

Obecnie miliony zestawów danych pochodzących z analiz transkryptomicznych (globalny pomiar ekspresji genów) pozostają zdeponowane w ogólnodostępnych bazach danych. Zestawy te, wytworzone celem uzyskania odpowiedzi na konkretne pytania dotyczące poziomu ekspresji genów, są zasobne w informacje, które nie zostały dotychczas wykorzystane w badaniach naukowych. Umiejętne przetworzenie tych zestawów danych pozwoli na odkrycie pozostałych, niewykorzystanych zasobów informacji, które mogą zostać dalej wykorzystane w stawianiu nowych hipotez badawczych, bez konieczności prowadzenia kolejnych kosztownych eksperymentów.

W niniejszym projekcie proponuje się zastosowanie analizy ICA (składowych niezależnych) oraz następujące po niej wyróżnienie iModulonów (niezależnie modulowanych zestawów genów) celem wydobywania ze zdeponowanych danych transkryptomicznych niewykorzystanych informacji. Strategia ta wymaga wykorzystania technik modelowania matematycznego oraz uczenia maszynowego, co zwiększy efektywność i potencjał odkrywczy takiego podejścia. W ten sposób będzie możliwa analiza i synteza setek zestawów danych, z których każdy zawiera informacje dotyczące tysięcy genów w sposób spójny, systematyczny i skoordynowany.

Planuje się przeprowadzenie analizy ICA-iModulon na zestawach danych transkryptomicznych dla gatunku drożdży *Yarrowia lipolytica*. Drożdże te stanowią wartościowy obiekt badań naukowych, oraz organizm wykorzystywany w praktyce przemysłowej (m.in. do produkcji pasz, substancji słodzącej, oraz rekombinowanych białek). Oprócz danych zdeponowanych w ogólnodostępnych bazach planuje się wytworzyć dodatkowe zestawy danych transkryptomicznych, aby wzmocnić zestawy wejściowe i możliwość wytypowania iModulonów indukowanych w konkretnych warunkach. Jest to ważny element tego projektu, ponieważ planuje się następnie wykorzystać wyniki analizy ICA-iModulon do opracowania syntetycznego promotora (elementu regulującego ekspresję genów) do wykorzystania przy kontrolowanej produkcji białek rekombinowanych. Białka rekombinowane stanowią wysoko cenny produkt biotechnologiczny wykorzystywany przy produkcji przemysłowych preparatów enzymatycznych, leków (testów diagnostycznych czy szczepionek), środków czystości, kosmetyków. Indukowalne elementy regulatorowe (wspomniane syntetyczne promotory) umożliwiające kontrolowaną produkcję białek rekombinowanych są wysoce pożądane w badaniach naukowych i praktyce przemysłowej.

Stąd realizacja niniejszego projektu prowadzi zarówno do zdobycia nowej wiedzy (charakterystyki gatunku drożdży wykorzystywanych w badaniach i przemyśle), ale także pozwoli na praktyczne wykorzystanie tej wiedzy.