinical application" (eng), *Cancer biology & medicine*, Jg. 16, Nr. 1, S. 4–10, 2019, doi: 10.20892/j.issn.2095-3941.2018.0055.

[3] F. Sanger, S. Nicklen und A. R. Coulson, "DNA sequencing with chainterminating inhibitors" (eng), *Proceedings of the National Academy of Sciences of the United States of America*, Jg. 74, Nr. 12, S. 5463–5467, 1977, doi: 10.1073/pnas.74.12.5463.

[4] D. P. Clark und N. J. Pazdernik, *Molekulare Biotechnologie: Grundlagen und Anwendungen*. Heidelberg: Spektrum Akad. Verl., 2009.

[5] Merck KGaA, Sanger-Sequenzierung – Schritte & Methode. [Online]. Verfügbar unter: <https://www.sigmaaldrich.com/DE/de/technical-documents/protocol/genomics/sequencing/sanger-sequencing> (Zugriff am: 3. Januar 2024).

[6] N. Gupta und V. K. Verma, "Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality" in *Microorganisms for Sustainability, Microbial Technology for the Welfare of Society*, P. K. Arora, Hg., Singapore: Springer Singapore, 2019, S. 313–341, doi: 10.1007/978-981-13-8844-6\_15.

[7] S. E. de Bruijn, Z. Fadaie, F. P. M. Cremers, H. Kremer und S. Roosing, "The Impact of Modern Technologies on Molecular Diagnostic Success Rates, with a Focus on Inherited Retinal Dystrophy and Hearing Loss" (eng), *International journal of molecular sciences*, Jg. 22, Nr. 6, 2021, doi: 10.3390/ijms22062943.

[8] Oxford Nanopore Technologies, How nanopore sequencing works, o.J. [Online]. Verfügbar unter: <https://www.youtube.com/watch?v=RcP85JHlmln>

[9] Y. Wang, Y. Zhao, A. Bollas, Y. Wang und K. F. Au, "Nanopore sequencing technology, bio-informatics and applications" (eng), *Nature biotechnology*, Jg. 39, Nr. 11, S. 1348–1365, 2021, doi: 10.1038/s41587-021-01108-x.

- [10] Oxford Nanopore Technologies, How is Oxford Nanopore used?, o.J. [Online]. Verfügbar unter: <https://www.youtube.com/watch?v=e9wtAlvPPxY>
- [11] VWR, MinION™ DNA/RNA Sequencing Device. [Online]. Verfügbar unter: <https://de.vwr.com/store/product/36202188/miniontm-dna-rna-sequencing-device> (Zugriff am: 27. Januar 2024).
- [12] M. Frenzel, M. Thielemann und B. Rentsch, "Charakterisierung der myogenen Differenzierung von C2C12-Mausmyoblasten als Modell der Skelettmuskelentwicklung", *Wissen im Markt*, 6. Jg., S. 69–77, 2022. [Online]. Verfügbar unter: [https://www.ba-sachsen.de/fileadmin/user\\_upload/WIM\\_2022\\_WEB\\_Frenzel\\_Thielemann\\_Rentsch.pdf](https://www.ba-sachsen.de/fileadmin/user_upload/WIM_2022_WEB_Frenzel_Thielemann_Rentsch.pdf)
- [13] P. Y. Lee, J. Costumbrado, C.-Y. Hsu und Y. H. Kim, "Agarose gel electrophoresis for the separation of DNA fragments" (eng), *Journal of visualized experiments : JoVE*, Nr. 62, 2012, doi: 10.3791/3923.
- [14] M. Takahashi und S. Kubota, "Exerciserelated novel gene is involved in myoblast differentiation" (eng), *Biomedical research (Tokyo, Japan)*, Jg. 26, Nr. 2, S. 79–85, 2005, doi: 10.2220/biomedres.26.79.
- [15] K. Xiong et al., "Duplex-Repair enables highly accurate sequencing, despite DNA damage" (eng), *Nucleic acids research*, Jg. 50, Nr. 1, e1, 2022, doi: 10.1093/nar/gkab855.
- [16] Oxford Nanopore Technologies, Nanopore Sequencing - guide MinION Mk1B. [Online]. Verfügbar unter: <https://community.nanoporetech.com/guides/minion/rna004/5/start-your-run?os=windows>.
- [17] Oxford Nanopore Technologies, Workflows and Analyse. [Online]. Verfügbar unter: <https://nanoporetech.com/documentation/results?category=analyse&topic=epi2me-labs>.
- [18] Oxford Nanopore Technologies, MinKNOW - Post-run alignment. [Online]. Verfügbar unter: <https://nanoporetech.com/document/experiment-companion-minknow#post-run-alignment=>.
- [19] J. T. Robinson et al., "Integrative genomics viewer" (eng), *Nature biotechnology*, Jg. 29, Nr. 1, S. 24–26, 2011, doi: 10.1038/nbt.1754.
- [20] Oxford Nanopore Technologies, Data analysis. [Online]. Verfügbar unter: <https://nanoporetech.com/document/data-analysis>.
- [21] A. Davey, How-to: NGS Quality Control. [Online]. Verfügbar unter: <https://frontlinegenomics.com/how-to-ngs-quality-control/>.
- [22] H. Alhakami, H. Mirebrahim und S. Lonardi, "A comparative evaluation of genome assembly reconciliation tools" (eng), *Genome biology*, Jg. 18, Nr. 1, S. 93, 2017, doi: 10.1186/s13059-017-1213-3.
- [23] Illumina, Inc., "Quality Scores for Next-Generation Sequencing: Assessing sequencing accuracy using Phred quality scoring", 2011. [Online]. Verfügbar unter: [https://www.illumina.com/documents/products/technotes/technote\\_Q-Scores.pdf](https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf).
- [24] PacBio, "Understanding accuracy in DNA sequencing", 2022. [Online]. Verfügbar unter: <https://www.pacb.com/wp-content/uploads/Informational-Guide-Understanding-Accuracy-in-DNA-Sequencing.pdf>.
- [25] N. Gurson, "When Do I Use Sanger Sequencing vs NGS?", Thermo Fisher Scientific, 2015. [Online]. Verfügbar unter: <https://www.thermofisher.com/blog/behindthebench/when-do-i-use-sanger-sequencing-vs-ngs-seq-it-out-7/>.
- [26] Y.-Z. Zhang et al., "Nanopore basecalling from a perspective of instance segmentation" (eng), *BMC bioinformatics*, Jg. 21, Suppl 3, S. 136, 2020, doi: 10.1186/s12859-020-3459-0.
- [27] C. Cheng, Z. Fei und P. Xiao, "Methods to improve the accuracy of next-generation sequencing" (eng), *Frontiers in bioengineering and biotechnology*, Jg. 11, S. 982111, 2023, doi: 10.3389/fbioe.2023.982111.
- [28] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high through-put" (eng), *Nucleic acids research*, Jg. 32, Nr. 5, S. 1792–1797, 2004, doi: 10.1093/nar/gkh340.
- [29] K. Sahlin und P. Medvedev, "Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis" (eng), *Nature communications*, Jg. 12, Nr. 1, S. 2, 2021, doi: 10.1038/s41467-020-20340-8.
- [30] D. Joshi, S. Mao, S. Kannan und S. Diggavi, "QAlign: aligning nanopore reads accurately using current-level modeling" (eng), *Bioinformatics (Oxford, England)*, Jg. 37, Nr. 5, S. 625–633, 2021, doi: 10.1093/bioinformatics/btaa875.
- [31] J. Shang, F. Zhu, W. Vongsangnak, Y. Tang, W. Zhang und B. Shen, "Evaluation and comparison of multiple aligners for next-generation sequencing data analysis" (eng), *BioMed research international*, Jg. 2014, S. 309650, 2014, doi: 10.1155/2014/309650.
- [32] R. Ogden, N. Vasiljevic und S. Prost, "Nanopore sequencing in non-human forensic genetics" (eng), *Emerging topics in life sciences*, Jg. 5, Nr. 3, S. 465–473, 2021, doi: 10.1042/ETLS20200287.
- [33] B. Birla, PacBio vs. Oxford Nanopore sequencing.