OXFORD

Briefings in Bioinformatics, 16(6), 2015, 974–986

doi: 10.1093/bib/bbv004
Advance Access Publication Date: 7 March 2015
Paper

Detecting and overcoming systematic bias in high-throughput screening technologies: a comprehensive review of practical issues and methodological solutions

Iurie Caraus, Abdulaziz A. Alsuwailem, Robert Nadon and Vladimir Makarenkov

Corresponding author. Vladimir Makarenkov, Département d'informatique, Université du Québec à Montréal, C.P. 8888 succ. Centre-Ville, Montreal, QC H3C 3P8, Canada. Tel.: 1-514-987-3000, ext. 3870; E-mail: makarenkov.vladimir@uqam.ca

Abstract

Significant efforts have been made recently to improve data throughput and data quality in screening technologies related to drug design. The modern pharmaceutical industry relies heavily on high-throughput screening (HTS) and high-content screening (HCS) technologies, which include small molecule, complementary DNA (cDNA) and RNA interference (RNAi) types of screening. Data generated by these screening technologies are subject to several environmental and procedural systematic biases, which introduce errors into the hit identification process. We first review systematic biases typical of HTS and HCS screens. We highlight that study design issues and the way in which data are generated are crucial for providing unbiased screening results. Considering various data sets, including the publicly available ChemBank data, we assess the rates of systematic bias in experimental HTS by using plate-specific and assay-specific error detection tests. We describe main data normalization and correction techniques and introduce a general data preprocessing protocol. This protocol can be recommended for academic and industrial researchers involved in the analysis of current or next-generation HTS data.

Key words: data correction methods; data normalization methods; high-content screening (HCS); high-throughput screening (HTS); systematic error

Introduction

There has been a growing interest in the development of high-throughput screening (HTS) technologies over the past few decades [1], largely because screening methods promoted by the pharmaceutical industry have played a key role in drug discovery. The increasing computing power and miniaturization of screening equipment now allow for carrying out HTS analyses even in small academic laboratories. The most popular screening technologies used in drug design are high-content screening (HCS) [2] and HTS [3]. Their different subcategories

Iurie Caraus is a PhD student at the Department of Computer Science of Université du Québec à Montréal (Canada). He is involved in the analysis of bioinformatics data related to RNA interference screening.
Abdulaziz A. Alsuwailem is a Master's student in the Department of Human Genetics at McGill University. He is involved in the analysis of high-content screening data.
Robert Nadon is an Associate Professor at the Department of Human Genetics of McGill University and Genome Quebec Innovation Centre (Montreal, Canada). His scientific interests include microarray expression, high-throughput screening of small molecule and RNAi data, image-based high-content screening and genome-wide mRNA translation.
Vladimir Makarenkov is a Full Professor and Director of a graduate Bioinformatics program at the Department of Computer Science of Université du Quebec à Montréal (Canada). His research interests are in the fields of bioinformatics, software engineering and data mining. They include design and development of bioinformatics software and databases, reconstruction of phylogenetic trees and networks and development of new statistical methods and related software for the analysis of high-throughput screening data.
Submitted: 11 November 2014; **Received (in revised form):** 26 December 2014

include small molecule [4], complementary DNA (cDNA) [5] and RNA interference (RNAi) [6] types of screening. In a typical HCS or HTS campaign, hundreds of terabytes of experimental data concerning molecule activity, specificity and physiological and toxicological properties can be generated. These data should be processed using appropriate data mining and statistical methods and protocols to identify promising drug candidates (i.e. hits). One of the key challenges that needs to be answered during the analysis of HCS and HTS data is the identification and successful elimination of bias (i.e. systematic error) in the measurements. In this review, we present the existing types of bias common to all HTS technologies and discuss their negative impact on the hit selection process. We underline the necessity of randomization of screened samples and indicate the advantages of using replicate measurements. We present the methods intended to detect systematic error and those designed to correct the data affected by it. We argue that the latter methods should be applied only when the presence of a specific type of systematic error in the data has been confirmed by a suitable statistical test [7]. Furthermore, we provide suggestions concerning which data normalization and correction techniques should be applied in various practical situations. Finally, we present a broad-spectrum data preprocessing protocol that can be used for the correction and analysis of screening data before assay quality estimation and hit selection steps. This protocol can also be used for detecting and removing bias in future HTS technologies involving sequential screening of multiple plates. To illustrate the results of our analyses, we examine publicly available HTS and HCS data generated at the Harvard's Medical School (Figure 1), McGill University HTS laboratory (Figure 2), McMaster University laboratory: Data screened for McMaster Data Mining and Docking Competition (Figure 3) as well as those provided by the largest public HTS/HCS database, ChemBank, maintained by Harvard University's Broad Institute (Figure 4).

Screening technologies and related biases

HTS and HCS technologies and their subcategories

In this review, we focus on the two most widely used screening technologies: HTS and HCS. In a typical HTS/HCS primary assay, the selected library of chemical compounds is screened against a specific biological target to measure the intensity of the related inhibition or activation signal [8]. The size of the compound library can vary from hundreds to millions of items. Compounds are allocated across disposable microtiter plates of different sizes, typically including 96, 384, or 1536 wells. Well locations within a plate follow a rectangular matrix pattern. Each compound is usually placed in a single well. A suitable biological target culture (e.g. cells or a bacterial enzyme) is then added to each well of the plate. It is common to conduct unreplicated HTS experiments, although, as we show next, it is much more appropriate to obtain at minimum duplicate measurements. Processing the assay plates by HTS robotic equipment consists of a number of experimental wet-lab steps, including incubation, rising and reagent additions to the biological culture of interest. Once the incubation period is over, the plates are scanned to obtain measures of biological activity characterizing the selected compounds. It is worth noting that the obtained raw activity levels depend not only on putative biological activity, but also on systematic and random errors affecting the given screen. Data analysis steps, including statistical procedures for data normalization and data correction, should then be carried out to identify hits.

The increasing capacity of computer storage devices and the improvements in automation have allowed the use of HTS technologies to achieve resolution at the cellular level [9]. This related technology is called HCS. HCS is a screening method with multiple readouts, which is based on microscopic imaging from a cell-based assay [10]. HCS obtains detailed information of cell structure by extracting multicolor fluorescence signals. HCS has three advantages relative to other screening techniques: (a) cell-based analysis achieves high physiological correspondence, especially regarding drug screening; (b) single-cell analysis captures the heterogeneity of cell populations as well as the related individual response to treatments; (c) HCS generally has low false-positive and false-negative rates [11]. Nowadays, HCS technologies are commonly used in all areas of contemporary drug discovery, including primary compound screening, post-primary screening capable of supporting structure–activity relationships, early evaluation of ADME properties and complex multivariate drug profiling [12]. The Mitocheck [13] and ChemBank [14] databases are the main online resources containing publicly available HCS data.

Different subcategories of HTS and HCS technologies exist, depending on the target of interest. They involve altering protein function using small molecules, increasing gene function using cDNA libraries and manipulating gene function using RNAi.

Small molecules

A 'small molecule', which can be either natural or artificial, is defined in pharmacology as a molecule associated with a particular biopolymer—for example, a nucleic acid or a protein [15]. There is currently a significant interest in extending efforts to discover small molecules targeting proteins encoded in the genomes of humans and pathogenic organisms [16]. Furthermore, small-molecule screening technologies have applications in other areas of drug discovery, such as target validation, assay development, secondary screening, pharmacological property assessment and lead optimization. The combination of principles of molecular pharmacology with modern high-throughput [4] and high-content [17] technologies is critical for the success of these discoveries.

cDNA library

High-quality full-length cDNA libraries are essential for identification and validation of novel drug targets in functional genomic applications [18]. The discovery of reverse transcriptase permitted the transformation of unstable mRNA molecules into stable cDNA molecules. A comprehensive review of cDNA HCS can be found in [19], and that of cDNA HTS in [5, 20].

RNA interference

In the past decade, RNAi has made great progress, evolving from a biological phenomenon into an effective method of drug discovery [21]. The two main advantages of RNAi screens compared to classical genetic screens are the following: (a) sequences of all identified genes are instantaneously identified and (b) lethal mutations are simple to determine because mutant recovery is not required [22]. The four types of RNAi reagents currently used in cell-based HTS the following are: dsRNAs, siRNAs, shRNAs and endoribonuclease-prepared siRNAs [23]. An important issue in genome-wide RNAi investigation is to combine both experimental and computational approaches to obtain high-quality RNAi HTS assays and to overcome off-target effects [24–26]. A recent review by Knapp and Kaderali focuses on the analysis of

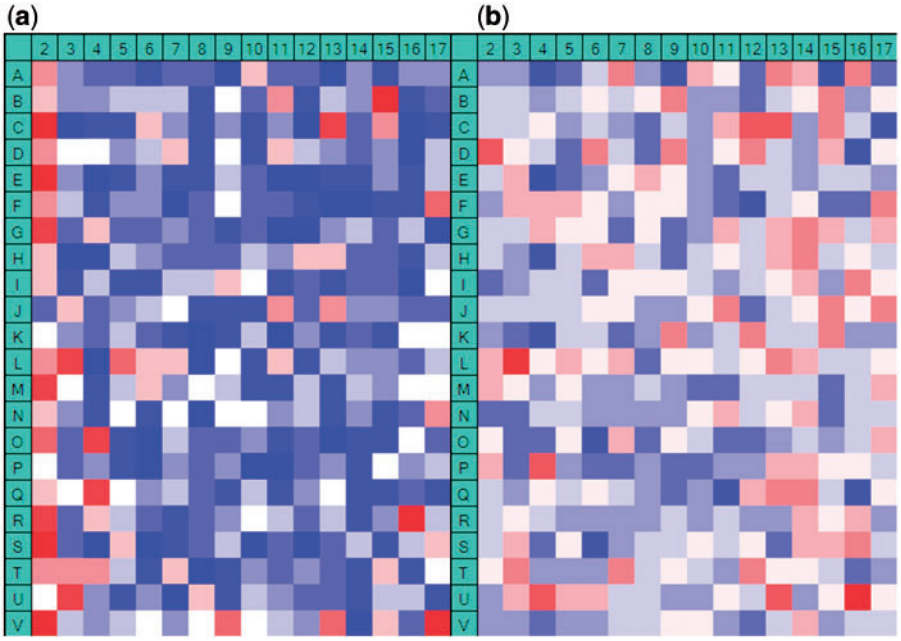

Figure 1. Systematic error in experimental HTS data (Harvard's 164-plate assay [41]). Hit distribution surfaces for the μ-2σ hit selection threshold are shown for: (**A**) Raw data; (**B**) B-score corrected data. Well, row and column positional effects are illustrated. The data are available at http://www.info2.uqam.ca/~makarenkov_v/HTS/home.php/downloads/Harvard_164.zip. A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.

RNAi HCS data and presents an approach for statistical processing of high-content microscopic screens [27].

Systematic error in screening technologies

As with all biotechnologies, screening data are prone to both 'random' and 'systematic errors'. Random error, which varies among measured HTS compounds, lowers screening precision and likewise affects false-positive and false-negative rates. Its adverse effects can be greatly minimized by obtaining at least duplicate measurements [28]. Systematic error (i.e. systematic or spatial bias) can be defined as the systematic under- or overestimation of measurements taken at the same plate or assay location [29]. Systematic errors can be the cause of non-specific phenotypes in specific well, row or column locations and thus lead to higher false-positive and false-negative rates [7, 30]. Its adverse effects can be minimized by the application of data correction methods and study design procedures such as randomization and blocking [8, 31].

Systematic error can be due to various technological and environmental factors, such as robotic failure, reader effect, pipette malfunctioning or other liquid handling anomalies, unintended differences in compound concentration related to agent evaporation, variation in the incubation time or temperature difference, as well as lighting or air flow abnormalities present over the course of the screening campaign [32, 33]. Thus, bias causing systematic under– or overestimation of biological activity measurements can cause some inactive compounds to be incorrectly identified as hits (i.e. 'false positives') and some active compounds to remain undetected (i.e. 'false negatives'). Systematic error can be well, row or column dependent. It can affect compounds placed either to the same well, row or column location over all plates of the assay (i.e. 'assay-specific error') or those located in a particular row or column of a single plate (i.e. 'plate-specific error') [34].

Some specific positional effects appearing in HTS/HCS screens as a consequence of bias are summarized below. One often overlooked hurdle of HTS technologies is the 'batch effect' [35]. A batch effect, i.e. bias present in some continuous subsets of the data and absent in others, occurs when some continuous groups of plates are affected by laboratory conditions that vary during the experiment. Although batch effects are hard to identify in low-dimensional assays, HTS technologies provide enough data to identify and remove them [35]. The 'edge effect', also called 'border effect', is another type of systematic error that consists in systematic under- or overestimation of the measurements located on the plate's edges. Carralot *et al.* [36] indicated that although most repetitive errors in RNAi HTS can be generally controlled, some biases, such as edge effects, cannot be easily corrected due to well-to-well discrepancies inherent in the spatial structure of the plate. The cause of this effect is often unclear but medium evaporation or uneven treatment of the entire plate surface might be contributing factors [37]. Similarly to the plate-specific edge effect, a more general assay-specific 'row', 'column' or 'well location effects' can occur in both HTS and HCS screens when the data located in a particular row, column or well location are systematically over- or underestimated across all the plates of the assay. On the other hand, a systematic 'intra-image bias', consisting of the microscope-related errors, arises while capturing images in HCS. One of the issues here is a nonuniformity of background light intensity distribution, which is a slowly varying and systematic change of the spatial distribution of light in images. Such an effect can add or subtract intensities at any pixel location, thus affecting cell segmentation and florescence measurements, which, in turn, affect data quantification and statistical analysis [38].

Cell population context can also create systematic bias in high-content cellular screens and thus significantly influence results of HCS campaigns [39]. A method allowing for normalizing and scoring statistically microscopy-based RNAi screens has been recently proposed [40]. This method exploits individual cell information of hundreds of cells per knockdown. The

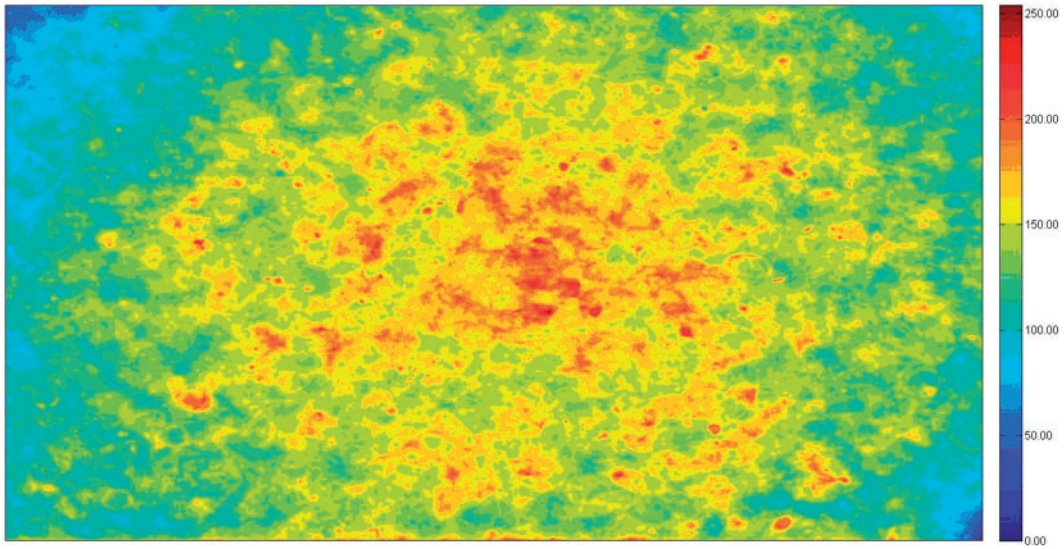

Figure 2. Non-smoothed foreground (non-uniformity bias) for images of a single (96-well × 4-field) HCS plate of microtubule polymerization status generated in the HCS laboratory of McGill University is shown. The data are available at http://nadon-mugqic.mcgill.ca. A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.

application of the proposed method and software [40] led to the identification of new host dependency factors of the hepatitis C and dengue viruses as well as to higher reproducibility of results of the screening experiments.

Figure 1A illustrates the presence of edge effects (e.g. the measurements in column 2 are systematically overestimated) in the Harvard 164-plate assay [29, 41]. This assay consists of a screen of compounds inhibiting the glycosyltransferase MurG function of *Escherichia coli*. Here, the binding effect of MurG to a fluorescent (fluorescein-labeled) analog of UDP-GlcNAc was estimated. In this example, the threshold of μ–2σ was applied to identify hits. The HTS Corrector software [42] was used to calculate raw (Figure 1A) and B-score corrected (Figure 1B) hit distribution surfaces (i.e. a hit distribution surface gives the number of hits per well location found over all plates of the assay). The edge effect observed in column 2, and partially in row V, in the raw data was successively eliminated by the B-score procedure [3].

Similarly, image nonuniformity bias in HCS can be approximated and corrected by combining multiple images to generate a single image with an expected random spatial distribution of intensity values [38]. Such an approximation represents the overall effect of bias on the imaging field estimated using an image-averaging technique [43]. This positional bias can be usually detected by comparing the center of the image with its edges. In most cases, there is at least a 2-fold increase in brightness between center and edges. Figure 2 illustrates nonuniformity bias present in a (96-well × 4-field) HCS plate of microtubule polymerization status screened in the HCS laboratory of McGill University.

Methods and results

Data randomization and use of controls

The primary aim of statistical practice consists in estimating experimental error, and in the case of systematic error, in reducing the negative effect of this error [44]. Experimental design and statistical methods should be applied to accomplish these objectives, although often underused in screening practice [31]. A fundamental approach for error reduction in experimental design must include control and randomization techniques [45]; R.A. Fisher introduced the concept of randomization in which experimental units are assigned to groups or treatment in a manner that the probability of assignment to any particular group or treatment is equal and unbiased [46]. The main advantage of randomization in screening technologies is that randomized experimental units can distribute the error in a way that does not introduce discrepancies to the experiment [31, 47, 48]. Thus, order of plate processing and compound placements both within each plate and across replicate plates of HTS/HCS assays should be randomized to reduce the impact of systematic bias on the outcome of screening experiments.

Controls contain compounds with well-known biological activity. Positive controls provide maximum possible activity measurements and negative controls provide minimum possible activity measurements. Controls are used in control-based normalization methods to render the screening data comparable across different plates and to establish assay background levels. Ideally, controls should be located randomly within plates, but in practice, only the first and the last columns of the plate are typically available for controls. The related systematic edge effect can be reduced by alternating the positive and negative controls in the available wells, so that they appear equally on each of the plate's rows and columns [8]. If the edge effect affects the control wells, it will also affect all of the plate's measurements because they are normalized relative to the control activities. Randomization of the position of compounds in the replicated experiments is also important, but unfortunately, is often limited due to practical considerations when automatic spotting approaches or some of the available statistical pipelines (e.g. cellHTS in BioConductor [49]) not supporting control randomization are used. RNAi controls generally exhibit more inter-well variability than small molecule controls because of variations in transfection efficiencies [50]. Cell-based biological controls are especially problematic because cell clumping or evaporation within different plate areas can lead to different growth conditions and thus to position-related bias [8, 36].

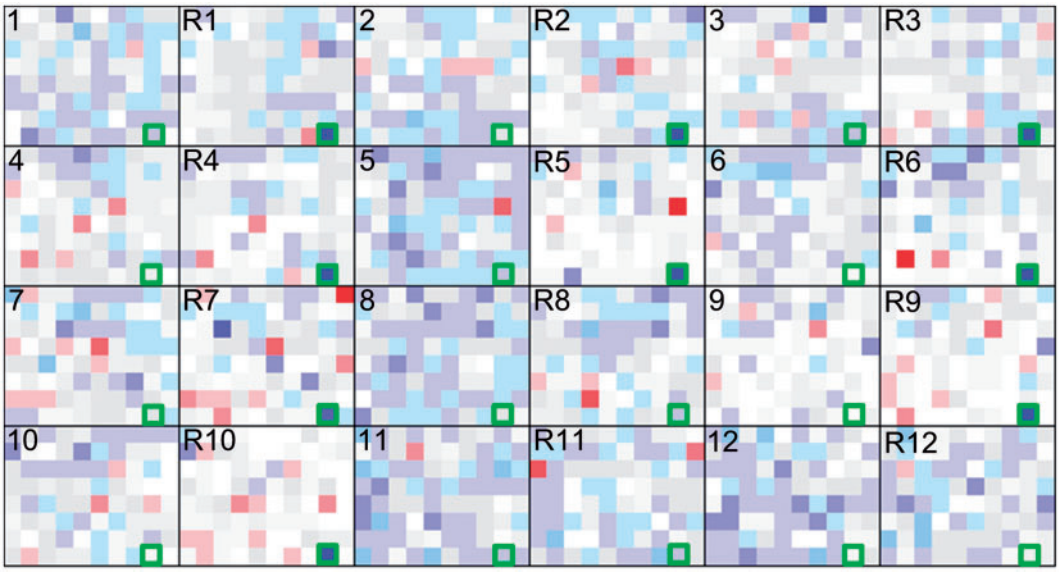

Figure 3. Batch positional effect appearing in the 'McMaster Test' assay screened during the McMaster Data Mining and Docking Competition [53]. The first 24 plates of the assay are shown (12 original and 12 replicates; the plate number is indicated on each plate; the replicates are indicated by the letter R). Each original plate is followed by its replicate. Hits are shown in blue. Green boxes emphasize well (8, 9) on each plate (i.e. well H10, according to the McMaster annotation) where the batch effect occurs. The data are available at http://www.info2.uqam.ca/~makarenkov_v/HTS/home.php/downloads/McMaster_1250.zip. A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.

Advantages of replicated measurements

Replicates offer the twin advantage of obtaining a greater precision of activity measurements and of estimating the measurements variability [8]. The use of replicates allows one to reduce the uncertainty associated to a single measurement (i.e. standard error of the mean), as indicated in Formula 1:

$$100 \times (1 - {}^{1}/_{\sqrt{n}})\%, \quad (1)$$

where n is the number of replicates. Thus, carrying out two replicated screens reduces imprecision by 29%, carrying out three replicated screens reduces impression by further 13% and carrying out four replicated screens reduces the impression by additional 8% (i.e. eliminating in total 50% of imprecision associated with a single measurement). Therefore, the replicates make minimally and moderately active compounds simpler to detect. Two types of replicates exist: technical and biological ones [51]. Technical replicates, which address the variability of the process, are repeated measurements of the same sample that represent independent measures of the random noise associated with equipment or protocols. Biological replicates, which mainly address the variability of the population but also reflect the variability of the process, are separate biological samples that were treated using the same protocol. When the sample population is unknown or has a higher variability, more biological replicates are needed. Increasing the number of technical replicates is important for a more variable technical protocol or when new screening equipment is used. Generally, biological variability is considerably greater than technical variability, so it is to our advantage to commit resources to sampling biologically relevant variables [51]. When planning for replication, researchers have to determine the proportion of variability induced by each experimental step to design statistically independent replicates and distribute the capacity for replication of the experiment across steps. Recognizing that obtaining even the minimal requirement of two replicates can be prohibitively expensive for some screens, Murie *et al.* [52] introduced the single assay-wide variance experimental design, which can generate statistical tests of biological activity based on replication of only a small subset of plates.

Figure 3 illustrates the presence of the batch effect in the 'McMaster Test dataset' including the original and replicated sets of plates [53]. The McMaster Test assay consisted of a sequence of 625 plates, each of which was screened twice (8×12-well plates were used; the first and the last columns of each plate contained controls—these columns are not displayed here; the remaining 80 wells contained different compounds meant to inhibit the *E. coli*'s dihydrofolate reductase). The well (8, 9) (i.e. well (H,10)—according to McMaster annotation; it is highlighted by a green box in Figure 3) displays a hit only in the replicated plates R1, R2, R3, R4, R5, R6, R7, R9 and R10, but not in the original plates 1, 2, 3, 4, 5, 6, 7, 9 and 10. This batch effect is absent in the replicated plate R8 and disappears starting from the replicated plate R11. Three of the hit compounds: MAC-0120363 (plates 1 and R1), MAC-0121481 (plates 3 and R3) and MAC-0121668 (plates 5 and R5) were initially recognized as 'Average Hits' by the McMaster competition organizers (the list of average hits contained 96 compounds whose average measurements, computed over the original and replicated screens, were lower than or equal to 75% of the reference control average), but all of them were then rejected as false positives when the dose-response relationship analysis of the selected compounds was carried out [53]. It is worth noting that only 96 of 50 000 screened compounds in this assay were recognized as 'Average Hits'.

Identification of hits

The identification of hits is the primary goal of any HTS/HCS campaign. Some screeners select as screening positives a fixed number, or a fixed percentage, of top-scoring compounds. Compounds whose activity exceeds a fixed percent-of-control threshold may also be considered as hits [8, 54]. A wide range of more sophisticated hit identification techniques is available nowadays. Birmingham and colleagues [50] reviewed the existing hit selection methods, which can be classified as small-molecule derived methods and RNAi-specific techniques. Small-molecule

derived methods include the following: selection of samples whose screening activity exceeds a fixed threshold, which usually equals mean − 3 SD for inhibition assays and mean + 3 SD for activation assays [8]; a robust to outliers improvement of the previous approach, using median instead of mean and median absolute deviation (MAD) instead of standard deviations [55]; for assays using replicated measurements, the difference in means between replicates for each condition can be assessed with multiple *t*-tests [50]; finally, the random variance model, which uses a weighted average of the compound-specific variance and an estimate of the typical variance of all of the compounds, has been shown to be appropriate for small molecule HTS data with performance superior to that of standard *t*-tests [56, 8]. RNAi-specific techniques include the following: quartile-based hit identification procedure, which establishes upper and lower hit selection thresholds based on number of interquartile ranges (i.e. above or below the first and third quartiles of the data) [57]; an accurate strictly standardized mean difference method, which computes the ratio between the difference of the means and the standard deviation of the difference between positive and negative controls [58]; and the redundant siRNA activity analysis method, designed for screeners interested in information about multiple RNAi reagents tested for each gene, which assigns *P*-values to all reagents of a single gene [59].

Data normalization techniques that correct for overall plate bias only

Data normalization in HTS and HCS consists in data transformation allowing for data comparability across different plates of the same assay [50]. The following simple types of data normalization, which do not correct for spatial systematic biases, are commonly used in screening technologies.

'Control Normalization' is a control-based normalization method using the measurements of both positive and negative controls (Formula 2):

$$\hat{x}_{ij} = \frac{x_{ij} - \mu_{neg}}{\mu_{pos} - \mu_{neg}} \tag{2}$$

where x_{ij} is the raw measurement of the compound located in well (i, j), $\hat{x}_{ij}$ is the normalized value of the raw measurement x_{ij}, μ_{pos} is the mean of positive controls of the plate and μ_{neg} is the mean of negative controls of the plate.

'Median percent inhibition' normalization is carried out as follows (Formula 3):

$$\hat{x}_{ij} = 100 \times \left(1 - \frac{x_{ij}}{med}\right) \tag{3}$$

where *med* is the median of all measurements of the plate.

'Z-score' normalization is defined as follows (Formula 4):

$$\hat{x}_{ij} = \frac{x_{ij} - \mu}{\sigma} \tag{4}$$

where μ and σ are, respectively, the mean and the standard deviation of all measurements of the plate.

'Robust Z-score' normalization can account for different scale and variability effects across HTS plates. It is less likely to produce biased scores because of outlying values of highly active compounds. Robust Z-score normalization is similar to Z-score except that the median is used instead of the mean and the median absolute deviation (*MAD*) is considered instead of the standard deviation to obtain the outlier resistant dispersion estimates (Formula 5):

$$\hat{x}_{ij} = \frac{x_{ij} - med}{MAD} \tag{5}$$

where *MAD* is the median absolute deviation of measurements of the plate.

Systematic error detection tests

Several error correction methods and software have been recently developed to minimize the impact of systematic bias [7]. These methods and software should, however, be used with caution. Makarenkov *et al.* [33] demonstrated that systematic error correction methods can introduce systematic bias when applied on error-free HTS data. The introduced bias may be less important as in the case of the well correction procedure [33] or important as in the case of the B-score method [3]. Thus, the presence or absence of systematic bias in raw HTS data must be first confirmed by the appropriate statistical tests [7, 60–62]. Systematic error detection tests that work well with screening data are summarized below.

Welch's t-test

This test is based on the classical two-sample Welch's *t*-test for the case of samples with various sizes and unequal variances [60]. Two variants of this test can be considered in the framework of HTS/HCS analysis. The first variant concerns its application to each row and each column of every plate of the assay. The second variant concerns its application to the assay's hit distribution surface. The measurements of the given plate (or of the hit distribution surface) are subdivided into two samples: the first sample contains the measurements of the tested row or column, while the second sample includes the remaining plate's measurements. The null hypothesis, H_0, here is that the considered row or column does not contain systematic error. For the two considered samples, S_1 with N_1 elements and S_2 with N_2 elements, the two sample variances, s_1^2 and s_2^2, are first calculated. Welch's *t*-test statistic can then be computed using Formula 6:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}, \tag{6}$$

where μ_1 is the mean of sample S_1 and μ_2 is the mean of sample S_2. The *t*-test value is then compared with the critical value corresponding to the chosen statistical significance level α to decide whether H_0 should be rejected. Welch's *t*-test is usually applied when the data are normally distributed but the sample variances may differ. However, for moderately large samples and a one-tailed test, this statistic is relatively robust to violations of the normality assumption.

χ^2 goodness-of-fit test

This test can be used to establish the presence or absence of systematic error in a hit distribution surface [7]. The null hypothesis H_0 here is the same as in Welch's *t*-test. The rejection region of H_0 is $P(\chi^2 > C_\alpha) > \alpha$, where C_α is the χ^2 distribution critical value, corresponding to the chosen parameter α and the number of degrees of freedom. For a hit distribution surface with N_R rows and N_C columns, one can test the presence of systematic error in a given row *r* by calculating the χ_r^2 statistic (Formula 7):

$$\chi_r^2 = \sum_{j=1}^{N_C} \frac{(x_{rj} - E)^2}{E} \quad (7)$$

where x_{rj} is the j^{th} value in row r, E is the hits count of the whole hit distribution surface divided by the number of wells (N_R x N_C). The number of degrees of freedom here is $N_R - 1$.

In the same way, the columns of the hit distribution surface affected by systematic error can be tested by computing the test statistic χ_c^2 (Formula 8):

$$\chi_c^2 = \sum_{i=1}^{N_R} \frac{(x_{ic} - E)^2}{E} \quad (8)$$

The number of degrees of freedom here is $N_C - 1$.

Systematic error affecting a particular well location (i, j) and appearing along all plates of the assay can be also identified by computing the χ^2 statistic [7] (Formula 9):

$$\chi^2 = \sum_{i=1}^{N_R} \sum_{j=1}^{N_C} \frac{(x_{ij} - E)^2}{E}. \quad (9)$$

The number of degrees of freedom here is $N_R \times N_C - 1$. The following main assumptions should be met for this test: (i) the observations are independent of each other and (ii) the expected hits count in each well location of the hit distribution surface should be at least 5.

Kolmogorov–Smirnov test preceded by Discrete Fourier Transform

This method consists of Discrete Fourier Transform (DFT) [63] signal analysis method followed by the Kolmogorov–Smirnov (KS) test [64]. It is included in some commercial software intended to detect systematic error in screening data (e.g. in the 'Array Validator' program described in [65]). The KS test is a non-parametric test having the advantage of making no assumption about the distribution of data.

As recently has been shown, Welch's t-test usually outperforms the χ^2 goodness-of-fit test and the KS test preceded by DFT in the context of HTS analysis [7]. A comprehensive simulation study involving artificially generated HTS data was carried out to compare the three aforementioned tests in a variety of practical situations. The success rate of the t-test was usually above 90%, regardless the plate size, the type and the magnitude of systematic error, whereas the values of Cohen's kappa coefficient for this test suggested its superior performance in the case of large plates and high level of systematic bias [7].

Mann–Whitney–Wilcoxon test

This test verifies whether two samples of measurements are identical. First, a suitable Type I error probability, α, is chosen for the test and the data in two samples of interest, X_1 and X_2, are ranked. The Mann–Whitney–Wilcoxon (MWW) test [61] is based on Formula 10:

$$z = \frac{W_1 - \frac{N_1 \times (N_1 + N_2 + 1)}{2} + C}{\sigma_W} \quad (10)$$

where N_1 and N_2 are the sizes of samples X_1 and X_2, and $W_1 = \sum_{k=1}^{N_1} Rank(X_{1k})$ is the sum of the ranks of the first sample measurements. The correction factor, C, equals 0.5, if the rest of the numerator of z is negative, or equals −0.5, otherwise. The standard deviation, σ_W, is determined using Formula 11:

$$\sigma_W = \sqrt{\frac{N_1 \times N_2 \times (N_1 + N_2 + 1)}{12}} \quad (11)$$

As this is a non-parametric test, it does not make assumptions about the underlying data distribution.

Rank products test

Consider the expression levels of n genes for k_1 independent replicates in sample X_1, and k_2 independent replicates in sample X_2. Let X_{ijm} be the expression level of the i^{th} gene in the j^{th} replicate of the m^{th} sample, where $1 \le i \le n$, $1 \le j \le k_m$, $1 \le m \le 2$. By ranking the expression levels X_{1jm}, X_{2jm}, …, X_{njm} within each replicate j, we form the vectors $R_{ijm} = rank(X_{ijm})$, where $1 \le R_{ijm} \le n$ and $1 \le m \le 2$. The suitable two-sample version of Breitling's Rank products statistic, RP, for the i^{th} gene can then be calculated by using Formula 12 [62, 66]:

$$RP_i = \left(\prod_{j=1}^{k_1} R_{ij1}\right)^{1/k_1} \Big/ \left(\prod_{j=1}^{k_2} R_{ij2}\right)^{1/k_2} \quad (12)$$

Genes associated with sufficiently large or small RP_i values are marked for further consideration. A few assumptions for this non-parametric test are the following [66]: (i) relevant expression changes affect only a minority of genes, (ii) measurements are independent between replicated plates (or screens); (iii) most changes are independent of each other and (iv) measurement variance is about equal for all genes. The MWW and Rank products tests have been successively applied in the RNAi screening [67, 68].

To estimate the magnitude of systematic bias in experimental HTS data, we carried out a series of tests using the data extracted from the largest public HTS/HCS database, ChemBank [14]. Figure 4A reports the average row and column systematic error rates in raw HTS measurements obtained from 41 HTS assays (735 plates in total) aimed at the inhibition of the *E. coli* bacterium. In this analysis, we considered all HTS assays related to the *E. coli* inhibition, which were available in ChemBank as of April 2014. The presented results were obtained by using Welch's t-test (Equation 6) with different values of the parameter $\alpha = 0.01$, 0.025, 0.05, 0.075 and 0.1. The null hypothesis here was that the considered row or column did not contain any systematic bias. Figure 4B illustrates the average hit distribution surface error rates for raw data. The presence of systematic errors in an assay can be determined through the analysis of its hit distribution surface depicting the total hit counts per well location over all plates of the assay [33]. Thus, we estimated over all assay's plates the number of measurements with values lower than the μ-$c\sigma$ threshold, where the mean value μ and the standard deviation σ were computed separately for each plate; the constant c was gradually set to 2.5, 3.0 and 3.5 to account for the most popular hit selection thresholds. Here also, Welch's t-test was used to determine the presence or absence of systematic error. Similarly, Figure 4C presents the average row and column error rates for the background-subtracted measurements for the same 41 HTS assays (background-subtracted data were also extracted from ChemBank), and Figure 4D shows the average hit distribution surface error rates for the background-subtracted data. The Matlab 8.2 package [69] was used in our computations. The presented graphics suggest that the row and column systematic bias is common to experimental HTS assays (i.e. plate-specific error)—at least 30% of rows and columns in the raw data and 20% of rows and columns in the background-subtracted data were affected by systematic bias (Figure 4A and

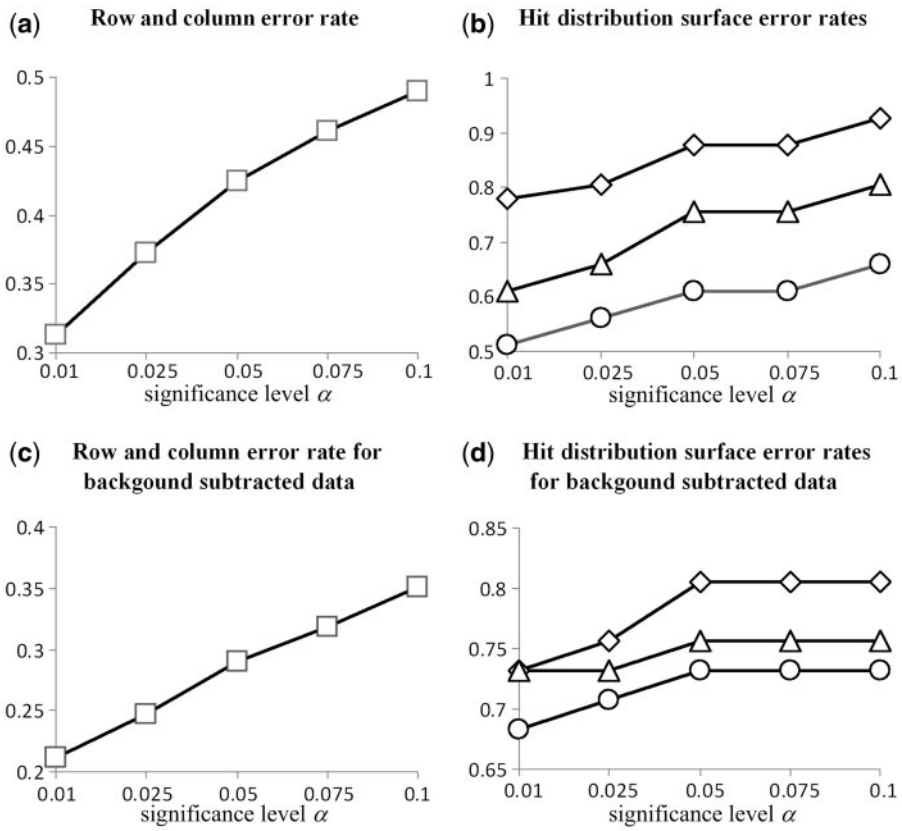

Figure 4. Proportion of rows and columns affected by systematic bias in 41 experimental HTS assays (735 plates in total; control wells were ignored) aiming at the inhibition of the *E. coli*. Experimental data were extracted from the Harvard University HTS databank (i.e. ChemBank [14]). Here we show: (**A**) Overall row and column error rate for raw data; (**B**) hit distribution surface error rates for raw data; (**C**) overall row and column error rate for background-subtracted data; (**D**) hit distribution error rate for background subtracted data. The following hit selection thresholds were used to identify hits and establish hit distribution surfaces of the assays: μ-2.5σ (◊), μ-3σ (Δ) and μ-3.5σ (○), where μ and σ are, respectively, the mean and SD of the plate's measurements.

C). Moreover, systematic error is even more visible when analyzing hit distribution surfaces (i.e. assay-specific error)—at least 50% of raw hit distribution surfaces and 65% of background-subtracted hit distribution surfaces were affected by systematic error (Figure 4B and D).

Data normalization techniques that correct for plate-specific and assay-specific spatial systematic biases

This section describes the statistical methods that are used for minimizing plate-specific and assay-specific (i.e. across-plate well-location bias) spatial systematic biases in screening technologies. Most of these methods allow the correction of overall plate bias as well.

R-scores

This plate-specific correction method [70] relies on Formula 13:

$$x_{ijp} = \mu_p + R_{ip} + C_{jp} + r_{ijp}, \tag{13}$$

where x_{ijp} is the compound measurement in row i and column j of plate p, μ_p is the mean of plate p, R_{ip} is the row bias affecting row i of plate p, C_{jp} is the column bias affecting column j of plate p and r_{ijp} is the residual in well (i, j) of plate p. These parameters can be estimated using, for example, the *rlm* function from the MASS package of the *R* language [71]. The R-scores are the model's residuals rescaled by dividing them by the standard deviation estimate of the regression function.

B-scores

This method corrects the raw plate measurements by iteratively eliminating possible row and column positional biases [3]. The statistical model for the raw measurement x_{ijp} is similar to Formula 13. The B-scores method relies on a two-way median polish (MP) procedure [72] carried out separately for each plate of the assay to obtain the estimates of x_{ijp}, μ_p, R_{ip} and C_{jp}. The residual r_{ijp} of the measurement in well (i, j) is then calculated as the difference between the raw measurement x_{ijp} and its fitted value $\tilde{x}_{ijp}$: $r_{ijp} = x_{ijp} - \tilde{x}_{ijp}$. Finally, the obtained residuals are divided by the MAD of plate p (Formula 14):

$$\text{B-score} = \frac{r_{ijp}}{MAD_p}, \text{where } MAD_p = med\{|r_{ijp} - med(r_{ijp})|\} \tag{14}$$

A variant of the B-scores method used in HCS [73] considers the mean true activity value, μ_{ijp}, in well (i, j) in Formula (13), instead of μ_p.

Well correction

This assay-specific correction method proceeds by data normalization along the well locations of the assay [33, 42]. At first, Z-score normalization (Formula 4) is performed within each plate of the assay. The following two steps are then carried out. First, a linear least-square approximation is performed for the measurements of each well location of the assay (this well-specific approximation is done across all plates of the assay). Second, Z-score normalization of the fitted measurements

obtained from regression is carried out independently for each well location of the assay (still across all plates of the assay).

Robust well correction

This is another assay-specific data correction procedure. Each plate is normalized using robust Z-scores (Formula 5) and then the entire set of plates is ordered by date of processing and a robust regression line is fit to the data. This fitting is carried out independently for each well location across all plates of the assay as in the well correction method. The obtained normalized residuals are considered as final corrected scores [31].

Diffusion model

This model is designed to eliminate the edge effect in the HTS RNAi screens [36]. The process-specific diffusion process is described by the following parabolic differential equation (Formula 15):

$$\frac{\partial \tilde{b}(i,j,t)}{\partial t} = c \times \Delta \times \tilde{b}(i,j,t) \tag{15}$$

where $\tilde{b}(i,j,t)$ is the evaluated spatiotemporal diffusion field in well (i, j) at time t (i.e. evaluated systematic bias), c is the diffusion coefficient and Δ is the Laplacian operator. The following boundary conditions are considered (Formula 16):

$$\begin{cases} \tilde{b}(i,j,t) = U_1, \forall (i,j) \in Z^2 \setminus \Gamma, \\ \tilde{b}(i,j,t=0) = U_0, \forall (i,j) \in \Gamma, \end{cases} \tag{16}$$

where U_0 and U_1 are the model's positive parameters; the model also assumes that:

- at the initial time $t=0$ of the dispensing, there is no edge effect on the given plate;
- the effect strength depends on a physical difference between the inside parameter Γ and outside parameter $Z^2 \setminus \Gamma$ of the given plate.

Loess correction method

The loess error correction method evaluates the plate's row and column effects by fitting a loess curve to each row and column of the given plate [74, 75]. The loess correction is defined by Formula 17:

$$\tilde{x}_{ij} = x_{ij} \times \left(\frac{\bar{r}_i}{r_{ij}}\right) \times \left(\frac{\bar{c}_j}{c_{ij}}\right) \tag{17}$$

where x_{ij} is the raw measurement in well (i, j), $\tilde{x}_{ij}$ is the adjusted measurement in this well, $\bar{r}_i$ is the mean of the fitted loess curve for row i, $\bar{c}_j$ is the mean of the fitted loess curve for column j, r_{ij} is the value of the fitted row loess curve for row i and column j and c_{ij} is the value of the fitted column loess curve for row i and column j.

Median filter

The median filter method [76] adjusts the intensity value of the given well (i, j) using the median of the intensity values of the nearby wells. First, a row median filter, whose filter window includes the wells located on the same row i, within k wells of well (i, j), is carried out. Second, a standard median filter procedure, its filter window includes the wells located within l wells of well (i, j), is applied. The constants k and l usually equal 3 for the 1536-well plates, and 1 and 2 for the 96-well plates. The method relies on Formula 18 to compute the adjusted measurements:

$$\tilde{x}_{ij} = x_{ij} \times \left(\frac{med_p}{med_{wij}}\right), \tag{18}$$

where med_p is the median intensity of plate p and med_{wij} is the median intensity of wells included in the filter window of well (i, j).

SPAtial and Well Normalization

This two-step method gradually applies a trimmed mean polish method on individual plates to minimize row and column systematic effects [77]. The considered statistical model relies on Formula 13. First, a well normalization step is carried out to determine spatial bias template, SBT_{ij}, which is the median of the scores at well location (i, j) computed over all plates of the assay. The spatial bias template scores are subtracted from the scores obtained by the median polish procedure: $\hat{r}_{ijp} = r_{ijp} - SBT_{ij}$. Finally, the resulting scores are rescaled by dividing them by the MAD of the plate. Thus, SPAtial and Well Normalization (SPAWN) corrects for both plate-specific and assay-specific biases.

Matrix error amendment and partial mean polish

These algebraic methods are designed to modify only those rows and columns of the given plate that are affected by systematic bias [34]. Matrix error amendment (MEA) and partial mean polish (PMP) methods rely on prior information concerning the presence and absence of systematic error in the rows and columns of the given plate. Such information can be obtained using a specific version of Welch's t-test or the χ^2 goodness-of-fit test (see previous section). One of the main advantages of the PMP method over MP and B-scores [3] is that PMP does not reduce the original data to residuals, keeping the corrected measurements on the same scale with the original ones.

Table 1 reports the discussed data normalization techniques recommended for the analysis of HTS and HCS data along with the underlying assumptions regarding their practical application.

Various plots that use robust statistical indices have been also suggested for detecting shifts and trends across time in large screening campaigns [3]. Systematic bias within plates can be detected with visualization methods such as two-dimensional heat maps and three-dimensional wire plots, although typical plate-specific bias patterns are more easily detected with autocorrelation plots that show the degree of correspondence between wells at various 'lags' (e.g. adjacent or separated by one well) [31]. Finally and somewhat counterintuitively, screens with few active compounds should show low correlations between replicate plates; for these screens, scatterplots that show high correspondence between replicate plates indicate across-plate well-specific bias rather than good biological reproducibility [31].

Discussion and conclusion

We reviewed current knowledge on systematic bias affecting raw data in HTS and HCS technologies. First, we discussed the causes of spatial systematic bias and its impact on the selection of correct hits in HTS and HCS experiments. The main steps of HTS and HCS screening protocols were presented along with

Table 1. Data normalization methods recommended for the analysis of HTS and HCS data, and classified according to the context in which they should be applied

Method	Overall plate bias correction	Spatial systematic bias correction	Removes plate-specific spatial bias	Removes assay-specific spatial bias	Data randomization assumption	Software available
Control normalization	Yes	No	No	No	Control placements are preferably randomized within plates	In many software; readily implemented in MS Excel and R
Median percent inhibition	Yes	No	No	No	No any	In many software; readily implemented in MS Excel and R
Z-score	Yes	No	No	No	No any	In many software; readily implemented in MS Excel and R
Robust Z-score	Yes	No	No	No	No any	SIGHTS [31]
R-score	Yes	Yes	Yes	Yes (RW function in SIGHTS)	Samples randomized within plates	rlm (MASS package) function in R, SIGHTS
B-score	Yes	Yes	Yes	No	Samples randomized within plates	BioConductor (cellHTS, RNAither) [49], HTS-Corrector [42], HCS-Analyzer [78], HTS-Helper [34]
Well correction	Yes	Yes	No	Yes	Samples randomized across different well locations of the assay	HTS-Corrector, HTS-Helper, SIGHTS
Robust well correction	Yes	Yes	No	Yes	Samples randomized across different well locations of the assay	SIGHTS
Diffusion model (HCS)	Yes	Yes	Yes	No	Samples randomized within plates	HCS-Analyzer
Loess correction	Yes	Yes	Yes	No	Samples randomized within plates	SIGHTS, BioConductor (RNAither)
Median filter	Yes	Yes	Yes	No	Samples randomized within plates	SIGHTS, HMF [76]
SPAWN	Yes	Yes	Yes	Yes (SPAWNW function in SIGHTS)	Samples randomized within plates and across different well locations	SIGHTS
MEA	No	Yes	Yes	No	Samples randomized within plates	HTS-Helper
PMP	No	Yes	Yes	No	Samples randomized within plates	HTS-Helper
IQEM (HCS)	No	Yes	Yes	No	Samples randomized within plates	IQEM code for MATLAB [38]

Note. The available software implementations are also indicated.

the subcategories of screening technologies, including small molecule, cDNA and RNAi screens. Positional bias effects characteristic of screening technologies, comprising batch effects, edge effects and well location effects, were discussed in detail. We highlighted that randomization of experimental units and use of replicates can significantly reduce the magnitude of systematic error. Data normalization techniques that correct for overall plate bias were presented, followed by the description of systematic error detection tests specific to screening technologies. Finally, we discussed error correction methods, indicating under which assumptions and for which kind of spatial bias each of them should be used. In particular, we underlined the distinction between the plate-specific and assay-specific systematic biases and pointed out that data correction methods should be applied only if the presence of systematic bias was confirmed by the appropriate statistical tests. Otherwise, an unwanted bias can be introduced into error-free data.

To summarize our presentation, we describe here a general data preprocessing and correction protocol (Figure 5), which could be used as a guide by academic and industrial researchers involved in the analysis of current or next-generation screening data. The first required step concerns general design of a screening campaign. The compound locations within each plate, as well as over all plates of the assays, should be randomized to reduce the impact of systematic bias on the outcome of screening experiments. Moreover, whenever the campaign funding allows, several replicates of the compound library should be screened. Replicated screens provide both a greater precision of activity measurements and the ability to assess measurement variability [8]. Once the assay measurements have been established, the appropriate data normalization procedure should be carried out to ensure the data comparability over different plates and screening conditions. Afterwards, systematic error detection tests should be carried out to confirm the presence or absence of systematic error in raw data (e.g. Welch's *t*-test or χ^2 goodness-of-fit test). In particular, these tests can be applied to identify the following: (i) positional effects of systematic error, including row, column and well location biases; (ii) error specificity, including plate, batch and assay-specific biases; (iii) type of systematic error, including

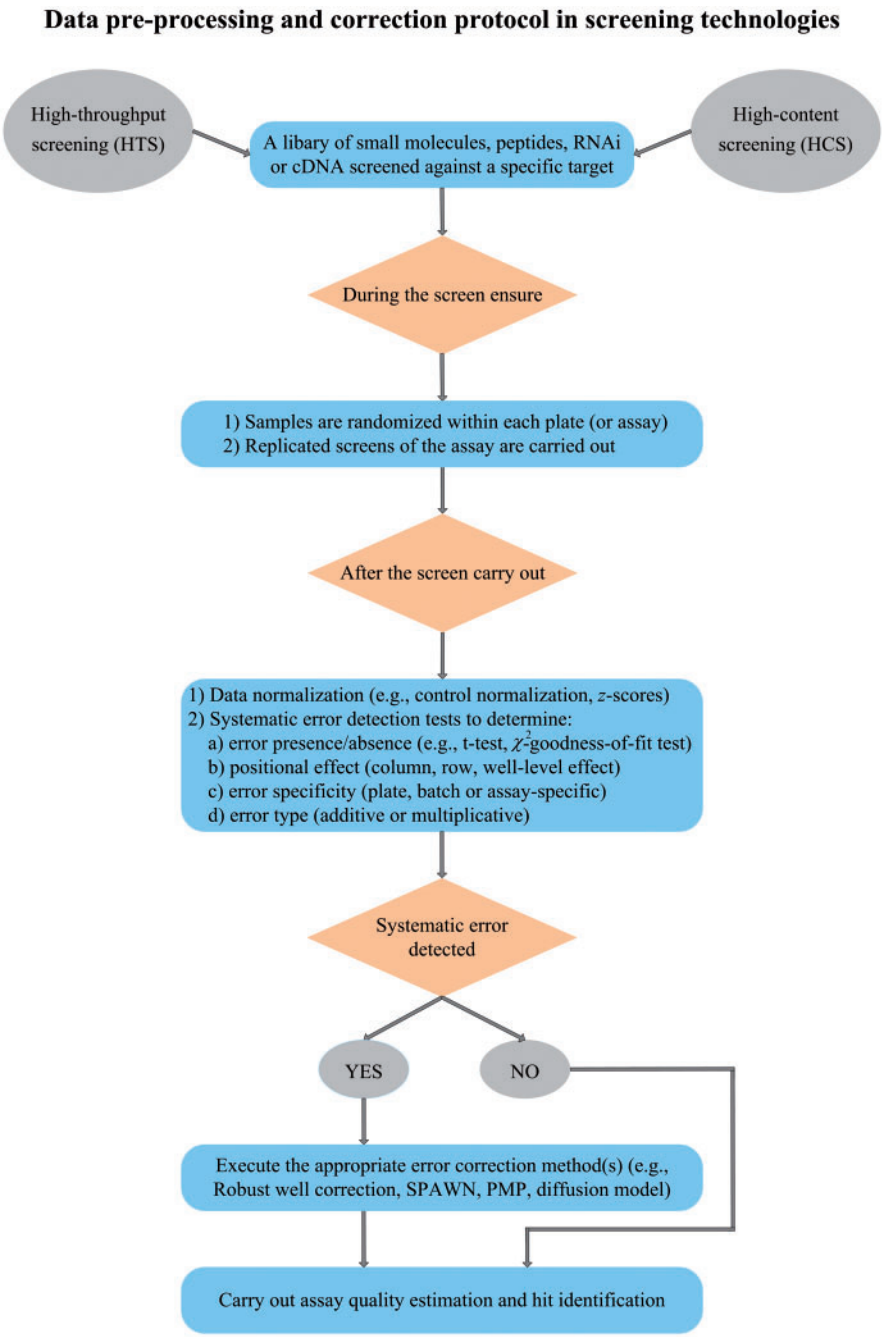

Figure 5. Recommended data preprocessing and correction protocol to be performed before the hit identification step in HTS and HCS. A colour version of this figure is available at BIB online: http://bib.oxfordjournals.org.

additive (e.g. Robust well correction, SPAWN or PMP methods can be applied to eliminate this type of bias) and multiplicative (e.g. diffusion model can be applied to eliminate this type of bias) biases. If systematic error was not detected in the data, then no correction method needs to be applied to them to avoid the risk of introduction of additional biases [7]. Otherwise, the appropriate error correction method, preferably including a success of control step, should be carried out. Once systematic bias is minimized, assay quality estimation and hit identification steps can be carried out. It is worth noting that the plate-specific correction methods (e.g. PMP) can sometimes be applied in combination with the assay-specific correction methods (e.g. Robust well correction). First, Welch's *t*-test can be carried out independently for each individual plate of the assay to detect the plate's rows and columns affected by systematic bias. The measurements affected by bias can be subsequently corrected by using the PMP method, which keeps the corrected data on the same scale with the original ones. Second, Welch's *t*-test can be performed over the hit distribution surface of the entire assay. If the test identifies the presence of systematic bias on the surface, the Robust well correction procedure can be carried out to remove the assay-specific bias. An alternative solution to this problem could be provided by the methods that correct for both plate-specific and assay-specific biases (e.g. SPAWN).

Funding

This work was supported by Le Fonds Québécois de la Recherche sur la Nature et les Technologies (grant no. 173878) and Natural Sciences and Engineering Research Council of Canada (grant no. 249644-2011).

Key Points

- We reviewed current knowledge on systematic bias affecting experimental HTS and HCS data.
- Study design issues and the way in which data are generated are crucial for providing unbiased screening results. Unfortunately, these key steps are often ignored by HTS practitioners.
- Data correction methods should be applied only if the presence of systematic error has been confirmed by the appropriate statistical tests.
- Discussed sources of systematic bias and presented statistical methods and software intended to correct experimental screening data provide a unifying framework when considering new screening technologies.
- We presented a general data preprocessing and correction protocol that can be used as a guide by academic and industrial researchers involved in the analysis of current or next-generation screening data.

References

1. Shelat AA, Guy RK. The interdependence between screening methods and screening libraries. *Curr Opin Chem Biol* 2007;**11**:244–51.
2. Giuliano KA, Haskins JR, Taylor DL. Advances in high content screening for drug discovery. *Assay Drug Dev Technol* 2003;**1**:565–77.
3. Brideau C, Gunter B, Pikounis W, *et al*. Improved statistical methods for hit selection in HTS. *J Biomol Screen* 2003;**8**:634–47.
4. Inglese J, Shamu CE, Guy RK. Reporting data from high-throughput screening of small-molecule libraries. *Nat Chem Biol* 2007;**3**:438–41.
5. Chiao E, Leonard J, Dickinson K, *et al*. High-throughput functional screen of mouse gastrula cDNA libraries reveals new components of endoderm and mesoderm specification. Genome Res, 2005;**15**:44–53.
6. Auer PL, Doerge RW. Statistical design and analysis of RNA sequencing data. *Genetics* 2010;**185**:405–16.
7. Dragiev P, Nadon R, Makarenkov V. Systematic error detection in experimental high-throughput screening. *BMC Bioinformatics* 2011;**12**:25.
8. Malo N, Hanley JA, Cerquozzi S, *et al*. Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006;**24**:167–75.
9. Noah JW. New developments and emerging trends in high-throughput screening methods for lead compound identification. *Int J High Throughput Screen* 2010;**1**:141–9.
10. Smellie A, Wilson CJ, Ng SC. Visualization and interpretation of high content screening data. *J Chem Inf Model* 2006;**46**:201–7.
11. Kozak K, Agrawal A, Machuy N, *et al*. Data mining techniques in high content screening: a survey. *J Comput Sci Syst Biol* 2009;**2**:219–39.
12. Zanella F, Lorens JB, Link W. High content screening: seeing is believing. *Trends Biotechnol* 2010;**28**:237–45.
13. Mitocheck HCS database, 2009. http://mitocheck.org.
14. ChemBank HTS/HCS database. http://chembank.broadinstitute.org.
15. Cram101. *E-Study Guide for: Polymers: Chemistry and Physics of Modern Materials: Chemistry, Materials sciences [Kindle edition]*, *CRAM101*, Ventura, CA, 2012.
16. Lazo JS, Brady LS, Dingledine R. Building a pharmacological lexicon: small molecule discovery in academia. *Mol Pharmacol* 2007;**72**:1–7.
17. Korn K, Krausz E. Cell-based high-content screening of small-molecule libraries. *Curr Opin Chem Biol* 2007;**11**:503–10.
18. Brown JC, Song C. High quality cDNA libraries for discovery and validation of novel drug targets. *Expert Opin Ther Targets* 2000;**4**:113–20.
19. Buchser W, Collins M, Garyantes T, *et al*. Assay development guidelines for image-based high content screening, high content analysis and high content imaging. In *Assay Guidance Manual*. [Internet], Bethesda (MD): *Eli Lilly & Company and the National Center for Advancing Translational Sciences*; 2012. Last Update: September 22, 2014. Available from: http://www.ncbi.nlm.nih.gov/books/NBK100913/.
20. Honma K, Ochiya T, Nagahara S, *et al*. Atelocollagen-based gene transfer in cells allows high-throughput screening of gene functions. *Biochem Biophys Res Commun* 2001;**289**:1075–81.
21. Sharma S, Rao A. RNAi screening: tips and techniques. *Nat Immunol* 2009;**10**:799–804.
22. Boutros M, Ahringer J. The art and design of genetic screens: RNA interference. *Nat Rev Genet* 2008;**9**:554–66.
23. Mohr S, Bakal C, Perrimon N. Genomic screening with RNAi: results and challenges. *Annu Rev Biochem* 2010;**79**:37–64.
24. Zhang XD, Espeseth AS, Johnson EN, *et al*. Integrating experimental and analytic approaches to improve data quality in genome-wide RNAi screens. *J Biomol Screen* 2008;**13**:378–89.
25. Buehler E, Khan AA, Marine S, *et al*. siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci Rep* 2012;**2**:428.
26. Amberkar S, Kiani NA, Bartenschlager R, *et al*. High-throughput RNA interference screens integrative analysis: towards a comprehensive understanding of the virus-host interplay. *World J Virol* 2013;**2**:18–31.
27. Knapp B, Kaderali L. *Statistical Analysis and Processing of Cellular Assays*. Technische Universität Dresden, Germany, iConcept Press, 2012.
28. Malo N, Hanley JA, Carlile G, *et al*. Experimental design and statistical methods for improved hit detection in high-throughput screening. *J Biomol Screen* 2010;**15**:990–1000.
29. Kevorkov D, Makarenkov V. Statistical analysis of systematic errors in high-throughput screening. *J Biomol Screen* 2005;**10**:557–67.
30. Ramadan N, Flockhart I, Booker M, *et al*. Design and implementation of high-throughput RNAi screens in cultured Drosophila cells. *Nat Protoc* 2007;**2**:2245–64.
31. Murie C, Barette C, Lafanechère L, *et al*. Improving detection of rare biological events in high-throughput screens. *J Biomol Screen* 2015;**20**(2):230–241.
32. Heyse S. Comprehensive analysis of high-throughput screening data. In: *International Symposium on Biomedical Optics. Proc. SPIE*, 2002;**4626**:535–47. San Jose, CA, International Society for Optics and Photonics.
33. Makarenkov V, Zentilli P, Kevorkov D, *et al*. An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics* 2007;**23**:1648–57.
34. Dragiev P, Nadon R, Makarenkov V. Two effective methods for correcting experimental high-throughput screening data. *Bioinformatics* 2012;**28**:1775–82.

35. Leek JT, Scharpf RB, Bravo HC, *et al*. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010;**11**:733–9.
36. Carralot J-P, Ogier A, Boese A, *et al*. A novel specific edge effect correction method for RNA interference screenings. *Bioinformatics* 2012;**28**:261–8.
37. Armknecht S, Boutros M, Kiger A, *et al*. High-throughput RNA interference screens in drosophila tissue culture cells. *Methods Enzymol* 2005;**392**:55–73.
38. Lo E, Soleilhac E, Martinez A, *et al*. Intensity quantile estimation and mapping—a novel algorithm for the correction of image non-uniformity bias in HCS data. *Bioinformatics* 2012;**28**:2632–9.
39. Snijder B, Sacher R, Rämö P, *et al*. Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* 2009;**461**:520–3.
40. Knapp B, Rebhan I, Kumar A, *et al*. Normalizing for individual cell population context in the analysis of high-content cellular screens. *BMC Bioinformatics* 2011;**12**:485.
41. Helm JS, Hu Y, Chen L, *et al*. Identification of active-site inhibitors of murg using a generalizable, high-throughput glycosyltransferase screen. *J Am Chem Soc* 2003;**125**:11168–9.
42. Makarenkov V, Kevorkov D, Zentilli P, *et al*. HTS-Corrector: software for the statistical analysis and correction of experimental high-throughput screening data. *Bioinformatics* 2006; **22**:1408–9.
43. Vassal E, Barette C, Fonrose X, *et al*. Miniaturization and validation of a sensitive multiparametric cell-based assay for the concomitant detection of microtubule-destabilizing and microtubule-stabilizing agents. *J Biomol Screen* 2006;**11**:377–89.
44. Box GE, Hunter JS, Hunter WG. *Statistics for Experimenters: Design, Innovation, and Discovery*. *Wiley-Interscience*, USA, 2005.
45. Hays WL. *Statistics*. Fort Worth, TX: Harcourt Brace Jovanovich, 1994.
46. Fisher RA. *Statistical Methods for Research Workers*. *Oliver & Boyd,* Edinburgh, Genesis Publishing Pvt Ltd, 1925.
47. Verdugo RA, Deschepper CF, Muñoz G, *et al*. Importance of randomization in microarray experimental designs with Illumina platforms. *Nucleic Acids Res* 2009;**37**:5610–18.
48. Box G.E. *Improving Almost Anything: Ideas and Essays*. *Wiley*, USA, 2006.
49. Boutros M, Bras LP, Huber W. Analysis of cell-based RNAi screens. *Genome Biol* 2006;**7**:R66.
50. Birmingham A, Selfors LM, Forster T, *et al*. Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* 2009;**6**:569–75.
51. Blainey P, Krzywinski M, Altman N. Points of significance: replication. *Nat Methods* 2014;**11**:879–80.
52. Murie C, Barette C, Lafanechère L, *et al*. Single assay-wide variance experimental (SAVE) design for high-throughput screening. *Bioinformatics* 2013;**29**:3067–72.
53. Elowe NH, Blanchard JE, Cechetto JD, *et al*. 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* 2005;**10**:653–7.
54. Gagarin A, Makarenkov V, Zentilli P. Using clustering techniques to improve hit selection in high-throughput screening. *J Biomol Screen* 2006;**11**:903–14.
55. Müller P, Kuttenkeule D, Gesellchen V, *et al*. Identification of JAK/STAT signalling components by genome-wide RNA interference. *Nature* 2005;**436**:871–5.
56. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* 2003;**19**:2448–55.
57. Zhang XHD, Yang XC, Chung N, *et al*. Robust statistical methods for hit selection in RNA interference high-throughput screening experiments. *Pharmacogenomics* 2006;**7**: 299–309.
58. Zhang XHD. A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics* 2007;**89**:552–61.
59. König R, Chiang CY, Tu BP, *et al*. A probability-based approach for the analysis of large-scale RNAi screens. *Nat Methods* 2007; **4**:847–9.
60. Welch BL. The generalization ofstudent's' problem when several different population variances are involved. *Biometrika* 1947;**34**:28–35.
61. Gibbons JD. *Nonparametric Methods for Quantitative Analysis*. Columbus, OH: American Sciences Press Columbus, 1985.
62. Koziol JA. The rank product method with two samples. *FEBS Letters* 2010;**584**:4481–4.
63. Cooley JW, Tukey JW. An algorithm for the machine calculation of complex Fourier series. *Math Comput* 1965;**19**: 297–301.
64. D'Agostino RB, Stephens MA. Goodness-of-Fit Technicques. *Marcel Dekker,* INC, New York: CRC Press, 1986.
65. Kelley B. Automated detection of systematic errors in array experiments. *JALA Charlottesv Va* 2003;**8**:24–6.
66. Breitling R, Armengaud P, Amtmann A, *et al*. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Letters* 2004;**573**:83–92.
67. Rieber N, Knapp B, Eils R, *et al*. RNAither, an automated pipeline for the statistical analysis of high-throughput RNAi screens. *Bioinformatics* 2009;**25**:678–9.
68. Rieber N, Kaderali L. RNAither, an automated pipeline for the statistical analysis of high-throughput RNAi screens, *Bioconductor*, Overview for open source software RNAither, 2014. http://www.bioconductor.org/packages/release/bioc/vignettes/RNAither/inst/doc/vignetteRNAither.pdf.
69. Gilat A. *MATLAB: An Introduction with Applications*, 5th edn (Solutions Manual), Wiley, 2014, 416.
70. Wu Z, Liu D, Sui Y. Quantitative assessment of hit detection and confirmation in single and duplicate high-throughput screenings. *J Biomol Screen* 2008;**13**:159–67.
71. Venables WN, Ripley BD, Venables W. *Modern Applied Statistics with S-PLUS*. New York: *Springer-Verlag,* 1994.
72. Tukey JW. *Exploratory Data Analysis*. Addison-Wesley, 1977.
73. Gosai SJ, Kwak JH, Luke CJ, *et al*. Automated high-content live animal drug screening using C. elegans expressing the aggregation prone serpin α1-antitrypsin Z. PloS One 2010;**5**: e15460.
74. Yu DN, Danku J, Baxter I, *et al*. Noise reduction in genome-wide perturbation screens using linear mixed-effect models. *Bioinformatics* 2011;**27**:2173–80.
75. Baryshnikova A, Costanzo M, Kim Y, *et al*. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat Methods* 2010;**7**:1017–24.
76. Bushway PJ, Azimi B, Heynen-Genel S, *et al*. Hybrid median filter background estimator for correcting distortions in microtiter plate data. *Assay Drug Dev Technol* 2010;**8**:238–50.
77. Murie C, Barette C, Lafanechère L, *et al*. Control-Plate Regression (CPR) normalization for high-throughput screens with many active features. *J Biomol Screen* 2013;**19**:661–71.
78. Ogier A, Dorval T. HCS-Analyzer: open source software for high-content screening data correction and analysis. *Bioinformatics* 2012;**28**:1945–46.