

Co to jest BIOINFORMATYKA?

Bioinformatyka jest interdyscyplinarną nauką zajmującą się zagadnieniami gromadzenia i przetwarzania informacji zawartej w danych uzyskiwanych w badaniach układów biologicznych prowadzonych na poziomie molekularnym. Do niedawna biologia oraz informatyka postrzegane były jako dwie całkowicie odrębne dziedziny wiedzy – biolodzy zajmowali się żywymi organizmami i mechanizmami ich funkcjonowania a informatycy – metodami przetwarzania informacji oraz układami, które te informacje przetwarzają (czyli m.in. komputerami).

Jednak na skutek rozwoju biologii molekularnej, wynalezienia nowych technik badawczych oraz wyodrębnienia się wielu specjalistycznych nauk o życiu, ilość oraz różnorodność generowanych danych biologicznych zwiększyła się na tyle, że konieczne było połączenie tych dwóch światów. Aby skutecznie analizować wzrastającą liczbę danych trzeba było m. in. zastosować zaawansowane, komputerowe metody obliczeniowe oraz wypracować sposoby przechowywania gigantycznej ilości informacji.

Co robią BIOINFORMATYCY?

- Budują i zarządzają bazami danych, w których gromadzone są informacje na temat sekwencji nukleotydowych (DNA, RNA), genomów, struktur przestrzennych makrocząsteczek, funkcji pełnionych przez geny i białka, szlaków metabolicznych, danych klinicznych, itp.
- Opracowują algorytmy do analizy sekwencji (składania fragmentów sekwencji w całość; wyszukiwania fragmentów kodujących białka, odnajdywania sekwencji regulujących ekspresję genów itp.).
- Tworzą programy komputerowe do porównywania sekwencji kwasów nukleinowych i białek a nawet całych genomów (m.in. w celu ustalenia pokrewieństwa, śledzenia ewolucji, znalezienia podobieństw dotyczących funkcji, umiejscowienia w komórce itp.).
- Konstruują komputerowe modele cząsteczek i układów cząsteczek o znaczeniu biologicznym (m. in. kwasów nukleinowych, białek, lipidów), dzięki czemu możliwe jest charakteryzowanie i przewidywanie ich struktury przestrzennej oraz funkcji pełnionej w organizmie.
- Prowadzą symulacje modeli układów biologicznych (badają *in silico* oddziaływania między cząsteczkami, szlaki metaboliczne i sygnałowe).
- Rozwijają techniki analizy obrazu (np. uzyskiwanego w wyniku tomografii, badań mikroskopowych, itd.).
- Opracowują metody wspomaganą komputerowo metaanalizy oraz eksploracji danych.

Z początku wyniki analiz bioinformatycznych były wykorzystywane jedynie przez biologów molekularnych, genetyków, ewolucjonistów i biochemików. Obecnie bioinformatyka coraz silniej wiąże się z medycyną, w tym w szczególności z tzw. medycyną precyzyjną (której celem jest dopasowanie rodzaju terapii do poszczególnego pacjenta) oraz medycyną prewencyjną.

Bioinformatycy zaangażowani są m.in. w poszukiwania czynników wirulencji, nowych szczepionek, związków bakteriobójczych, leków. Ich praca przyczynia się do lepszego diagnozowania i terapii wielu chorób. Równolegle, osiągnięcia bioinformatyki wykorzystywane są w biotechnologii, rolnictwie oraz ochronie środowiska.

Wydarzenia istotne dla rozwoju BIOINFORMATYKI

Podobnie jak w przypadku wielu innych gałęzi nauki, trudno jest wskazać precyzyjnie czas, w którym bioinformatyka jako samodzielna dyscyplina zaczęła funkcjonować w powszechnej świadomości w znaczeniu, w jakim postrzegana jest obecnie. Umownie przyjmuje się, że początki bioinformatyki przypadają na lata sześćdziesiąte XX wieku.

Lata pięćdziesiąte XX wieku

Tuż przed narodzinami bioinformatyki.

- 1950 – Per Edman opracowuje metodę sekwencjonowania peptydów
- 1952 – Martha Hershey i Alfred Chase ostatecznie potwierdzają, że to DNA, a nie białko jest nośnikiem informacji genetycznej
- 1953 – badania Jamesa Watsona, Francis Cricka, Rosalind Franklin i Maurice'a Wilkinsa doprowadzają do poznania struktury DNA
- 1955 – Frederic Sanger ustala kolejność występowania aminokwasów w insulynie – po raz pierwszy poznana zostaje sekwencja całego białka

Lata sześćdziesiąte XX wieku

Początki bioinformatyki związane są z analizą białek.

- 1962 – Margaret Deyhoff i Robert Ledley tworzą program komputerowy **COMPROTEIN** służący do składania sekwencji białek w oparciu o wyniki uzyskane z sekwencjonowania peptydów metodą Edmana
- 1966 – Margaret Deyhoff i Richard Eck publikują książkę pt. *Atlas of protein sequence and structure*. Jest to pierwsza w historii baza danych sekwencji biologicznych, póki co – analogowa
- 1966 – badania Roberta Holleya, Hara Gobinda Khorane, Heinricha Matthaei i Marshalla Nirenberga doprowadzają do złamania kodu genetycznego
- 1969 – Dennis Ritchie tworzy **C**, czyli język programowania wysokiego poziomu, który w kolejnych latach zrewolucjonizuje proces tworzenia systemów operacyjnych oraz oprogramowania użytkowego dla komputerów

Lata siedemdziesiąte XX wieku

Techniki biologii molekularnej przenoszą uwagę świata na DNA.

- 1970 – Saul Needleman i Christian Wunsch opracowują algorytm do porównywania sekwencji białek
- 1970 – Edgar Codd przedstawia relacyjny model baz danych, który w dziedzinie gromadzenia i zarządzania danymi odegra rolę podobną do języka programowania C w rozwijaniu oprogramowania komputerów
- 1972 – grupa badawcza Paula Berga tworzy pierwszy rekombinowany DNA – przy pomocy enzymów restrykcyjnych i ligazy DNA wirusa SV40 zostaje połączony z DNA faga, rozpoczyna się era klonowania
- 1973 – Robert Hahn i Winton Cerf opracowują pierwszą wersję protokołu TCP/IP służącego do przesyłania danych między komputerami połączonymi w sieć ARPANET (stworzoną dla Departamentu Obrony USA i będącą poprzedniczką Internetu)

1977 – niemal równolegle wynalezione zostają dwie różne metody sekwencjonowania DNA (metoda Maxama-Gilberta z czasem odchodzi do lamusa, metoda Fredericka Sangera używana jest do dziś)

1977 – Frederick Sanger publikuje sekwencję genomu bakteriofaga ϕ X174 (5386 pz)

1977 – na rynku pojawiają się pierwsze komputery dostosowane do użytku domowego (np. Commodore, Atari, Apple); ich system operacyjny ma wbudowany interpreter języka BASIC, który jest łatwy do opanowania dla programistów-amatorów

1978 – Margaret Dayhoff wyznacza rodzinę macierzy podstawień aminokwasowych PAM, których zastosowanie pozwala na prowadzenie ilościowych badań ewolucji molekularnej

1979 – powstaje pierwszy pakiet oprogramowania do analizy wyników sekwencjonowania metodą Sangera

Lata osiemdziesiąte i dziewięćdziesiąte XX wieku

Postępom w biologii molekularnej towarzyszy rozwój komputerów i algorytmów; upowszechnienie Internetu rewolucjonizuje komunikację, dostęp do informacji, sposób gromadzenia danych; bazy danych biologicznych oraz narzędzia do analizy danych stają się dostępne dla wszystkich.

1980 – w ramach *European Molecular Biology Laboratory* powstaje pierwsza biblioteka danych sekwencji nukleotydów (tzw. EMBL-Bank)

1985 – Kary Mullis opracowuje metodę PCR, dzięki której można namnażać DNA w próbówce, ma to m.in. wpływ na ilość danych pozyskiwanych z sekwencjonowania

1990 – rusza międzynarodowy projekt poznania sekwencji ludzkiego genomu (*Human Genome Project*)

1990 – Stephen Altschul publikuje algorytm heurystyczny BLAST (do dzisiaj rozwijany), który umożliwi szybkie wyszukiwanie sekwencji homologicznych w bazach danych sekwencji aminokwasowych i nukleotydowych

1991 – Tim Bernes-Lee udostępnia protokoły *World Wide Web* – zaczynają lawinowo powstawać rozmaite portale internetowe rewolucjonizujące dostęp społeczeństw do informacji

1995 – poznany zostaje genom pierwszego organizmu prokariotycznego – bakterii *Haemophilus influenzae* (1,8 mln pz)

1996 – pojawia się sekwencja genomu pierwszego jednokomórkowego organizmu eukariotycznego – drożdży *Saccharomyces cerevisiae* (ponad 12 mln pz)

1997 – zsekwencjonowany zostaje genom pierwszego wielokomórkowego organizmu eukariotycznego – *Caenorhabditis elegans* (ok. 100 mln pz)

Pierwsze dekady XXI wieku

Rozwój metod sekwencjonowania nowej generacji (NGS) uruchamia lawinowy wzrost liczby sekwencji nukleotydowych zdeponowanych w bazach danych, rośnie liczba poznanych struktur białkowych, rozwijają się nowe gałęzie nauk o życiu takie, jak: genomika, proteomika, transkryptomika, interaktomika, metabolomika.

2003 – ogłoszone zostaje zakończenie projektu sekwencjonowania ludzkiego genomu (w praktyce, do dzisiaj trwają prace nad uzupełnieniem brakujących danych)

2012 – w czasopiśmie *Cell* ukazuje się praca, w której za pomocą modelowania komputerowego na podstawie genotypu przewidziany zostaje fenotyp bakterii z rodzaju *Mycoplasma*.

Predyspozycje do studiowania BIOINFORMATYKI

Studia z bioinformatyki adresowane są do osób zainteresowanych najnowszymi technikami gromadzenia i przetwarzania informacji oraz mechanizmami działania organizmów żywych. Studia te powinny przypaść do gustu osobom lubiącym matematykę i programowanie oraz chemię i biologię (w szczególności działy biologii dotyczące genów i białek oraz molekularnych mechanizmów ewolucji).

Studia z BIOINFORMATYKI

Bioinformatycy muszą mieć dobre rozeznanie w wielu wybranych obszarach biologii. Na studiach uczą się więc m.in. o budowie komórki, DNA, RNA, genach, białkach, szlakach metabolicznych itp. Następnie poznają zagadnienia dotyczące sekwencji biopolimerów (kwasów nukleinowych, białek, cukrów) oraz metody pozwalające na odkrywanie, analizowanie i przetwarzanie informacji zawartych w tych sekwencjach.

Wspomniane metody przetwarzania informacji działają w oparciu o złożone algorytmy, których zrozumienie i stosowanie wymaga gruntownej znajomości matematyki i statystyki. W trakcie studiów przyszli bioinformatycy mają możliwość zapoznania się z takimi zagadnieniami jak uczenie maszynowe, sieci neuronowe czy eksploracja danych. Równocześnie, przez całe studia uczą się programowania.

Studia z bioinformatyki prowadzone są zwykle na wydziałach przyrodniczych (biologicznych, biotechnologicznych) lub na wydziałach matematyczno-informatycznych.

Jak można się domyślić, programy studiów na różnych uczelniach nie są identyczne. Spotykane rozbieżności dotyczą na przykład liczby zajęć obowiązkowych o charakterze ścisłym i biologicznym, doboru tematyki w przedmiotach specjalistycznych, możliwości rozwoju indywidualnych zainteresowań.

Typowy plan studiów z bioinformatyki zawiera kursy takie, jak:

biologia komórki

biologia molekularna

genetyka

genomika

wybrane zagadnienia fizyki i chemii

mechanizmy ewolucji

matematyka

statystyka

programowanie w różnych językach

bazy danych

BIOINFORMATYCY w akcji

Jan, 29 lat bioinformatyk, Polska

W czasie studiów poznałem metody wykorzystywane w symulacjach komputerowych i biologii obliczeniowej, nauczyłem się, jak analizować dane oraz zrozumiałem na czym polegają procesy zachodzące w układach biologicznych. Obecnie pracuję w wiodącej na rynku środkowoeuropejskim firmie bioinformatycznej i równocześnie piszę doktorat na Wydziale Biochemii, Biofizyki i Biotechnologii UJ. Moje główne zainteresowania dotyczą interakcji mikrobiomu jelitowego z nosicielem oraz nowoczesnych technologii sekwencjonowania kwasów nukleinowych (głównie metod nanoporowych).

Bożena, 28 lat data manager, Polska

Niedługo po studiach trafiłam do krakowskiej firmy bioinformatycznej, gdzie objęłam stanowisko młodszego bioinformatyka w dziale R&D. Obecnie pracuję na stanowisku data manager i zajmuję się zarządzaniem danymi genomicznymi oraz klinicznymi. W mojej codziennej pracy wykorzystuję publicznie dostępne bazy danych, w których gromadzone są wyniki sekwencjonowania nowej generacji (NGS) oraz informacje o prowadzonych badaniach klinicznych. Mam także kontakt z polskimi oraz zagranicznymi ośrodkami onkologicznymi. Dzięki wynikom mojej pracy naukowcy z firmy mogą tworzyć nowe technologie oparte o sztuczną inteligencję, których celem jest skuteczniejsza immunoterapia chorób nowotworowych oraz spersonalizowanie leczenia pacjentów. Daje mi to dużą satysfakcję i poczucie, że jestem zaangażowana w ważną sprawę, jaką jest walka z rakiem.

Rafał, 28 lat inżynier informacji naukowej, Polska

Na studiach wziąłem udział w programie stażowym Pelikan IT, w ramach którego, pracując na stanowisku młodszego chemika obliczeniowego, byłem zaangażowany w opracowanie aplikacji webowej do analizy cząsteczek pod kątem ich przydatności w medycynie. Po zakończeniu stażu, jako naukowiec, w tej samej krakowskiej firmie biotechnologicznej, współtworzyłem aplikację do analizy krzywych dawko-zależnych i równolegle zajmowałem się programem do składowania struktur i wyników eksperymentów. W kolejnych latach moje obowiązki miały coraz bardziej biznesowy charakter. Obecnie na powrót zajmuję się programowaniem. Uważam bioinformatykę za wspaniałą naukę, która dzięki szerokiemu spojrzeniu na otaczającą nas przyrodę pozwala wyciągnąć z niej mnóstwo użytecznych informacji, bez konieczności ruszania się sprzed komputera.

Gdzie można wykorzystać wiedzę i umiejętności nabyte w trakcie studiów z BIOINFORMATYKI?

Absolwenci Bioinformatyki posiadają umiejętności właściwe absolwentom studiów informatycznych: biegłość w programowaniu, tworzeniu i obsłudze baz danych, znajomość algorytmów i struktur danych, oraz wyróżniają się dobrze ugruntowaną wiedzą z zakresu nauk o życiu, w szczególności: biochemii, biologii komórki i genomiki.

Dzięki tej interdyscyplinarności znajdują zatrudnienie w wielu różnego rodzaju instytucjach.

Firmy farmaceutyczne

działy badań i rozwoju (R&D), projektowanie/ modelowanie nowych leków, metaanaliza danych w badaniach molekularnych podstaw chorób i mechanizmów działania leków

Firmy z branży *life science*

wielkoskalowa analiza danych generowanych w rozmaitych dyscyplinach nauk o życiu, metaanaliza danych, tworzenie specjalistycznych narzędzi (programów) do analizy danych biologicznych

Instytuty badawcze

stanowiska naukowe, stanowiska specjalistyczne, praca wdrożeniowo-rozwojowa, analiza danych, gromadzenie i zarządzanie danymi

Uniwersytety

stanowiska naukowe i dydaktyczne, praca badawcza, nauczanie

Branża IT

praca na stanowisku informatyka/programisty

Firmy prawnicze, firmy konsultingowe, biura patentowe, instytucje rządowe

doradztwo specjalistyczne, metaanaliza danych



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

Wydział Biochemii, Biofizyki i Biotechnologii

ul. Gronostajowa 7, 30-387 Kraków

tel.: 12 664 60 00, 12 664 54 51

e-mail: student.wbbib@uj.edu.pl

wbbib.uj.edu.pl
rekrutacja.uj.edu.pl
irk.uj.edu.pl

W przygotowanie Informatora zaangażowani byli pracownicy Wydziału Biochemii, Biofizyki i Biotechnologii Uniwersytetu Jagiellońskiego. Osobom zainteresowanym szerszym wprowadzeniem w zagadnienia z zakresu bioinformatyki polecamy dalszą lekturę, w szczególności:

- P.G. Higgs, T. K. Attwood, *Bioinformatyka i ewolucja molekularna*, PWN 2008
- J. Gauthier, A. T. Vincent, S. J. Charette, N. Derome, *A brief history of bioinformatics*, Briefings in Bioinformatics 2018: 1-16, doi: 10.1093/bib/bby063