



## Ganzgenom-Sequenzierungen von SARS-CoV-2 mittels NGS

In den letzten Jahren kam es zu einer extrem spannenden Entwicklung bezüglich neuer Verfahren zur Sequenzierung der Erbinformation von Organismen und Viren, die unter dem Begriff „Next-Generation Sequencing“ (NGS) zusammengefasst werden. Hierbei handelt es sich um die Möglichkeit des parallelen Sequenzierens von sehr vielen DNA-Fragmenten in einem einzigen Sequenzierlauf. Erst wurden diese Verfahren in der Forschung und für die Identifizierung krankheitsassoziierter Gene eingesetzt. Mittlerweile hat die Anwendung der NGS-Analytik Einzug in die molekulargenetische Routinediagnostik gehalten.

Um die in Deutschland zirkulierenden SARS-CoV-2-Varianten zu ermitteln, trat am 18. Januar 2021 die Corona-Surveillanceverordnung (CorSurV) in Kraft. Darin werden Laboratorien verpflichtet, Ganzgenom-Sequenzierungen von SARS-CoV-2-Isolaten vorzunehmen bzw. diese von einem Sequenzierungslabor durchführen zu lassen.

Der Vorteil des Einsatzes von NGS-basierter Diagnostik gegenüber der konventionellen Sequenzierung mittels Sanger zeigt sich in der extrem hohen Sequenzierkapazität und Sequenztiefe gleich mehrerer Proben in einem Ansatz.

### Durchführung des Next-Generation Sequencing

Um die Sequenzierung des SARS-CoV-2-Genoms durchführen zu können, muss die zu untersuchende Probe positiv auf SARS-CoV-2-RNA getestet worden, der CT-Wert sollte bei 27 liegen bzw. 100 bzw. 1000 Genome pro ml Probe einhalten. Diese Menge an Virusgenomen ist zur Erstellung einer sog. Probenspezifischen DNA-Bibliothek („Library“) erforderlich, welche am Ende der Sequenzierung dann vollständige Virusgenome liefert. Diese Library stellt die notwendige Grundlage für die eigentliche Sequenzierungsreaktion im Sequenzierungs-Gerät dar. Für eine „Library“ werden DNA Fragmente in mehreren aufwendigen Reaktionsschritten mit Erkennungselementen für die eigentliche Vermehrung (Amplifikation) im Sequenzierer versehen s. Abb. 1.

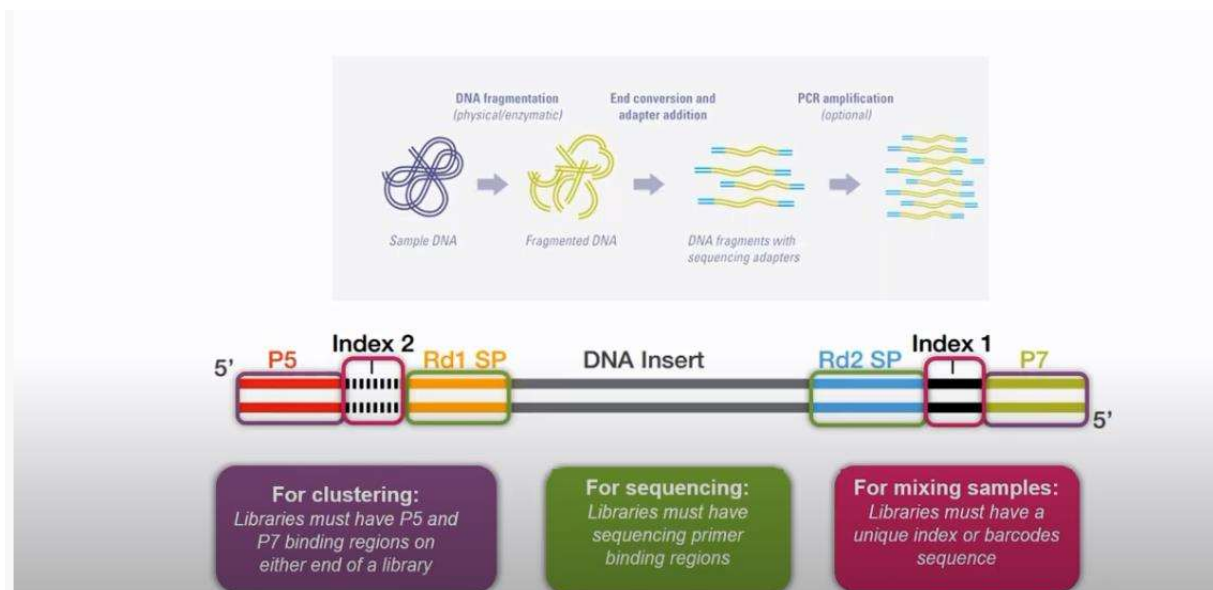


Abbildung 1 Die Vorbereitung der DNA Fragmente für die Sequenzierung (Library Preparation) ist der entscheidende Schritt zum verwertbaren Ergebnis. Das zu sequenzierende DNA Fragment (DNA Insert) wird mit Primer Bindungsstellen für die Sequenzierungsprimer Read 1 (Rd1 SP) und Read 2 (Rd2 SP) versehen, Barcode Index 1 und 2 zur Erkennung verschiedener Proben im Sequenzierungscluster und 2 Adapter zur Brücken-Bindung an die Durchflusszelle (P5 und P7) werden angefügt



In unseren molekulardiagnostischen Laboren setzen wir MiSeq-Sequenzierer ein, welche die DNA-Sequenz bei der sogenannten „Sequencing-by-Synthesis“ (SBS)-Methode analysieren. Bei dieser Methode wird die zuvor aus der Patientenprobe isolierte RNA in DNA umgeschrieben, anschließend wird die DNA vervielfältigt (amplifiziert) und über spezifische Adaptoren kovalent an einen Glasobjektträger (Fließzelle oder engl. "FlowCell") gebunden, auf dem dann die eigentliche Sequenzierungsreaktion stattfindet.

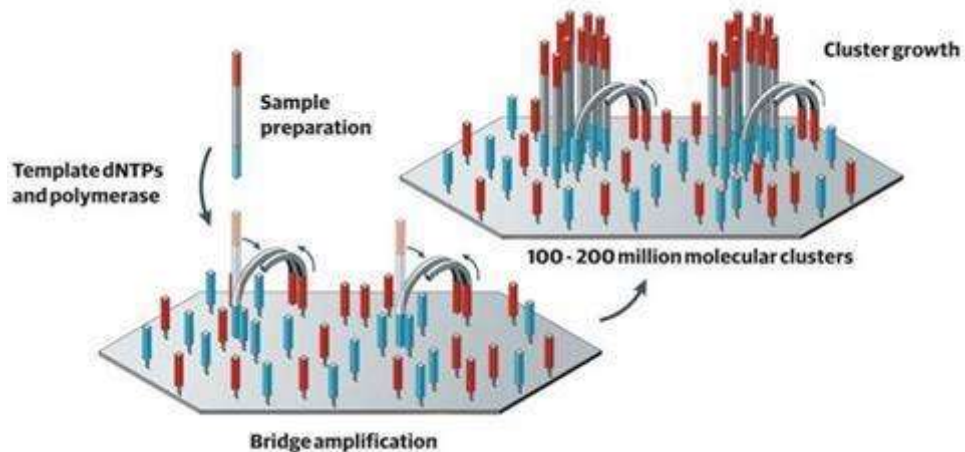


Abbildung 2 Brücken-Bindung der DNA Fragmente auf der Oberfläche der Durchflusszelle („Flow Cell“) und Vermehrung der Fragmente mittels „bridge-amplifikation“

Von einem auf dieser Durchflusszelle gebundenen Startmolekül ausgehend werden durch einen PCR-ähnlichen Schritt namens „bridge-amplifikation“ sehr viele DNA-Cluster gebildet. Ein Cluster ist eine Gruppe von nebeneinander auf der Fließzelle befindlichen DNA-Sequenzabschnitten und enthält tausende DNA-Fragmente der gleichen Sequenz (Template-DNA). Die Cluster sollten ausreichend viele DNA-Fragmente enthalten und bei der anschließenden eigentlichen Sequenzierung ausreichend Signale liefern. Bei der Sequenzierung werden die vier verschiedenen Nukleotide A, C, G, und T jeweils als einzelne Nukleotidlösung über die Glasplatte gespült und somit der Polymerase zur Strangsynthese angeboten. In jedem Sequenzierzyklus werden ein bis mehrere Nukleotide komplementär zu der Template-DNA eingebaut A bei T, T bei A, C bei G und G bei C.

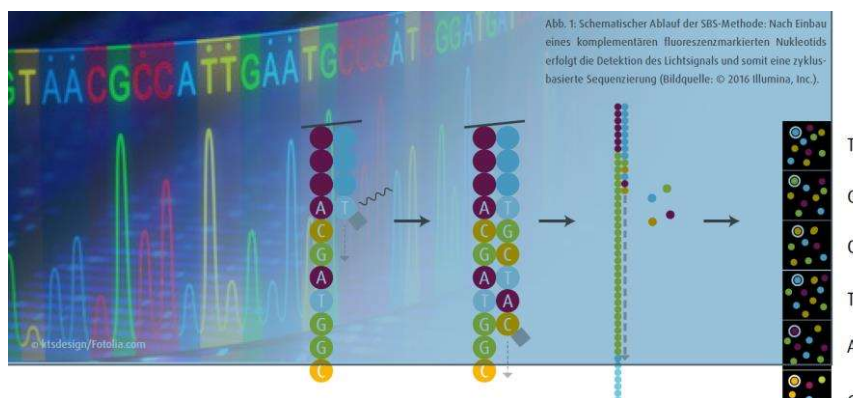


Abbildung 3 Einbau der Nukleotide und Generierung von Fluoreszenzsignalen



Während des Einbaus passender komplementärer Basen in den DNA-Strang werden Fluoreszenzsignale freigesetzt und photographisch detektiert. Werden keine passenden Nukleotide in die Template-DNA eingebaut, wird auch für dieses Cluster kein Fluoreszenzsignal freigesetzt. Es werden also während der Strangsynthese eine Abfolge an Bildern generiert – daher der Begriff „Sequencing By Synthesis“. Die DNA-Sequenzfolge setzt sich somit aus der Kenntnis der aufgespülten Nukleotidlösungen und der Reihenfolge der aufgespülten Flüssigkeiten und den dazu aufgenommenen Bildern pro Cluster zusammen.

Die übrigen nicht eingebauten Nukleotide werden durch darauffolgende Waschschriffe entfernt.

**MiSeq der Firma Illumina®:** Die Durchsatzrate dieses Gerätes liegt je nach verwendetem Reagenzienkit und Probenanzahl bei etwa 4 bis 15 Gigabasen (das entspricht einem Output von 4–15 x 10<sup>9</sup> einzelnen Basen). Während einer Sequenzierreaktion im Gerät können bis zu 50 Millionen einzelne Cluster analysiert werden. Nach der Sequenzierung der SARS-CoV-2 Genompanel erfolgt eine umfassende bioinformatische Auswertung der erhaltenen großen Menge an Sequenz-Daten sowie die Erstellung einer Consensus-Sequenz, welche im FASTA-Format in die RKI DESH-Datenbank eingespielt wird.

### **Auswertung (Bioinformatik)**

Direkt nach dem Ende der Sequenzierung werden die Daten der verschiedenen Proben, die simultan in einem Lauf analysiert wurden, wieder getrennt (Demultiplexing). Dadurch werden sogenannte „fastq“-Dateien generiert. Sie setzen sich aus den detektierten Sequenzen und der Qualität (Zuordnungssicherheit) zusammen. Im nächsten Schritt werden die vielen Millionen Sequenzabschnitte („Reads“) auf verschiedene Weise gefiltert, um die nicht so hochwertigen Reads zu entfernen und aus der weiteren Datenanalyse auszuschließen. Auch müssen noch die in der Sequenz vorhandene Adapter- und Primer-Sequenzen entfernt werden. Erst jetzt kann jeder Read mit einem festgelegten Referenzgenom (vom RKI vorgegeben) verglichen werden. Nur bei einer ausreichenden Übereinstimmung werden diese Reads für die weitere Analyse herangezogen. Weiterhin muss nun überprüft werden, an welche Positionen des Vergleichsgenoms diese Reads passen und es werden statistische Kennwerte erzeugt, um zu sehen, ob es Abweichungen zu dem Genom des Referenz-Corona-Virus gibt und ob diese statistisch signifikant sind. Aus diesen Informationen wird im letzten Schritt der bioinformatischen Analyse durch einen komplexen mathematischen Algorithmus die Übereinstimmungssequenz („Consensus-Sequenz“) mit dem Referenzgenom erstellt. Sie repräsentiert am Ende der Analyse die dominierende RNA-Sequenz in der vorliegenden Probe.

### **Analysedauer:**

Die Bearbeitungszeit des Untersuchungsauftrages beträgt i.d.R. 7-10 Tage.