Hit Dexter: A Machine-learning Model for the Prediction of Frequent Hitters

Conrad Stork,[1] Johannes Wagner,[1] Nils-Ole Friedrich,[1] Christina de Bruyn Kops,[1] Martin Šícho,[1,2] Johannes Kirchmair[1]*

[1] Universität Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Department of Computer Science, Center for Bioinformatics, Hamburg, 20146, Germany

[2] CZ-OPENSCREEN: National Infrastructure for Chemical Biology, Laboratory of Informatics and Chemistry, Faculty of Chemical Technology, University of Chemistry and Technology Prague, 166 28 Prague 6, Czech Republic

**J. Kirchmair. E-mail: kirchmair@zbh.uni-hamburg.de. Tel.: +49 (0)40 42838 7303.*

High-throughput screening is a key technology in early drug design that enables the screening of tens of thousands of compounds per day. [1,2] However, false-positive signals triggered by badly behaving compounds (frequent hitters, pan-assay interference compounds, aggregators and others) continue to pose a major challenge in early drug discovery and still lead to a substantial number of false hits reported in the scientific literature. [3] Few computational approaches that allow the identification of badly behaving compounds exist, and the applicability of these existing methods is limited. We present Hit Dexter, a web service that allows the identification of frequent hitters with high accuracy. [4] Hit Dexter is based on two extremely randomized tree classifiers trained on a well-prepared subset of the PubChem Bioassay database containing 311k compounds tested on at least 50 proteins each. We show that Hit Dexter is able to discriminate non-promiscuous from promiscuous and highly-promiscuous compounds of large external test sets with MCC and AUC values of up to 0.67 and 0.96, respectively. Hit Dexter is available as a free web service at http://hitdexter.zbh.uni-hamburg.de.

[1] P. Szymański, M. Markowicz, E. Mikiciuk-Olasik, *Int. J. Mol. Sci.* **2012**, *13*, 427–452.
[2] R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, et al., *Nat. Rev. Drug Discov.* **2011**, *10*, 188–195.
[3] J. Baell, M. A. Walters, *Nature* **2014**, *513*, 481–483.
[4] C. Stork, J. Wagner, N.-O. Friedrich, C. de Bruyn Kops, M. Šícho, J. Kirchmair, *ChemMedChem* **2017**, DOI 10.1002/cmdc.201700673.