

Uniknąć losu zbójcy Gębona



PAWEŁ GOLIŁ

Instytut Genetyki i Biotechnologii,
Wydział Biologii, Uniwersytet Warszawski
Instytut Biochemii i Biofizyki
Polska Akademia Nauk, Warszawa
pgolik@igib.uw.edu.pl
Prof. Paweł Golik jest dyrektorem Instytutu Genetyki i Biotechnologii, szefem grupy mitochondrialnej obróbki RNA, popularyzuje genetykę.

„... i wrzasnął wielkim głosem, bo miał dość, lecz już go Informacja trzystu tysiącami mil papierowych spowiła i spętała, że nie mógł się ruszyć i musiał czytać dalej...”

Stanisław Lem, „Cyberiada. Wyprawa szósta, czyli jak Trurl i Klapaucjusz demona drugiego rodzaju stworzyli, aby zbójcę Gębona pokonać”

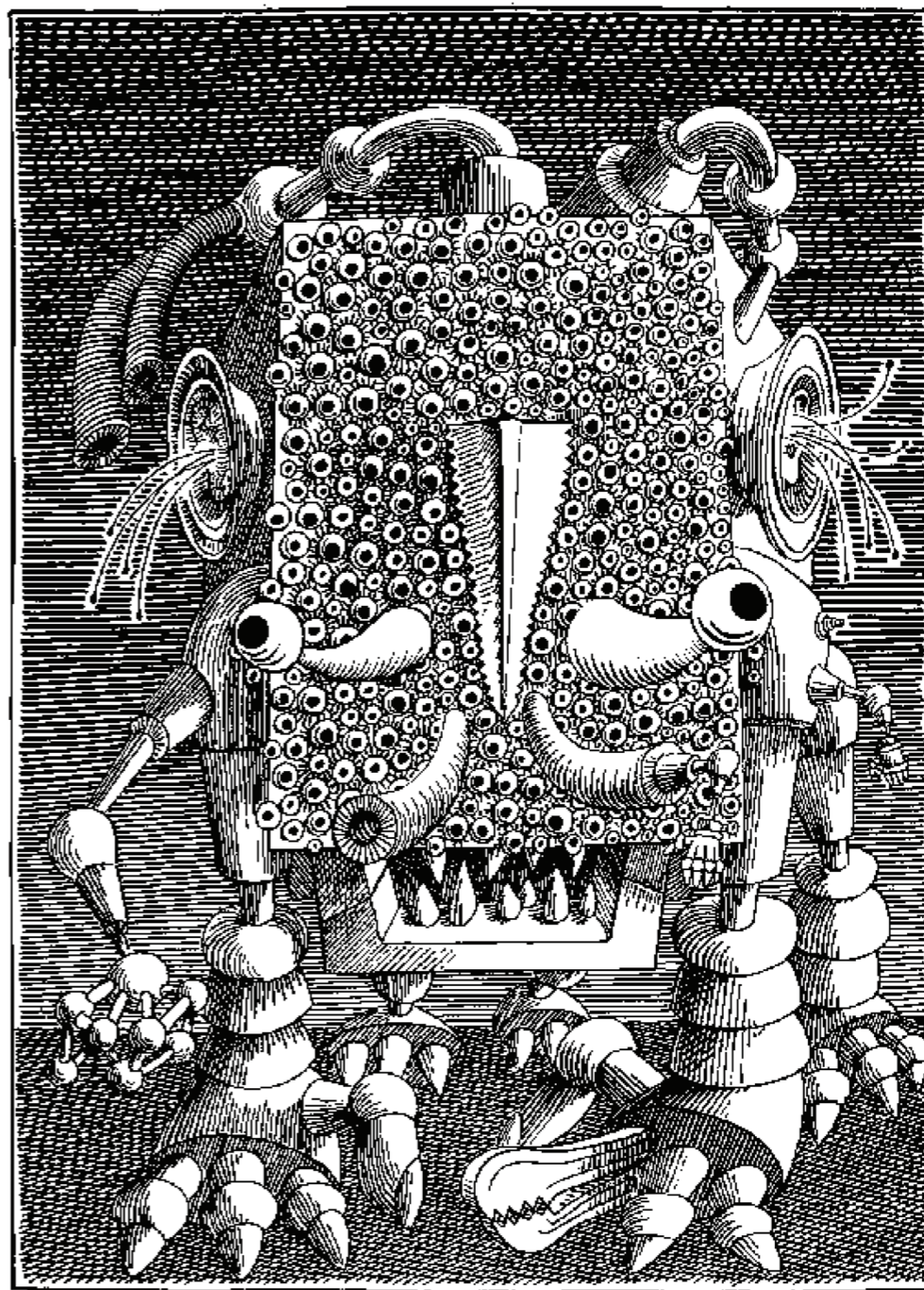
Kto nie czytał zacytowanego opowiadania Lema, ten czym prędzej winien nadrobić zaległości, gdyż żadne streszczenie nie odda uroku tej prozy. Tytułowego zbójcę od zwykłych rabusiów odróżniało to, że rabował nie kosztowności, lecz poszukiwał wiedzy i informacji „wszelkiej, byle prawdziwej”. Bohaterowie skonstruowali mu zatem urządzenie, które dostarczyło mu tego, czego pragnął – potrafiło z przypadkowych ruchów cząstek powietrza wydobyć wszystkie prawdziwe informacje, jakie znał Wszechświat. Szybko jednak się okazało, że nieograniczony dostęp do informacji stał się dla zbójcy bardziej przekleństwem, niż zaspokojeniem pragnień. Utonął „przywalony lawiną informacyjną”, nie mogąc w zalewie błahych i trywialnych informacji odnaleźć wiedzy, która rzeczywiście byłaby przydatna.

Co ma wspólnego napisana przed ponad pół wiekiem filozoficzna przypowieść wielkiego pisarza ze współczesną biologią? Więcej, niż mogłoby się wydawać! Odnalezienie sensu w zalewie danych i informacji jest bowiem jednym z głównych wyzwań stających przed tą dziedziną wiedzy.

Lawinowy wzrost danych

Przełom lat 70. i 80. XX wieku przyniósł opracowanie metod pozwalających na odczytywanie sekwencji (kolejności) nukleotydów tworzących zapis informacji genetycznej w DNA (odkrycie to zostało w 1980 roku uhonorowane Nagrodą Nobla w dziedzinie chemii dla Waltera Gilberta i Fredericka Sanger). Już pod koniec lat 80. pojawiły się pierwsze zautomatyzowane sekwenatory, które ułatwiły tę początkowo żmudną i niełatwą procedurę, a genetycy zaczęli coraz śmielej mówić o poznawaniu sekwencji DNA już nie tylko pojedynczych genów, ale też całych genomów – czyli kompletnego zapisu informacji genetycznej danego organizmu. Narodziła się genomika. Tempo jej rozwoju zaskoczyło nawet najśmielszych optymistów. Połowa lat 90. przyniosła pierwsze kompletne sekwencje genomów prostych organizmów – bakterii i drożdży, a już w 2001 roku opublikowano wstępną wersję liczącego około 3 miliardów nukleotydów genomu człowieka. Początek XXI wieku to kolejny etap rozwoju technik sekwencjonowania DNA – tzw. sekwencjonowanie wysokoprzepustowe, zwane też sekwencjonowaniem nowej generacji (ang. NGS – Next Generation Sequencing). Postępem wyspecjalizowanego sprzętu do badania DNA towarzyszył, znany każdemu z codziennego życia, szybki rozwój mocy obliczeniowej komputerów niezbędnych do tego, by z krótkich fragmentów odczytywanych przez sekwenatory złożyć kompletną sekwencję. W ciągu 10 lat, między 1999 a 2009 rokiem koszt sekwencjonowania DNA obniżył się około 14 tys. razy, a prędkość odczytywania sekwencji wzrosła między 2000 a 2010 r. 50 tys. razy!

Nowoczesne sekwenatory odczytujące miliardy nukleotydów w pojedynczym eksperymencie znajdują się już w wyposażeniu wielu laboratoriów, także w Polsce. Do tego dodać należy bogaty i także szybko rozwijający się repertuar zautomatyzowanych technik genomiki funkcjonalnej, badającej, jak informacja genetyczna jest wyrażana w komórce, jakie powstają w niej białka, jaki jest efekt mutacji w każdym z genów itp. Każdy numer specja-



Czy biologia uniknie losu zbójcy Gębona z „Cyberiady” Stanisława Lema?

listycznych periodyków biologicznych przynosi kolejne doniesienia o zsekwencjonowanych genomach i innych analizach wysokoprzepustowych, a dyski twarde komputerów badaczy zapewniają się terabajtami danych. Zdawać by się mogło, że biologia niebawem odkryje przed nami wszystkie tajemnice. Wypowiedzi w takim tonie można było zresztą usłyszeć i przeczytać po ogłoszeniu sekwencji genomu człowieka – skoro znamy zapis naszego DNA, to pełne zrozumienie naszej biologii jest na wyciągnięcie ręki.

Szybko zdaliśmy sobie sprawę z tego, że odczytanie DNA to zaledwie pierwszy krok na bardzo długiej drodze. Ale czy w lawinowo rosnącej powodzi danych nie czeka nas los zbójcy Gębona z opowiadania Stanisława Lema?

Przepis w języku DNA

By zrozumieć wyzwania stojące przed genetyką i genomiką należy chwilę poświęcić na zastanowienie się nad tym, czym naprawdę jest informacja genetyczna. Niekiedy porównuje się ją do inżynierskiego planu budowy, schematu konstrukcyjnego (ang. *blueprint*), jest to jednak metafora myląca. Nie znajdziemy bowiem w DNA prostego odwzorowania tego, jak będzie działać i wyglądać komórka czy organizm. Informacja genetyczna jest językiem, instrukcją mówiącą o tym, jakie białka i RNA mają być przez komórkę produkowane, a także kiedy produkcja każdego z nich ma być uruchamiana. Białka (a także RNA) zaś prowadzą reakcje chemiczne, budują struktury komórkowe, re-

Biologia w epoce genomiki

gulują działanie genów i uczestniczą w przekazywaniu sygnałów.

Właściwszą metaforą będzie tu nie schemat konstrukcyjny, ale przepis z książki kucharskiej. To bardzo istotna różnica – oglądając schematyczny rysunek, możemy bowiem łatwo wyobrazić sobie, jak wygląda gotowe urządzenie, które on przedstawia. Tymczasem lektura przepisu nie da osobie bez doświadczenia w kuchni wyobrażenia o smaku gotowego dania. Jeśli ktoś nigdy nie piekł ciasta (albo nie kosztował czyjś wypiek), to nie będzie w stanie wyobrazić sobie, jaki efekt da połączenie mąki, wody, cukru i jajek, a następnie podgrzanie do odpowiedniej temperatury, nawet jeżeli wie, jak wyglądają i smakują poszczególne składniki. Dowie się tego dopiero wtedy, gdy skosztuje przygotowanego według przepisu ciasta. Podobnie w biologii, nawet jeżeli odczytamy sekwencję wszystkich genów w genomie komórki, poznamy sekwencje kodowanych przez nie białek i cząsteczek RNA, a nawet dowiemy się, kiedy każde z nich będzie wytwarzane, to nie wystarczy nam to do zrozumienia funkcjonowania komórki. Musimy geny, komórki i organizmy poznać w działaniu, badać nie tylko poszczególne elementy, ale także całe złożone układy. Tu na razie nie wyręczy nas maszyny, a kolejne pokolenia biologów będą miały przed sobą fascynujące wyzwania.

Rzeczywistość jest złożona

60 lat po odkryciu struktury DNA przez Rosalind Franklin, Jamesa Watsona i Francis Cricka dosyć dobrze rozumiemy już, jak funkcjonują geny i potrafimy prześledzić drogę, jaką przechodzi informacja genetyczna od zapisu w DNA do gotowego, funkcjonującego w komórce produktu (choć oczywiście wciąż napotykamy niespodzianki i nowe odkrycia). Skoro wiemy, jak działa gen, a dzięki genomice możemy poznać wszystkie geny danego organizmu, to droga do pełnego zrozumienia biologii wydaje się prosta. Tak jednak nie jest, a zdobywane dzięki nowoczesnym technikom dane dopiero pokazują nam, jak wiele jeszcze nie rozumiemy.

Przez dziesiątki lat klasyczna i molekularna genetyka koncentrowała się na pojedynczych genach. I choć znany ze szkolnych zadań uproszczony paradygmat „jeden gen – jedna cecha”, jak w pionierskich pracach Grzegorza Mendla, gdzie rzeczywiście warianty pojedynczego genu decydowały

o tym, czy kwiaty groszku miały barwę białą, czy czerwoną, szybko okazał się nadmiernym uproszczeniem, to długo dominował w podejściu do badań genetycznych. Gdy automatyzacja technik laboratoryjnych pozwoliła na wyszukiwanie różnic w DNA pomiędzy poszczególnymi ludźmi, zaczęło się poszukiwanie genetycznego podłoża różnych cech, a czasopisma (głównie popularne) zaczęły donosić o „genie inteligencji”, „genie skłonności do alkoholizmu”, „genie schizofrenii” i wielu innych. Rzeczywistość okazała się bardziej złożona.

Większość cech organizmów takich jak człowiek to tzw. cechy wieloczynnikowe, gdzie o fenotypie (czyli tym, co obserwujemy) decydują warianty wielu różnych genów oraz wpływ środowiska. Gdy dane o różnicach w sekwencjach DNA poszczególnych ludzi zaczęły płynąć szerokim strumieniem, badacze zaczęli szukać statystycznych korelacji pomiędzy poszczególnymi wariantami a prawdopodobieństwem wystąpienia danej cechy. Wkrótce jednak okazało się, że genomowe poszukiwania nie przynoszą wyjaśnienia genetyki cech wieloczynnikowych. Stworzono termin „brakująca odziedziczalność” (ang. *missing heritability*) – wiemy (np. z badań nad bliźniętami i bliskimi krewnymi), że dana cecha ma istotne podłoże dziedziczne, ale nie potrafimy odnaleźć go w sekwencji genomu. Spośród różnych możliwych wyjaśnień tego zjawiska szczególną uwagę warto zwrócić na problem oddziaływania genów między sobą.

Sieć oddziaływań

Oddziaływania genetyczne to zagadnienie znane genetykom od dosyć dawna. Z prac nad prostymi organizmami modelowymi wiemy, że czasami mutacje w zupełnie innym genie mogą znieść efekt mutacji w jakimś genie (supresja), albo przeciwnie – wzmocnić go (tzw. oddziaływania syntetyczne). Genetycy pracujący z drożdżami (jak autor tego tekstu) doskonale znają sytuację, gdy zaburzenie działania jakiegoś genu pojedynczo nie daje dających się zaobserwować skutków, a dopiero uszkodzenie dwóch genów naraz powoduje wystąpienie wyraźnego efektu fenotypowego. Automatyzacja pracy laboratoryjnej pozwala nam nawet, przynajmniej u prostych organizmów modelowych, analizować takie oddziaływania genetyczne na skalę całych genomów. Proste układy „jeden gen – jedna cecha”,



Przez dziesiątki lat klasyczna i molekularna genetyka koncentrowała się na pojedynczych genach. Rzeczywistość jest bardziej złożona

jak w doświadczeniach Mendla, są u organizmów takich jak człowiek raczej wyjątkiem niż regułą. Za większość cech złożonego organizmu (wyjątkiem są rzadkie jednogenowe choroby genetyczne) odpowiadają interakcje licznych genów. Dopiero od niedawna zaczynamy zdawać sobie sprawę z tego, jak licznych. Wyniki opublikowanego w 2010 roku przez międzynarodowe konsorcjum badania wykazały, że na wzrost człowieka wpływają allele setek różnych genów (co najmniej 180). Żaden z tych genów pojedynczo nie przybliżył nas do zrozumienia genetycznych podstaw różnic we wzroście człowieka, dopiero poznanie ich oddziaływań da nam odpowiedź. Jest to wielkie wyzwanie – klasyczna genetyka dobrze radzi sobie z oddziaływaniami dwóch czy najwyżej kilku genów ze sobą, wobec systemu złożonego z setek genów jest jednak na razie bezradna. Dopiero dowiadujemy się, jak wielkie wyzwanie przed nami stoi.

Klasyczny obraz pojedynczych genów ustępuje miejsca niezwykle skomplikowanej sieci wzajemnych oddziaływań. Złożoność systemu nie wynika jedynie z liczby elementów składowych, ale przede wszystkim z liczby powi-

zań między nimi. Sekwencjonowanie genomu można porównać do stworzenia katalogu tych elementów, a droga do zrozumienia działania układu wieść musi od tego katalogu do poznania i zrozumienia powiązań między nimi. Postęp technik masowego sekwencjonowania i analizy genów dostarcza nam wielkiej masy danych, jednak bez zrozumienia systemowych właściwości komórek i organizmów czeka nas los Gębona – zginiemy przytłoczeni terabajtami danych, z których niewiele pożytecznego dla nas wyniknie.

W pułapce informacji

Jak biologia może uniknąć pułapki gromadzenia i katalogowania masy informacji, która sama jednak w niewielkim stopniu przybliża ją do zrozumienia funkcjonowania organizmów żywych? Opisowy aparat klasycznej biologii nie da sobie rady z tak złożonym problemem, nie wystarczy też dobrze znana biologom opisowa statystyka. Biologia musi skorzystać z doświadczeń fizyki i matematyki, które z problemami bardzo złożonych systemów zmagają się nieco dłużej. Powstaje nowa, interdyscyplinarna dziedzina, zwana biologią

Biologia w epoce genomiki

Właściwszą metaforą będzie tu nie schemat konstrukcyjny, ale przepis z książki kucharskiej

APS Image Gallery



systemów, która stawia sobie za cel stworzenie całościowego opisu złożonych układów biologicznych. Wymaga to opracowania matematycznych modeli, które pozwolą na opisanie złożonych sieci interakcji genów i ich produktów i skorelowania tych modeli z danymi eksperymentalnymi. To wyzwanie dla teoretyków i eksperymentatorów. Niezbędnym narzędziem będzie tu oczywiście komputer – nie można wyobrazić sobie analizy dużych zbiorów danych biologicznych bez odpowiednich narzędzi bioinformatyki.

Biologia systemów stawia dopiero pierwsze kroki. Tworzone są pierwsze modele matematyczne szlaków metabolicznych, sieci przekazywania sygnału w komórce czy interakcji między genami. W 2010 roku opublikowano mapę oddziaływań genetycznych około 6 tys. genów tworzących genom drożdży, na razie badając tylko interakcje genów parami. Matematyczna analiza takiej sieci dostarcza już wielu ciekawych wniosków, a jej niektóre właściwości przypominają nieco mapy systemów, takich jak połączenia serwerów w Internecie, siatki połączeń lotniczych czy sieci znajomości międzyludzkich. Znajdziemy tam ośrodki koncentrujące dziesiątki połączeń, i peryferie połączone tylko kilkoma więzami z ośrodkami centralnymi. Modelując takie sieci można analizować różne właściwości systemu, np. odporność na zaburzenia, zdolność do zmian i adaptacji. Lokując niezbadany dotąd gen na takiej mapie możemy na podstawie otoczenia, w którym się znalazł, postawić hipotezę odnośnie do jego funkcji. Możemy próbować tworzyć modele tego, jak taki układ odpowiada na bodźce z zewnątrz i jak ewoluuje.

Musimy jednak pamiętać o tym, czego nauczyło nas badanie złożonych systemów przez fizyków. Złożone nieliniowe systemy są tak wrażliwe na niewielkie zaburzenia parametrów, że w dalszej perspektywie zachowują się w sposób praktycznie niemożliwy do przewidzenia. Dobrym przykładem jest pogoda. Problemy związane z jej przewidywaniem są nam doskonale znane. Za wcześnie jeszcze, by stwierdzić, jakie są pod tym względem systemy biologiczne. Z jaką dokładnością sekwencja DNA pozwoli na przewidywanie cech człowieka zależnych od interakcji setek genów – tego wciąż nie wiemy.

Przyszłość biologii systemów

Wiemy jednak, że bez nowych podstaw teoretycznych, bez współpracy z informatyką, matematyką i fizyczną teorią złożoności, grozi nam zgubienie się w nawałnicy danych spływających każdego dnia z wysokoprzepustowej aparatury laboratoryjnej. W przyszłości nauk biologicznych wciąż jest miejsce dla „klasycznego” badania funkcji pojedynczych genów i białek, dla gromadzenia ogromnych zbiorów danych i dla interdyscyplinarnych poszukiwań teoretycznych podstaw działania systemów biologicznych. Wielka przygoda genomiki dopiero przed nami. ■

Chcesz wiedzieć więcej?

The human genome at ten. *Nature*, www.nature.com/humangenome (zbiór artykułów z okazji 10-lecia sekwencji genomu ludzkiego).

Maier B., (2008). Personal genomes: The case of the missing heritability. *Nature*, 456, 18-21.

Lango Allen H., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467, 832-838.

Costanzo M., et al. (2010). The genetic landscape of a cell. *Science*, 327, 425-431.