

Wprowadzenie do biologii molekularnej: podstawowe terminy i definicje

Adam SZPECHCIŃSKI

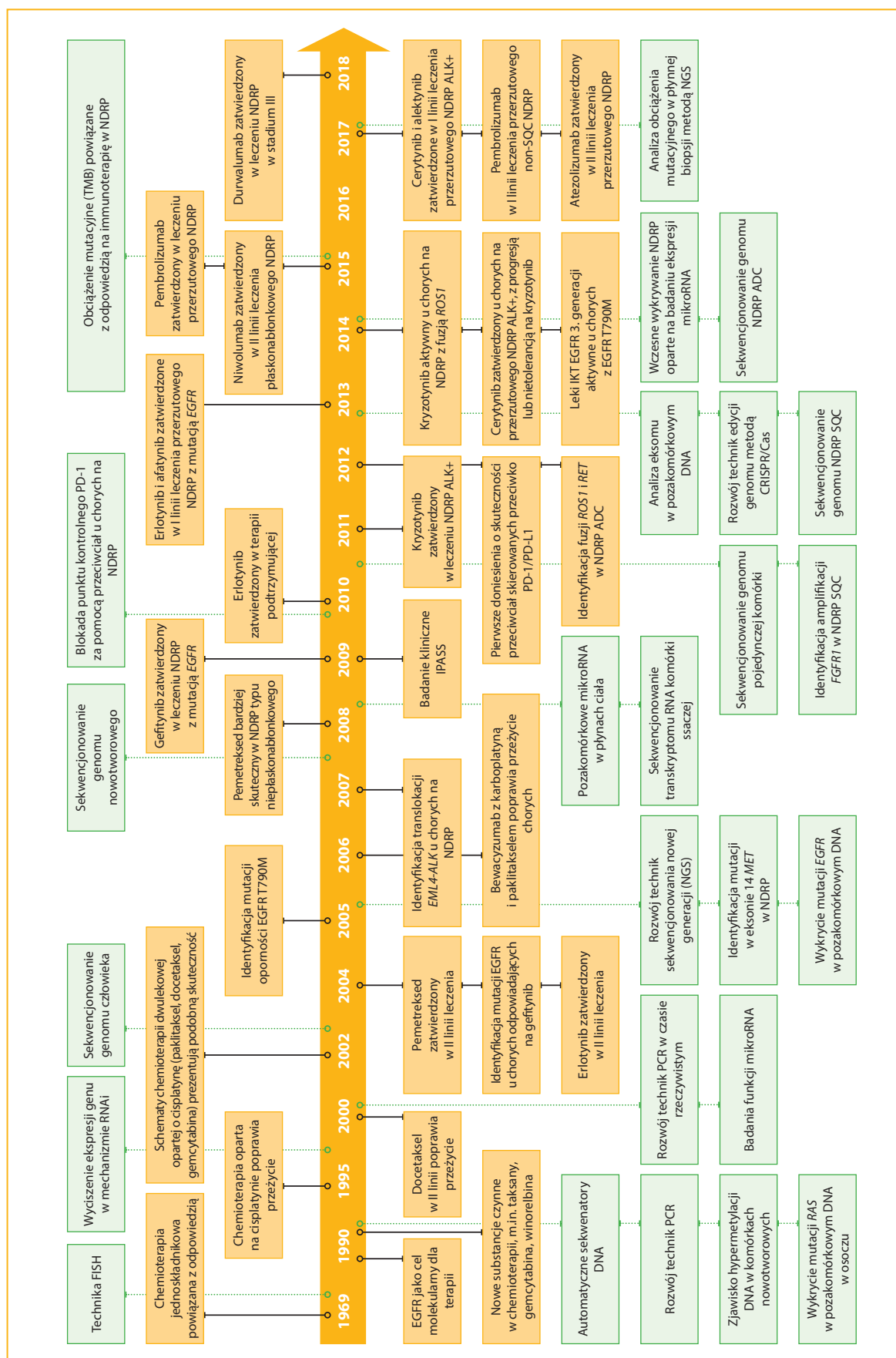
Zakład Genetyki i Immunologii Klinicznej,
Instytut Gruźlicy i Chorób Płuc,
Warszawa

Wprowadzenie

W ostatnich 25 latach nastąpił intensywny rozwój biologii molekularnej, który przełożył się na szybki postęp badań podstawowych i translacyjnych nad mechanizmami kancerogenezy, co z kolei zaowocowało opracowaniem nowych strategii terapeutycznych z wykorzystaniem leków ukierunkowanych na konkretne cele molekularne w komórce nowotworowej. W przypadku nowotworów płuc przełomem były zidentyfikowanie różnic między typami i podtypami histologicznymi również na poziomie molekularnym oraz koncepcja raka płuca jako choroby o podłożu genetycznym. Od tej pory nieodłącznym partnerem onkologii klinicznej pozostaje diagnostyka molekularna, która właśnie przeżywa renesans związany z wdrażaniem nowej generacji technologii analitycznych (ryc. 1). Każdego roku ukazują się tysiące artykułów naukowych prezentujących odkrycia z zakresu biologii komórki na poziomie molekularnym, weryfikowane są stare hipotezy i tworzone są nowe, które stawiają wyzwania przed kolejnymi grupami badawczymi. Aktualizowana jest także wiedza stanowiąca fundament biologii molekularnej – definicje podstawowych pojęć dotyczących kwasów nukleinowych i procesów z nimi związanych, struktury i funkcji genu czy zjawisk genetycznych i epigenetycznych regulujących syntezę białka w komórce prawidłowej i nowotworowej.

B iologia molekularna to dziedzina biologii, która bada przebieg i regulację podstawowych procesów życiowych komórki na poziomie cząsteczek, czyli molekuł, ze szczególnym naciskiem na oddziaływanie między kwasami nukleinowymi (DNA i RNA) a białkami, jak również na ich biosyntezę i funkcjonowanie. Naukowcy prowadzący badania z zakresu biologii molekularnej stosują szereg metod typowych dla swojej dziedziny: techniki PCR (ang. *polymerase chain reaction*, reakcja łańcuchowa polimerazy), sekwencjonowanie DNA, elektroforezę żelową. Bardzo często korzystają także z technik analitycznych oraz wiedzy z innych dziedzin, przede wszystkim genetyki, biochemii i biofizyki. Ostatnie lata to okres wyłaniania się specjalizacji biologii molekularnej – poświęconych badaniu określonych struktur, zjawisk czy obszarów tematycznych, przykładowo genetyki molekularnej, inżynierii genetycznej, genomiki, transkryptomiki, epigenetyki, proteomiki, bioinformatyki. Rozwój technik biologii molekularnej pozwolił ponadto na wyodrębnienie nowej gałęzi diagnostyki medycznej – diagnostyki molekularnej.

D iagnostyka molekularna opiera się przede wszystkim na analizie struktury kwasów nukleinowych. Pozwala rozpoznawać choroby uwarunkowane nieprawidłową budową i funkcją genów (np. obecnością mutacji genowych) i oceniać predyspozycje do wystąpienia pewnych schorzeń w ciągu życia (np. dziedziczna predyspozycja do raka piersi). Pomaga też podejmować decyzje kliniczne, dotyczące wyboru i przebiegu leczenia na podstawie statusu ilościowego i/lub jakościowego określonych markerów genetycznych (np. terapia inhibitorami kinazy tyrozynowej EGFR u chorych na NDRP – niedrobnokomórkowego raka płuca), czyli miejsc w sekwencji



Ryc. 1. Historia najważniejszych odkryć i dokonań w zakresie leczenia raka płuca (pola pomarańczowe) oraz towarzyszący im rozwój biologii molekularnej nowotworów i metod diagnostyki *in vitro* (pola zielone)

jądrowego DNA. Diagnostyka molekularna czerpie z metodyki i rezultatów badań podstawowych i translacyjnych z zakresu biologii molekularnej oraz nowych technologii analitycznych, które są optymalizowane i walidowane w warunkach laboratoryjnych z myślą o zastosowaniu w przypadkach konkretnych patologii celem uzupełnienia klinicznego rozpoznania choroby. Wobec dostępności nowych terapii ukierunkowanych molekularnie (terapii celowanych) w szeregu chorób nowotworowych, włączając NDRP, diagnostyka molekularna wraz z rozpoznaniem patomorfologicznym odgrywają kluczową rolę w kwalifikacji pacjentów do różnych schematów terapeutycznych i określaniu mechanizmu nabytej oporności lekowej w trakcie progresji choroby, który ma znaczenie przy wyborze kolejnej linii leczenia. Przedmiotem molekularnego badania diagnostycznego jest struktura i/lub funkcja genów o szczególnym znaczeniu w patomechanizmie danej choroby, co można wykorzystać jako biomarker diagnostyczny, predykcyjny lub prognostyczny.

Gen to pojęcie używane powszechnie w biologii i medycynie, niemniej jego definicja nie jest jednoznaczna ani ostatecznie ustalona. Termin „gen” wprowadził botanik Wilhelm Johannsen w 1909 roku – jako pojęcie teoretyczne związane z koncepcją dziedziczenia cech opracowaną przez Grzegorza Mendla w roku 1866. Mendel pisał o jednostkach dziedziczności i zauważał, że gamety mają takich jednostek po 50%, a jednostki mogą mieć charakter dominujący, recesywny bądź mieszany. Johannsenowski termin „gen” zastąpił wszystkie wcześniej stosowane określenia: „jednostki”, „elementy”, „czynniki wpływające na dziedziczenie cech” i inne o podobnym znaczeniu. Obecnie definicja genu rozwijana jest na dwóch poziomach – jako gen ewolucyjny oraz gen molekularny.

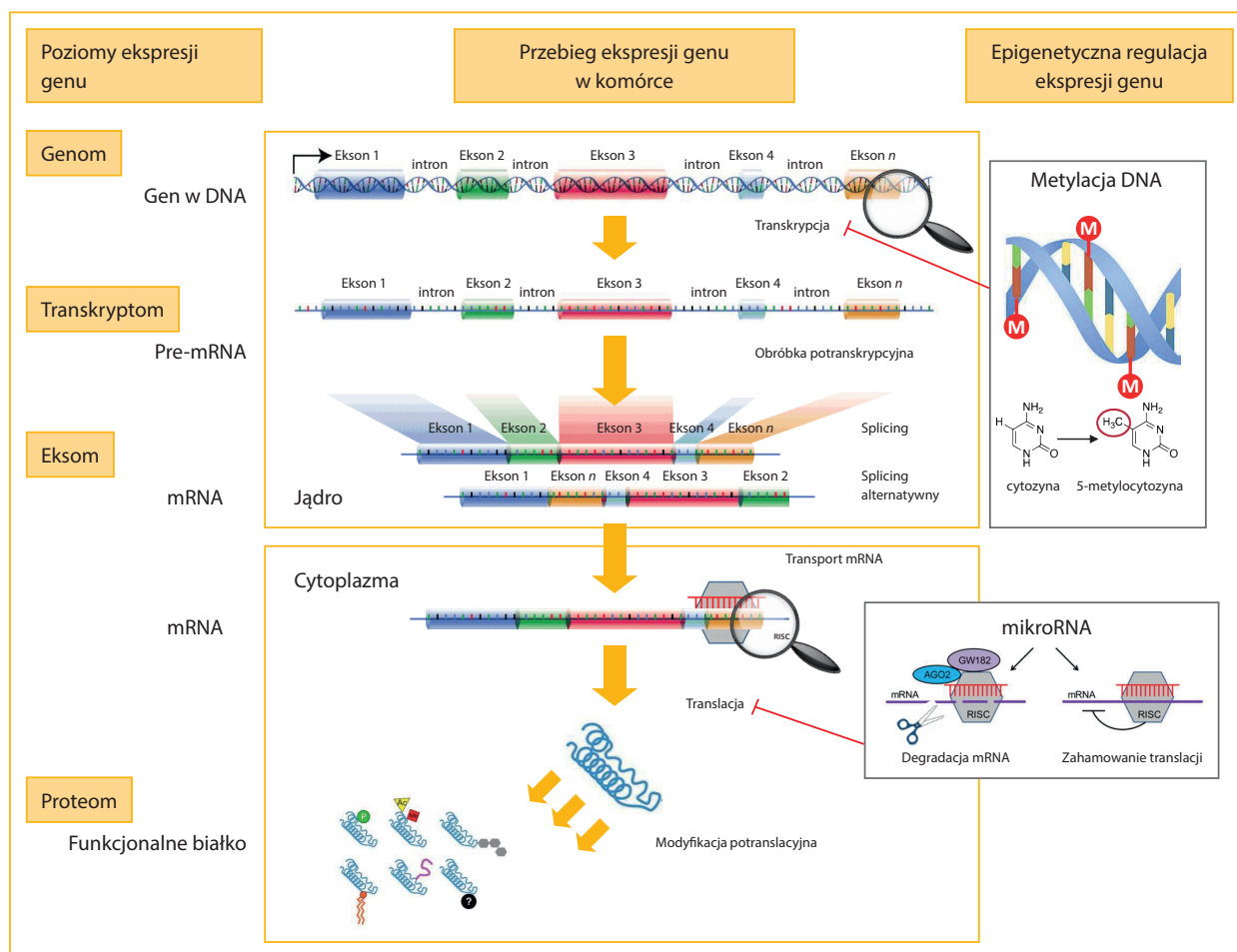
W ujęciu ewolucyjnym geny to podstawowe jednostki dziedziczności, będące konkurującymi odcinkami kwasu nukleinowego i zajmujące ściśle określone miejsca w chromosomie, czyli *locus* (przykładowo zapis „*locus* 6p21.3” należy rozumieć następująco: 6 – numer chromosomu, p – krótkie ramię chromosomu, 2 – region na danym ramieniu chromosomu, 1 – prążek w danym regionie, widoczny pod mikroskopem, 3 – pozycja subprążka w badaniu cytogenetycznym), których reprezentacja w kolejnych pokoleniach ulega zmianie (zgodnie z prawami Mendla). W określonym *locus* mogą się znaleźć różne warianty związanego z nim genu (różniące się sekwencją nukleotydów w DNA), nazywane allelami. Istnienie więcej niż jednej wersji danego genu określa się jako polimorfizm. Dzięki zdegenerowaniu kodu genetycznego (jeden

aminokwas może być kodowany przez kilka kodonów, inaczej trypletów nukleotydowych, np. kodony GGC, GGA, GGG kodują glicynę w sekwencji DNA i mRNA), tylko część wspomnianych różnic przekłada się na różnice w budowie kodowanych białek. Wpływa to na zróżnicowanie właściwości cząsteczek białka kodowanego przez różne allele jednego genu. Gen ewolucyjny jest przedmiotem badań genetyki klasycznej, populacyjnej i medycznej, która bada przyczyny i sposób przekazywania chorób dziedzicznych.

W diagnostyce molekularnej chorób nowotworowych za geny uważa się odcinki DNA ulegające ekspresji realizowanej przez syntezę RNA lub białek, a więc uznawana jest tu przede wszystkim koncepcja genu molekularnego. Większość genów eukariotycznych znajduje się w jądrze komórkowym, w DNA chromosomów, istnieją jednak geny zawarte w DNA mitochondriów (u ssaków DNA mitochondrialny jest dziedziczony w linii matczynej, co powoduje segregację genów niezgodną z prawami Mendla – dziedziczenie pozachromosomowe). Ze względu na funkcje rozróżnia się geny struktury (strukturalne), zawierające informację o syntezie białek, i geny regulacyjne (regulatory), regulujące aktywność genu struktury.

U organizmów eukariotycznych sekwencja genu często jest nieciągła i składa się z odcinków kodujących – egzonów (inaczej eksonów), zawierających właściwą informację o budowie białka w postaci kolejno ułożonych kodonów, odpowiadających sekwencji aminokwasów w białku – które przedzielone są niekodującymi odcinkami DNA, intronami. Najnowsze badania wskazują na udział intronów w regulacji procesu syntezy białek.

Ekspresja genu (ang. *gene expression*) to wieloetapowy, złożony proces biochemiczny, w którym informacja genetyczna zawarta w genie zostaje odczytana i przepisana na jego produkty – białka lub różne formy RNA. W pierwszym etapie ekspresji genu, podczas transkrypcji, zachodzi przepisywanie informacji zawartej w jądrowym (lub mitochondrialnym) DNA na RNA, który stanowi roboczą kopię informacji genetycznej, transportowaną z jądra do cytoplazmy komórki, gdzie zachodzą dalsze etapy syntezy białka (ryc. 2). Zanim jednak RNA opuści jądro komórkowe, pierwotny transkrypt, nazywany pre-mRNA i będący wierną kopią sekwencji danego genu w jądrowym DNA (zawiera zarówno egzony, jak i introny), poddawany jest szeregowi modyfikacji enzymatycznych, w których wyniku introny są usuwane, a egzony – łączone w ciągłą nić mRNA w procesie splicingu. Niekiedy egzony z pre-mRNA łączą się ze sobą na różne sposoby, niekoniecznie po kolei (według genu), z pominięciem



Ryc. 2. Przebieg ekspresji genu w komórce z uwzględnieniem najważniejszych epigenetycznych mechanizmów regulacyjnych

niektórych egzonów albo z zachowaniem niektórych intronów (splicing alternatywny), dzięki czemu z jednego genu może powstać więcej niż jedna cząsteczka mRNA, a w efekcie – wiele różnych białek.

Kolejnym kluczowym etapem ekspresji genu jest translacja, czyli przepisanie informacji genetycznej zawartej w roboczej kopii kodu genetycznego na sekwencję aminokwasów w łańcuchu polipeptydowym tworzących konkretną strukturę białka. Nowo powstałe białka poddawane są licznym potranslacyjnym modyfikacjom biochemicznym, które warunkują ich właściwości chemiczne, fizyczne, stabilność i aktywność enzymatyczną. Modyfikacja potranslacyjna jako następny – po alternatywnym splicingu – sposób na zwiększenie liczby różnych białek kodowanych przez jeden gen, może obejmować następujące procesy: proteolizę (proteazy usuwają zbędne fragmenty), N-acetylację, N-metylację, N-formylację (dołączenie grupy acetylowej, metylowej i formylowej), hydroksylację (dołączenie grupy hydroksylowej –OH),

fosforylację (aktywacja białka przez dołączenie grupy fosforanowej), glikozylację (enzymatyczne przyłączenie reszt cukrowych) czy ubiquitylację (przyłączenie ubiquityny i późniejsza degradacja). Zakłada się, że sekwencja jednego genu może być źródłem aż 100 różnych białek.

Ekspresja genu może podlegać regulacji na każdym z etapów za pomocą rozmaitych mechanizmów, np. poprzez sekwencje regulatorowe genu, białka wiążące DNA, metylację DNA, niekodujący RNA (mikroRNA, lncRNA i inne). Dwa ostatnie mechanizmy regulatorowe są przedmiotem szczególnego zainteresowania biologów molekularnych w kontekście zjawisk epigenetycznych związanych z procesem kancerogenezy – zarówno w badaniach podstawowych, mających na celu poznanie roli tych mechanizmów w wyciszaniu ekspresji genów supresji nowotworowej (ang. *tumor suppressor gene*, TSG), jak i w badaniach translacyjnych, ukierunkowanych na rozwój nowych biomarkerów molekularnych, specyficznych względem danego typu nowotworu.

Genom to termin określający jądrowy materiał genetyczny zawarty w podstawowym (haploidalnym) zespole chromosomów. Genom obejmuje nie tylko geny (regiony kodujące białka), ale też całość niekodującego DNA zlokalizowanego między genami. Niewielka część mitochondrialnego DNA, która koduje 37 genów odpowiedzialnych za prawidłowe funkcjonowanie łańcucha oddechowego w komórce, zwykle nazywana jest genomem mitochondrialnym. Analizę genomu danego organizmu zajmuje się dziedzina biologii molekularnej zwana genomiką. Podstawowym narzędziem analizy genomu jest metoda sekwencjonowania DNA. Obecnie wykorzystuje się przede wszystkim techniki sekwencjonowania nowej generacji (ang. *next-generation sequencing*, NGS), pozwalające na odczytanie sekwencji całego genomu w ciągu jednego cyklu pracy urządzenia. Wyniki sekwencjonowania DNA poddawane są analizie bioinformatycznej.

Transkryptom to zestaw wszystkich cząsteczek RNA (transkryptów) obecny w określonym momencie w komórce, grupie komórek lub organizmie, **eksom** – ogół sekwencji kodujących genów, czyli egzonów w jądrowym DNA, ulegający transkrypcji do mRNA (po usunięciu intronów) i translacji do postaci funkcjonalnych białek, a **proteom** – zestaw białek występujących w komórce, białkowy

odpowiednik genomu. W przeciwieństwie do genomów, transkryptomy i proteomy nieustannie zmieniają się pod względem ilości i składu w odpowiedzi na różne czynniki. Badaniem transkryptomu i proteomu zajmują się dziedziny nauki nazywane odpowiednio transkryptomiką i proteomiką. Badania transkryptomiczne polegają najczęściej na określaniu poziomu ekspresji genów – ilości ich transkryptów w komórkach. Do badań wykorzystuje się często techniki pozwalające na wykrywanie i ocenianie tysięcy typów cząsteczek RNA, pochodzących z różnych genów, np. mikromacierze RNA (ang. *RNA microarray*) oraz techniki NGS zwane RNAseq. W diagnostyce molekularnej nowotworów wykrywa się i analizuje zaburzenia molekularne w genach na wszystkich poziomach ich ekspresji – od DNA, przez mRNA, do poziomu białka; coraz częściej stosowane są tu wysokoprzepustowe techniki sekwencjonowania genomowego (ang. *whole-genome sequencing*, WGS) i eksomowego (ang. *whole-exome sequencing*, WES), oparte na metodzie NGS (tabela 1).

Projekt poznania ludzkiego genomu (ang. Human Genome Project; <https://www.genome.gov/>) to ukończony w 2003 roku, międzynarodowy program naukowy, mający na celu poznanie sekwencji wszystkich komplementarnych par zasad tworzących ludzki genom. Celem projektu było nie tylko odczytanie miliardów par

Poziom ekspresji genu	Zaburzenia molekularne		
	Mutacje genowe	Rearanżacje genów	Amplifikacja genów
gen	<i>EGFR, BRAF, KRAS, ALK</i>	<i>ALK, ROS1, RET, FGFR</i>	<i>MET, FGFR</i>
DNA	sekwencjonowanie DNA real-time PCR (qPCR) NGS (<i>targeted sequencing</i>)	FISH NGS	FISH MLPA real-time PCR NGS
RNA	RT-qPCR NGS (RNA-Seq)	RT-qPCR (transkrypty fuzyjne) NGS (RNA-Seq)	RT-qPCR (nadekspresja mRNA)
białko	IHC (przeciwciała specyficzne względem zmutowanego białka) spektrometria mas	IHC (ekspresja białka fuzyjnego)	IHC (nadekspresja białka)

Tabela 1. Metody diagnostyki molekularnej raka płuca z podziałem na poziomy ekspresji genu oraz typ wykrywanych zaburzeń genetycznych

komplementarnych zasad składających się na DNA człowieka z minimalnym prawdopodobieństwem błędu, lecz także zidentyfikowanie i zliczenie funkcjonalnych genów (m.in. dzięki charakterystycznym sekwencjom promotorowym, zlokalizowanym przed regionem kodującym w genie). Liczbę genów kodujących białka oszacowano wtedy na około 22 300, a wielkość genomu ludzkiego – na około 3,2 mld par zasad. Ponadto odkryto, że w ludzkim genomie jest znacznie więcej tzw. segmentowych duplikacji DNA (powtarzających się fragmentów DNA o prawie identycznej sekwencji), niż pierwotnie zakładano. Ukończenie Human Genome Project przez dwa niezależne zespoły badawcze i opublikowanie danych nie tylko stanowiło naukowy przełom w dziedzinie genetyki człowieka, ale także znacząco przyczyniło się do rozwoju metod biologii molekularnej oraz stworzyło fundamenty współczesnej genomiki i bioinformatyki. Rozpoczęto wówczas szereg kolejnych ambitnych przedsięwzięć naukowych, mających na celu zweryfikowanie poprzednich wyników, pogłębienie uzyskanej wiedzy dzięki nowocześniejszym, wydajniejszym i tańszym w eksploatacji technikom badawczym (rozwój technik sekwencjonowania genomowego, nowe metody edycji genomu) oraz rozłożenie działań na kilka równoległych projektów – skupionych wokół zmienności genomu ludzkiego w światowych populacjach, analizy proteomu czy poznania różnic między genomami osób zdrowych i chorych na różne typy nowotworów.

Międzynarodowa inicjatywa badawcza The 1000 Genomes Project (w skrócie: 1KGP; <http://www.internationalgenome.org/>), rozpoczęta w 2008 roku, ma na celu zsekwencjonowanie genomów co najmniej tysiąca anonimowych uczestników, przedstawicieli różnych grup etnicznych, m.in. ludu Jorubów z Nigerii, kenijskich Masajów, Japończyków, Chińczyków, Peruwianczyków czy Włochów zamieszkujących Toskanię. Głównym celem projektu jest stworzenie mapy zmienności genetycznej człowieka, którą będzie można wykorzystywać w badaniach nad genetycznymi uwarunkowaniami wielu chorób i częstością ich występowania w populacjach. Jeszcze większym wyzwaniem wydaje się The 100.000 Genomes Project (<https://www.genomicsengland.co.uk/>), prowadzony z inicjatywy rządu Wielkiej Brytanii od 2012 roku projekt badawczy, którego celem jest zsekwencjonowanie ponad 100 tys. genomów osób cierpiących na różnego rodzaju choroby, przede wszystkim rzadkie, nowotworowe i zakaźne. W grudniu 2018 roku ogłoszono ukończenie sekwencjonowania ostatniego ze 100 tys. genomów. Z kolei The Human Proteome Project (w skrócie: HPP) to zbiorcza nazwa wielu przedsięwzięć badawczych koordynowanych przez Human Proteome Organization.

Mają one na celu lepsze poznanie ludzkiego proteomu, szczególnie pod kątem funkcji białek, miejsca ich syntezy i działania w organizmie (np. w wątrobie, mózgu, krwi) oraz związku z patologiami. Obecnie wielkość ludzkiego proteomu referencyjnego szacuje się na 73 931 białek, z których ponad 2/3 to produkty alternatywnego splicingu (co tłumaczy, dlaczego liczba białek jest kilkakrotnie większa od liczby genów).

Piśmiennictwo

1. Alberts B et al. *Molecular Biology of the Cell*. 6th edition. New York: Garland Science, 2014: 1–10.
2. Poste G. Molecular diagnostics: a powerful new component of the healthcare value chain. *Expert Rev Mol Diagn* 2001; 1(1): 1–5.
3. Heller M, Pabjan T. *Elementy filozofii przyrody*. Copernicus Center Press; 2014: 240.
4. Chorąży M. Gen strukturalny – ewolucja pojęcia i dylematy. *Nauka* 2009; 3: 57–108.
5. Rearick D et al. Critical association of ncRNA with introns. *Nucleic Acids Res* 2011; 39(6): 2357–2366.
6. Bicknell AA et al. Introns in UTRs: why we should stop ignoring them. *BioEssays* 2012; 34(12): 1025–1034.
7. Cenik C et al. Genome analysis reveals interplay between 5'UTR introns and nuclear mRNA export for secretory and mitochondrial genes. *PLoS Genet* 2011; 7(4): e1001366.
8. Brosius J. The fragmented gene. *Ann N Y Acad Sci* 2009; 1178: 186–193.
9. Szpechciński A. Wykorzystanie sekwencjonowania nowej generacji (NGS) w diagnostyce molekularnej niedrobnokomórkowego raka płuca: aktualne możliwości i perspektywy na przyszłość. *Wiadomości Onkologiczne w Pigulce. Rak Płuca* 2017; 2: 13–24.
10. Wang Z et al. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev Genet* 2009; 10(1): 57–63.
11. Ng SB et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009; 461(7261): 272–276.
12. Wilkins M. Proteomics data mining. *Expert Rev Proteomics* 2009; 6(6): 599–603.
13. Shim HS et al. Korean Cardiopulmonary Pathology Study Group; Korean Molecular Pathology Study Group. Molecular testing of lung cancers. *J Pathol Transl Med* 2017; 51(3): 242–254.
14. International Human Genome Sequencing Consortium (IHGSC). Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431(7011): 931–945.
15. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; 467(7319): 1061–1073.

16. UK Government, Department of Health and Social Care. DNA mapping to better understand cancer, rare diseases and infectious diseases. July 2013.
17. The Uni Prot Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2018; 46(5): 2699.



Adres do korespondencji

dr n. med. Adam Szpechciński

Zakład Genetyki i Immunologii Klinicznej

Instytut Gruźlicy i Chorób Płuc

ul. Płocka 26

01-138 Warszawa

tel.: 22 431 21 58

e-mail: a.szpehcinski@igichp.edu.pl