Application of an in silico Mechanism-of-Action Protocol to High-Content Cytotoxicity Screening Data utilizing WikiPathway data extracted from Open PHACTS

Lewis H. Mervin[1], Qing Cao[2], Ian P. Barrett[3], Mike A. Firth[3], David Murray[4], Lisa McWilliams[4], Ola Engkvist[2] and Andreas Bender[1]

[1] Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, UK
[2] Discovery Sciences, Chemistry Innovation Centre, AstraZeneca R&D, Mölndal, Sweden
[3] Discovery Sciences, AstraZeneca R&D, Cambridge Science Park, UK
[4] Discovery Sciences, Innovative Medicines, Alderley Park, Manchester, UK

The cytotoxic tendencies of compounds are a leading cause of attrition for many drug candidates. Although cell viability assays are capable of quantifiably measuring even low-level cytotoxicity, such experiments require subsequent mechanism-of-action (MoA) studies to identify cellular targets and pathways lending to the observed biological response. In this study, we developed an *in silico* MoA analysis protocol, comprising Bernoulli Naïve Bayes compound bioactivity profiling utilizing 9.5 million active and 600 million inactive data points, annotation of predicted targets with pathways from Open PHACTS v2.0, and calculation of enrichment metrics to highlight targets and pathways likely to be implicated in the toxic phenotype. This protocol was applied to a set of 6,800 cytotoxic compounds from an in-house cell-based screen. Two different compound reference sets are used to describe inactive chemical space to correct for the over or under-prediction of targets; 6.8 million putative non-toxic compounds from PubChem and 320,000 confirmed non-toxic compounds from in-house cytotoxicity screens. PubChem enrichment results highlight the diverse targets and pathways extracted from WikiPathways implicated in the mechanisms of cytotoxicity, with many processes associated with the fidelity of Mitotic Telophase, Nucleosome assembly and mRNA Capping. Comparison against confirmed non-toxic compounds found increased resolution for specific pathways, for example the PRC2, RHO and TLR3 signaling cascades. A random forest algorithm was subsequently trained on the cytotoxic compounds, and used to predict cytotoxicity for a library of untested compounds at a range of probabilities. Subsequent assessment of 1,000 compounds in a cell-based viability screen enabled the gradual evaluation of the applicability domain and assessment of the MoA hypothesis for the toxic phenotype.