An Integrative Machine Learning Approach in Detecting Gene Networks Predicting Liver Cancer Incidence and Survival

Stephan Komladzei[1], Murshalina Akhter[1], Md Tamzid Islam[1], Mihaela Sardiu[1,2], Yanming Li[1,2]
[1]University of Kansas Medical Center, Kansas City, KS, Department of Biostatistics & Data Science
[2]University of Kansas Cancer Center, Kansas City, KS

Received Aug. 21, 2024; Accepted for publication Aug. 26, 2024; Published online Aug. 27, 2024
https://doi.org/10.17161/kjm.vol17.22665

Introduction. Liver cancer (LC) is one of the most common types of cancer and the third largest cause of cancer-related deaths globally. Although there is no known cause of LC, the disease has some associated risk factors. Studies have also linked genetic mutations to LC. Patients who receive early diagnosis have a five-year relative survival rate of ~36%, whereas ~3% for late diagnoses.

Methods. In this study, we employed a novel machine learning technique called netLDA which initially identified predictive strong DE genes and used as them hubs genes to identify coregulating genes that may only exhibit marginally weak DE effects. The detected genes form predictive gene networks (PGNs), which are used to make predictions.

Results. Our study detected previously identified LC occurrence susceptibility strong genes; *GPC3, STMN1* and *TCF19.* Potential biomarker genes for LC occurrence such as *MT1E* and *PLVAP* were detected as co-regulating weak genes. In our one-year survival study, *SPP1* and *ADH4* were detected as strong genes whilst *MFAP4* was identified as novel co-regulating weak gene. The PGNs achieved an AUC of about 1.00 in our case-control study and 0.77 in our one-year survival study. Some of the detected genes overlapped with some significant LC related pathways to provide additional layers of biological significance.

Conclusions. With our study, the identified predictive genes can propel studies into understanding the underlying mechanisms for LC development and prognosis. These identified biomarker genes can be used as new therapeutic targets by industries in the development of new drugs for LC.

The authors declare no conflict of interest.
The research reported in this publication was supported by the University of Kansas pilot project. The high-performance computing resources used in this study was supported by the K-INBRE Bioinformatics Core, which is supported in part by the National Institute of General Medical Science award (P20 GM103418), the Biostatistics and Informatics Shared Resource, supported by the National Cancer Institute Cancer Center Support Grant (P30 CA168524), and the Kansas Institute for Precision Medicine COBRE, supported by the National Institute of General Medical Science award (P20 GM130423).