

mgr inż. Anna Woźniak  
mgr Michał Boroń  
dr n. med. Renata Zbieć-Piekarska  
dr hab. n. med. Magdalena Spólnicka  
*Centralne Laboratorium Kryminalistyczne Policji*

## Zastosowanie technologii wysokoprzepustowego sekwencjonowania DNA w genetyce sądowej

### Streszczenie

Przełom XX i XXI wieku to początek wysokoprzepustowych metod sekwencjonowania DNA, które dzięki coraz większej skuteczności i stopniowej redukcji kosztów doprowadziły do zrewolucjonizowania badań w naukach biomedycznych. W niniejszym artykule omówiono najbardziej popularne technologie sekwencjonowania następnej generacji oraz ich praktyczne zastosowanie w analizach genetycznych w kryminalistyce.

**Słowa kluczowe:** sekwencjonowanie następnej generacji, NGS, MPS, sekwencjonowanie przez syntezę, zastosowanie MPS w kryminalistyce

### Wstęp

W latach osiemdziesiątych XX wieku opracowano pierwsze metody sekwencjonowania DNA. Zespół A.M. Maxama i W. Gilberta zaproponował metodę chemiczną (Maxam, Gilbert, 1977), a zespół F. Sangera i A.R. Coulsona metodę terminacji łańcucha (Sanger, Nicklen, Coulson, 1977). Z uwagi na większą prostotę, mniejszą toksyczność oraz duży potencjał dalszego usprawnienia technika opracowana przez zespół Sangera weszła do powszechnego użycia i przez długi czas stanowiła tzw. złoty standard sekwencjonowania DNA. Udoskonalenia oryginalnego protokołu polegały m.in. na zastąpieniu znakowania dideoksynukleotydów za pomocą izotopów na rzecz znaczników fluorescencyjnych i automatyzacji procesu analizy produktów sekwencjonowania (Franca, Carrilho, Kist, 2002; Heather, Chain, 2016). Dzięki tej metodzie, obecnie zaliczanej do metod sekwencjonowania pierwszej generacji, dokonano przełomowych badań z zakresu genetyki, w tym również genetyki człowieka, a w szczególności poznano pierwsze pełne sekwencje genomów mitochondrialnego i jądrowego (Anderson i in., 1981; Van Dijk i in., 2014). Technologia sekwencjonowania Sangera należy do najbardziej wiarygodnych metod analizy sekwencji DNA, ale jej istotnym ograniczeniem jest niska przepustowość. Rosnące zapotrzebowanie na analizę sekwencji DNA na dużą skalę doprowadziło do opracowania kilku technologii

określanych wspólnym terminem: sekwencjonowanie następnej generacji – NGS (ang. *Next Generation Sequencing*) lub (obecnie coraz częściej) masowe równoległe sekwencjonowanie – MPS (ang. *Massively Parallel Sequencing*). Sekwencjonowanie następnej generacji może dotyczyć analizy grupy powielonych wyjściowych fragmentów DNA, co określane jest jako sekwencjonowanie drugiej generacji – SGS (ang. *Second Generation Sequencing*), lub analizy pojedynczych cząsteczek DNA (bez potrzeby ich wcześniejszej amplifikacji), nazywanej sekwencjonowaniem trzeciej generacji – TGS (ang. *Third Generation Sequencing*) (Kotowska, Zakrzewska-Czerwińska, 2010; Piątkowski i in., 2013).

### Obszary zastosowania w analizie DNA

Mając na uwadze skalę sekwencjonowania oraz obszar, którego sekwencja jest przedmiotem analizy, można wyróżnić trzy podstawowe technologie sekwencjonowania: całego genomu, egzomu oraz określonych fragmentów genomu.

#### 1. Sekwencjonowanie genomu

Technologia sekwencjonowania całego genomu – WGS (ang. *Whole Genome Sequencing*) jest najbardziej kompleksową metodą identyfikowania chorób dziedzicznych czy charakteryzowania niedziedzicznych mutacji powodujących rozwój nowotworów

u ludzi. Można w ten sposób sekwencjonować genom nie tylko ludzki, lecz także innych gatunków zwierząt, roślin, grzybów czy mikroorganizmów, zarówno tych, które już są znane, dzięki wykorzystaniu dostępnych genomów referencyjnych, jak i gatunków niezbadanych. W sekwencjonowaniu *de novo* genom konstruuje się od podstaw, nie korzystając z sekwencji referencyjnej. Stosuje się w tym celu specjalne narzędzia bioinformatyczne.

## 2. Sekwencjonowanie egzomu

W przypadku sekwencjonowania egzomu (ang. *Exome Sequencing*) obszarem objętym analizą są egzony, czyli rejony genomu tworzące geny. Zajmują one jedynie ok 1,5% całego DNA genomowego, ale w ich obrębie znajduje się większość wariantów odpowiedzialnych za stany chorobowe. Koszty sekwencjonowania są stosunkowo niskie, a ilość danych, które można w ten sposób pozyskać – ogromna ([www.genome.gov](http://www.genome.gov)).

## 3. Resekwencjonowanie celowane

W metodyce resekwencjonowania celowanego (ang. *Targeted Sequencing*) sekwencjonowane są tylko konkretne zestawy genów lub regionów genomu. Podejście takie pozwala zaoszczędzić czas i pieniądze dzięki analizie danych dotyczących tylko specyficznych rejonów DNA. Mogą nimi być egzony, wybrane geny, fragmenty niekodującego DNA lub DNA mitochondrialny. Sekwencjonowanie celowane stosuje się również w celu analizy rejonów sprzężonych z cechami fenotypowymi. Metoda ta umożliwia uzyskanie większego pokrycia analizowanych sekwencji i identyfikację rzadkich wariantów lub różnych typów mutacji przy zmniejszonych kosztach analizy.

## Podstawowe metody sekwencjonowania następnej generacji

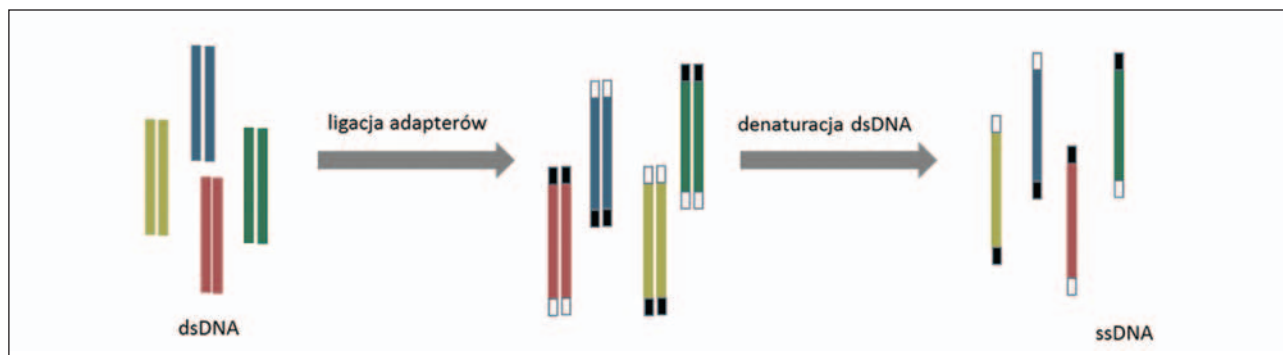
### 1. Sekwencjonowanie drugiej generacji

W technologiach tych masowo sekwencjonowane są tysiące lub miliony zamplifikowanych fragmentów DNA, a włączanie nukleotydów do nowo powstających nici zachodzi cyklicznie w czasie równoległych reakcji. Metody SGS są relatywnie szybkie, ponieważ sekwencjonowanie i detekcja zachodzą równocześnie. Aparatura jest czuła na ilościowy odczyt sygnału płynącego od dużej liczby identycznych cząsteczek przymocowanych do konkretnej lokalizacji. Producenci zapewniają specjalne serwery i oprogramowanie pozwalające na odczyt i analizę dużej liczby złożonych danych pomiarowych, a większość sekwenatorów ma obecnie kompaktowe rozmiary. Dostępne na rynku technologie i platformy sekwencjonowania drugiej generacji różnią się między sobą zastosowanymi rozwiązaniami technicznymi oraz sposobem detekcji. Przekłada się to m.in. na różnice w długości odczytów, wydajności procesu, prędkości reakcji, kosztach eksperymentu oraz typie i częstotliwości indukowanych w trakcie odczytu błędów. Metody drugiej generacji mają

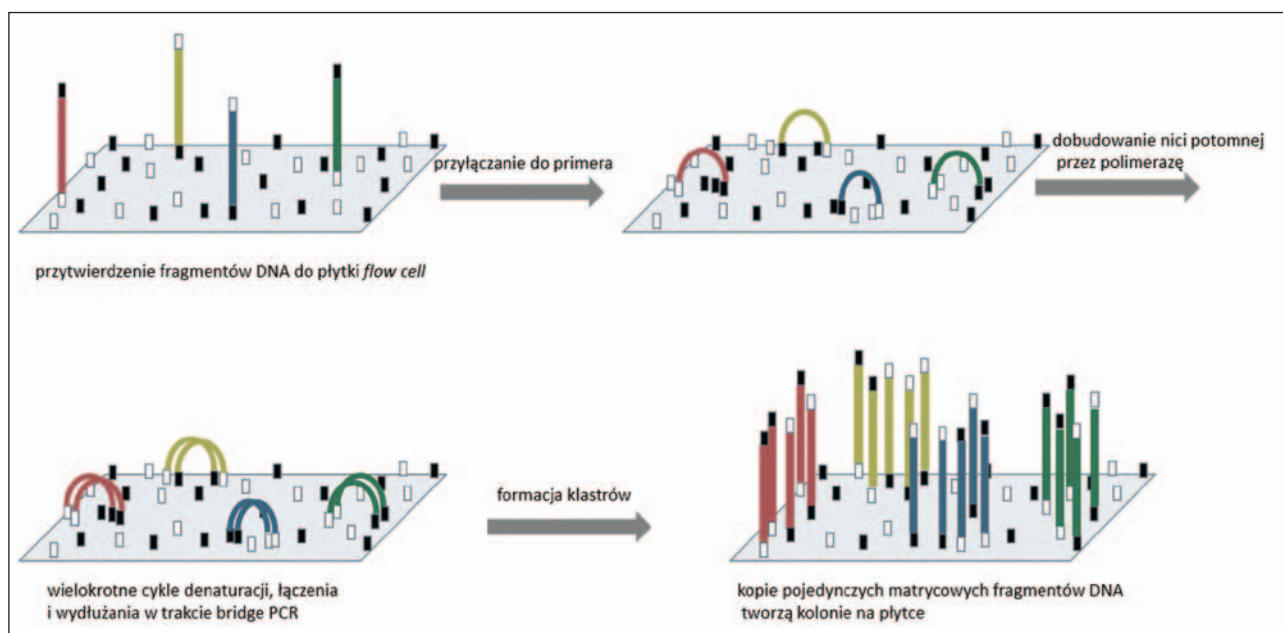
ograniczenia wynikające z użycia PCR oraz wymagają zastosowania nowych algorytmów do układania i składania sekwencji (ang. *alignment and assembly*). Do rzetelnej analizy konieczne jest także otrzymanie odpowiedniego pokrycia, czyli minimalnej liczby odczytów niezbędnej do prawidłowej analizy badanych fragmentów (amplikonów) (Yang, Xie, Yan, 2014; Płoski, 2016; Gupta, Gupta, 2014). W trakcie udoskonalania technik SGS pojawiały się metody i aparatura, które obecnie są już wycofane, np. pirosekwencjonowanie (firma Roche) czy metoda ligacji oligonukleotydów SOLID. Na rynek wkraczają też nowe firmy oferujące inne rozwiązania. Od 2016 r. wysokoprzepustowe sekwenatory oferuje firma BGI z Chin. Obecnie wśród technologii sekwencjonowania MPS najczęściej używane są metody sekwencjonowania przez syntezę z detekcją opartą na pomiarach fluorescencji lub ilości uwalnianych w trakcie reakcji jonów wodoru.

### a. Sekwencjonowanie przez syntezę z detekcją opartą na fluorescencji

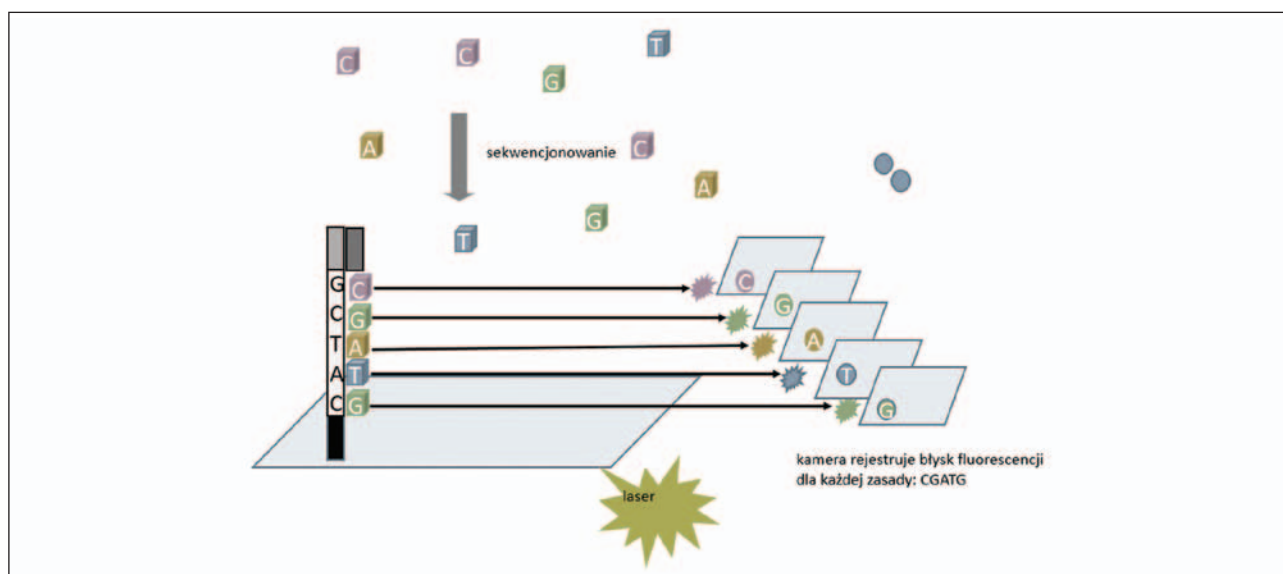
W przypadku tego rodzaju sekwencjonowania przygotowanie bibliotek jest wieloetapowe. Preparowanie kwasu nukleinowego do sekwencjonowania MPS, nazywane konstruowaniem bibliotek MPS, zastąpiło wcześniej wykorzystywane pracochłonne i czasochłonne klonowanie DNA do wektorów bakteryjnych. Na początku DNA poddaje się reakcji amplifikacji z użyciem zestawu specjalnych starterów. Podczas tego procesu fragmenty DNA są namnażane i wydłużane o syntetyczne oligonukleotydy zwane znacznikami (ang. *tag sequences*) lub adapterami, które pełnią funkcje identyfikujące i kontrolne na etapie sekwencjonowania. Następnie biblioteki z użyciem kulek magnetycznych oczyszczają się i normalizują, czyli selekcjonuje według wielkości fragmentów, jednocześnie wyrównując stężenia poszczególnych bibliotek względem siebie. Otrzymane biblioteki łączy się razem, denaturuje i odpowiednio rozcieńcza (ryc. 1a). Przygotowane i zebrane razem biblioteki (ang. *pooled libraries*) umieszczane są w sekwenatorze. Dochodzi tam do hybrydyzacji jednoniciowych fragmentów DNA do komplementarnych oligonukleotydów przyczepionych do dołków w powierzchni szklanej płytki (ang. *flow cell*), a następnie amplifikacji DNA. W reakcji mostkowego PCR (ang. *bridge PCR*) po dodaniu nukleotydów polimeraza dobudowuje nowo powstającą nić. Po etapie denaturacji następuje wyptukanie nici matrycowej. Cykle łączenia, wydłużania i denaturacji powtarzają się wielokrotnie, aż powstaną klastry zawierające tysiące kopii wyjściowego fragmentu. Ostatni cykl PCR kończy się wyptukaniem jednej nici DNA oraz zablokowaniem dalszej elongacji przez dodanie do końca 3' blokującego dideoksynukleotydu (ddNTP) (ryc. 1b). Maszyna po przygotowaniu klastrów z kopiami wyjściowych cząsteczek DNA przystępuje do sekwencjonowania, w którym wstawiane są fluorescencyjnie znakowane nukleotydy. Każdy kolor znacznika jest przypisany konkretnej zasadzie i działa jak odwracalny



Ryc. 1a. Przygotowanie bibliotek NGS w technologii Illumina®.



Ryc. 1b. Mostkowy PCR i formowanie klastrow na płytce.



Ryc. 1c. Sekwencjonowanie przez syntezę w technologii Illumina®. Cztery wyznakowane fluorescencyjnie nukleotydy są obecne w roztworze. Fluorescencję rejestruje kamera CCD po inkorporacji komplementarnego nukleotydu w każdym cyklu.

terminator. Po włączeniu nukleotydu ze znacznikiem następuje detekcja fluorescencji przy użyciu kamery CCD. Dzięki dużej liczbie kopii DNA przytwierdzonych w jednym obszarze sygnał z klastra jest wystarczająco intensywny, by aparat mógł go wychwycić. Wyptukanie grupy fluorescencyjnej odblokowuje możliwość przyłączenia kolejnego nukleotydu. Cykle powtarzane są wielokrotnie (kilkaset razy), w każdym z nich następuje inkorporacja jednego z czterech obecnych w roztworze nukleotydów (ryc. 1c). Ten typ sekwencjonowania wykorzystywany jest w aparatach firmy Illumina (Heather, Chain, 2016; Van Dijk i in., 2014; Kotowska, Zakrzewska-Czerwińska, 2010; Piątkowski i in., 2013; Płoski, 2016; Gupta, Gupta, 2014).

b. Sekwencjonowanie przez syntezę z detekcją opartą na półprzewodnikach i pomiarze pH

Konstrukcja bibliotek wygląda podobnie jak przy poprzedniej technologii. Inaczej zaprojektowany jest etap amplifikacji bibliotek. Tutaj przeprowadza się go poza sekwenatorem w formie PCR emulsyjnego. W emulsji oleju z wodą powstają krople zawierające kulkę, polimerazę DNA i inne odczynniki niezbędne dla reakcji. Amplifikacja DNA odbywa się na kulkach wg zasady: jeden fragment DNA – jedna kulka (ryc. 2a). Następnie nieopłaszczone, puste kulki są w dużej mierze odptukiwane (tzw. *enrichment*), a pełne kulki przenoszone na specjalną płytkę zwaną chipem. Kompleksy kulka – namnożone fragmenty DNA pasują rozmiarem do dołków chipa, pod którymi znajdują się czujniki mierzące zmiany pH. Po manualnym umieszczeniu chipa w sekwenatorze odbywa się proces sekwencjonowania. Nukleotydy dodawane są pojedynczo, a w przypadku wbudowania nukleotydu komplementarnego do

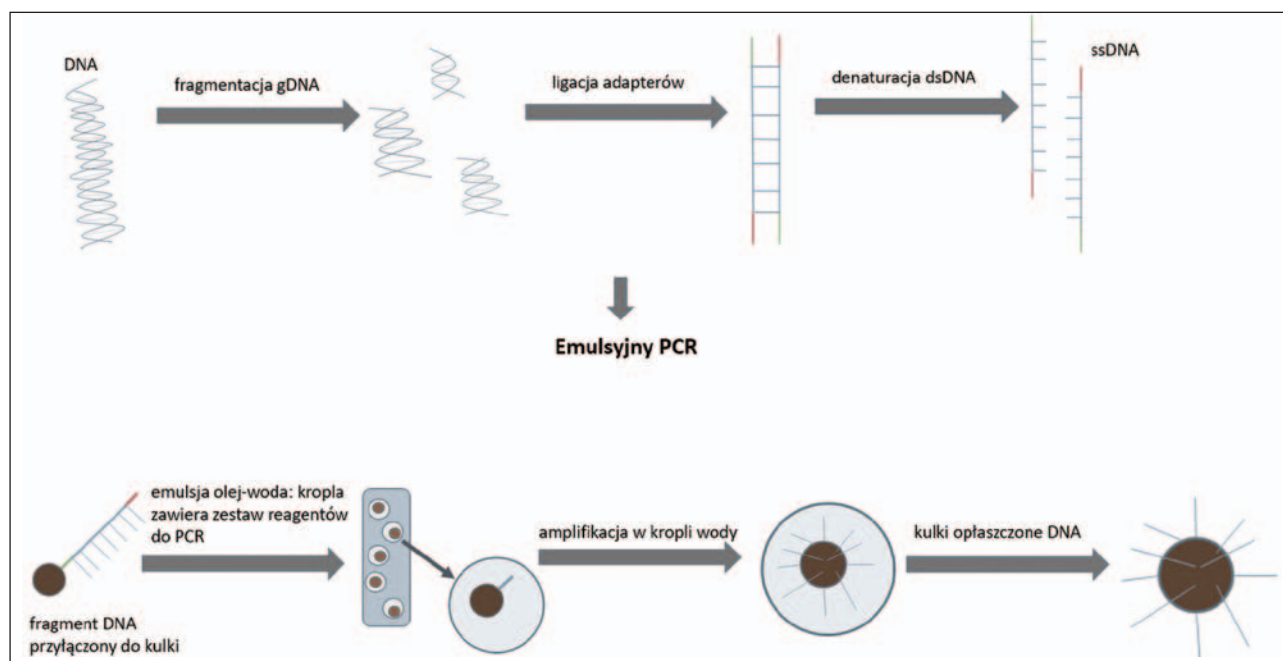
nici matrycowej następuje pomiar uwalnianych jonów wodoru (ryc. 2b). Aparat odczytuje sygnał, a jego siła zależy od liczby nukleotydów przyłączonych w cyklu. W dołkach, w których kolejny nukleotyd nie jest komplementarny do aktualnie obecnego w roztworze, nie zachodzi inkorporacja ani zmiana pH. Wyzwaniem w tej metodzie jest prawidłowa ocena liczby wbudowanych zasad w długich ciągach homopolimerowych, a zaletą – wyeliminowanie części aparatury rejestrującej (laser, kamera), uproszczenie procesu, skrócenie czasu odczytu i obniżenie kosztów sekwencjonowania. Ten typ sekwencjonowania wykorzystywany jest w aparatach firmy Ion Torrent: PGM, Ion Proton oraz najnowszym S5, współpracującym z kompatybilnym robotem, przygotowującym biblioteki i ładującym próbki na chip, nazwanym Ion Chef (Van Dijk i in., 2014; Piątkowski i in., 2013; Płoski, 2016).

## 2. Sekwencjonowanie trzeciej generacji

Sekwencjonowanie trzeciej generacji polega na odczycie sekwencji tylko jednej molekuly DNA bez namnażania cząsteczek w reakcji PCR. Sekwencjonowanie zachodzi w czasie rzeczywistym, jest szybsze niż sekwencjonowanie przez syntezę, a odczytane sekwencje są wielokrotnie dłuższe. Wadą tych technologii są na razie odczyty obciążone dużymi błędami.

### a. Sekwencjonowanie SMRT w dołkach światłowodowych ZMW

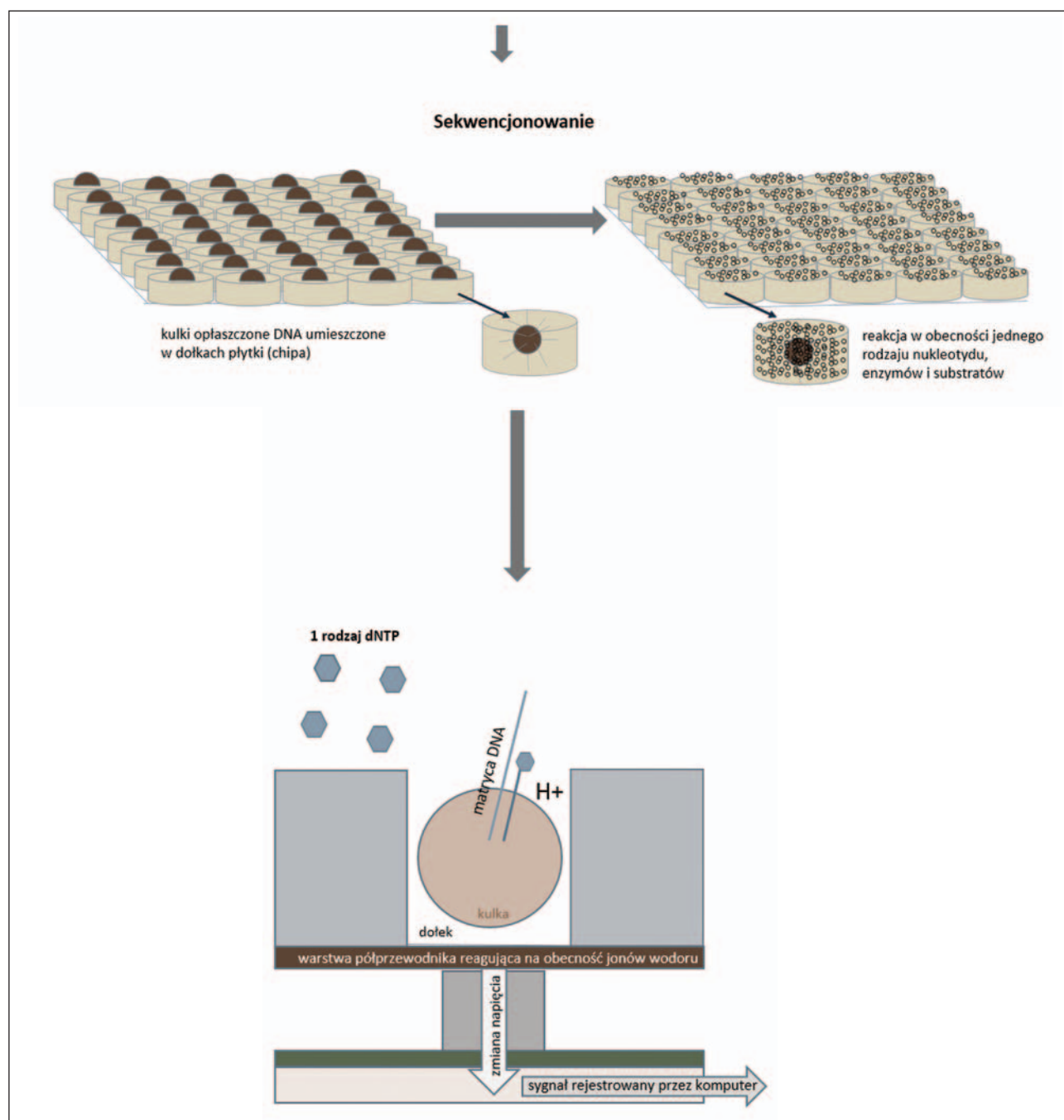
Przygotowanie DNA do sekwencjonowania odbywa się w kilku etapach. Pierwszym jest fragmentacja DNA na odcinki o długości 10 kb, a następnym naprawa uszkodzeń powstałych w trakcie fragmentacji, naprawa końców DNA i jego oczyszczenie. W kolejnym kroku



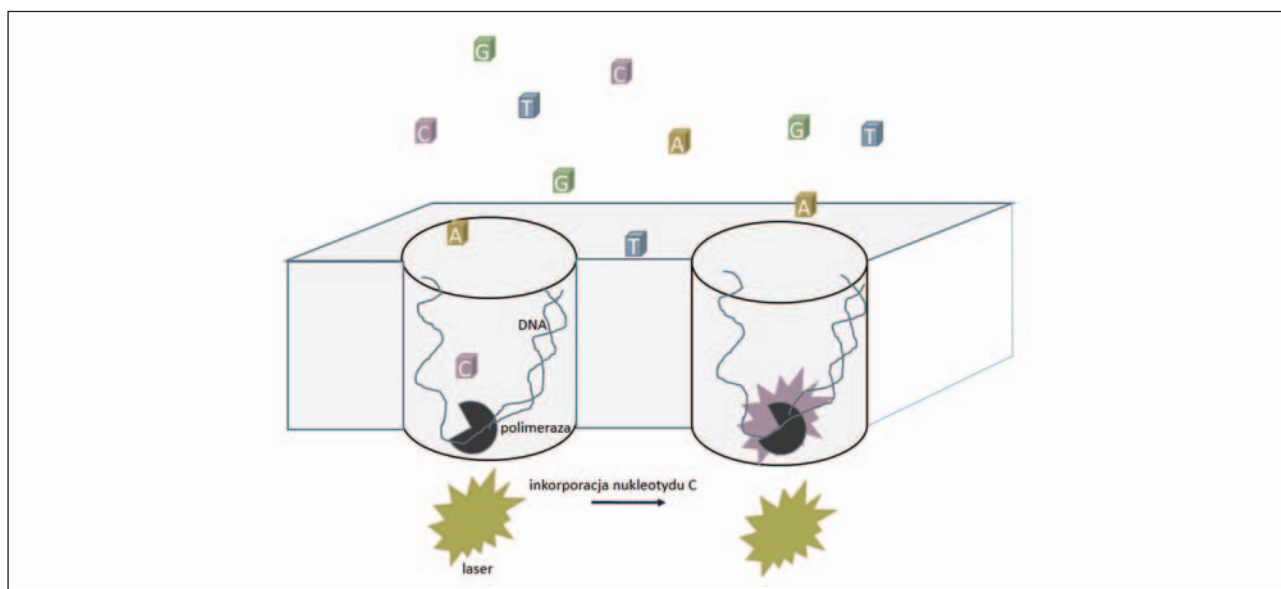
Ryc. 2 a. Przygotowanie bibliotek NGS w technologii Ion Torrent™ i klonalna amplifikacja na kulkach (emulsyjny PCR).

do otrzymanych tzw. tępych końców dodawane są dwa adaptory typu spinka do włosów (pozwala to na odczyt obu nici wiele razy) oraz przyłączane startery sekwencyjne do powstałych kompleksów SMRTbell™ Templates. Na końcu do każdego kompleksu dołączana jest cząsteczka polimerazy. Sekwencjonowanie odbywa się na płytce SMRT (ang. *Single Molecule Real Time Sequencing*) z dołkami ZMW (ang. *Zero-Mode Waveguides*), dokąd kompleksy dostają się na drodze dyfuzji lub za pośrednictwem kulek. Rozmiar dołka pozwala na umieszczenie w nim tylko jednego kompleksu. ZMW to otwór wykonany w metalowej membranie grubości 100 nm założonej na szkło. Mała

średnica otworu nie pozwala światłu lasera znajdującego się pod dołkiem na przejście przez całą jego długość. Ulega rozproszeniu na 1/3 głębokości, w miejscu przymocowania polimerazy. Każdy z czterech nukleotydów wyznakowany jest fluoroforem, który po wzbudzeniu światłem lasera generuje błysk światła. Wszystkie nukleotydy są obecne w roztworze, w którym zanurzona jest płytka. Wpływają do dołków na zasadzie dyfuzji i wypływają w razie braku komplementarności. Laser nie wzbudza znaczników, dopóki nie dotrą one do dna. Inkorporacja nukleotydu przez polimerazę wydłuża czas jego ekspozycji na światło lasera, co skutkuje silniejszą emisją światła w odpowiednim



**Ryc. 2 b.** Sekwencjonowanie w technologii Ion Torrent™ z detekcją opartą na pomiarze zmiany pH.



**Ryc. 3.** Sekwencjonowanie pojedynczej cząsteczki DNA w technologii PacBio (SMRT). Każda zasada emituje światło innego koloru po wzbudzeniu światłem lasera.

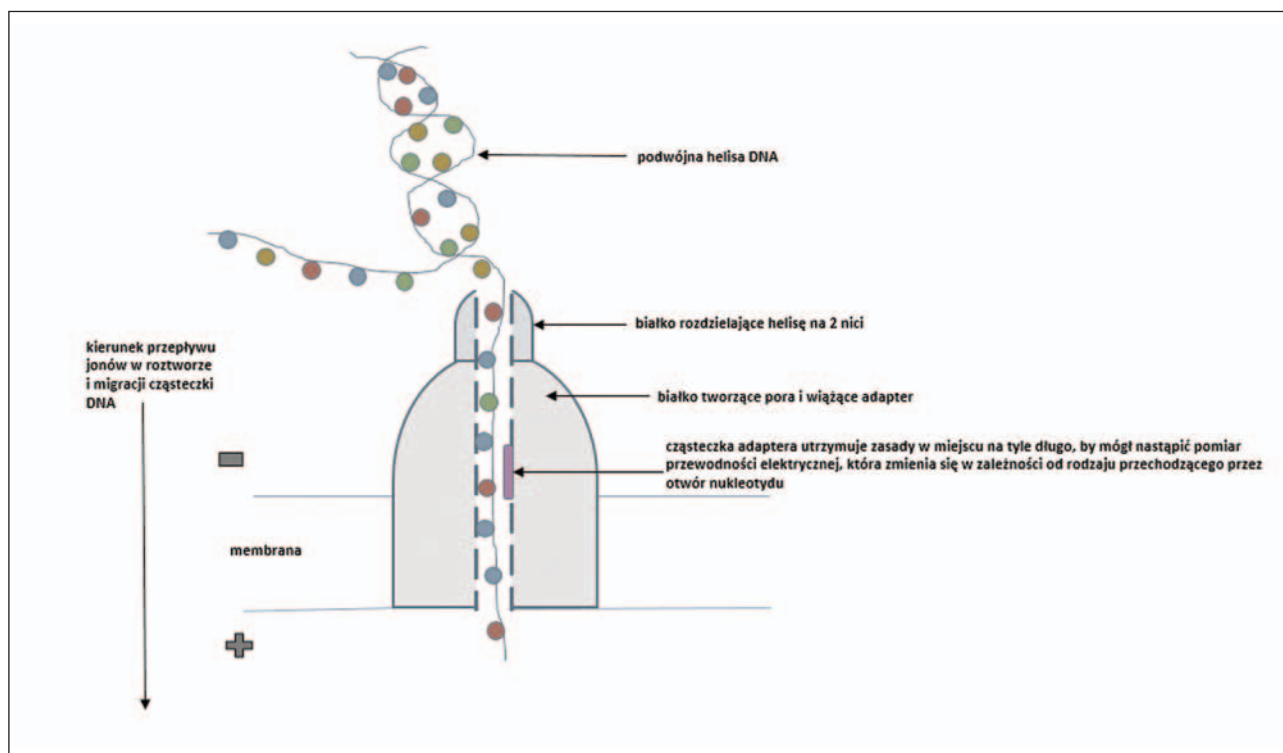
dla danej zasady kolorze (ryc. 3). Błysk fluorescencji rejestruje kamera w czasie rzeczywistym. Znacznik jest uwalniany po inkorporacji, co odblokowuje możliwość wstawienia kolejnej zasady przez polimerazę. Ten typ sekwencjonowania został wprowadzony przez firmę Pacific Biosciences. Pozwala na odczyty rzędu kilkudziesięciu tysięcy par zasad w ciągu kilku godzin, jak również na wykrywanie modyfikacji epigenetycznych w DNA (Heather, Chain, 2016; Van Dijk i in., 2014; Kotowska, Zakrzewska-Czerwińska, 2010; Piątkowski i in., 2013; Yang, Xie, Yan, 2014; Płoski, 2016; Gupta, Gupta, 2014; Rhoads, Au, 2015).

#### b. Sekwencjonowanie w nanoporach

Preparatyka próbek DNA polega na fragmentacji DNA, reperfacji końców oraz ich adenyacji. Po etapie oczyszczenia następuje ligacja adaptera o strukturze spinki do włosów do ogona poli A w dwuniciowym DNA. Do nici DNA dołączane są też: tzw. silniczek molekularny rozplątujący podwójną helisę, który zapewnia optymalny czas przejścia przez nanopor dzięki unieruchomieniu jednej nici, oraz molekula mocująca (ang. *tether*) wolny koniec nici DNA blisko otworu w membranie. Technologia ta wykorzystwała znane w przyrodzie zjawisko transportu cząsteczek przez membrany, np. błony komórkowe z kanałami jonowymi. Membrany mogą być pochodzenia biologicznego (białkowe, np. phi29, alfa-hemolizyna), syntetyczne lub mieszane. Syntetyczne nanopory to otwory o średnicy 1–100 nm tworzone przez trawienie chemiczne, bombardowanie jonami lub wiązką z elektronowego mikroskopu transmisyjnego cienkich warstw materiałów (np. polimery, silikon, związki krzemu, glinu, molibdenu lub grafen). Nanopory mają różne właściwości: grubość, tempo translokacji cząsteczek, lepkość do DNA etc.

Membrana jest tak dobrana, aby była jak najcieńsza, co daje możliwość odczytu jednej zasady w danym czasie. Nieprzepuszczalna elektrycznie membrana zanurzona jest w roztworze, a napięcie elektryczne umożliwia przepływ jonów przez otwór. Gradient stężeń nakierowuje cząsteczki do nanopora. Im bliżej otworu, tym wyższe stężenie i wyższa przewodność elektryczna. Cząsteczki mające ładunek elektryczny przemieszczają się w kierunku elektrody o wartości przeciwnej, a nić DNA ze względu na swój ładunek przesuwa się w roztworze w kierunku elektrody o ładunku dodatnim (ryc. 4). Gdy cząsteczka DNA wejdzie do nanopora, ruch jonów zostaje częściowo zablokowany i wartość mierzonego prądu maleje. Tempo przechodzenia cząsteczki przez por można modyfikować. Spowalnia je obniżenie temperatury lub natężenia prądu oraz wzrost stężenia soli. Szybszy jest przepływ cząsteczek przez otwory o mniejszej średnicy. Kinetyka przechodzenia DNA zależy od długości i konformacji nici. Każda zasada ma inną przewodność, odczyt sekwencji zachodzi więc na podstawie jej pomiarów (Feng i in., 2015).

W ramach jednego eksperymentu można sekwencjonować jedną lub wiele cząsteczek DNA. Odczyty mogą mieć długość do kilkuset tysięcy pz. Ten typ sekwencjonowania wykorzystywany jest w aparatach firmy Oxford Nanopores. Urządzenia, w których odbywa się proces sekwencjonowania, to chipy wielkości pudełka zapatek z czujnikami mierzącymi potencjał, elektrodami i jedną lub wieloma płytkami z nanoporami noszącymi nazwy MinION i PromethION. Metodą tą można także identyfikować zmiany epigenetyczne w DNA (Heather, Chain, 2016; Kotowska, Zakrzewska-Czerwińska, 2010; Gupta, Gupta, 2014; Laver i in., 2015; Regalado, 2014).



Ryc. 4. Sekwencjonowanie pojedynczej cząsteczki DNA w technologii Oxford Nanopores.

### Zastosowanie metod wysokoprzepustowego sekwencjonowania DNA w genetyce sądowej

Technologia MPS wprowadza gamę nowych możliwości badania DNA. Część z nich znajduje bezpośrednie zastosowanie w kryminalistyce, ponieważ stanowi rozszerzenie oraz połączenie stosowanych dotychczas metod identyfikacji osobniczej. Obejmuje to profilowanie autosomalnych oraz położonych na chromosomach płci markerów STR (ang. *Short Tandem Repeat*), markerów SNP (ang. *Single Nucleotide Polymorphism*) i analizę mitochondrialnego DNA (mtDNA). Szczególnie obiecujące jest zastosowanie sekwencjonowania wysokoprzepustowego na etapie dochodzeniowo-sledczym do tzw. kryminalistycznego fenotypowania DNA (FDP). Metody predykcyjnej analizy DNA wymagają analizy setek, a nawet tysięcy polimorfizmów rozsianych w całym genomie człowieka i dlatego rozwój tych metod ściśle związany jest z wdrożeniem rozwiązań genomicznych w genetyce sądowej. Fenotypowanie DNA odnosi się przede wszystkim do przewidywania wyglądu fizycznego, ale często włącza się również tutaj analizę pochodzenia biogeograficznego oraz wieku człowieka. Wszystkie te informacje mogą mieć znaczenie wywiadowcze. Zwiększenie informatywności badanego materiału biologicznego o dodatkowe dane, nawet w przypadku próbek trudnych i zdegradowanych, jest jedną z kluczowych zalet nowoczesnych technologii sekwencjonowania i w przyszłości NGS może być rutynowo wykorzystywane w działaniach Policji.

### 1. Wykorzystanie NGS

#### a. Identyfikacja osobnicza: markery STR i SNP

Do identyfikacji genetycznej wykorzystuje się obecnie analizę zmienności niekodujących fragmentów DNA, głównie sekwencji mikrosatelitarnych zwanych markerami STR. Profilowanie DNA dotyczyć może *loci* STR zlokalizowanych w autosomach oraz w chromosomach płci: X i Y. Autosomalne markery STR są wykorzystywane do identyfikacji genetycznej w laboratoriach kryminalistycznych przy użyciu technologii multiplexowego PCR i elektroforezy kapilarnej. Profile uzyskane z badań wprowadzane są do krajowych baz danych DNA i przeszukiwane na podstawie krajowych i międzynarodowych regulacji prawnych. Podstawowym zestawem STR jest układ 13 *loci* z systemu CODIS (ang. *Combined DNA Index System*) opracowanego przez FBI wraz z markerem genu amelogeniny pozwalającym na określenie płci. Analiza Y-STR jest szczególnie przydatna w przypadku ustalania pokrewieństwa w linii męskiej lub wykrywania męskiego komponentu w przestępstwach na tle seksualnym. Możliwość przypisania danego osobnika do odpowiedniej haplogrupy dają wysoce konserwatywne mutacje chromosomu Y dziedziczone w kolejnych pokoleniach. Światowa referencyjna baza danych Y-STR Haplotype Reference Database (Y-HRD) zawiera informacje o zidentyfikowanych haplotypach oraz częstościach ich występowania w różnych populacjach. Profilowanie z wykorzystaniem technologii NGS już dzisiaj pozwala na jednoczesne badanie kilkudziesięciu markerów STR oraz kilkuset markerów SNP w jednej reakcji dla kilkudziesięciu

próbek. Dzięki uzyskaniu danych na temat liczby powtórzeń oraz sekwencji DNA badanych markerów dysponujemy większą siłą dyskryminacji. Spowodowane jest to występowaniem izoalleli, czyli alleli o identycznej długości, różniących się jednak sekwencją DNA. Liczba ta zwiększa się także dzięki analizie polimorfizmu DNA regionów flankujących markerów STR. Fakt występowania izoalleli może usprawnić proces dekonwolucji mieszanin, a także umożliwić ocenę pochodzenia mutacji od matki lub od ojca w czasie badań identyfikacyjnych lub badań ojcostwa. Zestawy markerów identyfikacyjnych typu SNP znajdują swoje zastosowanie szczególnie w sytuacji, kiedy DNA jest zdegradowany. Zestaw 50–100 SNP ma siłę dyskryminacyjną równą tej uzyskiwanej dla 10–16 *loci* STR (Loveliness i in., 2017; Pakstis i in., 2009; Gill i in., 2004).

#### b. Analiza mitochondrialnego DNA

Technologia NGS znajduje zastosowanie również przy analizie mtDNA, szczególnie w przypadku małych ilości materiału czy zdegradowanego DNA. Sekwencjonując cały genom mitochondrialny zamiast jedynie hiperzmiennych polimorficznych regionów HVI/HVII, które zazwyczaj bada się technologią Sangera, zwiększa się siłę dyskryminacyjną oraz niweluje problem wspólnych regionów hiperzmiennych w populacjach wywodzących się od tego samego przodka. Dodatkowo dzięki sekwencjonowaniu następnej generacji wykrywalność zjawiska heteroplazmii podwyższa się ponad dwukrotnie w stosunku do sekwencjonowania techniką Sangera. Na podstawie analizy mtDNA można ustalić pokrewieństwo w linii żeńskiej, co jest szczególnie użyteczne przy porównywaniu próbek pochodzących od ofiar katastrof masowych lub NN osób z żyjącymi krewnymi ze strony matki. Ujawniono obecność licznych haplogrup mtDNA, z których wiele wykazuje rozkład kontynentalny i stanowi źródło danych biogeograficznych. Populacyjna baza danych regionu kontrolnego mtDNA EMPOP zawiera dane dotyczące regionów hiperzmiennych, regionu kodującego i całego genomu mitochondrialnego. Baza ta daje możliwość weryfikacji poprawności otrzymanych wyników oraz przyporządkowania zbadanej sekwencji do określonej haplogrupy (Holland, Makova, McElhoe, 2018; Just, Irwin, Parson, 2015; Shih i in., 2018).

#### c. Przewidywanie wyglądu i pochodzenia biogeograficznego

Niezwykle obiecującym i jednym z ważniejszych obecnie kierunków rozwoju jest predykcja cech wyglądu fizycznego (ang. *DNA phenotyping*) oraz pochodzenia biogeograficznego (ang. *ancestry*) dawcy próbek DNA. Narzędzia, które do tego służą, w zdecydowanej większości opierają się na analizie markerów SNP. Opracowano i opublikowano specjalne panele najlepszych markerów tych cech służących przewidywaniu zarówno wyglądu, jak i pochodzenia. W przypadku przewidywania fenotypu badanie polega na

genotypowaniu markerów sprzężonych z genami odpowiadającymi za wygląd człowieka oraz budowaniu modeli matematycznych pozwalających na analizę otrzymanych danych. Do chwili obecnej najdokładniej poznane są markery cech pigmentacyjnych. Zabarwienie tęczówki oka, włosów czy odcień skóry to wynik interakcji kilku genów charakteryzujących się wysoką odziedziczalnością. Należą do nich m.in.: *HERC2*, *OCA2*, *SLC24A4*, *SLC45A2*, *SLC24A5*, *TYR*, *IRF4*, *MC1R*, *TYRP1*, *TPCN2*, *ASIP*, *KITLG*, *EXOC2*. Geny te nie są równoważne i jedno odpowiada za większy procent zmienności danej cechy niż inne, np. *HERC2*, *OCA2* za pigmentację oczu, a *MC1R* za pigmentację włosów. Na potrzeby kryminalistyki opracowano i zwalidowano kilka systemów predykcyjnych. Pozwalają one, po podaniu genotypu dla konkretnych markerów, na predykcję pigmentacji oczu, włosów czy nawet odcienia skóry, np. IrisPlex (predykcja koloru oczu: niebieski, pośredni, brązowy), HIrisPlex (predykcja koloru oczu: niebieski, pośredni, brązowy; koloru włosów: blond, rudy, brąz, czarny; odcienia włosów: jasny, ciemny) czy Snipper (kolor oczu: brąz, niebieski, zielonoorzechowy; włosów: blond, rudy, brąz, czarny; odcienia włosów: jasny, ciemny; odcienia skóry: jasny, średni, ciemny). Trwają prace nad zidentyfikowaniem genetycznym takich cech jak męskie łysienie (m.in. geny *AR/EDA2R* na chromosomie X), siwienie, morfologia włosów (m.in. geny *EDAR*, *FGFR2*, *TCHH*), kształt twarzy (m.in. geny *PAX3*, *PRDM16*, *TP63*) czy wzrost (Kayser, 2015). Ocena pochodzenia biogeograficznego człowieka opiera się głównie na analizie markerów AIMs (ang. *Ancestry Informative Markers*) typu SNP i InDel. Analiza pochodzenia na podstawie markerów pozyskanych z chromosomu Y czy mtDNA pozwala na rekonstrukcję migracji i rozprzestrzeniania się populacji, a w odniesieniu do jednostki – na połączenie cech wyglądu z jej pochodzeniem w zależności od procentu domieszki innej populacji.

Komersyjne zestawy do badań kryminalistycznych produkowane przez wiodące na rynku firmy ThermoFisherScientific czy Illumina pozwalają na identyfikację genotypu i/lub predykcję pigmentacji włosów, oczu oraz pochodzenia przy wykorzystaniu technologii NGS. Przykładowo zestaw FGx firmy Illumina pozwala na uzyskanie w ciągu kilku dni informacji na temat 27 autosomalnych markerów STR, 24 markerów STR chromosomu Y, 7 markerów STR chromosomu X, 94 identyfikacyjnych markerów typu SNP, 22 markerów SNP wykorzystywanych do określenia koloru oczu i włosów oraz 56 markerów SNP wykorzystywanych do określenia pochodzenia (<https://www.illumina.com/>).

#### d. Przewidywanie wieku człowieka

Dzięki zastosowaniu MPS możliwe jest obecnie ustalenie wieku osoby na podstawie pozostawionych śladów biologicznych z dokładnością do kilku lat. Ma to znaczenie tym większe, że wiek wpływa na wygląd zewnętrzny człowieka, a tym samym rezultat analizy predykcyjnej



cech fenotypowych jest związany z takimi zmianami jak np. łysienie czy siwienie włosów. Do badań wykorzystuje się markery epigenetyczne – rejony genomu, które z wiekiem zmieniają swój profil metylacyjny. Na podstawie analizy tych zmian i modelowania matematycznego przewiduje się wiek chronologiczny człowieka. Markery wieku zazwyczaj zlokalizowane są w obrębie wysp CpG w rejonach promotorowych genów i wykazują tendencje do hipometylacji lub hipermetylacji wraz z upływem czasu. Opublikowano kilkanaście potencjalnie użytecznych markerów epigenetycznych. Niektóre z tych markerów mają zastosowanie podczas szacowania wieku biologicznego, inne natomiast podczas szacowania wieku chronologicznego. Do analiz najczęściej wykorzystuje się zestawy złożone z kilku markerów, które pozwalają na szacowanie wieku człowieka z dokładnością średnio ok.  $\pm 3$  do 5 lat z próbek krwi, śliny lub spermy (Zbieć-Piekarska i in., 2015).

e. Identyfikacja tkanek i płynów ustrojowych człowieka: markery mRNA i miRNA

Płyny ustrojowe oraz tkanki odróżnia się od siebie przez sekwencjonowanie biomarkerów mRNA, co pozwala uzyskać tzw. profil mRNA. Wykorzystując fakt zróżnicowanej ekspresji poszczególnych genów w komórkach różnych narządów (np. płuca, mózg, nerki, skóra) lub w substancjach interesujących z punktu widzenia badań kryminalistycznych (np. krew, ślina, nasienie, krew menstruacyjna, wydzielina z pochwy), zidentyfikowano szereg markerów mRNA pozwalających na ich identyfikację. Co ważne, jednoczesna ekstrakcja RNA i DNA z próbek pozwala na przeprowadzenie zarówno identyfikacji mRNA, jak i klasycznego profilowania STR. Próbkę można multipleksować, co przyspiesza czas analizy, a procedura jest odpowiednia także do zdegradowanych próbek dowodowych.

Jako potencjalnie użyteczne do różnicowania substancji takich jak mocz, krew czy ślina pod uwagę brane są także biomarkery miRNA. Cząsteczki miRNA to krótkie (20–25 pz) niekodujące struktury RNA bardziej odporne na degradację, co w przypadku materiału dowodowego pochodzącego z odmiennych środowisk i przechowywanego w różnych warunkach ma szczególnie duże znaczenie. Zidentyfikowano szereg specyficznych tkankowo miRNA, jednakże metodyka wymaga dalszego dopracowania, by uzyskiwane wyniki były powtarzalne (Hanson i in., 2018; Silva i in., 2014).

f. Mikrobiom

Sekwencjonowanie RNA oraz DNA mikroorganizmów stanowi źródło informacji na temat biogeograficznego pochodzenia badanych dowodów i może być wykorzystane w śledztwie. Na podstawie wiedzy o ludzkim mikrobiomie możliwa jest identyfikacja człowieka, identyfikacja miejsca na ciele, z którego pochodzi dowód, oraz oszacowanie czasu zgonu. Sekwencjonowanie materiału genetycznego pozwala także na identyfikację

rodzaju bakterii czy wirusów w przypadku spraw z użyciem broni biologicznej (Lilje i in., 2013).

### Podsumowanie

W dziedzinie badań genetycznych i analiz molekularnych nowe technologie stosunkowo szybko zastępowane są jeszcze nowszymi. Z jednej strony powstają sekwenatory wysokoprzepustowe mające możliwość analizy wielu genomów jednocześnie, z drugiej strony badacze szukają pojedynczych zmian w sekwencji DNA dających się połączyć z konkretnymi cechami wyglądu człowieka. W ostatnim czasie znaczna część elementów protokołu sekwencjonowania i przygotowania bibliotek została automatyzowana, co pozwala na wykonanie analiz z nikłym udziałem człowieka. Rośnie precyzja odczytu pojedynczych cząsteczek DNA, a małe przenośne urządzenia sprawiają, iż coraz bardziej prawdopodobna jest wizja mobilnych laboratoriów, które na miejscu zdarzenia wykonają najbardziej skomplikowane analizy, umożliwiające określenie wyglądu i wieku podejrzanego, a w ten sposób dodanie istotnych informacji dla śledztwa.

**Źródło rycin:** autorzy

### Bibliografia

1. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coilsen, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806).
2. Feng, Y., Zhang, Y., Ying, C., Wang, D., Du, C. (2015). Nanopore-based Fourth-generation DNA Sequencing Technology. *Genomics, Proteomics & Bioinformatics*, 13(1).
3. Franca, L.T.C., Carrilho, E., Kist, T.B.L. (2002). A review of DNA sequencing techniques. *Quarterly Reviews of Biophysics*, 35(2).
4. Gill, P., Werrett, D.J., Budowle, B., Guerrieri, R. (2004). An assessment of whether SNPs will replace STRs in national DNA databases – joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGAM). *Science and Justice: Journal of the Forensic Science Society*, 44(1).
5. Gupta, A.K., Gupta, U.D. (2014). Next generation sequencing and its applications. W: A.S. Verma (red.), *Animal Biotechnology: Models in Discovery and Translation*. Amsterdam–Boston: Elsevier.
6. Hanson, E., Ingold, S., Haas, C., Ballantyne, J. (2018). Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing. *Forensic Science International: Genetics*, 34.

7. Heather, M., Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1).
8. Holland, M.M., Makova, K.D., McElhoe, J.A. (2018). Deep-coverage MPS analysis of heteroplasmic variants within the mtGenome allows for frequent differentiation of maternal relatives. *Genes* (Basel), 9(3).
9. Just, R.S., Irwin, J.A., Parson, W. (2015). Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Science International: Genetics*, 18.
10. Kayser, M. (2015). Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics*, 18.
11. Kotowska M., Zakrzewska-Czerwińska J. (2010). Kurs szybkiego czytania DNA – nowoczesne techniki sekwencjonowania. *Biotechnologia*, 4(91).
12. Laver, T., Harrison, J., O'Neil, P.A., Moore, K., Farbos, A., Paszkiewicz, K., Studholme, D.J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification*, 3.
13. Lilje, L., Lillsaar, T., Rätsep, R., Simm, J., Aaspõllu, A. (2013). Soil sample metagenome NGS data management for forensic investigation. *Forensic Science International: Genetics*, 4(1).
14. Loveliness, D., Dennis, S., Salvador, J., Calacal, G., De Ungria, M. (2017). Comparison of two massively parallel sequencing platforms using 83 single nucleotide polymorphisms for human identification. *Scientific Reports*, 7, Article number: 398 (2017); doi:10.1038/s41598-017-00510-3.
15. Maxam, A.M., Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74.
16. Pakstis, A.J., Speed, W.C., Fang, R., Hylan, F.C.L., Furtado, M.R., Kidd, J.R., Kidd, K.K. (2009). SNPs for a universal individual identification panel. *Human Genetics*, 127(3), doi: 10.1007/s00439-009-0771-1.
17. Piątkowski, J., Skalniak, A., Bodzioch, M., Pach, D., Hubalewska-Dydejczyk, A. (2013). Wprowadzenie do sekwencjonowania ludzkiego genomu w diagnostyce. *Przegląd Lekarski*, 70(7).
18. Płoski, R. (2016). Next generation sequencing – general information about the technology, possibilities, and limitations. W: U. Demkow, R. Płoski (red.), *Clinical Applications for Next-Generation Sequencing*. London: Academic Press.
19. Regalado, A. (2014). Radical new DNA sequencer finally gets into researchers hands. *Biomedicine*, 17 września.
20. Rhoads, A., Au, K. (2015). PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, 13(5).
21. Sanger, F., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74.
22. Shih, S.Y., Bose, N., Goncalves, A., Erlich, H., Calloway, C. (2018). Applications of probe capture enrichment next generation sequencing for whole mitochondrial genome and 426 nuclear SNPs for forensically challenging samples. *Genes* (Basel), 9(1).
23. Silva, S., Lopes, C., Teixeira, A.L., Sousa, M., Medeiros, R. (2014). Forensic miRNA: Potential biomarker for body fluids. *Forensic Science International: Genetics*, 14, <https://doi.org/10.1016/j.fsigen.2014.09.002>
24. Van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9).
25. Yang, Y., Xie, B., Yan, J. (2014). Application of next-generation sequencing technology in forensic science. *Genomics, Proteomics & Bioinformatics*, 12(5).
26. Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Ż., Pałeczka, A., Kucharczyk, K., Płoski, R., Branicki, W. (2015). Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Science International: Genetics*, 17.