

STRESZCZENIE

Sekwencjonowanie nanoporowe należy do technik tzw. trzeciej generacji. Polega ono na selektywnym transporcie kwasów nukleinowych przez kanały białkowe o nanometrowych rozmiarach (nanopory) i identyfikacji ich sekwencji na podstawie zmian mierzonego sygnału elektrycznego. Technika nanoporowa pozwoliła na znaczące wydłużenie odczytów sekwencyjnych oraz bezpośrednie sekwencjonowanie natywnych cząsteczek DNA i RNA. W efekcie, w ciągu zaledwie kilku lat od czasu wprowadzenia pierwszego sekwenatora nanoporowego, stała się ona jedną z wiodących technik sekwencjonowania. Zakres zastosowań techniki nanoporowej jest bardzo szeroki i obejmuje w szczególności asemblację genomów, badanie zmienności strukturalnej, identyfikację modyfikacji kwasów nukleinowych oraz badanie alternatywnego składania transkryptów. Dzięki swojej dostępności i prostocie, sekwencjonowanie nanoporowe jest również coraz powszechniej stosowane w badaniach epidemiologicznych, do szybkiej identyfikacji i monitorowania rozprzestrzeniania się chorobotwórczych bakterii oraz wirusów.

TRZY GENERACJE TECHNIK SEKWENCJONOWANIA

Poznanie informacji zakodowanej w sekwencji DNA stanowi klucz do zrozumienia mechanizmów przechowywania, odczytywania i przekazywania tej informacji przez komórki, a także poznania zmienności genetycznej i jej wpływu na różnice fenotypowe pomiędzy organizmami. Wczesne metody sekwencjonowania, opracowane w latach 70-tych ubiegłego wieku, znane są dziś pod nazwą sekwencjonowania pierwszej generacji [1]. Spośród nich, największe uznanie zdobyła metoda terminacji łańcucha, opracowana przez zespół Fredericka Sangera z Uniwersytetu w Cambridge [2]. U podstaw tej metody leży reakcja wydłużania nici DNA przez enzym – polimerazę DNA, która dodaje kolejno pojedyncze nukleotydy obecne w mieszaninie reakcyjnej. W skład mieszaniny nukleotydów wchodzi również ich zmodyfikowane chemicznie wersje, pełniące rolę terminatorów reakcji. Przyłączenie ich do tworzonego łańcucha DNA uniemożliwia jego dalsze wydłużanie. Terminatory są losowo wbudowywane w wydłużaną cząsteczkę, w wyniku czego powstają fragmenty DNA o różnej długości. Pierwotnie sekwencja odczytywana była poprzez równoległe elektroforetyczne rozdzielanie produktów z czterech mieszanin reakcyjnych, zawierających poszczególne rodzaje terminatorów. Obecnie odczytywanie sekwencji bazuje na znacznikach fluorescencyjnych dołączanych do poszczególnych rodzajów terminatorów, umożliwiających ich identyfikację w jednej próbówce, co czyni sekwencjonowanie dużo szybszym i łatwiejszym w wykonaniu.

Zastosowanie metody Sangera na szeroką skalę doprowadziło do poznania sekwencji genomów wielu gatunków, w tym człowieka [3,4]. Jest to metoda wciąż używana jako kontrolna w rutynowej pracy laboratoryjnej, z uwagi na wysoką dokładność (99.99%) i umiarkowaną długość wyników sekwencji – odczytów – zazwyczaj od 500 do 1 000 nukleotydów (nt). Jej główne ograniczenie stanowi brak możliwości multipleksowania (w jednej reakcji może być czytana tylko jedna matryca DNA) i wynikające z tego wysokie koszty sekwencjonowania. Dążenie do rozszyfrowania genomów kolejnych gatunków stymulowało opracowanie nowych sposobów sekwencjonowania, pozwalających na zwiększenie wydajności tego procesu (Ryc. 1). Jeden z nich polegał na unieruchomieniu krótkich fragmentów DNA na płytce, ich namnożeniu oraz równoczesnym sekwencjonowaniu – podejście to nazwano sekwencjonowaniem drugiej generacji (NGS, ang. *next generation sequencing*) [5]. W latach 2005-2007 wprowadzono na rynek trzy rodziny sekwenatorów NGS, w których zastosowano różne metody odczytu krótkich sekwencji: pirosekwencjonowanie, sekwencjonowanie przez ligację lub sekwencjonowanie przez syntezę. To ostatnie podejście okazało się najbardziej popularną techniką NGS. Sekwencjonowanie przez syntezę bazuje na metodzie Sangera i charakteryzuje się podobnie wysoką dokładnością, generując w tym samym czasie wielokrotnie więcej danych wyjściowych w po-

mgr Anastasiia Satyr,

dr Agnieszka Żmieńko✉

Zakład Biologii Molekularnej i Systemowej,
Instytut Chemii Bioorganicznej Polskiej Akademii Nauk

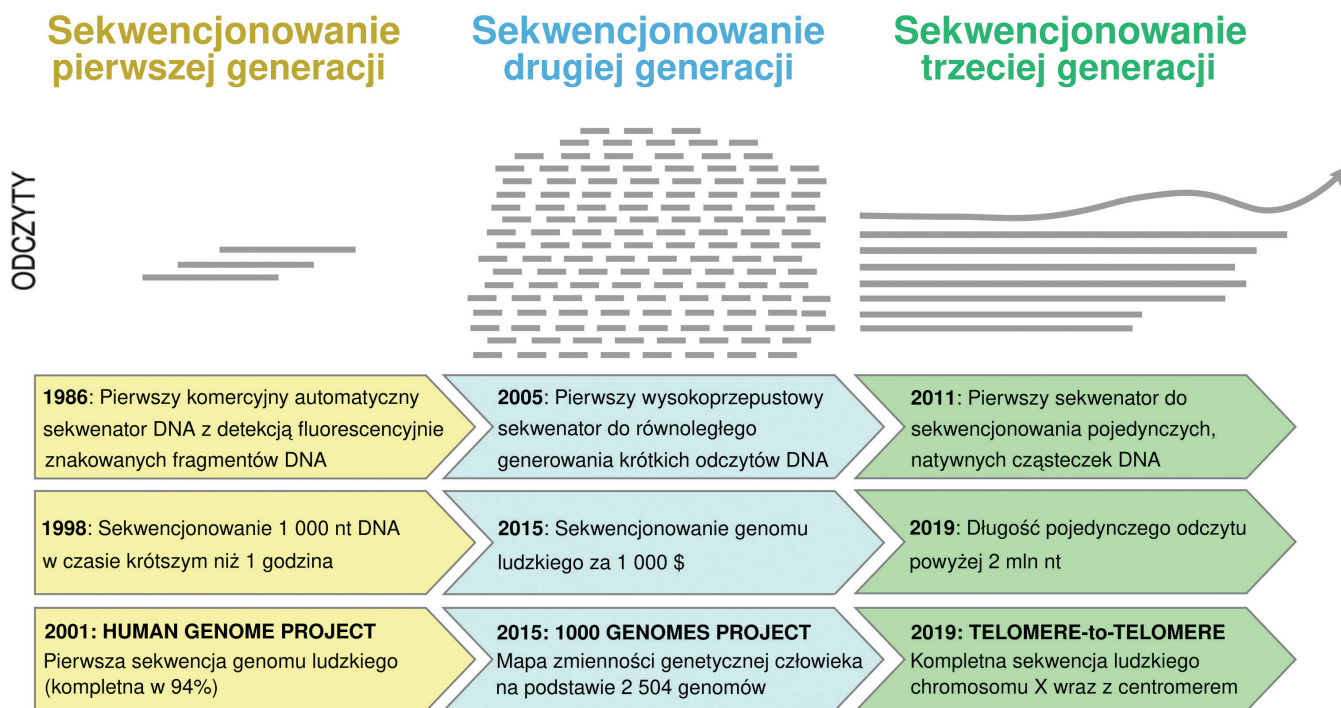
https://doi.org/10.18388/pb.2020_328

✉ autor korespondujący: akisiel@ibch.poznan.pl

Słowa kluczowe: sekwencjonowanie genu, technologia nanoporowa, długie odczyty, asemblacja *de novo*, zmienność strukturalna, badania transkryptomyczne, modyfikacje zasad, diagnostyka chorób

Wykaz skrótów: cDNA – komplementarny DNA (ang. *complementary DNA*), NGS – sekwencjonowanie następnej generacji (ang. *next generation sequencing*), nt – nukleotyd, SMRT – sekwencjonowanie pojedynczej cząsteczki w czasie rzeczywistym (ang. *single-molecule real-time sequencing*)

Podziękowania: Badania prowadzone przez autorki niniejszej pracy przeglądowej są finansowane przez Narodowe Centrum Nauki w ramach projektu SONATA (2017/26/D/NZ2/01079).



Rycina 1. Rozwój technik sekwencjonowania. Na rysunku zobrazowano różnicę pomiędzy ilością i długością odczytów pierwszej, drugiej i trzeciej generacji (szczegóły w tekście). Schemat przedstawia milowe osiągnięcia w rozwoju technologii sekwencjonowania oraz ich wpływ na rozwój wiedzy, na przykładzie poznawania sekwencji genomu człowieka.

staci krótkich odczytów DNA. Początkowo ich długość wynosiła 36 nt, z czasem zaś wzrosła do 150 nt, a w przypadku niektórych sekwenatorów – nawet do 300 nt. W momencie wprowadzenia technik NGS koszty sekwencjonowania całych genomów znacząco spadły, co spowodowało prawdziwy przełom w wielu dziedzinach biologii i medycyny.

Techniki NGS mają ogromne spektrum zastosowań. Umożliwiły one rozszyfrowanie sekwencji genomowego DNA gatunków istotnych dla nauki i przemysłu, dały początek badaniom populacyjnym, a także znalazły się na liście metod diagnostycznych, wykrywających choroby genetyczne [6]. Dzięki opracowaniu odpowiednich metod przygotowania bibliotek sekwenacyjnych stało się także możliwe sekwencjonowanie cząsteczek RNA na dużą skalę [7]. To z kolei umożliwiło analizę całego transkryptomu, czyli wszystkich produktów ekspresji genów obecnych w komórce w danych warunkach oraz pozwoliło na identyfikację nowych typów krótkich niekodujących RNA, pełniących w komórkach funkcje regulatorowe. Z drugiej strony, zarówno sekwenatory, jak i odczytyniki do pracy z NGS pozostają dość drogie i dostępne są raczej dla większych laboratoriów i zespołów badawczych. W dodatku odtworzenie wyjściowych sekwencji z olbrzymiej ilości krótkich odczytów wymaga wieloetapowej analizy, wykonywanej przez wykwalifikowanych specjalistów – bioinformatyków, z wykorzystaniem odpowiednich zasobów komputerowych. W wielu przypadkach sekwencja odtworzona z krótkich odczytów może być niepełna lub zawierać błędy. Te ograniczenia wskazywały na konieczność zwiększenia długości odczytów, przy zachowaniu wysokiej dokładności, wydaj-

ności i krótkiego czasu sekwencjonowania. Odpowiedź na to zapotrzebowanie przyniósł rozwój technik sekwencjonowania trzeciej generacji, spośród których warto wymienić szczególnie sekwencjonowanie pojedynczej cząsteczki w czasie rzeczywistym (SMRT, ang. *single-molecule real-time sequencing*) oraz sekwencjonowanie nanoporowe [8]. Ich wspólną cechą jest brak konieczności amplifikacji materiału wyjściowego – do detekcji i odczytania danej sekwencji wystarczający jest sygnał pochodzący od jednej cząsteczki RNA lub DNA. Pozwoliło to na znaczące uproszczenie procedur przygotowania bibliotek sekwenacyjnych, a co za tym idzie – wydłużenie odczytów.

Technologia SMRT została wprowadzona w 2011 roku. Polega ona na równoległym odczycie tysięcy sekwencji DNA w trakcie syntezy prowadzonej przy zastosowaniu fluorescencyjnie znakowanych nukleotydów. Bibliotekę sekwenacyjną stanowi mieszanina cząsteczek DNA (matryc) poddanych cyrkularyzacji przy pomocy specjalnych adapterów. Obecność adaptera umożliwia wiązanie się DNA ze starterem oraz przyłączoną do niego pojedynczą cząsteczką polimerazy DNA, unieruchomioną w specjalnych zagłębieniach płytki sekwenacyjnej. W trakcie procesu syntezy zagłębienie oświetlane jest promieniem lasera [8]. Rozmiar zagłębienia jest na tyle mały, że oświetlany jest wyłącznie wbudowywany nukleotyd, a w efekcie syntezy znacznik fluorescencyjny jest odcinany i dyfunduje poza obszar detekcji. Dzięki temu sekwencjonowanie zachodzi bezpośrednio w momencie wydłużania nici. Długość odczytów SMRT jest znacząco większa niż w przypadku odczytów NGS i może wynosić nawet 50 000 nt (średnio 20 000 nt). Technologia

SMRT używana jest między innymi do badań wariantów strukturalnych czy składania dużych i skomplikowanych genomów [9]. Pomimo zalet związanych z długością odczytów, przepustowość i dokładność technologii SMRT nadal zauważalnie ustępuje metodom NGS. Poza tym, z uwagi na wysoki koszt aparatury, sekwenatory SMRT, podobnie jak wcześniej sekwenatory NGS, funkcjonują przede wszystkim w dużych laboratoriach badawczych bądź usługowych.

Pod względem dostępności, prawdziwą rewolucję przyniosła kolejna metoda sekwencjonowania, oparta na innowacyjnej technologii nanoporów [10]. Jej komercyjna implementacja w postaci przenośnego sekwenatora MinION, wprowadzonego na rynek w 2014 roku przez firmę Oxford Nanopore Technologies, odniosła błyskawiczny sukces dzięki małym rozmiarom, prostocie obsługi i przystępnej cenie. W kolejnych rozdziałach skupimy się na przedstawieniu zasady sekwencjonowania nanoporowego, scharakteryzujemy najważniejsze własności i główne rodzaje białek nanoporowych stanowiących „centra sekwencyjne” oraz zaprezentujemy dostępne sekwenatory, które bazują na tej technologii. Ostatnią część pracy poświęcimy omówieniu obecnych zastosowań sekwencjonowania nanoporowego, a także perspektywom jego dalszego rozwoju.

IDEA SEKWENCJONOWANIA NANOPOROWEGO

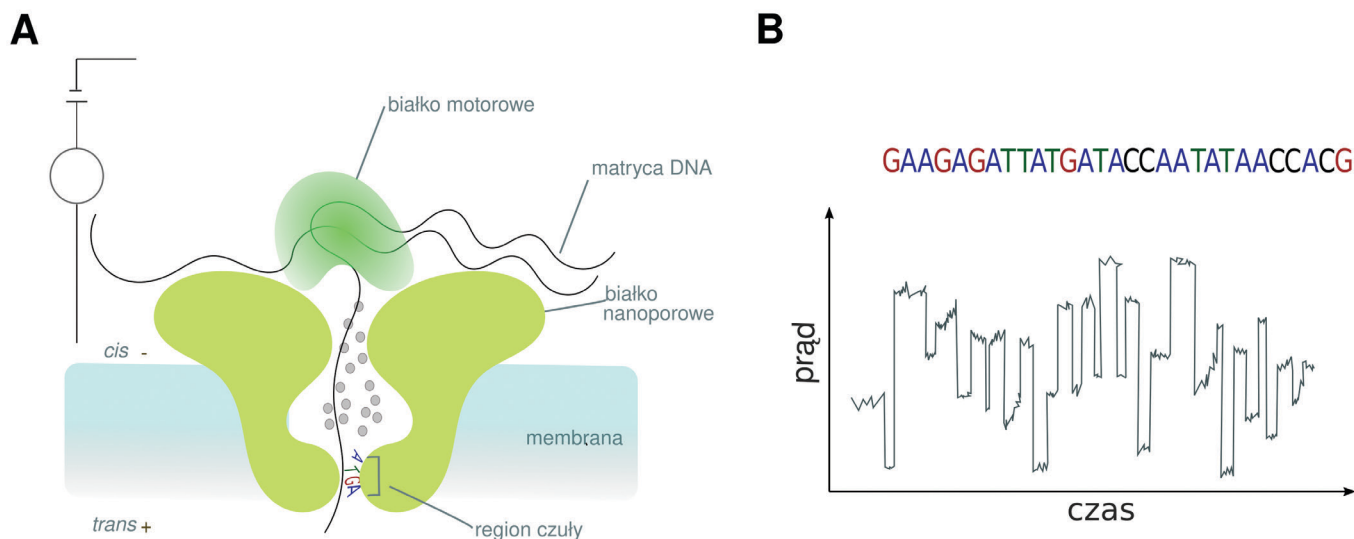
Pomysł wykorzystania nanoporów do sekwencjonowania DNA sięga lat 80-tych ubiegłego wieku, kiedy to prace nad tą technologią rozpoczęło kilka laboratoriów na świecie [11]. W 1995 roku opisano po raz pierwszy udany eksperyment elektroforezy kwasu nukleinowego przez kanał jonowy w membranie [12]. Jako kanał wykorzystano wówczas α -hemolizynę – białkową toksynę pochodzącą z bakterii *Staphylococcus aureus*, zdolną do tworzenia porów w błonach eukariotycznych. Badacze umieścili α -hemolizynę w dwuwarstwie lipidowej, rozdzielającej na dwie części pojemnik wypełniony buforem. Białko tworzyło kanał, otwierający się pod wpływem przyłożonego napięcia. Średnica kanału w najwęższym miejscu wynosiła 1,5 nm, co umożli-

wiało przejście rozciągniętej jednoniciowej cząsteczki RNA lub DNA na drugą stronę membrany, powodując jednocześnie czasowe blokowanie kanału i spadek przepływu prądu jonowego. Czas trwania spadku był proporcjonalny do długości translokowanej cząsteczki. W późniejszych eksperymentach pokazano, że śledzenie charakterystyki zmian przepływu prądu (czasu i amplitudy) związanych z przechodzeniem DNA przez nanopor umożliwia identyfikację nukleotydów, np. odróżnienie ciągów puryn od pirymidyn [13]. Dalsze prace w tym kierunku doprowadziły do stworzenia algorytmów tłumaczących zmiany potencjału elektrycznego na sekwencję DNA.

Poruszanie się DNA w nanoporze pod wpływem napięcia elektrycznego jest niezwykle szybkim procesem – czas przebywania jednego nukleotydu w najbardziej wąskim punkcie nanopora, stanowiącym obszar detekcji sygnału, wynosi mniej niż 10 μ s. Zwiększenie stosunku sygnału do szumu, niezbędne do poprawnej identyfikacji nukleotydów, wymagało wydłużenia czasu detekcji co najmniej dziesięciokrotnie [14]. Ponadto, DNA zazwyczaj występuje w formie dwuniciowej, podczas gdy rozmiar otworu białka nanoporowego i sam proces sekwencjonowania wymagają przejścia jednoniciowej cząsteczki kwasu nukleinowego. Rozwiązanie obu tych problemów przyniosło zastosowanie pomocniczego białka, zwanego białkiem motorowym (ang. *motor protein*) [11]. Jest to enzym o właściwościach helikazy, który wiąże się z DNA i umożliwia jego rozplecenie oraz skierowanie tylko jednej nici do otworu. Powoduje to jednocześnie spowolnienie i uregulowanie czasu przejścia DNA przez nanopor. Połączenie w całość wymienionych elementów doprowadziło do stworzenia sekwenatora nanoporowego w obecnym wydaniu.

PRZEBIEG EKSPERYMENTU SEKWENCJONOWANIA DNA W NANOPORACH

MinION – pierwszy i podstawowy sekwenator nanoporowy to niepozorne urządzenie ważące 90 gramów i mieszczące się na dłoni. Na czas sekwencjonowania sekwenator



Rycina 2. Zasada działania sekwenatora nanoporowego. A – białko motorowe rozplata nić DNA podczas jej przechodzenia przez nanopor w trakcie sekwencjonowania. B – schemat przekształcenia sygnału elektrycznego w sekwencję nukleotydów.

jest podłączany do laptopa i sterowany poprzez dedykowane oprogramowanie. Główne części funkcjonalne urządzenia MinION to wymienna płytkę sekwencyjną (ang. *flow cell*) i detektor. Na płytce umiejscowiona jest elektrooporowa membrana z syntetycznego polimeru, w którą wbudowane są białka – nanopory (Ryc. 2A). Membrana trwale oddziela dwa przedziały wypełnione buforem, które nazywamy stronami *cis* i *trans*. Rozpoczynając sekwencjonowanie, przygotowaną wcześniej bibliotekę sekwencyjną podajemy na stronę *cis*. Samo przygotowanie bibliotek sekwencyjnych jest stosunkowo proste i polega na przyłączeniu do obu końców DNA sekwencji adapterowych, które wiążą białko motorowe i kierują DNA w stronę nanopora. Co ważne, białko motorowe zostaje aktywowane dopiero przy kontakcie z nanoporem, a więc DNA w roztworze pozostaje w formie dwuniciowej. Tak przygotowana biblioteka jest gotowa do sekwencjonowania.

Przyłożenie napięcia elektrycznego do systemu inicjuje stochastyczny ruch jonów w roztworze i otwarcie nanoporów, co uruchamia proces sekwencjonowania. Naładowane ujemnie cząsteczki kwasu nukleinowego poruszają się w kierunku zgodnym z przyłożonym napięciem i po rozpleceniu przez białko motorowe przechodzą przez nanopor na stronę *trans* układu. Detektor wykrywa wahania potencjału elektrycznego wywołane przechodzeniem cząsteczek DNA przez obszar czuły nanopora i przekazuje je do komputera, gdzie zapisywane są za pomocą dwóch parametrów – natężenia prądu i czasu. Można te wahania przedstawić w formie wykresów zygzakowatych (ang. *squiggle plot*), o przebiegu charakterystycznym dla danej sekwencji nukleotydów (Ryc. 2B). Te informacje są przetwarzane na dalszym – bioinformatycznym etapie sekwencjonowania.

Surowe sygnały wychodzące z sekwenatora zapisywane są w formacie FAST5. Następnie algorytm obliczeniowy dokonuje transformacji tych nieczytelnych dla człowieka danych na sekwencję DNA, w procesie zwanym po angielsku *basecalling* i przekształca pliki FAST5 na format FASTQ. Ten ostatni jest standardowym formatem przechowywania informacji o sekwencji i jej jakości, wprowadzonym wraz z technologią NGS i rozpoznawanym praktycznie przez wszystkie programy do dalszych etapów analizy danych sekwencyjnych.

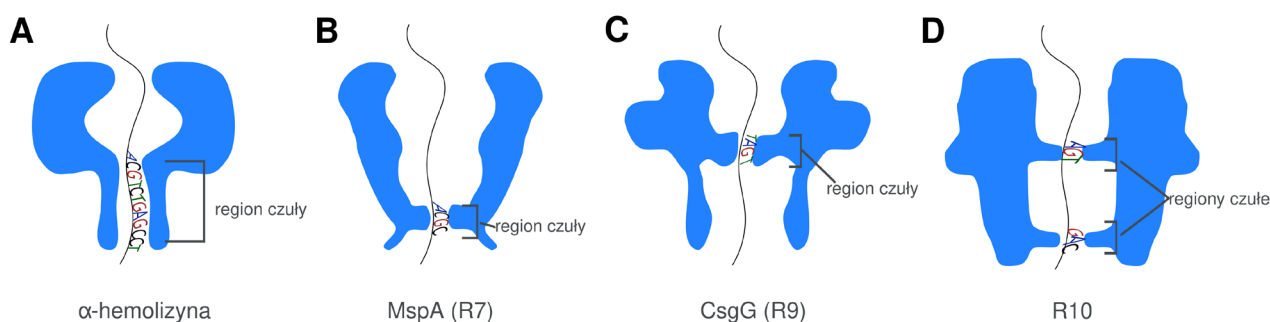
Ponieważ rozmiar i czułość obecnych nanoporów nie pozwalają na wykrywanie sygnałów pochodzących od

pojedynczych nukleotydów, interpretacja danych przez program do odczytu nukleotydów zachodzi na poziomie „słów” – segmentów złożonych z czterech nukleotydów. Unikatowe kombinacje słów zawarte są w specjalnym zestawie danych – modelu. Korzystając z takiego modelu można przekształcić otrzymane z sekwenatora sygnały na sekwencje DNA. Poprawność tego procesu ma integralne znaczenie dla dokładności sekwencjonowania. Co za tym idzie, oprogramowanie do identyfikacji sekwencji jest nieustannie rozwijane i obecnie istnieje już ponad 20 programów, bazujących na różnych typach algorytmów. Jednym z bardziej rozpowszechnionych jest Guppy Basecaller, stworzony w oparciu o sieci neuronowe, będący narzędziem o bardzo dużej dokładności i szybkości przetwarzania danych [15].

BIOLOGICZNE NANOPORY

Funkcjonalność nanopora ma kolosalny wpływ na kształt systemu sekwencyjnego i zależy od kilku jego cech. Najważniejsze z nich to struktura kanału – jego rozmiar i kształt powinny pasować do struktury jednoniciowego DNA/RNA – oraz czułość, którą można określić jako unikalność zaburzeń przepływu prądu przez poszczególne nukleotydy. Nanopor powinien zapewniać rozpoznawanie i odróżnianie możliwie dużej ilości sygnałów wywołanych przejściem kwasu nukleinowego, a jednocześnie nie generować szumów, powodowanych np. wykrywaniem nukleotydów znajdujących się poza obszarem detekcji. W celu zapewnienia obu tych cech, testuje się zarówno nanopory biologiczne jak i te wytwarzane na bazie materiałów nieorganicznych, takich jak tlenek krzemu czy grafen [10]. W porównaniu z biologicznymi, nieorganiczne nanopory zachowują większą stabilność w warunkach napięcia elektrycznego i mogą tworzyć na tyle krótki kanał, aby generować sygnał dla 1–2 nukleotydów. Ich wadą jest to, że są słabo modyfikowalne pod kątem własności chemicznych, mniej kompatybilne z białkami motorowymi i produkują słabo czytelny sygnał [16]. Dlatego w sekwenatorach DNA zaimplementowano systemy bazujące na nanoporach biologicznych. Są to zmodyfikowane formy białek kanałowych, w naturze zdolnych do tworzenia porów w błonach komórkowych.

Wspomniana wcześniej α -hemolizyna, będąca pierwszą skuteczną implementacją nanopora, była dobrym punktem odniesienia dla opracowania całego systemu. Ma ona jednak pewne niedoskonałości. Region czuły nanopora czyli



Rycina 3. Białka nanoporowe wykorzystywane w sekwencjonowaniu kwasów nukleinowych. A – α -hemolizyna, pierwsze białko nanoporowe; B – MspA, białko nanoporowe w wersji płytek sekwencyjnych R7; C – CsgG, białko nanoporowe w wersji płytek sekwencyjnych R9; D – białko nanoporowe z dwoma regionami czułymi w wersji płytek sekwencyjnych R10.

ta część kanału, w której nukleotydy znajdują się podczas detekcji, musi być możliwie wąski oraz mieścić jak najmniej nukleotydów jednocześnie [11]. W regionie czułym α -hemolizyny mieści się około 10 nukleotydów, które jednocześnie zaburzają przepływ prądu, co utrudnia ich identyfikację i powoduje stworzenie dodatkowego szumu podczas detekcji (Ryc. 3A). W nanoporowych płytkach sekwencyjnych wersji R7, wprowadzonych wraz z sekwenatorem MinION, zastosowano inne białko, porynę A, pochodzącą od bakterii *Mycobacterium smegmatis* (MspA). Białko MspA stabilniej niż α -hemolizyna utrzymuje stan otwarty pod wpływem przyłożonego napięcia, a w jego regionie czułym mieszczą się około 4 nukleotydy, co pozwala na dużo lepszą identyfikację sekwencji przeciąganej nici DNA w oparciu o charakterystykę zmian przepływu prądu w czasie (Ryc. 3B).

Najbardziej skutecznym jak dotąd nanoporem biologicznym okazało się zmodyfikowane białko CsgG (ang. *Curli production assembly/transport component*) z *Escherichia coli*. W naturalnych warunkach CsgG jest lipoproteiną wbudowaną w zewnętrzną błonę bakteryjną, gdzie pełni funkcję kanału wyprowadzającego na zewnątrz komórek amyloidy, uczestniczące w tworzeniu biofilmu [17]. Nanopory CsgG zostały zaimplementowane w produkcji płytek sekwencyjnych w wersji R9, udostępnionych w październiku 2016 roku [18]. Nanopory R9 mają podobną do R7 długość regionu czułego, ale wykazują jeszcze lepszą stabilność, selektywność oraz kompatybilność z innymi elementami systemu nanoporowego (Ryc. 3C). Dalsze udoskonalanie nanoporów doprowadziło do wydania kolejnej wersji płytek sekwencyjnych, R10, wprowadzonych do sprzedaży na początku 2020 r. W tej wersji otwór nanopora jest wydłużony i posiada dwa regiony czułe, kolejno po sobie zbierające sygnały generowane przez przechodzącą cząsteczkę DNA (Ryc. 3D). W efekcie prowadzi to do większej dokładności odczytu i zmniejsza ryzyko błędów związanych z sekwencjonowaniem regionów zawierających ciągi identycznych nukleotydów.

ZALETY I WADY SEKWENCJONOWANIA NANOPOROWEGO

W porównaniu z metodami NGS, sekwencjonowanie nanoporowe wyróżnia szereg cech, które zadecydowały o szybkim rozpowszechnieniu się tej techniki. Należą do nich:

Identyfikacja DNA w natywnej formie. Źródłem sygnału jest samo przechodzenie DNA przez nanopor, a nie proces jego amplifikacji. Znacząco upraszcza to i przyspiesza przygotowanie biblioteki sekwencyjnej. Podczas tego procesu, adaptory są przyłączane do DNA przy udziale enzymu ligazy lub transpozazy, zależnie od zastosowanego protokołu (<https://store.nanoporetech.com/eu/sample-prep.html>). W tym drugim przypadku cała procedura przygotowania biblioteki trwa zaledwie 10 min;

Nieobowiązkowość reakcji amplifikacji DNA. Amplifikacja DNA jest potrzebna w konkretnych przypadkach, kiedy ilość materiału wyjściowego jest czynnikiem limitującym lub gdy priorytetem jest wydajność sekwencjonowania, a nie długość odczytów. Jednak w przeciwieństwie do technik NGS, PCR nie jest niezbędnym etapem przygotowania

biblioteki do sekwencjonowania nanoporowego. Eliminując etap amplifikacji ograniczamy ryzyko błędów technicznych, np. wprowadzenia zmian w sekwencji czy nierównego namnożenia fragmentów DNA o różnej zawartości par GC. Przekłada się to na lepszą reprezentację stężeń transkryptów w analizach ekspresji genów oraz bardziej równomierne pokrycie odczytami sekwencjonowanych genomów [19];

Brak ograniczenia długości odczytów. Odczytywanie sekwencji DNA może trwać tak długo, jak długo aktywne jest białko nanoporowe [20]. Obecne wersje nanoporów pozwalają na sekwencjonowanie każdej nici w tempie ok. 400 nukleotydów na sekundę, a sam proces sekwencjonowania zazwyczaj trwa do 48 h. Oznacza to w praktyce brak technicznych ograniczeń co do długości sekwencji – jest ona determinowana wyłącznie długością przechodzącej przez nanopor nici DNA. W istocie, najdłuższy odnotowany dotąd odczyt nanoporowy, pochodzący od genomowego DNA ludzkiego wynosił 2272 580 nt [21]. Dla porównania, ta długość równa się połowie długości genomu *E. coli* albo ósmej części chromosomu 4 u rzodkiewnika (*Arabidopsis thaliana*);

Krótki czas sekwencjonowania. W przypadku technik NGS detekcja sygnału wymaga zatrzymania procesu amplifikacji po każdym cyklu i jest najbardziej czasochłonnym etapem sekwencjonowania. W sekwencjonowaniu nanoporowym detekcja sygnału zachodzi w czasie rzeczywistym, w momencie kontaktu DNA z nanoporem [11,22];

Skromne wymagania sprzętowe. Przygotowanie biblioteki sekwencyjnej wymaga wyłącznie standardowych urządzeń laboratoryjnych, takich jak termoblok czy aparat do analizy stężenia i długości DNA. W technice nanoporowej po raz pierwszy nie jest wymagany specjalistyczny sprzęt do odczytywania sygnałów fluorescencji. Sam sekwenator jest podłączany do komputera przez standardowy port USB 3.0, dlatego sekwencjonowanie można przeprowadzić nawet w warunkach polowych, używając laptopa wyposażonego w odpowiednią ilość pamięci operacyjnej i pojemny dysk SSD, przystosowany do bardzo szybkiego zapisu danych [23]. Większe wymagania (bardzo wydajny procesor graficzny) muszą być spełnione, jeżeli planowany jest równoczesny odczyt nukleotydów i podgląd wyników sekwencjonowania w czasie rzeczywistym. Oprogramowanie sekwenatora daje taką możliwość, jednak zwykle ten i kolejne etapy analiz bioinformatycznych przeprowadza się już po zakończeniu sekwencjonowania, na odpowiednio przystosowanych komputerach i serwerach;

Możliwość identyfikacji zmodyfikowanych nukleotydów. Jako że w sekwencjonowaniu nanoporowym wykrywanie nukleotydów zachodzi bezpośrednio, istnieje możliwość opracowania modeli detekcji, które pozwolą poszerzyć zakres wykrywanych zasad ponad cztery podstawowe. Na przykład, Wescoe i wsp. opracowali algorytm wykrywania pięciu wariantów epigenetycznie zmodyfikowanej cytozyny [24];

Możliwość bezpośredniego sekwencjonowania RNA. System nanoporowy umożliwia sekwencjonowanie cząstek RNA bezpośrednio, bez konieczności wcześniejszego etapu przepisania na sekwencję komplementarnego DNA

(cDNA). W połączeniu z udoskonalanymi ciągle algorytmami do odczytu nukleotydów daje to nadzieję na lepsze poznanie właściwości cząsteczek RNA, takich jak modyfikacje nukleotydów. Dodatkowo, możliwe jest odtworzenie pełnych sekwencji transkryptów na poziomie pojedynczych odczytów, co jest korzystne dla lepszego zrozumienia mechanizmów alternatywnego składania transkryptów (ang. alternative splicing) [25];

Skalowalność eksperymentu. Na pojedynczej płytce sekwencyjnej stosowanej w aparacie MinION mieści się 2 048 nanoporów. Są one zgrupowane po cztery, tworząc 512 niezależnych kanałów detekcji, z których jednocześnie pozyskiwane są informacje. W obrębie danego kanału nieaktywny nanopor może być zastąpiony innym, co pozwala na optymalizację czasu i wydajności sekwencjonowania. W jednym eksperymencie sekwencjonowania na urządzeniu MinION można wygenerować odczyty DNA o łącznej długości do 30 miliardów nt bądź do 12 milionów odczytów RNA. Na rynku dostępne są jednak również inne sekwencjatory i urządzenia pomocnicze, umożliwiające zwiększenie lub zmniejszenie skali eksperymentu (<https://store.nanoporetech.com/eu/devices>). Największym z nich jest obecnie PromethION 48. Działa on na tej samej zasadzie co MinION, ale może pomieścić do 48 niezależnych płytek sekwencyjnych. Każda z nich wyposażona jest w 3 000 kanałów i 12 000 nanoporów, i każda może jednorazowo wygenerować odczyty o łącznej długości pozwalającej na 20–40-krotne pokrycie genomu człowieka [26]. Do sekwencjonowania można jednocześnie użyć dowolnej liczby płytek, a ich ilość można zwiększać lub zmniejszać wg zapotrzebowania, w trakcie pracy sekwencjatora.

Jak każda technologia, sekwencjonowanie nanoporowe ma również swoje ograniczenia. Wydajność eksperymentu zależy od kilku czynników, przez co może być bardzo zmienna. Jednym z elementów ograniczających jest wyjściowa ilość DNA, przy czym zależy ona od jego długości – im wyższa masa cząsteczkowa DNA, tym więcej należy go użyć do sekwencjonowania. Dodatkowo długie cząsteczki DNA wolniej migrują w roztworze sekwencyjnym i rzadziej wchodzi w interakcje z nanoporem. Zatem należy się liczyć z tym, że – o ile nie zostanie przeprowadzona selekcja DNA pod kątem długości na etapie przygotowania biblioteki – w wyjściowym zbiorze danych większość odczytów będzie miała małą lub średnią długość. Charakterystyczną cechą eksperymentów nanoporowych jest również to, że duży odsetek takich krótkich odczytów przy sekwencjonowaniu bibliotek genomowego DNA ma bardzo niską jakość i powinien zostać odfiltrowany na etapie analizy (<https://nanoporetech.com/resource-centre/maize-genome-complexity-traversed-oxford-nanopore-technology>).

Kolejnym czynnikiem wpływającym na wydajność eksperymentu jest stan nanoporów w płytce sekwencyjnej. Ponieważ są to białka, ich żywotność jest ograniczona i zależy między innymi od czasu i warunków przechowywania płytki. Jedyną oceną oczekiwanej wydajności jest testowanie płytki pod kątem liczby aktywnych nanoporów w początkowym momencie sekwencjonowania. Dodatkowym utrudnieniem jest zjawisko zatykania nanoporów przez DNA, przez co stają się niedostępne dla kolejnych

cząsteczek. Oprogramowanie sekwencjatora ma zdolność wykrywania i odblokowywania takich nanoporów poprzez zastosowanie odwróconego prądu i „wycyfowanie” utkniętej cząsteczki DNA. Ten proces jest jednak daleko mniej skuteczny w przypadku sekwencjonowania długich cząsteczek, przez co wydajność sekwencjonowania genomowego DNA maleje wraz z czasem.

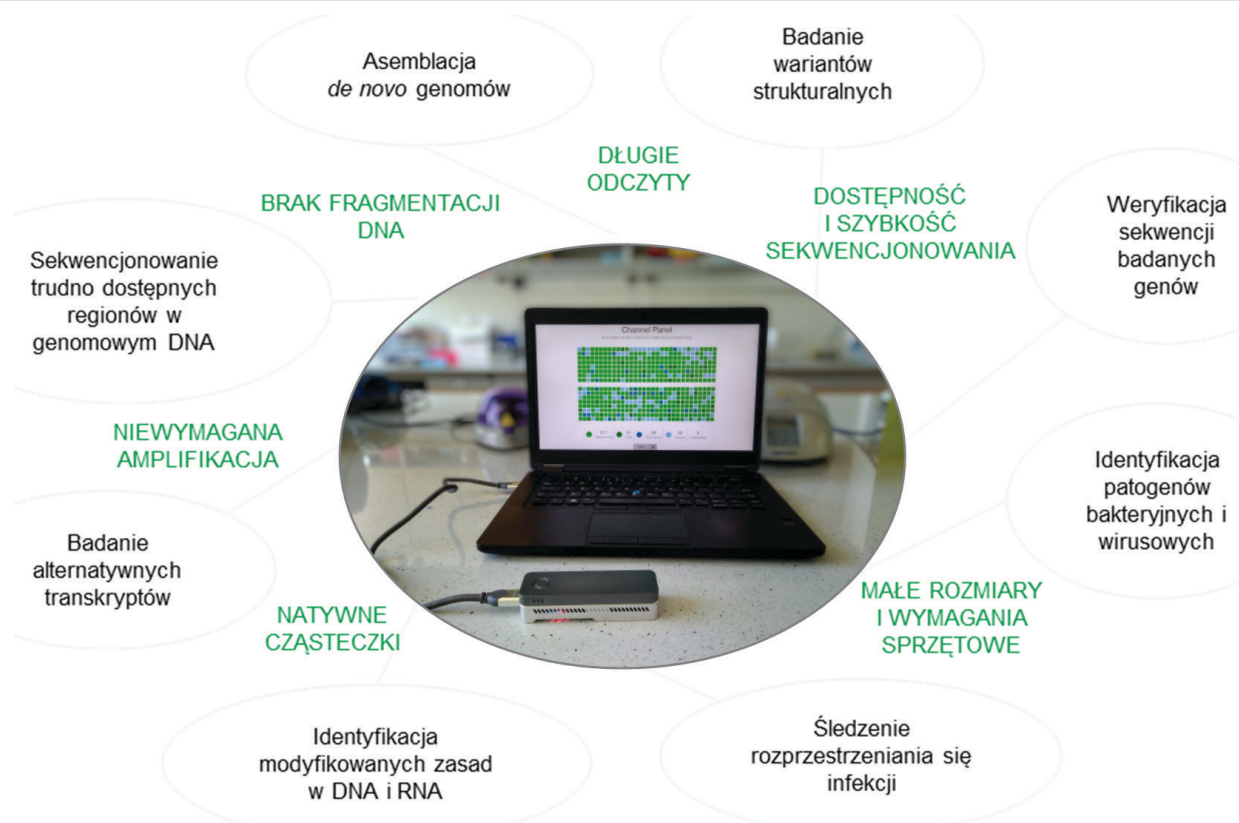
Ważnym ograniczeniem sekwencjonowania nanoporowego w porównaniu z technikami NGS pozostaje niższa dokładność pojedynczych odczytów, definiowana jako procent zgodności sekwencji nukleotydów w odczycie ze znaną referencją. Pierwsze sekwencjatory nanoporowe miały dokładność rzędu 66%. Błędy w odczytach nanoporowych zdarzają się stochastycznie i są głównie efektem szumu termodynamicznego w trakcie poruszania się cząsteczek w nanoporach [27]. Ze względu na losowość występowania tych błędów w sekwencji, dokładność wyników można znacząco poprawić, bazując na odpowiednio dużej liczbie odczytów (wysokim stopniu pokrycia genomu), z wykorzystaniem komputerowych algorytmów [28]. Często stosowaną metodą jest także wykorzystanie krótkich odczytów uzyskanych techniką NGS do poprawy jakości odczytów nanoporowych, na różnych etapach analizy [29]. Należy jednak zauważyć, że dzięki systematycznemu udoskonalaniu zarówno sekwencjatora jak i programów do analizy danych nanoporowych, a w szczególności zastosowanych nanoporów biologicznych, dokładność odczytów sekwencji DNA systematycznie rośnie i obecnie wynosi już 95% [30].

ZASTOSOWANIA SEKWENCJONOWANIA NANOPOROWEGO

Zastosowanie sekwencjonowania nanoporowego zniosło wcześniejsze ograniczenia nałożone przez niską długość odczytów uzyskanych technikami NGS, przyczyniając się do szybkiego rozwoju wielu dziedzin biologii i medycyny (Ryc. 4). Poniżej przedstawiono przykładowe osiągnięcia, obrazujące skalę tego postępu.

SEKWENCJONOWANIE I ASEMBLACJA GENOMÓW

Jednym z głównych zastosowań sekwencjonowania DNA pozostaje poznawanie nowych genomów. Odtworzenie sekwencji genomu wyłącznie na podstawie odczytów sekwencyjnych nazywa się asemblacją *de novo*. W dużym uproszczeniu polega ona na poszukiwaniu nakładających się odczytów i połączeniu ich w kontigi – ciągle sekwencje konsensusowe. Idealnie złożony kontig jest reprezentacją całego chromosomu. W praktyce, otrzymujemy pewną liczbę kontigów o różnej długości, których wzajemne ułożenie ustala się na kolejnych etapach asemblacji. Wynik asemblacji zależy od początkowej długości i jakości odczytów oraz rozmiaru i złożoności badanego genomu. Przykładowo, w genomach często obecne są ciągi powtórzeń, przekraczające długość pojedynczych odczytów. Takie struktury prowadzą do powstawania wielu alternatywnych wariantów połączeń kontigów o nakładających się końcach [31]. Pod tym względem najłatwiejsze do odtworzenia są sekwencje genomów bakteryjnych. Są one względnie małe, mają formę chromosomu kolistego, a długość ich regionów powtórzonych sięga najwyżej 7 000 par zasad, co leży w zakresie



Rycina 4. Właściwości i zastosowania sekwencjonowania nanoporowego. Zdjęcie przedstawia sekwenator MinION połączony ze sterującym nim laptopem.

długości odczytów trzeciej generacji. Po raz pierwszy asemblację genomu bakteryjnego wyłącznie z długich odczytów wykonano techniką SMRT w 2013 roku [32], a z odczytów nanoporowych – dwa lata później [28].

Asemblacja *de novo* umożliwiła odtworzenie sekwencji genomów eukariotycznych wielu nowych gatunków. W przypadku dużych i złożonych genomów, asemblację wykonane w oparciu o długie odczyty są udokładniane i korygowane przy pomocy danych dodatkowych, pochodzących z sekwencjonowania NGS, badania fizycznych interakcji chromatyny, informacji o położeniu markerów genetycznych (dane cytogenetyczne) czy rozkładzie miejsc restrykcyjnych w genomie (mapowanie optyczne) [29,33]. Pod względem złożoności asemblacji, genomy roślinne stanowią szczególne wyzwanie, z uwagi na dużą rozpiętość ich rozmiarów, stopień ploidalności oraz olbrzymią zawartość regionów powtórzonych i transpozonów [34–36]. Dzięki sekwencjonowaniu długich odczytów uzyskano bardziej kompletne i ciągłe wersje szeregu genomów roślinnych – rzodkiewnika, sorgo, pomidora, kapusty czy banana [29,33,37,38]. Te osiągnięcia stwarzają realną perspektywę poznania bogactwa genetycznego organizmów eukariotycznych na Ziemi, czego przejawem jest zainicjowanie w 2018 r. projektu o nazwie Earth Biogenome Project, którego celem jest zsekwencjonowanie jak największej ilości gatunków w ciągu dziesięciu lat [39].

Ze względu na swoją długość, odczyty nanoporowe dają możliwość wypełnienia przerw istniejących w wielu genomach referencyjnych, np. w genomie człowieka. Pierwsza, mocno pofragmentowana wersja tego genomu została opu-

blikowana w 2001 roku, w wyniku realizacji ogromnego projektu, który trwał ponad dziesięć lat. W kolejnych latach ta wersja była udokładniana, dzięki czemu poznano skład i strukturę ludzkich genów, mechanizmy ich ekspresji i regulacji, jak również odkryto genetyczne przyczyny wielu chorób oraz opracowano nowe sposoby ich diagnostyki i leczenia. Jednak nawet po dwudziestu latach od momentu opublikowania, genom referencyjny człowieka wciąż zawiera przerwy w regionach trudnych do zsekwencjonowania, takich jak tandemowe powtórzenia. Niedawno grupa badaczy tworzących konsorcjum T2T (ang. *telomere-to-telomere*, czyli „od telomeru do telomeru”) przedstawiła sekwencję genomu ludzkiej linii komórkowej CHM13, złożonego z połączenia odczytów nanoporowych, odczytów SMRT oraz danych uzyskanych innymi technikami [40]. Bazując na tej asemblacji, badacze wykonali następnie manualne korekty i eksperymenty walidacyjne, uzyskując w efekcie pełną sekwencję chromosomu X. Zlikwidowali tym samym 29 dotychczasowych przerw w sekwencji tego chromosomu, obejmujących łącznie ponad milion nukleotydów.

BADANIE ZMIENNOŚCI STRUKTURALNEJ

Zmianami strukturalnymi określa się zmiany w sekwencji DNA obejmujące duże obszary genomu (zwykle kilkaset par zasad lub więcej). Do tego opisu pasuje wiele różnych wariantów, takich jak zmiany liczby kopii fragmentów DNA (duplikacje i delecje), inwersje, translokacje wewnątrz- i międzychromosomalne, insercje mobilnych elementów genetycznych, zmiany liczby powtórzeń tandemowych, czy wydłużanie ciągów satelitarnego DNA. Wobec tej różnorodności, zmienność strukturalna genomów jest szczegól-

nie trudnym do badania zjawiskiem. Dlatego dla poznania skomplikowanych rearanżacji genomowych korzystne jest zastosowanie dłuższych odczytów, o bardziej unikalnych sekwencjach. Długie odczyty umożliwiają również wykorzystanie asemblacji *de novo* w badaniu i diagnostyce schorzeń genetycznych wynikających ze zmian strukturalnych DNA, podczas gdy analizy z wykorzystaniem danych NGS bazują przede wszystkim na poszukiwaniu niezgodności w mapowaniu krótkich odczytów do genomu referencyjnego. Przyczyną niektórych schorzeń może być zjawisko katastrofy chromosomalnej (ang. *chromothripsis*), które polega na silnym pofragmentowaniu jednego lub kilku chromosomów, a następnie ich rearanżacji, prowadzących do takich zmian jak mutacje genowe czy powstawanie genów fuzyjnych [41]. W przypadku tak skomplikowanych przedstawień, sekwencjonowanie długich odczytów jest niemal jedyną możliwością dokładnego ich zbadania, dlatego jest coraz częściej stosowane. W porównaniu z techniką NGS, pozwala ono na wykrycie większej liczby rearanżacji, przy mniejszym pokryciu genomu oraz zmniejsza ryzyko błędnego określenia miejsca ich występowania [42]. Dodatkową zaletą wynikającą z zastosowania technologii nanoporowej jest także ułatwienie fazowania (ang. *phasing*), czyli prawidłowego rozdzielania alleli pochodzących z chromosomu matczyngo i ojcowskiego w DNA heterozygoty.

Całogenomowe sekwencjonowanie DNA metodą długich odczytów wykorzystano również do badania translokacji zrównoważonych i inwersji u człowieka. Podczas mejozy obecność tego typu rearanżacji może skutkować nieprawidłową segregacją chromosomów, a w konsekwencji nieprawidłowym wykształceniem gamet, objawiającym się zaburzeniami płodności u nosiciela. Tradycyjnie translokacje i inwersje bada się za pomocą oznaczania kariotypu lub – wykazującej nieco wyższą rozdzielczość – hybrydyzacji fluorescencyjnej *in situ*. Żadna z tych metod nie daje szczegółowego wglądu w sekwencję regionów granicznych rearanżacji, przez co trudno jest ocenić jej możliwy wpływ na strukturę genów pokrywających się z tymi regionami. Hu i wsp. przebadali genomy sześciu osób z potwierdzonymi translokacjami zrównoważonymi i jednej osoby z inwersją, stosując sekwencjonowanie nanoporowe oraz mapowanie odczytów do genomu referencyjnego [43]. W efekcie prawidłowo zidentyfikowali wszystkie rearanżacje, a dodatkowo określili ich koordynaty genomowe. Dzięki temu w czterech przypadkach wykryto nieujawnione wcześniej rozbieżności struktury genów leżących na granicy obszaru rearanżacji, poprzez ich częściową translokację. Wskazuje to na perspektywę zastosowania sekwencjonowania nanoporowego do klinicznej diagnostyki translokacji zrównoważonych, przy stosunkowo niewielkim pokryciu genomu, a więc niskich kosztach.

Jeżeli w kontekście genomów ludzkich badania zmienności strukturalnej najczęściej mają podłoże medyczne, to w przypadku roślin, zmienność strukturalna genomów analizowana jest pod kątem roli w przystosowaniu do nowych warunków środowiskowych, adaptacji na stres, procesów ewolucyjnych czy ważnych cech użytkowych. Doskonałym przykładem obrazującym korzyści z zastosowania długich odczytów w badaniach genomów roślinnych, jest kompleksowa praca opisująca asemblację *de novo* i porównanie se-

kwencji genomów dwóch odmian ryżu (*Oryza sativa*): basmati (linia Basmati 334 z Pakistanu) i sadri (linia Dom Sufid z Iranu) [44]. Obie odmiany mają wysokie walory smakowe i reprezentują ostatnią i najmniej poznaną spośród czterech głównych ewolucyjnych grup ryżu: japońska, indica, aus i circum-basmati. Autorzy wspomnianej pracy złożyli genomy referencyjne dla obu odmian na podstawie odczytów nanoporowych, uzyskując asemblacje o wysokiej jakości i ciągłości. Następnie porównali je między sobą oraz ze znanymi genomami innych grup ryżu, identyfikując warianty strukturalne, specyficzne dla badanych odmian, bądź całej grupy circum-basmati, takie jak polimorfizmy pojedynczego nukleotydu, ekspansje sekwencji retrotranspozonów oraz insercje i delekcje genów. Oprócz badań kompleksowych, odczyty nanoporowe stosuje się do szybkiej analizy sekwencji konkretnych regionów w genomie. Debladis i wsp. zbadali metodą ilościowego PCR aktywność transkrypcyjną retrotranspozonu Evadé w rekombinacyjnej linii epiRIL12 rzodkiewnika, wykazującej obniżony poziom metylacji genomu [45]. Wyniki analizy wskazywały na obecność dodatkowych kopii Evadé w genomie. Aby potwierdzić te przypuszczenia, autorzy przeprowadzili sekwencjonowanie nanoporowe genomu epiRIL12, a poszczególne odczyty porównywali z sekwencją genomu referencyjnego. W ten sposób, bez długiej i skomplikowanej analizy danych, a jedynie dzięki uzyskaniu odpowiednio długich odczytów, obejmujących zarówno sekwencję transpozonu jak i miejsc flankujących, autorzy wykryli i zlokalizowali w genomie 9 nowych kopii Evadé. Zatem, dzięki niskim nakładom kosztów, sekwencjonowanie nanoporowe może służyć jako względnie prosty sposób weryfikacji hipotez, sformułowanych na podstawie wcześniejszych badań.

W badaniach ograniczonych do konkretnych regionów DNA, sekwencjonowanie całego genomu nie jest wymagane. Bardziej odpowiednim podejściem jest sekwencjonowanie ukierunkowane biblioteki DNA wzbogaconej w interesujące fragmenty genomu. Techniki wzbogacenia DNA często wykorzystywane są w badaniach medycznych i diagnostyce, gdzie ważne jest wykrycie i określenie rodzaju mutacji w jednym genie lub zestawie genów, u wielu indywidualnych pacjentów. Sekwencjonowanie ukierunkowane bardzo dobrze sprawdziło się również w badaniu zmienności genów odporności R (ang. *resistance genes*) u psianki *Solanum americanum* – dziko rosnącej rośliny z rodziny psiankowatych, wykazującej dużą odporność na patogena *Phytophthora infestans*, wywołującego m. in. zarazę ziemniaka [46, 47, 48]. W genomach roślin geny R o prawie identycznych sekwencjach często są ułożone w klastry, co utrudnia odtworzenie sekwencji poszczególnych genów w oparciu o krótkie odczyty NGS. W przypadku psianki przygotowano wzbogaconą bibliotekę DNA poprzez „wychwycenie”, a następnie amplifikację fragmentów genomowego DNA zawierających sekwencje genów R. Do scharakteryzowania tak wzbogaconej biblioteki wykorzystano sekwencjonowanie trzeciej generacji (nanoporowe i/lub SMRT) oraz techniki asemblacji *de novo*. Pozwoliło to między innymi na zidentyfikowanie i sklonowanie genu Rpi-amr3i, odpowiedzialnego za odporność psianki na *P. infestans* [48]. Z kolei metoda przygotowania biblioteki sekwencyjnej polegająca na selektywnym wycięciu interesujących fragmentów genomu z użyciem enzymu Cas9, umożliwia przeprowadzenie

wzbogacenia bezpośrednio na natywnym DNA [49]. Zatem, sekwencjonowanie ukierunkowane jest dobrym sposobem na zwiększenie pokrycia odczytami regionów pożądaných i uproszczenie analiz, wraz ze zmniejszeniem kosztów badania.

ANALIZA TRANSKRYPTOMU

Możliwość badania transkryptomów przez sekwencjonowanie pozwala na ilościową i porównawczą ocenę poziomów ekspresji genów, identyfikację produktów alternatywnego składania transkryptów oraz identyfikację nowych niekodujących RNA. Zastosowanie do tego celu technik NGS wiąże się z koniecznością konwersji RNA na cDNA oraz – podobnie jak w przypadku analizy genomowego DNA – wymusza składanie końcowych sekwencji z krótkich odczytów. Bioinformatyczne przewidywanie alternatywnych transkryptów na podstawie odczytów NGS może generować niepoprawne złożenia lub pomijać transkrypty rzadko występujące w badanym materiale [50]. Znamienny jest fakt, że zestawy alternatywnych transkryptów uzyskane przez różne programy, działające w oparciu o odczyty NGS, znacząco różnią się między sobą [51]. Wpływa to również na precyzję na dalszych etapach oceny. W tym kontekście oczywiste korzyści niesie ze sobą możliwość odtworzenia całej sekwencji transkryptu w pojedynczym odczycie. Zastosowanie sekwencjonowania trzeciej generacji pozwala na identyfikację większej ilości alternatywnych form transkryptów niż NGS, nawet przy niskiej głębokości pokrycia [52].

Unikalną zaletą sekwencjonowania nanoporowego jest możliwość bezpośredniego sekwencjonowania cząsteczek RNA [53]. Mimo że podczas przygotowania biblioteki do jednoniciowych transkryptów RNA dobudowywana jest nić cDNA, to służy ona tylko utrzymaniu stabilności cząsteczki, ale nie zostaje poddana sekwencjonowaniu. Wśród coraz liczniejszych przykładów zastosowania bezpośredniego sekwencjonowania RNA warto wymienić analizę porównawczą transkryptomu modelowego nicienia *Caenorhabditis elegans* w trzech stadiach rozwojowych: embrionu, larwy i dorosłego osobnika [54]. W wyniku sekwencjonowania uzyskano odczyty o średniej długości około 1 000 nt, przy czym blisko połowa z nich reprezentowała transkrypty pełnej długości. Analiza ich sekwencji pozwoliła na zidentyfikowanie większości dotychczas poznanych transkryptów alternatywnych, a ponadto ujawniła łącznie 57 000 nowych izoform, reprezentujących prawie 12 000 genów. Co ciekawe, zupełnie inne geny wykazywały różnicowanie pomiędzy badanymi stadiami rozwojowymi pod względem ilościowym (poziom ekspresji), a inne pod względem jakościowym (różne izoformy). Sugeruje to istnienie bardzo subtelnych mechanizmów regulacji poprzez alternatywne składanie transkryptów.

Warto zauważyć, że bezpośrednie sekwencjonowanie RNA jest metodą nową i jej zastosowanie nadal wiąże się z szeregiem ograniczeń. Do tych ograniczeń należą obniżona wydajność sekwencjonowania, jak również niższa jakość i dokładność (około 85%) pojedynczych odczytów RNA w porównaniu z sekwencjonowaniem cDNA. Wynika to przede wszystkim z mniejszej stabilności cząsteczek RNA i odpowiednio większego szumu podczas detekcji. Niemniej

jest to jedyna metoda analizy sekwencji RNA w natywnej postaci, dająca również wgląd w charakter modyfikacji chemicznych tych cząsteczek.

BADANIE MODYFIKACJI CHEMICZNYCH W DNA I RNA

Chemiczne modyfikacje zasad naturalnie występują w kwasach nukleinowych. Dotychczas opisano blisko 20 rodzajów modyfikacji DNA w genomach bakteryjnych i eukariotycznych oraz kilkakrotnie więcej rodzajów modyfikacji RNA [55,56]. Najlepiej poznany i najczęściej badany typem modyfikacji zasad w genomowym DNA jest metylacja – dynamiczny proces, modyfikujący strukturę chromatyny na różnych etapach rozwoju organizmu oraz pod wpływem działania czynników stresowych. Zmieniając dostępność chromatyny dla czynników transkrypcyjnych i modyfikując ich zdolność wiązania do danej sekwencji, metylacje kształtują profil ekspresji genów w komórkach. W kodujących i niekodujących RNA, występowanie różnorodnych modyfikacji jest często kluczowe dla ich stabilności i prawidłowego działania. Zatem poznanie modyfikacji zasad w DNA i RNA ma istotne znaczenie.

Metylowane nukleotydy są niewykrywalne za pomocą konwencjonalnego sekwencjonowania NGS, ponieważ na skutek amplifikacji zostają zastąpione niemetylowanymi odpowiednikami, przez co tracona jest informacja o ich występowaniu [53,57]. Dlatego niezbędne jest zastosowanie specyficznych podejść, takich jak sekwencjonowanie po reakcji z wodorosiarczynem sodu (ang. *sodium bisulfite sequencing*). Metoda ta polega na przekształceniu niemetylowanej cytozyny na uracyl, a następnie amplifikacji, podczas której uracyl jest zamieniany na tyminę. Porównując odczyty sekwencyjne z genomem referencyjnym można odróżnić miejsca metylowane od niemetylowanych. Z kolei wykrywanie modyfikacji RNA bazuje na immunoprecypitacji z wykorzystaniem przeciwciał, rozpoznających metylowane nukleotydy, jednak specyficzność takiego rozpoznawania nie jest doskonała. Obie te metody są czasochłonne i specyficzne dla konkretnych rodzajów modyfikacji.

Rozwój sekwencjonowania trzeciej generacji przyniósł znaczący postęp w badaniach modyfikacji chemicznych zasad w DNA i RNA. Technika SMRT daje ograniczoną możliwość ich detekcji, na podstawie zmian zachowania polimerazy podczas syntezy, wykrywalnych jako wydłużenie czasu trwania odczytu fluorescencyjnego. Z kolei sekwencjonowanie nanoporowe umożliwia bezpośrednią detekcję modyfikowanych zasad w nanoporze i ich jednoznaczną identyfikację za pomocą algorytmów obliczeniowych. Obecnie stosowane algorytmy do analizy odczytów nanoporowych są już w stanie rozpoznać kilka rodzajów metylowanych zasad w genomowym DNA i RNA. Wykorzystano je między innymi do analizy rozkładu metylowanych cytozyn w obrębie chromosomu 20 genomu człowieka [30]. Uzyskane wyniki wykazywały wysoką zgodność z wcześniejszymi wynikami pochodzącymi z sekwencjonowania po reakcji z wodorosiarczynem sodu. Podobnie, we wspomnianej wcześniej pracy dotyczącej badania transkryptomu *C. elegans* na różnych etapach rozwoju, informację o transkryptach połączono z identyfikacją i analizą obecnych w nich modyfikacji chemicznych. Wykazano między innymi, że regiony kodujące transkryp-

tów zawierają więcej modyfikacji niż regiony niekodujące, co sugeruje ich wpływ na regulację procesu translacji [54].

DIAGNOSTYKA PATOGENÓW BAKTERYJNYCH I WIRUSOWYCH

Korzyści z zastosowania sekwencjonowania nanoporowego dla opisanych wyżej aplikacji wynikają przede wszystkim z istotnego zwiększenia długości odczytów, w porównaniu z wcześniejszymi technikami sekwencjonowania. Z kolei w przypadku diagnostyki i monitorowania zakażeń wirusowych kluczowe okazują się prostota obsługi sekwenatora MinION i łatwość jego transportu oraz krótki czas uzyskania wyników. Możliwość sekwencjonowania „na miejscu” próbek pobranych od pacjentów, nawet w przypadku, gdy ogniska choroby są oddalone od dużych ośrodków naukowych i centrów diagnostycznych, zwiększa szanse na podjęcie skutecznej terapii i ograniczenie rozprzestrzeniania się choroby. Dodatkowo, szybkie uzyskanie sekwencji genomu wirusa podczas wybuchu epidemii ma znaczenie dla wyjaśnienia tempa i kierunku jego ewolucji oraz badania łańcuchów transmisji. Jako pierwszy te korzyści ocenił zespół Lomana z Uniwersytetu Birmingham podczas epidemii wirusa Ebola w Afryce w 2015 r., organizując w Gwinei polowe laboratorium sekwencjonowania [23]. W czasie funkcjonowania laboratorium przeanalizowano 142 próbki od zakażonych pacjentów. Typowy czas przygotowania próbki i jej analizy wynosił 2 dni, przy czym w większości przypadków już po 1 godzinie sekwencjonowania otrzymywano ilość odczytów wystarczającą do potwierdzenia obecności wirusa. Opracowany wówczas protokół amplifikacji i sekwencjonowania wirusowego RNA zastosowano następnie w mobilnym laboratorium genomicznym, badającym transmisję wirusa Zika w krajach Ameryki Północnej i Południowej [58,59].

Szybka diagnostyka infekcji jest również istotna dla rolnictwa, znacząco wpływając na ograniczenie strat w uprawach wywoływanych chorobami wirusowymi. Potencjał sekwencjonowania nanoporowego do detekcji i identyfikacji wirusów porażających plantacje manioku w Tanzanii, Ugandzie i Kenii zaprezentowali Boykin i wsp., działający w ramach akcji Cassava Virus Action Project [60]. Dzięki zastosowaniu przenośnego urządzenia do ekstrakcji kwasów nukleinowych oraz aparatu MinION, byli oni w stanie przeprowadzać analizy bezpośrednio na farmach. Wyniki uzyskiwali w ciągu 3 godzin, a czułość systemu pozwalała na detekcję wirusów nawet w roślinach nie wykazujących jeszcze objawów infekcji.

Inna grupa pokazała skuteczność metody nanoporowej w identyfikacji wirusów i bakterii, zarówno w tkankach roślin jak i owadów stanowiących wektory dla patogenów, poprzez analizę całkowitego transkryptomu [61]. Badania prowadzono na układzie modelowym typu patogen-gospodarz (wirus ospowatości śliwy PPV – brzoskwinia *Prunus persica*) oraz dwóch układach modelowych typu patogen-wektor (wirus ospowatości śliwy PPV – mszyca brzoskwiniowo-ziemniaczana *Myzus persicae* i gram-ujemna bakteria *Candidatus Liberibacter asiaticus* – miódówka *Diaphorina citri*). Wyekstrahowany RNA był poddawany amplifikacji połączonej z przepisaniem na cDNA i sekwencjonowany na aparacie MinION przez 24 godziny. Chociaż w każdym badanym układzie

odsetek odczytów reprezentujących poszukiwane patogeny był niski (5-7%), to ich identyfikacja poprzez porównanie z bazami danych genomowych była jednoznaczna, a pierwsze odczyty reprezentujące te patogeny uzyskano już po kilkadziesiąt sekund od rozpoczęcia sekwencjonowania. Dodatkowo, w próbkach owadów zidentyfikowano odczyty pochodzące z bakterii endosymbiotycznych.

Biorąc pod uwagę skalę zastosowań sekwencjonowania nanoporowego, nie mogło wśród nich zabraknąć badań nad wirusem SARS-CoV-2 i przebiegiem wywołanej przez niego pandemii choroby COVID-19, która w 2020 r. w ciągu zaledwie kilku miesięcy rozprzestrzeniła się praktycznie na cały świat. COVID-19 to choroba układu oddechowego, która najczęściej ma łagodny bądź bezobjawowy przebieg, jednak w niektórych przypadkach, szczególnie przy chorobach współistniejących, może doprowadzić do ostrej niewydolności oddechowej i śmierci. Dlatego od momentu opublikowania pierwszego genomu wirusa SARS-CoV-2 rozpoczęły się intensywne badania nad poznaniem mechanizmów jego transmisji. SARS-CoV-2 należy do rodziny *Coronaviridae*, o genomie złożonym z jednoniciowej cząsteczki RNA o dodatniej polaryzacji, a więc będącej jednocześnie jego transkryptem. Rozmiar genomu SARS-CoV-2 wynosi prawie 30 000 nt, co mieści się w zakresie odczytów nanoporowych. Konsekwentnie, tylko w okresie od marca do końca maja 2020 r. w bazie bibliograficznej Pubmed oraz repozytoriach MedRxiv i BiorXiv pojawiło się ponad 30 doniesień naukowych, gdzie sekwencjonowanie nanoporowe zostało wykorzystane w badaniach tego wirusa. Spośród nich warto wymienić pracę, w której badano skład i strukturę transkryptomu wirusa, stosując bezpośrednio sekwencjonowanie RNA [62]. Doprowadziło to między innymi do odkrycia fuzyjnych transkryptów, kodujących nowe sekwencje potencjalnych białek. Zidentyfikowano również miejsca i rodzaje modyfikacji chemicznych transkryptomu. To i inne badania genomu wirusa SARS-CoV-2, stanowią punkt wyjścia do szczegółowego poznania mechanizmów patogenezы COVID-19 i otwierają nowe kierunki do opracowania terapii tej choroby.

PRZYSZŁOŚĆ TECHNOLOGII NANOPOROWEJ

Metody sekwencjonowania trzeciej generacji spowodowały olbrzymi postęp technologiczny. W szczególności sekwencjonowanie nanoporowe, od czasu jego wprowadzenia, udoskonalono w niemal każdym aspekcie działania, zaczynając od zwiększenia stabilności białka nanoporowego, poprzez ulepszenie komponentów systemu sekwencyjnego i kończąc na opracowaniu coraz bardziej dokładnych algorytmów obliczeniowych. Długie odczyty nanoporowe całkowicie zmieniły możliwości sekwencjonowania i asembblacji skomplikowanych genomów, z dużych i kosztownych projektów badawczych na eksperymenty możliwe do przeprowadzenia w małych zespołach. Umiarkowany koszt sprzętu, „kieszonkowe” rozmiary sekwenatora MinION i szybkość przygotowania biblioteki sekwencyjnej umożliwiły przeniesienie badań dotyczących poznania żywych gatunków bezpośrednio na miejsce ich występowania w naturze. Ponadto, dając możliwość sekwencjonowania natywnych cząsteczek, jest to obecnie jedyna dostępna metoda bezpośredniej analizy sekwencji RNA.

Dzięki swoim unikalnym właściwościom technika nanoporowa ma szansę stać się w niedalekiej przyszłości podstawą ekspresowych testów diagnostycznych w medycynie spersonalizowanej, dostępnej dla każdego pacjenta. Z drugiej strony, ponieważ systemy nanoporowe wykazują wysoką selektywność przy identyfikacji zasad w kwasach nukleinowych, w przyszłości ich wykorzystanie można rozszerzyć do detekcji innych cząsteczek, takich jak białka czy metabolity. Takie próby były już podejmowane równoległe z pracami nad sekwencjonowaniem DNA, ale – jak dotąd – z mniejszym sukcesem. Niemniej, biorąc pod uwagę potencjał i obecną skalę zastosowań sekwencjonowania nanoporowego należy oczekiwać, że w najbliższych latach będzie ono jedną z wiodących technik biologii molekularnej.

PIŚMIENNICTWO

1. Heather JM, Chain B (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* 107: 1-8
2. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* 74: 5463-5467
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. (2001) The sequence of the human genome. *Science* 291: 1304-1351
5. Metzker ML (2010) Sequencing technologies – the next generation. *Nat Rev Genet* 11: 31-46
6. Werner T (2010) Next generation sequencing in functional genomics. *Brief Bioinform* 11: 499-511
7. Żmienieko A, Jackowiak P, Figlerowicz M (2011) Transcriptome sequencing: next generation approach to RNA functional analysis. *BioTechnology* 4: 311-319
8. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C (2018) The Third Revolution in Sequencing Technology. *Trends Genet* 34: 666-681
9. Rhoads A, Au KF (2015) PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* 13: 278-289
10. Wang Y, Yang Q, Wang Z (2014) The evolution of nanopore sequencing. *Front Genet* 5: 449
11. Deamer D, Akeson M, Branton D (2016) Three decades of nanopore sequencing. *Nat Biotechnol* 34: 518-524
12. Kasianowicz JJ, Brandin E, Branton D, Deamer DW (1996) Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A* 93: 13770-13773
13. Akeson M, Branton D, Kasianowicz JJ, Brandin E, Deamer DW (1999) Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophys J* 77: 3227-3233
14. Caldwell CC, Spies M (2017) Helicase SPRNting through the nanopore. *Proc Natl Acad Sci U S A* 114: 11809-11811
15. Wick RR, Judd LM, Holt KE (2019) Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20: 129
16. de Lannoy C, de Ridder D, Risse J (2017) The long reads ahead: de novo genome assembly using the MinION. *F1000Research* 6: 1083
17. Goyal P, Krasteva PV, Van Gerven N, Gubellini F, Van den Broeck I, Troupiotis-Tsaïlaki A, Jonckheere W, Péhau-Arnaudet G, Pinker JS, Chapman MR, Hultgren SJ, Howorka S, Fronzes R, Remaut H (2014) Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* 516: 250-253
18. Carter J-M, Hussain S (2018) Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. *Wellcome Open Res* 2: 23
19. Lu H, Giordano F, Ning Z (2016) Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* 14: 265-279
20. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS (2018) Long reads: their purpose and place. *Hum Mol Genet* 27: R234-R241
21. Payne A, Holmes N, Rakyan V, Loose M (2019) BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35: 2193-2198
22. Stephens Z, Wang C, Iyer RK, Kocher J-P (2018) Detection and visualization of complex structural variants from long reads. *BMC Bioinformatics* 19: 508
23. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. (2016) Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530: 228-232
24. Wescoe ZL, Schreiber J, Akeson M (2014) Nanopores discriminate among five C5-cytosine variants in DNA. *J Am Chem Soc* 136: 16582-16587
25. Drexler HL, Choquet K, Churchman LS (2020) Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol Cell* 77: 985-998
26. De Coster W, De Rijk P, De Roeck A, De Pooter T, D'Hert S, Strazisar M, Slegers K, Van Broeckoven C (2019) Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res* 29: 1178-1187
27. Mikheyev AS, Tin MMY (2014) A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour* 14: 1097-1102
28. Loman NJ, Quick J, Simpson JT (2015) A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* 12: 733-735
29. Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker JR, et al. (2018) High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. *Nat Commun* 9: 541
30. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 36: 338-345
31. Koren S, Phillippy AM (2015) One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol* 23: 110-120
32. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10: 563-569
33. Hoang PNT, Michael TP, Gilbert S, Chu P, Motley ST, Appenroth KJ, Schubert I, Lam E (2018) Generating a high-confidence reference genome map of the Greater Duckweed by integration of cytogenomic, optical mapping, and Oxford Nanopore technologies. *Plant J* 96: 670-684
34. Dodsworth S, Leitch AR, Leitch IJ (2015) Genome size diversity in angiosperms and its influence on gene space. *Curr Opin Genet Dev* 35: 73-78
35. Kyriakidou M, Tai HH, Anglin NL, Ellis D, Strömvik MV (2018) Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front Plant Sci* 9: 1660
36. Schatz MC, Witkowski J, McCombie WR (2012) Current challenges in de novo plant genome sequencing and assembly. *Genome Biol* 13: 243
37. Jiao W-B, Schneeberger K (2017) The impact of third generation genomic technologies on plant genome assembly. *Curr Opin Plant Biol* 36: 64-70
38. Belsler C, Istace B, Denis E, Dubarry M, Baurens F-C, Falentin C, et al. (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat Plants* 4: 879-887
39. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. (2015) Earth BioGenome Project: Sequencing life for the future of life. *Proc Natl Acad Sci U S A* 115: 4325-4333
40. Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, et al. (2020) Telomere-to-telomere assembly of a complete human X chromosome. *Nature* Jul. 14. doi: 10.1038/s41586-020-2547-7

41. Koltsova AS, Pendina AA, Efimova OA, Chiryayeva OG, Kuznetsova T V, Baranov VS (2019) On the Complexity of Mechanisms and Consequences of Chromothripsis: An Update. *Front Genet* 10: 393
42. Cretu Stancu M, van Roosmalen MJ, Renkens I, Nieboer MM, Middelekamp S, de Ligt J, Pregno G, Giachino D, Mandrile G, Valle-Inclan JE, Korzelius J, de Bruijn E, Cuppen E, Talkowski ME, Marschall T, de Ridder J, Kloosterman WP (2017) Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun* 8: 1326
43. Hu L, Liang F, Cheng D, Zhang Z, Yu G, Zha J, Wang J, Xia Q, Juan D, Tan Y (2020) Location of Balanced Chromosome-Translocation Breakpoints by Long-Read Sequencing on the Oxford Nanopore Platform. *Front Genet* 10:1313
44. Choi JY, Lye ZN, Groen SC, Dai X, Rughani P, Zaaier S, Harrington ED, Juul S, Purugganan MD (2020) Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. *Genome Biol* 21: 21
45. Debladis E, Llauro C, Carpentier M-C, Mirouze M, Panaud O (2017) Detection of active transposable elements in *Arabidopsis thaliana* using Oxford Nanopore Sequencing technology. *BMC Genomics* 18: 537
46. Minervini CF, Cumbo C, Orsini P, Anelli L, Zagaria A, Specchia G, Albano F (2020) Nanopore Sequencing in Blood Diseases: A Wide Range of Opportunities. *Front Genet* 11:76
47. Giolai M, Paajanen P, Verweij W, Witek K, Jones JDG, Clark MD (2017) Comparative analysis of targeted long read sequencing approaches for characterization of a plant's immune receptor repertoire. *BMC Genomics* 18: 564
48. Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG (2016) Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat Biotechnol* 34: 656-660
49. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ, Timp W (2020) Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat Biotechnol* 38: 433-438
50. Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nat Rev Genet* 12: 671-682
51. Steijger T, Abril JF, Engström PG, Kokocinski F, Hubbard TJ, Guigó R, Harrow J, Bertone P (2013) Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods* 10: 1177-1184
52. Li Y, Dai C, Hu C, Liu Z, Kang C (2017) Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. *Plant J* 90: 164-176
53. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, et al. (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 15: 201-206
54. Li R, Ren X, Ding Q, Bi Y, Xie D, Zhao Z (2020) Direct full-length RNA sequencing reveals unexpected transcriptome complexity during *Caenorhabditis elegans* development. *Genome Res* 30: 287-298
55. Raiber E-A, Hardisty R, van Delft P, Balasubramanian S (2017) Mapping and elucidating the function of modified bases in DNA. *Nat Rev Chem* 1: 0069
56. Helm M, Motorin Y (2017) Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat Rev Genet* 18: 275-291
57. Xu L, Seki M (2020) Recent advances in the detection of base modifications using the Nanopore sequencer. *J Hum Genet* 65: 25-33
58. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. (2017) Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 12: 1261-1276
59. Faria NR, Quick J, Claro IM, Théze J, de Jesus JG, Giovanetti M, et al. (2017) Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546: 406-410
60. Shaffer L (2019) Inner Workings: Portable DNA sequencer helps farmers stymie devastating viruses. *Proc Natl Acad Sci U S A* 116: 3351-3353
61. Bronzato Badial A, Sherman D, Stone A, Gopakumar A, Wilson V, Schneider W, King J (2018) Nanopore Sequencing as a Surveillance Tool for Plant Pathogens in Plant and Insect Tissues. *Plant Dis* 102: 1648-1652
62. Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H (2020) The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181: 914-921

Nanopore sequencing and its application in biology

Anastasiia Satyr, Agnieszka Żmieńko 

Department of Molecular and Systems Biology, Institute of Bioorganic Chemistry Polish Academy of Sciences, Poznan

 e-mail: akisiel@ibch.poznan.pl

Key words: genome sequencing, nanopore technology, long reads, *de novo* assembly, structural variation, transcriptome research, base modifications, pathogen diagnosis

SUMMARY

Nanopore sequencing is a third generation sequencing technique. It involves the electrophoretic transport of nucleic acids through the protein channels of nanometer size, called nanopores, followed by deciphering their nucleotide sequence, based on the changes in the measured electrical signal. The nanopore technique allowed for remarkable extending the sequencing read lengths and enabled direct sequencing of native DNA and RNA molecules. As a result, within just a few-year period since the release of the first nanopore sequencer, it has become one of the leading sequencing technologies. The broad scope of nanopore sequencing applications includes, in particular, genome assembly, structural variation studies, identification of nucleic acid chemical modifications and transcript alternative splicing studies. Portable, easy to operate, nanopore sequencers are also increasingly used in epidemiological research to quickly identify and monitor the spread of pathogenic bacteria and viruses.