BioVision: A Novel in silico Drug Discovery Pipeline to Predict Drug Candidates for Non-Small Cell Lung Cancer

McKenna, Patrick (School: College Park High School)

Globally, cancer is the leading cause of death, with 1.9 million new diagnoses and 600,000 deaths annually attributed to lung cancer alone. Non-small cell lung cancer (NSCLC) represents approximately 85% of all lung cancer cases, making it a predominant and deadly form of cancer worldwide. Currently, there are no clinically approved drugs successful in treating high-risk NSCLC due to cancerous side effects and transcription factors that are undruggable. Furthermore, current drug design methods are antiquated; they are not only costly and time-consuming but also lack general protein inhibition capabilities and efficient testing protocols for drug candidates. This study introduces a novel, robust, and production-ready in silico deep learning pipeline for de novo drug design, employing an ensemble of machine learning techniques specifically targeting NSCLC. The pipeline encompasses: (1) identification of overexpressed genes through graph neural networks; (2) construction of an accurate protein model; (3) large-scale drug screening; (4) development and optimization of drug-like compounds post cryptic pocket discovery; and (5) validation of compounds using molecular simulations and QSAR modeling. This method significantly streamlines the drug screening process, reducing time and costs by upwards of 90% and 80%, respectively. The results confirm the efficacy of this approach, positioning this work as a potent tool for efficient oncological therapeutic discovery and screening.