04-3

Graph Convolutional Networks for Ligand-based Virtual Screening against the Androgen Receptor

Romeo Cozac[1] romeo.cozac@elix-inc.com
Nazim Medzhidov[1] nazim.medzhidov@elix-inc.com
Casey Galvin[1] casey.galvin@elix-inc.com
Shinya Yuki[1] shinya.yuki@elix-in.com

[1] Elix Inc., Daini Togo Park Building 3F, 8-34 Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan

Keywords: Androgen Receptor, Machine learning, Graph convolutional Networks, drug repurposing

Androgen receptor (AR) is a ligand-dependent transcription factor that belongs to the family of steroid hormone nuclear receptors. Androgens bind to the ligand binding domain of the AR with strong affinity and are capable of regulating transcription of AR-regulated genes. AR signaling has implications in pancreatic cancer as well as tumors in the lungs, kidney, liver, and bladder. The standard treatment approach for patients with prostate cancer is to lower testosterone levels in the body, however, this does not always prove effective since some patients do not respond to this form of treatment. Therefore alternative treatment options are necessary. Small molecule antagonists that interfere with androgens binding to AR have been under active investigation. In this study we utilize a graph based Machine Learning model to identify small molecule AR antagonists. In our approach, we design a flexible architecture that supports different graph convolutional layers. We used Bayesian optimization to find the best-performing graph kernel and hyperparameters, and we applied MC Dropout to measure the variance and confidence of the predicted values. The trained model was used to screen three datasets of commercially available compounds: the ZINC dataset, the eMolecules dataset, and a list of FDA-approved drugs. Using a high confidence threshold for the predicted activity and confidence, we reduced the original set of 364K molecules to 55 hits which were not present in our training set. Following a patent search on the 55 hits, we noticed that 36% had at least one relevant patent describing high activity against the Androgen Receptor, proving that graph-based predictive models can be efficient tools for virtual screening.

[1]Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, et al. Conformations of immunoglobulin hypervariable regions. Nature. 1989 Dec 21-28;342(6252):877-83. doi: 10.1038/342877a0. PMID: 2687698.

[2]Jeffrey A Ruffolo, Carlos Guerra, Sai Pooja Mahajan, Jeremias Sulam, Jeffrey J Gray, Geometric potentials from deep learning improve prediction of CDR H3 loop structures, Bioinformatics, Volume 36, Issue Supplement_1, July 2020, Pages i268–i275

[3] Alexander Rives, Siddharth Goyal, Joshua Meier, Demi Guo, Myle Ott, C. LawrenceZitnick, Jerry Ma, Rob Fergus, Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences, *Proceedings of the National Academy of Sciences* doi: 10.1073/pnas.2016239118