

 HIGH-CONTENT SCREENING

Integrating information

High-content screening (HCS), which combines automated microscopy with image analysis, enables phenotypic profiling of compounds on the basis of cellular activity. However, huge data sets are generated — many of which have unclear biological meaning — and so the huge potential of HCS to inform on biological effects relevant to therapeutics and toxicity is largely untapped. Now, Feng and colleagues describe a methodology for the integration of HCS data with chemical structure information and computational target predictions.

First, the authors used a cellular-proliferation assay that measured 36 nuclear cytological features to screen a chemically diverse library (that included known bioactives) of more than 6,000 compounds. Factor analysis — a well-defined method for analysing multidimensional data sets that allows data reduction and quantifies common factors or phenotypes — was then introduced to mine the data sets. This enabled the cytological features to be mapped into six underlying attributes such as nuclear size, nuclear shape and DNA replication.

By calculating mean response scores for each factor for each compound, 211 hits were identified, which were then profiled for biological activity using hierarchical clustering of the factor scores. This revealed seven primary clusters, termed phenotypes, which could be interpreted as having biological meaning, such as mitotic arrest and apoptosis.

Phenotypic factors were then compared with either chemical structure information or computationally predicted targets. When computational methods and analyses compared phenotypic data with structural information, the majority of compounds with similar structures showed similar phenotypic readouts. However, 4% of compounds that had small changes in structure showed large changes in function.

The authors then chose several structure–activity relationships (SARs) to be examined in more detail; this revealed the capability of HCS combined with factor analysis to make subtle phenotypic distinctions. For example, the effects of corticosteroid-like and progesterone-like steroids could be readily discriminated, even though both cause cells to stop proliferating at the same stage of the cell cycle.

To investigate how distinct structural classes of compounds could produce similar phenotypes, the authors implemented a statistical, ligand-based target prediction method. Statistical models of substructural features were combined with an annotated chemogenomics database that associated molecular structures with their cognate biological targets. This was then used to predict the targets of the active compounds. The analysis revealed that phenotypes correlated better with predicted compound targets than with the compound structures themselves, indicating that the observed divergence in SARs could in part be accounted for



PHOTOALTO

by structurally different compounds having common targets.

Last, the authors investigated whether the methodology could predict a particular target. They studied results from a mitotic arrest phenotype — which contained four distinct groups of structurally related compounds — for which the structure-based method predicted multiple targets for each compound. By focusing on the top five predicted targets, it was found that most of the compounds were predicted to target tubulin, and indeed, when cytoskeletal morphology was examined experimentally, this was found to be the case.

The integration of complex imaging data with additional data sets represents a broadly applicable, easily transferable way to gain maximal insight into HCS data. Also, the use of primary cells and more disease-relevant probes will further increase the resolution of this methodology in areas relevant to drug discovery.

Charlotte Harrison

ORIGINAL RESEARCH PAPER Young, D. W. *et al.* Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nature Chem. Biol.* **4**, 59–68 (2008)