

Zastosowanie sekwencjonowania NGS w technologii Illumina jako metody badania zmienności genetycznej wirusa SARS-CoV-2

The use of NGS sequencing in Illumina technology as a method for examining the genetic variability of the SARS-CoV-2 virus

Elżbieta Bartkowiak¹ (ORCID: 0009-0003-0611-683X), Beata Rozwadowska¹ (ORCID: 0000-0002-2288-114X), Marta Albertyńska¹ (ORCID: 0000-0003-4729-9823), Adrian Miara^{1,2} (ORCID: 0000-0001-6606-8444), Grzegorz Hudzik¹ (ORCID: 0009-0008-0816-0291)

¹Interdyscyplinarna Pracownia Diagnostyki Molekularnej, Wojewódzka Stacja Sanitarno-Epidemiologiczna w Katowicach, Polska

²Wyższa Szkoła Inżynierii i Zdrowia w Warszawie, Polska

Streszczenie

Sekwencjonowanie nowej generacji NGS (ang. *Next-Generation Sequencing*) to zaawansowana technologia, która umożliwia szybkie i precyzyjne analizy sekwencji DNA i RNA. Zastosowanie platformy NGS Illumina MiniSeq (Illumina, USA) do sekwencjonowania genomu koronawirusa SARS-CoV-2 pozwala na dokładną identyfikację wirusa, analizę wariantów genetycznych oraz monitorowanie jego ewolucji. To cenne narzędzie wspierające badania epidemiologiczne, opracowywanie skutecznych strategii zwalczania wirusa i lepsze zrozumienie choroby COVID-19. Nadzór molekularny nad zmiennością genetyczną wirusa SARS-CoV-2 z wykorzystaniem tej technologii prowadzony jest w Interdyscyplinarnej Pracowni Diagnostyki Molekularnej w Wojewódzkiej Stacji Sanitarno-Epidemiologicznej w Katowicach od września 2021 roku. Sekwencjonowanie NGS umożliwia identyfikację oraz śledzenie nowych wariantów i subwariantów wirusa SARS-CoV-2 na terenie województwa śląskiego, małopolskiego i opolskiego. W trakcie badań w laboratorium WSSE w Katowicach udokumentowane zostały warianty SARS-CoV-2 o różnym poziomie zakaźności i konsekwencjach dla zdrowia publicznego, tj. Alpha (B.1.1.7), Gamma (P.1), Delta (B.1.617.2) czy Omicron (B.1.1.529).

Słowa kluczowe: COVID-19, Illumina, NGS, SARS-CoV-2

Abstract

Next-Generation Sequencing (NGS) is an advanced technology that enables quick and precise analyzes of DNA and RNA sequences. The use of the Illumina MiniSeq NGS platform (Illumina, USA) for sequencing the SARS-CoV-2 coronavirus genome allows for accurate identification of the virus, analysis

Received: 13.09.2024

Accepted: 28.01.2025

Published: 01.06.2025

DOI: 10.5604/01.3001.0054.9880

Corresponding author:

dr n. med. Beata Rozwadowska, Interdyscyplinarna Pracownia Diagnostyki Molekularnej, Wojewódzka Stacja Sanitarno-Epidemiologiczna, Katowice 40-074, ul. Raciborska 39, tel. +48 32 351 23 00, wew. 222, e-mail: rozwadowska.beata@sanepid.gov.pl

Cite the article as:

Bartkowiak E, Rozwadowska B, Albertyńska M, Miara A, Hudzik G. The use of NGS sequencing in Illumina technology as a method for examining the genetic variability of the SARS-CoV-2 virus. *Diagn Lab.* 2025; 61(2): 101–112



Open access

The content of the journal is available in Open Access formula which means free access to scientific data for researchers and readers.



Copyright

This material is available under the Creative Commons – Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). The full terms of this license are available on: <https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Licence

Some right reserved: Polish Society of Laboratory Diagnostics. Published by Index Copernicus Sp. z o.o.

Conflict of interests

The authors declare that they have no competing interests.

of its genetic variants and monitoring of its evolution. It is a valuable tool supporting epidemiological research, developing effective strategies to combat the virus and better understanding the COVID-19 disease it causes. Molecular surveillance of the genetic variability of the SARS-CoV-2 virus using this technology has been carried out at the Interdisciplinary Molecular Research Laboratory at the Provincial Sanitary and Epidemiological Station in Katowice since September 2021. NGS sequencing is used to identify and track new variants and subvariants of the SARS-CoV-2 virus in the Silesian, Lesser Poland and Opole Voivodeships. During research at the WSSE laboratory in Katowice, many SARS-CoV-2 variants with different levels of infectivity and consequences for public health were documented, i.e. Alpha (B.1.1.7), Gamma (P.1), Delta (B.1.617.2) or Omicron (B.1.1.529).

Keywords: COVID-19, Illumina, NGS, SARS-CoV-2

WSTĘP

Ludzki koronawirus SARS-CoV-2 (ang. *severe acute respiratory syndrome coronavirus 2*), wywołujący zespół ostrej niewydolności oddechowej, jest odpowiedzialny za pandemię choroby COVID-19 (ang. *coronavirus disease 2019*) i w znaczący sposób wpłynął na globalną gospodarkę oraz zdrowie publiczne na całym świecie [1]. Światowa Organizacja Zdrowia (WHO; ang. *World Health Organization*) ogłosiła 11 marca 2020 roku stan pandemii wywołanej koronawirusem SARS-CoV-2 [2, 3]. Choroba miała niszczący wpływ na populację świata i do połowy 2024 roku dotknęła prawie 780 milionów ludzi, powodując ponad 7 milionów zgonów na całym świecie [4]. Pandemia COVID-19 stała się największym globalnym kryzysem zdrowotnym od czasu pandemii grypy w 1918 roku. Odkąd wirus został po raz pierwszy zidentyfikowany w grudniu 2019 roku w chińskim Wuhan, naukowcy z całego świata prowadzą badania nad jego budową, mutacjami i wariantami molekularnymi, aby zrozumieć biologię wirusa i opracować skuteczne strategie zwalczania COVID-19 [1, 5-8].

SARS-CoV-2 należy do rodziny koronawirusów *Coronaviridae*, które charakteryzują się obecnością dużej wirionowej cząsteczki otoczonej osłonką. Podobnie jak inne beta-koronawirusy, tj. SARS-CoV i MERS-CoV, cechuje się dużym potencjałem epidemicznym [9, 10]. Genom SARS-CoV-2 składa się z jednonicowego RNA (ssRNA; ang. *single strand RNA*) o długości prawie 30 tys. zasad (29,9 Kb) [3, 6, 7]. W genomie zakodowanych jest około 29 białek, w tym kluczowe białko kolca „S” (ang. *spike protein*) i inne strukturalne białka wirusowe oraz białka regulujące replikację, transkrypcję,

a także białka odpowiedzialne za interakcje wirusa z komórkami gospodarza [8, 11]. Jako wirus RNA, SARS-CoV-2 nieustannie zmienia się poprzez mutacje, które mogą wpływać na jego właściwości biologiczne, w tym zakaźność, zdolność do unikania nadzoru układu odpornościowego gospodarza czy skuteczność terapii. W miarę jak wirus rozprzestrzenia się globalnie, pojawiają się różne warianty i subwarianty, z których niektóre są bardziej zaraźliwe i groźne niż pierwotny szczep [12]. Obecnie udokumentowano na świecie wiele wariantów SARS-CoV-2, z których część ma istotne znaczenie epidemiologiczne, tj. Alpha (B.1.1.7), Delta (B.1.617.2), Gamma (P.1) czy dominujący aktualnie Omicron (B.1.1.529) [12-15].

W związku z pojawianiem się nowych wariantów wirusa SARS-CoV-2 istotne stało się poznanie poszczególnych zmian w genomie, co jest możliwe dzięki jego zsekwencjonowaniu.

Początki technik umożliwiających poznanie sekwencji kwasów nukleinowych sięgają 1975 r., kiedy Sanger i wsp. opracowali metodę sekwencjonowania, która po licznych modyfikacjach (np. metoda dideoksy) jest wykorzystywana do dziś. Metoda Sangera określana jest jako sekwencjonowanie I generacji. Ze względu na niską przepustowość i brak możliwości pełnej automatyzacji w latach 90. XX w. zaczęto poszukiwać bardziej nowoczesnych metod, które pozwoliłyby na zwiększenie liczby analizowanych matryc oraz długości odczytywanych sekwencji, ich pulowanie i – tym samym – optymalizację kosztów prowadzonych badań. To dało początek metodom sekwencjonowania II generacji, m.in. pirosekwencjonowaniu czy sekwencjonowaniu

przez ligację, które pozwoliły na zwiększenie precyzji odczytu i przepustowości, jednakże często wymagały skomplikowanych systemów detekcji. Z tego powodu zostały później wyparte prawie całkowicie przez technologie firmy Illumina, które oparte są na sekwencjonowaniu przez syntezę na płytce przepływowej. Kolejną generacją sekwencjonowania były techniki, które umożliwiły prowadzenie odczytu długich (od kilku tysięcy do ponad stu tysięcy nukleotydów), pojedynczych cząsteczek kwasów nukleinowych w czasie rzeczywistym, takie jak sekwencjonowanie nanoporowe (Oxford Nanopore Technologies) czy sekwencjonowanie pojedynczej cząsteczki w czasie rzeczywistym (PacBio).

Jedną z najnowocześniejszych i najczęściej stosowanych do tego celu technik jest sekwencjonowanie nowej generacji NGS w technologii Illumina. Umożliwia ona jednocześnie sekwencjonowanie wielu tysięcy, a nawet milionów fragmentów DNA lub RNA w trakcie pojedynczego doświadczenia, w przeciwieństwie do metody Sangera opartej na sekwencjonowaniu tylko pojedynczych fragmentów. Dzięki technologii NGS można znacznie szybciej otrzymać nieporównywalnie większe ilości danych niezbędnych do złożenia pełnej sekwencji genomu, aniżeli w technice terminacji łańcucha. Jest to bardzo korzystne z uwagi na obniżenie kosztów, ponieważ wygenerowanie odpowiedniej ilości danych przez sekwenator NGS Illumina odbywa się w trakcie jednej reakcji sekwencjonowania [14, 16, 17].

Wszystkie metody sekwencjonowania NGS wymagają przygotowania biblioteki, będącej zestawem odpowiednio przygotowanych fragmentów DNA [18, 19]. Wyniki sekwencjonowania są generowane równocześnie dla wielu fragmentów i następnie składane w pełne sekwencje, dzięki czemu sekwencjonowanie dużych ilości danych genetycznych jest szybkie i efektywne.

Fragmenty w bibliotece mają zwykle długość w przedziale 100 – 500 bp, a ich dokładna wielkość zależy od długości pojedynczej sekwencji, którą można uzyskać za pomocą stosowanej metody sekwencjonowania nowej generacji [20]. Bardzo istotna jest również losowa fragmentacja DNA, gdyż każdy fragment biblioteki sekwencjonowany jest od dwóch końców – tzw. *paired-end sequencing*. Odczyty z obu stron mogą, ale nie muszą, nakładać się na siebie, natomiast wszelkie odchylenia w odczytach mogą świadczyć

o pojawiających się zmianach strukturalnych w badanym genomie [16, 21].

Do unieruchamiania fragmentów bibliotek w technologii NGS Illumina wykorzystuje się stałe podłoże wykonane ze szkła, pokrytego wieloma kopiami krótkich oligonukleotydów, do których przyłączają się komplementarne końce 5' i 7' cDNA (ang. *complementary DNA*) z biblioteki. Jest to tzw. płytka *flow cell*. Końce fragmentów biblioteki są ligowane z adaptorami, czyli krótkimi dwuniciowymi kawałkami DNA, których sekwencje są komplementarne do oligonukleotydów związanych z płytką *flow cell*. Zawierają one przyłączone indeksy 1 i 2, unikalne dla każdej próbki. Po przeprowadzeniu denaturacji biblioteki otrzymuje się jednoniciowe cząsteczki, przyłączające się do powierzchni płytki *flow cell* dzięki tworzeniu par zasad między sekwencjami ich adaptorów a unieruchomionych nukleotydów [17, 22]. W trakcie reakcji sekwencjonowania następuje amplifikacja unieruchomionych fragmentów DNA poprzez przeprowadzenie łańcuchowej reakcji polimerazy PCR (ang. *polimerase chain reaction*) i dzięki temu można otrzymać wystarczającą liczbę identycznych kopii matrycy do sekwencjonowania. Jest to proces nazywany amplifikacją mostkową, w wyniku czego powstają odrębne i unieruchomione klastry identycznych, monoklonalnych fragmentów DNA, zarówno w orientacji *forward*, jak i *reverse* [10, 23]. Następnie klastry są modyfikowane poprzez proces przemysławiania, który w początkowym etapie sekwencjonowania pozostawia tylko fragmenty w orientacji *forward*. Po odczytaniu sekwencji *forward* proces jest odwracany i następuje odczyt sekwencji w orientacji *reverse*. Platforma MiniSeq Illumina wykorzystuje płytki *flow cell*, na których powstaje losowa wielkość klastrów (ang. *random cluster*) o nieokreślonych kształtach, nieregularnych odstępach i rozmiarach o średnicy od 1 do 2 mikrometrów, dlatego też istotnym czynnikiem optymalizacji jakości danych i wydajności sekwencjonowania jest monitorowanie gęstości klastrów [23, 24].

W sekwenatorach Illumina zastosowano technologię krótkich odczytów sekwencji, bazującą na sygnałach fluorescencyjnych oraz na metodzie sekwencjonowania przez syntezę – SBS (ang. *sequencing by synthesis*) [18, 25]. Sekwencjonowanie przez syntezę opiera się na metodzie Sangera i charakteryzuje się wysoką dokładnością, generując w tym samym czasie wielokrotnie

więcej danych wyjściowych, co czyni je znacznie szybszym i łatwiejszym w wykonaniu. Każdy z amplifikowanych fragmentów w klastrze jest przedłużany poprzez dodawanie kolejnych zmodyfikowanych nukleotydów podczas reakcji syntezy. Są to nukleotydy posiadające znaczniki fluorescencyjne w różnych kolorach, co umożliwia ich identyfikację. Podczas każdego cyklu syntezy naci tylko jeden nukleotyd jest dodawany, a następnie reakcja jest wstrzymywana. W tym czasie układ optyczny w sekwenatorze rejestruje emisję sygnału fluorescencyjnego z całej płytki *flow cell*, identyfikując w ten sposób przyłączane nukleotydy. Po zarejestrowaniu sygnałów fluorescencyjnych klastry są przygotowywane do kolejnej rundy poprzez usunięcie znaczników fluorescencyjnych. Proces ten nazywany jest sekwencjonowaniem metodą odwracalnej terminacji, ponieważ jest wielokrotnie powtarzany, aż do momentu uzyskania pełnej sekwencji DNA. Oprogramowanie w sekwenatorze interpretuje uzyskane dane sekwencyjne na podstawie intensywności sygnałów fluorescencyjnych. Metoda ta umożliwia równoczesną analizę wielu klastrów, co przyspiesza proces sekwencjonowania i czyni go bardziej efektywnym [26]. W aparacie MiniSeq Illumina zastosowano 2-kanalowe obrazowanie, dzięki czemu tylko dwa barwniki fluorescencyjne i dwa obrazy wykorzystywane są do określenia włączenia wszystkich czterech nukleotydów w każdym cyklu sekwencjonowania. Umożliwia to szybsze sekwencjonowanie i wydajniejsze przetwarzanie danych [26-28]. Urządzenie MiniSeq Illumina może efektywnie analizować dane każdego klastra jednocześnie i odczytywać miliony stosunkowo krótkich sekwencji o wielkości 300 bp, co sprawia, że łącznie podczas sekwencjonowania można uzyskać 25 mln odczytów oraz 7,5 Gb danych sekwencyjnych [23].

Podsumowując, zasadniczymi kryteriami przy wyborze technologii sekwencjonowania NGS są szybkość i przepustowość oraz precyzyjne i kosztowo efektywne sekwencjonowanie dużych ilości DNA i RNA. Istotne znaczenie ma także poszerzanie zastosowań i dostępności. To właśnie dzięki temu sekwencjonowanie NGS w technologii Illumina jest najczęściej wybieraną metodą NGS w badaniach w dziedzinie biologii molekularnej, medycyny oraz nauk związanych z genomiką. Jest również metodą coraz powszechniej stosowaną w badaniach epidemiologicznych, która umożliwia identyfikację i monitorowanie rozprzestrzeniania się

chorobotwórczych wirusów, bakterii, grzybów i pasożytów [29, 30].

W ramach programu „Molekularnego nadzoru epidemiologicznego nad COVID-19” we wrześniu 2021 r., w Interdyscyplinarnej Pracowni Diagnostyki Molekularnej w Wojewódzkiej Stacji Sanitarno-Epidemiologicznej w Katowicach wdrożono sekwencjonowanie nowej generacji NGS w technologii Illumina. Badania trwają do dziś i obejmują teren województwa śląskiego, małopolskiego oraz opolskiego. Próbkę kliniczne od pacjentów z dodatnim wynikiem badania RT-PCR, spełniające kryteria epidemiologiczne oraz laboratoryjne, trafiają do laboratorium zgodnie z przyjętym harmonogramem. Dane sekwencyjne uzyskane z platformy Illumina są przekazywane do bazy GISAID (ang. *Global Initiative on Sharing All Influenza Data*), w której każdej sekwencji przypisywany jest unikalny kod EPI_ID. GISAID jest platformą gromadzącą informacje o genomach wirusów grypy, SARS-CoV-2, małpiej ospy czy arbowirusów, co umożliwia naukowcom monitorowanie mutacji i ich ewolucji, a także wspiera badania epidemiologiczne. Pierwszą kompletną sekwencję genomu wirusa SARS-CoV-2 zdeponowano w bazie GISAID 10 stycznia 2020 r. [10, 31, 32].

Sekwencjonowanie genomu wirusa SARS-CoV-2 przeprowadzane jest zgodnie z protokołem Illumina COVIDSeq (Illumina, USA), który został zatwierdzony przez Amerykańską Agencję ds. Żywności i Leków (FDA; ang. *Food and Drug Administration*) do użytku klinicznego. Jest to zestaw oparty na amplikonach, zawierający startery 2019-nCoV, przeznaczony do wykrywania mutacji i charakterystyki RNA wirusa SARS-CoV-2. Zestaw przeznaczony jest dla laboratoriów prowadzących badania w zakresie identyfikacji oraz kierunków transmisji nowych wariantów SARS-CoV-2 [33, 34]. W zestawie COVIDSeq wykorzystano zmodyfikowaną wersję powszechnie dostępnego protokołu ARTIC multiplex PCR. Protokół umożliwia w sekwenatorze MiniSeq Illumina skalowanie liczby próbek do 96 w jednej reakcji sekwencjonowania i uzyskanie ostatecznych oraz potwierdzonych wyników w trakcie 3 dni roboczych [16, 35, 36].

PRZEBIEG PROCESU SEKWENCJONOWANIA SARS-CoV-2

Cały proces sekwencjonowania genomu wirusa SARS-CoV-2 w technologii Illumina z wykorzystaniem

protokołu laboratoryjnego COVIDSeq można podzielić na 4 główne etapy, które obejmują kolejno:

1. Przygotowanie próbek do sekwencjonowania poprzez ekstrakcję materiału genetycznego RNA SARS-CoV-2.
2. Przygotowanie biblioteki, obejmujące obróbkę RNA wraz z przeprowadzeniem reakcji odwrotnej transkrypcji, a następnie wzbogacanie cDNA w badane sekwencje oraz przyłączanie specjalnych adaptorów z indeksami do obu końców, tak aby mogło zajść sekwencjonowanie (amplifikacja i tagmentacja amplikonów). Następnie przeprowadza się pulowanie i oczyszczanie biblioteki, która finalnie poddawana jest kontroli jakości uzyskanego materiału, obejmującej pomiar jej stężenia i ocenę długości fragmentów biblioteki. W końcowym etapie wykonywane jest precyzyjne rozcieńczanie i denaturacja biblioteki sekwencjonowania oraz biblioteki kontrolnej PhiX.
3. Sekwencjonowanie w aparacie MiniSeq Illumina, które jest etapem całkowicie zautomatyzowanym i polega na załadowaniu do sekwenatora płytki *flow cell* oraz kasety reakcyjnej z odczynnikami, w której umieszcza się przygotowaną bibliotekę. Następnie uruchamiany jest proces sekwencjonowania przez syntezę, rozpoczynający się od generowania klastrów, aż po dwukrotne odczyty indeksów i sekwencji odcinków o długości 2 x 150 bp, kończący się na detekcji sygnałów fluorescencyjnych z każdego klastra.
4. Bioinformatyczna analiza danych, która odbywa się po zakończonym procesie sekwencjonowania, kiedy dane są przetwarzane w celu odczytu sekwencji nukleotydów w każdym klastrze. Proces ten polega na ocenie parametrów sekwencjonowania, interpretacji wyników i analizie wariantów przy wykorzystaniu odpowiednich do tego baz danych oraz specjalistycznego oprogramowania bioinformatycznego.

PRZYGOTOWANIE PRÓBEK DO SEKWENCJONOWANIA

Materiałem poddanym sekwencjonowaniu jest RNA SARS-CoV-2, wyekstrahowany z materiału pobranego od osób chorych na COVID-19 (wymazy z gardła i nosa, wymazy z nosogardzieli, popłuczyny oskrzelowo-pęcherzykowe

lub aspiraty tchawicze). Ekstrakcja RNA może być przeprowadzona za pomocą zestawów manualnych lub z wykorzystaniem systemów automatycznych. W laboratorium WSSE w Katowicach wykorzystywane są systemy do izolacji automatycznej MagNA Pure 24 i MagNA Pure 96 (Roche, Niemcy) oraz MagEx STARlet (Hamilton, USA) oraz zestawy do ręcznej izolacji RNA, np. GeneProof PathogenFree RNA Isolation Kit (GeneProof, Czechy). W postępowaniu z próbkami istotne jest zachowanie czystości próbki, unikanie kontaminacji oraz przechowywanie izolatów w odpowiednich warunkach (temperatura -70°C), aby uniknąć degradacji materiału genetycznego. Ponadto wartość C_t próbki zakwalifikowanej do sekwencjonowania nie powinna być wyższa niż 30 [1].

PRZYGOTOWANIE BIBLIOTEKI SEKWENCJONOWANIA NGS

Illumina opracowała wieloetapowy protokół COVIDSeq, służący do przygotowania biblioteki NGS, dzięki któremu otrzymuje się finalnie 50 μl spulowanej i oczyszczonej biblioteki, zawierającej fragmenty DNA z wszystkich próbek poddanych sekwencjonowaniu [34, 35].

Biblioteka sekwencjonowania NGS to zbiór fragmentów DNA, które mają odpowiednią długość oraz oznakowanie umożliwiające ich identyfikację i przypisanie do właściwej próbki. Do skonstruowania biblioteki DNA wykorzystuje się zestaw odczynników Illumina COVID-Seq Assay, zawierający startery 2019-nCoV do wykrywania sekwencji charakterystycznych dla RNA wirusa, które umożliwiają równoległe sekwencjonowanie do 96 próbek w jednej reakcji. Protokół NGS z wykorzystaniem zestawu Illumina COVIDSeq Assay składa się z niżej wymienionych etapów [37, 38]:

1. Annealing RNA

W przypadku bibliotek RNA, w których sekwencje są często zanieczyszczone lub trudne do odczytania, w początkowym etapie przeprowadza się annealing RNA. Polega on na wstępnym oczyszczeniu i oznaczeniu RNA za pomocą znaczników, takich jak oligo-dT, które wiążą się z wybranymi regionami RNA.

2. Odwrotna transkrypcja

Kolejnym etapem podczas przygotowywania biblioteki jest synteza komplementarnych cDNA na matrycy cząsteczek RNA. Ten proces to odwrotna transkrypcja RT

(ang. *reverse transcription*), czyli konwersja RNA do cDNA z wykorzystaniem enzymu odwrotnej transkryptazy. Reakcja przeprowadzana jest po dodaniu do próbek odpowiedniej ilości enzymu oraz mieszaniny reakcyjnej FSM (ang. *First Strand Mix*), zawierającej specyficzne startery (oligonukleotydy) wiążące się z regionem docelowym genomu wirusa SARS-CoV-2 oraz nukleotydy budujące nowy łańcuch DNA.

3. Amplifikacja cDNA

Amplifikacja cDNA polega na wielokrotnym powielaniu fragmentów DNA w celu uzyskania większej ilości materiału genetycznego wirusa do analizy. Wykorzystuje się do tego łańcuchową reakcję polimerazy – PCR – która umożliwia powielenie określonych fragmentów DNA w sposób selektywny i kontrolowany. Ten krok zgodnie z protokołem COVIDSeq wymaga użycia dwóch nowych 96-dołkowych płytek PCR oraz mieszaniny IPM (ang. *Illumina PCR Mix*), a także specyficznych starterów CPP1 i CPP2 (ang. *COVIDSeq Primer Pool 1 i 2*), komplementarnych do sekwencji docelowych w genomie wirusa SARS-CoV-2. Jest to najdłuższy etap przygotowania biblioteki, trwający w termocyklerze około 3 godzin i obejmujący 35 cykli PCR.

4. Tagmentacja amplikonów PCR

Po amplifikacji cDNA otrzymuje się wiele kopii fragmentów DNA, pochodzących z genomu wirusa SARS-CoV-2. Te fragmenty to amplikony PCR, zawierające sekwencje docelowe, które mają być zsekwencjonowane. Dla efektywnego sekwencjonowania na platformie Illumina MiniSeq konieczne jest przygotowanie odpowiednio krótkich fragmentów, które są niezbędne w technologii sekwencjonowania przez syntezę. Tagmentacja polega na jednoczesnej fragmentacji i przyłączaniu sekwencji adaptorowych przy użyciu kulek magnetycznych opłaszczonych transpozomami (EBLTS; ang. *Enrichment BLT*). Jest kontrolowana i optymalizowana pod kątem długości fragmentów i ilości DNA. W przypadku wirusa SARS-CoV-2 należy uzyskać równomierne i odpowiednio długie tagmenty o długości 300 – 400 bp [38].

5. Oczyszczanie po tagmentacji

Tagmenty są stabilizowane i oczyszczane, aby usunąć niepożądane produkty reakcji, inhibitory enzymów i inne zanieczyszczenia. Możliwe jest to dzięki właściwościom magnetycznym kulek EBLTS przy użyciu separatora magnetycznego.

6. Amplifikacja tagmentowanych amplikonów

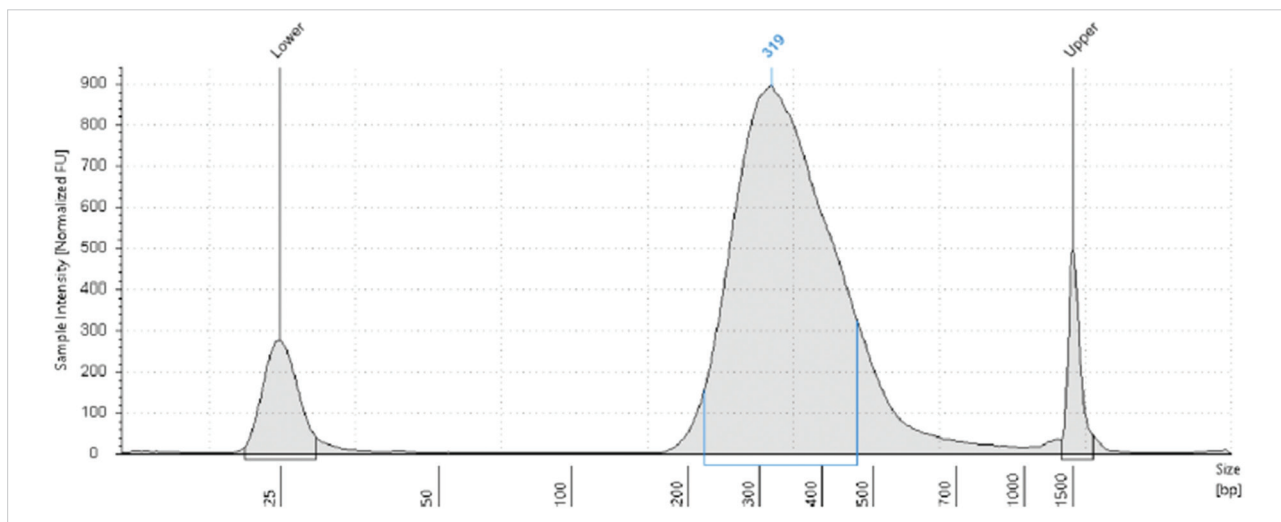
Ztagmentowane fragmenty DNA w kolejnych etapach są amplifikowane, oczyszczane i indeksowane z wykorzystaniem zestawów indeksów IDT for Illumina Nextera UD Indexes Set A, B, C, D (384 indexes, 384 samples) [39]. Zawierają one różne unikalne sekwencje DNA, o długości 10 nukleotydów, które są przypisywane każdej próbce wybranej do sekwencjonowania. Indeksowanie próbek w trakcie przygotowania biblioteki NGS umożliwia multipleksowanie, czyli jednoczesne sekwencjonowanie wielu próbek w jednym przebiegu sekwencjonowania.

7. Pulowanie i oczyszczanie biblioteki

Po zakończeniu procesu amplifikacji, na separatorze magnetycznym kulki EBLTS przyłączają się do magnesu, a zamplifikowana biblioteka znajduje się w supernatancie. W celu spulowania biblioteki należy przenieść po 5 µl biblioteki z każdego z 96 dołków płytki do nowej, wspólnej probówki. Następnie biblioteka oczyszczana jest przy użyciu kulek magnetycznych ITB (ang. *Illumina Tune Beads*) poprzez usuwanie niewłaściwych fragmentów i zanieczyszczeń, co zwiększa jakość biblioteki oraz umożliwia uzyskanie dokładniejszych i bardziej wiarygodnych wyników sekwencjonowania. Kulki te łączą się z fragmentami DNA, które są odpowiedniej wielkości lub które spełniają określone kryteria jakościowe. Jest to kluczowy etap przygotowania biblioteki do sekwencjonowania NGS. Dwukrotne płukanie kulek przeprowadzane jest przy użyciu świeżo przygotowanego 80% etanolu. Następnie należy bardzo dokładnie usunąć alkohol i odczekać do całkowitego odparowania. Po zawieszeniu biblioteki w buforze RSB (ang. *Resuspension Buffer*) należy usunąć kulki ITB, wykorzystując separator magnetyczny i oczyszczoną bibliotekę przenieść do nowej probówki. Finalnie uzyskuje się 50 µl spulowanej i oczyszczonej biblioteki, złożonej z 96 próbek wybranych do sekwencjonowania.

8. Ilościowe określanie i normalizacja biblioteki NGS

Kolejnym istotnym kontrolnym krokiem podczas przygotowania biblioteki NGS jest ilościowa ocena jej jakości, aby upewnić się, że biblioteka została odpowiednio przygotowana do sekwencjonowania. W tym celu przeprowadza się pomiar stężenia DNA oraz pomiar długości fragmentów DNA, co jest niezbędne do określenia wyjściowego poziomu molarności spulowanej biblioteki.



Rycina 1. Raport z pomiaru długości otrzymanych fragmentów DNA techniką elektroforezy kapilarnej przy użyciu aparatu TapeStation w przykładowej bibliotece NGS dla próbek SARS-CoV-2.

Pomiar stężenia biblioteki NGS ma na celu określenie ilości materiału genetycznego. Ilościowa ocena DNA przeprowadzana jest przy użyciu fluorymetru Quantus (Promega Corporation, USA) i zestawu odczynników QuantiFluor ONE dsDNA System, zawierających barwnik fluorescencyjny. Fluorymetryczny pomiar stężenia DNA charakteryzuje się wysoką czułością i specyficznością oraz szerokim zakresem liniowości. Wyniki odczytu podawane są w [ng/μl]. Jeżeli uzyskane wartości znajdują się poza zakresem liniowości urządzenia, należy rozcieńczyć bibliotekę w stosunku 1:10 i powtórzyć pomiar. Pomiar długości uzyskanych fragmentów w bibliotece NGS Illumina odbywa się przy użyciu aparatu TapeStation 4150 (Agilent Technologies, USA), który umożliwia automatyczną elektroforezę kapilarną DNA i RNA. W tym celu wykorzystywane są odpowiednie taśmy High Sensitivity D1000 ScreenTape, zawierające 16 oddzielnych kanałów z żelami poliakrylamidowymi, które zapewniają rozdzielanie fragmentów biblioteki oraz odczynniki High Sensitivity D1000 Reagents w postaci buforu i drabinki, umożliwiające analizę długości cząsteczek DNA o wielkości 35 – 1000 bp. Elektroforeza jednej próbki trwa około jednej minuty. Gotowe zestawy odczynników oraz pełna automatyzacja elektroforezy zapewniają łatwość obsługi i wysoką powtarzalność wyników. Najbardziej optymalna długość fragmentów DNA w otrzymanej bibliotece, w przypadku protokołu Illumina COVIDSeq przeznaczonego dla aparatu MiniSeq, powinna mieścić się w przedziale 300 – 400 bp (ryc. 1).

9. Denaturacja i normalizacja biblioteki NGS oraz biblioteki PhiX

Końcowy etap przygotowania biblioteki NGS ma na celu doprowadzenie biblioteki do odpowiedniego stężenia molowego, wynoszącego 1,2 pM zgodnie z wytycznymi zawartymi w protokole producenta MiniSeq System Denature and Dilute Libraries Guide. Dysponując wynikami pomiarów stężenia oraz długości fragmentów DNA procesowanej biblioteki, należy obliczyć molarność zgodnie ze wzorem:

$$\text{Molarność [nM]} = \frac{\text{stężenie biblioteki} \left[\frac{\text{ng}}{\mu\text{l}} \right]}{660 \left[\frac{\text{g}}{\text{mol}} \right] \times \text{długość biblioteki [bp]}} \times 10^6$$

W zależności od uzyskanej molarności bibliotekę należy rozcieńczyć i doprowadzić stopniowo do stężenia 1 nM. Do normalizacji i rozcieńczeń biblioteki wykorzystuje się czystą wodę przeznaczoną do PCR, 0,1 N NaOH oraz 200 mM Tris-HCl o pH = 7,0. NaOH stosowany do denaturacji próbki powinien być świeżo rozcieńczony z wyjściowego stężenia 1 N do 0,1 N, do uzyskania optymalnej wartości pH > 12,5. Spadek pH osłabia denaturację DNA i tym samym powoduje niższą gęstość klastrów w trakcie procesu sekwencjonowania [24]. W kolejnym kroku, do zdenaturowanej biblioteki, należy dodać schłodzony bufor HT1 (ang. *Hybridization Buffer*), uzyskując 1 ml rozcieńczonej biblioteki o stężeniu 5 pM. Po dalszym rozcieńczeniu buforem HT1 otrzymuje się 1,2 pM bibliotekę NGS przygotowaną do przeprowadzenia właściwego procesu sekwencjonowania.

Oprócz przygotowania biblioteki NGS należy również zdenaturować i znormalizować bibliotekę kontrolną PhiX. Biblioteka PhiX Control v3 (powszechnie określana jako PhiX) pochodzi z małego, dobrze scharakteryzowanego genomu bakteriofaga PhiX. Jest to skoncentrowana biblioteka Illumina (10 nM w 10 µl), która charakteryzuje się średnią długością fragmentów DNA 500 bp i zrównoważonym składem oraz proporcją zasad ~45% GC do ~55% AT. Bibliotekę kontrolną PhiX rozcieńcza się i denaturuje analogicznie jak w przypadku biblioteki NGS [41, 42].

W końcowym etapie do biblioteki NGS o stężeniu 1,2 pM należy dodać 2% przygotowanej biblioteki kontrolnej PhiX, również o stężeniu 1,2 pM. Tak przygotowaną bibliotekę NGS w ilości 500 µl umieszcza się w specjalnym zbiorniku w kasie odczynnikowej MiniSeq High/Mid Output Reagent Cartridge [40].

SEKWENCJONOWANIE NGS
W APARACIE MINISEQ ILLUMINA

Parametry przebiegu sekwencjonowania, informacje o próbkach i indeksach są wprowadzane do oprogramowania Local Run Manager, dostępnego w aparacie MiniSeq Illumina. Należy wybrać długość i rodzaj odczytów w trakcie sekwencjonowania – w tym przypadku sekwencjonowanie obu końców 2 x 151 bp. Oznacza to, że pojedyncze odczyty w każdej rundzie sekwencjonowania mają długość 151 bp i rozpoczynają się od dwóch końców cząsteczki – najpierw w orientacji *forward*, a następnie *reverse*. W związku z tym każdorazowo występuje także podwójny odczyt indeksów (ang. *Dual Index Reads*) – index1 i index2, po 10 cykli na każdej z nich. To sprawia, że w aparacie MiniSeq zachodzą łącznie 322 cykle w trakcie sekwencjonowania genomu SARS-CoV-2. Wszystkie te informacje zostają zapisane

na platformie BaseSpace Sequence Hub Illumina, która zapewnia odpowiednie miejsce do przechowywania dużych ilości generowanych danych NGS oraz ich analizę.

Postępując zgodnie z instrukcjami oprogramowania sterującego, w aparacie MiniSeq umieszcza się płytkę reakcyjną MiniSeq Flow Cell oraz kasetę odczynnikową MiniSeq w wersji High lub Mid Output Reagent Cartridge. W przypadku SARS-CoV-2 wersja kasety High przeznaczona jest do sekwencjonowania maksymalnej liczby 96 próbek w czasie około 24 godzin, natomiast wersja Mid dla 24 próbek w czasie 17 godzin. Zestawienie parametrów sekwencjonowania na platformie MiniSeq Illumina oraz specyfikacji dostępnych zestawów do sekwencjonowania przedstawiono w tabeli I. Rzeczywiste parametry wydajności mogą się różnić w zależności od typu próbki, jej jakości i wartości PF% (ang. *Passing Filter*) dla klastrów [18, 39, 43-45].

BIOINFORMATYCZNA ANALIZA
DANYCH NGS

Po zakończeniu sekwencjonowania uzyskuje się dane w postaci surowych odczytów (ang. *raw reads*), które zawierają sekwencje nukleotydów. W celu analizy tych danych konieczne jest przeprowadzenie operacji przetwarzania, która obejmuje początkowo ocenę parametrów sekwencjonowania NGS oraz jakości surowych danych. Kolejny krok to interpretacja wyników i analiza otrzymanych wariantów z wykorzystaniem baz danych i narzędzi bioinformatycznych, przeznaczonych dla wybranego eksperymentu. W przypadku sekwencjonowania NGS Illumina analizę bioinformatyczną przeprowadza się na platformie BaseSpace Sequence Hub.

Po przetworzeniu danych sekwencyjnych otrzymuje się informację na temat tzw. pokrycia sekwencji (ang.

Tabela I. Specyfikacja i zestawienie parametrów wydajności sekwencjonowania na platformie MiniSeq Illumina.

	MiniSeq System High-Output Kit			MiniSeq System Mid-Output Kit
Długość odczytu	2 × 150 bp	2 × 75 bp	1 × 75 bp	2 × 150 bp
Quality Scores (Q30)	Q30 > 80%	Q30 > 85%	Q30 > 85%	Q30 > 80%
Output (ilość danych NGS)	6,6 – 7,5 Gb	3,3 – 3,75 Gb	1,65 – 1,875 Gb	2,1 – 2,4 Gb
Czas całkowity*	~24 godziny	~13 godzin	~7 godzin	~17 godzin



sequence coverage), co oznacza liczbę odczytów nukleotydów, które pokrywają daną pozycję w genomie SARS-CoV-2. Poziom pokrycia jest analizowany w celu oceny wydajności sekwencjonowania, która wyrażona jest jako średnia liczba odczytów dla danego nukleotydu w odniesieniu do wartości docelowej. Wartość docelowa to minimalny poziom pokrycia, który jest akceptowalny w konkretnej analizie. Może to być różna wartość (np. 30x, 50x, 100x), w zależności od rodzaju badania. Niskie pokrycie może oznaczać, że niektóre obszary genomu nie zostały wystarczająco dobrze odczytane, co może prowadzić do błędów w analizach genetycznych. W badaniach związanych z identyfikacją wariantów genetycznych, takich jak mutacje czy polimorfizmy, wysoki poziom pokrycia jest istotny, ponieważ pozwala na zwiększenie prawdopodobieństwa identyfikacji tych zmian genetycznych [46, 47].

1. Ocena parametrów sekwencjonowania

Po zakończeniu sekwencjonowania należy dokonać oceny następujących parametrów reakcji: *Clusters Passing Filter* (PF%), *Clusters density* [K/mm²], tj. gęstość klastrów, *Quality Scores* (Q30%), czyli jakość składanej sekwencji oraz procentowy udział w sekwencjonowaniu biblioteki kontrolnej PhiX [24].

Clusters Passing Filter (PF%) w sekwencjonowaniu NGS Illumina odnosi się do procentowego udziału odczytów sekwencji, które przeszły przez określone kryteria jakościowe i zostały zaklasyfikowane jako prawidłowe. Oznacza to, że odczyty te spełniają minimalne wymagania jakościowe i są uważane za użyteczne do analizy. W rzeczywistości nie zawsze jest możliwe osiągnięcie PF = 100%, ponieważ mogą występować różne czynniki zakłócające, takie jak błędy w kolejności zasad, fragmenty o niskiej jakości lub inne problemy techniczne. Przyjmuje się, że wartości PF% na poziomie 80% lub wyższym są uważane za akceptowalne i wskazują na dobre jakościowo wyniki sekwencjonowania w technologii Illumina, natomiast PF% nie może spaść poniżej poziomu 63%.

Clusters density [K/mm²] dotyczy gęstości klastrów w sekwencjonowaniu NGS Illumina i odnosi się do procesu grupowania fragmentów DNA na powierzchni płytki *flow cell* przed rozpoczęciem procesu sekwencjonowania właściwego. Gęstość klastrów ma kluczowe znaczenie, ponieważ wpływa na jakość i efektywność

sekwencjonowania. Optymalny proces *clusteringu* jest ważny dla uzyskania dokładnych i wiarygodnych danych sekwencji. Dla platformy MiniSeq Illumina za akceptowalne uważa się wartości w zakresie od 170 do 220 tysięcy klastrów na milimetr kwadratowy [K/mm²]. Nadmierne przekładowanie powoduje problemy z analizą odczytanego przez sekwenator obrazu, co skutkuje niejednoznacznością podczas procesu przypisywania zasad, a to z kolei prowadzi do spadku jakości danych.

Quality Scores (Q30%) – wskaźnik ten oznacza jakość składania sekwencji, co jest kluczowe dla dokładności odczytu. Wartość Q30 oznacza, że prawdopodobieństwo błędnego odczytu jednej konkretnej zasady w sekwencji wynosi 1/1000, co jest równoważne dokładności na poziomie 99,9%. Q30 mierzony jest na podstawie jakości odczytów w każdym cyklu sekwencjonowania i stanowi przekształcenie logarytmiczne, wyrażające stosunek intensywności sygnału zarejestrowanego przez sekwenator do intensywności sygnału referencyjnego. Im wyższa wartość Q30, tym lepsza jakość odczytów sekwencji, co przekłada się na bardziej pewny i dokładny wynik sekwencjonowania. Platforma MiniSeq Illumina zapewnia wartość Q30 na poziomie większym od 80%.

Biblioteka kontrolna PhiX jest używana jako próbka kontrolna lub referencyjna do kalibracji i monitorowania jakości sekwencjonowania ze względu na zrównoważony skład nukleotydów. Wartość procentowa wskaźnika PhiX może różnić się w zależności od konkretnego eksperymentu, ale zazwyczaj rekomenduje się, aby PhiX stanowił ≥1% całkowitej objętości klastra. Zbyt niski procentowy udział PhiX może utrudnić kalibrację i monitorowanie jakości, podczas gdy zbyt wysoki procentowy udział PhiX może obniżyć efektywność sekwencjonowania próbki docelowej. W protokole COVIDSeq optymalna ilość PhiX powinna być na poziomie zbliżonym do 2% [41, 42].

2. Interpretacja wyników i analiza wariantów SARS-CoV-2

Właściwa interpretacja wyników jest kluczowym krokiem związanym z analizą bioinformatyczną danych NGS. W przypadku SARS-CoV-2 przebieg analizy danych obejmuje demultipleksację indeksów, generowanie plików FASTQ/FASTA oraz weryfikację i raportowanie wyników.

Demultipleksacja indeksów (ang. *demultiplexing*) wykonywana jest w początkowym etapie analizy i ma na celu wyodrębnienie sekwencji należących do różnych próbek na podstawie ich indeksów. Każda próbka jest reprezentowana jako osobny zestaw danych. Po demultipleksacji indeksy są usuwane z danych sekwencyjnych, pozostawiając jedynie sekwencje [16, 44].

Generowanie plików FASTQ ma na celu zapisanie w nich danych zawierających sekwencje nukleotydowe i informacje o jakości odczytu. Pliki FASTQ są powszechnie używane do przechowywania wyników sekwencjonowania [47].

Identyfikacja wariantów genetycznych SARS-CoV-2 polega na analizie sekwencji, przeprowadzanej w oprogramowaniu Dragen Covid Lineage na platformie BaseSpace Sequence Hub Illumina [48]. Oprogramowanie to jest narzędziem używanym do określania linii ewolucyjnych oraz mutacji SARS-CoV-2. Dopasowuje odczyty do wybranego przez użytkownika genomu referencyjnego. Algorytmy porównawcze uwzględniają różnice w długości sekwencji, insercje, delecje i substitucje nukleotydów.

Generowanie plików FASTA odbywa się na platformie BaseSpace Sequence Hub Illumina po zakończonej analizie sekwencji przeprowadzanej w oprogramowaniu Dragen Covid Lineage. Pliki FASTA są używane w analizach bioinformatycznych do przechowywania, udostępniania i przetwarzania sekwencji genetycznych.

Weryfikacja i raportowanie wyników polega na przesłaniu plików FASTA wraz z metadanymi do globalnej bazy danych GISAID w celu otrzymania ID sekwencji i jej weryfikacji. ID sekwencji to unikalny kod EPI – identyfikator przydzielany każdej sekwencji genomowej umieszczonej w GISAID. Dzięki temu możliwe jest jednoznaczne rozróżnienie każdej sekwencji oraz zidentyfikowanie próbek pacjenta. Pozwala to naukowcom na śledzenie szlaków transmisji, identyfikację ognisk zakażeń oraz analizę rozprzestrzeniania się choroby. Baza GISAID umożliwia globalną współpracę naukową poprzez dzielenie się danymi i wynikami badań między naukowcami z różnych regionów świata. Jest to istotne w kontekście pandemii, gdzie szybka wymiana informacji genetycznych jest kluczowa dla skutecznych działań ograniczających szerzenie się drobnoustroju [49].

PODSUMOWANIE

Sekwencjonowanie NGS stało się kluczową technologią w dziedzinie biologii molekularnej i genomiki, umożliwiając badaczom analizę sekwencji genetycznych chorobotwórczych mikroorganizmów. W walce z chorobą zakaźną COVID-19 sekwencjonowanie NGS to potężne narzędzie służące do zrozumienia wirusa SARS-CoV-2, diagnozowania przypadków, monitorowania pojawiania się nowych szczepów oraz modelowania filodynamiki w epidemiologii molekularnej. Modelowanie filodynamiki łączy dane sekwencji genetycznych z informacjami epidemiologicznymi w celu odtworzenia historii ewolucyjnej i dynamiki transmisji patogenów. NGS ułatwia gromadzenie dużej ilości danych genomowych, co pozwala badaczom analizować relacje ewolucyjne między różnymi próbkami wirusa SARS-CoV-2.

Sekwencjonowanie genomu SARS-CoV-2 jest niezwykle ważne dla zrozumienia jego zmienności genetycznej, rozprzestrzeniania się wirusa oraz identyfikacji nowych wariantów, które mogą wpływać na skuteczność terapii i szczepień. Tradycyjne metody diagnostyczne, takie jak testy PCR, mają na celu wykrywanie określonych regionów genetycznych wirusa. Natomiast NGS umożliwia zsekwencjonowanie całego genomu wirusa, co daje kompleksowy obraz genetycznego jego składu. Dzięki tej zdolności można identyfikować konkretne warianty i subwarianty wirusa oraz wykrywać potencjalne mutacje, które mogą wpływać na skuteczność testów diagnostycznych lub interwencji terapeutycznych.

Stosowane obecnie platformy, służące do sekwencjonowania genomu SARS-CoV-2, m.in. MiniSeq/MiSeq (Illumina), MinION/GridION (Oxford Nanopore Technologies) oraz Sequel I/II (PacBio), różnią się przepustowością, dokładnością odczytu i stopniem pokrycia genomu, czasem trwania analizy oraz kosztami badania pojedynczej próbki. Wyniki przeprowadzonych przez Carbo i wsp. [50] badań porównawczych wskazują, że sekwencjonowanie (96 próbek) z wykorzystaniem technologii Illumina jest dłuższe (ok. 24 h) w porównaniu do sekwencjonowania nanoporowego (ok. 13 h), a całkowite czasy obrotu próbki (tTAT; ang. *total turn-around time*) wynoszą – odpowiednio – ok. 3 dni (systemy Illumina oraz PacBio) oraz ok. 1 dzień (systemy Oxford Nanopore Technologies). Koszt badania w przeliczeniu na pojedynczą próbkę dla obu metod wyniósł

ok. 50 – 100 euro. Zgodnie z wynikami badań Rana-singhe i wsp. [51] ww. metody różnią się wskaźnikiem oceny jakości danych surowych. Prawdopodobieństwo nieprawidłowego odczytu zasady dla technologii Illumina wynosi od 1/1000 do 1/10000 (dokładność rzędu 99,9 – 99,99%), natomiast w przypadku technologii ONT od 1/10 do 1/100, co przekłada się na dokładność między 90% a 99%. Jednocześnie sekwencjonowanie z wykorzystaniem systemów PacBio umożliwia sekwencjonowanie zapewniające odczyty długich fragmentów, jednakże częstość występowania błędów w odczycie nadal jest wyższa niż w przypadku technologii sekwencjonowania przez syntezę czy sekwencjonowania nanoporowego, szczególnie w przypadkach występujących błędów insercji lub delecji [52, 53].

Technologia sekwencjonowania NGS Illumina oferuje skuteczne narzędzie do badania wirusa SARS-CoV-2, również ze względu na swoją wysoką wydajność, precyzję i możliwość jednoczesnego sekwencjonowania

wielu próbek. Odpowiedniej jakości dane sekwencyjne uzyskane z użyciem Illumina MiniSeq wymagają jednak odpowiedniego przygotowania próbek oraz precyzyjnej, wieloetapowej manualnej pracy związanej z właściwym przygotowaniem biblioteki NGS. Bardzo istotna jest także odpowiednia analiza i interpretacja wyników sekwencjonowania, w tym usuwanie i filtracja odczytów o niskiej jakości. Analiza tych danych za pomocą odpowiedniego oprogramowania bioinformatycznego na platformie Illumina umożliwia identyfikację wariantów genetycznych i mutacji, które mają kluczowe znaczenie dla monitorowania dynamiki zachorowań na COVID-19 i dostosowywania skutecznych strategii zdrowotnych. Standaryzacja protokołów oraz inicjatywy dzielenia się danymi NGS są istotne dla harmonizacji technologii NGS i ułatwienia globalnej współpracy w badaniach nad COVID-19. Z tych wszystkich względów identyfikacja i monitorowanie tych wariantów są kluczowe dla skutecznej kontroli choroby COVID-19.

PIŚMIENNICTWO

- Pillay S, Giandhari J, Tegally H, et al. Whole Genome Sequencing of SARS-CoV-2: Adapting Illumina Protocols for Quick and Accurate Outbreak Investigation during a Pandemic. *Genes (Basel)*. 2020; 11(8): 949. DOI: 10.3390/genes11080949.
- Niemi MEK, Daly MJ, Ganna A. The human genetic epidemiology of COVID-19. *Nat Rev Genet*. 2022; 23(9): 533–546.
- Rusňáková D, Sedláčková T, Radvák P, et al. Systematic Genomic Surveillance of SARS-CoV-2 Virus on Illumina Sequencing Platforms in the Slovak Republic – One Year Experience. *Viruses*. 2022; 14(11): 2432. DOI: 10.3390/v14112432.
- WHO COVID-19 Dashboard data. World Health Organization, 2024, <https://data.who.int/dashboards/covid19>, [06.08.2024].
- John G, Sahajpal NS, Mondal AK, et al. Next-Generation Sequencing (NGS) in COVID-19: A Tool for SARS-CoV-2 Diagnosis, Monitoring New Strains and Phylodynamic Modeling in Molecular Epidemiology. *Curr Issues Mol Biol*. 2021; 43(2): 845–867.
- Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020; 579(7798): 265–269.
- Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. *Gene Rep*. 2020; 19: 100682. DOI: 10.1016/j.genrep.2020.100682.
- Udugama B, Kadhiresan P, Kozłowski HN, et al. Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano*. 2020; 14(4): 3822–3835.
- Hu B, Guo H, Zhou P, et al. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021; 19(3): 141–154.
- Pan B, Ji Z, Sakkiath S, et al. Identification of Epidemiological Traits by Analysis of SARS-CoV-2 Sequences. *Viruses*. 2021; 13(5): 764. DOI: 10.3390/v13050764.
- Bader W, Delerice J, Aherfi S, et al. Quasispecies Analysis of SARS-CoV-2 of 15 Different Lineages during the First Year of the Pandemic Prompts Scratching under the Surface of Consensus Genome Sequences. *Int. J. Mol. Sci*. 2022; 23(24): 15658. DOI: 10.3390/ijms232415658.
- Lauring AS, Hodcroft EB. Genetic Variants of SARS-CoV-2 – What Do They Mean? *JAMA*. 2021; 325(6): 529–531.
- Aleem A, Akbar Samad AB, Vaqar S. Emerging Variants of SARS-CoV-2 and Novel Therapeutics Against Coronavirus (COVID-19). In: StatPearls. StatPearls Publishing, Treasure Island (FL). 2024, <https://www.ncbi.nlm.nih.gov/books/NBK570580/>.
- Carpenter RE, Tamrakar V, Chahar H, et al. Confirming Multiplex RT-qPCR Use in COVID-19 with Next-Generation Sequencing: Strategies for Epidemiological Advantage. *Glob Health Epidemiol Genom*. 2022; 2022: 2270965. DOI: 10.1155/2022/2270965.
- Tao K, Tzou PL, Nouhin J, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet*. 2021; 22(12): 757–773.
- Materiały dla użytkowników systemów Illumina, An introduction to Next-Generation Sequencing Technology, 2017, <https://www.illumina.com/>.
- Materiały dla użytkowników systemów Illumina, Illumina Sequencing Technology, 2010, <https://www.illumina.com/>.
- Singh RR, Luthra R, Routbort MJ, et al. Implementation of next generation sequencing in clinical molecular diagnostic laboratories: advantages, challenges and potential. *Expert Review of Precision Medicine and Drug Development*, 2016; 1(1): 109–120.
- Tejari N, Mohak S, Lakshminarayanan CP, et al. High-quality single amplicon sequencing method for illumina MiSeq platform using pool of 'N' (0-10) spacer-linked target specific primers without PhiX spike-in. *BMC Genomics*. 2023; 24(1): 141. DOI: 10.1186/s12864-023-09233-4.
- Brown TA. Sekwencjonowanie genomów. In: Węgliński P, (eds.). *Genomy*. Wyd. Naukowe PWN, Warszawa. 2019: 79–110.

21. Liu T, Chen CY, Chen-Deng A, et al. Joining Illumina paired-end reads for classifying phylogenetic marker sequences. *BMC Bioinformatics*. 2020; 21: 105. DOI: 10.1186/s12859-020-3445-6.
22. Buermans HPJ, Den Dunnen JT. Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta*. 2014; 842(10): 1932–1941.
23. Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol*. 2018; 122(1): e59. DOI: 10.1002/cpmb.59.
24. Materiały dla użytkowników systemów Illumina, Optimizing Cluster Density on Illumina Sequencing Systems, 2016, <https://www.illumina.com/>.
25. Barbian HJ, Kittner A, Teran R, et al. A response playbook for early detection and population surveillance of new SARS-CoV-2 variants in a regional public health laboratory. *BioRxiv* 2023.09.01.23293209 [Preprint]. 2023. DOI: 10.1101/2023.09.01.23293209.
26. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016; 17(6): 333–351.
27. Materiały dla użytkowników systemów Illumina, CMOS Chip and One-Channel SBS Chemistry, 2018, <https://www.illumina.com/>.
28. Materiały dla użytkowników systemów Illumina, Faster sequencing and data processing. 2-channel SBS generates data faster than 4-channel SBS, while maintaining quality and accuracy, 2018, <https://www.illumina.com/>.
29. John G, Sahajpal NS, Mondal AK, et al. Next-Generation Sequencing (NGS) in COVID-19: A Tool for SARS-CoV-2 Diagnosis, Monitoring New Strains and Phylodynamic Modeling in Molecular Epidemiology. *Curr Issues Mol Biol*. 2021; 43(2): 845–867.
30. Cheng C, Fei Z, Xiao P. Methods to improve the accuracy of next-generation sequencing. *Front Bioeng Biotechnol*. 2023; 11: 98211. DOI: 10.3389/fbioe.2023.982111.
31. Khare S, Gurry C, Freitas L, et al. GISAID's Role in Pandemic Response. *China CDC Wkly*. 2021; 3(49): 1049–1051.
32. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro Surveill*. 2017; 22(13): 30494. DOI: 10.2807/1560-7917.ES.2017.22.13.30494.
33. Materiały dla użytkowników systemów Illumina, Improved performance using the Illumina COVIDSeq™ Assay (96 samples), Technical note, 2021, <https://www.illumina.com/>.
34. Rekomendacje FDA-U.S. Food and Drug Administration, Illumina COVIDSeq Test Instructions for Use, <https://www.fda.gov/>.
35. Carpenter RE, Tamrakar VK, Almas S, et al. Optimization of the Illumina COVIDSeq™ protocol for decentralized, cost-effective genomic surveillance. *Pract Lab Med*. 2023; 34: e00311. DOI: 10.1016/j.plabm.2023.e00311.
36. Materiały dla użytkowników systemów Illumina, Illumina COVID-Seq RUO Kits Documentation, 2021, <https://www.illumina.com/>.
37. Illumina DNA Prep. A fast, integrated library prep workflow for a wide range of sequencing applications, Data sheet, 2023, <https://www.illumina.com/>.
38. Head SR, Komori HK, LaMere SA, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*. 2014; 56(2): 61. DOI: 10.2144/000114133.
39. Materiały dla użytkowników systemów Illumina, IDT for Illumina Nextera DNA UD Indexes Files for Local Run Manager, 2021, <https://www.illumina.com/>.
40. Winston LCK, Si EP, Chun KL, et al. Towards a Rapid-Turnaround Low-Depth Unbiased Metagenomics Sequencing Workflow on the Illumina Platforms. *Bioengineering (Basel)*. 2023; 10(5): 520. DOI: 10.3390/bioengineering10050520.
41. Curini V, Ancora M, Jurisic L, et al. Evaluation of next generation sequencing approaches for SARS-CoV-2. *Heliyon*. 2023; 9: e21101. DOI: 10.1016/j.heliyon.2023.e21101.
42. Materiały dla użytkowników systemów Illumina, Using a PhiX Control for HiSeq® Sequencing Runs, Technical not, 2013, <https://www.illumina.com/>.
43. Ray M, Sable MN, Sarkar S, et al. Essential interpretations of bioinformatics in COVID-19 pandemic. *Meta Gene*. 2021; 27: 100844. DOI: 10.1016/j.mgene.2020.100844.
44. Goswami C, Sheldon M, Bixby C, et al. Identification of SARS-CoV-2 variants using viral sequencing for the Centers for Disease Control and Prevention genomic surveillance program. *BMC Infect Dis*. 2022; 22(1): 404. DOI: 10.1186/s12879-022-07374-7.
45. Santacruz D, Enane FO, Fundel-Clemens K, et al. Automation of high-throughput mRNA-seq library preparation: a robust, hands-free and time efficient methodology. *SLAS Discovery*. 2022; 27: 140–147. DOI: 10.1016/j.slasd.2022.01.002.
46. Materiały dla użytkowników systemów Illumina. Coverage Depth Recommendations. <https://emea.illumina.com/science/technology/next-generation-sequencing/plan-experiments/coverage.html>.
47. Lowry K, Bauer MJ, Buckley C, et al. Evaluation of Illumina® COVIDSeq™ as a tool for Omicron SARS-CoV-2 characterisation. *J Virol Methods*. 2023; 322: 114827. DOI: 10.1016/j.jviromet.2023.114827.
48. Materiały dla użytkowników systemów Illumina. DRAGEN COVID Lineage, The Illumina® DRAGEN Microbial Amplicon App uses DRAGEN and 3rd-party tools to detect viruses and generate consensus sequences. <https://emea.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/dragen-covid-lineage.html>.
49. Donzelli S, Ciuffreda L, Pontone M, et al. Optimizing the Illumina COVIDSeq laboratorial and bioinformatics pipeline on thousands of samples for SARS-CoV-2 Variants of Concern tracking. *Comput Struct Biotechnol J*. 2022; 20: 2558–2563.
50. Carbo EC, Mourik K, Boers S, et al. A comparison of five Illumina, Ion Torrent, and nanopore sequencing technology-based approaches for whole genome sequencing of SARS-CoV-2. *Eur J Clin Microbiol Infect Dis* 2023; 42: 701–713. DOI: doi.org/10.1007/s10096-023-04590-0.
51. Ranasinghe D, Jayadas TTP, Jayathilaka D, et al. Comparison of different sequencing techniques for identification of SARS-CoV-2 variants of concern with multiplex real-time PCR. *PLoS One* 2022; 17(4): e0265220. DOI: 10.1371/journal.pone.0265220.
52. Furlani B, Kouter K, Rozman D, et al. Sequencing of Nucleic Acids: from the First Human Genome to Next Generation Sequencing in COVID-19 Pandemic. *Acta Chim Slov* 2021; 68(2): 268–278.
53. Marcinkowska-Swojak M, Rakoczy M, Podkowiński J, et al. Od Sangera do sekwencjonowania genomów – przegląd technologii sekwencjonowania DNA. *Postepy Biochem* 2024; 70(2): 173–189. DOI: 10.18388/pb.2021_534.