

## Wybrane metody analizy danych genomicznych

dr inż. Wojciech Lesiński, Uniwersytet w Białymstoku

Analiza różnego rodzaju danych genomicznych (ekspresja genów, sekwencjonowanie RNA, Copy Number Variation, Single Nucleotide Polymorphism) jest jednym z głównych zadań bioinformatyki. Często celem takiej analizy jest identyfikowanie genów powiązanych z danym zjawiskiem. Cechą charakterystyczną takich danych jest olbrzymi rozmiar wektora cech (kilka, kilkadziesiąt, a nawet kilkaset cech tysięcy cech opisujących daną obserwację) oraz niewielka (przeważnie) liczba obserwacji. W takich danych przeważająca liczba cech nie niesie ze sobą informacji, które nas interesują. Z tego powodu konieczne jest prowadzenie dokładnej selekcji cech. Kolejną właściwością jest znaczna wrażliwość danych na efekty próby, co wymusza normalizację. W ramach seminarium przedstawione zostaną wybrane metody analizy tego typu danych na przykładzie odpowiedzi immunologicznej kurcząt, sekwencjonowanie RNA w klasyfikacji neuroblastomy oraz ekspresji genów dla nowotworu piersi oraz prostaty.