A novel approach for predicting breast cancer metastasis through measurement of global transcriptional dysregulation

Tanzira Najnin*
tanzira.najnin@utsa.edu
Department of Computer Science
The University of Texas at San Antonio
Texas, USA

Sakhawat Hossain Saimon*
Department of Computer Science
The University of Texas at San Antonio
Texas, USA

Maryam Zand*
Department of Psychiatry, University of California San Diego
San Diego, USA

Nahim Adnan
Department of Computer Science
The University of Texas at San Antonio
Texas, USA

Zhijie Liu
Department of Molecular Medicine, UT Health San Antonio
Texas, USA

Tim Hui-Ming Huang
Department of Molecular Medicine, UT Health San Antonio
Texas, USA

Jianhua Ruan
jianhua.ruan@utsa.edu
Department of Computer Science
The University of Texas at San Antonio
Texas, USA

ABSTRACT

Understanding why some breast cancers become metastatic and predicting the metastatic risk at an earlier cancer stage are two major goals of breast cancer research. While improvements in one goal are expected to benefit the other, this synergy is rarely achieved. On one hand, studies focusing on knowledge discovery have helped identify key biological pathways associated with metastasis, but a holistic understanding of how they interplay remains elusive, and using these pathways as metastasis-predicting features resulted in unsatisfactory accuracy. On the other hand, complex machine learning models can achieve high prediction accuracy, but the predictions are challenging to explain in biological terms. Here, we propose a novel computational approach that helps achieve both goals simultaneously. To support knowledge discovery, we construct and compare gene regulatory networks (GRNs) to identify metastasis-associated regulatory changes. To enable explainable metastasis prediction, we introduce a simple metric called the dysregulation score, which relies on the reconstructed GRNs and allows for direct metastasis prediction. Experimental results demonstrate that our method not only identified significant GRN changes for known metastasis-related genes and transcription factors but also unveiled valuable systems-level insights, including the loss of co-regulation between key biological processes. Importantly, leveraging the dysregulation score, our model-free classifier outperformed several complex machine learning models, based on rigorous within- and cross-dataset evaluation procedures and multiple performance metrics. In summary, this work bridges a gap between systems-level knowledge discovery and accurate yet explainable phenotype prediction by employing carefully designed knowledge representations and knowledge-based prediction models.

*All three authors contributed equally to this research.

This work is licensed under a Creative Commons Attribution International 4.0 License.

BIOKDD '24, August 26, 2024, Barcelona, Spain
© 2024 Copyright held by the owner/author(s).