**PaccMannRL: Designing anticancer drugs from transcriptomic data via reinforcement learning**

Jannis Born\*Matteo Manica\*Ali Oskooei\*Joris CadowMaría Rodríguez Martínez{jab,tte,osk,dow,mrm}@zurich.ibm.com

IBM Research, 4 Säumerstrasse, Rüschlikon, Switzerland 8803

This research contributes towards enhancing the efficacy of candidate anticancer drug compounds by ensuring their desired chemical properties are specifically tailored to the specific cellular environment or gene expression profile of a tumor cell that the compounds intend to act. The research presents a deep RL-based model for anticancer molecule generation that, for the first time, enables the generation of novel anticancer compounds while taking into account the disease context, which is encoded in the specific gene expression profile (GEP) of a tumor cell. The research demonstrates the possibility of optimizing, using RL optimizing framework, the proposed anti-cancer compound generative model to produce candidate compounds with high predicted efficacy (IC50-the micromolar concentration necessary to inhibit 50% of the cells in a sample) against a given target GEP. The proposed model is reported to have a capacity of proposing novel molecular structures from a chemical space of about 1030 to 1060 molecules. The research successfully validated an estimated 96.2 percent of the generated molecules. The research model’s input are: RNA-Sequence gene expression profiles from cancerous and healthy human tissues (transcriptomic data-https://www.ncbi.nlm.nih.gov/pubmed/20972753), test and train molecular datasets from credible databases, IC50 drug sensitivity data, as well as bioactive drug-like molecular data. On the other hand, the model’s outputs include generated molecular structures, Morgan fingerprints (ECFP) of the generated molecules, and a candidate compound. The most crucial elements of the presented research model is a conditional generator framework, which entails a bio molecular profile VAE (variational autonencoder) (PVAE) and a sequential compound generator VAE (SVAE), an RL optimization framework, Tanimoto similarity model, and a multimodal IC50 prediction model (PaccMann). This research significantly relates to the overall mission of Emergent Dynamics, Inc., as it develops efficacy curative solutions through deep generative models for Biomedicine. An optimized generative model can design effective drugs closely following cell tissue adaptive sequences and bioactive small molecular structural patterns. The research demonstrates the development of an effective anticancer drug design model through reinforcement learning.

https://pdfs.semanticscholar.org/0b9b/36c2ebe192d5c290f9f2f2aa9466f33b9357.pdf