

**dr n. med.****Beata Małachowska**

Badaczka zajmująca się medycyną, biostatystyką, bioinformatyką i biologią obliczeniową. Pracuje w Albert Einstein College of Medicine w Nowym Jorku jako postdoc i analityk danych w Department of Radiation Oncology. W swojej pracy naukowej skupia się na analizie danych sekwencjonowania pojedynczych komórek ze szczególnym naciskiem na takie obszary jak popromienne zapalenie jelit oraz rola mimetyków trombopoetyny w leczeniu zaburzeń popromiennych. Laureatka wielu wyróżnień naukowych, w tym ISPAD-JDRF Fellowship Award, stypendium START Fundacji na rzecz Nauki Polskiej, Złotego Otisa oraz nagrody L'Oreal-UNESCO dla Kobiet. b.e.malachowska@gmail.com

POJEDYNCZA KOMÓRKA PEŁNA INFORMACJI

Rozwój nowej dziedziny nauki – bioinformatyki
– przyczynił się do dokładnego poznania ludzkiego
DNA i RNA.

Beata Małachowska

Department of Radiation Oncology
Albert Einstein College of Medicine w Nowym Jorku

Gdy w 1990 roku rozpoczął się projekt poznania ludzkiego genomu (ang. *Human Genome Project* – HGP), zakładano, że identyfikacja sekwencji całego genomu (DNA) zajmie naukowcom 15 lat. Genom człowieka zawiera około 3 mld liter, przy czym zakłada się, że informacja ta jest stała i jednako- wa w każdej komórce danego organizmu. W HGP było zaangażowanych 20 laboratoriów z sześciu krajów. Obecnie, czyli ponad 30 lat później, możemy już poznawać nie tylko sekwencję całego DNA danego człowieka, lecz także ilość i jakość dynamicznie zmieniającego się RNA produkowanego na podstawie statycznego DNA. Możemy to osiągnąć w ciągu kilkunastu godzin, w jednym laboratorium, dla tysięcy indywidualnych komórek jednocześnie. Problemem nie jest już otrzymanie tej informacji czy nawet finansowanie takich projektów – sekwencję pojedynczego genomu możemy poznać za mniej więcej 1 tys. dolarów. Problemem jest opracowanie setek gigabajtów danych pochodzących z pojedynczych eksperymentów biologicznych w sposób zrozumiały i użyteczny dla człowieka.

Każda żywa komórka człowieka ma w swoim jądrze 46 cząsteczek DNA. Gdyby ułożyć je jedna za drugą, długość takiego łańcucha wynosiłaby około dwóch

metrów. Żeby mogły się one zmieścić w małej komórce, muszą być bardzo silnie poskręcane i poukładane tak, by w razie potrzeby komórka miała dostęp do potrzebnych jej fragmentów DNA. Taka silnie poskręcana, niedostępna część DNA jest nazywana heterochromatyną i uznaje się, że w danej komórce nie powstaje z niej RNA. Pozostała część nazywa się euchromatyną i jest dostępna do produkcji RNA. Część z naszego DNA (tzw. geny) jest przepisywana na cząsteczki RNA, które mogą następnie służyć jako matryce do tworzenia białek. Jak dużo białka powstanie, zależy m.in. od tego, jak wiele matryc RNA zostanie wyprodukowanych na podstawie DNA.

Różne przeznaczenie

Mimo że każda żywa komórka ma takie samo DNA, ich kształt i funkcje różnią się od siebie znacząco. Komórka układu immunologicznego ma umiejętność rozpoznawania bakterii w naszym organizmie, wykorzystuje do tego specjalne receptory (białka) na swojej powierzchni, które identyfikują fragmenty bakterii jako obce. Z kolei neurony, komórki układu nerwowego, błyskawicznie przesyłają sygnały z jednego końca naszego ciała na drugi, wykorzystując sygnał elektryczny. Robią to za pomocą specjalnych kanałów (białek) na swojej powierzchni. Jak wynika z przytoczonych przykładów, o tożsamości komórki i o jej funkcji decydują białka, które może ona wyprodukować, a to dzieje się na podstawie dostępnego RNA (a dokładniej mRNA, rodzaju RNA wykorzystywanego jako matryca informacyjna).

Dzięki technice sekwencjonowania nowej generacji (dostępnej komercyjnie od 2005 roku) odzyskiwanie informacji z DNA/RNA na masową skalę stało się dużo prostsze i z czasem coraz tańsze. Prace w HGP trwały ostatecznie 13 lat dzięki wprowadzeniu techniki *shotgun sequencing*. Technika ta polega na rozbijaniu długich cząsteczek DNA na mniejsze fragmenty i poznawaniu ich sekwencji (stąd nazwa „sekwencjonowanie”), a następnie na komputerowym składaniu zawartych w nich informacji. Biorąc pod uwagę ilość informacji, którą niesie DNA pojedynczego człowieka (3,4 GB), ręczne porównywanie sekwencji i nakładanie ich na siebie tak, by tworzyły spójną liniową całość, byłoby zajęciem niezwykle czasochłonnym. Komputerowe przetwarzanie informacji stało się niezbędne w pracy biologów, a dziedzina zwana bioinformatyką zaczęła rozkwitać. Powstały tzw. laboratoria suche (ang. *dry lab*), które w odróżnieniu od klasycznych „mokrych” laboratoriów biologicznych (ang. *wet lab*) zajmowały się wyłącznie analizą danych z eksperymentów biologicznych. W laboratoriach tych pracują ludzie o bardzo różnym wykształceniu, m.in. bioinformatycy, biolodzy, biotechnolodzy, medycy, którzy zaczęli interesować się bioinformatyką na potrzeby swoich badań, a także informatycy, inżynierowie oprogramowania, którzy wyspecjalizowali się

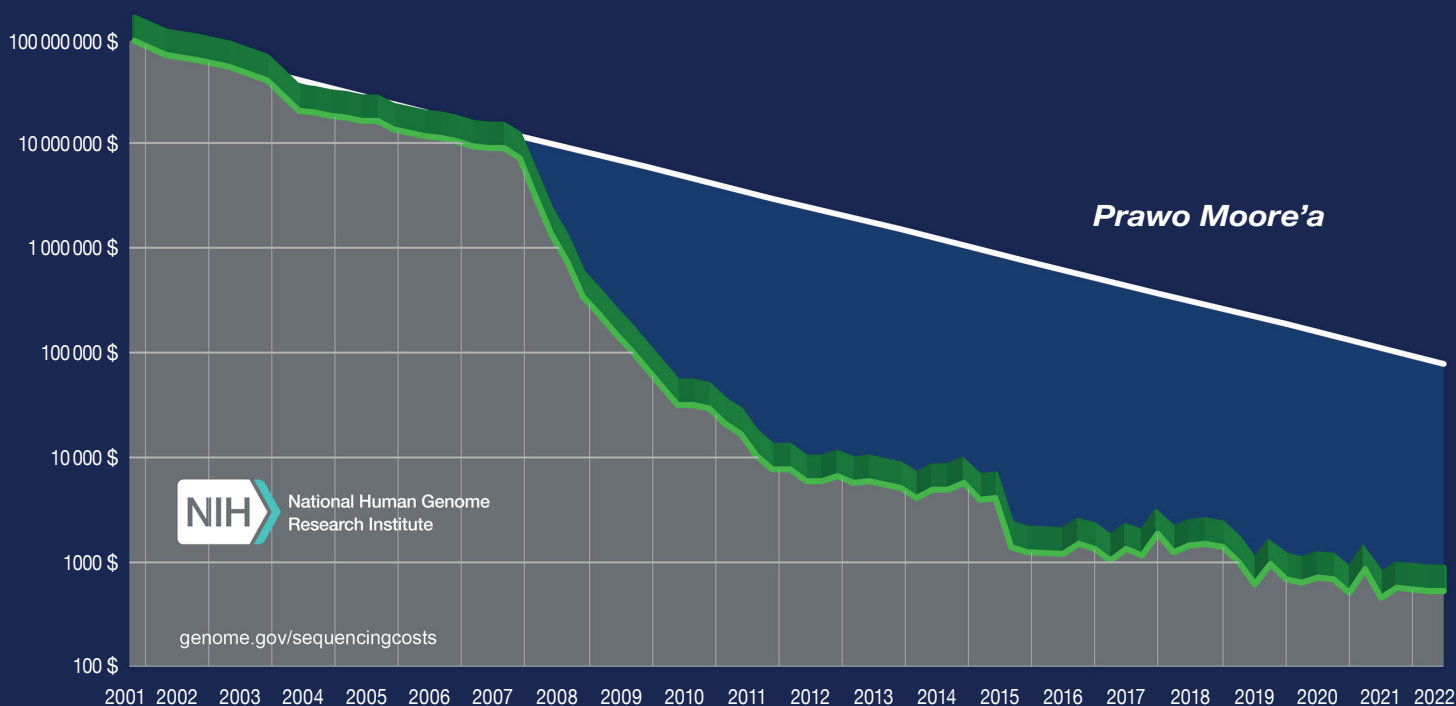
Zapis DNA

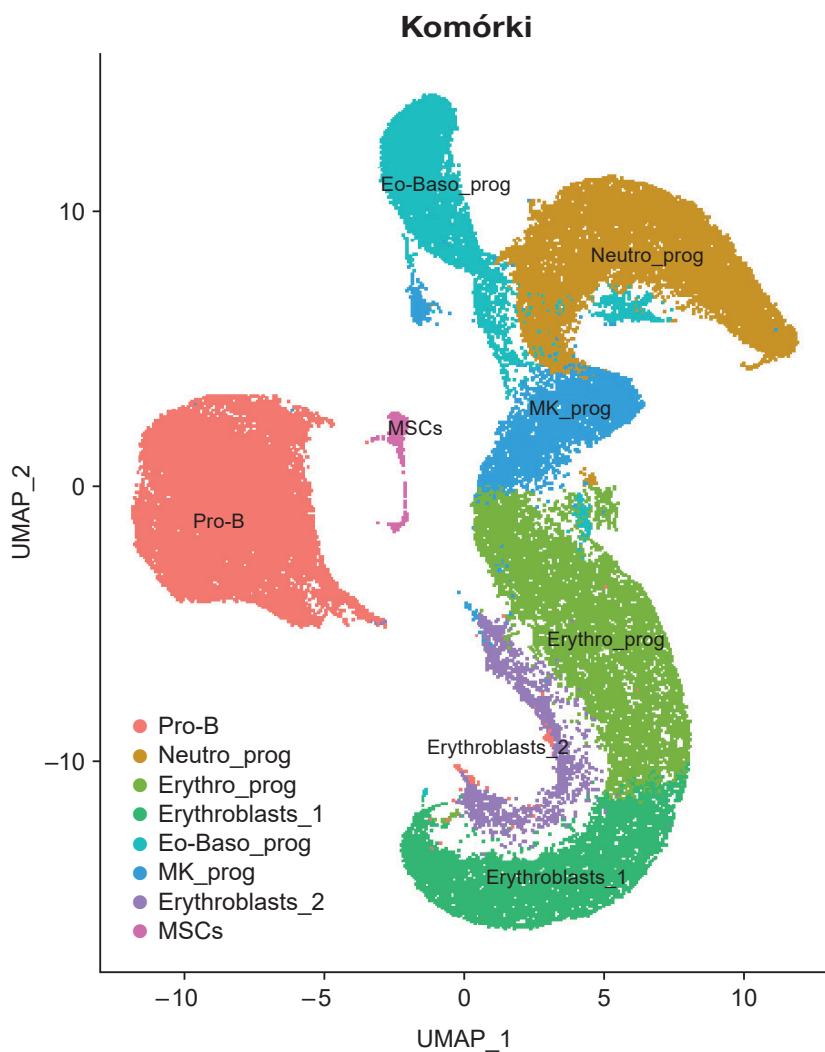
Mutacje w organizmie człowieka powodują, że nie wszystkie komórki mają identyczny zapis DNA. Szczególnie powszechnie zjawisko to ujawnia się w komórkach nowotworowych, gdzie modyfikacja zapisu DNA to nie tylko zmiana pojedynczych liter, lecz także zwiększenie liczby kopii poszczególnych zapisów, utrata innych, a także przekształcenia strukturalne, jak odwrócenie kierunku zapisu danego fragmentu.

w kierunku analiz bioinformatycznych. Ten zlepek ludzi i doświadczeń tworzy kuźnię do nieustannego rozwijania nowych algorytmów analiz danych. Przyciąga on szczególnie młodych ludzi, którzy wywodzą się z pokolenia wychowanego z komputerem.

Wróćmy jednak do DNA. Już wiemy, jak poznać jego sekwencję dla danego organizmu. Dzięki ukończeniu HGP znamy już „podstawowy” zapis DNA człowieka. Co ciekawe, technika sekwencjonowania DNA może posłużyć również do poznania RNA, gdyż może zostać ono przepisane w prosty sposób na komplementarne (jak lustrzane odbicie) do niego DNA. Jednakże RNA, jak wcześniej wspomnieliśmy, nie jest tożsame dla każdej komórki. W zależności od jej funkcji,

Spadek kosztów sekwencjonowania genomu pojedynczego człowieka na przestrzeni lat





Graficzna prezentacja wyników z analizy sekwencjonowania pojedynczej komórki. Każdy punkt stanowi indywidualną komórkę, a odległość między punktami stanowi miarę podobieństwa profilu RNA między komórkami

kształtu, stanu chorobowego czy działania leków – rodzaj i ilość RNA może ulegać zmianie. Jedne geny mogą zmniejszać swoją ekspresję (ilość wyprodukowanego RNA) lub całkowicie ją wygaszać, inne mogą aktywować swoją ekspresję lub zwiększać ją do wyższego poziomu. Obecnie szacuje się, że w ludzkim genomie jest od 20 do 25 tys. genów, a kombinacja rodzaju i poziomu ich ekspresji jest źródłem zmienności biologicznej między komórkami, tkankami, a nawet całym organizmem. Ludzki organizm posiada około 200 rodzajów komórek, których profil ekspresji może zmieniać się w zależności od wielu czynników, m.in.

Niekodujące DNA

Nie wszystkie geny – fragmenty ulegające przepisaniu na RNA – prowadzą do powstania białek. Część tak powstałego RNA pełni funkcję regulacji ekspresji innych genów, stabilizuje genom czy też asystuje przy procesie powstawania białka, lecz nie jest używana jako matryca do jego zapisu.

płci, stanu chorobowego, a także pory dnia, diety czy temperatury otoczenia, co powoduje, że każda komórka w naszym organizmie może mieć swój indywidualny profil ekspresji.

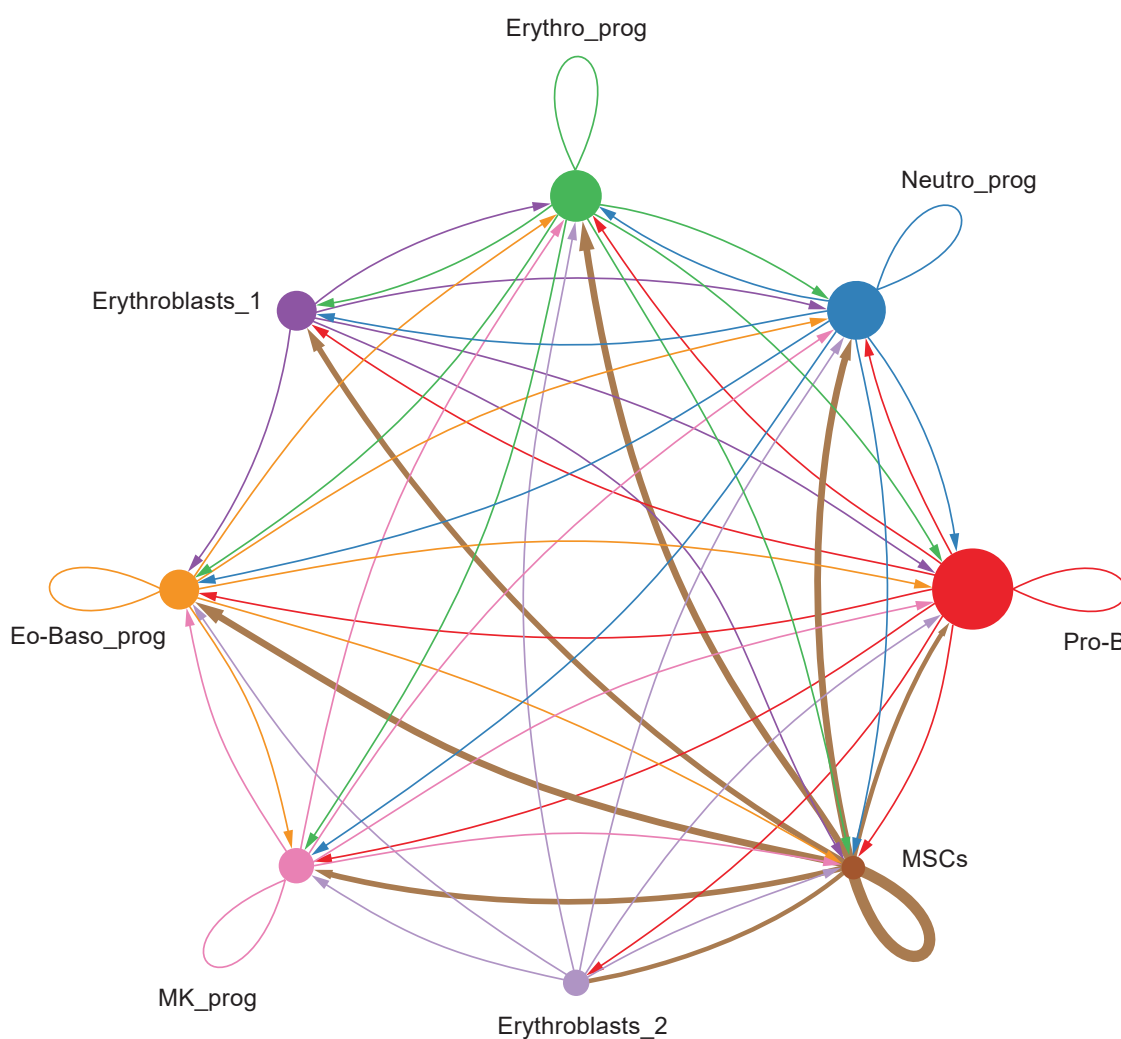
Dzięki dalszemu rozwojowi technik sekwencjonowania nowej generacji mamy obecnie możliwość uzyskania informacji o rodzaju i ilości RNA produkowanego przez pojedynczą komórkę (scRNAseq – ang. *single cell RNA sequencing*). Szacuje się, że pojedyncza komórka zawiera około 360 tys. cząsteczek mRNA, z czego 12 tys. ma unikatową sekwencję (pojedyncze transkrypty, czyli rodzaj mRNA), a pozostałe stanowią ich kopię (ilość mRNA). W typowym eksperymencie tego typu poznaje się sekwencję około 10–20 tys. komórek pochodzących z jednej próbki, a próbek może być wiele w jednym eksperymencie. Wąskim gardłem jest analiza tak dużej ilości danych.

Skąd się biorą trudności

Musimy najpierw zrozumieć, jak powstają dane z sekwencjonowania pojedynczej komórki. Najpierw pojedyncze komórki z danej próbki muszą być oddzielone od siebie i zanurzone w kropelkach zawierających znaczniki dla RNA oraz odpowiednie odczynniki. Te kropelki to nijako miniaturowe próbki, w których zachodzą reakcje właściwe dla danej komórki. RNA z danej komórki jest dzielone na krótkie fragmenty, do których są przyłączane znaczniki unikatowe dla danej komórki. Żeby zapewnić, że każda komórka znajduje się w osobnej kropelce z indywidualnym znacznikiem, większość kropelek nie zawiera komórek. Następnie sekwencjonujemy około miliona kropelek, by uzyskać informację z 10–20 tys. komórek.

Pierwszym krokiem analizy jest ustalenie, które kropelki zawierają prawdziwe komórki, a które jedynie zanieczyszczenia. Następnie fragmenty RNA są porównywane z genomem, by poskładać fragmenty pochodzące z pojedynczego genu i określić, które geny są aktywne w danej komórce i na jakim poziomie. Ze względu na ogromną liczbę kropelek i fragmentów RNA do porównania (np. 450 mln dla 20 tys. komórek) proces ten wymaga ogromnej mocy obliczeniowej i trwa kilka dni. Po ustaleniu, które dane pochodzą z prawdziwych komórek, przechodzimy do właściwej analizy danych. Najpierw musimy przefiltrować i przekształcić dane, by pozbyć się wszelkich artefaktów technicznych. W kolejnym etapie na podstawie naszej wiedzy biologicznej identyfikujemy typy komórek, grupując je i znajdując charakterystyczne geny dla każdej grupy.

Następnie jest przeprowadzana analiza właściwa, specyficzna dla danego projektu naukowego. Obecnie istnieją tysiące algorytmów do analizy tego typu danych, dobranie odpowiedniego do danego projektu naukowego jest trudne, a poprawne jego użycie jest jeszcze trudniejsze. Na przykład dane o profilu RNA



pojedynczej komórki umożliwiają określenie, w jaki sposób poszczególne typy komórek reagują na leczenie, czy wyjaśnienie mechanizmu działania nowych leków na poziomie molekularnym, co pozwala przewidzieć optymalne warunki ich zastosowania i potencjalne działania niepożądane. Możemy również opisać, jak komórki informują się wzajemnie o niebezpieczeństwie (infekcji, uszkodzeniu, potrzebie naprawy). Dzięki zdobytej wiedzy możemy symulować sygnały, które wydają pojedyncze komórki, używając odpowiednich substancji chemicznych po to, żeby nakłonić komórki do wykonywania pracy na rzecz organizmu (np. naprawy uszkodzonych tkanek). W badaniach, które prowadzimy w Department of Radiation Oncology w Albert Einstein College of Medicine, m.in. poszukuje się nowych leków stymulujących regenerację poszczególnych tkanek po narażeniu na promieniowanie jonizujące (np. w przypadku incydentów radiacyjnych).

W dobie nowoczesnej medycyny i biologii analiza bioinformatyczna odgrywa kluczową rolę. Pozwala ona na dokładne zrozumienie aktywności genów

na poziomie pojedynczej komórki, co jest nieocenione w odkrywaniu subtelnych różnic w funkcjonowaniu komórek w różnych stanach zdrowia i choroby. Dzięki scRNAseq możemy identyfikować specyficzne podtypy komórek, zrozumieć ich funkcje oraz interakcje w sposób, który wcześniej był niemożliwy.

Takie analizy dostarczają cennych informacji, które przyczyniają się do rozwoju spersonalizowanej medycyny, umożliwiając precyzyjne diagnozy i dostosowane terapie. W kontekście chorób takich jak nowotwory wiedza o zaburzonych profilach ekspresji genów w pojedynczych komórkach może prowadzić do opracowania skuteczniejszych metod leczenia. Ponadto zrozumienie dynamiki RNA w komórkach jest kluczowe dla badań nad różnymi procesami biologicznymi, od rozwoju embrionalnego po odpowiedzi immunologiczne.

Dlatego też w czasach, gdy dysponujemy ogromnymi ilościami danych biologicznych, bioinformatyczna analiza staje się niezbędnym narzędziem do przekształcania tych danych w użyteczną wiedzę, przyspieszając tym samym postępy w nauce i medycynie. ■

Chcesz wiedzieć więcej?

Kasprzyk M., Mleczko A.M., Celichowski P., Bąkowska-Żywicka K., *Przetwarzanie RNA – niezwykle mechanizm powstawania nowych klas niekodujących RNA z funkcjonalnych RNA*, „Postępy Biochemii” 2014.

Loewe L., Hill W.G. *The population genetics of mutations: good, bad and indifferent*, „Philos Trans R Soc Lond B Biol Sci” 2010, <https://doi.org/10.1098/rstb.2009.0317>

Vercellino J., Małachowska B., Kulkarni S. et al. *Thrombopoietin mimetic stimulates bone marrow vascular and stromal niches to mitigate acute radiation syndrome*, „Stem Cell Res Ther” 2024, <https://doi.org/10.1186/s13287-024-03734-z>