5α-reductase inhibitor classification model development based on machine learning and deep learning algorithms

Hyun Kil Shin(hyunkil.shin@kitox.re.kr)[1], Yong Oh Lee[2], Young Jun Kim[2]

[1]Korea Institute of Toxicology, [2]KIST Europe

Abstract

Purpose: Steroidal 5α-reductase is a nuclear membrane-bound enzyme converting endogenous testosterone (T) to dihydrotestosterone (DHT) in the presence of cofactor NADPH [1]. DHT provokes benign prostatic hyperplasia and rapid prostate enlargement in later adulthood by stimulating cell growth in the tissue and modulating reproductive endpoints [2]. Due to the significance of 5α-reductase in controlling the concentration of DHT in cellular level in terms of reproductive toxicity, classification models for 5α-reductase inhibitors were developed based on machine learning and deep learning algorithms to screen potential endocrine disruptors.

Methods: In this study, Quantitative Structure-Activity Relationship (QSAR) models were developed for predicting inhibitors of 5α-reductase. In machine learning (ML) model, MACCS fingerprint[3] was used with a feature selection algorithm to subset significant molecular features for 5 α-reductase inhibition. For deep learning model, convolutional neural network (CNN) architecture was implemented with SMILES feature matrix [4]. Data set for model development was collected from ChEMBL in which chemical structures and their assay values were extracted from the literatures (https://www.ebi.ac.uk/chembl/). After data curation, 397 chemicals were secured with sufficient quality for classification model development. 100nM of IC50 value was used as the criterion; hence, chemicals with lower than 100nM of IC50 were labeled as an inhibitor, and others as non-inhibitor.

Results: Our internal validation results showed that the predictive accuracies were 82% (ML) and 93% (CNN), respectively which seems pertinent predictive performance in screening 5α-reductase inhibitors as reproductive toxicants.

Keywords: *QSAR, machine learning, 5-alpha reductase inhibitor*

References

[1] Jarman, Michael, H. John Smith, Paul J. Nicholls, and Claire Simons. Inhibitors of enzymes of androgen biosynthesis: cytochrome P450 17α and 5 α-steroid reductase. Natural product reports 15, no. 5 (1998): 495-512.

[2] Martyniuk, Christopher J., Sonja Bissegger, and Valérie S. Langlois. Reprint of "Current perspectives on the androgen 5 alpha-dihydrotestosterone (DHT) and 5 alpha-reductases in teleost fishes and amphibians". General and comparative endocrinology 203 (2014): 10-20.

[3] Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Garcia-Vallvé, S., & Pujadas, G. (2015). Molecular fingerprint similarity search in virtual screening. Methods, 71, 58-63.

[4] Hirohara, Maya, Yutaka Saito, Yuki Koda, Kengo Sato, and Yasubumi Sakakibara. Convolutional neural network based on SMILES representation of compounds for detecting chemical motif. BMC bioinformatics 19, no. 19 (2018): 526.

Biography

Hyun Kil Shin, Ph.D in Biotechnology, is a researcher at KIT. Main research topic is development of prediction models for toxicity.

Yong Oh Lee, Ph.D in Computer Engineering, is a senior researcher at KIST Europe. His research interest is deep learning and its application to various engineering fields.

Young Jun Kim, Ph.D in Biophysics is a senior researcher at KIST Europe. His research interest is eco-toxicity screening tools in environmental organism for environmental chemical assessments.