

Silesian University of Technology
Faculty of Automatic Control, Electronics
and Computer Science



Silesian
University
of Technology

**Skipping batch effect correction:
clustering-based methods for analyzing
confounded single-cell RNA-sequencing data**

PhD Thesis

Author: **Tomasz Kujawa**
Supervisor: **prof. dr hab. inż. Joanna Polańska**
Co-supervisor: **dr inż. Michał Marczyk**

Gliwice, June 2023

STRESZCZENIE

Efekt paczki jest nieuniknionym zjawiskiem w przypadku wysokoprzepustowych i wielkoskalowych eksperymentów, gdzie ograniczenia logistyczne wymagają generowania danych w różnym czasie i przy zaangażowaniu wielu laboratoriów, często wyposażonych w odmienne platformy sprzętowe, wykorzystujących różne partie odczynników i przy udziale zróżnicowanego personelu badawczego.

Wspólna analiza takich danych jest niewykonalna, ponieważ efekty paczki przesłaniają badaną zmienność biologiczną. Takie dane należy skorygować, aby konkluzje wyciągnięte z ich analizy były wiarygodne. Niestety sam proces korekty wiąże się z kilkoma negatywnymi konsekwencjami. Korekta zniekształca bowiem pierwotną naturę oraz dystrybucję danych. Ponadto brakuje miary do ilościowego szacowania niepewności tego procesu. Co więcej, korekta z wykorzystaniem narzędzi bioinformatycznych jest praktycznie niemożliwa w przypadkach, gdzie badana zmienna biologiczna jest całkowicie skorelowana ze zmienną techniczną.

W niniejszej pracy zaproponowano podejście umożliwiające skonsolidowaną analizę zestawów danych pochodzących z eksperymentów scRNAseq i prezentujących silny efekt paczki. Proponowane rozwiązanie opiera się na iteracyjnej metodzie grupowania z selekcją cech połączonej z analizą funkcjonalną zestawów genów. Po analizie funkcjonalnej otrzymane klastry są łączone między paczkami na podstawie ich funkcjonalnego podobieństwa. Celem takiego podejścia jest złagodzenie negatywnego wpływu efektu paczki bez konieczności jego korekty i związanymi z tym konsekwencjami.