Dose-Response, 13:1, 2015
Formerly Nonlinearity in Biology, Toxicology, and Medicine
Copyright © 2014 University of Massachusetts
ISSN: 1559-3258
DOI: 10.2203/dose-response.13-045.Fujii

ISOTONIC REGRESSION BASED-METHOD IN QUANTITATIVE HIGH-THROUGHPUT SCREENINGS FOR GENOTOXICITY

Yosuke Fujii[1], Takeo Narita[2], Raymond Richard Tice[3], Shunich Takeda[2], Ryo Yamada[1] □ [1]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Japan; [2]Department of Radiation Genetics, Kyoto University Graduate School of Medicine, Japan; [3]Division of the National Toxicology Program, National Institute of Environmental Health Sciences, USA

□ Quantitative high-throughput screenings (qHTSs) for genotoxicity are conducted as part of comprehensive toxicology screening projects. The most widely used method is to compare the dose-response data of a wild-type and DNA repair gene knockout mutants, using model-fitting to the Hill equation (HE). However, this method performs poorly when the observed viability does not fit the equation well, as frequently happens in qHTS. More capable methods must be developed for qHTS where large data variations are unavoidable. In this study, we applied an isotonic regression (IR) method and compared its performance with HE under multiple data conditions. When dose-response data were suitable to draw HE curves with upper and lower asymptotes and experimental random errors were small, HE was better than IR, but when random errors were big, there was no difference between HE and IR. However, when the drawn curves did not have two asymptotes, IR showed better performance ($p < 0.05$, exact paired Wilcoxon test) with higher specificity (65% in HE vs. 96% in IR). In summary, IR performed similarly to HE when dose-response data were optimal, whereas IR clearly performed better in suboptimal conditions. These findings indicate that IR would be useful in qHTS for comparing dose-response data.

Key words: Isotonic regression, Hill equation, quantitative high-throughput screening, genotoxicity

INTRODUCTION

Over the last few decades, the number of chemical compounds in commercial use has grown rapidly, as indicated in the increase in CAS registry numbers from 40 million in 2008 to 75 million in 2013. However, comprehensive toxicological profiles for many of these compounds are lacking (NRC, 2007). In response to this deficiency of toxicological information, the original members of the U.S. Tox21 community have been evaluating the utility of quantitative high-throughput screening (qHTS) (Collins *et al.*, 2008; Shukla *et al.*, 2010; Inglese *et al.*, 2006; Thomas *et al.*, 2009). They have focused on various types of toxicity including reproductive and developmental toxicity, and immunotixicity. One of the toxicological endpoints of interest to U.S. Tox21 is genotoxicity.

Address correspondence to Ryo Yamada, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, South Research Bldg. No.1, Room 515, Syogoinkawaharamachi, Kyoto Sakyo-ku, Kyoto, 606-8507, Japan; TEL: +81-75-366-7403; FAX: +81-75-751-4168; ryamada@genome.med.kyoto-u.ac.jp

Traditional toxicological evaluation has largely relied on animal models. They are, however, expensive, low-throughput, and sometimes inconsistently predictive of human biology and pathophysiology (Shukla *et al.*, 2010). Recently, the NIH Chemical Genomics Center (NCGC) conducted evaluations of a novel qHTS approach for identifying genotoxicity potential based on the detection of increased cytotoxicity in isogenic DT40 DNA repair-deficient cell lines, compared to that exhibited by the repair-proficient parental cell line under the same exposure conditions (Yamamoto *et al.*, 2011; Ji *et al.*, 2009). Miniaturized assay volumes (< 10 μL/well) in a 1536 well-plate format provide the high-throughput required to generate dose-response data for every compound library member tested. The premise behind this strategy is that a decrease in DNA repair competency will increase the sensitivity of cells to being killed by genotoxic chemical compounds. The DNA repair-proficient parental cell line serves as a control in this screening, providing high sensitivity and specificity (Evans *et al.*, 2010). Furthermore, the use of a panel of DNA repair-deficient cell lines allows for characterization of the nature of DNA lesions caused by genotoxicants (Mizutani *et al.*, 2004; Nojima *et al.*, 2005; Wu *et al.*, 2006).

There is much room for improvement in analyzing dose-response data from qHTS, even though it gives us a large amount of information on compounds. In a previous qHTS study (Yamamoto *et al.*, 2011), screening was conducted for 2864 compounds from the phase I library provided by the National Toxicology Program (Tice *et al.*, 2012). Dose-response data were obtained from one wild-type clone and seven mutant clones deficient in different DNA repair factors. The dose-response data were analyzed by fitting to Hill equation (HE) sigmoid curves, according to the standard NCGC protocol (NTP, 2010). This protocol classified the estimated curves for their genotoxic evidence and the quality of the estimated curves. The qualities of the curves were assessed by the presence of asymptotes and inflection, indicating their appropriateness as HE curves. Further details are provided in Figure 1. The curves were called “complete” if they had upper and lower asymptotes and an inflection; otherwise, they were classified as “incomplete” (Figure 1BC, Inglese *et al.*, 2006). Based on these criteria, complete curves were obtained for only 6.72% of all records (1539/22912). This situation was considered likely to produce false judgments of genotoxicants. One reason for this low rate of “complete” curves was that no experimental condition could be designed compatible with allowing a large number of compounds to be screened together at a high-throughput scale due to the heterogeneity of the compounds. Another reason was that dose-response data did not always fit to the theoretical Hill equation due to the wide differences in chemical characteristics of the compounds (Jiang *et al.*, 2011).

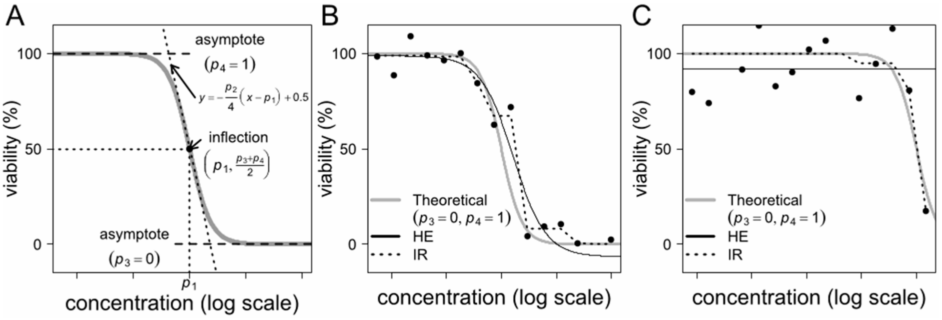

FIGURE 1. Theoretical and estimated dose-response curves and classification of estimated curves. (A) The theoretical HE curve has two asymptotes at 0 (0%) and 1 (100%) viabilities, and the inflection in the middle of the slope. The y-coordinates of the two asymptotes, the coordinates of the inflection, and the equation of the tangent line at the inflection, are indicated using the parameters in equation (1). (B) When observing concentrations covered by the range including the two asymptotes, the HE estimation curve (solid line) had two asymptotes and an inflection, so the estimated curve was called "complete" based on the NCGC classification criteria. Both HE and IR estimation curves (solid and dotted, respectively) were close to the theoretical curve (grey). Only one curve is drawn for simplicity. (C) On the other hand, when observing concentrations that only partially covered the range, the HE estimation curve had only one asymptote with an inflection, and the estimated curve was called "incomplete" based on the criteria. The estimated IR curve (dotted) had only one asymptote as well, but it fit better to the theoretical curve than the HE curve.

DT40 DNA repair-deficient cell lines have been studied in detail as a method to measure the genotoxicity of chemical compounds by comparing the dose-survival curves of damaged cell lines with the curves of intact cell lines under refined experimental conditions. The cell culture conditions, the dose of each compound, and time points of observation are optimized for individual chemicals and cell lines. The experiments are performed on a small scale at a regular lab bench and are repeated as necessary. Expansion of this system to a Tox21 high-throughput setting, where 1536-well plates were used and the experimental conditions for the wells and plates were unified, meant both that the experimental conditions were not optimized for individual chemicals and that the repetition number was fixed and limited. This produced two major statistical problems that need to be solved: (1) the dose-survival curves did not always fit within the dosage range, and (2) the random errors were not always appropriate to allow dose-survival curves to be generated. The conventional sigmoid curve fitting method, i.e., the HE method, is a parametric method that assumes the presence of two horizontal asymptotes with a symmetric slope between them. It performs well when the data fit with the parametric assumptions, but it is not reliable when data deviate significantly from the assumptions. In contrast, nonparametric methods are much more adaptable to variable data in general. Besides of data variability, monotonicity should be assumed for our dose-response curves. Actually, there are two types of monotonicity. One is the monotonic decrease in viability both of wild-type and mutant cell lines with an increase in chemical concentration. The other is that the curve for the wild-type cell should be higher than the curve for the mutant at any concentration. Therefore, a nonparametric method that could incorporate those two monotonic restrictions should be appropriate for these data. The isotonic regression method (IR) meets these conditions. Consequently, we compared the performance of HE and IR on simulated qHTS data using the DT40 system.

We examined the problems in evaluating genotoxicity with dose-response data from the high-throughput system and divided them into the following three components: (1) the magnitude of random errors in observed viability and their effect size of genotoxicity, (2) the inadequate coverage of the curves by the experimental concentrations used and (3) the variations in the gradients of the curves (i.e., two dose-response curves are not necessarily parallel in the real data).

We compared our new method, based on isotonic regression (IR), with a conventional method based on the HE, for their performance in terms of these three problems using simulation data.

METHODS

Experimental design of qHTS

With qHTS, chemical compounds were screened for their genotoxicity by comparing dose-response curves of multiple DNA repair gene mutants, *m,* against the curve of one wild-type, *w.* Any compound for which any one of the mutant curves indicated a higher death rate compared with the wild-type curve was judged to be a genotoxicant. In later parts of this report, we focus on comparisons between one mutant and one wild-type for simplicity.

The wild-type and mutant were exposed to a compound at multiple concentrations evenly spaced on a logarithmic scale $\boldsymbol{c} = (c_i = c_1 r^{i-1})$; $i = \{1, 2, ..., n\}$ (Figure 2A). We observed cell viability at each c_i for the two cell lines. The viabilities were standardized to range from 0 to 1 (0% to 100%) based on the experimental signal intensity of positive and negative controls (Xia *et al.*, 2008). Because the strength of the genotoxicity of chemical compounds varied and $\boldsymbol{c}$ was fixed for all compounds in the qHTS system, the dose-response curves of some compounds were not adequately covered by this concentration range (Figure 2B). The nomenclature and variables used in this paper are summarized in Supplementary Data.

Generation of simulation datasets

Theoretical dose-response data for the wild-type and mutant were generated from a log-logistic sigmoid curve function (Ritz, 2010; Finney, 1979) on the logarithmic scale $\boldsymbol{x} = log_{10}\boldsymbol{c}$.

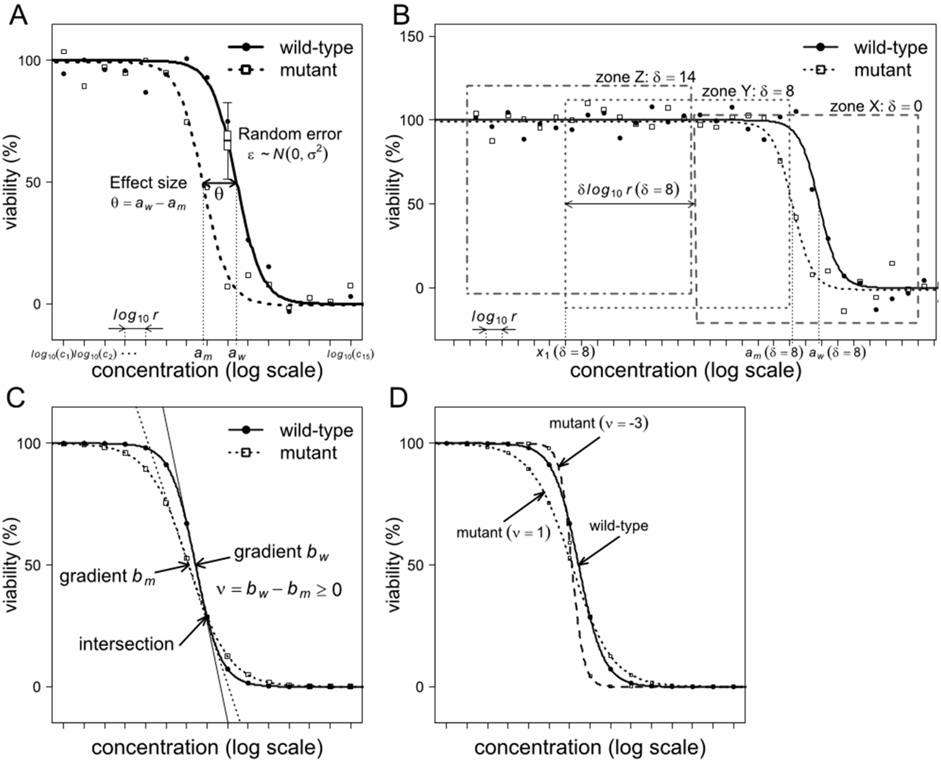

FIGURE 2. Simulation designs. (A) For the theoretical dose-response curves, we fixed two parameters, p_3 and p_4, out of the four parameters in equation (1) and rewrote equation (2) with two parameters, $a = p_1$ and $b = p_2$. The viability was a function of concentration provided on a logarithmic scale. When the compound was a typical genotoxicant via the mutated gene function, the two dose-response curves of wild-type (solid) and mutant (dotted) were parallel with a horizontal shift. The horizontal shift of the mutant line to the left meant that the mutant cells were less viable as lower concentrations of the compound. The effect size of the genotoxicity was defined as the horizontal shift, $\theta = a_w - a_m$, which is indicated as "Effect size" on the graph. The experimentally observed viabilities deviated from the curve with random errors, as indicated by filled circles (wild-type) and rectangles (mutant). Random errors are indicated by a box plot in the panel. This dataset was generated with $\theta / \sigma = 0.167$. (B) This plot describes suboptimal conditions where the experimental concentrations did not cover the dose-response curves well. When the coverage was suboptimal, the estimated curve did not have two asymptotes and it was classified as "incomplete". The parameter δ controlled the location of the experimental concentrations relative to the theoretical curves. Each dotted rectangle corresponds to $\delta = 0$, 8, and 14 as it shifts to the left. We observed cell viability from the minimum concentration to the maximum concentration for each simulation. Zone X, Y, and Z are described in Figure 5. (C) This panel describes a compound whose dose-response curves in wild-type and mutant were not parallel and had an intersection in the middle of the slopes. This indicated that the compound had a greater effect lowering the viability of the mutant cells when its concentration was low, but when its concentration was high, the viability of wild-type cells was lower. Although it is not easy to explain this phenomenon by simple biological models, the estimated curves based on observation sometimes fit this pattern. (D) This panel describes how v changed the shapes of curves. When $v = 0$, the wild-type and mutant curves were parallel, but when $v > 0$, the slope of the mutant curve was steeper, and when $v < 0$, the slope of the wild-type curve was steeper. The two mutant curves represented the largest and smallest v, from -3 to 1, which was the range we evaluated in this report.

$$f_*(x) = p_3 + \frac{p_4 - p_3}{1 + e^{p_2(x-p_1)}}, \ *: w,m. \tag{1}$$

The parameters p_3 and p_4 stand for the horizontal asymptotes (Figure 1A). For our models, p_3 and p_4 were fixed at 0 and 1, respectively. The parameter p_1 indicated the concentration of inflection ($log_{10}\mathrm{EC}_{50}$) and parameter p_2 determined the gradient of the slope (Figure 1A). We replaced p_1, p_2, p_3 and p_4 with a, b, 0 and 1, respectively, to express the theoretical curve function and to distinguish between wild-type and mutant, and a and b were indicated by the subscripts w and m. Subsequently, experimental data were generated with random Gaussian errors ε as shown in equation (2).

$$f_*(x) \sim \frac{1}{1 + e^{b_*(x-a_*)}} + \varepsilon, \ *: w,m. \tag{2}$$

In this system, genotoxicity was defined as an increased death rate of the mutant cell line. Genotoxicity was measured as the shift of the mutant curve to the left. The size of the shift was expressed as $\theta = a_w - a_m$ (Figure 2A).

Conditions to generate simulated data for method comparison

We parameterized equation (2) for the compounds according to the experimental conditions as shown in Figure 2. As shown in Figure 2A, the observed viability deviated from the theoretical curves. The deviation was given as random errors following a normal distribution, $\varepsilon \sim N(0, \sigma^2)$. The effect size of genotoxicity of the positive control compounds, $\theta = a_w - a_m$, was also parameterized as shown in Figure 2A. We selected the values of σ and θ to be $\sigma = \{3, 6, \ldots, 18\}$ and $\theta = \{0.2, 0.4, \ldots, 2\}$ based on a previous study by Yamamoto et al. (2011). The dataset in Figure 2A was generated with $\theta = 1$ and $\sigma = 6$. As shown in Figure 2B, the experimental concentration window sometimes covered the theoretical curves well, while sometimes the coverage was inadequate. To parameterize the coverage, we introduced a parameter, $\delta = \frac{(a_w + a_m)/2 - (x_1 + x_n)/2}{log_{10}r}$ that determined the relative location of the curves in the experimental window. We set $\delta = \{-14, -13, \ldots, -1, 0, 1, \ldots, 14\}$. The last parameter we generated was the difference between the gradients of the curves, $v = b_w - b_m$ (Figure 2C). For the majority of the simulations, we set $v = 0$, and when we specifically evaluated the effect of v, we set v in the range from -3 to 1 (Figure 2D).

With these parameters, six σ, ten θ, twenty nine δ, and twenty one v were combined, providing 36540 conditions in total. For each condition, we generated 1000 simulation datasets.

Statistics to measure genotoxicity

We compared two methods, the conventional HE-based method and our proposed IR-based method, for their ability to detect genotoxicity. The differences in the statistics between wild-type and mutant, $\Delta S_{HE,\ w} = S_{HE,\ w} - S_{HE,\ m}$ and $\Delta S_{IR} = S_{IR,\ w} - S_{IR,\ m}$, were used as the measure of genotoxicity as described below.

Hill equation and the difference between EC$_{50}$ values

As shown in Figure 3A, two curves were drawn using dose-response datasets of wild-type and mutant cell lines. Each curve was expressed as equation (1), and its four parameters (p_1, p_2, p_3, and p_4) were estimated, so that the difference between the observed values and the values on the estimated curve should be a minimum (Hill, 1910). For the estimation of one curve, the dataset of the corresponding cell line was used to estimate the four parameters of one curve and the dataset of the other cell line was used separately to estimate the parameters of the other curve. In technical terms, the parameters were estimated using the “drm” function in the “drc” package in R (http://cran.r-project.org/). These function estimated parameter values were identified by maximizing their likelihood using the Limited-memory BFGS optimization procedure (Byrd *et al.*, 1994).

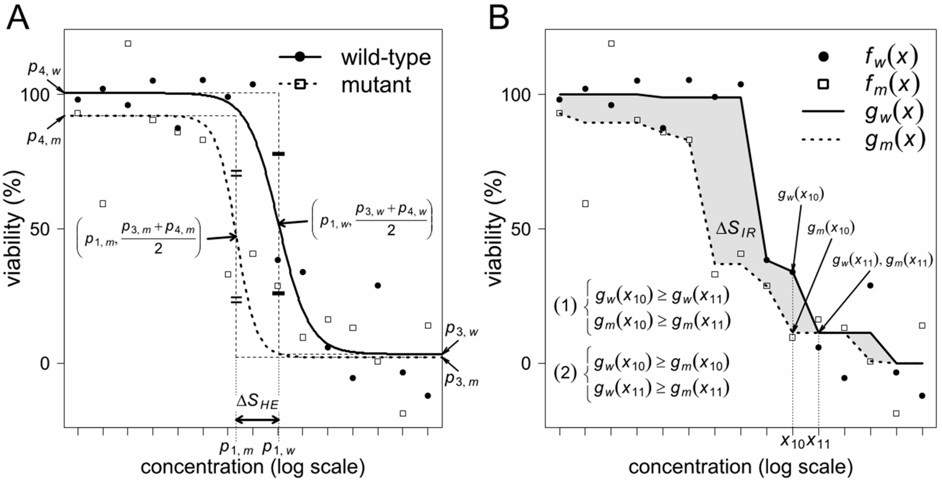

FIGURE 3. The estimated curves obtained using the Hill equation and isotonic regression methods, and the statistics for the genotoxicity results. (A) The filled circles and rectangles represent observed viability values for the wild-type and mutant cell lines, respectively. Two curves were estimated with HE. The solid curve is of the wild-type cell line and the dotted curve is of the mutant cell line. The statistics for the genotoxicity results, ΔS_{HE}, is the distance between the horizontal coordinates of p_1 (log_{10}EC$_{50}$) of the two curves. (B) For the same dataset, isotonic regression (IR) provided two lines, the wild-type line (solid) and the mutant line (dotted). The points $g_w(\boldsymbol{x})$ are on the solid line and the points $g_m(\boldsymbol{x})$ are on the dotted line. Both solid and dotted lines decreased monotonically, and the solid line was above the dotted line throughout. Four inequality restrictions for x_{10} and x_{11} are indicated. The statistics for the genotoxicity results, ΔS_{IR}, is the area between two lines, which is shadowed.

Because the theoretical genotoxicity was $\theta = p_{1,\,w} - p_{1,\,m} = a_w - a_m$, its estimate $\hat{\theta}_{HE}$ was given as shown below using the estimated p_1 values of two lines, $S_{HE,\,w}$ and $S_{HE,\,m}$. This is in accordance with the conventional index of genotoxicity (Yamamoto *et al.*, 2011) (Figure 3A) and is related to the EC_{50} of the two cell lines.

$$S_{HE,*} = p_{1,*} = \hat{a}_*, * : w, m. \tag{3}$$

$$\hat{\theta}_{HE} = S_{HE} = S_{HE,\,w} - S_{HE,\,m} = log_{10}\frac{EC_{50,w}}{EC_{50,m}} \tag{4}$$

Isotonic regression and the area between two estimated lines

As shown in Figure 3B, two lines were drawn, one for each of the two cell lines. The two lines met two restrictions: (1) both lines were composed of line segments and both decreased monotonically and (2) the wild-type line (solid) was above the mutant line (dotted) for all the concentrations tested. IR generated two lines at the same time using the dose-response datasets of the two cell lines together. IR searched the two lines so that they met the two restrictions and so that the deviations between the observed values and the estimated values were minimal. Specifically, IR is a nonparametric approach for building models whose fits are monotonic. In the dose-response curve scenario, an observed viability vector, $F(\boldsymbol{x})$, to another estimated viability vector, $G(\boldsymbol{x})$, carries a weight vector, $\boldsymbol{h}$, that imposes inequality restrictions over the values of $G(\boldsymbol{x})$, such as $G(x_i) \geq G(x_j)$. Thus, to solve equation (5),

$$minimize\sum_{i=1}^{n} h_i\{F(x_i) - G(x_i)\}^2 \tag{5}$$

subject to the inequality restrictions (Best and Chakravarti, 1990). This is a very powerful method to analyze a monotonic trend in nonlinear data. In our case, $F(\boldsymbol{x})$ consisted of the observed viabilities of both cell lines, where $F(\boldsymbol{x}) = \{f_w(\boldsymbol{x}), f_m(\boldsymbol{x})\}$. Similarly, $G(\boldsymbol{x})$ consisted of the estimates of the viabilities of two cell lines, where $G(\boldsymbol{x}) = \{g_w(\boldsymbol{x}), g_m(\boldsymbol{x})\}$. See Figure 3B for $f_w(\boldsymbol{x}), f_m(\boldsymbol{x})$ and $g_w(\boldsymbol{x}), g_m(\boldsymbol{x})$. There were two types of constraints on $G(\boldsymbol{x}) = \{g_w(\boldsymbol{x}), g_m(\boldsymbol{x})\}$. First, viability should monotonically decrease in the range from 0 to 1 both in the wild-type and in the mutant ($g_*(x_i) \geq g_*(x_j)$; $x_i \leq x_j$, for all *i-j* pairs, where * stands for *w* or *m*). Second, the wild-type viability should be equal to or greater than the mutant viability at the same concentration ($g_w(x_i) \geq g_m(x_i)$; for all *i*). In our case, all data points were given equal weight ($h_i = 1$).

We used the area between the two estimated $g_w(\boldsymbol{x})$ and $g_m(\boldsymbol{x})$ lines (trapezoidal rule for approximating an integral, Figure S2 in Supplementary Data) as the index of genotoxicity.

$$S_{IR,*} = \sum_{i=1}^{n-1} \{g_*(x_{i+1}) + g_*(x_i)\} \frac{log_{10} r}{2}, * : w,\ m. \tag{6}$$

$$\Delta S_{IR} = S_{IR,\,w} - S_{IR,\,m}. \tag{7}$$

We performed IR using the "quadprog" package in R and the "cvxopt" module in Python (Martin *et al.*, 2012, 2013) (Figure 3B).

Assessment of performance via receiver operating characteristic curves

As we described in the Introduction, there were two problems with the qHTS data from the genotoxicity screening system: (1) the dose-survival curves did not always fit within the dosage range, and (2) the random error variations were not always appropriate to allow drawing of the dose-survival curves. In order to compare the two sets of statistics based on HE and IR for datasets with these problems, we generated simulation data using four parameters: (i) the difference in the magnitude of genotoxicity between the two cell lines, θ, (ii) the size of the random errors, σ, (iii) the observation window, δ and (iv) the difference between the gradient of the curves from the two cell lines, v. The main objective of our study was to evaluate which factor(s) affects the performance of the two methods in genotoxicity screening.

Basically, the performance of the two methods in terms of discriminating genotoxicants from non-genotoxicants was evaluated by sensitivity and specificity. The changes in sensitivity and specificity, along with various cut-off values, were evaluated using the receiver operating characteristic (ROC) curve, and the area under the receiver operating characteristic curve (ROC-AUC) was used to quantitate their performance of screening (Swets, 1988). The relations between the conditional parameters and ROC-AUC or sensitivity/specificity were evaluated. Actually, the effects of (i) the true effect size of genotoxicity between the two cell lines, θ, and (ii) the size of the random errors, σ, on the performance of the two methods, were better interpreted using their ratio θ / σ, as is the case with the majority of statistical tests. We tested the relation between θ / σ and ROC-AUC using the DeLong test (Delong *et al.*, 1988) when testing individual θ / σ, and we used the exact paired Wilcoxon test when testing for a set of θ / σ.

Evaluation of the effect of δ on sensitivity and specificity when screening thresholds were fixed

In the qHTS system, one threshold value is applied to all compounds, which separates the compounds into two groups: genotoxicity-positives and genotoxicity-negatives. We therefore applied a fixed threshold value for each method and evaluated the effect of δ on their sensitivity and specificity with $\sigma = 18$, $\theta = 1$, and $v = 0$. Sensitivity and specificity were tested using a binomial test. We determined the threshold value for each method by applying Youden's J statistic (Perkins and Schisterman, 2006; Youden, 1950) to our ROC-AUC for datasets with $\sigma = 18$, $\theta = 1$, $\delta = 0$, and $v = 0$ ("pROC" package in R. Robin *et al.*, 2011).

RESULTS

Effects of ratio of the true effect size of genotoxicity to random errors (θ / σ)

We evaluated the effects of the variance of random errors and of the effect size of genotoxicity on the performance of HE and IR because the efficiency to detect meaningful genotoxicity depended on the relative size of the genotoxicity value to the random errors, as is true for statistical methods in general. As shown in Figure 2A, the effect size was the length between two theoretical curves; the data points deviated from the theoretical curves due to the random errors. When the random errors were bigger, the precision of the estimated curves was lower. Because we considered the compound positive for genotoxicity when the length between the wild-type and mutant curves was larger than a threshold value, the sensitivity of the test was lower when the random errors were bigger. When the true effect size was bigger, the effect of random errors on the sensitivity was smaller. Therefore, we adopted the ratio of the true effect size of genotoxicity to random errors as a parameter to evaluate the performance of HE and IR. We compared the two methods when the curves were parallel ($v = 0$) and largely within the experimental concentrations ($\delta = 0$).

The ratio of genotoxicity to the variance of the random errors, θ / σ, and ROC-AUC showed a strong relationship, as shown in Figure 4. In the right half of Figure 4, the ROC-AUC of HE and IR was almost 1, indicating that both methods could effectively detect genotoxicity when the ratio of the strength of genotoxicity to the variance of the random errors was large. On the other hand, in the left half of Figure 4, the ROC-AUC of both methods was positively correlated with the ratio, and the ROC-AUC of HE was higher than that of IR. The overall ROC-AUC of the two methods differed significantly ($p < 0.05$, exact paired Wilcoxon test). The difference was due to HE's better performance θ / $\sigma \leq 0.067$ (the left half of Figure 4). Most of this difference was statistically significant ($p < 0.05$, DeLong test).

Effects of the coverage of dose-response curves with observation windows (δ)

We evaluated the effects of the observation windows on the performance of HE and IR. Figure 2B shows the dose-response curves for the con-

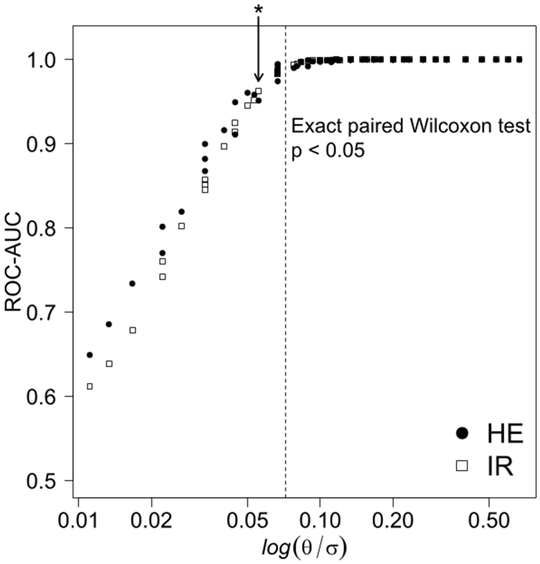

FIGURE 4. The relationship between θ / σ and ROC-AUC. When the ratio of the true effect sizes to the variance of the random errors was smaller than the vertical dotted line (the left side of Figure), the ROC-AUC of HE was better than that of IR. On the other right side, the performance of the two methods was not different. The two methods were tested for differences for all θ / σ values as a set, which was statistically significant ($p < 0.05$, exact paired Wilcoxon test). The majority of θ / σ on the left side were also significant when compared individually ($p < 0.05$, DeLong test). The vertical dotted line corresponds to θ / $\sigma = 0.067$. The asterisk (*) and an arrow indicate θ / $\sigma = 0.056$, used in the data generation of Figure 5 ($\sigma = 18$, $\theta = 1$).

ditions at $\delta = 0$, 8, and 14, and we evaluated for $\delta = \{-14, -13, \ldots, -1, 0, 1, \ldots, 14\}$. ROC-AUC, sensitivity and specificity were plotted along with δ (Figure 5). The horizontal axis was zoned into three segments according to curve classes: in zone X, the asymptotes of both curves were covered by the observation window. Therefore, it was likely that the estimation using HE was easy and was thus classified as "complete". In zone Y, one of the two asymptotes was included in the window, and the slope was only partially covered. In this zone, the estimation with HE was likely to be "incomplete" but the estimation using IR could be informative. In zone Z, the observation window did not cover the slope and neither method was useful. Please see the top panel of Figure 5. In zone X, the ROC-AUC was over 0.85 for both HE and IR. IR barely decreased ROC-AUC, whereas HE slightly decreased ROC-AUC. In zone Z, the ROC-AUC of both methods was 0.5, indicating that both methods were not informative. In zone Y, the ROC-AUC obtained for the two methods showed different behavior, namely, HE decreased more rapidly than IR as δ increased. Of note, the ROC-AUC curve of HE in zone Y decreased irregularly (an increase in ROC-AUC at $\delta = 9$) (Figure 5 top). This point corresponded to the condition where the edge of the observation window was very close

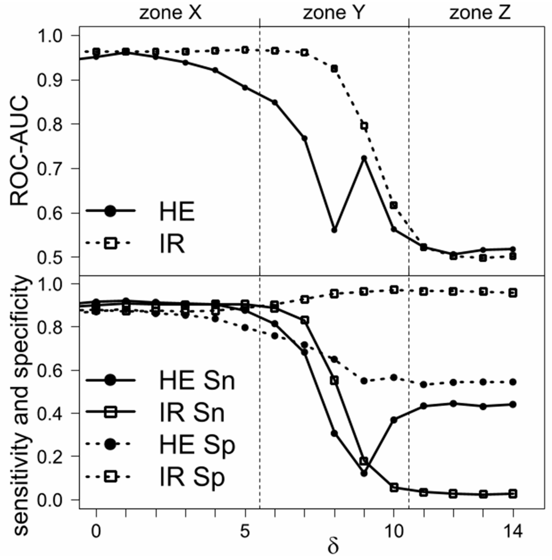

FIGURE 5. ROC-AUC (top), and sensitivity and specificity (bottom) with various δ, with $\sigma = 18$, $\theta = 1$ and $v = 0$. θ / σ of this experiment is provided in Figure 4. δs were divided into 3 zones, Zone X, Y, and Z, as described in the main text. (Top) In zone X, both methods had similar performance, but HE was more subject to δ. In zone Y, the reliability of both methods decreased, but the decrease in HE was more rapid than that of IR. In zone Z, both ROC-AUCs were 0.5. (Bottom) Sensitivity and specificity. In zone Z, the sensitivity and specificity of HE were 0.5, while the sensitivity and specificity of IR were almost 0 and 1, respectively. Both plots share the horizontal axis, δ. Abbreviations: Sn = sensitivity, Sp = specificity.

to the inflection point of the theoretical curves (data not shown). In zone Y, the superior performance of IR over HE was confirmed for all θ / σ ($p < 0.05$, exact paired Wilcoxon test).

Effects of the difference in gradients of curves (*v*)

We evaluated the situations where $v \neq 0$, or where the dose-response curves were not parallel and thus intersected with each other. We used datasets with $\sigma = \{3, 6, \ldots, 18\}$, $\theta = \{0.2, 0.4, \ldots, 2\}$, and $\delta = 0$ so that the curves should be "complete". The parameter indicating the difference of the curve gradients between the wild-type and the mutant, v, varied in this evaluation (Figure 2D).

The following are some facts regarding the genotoxicity indices, ΔS_{HE} and ΔS_{IR}. Supplementary Data and Figure S1 provides the proofs.

(a) When $v = 0$, $\Delta S_{HE} = a_w - a_m$, was the same as the area between two theoretical curves. This is because the height of the area between the curves was 1, the curves were parallel, and the area between the curves was the same. The area of the rectangle (or parallelogram) had a width of a_w - a_m and a length of 1 (Figure 6A).

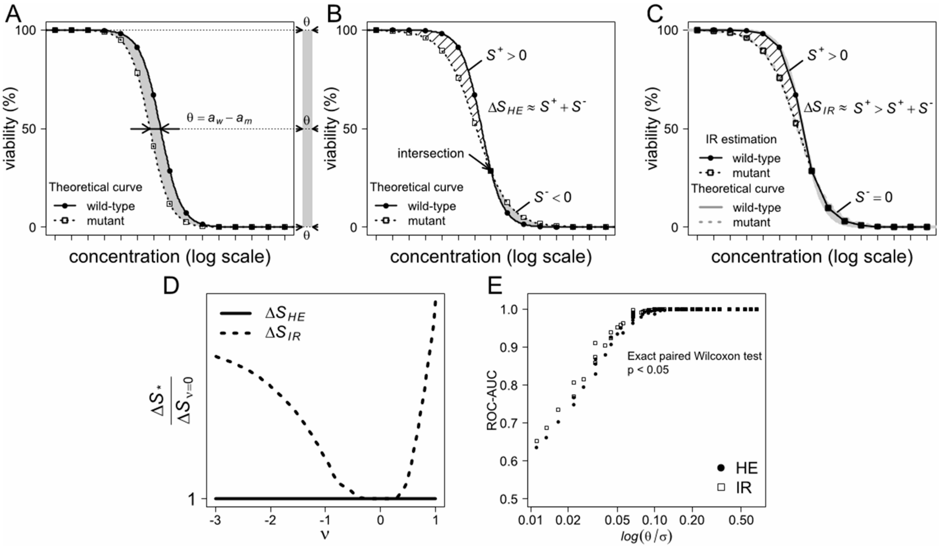

FIGURE 6. The effect of v on the statistics of HE and IR (A) When dose-survival curves of the wild-type and the mutant were parallel ($v = 0$), the area between the two curves was identical to $\theta = a_w - a_m$, indicated as a grey rectangle. (B) When the two curves were not parallel ($v \neq 0$), the area between the two curves was divided into S^+ and S^-. In this case, the area was also identical to $\theta = a_w - a_m$, namely, ΔS_{HE}. (C) IR estimated the area S^- to be zero because of the restriction of $g_w(x_i) \geq g_m(x_i)$. (D) The effect of v on ΔS_{HE} and ΔS_{IR}. The ratio of ΔS_* (*: HE, IR) to ΔS_* at $v = 0$ was plotted. v did not affect ΔS_{HE} while it affected ΔS_{IR}. (E) The relationship between θ / σ and ROC-AUC at $v = 1$. In contrast to Figure 4, the overall ROC-AUC of IR was better than that of HE ($p < 0.05$, exact paired Wilcoxon test) and the majority of θ / σ on the left side were also significant when compared individually ($p < 0.05$, DeLong test).

(b) When $v = 0$ and when the number of observing points was large, the lines obtained using IR could be identical with their theoretical curves. In this case, the area between the lines of IR, ΔS_{IR}, was the same as the area between the theoretical curves, which was the same as $\Delta S_{HE} = a_w - a_m$ as described in (a) (Figure 6A).

(c) When $v \neq 0$, the area between the theoretical curves was again the same as $\Delta S_{HE} = a_w - a_m$. Note that the curves had the intersected when $v \neq 0$ and the area between the curves was divided into two parts: one was the area where the wild-type curve was higher than the mutant curve (the area S^+), and the other was the area where the wild-type curve was lower than the mutant curve (the area S^-; Figure 6B). The area which was the same as $\Delta S_{HE} = a_w - a_m$ was the sum of the areas S^+ and S^-.

(d) When $v \neq 0$, the area between the theoretical curves was divided into two parts as mentioned above. At concentrations where the wild-type curve was higher than the mutant curve, the IR method estimated two lines that should fit to the theoretical curves. In contrast, at concentrations where the wild-type curve was lower than the mutant curve, the IR estimated two lines as the same line, because the line for the wild-type should not be below the line for the mutant (Figure 6C). This meant that ΔS_{IR} tended to be larger than the area between the theoretical curves, which is the difference between the areas S^+ and S^-, as explained in (c). Therefore, when $v \neq 0$, ΔS_{IR} tended to be larger than ΔS_{HE}.

Figure 6D shows the effect of v on ΔS_{HE} and ΔS_{IR} under the condition where the observational points were adequate and with no random errors. v did not affect ΔS_{HE}, as described in (c), while it did affect ΔS_{IR}. ΔS_{IR} became larger both when $v < 0$ and when $v > 0$ than the value when $v = 0$.

Because of this phenomenon, ROC-AUC behaved differently when $v \neq 0$. Figure 6E showed the relation between θ / σ and ROC-AUC when $v = 1$ and with zone X condition. When θ / σ was big (in the right half of Figure 6E), the ROC-AUCs of HE and IR were almost identical, but when θ / σ was small (in the left half of Figure 6E), the ROC-AUCs of IR were bigger than those of HE. This was contrary to the results when $v = 0$ (Figure 4). The difference was statistically significant ($p < 0.05$, exact paired Wilcoxon test). For all $v \geq 0.8$, the ROC-AUCs of IR were better than those of HE ($p < 0.05$, exact paired Wilcoxcon test; data not shown).

Sensitivity and specificity with various observation windows (δ)

We observed the relationship between δ and the sensitivity and specificity of the two methods (Figure 5 bottom). In zone X, the specificity of HE decreased with an increase in δ, and appeared to be the main source of the decrease in ROC-AUC in HE. In zone Z, ROC-AUC was approximately 0.5 for both methods, but the sensitivity and specificity behaved differently. In HE, both the sensitivity and specificity were close to 0.5, indicating that HE judgments would not be useful. On the other hand, the IR specificity was very high but with very low sensitivity. IR thus provided high false negatives but low false positives, even in zone Z. Zone Y had features intermediate between zones X and Y. In zone Y, the sensitivity of HE and IR decreased, with IR showing some advantages over HE based on ROC-AUC. In contrast, the specificity of HE and IR differed significantly: HE specificity became smaller, but IR specificity remained high. At $\delta = 8$ (Figure 2B), the specificity of HE and IR was 65% and 96%, respectively ($p < 0.05$, binomial test) (Figure 5 bottom). Thus, the difference in specificity between the two methods appeared to be the source of the difference in ROC-AUC.

DISCUSSION

We found that quantitative high-throughput systems used to screen compounds by comparing two dose-response curves were affected by the screening conditions. Specifically, considerable parts of the dose-response curve data were not analytically optimal because of random errors or shifts in the informative parts of the curves. These inevitable phenomena mean that that the utility of HE is limited for qHTS. This is in agreement with a previous study showing that only 6.72% of all records were appropriate for HE curve fitting (Yamamoto *et al.*, 2011). Accordingly, we applied IR and compared its performance with HE.

First, IR performed comparably with HE when the two curves were completely observed within the experimental concentrations; the random errors were acceptable compared with effect sizes. In contrast, when random errors were very large, the performance of IR was slightly worse than that of HE. However, such data may not allow conclusions to be drawn since the effect size of the variables in question is very small. Consequently, the utility of the data for practical genotoxicity is limited, since the random errors appear to be too large to be acceptable from the standpoint of experimental quality control.

Besides the random errors, we frequently observed suboptimal dose-response data which poorly fit the HE curve within the experimental concentrations. The accuracy of HE was affected by the poor-fitness when the slope part of one of the two curves was at the boundary of the experimental concentrations used. This is because the accuracy of HE depends on the estimation of *a* (log_{10}EC$_{50}$), which corresponds to the inflection point of the curve. Indeed, inclusion of this point critically affected the performance of HE (data not shown). This indicated that the effect of partial observation of the informative part of the curves was critical. In contrast, IR was not affected by these particular points. We demonstrated that partial observation of the informative parts of the curves did not affect the performance of IR as much as that of HE. When the experimental concentrations were almost outside the informative part of the curves, the sensitivity and specificity of HE were close to 0.5, indicating that it would be inadequate for judging compounds. In the same conditions, IR showed very low sensitivity, but its specificity was very high. This indicates that the results from IR may be useful for screening, although their use should be further evaluated in the context of the particular purpose of each individual screening system.

In addition to the conditions where the two curves are parallel, in this study we also simulated compounds for which the dose-response curve of the mutant was not parallel to the curve of the wild-type. This issue is related to the definition of genotoxicity, as described in Supplementary Data 2. When two dose-response curves are parallel, the two curves do not intersect (Figure 2A) and the area between them is equal to the difference of the logarithmic scale of EC$_{50}$ (following appropriate scale and unit adjustments). This equality means that the performance of the two methods is similar for compounds with parallel curves. On the other hand, when the two curves are not parallel, the mutant curve is at least partially above the wild-type curve (Figure 2C). This means that the compounds cause a higher death rate of the mutant when their concentration is below the concentration at the intersection, and a higher death rate of the wild-type when their concentration is above the concentration, or vice versa. It would be useful to detect any compound that may cause a higher death rate of mutants that are susceptible to DNA damage, even at a limited range of concentrations, particularly in the screening settings. This is because the biological effects of chemical compounds may differ completely depending on the concentration of the chemical compound. Such chemical compounds can be successfully detected by IR, since IR does not measure the amount of horizontal shift of the two curves but rather measures the area representing the increased death rate of the cells. Additionally, we simulated various effects of the genotoxicants on the mutant cell line, such as the non-horizontal shift of the dose-response curves. The evaluation of the potential correlation between the patterns of the shifts and the physico-chemical features of the compounds might produce new insights into the mechanism of genotoxicity, although we have not studied the real compounds with our method.

In this paper we are proposing a relatively new method based on IR to quantitatively evaluate suboptimal dose-response data. We showed that IR performed better than HE when dose-response data were not obtained under optimal conditions. Although the utility of HE has been long established, particularly in conventional small-scale experiments, the use of IR along with HE would be highly beneficial for the evaluation of high-throughput data. IR has been shown to be useful in microarray analyses, where the dose-response data were regarded as nonlinear, monotonic functions (Hu *et al.*, 2005; Harrill *et al.*, 2008). The results described above provide another example showing that IR is useful for dose-response evaluation.

ACKNOWLEDGMENT

This work was supported in part by the Japanese Society for the Promotion of Science (JSPS) KAKENHI [the Grant-in-Aid for Scientific Research (S) No. 23221005].

REFERENCES

Best MJ and Chakravarti N. 1990. Active set algorithms for isotonic regression; a unifying framework. Mathematical Programming 47:425-39.

Byrd RH, Byrd RH, Lu P, Lu P, Nocedal J, Nocedal J, and Zhu C. 1994. A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing 16:1190-208.

Collins FS, Gray GM, and Bucher JR. 2008. Toxicology. transforming environmental health protection. Science 319(5865):906-7.

DeLong ER, DeLong DM, and Clarke-Pearson DL. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics Sep(44(3)):837-45.

Evans TJ, Yamamoto KN, Hirota K, and Takeda S. 2010. Mutant cells defective in DNA repair pathways provide a sensitive high-throughput assay for genotoxicity. DNA Repair (12):1292-8.

Finney DJ. 1979. Bioassay and the practise of statistical inference. Int. Statist. Rev. 47:1-12.

Harrill JA, Li Z, Wright FA, Radio NM, Mundy WR, Tornero-Velez R, and Crofton KM. 2008. Transcriptional response of rat frontal cortex following acute in vivo exposure to the pyrethroid insecticides permethrin and deltamethrin. BMC Genomics 9:546, 2164-9-546.

Hill AV. 1910. The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. The Journal of Physiology 40(Suppl).

Hu J, Kapoor M, Zhang W, Hamilton SR, and Coombes KR. 2005. Analysis of dose-response effects on gene expression data with comparison of two microarray platforms. Bioinformatics 21(17):3524-9.

Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, and Austin CP. 2006. Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries.. Proc Natl Acad Sci U S A 1;103(31):11473.

Ji K, Kogame T, Choi K, Wang X, Lee J, Taniguchi Y, and Takeda S. 2009. A novel approach using DNA-repair-deficient chicken DT40 cell lines for screening and characterizing the genotoxicity of environmental contaminants. Environ Health Perspect 117(11):1737-44.

Jiang X, Osl M, Kim J, and Ohno-Machado L. 2011. Smooth isotonic regression: A new method to calibrate predictive models. AMIA Summits Transl Sci Proc 2011:16-20.

Martin A, Joachim D, and Lieven V. 2012. Interior-point methods for large-scale cone programming. MIT Press.

Martin A, Joachim D, and Lieven V. 2013. CVXOPT: A Python package for convex optimization. Available at http://cvxopt.org/

Mizutani A, Okada T, Shibutani S, Sonoda E, Hochegger H, Nishigori C, Miyachi Y, Takeda S, and Yamazoe M. 2004. Extensive chromosomal breaks are induced by tamoxifen and estrogen in DNA repair-deficient cells. Cancer Res 64(9):3144-7.

National Research Council (NRC). 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. National Research Council of the National Academies, Washington, D.C.

National Toxicology Program (NTP). 2010. Review of the biomolecular screening branch by the NTP board of scientific counselors.

Nojima K, Hochegger H, Saberi A, Fukushima T, Kikuchi K, Yoshimura M, Orelli BJ, Bishop DK, Hirano S, Ohzeki M, Ishiai M, Yamamoto K, Takata M, Arakawa H, Buerstedde JM, Yamazoe M, Kawamoto T, Araki K, Takahashi JA, Hashimoto N, Takeda S, and Sonoda E. 2005. Multiple repair pathways mediate tolerance to chemotherapeutic cross-linking agents in vertebrate cells. Cancer Res 65(24):11704-11.

Perkins NJ and Schisterman EF. 2006. The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 163(7):670-5.

Ritz C. 2010. Toward a unified approach to dose-response modeling in ecotoxicology. Environ Toxicol Chem 29(1):220-9.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, and Muller M. 2011. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12:77,2105-12-77.

Shukla SJ, Huang R, Austin CP, and Xia M. 2010. The future of toxicity testing: A focus on in vitro methods using a quantitative high-throughput screening platform.. Drug Discov Today Dec 15:997-1007.

Swets JA. 1988. Measuring the accuracy of diagnostic systems. Science Jun 3(240(4857)):1285-93.

Thomas CJ, Auld DS, Huang R, Huang W, Jadhav A, Johnson RL, Leister W, Maloney DJ, Marugan JJ, Michael S, Simeonov A, Southall N, Xia M, Zheng W, Inglese J, and Austin CP. 2009. The pilot phase of the NIH chemical genomics center. Curr Top Med Chem 9(13):1181-93.

Tice RR, Austin CP, Kavlock RJ, and Bucher JR. 2012. Transforming public health protection: A U.S. Tox21 progress report. Environ Health Perspect. 121: 756-765.

Wu X, Takenaka K, Sonoda E, Hochegger H, Kawanishi S, Kawamoto T, Takeda S, and Yamazoe M. 2006. Critical roles for polymerase zeta in cellular tolerance to nitric oxide-induced DNA damage. Cancer Res 66(2):748-54.

Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, Jadhav A, Smith CS, Inglese J, Portier CJ, Tice RR, and Austin CP. Compound cytotoxicity profiling using quantitative high-throughput screening. Environ Health Perspect 116(3):284-91.

Yamamoto KN, Hirota K, Kono K, Takeda S, Sakamuru S, Xia M, Huang R, Austin CP, Witt KL, Tice RR. 2011. Characterization of environmental chemicals with potential for DNA damage using isogenic DNA repair-deficient chicken DT40 cell lines. Environ Mol Mutagen 52(7):547-61.

Youden WJ. 1950. Index for rating diagnostic tests. Cancer 3(1):32-5.

SUPPLEMENTARY DATA

Notation and parameter characteristics

Notation	**Description**
HE	Hill equation
IR	Isotonic regression
w	Wild-type
m	Mutant
n	Concentration number in the experiment
c_i	The^i i-th concentration in the experiment
r	Base of concentration in the experiment
x	Concentration $log_{10}c$
p_1, p_2, p_3, p_4	Parameters for Hill equation
EC$_{50}$	Efficacy concentration at 50% maximal
a	Parameter for EC$_{50}$. $a = p_1$
b	Parameter for Hill coefficient. $b = p_2$
ε	Experimental random errors
σ	Standard variation of random error
θ	$a_w - a_m = p_{1,\,w} - p_{1,\,m} = \log_{10}\dfrac{EC_{50,w}}{EC_{50,m}}$
δ	Shift of observation window
v	$b_w - b_m$
ΔS	Difference statistics S between S_w and S_m
ROC	Receiver operating characteristic curve
ROC-AUC	Area under the receiver operating characteristic curve

The definition of genotoxicity and the area between curves.

Each compound, d_j, had dose-survival functions for the wild-type, *w,* and the mutant, *m,* given by $f_w(\boldsymbol{x})$ and $f_m(\boldsymbol{x})$, respectively. For the sake of simplicity, we used a two-parameter HE, $f(y \mid a, b) = \dfrac{1}{1+e^{b(y-a)}}$. We modeled the dose-survival functions of two cell lines with HE as $f_w = f(y \mid a_w, b_w)$ and $f_m = f(y \mid a_m, b_m)$. When $b_w = b_m$, the two curves were parallel and did not intersect, and when $b_w \neq b_m$, $a_w \neq a_m$ they had one intersection at $y_{cross} = \dfrac{a_w b_w - a_m b_m}{b_w b_m}$.

The qHTS system identified genotoxicity from a higher death rate of the mutant compared with that of the wild-type. Therefore, when $b_w = b_m$, the area between two curves,

$$\int_{-\infty}^{\infty} \{f_w(x) - f_m(x)\}\, dx \qquad \text{(S1)}$$

represents the strength of genotoxicity, and this corresponds to $\theta = a_w - a_m$. The area between two curves is identical, as shown by the rectangle on the right side of Supplementary Figure S1A.

When $b_w \neq b_m$, there appeared to be two definitions of genotoxicity strength.

One definition was the same as that provided by equation (S1), when $b_w = b_m$,

$$\int_{-\infty}^{\infty} \{f_w(x) - f_m(x)\}\, dx\,. \qquad \text{(S2)}$$

where a compound resulting in the mutant-specific decrease in viability was counted as a genotoxic compound, and the mutant-specific increase

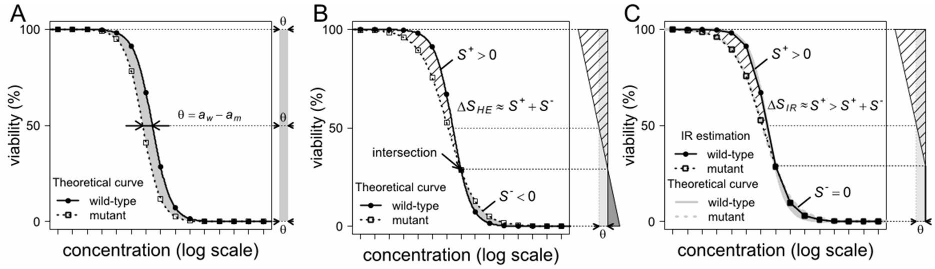

FIGURE S1. The area between the wild-type and mutant curves. (A) When the two curves are parallel, the area is indicated as a grey area. This is identical to the grey rectangle whose width is $\theta = a_w - a_m$, located at the right side of the Figure. (B) When the two curves intersected, the area was divided into S^+ where the wild-type curve was higher than the mutant curve, and S^- where the wild-type curve was lower than the mutant curve. Area S^+ corresponds to the mesh triangle and area S^- corresponds to the dark grey triangle. S^- has a negative value, but S^+ increased by the same amount as S^-. That is, $S^+ + S^-$ is identical to $\theta = \Delta S_{HE}$ (light grey rectangle). (C) One of the restrictions of IR, that the line for the wild-type should not be below the line for the mutant, made the area S^- zero. S^- is shown as a dark grey triangle in Figure S1B, while here, S^- is indicated as a line. That is, $\Delta S_{IR} = S^+$ is larger than $\theta = \Delta S_{HE}$.

in viability was counted as an inverse effect. This condition is shown in Supplementary Figure S1B. Again, the area can be transformed into simple shapes. The area is the upper bigger triangle minus the bottom smaller triangle. This is the same rectangle as the one drawn in panel A that is shadowed. When the upper white triangle is rotated and moved down, it covers the part of the rectangle that was not covered by this triangle and the lower, smaller, dark-shadowed triangle. This explains that the sum of S^+ and S^- is identical with $\theta = a_w - a_m$.

The other definition only counted the mutant-specific decrease in viability and ignored the increase in viability, which can be expressed as:

$$\int_{-\infty}^{\infty} \max\{0, f_w(x) - f_m(x)\}dx \,. \qquad \text{(S3)}$$

Because this only counts the positive triangle (Figure S1C), the area is larger than $\theta = a_w - a_m$.

Calculation of the statistics for isotonic regression (IR) via the trapezoidal rule.

The grey area under the broken lines after IR estimation, S_i, was calculated by the trapezoidal rule and summed through to iteration *i*. The two parallel bases and height correspond to $g(x_i)$, $g(x_{i+1})$, and $log_{10}r$, respectively (Figure S2).

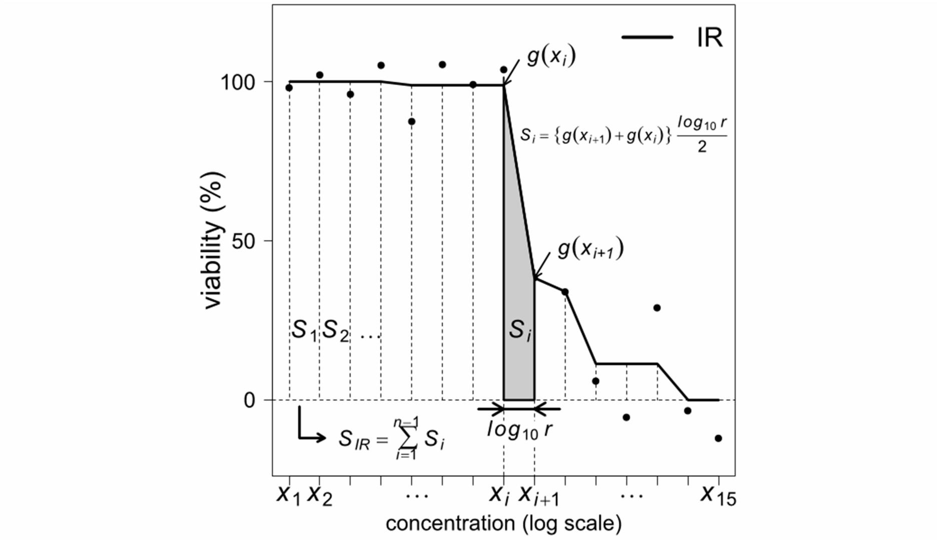

FIGURE S2. Calculation of the statistics for isotonic regression (IR) via the trapezoidal rule. The grey area under the dashed lines after IR estimation, S_i, was calculated by the trapezoidal rule and summed through to iteration *i*. The two parallel bases and the height correspond to $g(x_i)$, $g(x_{i+1})$, and $log_{10}r$, respectively.