

Metody analizy DNA

– laboratorium genetyczne cz. I

STRESZCZENIE | W pracy omówione zostały wybrane metody molekularne stosowane rutynowo w diagnostyce chorób genetycznie uwarunkowanych, w tym metody podstawowe takie jak reakcja łańcuchowej polimerazy (PCR), sekwencjonowanie metodą Sangera czy hybrydyzacja kwasów nukleinowych. Uwzględniono także nowoczesne metody molekularne takie jak multiplexowa amplifikacja sond zależna od ligacji (MLPA), porównawcza hybrydyzacja genomowa do mikromacierzy (aCGH) czy sekwencjonowanie następnej generacji.

SŁOWA KLUCZOWE

choroby genetycznie uwarunkowane, reakcja łańcuchowa polimerazy, sekwencjonowanie, sekwencjonowanie następnej generacji, hybrydyzacja, MLPA, aCGH

SUMMARY | The paper discusses selected molecular methods routinely used in the diagnosis of inherited diseases, including basic methods such as polymerase chain reaction (PCR), Sanger sequencing or nucleic acid hybridization. Also modern molecular methods such as Multiplex ligation-dependent probe amplification (MLPA), array-comparative genomic hybridization (aCGH) and next-generation sequencing (NGS).

KEY WORDS | inherited diseases, polymerase chain reaction, sequencing, next-generation sequencing, hybridization, MLPA, aCGH

Monika Gos, Agnieszka Charzewska, Katarzyna Wertheim-Tysarowska, Sylwia Rzońca, Beata Nowakowska, Dorota Hoffman-Zacharska

IZAKŁAD GENETYKI MEDYCZNEJ, INSTYTUT MATKI I DZIECKA, WARSZAWA

Główną istotą działalności laboratoriów genetycznych jest diagnostyka chorób/zespołów chorobowych, u których podłoża znajdują się zmiany w materiale genetycznym człowieka. Są to badania jednoznaczne, ostatecznie potwierdzające rozpoznanie kliniczne, które w założeniu wykonywane są jednorazowo dla konkretnego pacjenta, a często i dla członków jego rodziny. Badania prowadzone w laboratoriach genetycznych skupiają się na identyfikacji mutacji – aberracji chromosomowych lub zmian w sekwencji DNA. Identyfikacja konkretnej zmiany jest wskazówką co do dalszego postępowania terapeutycznego, a także umożliwia ocenę ryzyka zachorowania i poradnictwo genetyczne.

Metody stosowane w diagnostyce chorób genetycznie uwarunkowanych można podzielić na dwie grupy: badania cytogenetyczne, w trakcie których analizuje się liczbę i strukturę chromosomów, i badania molekularne, skupiające się raczej na badaniu mutacji sekwencji DNA. W ostatnim czasie podział ten stracił na znaczeniu, gdyż metody wykorzystywane do tej pory jedynie w badaniach molekularnych (np. hybrydyzacja, ligacja czy PCR) zostały zaadaptowane na potrzeby analizy aberracji chromosomowych. Niniejszy artykuł jest prezentacją wybranych metod molekularnych stosowanych w Zakładzie Genetyki Medycznej Instytutu Matki i Dziecka w laboratorium w codziennej praktyce diagnostycznej.

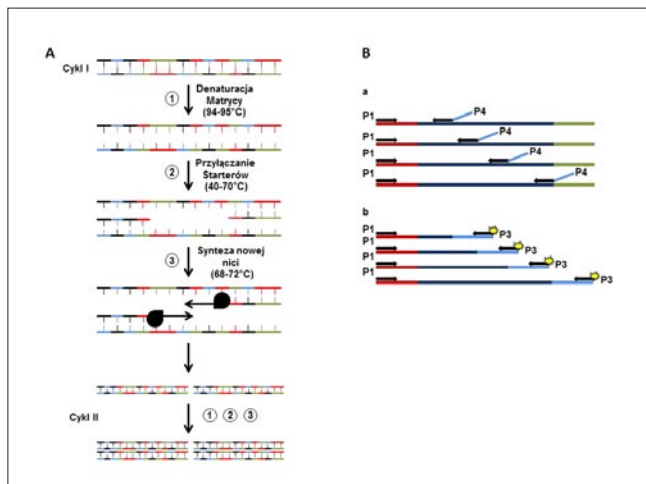
REAKCJA ŁAŃCUCHOWA POLIMERAZY PCR

Reakcja łańcuchowa polimerazy (PCR, ang. *polymerase chain reaction*) jest techniką rutynowo wykorzystywaną

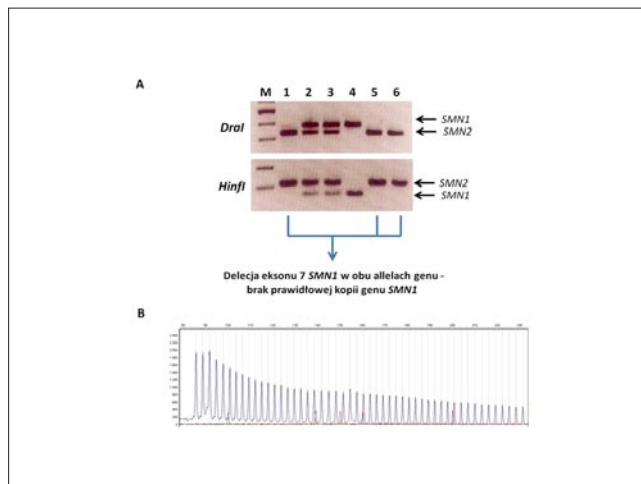
w laboratoriach genetycznych. Metoda PCR polega na specyficznym powieleniu (amplifikacji) wybranego fragmentu genomu w warunkach *in vitro*. Po raz pierwszy została opisana w roku 1986, a jej twórca Kary Mullis za swoje odkrycie otrzymał Nagrodę Nobla (1993 rok). Technika PCR w istotnym stopniu przyspieszyła i udoskonała badania nad strukturą oraz funkcją genów. W tej chwili w laboratoriach wykorzystuje się zarówno jej klasyczną wersję, jak i liczne modyfikacje. Przeprowadzenie reakcji PCR stanowi punkt wyjściowy wielu analiz prowadzonych w diagnostyce chorób genetycznie uwarunkowanych, m.in. sekwencjonowania modyfikowaną metodą Sangera czy analizy sekwencji mikrosatelitarnych. Ponadto technika ta umożliwia jakościową (RT-PCR) oraz ilościową (PCR w czasie rzeczywistym) ocenę ekspresji genów.

Reakcja PCR prowadzona jest w warunkach cyklicznie powtarzających się trzech etapów (denaturacji DNA, przyłączania starterów i wydłużania – dobudowywania drugiej nici DNA), a do jej wykonania potrzebne są chociaż częściowa znajomość sekwencji badanego fragmentu kwasu nukleinowego, odczynnik, w tym odpowiednio dobrana polimeraza DNA, oraz termocykler. Matrycą do reakcji PCR może być DNA lub RNA. W przypadku, gdy powielany jest fragment RNA, zazwyczaj konieczne jest przeprowadzenie reakcji odwrotnej transkrypcji (RT-PCR, ang. *reverse transcription PCR*), czyli przepisania sekwencji RNA na cDNA z zastosowaniem odwrotnej transkryptazy.

Specyficzne powielenie wybranego fragmentu jest możliwe dzięki zastosowaniu komplementarnej do nici DNA/



Rys. 1. Schemat powielania fragmentu DNA klasyczną metodą PCR (A) oraz TP-PCR (B)



Rys. 2. Przykładowe wyniki diagnostyki molekularnej z wykorzystaniem metody PCR-RFLP (A) oraz TP-PCR (B)

RNA pary starterów, których miejsce przyłączenia wyznacza lokalizację i wielkość analizowanego fragmentu kwasu nukleinowego. Starterami w reakcji PCR są krótkie (zwykle około 20 nukleotydów) syntetyczne sekwencje jednoniciowego DNA. W reakcji wykorzystywane są dwa startery: starter sensowny (ang. *forward*), którego sekwencja jest taka sama jak sekwencja powielana, oraz starter antysensowny (ang. *reverse*), którego sekwencja jest komplementarna do sekwencji powielanej (ryc. 1). Synteza na matrycy DNA lub cDNA katalizowana jest przez polimerazę DNA, a substratem reakcji jest mieszanina czterech dNTP (dATP, dCTP, dGTP, dTTP). Obecnie na rynku dostępnych jest wiele polimeraz różniących się wydajnością oraz dokładnością, a ich wspólną cechą jest zachowanie aktywności w szerokim zakresie temperatur (37-94°C). Optymalne warunki dla aktywności polimerazy DNA zapewnia bufor do reakcji oraz odpowiednio dobrane stężenie jonów Mg^{2+} . Optymalizacja warunków reakcji PCR zarówno w zakresie stężeń wszystkich składników mieszaniny reakcyjnej, jak i doboru warunków procesu amplifikacji jest kluczowym elementem prawidłowo wykonanej diagnostyki z zastosowaniem tej metody.

MODYFIKACJE METODY PCR

Pomimo rozwoju wielu nowych metod analizy kwasów nukleinowych łańcuchowa reakcja polimerazy jest jedną z najważniejszych technik wykorzystywanych w diagnostyce molekularnej. Na bazie jej klasycznej wersji PCR powstało wiele modyfikacji, które z powodzeniem zostały zaadaptowane na potrzeby diagnostyki

chorób genetycznie uwarunkowanych. Przykładowe modyfikacje tej metody to: analiza polimorfizmu długości fragmentów restrykcyjnych po reakcji PCR (PCR-RFLP, ang. *PCR Restriction Fragments Length Polymorphism*), TP-PCR (ang. *Triplet Repeat Primed PCR*) czy PCR w czasie rzeczywistym (qRT-PCR, ang. *quantitative Real Time PCR*).

Analiza polimorfizmu długości fragmentów restrykcyjnych – PCR (PCR-RFLP)

Metoda PCR-RFLP została zastosowana zaraz po wprowadzeniu do laboratoriów klasycznej techniki PCR i nadal jest jedną z technik stosowanych w diagnostyce molekularnej. W technice tej analizuje się długości fragmentów DNA powstałych w wyniku trawienia enzymami restrykcyjnymi (restryktazami) produktu PCR. Trawienie restryktazami jest procesem, w którego wyniku dochodzi do przecięcia podwójnej nici DNA w ściśle określonym miejscu, wyznaczanym przez specyficzną 4-8-nukleotydową sekwencję rozpoznaną przez enzymy restrykcyjne. Obecność mutacji może spowodować zanik miejsca cięcia dla enzymu restrykcyjnego lub pojawienie się nowego, co powoduje zmiany w długości fragmentów DNA uzyskanych po trawieniu produktu PCR danym enzymem restrykcyjnym (ryc. 2a). Analiza polimorfizmu długości fragmentów restrykcyjnych umożliwia rozróżnienie w materiale badanym poszczególnych alleli, w tym również wariantów, które mogą mieć charakter patogeniczny. Znacznym ograniczeniem techniki PCR-RFLP jest możliwość analizy jedynie opisanych wcześniej zmian w sekwencji DNA.

Metoda TP-PCR

Metoda TP-PCR (ang. *Triplet Repeat Primed PCR*) została opracowana dla potrzeb diagnostyki molekularnej jednostek chorobowych z grupy tzw. chorób powtórzeń trójnukleotydowych (TREDs, ang. *Trinucleotide Repeats Disorders*). Podłoże molekularne tych chorób związane jest z niestabilnością trójnukleotydowych sekwencji mikrosatelitarnych $[GAA]_n/[CAG]_n/[CGG]_n$ w określonych genach (np. w przypadku choroby Huntingtona jest to gen HTT). Niestabilność ta zwiększa prawdopodobieństwo dalszego wydłużania się obszaru podlegającego ekspansji, aż do wartości uznawanej za patogenną dla danej choroby. Metoda TP-PCR umożliwia wykazanie obecności długich, uznawanych za patogeniczne, ciągów powtórzeń nukleotydowych, trudnych do amplifikacji w standardowych warunkach PCR.

Testy molekularne w badaniach rutynowych w pierwszym etapie obejmują zazwyczaj amplifikację regionu powtórzeń klasyczną metodą PCR, która umożliwia ustalenie na podstawie wielkości uzyskanych amplikonów, czy liczba powtórzeń zawiera się w zakresie prawidłowym czy patogenicznym. W przypadku chorób takich jak dystrofia miotoniczna typu 1 i 2 (DM1, DM2), zespół kruchego chromosomu X (FXS), ataksja Friedreicha (FRDA) czy wczesna postać choroby Huntingtona (HD) liczba powtórzeń w zakresie patogenicznym jest jednak zbyt duża i analizowany fragment DNA nie ulega powieleniu w standardowych warunkach. Możliwa jest zatem tylko analiza amplikonów dla zakresu prawidłowego.

Jedną z metod rozwiązania tego problemu jest stosowanie protokołu tzw. *long range PCR* (LR-PCR), czyli reakcji PCR z zastosowaniem polimerazy *Expand Long*, która umożliwia syntezę długich amplikonów. Teoretycznie metoda ta pozwala na uwidocznienie obecności silnych ekspansji, z doświadczeń szeregu laboratoriów wynika jednak, że może prowadzić do uzyskiwania wyników niejednoznacznych.

Technika TP-PCR oparta jest o rozdział elektroforetyczny w sekwenatorze kapilarnym produktów reakcji powielenia specyficznych fragmentów zawierających powtórzenia nukleotydowe. Do ich powielenia wykorzystuje się trzy startery: wiążący się na 5' końcu badanego obszaru genu (P1, specyficzny dla *locus*), wiążący się losowo w różnych miejscach obszaru powtórzeń (o sekwencji komplementarnej dla 7-8 powtórzeń badanego rejonu, P2) oraz wiążący się do sekwencji startera P2 na jego 3' końcu (P3). Analiza PCR jest dwuetapowa, najpierw następuje amplifikacja produktów ograniczonych starterami P1 i P2. Uzyskane w ten sposób fragmenty stanowią matryce do namnożenia produktów poddawanych analizie elektroforetycznej z wykorzystaniem znakowanego fluorescencyjnie startera P3 i P1 (ryc. 1). W wyniku reakcji TP-PCR powstaje seria amplikonów, których wielkość związana jest z różnym miejscem przyłączenia startera P2 w obrębie sekwencji powtórzonej. W przypadku obecnej ekspansji otrzymujemy serię produktów/pików coraz „słabszych” w miarę wzrostu ich długości. Taki wynik uzyskujemy już przy obecności jednego nieprawidłowego allela (ryc. 2b), co w przypadku chorób o dziedziczeniu recesywnym (taką chorobą jest ataksja Friedreicha) jest wskazaniem do oceny, czy mutacja dotyczy jednego allela czy obu. Wykorzystywaną w tym przypadku metodą jest standardowa metoda PCR – obecność produktu w tej reakcji będzie świadczyć o nosicielstwie choroby, jego brak zaś będzie stanowić potwierdzenie rozpoznania klinicznego choroby. W przypadku chorób o dziedziczeniu dominującym (DM1, DM2 czy HD) obecność nieprawidłowego allela jest potwierdzeniem diagnozy klinicznej.

Ilościowy PCR w czasie rzeczywistym

Podstawową różnicą pomiędzy konwencjonalną metodą PCR a ilościowym PCR

w czasie rzeczywistym (qRT-PCR, ang. *quantitative real-time PCR*) jest możliwość ilościowej oceny powstającego produktu reakcji w trakcie jej trwania. Monitorowanie przyrostu liczby kopii badanej sekwencji w czasie rzeczywistym jest możliwe dzięki zastosowaniu znaczników w postaci barwników fluorescencyjnych (fluorochromów). W zależności od zastosowanej metodyki za pomocą fluorochromów znakowane mogą być startery, specyficzne sondy oligonukleotydowe (np. TaqMan) bądź dwuniciowy produkt amplifikacji (barwnik SYBRGreen). Wykonanie badania z zastosowaniem techniki qRT-PCR wymaga specjalistycznego sprzętu – termocyklerów sprzężonych z detektorem fluorescencji.

W zależności od kontekstu prowadzonych badań w metodzie PCR w czasie rzeczywistym stosuje się dwie metody analizy: bezwzględną (metoda absolutna) oraz względną. Bezwzględna ocena ilościowa produktu w próbce badanej prowadzona jest w oparciu o krzywą wzorcową (standardową) sporządzoną dla serii rozcieńczeń DNA lub RNA. Analiza względna prowadzona jest względem próbki lub genu referencyjnego i ma charakter analizy porównawczej.

Jedną z podstawowych zalet techniki RT-PCR jest jej wysoka czułość, która umożliwia wykrywanie nawet bardzo małych ilości materiału w próbce badanej. Ze względu na tę właściwość qRT-PCR może być wykorzystany do wykrywania DNA obcego pochodzenia (bakterii, wirusów, grzybów) i stał się metodą rutynowo stosowaną w diagnostyce mikrobiologicznej i wirusologicznej. W badaniach molekularnych chorób genetycznie uwarunkowanych PCR w czasie rzeczywistym znalazł zastosowanie jako metoda, za pomocą której można zweryfikować wyniki uzyskane innymi technikami takimi jak MLPA czy mikromacierze aCGH.

SEKWENCJONOWANIE KLASYCZNĄ METODĄ SANGERA

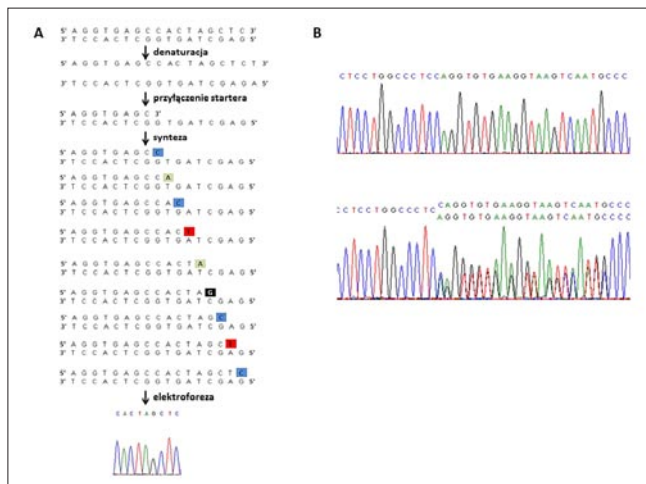
Sekwencjonowanie DNA umożliwia określenie kolejności nukleotydów w cząsteczce DNA. Pierwsze techniki sekwencjonowania DNA opracowane zostały w połowie lat 70. XX wieku. Jedną z nich była metoda terminacji łańcucha opracowana przez Sangera, której zmodyfikowaną wersję stosuje się rutynowo w laboratoriach diagnostycznych na ca-

łym świecie do dzisiaj. Metoda ta zakłada przeprowadzenie reakcji syntezy komplementarnego łańcucha DNA z wykorzystaniem jako substratów reakcji mieszaniny deoksynukleotydów (dNTP) oraz dideoksynukleotydów (ddNTP) znakowanych fluorescencyjnie. Losowe włączenie ddNTP zamiast dNTP do powstającego łańcucha powoduje zatrzymanie procesu syntezy, co prowadzi do powstania mieszaniny cząsteczek DNA o różnej długości, kończących się ddNTP znakowanymi różnymi fluorochromami. Otrzymane fragmenty poddaje się rozdzielaniu elektroforetycznemu w sekwenatorze kapilarnym, a następnie odczytuje otrzymaną sekwencję DNA. Zastosowanie fluorochromów i sekwenatorów kapilarnych pozwoliło na zautomatyzowanie procesu odczytywania sekwencji i znaczną redukcję czasu analizy.

W praktyce metoda Sangera umożliwia odczyt około 800 nukleotydów w jednej reakcji. Usprawnienia te w połączeniu z bardzo wysoką (99%) czułością metody sprawiły, że w laboratoriach diagnostycznych sekwencjonowanie DNA zaczęło konkurować z innymi metodami biologii molekularnej. Niewątpliwie dodatkową zaletą sekwencjonowania jest także fakt, że umożliwia ono zarówno wykrycie znanych mutacji patogenicznych, jak i identyfikację nowych zmian nieopisywanych wcześniej w piśmiennictwie. Dzięki temu badając np. ekson 73 genu *COL7A1* (mutacje tego genu odpowiedzialne są za dystroficzną postać pęcherzowego oddzielania się naskórka) celem identyfikacji częstszej mutacji p.Gly2043Arg, możliwe jest jednoczesne stwierdzenie bądź wykluczenie obecności 44 innych mutacji opisanych w tym eksonie genu, jak również wykrycie nowych zmian.

W przypadku większości chorób o podłożu genetycznym, za które są odpowiedzialne mutacje specyficznych genów, nie są dostępne komercyjnie żadne testy diagnostyczne umożliwiające identyfikację najczęstszych zmian. Co więcej, nawet gdy takie testy zostały opracowane, często przeszkodą w ich wykorzystaniu jest fakt, iż przygotowano je dla populacji innych niż polska, i w związku z tym panel mutacji występujących najczęściej w naszym kraju może znacznie odbiegać od zawartego w komercyjnym zestawie.

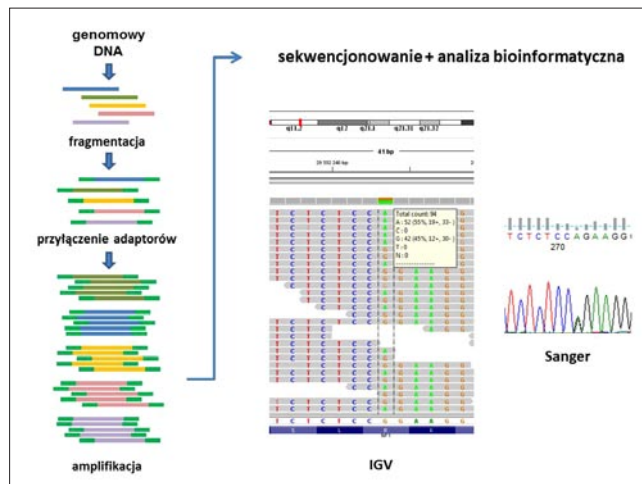
OD REAKCJI PCR DO FLUOROGRAMU



Rys. 3. Schemat reakcji sekwencjonowania metodą Sangera z zastosowaniem znakowanych fluorescencyjnie dideoksynukleotydów (A) oraz przykładowe fluorogramy (B)

Pierwszym etapem analizy molekularnej mającej na celu sekwencjonowanie próbki DNA pacjenta jest, podobnie jak w większości innych metod stosowanych w badaniach genetycznych, wykonanie reakcji PCR. Następnie produkt reakcji amplifikacji jest oczyszczany i poddawany reakcji sekwencjonowania. Reakcję sekwencyjnego PCR przeprowadza się w termocyklerze, a sam schemat reakcji przypomina tradycyjny program PCR (denaturacja, przyłączanie starterów, wydłużenie fragmentu DNA). Jedyną różnicą jest obecność w reakcji tylko jednego startera, w związku z czym w reakcji powstaje mieszanina znakowanych fluorescencyjnie produktów różnej wielkości zawierających na końcu 3' określony dla danej pozycji nukleotydowej rodzaj ddNTP (ryc. 3). Po zakończeniu reakcji próbkę oczyszcza się (usunięte zostają niewykorzystane substraty reakcji), denaturuje termicznie i poddaje się elektroforezie kapilarnej w sekwenatorze. Wynik sekwencjonowania stanowi fluorogram, na którym widoczne są sygnały od barwników fluorescencyjnych, którymi wyznakowane były poszczególne ddNTP.

Analiza fluorogramu wykonywana na potrzeby diagnostyki jest etapem wymagającym wiedzy i doświadczenia z zakresu biologii molekularnej i genetyki. Analiza ta ma na celu określenie, czy w analizowanym fragmencie DNA znajduje się zmiana w sekwencji nukleotydów względem sekwencji referencyjnej oraz – w przypadku stwierdzenia takiej zmiany – prawidłowe jej nazwanie i interpretowanie w kontekście klinicznym. Analizę fluorogramów należy wykonywać bardzo dokładnie, najlepiej posilując się



Rys. 4. Schemat przedstawiający najważniejsze etapy metody sekwencjonowania następnej generacji (przedstawiono wynik graficzny z programu Integrative Genomics Viewer uzyskany dla zmiany w genie *NF1* oraz fluorogram potwierdzający obecność zmiany)

programami bioinformatycznymi i odnosząc się do odpowiedniej sekwencji referencyjnej.

INTERPRETACJA WYNIKÓW SEKWENCJONOWANIA

Seqwencjonowanie umożliwia wykrycie substytucji nukleotydowych, insercji, delecji czy zmian złożonych o różnej wielkości (od 1 pary zasad do nawet kilkuset). W przypadku identyfikacji jakiegokolwiek zmiany niezbędne jest jej prawidłowe nazwanie. Zastosowanie prawidłowej nomenklatury jest także niezwykle istotne ze względu na fakt, że analizy genetyczne są przeważnie kontynuowane u krewnych pacjenta (np. badanie nosicielstwa). Nie zawsze analizy te przeprowadzane są w tym samym laboratorium, w którym badano probanda, co oznacza, iż laboratorium to może nie dysponować materiałem porównawczym. Jeśli zatem podana nazwa mutacji jest nieprawidłowa, istnieje ryzyko, że u krewnych pacjenta wykonane zostanie badanie innego regionu genu, czego konsekwencją może być uzyskanie wyniku niezgodnego ze stanem faktycznym. Aktualnie obowiązujące reguły dotyczące nazewnictwa mutacji zebrane są na stronie internetowej organizacji Human Genome Variation Society.

Pośród szeregu wytycznych ogromne znaczenie ma sposób numerowania kolejnych nukleotydów w sekwencji DNA. Nukleotydem nr 1 jest zawsze nukleotyd A z pierwszego kodonu (a więc ATG) sekwencji kodującej.

Dalsze etapy analizy mają na celu określenie statusu patogenności zidentyfikowanej mutacji. Określenie znaczenia danej zmiany w kontekście klinicznym

jest często wyzwaniem bardzo trudnym i zdarzają się przypadki niejednoznacznej interpretacji. Pomocny na tym etapie jest szereg baz danych, m.in. Human Gene Mutation Database, Ensembl, The 1000 Genomes Project, jak również – mające jednakże znaczenie wyłącznie indykacyjne – niektóre programy bioinformatyczne, np. SIFT, Mutation Taster czy PolyPhen2.

SEKWENCJONOWANIE NASTĘPNEJ GENERACJI

Zmniejszające się koszty sekwencjonowania metodą Sanger, możliwość analizy większej liczby próbek oraz zwiększenie dostępności do aparatury sprawiło, że sekwencjonowanie tzw. pierwszej generacji jest obecnie techniką wykorzystywaną rutynowo w badaniach diagnostycznych prowadzonych przez laboratoria genetyczne. Nie zmienia to faktu, że metoda ta, mimo wprowadzenia sekwenatorów, umożliwiających jednoczesną analizę 96 amplikonów, jest w dalszym ciągu techniką niskoprzepustową, wymagającą skupienia badań molekularnych na konkretnym regionie DNA. Tym niemniej technika ta została z sukcesem zastosowana do zsekwencjonowania całego genomu człowieka w ramach Human Genome Project (wyniki zostały opublikowane w 2001 roku); koszt badań wynosił 2,7 miliarda dolarów, zaś samo przeprowadzenie analizy zajęło 13 lat. Od tego czasu, dzięki intensywnemu rozwojowi technik molekularnych, obserwujemy znaczne usprawnienie i obniżenie kosztów sekwencjonowania genomu człowieka. Nowe techniki, które określa się wspólną nazwą sekwencjonowania następnej ge-

neracji (ang. *Next Generation Sequencing*, NGS) umożliwiają zsekwencjonowanie całego genomu człowieka w ciągu kilku dni, zaś koszt tej analizy jest niższy niż 10 tys. dolarów i stale maleje, zbliżając się do kwoty 1000 dolarów.

Sekwencjonowanie następnej generacji jest głównie utożsamiane z sekwencjonowaniem DNA genomowego. Jednakże technika ta znajduje także zastosowanie w analizie transkryptomu (sekwencjonowanie całego RNA znajdującego się w komórkach tzw. RNA-seq), epigenomu (analiza modyfikacji epigenetycznych genomu – zmian w metylacji sekwencji) czy mikrobiomu (analiza zmienności składu flory bakteryjnej). Wykorzystane jest także w analizie oddziaływań DNA z białkami po uprzedniej immunoprecypitacji chromatyny (ang. *chromatin immunoprecipitation sequencing*, ChIP-seq). Istotnym zastosowaniem techniki NGS jest również sekwencjonowanie *de novo* genomów różnych organizmów. NGS, podobnie jak inne techniki analizy kwasów nukleinowych, szybko znajduje swoje praktyczne zastosowanie w diagnostyce genetycznej.

Technika sekwencjonowania następnej generacji polega na jednoczesnym sekwencjonowaniu wielu milionów cząsteczek kwasów nukleinowych, pojedynczych lub poddanych amplifikacji techniką PCR, oddzielonych od siebie przestrzennie, z wykorzystaniem superczułych detektorów, np. światła fluorescencyjnego czy zmian pH. Dzięki tej technologii możliwe jest zsekwencjonowanie całego genomu (ang. *Whole Genome Sequencing*, WGS) – sekwencji kodujących i niekodujących, co prowadzi do uzyskania dużej ilości danych, których jeszcze nie jesteśmy w stanie w pełni właściwie zinterpretować. Analizę danych znacznie ułatwia ograniczenie badanego obszaru genomu jedynie do sekwencji kodującej tzw. eksomu, który stanowi jedynie 1% całości DNA w komórce. Podejście takie jest określane terminem sekwencjonowania eksomowego (ang. *Whole Exome Sequencing*, WES). Ponadto metoda NGS pozwala na jednoczesną analizę jednego lub kilku genów, co jest szczególnie przydatne w przypadku chorób o dużej heterogenności genetycznej (mutacje w kilku różnych genach mogą być odpowiedzialne za wystąpienie specyficznych objawów choroby) i dobrze scharakteryzowanym podłożu molekularnym. Podejście takie

określa się mianem sekwencjonowania celowanego (ang. *targeted sequencing*), w trakcie którego będzie wykonywana jest analiza jedynie tych genów, których mutacje związane są z patogenizacją danej choroby.

! OGÓLNY ZARYS PROCEDURY

Sekwencjonowanie następnej generacji jest procedurą wieloetapową, rozpoczynającą się od przygotowania matrycy tzw. biblioteki do sekwencjonowania, po czym następuje właściwa reakcja sekwencjonowania, która może trwać od kilku godzin do kilkunastu dni – w zależności od typu aparatu i zastosowanej techniki sekwencjonowania. Najtrudniejszym etapem NGS jest bioinformatyczna analiza otrzymanych danych, na podstawie których otrzymuje się listę wariantów – zmian w sekwencji DNA specyficznych dla danego pacjenta.

Pierwszym etapem, jak wspomniano wyżej, jest przygotowanie biblioteki do sekwencjonowania. Materiał do analizy stanowi zazwyczaj DNA, ewentualnie RNA, który poddaje się fragmentacji z wykorzystaniem metod enzymatycznych, chemicznych, sonikacji lub nebulizacji. Do powstałych fragmentów przyłączane są specyficzne oligonukleotydowe adaptory, które wykorzystywane są w reakcjach amplifikacji lub na późniejszych etapach analizy służą do identyfikacji sekwencji dla poszczególnych pacjentów. Na tym etapie procedury możliwe jest także przeprowadzenie wzbogacania sekwencji, jeśli chcemy analizować eksom (tzw. *exome enrichment*) lub wybrane geny (tzw. *target gene enrichment*).

Kolejnym etapem procedury jest amplifikacja fragmentów DNA w przypadku technik NGS wymagających powielenia materiału genetycznego, określanego jako sekwencjonowanie tzw. drugiej generacji. Na tym etapie powielane są fragmenty DNA z wykorzystaniem starterów komplementarnych do sekwencji adaptorowych. Sama amplifikacja może odbywać się z wykorzystaniem techniki emulsyjnego PCR – reakcja amplifikacji pojedynczej cząsteczki DNA odbywa się w mikroreaktorze powstałym w wyniku zawieszenia specjalnego oleju w wodzie. Alternatywą do tej metody może być amplifikacja mostkowa, w której fragment DNA wiąże się z unieruchomionym na płytce oligonukleotydem komplementarnym do zastosowanych adapto-

rów, a następnie jest powielany, tworząc zgrupowanie identycznych fragmentów DNA. Właściwa reakcja sekwencjonowania amplikonów unieruchomionych na nośniku (mikropłytki) może być także prowadzona z wykorzystaniem różnych metod. Do najczęściej stosowanych należą sekwencjonowanie przez syntezę z wykorzystaniem odwracalnych terminatorów reakcji, pirosekwencjonowanie oraz sekwencjonowanie przez syntezę z wykorzystaniem technologii półprzewodnikowej.

Ostatnim, najtrudniejszym, etapem metody NGS jest analiza bioinformatyczna wyników, która obejmuje: dopasowanie otrzymanych sekwencji do tzw. genomu referencyjnego, analizę sekwencji pod kątem obecności zmian – podstawień pojedynczych nukleotydów lub insercji/delekcji. Kolejną czynnością jest opisanie zidentyfikowanych zmian, m.in. ich dokładnej lokalizacji w genomie/genie, określenie ich zygoticzności, odniesienie zmiany do istniejących baz danych, np. baz zidentyfikowanych już polimorfizmów (zmian w sekwencji DNA, które nie będą unikatowe dla danego pacjenta np. dbSNP, 1000 Genomes), określenie charakteru zmiany i jej ewentualnego wpływu na produkt białkowy danego genu, w przypadku zmian zidentyfikowanych w sekwencjach kodujących.

Wynikiem końcowym sekwencjonowania następnej generacji nie jest, jak w przypadku klasycznego sekwencjonowania Sangera, fluorogram lub tekstowy zapis sekwencji DNA, tylko zazwyczaj tabela, zawierająca zmiany nukleotydowe zidentyfikowane u pojedynczego pacjenta. Możliwe jest wykorzystanie specjalnych programów np. Integrative Genomics Viewer do graficznego przedstawienia zidentyfikowanych zmian, jednak ich liczba jest zazwyczaj zbyt duża, aby manualnie przejrzeć otrzymane wyniki. Szacuje się, że sekwencjonowanie całego genomu pacjenta umożliwia identyfikację 3,5 mln zmian w sekwencji DNA, podczas gdy ograniczenie analizy wyłącznie do sekwencji kodujących zmniejsza tę liczbę do 20 tys., z czego 100-200 wariantów jest unikatowe dla badanej osoby.

! NGS W DIAGNOSTYCE CHOROÓB MONOGENOWYCH

Znaczącą zaletą sekwencjonowania następnej generacji w odniesieniu do klasycznego sekwencjonowania metodą

Sangera jest to, że w analizie nie musimy skupiać się na konkretnych regionach chromosomowych. Nawet jeśli nie znamy genu, którego mutacje są odpowiedzialne za wystąpienie objawów choroby, sekwencjonowanie genomu czy eksomu może pomóc w identyfikacji defektu molekularnego w nowym genie. Podejście to umożliwiło identyfikację wielu nowych genów, których mutacje odpowiedzialne są za choroby monogenowe. Pierwsze z opublikowanych doniesień opisujących identyfikację nowych genów z wykorzystaniem techniki NGS dotyczyły: zespołu Freemana-Sheldona (gen *MYH3*), zespołu Millera (gen *DHODH*), zespołu Schinzla-Giediona (gen *SETBP1*), zespołu Kabuki (gen *MLL2*) czy nieswoistego zapalenia jelit przypominającego zespół Crohna (gen *XIAP*). W chwili obecnej pojawia się coraz więcej doniesień dotyczących techniki NGS i identyfikacji molekularnych przyczyn chorób genetycznie uwarunkowanych. Ponadto w niektórych krajach (USA, Kanada) sekwencjonowanie eksomu jest wdrażane do rutynowej diagnostyki, gdyż pozwala na szybką identyfikację defektu molekularnego, a tym samym – szybsze postawienie odpowiedniej diagnozy klinicznej, co z kolei umożliwia wdrożenie odpowiedniego leczenia czy rozpoczęcie rehabilitacji. Nie zmienia to faktu, że sekwencjonowanie następnej generacji w dalszym ciągu pozostaje techniką kosztowną, wymagającą zakupu specjalnej aparatury oraz posiadania sprzętu komputerowego i personelu, który będzie w stanie przeprowadzić niezbędną analizę bioinformatyczną otrzymanych danych. W Polsce, technika sekwencjonowania następnej generacji jest wykorzystywana na razie głównie w badaniach naukowych i wydaje się, że już wkrótce może zostać wykorzystana także w rutynowej diagnostyce chorób genetycznie uwarunkowanych.

**Piśmiennictwo dostępne na stronie
www.laboratorium.elamed.pl**