PREDICTING GENE DEPENDENCIES OF CANCER FROM INTEGRATED GENOMIC PROFILES BY DEEP LEARNING

Yu-Chiao Chiu[1], Siyuan Zheng[1,2], Manjeet K. Rao[1,3], Yufei Huang[2,4*], Yidong Chen[1,2*]
[1]Greehey Children's Cancer Research Institute, [2]Department of Epidemiology and Biostatistics, [3]Department of Cell Systems and Anatomy, UT Health San Antonio, [4]Department of Electrical and Computer Engineering, UT San Antonio

Cancer dependency, the study of essential genes for cancer cell proliferation and survival, is a crucial component of the development of novel therapies of cancer. Recently, the DepMap project utilized CRISPR-Cas9 loss-of-function screens to construct a genome-wide landscape of gene essentialities of cancer cell lines (CCLs). However, the need to utilize the cumulated genomic data to accurately predict gene dependencies for unscreened CCLs and translate the findings to tumors is largely unmet. Taking advantage of the emerging technology of deep learning, here we propose a model that learns data embeddings of complex genomics profiles that lead to predictions of gene dependencies. The model predicts a dependency score of a gene of interest in a CCL or tumor given its genomics profiles. It contains three sub-networks: i) an encoder network to reduce the dimension of each genomics data, ii) an encoder network for embedding the fingerprints of a gene dependency of interest, and iii) a prediction network that transforms outputs of networks i and ii into a dependency score. We trained the model using the dependency profiles of CCLs collected from DepMap. Our method achieved an accurate prediction (correlation coefficients between original and predicted dependency scores = 0.89 and P ~0) and outperformed conventional machine learning methods. In order to translate the findings to real tumors, we applied the model to generate a pan-cancer dependency map of ~8,000 tumors of which genomics data were profiled by The Cancer Genome Atlas. Our data suggested that gene dependencies are dominated by individual or multiple genomics and are not cancer type dependent. The finding is in line with the premise of precision oncology that the genomics may play a more critical role in determining the tumor's response to treatments than the cancer type. Overall, the proposed deep learning method is capable of integrating genomics profiles and predicting gene dependencies for CCLs and real tumors. We expect the model to evolve with rapidly developing dependency screens and facilitate the prioritization of therapeutic targets of cancers.

Funding: NCI Cancer Center Shared Resources, NIH, CPRIT, SALSI Innovation Challenge Award, and SALSI Postdoctoral Research Fellowship.