Using Leadscope to Analyze the ToxCast Data

Glenn J. Myatt*, David Bower, Kevin P. Cross, Donald P. Quigley, Leadscope, Inc, Columbus, Ohio 43215, USA

Abstract

In the first phase of the ToxCast project, over 300 compounds with known in vivo data have been screened across a series of in vitro assays. Using the in vitro data, alongside other information related to the chemical structures, in silico models are to be built to predict in vivo toxicity. This poster presents an analysis of the ToxCast data to identify relationships between the in vitro screens and the in vivo endpoints. An analysis was also performed exploring the relationships between chemical fragments and the in vivo endpoints. Classification models were built using partial logistic regression for all in vivo endpoints with a sufficient active/inactive ratio. Initially, a set of in vitro assays were used as independent variables for each in vivo endpoint model. Combinations of in vitro assays were selected that produced the highest model concordance. Using these results, a further analysis was performed to understand what types of chemical structures are poorly predicted. Specific chemical fragment fingerprints are then added as independent variables to the model to improve the model's specificity or sensitivity.

Introduction

As a result of the high cost associated with assessing the toxicity of the thousands of chemicals in commerce, the US EPA National Center of Computational Toxicology has initiated the ToxCast project. In phase I, the goal is to identify signatures from hundreds of in vitro assays to use in combination with other data, such as structural fragments, that are predictive of in vivo endpoints.

Chemical Structures

In phase I, approximately 300 chemicals were selected for testing where sufficient in vivo data had already been collected. These chemicals were clustered to illustrate the chemical diversity using the Leadscope software (agglomerative hierarchical clustering, complete linkage, at a distance of 0.7)[1]. This resulted in the following clusters:

9-membered clusters

7-membered clusters

6-membered clusters

5-membered clusters

4-membered clusters

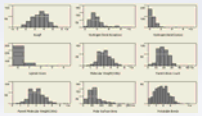
3-membered clusters

2-membered clusters

Singletons

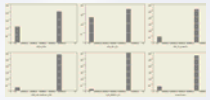
* Email: gmyatt@leadscope.com

The following illustrates the frequency distribution of nine calculated properties in Leadscope:



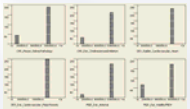
In Vitro Data

In vitro data has been generated from a number of laboratories for over 400 HTS endpoints. The data shows the lowest concentration at which significant activity was observed, with 1000000 used to indicate an inactive result. A portion of the table of results is displayed below. In addition, the frequency distribution for 6 randomly selected vitro endpoint is shown. It is typical that the majority of the values for a single assay are assigned as inactive (1000000), with a smaller number of values typically in the range 0-2000.



In Vivo Data

The in vivo data covers chronic, development, and multiple generation reproductive studies. Individual values are in mg/kg/day and are always LELs, with 1000000 representing an inactive result. Not all structures have values for all endpoints. The data covers specific rat, mouse and rabbit endpoints. The following illustrates a portion of the in vivo data table, along with frequency distributions for six selected endpoints.

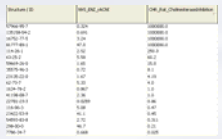


Data Analysis

Certain compounds have been tested in duplicate or triplicate. The in vitro results for these compounds are consolidated using the following rules: if the majority of results are inactive the value is assigned 1000000 (inactive); if the majority of results are active then the average value is calculated; otherwise the value is not used. To facilitate data analysis, two additional variables are created from the original data. A variable based on a five-point grading scheme is generated where 1000000 is 1, active data >500 is 2, 50-500 is 3, 10-50 is 4, and below 10 is 5. Binary variables are also generated with 1000000 assigned inactive and all other active values assigned as active.

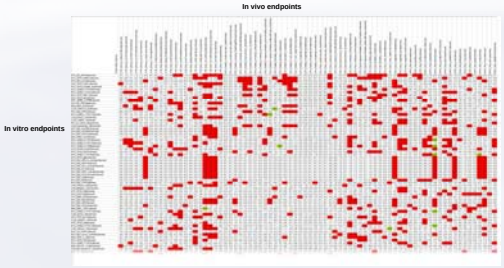
A z-score[1] is calculated between all in vitro endpoints (transformed to binary) and all in vivo endpoints (5-point grade). Similarly, a z-score is calculated between all Leadscope features and all in vivo endpoints. Higher positive values indicate that there is an unusually high average in vivo activity level associated with either the active in vitro endpoint or the presence of a chemical feature. Similarly, negative z-scores indicate a lower than expected level of in vivo activity. Red is used to highlight z-scores greater than 2 and green less than -2. Two examples are shown below:

The average activity based on the 5-point grading scheme for CHR_RAT_CholinesteraseInhibition is 1.63. The average activity for the subset of compounds that are active in the NVS_ENZ_rAChE assay is 3.83. This unusually high activity level is shown as a high positive z-score of 6.9. The 18 compounds in this subset are shown below.

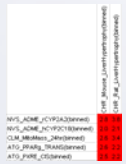


Analysis of in vitro data

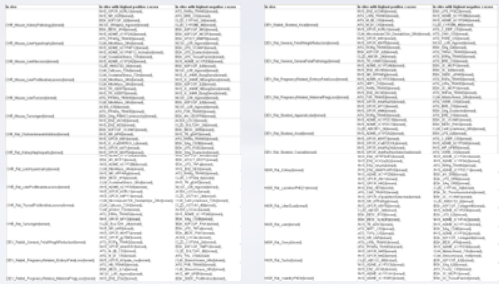
The chart below illustrates the z-scores calculated for all in vitro assays against all in vivo endpoints. It is sorted on the sum of the row z-scores, with only the top rows shown.



Organizing and filtering the table in various ways can help answer specific questions, such as what in vitro assays have high z-scores in CHR_Mouse_LiverHypertrophy and CHR_Rat_LiverHypertrophy?



The table below shows for all in vivo endpoints with greater than 15% actives, the four in vitro assays with the highest positive z-scores and the lowest negative z-scores. This initial prioritization of the in vitro assays will be used as a starting point for modeling.

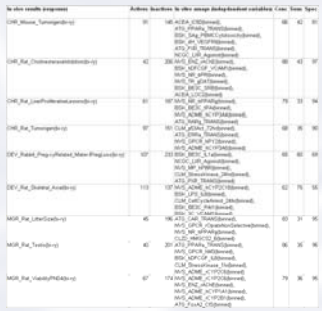


Analysis of Structural Features

The chart below illustrates the z-scores calculated for all of Leadscope's 27,000 specific features that are present in at least three compounds for all in vivo endpoints. It is sorted on the sum of the row's z-scores and only the top rows are shown.

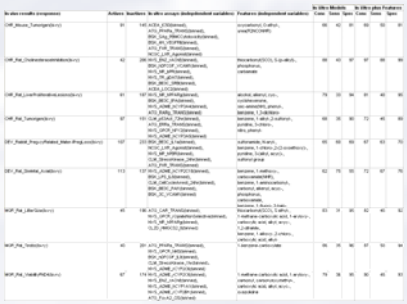
In Vitro Model Results

For the 8 prioritized in vitro assays (according to their z-score), 255 models from all combinations of these assays were built. The model with the highest accuracy was selected for each in vivo endpoint. The following table presents the results for all models built where (1) the active/inactive ratio was greater than 0.15 and the sensitivity of the model was greater than 30%. The models were all built using the Leadscope predictive model builder, which uses a partial logistic regression model [2]. The cross validated (5%) results are presented in the following table:



In Vitro Plus Chemical Features Model Results

Where the models have a low sensitivity (or specificity), the false negatives (or false positives) were analyzed against all Leadscope features. Features with an unusually high number present in the false negative (or false negative) set, that also had a z-score greater than two were selected and used to build a series of new models (in addition to the in vitro data). The cross validated (5%) results are presented below:



Introducing the chemical features as independent variables (alongside the in vitro assays) increased the average sensitivity of the nine models from 43.2% to 54.2%. Using the average of the sensitivity and specificity as a measure of the models' performance, the average over the nine models increased from 64.5% to 82.2%.

Conclusions

The diversity of the ToxCast chemical structures has been visualized by clustering the compounds. Approximately, one third are members of clusters of size 4 or greater, with another third assigned as singletons.

A systematic approach was adopted to create a series of classification models. Each in vivo endpoint (transformed to binary) was used as the response variable and the selected in vitro assays were used as the independent variables. Steps included identifying the best combinations of in vitro assays for use in modeling each in vivo endpoint, analyzing the compounds that do not model well, and identifying statistically important structural features that are under-represented in the set of false positives or false negatives.

Classification models can be built using a small series in vitro assays as independent variables. Adding selected chemical features as additional independent variables increases the performance of the model.

References

[1] Yang C., Richard A.M., Cross, K.P. The Art of Data Mining the Minefields of Toxicity Databases to Link Chemistry to Biology. *Current Computer-Aided Drug Design* **2006**;2:135-150.

[2] Yang, C., Cross, K., Myatt, G.J., Blower, P.E., and Rathman, J.F. Building predictive models for protein tyrosine phosphatase 1B inhibitors based on discriminating structural features by reassembling medicinal chemistry building blocks. *J. Med. Chem.*, **2004**; 47:5984-5994