Published in final edited form as:

Chem Res Toxicol. 2019 July 15; 32(7): 1384–1401. doi:10.1021/acs.chemrestox.9b00053.

Identifying Compounds with Genotoxicity Potential Using Tox21 High Throughput Screening Assays

J-H Hsieh¹, SL Smith-Roe², R Huang³, A Sedykh⁴, KR Shockley⁵, SS Auerbach², BA Merrick², M Xia³, RR Tice⁶, KL Witt^{2,*}

¹Kelly Government Solutions, Research Triangle Park, NC

²Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC

³National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD

⁴Sciome, LLC, Research Triangle Park, NC

⁵Division of Intramural Research, National Institute of Environmental Health Sciences, Research Triangle Park, NC

⁶RTice Consulting, Durham, NC

Abstract

Genotoxicity is a critical component of a comprehensive toxicological profile. The Tox21 Program used five quantitative high throughput screening (qHTS) assays measuring some aspect of DNA damage/repair to provide information on the genotoxic potential of over 10,000 compounds. Included were assays detecting activation of p53, increases in the DNA repair protein ATAD5, phosphorylation of H2AX, and enhanced cytotoxicity in DT40 cells deficient in DNA repair proteins REV3 or KU70/RAD54. Each assay measures a distinct component of the DNA damage response signaling network; >70% of active compounds were detected in only one of the five assays. When qHTS results were compared to results from three standard genotoxicity assays (bacterial mutation, *in vitro* chromosomal aberration, and *in vivo* micronucleus), a maximum of 40% of known, direct-acting genotoxicants were active in one or more of the qHTS genotoxicity assays, indicating low sensitivity. This suggests that these qHTS assays cannot in their current form be used to replace traditional genotoxicity assays. However, despite the low sensitivity, ranking chemicals by potency of response in the qHTS assays revealed an enrichment for genotoxicants up to 12-fold compared with random selection, when allowing a 1% false positive rate. This finding indicates these qHTS assays can be used to prioritize chemicals for further investigation, allowing resources to focus on compounds most likely to induce genotoxic effects. To refine this prioritization process, models for predicting genotoxicity potential of chemicals that were active in Tox21 genotoxicity assays were constructed using all Tox21 assay data, yielding a prediction accuracy up to 0.83. Data from qHTS assays related to stress response pathway signaling (including genotoxicity) were the most informative for model construction. By using the results from qHTS genotoxicity assays, predictions from models based on qHTS data, and

*To whom correspondence should be addressed: witt@niehs.nih.gov.

predictions from commercial bacterial mutagenicity QSAR models, we prioritized Tox21 chemicals for genotoxicity characterization.

Introduction

Critical to any comprehensive compound toxicity profile is an assessment of genotoxicity potential. This is due in part to the mechanistic linkage between genotoxicity and numerous adverse health outcomes including cancer, neurodegenerative diseases, and birth defects¹. Historically, genotoxicity profiles for substances have been established using a variety of *in vivo* and *in vitro* assays, including bacterial mutagenicity, chromosomal aberration, micronucleus, DNA damage, and gene mutation assays. Each assay detects a specific type of genetic damage (e.g., point mutations, chromosome breakage, DNA adducts). All have been well-characterized and validated, and all are governed by international testing guidelines (e.g., Organization for Economic and Community Development [OECD], International Council for Harmonization [ICH]). Data produced from these assays are accepted by regulatory agencies as part of a comprehensive data package that defines the genotoxic potential of a compound and provides information suitable for human risk assessment. Extensive sets of consensus reference compounds (both positives as well as negatives) have been compiled for each assay². All *in vitro* assays are conducted with and without an exogenous metabolic activation system, generally an induced rat liver S9 microsome mix. Positive responses in these assays are highly predictive of cancer in rodent studies (high positive predictivity), but negative results are not informative for carcinogenicity predictions, as each assay detects a distinct subgroup of compounds (e.g., point mutagens or clastogens), and not all carcinogens are genotoxic. No single test, if negative, can define a compound as non-genotoxic. To ensure maximal opportunity to identify a genotoxicant or to conclude that a compound has no genetic liability, a battery of tests covering a variety of genotoxicity endpoints is routinely used. Although there are benefits to using each of these traditional assays to characterize compound genotoxicity, there are also limitations (e.g., sensitivity, cost, time). Each assay as currently used tends to produce a phenotypic response (yes/no) rather than provide highly quantitative data from detailed dose response curves that might help in better defining risk by supporting benchmark dose and point-of-departure determinations, which are increasingly sought by regulatory scientists to enhance risk assessment determinations.

Although genetic toxicologists have relied upon these well-characterized assays for decades, there exists an enormous backlog of thousands of compounds for which insufficient, or no, genotoxicity data have been generated. Due to limited resources in time, personnel, and finances, the backlog of testing cannot be reasonably eliminated by testing all of these data-poor compounds in traditional assays, despite the need for such data (e.g., Government of Canada, Chemicals Management Plan³). Thus, there is a demonstrated need for alternative methods to characterize compound genotoxicity and/or to prioritize the thousands of inadequately tested compounds for evaluation in traditional genotoxicity assays, to better focus available resources where the need is greatest. One such alternative method is the use of *in silico* quantitative structure-activity relationship (QSAR) models for predicting bacterial mutagenicity⁴. However, genotoxicity is not limited to the induction of point

mutations in bacteria, but also includes other types of genetic damage including chromosomal changes (both numerical and structural), and DNA damage (e.g., DNA adducts, DNA-DNA crosslinks). Although additional predictive models for other genotoxicity biomarkers are under construction, they are not yet fully developed. Thus, to help in addressing the backlog of data-poor compounds, the U.S. Tox21 Program has been actively investigating alternative *in vitro* assays for use in predicting genotoxicity and for prioritizing compounds identified as likely to be genotoxic for further evaluation in more traditional assays to confirm genotoxicity predictions.

The U.S. Tox21 Program is an interagency effort that includes the National Toxicology Program (NTP), the U.S. Environmental Protection Agency (EPA), the NIH Center for Advancing Translational Sciences (NCATS), and the Food and Drug Administration (FDA). It was established to achieve three main goals: 1) develop rapid and accurate *in vitro* methods to evaluate the toxicity of thousands of compounds with insufficient toxicity data to support risk assessments, 2) develop toxicity profiles for these compounds, and 3) prioritize compounds, based on their toxicity profiles, for further in-depth toxicological characterization⁵. The latter goal is of particular importance, given the limited resources available for evaluating the tens of thousands of compounds in commerce for which little or no toxicity information exists.

An initial Tox21 effort involved compiling a large library⁶ of more than 10,000 commercially available compounds including pesticides, industrial chemicals, natural food products, and drugs, with characteristics amenable for qHTS (e.g., low volatility, soluble in dimethyl sulfoxide (DMSO)). To rapidly provide comprehensive compound toxicity profiles, a testing approach using what are referred to as quantitative (detailed concentration-response curves) high throughput screening (qHTS) assays was used. Over 70 such assays were developed. These qHTS assays use 1536-well plates on a robotic platform and a simple homogeneous (add, mix, measure) format with no aspiration steps due to the small volume per well (~5 μ L total volume). Aspiration with such a small assay volume promotes high well-to-well variation, compromising data quality. This inability to aspirate precludes the use of rat liver S9 mix for metabolic activation. On the order of 1000 – 3000 cells are plated per well in these assays and results are often measured within hours, so the events that are measured must be high frequency events that produce robust signals for detection by automated plate readers.

Some 70% of compounds within the Tox21 library have not been tested for genotoxicity in any traditional assays (e.g., bacterial mutagenicity, chromosomal aberration, micronucleus, or gene mutation assays). To aid in providing this much-needed genotoxicity data, the Tox21 assay portfolio included five qHTS assays that measure some aspect of DNA damage or repair. These are a 1) p53 response element β -lactamase (p53RE-bla) reporter assay in HCT-116 cells that detects activation of p53 in response to DNA damage and other cellular stressors⁷; 2) an ATAD5-luc assay in HEK293T cells to measure increased levels of a luciferase-tagged ATAD5 protein that localizes to the site of stalled replication forks resulting from DNA damage in replicating cells⁸; 3) a phosphorylated H2AX (γ H2AX) assay in Chinese hamster ovary (CHO)-K1 cells that detects DNA double strand breaks; and two differential cytotoxicity assays using a wild-type DT40 chicken lymphoblastoid cell

line, including 4) an isogenic knockout cell line deficient for REV3, a polymerase involved in translesion synthesis at stalled replication forks, and 5) an isogenic double knockout cell line deficient for KU70 and RAD54, proteins involved in non-homologous end joining and homologous recombination, respectively^{9,10}. Together, these 5 assays detect whether chemical exposure induces stalled replication forks and/or DNA double-strand breaks, indicative of genotoxicity (Figure 1).

It should be noted that there is some precedent for using qHTS assays for informing toxicological decision making. HTS data for compound profiling has recently been successfully applied to predicting endocrine disrupting activity¹¹, and acute toxicity effects in rats¹², and in understanding mechanisms of induced liver toxicity¹³.

There are challenges in using qHTS data for predicting genotoxicity, however. Each of the five qHTS DNA damage assays employed by Tox21 uses a different protocol (e.g., cell line, exposure duration, signal generation, read-out), but for each, positive and DMSO controls are run concurrently. None of these qHTS assays is governed by uniform performance guidelines, although assay technical performance must meet pre-set standards. Predictivity for human health impacts (e.g., carcinogenicity) has not been established. These qHTS assays can only attempt to assess effects of the parent compound, since S9 mix to support metabolic activation of an inactive parent molecule to reactive metabolites is currently precluded in the assay design. Keeping in mind these caveats, to determine the value of the qHTS DNA damage assay data in informing genotoxicity potential, we analyzed the results of each of the five assays independently and in combinations of 2, 3, 4, and 5 assays. Analyses were then compared to the data available from three traditional genotoxicity tests (bacterial mutagenicity, *in vitro* chromosomal aberration, and *in vivo* rodent erythrocyte micronucleus assays); only data produced in the absence of exogenous metabolic activation (S9) were used in the comparison, and the *in vivo* micronucleus assay data were curated to remove any chemical with a suspected requirement for S9 activation for a positive response in the assay. These three traditional assays were selected for this exercise because each has a sizable database available for comparison, and the data from each of these three assays is recognized by regulatory agencies as a valid indicator of genotoxicity; thus, they represent “gold standards” in genotoxicity testing. This effort was aimed at determining if the qHTS assay results can be used to accurately predict the outcomes achieved in these three traditional genetic toxicity assays, and therefore serve as replacements for the traditional assays, or if results from these qHTS assays are better suited for identifying potential genotoxicants, based on strength and breadth of response, and then prioritizing these potential genotoxicants for further in-depth characterization and confirmation in traditional assays.

We also investigated the potential benefits of incorporating data from additional Tox21 qHTS assays that measure other types of biological activity (e.g., nuclear receptor signaling, heat shock protein upregulation, mitochondrial damage) for predicting genotoxicity. Finally, predictions derived from chemical structure-based genotoxicity models were overlaid on the comprehensive analysis of the Tox21 qHTS data to rank compounds by genotoxicity potential, thereby prioritizing them for follow-up confirmation studies. Subsequently, seven of the highest ranked compounds were experimentally confirmed as genotoxicants. Based on

the results of these analytical approaches, we conclude that qHTS data have a promising role in the efforts to provide genotoxicity information to help inform regulatory decisions for the large number of compounds in commerce that currently lack such vital information. Further refinement of the assays and additional model development are planned to further characterize the role these qHTS assays, or the endpoints they measure, have for characterizing data-poor compounds.

Materials and Methods

Tox21 genotoxicity assays

Five assays were conducted to generate data to help define the DNA damaging potential for compounds in the Tox21 library. These included the p53RE-bla assay in human HCT-116 cells (Invitrogen, Carlsbad, CA), the ATAD5-luc assay in human HEK293 cells (laboratory-developed test⁸), the γ H2AX assay in CHO-K1 cells (Cisbio, Bedford, MA), and two differential cytotoxicity assays in isogenic DT40 chicken lymphoblastoid knockout cell lines deficient for DNA repair proteins REV3 or KU70/RAD54 (laboratory-developed test^{10,14-16}). Assay descriptions and performance characteristics are provided in Tables 1 and 2, respectively. Each of these five assays assesses chemical activity within a specific DNA repair signaling pathway, although some degree of overlap among pathways occurs due to the inherent redundancy in mammalian DNA repair mechanisms. Of note, the p53RE-bla assay may respond to a number of additional cellular stressors other than DNA damage, although most substances that induce p53 activation do so via induction of DNA damage.¹⁷

Based on the five Tox21 qHTS genotoxicity assays (p53RE-bla, ATAD5-luc, γ H2AX, and two DT40 knockouts [*REV3*^{-/-} and *KU70*^{-/-}/*RAD54*^{-/-}]; Figure 1), five genotoxicity-related endpoints were generated. Each assay endpoint represents a signed and directed effect (e.g., increase or decrease) of a biological process (e.g., transcriptional factor activity) for a genotoxicity target-of-interest (e.g., p53). In the following text, we used these annotations (KU70/RAD54 $\uparrow$, REV3 $\uparrow$, ATAD5 $\uparrow$, γ H2AX $\uparrow$, p53 $\uparrow$) to represent the five genotoxicity targets derived from the five assays. The tabular information associated with the assays, including targets, methods for target quantification, and experiment formats, are shown in Table 1.

Methods for data interpretation among the five assays have variable levels of complexity. The ATAD5-luc assay has a single channel luciferase readout in which intensity of the luminescent signal is directly proportional to the amount of protein present in the cell. The p53RE-bla and γ H2AX assays are two-channel assays, including a reporter channel and a background channel (for p53RE) or two anti-H2AX antibodies (one that detects H2AX and one that detects γ H2AX) to measure time resolved fluorescence resonance energy transfer (TR-FRET). For the p53RE assay, the ratio of the reporter gene channel to the background channel was used for data analysis. However, it was also necessary to ensure that similar responses were seen in both the reporter gene channel and the calculated ratio data, as cytotoxicity in the absence of a change in the reporter gene channel can produce an increase in the ratio that mimics reporter gene activity^{18,19}. For the two DT40 knockout assays, the activity endpoints cannot be derived directly from the outputs of the assay, but rather, the final endpoint must be quantified by the degree of increased cytotoxicity observed in the

isogenic knockout cell line (e.g., *REV3*^{-/-}) that has compromised DNA repair capacity compared with the DNA repair competent wild-type cell line (see Tox21 qHTS data analysis). The degree of separation between the concentration response curves (wild-type versus knockout) for differential cytotoxicity impacts the proportion of compounds that are identified as active.

Tox21 10K compound library

The Tox21 compound library currently consists of ~13,100 (~9000 unique) compounds procured from commercial sources by the EPA, NIEHS/NTP, and NCATS. The library consists of a wide variety of chemicals, including pesticides, industrial chemicals, natural food products, and drugs. The latter category includes failed drugs that were never marketed, drugs that are no longer marketed, and drugs that are marketed currently. The list of unique compound substances, including chemical names and Chemical Abstracts Service Registry Numbers (CASRNs), as well as curated chemical structures and auto-generated structure identifiers (formula, systematic names, SMILES, desalted SMILES, InChI) can be downloaded from the EPA Chemical Dashboard website (https://comptox.epa.gov/dashboard/chemical_lists/tox21sl). Each substance was assigned a Tox21 ID. Substances were prepared as stock solutions (generally at 20 mM) in dimethyl sulfoxide (DMSO), and were serially diluted in 1536-well microplates to yield 15 concentrations generally ranging from 1 nM to 92 μM (final concentrations in the wells).

Eighty-eight duplicate compounds were intentionally included on each of the screening plates to evaluate technical variability across plates and runs. The Tox21 compound library stock solutions have been undergoing chemical analysis to assess compound sample identity, purity, concentration, and stability under conditions of use (at initial plating, and again after 4 months of use). The currently available data are available at <https://tripod.nih.gov/tox21/samples>. Substances with suboptimal purity ratings (grade = F [incorrect MW], Fnc [no sample detected], or Fc [very low concentration, <5% of expected value]) at either of the time points measured (0 or four months) were excluded from the analyses conducted in this study.

For the following analyses related to evaluating the predictivity of Tox21 genotoxicity assays, we only focused on chemicals that were screened in all five assays. In total, 6831 unique chemicals with acceptable purity or under analytical analysis were screened in all five assays; the data derived from the qHTS assays were used in evaluating the concordance of qHTS assay results with results from traditional genotoxicity assays to determine feasibility of using qHTS assay data for predicting genotoxicity. In contrast, for the analyses used to prioritize chemicals for additional assessment of genotoxicity potential, the data from all of the 8579 unique compounds in the Tox21 library with acceptable purity or under analytical analysis (2200 compounds) that had been tested in any of the five genotoxicity qHTS assays were used.

Experimental genotoxicity data—Experimental genotoxicity data for Tox21 compounds were retrieved from the Leadscape® SAR Genetox Database (http://www.leadscope.com/product_info.php?products_id=77) as described previously⁷. In total,

three data sets with three endpoints were created for comparison with Tox21 genotoxicity data: bacterial mutagenicity without S9 (BM; # positive/# negative = 585/756), *in vitro* chromosome aberration without S9 (CA; # positive/# negative = 378/458), and *in vivo* micronucleus (MN; # positive/# negative = 142/469). Only traditional data generated in the absence of exogenous metabolic activation (S9 mix) were utilized, because the qHTS assays do not incorporate metabolic transformation beyond what limited metabolism capability may be present in each of the cell lines. For the *in vivo* erythrocyte micronucleus test data, compounds inferred from *in vitro* assay data (BM or *in vitro* CA data) to have activity dependent on metabolic activation were excluded. In addition, we generated another data set for the BM endpoint, based on positive results achieved with compounds tested at concentrations comparable to those in the qHTS assays and negative results for compounds that were seen at concentrations higher than those in the qHTS assays (“BM (dose-aware)”⁷). The data in these four datasets (BM, CA, MN, and BM (dose-aware)) were used for comparing test outcomes between traditional genotoxicity assays and the Tox21 qHTS genotoxicity assays.

Predicted genotoxicity data—In addition to the experimental genotoxicity data obtained from Leadscape, *in silico* QSAR models provided by MultiCASE, Inc. (<http://www.multicase.com>) were used to predict the BM of the Tox21 compounds. The Tox21 compounds were virtually screened against two commercial QSAR models for BM: the *Salmonella* model [*PHARM_SALM* (v1.5.1.8)] and the *E. coli*/*Salmonella* TA102 model [*PHARM_ECOLI* (v1.5.1.8)]. These models were constructed with all available data (i.e., irrespective of exogenous metabolic activation). Compounds predicted to be positive within the applicability domain in either model were assigned a positive call; compounds predicted to be negative within the applicability domain in both models were considered negative. All other compounds (e.g., those that were negative only in one model, out of applicability domain, or lacked chemical descriptors) were labeled as inconclusive. The dataset was named “*in silico* BM”. This dataset was used for prioritization of chemicals with unknown genotoxicity potential (see ‘Chemical prioritization scheme’ below).

Tox21 qHTS data analysis—The raw plate reads for each titration point were first normalized relative to the positive control compound and DMSO-only wells (0%) as follows: % Activity = $[(V_{\text{compound}} - V_{\text{DMSO}})/(V_{\text{pos}} - V_{\text{DMSO}})] \times 100$, where V_{compound} denotes the compound well values, V_{pos} denotes the median value of the positive control wells, and V_{DMSO} denotes the median values of the DMSO-only wells. The positive control chemicals used in the normalization were tetra-N-octylammonium bromide (the two DT40 knockout cell assays), 5-fluorouridine (ATAD5-luc assay), etoposide (γ H2AX assay), and mitomycin C (p53RE-bla assay). The nominal concentrations of these positive controls in the wells were 46 μ M (both DT40 knockout cell assays), 36.8 μ M (ATAD5-luc assay), 153 μ M (γ H2AX assay), and 11.5 μ M (p53RE-bla assay). Tetra-N-octylammonium bromide is a cytotoxic compound, while the other three positive control chemicals (5-fluorouridine, etoposide, and mitomycin C) are known genotoxic compounds. The % of Activity was rescaled so that the baseline value was 0%. The data set was then corrected using the DMSO-only compound plates at the beginning and end of the compound plate stack by applying an NCATS in-house pattern correction algorithm²⁰.

The normalized concentration–response data for each substance at each run (three runs in total) were applied to a qHTS noise filtering algorithm, Curvexp^{21,22}, with noise level derived from the response variation seen in the 88 technical replicates²³. Four activity parameters, including weighted area-under-curve (wAUC, total activity), point-of-departure (POD, concentration at which the response is equivalent to the noise threshold), EC₅₀ (half maximal effect concentration), and E_{max} (maximal response), were reported for each curve. The wAUC metric is the product of the POD and the AUC, normalized by the test concentration range¹⁹. The design of the metric is to capture both potency and efficacy of the response (i.e., AUC) as well as can be used to compare compounds tested over different concentration ranges (through weighting by POD and concentration range normalization).

Curves with absolute wAUC > 0 were considered as having significant responses; curves with wAUC = 0 were considered as having no significant responses. Substances with >50% of curves with significant responses were considered active. The other activity parameters (POD, EC₅₀, and E_{max}) were summarized using the median value based on data from three runs. Potency values were not assigned for inactive chemicals. For the relatively weaker active substances (wAUC < 25th percentile of active compounds), significant responses were required in all 3 runs. If not all concentration response curves for a substance were significant, a flag was applied ('weak/noisy in repeats').

Additionally, active results that might be due to assay interference, including 'auto-fluorescence' and 'no signal readout support', and 'cytotoxicity', were flagged¹⁹. The 'auto-fluorescence' and 'inconsistent channel responses in assays' flags were for the γ H2AX and p53RE-bla assays due to their fluorescence output and two-channel format. The 'cytotoxicity' flag was specifically for the DT40 assay, indicating that there was no significant increase in cytotoxicity observed in the knockout cell line vs the wild-type cell line. Significance was defined as having a POD value of $p < 0.05$ using the one-tailed Welch's t-test and a wAUC fold-change > 1.5.

Results for active substances with flags were labeled as inconclusive. To simplify the comparison with traditional genotoxicity data in the databases we used, where compounds are identified by CASRN, the substance results were further collapsed by CASRN. Thus, active compounds are defined as those chemicals having significant responses for >50% of the non-flagged substances of the same CASRN. Activity values assigned to each compound were the average activity values derived from the non-flagged substances.

Statistical methods—The three major analytical tasks performed in this study are summarized in Table 3: a) evaluation of the predictivity of the Tox21 genotoxicity assays for known genotoxicity, b) evaluation of the ability of the Tox21 genotoxicity assays to prioritize compounds based on activity levels, c) and prediction/prioritization of chemicals with unknown genotoxicity potential using Tox21 qHTS data. For each of the tasks, we applied specific statistical approaches: contingency table, receiver operating characteristic (ROC) enrichment, and elastic net regularization, respectively. The details of these approaches are explained in the following sections.

Contingency table: The results (active and inactive) for each of the five Tox21 genotoxicity assay endpoints were compared with the results (positive and negative) for each of the four traditional genotoxicity assay endpoints (BM, CA, MN, *in silico* BM) using the contingency table approach, from which sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were derived. The 95% confidence intervals (percentile method) for each parameter were calculated using bootstrap statistics (R *boot* package²⁴, number of bootstrap replicates = 2000).

In addition, we investigated if combining results from the five Tox21 assays would improve predictivity. For this task, the active results from the five Tox21 qHTS genotoxicity endpoints were combined using either the 'union' or the 'intersection' method. For the 'union' method, any chemical active in any one of the 5 assays was included as active; for the 'intersection' method, only chemicals that were active in all of the assays under consideration (2 – 5 assays) were included. The inactive results were also combined; however, only the 'intersection' method was applied for the inactive results (a chemical was considered inactive if it was inactive in all of the Tox21 genotoxicity endpoints used). Compounds that did not meet the established criteria for grouping were classified as inconclusive and were not used in constructing the contingency tables. The same statistics were applied for calculating the confidence intervals.

Receiver operating characteristic (ROC) enrichment: The contingency table allows us to understand the overall predictivity (based on binary calls) of the Tox21 genotoxicity assays but in the case of low predictivity, the assays could still be useful for prioritization by potency ranking, if the top ranked chemicals tended to be genotoxic. Therefore, we used the ROC enrichment metric to evaluate the performance of prioritization²⁵.

For active compounds in each of the five Tox21 qHTS genotoxicity endpoints, logistic transformation was applied to the wAUC value to convert it into an activity score with a value between 0 and 1²⁶. For actives in the two DT40 knockout assays, the difference in wAUC between the knockout cell line and the wild-type cell line was used to determine activity score. Inactive compounds had a score equivalent to 0. A score equivalent to 0.0001 was set for the inconclusive results. Chemicals were ranked based on the activity score in each of the five Tox21 qHTS genotoxicity endpoints and the result was compared with each of the four traditional genotoxicity assay endpoints (BM, CA, MN, *in silico* BM). The ROC enrichment (ROCE) at 1% was calculated based on the ROC curve. The ROCE at 1% (ROCE@1%) represents the slope (sensitivity/false positive value) of the ROC curve when 1% of known negatives were included in the top portion of the ranking list (i.e., 1% false positive rate). The 95% confidence interval (percentile method) for each of the parameters was also calculated using the bootstrap statistics (R *pROC* package²⁷, number of bootstrap replicates = 2000).

In addition, we evaluated the prioritization performance based on the combined results (2 – 5) by direct summation of the activity scores. To ensure that compounds that were active in multiple assays were ranked at the top of the priority list, a weight factor (1) was added to the activity score if the chemical was active. Similarly, the ROCE@1% was calculated

accordingly, based on the ranking list from the combined score. The same statistical approach was applied to confidence intervals.

Elastic net regularization: We intended to use all of the available Tox21 data, not just the Tox21 genotoxicity assays, to prioritize/predict chemicals with unknown genotoxicity. Therefore, we applied a linear regression method (elastic net regularization, R package, *glmnet*²⁸) to build models for each of the three traditional genotoxicity endpoints (BM, CA, MN) using all (or a subset) of Tox21 assay data (see below, ‘modeling sets’). Elastic net regularization is designed to deal with collinearity and it is also a method of variable selection, which forces coefficients for some/many variables to be zero to provide good estimates of the coefficients of the remaining variables²⁹. The dependent variable was a positive/negative activity call in the traditional genotoxicity endpoints, while the independent variables were the activity values (i.e., wAUC) of the endpoints from Tox21 assay screens. The wAUC values were transformed by using $\log_{10}(\text{wAUC}+1)$ function (+1 shift was used to avoid the infinity value for the inactive activities). The inconclusive activity calls from the Tox21 qHTS assays were set as inactive (wAUC = 0). The independent variable matrix was Z-score scaled prior to model training. A five-fold cross validation (CV) was applied to optimize the lambda with different alpha values from 0 to 1 (0.1 as increment) as input using the ROC Area Under Curve (ROC AUC) performance metric from the ROC curve based on genotoxic probability (i.e., either BM, CA, or MN) as the output. The procedure was repeated independently 10 times. The information for the optimized lambda at a given alpha producing maximum ROC AUC in the CV was stored. In total, 110 models were generated for each dataset based on optimized lambda. To gain insights into the variable (i.e., Tox21 assays) contributions to BM, CA, or MN endpoint prediction, we reported the average coefficient values from all the models that had ROC AUC values higher than the lower bound of the confidence interval from the parameter set (10 runs with alpha fixed) producing the best performance. These models (designated htsBM, htsCA, htsMN) were also applied to predict the genotoxicity potential of chemicals. The average probability results were reported and used in the chemical prioritization scheme (see below ‘Chemical prioritization scheme’).

Modeling sets: The datasets of independent variables were summarized in Table 3, including i) only data from genotoxicity stress response assays (n = 5), ii) only data from stress response pathway assays (n = 13), iii) data from all Tox21 assays (n = 78). For dependent variables, we also used three different types for building models with different purposes: a) all chemicals with a positive or negative call in each of the traditional genotoxicity endpoints (BM, CA, or MN), b) a subset of chemicals in a) that were active in one of the five Tox21 genotoxicity assays; c) a subset of chemicals in b) that were positive using concentrations comparable to the ones in qHTS and negative using concentrations higher than the ones in qHTS (“dose-aware”⁷, BM-only). The a) model was designed for genotoxicity prediction and b, c) models were designed for genotoxicity prioritization.

Chemical prioritization scheme—Chemicals that were active in at least one of the five Tox21 qHTS genotoxicity endpoints were identified and binned based on the results (positive, negative, inconclusive) from the *in silico* BM prediction. For each of the three

categories of results, the chemicals were further binned based on the availability of data from the bacterial mutagenicity (either +S9 or -S9) endpoint and *in vivo* micronucleus endpoint: no data for either of the endpoints (Tier 1), data available for one of the endpoints (Tier 2), or data available for both endpoints (Tier 3). The *in vitro* chromosomal aberration data were not used for this purpose because those data were less decisive than the bacterial mutagenicity and *in vivo* micronucleus data. The genotoxicity models constructed using the Tox21 qHTS assay results (htsBM, htsCA, htsMN) were then used to predict the genotoxic probability of the chemicals for each of the three traditional genotoxicity endpoints. The genotoxic probability for three prediction endpoints (htsBM, htsCA, htsMN) and the logistic wAUC score for five Tox21 qHTS genotoxicity endpoints were imported into the ToxPi software³⁰ (<http://comptox.unc.edu/toxpi.php>, v1.3) to create the pie charts.

Experimental validation—The BM assay and *in vitro* MN assay were used to validate a subset of highly ranked compounds selected from the prioritized lists. These are standard tests with regulatory acceptance that are routinely conducted by NTP for characterizing the genotoxicity of compounds. The details of the protocols are included in the Supplemental Methods. The results of both the BM and *in vitro* MN studies are available in the NTP database (Chemical Effects in Biological Systems: <https://manticore.niehs.nih.gov/cebssearch/>)

Results

qHTS assay performance

Three traditional qHTS assay parameters were used to characterize assay performance (S/B, CV%, or Z' factor, see Table 2). All five Tox21 genotoxicity assays had acceptable or excellent performance for at least one of the three parameters, and overall performance for each assay was good ((S/B >3, CV(%) <10%, Z' factor >0.5, and Cohen's kappa >0.7). In addition, several types of replicates, including the assay positive control (reference compound with defined activity) that was tested using 16 concentrations in duplicate on each plate, duplicate testing of the 88 technical duplicates included on each plate, and triplicate testing of the whole Tox21 library, have been integrated in the Tox21 qHTS assays for quality control. Some empirical assay performance parameters were calculated based on the replicate data, including the standard deviation (SD) of the potency value of the assay positive control on plates, the SD of the potency value of 88 technical duplicates on plates, and Cohen's kappa value for evaluating response hit call concordance between triplicate testing. Good concordance (>0.7) of hit calls and/or potency fold change 2 are seen in these five assays.

Activity overview

The results from the five Tox21 genotoxicity assays were processed as described in Materials and Methods. Active compounds are chemicals that show significant effects for a genotoxicity target of interest; inactive compounds are chemicals that show no effects in an assay; inconclusive compounds are chemicals for which significant effects may not result from action against the target of interest (e.g., assay interference). For the five Tox21 qHTS genotoxicity endpoints, the percentage of each activity call (i.e., active, inactive, and

inconclusive) was calculated based on chemicals that were tested in all five genotoxicity assays (n = 6831) (Figure 2a). Among the five Tox21 qHTS genotoxicity endpoints, 'γH2AX ↑' had the highest percentage of actives (~7.3%) and 'Rev3 ↑' had the lowest (~1.4%). The inconclusive activity calls for the two DT40 cell genotoxicity endpoints ('Rev3 ↑' and 'Ku70/Rad54 ↑') had markedly higher percentages (~35% for each assay) compared with the other three qHTS genotoxicity endpoints (<10%), indicating high response signals but low confirmed activity in these assays.

The reasons underlying the inconclusive calls (i.e., flag type) for each of the five Tox21 qHTS genotoxicity assay endpoints were investigated (Figure 2b). The percentage of flag types for the inconclusive calls (i.e., significant effects with flags) was calculated by including the active calls (i.e., significant effects without flags). In terms of the flag types, more than 70% of the potential active calls in the DT40 cell assays were flagged for 'cytotoxicity'; active calls only accounted for 3.9% ('Rev3 ↑') or 7% ('Ku70/Rad54 ↑') of the potential active calls in these two assays. For both of the two-channel assays ('p53 ↑' and 'γH2AX ↑'), which used a ratio value (the ratio of signal channel to background channel) to quantify the response, the major confounding factor is a flag for 'no signal readout support', which covered 43.8% and 17.7% of the potential active calls in 'p53 ↑' and 'γH2AX ↑', respectively. The flag for 'weak/noisy in repeats', was generally limited to about 15% of the actives, except for the 'ATAD5 ↑' endpoint, where 31.6% of potential active calls had this flag. Among the five genotoxicity assay endpoints, only two ('γH2AX ↑' and 'ATAD5 ↑') had >50% of the potential active calls (active calls + inconclusive calls) free of flags, indicating that careful attention should be paid to assay interference when evaluating Tox21 qHTS genotoxicity data.

Overlapping among active calls in two or more assays was examined using either the Venn Diagram (for overlaps between two or three Tox21 genotoxicity assay endpoints)³¹, or the UpSet (for overlaps among all five Tox21 genotoxicity assay endpoints) plots³² (Figure 3). In Figure 3a, the Tox21 genotoxicity assay endpoints associated with three different types of events were compared, including phosphorylation of ATM (ataxia-telangiectasia mutated)/ATR (ataxia-telangiectasia and Rad3 related) targets ('ATAD5 ↑', 'p53 ↑', and 'γH2AX ↑'), stalled replication fork events ('ATAD5 ↑' and 'Rev3 ↑'), and DNA doubled stranded break events ('γH2AX ↑' and 'Ku70/Rad54 ↑'). Generally, the amount of overlap among active compounds within each target event is low – 4.7% (38/811), 4.3% (13/298), and 8% (50/620) for phosphorylated ATM/ATR targets, stalled replication forks, and doubled stranded DNA breaks, respectively – thus highlighting the uniqueness of each Tox21 genotoxicity assay.

When the overlaps among active calls for all possible combinations among the five Tox21 genotoxicity assays (Figure 3b) were compared, very few chemicals were found to be active in more than three assays, and none were active in all five. This finding underscores the distinct nature of the Tox21 genotoxicity assay endpoints. Dividing the number of unique active calls (i.e., compound was uniquely active in only a single assay out of the five assays) in each Tox21 qHTS genotoxicity assay (the right most five vertical bars) by the total number of active calls identified in the assay (the horizontal bars) reveals that 'γH2AX ↑' had the highest percentage of unique active calls (308/496 = 62.1%), followed by 'Rev3 ↑'

(53/95 = 55.8%), 'Ku70/Rad54 ↑' (91/174 = 52.3%), 'p53 ↑' (148/316 = 46.8%), and 'ATAD5 ↑' (82/216 = 38%). For chemicals that were active in at least one HTS genotoxicity assay (n = 942), >70% of them were active in just one assay (682/942). This demonstrates that multiple assays, using different approaches for detecting DNA damage and different time points, are needed to capture genotoxic activity.

The results from the *in silico* prediction of the BM endpoint were also overlaid on Figure 3b for comparison. Overall, the ratio of positive to negative predicted by the *in silico* BM model is increased and the number of actives in the five qHTS assays is increased (Figure 3b). For chemicals active in only one qHTS assay, the ratio of positive to negative is 0.45 (n = 563). The ratio is increased to 0.76 (n = 145) for chemicals active in two qHTS assays. The ratio is further increased to 1.29 (n = 48) for chemicals that were active in three qHTS assays. And for those chemicals that were active in four qHTS assays, all of them are predicted to be positive (n = 13). Thus, we expect that prediction accuracy (i.e., positive predictive value) may be increased for chemicals that were active in multiple Tox21 qHTS genotoxicity assays.

Retrospective validation of Tox21 genotoxicity assays

The data from four datasets (BM, BM (dose-aware), CA, and MN) were collected and the results were compared with the data from the five Tox21 qHTS genotoxicity assays individually and in combination. Two statistical methods, contingency table and ROC enrichment, were applied to evaluate the retrieval/prioritization performance. A contingency table allows the evaluation of the overall accuracy based on the most confident Tox21 genotoxicity data (inconclusive calls are excluded) (Figure 4a), while ROC enrichment focuses on the prioritizing ability, highlighting the degree of enrichment for genotoxicants at the top portion of the ranking list (Figure 4b).

Contingency table—For each individual Tox21 genotoxicity assay, the average [maximum] sensitivity for identifying known BM, BM (dose-aware), CA, or MN actives was 9.6% [13.1%], 13.7% [17.2%], 9.9% [17.3%], and 12% [17.4%] respectively, meaning <20% of known genotoxicants were captured by any single Tox21 qHTS genotoxicity assay. When combining Tox21 qHTS genotoxicity assay data stepwise using the 'union' method, the sensitivity increased as would be expected (Figure 4a). The sensitivity [upper bound of 95% confidence interval] to the BM, BM (dose-aware), CA, or MN endpoints was 29.5% [34.2%], 40% [50%], 31.7% [37.5%], and 35.2% [44.8%] respectively, when combining all five Tox21 genotoxicity assay endpoints. Thus, when taking the tested concentrations into consideration, a greater proportion of known mutagens was captured with the union of all 5 Tox21 qHTS genotoxicity assays (40% vs 29.5%).

Although combining Tox21 qHTS genotoxicity assay data using the 'union' method increased the sensitivity with the addition of each new dataset, the respective positive prediction values (PPV) dropped accordingly, compared with the PPV for each single Tox21 genotoxicity assay endpoint. In contrast, combining Tox21 qHTS genotoxicity assay endpoints using the 'intersection' method greatly increased the PPV, with an average PPV [maximum PPV] for the BM, BM (dose-aware), CA, or MN endpoints of 77.4% [100%],

77.3% [100%], 74.1% [100%], and 63.6% [100%] respectively, at the cost of decreasing sensitivity. However, it should be emphasized that the calculations based on intersecting actives were based on ~18 chemicals in BM, ~6 in BM (dose-aware), ~10 in CA, and ~7 in MN, and this small number of chemicals resulted in the very wide confidence intervals observed. Also, the NPV values in the BM (dose-aware) and MN endpoints are higher than the ones in BM and CA endpoints, indicating more compounds inactive in qHTS assays turned out to be positive in the BM and CA endpoints using this method. Overall, this analysis demonstrates that the five Tox21 genotoxicity assays cannot be used to replace traditional genotoxicity assays. In the following section, we investigate if these assays can be used for chemical prioritization.

ROC enrichment—By using the wAUC score of chemicals in each individual Tox21 genotoxicity assay to rank chemicals, the average [maximum] fold enrichment when allowing a 1% false positive rate (i.e., considering only the top part of the ranked list, containing 1% of known negatives; the results of using 0.5%, 2%, 5%, or 10% of known negatives are available in Supplemental Data) was 2.4 [3.5], 3.9 [10.0], 3.5 [5], and 5 [6.5] for the BM, BM (dose-aware), CA, or MN endpoints, respectively. Stepwise combination of the Tox21 qHTS genotoxicity assay data generally increased the enrichment for every traditional endpoint except the BM, where enrichment dropped slightly when combining more than three Tox21 qHTS genotoxicity endpoints. This is likely due to the fact that mutagenesis (i.e., BM assay) is a more specific mechanism than other genotoxicity events (e.g., chromosomal damage). Based on the bootstrap statistics (50th percentile), the best enrichment achieved is 5.5, 12.6, 6.4, 12.1, for the BM, BM (dose-aware), CA, or MN endpoints, respectively, suggesting that the ranked list is at least six-fold enriched with genotoxicants compared to random when allowing a 1% false positive rate. The best enrichment was usually achieved when the results from more than three endpoints were combined except for the BM (dose-aware) endpoint, where the best enrichment was achieved when combining ‘Rev3 ↑’ and ‘Ku70/Rad54 ↑’. These results indicate that these Tox21 qHTS genotoxicity assays can be used for chemical prioritization.

Genotoxicity models for prioritization—In the previous analysis (ROC enrichment), we demonstrated that results from the Tox21 genotoxicity assays can be used to prioritize chemicals with unknown genotoxicity potential along a continuum ranging from ‘highly likely’ to ‘very unlikely’ to be genotoxic. Here, we used a more comprehensive set of qHTS assays to see if additional data would be helpful in improving or refining this prioritization for genotoxicity. Datasets with different selections of Tox21 qHTS assay endpoints and chemicals were used to build models for the three traditional genotoxicity endpoints (BM, CA, or MN) using the procedure described in the Material and Methods and Table 3. The performance is plotted in Figure 5a.

Models built using the ‘baseline’ dataset (blue dot in Figure 5a), consisting of all Tox21 data (78 assay endpoints) and all available traditional genotoxicity data (endpoint: [# positive/# negative], BM: [513/653], CA: [343/399], and MN: [124/402]), had the worst performance. The ROC AUC was 0.65, 0.63, and 0.60 for the BM, CA, and MN endpoints, respectively. This result represents the ‘baseline’ performance for using Tox21 data for the prediction of

genotoxicity. The poor performance of this all-inclusive model is not surprising, since in previous analyses (Figure 4a), we learned that at best, only about 35% of known genotoxicants were identified in the Tox21 genotoxicity assays. Therefore, when constructing additional models, we only used data from assays that included chemicals that were active in at least one of the five qHTS genotoxicity assays. In addition, we also varied the number of independent variables, starting with the Tox21 qHTS genotoxicity assays only, and then expanding to include all qHTS stress response pathway related assays, and then all Tox21 qHTS assays (black dot in Figure 5a) (BM: [114/101], CA: [83/61], and MN: [32/69]). This approach resulted in substantial improvement in the performance of the models, except for the CA endpoint. In general, we found that increasing the number of Tox21 qHTS assay calls used for model building increases model performance, but also increases model complexity. The best achievable AUC was 0.74, 0.65, and 0.76, for the BM, CA, and MN endpoints, respectively. Additionally, for the BM endpoint, we could build models using a refined dataset (BM: [40/36]) that took into consideration the concentration ranges of both the traditional BM assay and the Tox21 qHTS genotoxicity assays. For these models, we included only cases where positive calls in the BM assay were observed with doses that were comparable to the concentration ranges used in the qHTS assays and cases where negative calls in the BM assay were reported at concentrations that exceeded the range used in the qHTS assays. When dose ranges were considered in this way, the highest achievable AUC was about 0.83 (red dot in Figure 5a). The constructed models using all Tox21 assay results demonstrated acceptable accuracy that can be applied for the genotoxicity prioritization.

The coefficients of variables (i.e., Tox21 qHTS assay endpoints) in the models were collected as described in Materials and Methods. The maximum and minimum coefficient value was 0.66 and -0.41. To focus on relatively more informative variables when plotting, the list of the variables was further filtered (based on arbitrary cutoffs) to ensure the coefficients had an absolute value larger than 0.3 and were generated based on results from >20 models as well as >10 active calls in the modeling sets. The results based on the filtered list were plotted in Figure 5b. The complete list of variables with coefficients can be found in the Supplemental Data.

All but 'ATAD5 ↑' of the five Tox21 genotoxicity assay endpoints were found to have positive coefficient values in one of the traditional genotoxicity endpoints. The 'ATAD5 ↑' call was found to have a contribution close to 0 for BM and MN, and 0.15 for CA. For BM, the 'Rev3 ↑' call had the highest positive contribution (0.66) followed by two other calls related to activation of stress response pathways ('Nrf2 ↑' & 'heat shock factor (HSF) ↑'), with coefficient values of 0.45 and 0.44. The 'mitochondrial membrane potential (MMP) ↓' was found to have negative contribution (-0.35) to the BM endpoint, indicating that chemicals that disrupt MMP may be less likely to be genotoxic. For the CA endpoint, only the 'γH2AX ↑' call passed the filter with a positive value of 0.39. For MN, all but 'ATAD5 ↑' of Tox21 genotoxicity assay endpoints were found to have positive coefficient values, where 'p53 ↑' has the highest value (0.54). Interestingly, activation of constitutive androstane receptor (CAR) transcriptional factor or 'NR1I3 ↑' was found to have a negative contribution to MN (-0.41) and 'NR1I3 ↓' has a positive contribution (0.30). These two findings are complementary with each other. One of the cytotoxicity counter screens present

in the Tox21 qHTS assays (i.e., the cell viability assay multiplexed with PPAR δ agonist-mode assay) was found to have a negative contribution to the MN endpoint (−0.34). Despite this case, overall the most positive contributing factors in predicting genotoxicity remained the Tox21 genotoxicity assays. Other stress response pathway assays contributed in predicting BM and particularly the CAR assay was found to contribute to the prediction of MN.

Chemical prioritization—In previous analyses, the prioritization capacity of the Tox21 genotoxicity assays was validated and predictive models for genotoxicity based on all Tox21 qHTS assays were constructed. We then combined the outputs from these two analyses to prioritize chemicals with unknown genotoxicity potential. The ToxPi software was selected for visualization and the logistic wAUC score [0-1] from each of the five Tox21 qHTS genotoxicity assay endpoints and the probability from each of the three predictive models were used as the input. Furthermore, the QSAR model for *in silico* BM is known to have a high prediction accuracy⁴. Thus, we included the QSAR information to help categorize the chemicals. We wanted to explore, for example, whether it might be more worthwhile to test chemicals whose genotoxicity could not be predicted based on the QSAR model (out of domain) or, if it would be more useful to test chemicals that were predicted by QSAR to be negative but that were ranked high in the ToxPi activity chart.

The flowchart in Figure 6 shows this scheme and includes the number of chemicals in each category. In total, there were 1082 chemicals active in at least one of the genotoxicity endpoints: 313 with a positive label, 565 with a negative label, and 204 with an inconclusive label based on QSAR model predictions. For each of the categories, the chemicals were further binned based on the availability of data for the bacterial mutagenicity (either +S9 or -S9) endpoint and the *in vivo* micronucleus endpoint: no data for either endpoint (Tier 1), data for only one endpoint (Tier 2), or data for both endpoints (Tier 3). For Tier 1, there are 148 with a positive label, 373 with a negative label, and 147 with an inconclusive label based on *in silico* BM QSAR model predictions. The ToxPi values for the top 10 chemicals in each bin (Tier 1) are shown in Figure 7. The spreadsheets for creating the pie charts using ToxPi software and the background data for the 1082 chemicals used for the QSAR model predictions can be found in Supplemental ToxPi files and Supplemental Data.

Despite the absence of genotoxicity data in commercial databases for the prioritized compounds, by searching through research publications and package inserts for drugs, genotoxicity data were found for some of the chemicals. Of the top 10 compounds predicted by the *in silico* BM model to be genotoxic, 9 were reported to have demonstrated genotoxic activity and for 1 (TNP-470), no data are available (Table 4). Of the 10 compounds predicted to be negative by the *in silico* BM model, 7 have demonstrated evidence of genotoxicity in the *in vivo* MN assay and one compound (dipentaerythritol pentaacrylate) was judged to be equivocal in the *in vitro* MN assay (Supplemental Table 1); only one (2-chloro-N-(2-methyl-4-bromophenyl)acetamide) of the 10 has been shown to be positive in the BM assay, and that positive result was seen only in the presence of rat liver S9. Vinblastine sulfate, a genotoxic compound, was correctly predicted to be negative by the BM model because it is an aneugen (microtubule disrupting agent); it was active in the ATAD5-luc, p53RE-bla, htsCA, and htsMN assays. Thus, these results provide evidence for the usefulness of the *in*

silico BM model for binning the prioritized compounds by mechanism of action: compounds positive in the MN assay but negative in the *in silico* BM model may induce chromosomal damage (clastogenicity or aneugenicity) but not point mutations. Thus, results are consistent with both the positive predictivity of the qHTS DNA damage assays as well as the BM prediction model. Of the 10 compounds found to be inconclusive by the *in silico* BM model, 6 have demonstrated genotoxicity and 1 (parbendazole) may be found to induce MN, based on the chemical class to which it belongs (tubulin binding agents, closely related to oxibendazole). The two organotins in this group may prove to be genotoxic if they gain entry into the cell³³ (Supplemental Table 2).

To summarize these results, of these 30 Tier 1 compounds, 22 are demonstrated genotoxicants, 1 is equivocal based on available data, and 5 compounds may be genotoxic based on chemical class (2 acrylates, 2 organotins, and 1 benzimidazole). Of the 22 demonstrated genotoxicants, 14 were active in the γ H2AX qHTS assay. Of the 8 compounds that were genotoxic but inactive in γ H2AX, 7 were active in htsCA and the p53RE and/or ATAD5-luc assays. For example, SR271425 is a thioxanthone analog (thioxanthone is positive in bacterial mutation assays, is a DNA alkylator and cross-linker, and is also active in the REV3 assay); SAHA (Vorinostat), positive in the BM, CA, and MN assays as per the manufacturer's package insert (Merck & Co, Inc.), is an HDAC inhibitor, and was active in the ATAD5-luc, p53RE-bla, and htsCA assays; oxibendazole is a tubulin-binding agent that was active in the p53RE-bla, ATAD5-luc, htsCA, and htsMN assays. Interestingly, the two organotins and the two acrylates, which we speculated might show genotoxic activity under appropriate test conditions, were active in the γ H2AX assay. This analysis demonstrates that, although we did not find evidence of genotoxicity in the commercial databases that we used as our source, most of the top ranked Tier 1 compounds are highly likely to be genotoxic, indicating that we can prioritize other top ranked compounds that lack searchable genotoxicity data for further testing.

Prospective validation of Tox21 genotoxicity assays—To investigate the genotoxicity potential of some “true unknowns”, seven top ranked chemicals (Table 5, four of them were also listed in Supplemental Table 1 & 2) without existing genotoxicity data were obtained for experimental confirmation in traditional BM and *in vitro* MN tests. All seven compounds were active in at least two of the Tox21 qHTS assays that measured activity indicative of chromosomal damage: γ H2AX, ATAD5-luc, p53RE-bla, and DT40-DSB (*KU70*^{-/-}/*RAD54*^{-/-}). Six of the seven compounds induced MN in human lymphoblastoid TK6 cells *in vitro* in the absence of S9, while one additional compound gave responses judged to be equivocal, based on the magnitude of the response. Three of the seven compounds (2-chloro-N-(2-methyl-4-bromophenyl)acetamide, 2-(thiocyanomethylthio)-benzothiazole, and 1H-isoindole-1,3(2H)-diimine) were positive in bacterial mutagenicity tests, although interestingly, two of the compounds (2-chloro-N-(2-methyl-4-bromophenyl)acetamide and 1H-isoindole-1,3(2H)-diimine) required S9 for mutagenic activity in the bacterial tester strains. All of the compounds that induced MN in TK6 cells *in vitro* were active in the γ H2AX qHTS assay and either the p53RE-bla or ATAD5-luc qHTS assays, or both. All inducers of MN *in vitro* were active in three qHTS genotoxicity assays, indicating that activity in multiple qHTS DNA damage assays increases

the likelihood that the compound will be shown to be genotoxic in a traditional assay. Only 2 of the 7 compounds (potassium dicyanoaurate and 2-chloro-N-(2-methyl-4-bromophenyl)acetamide) were detected in the DT40-DSB (*KU70*^{-/-}/*RAD54*^{-/-}) assay, and one of those (2-chloro-N-(2-methyl-4-bromophenyl)acetamide) was positive in the Ames assay in the presence of S9, but not without S9. The three Ames positive compounds (2-chloro-N-(2-methyl-4-bromophenyl)acetamide, 2-(thiocyanomethylthio)-benzothiazole, and 1H-isoindole-1,3(2H)-diimine) were each active in a different bacterial strain with a unique genetic target. The positive findings from the experimental testing results demonstrate that Tox21 genotoxicity assays can not only be used for prioritizing chemicals for genotoxicity potential, but also provide insights into genotoxicity mechanisms.

Discussion

In this study, we presented and analyzed the data from five Tox21 qHTS genotoxicity assays: ATAD5-luc, p53RE-bla, γ H2AX, and two DT40 cell knockouts (*REV3*^{-/-} and *KU70*^{-/-}/*RAD54*^{-/-}). The results of each of the five assays were independently and in combinations of 2, 3, 4, and 5 assays compared to traditional genotoxicity tests (BM, CA, MN). Only genotoxicants that were active without metabolic activation and at test doses comparable to concentrations used in the qHTS assays (based on the BM data, only) were used in the comparison. However, despite adjusting for dose and activation requirements, sensitivity of the qHTS assays for known genotoxicants included in the Tox21 library was at best around 40%. A number of additional factors may have played a role in this low sensitivity, including, e.g., differences in duration of exposure (e.g., 3 - 16 h for the mammalian cell DNA damage assays), marked differences in the measured endpoint (mutation or chromosomal damage versus reporter gene activity), methods of read-out or data collection (e.g., microscopic examination, colony counting, fluorescent or luminescent signal), the relatively small number of cells in each well of a 1536-well plate (<3,000) that may have provided a limited number of targets for inducing measurable effects, and cell type in which the assay is conducted (variable DNA repair capabilities). A controlled variable study is needed to further investigate these additional data factors. Because the genotoxicity classification of a compound is currently based on the results of traditional genotoxicity assays (bacterial mutation, *in vitro* chromosomal damage, and *in vivo* micronucleus assays, for example), we chose to use these three traditional assays, for which sizable databases are available, as the “gold standard” benchmark for genotoxicity, despite their lower throughput format and generally higher test article concentrations. Whether conducting these five qHTS assays in lower throughput (e.g., 96-well plate format) and with higher compound concentrations would alter assay sensitivity is an open question, but one that deserves further exploration.

To help to explore the impact of the dose differential between the qHTS assays and traditional BM assays, and to assist in future similar investigations, we identified a set of 25 “common mutagens” (or consensus mutagens) that were consistently positive across traditional genotoxicity studies in the absence of S9 and that had a lowest effective concentration <100 μ M in the BM assay⁷. We think that this collection of compounds might serve as a useful reference set for researchers in defining the sensitivity of novel, new high throughput *in vitro* approaches for identifying genotoxic compounds in large compound

libraries. Twenty of these 25 consensus mutagens had defined calls in the qHTS assays (the remaining five were labeled inconclusive); 11 of the 20 were active in at least one of the five Tox21 genotoxicity assays and 9 were inactive in all 5 assays (Supplemental Table 3). The nine mutagens that were inactive in all 5 Tox21 genotoxicity assays were epichlorohydrin, 2,4-dinitrotoluene, sodium azide (a positive control for the BM assay), metronidazole, glycidol, 3-methylcholanthrene, methyl methanesulfonate, 1-ethyl-1-nitrosourea (a positive control in a variety of traditional genetic toxicity tests), and dimethylcarbamoyl chloride. Consistent with the negative results in the qHTS assays for DNA damage, both methyl methanesulfonate and glycidol were found to be negative when tested at a concentration of 100 μ M in a high throughput *in vitro* comet assay designed to detect DNA damage (CometChip®); however, 1-ethyl-1-nitrosourea was positive in this same assay³⁴. Additional orthogonal tests (e.g., MultiFlow® DNA Damage assay, Litron Laboratories) are planned to further characterize the modes of action and activity profiles of the nine compounds inactive in all qHTS genotoxicity assays. Furthermore, we are exploring methods to use rat and human liver S9 preparations, in non-toxic concentrations, to improve the metabolic competency of the cells used in the five DNA damage assays.

Although the current Tox21 genotoxicity assays cannot be used to predict the outcomes of traditional genotoxicity assays (-S9) and thus, cannot replace them, we demonstrated that they have utility in prioritizing chemicals with unknown genotoxicity potential for confirmatory testing. This ability to prioritize compounds for testing allows resources to be focused initially on those compounds most likely to represent a hazard. In addition, by building elastic net regularization models using all Tox21 qHTS assays for prediction of traditional genotoxicity endpoints (BM, CA, and MN), we found that, in general, including the results from certain additional Tox21 qHTS assays improved the model performance compared with models using the Tox21 genotoxicity assays alone. The qHTS assays identified as the top contributors to the models were consistent with expectation. For example, 'Rev3 $\uparrow$ ', targeting stalled replication fork damage and related to the BM endpoint, was found to be the top positive contributor in the htsBM model. Other stress response pathways (HSF and Nrf2) were also found to be the top contributors in the htsBM model. Similarly, ' γ H2AX $\uparrow$ ', targeting DNA doubled-strand breaks and related to the CA endpoint, was found to be the top positive contributor in the htsCA model. This observation is consistent with the recent finding that despite differing modes of action of clastogens, all of them showed a strong correlation with phosphorylation of H2AX in HepG2 cells³⁵. For MN, a more complex endpoint, all except 'ATAD5 $\uparrow$ ' were considered top positive contributors to the htsMN model. The other identified positive contributor to the model, CAR transcription factor, remains to be explored³⁶. In addition, we cannot offer an explanation for the high negative contribution to the MN endpoint of the counter-screen viability data in the PPAR δ agonist-mode assay. To confirm the results of the initial analysis for all endpoints, the analysis was repeated using a different random seed; all the assays initially identified as the highest positive and negative contributors to the models were confirmed.

The high inconclusive rate and the low active rate for the 2 DT40 knockout cell assays ('Rev3 $\uparrow$ ' and 'Ku70/Rad54 $\uparrow$ ' endpoints), coupled with their lower QC values (but still meeting the qHTS standards) provide evidence of the reduced influence of the DT40 knockout assays within our overall process of identifying and prioritizing compounds for

genotoxicity potential. However, despite these considerations, the two qHTS genotoxicity endpoints generated from these DT40 knockout cell assays have a high positive contribution in the models for predicting traditional genotoxicity (i.e., low sensitivity but high predictivity). This pattern of activity is seen with a number of traditional genotoxicity assays (e.g., *in vivo* micronucleus and *Drosophila* sex-linked recessive lethal mutation assays) that are also characterized by low sensitivity but high positive predictivity for rodent carcinogenicity^{37,38}. A high level of positive predictivity can be valuable in a prioritization scheme but is not particularly useful in screening large compound libraries for a specific activity when accompanied by a low hit rate. Thus, assay performance characteristics with regard to reference compounds (sensitivity, predictivity) need to be carefully considered in determining appropriate applications.

In contrast to the pattern of activity seen with the DT40 assays, the 'ATAD5 ↑' assay endpoint, despite its relatively low inconclusive rate and medium level active rate, does not provide a significant contribution to the models for predicting traditional genotoxicity. Perhaps related to this reduced contribution of ATAD5-luc assay to the models, we discovered, while investigating the actives with high wAUC values, that some portion of the ATAD5-luc active compounds (e.g., resveratrol, bortezomib [<https://druginfo.nlm.nih.gov/drugportal/rn/179324-69-7>], carfilzomib [<https://druginfo.nlm.nih.gov/drugportal/rn/868540-17-4>]) were protease inhibitors, not inducers of DNA damage. Protease inhibitors may cause an increase in the ATAD5-luc signal, suggesting upregulation of ATAD5-luc levels, by interfering with the normal breakdown and elimination of the ATAD5-luc protein as part of the cellular feedback pathway for maintaining low levels of the protein in the absence of DNA damage³⁹. Given this finding, and the surprisingly low contribution of the assay to the models, it is possible that there are additional artifacts in this assay that cannot be excluded using the current standard data analysis pipeline.

The p53RE-bla assay ('p53 ↑') and γH2AX assay ('γH2AX ↑'), both of which are commercially available assays, appear to be more suitable for routine qHTS screening due to their relatively higher active and lower inconclusive rates, as well as their good QC scores. Further, our results suggest that compounds showing activity in both of these assays (intersection of results) have a high probability of being active in traditional genotoxicity assays (Table 4; Supplemental Tables 1, 2). And, in fact, the value in assessing activity against these two endpoints in combination has been recognized by others, notably the developers of the MultiFlow® DNA Damage assay that provides information on genotoxicity potential and mechanism of action for compounds that induce micronuclei, based on data integration from multiple biomarkers, including p53 and γH2AX⁴⁰. A subset of Tox21 compounds selected for activity patterns in the p53RE-bla and γH2AX qHTS assays were tested in the MultiFlow assay and results compared well with the results of the two Tox21 screens⁴⁰. However, there are still acknowledged limitations with these two qHTS assays. For example, we previously reported that chemicals that, based on traditional genetic toxicity test data, would be expected to activate the p53 transcriptional factor signaling pathway did not show activity in the p53RE-bla assay⁷. Thus, sensitivity to known genotoxicants is low for these two assays, also, in the high throughput format (1536-well plate) that was used in the Tox21 Program.

By using the qHTS genotoxicity assay results, models based on qHTS data, and bacterial mutagenicity QSAR model predictions, we prioritized Tox21 chemicals for genotoxicity characterization. As a retrospective validation, the top ranked chemicals without genotoxicity information in our databases were further searched for existing genotoxicity data (e.g., published literature) and we found that >60% of these chemicals have been identified as genotoxicants. Although this result demonstrates the usefulness of qHTS data for genotoxicity prioritization, it also suggests that genotoxicity data for certain Tox21 chemicals, particularly pharmaceutical agents, are not available in the current databases. Nevertheless, seven top ranked “true unknowns” were selected for assessing their activity in traditional BM and *in vitro* MN tests. All seven compounds had shown activity in at least two of the Tox21 genotoxicity assays that measured responses associated with chromosomal damage events: γ H2AX, ATAD5-luc, and p53RE-bla. Therefore, we hypothesized that these seven compounds, if they demonstrated activity in the traditional assays, would most likely show activity in the *in vitro* MN assay (because MN are biomarkers of chromosomal damage, and γ H2AX and ATAD5-luc are endpoints directly associated with chromosomal damage). Results supported this hypothesis (Table 5). Additional follow-up studies designed to investigate mode of action of these seven compounds along with comprehensive genotoxicity test data from other top-ranked compounds in the priority listing generated from the models described herein, will be the subject of a future manuscript.

We acknowledge that the data reported herein reveal significant gaps in the coverage provided by this set of five qHTS DNA damage assays, based on their low sensitivity for detecting known genotoxicants. Therefore, we encourage the development of novel, new *in vitro* screening methods to broaden the capability to detect genotoxic compounds in large compound libraries in order to help in reducing the backlog of uncharacterized compounds. For our part, we will continue work to better define the combinations of *in vitro* qHTS genotoxicity endpoints and activity potencies that are most predictive of the outcomes of traditional assays with regulatory acceptance. Further refinement of these Tox21 assays coupled with orthogonal testing using closely aligned analytical tools such as the TGx-DDI biomarker⁴¹ and the MultiFlow™ DNA Damage⁴⁰ assay are planned to better understand the role these qHTS assays, or the endpoints they measure, have for characterizing data-poor compounds. As one example of refinement of qHTS assay protocols, we are exploring methods to use rat and human liver S9 preparations, in non-toxic concentrations, to improve the metabolic competency of the cells used in these Tox21 assays when they are conducted using a 1536-well plate format. Additionally, we plan to investigate the effect of using a lower throughput format (e.g., 96-well plate) on assay sensitivity, given the realization that few laboratories conduct assays using a 1536-well plate format on a robotic platform. Finally, after the completion of additional confirmation studies for compounds identified as potential genotoxicants, our predictive models will be updated, and we hope to offer more definitive recommendations for using assays based on the endpoints measured in these five qHTS assays. In the future, high throughput toxicogenomic signatures of qHTS actives may provide additional confirmation of the ability of these screens to identify true genotoxicants, and provide support for conducting lower throughput versions of these assays as a part of a routine genotoxicity testing scheme to characterize compounds of unknown potential.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank MultiCASE and Leadscope for graciously making available to us their bacterial mutagenicity predictive models and SAR database, Nisha Sipes and Keith Houck for critical review of the manuscript, and Richard Paules for his encouragement throughout this process.

Funding

National Institute of Environmental Health Sciences/Division of the National Toxicology Program

NIH Center for Advancing Translational Sciences, National Institutes of Health. Interagency Agreement: NTR12003-001-0100

References

- (1). Turkez H; Arslan ME; Ozdemir O Genotoxicity Testing: Progress and Prospects for the next Decade. *Expert Opin. Drug Metab. Toxicol* 2017, 13 (10), 1089–1098. 10.1080/17425255.2017.1375097. [PubMed: 28889778]
- (2). Kirkland D; Kasper P; Martus H-J; Müller L; van Benthem J; Madia F; Corvi R Updated Recommended Lists of Genotoxic and Non-Genotoxic Chemicals for Assessment of the Performance of New or Improved Genotoxicity Tests. *Mutat. Res. Toxicol. Environ. Mutagen* 2016, 795, 7–30. 10.1016/j.mrgentox.2015.10.006.
- (3). Chemicals Management Plan. Government of Canada.
- (4). Honma M; Kitazawa A; Cayley A; Williams RV; Barber C; Hanser T; Saiakhov R; Chakravarti S; Myatt GJ; Cross KP; et al. Improvement of Quantitative Structure-Activity Relationship (QSAR) Tools for Predicting Ames Mutagenicity: Outcomes of the Ames/QSAR International Challenge Project. *Mutagenesis* 2018 10.1093/mutage/gy031.
- (5). Kavlock RJ; Austin CP; Tice RR Toxicity Testing in the 21st Century: Implications for Human Health Risk Assessment. *Risk Anal. Off. Publ. Soc. Risk Anal* 2009, 29 (4), 485–487; discussion 492–497. 10.1111/j.1539-6924.2008.01168.x.
- (6). Attene-Ramos MS; Miller N; Huang R; Michael S; Itkin M; Kavlock RJ; Austin CP; Shinn P; Simeonov A; Tice RR; et al. The Tox21 Robotic Platform for the Assessment of Environmental Chemicals – from Vision to Reality. *Drug Discov. Today* 2013, 18 (15), 716–723. 10.1016/j.drudis.2013.05.015. [PubMed: 23732176]
- (7). Witt KL; Hsieh J-H; Smith-Roe SL; Xia M; Huang R; Zhao J; Auerbach SS; Hur J; Tice RR Assessment of the DNA Damaging Potential of Environmental Chemicals Using a Quantitative High-Throughput Screening Approach to Measure P53 Activation. *Environ. Mol. Mutagen* 2017, 58 (7), 494–507. 10.1002/em.22112. [PubMed: 28714573]
- (8). Fox JT; Sakamuru S; Huang R; Teneva N; Simmons SO; Xia M; Tice RR; Austin CP; Myung K High-Throughput Genotoxicity Assay Identifies Antioxidants as Inducers of DNA Damage Response and Cell Death. *Proc. Natl. Acad. Sci* 2012, 109 (14), 5423–5428. 10.1073/pnas.1114278109. [PubMed: 22431602]
- (9). Yamamoto KN; Hirota K; Kono K; Takeda S; Sakamuru S; Xia M; Huang R; Austin CP; Witt KL; Tice RR Characterization of Environmental Chemicals with Potential for DNA Damage Using Isogenic DNA Repair-Deficient Chicken DT40 Cell Lines. *Environ. Mol. Mutagen* 2011, 52 (7), 547–561. 10.1002/em.20656. [PubMed: 21538559]
- (10). Nishihara K; Huang R; Zhao J; Shahane SA; Witt KL; Smith-Roe SL; Tice RR; Takeda S; Xia M Identification of Genotoxic Compounds Using Isogenic DNA Repair Deficient DT40 Cell Lines on a Quantitative High Throughput Screening Platform. *Mutagenesis* 2016, 31 (1), 69–81. 10.1093/mutage/gev055. [PubMed: 26243743]

- (11). Judson RS; Houck KA; Watt ED; Thomas RS On Selecting a Minimal Set of in Vitro Assays to Reliably Determine Estrogen Agonist Activity. *Regul. Toxicol. Pharmacol* 2017, 91, 39–49. 10.1016/j.yrtph.2017.09.022. [PubMed: 28993267]
- (12). Zhang J; Hsieh J-H; Zhu H Profiling Animal Toxicants by Automatically Mining Public Bioassay Data: A Big Data Approach for Computational Toxicology. *PLOS ONE* 2014, 9 (6), e99863 10.1371/journal.pone.0099863. [PubMed: 24950175]
- (13). Kim MT; Huang R; Sedykh A; Wang W; Xia M; Zhu H Mechanism Profiling of Hepatotoxicity Caused by Oxidative Stress Using Antioxidant Response Element Reporter Gene Assay Models and Big Data. *Environ. Health Perspect* 2016, 124 (5), 634–641. 10.1289/ehp.1509763. [PubMed: 26383846]
- (14). Okada T; Sonoda E; Yoshimura M; Kawano Y; Saya H; Kohzaki M; Takeda S Multiple Roles of Vertebrate REV Genes in DNA Repair and Recombination. *Mol. Cell. Biol* 2005, 25 (14), 6103–6111. 10.1128/MCB.25.14.6103-6111.2005. [PubMed: 15988022]
- (15). Wei C; Skopp R; Takata M; Takeda S; Price CM Effects of Double-Strand Break Repair Proteins on Vertebrate Telomere Structure. *Nucleic Acids Res.* 2002, 30 (13), 2862–2870. 10.1093/nar/gkf396. [PubMed: 12087170]
- (16). Winding P; Berchtold MW The Chicken B Cell Line DT40: A Novel Tool for Gene Disruption Experiments. *J. Immunol. Methods* 2001, 249 (1), 1–16. 10.1016/S0022-1759(00)00333-1. [PubMed: 11226459]
- (17). Corton C; Witt KL; Yauk C Identification of P53 Activators in a Human Microarray Compendium.
- (18). Huang R; Sakamuru S; Martin MT; Reif DM; Judson RS; Houck KA; Casey W; Hsieh J-H; Shockley KR; Ceger P; et al. Profiling of the Tox21 10K Compound Library for Agonists and Antagonists of the Estrogen Receptor Alpha Signaling Pathway. *Sci. Rep* 2014, 4 10.1038/srep05664.
- (19). Hsieh J-H; Sedykh A; Huang R; Xia M; Tice RR A Data Analysis Pipeline Accounting for Artifacts in Tox21 Quantitative High-Throughput Screening Assays. *J. Biomol. Screen* 2015, 1087057115581317. 10.1177/1087057115581317.
- (20). Wang Y; Huang R Correction of Microplate Data from High-Throughput Screening In High-Throughput Screening Assays in Toxicology; Zhu H, Xia M, Eds.; *Methods in Molecular Biology*; Springer New York, 2016; pp 123–134. 10.1007/978-1-4939-6346-1_13.
- (21). Sedykh A; Zhu H; Tang H; Zhang L; Richard A; Rusyn I; Tropsha A Use of in Vitro HTS-Derived Concentration–Response Data as Biological Descriptors Improves the Accuracy of QSAR Models of in Vivo Toxicity. *Environ. Health Perspect.* 2011, 119 (3), 364–370. 10.1289/ehp.1002476. [PubMed: 20980217]
- (22). Sedykh A CurveP Method for Rendering High-Throughput Screening Dose-Response Data into Digital Fingerprints In High-Throughput Screening Assays in Toxicology; Zhu H, Xia M, Eds.; *Methods in Molecular Biology*; Springer New York: New York, NY, 2016; pp 135–141. 10.1007/978-1-4939-6346-1_14.
- (23). Hsieh J-H Accounting Artifacts in High-Throughput Toxicity Assays In High-Throughput Screening Assays in Toxicology; Zhu H, Xia M, Eds.; *Methods in Molecular Biology*; Springer New York, 2016; pp 143–152. 10.1007/978-1-4939-6346-1_15.
- (24). Cauty A; Ripley B Boot: Bootstrap R (S-Plus) Functions; 2017.
- (25). Jain AN; Nicholls A Recommendations for Evaluation of Computational Methods. *J. Comput. Aided Mol. Des* 2008, 22 (3–4), 133–139. 10.1007/s10822-008-9196-5. [PubMed: 18338228]
- (26). Hsieh J-H; Sedykh A; Huang R; Xia M; Tice RR A Data Analysis Pipeline Accounting for Artifacts in Tox21 Quantitative High-Throughput Screening Assays. *J. Biomol. Screen* 2015, 1087057115581317 10.1177/1087057115581317.
- (27). Robin X; Turck N; Hainard A; Tiberti N; Lisacek F; Sanchez J-C; Müller M PROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 2011, 12 (1), 77 10.1186/1471-2105-12-77. [PubMed: 21414208]
- (28). Friedman J; Hastie T; Tibshirani R Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw* 2010, 33 (1), 1–22. [PubMed: 20808728]

- (29). Zou H; Hastie T Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol* 2005, 67 (2), 301–320.
- (30). Reif DM; Martin MT; Tan SW; Houck KA; Judson RS; Richard AM; Knudsen TB; Dix DJ; Kavlock RJ Endocrine Profiling and Prioritization of Environmental Chemicals Using ToxCast Data. *Environ. Health Perspect* 2010, 118 (12), 1714–1720. 10.1289/ehp.1002180. [PubMed: 20826373]
- (31). Hulsen T; de Vlieg J; Alkema W BioVenn – a Web Application for the Comparison and Visualization of Biological Lists Using Area-Proportional Venn Diagrams. *BMC Genomics* 2008, 9, 488 10.1186/1471-2164-9-488. [PubMed: 18925949]
- (32). Conway JR; Lex A; Gehlenborg N UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties. *Bioinformatics* 2017, 33 (18), 2938–2940. 10.1093/bioinformatics/btx364. [PubMed: 28645171]
- (33). Dopp E; Hartmann LM; von Recklinghausen U; Florea AM; Rabieh S; Shokouhi B; Hirner AV; Obe G; Rettenmeier AW The Cyto- and Genotoxicity of Organotin Compounds Is Dependent on the Cellular Uptake Capability. *Toxicology* 2007, 232 (3), 226–234. 10.1016/j.tox.2007.01.014. [PubMed: 17316952]
- (34). Sykora P; Chiari Y; Heaton A; Moreno N; Glaberman S; Sobol RW Application of the CometChip Platform to Assess DNA Damage in Field-Collected Blood Samples from Turtles. *Environ. Mol. Mutagen* 2018, 59 (4), 322–333. 10.1002/em.22183. [PubMed: 29536573]
- (35). Kopp B; Dario M; Zalko D; Audebert M Assessment of a Panel of Cellular Biomarkers and the Kinetics of Their Induction in Comparing Genotoxic Modes of Action in HepG2 Cells. *Environ. Mol. Mutagen* 2018, 59 (6), 516–528. 10.1002/em.22197. [PubMed: 29668064]
- (36). Kamino H; Negishi M The Nuclear Receptor Constitutive Active/Androstane Receptor Arrests DNA-Damaged Human Hepatocellular Carcinoma Huh7 Cells at the G2/M Phase. *Mol. Carcinog* 2012, 51 (2), 206–212. 10.1002/mc.20783. [PubMed: 21557330]
- (37). Witt KL; Knapton A; Wehr CM; Hook GJ; Mirsalis J; Shelby MD; MacGregor JT Micronucleated Erythrocyte Frequency in Peripheral Blood of B6C3F(1) Mice from Short-Term, Prechronic, and Chronic Studies of the NTP Carcinogenesis Bioassay Program. *Environ. Mol. Mutagen* 2000, 36 (3), 163–194. [PubMed: 11044899]
- (38). Foureman P; Mason JM; Valencia R; Zimmering S Chemical Mutagenesis Testing in *Drosophila*. X. Results of 70 Coded Chemicals Tested for the National Toxicology Program. *Environ. Mol. Mutagen* 1994, 23 (3), 208–227. [PubMed: 8162896]
- (39). Sikdar N; Banerjee S; Lee K; Wincovitch S; Pak E; Nakanishi K; Jasin M; Dutra A; Myung K DNA Damage Responses by Human ELG1 in S Phase Are Important to Maintain Genomic Integrity. *Cell Cycle Georget. Tex* 2009, 8 (19), 3199–3207. 10.4161/cc.8.19.9752.
- (40). Bryce SM; Bernacki DT; Smith-Roe SL; Witt KL; Bemis JC; Dertinger SD Investigating the Generalizability of the MultiFlow ® DNA Damage Assay and Several Companion Machine Learning Models With a Set of 103 Diverse Test Chemicals. *Toxicol. Sci* 2018, 162 (1), 146–166. 10.1093/toxsci/kfx235. [PubMed: 29106658]
- (41). Corton JC; Williams A; Yauk CL Using a Gene Expression Biomarker to Identify DNA Damage-Inducing Agents in Microarray Profiles. *Environ. Mol. Mutagen* 2018, 59 (9), 772–784. 10.1002/em.22243. [PubMed: 30329178]
- (42). Naoe Y; Inami M; Matsumoto S; Nishigaki F; Tsujimoto S; Kawamura I; Miyayasu K; Manda T; Shimomura K FK317: A Novel Substituted Dihydrobenzoxazine with Potent Antitumor Activity Which Does Not Induce Vascular Leak Syndrome. *Cancer Chemother. Pharmacol* 1998, 42 (1), 31–36. 10.1007/s002800050781. [PubMed: 9619755]
- (43). McCann J; Choi E; Yamasaki E; Ames BN Detection of Carcinogens as Mutagens in the Salmonella/Microsome Test: Assay of 300 Chemicals. *Proc. Natl. Acad. Sci. U. S. A* 1975, 72 (12), 5135–5139. [PubMed: 1061098]
- (44). Lockhart AC; Calvo E; Tolcher AW; Rowinsky EK; Shackleton G; Morrison JG; Rafi R; Vermeulen W; Rothenberg ML A Phase I Dose-Escalation Study of Sr271425, an Intravenously Dosed Thioxanthone Analog, Administered Weekly in Patients With Refractory Solid Tumors. *Am. J. Clin. Oncol* 2009, 32 (1), 9–14. 10.1097/COC.0b013e318178331b. [PubMed: 19194116]

Assay characteristics:

- 1536-well plate format, robotic platform, homogeneous assays (add, mix, measure)
- ~5 μ L total assay volume/well; all DMSO soluble compounds
- 15-point concentration range, from 5 nM to 92 μ M
- 1000 – 3000 cells/well

Figure 1:
DNA repair pathway coverage provided by the five Tox21 genotoxicity assays

Figure 2: Activity type and flag type distribution for the five Tox21 genotoxicity assay endpoints. a) Active, inactive, and inconclusive rate percentage. b) Flag type percentage.

Figure 3:

Overlap among active compounds for each Tox21 genotoxicity assay endpoint. a) Degree of overlap based on grouping by similar molecular targets of functions. b) Degree of overlap among varying activity patterns among all 5 assays; black dots denote activity pattern across assays, the height of the red bar/gray triangle represents the number of positive/negative predictions by the *in silico* BM model, and the width of the black bar on the left side of the figure (with yellow text) represents the number of active chemicals in each of the 5 assays.

Figure 4:

Predictivity of Tox21 genotoxicity assay endpoints for traditional genotoxicity endpoints. a) Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) based on contingency tables; the blue/magenta colors represent the 50th percentile from the bootstrap statistics and the gray color represents the 95% of confidence interval (dot: average; range: maximum and minimum). b) ROC enrichment allowing a 1% false positive rate; the black color presents the 50th percentile from the bootstrap statistics and the gray color represents the 95% of confidence interval (point: average; range: maximum and minimum); the red line indicates performance expected by random guess.

Figure 5:

Genotoxicity prediction based on Tox21 qHTS data. a) model performance. Dot: average performance; range: 95% confidence interval; blue: models were constructed using all available traditional genotoxicity data and Tox21 qHTS data; black: models were constructed based on chemicals with available traditional genotoxicity data and also active in one of the Tox21 genotoxicity assay endpoints; red: models were constructed based on chemicals with available traditional genotoxicity data, where calls were reported with doses comparable to qHTS concentration range (positive) or beyond the range (negative), and also active in one of the Tox21 genotoxicity assay endpoints; b) coefficients in the best predictive models. Error bars represent the 95% confidence interval.

Figure 6:

Chemical prioritization for genotoxicity potential. A flowchart to put chemicals into bins based on *in silico* BM predictions and to arrange them into tiers based on the data availability in the BM endpoint (either +S9 or -S9) and the *in vivo* micronucleus endpoint. Tier 1: no data for either of the endpoints; Tier 2: data available for one of the endpoints; Tier 3: data available for both endpoints.

Figure 7:

Top 10 chemicals in the first tier of the prioritization scheme (presented using ToxPi software). The eight pies represent a chemical's activities in Tox21 genotoxicity assays (Ku70/Rad54, Rev3, ATAD5, γ H2AX, p53) and its probability for genotoxicity (htsBM, htsCA, htsMN) based on linear regression models using Tox21 qHTS data.

Table 1:

Tox21 qHTS genotoxicity assay endpoint tabular descriptions

Assay name	KU70 ^{-/-} /RAD54 ^{-/-}	REV ^{-/-}	ATAD5-luc	γH2AX	p53RE-bla
Assay endpoint ¹	KU70/RAD54 ↑	REV3 ↑	ATAD5 ↑	γH2AX ↑	p53 ↑
Gene Entrez ID	395767 424611	428622	79915	100757539	7157
Gene Entrez symbol	XRCC6 RAD54L	REV3L	ATAD5	LOC100757539	TP53
Genus/species	<i>Gallus gallus</i>	<i>Gallus gallus</i>	<i>Homo sapiens</i>	<i>Cricetulus griseus</i>	<i>Homo sapiens</i>
Incubation time (hour)	40	40	16	3	16
Cell line	DT40 clone 100	DT40 clone 657	HEK293	CHO-K1	HCT-116
Endpoint detection	Differential cytotoxicity	Differential cytotoxicity	Reporter gene transcription	Fluorescent antibody binding	Reporter gene transcription
Endpoint detection method	Differential cytotoxicity between wild-type (clone 653) and knockout cell line	Differential cytotoxicity between wild-type (clone 653) and knockout cell line	Luciferase level	Europium cryptate/d2 antibodies level	β-lactamase level
Readout	Luminescence	Luminescence	Luminescence	Fluorescence	Fluorescence
PubChem AID	NA ²	NA ²	720516	1224896	720552

¹ Assay endpoint represents a signed and directed effect (increase) of a biological process (e.g., transcriptional factor activity or induction of DNA damage) for a genotoxicity target-of-interest (e.g., p53 activity or cytotoxicity in the REV3 knock-out cell line).

² Only results from individual cell lines are available in PubChem (<https://www.ncbi.nlm.nih.gov/pcassay/?term=tox21+dt40>)

Table 2.

Tox21 qHTS genotoxicity assay performance

Assay name	<i>KU70</i> ^{-/-} / <i>RAD54</i> ^{-/-}	<i>REV</i> ^{-/-}	ATAD5-luc	γ H2AX	p53RE-bla
General					
S/B	40.19 $\pm$ 4.12	39.58 $\pm$ 8.88	6.02 $\pm$ 0.90	4.58 $\pm$ 0.63	3.13 $\pm$ 0.43
CV(%)	5.63 $\pm$ 2.06	6.33 $\pm$ 3.33	6.40 $\pm$ 0.51	6.03 $\pm$ 0.41	6.01 $\pm$ 2.27
Z' factor	0.78 $\pm$ 0.15	0.79 $\pm$ 0.09	0.73 $\pm$ 0.04	0.54 $\pm$ 0.09	0.61 $\pm$ 0.17
Performance measures for reference compounds (positive controls)					
Compound	TOAB *	TOAB *	5-Fluorouridine	Etoposide	Mitomycin C
EC ₅₀ (μ M)	0.42	0.43	2.05	4.98	1.80
SD of EC ₅₀	1.61	1.44	2.05	2.95	1.64
88 Duplicates					
SD of POD	2.01	1.91	2.02	1.61	1.49
Three runs					
Cohen's kappa	0.83	0.75	0.87	0.79	0.85

S/B: signal to background ratio, marginal ([2,3]), excellent (>3); CV(%): covariance of raw reads, acceptable (< 10%); Z': Z-factor, marginal ([0,0.5]), excellent (>0.5); EC₅₀: half maximal effect concentration; POD: point-of-departure; Cohen's kappa, moderate ([0.4,0.6], good ([0.6,0.8]), very good ([0.8,1]); mean plus standard deviation values were presented for the S/B, CV(%), and Z' factor; mean value was presented for the other parameters.

* Tetra-N octylammonium bromide, a cytotoxicant, not a genotoxic compound. The two DT40 assays are differential cytotoxicity assays that, by inference, to identify specific classes of DNA damaging compounds.

Table 3.

A summary of statistical methods and applied datasets

Task	Evaluation of qHTS genotoxicity assay predictivity		Evaluation of qHTS genotoxicity assay prioritization ability		Prediction/prioritization for genotoxicity using Tox21 qHTS data	
	Data evaluated	Reference data	Data evaluated	Reference data	Data evaluated	Reference data
Data	a) qHTS genotoxicity assay endpoints (n = 5) b) Union or intersection combinations of a)	i) Traditional genotoxicity endpoints (n = 3) ii) <i>in silico</i> genotoxicity endpoint (n = 1)	a) qHTS genotoxicity assay endpoints (n = 5) b) linear addition of combinations of a)	i) Traditional genotoxicity endpoints (n = 3) ii) <i>in silico</i> genotoxicity endpoint (n = 1)	a) qHTS genotoxicity assay endpoints (n = 5) b) qHTS stress response assay endpoints (n = 13) c) qHTS assay endpoints (n = 78)	i) Traditional genotoxicity endpoints (n = 3) ii) chemicals in a) active in any of 5 qHTS genotoxicity assay endpoints iii) chemicals in b) with activity in qHTS comparable testing range
Activity type	Binary call (active, inactive)	Binary call (positive, negative)	Logit-transformed activity ([0-1]), inconclusive = 0.0001	Binary call (positive, negative)	Log-transformed activity (0 to ~), inconclusive = 0	Binary call (positive, negative)
Statistical method	Contingency table		ROC enrichment		Elastic net regularization	

Table 4.Literature search of top 10 prioritized chemicals predicted as positive in *in silico* BM models

Chemical name	CASRN	Top 3 – 4 influences	Information from the literature and other sources	Notes/references
Idarubicin hydrochloride	57852-57-0	γ H2AX, ATAD5, p53, REV3	Anthracycline, Topo II inhibitor, MN+, BM+	Idamycin package insert (Pfizer)
Pirarubicin	72496-41-4	γ H2AX, ATAD5, p53, REV3	Anthracycline, Topo II inhibitor, MN+, BM+	
Oxiranemethanamine, N-[4-(oxiranylmethoxy)phenyl]-N-(oxiranylmethyl)-	5026-74-4	ATAD5, htsBM, htsCA, htsMN	Epoxide, CA+, BM+	aka: TK 12759 Data fro ECHA
FR073317	102409-92-7	P53, γ H2AX, htsCA	Epoxide, a dihydrobenzoxazine, DNA-DNA cross-linker more potent than mitomycin C	Naoe Y. et al. ⁴²
Carfilzomib	868540-17-4	P53, ATAD5, htsBM	BM–, CA+, <i>in vivo</i> MN–; is a protease inhibitor	Kyprolis package insert (Amgen, Inc.)
TNP-470	129298-91-5	KU70/RAD54, htsBM, htsMN	Angiogenesis inhibitor	
N-Butyl-N'-nitro-N-nitrosoguanidine	13010-08-7	htsBM, htsCA, htsMN	BM+	McCann et al. ⁴³
Copper dimethyldithiocarbamate	137-29-1	γ H2AX, ATAD5, htsBM, htsCA	BM+	ECHA
SR271425	155990-20-8	P53, htsCA, htsMN	DNA-reactive thioxanthone; alkylator and cross-linker; BM+	Lockhart AC. Et al. ⁴⁴
SAHA (Vorinostat)	149647-78-9	ATAD5, p53, htsCA	HDAC inhibitor; BM+, <i>in vitro</i> CA+, <i>in vivo</i> MN+; forms DNA adducts	Zolinza package insert (Merck & Co., Inc.)

HDAC: histone deacetylase; Topo II: Type II topoisomerase; MNMG: methylnitronitrosoguanidine; CASRN: CAS (Chemical Abstracts Service) Registry Number; ECHA: European Chemicals Agency

Table 5.

Experimental testing (BM and *in vitro* MN assays) with the seven top ranked Tox21 chemicals with unknown genotoxicity potential

Compound	CASRN	Compound use	<i>In silico</i> BM	γ H2AX	P53RE	ATA D5-luc	KU70 ^{-/-} /RAD54 ^{-/-}	REV3 ^{-/-}	<i>In vitro</i> MN ¹	BM ²
Potassium dicyanoaurate	13967-50-5	Dental gold alloy	inconclusive	+		+	+		W+	–
Dipyrrithione	3696-28-4	Antibacterial, antifungal	inconclusive	+	+	+			+	–
Dipentaerythritol pentaacrylate	60506-81-2	Acrylate	negative	+	+	+			E	–
4-(Hexyloxy)phenol	18979-55-0	Anticancer, anti-inflammatory	negative	+	+	+			+	–
2-Chloro-N-(2-methyl-4-bromophenyl)acetamide	96686-51-0	Biocide	negative	+	+		+		W+	TA100+ (with S9)
2-(Thiocyanomethylthio)-benzothiazole	21564-17-0	Pesticide	positive	+	+	+			W+	E. coli+ (with/without S9)
1H-Isindole-1,3(2H)-diimine	3468-11-9	Dye	negative	+	+	+			+	TA98+ (with S9)
Total number of active calls			1	7	6	6	2	0	7	3

¹W+, weak positive; E, equivocal

²Bacterial tester strains, genetic targets, MOA: TA100, G-G-G-, base-pair substitution; TA98, C-G-C-G-C-G-, frameshifts; *Escherichia coli*, UAA (ochre), transitions/transversions (all types of base-pair substitutions)