

POPULARNONAUKOWE STRESZCZENIE PROJEKTU

Tematyka projektu dotyczy rozwijania nowych metod obliczeniowych oraz odpowiadających im narzędzi bioinformatycznych dla zastosowań w obrazowaniu molekularnym tkanek nowotworowych. Bioinformatyka jest interdyscyplinarną dziedziną nauki skupiającą się na zastosowaniach metod i narzędzi informatycznych do rozwiązywania problemów z zakresu biologii i medycyny. Obrazowanie molekularne tkanek jest nową techniką eksperymentalną, która pozwala na badanie struktury przestrzennej i składu molekularnego tkanek ze znacznie lepszą dokładnością niż wcześniejsze metody. Skrawki tkanek otrzymane z usuniętych guzów nowotworowych poddawane są utrwaleniu a na ich powierzchni planuje się siatkę punktów pomiarowych. Następnie każdy z punktów siatki staje się celem dla setek uderzeń komputerowo sterowanej wiązki laserowej. Uderzenia wiązki laserowej zogniskowane w punkty siatki powodują odparowanie materiału biologicznego tkanki i powstanie „pióropusza” zbudowanego z wyrzucanych, zjonizowanych biomolekuł. Biomolekuły kierowane są do komory próżniowej spektroskopu masowego i poruszają się w polu elektrostatycznym w kierunku detektora jonów. Dokonywany jest pomiar czasu pomiędzy momentem włączenia (impulsu) pola elektrostatycznego oraz momentem wykrycia uderzenia w detektor jonów. Pozwala to na „zważenie” molekuł wyrzucanych przez wiązkę lasera przez wykorzystanie faktu, że przyspieszenia (pojedynczo) zjonizowanych molekuł w polu elektrostatycznym są odwrotnie proporcjonalne do ich mas. Wykres zależności liczby (sygnału intensywności) wykrytych molekuł, uzyskany przez zliczenie impulsów z detektora jonów, od mas molekuł nazywa się widmem masowym. Taką technologią pomiarową określa się pochodzącym z języka angielskiego skrótem MALDI-IMS (Matrix-Assisted Laser Desorption Ionization – Imaging Mass Spectrometry).

Widma masowe uzyskane dla siatki punktów na powierzchni badanego skrawka tkanki nowotworowej są często wykorzystywane do zbudowania dwuwymiarowych obrazów (map) będących wizualizacją przestrzennej dystrybucji intensywności wybranej biomolekuły. Spektrometry MALDI-IMS nowej generacji, wykorzystując siatki 100 x 100 punktów odległych od 10 do 100 μm . Siatka o takiej rozdzielczości pozwala już na obserwację wielu szczegółów dotyczących biologii rozwoju nowotworów. Widmo masowe uzyskane dla każdego punktu siatki zawiera ponad 100 tys. różnych wartości mas oraz tyle samo odpowiadających im intensywności. Jeden eksperyment kliniczny może dotyczyć analiz dziesiątek lub setek skrawków tkanek, co ostatecznie prowadzi do tworzenia wielkich zbiorów danych, których objętość przekracza setki gigabajtów, a może sięgać nawet terabajtów. Przetwarzanie zbiorów danych o takich rozmiarach jest wyzwaniem dla bioinformatyków, zarówno w aspekcie rozwijania odpowiednich algorytmów dla przechowywania, przesyłu, wymiany oraz analizy jak też w aspekcie wspierania procesu obliczeniowego przez systemy sprzętowe o odpowiedniej mocy i wydajności. Wprawdzie ostatnimi laty obserwuje się dynamiczny rozwój algorytmów do analizy tego typu danych, jednak z powodu problemów wynikających z wielkich rozmiarów zbiorów danych oraz z ich skomplikowanej struktury, obecnie narzędzia te wykazują jeszcze szereg ograniczeń. Rozwój narzędzi dla przetwarzania danych MALDI-IMS obecnie jest na etapie rozpoznania podstawowych problemów, które muszą zostać rozwiązane aby mogły być użyteczne w wspieraniu obliczeniowym eksperymentów obrazowania molekularnego. Problemy te obejmują potrzebę opracowania odpowiednich metod kompresji danych MALDI-IMS, potrzebę znalezienia metod definiowania/określania najważniejszych cech obrazów molekularnych oraz metod ich obliczania, a także potrzebę opracowania odpowiednich testów statystycznych dla wnioskowania bazującego na danych MALDI-IMS. Narzędzia do przetwarzania danych MALDI-IMS powinny być także udostępnione jako środowiska obliczeniowe dla wygody ich używania.

Oryginalność proponowanego w projekcie podejścia polega na zaplanowaniu efektywnych metod i rozwiązań prowadzących do wyeliminowania wielu z obecnych ograniczeń algorytmów przetwarzania danych MALDI-IMS. W naszych wcześniejszych pracach rozwinęliśmy efektywne metody modelowania matematycznego widm masowych. Wykorzystanie modeli matematycznych mieszanin gaussowskich pozwoli nam na (i) skuteczne wyliczenie najważniejszych cech obrazów molekularnych, (ii) 20-30 krotną kompresję danych MALDI-IMS. Planujemy rozwinąć algorytmy grupowania oraz metody analiz statystycznych oraz stworzyć środowisko obliczeniowe pozwalające na przetwarzanie oraz na wizualizację wyników przetwarzania danych obrazowania molekularnego. Planujemy także praktyczną weryfikację opracowanych metod dla szeregu rzeczywistych zbiorów danych MALDI-IMS dla nowotworów gardła i szyi, piersi, tarczycy i prostaty.

Wspieranie technik MALDI-IMS przez rozwijanie odpowiednich algorytmów obliczeniowych przyczyni się do usprawnień w diagnostyce i leczeniu nowotworów. Rozwijanie technik obliczeniowych i metod modelowania matematycznego dla danych MALDI-IMS pozwoli przeprowadzić wszechstronne porównania pomiędzy tkankami nowotworowymi i zdrowymi, a także na dokładną charakteryzację specyficznych dla tkanki nowotworowej profili metabolicznych oraz na odwzorowanie niejednorodnych obszarów guzów nowotworowych o najprawdopodobniej różnej radiowrażliwości. To z kolei może przyczynić się do lepszego zrozumienia mechanizmów biologicznych rozwoju nowotworów, a także pozwolić na lepsze planowanie terapii zwiększając jej skuteczność i minimalizując koszty społeczne choroby nowotworowej.