Discovering novel bispecific antibody targets for limited stage small cell lung cancer using machine learning at the single-cell RNA sequencing level

Gyo-Jin Choi[1], Sang-Beom Bang[1], Jieun Park[2], Young Kee Shin[1,3,*]

[1] *Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, Seoul National University*
[2] *Research Institute of Pharmaceutical Science, Department of Pharmacy, Seoul National University*
[3] *College of Pharmacy, Seoul National University*
*Corresponding author: ykeeshin@snu.ac.kr

Despite numerous clinical trials, chemotherapy alone is still used as first-line therapy because no targeted therapy has been approved for limited-stage small cell lung cancer (LS-SCLC). Therefore, the purpose of the study was to identify novel bispecific antibody targets with significant expression differences between LS-SCLC and normal lung cells using machine learning methods on single-cell RNA sequencing data. Three normal lung samples from the GSE196638 dataset and five LS-SCLC samples from the HTAN-MSK dataset were analyzed, with a total cell count of approximately 38,000 cells. The genes expressed on the cell surface were ranked using random forest, a machine learning technique. And then we excluded genes with a higher average gene expression in normal lung cells than in tumor cells. After selecting the top 50 genes, we generated all possible bispecific combinations and ranked them using convolutional neural networks, a deep learning technique. As a result, we found several combinations where the percentage of tumor cells expressing both candidate target genes was high, while the percentage of normal lung cells was low. We also identified several combinations where the percentage of tumor cells expressing at least one target gene was high, while the percentage of normal lung cells was low. Based on these findings, we are able to propose several bispecific antibody combinations that could target LS-SCLC much more specifically than a single target antibody while reducing side effects on normal lung cells. In the near future, we will examine the protein expression level of the candidate genes in the SCLC cell line.