Identification of Compounds Interfering with HTS Assay Technology[1]

David, L.* [1,2], Walsh, J. [3], Sturm N. [4], Feierberg I. [5], Nissink J.W.M. [6], Chen H. [2], Bajorath J. [7], Engkvist O. [2]

* Lead presenter

[1] laurianne.david@astrazeneca.com

[2] *Hit Discovery, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca Gothenburg, Pepparedsleden 1, 43183, Mölndal, Sweden*

[3] *Hit Discovery, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca Macclesfield, Alderley Park, Macclesfield, SK10 4TG, Cambridge, UK*

[4] *Data Science and AI, Drug Safety & Metabolism, BioPharmaceuticals R&D, AstraZeneca Gothenburg, Pepparedsleden 1, 43183, Mölndal, Sweden*

[5] *Hit Discovery, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca Boston, 35 Gatehouse Drive, Waltham, Massachusetts 02451, US*

[6] *Computational Chemistry, Oncology R&D, AstraZeneca Cambridge, Cambridge Science Park, Milton Road, Cambridge CB4 0WG, UK*

[7] *Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Endenicher Allee 19c, 53115, Bonn, Germany*

A major challenge in High-Throughput Screening (HTS) studies is given by the presence of frequent-hitters[2], which include non-selective compounds and molecules with false positive signals in many screens. HTS triaging aims at the identification and the eventual removal of such compounds. False positive compounds can be **C**ompounds that **I**nterfere with an **A**ssay **T**echnology (CIATs). CIATs can impede research and waste resources by being investigated in follow-up studies. An existing approach for CIATs identification is BSF[3], a structure-independent score based on historical HTS data. Furthermore, different substructure filters are available to identify CIATs, the most used nowadays being the Pan-Assay Interference Compounds filter (PAINS)[4]. This filter is a set of 480 substructures that were derived from a dataset based on a specific assay technology.

In our study, we developed a machine-learning model to predict CIATs in three assay technologies. Experimentally validated CIATs and non-CIATs were collected from the AstraZeneca in-house HTS collection. Compounds were represented by their ECFP4 fingerprints and were used to train the model. Our model provided successful prediction of CIATs for existing and novel compounds and a complementary and wider set of predicted CIATs compared to BSF and to the PAINS substructural filters. Our analysis showed the importance of the applicability domain of filtering rules in HTS triaging.

Bibliography:

[1] David L.; Walsh J. et al. ChemMedChem, 2019, doi:10.1002/cmdc.201900395
[2] Roche O.; Schneider P. et al. J. Med. Chem., 2002, 45, 137-142.
[3] Nissink J.W.M.; Blackburn S. Future Med. Chem., 2014, 6, 1113-1126.
[4] Baell J.B.; Holloway G.A. J. Med. Chem., 2010, 53, 2719-2740.