Automated machine learning for breast cancer methylation biomarker discovery: a data-driven approach to be implemented in liquid biopsy

Abstract Submitter: Maria Panagopoulou Pantazi, Greece*

Co-Authors: Maria Papadaki, Makrina Karaglani, Theodosis Theodosiou, Kleita Michaelidou, Ioannis Tsamardinos, Stylianos Kakolyris, Sofia Agelaki, Ekaterini Chatzaki

*Laboratory of Pharmacology, Medical School, Democritus University of Thrace

Abstract

Background: Despite significant progress, breast cancer remains a devastating disease needing novel approaches for effective clinical management. Recently, increasing attention has been given to epigenetic biomarkers assessed in cell-free DNA (cfDNA) as potential tools for clinical applications.

Objectives: This study introduces a three-step, Automated Machine Learning (AutoML)-based, data-driven biomarker discovery pipeline, utilizing genome-wide DNA methylomes and laboratory validation in patient cfDNA. The goal is to identify robust biomarkers and generate high-performance methylation biosignatures of clinically significant end-points to be implemented as liquid biopsy diagnostics.

Methods: Publicly available methylomes were analyzed using AutoML to compare malignant and benign breast tissue and identify breast cancer-specific methylation biomarkers. Next, the methylation status of the identified biomarkers was evaluated in cfDNA from 195 breast cancer patients and 135 healthy individuals by Methylation Specific qPCR (qMSP). Finally, AutoML was applied to experimental and clinical data to develop optimized biosignatures for significant end-points.

Results: AutoML identified three novel breast cancer-specific methylation biomarkers in the promoters of CLDN15, MRGPRD and ZNF430. Bioinformatic literature mining demonstrated limited knowledge in regards to their implication with cancer biology. Laboratory validation confirmed elevated methylation levels in cancer patients as compared to healthy individuals for all three promoters. Furthermore, promoter methylation correlated with poor prognostic and predictive parameters. Using AutoML on experimental laboratory methylation measurements and patients' clinical data, we developed four biosignatures: a diagnostic biosignature distinguishing breast cancer from healthy individuals (AUC 0.798), a classification biosignature differentiating breast cancer subtypes (AUC 0.678), a prognostic biosignature predicting relapse risk (AUC of 0.785), and a predictive biosignature assessing treatment response in metastatic patients (AUC 0.861).

Conclusion: Our data-driven pipeline successfully identified 3 novel breast cancer specific methylation biomarkers. Their implication with pathology needs further attention. Moreover, we built biosignatures demonstrating strong predictive performance against clinically significant end-points. The low number of features and the minimally invasive nature of liquid biopsy highlight the potential for clinical implementation. This AutoML-based approach can be broadly applied across various cancer settings to provide readily available diagnostic solutions.

Do you have any conflicts of interest?

Yes, I have a conflict of interest.

Conflict of interest: Th. Theodosiou and E. Chatzaki are co-founders of ABCureD PC.