

## **MLGenSig (Machine Learning Genetic Signatures) Metody uczenia maszynowego w budowie zintegrowanych sygnatur genetycznych**

Celem naukowym projektu jest opracowanie metodologii budowy zintegrowanych sygnatur genetycznych w oparciu o heterogeniczne źródła danych dotyczące jednocześnie profili opartych o sekwencje DNA / ekspresje mRNA / ekspresje białek i profili metylacji DNA. Jest to bardzo ważny problem badawczy. W wielu analizach bioinformatycznych zachodzi potrzeba albo wskazania zbioru cech, które charakteryzują określony stan komórki, albo zbudowania metody oceniającej czy komórka jest w określonym stanie.

Jednym z zastosowań jest molekularne profilowanie komórki pod kątem reakcji na określony rodzaj leczenia. Jeżeli mamy dwie grupy próbek, jedna zawierająca komórki odporne na np. chemioterapię, druga podatna na leczenie, to celem sygnatury genetycznej jest charakteryzacja cech genetycznych komórek podatnych na leczenie. Zastosowań takiej metodologii jest wiele, wymienić można przykładowo charakterystyka czy komórki nowotworowe nabyły cech macierzystych komórek nowotworowych, charakterystyka czy komórki nowotworowe nabyły cech molekularnych guzów o gorszych rokowaniach.

Dostępność nowych, wysokoprzepustowych metod badania procesów biologicznych, pozwala na identyfikacje coraz subtelniejszych cech genetycznych komórki. Coraz więcej badań patrzy jednocześnie na procesy zachodzące w komórce z różnych perspektyw czy to poprzez pomiary DNA-seq (w tym mutacje typu SNP czy copy number variation), mRNA-seq (w tym ekspresja genów lub alternatywny splicing), czy profile metylacji lub profile ekspresji białek (mikromacierze RRPA). Dostępność metod genotypowania doprowadziła szybko do powstania metodologii budowy sygnatur genetycznych w oparciu o dane z RNA-seq, DNA-seq, czy ekspresji białek. Brak jednak zintegrowanego podejścia, które uwzględniałoby wzajemne zależności pomiędzy DNA, transkryptami i białkami. Obecnie budowane sygnatury analizują każdą z cech biologicznych niezależnie.