Supplemental file S1.

A random forest analysis (known as Bootstrap forest in JMP software), was used to analyze training set data: 85 chemicals assigned to 1 of 3 groups: clastogenic, aneugenic, or non-genotoxicic. Random forests is a machine learning statistical method that can be used to identify the most important factors required to make a prediction, in this case, classifying mode of action group membership. Decision trees for group membership are constructed with randomly selected subsets of chemicals and variables. Forests of these decision trees are built, which together make a prediction for each chemical. These results are used to identify and validate variables most important to the prediction.

• JMP PRO, v13

• Bootstrap forest platform applied to MultiFlow assay data derived from 85 chemical training set

• 10 candidate biomarkers (listed under the heading ‘Term’, below) evaluated for their ability to discriminate among 3 groups, clastogenic, aneugenic, or non-genotoxic

• Weighted on %cytotoxicity at 24 h

• 300 tree forest, 3 factors per split

