in silico prediction of drug-induced phospholipidosis

Hao Wang

Institute of Chemistry and Biochemistry, Freie Universität, Berlin, Germany;
Computational Systems Biology, Humboldt-Universität zu Berlin, Berlin, Germany (associate)

Drug-induced phospholipidosis (PLD), a side effect of drugs, is a phospholipid storage disorder characterized by the excess accumulation of PLD and the inducing drug in the lysosomes of the affected tissues [1]. Such side effects must be considered as early as possible in the drug discovery process. An *in silico* prediction model is a low-cost method and can effectively accelerate drug development, in particular in an early stage. Currently, numerous *in silico* models [2] have been designed to predict PLD. However, the performance of any *in silico* prediction model can not be better than the quality of the experimental data set, which is used for machine learning. In this study, we employ a new *in silico* method, DemFeature, to predict PLD. DemFeature is a supervised machine learning model, which is developed from DemPred [3]. The advantage of DemFeature is to construct an individual training subset for each considered test molecule. This method ignores in the training phase molecules of the learning set that are not sufficiently similar or dissimilar to the molecule to be classified. Based on this strategy, ideally, the considered molecule can be more accurately interpreted based on the chemical and biological information of known molecules. The Goracci dataset [2] is used for training and validating because it is the most reliable composed by comparing 7 PLD datasets. Finally, we compare our prediction results with other models used for predicting PLD based on the same dataset. Our model shows the best prediction performance.

References

[1] Reasor, M. J.; Hastings, K. L.; R.G. Expert Opin. *Drug Saf.* 2006, 5, 567-83.
[2] Goracci, L.; Ceccarelli, M.; Bonelli, D and Cruciani G. *J. Chem. Inf. Model.*, 2013, 53 (6), 1436–1446.
[3] Kavuk, O.D.; Kamada, M.; Akutsu, T and Knapp, E.W., BMC Bioinformatics 2011,12 412-421.