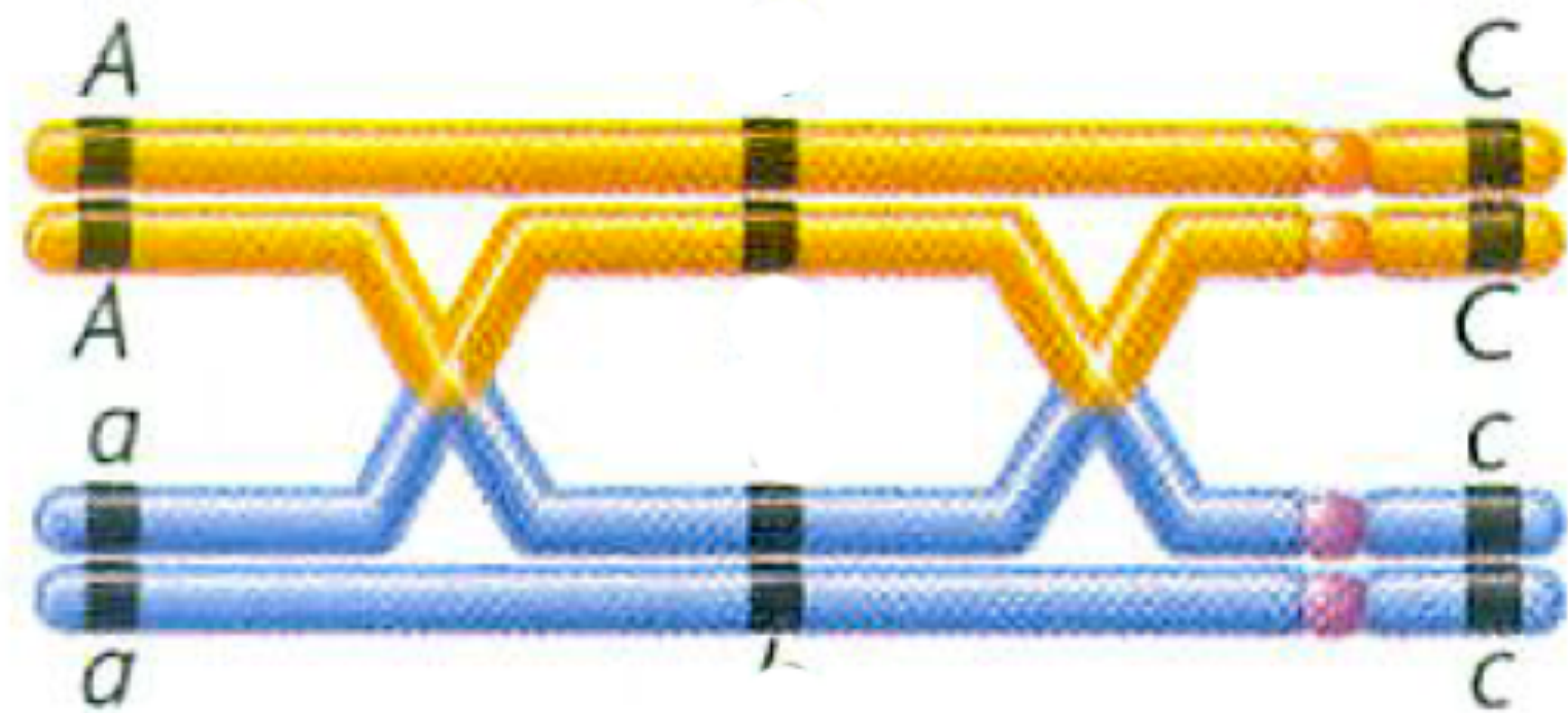


Metody genetyki i genomiki

Badanie i modyfikacje genomów

Mapowanie

- Jednostka cM (centymorgan) = 1% rekombinacji
- W rzeczywistości zależność nie jest liniowa
- Podwójny crossing-over – gamety typu rodzicielskiego
- Interferencja – zajście crossing-over w danym miejscu wpływa na prawdopodobieństwo zajścia kolejnego w pobliżu



Funkcja mapowa

- **Zależność odległości genetycznej (d) od obserwowanej częstości rekombinantów (θ)**
- Funkcja mapowa Haldane'a
 - wielokrotne c-o, bez interferencji
- Funkcja Kosambi'ego
 - uwzględnia też interferencję, szeroko stosowana
- Dla małych θ : $d \approx \theta$

$$d = \frac{\ln(1 - 2\theta)}{2}$$

$$d = \frac{\ln\left(\frac{1 + 2\theta}{1 - 2\theta}\right)}{4}$$

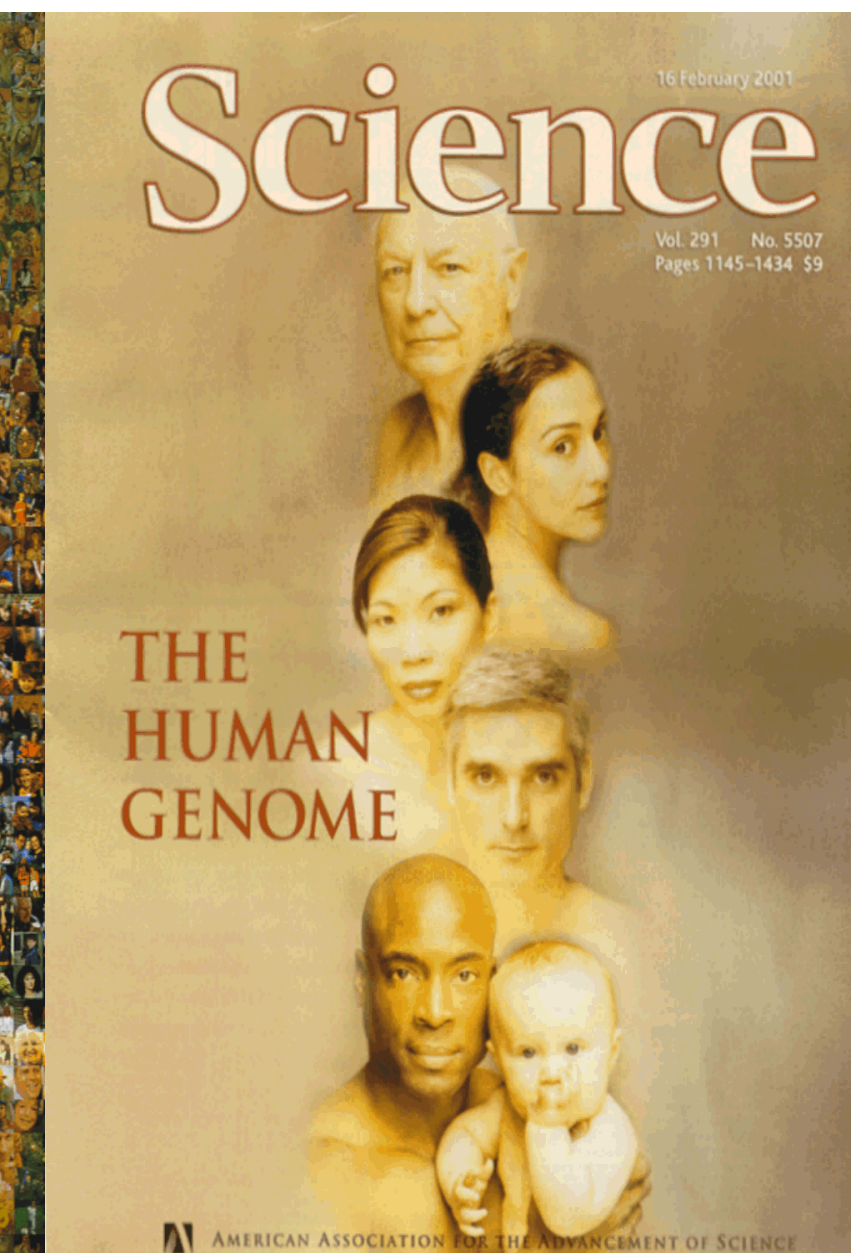
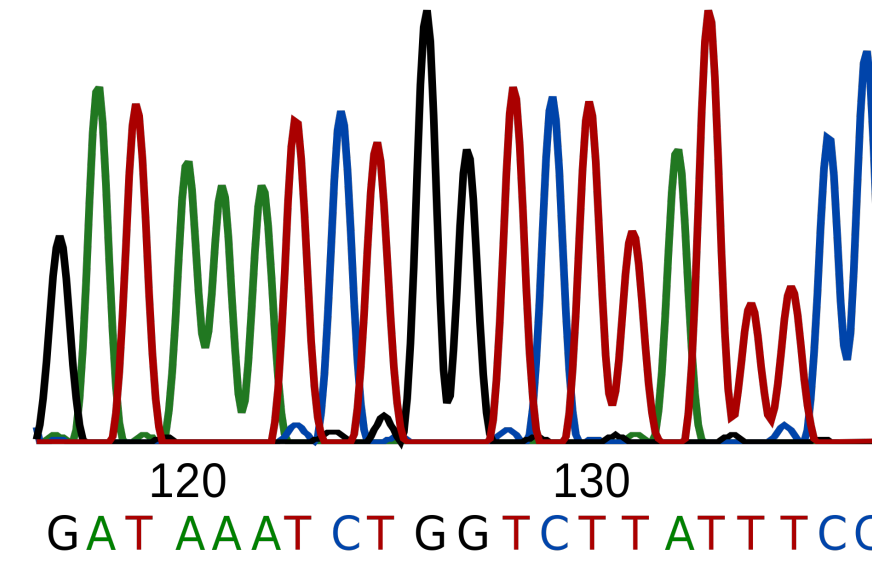
Po co mapowanie

- Mapowanie straciło znaczenie jako technika poznawania genomów
- Ciągłe jest ważną metodą identyfikowania genów odpowiadających za określony fenotyp
 - porównanie sekwencji genomów (np. zdrowego i chorego) da wiele różnic - która jest tą właściwą?
 - jakie zmienne elementy genomu wykazują w krzyżówkach/rodowodach sprzężenie z badanym fenotypem

Od genetyki do genomiki

Co już wiemy?

- Idea genu jako jednostki dziedziczności (początek XX w., na podstawie zapomnianych prac Mendla)
- druga połowa XX w. - poznanie mechanizmów działania genów
- 1977 - odczytywanie sekwencji DNA
- 1995 - pierwsze sekwencje całych genomów (bakterii)
- 2001 - genom człowieka - znamy wszystkie geny
- XXI. wiek - sekwencjonowanie nowej generacji
 - szybkie i niedrogi poznawanie genomów tysięcy ludzi
 - coraz lepiej rozumiemy, jak działa gen

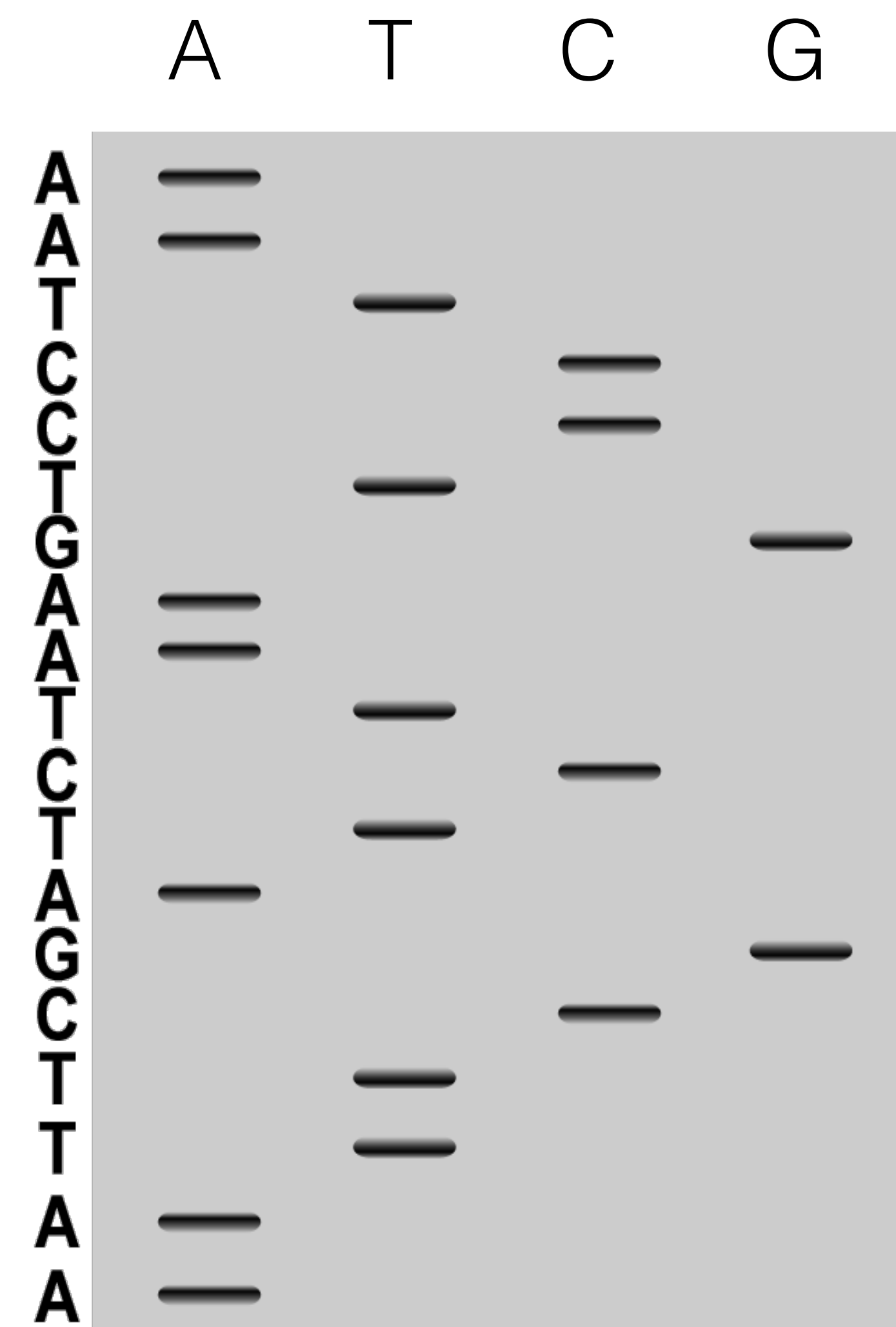
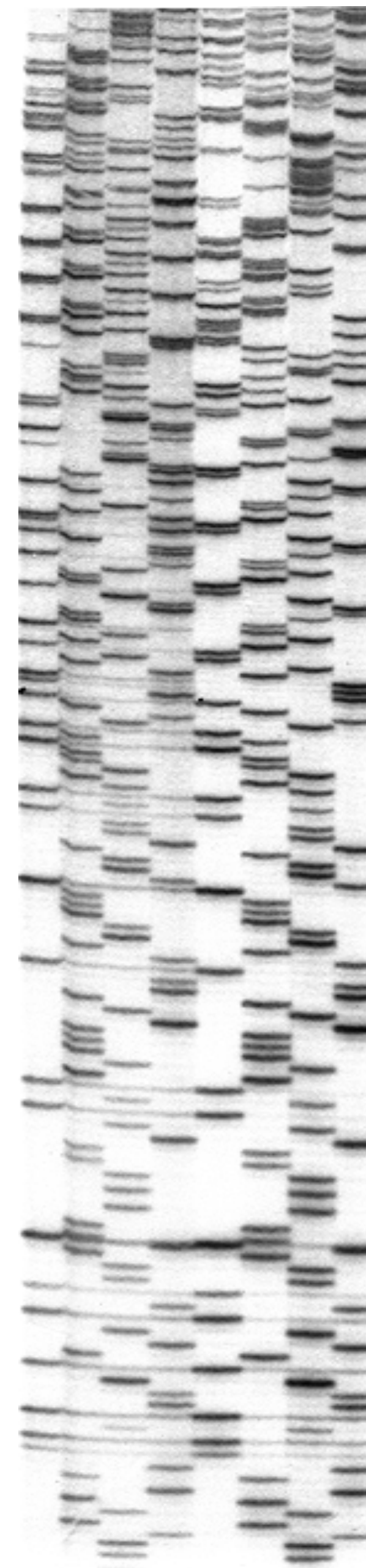


Lektura

- Allison “Podstawy biologii molekularnej”, rozdział 8 i 9

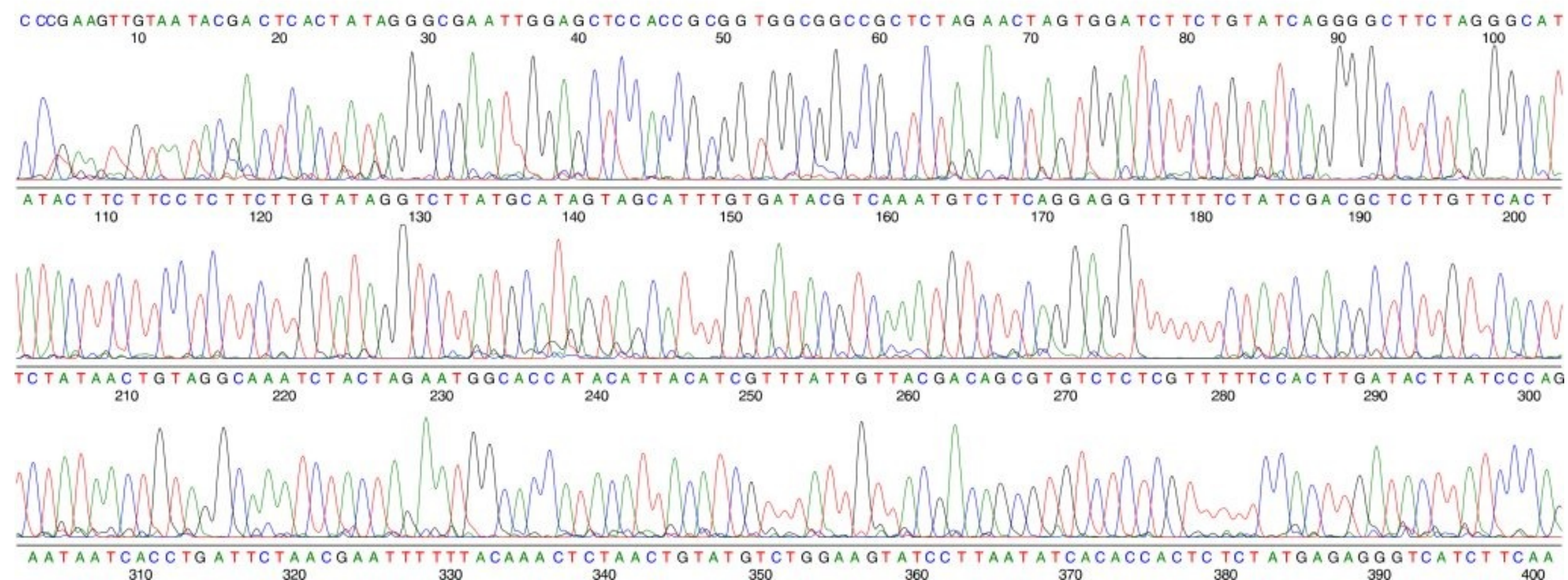
Tradycyjny odczyt sekwencji

- Metoda Sangera (1977)
- Synteza DNA w obecności analogów nukleotydów (forma dideoksy) terminujących reakcję na określonym nukleotydzie
- Znakowanie radioaktywne, osobne reakcje
- Konkurencyjna metoda Maxama-Gilberta, oparta na degradacji DNA nie przyjęła się (trudniejsza)



Sekwencjonowanie automatyczne - 2. generacja

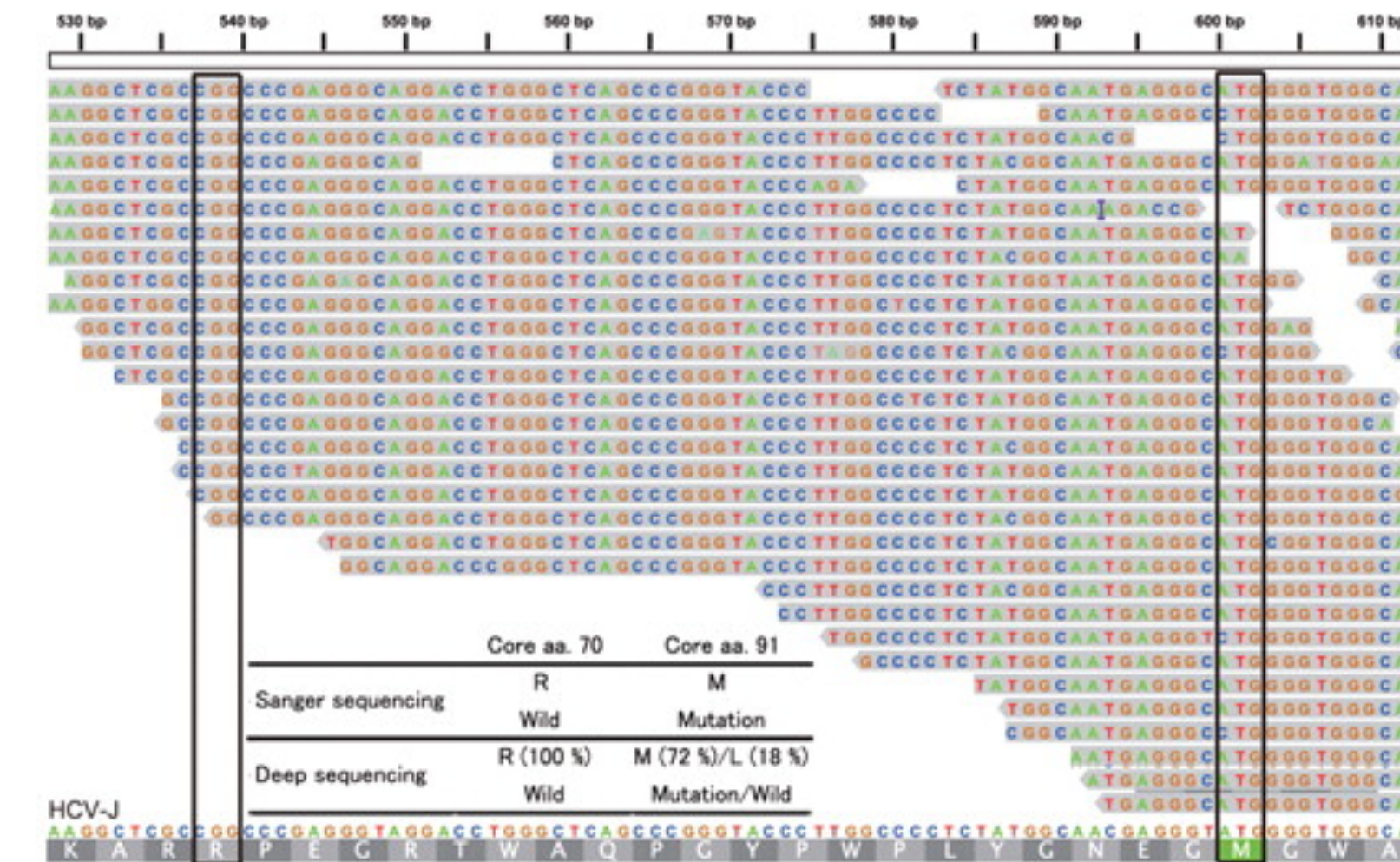
- Lata 90.
- Dideoksynukleotydy znakowane fluorescencyjnie (4 kolory)
- Elektroforeza kapilarna



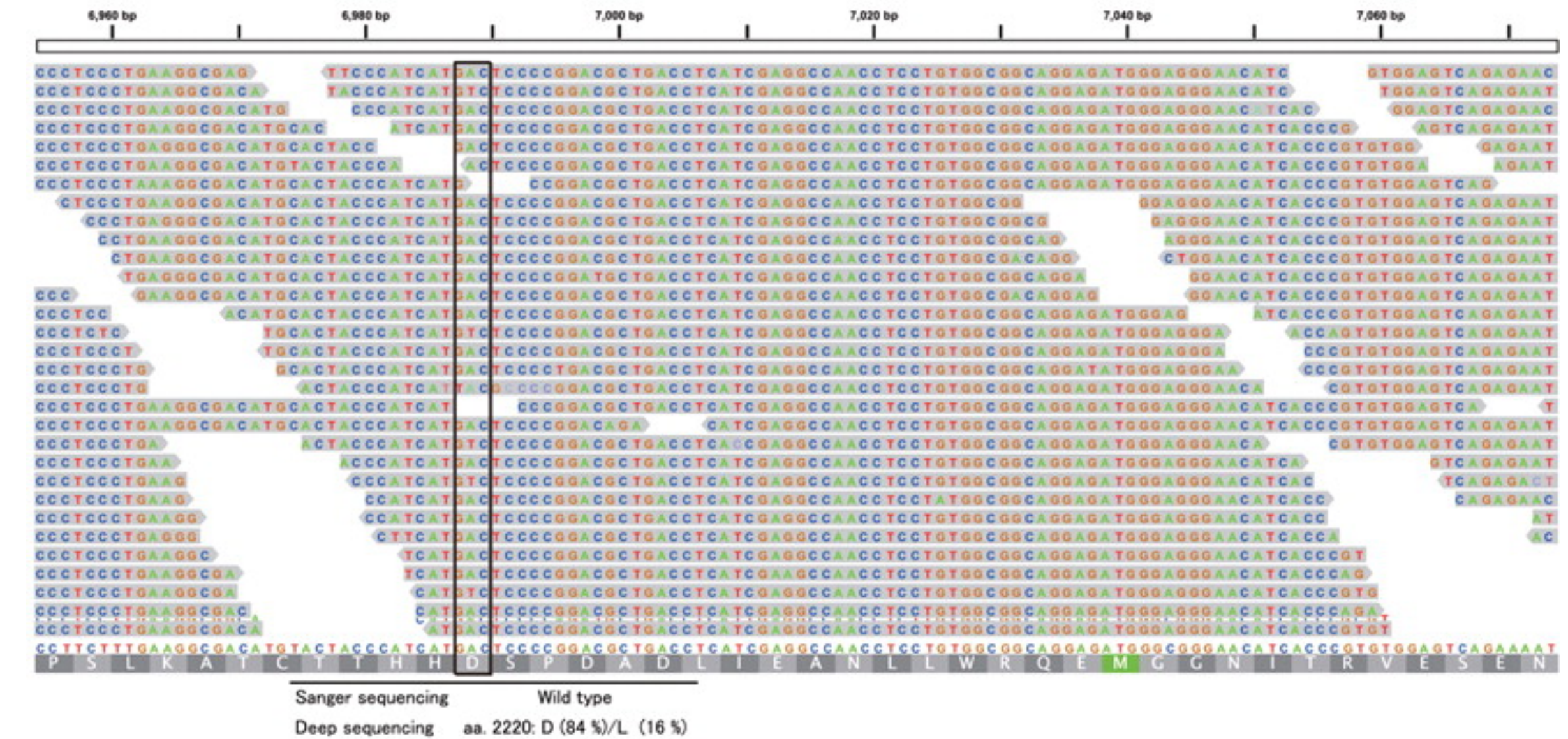
Sekwencjonowanie wysokoprzepustowe

- Tzw. *deep sequencing*, sekwencjonowanie nowej generacji (NGS)
- Generowanie w jednym przebiegu milionów niezależnych odczytów
- Pojedyncze odczyty krótkie (25-400 bp)
- Sekwencję odtwarza się przez nałożenie na siebie bardzo wielu krótkich odczytów
- standard: pokrycie 30-50x uznawane za "wysokie", 10x za niskie

A) Core aa. 70 and aa. 91

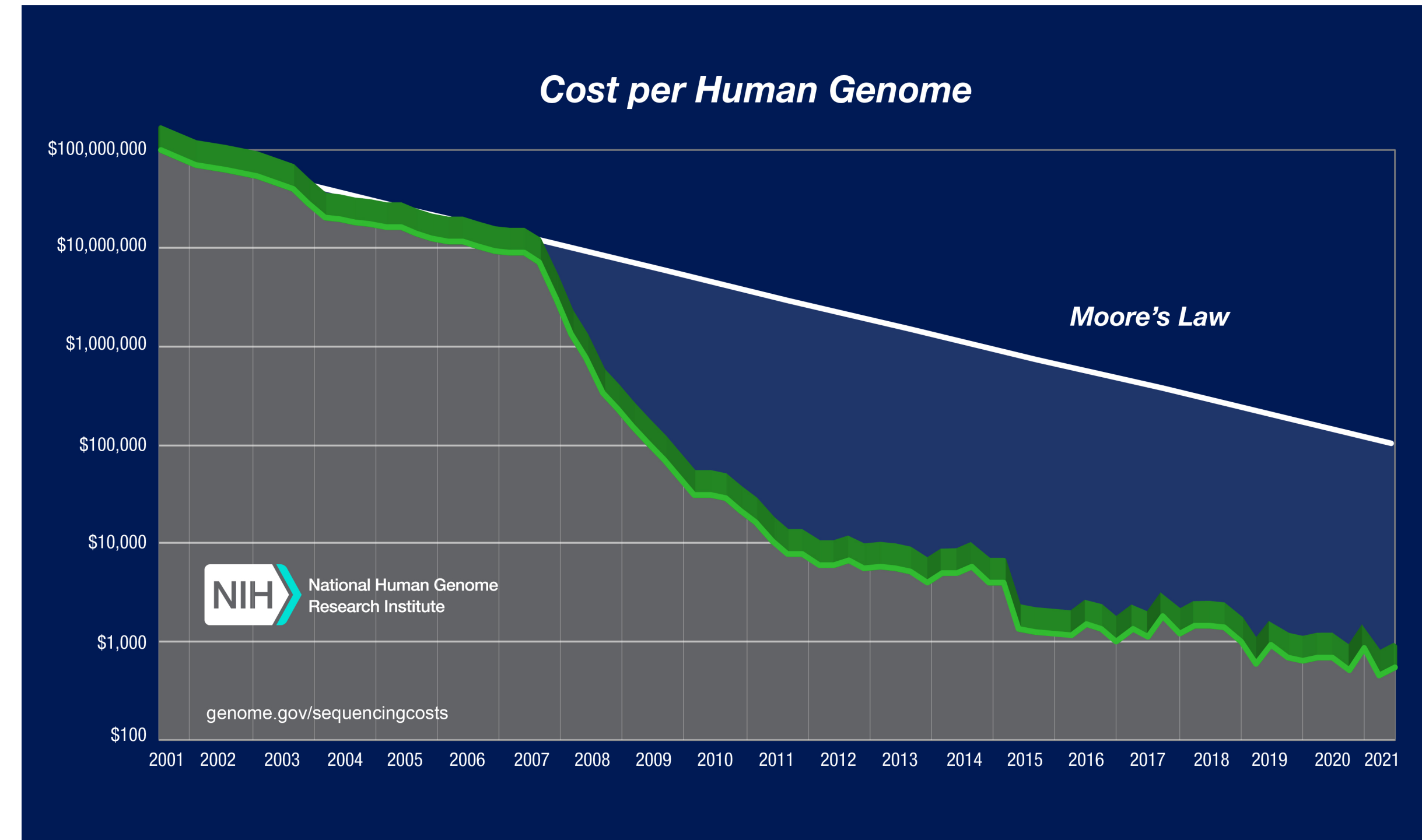


B) NS5A-ISDR



Sekwencjonowanie obecnie

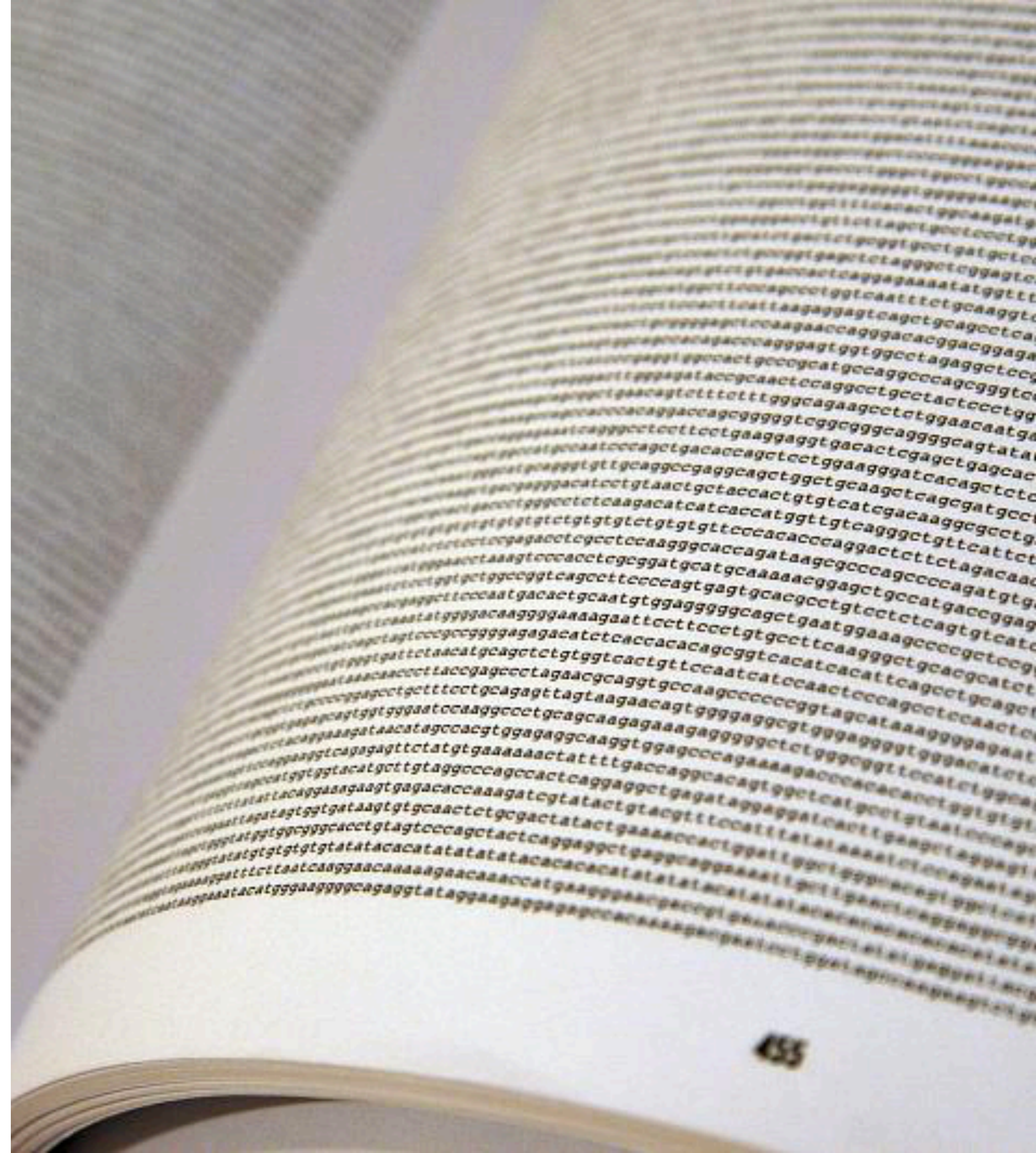
- Różne metody
 - sekwencjonowanie przez syntezę
 - sekwencjonowanie bezpośrednio
- Krótkie odczyty (25-400 bp), ale dużo i dokładne (Illumina, IonTorrent)
- Długie odczyty (tysiące bp), ale mniej i więcej błędów (Oxford Nanopore, Pacific Biosystems)



<https://www.genome.gov/>

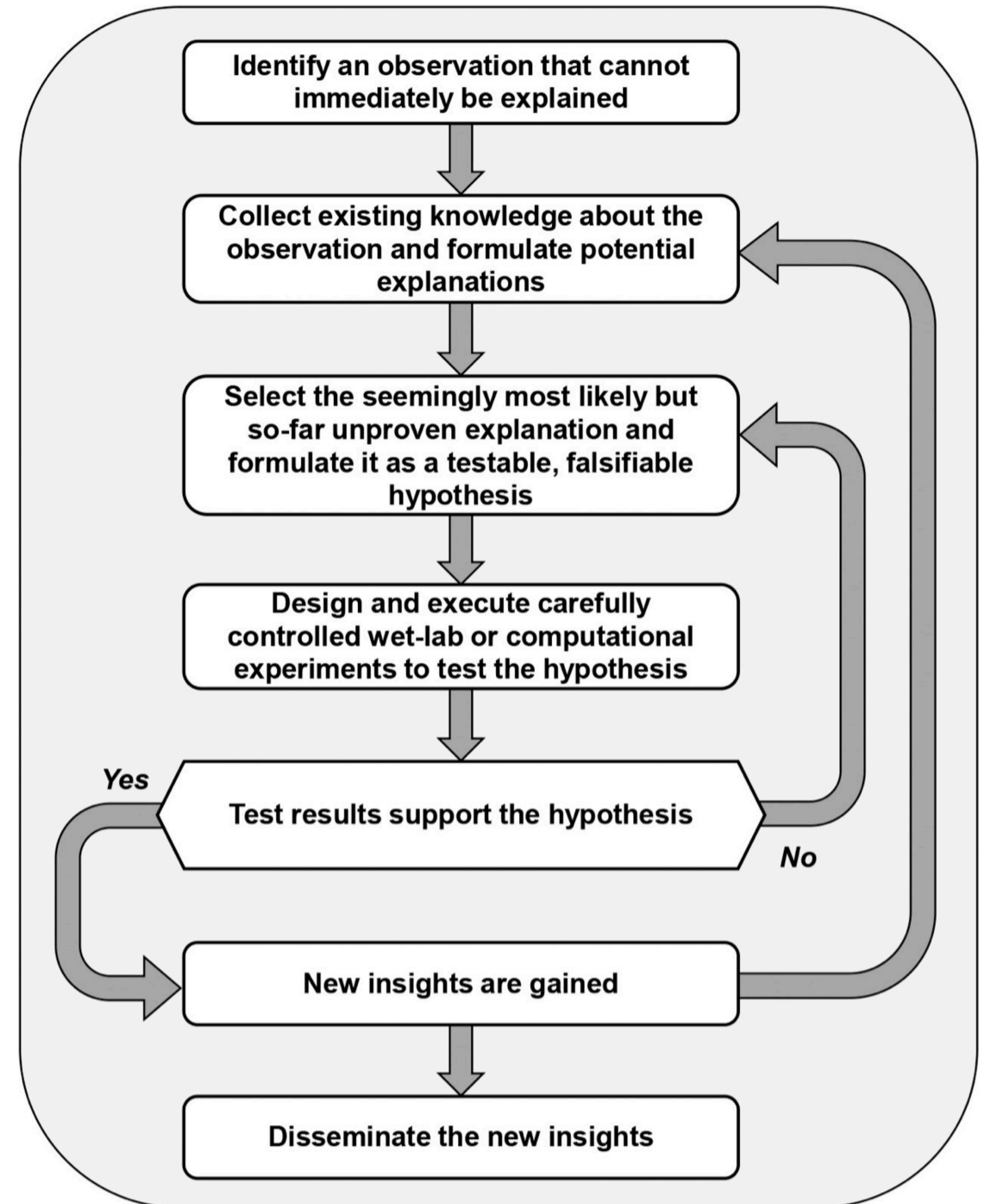
Reakcja na postęp metod genomiki

- "business as usual": to nadal taka sama biologia i genetyka, tylko na większą skalę
- "data-driven research": nowe podejście do metody naukowej



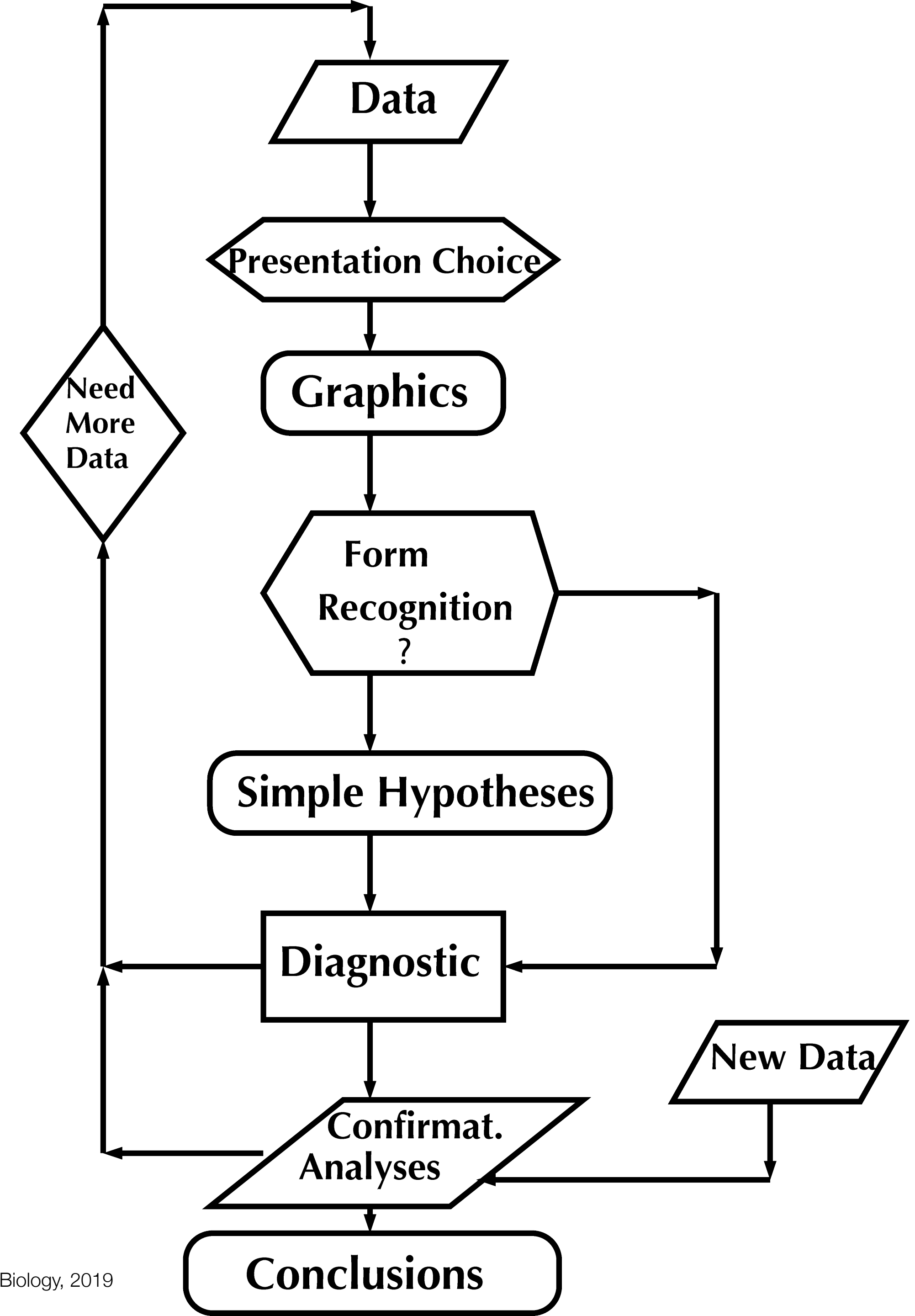
Współczesna metoda hipotetyczno-dedukcyjna

- Identyfikacja problemu badawczego
- Hipoteza robocza sformułowana **przed zbieraniem danych** z eksperymentów i obserwacji
- Eksperymenty i obserwacje
- Wariant formalny: hipoteza zerowa, czy dane falsyfikują hipotezę zerową
- Niezależnie od dyskusji filozoficznych - tak działa obecnie nauka (nauki empiryczne)



Metoda “*data-driven*”

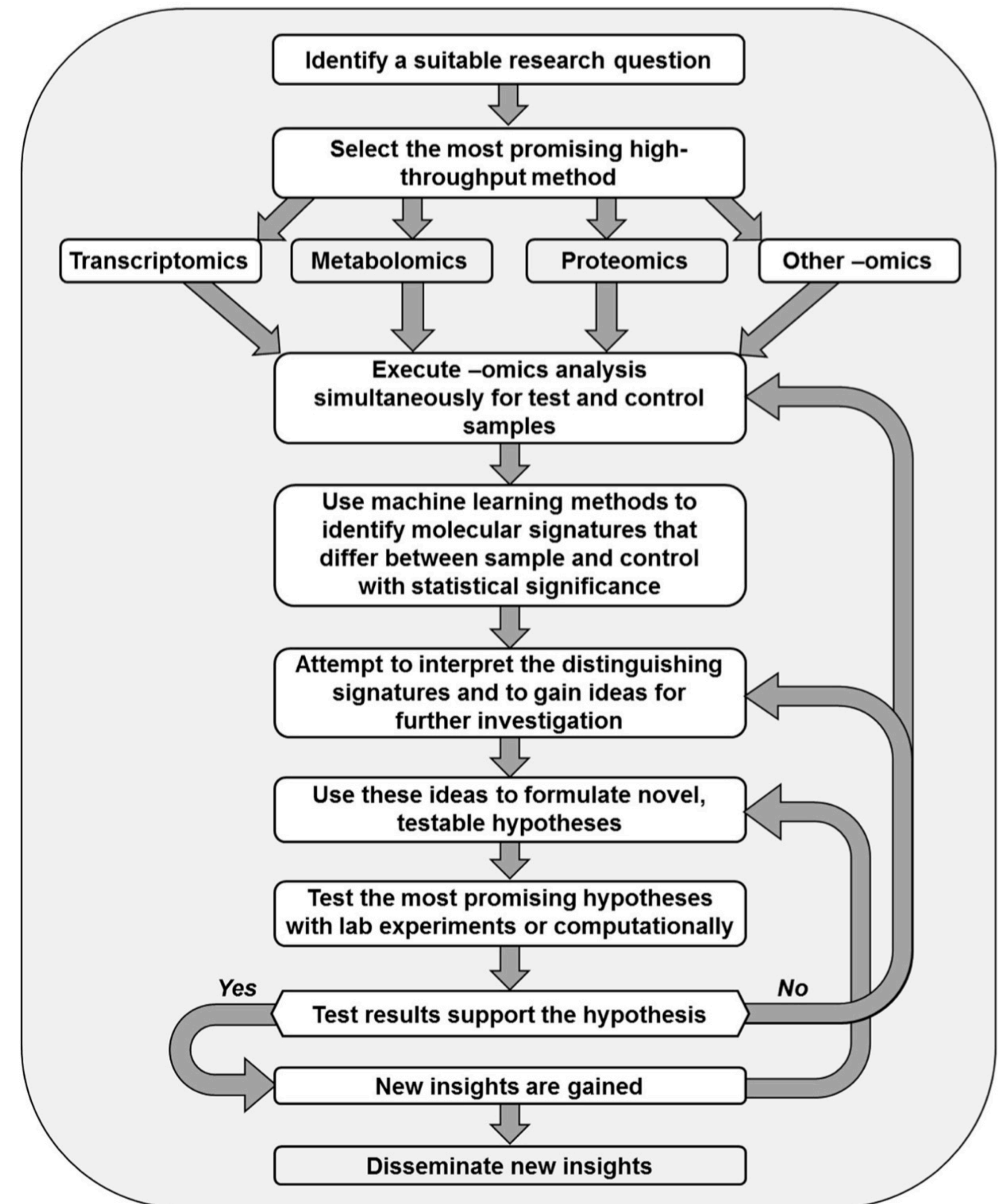
- Poszukiwania prawidłowości w danych **przed sformułowaniem hipotezy**
- Hipotezy formułowane po analizie danych, a nie przed
- Pierwszy krok *exploratory data analysis (EDA)* - analizy eksploracyjne, w oparciu o proste narzędzia, bez złożonych podstaw teoretycznych (Tukey, 1977)
- Kolejny krok: *confirmatory data analyses (CDA)* - analizy potwierdzające, w oparciu o koncepcje teoretyczne uzyskane dzięki analizom eksploracyjnym



Metoda “*data-driven*” wersja współczesna

- Pojawiła się w wyniku rozwoju genomiki i pokrewnych dziedzin
- Biologia zaczęła wytwarzać duże ilości danych ilościowych
- Opiera się na zbieraniu dużych (dążących do kompletności) zbiorów danych
- Poszukiwania prawidłowości w danych **przed sformułowaniem hipotezy**
- Wariant - analiza eksploracyjna z wykorzystaniem metod uczenia maszynowego

Citation: Voit EO (2019) Perspective: Dimensions of the scientific method. PLoS Comput Biol 15(9): e1007279. <https://doi.org/10.1371/journal.pcbi.1007279>



Koniec hipotezy?

- Podejście zakładające poszukiwanie prawidłowości w dużych zbiorach danych, zbieranych bez wstępnych założeń, może być produktywnie
- Ale niesie też (dobrze znane w literaturze) ryzyko
 - Lem, S. Cyberiada, Wyprawa szósta: czyli jak Trurl i Klapaucjusz demona drugiego rodzaju stworzyli, aby zbójcę Gębona pokonać., 1965.
 - Niby jakiej to autentycznej i drogocennej informacji pożądasz? – pyta Klapaucjusz.
 - Wszelkiej, byle prawdziwej – tamten na to. – Każda może się przygodzić na okoliczność życiową. Mam już wadoły moje i lochy pełnawe, ale się jeszcze drugie tyle pomieści. Gadajcie, co wiecie i umiecie, a ja sobie zapisze. Tylko predko!

go. Ów zaś siedział o beczkę oparty i w popiskiowaniu pisaka brylantowego, którym Demon spisywał na wstędze papierowej wszystko, czego się od atomów drgających dowiadywał, czytał o tym, jak się wije arlebardzkie wija i że córka króla Petrycego z Labaudii zwała się Garbunda, i co jadł na drugie śniadanie Fryderyk II, król bladawców, nim wojnę wypowiedział Gwendolinom, i ile powłok elektronowych liczyłby sobie atom termionolium, gdyby taki pierwiastek był możliwy, i jakie są wymiary dziurki tylnej małego ptaszka, zwanego kurkucielem, którego na swych rozamforach malują Marłajowie Wabędzcy, jak również o trzech smakach poliwonnych szlamu oceanicznego na Wodocji Przyzroczej, i o kwiatku Łubuduku, który myśliwych staromalfandzkich wali siarczyście na odlew, świtem wzruszony, i jak wyprowadzić wzór na dostawę kąta podstawy wieloboku, ikoseadrem zwanego, i kto był jubilerem

“[...] i pomału diabli go zaczęli brać, bo już mu świtało, że wszystkie owe całkiem prawdziwe i ze wszech miar sensowne informacje zupełnie nie są mu potrzebne, gdyż robił się z tego groch z kapustą, od którego głowa pękała, a nogi drżały.”

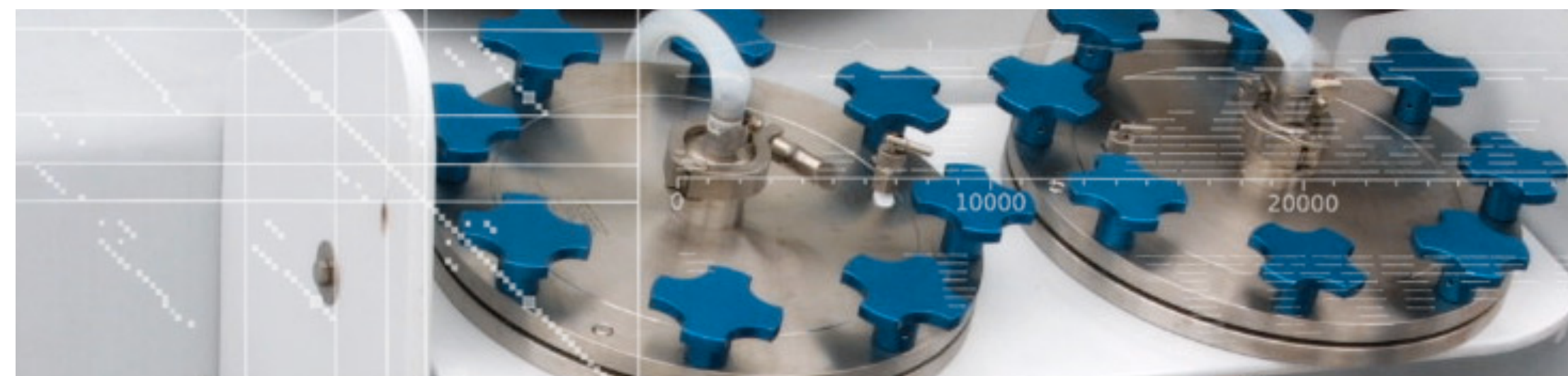
–Stanisław Lem, “Wyprawa szósta [...]”

Metagenomika

- Izolacja DNA ze środowiska i sekwencjonowanie
- Jedyne sposoby badania mikroorganizmów, które nie dają się hodować

Metagenomika

Analiza sekwencji całości DNA
wyizolowanego ze zbiorowiska organizmów



Odkrycia dzięki sekwencjonowaniu

- UCYN-A
 - Sinica (cyjanobakteria)
 - Niewielki genom (1,4mln par zasad, 1200 genów)
 - Brak zdolności fotosyntezy, cyklu Krebsa, syntezy niektórych aminokwasów
 - Zdolność asymilacji azotu
 - Symbioza (gospodarz - *Haptophyta*, *Primnesiophyta* - jednokomórkowe glony)
 - Jeden z głównych czynników asymilacji azotu w środowisku morskim

Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium

H. James Tripp¹, Shellie R. Bench¹, Kendra A. Turk¹, Rachel A. Foster¹, Brian A. Desany², Faheem Niazi², Jason P. Affortit² & Jonathan P. Zehr¹

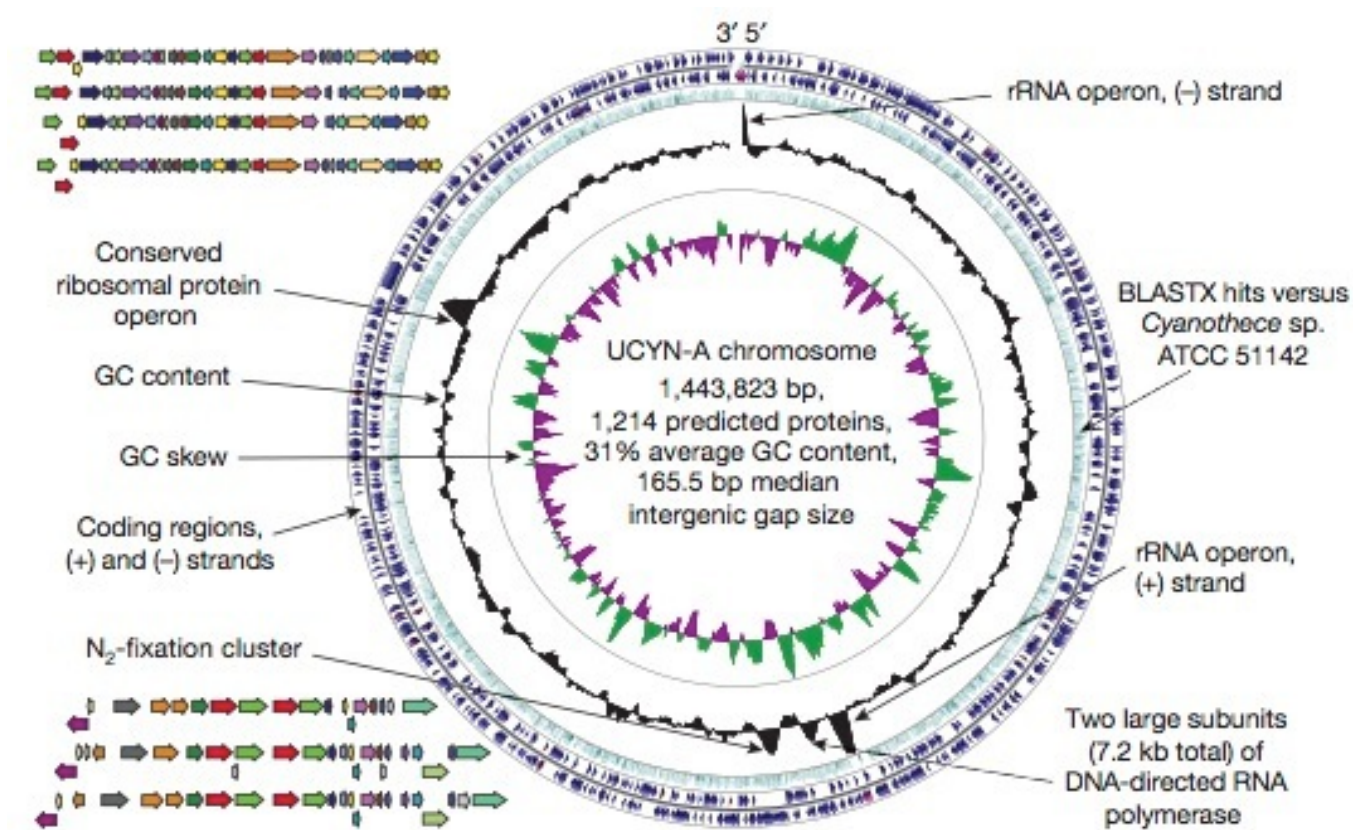
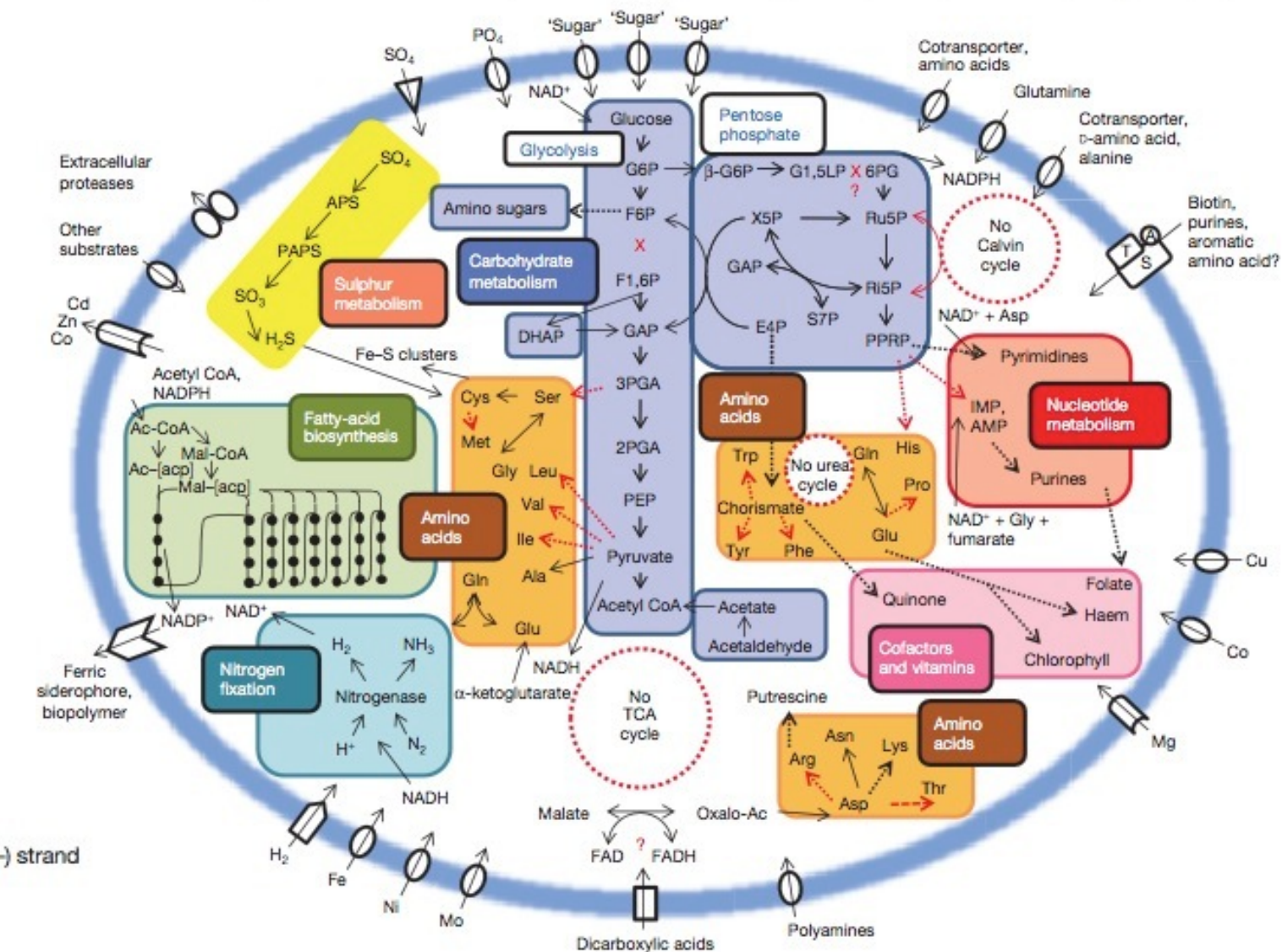
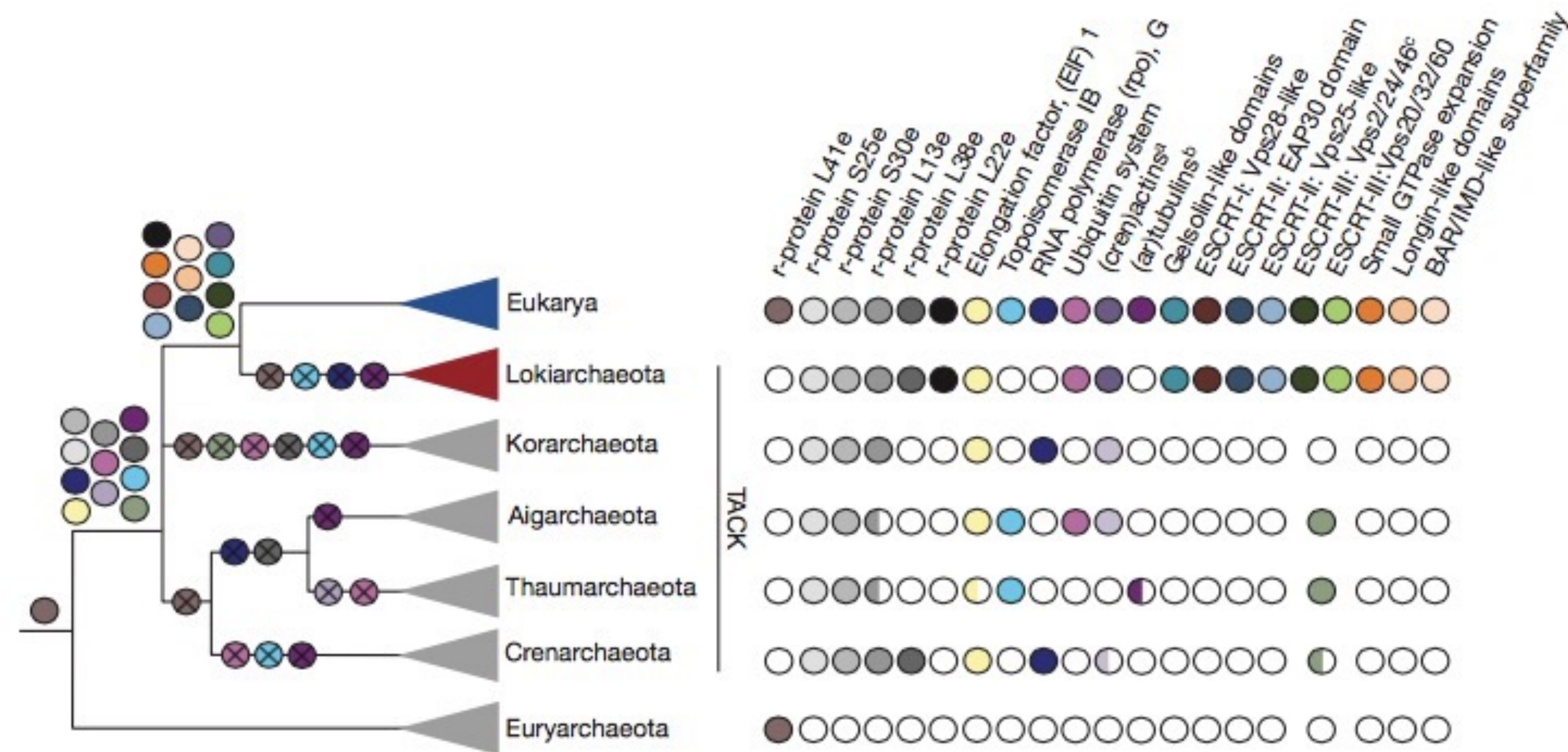


Figure 1 | Circular map of UCYN-A chromosome compared with Cyanotheca sp. ATCC 51142. The outer two rings show the two strands of the

Candidatus Atelocyanobacterium thalassa

Lokiarcheota

- Typ w domenie Archaea
- Zidentyfikowany na podstawie analiz metagenomowych (2015)
- Najbliżej spokrewniona z Eukaryota
- Posiada geny kodujące białka umożliwiające tworzenie złożonych struktur błonowych
- Tak mógł wyglądać gospodarz endosymbiozy, która dała początek Eukaryota



ARTICLE

doi:10.1038/nature14447

Complex archaea that bridge the gap between prokaryotes and eukaryotes

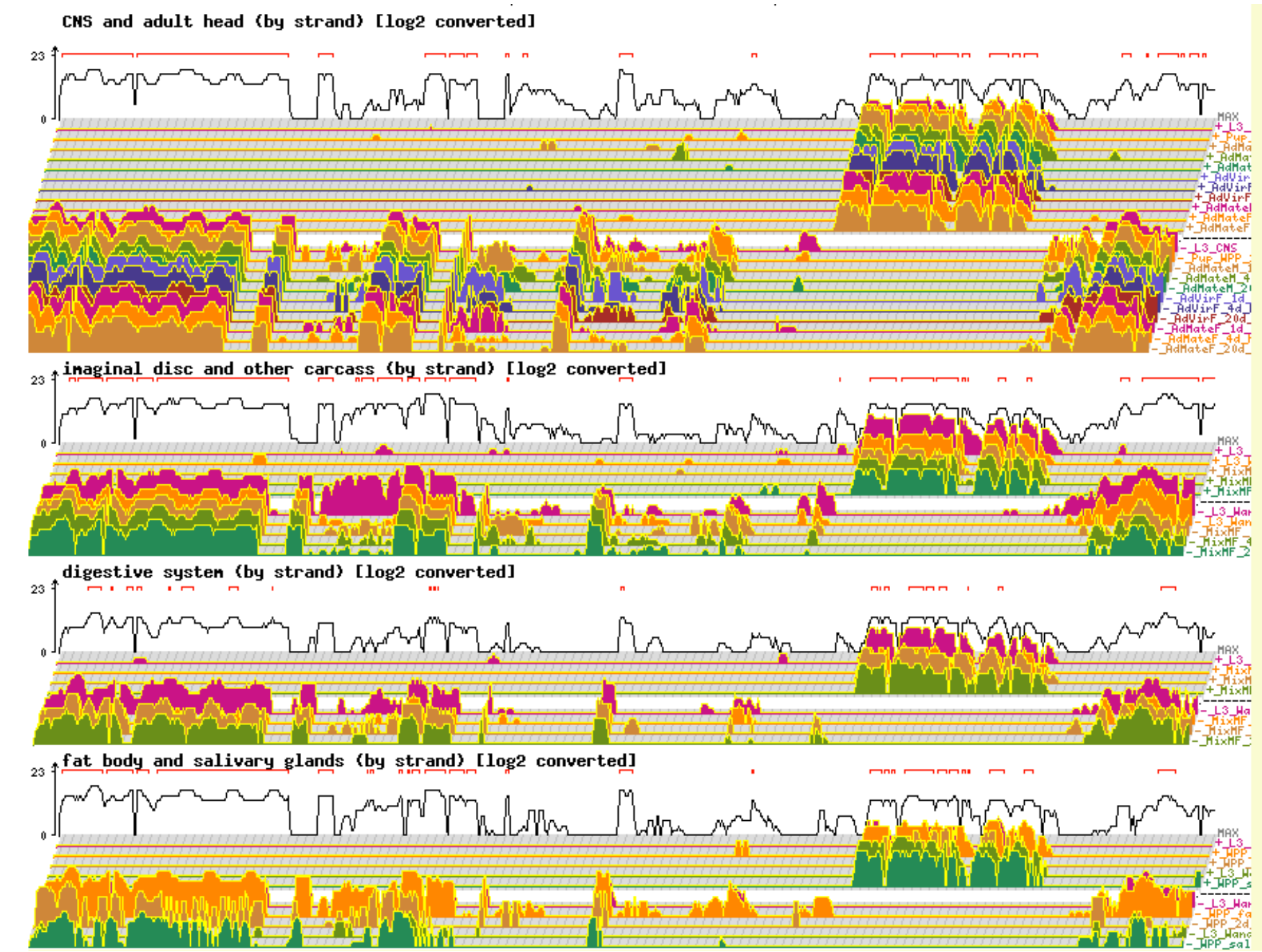
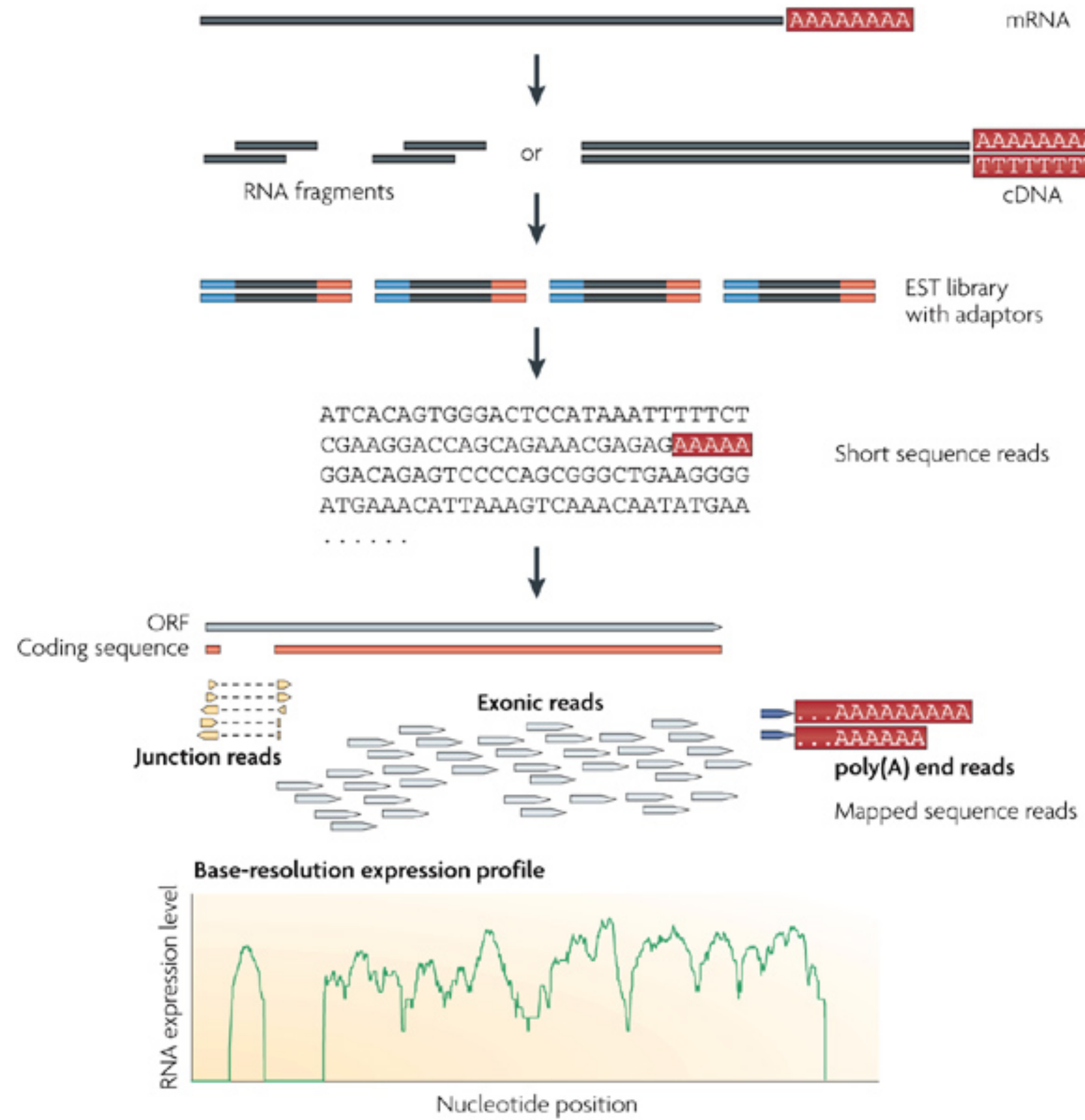
Anja Spang^{1*}, Jimmy H. Saw^{1*}, Steffen L. Jørgensen^{2*}, Katarzyna Zaremba-Niedzwiedzka^{1*}, Joran Martijn¹, Anders E. Lind¹, Roel van Eijk[†], Christa Schleper^{2,3,1}

©2015 Macmillan Publishers Limited. All rights reserved

Wielkie projekty

- Projekt 1000 genomów - różnorodność genetyczna człowieka
- Metagenomika mikrobiomu przewodu pokarmowego człowieka
- Genomy wymarłych gatunków (np. Neandertalczyk)

RNA-seq



Sekwencjonowanie nowej generacji – wyzwanie dla bioinformatyki

- Krótkie odczyty (50-150 nt)
 - pojedyncze
 - “paired-end”
- Problem identyfikacji i składania sekwencji
- Indeksowanie i multipleks

Genomika funkcjonalna

- Wysokoprzepustowe analizy:
 - ekspresji genów (mikromacierze, RNA-seq)
 - proteomu
 - interakcji genetycznych i fizycznych
 - fenotypów



Co to znaczy? Genomika funkcjonalna.

TCACAATTTAGACATCTAGTCTTCCACTTAAGCATATTTAGATTGTTTCCAGTTTTCCAGCTTTTATGACTAAATCTTCTAAAATTGTTTTCCCTAAATGTATATTTAATTTG
TCTCAGGAGTAGAATTTCTGAGTCATAAAGCGGTCATATGTATAAATTTAGGTGCCTCATAGCTCTTCAAATAGTCATCCCATTTTATACATCCAGGCAATATATGAGAG
TTCTTGGTGCTCCACATCTTAGCTAGGATTTGATGTCAACCAGTCTCTTTAATTTAGATATTCTAGTACATACAAAATAATACCTCAGTGTAACCTCTGTTTGTATTTCCCT
TGATTAAGTATGCTGAGCACATCTTCATGTGCTTATTGACCATTAATTAGTCTTATTTGTTAAATGTCTCAAATATTTTATACAGTTTTACATTGTGTTATTCATTTTTTAAA
AAATTCATTTTAGGTTATATGTATGTGTGTGTCAAAGTGTGTGTACATCTATTTGATATATGTATGTCTATATATTCTGGATACCATCTCTGTTTCATGCATTGCATATATATTT
GCCTATTTAGTGGTTTATCTTTTCATTTTCTTTTGGTATCTTTTCATTAGAAATGTTATTTATTTGAGTAAGTAACATTTAATATATTCTGTAACATTTAATGAATCATTTTATG
TTATGTTTAGTATTAATTTCTGAAAACATTCTATGTATTCTACTAGAATTGTCATAATTTTATCTTTTATATACATTGATATTTTATGTCAAATATGTAGGTATGTGATATTATG
CACATGGTTTTAATTCAGTTAATTGTTCTTCCAGATGTTTGTACCATTCCAACATCATTTAAATCATTAAATGAAAAGCCTTTCCCTACTAGCTAGCCAGCTTTGAAAATC
CATTCATAGGGTTTGTGTTAATATATTTTGTCTTTTTTTCCCTTCTACTGATCTCTTATATTAATACCTACTGTGGCTTTATATGAAGTCATGGAATAATACGTAGTAAG
CCCTCTAACACTGTTCTGTTACTGTTGTTATTGTTTTCTCAGGGTACTTTGAAATATTCGAGATTTTATTATTTTTTAGTAGCCTAGATTTCAAGATTGTTTTGACGATCAAT
TTTTGAATCAATTGTCAATATTTTAGTAATAAAATGATGATTTTATTGTTGAAATACATTAAATCTATAAGCCAAATTGGAGATTATTGATATATTAACAAAAATGAGTTTTCC
AGTCCATGAATGTATGCACATTATAAAATTCATTCTTAAGTATGTCATTTTTAAGTTTTAGTTTTCAGCAGTATATGTTTGTACATAGGTAAACTCCTGTCATGGGGGTTA
GTTGTACAGGTTATTTTATCATCCAGGCATAAAGCCCAGTACCCAGTAGTTATCTTTTCTGCTCCTCTCCCTCCTGTCACCCTCCACTCTCAAGTAGACCCCAGTTTC
TGTTGTTCTCTTCTTTGCATTAATGACTTCTCATCATTTAGATTGCACTTGTAAGTGAGAACAGGACGTATGTGGTTTTCTACTCCTGTGTTAGTTTGCTAAGGATAACC
ACCTCCATCTCCATCCATGTTCCCACAAAAGACATGATCTCCTTTTTTATGGCTGCATATTATTCCATGGTATATATGTACCACATTTTCTTTATCCAATCTGTCATTGATG
GACATTTAGGTTGTTTCCACATCATTGCCGTTGTAATACTGCTGCAGTGAATATTCGTGTGTATGTCTTTATGGTAGAATGATTATATTCCTCTGGGTATATTTCCAAGT
AATGGGATGGTTGGGTCAAATGGTAATTCTGCTTTTAGCTTTTTGAGGAATTGCCATATTGCCTTTACAAACGGTTGAACTAATTTATACTCCCAAGAGTGTATAAGTTG
TTCCTTTTTCTCTGCAACCTCGACATCACCTGTTATTTATGACTTTTTATATAATAGCCATTCTGCTGGTCTGAGATGGTATCTCATTATGATTTTATTGATTTGCATTTCTCTAAT
GCTCAGTGATATTGAGCTTGGCTGCATATATGTCTTCTTTTAAAATATCTGTTTATGTCCTTTGCCTAATTTATAACGGGGTTGTTTGTTTTTCTCTTGTAATTTGTTTAA
GTTCCCTTATAGATTCTAGGTATTAACCTTTTTTTCAGAGGCGTGGCTTGCAAATATTTTCTCCCATTCTATAGGTTGTCTGTTTATTCTGTTGATAGTTTCCCTTGCTGTG
CAGAAGCTCTTAACTTTAATTAGATCCGACTTGTCAATTTTTGCTTTGGTCGCAATTGCTTTTATGTTATTGTCGTGAAATCTTTGCTAGTTCTTAGGTCCAGGATGATA
TTGCCAAGTTGTCTTCCAGGGCTTTATAATTTGGATTTTACATTTAAGTCTTAATATATTTATTAATTTGTTAGGGTTTTCAGGATACAAGGACAATATAGCAGCAAAC
AATGTAAGTAAAATCTGAAAATAATAGAAAACAGTTAATTGAACACTTTACCATTATGTAATGCCCTTCTTTGTCTTTCCCTGATCTTTGTTGGTTTGAAGTTCAAAAA
AGACAACTTAATGGTACAATAGGTATTGTAGATTTTCCAGGACTTTCTGTATAAAATATTTTGTATATATGAATAGATCATTTTTTATTCCAGTCTTTAAACATTTTCTTAACAT
TTTCTTCTATTGCTTCACTTCACTCGCTAGGACCATCAGGACAGTGTGAAACAGAAATTGTCAGACTGATCATCACAACTTTTTCTAGATTTTAGAAGGAAATTTTTCTT
TATTTCAACATAAAGCAGCATGTTAATGCCAAGTTTTAATATGTGTTATCAGATTGAAATTTTTTGTATATTTCTACATTACCAAGAATTTTTAGCAAGAGTTTTTGTGAG
TTTTAATTTAAAATCATTTGTTAATTTTCTGATTTTTTATTCTCTTTTTACCTAAGAGATTAACACTGACTACAGATTGAATATAAACAAACAAACAAACAAAA
CTCTAAAATGCTGTGGATCAACACCACTTAGTAATTTGTATACTTGGATTCAATTTGCTGAAATTTTGTAGACATTTTTCGTCGATATTTATGAGGGATGTTGATCTGT
AAAAGTATTAATGCTTTTGCAGATAGTGTACCCATATAAAAACTTTGAACAAAATCAGATTATCACTGTGGATATTTCTATTTTGAACAACTTAGATGATAATTT
AATCTATATCCTAGATGAACT

Mały fragment chromosomu 21

Odwrotna genetyka – od genu do funkcji

Genetyka tradycyjna

Funkcja (mutacja, fenotyp)



Klonowanie genu



Analiza sklonowanego genu

Genetyka „odwrotna”

Gen (z sekwencji całego genomu)

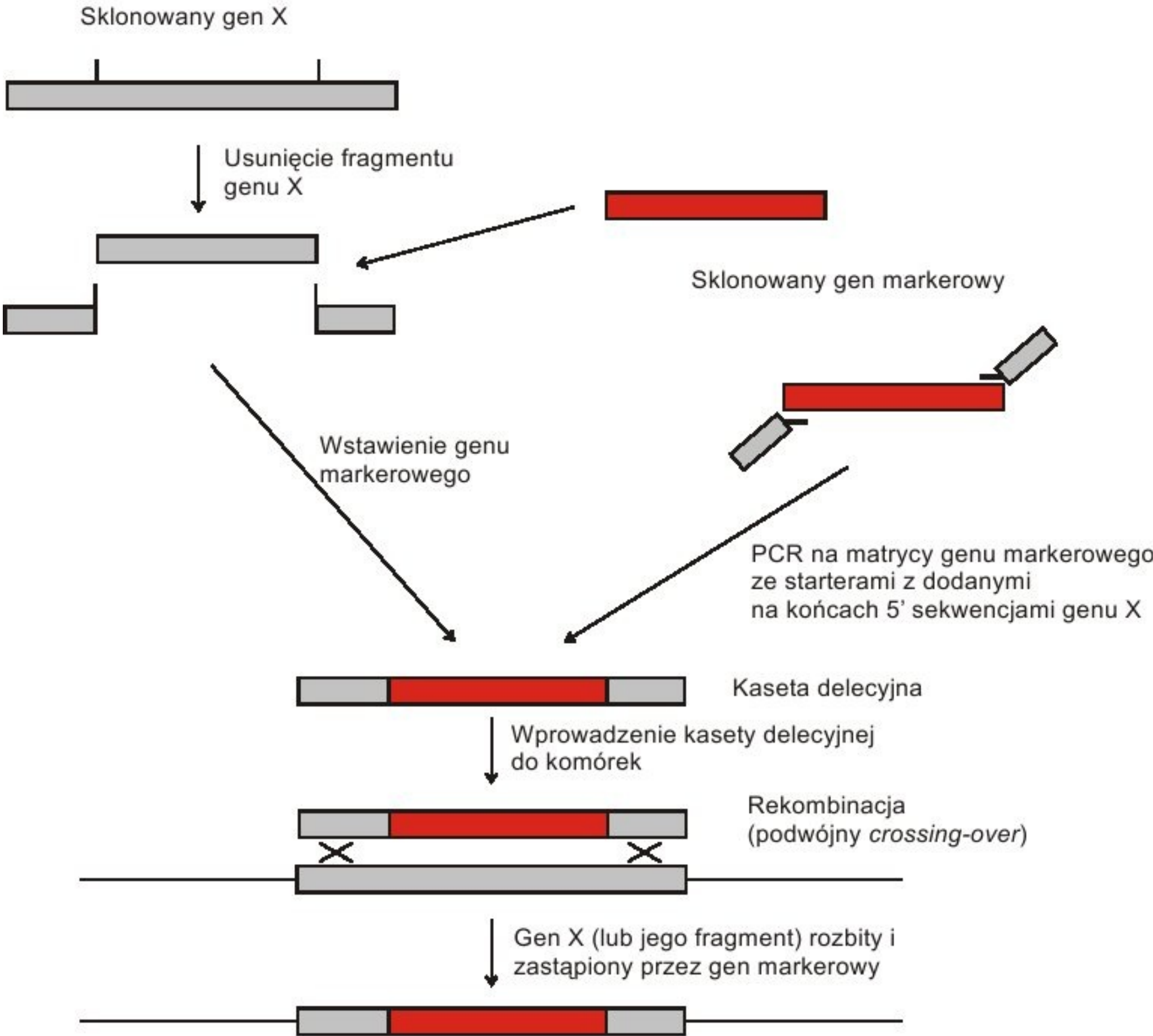


Inaktywacja genu



Analiza uzyskanego fenotypu

Odwrotna genetyka – inaktywacja przez rekombinację



Odwrotna genetyka – interferencja RNA



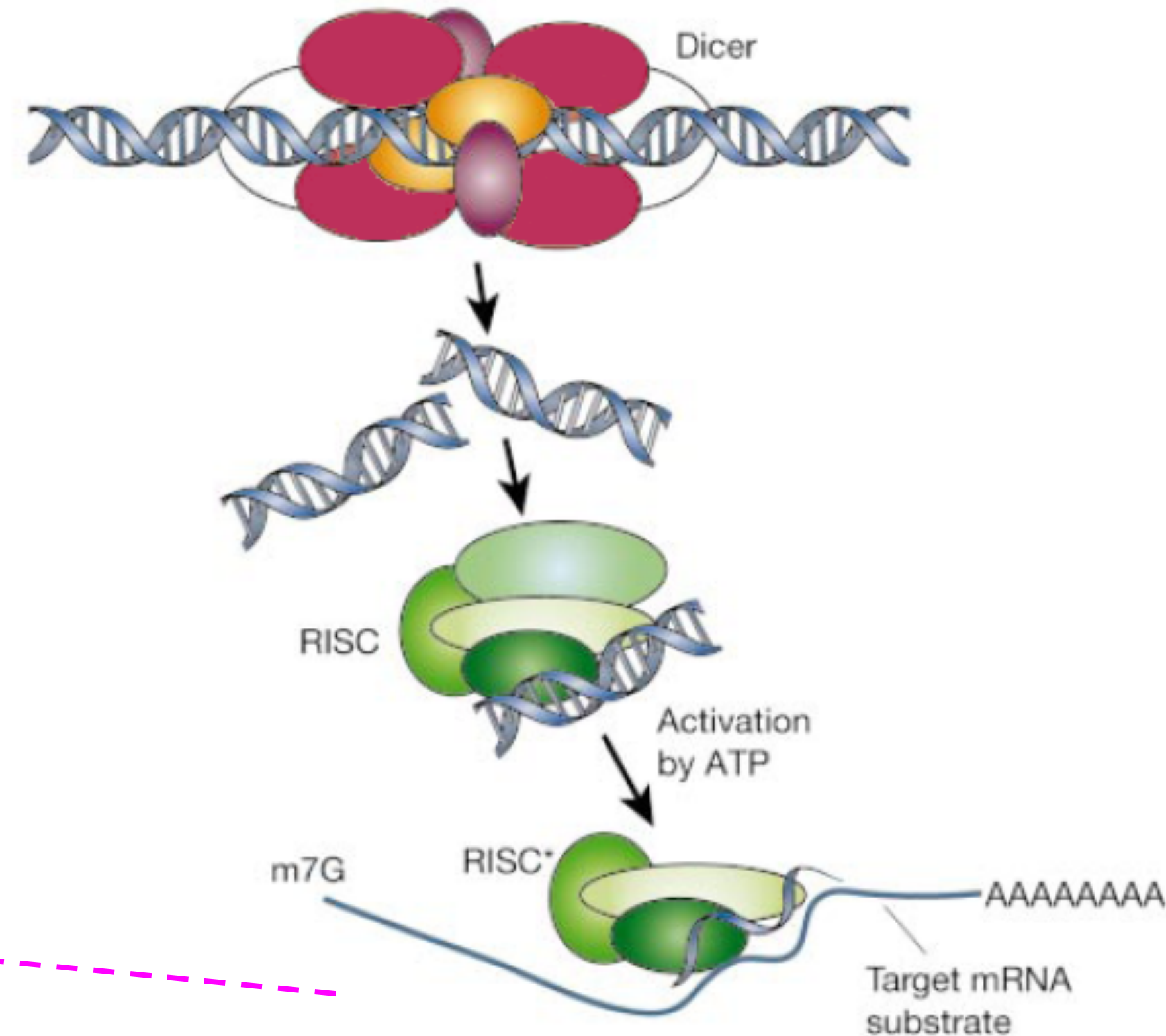
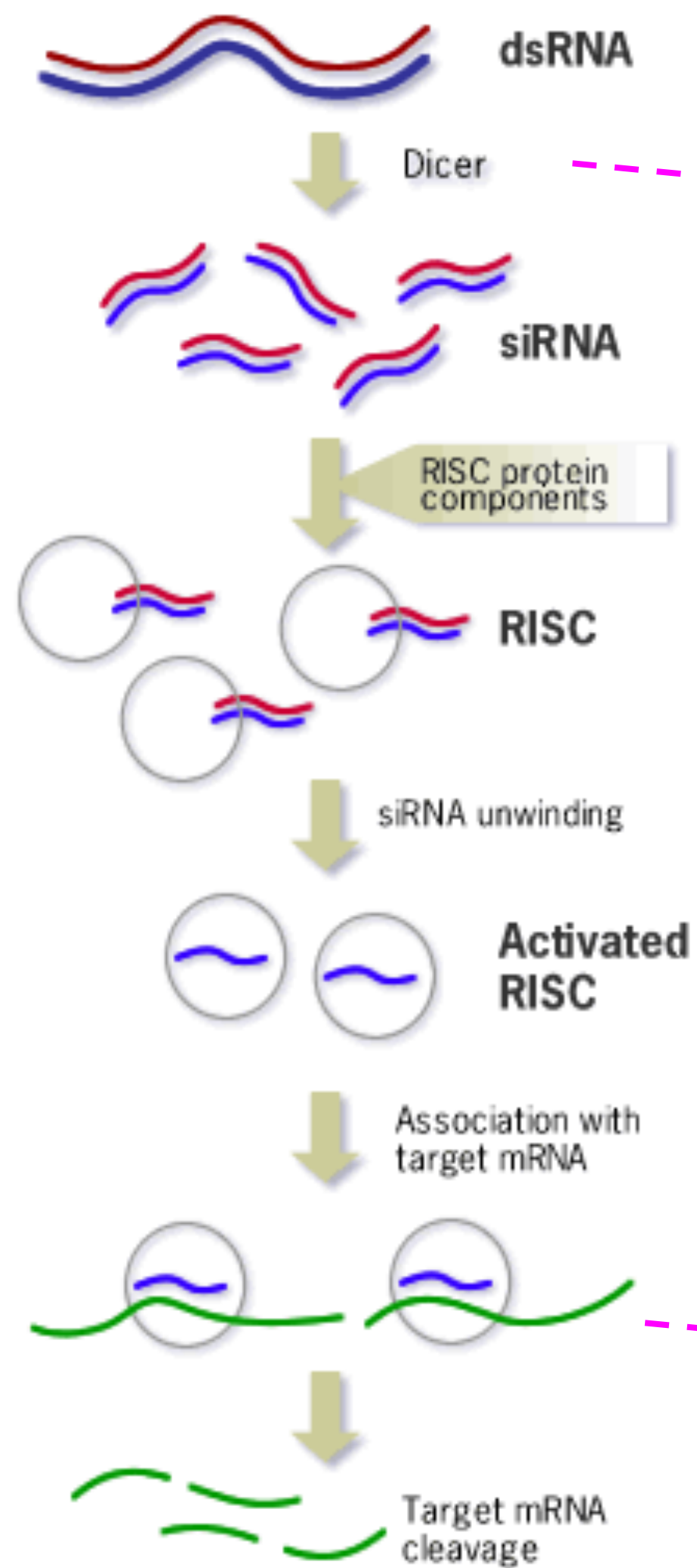
Odkrycie roku 2002 –
regulacyjna rola małych RNA

Nagroda Nobla w dziedzinie
medycyny 2006, za odkrycie
mechanizmu interferencji

RNA

A. Fire i C. Mello

siRNA - jak to działa?



Efekt – degradacja mRNA

Modyfikacje genomu: od inżynierii genetycznej do biologii syntetycznej

Zastosowania

- Badania podstawowe
- Biotechnologia

Granica między badaniami podstawowymi a stosowanymi jest płynna, stosowane techniki są podobne, różnice dotyczą głównie skali.

Podstawowe techniki

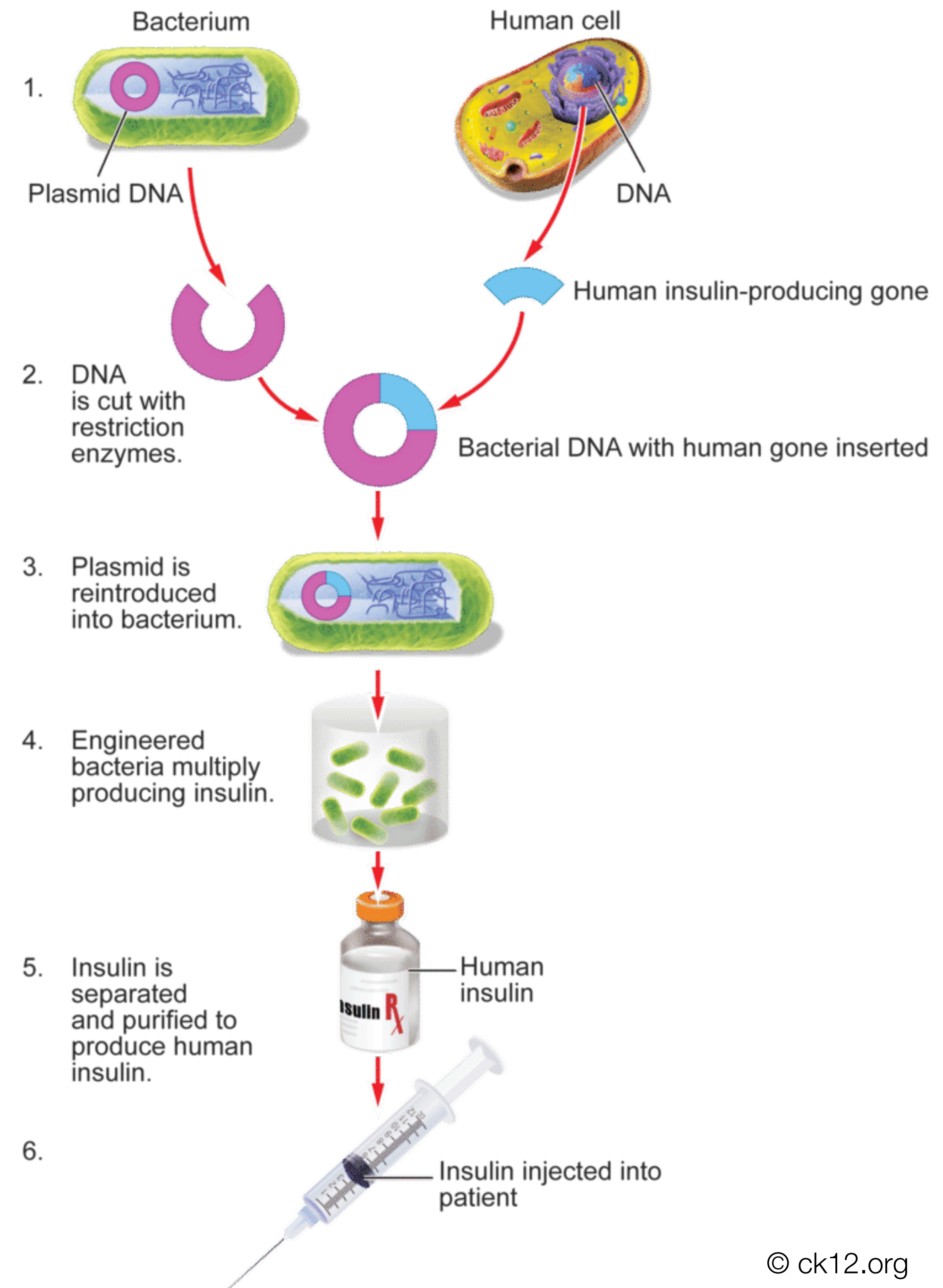
- Izolacja DNA lub RNA
 - cDNA – izolacja RNA i przepisanie na DNA
 - Chemiczna synteza DNA *de novo*
- PCR
- Klonowanie DNA
- Mutageneza losowa i ukierunkowana
 - w tym wprowadzanie modyfikacji do genomu
- Wykrywanie DNA, RNA i białek
- Sekwencjonowanie

Lektura

- Allison “Podstawy biologii molekularnej”, rozdział 8 i 9

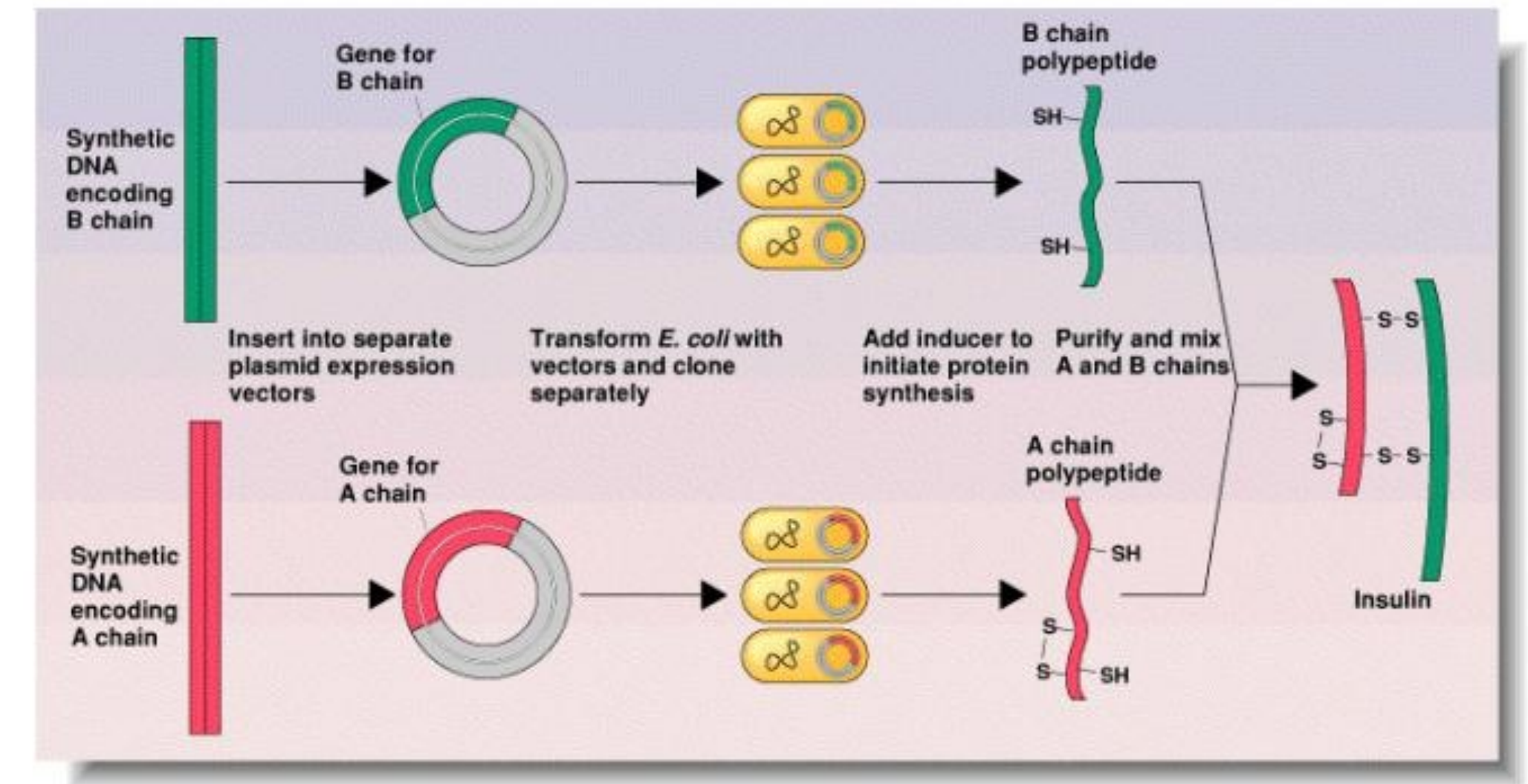
Inżynieria genetyczna

- Zbiór metod w użyciu od lat 70.
- Dla zmodyfikowania genomu należy:
 - wyizolować DNA
 - zmodyfikować DNA *in vitro*
 - wprowadzić zmodyfikowany DNA do komórek (tego samego lub innego organizmu)
- Narzędziami są (często zmodyfikowane) enzymy: restryktazy, ligazy, polimerazy, itp.



Ekspresja heterologiczna

- Wytwarzanie w komórkach białka naturalnie w nich niewystępującego dzięki wprowadzeniu odpowiedniego genu
- Np. bakterie wytwarzające ludzką insulinę, i wiele innych
- Leki wykorzystujące produkowane w komórkach zmodyfikowane przeciwciała
- Fuzje białkowe - łączenie różnych fragmentów białek przez łączenie sekwencji DNA



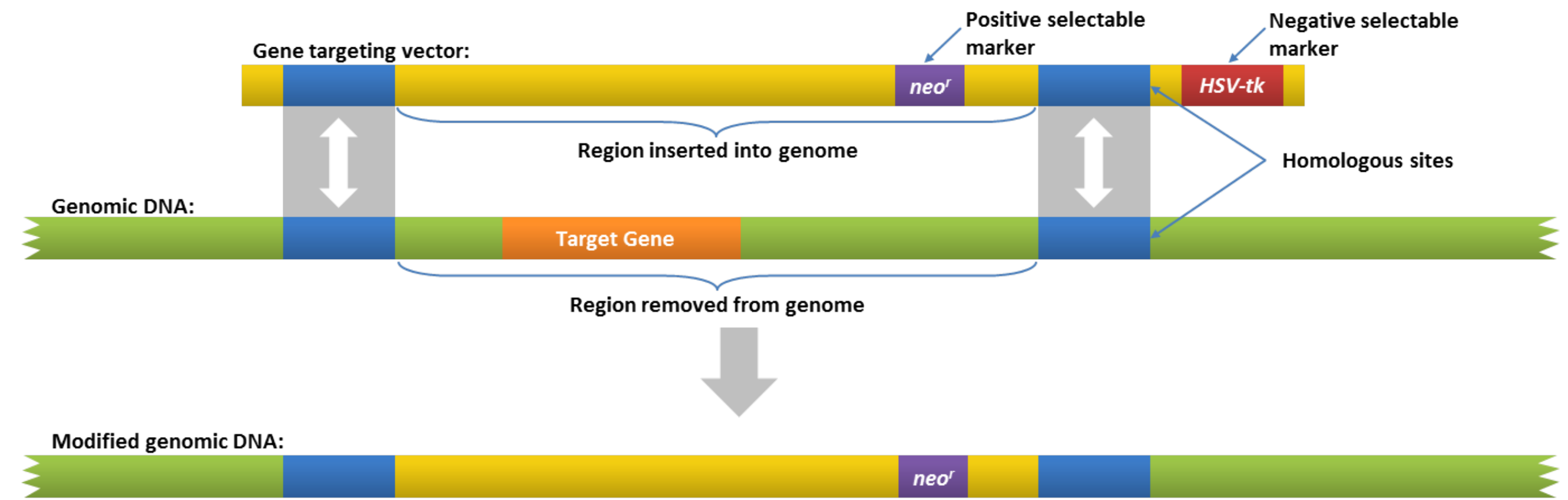
Harcourt Brace & Company



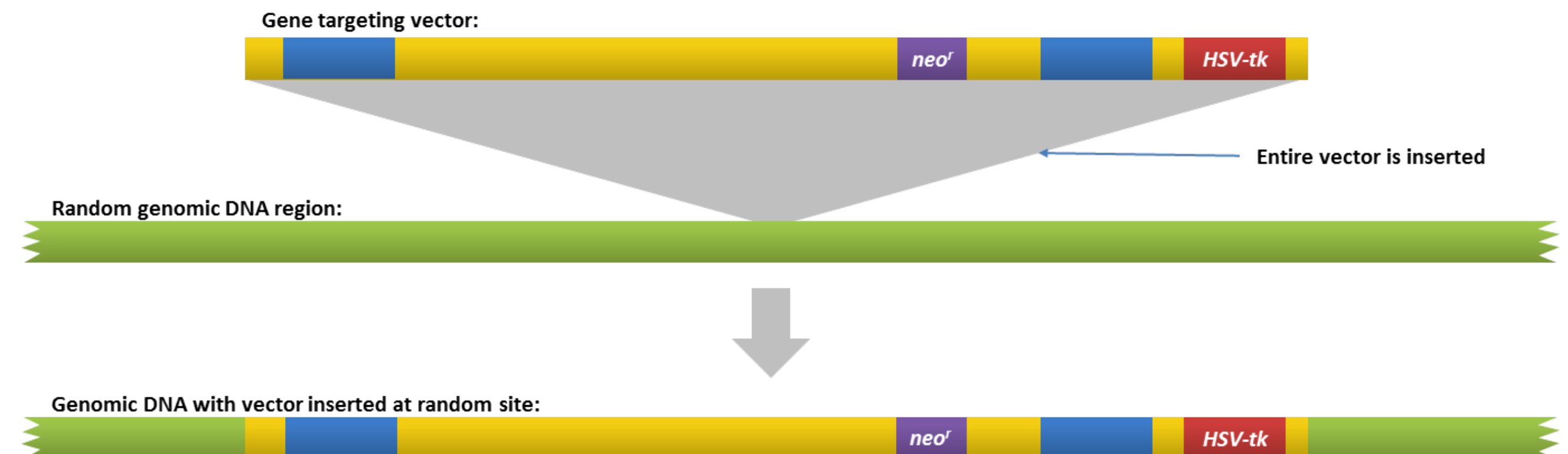
Główne wyzwania inżynierii genetycznej

- Wprowadzenie zmienionego DNA do genomu gospodarza
- W ściśle określone miejsce (targeting)
- Wydajność różna u różnych organizmów
 - np. drożdże - b. wysoka
 - człowiek b. niska
- Dodatkowe kroki potrzebne by nie pozostawić śladów w genomie (np. markerów selekcyjnych)

a) Homologous recombination:



b) Random integration:



Redagowanie genomu

- Wprowadzenie zmiany w określonym miejscu genomu
- *In situ* - bez izolacji DNA i ponownego wprowadzenia zmienionej sekwencji!
- Do komórki wprowadzany jest system modyfikujący DNA

