MultiFlow®

DNA Damage Kit – p53, γH2AX, Phospho-Histone H3

App Note 2:
Train and Test

Build your foundation for a powerful assay

Introduction

Once the MultiFlow Installation has been performed (App Note 1), you are now ready to begin analyzing reference compounds. This data will serve as the basis for profiling unknown compounds.

Machine Learning

Machine Learning may sound intimidating, but it is really just a form of pattern recognition performed by the computer. Responses from compounds with well-established genotoxicity profiles are entered into specialized software. This serves as a library of prototypical responses for the various endpoints. The software then refers to this library when it encounters an unknown compound.

By comparing the responses of the unknown compound to those of known compounds, the software can predict the Mode of Action for that unknown compound. The combination of a strong training set with several different models is an effective approach to predict a new compound's class.

Training Set

The initial Training Set consists of 12 compounds with varied genotoxic mechanisms (Table 1). TK6 cells are exposed to a large concentration range: low exposure with little to no response, and higher concentrations with larger responses. Four compounds with nineteen concentrations in singlet are studied per 96 well plate, along with positive and negative controls. Samples are collected and analyzed at 4 and 24 hours of exposure to provide data over two time points. With MultiFlow's simple add-and-read method, the training and testing can be done in a matter of days.

Table 1: Chemical List

Compound	Class	Group
AMG-900	aneugen	train
Carbendazim	aneugen	train
Crizotinib	aneugen	train
Griseofulvin	aneugen	train
5-Fluorouracil	clastogen	train
Menadione	clastogen	train
Methyl methanesulfonate	clastogen	train
Stauvudine	clastogen	train
Brefeldin A	nongenotox	train
CCCP	nongenotox	train
Cycloheximide	nongenotox	train
Tunicamycin	nongenotox	train
Mebendazole	aneugen	test
Nocodozole	aneugen	test
Tozasertib	aneugen	test
ZM-447439	aneugen	test
Etoposide	clastogen	test
Hydroxyurea	clastogen	test
Methotrexate	clastogen	test
MNNG	clastogen	test
Clofibrate	nongenotox	test
Imatinib mesylate	nongenotox	test
Hexachloroethane	nongenotox	test
Phenanthrene	nongenotox	test

The data from the initial Training Set are checked using Litron's MultiFlow Quality Control process. From these data, three Machine Learning models (currently Random Forest, Artificial Neural Network, and Logistic Regression) are created. The ability of these models to correctly identify the compounds according to their *a priori* classification is verified using several metrics.

Test Set

Once the models have been verified, the next step is to perform the assay with the Test Set of 12 compounds (Table 1). The resulting data are analyzed using the models prepared above. Once these Machine Learning models accurately identify the Test Set compounds, they are likely to predict the genotoxic Mode of Action for unknown compounds.

Building the Models

The data from both the Training and Test Sets are combined into a 24 compound dataset. This will serve as the lab's training set for future analyses. The Machine Learning algorithms are generated again using this new, larger dataset. Each of the 24 compounds are verified against these new models (Figure 1).

In addition to Machine Learning models, labs are also encouraged to apply MultiFlow's Global Evaluation Factors (GEFs). GEFs can effectively classify compounds with new response patterns that the models may miss.

Conclusions

This completes the process of training and testing the predictive abilities of the MultiFlow assay. Your lab is now ready to study new compounds. The third App Note in the series covers the data analytics and report generation services available at Litron.

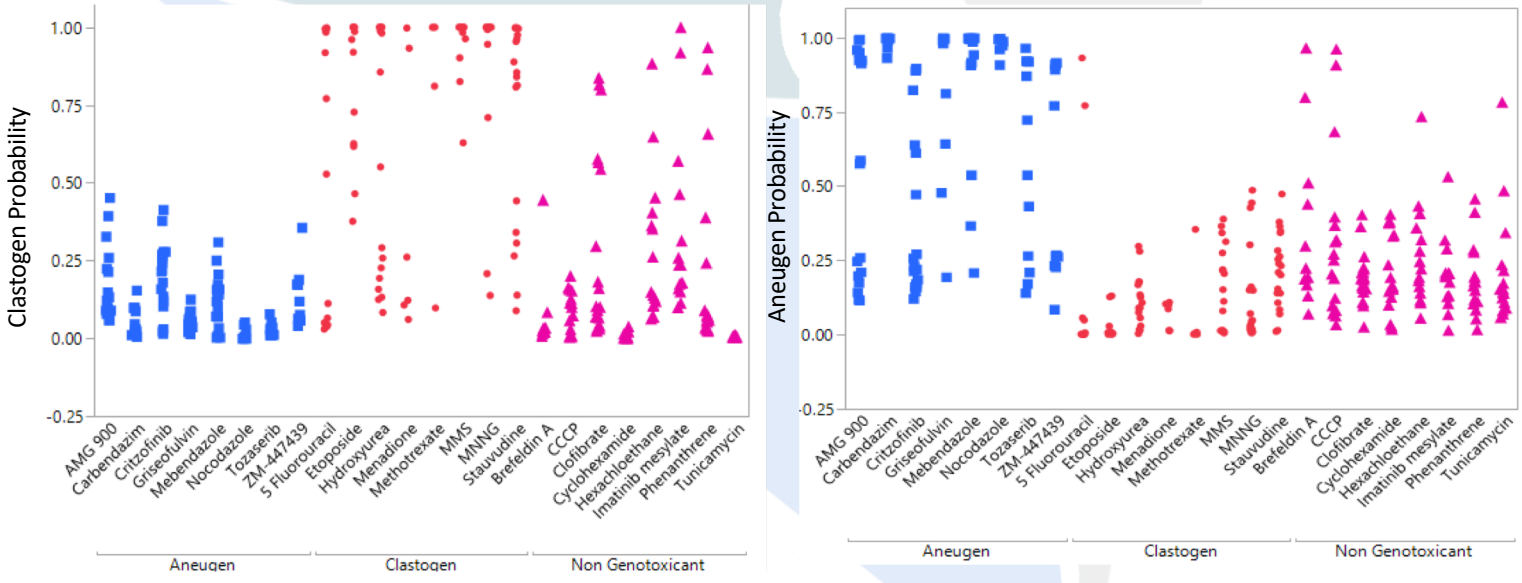

Figure 1: Manhattan plots showing clastogens (red) and aneugens (blue) from the 24 compound set being correctly identified.