BIOINFORMATICS ORIGINAL PAPER Vol. 28 no. 13 2012, pages 1775–1782
doi:10.1093/bioinformatics/bts262

Data and text mining

Advance Access publication May 3, 2012

Two effective methods for correcting experimental high-throughput screening data

Plamen Dragiev[1,2], Robert Nadon[2,3] and Vladimir Makarenkov[1,*]

[1]Département d'Informatique, Université du Québec à Montréal, C.P.8888, s. Centre-Ville, Montréal, QC, H3C-3P8, [2]Genome Quebec Innovation Centre, 740 Dr. Penfield Ave., Montreal, QC, H3A-1A4 and [3]McGill University, Department of Human Genetics, 1205 Dr. Penfield Ave., N5/13, Montreal, QC, Canada, H3A-1B1

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Rapid advances in biomedical sciences and genetics have increased the pressure on drug development companies to promptly translate new knowledge into treatments for disease. Impelled by the demand and facilitated by technological progress, the number of compounds evaluated during the initial high-throughput screening (HTS) step of drug discovery process has steadily increased. As a highly automated large-scale process, HTS is prone to systematic error caused by various technological and environmental factors. A number of error correction methods have been designed to reduce the effect of systematic error in experimental HTS (Brideau *et al.*, 2003; Carralot *et al.*, 2012; Kevorkov and Makarenkov, 2005; Makarenkov *et al.*, 2007; Malo *et al.*, 2010). Despite their power to correct systematic error when it is present, the applicability of those methods in practice is limited by the fact that they can potentially introduce a bias when applied to unbiased data. We describe two new methods for eliminating systematic error from HTS data based on a prior knowledge of the error location. This information can be obtained using a specific version of the t-test or of the χ^2 goodness-of-fit test as discussed in Dragiev *et al.* (2011). We will show that both new methods constitute an important improvement over the standard practice of not correcting for systematic error at all as well as over the B-score correction procedure (Brideau *et al.*, 2003) which is widely used in the modern HTS. We will also suggest a more general data preprocessing framework where the new methods can be applied in combination with the Well Correction procedure (Makarenkov *et al.*, 2007). Such a framework will allow for removing systematic biases affecting all plates of a given screen as well as those relative to some of its individual plates.

Contact: makarenkov.vladimir@uqam.ca

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on January 26, 2012; revised on April 3, 2012; accepted on April 30, 2012

1 INTRODUCTION

A typical drug development project starts with a candidate identification phase in which a large chemical compound library is tested against a given biological target (Malo *et al.*, 2006). Complex high-throughput screening equipment is employed at this stage to obtain precise estimates of compound activity levels. The collected data are then used to identify the compounds that show the most promising 'drug-like' activity behavior (Brideau *et al.*, 2003; Malo *et al.*, 2006). The selected compounds, called hits, typically undergo further testing to confirm their reproducibility and suitability for drug development. Depending on the nature of the study, the hits may be compounds with the highest activation capacity (i.e. activation assays), inhibition capacity (i.e. inhibition assays) or both. The hit selection process assumes that the measurements taken by HTS equipment accurately represent the activity levels of the tested compounds. An important consideration for this to be true is that experimental conditions are the same for all compounds of the screen. Biases in the measurements can nonetheless appear, due to inconsistencies in the environmental factors, such as electricity, temperature, humidity or lighting changes (Heyse, 2002; Makarenkov *et al.*, 2007). Organizational factors can also have a significant systematic impact on the results of an HTS campaign. For example, differences in the incubation time allow the solvent evaporation to cause unintended variations in the solution concentrations. Highly sensitive readers in particular can detect subtle differences among the tested molecules which misdirect follow-up efforts when they are due to bias rather than to biology.

As a result of systematic bias causing under- or over-estimation of biological activity, inactive compounds may be incorrectly selected as hits (false positives), whereas promising (active) compounds may remain undetected (false negatives). In HTS, systematic error is usually column or row dependent (Brideau *et al.*, 2003; Makarenkov *et al.*, 2007). It is important to note that systematic error can either affect compounds placed in the same well, column or row location in all plates of the screen (i.e. screen-specific error) or affect a column or row of a specific single plate of the screen (i.e. plate-specific error).

Figure 1 illustrates the presence of positional effects in two publicly available experimental HTS datasets: McMaster Test dataset, used as a benchmark for the McMaster Data Mining and Docking Competition (Elowe *et al.*, 2005; it contained the compounds intended to inhibit the *Escherichia coli* Dihydrofolate reductase, DHFR) and a dataset provided by the Chemistry Department of Princeton University and consisting of a screen of compounds meant to inhibit the glycosyltransferase *MurG* function of *E. coli* (Helm *et al.*, 2003). Figures 1a, c show activity levels averaged across all plates (i.e. assay background surfaces), whereas Figures 1b, d show the activity levels of two selected single plates (from the McMaster and Princeton datasets, respectively). These

*To whom correspondence should be addressed.

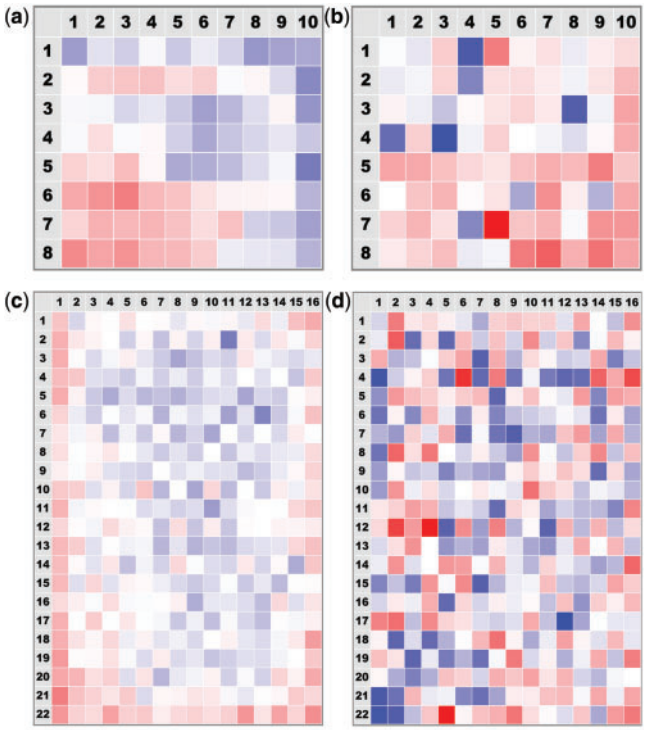

Fig. 1. Hit maps showing the presence of positional effects in the McMaster 1250-plate assay (Elowe *et al*. 2005)—(**a**) whole assay background surface, (**b**) plate 1036 measurements; and in the Princeton 164-plate assay (Helm *et al*. 2003)—(**c**) whole assay background surface, (**d**) plate 144 measurements. Color intensity is proportional to the compounds' signal levels (higher signals, i.e. potential target inhibitors, are shown in red)

examples demonstrate that systematic biases in HTS may have different screen-specific and plate-specific systematic deviations. For instance, in the McMaster dataset, the measurements in the column 10 are globally over-estimated (Fig. 1a), but in plate 1036 they are rather under-estimated (Fig. 1b). Similarly, Figure 1c reveals apparent 'edge effects' in the Princeton dataset with the values of the outer rows and columns being below the screen average. This effect was not observed, however, for all plates of the Princeton screen, with an evident over-estimation of the first column measurements detected in plate 144 (Fig. 1d). Thus, systematic error correction methods should be able first to recognize the character of systematic error affecting the data at hand and then remove it either from the whole assay and/or only from the specific plates where it was detected. In this article, we describe two new methods for eliminating plate-specific systematic error and show how these methods can be applied in a more general correction framework that also includes the Well Correction procedure (Makarenkov *et al.*, 2007) which allows for removing screen-specific systematic biases.

2 METHODS

2.1 Data preprocessing in HTS

To analyze experimental HTS assays, a data preprocessing treatment should be performed before the hit selection. Several data normalization and correction techniques, including the step of the quality control, have been proposed to preprocess experimental HTS data (Brideau *et al.*, 2003; Carralot *et al.*, 2012; Dragiev *et al.*, 2011; Kevorkov and Makarenkov, 2005; Makarenkov *et al.*, 2007; Malo *et al.*, 2006, 2010; Shun *et al.*, 2011; Zhang, 2008; Zhang *et al.*, 1999). The most popular data normalization procedures used in HTS are as follows: Percent of control that normalizes the given compound measurements relative to the mean value of the plate's positive controls; Normalized percent inhibition in which the normalization is carried out relative to both positive and negative controls; and Z-score that consists in a zero mean and unit SD normalization of the plate's measurements (Malo *et al.*, 2006). Regarding data correction, mention the B-score (Brideau *et al.*, 2003) and Well Correction (Makarenkov *et al.*, 2006, 2007) methods which will be considered in this study. Their main steps of these methods are as follows:

B-score (Brideau *et al.*, 2003) is a robust normalization procedure commonly used in experimental HTS. Similarly to the above-mentioned normalizations, B-score sensibly handles plate-to-plate variability. In addition, it also corrects the raw plate measurements by removing the existing row and column positional effects. It assumes the following statistical model of HTS measurements [Equation (1)]:

$$x_{ijp} = \mu_p + R_{ip} + C_{jp} + \varepsilon_{ijp}, \tag{1}$$

where x_{ijp} is the raw measurement of the compound in well (i, j) of a given plate p, μ_p is the plate average, R_{ip} is the systematic error affecting row i, C_{jp} is the systematic error affecting column j and ε_{ijp} is the random noise affecting well (i, j) of this plate. B-score first uses a two-way median polish (MP) procedure (Tukey, 1977) to obtain the estimated values of x_{ijp}, μ_p, R_{ip} and C_{jp} [Equation (2)]:

$$\hat{x}_{ijp} = \hat{\mu}_p + \hat{R}_{ip} + \hat{C}_{jp}. \tag{2}$$

The residual, r_{ijp}, for the measurement in well (i, j) is then calculated as the difference between the raw measurement x_{ijp} and its fitted value $\hat{x}_{ijp}$: $r_{ijp} = x_{ijp} - \hat{x}_{ijp}$. Finally, the raw compound measurement is replaced with the corresponding residual adjusted by the plate's median absolute deviation [MAD_p, Equation (3)]:

$$x'_{ijp} = \frac{r_{ijp}}{\mathrm{MAD}_p}, \quad \mathrm{MAD}_p = \mathrm{median}\,\{|r_{ijp} - \mathrm{median}\,(r_{ijp})|\}, \tag{3}$$

where x'_{ijp} is the normalized measurement value.

Well Correction (Makarenkov *et al.*, 2006, 2007) is another combined data normalization and correction method designed to compensate for positional effects affecting rows, columns or individual wells, and appearing in all plates of the screen (i.e. screen-specific error). Well Correction includes the two following steps:

(1) For each well location of the screen, a linear or polynomial least-squares approximation is carried out for the compound measurements located in that well over all plates of the screen. This approximation is performed separately for each well location.

(2) The approximated entities within the same well location are then normalized over all plates of the screen using Z-score. This normalization is performed separately for each well location.

Once the data normalization and correction steps are completed, a hit selection procedure, meant to identify the compounds that will be promoted to leads, is carried out. The most popular strategy for hit selection proceeds by the identification of the compounds whose activity levels exceed a predefined threshold (Malo *et al.*, 2006). Typically, the hit selection threshold is expressed in terms of the mean, μ, and the SD of the observed measurements. A commonly used approach selects as hits the compounds whose activity levels deviate from the mean value μ for >3SD.

Despite their ability to eliminate systematic error, HTS preprocessing techniques cannot guarantee the recovery of correct hits. In our previous works (Dragiev *et al.*, 2011; Makarenkov *et al.*, 2007), we showed that a misapplication of error correction methods on error-free HTS data introduces a significant bias that affects very negatively the accuracy of the hit selection process. For instance, a simulation study described in Makarenkov *et al.*

(2007) suggests that the B-score method is unable to cope with screen-specific systematic error (Figs 2 and 3 in the latter study) and that the Well Correction method is not suited for eliminating plate-specific systematic error (Fig. 4 in the latter study). Hence, error correction methods should be used with caution and only when the presence of systematic noise in the data has been confirmed by statistical tests. In our recent work (Dragiev *et al.*, 2011), we described how individual HTS plates can be assessed for presence of systematic error, thus facilitating the decision regarding the application of data correction techniques.

2.2 Two new data correction methods

Here we present two new methods for HTS systematic error correction, called Matrix Error Amendment (MEA) and Partial Mean Polish (PMP). Both methods rely on prior information concerning the location of rows and columns of individual plates that are systematically over- or under-estimated. Such information might be available through the analysis of an individual plate (or entire screen) background (Kevorkov and Makarenkov, 2005) or can be acquired using a specific version of the t-test or of the χ^2 goodness-of-fit test (Dragiev *et al.*, 2011; see also the Supplementary Materials section for the application of these tests in the HTS context). Both MEA and PMP methods are applied on a plate-by-plate basis.

Let X be a plate of HTS measurements with m rows and n columns. Let x_{ij} be the measurement of the compound located in well (i, j) of X and let μ be the mean value of all measurements of plate X that are not affected by systematic error.

In the case when plate X is free of systematic error, we can expect that the mean of the values in a given row i ($i=1, \ldots, m$) does not deviate substantially from μ, which in this case is the mean of all measurements on the plate: $\sum_{j=1}^{n} x_{ij} \approx n\mu$. Similarly, for a given column j ($j=1, \ldots, n$) of X, we expect that: $\sum_{i=1}^{m} x_{ij} \approx m\mu$.

Assume that X is affected by systematic error. Let $r_1, r_2, \ldots, r_p$ ($p<m$) be the set of rows of X, and $c_1, c_2, \ldots, c_s$ ($s<n$) be the set of columns of X, where the presence of systematic error has been confirmed. It is worth noting that the set $r_1, r_2, \ldots, r_p$ can represent any subset of the complete set of rows 1, 2, …, m and the set $c_1, c_2, \ldots, c_s$ can represent any subset of the complete set of columns 1, 2, …, n of plate X. The only necessary condition for the application of the new methods is the presence in X of at least one row and at least one column not affected by systematic error. Let e_{r_i} be the unknown value of systematic error affecting row r_i and e_{c_j} be the unknown value of systematic error affecting column c_j. The following 4-fold set of linear equations can be composed:

$$\sum_{j=1}^{n} x_{r_i j} - n e_{r_i} - \sum_{j=1}^{s} e_{c_j} = n\mu, \tag{4}$$

$$\sum_{i=1}^{m} x_{i c_j} - m e_{c_j} - \sum_{i=1}^{p} e_{r_i} = m\mu, \tag{5}$$

$$\sum_{j=1}^{n} x_{ij} - \sum_{j=1}^{s} e_{c_j} = n\mu, \tag{6}$$

$$\sum_{i=1}^{m} x_{ij} - \sum_{i=1}^{p} e_{r_i} = m\mu, \tag{7}$$

where Equation (4) corresponds to rows $r_1, r_2, \ldots, r_p$ affected by row systematic error, Equation (5) to columns $c_1, c_2, \ldots, c_s$ affected by column systematic error, Equation (6) to rows not affected by row systematic error and Equation (7) to columns not affected by column systematic error.

MEA method

Systematic error in HTS does not typically affect all the columns and rows of a plate. The affected columns and rows are often those located on the plate edges (Brideau *et al.*, 2003; Kevorkov and Makarenkov, 2005). Thus, typically, p is much smaller than m and s is much smaller than n. The presence of rows and columns not affected by systematic error allows us to estimate μ and leaves e_{r_i} and e_{c_j} the only unknowns in the linear system of equations (4)–(7), which have $m+n$ equations and fewer than $m+n$ unknowns.

The MEA method consists of the two following steps:

(1) Estimate the values of the row and column systematic errors $\hat{e}_{r_i}$ and $\hat{e}_{c_j}$ ($i = 1, \ldots, p$ and $j=1, \ldots, s$), independently for every plate of the assay, by solving the system of linear equations (4)–(7).

(2) Adjust the measurements of all compounds located in rows and columns of the plates affected by systematic error using the error estimates $\hat{e}_{r_i}$ and $\hat{e}_{c_j}$ determined in Step 1.

Two approaches of solving the system of linear equations (4)–(7) were tested in our study. First, by combining all Equations (4)–(7), we composed an over-determined system of linear equations **Ae=b** with $m+n$ equations and fewer than $m+n$ unknowns, where **A** was the matrix of the coefficients for the unknowns e_{r_i} and e_{c_j} ($i=1, \ldots, p$ and $j=1, \ldots, s$) combined in the vector **e** of size $p+s$, and **b** was the vector of free terms. We found that in all cases the matrix $\mathbf{A^T A}$ was singular, thus rendering inapplicable the standard least-square approximation method for solving over-determined systems of linear equations. We were able, however, to find an approximate solution of this system by using the singular value decomposition (SVD) method. Second, we also tested a simpler and computationally less intensive approach consisting of combining only Equations (4) and (5) into the linear system (8), having exactly $m+n$ equations and $m+n$ unknowns. When $m+n>5$, the system (8) always has a unique solution which can be found using standard methods for solving linear equations systems (e.g. Gaussian elimination).

$$\begin{pmatrix} n & 0 & \ldots & 0 & 0 & 1 & 1 & \ldots & 1 & 1 \\ 0 & n & \ldots & 0 & 0 & 1 & 1 & \ldots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & n & 0 & 1 & 1 & \ldots & 1 & 1 \\ 0 & 0 & \ldots & 0 & n & 1 & 1 & \ldots & 1 & 1 \\ 1 & 1 & \ldots & 1 & 1 & m & 0 & \ldots & 0 & 0 \\ 1 & 1 & \ldots & 1 & 1 & 0 & m & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \ldots & 1 & 1 & 0 & 0 & \ldots & m & 0 \\ 1 & 1 & \ldots & 1 & 1 & 0 & 0 & \ldots & 0 & m \end{pmatrix} \begin{pmatrix} e_{r_1} \\ e_{r_2} \\ \vdots \\ e_{r_{p-1}} \\ e_{r_p} \\ e_{c_1} \\ e_{c_2} \\ \vdots \\ e_{c_{s-1}} \\ e_{c_s} \end{pmatrix} = \begin{pmatrix} b_{r_1} \\ b_{r_2} \\ \vdots \\ b_{r_{p-1}} \\ b_{r_p} \\ b_{c_1} \\ b_{c_2} \\ \vdots \\ b_{c_{s-1}} \\ b_{c_s} \end{pmatrix}, \tag{8}$$

where $b_{r_i} = \sum_{j=1}^{n} x_{r_i j} - n\mu$ and $b_{c_j} = \sum_{j=1}^{m} x_{i c_j} - m\mu$.

According to our simulation study, the second approach, which requires less computer power, generally provided better results in terms of systematic error identification (i.e. it yielded a higher hit detection rate, see Section 3.1). Thus, its detailed results are presented in Section 3.

The final step of the MEA method proceeds by subtracting the obtained systematic error estimates $\hat{e}_{r_i}$ and $\hat{e}_{c_j}$ from the raw plate measurements [Equations (9)–(10)]. For all rows r_i ($i=1, \ldots, p$) affected by systematic error, we have:

$$x'_{r_i j} = x_{r_i j} - \hat{e}_{r_i}, \text{ for all } j\colon 1 \leqslant j \leqslant n, \tag{9}$$

and for all columns c_j ($j=1, \ldots, s$) we have:

$$x'_{i c_j} = x_{i c_j} - \hat{e}_{c_j}, \text{ for all } i\colon 1 \leqslant i \leqslant m. \tag{10}$$

PMP method

Denote by μ_i the mean value of all measurements in row i and by μ_j the mean value of all measurements in column j of plate X:

$$\mu_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \text{ and } \mu_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}.$$

Equations (4) and (5) can be rewritten as Equations (11) and (12):

$$n e_{r_i} = \sum_{j=1}^{n} x_{r_i j} - n\mu - \sum_{j=1}^{s} e_{c_j}, \tag{11}$$

$$m e_{c_j} = \sum_{i=1}^{m} x_{i c_j} - m\mu - \sum_{i=1}^{p} e_{r_i}, \tag{12}$$

where μ is the mean value of all measurements of X not affected by systematic error.

Dividing Equations (11) and (12) by n and m, respectively, we obtain:

$$e_{r_i} = \mu_{r_i} - \mu - \frac{1}{n}\sum_{j=1}^{s} e_{c_j}, \quad (13)$$

$$e_{c_j} = \mu_{c_j} - \mu - \frac{1}{m}\sum_{i=1}^{p} e_{r_i}. \quad (14)$$

Since systematic error usually affects only a few columns and rows of HTS plates [e.g. row and column measurements on plate edges are often biased; for more details see Brideau *et al.* (2003) or Kevorkov and Makarenkov (2005)] and causes an over- or under-estimation of the affected measurements (i.e. the error values can be negative or positive), we can assume that the term consisting of the total column error divided by the number of columns has a negligible impact compared with the other terms in Equation (13) and thus that the row systematic error of row r_i can be estimated as the difference between the mean value of the entities in that row and the mean value μ of the plate measurements that are not affected by systematic error:

$$\hat{e}_{r_i} = \mu_{r_i} - \mu. \quad (15)$$

Similarly, for the column c_j, we can expect that:

$$\hat{e}_{c_j} = \mu_{c_j} - \mu. \quad (16)$$

Based on the assumptions above, we can formulate the PMP iterative procedure (only a part of the plate's rows and columns, i.e. those affected by systematic bias, will be 'polished' by the method). The means in this procedure can be easily replaced by the medians giving rise to the Partial Median Polish method which could be viewed as an extension of a well-known Median Polish procedure by Tukey (1977) for the case when the error locations are known.

The main steps of the PMP method are the following:

(1) Compute the mean value μ of all entities of the given plate that are not affected by systematic error: $\mu = \frac{\sum_{i\notin R, j\notin C} x_{ij}}{(m-p)(n-s)}$, where $R=\{r_1, r_2, \ldots, r_p | 0 \le p < m\}$ is a set of rows of X affected by systematic error and $C=\{c_1, c_2, \ldots, c_s | 0 \le s < n\}$ is a set of columns of X affected by systematic error.

(2) For each i $(1 \le i \le p)$, compute the mean value, μ_{r_i}, of row r_i as $\mu_{r_i} = \frac{1}{n}\sum_{j=1}^{n} x_{r_i j}$, and then, using Equation (15), the estimate of the row bias $\hat{e}_{r_i}$ as: $\hat{e}_{r_i} = \mu_{r_i} - \mu$.
For each j $(1 \le j \le s)$ compute the mean value, μ_{c_j}, of column c_j as $\mu_{c_j} = \frac{1}{m}\sum_{i=1}^{m} x_{ic_j}$, and then, using Equation (16), the estimate of the column bias $\hat{e}_{c_j}$ as: $\hat{e}_{c_j} = \mu_{c_j} - \mu$.

(3) For all rows affected by systematic bias, adjust their measurements using the error estimates determined in Step 2, i.e. for each i $(1 \le i \le p)$, and for each j $(1 \le j \le n)$: $x_{rij} = x_{rij} - \hat{e}_{r_i}$.
For all columns affected by systematic error, adjust their measurements using the error estimates determined in Step 2, i.e. for each j $(1 \le j \le s)$, and for each i $(1 \le i \le m)$: $x_{ic_j} = x_{ic_j} - \hat{e}_{c_j}$.

(4) Compute the value of the convergence parameter δ: $\delta = \sum_{i=1}^{p} |\hat{e}_{r_i}| + \sum_{j=1}^{s} |\hat{e}_{c_j}|$.

(5) If $\delta < \varepsilon$, where ε is a selected convergence threshold, or if a fixed maximum number of iterations has been already carried out, then return X, otherwise, repeat Steps 2–5.

3 RESULTS AND DISCUSSION

To evaluate the performances of the two introduced systematic error correction methods we first carried out simulations with artificially generated HTS measurements. We also applied both MEA and PMP methods to analyse the 1250-plate HTS screen produced at the HTS Laboratory of McMaster University (i.e. the Test dataset proposed as a benchmark for the McMaster Data Mining and Docking Competition, see Fig. 1 and Elowe *et al.*, 2005).

3.1 Simulation study

The simulated data also consisted of 1250-plate assays. Plate sizes were 96-well plates (8 rows × 12 columns), 384-well plates (16 rows × 24 columns) and 1536-well plates (32 rows × 48 columns). Inactive compound measurements were generated according to the standard normal distribution. Active compounds (hits) were added randomly to the plates to form assays with the following hit percentages: 0, 0.5, 1, 2, 3, 4 and 5%. Hit locations were chosen randomly within each plate (i.e. the probability that a given well contained a hit compound was the same for all wells of the plate, regardless of the well location within the plate). The hit measurements were generated according to the normal distribution with parameters $\sim N(\mu\text{–}5\text{SD}, \text{SD})$, where μ and SD were the mean and standard deviation of the original dataset (obtained before the addition of hits; i.e. $\mu = 0$ and SD = 1). Systematic row and column errors were added to randomly selected rows and columns of each plate. The rows and columns affected by systematic error were selected separately for each plate, and thus their locations differed from plate to plate. The values of systematic bias followed a normal distribution with parameters $\sim N(0, C)$. The following values of the error, C, were considered to generate assays affected by different degree of systematic error: 0, 0.6, 1.2, 1.8 and 2.4SD. To mimic empirical HTS data, in our first simulations the effect of systematic error was limited to a few rows and columns only. Thus, at most 2 rows and 2 columns for 96-well plates, at most 4 rows and 4 columns for 384-well plates and at most 8 row and 8 columns for 1536-well plates were affected by systematic bias. A small random error was also added to both hit and non-hit measurements. The random error in all datasets followed a normal distribution with parameters $\sim N(0, 0.6\text{SD})$.

Equation (17) specifies the model we used to generate an error-affected measurement of the compound located in well (i, j) of plate p:

$$x'_{ijp} = x_{ijp} + e_{r_{ip}} + e_{c_{jp}} + \text{rand}_{ijp}, \quad (17)$$

where x'_{ijp} is the resulting measurement value, x_{ijp} is the original error-free measurement, $e_{r_{ip}}$ is the systematic error affecting row i of plate p, $e_{c_{jp}}$ is the systematic error affecting column j of plate p and rand_{ijp} is the random error in well (i, j) of plate p.

Six data correction/hit selection methods were tested in our simulations. All tested methods comprised an identical hit selection step, but differed in the way the data were processed before the hit selection. The hits were selected globally for each assay using the hit selection threshold of $\mu_{hs} - 3\text{SD}_{hs}$ (i.e. all compounds with the measurements lower than $\mu_{hs} - 3\text{SD}_{hs}$ were declared hits, where μ_{hs} and SD_{hs} were, respectively, the mean and SD of the entire assay after the addition of hits and systematic error). The six methods evaluated in our simulation study were the following:

- Original data processing without any data correction;
- B-score correction method (Brideau *et al.*, 2003);
- MEA method performed under the assumption that the exact locations of the error-affected rows and columns on each plate of the assay are known;
- MEA method performed for the rows and columns where systematic error was detected by the t-test (for more details, see Dragiev *et al.*, 2011);

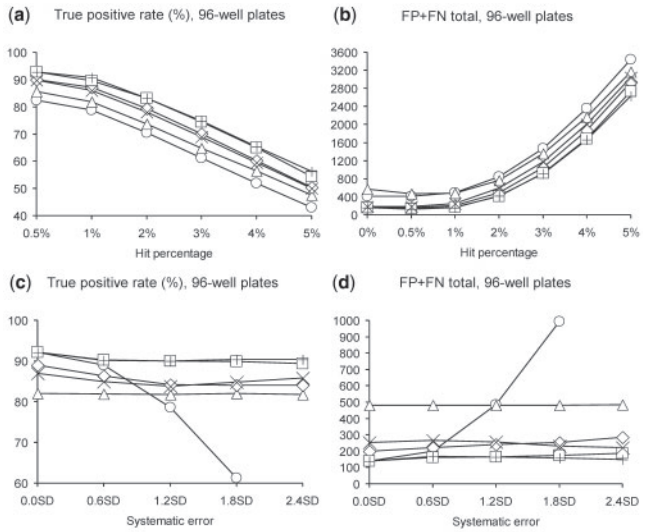

Fig. 2. True positive rate and total number of false positive and false negative hits (i.e. total number of false conclusions) per assay for 96-well plate assays estimated under the condition that at most two rows and two columns of each plate were affected by systematic error. Panels (**a**) and (**b**) present the results obtained for datasets with the fixed systematic error SD of 1.2SD. Panels (**c**) and (**d**) present the results for datasets with the fixed hit percentage rate of 1%. Methods legend: no-correction (◯), B-score (Δ), MEA (□), *t*-test and MEA (◇), PMP (+), *t*-test and PMP (×)

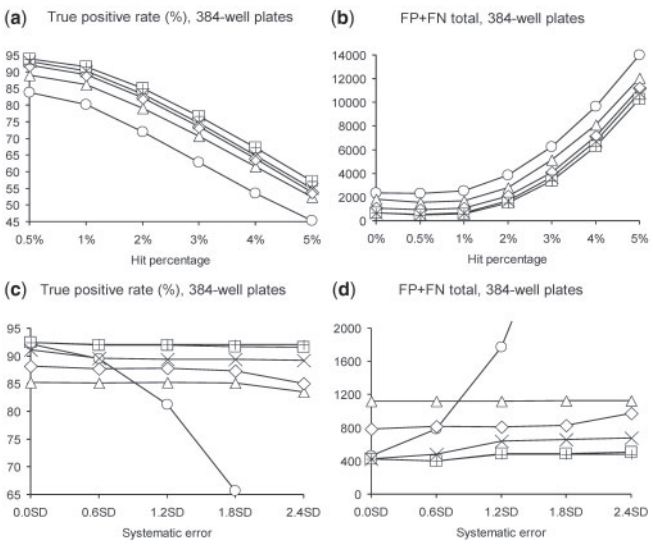

Fig. 3. True positive rate and total number of false positive and false negative hits per assay for 384-well plate assays estimated under the condition that at most four rows and four columns of each plate were affected by systematic error. Figure 2 panel description applies here

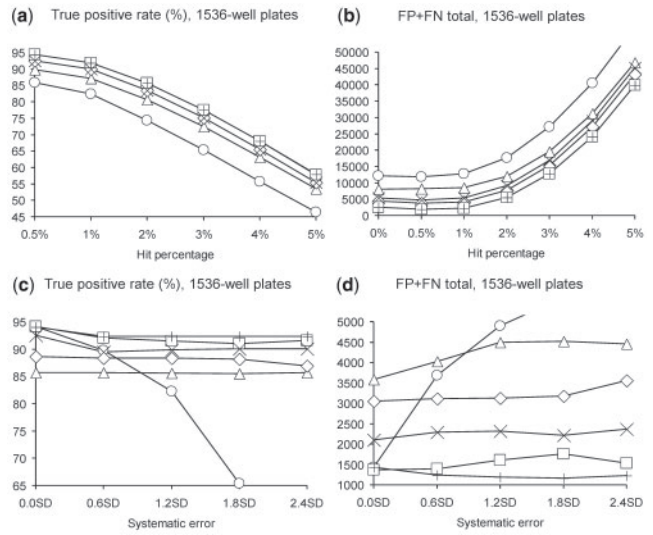

Fig. 4. True positive rate and total number of false positive and false negative hits per assay for 1536-well plate assays estimated under the condition that at most of eight rows and eight columns of each plate were affected by systematic error. Figure 2 panel description applies here

- PMP method performed under the assumption that the exact locations of the error-affected rows and columns on each plate of the assay are known;
- PMP method performed for the rows and columns where systematic error was detected by the *t*-test (for more details, see Dragiev *et al.*, 2011).

In all experiments, we assessed the performances of the six data preprocessing methods by measuring the total number of false positives and false negatives, and by estimating the methods hit detection rate (i.e. true positive rate).

We conducted two series of experiments to evaluate the methods performances depending on the hit percentage and the variance of systematic error. The first series of experiments used datasets with the fixed systematic error of 1.2SD and the hit percentage rate varying from 0% to 5% (there are no true positives for the case of 0% of hits; see Figs 2–4a).

The second series of experiments considered datasets with the fixed hit percentage of 1% and the systematic error varying from 0 to 2.4SD. Some 500 datasets were generated for both series of experiments and for each parameter combination. Figures 2–4 present the average results obtained for the two series of experiments for the 96-well, 384-well and 1536-well plates, respectively.

Furthermore, we conducted additional simulations to assess the performances of the MEA and PMP methods in the situation when up to 50% of the plates' rows and columns were affected by systematic bias. The graphics depicting relative performances of the MEA, PMP, B-score and no-correction strategies in this case are presented in Supplementary Figures 1S–3S.

The simulation results suggest that both proposed methods outperformed the B-score and no-correction procedures when the number of the plate's rows and columns affected by systematic error was low (e.g. in case of commonly observed edge effects), regardless of plate size, hit rate and systematic error variance (see Figs 2–4). In the situations when the number of affected rows and columns of each plate affected by systematic bias could attain 50%

of the plate's total number of rows and columns (see Supplementary Figs 1S–3S), the MEA and PMP methods generally yielded better results than B-score when the hit percentage was under 3% (see Supplementary Figs 1S–3S, cases a and b) or when the level of systematic error was under 1.8SD (see Supplementary Figs 1S–3S, cases c and d). However, in the situations when the hit percentage or systematic error variance was high, the B-score procedure generally showed a more stable behavior than the new methods. This was largely due to the fact that the performance of the *t*-test, carried out prior to MEA and PMP, decreases as the amount of data affected by systematic error grows (Dragiev *et al.*, 2011). In general, the MEA method turned out to be the best performing method for correcting systematic error within 96-well plates when the systematic error variance or the hit percentage was low (see Fig. 2 and Supplementary Fig. 1S), whereas the PMP method provided better results than MEA for the 96-well plates when the systematic error variance or the hit percentage was elevated as well as for the 384- and 1536-well plates (see Figs 3 and 4; Supplementary Figs 2S and 3S). It is worth noting that the B-score method was very prone to generating false positives.

3.2 Analysis of the McMaster Test assay

We carried out the MEA and PMP methods to analyse the McMaster Data Mining and Docking Competition Test assay (see Elowe *et al.*, 2005 and Fig. 1a and b). We examined their impact on the hit identities determined during the HTS phase of the project. This dataset consisted of 625, 96-well plates (with 8 rows and 12 columns) screened in duplicate. Columns 1 and 12 of all plates contained controls and thus were not considered in our study. The assay conditions were identical for all plates. They were as follows: Each 200 μl reaction mixture contained 40 μM NADPH, 30 μM DHF, 5 nM DHFR, 50 mM Tris (pH 7.5), 0.01% (w/v) Triton and 10 mM β-mercaptoethanol. The compounds from the screening library were added to the reaction before initiation by enzyme at a final concentration of 10 μM. All measurements were taken at 25°C.

The threshold of μ–2.29SD was used to identify hits. This threshold led to the identification of 96 average hits which were reported by the competition organizers (Elowe *et al.*, 2005). Our previous works showed that the measurements in the McMaster Test dataset were affected by systematic error (Dragiev *et al.*, 2011; Makarenkov *et al.*, 2007), especially when some higher hit selection thresholds were used (e.g. μ–SD or μ–2SD). The hit sets provided by the six following methods were compared: uncorrected data processing, B-score, and the introduced MEA and PMP methods applied as such and in the combination with the Well Correction procedure (Makarenkov *et al.*, 2007) allowing for removing screen-specific systematic error. Both MEA and PMP methods were carried out on a plate-by-plate basis and were preceded by the *t*-test, which was necessary to recover systematic error row and column locations. The *t*-test was performed with the α parameter value set to 0.01 (see Supplementary Materials). As the McMaster Test dataset contained replicates, the hit selection procedure was adjusted to search for average hits (i.e. the average of the two measurements of every compound was calculated and the obtained result was supplied to the hit selection procedure). The totals of hits retraced by the six considered methods are presented in Table 1 and Supplementary Tables S1–S12 (the detailed results).

Both proposed methods identified more potential hits (100 for MEA and 115 for PMP) than the organizers of the McMaster competition (i.e. 96 hits for the uncorrected dataset), while rejecting a few of the original hits as false positives. The MEA method found 8 extra hits, while rejecting 4 of the original hits as false positives. The PMP method extended the set of original hits with 24 new hits, while rejecting only 5 of them. In contrast, the B-score method rejected 28 original hits, and provided 118 new potential hits (according to our simulation results, many of those new hits can be in fact false positives). The total overlap of all the six considered methods consisted in 55 consensus average hits that could be recommended for further testing including the structure-activity relationships (SARs) analysis and various clinical trials. As shows the example of the consensus hits set of the McMaster Test assay [see Elowe *et al.* (2005) or Table 9sm in Makarenkov *et al.* (2007)], consensus hits can also contain an important percentage of false negatives and false positives. The consensus hits list of this assay, which included 42 hit compounds in total, comprised only 14 of 26 hit compounds confirmed by the SAR analysis conducted by the McMaster competition organizers (i.e. 12 of 26 confirmed hits were false negatives and 28 of 42 consensus hits were false positives). Thus, SAR investigations should be always conducted in conjunction with data correction and hit selection techniques to confirm the selected hits.

Table 1. Number of hits selected by the six data correction methods for the McMaster Test dataset

Data correction method	Number of hits
No-correction	96
B-score	186
Matrix Error Amendment (MEA)	100
Partial Mean Polish (PMP)	115
Well Correction + MEA	109
Well Correction + PMP	109

The hit selection threshold of μ–2.29SD was used.

It is worth also noting that MEA and PMP agreed on most of the hits they selected (i.e. 92 of the hits identified by MEA were also detected by PMP). Furthermore, after the application of Well Correction, the MEA and PMP methods provided an identical set of 109 hits. Figure 5 and Supplementary Tables 1S–12S present the hit distribution surfaces (i.e. hit totals obtained for each well location and computed over all plates of the given assay) of the Master Test assay obtained for the hit selection thresholds μ–SD and μ–2.29SD.

The consecutive application of two data correction methods: Well Correction and MEA (Fig. 5i and j) or Well Correction and PMP (Fig. 5k and l), allowed us to eliminate screen-specific systematic error, first, and plate-specific systematic error, second (see also Supplementary Tables 9S to 12S). For instance, the MEA and PMP hit distribution surfaces provide better fits to the corresponding plain surfaces (which represent a perfect uniform distribution of the assay hits across all well locations) when Well Correction is applied beforehand (Fig. 5i and k). After the application of Well Correction, the hit distribution surface χ^2 goodness-of-fit statistic for the hit selection thresholds μ–SD decreased from 1178.53 (Fig. 5e and Supplementary Table S5) to 203.18 for MEA (Fig. 5i and Supplementary Table 9S) and from 994.27 (Fig. 5g and Supplementary Table 7S) to 198.68 for PMP (Fig. 5k and Supplementary Table 11S).

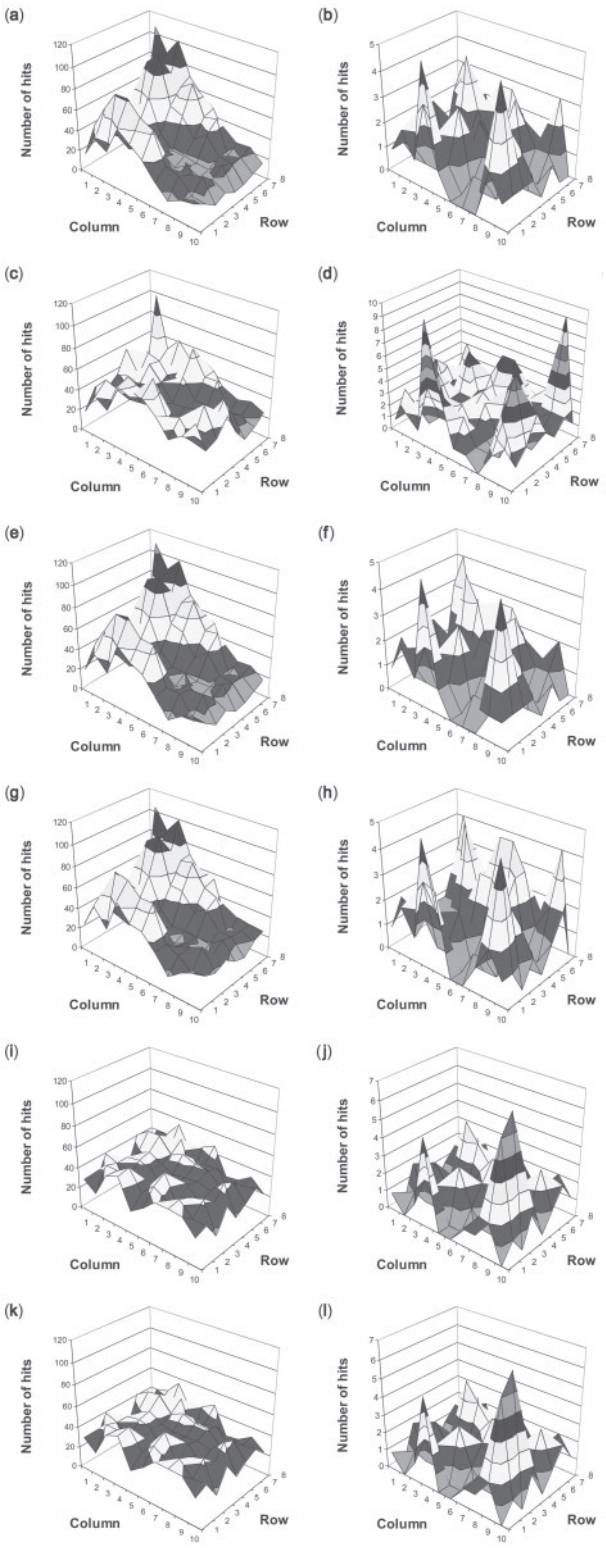

Fig. 5. Hit distribution surfaces of the McMaster Test dataset for the hit selection thresholds μ–SD (cases **a**, **c**, **e**, **g**, **i** and **k**) and μ–2.29SD (cases **b**, **d**, **f**, **h**, **j** and **l**) obtained for: the raw (i.e. uncorrected) data (a, b), and the data corrected by B-score (c, d), MEA (e, f), PMP (g, h), Well Correction + MEA (i, j) and Well Correction + PMP (k, l)

4 CONCLUSION

We described two new methods, called MEA and PMP, allowing for elimination of plate-specific systematic error from experimental HTS data. Both methods rely on the prior information concerning the location of the rows and columns of the given plate affected by systematic bias. Such information can be obtained by using the methodology described in Dragiev *et al.* (2011).

We conducted a simulation study with different HTS plate sizes, hit percentages and systematic error magnitudes. In this study, the MEA and PMP methods were compared with the B-score (Brideau *et al.*, 2003) and no-correction strategies. Both new methods always outperformed the B-score and no-correction procedures when the number of the plate's rows and columns affected by systematic error was low (Figs 2–4). In the simulations where the number of rows and columns affected by systematic error could reach 50% of the plate's total number of rows and columns (Figs 1S–3S), the MEA and PMP methods generally yielded better results than B-score when the hit percentage was under 3% (in a typical HTS campaign the hit percentage is usually under 1%) or when the level of systematic error was under 1.8SD. The B-score method showed a more stable behavior than MEA and PMP only when the number of rows and columns affected by systematic error, hit percentage and systematic error variance were high (mainly due to a mediocre performance of the *t*-test in this case). MEA was generally the best method for correcting systematic error within 96-well plates, whereas PMP performed better for 384 and 1536-well plates.

The analysis of the McMaster Data Mining and Docking Competition Test assay (Elowe *et al.*, 2005) showed that the new methods can be also applied in the combination with the Well Correction technique (Makarenkov *et al.*, 2007) aiming to remove screen-specific systematic error. Hence, a general data correction phase in HTS, permitting for the elimination of both screen- and plate-specific systematic biases, can be conducted in the following way:

(1) **Normalize** the raw measurements using Percent of control, Normalized percent inhibition or Z-score transformation. This normalization step can be carried out either on a plate-by-plate basis or for all assay measurements together (i.e. when all plates have been processed under the same experimental conditions);

(2) **Perform** the *t*-test or χ^2 goodness-of-fit test on the hit distribution surface for the selected hit selection threshold; **if** systematic error is detected **then** carry out the Well Correction method;

(3) **Perform** the *t*-test or χ^2 goodness-of-fit test on each individual plate of the assay to identify its rows and columns affected by systematic error as well as the error locations;

(4) **For** all plates where systematic error is detected: Correct the plate measurements by carrying out the PMP or MEA method (or, alternatively, the B-score procedure).

In this study, we addressed the issue of the commonly considered additive systematic artifact that can be described using Equation (17). It is worth noting that the multiplicative type of systematic bias affecting well (i,j) of plate p and defined by

Equation (18):

$$x'_{ijp} = x_{ijp} \times e_{r_{ip}} \times e_{c_{jp}} + \text{rand}_{ijp}, \quad (18)$$

can be also treated using the proposed methods. Whereas the MEA method should undergo substantial changes to treat multiplicative type of systematic error because the linear equations systems (4)–(7) and (8) will be transformed into the corresponding nonlinear equations systems, the PMP method can be easily adapted for the identification and correction of multiplicative bias by adding the following equations: $\hat{e}_{r_i} = \mu_{r_i}/\mu$ and $\hat{e}_{c_j} = \mu_{c_j}/\mu$ to Step 2, and then $x_{r_i j} = x_{r_i j}/\hat{e}_{r_i}$ and $x_{ic_j} = x_{ic_j}/\hat{e}_{c_j}$ to Step 3, of the method instead of the corresponding equations containing the subtraction sign.

A version of the PMP method, in which a median is used instead of the mean, could be viewed as a direct extension of the well-known median polish algorithm (Tukey, 1977), applicable in the situations when the exact error location is known (the traditional MP assumes that systematic error is present in all rows and columns of the given matrix). Another advantage of the PMP method over MP and its B-score analog is that our method does not reduce the original data to residuals, keeping the corrected data on the same scale with the original ones and not modifying the unbiased data at all. Moreover, both of the proposed methods could be interesting in general, from the statistical point of view, and applied as data correction methods in any other field.

A new program implementing the two data correction methods described in this article, and including also the Well Correction, B-score and Z-score procedures, is freely available at the following URL: http://www.info2.uqam.ca/~makarenkov_v/HTS_Helper.

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada [NSERC-249644-2011] and individual stipend to Plamen Dragiev given by the Nature and Technologies Research Funds of Quebec [FQRNT].

REFERENCES

Brideau,C. *et al*. (2003) Improved statistical methods for hit selection in HTS. *J. Biomol. Screen.*, **8**, 634–647.

Carralot,J.P. *et al*. (2012) A novel specific edge effect correction method for RNA interference screenings. *Bioinformatics*, **28**, 261–268.

Dragiev,P. *et al*. (2011) Systematic error detection in experimental high-throughput screening. *BMC Bioinformatics*, **12,** 25.

Elowe,N.H. *et al*. (2005) Experimental screening of dihydrofolate reductase yields a 'Test Set' of 50,000 small molecules for a computational data-mining and docking competition. *J. Biomol. Screen.*, **10**, 653–657.

Helm,J.S. *et al*. (2003) Identification of active-site inhibitors of MurG using a generalizable, high-throughput glycosyltrans-ferase screen. *J. Am. Chem. Soc.*, **125**, 11168–11169.

Heyse,S. (2002) Comprehensive analysis of high-throughput screening data. *Proceedings of SPIE*; 2002; Bellingham, WA. 2002, 4626 pp. 535–547.

Kevorkov,D. and Makarenkov,V. (2005) Statistical analysis of systematic errors in HTS. *J. Biomol. Screen.*, **10**, 557–567.

Makarenkov,V. *et al*. (2006) HTS-Corrector: new application for statistical analysis and correction of experimental data. *Bioinformatics*, **22**, 1408–1409.

Makarenkov,V. *et al*. (2007) Statistical analysis of systematic errors in HTS. *Bioinformatics*, **23**, 1648–1657.

Malo,N. *et al*. (2006) Statistical practice in high-throughput screening data analysis. *Nat. Biotechnol.*, **24**, 167–175.

Malo,N. *et al*. (2010) Experimental design and statistical methods for improved hit detection in high-throughput screening. *J. Biomol. Screen.,* **15**, 990–1000.

Shun,T.Y. *et al*. (2011) Identifying actives from HTS data sets: practical approaches for the selection of an appropriate HTS data processing method and quality control review. *J. Biomol. Screen.*, **16**, 1–14.

Tukey,J.W. (1977) *Exploratory Data Analysis.* Addison Wesley, Cambridge, MA.

Zhang,X.D. (2008) Novel analytic criteria and effective plate designs for quality control in genome-scale RNAi screens. *J. Biomol. Screen.*, **13**, 363–377.

Zhang,J. *et al*. (1999) A simple statistical parameter for use in evaluation and validation of high-throughput screening assays. *J. Biomol. Screen.*, **4**, 67–73.