[BIOINFO 2025]

Genotype-to-Drug Diffusion for Tailored Anti-cancer Hit-like Small Molecules

Hyunho Kim[1,2], and Hojung Nam[1,*]

[1] Department of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Buk-gu, Gwangju, 61005, Republic of Korea

[2] Division of Advanced Predictive Research, Korea Institute of Toxicology (KIT), 141 Gajeong-ro, Yuseong-gu, Daejeon, 34114, Republic of Korea

*Corresponding author: hjnam@gist.ac.kr

Developing effective cancer therapeutics remains a challenge because of tumor heterogeneity and the lack of clearly defined drug targets. To address this, we introduce Genotype-to-Drug Diffusion (G2D-Diff), a generative framework designed to create anti-cancer small molecules conditioned on tumor genotypes and drug response profiles. Unlike conventional generative approaches, G2D-Diff learns the conditional distribution directly from large-scale drug response data, which enables the generation of diverse and drug-like compounds aligned with specific response outcomes. The framework combines a chemical variational autoencoder with a conditional latent diffusion model, supported by a contrastive pre-trained condition encoder that improves generalization to unseen genotypes. In our evaluations, G2D-Diff consistently outperformed existing methods in molecular diversity, feasibility, and condition fitness. Notably, the model offers interpretability through its attention mechanism by providing genotype-response specific potential cancer targets and pathways. When applied to triple-negative breast cancer in case studies, the model created promising hit-like molecules by highlighting condition-specific relevant genes and pathways. Taken together, G2D-Diff can accelerate early-stage cancer drug discovery by suggesting genotype-guided de novo anti-cancer molecules, toward personalized anti-cancer therapeutics.