



LEOWOLFERT/SHUTTERSTOCK.COM

**dr hab.****January Weiner**

Jest z wykształcenia biologiem, a z zawodu bioinformatykiem.

Od kilkunastu lat zajmuje się funkcjonalną analizą odpowiedzi transkrypcyjnych u ludzi i organizmów modelowych, zwłaszcza w kontekście chorób zakaźnych i szczepień.
january.weiner@bih-charite.de

REWOLUCJA

Mało która technologia tak odmieniła oblicze biologii jak sekwencjonowanie – czyli odczytywanie z fragmentu DNA kolejnych par zasad, które tworzą nasz genom.

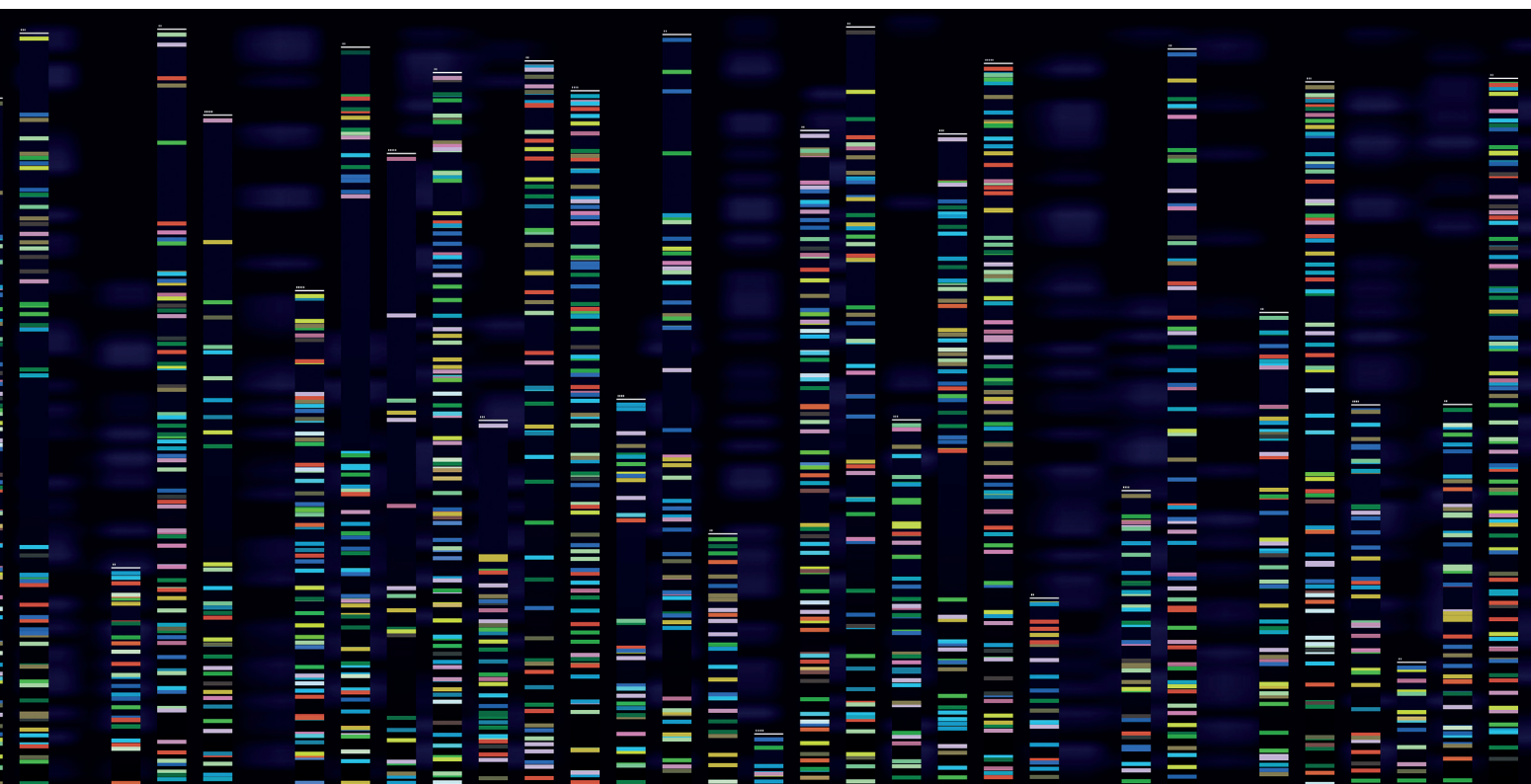
January Weiner

Berlin Institute of Health at Charité
 – Universitätsmedizin Berlin

Pierwsze dwa ludzkie genomy zsekwencjonowano ponad 20 lat temu, do dziś odczytano kolejne, należące do kilkuset tysięcy różnych osób. Rzecz w tym, że nasze geny występują w wielu wariantach (w genetyce nazywanych allelami) i każdy z nas ma ich unikatowy zestaw. W sekwencjonowaniu kolejnych ludzkich genomów chodzi nie tylko o naukowe wyjaśnienie różnic między ludźmi, pochodzenia naszych

cech wspólnych, tego, jakie geny kodują nasz wzrost, a jakie inteligencję – lecz także o zastosowania biomedyczne, czyli poznawanie genetycznego podłoża wielu schorzeń. Ma to zastosowanie zarówno w diagnostyce, jak i poszukiwaniu nowych metod leczenia. Sekwencjonowanie genomu niesie też nadzieję na rozwinięcie medycyny spersonalizowanej, w której leczenie jest dostosowane do organizmu pacjentów, w tym ich genomu.

Sekwencjonowanie ma jednak o wiele więcej zastosowań niż tylko odczytanie DNA. Oprócz alleli, które odziedziczyliśmy po rodzicach, i nowych mutacji, które posiadają wszystkie nasze komórki, w naszych organizmach ciągle powstają nowe warianty DNA. Po pierwsze, występująca u wszystkich kręgowców swoista odpowiedź immunologiczna polega



TARTILA/SHUTTERSTOCK.COM

BIOMEDYCZNA

na tworzeniu nowych przeciwciał w drodze rekombinacji i selekcji. W związku z tym genomy komórek wytwarzających przeciwciała różnią się od genomów innych naszych komórek. Dzięki sekwencjonowaniu możemy lepiej zrozumieć proces wytwarzania odporności na infekcje, a także choroby autoimmunologiczne. Po drugie, nowe, niezamierzone przez ewolucję mutacje zachodzą ciągle przy wszystkich podziałach naszych komórek. To właśnie one są przyczyną chorób nowotworowych. Sekwencjonowanie genomu komórek rakowych umożliwia zidentyfikowanie kluczowych mutacji i dobranie skutecznej terapii.

Ekspresja genów

Kolejnym zastosowaniem sekwencjonowania jest transkryptomika. Każdy nasz gen ulega ekspresji, to znaczy – jest przepisywany z DNA na RNA, a później, jeśli jest to gen kodujący sekwencję białkową – na białko. Nasze komórki reagują na otoczenie i komunikują się między sobą w dużej mierze przez regulację ekspresji genów: „włączanie” i „wyłączanie” przepisywania genów na RNA w zależności od kon-

tekstu. Na przykład w wyniku infekcji wirusowej w organizmie komórka może otrzymać sygnał stymulujący ją do produkcji białek hamujących infekcję wirusową. Sekwencjonowanie transkryptomu umożliwia wykrycie i scharakteryzowanie odpowiedzi immunologicznej. W ostatnich latach rozpowszechniły się metody pozwalające na badanie transkryptomów nawet pojedynczych komórek (*single-cell RNA-seq* – *sc-RNA-seq*).

Sekwencjonowanie DNA jest jedną z podstawowych technologii do badania organizmów i komórek na poziomie molekularnym. W przeciwieństwie do innych metod, np. opartych na przeciwciałach albo spektrometrii masowej, umożliwia przebadanie olbrzymiej liczby próbek w niezwykle krótkim czasie. Z początkiem XXI wieku pojawiło się bowiem wiele nowych metod sekwencjonowania DNA – czasem mówi się o sekwencjonowaniu drugiej generacji, a najczęściej o sekwencjonowaniu wysokoprzepustowym. Szczególną popularność zyskały metody sekwencjonowania krótkich odczytów – to oznacza, że pojedyncza odczytana sekwencja ma od kilkudziesięciu do 150 par zasad. W przypadku wielu wymienionych

wyżej zastosowań sekwencjonowanie krótkich odczytów ma olbrzymie zalety: z jednej strony odczyty są wystarczająco długie, by jednoznacznie zidentyfikować np. większość matrycowego RNA (które koduje nasze białka), z drugiej – pozwala na jednoczesne sekwencjonowanie setek milionów molekuł w czasie zaledwie kilku godzin.

Dlatego warto zwrócić uwagę na dwie nowe rodziny technologii związanych z sekwencjonowaniem: sekwencjonowanie długich fragmentów DNA oraz transkryptomikę przestrzenną.

Sekwencjonowanie długich fragmentów

U schyłku XX wieku zautomatyzowane sekwenatory potrafiły odczytać kilkaset par zasad w kilkunastu bądź kilkudziesięciu fragmentach DNA jednocześnie. Za pomocą takiej technologii z początkiem XXI wieku opublikowano pierwsze niemal kompletne sekwencje ludzkiego genomu. Chociaż zsekwencjonowano wów-

turalnych, czyli wielkoskalowych mutacji w naszym genomie, co oczywiście utrudnia ich wykrywanie.

W ciągu ostatnich kilku lat dokonał się olbrzymi postęp, jeśli chodzi o sekwencjonowanie długich odczytów. Na uwagę zasługują zwłaszcza dwie technologie: sekwencjonowania pojedynczych molekuł w czasie rzeczywistym (*single molecule real time* – SMRT) i sekwencjonowania za pomocą nanoporów (*nanopore sequencing*). Obie umożliwiają sekwencjonowanie bardzo długich fragmentów DNA – o rzędy wielkości dłuższych niż wszystkie inne metody.

SMRT bardzo przypomina klasyczne sekwencjonowanie metodą Sanger'a o tyle, że podstawowym mechanizmem sekwencjonowania jest również synteza komplementarnej nici DNA. W syntezie biorą też udział nukleotydy znakowane fluorescencyjnie, co umożliwia detekcję nukleotydów wbudowywanych w syntetyzowaną nić DNA. Zasadnicza różnica polega jednak na tym, że w czasie rzeczywistym obserwuje się pojedyncze molekuły. Dzięki temu polimeraza DNA pracuje nieustannie i może zsintetyzować nawet cząstki długości kilkudziesięciu tysięcy zasad.

Na zupełnie innym mechanizmie opiera się sekwencjonowanie za pomocą nanoporów. Inaczej niż w przypadku większości sprawdzonych praktycznie technologii sekwencjonowania detekcja kolejnych nukleotydów nie odbywa się przez proces syntezy nowej nici DNA. W tej technologii pod wpływem pola elektrycznego przesuwają się one przez mikroskopijny otwór (nanopor) w membranie. Towarzyszące zmiany napięcia zależą od tego, jaki nukleotyd w danym momencie przez nią przechodzi. Monitorowanie zmian napięcia umożliwia odczytanie sekwencji, której długość teoretycznie jest ograniczona wyłącznie długością sekwencjonowanej nici DNA. W praktyce, choć najdłuższe uzyskane sekwencje miały ponad milion zasad długości, średnia długość odczytu to od kilkunastu do ponad 20 tys. zasad.

Obie metody zostały wykorzystane niedawno do zsekwencjonowania – po raz pierwszy w historii – kompletne genomu człowieka, wliczając w to telomery, centromery i wszystkie inne sekwencje, których do tej pory nie udało się odczytać. Nie ulega wątpliwości, że jest to początek nowej ery zarówno w diagnostyce medycznej, jak i nauce podstawowej. Chwilowo obie metody są dość drogie, jednak koszty sekwencjonowania spadają bardzo szybko. Już w tej chwili koszt sekwencjonowania pojedynczego ludzkiego genomu spadł do kilkuset dolarów – dla porównania, na zsekwencjonowanie pojedynczego ludzkiego genomu jeszcze 15 lat temu trzeba było wydać miliony dolarów.

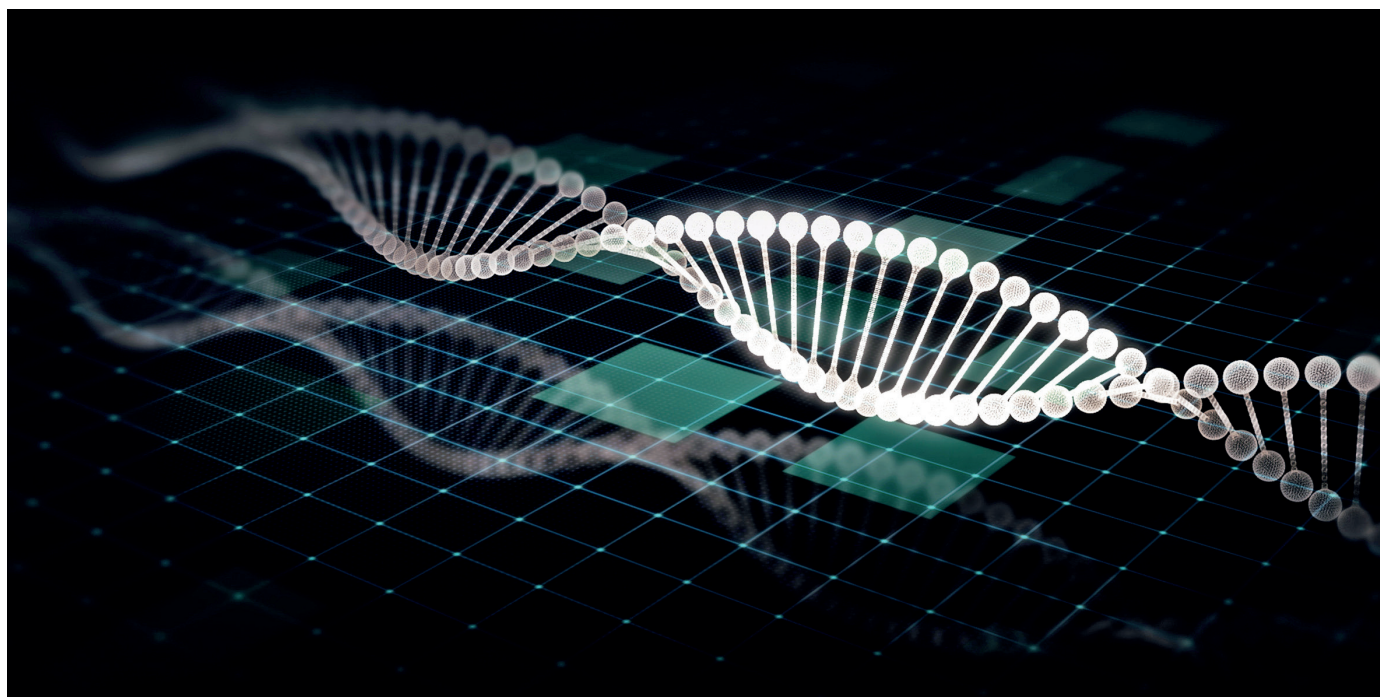
Monitorowanie zmian

Wśród zastosowań związanych z sekwencjonowaniem transkryptomika zajmuje poczesne miejsce. Z wszystkich metod służących do monitorowania zmian w pro-

W sekwencjonowaniu kolejnych ludzkich genomów chodzi nie tylko o naukowe wyjaśnienie różnic między ludźmi, pochodzenia naszych cech wspólnych – lecz także o zastosowania biomedyczne, czyli poznawanie genetycznego podłoża wielu schorzeń.

czas niemal wszystkie geny i ponad 90 proc. całego ludzkiego DNA, wiele obszarów naszego genomu nie poddawało się dobrze ówczesnej technologii sekwencjonowania ani nawet nowszym technologiom. Problem stanowiły obszary powtórzeń sekwencji DNA, w których ten sam ciąg par zasad jest powielony wiele razy. Niektóre rejony takich sekwencji mogą obejmować nawet miliony par zasad. Na przykład fragmenty znajdujące się na każdym końcu wszystkich naszych chromosomów – telomery – składają się z setek lub tysięcy powtórzeń krótkich sekwencji.

I tu pojawia się fundamentalny problem: jeśli odczytana sekwencja jest krótsza niż rejon powtórzenia, to często nie da się jednoznacznie powiedzieć, skąd dokładnie ta sekwencja pochodzi. Za pomocą sekwencjonowania krótkiego odczytu, a nawet technologii pozwalających na odczytanie kilkuset par zasad nie można zsekwencjonować telomerów. To samo dotyczy wspomnianych już wariantów struk-



cesach zachodzących w żywych komórkach jest niewątpliwie najprostsza i najtańsza, jest więc rutynowo stosowana już od dwóch dekad. Znalazła olbrzymią liczbę zastosowań we wszystkich chyba dziedzinach nauk biologicznych i medycznych. Pozwala na precyzyjne śledzenie zmian w ekspresji genów w całych organizmach, tkankach, wyselekcjonowanych typach komórek, a nawet w pojedynczych komórkach. Teraz możemy zaobserwować kolejny przełom, którym jest transkryptomika przestrzenna.

Metody mikroskopowe od dawna umożliwiały badanie procesów zachodzących w naszych tkankach wraz z wizualizacją pojedynczych genów i białek. W ten sposób można było sprawdzić, gdzie w tkance czy komórce następuje ekspresja konkretnego genu. Obecny rozwój technologiczny umożliwił połączenie transkryptomiki (wysokoprzepustowego badania ekspresji wielu genów jednocześnie) i mikroskopowej wizualizacji organizacji tkanek. Dzięki temu można ustalić nie tylko, jakie komórki w jaki sposób regulują ekspresję genów, lecz także w którym miejscu badanej tkanki się znajdują i jak zmienia się ekspresja genów na granicy między zdrową a chorą tkanką.

Istnieją z grubsza dwa podejścia do transkryptomiki przestrzennej. Pierwszy rodzaj metod opiera się na przygotowaniu specjalnych szkiełek podstawowych, w których każdy fragment powierzchni jest specjalnie oznakowany za pomocą krótkich sekwencji DNA (oligonukleotydów). Na szkiełko jest nakładany preparat tkankowy. Kiedy na podstawie RNA z komórek preparatu jest syntetyzowane cDNA, dołączają do niego oligonukleotydy. Następnie cDNA z całego preparatu jest zbierane i zupełnie zwyczajnie sekwen-

cjonowane. Dzięki znakowaniu można jednak później rozpoznać, z którego miejsca na szkiełku pochodziła dana sekwencja, i odtworzyć przestrzenną strukturę preparatu, a potem nałożyć ją na zdjęcia mikroskopowe. Zaletą tych metod jest stosunkowo duża „głębokość”, czyli liczba różnych molekuł RNA, które za ich pomocą mogą być zidentyfikowane. Wadą jest stosunkowo niska rozdzielczość.

Drugi rodzaj metod opiera się na detekcji *in situ*, w podobny sposób, jak robiono to już od wielu dekad, jednak na większą skalę. Preparat jest znakowany za pomocą fluorescencyjnych sond odpowiadających różnym sekwencjom mRNA. Często polega to na znakowaniu krótkich, charakterystycznych sekwencji DNA markerami fluorescencyjnymi. Takie sondy w specyficzny sposób wiążą się z mRNA, które dzięki nim można następnie bezpośrednio zwizualizować na preparacie. Istnieje bardzo dużo wariantów w tej rodzinie metod transkryptomiki przestrzennej, jednak większość z nich umożliwia tylko wykrywanie zdefiniowanego *a priori* repertuaru genów.

Rzeczywiste nowości w biologii i medycynie trwają lata, a nawet dziesięciolecia. Metody, które 10 lat temu były przede wszystkim eksperymentalne i z rzadka pojawiały się na łamach czasopism z najwyższej półki, dziś są udoskonalane i rutynowo stosowane w laboratoriach na całym świecie, a ich koszty, choć niebanalne, nie są już zaporowe. Sekwencjonowanie długich fragmentów i transkryptomika przestrzenna właśnie wkraczają w ten drugi etap. Prawdopodobnie ich praktyczne znaczenie zarówno dla nauk podstawowych, jak i biomedycyny znacząco wzrośnie w ciągu najbliższych lat. ■

Chcesz wiedzieć więcej?

Deamer D., Akeson M., Branton D., *Three Decades of Nanopore Sequencing*, „Nature Biotechnology” 2016.

Lightbody G., Haberland V., Browne F., Taggart L., Zheng H., Parkes E., Blayney J.K., *Review of Applications of High-Throughput Sequencing in Personalized Medicine: Barriers and Facilitators of Future Progress in Research and Clinical Application*, „Briefings in Bioinformatics” 2019.

Logsdon G.A., Vollger M.R., Eichler E.E., *Long-Read Human Genome Sequencing and Its Applications*, „Nature Reviews Genetics” 2020.