Novel Analytic Criteria and Effective Plate Designs for Quality Control in Genome-Scale RNAi Screens

Xiaohua Douglas Zhang

J Biomol Screen 2008 13: 363
DOI: 10.1177/1087057108317062

Novel Analytic Criteria and Effective Plate Designs for Quality Control in Genome-Scale RNAi Screens

XIAOHUA DOUGLAS ZHANG[1]

One of the most fundamental challenges in genome-wide RNA interference (RNAi) screens is to glean biological significance from mounds of data, which relies on the development and adoption of appropriate analytic methods and designs for quality control (QC) and hit selection. Currently, a Z-factor-based QC criterion is widely used to evaluate data quality. However, this criterion cannot take into account the fact that different positive controls may have different effect sizes and leads to inconsistent QC results in experiments with 2 or more positive controls with different effect sizes. In this study, based on a recently proposed parameter, strictly standardized mean difference (SSMD), novel QC criteria are constructed for evaluating data quality in genome-wide RNAi screens. Two good features of these novel criteria are: (1) SSMD has both clear original and probability meanings for evaluating the differentiation between positive and negative controls and hence the SSMD-based QC criteria have a solid probabilistic and statistical basis, and (2) these QC criteria obtain consistent QC results for multiple positive controls with different effect sizes. In addition, I propose multiple plate designs and the guidelines for using them in genome-wide RNAi screens. Finally, I provide strategies for using the SSMD-based QC criteria and effective plate design together to improve data quality. The novel SSMD-based QC criteria, effective plate designs, and related guidelines and strategies may greatly help to obtain high quality of data in genome-wide RNAi screens. (*Journal of Biomolecular Screening* 2008:363-377)

Key words: strictly standardized mean difference, Z factor, quality control, plate design, RNAi high-throughput screening

INTRODUCTION

RNA interference (RNAi) is a natural regulatory mechanism in many organisms in which double-stranded RNA directs the posttranscriptional silencing of target genes in a sequence-specific manner.[1] The "functional" component of RNAi is small interfering RNA (siRNA). RNAi offers a potential safe and effective way of turning off a gene[1,2] and has been seen as the 3rd class of drugs, after small molecules and proteins.[3,4] Genome-wide RNAi researches using RNAi high-throughput screening (HTS) biotechnology open new avenues for elucidating gene functions and have revealed rich, often unappreciated insights into many biological processes.[5-11] As in any high-throughput platforms, one of the most fundamental challenges in HTS is to glean biological significance from mounds of data, which relies on the development and adoption of appropriate statistical designs and analytic methods for quality control (QC) and hit selection.[12]

[1]Biometrics Research, Merck Research Laboratories, West Point, Pennsylvania.

Received Nov 2, 2007, and in revised form Feb 14, 2008. Accepted for publication Feb 15, 2008.

Journal of Biomolecular Screening 13(5); 2008
DOI: 10.1177/1087057108317062

An RNAi HTS experiment is usually conducted in 96-well, 384-well, or 1536-well plates in which the cells in a well are treated with a unique siRNA or a control. A typical RNAi HTS project starts with a primary screen of about 20,000 to 50,000 siRNAs most of which have no replicate. The siRNAs identified (called "hits") in the primary screen are further investigated using 1 or more confirmatory screens in which each siRNA has replicates. Currently, a typical primary screen has fifty to two hundred 384-well plates and a typical confirmatory screen has three to twenty 384-well plates. The measured response is usually the intensity emitted by labeled particles such as fluorescent dyes.[5] For simplicity, in this study, the term "intensity of an siRNA" is used to refer to "the measured intensity of a phenotype corresponding to the treatment of an siRNA," which may be light intensity, the intensity emitted by a dye, or the ratio of intensity emitted by 2 dyes depending on individual experiments. There are, at least, a positive control with specific knockdown effects (usually occupying 4, 8, or 16 wells) and at least a negative control with unspecific siRNA knockdown effects (usually occupying 4, 6, 8, 16, or 20 wells) in each plate. The quality of data in a screen is usually indicated by the differentiation of measured intensity between the positive control(s) and the negative control in a plate.

As analytic metrics, signal-to-background ratio, signal-to-noise ratio, signal window, assay variability ratio, and Z factor

have been adopted to evaluate data quality.[11,13-22] Signal-to-background ratio and signal-to-noise ratio have problems in capturing data variability; consequently, QC results reached using signal-to-background ratio and signal-to-noise ratio are misleading.[15,19,21,23] Signal window, assay variability ratio, and Z factor all capture data variability. Simulation studies have shown that Z factor is better than signal window in both accuracy and precision.[21] Assay variability ratio is equivalent to Z factor. Strictly standardized mean difference (SSMD) has recently been proposed for assessing data quality in RNAi HTS assays.[23] Compared with Z factor, SSMD is more favorable in terms of both probabilistic interpretation and statistical estimation and inference. This advantage of SSMD is important because of the stochastic feature of readouts in HTS assays.

Quality assessment metrics measure the differentiation between 1 or more positive controls and a negative reference. The degree of differentiation between a positive control and a negative reference also depends on both the quality of an assay and the size of positive control. Among the assays that have positive controls with the same effect size, the larger the degree of differentiation is, the better quality the assay has. Meanwhile, among different positive controls in the assays with the same quality, the greater the size of a positive control is, the larger the degree of differentiation is. Consequently, in the situation where different positive controls are used in an assay, the degree of differentiation between a strong positive control and a negative reference will, by definition, be larger than between a moderate positive control and a negative reference. Therefore, QC criteria for the strong positive control should be different from those for the moderate positive control.

Currently, the popularly used Z-factor-based QC criterion[19] is most suitable for the strong or very strong positive controls. So is the recently proposed simple SSMD-based criterion[23] in which an assay with SSMD ≥ 3 (SSMD ≤ -3) passes QC and fails otherwise. In an RNAi HTS assay, a moderate or fairly strong positive control is usually more instructive than the strong or very strong positive control, because the effectiveness of this control is more similar to HTS hits of interest. The simple Z-factor-based QC criterion or the simple SSMD-based QC criterion to moderate or fairly strong positive control may not work effectively in helping to identify true hits of interest. Furthermore, applying the simple Z-factor-based QC criterion or the simple SSMD-based QC criterion to multiple positive controls with different effect sizes will lead to inconsistent QC results: passes QC for the strong or very strong positive control and fails QC for the moderate or fairly strong positive control. In this study, I prove the links between d^+-probability[23-25] and SSMD, and demonstrate that the population value of SSMD effectively quantify data quality in a plate. Based on both the original and the probabilistic meanings of SSMD, I propose detailed criteria of using SSMD for QC in an HTS assay and present strategies for using analytic metrics for quality control in RNAi HTS experiments.

The arrangement of control and sample wells in a plate is called "plate design." The existence of systematic errors of measurement is not uncommon in HTS experiments.[13,26-28] A good plate design helps to identify systematic errors (especially those linked with well position) and to determine what normalization we should use to adjust the data so that we can remove/reduce the impact of systematic errors on both QC and hit selection. Currently, the commonly used plate designs cannot effectively identify and adjust for systematic errors.[13,26,27,29] A plate design has been described for adjusting for column effects in screens conducted in 96-well plates.[13] In this study, I propose plate designs that more effectively display and adjust for systematic errors in screens conducted in 96-well, 384-well, or 1536-well plates. Using simulation studies, I demonstrate the usefulness of these designs to adjust for column effects, row effects, and edge effects and present basic guidelines for the adoption of plate designs in RNAi HTS experiments.

NOVEL QUALITY ASSESSMENT CRITERIA

Quality assessment metrics

In a typical HTS experiment, a clear distinction between a positive control and a negative reference such as a negative control is an index for good quality. As described in the introductory section, many quality assessment measures have been proposed to measure the degree of differentiation between a positive control and a negative reference. The following formulas for the estimates of these measures show the commonalities and differences of these measures. Let $\bar{X}_P$ and s_P be the sample mean and standard deviation of a positive control respectively and $\bar{X}_N$ and s_N be the sample mean and standard deviation of a negative reference respectively. Then the estimates of quality assessment measures are:

$$\text{signal-to-background ratio} = \frac{\bar{X}_P}{\bar{X}_N},$$

$$\text{signal-to-noise ratio} = \frac{\bar{X}_P - \bar{X}_N}{s_N},$$

$$\text{signal window} = \frac{|\bar{X}_P - \bar{X}_N| - 3(s_P + s_N)}{s_N},$$

$$\text{assay variability ratio} = \frac{3(s_P + s_N)}{|\bar{X}_P - \bar{X}_N|},$$

$$Z \text{ factor} = \frac{|\bar{X}_P - \bar{X}_N| - 3(s_P + s_N)}{|\bar{X}_P - \bar{X}_N|} = 1 - \frac{3(s_P + s_N)}{|\bar{X}_P - \bar{X}_N|} \quad (1)$$

and

$$\text{SSMD} = \frac{\bar{X}_P - \bar{X}_N}{\sqrt{s_P^2 + s_N^2}}. \quad (2)$$

SSMD and signal-to-noise ratio have the same numerator but different denominators. Signal-to-noise ratio incorporates variability only in the negative reference where SSMD accounts for variability in both controls. Consequently, signal-to-noise ratio leads to misleading results when the 2 controls have different variability whereas SSMD does not. Clearly, assay variability ratio = 1 + Z factor; thus assay variability ratio and Z factor are equivalent. Iversen et al[21] show that Z factor is better than signal window in terms of accuracy and precision. SSMD is similar to the reverse of assay variability ratio. Both take into account variability in both controls; however, the way to combine the information about the variability in the 2 controls is different: assay variability ratio and Z factor directly sum up the standard deviations in the 2 controls whereas SSMD adopts the standard deviation of the difference between the positive and negative controls. As a result, assay variability ratio and Z factor are conveniently interpreted using graphics or "control chart" whereas SSMD is more favorable in terms of both probabilistic interpretation and statistical estimation and inference. As proved in the Appendix, SSMD has a direct relationship with the probability of the difference between the 2 controls being positive. This relationship offers a strong base for a probabilistic interpretation to SSMD-based criteria. By contrast, there is no direct relationship between Z factor and related probability although Sui and Wu[22] roughly give a power interpretation to Z-factor-based criteria under certain assumptions and limitations and Zhang[23] shows that Z factor > 0 is a subset of |SSMD| > 3 and Z factor > 0.5 is a subset of |SSMD| > 6. SSMD values are also easily interpreted by the number of folds of average value to variability of the difference between the 2 controls. The absolute sign in Z factor makes it nontrivial to derive the estimation and inference about Z factor.

It is trivial to derive the estimation and confidence interval of SSMD from a complete statistical basis whereas it is nontrivial to get the confidence interval of Z factor mathematically. When the positive control and the negative reference have equal variance, a better estimate (namely, minimal variance unbiased estimate) of SSMD is

$$\begin{aligned} \text{SSMD} &= \frac{\bar{X}_P - \bar{X}_N}{\sqrt{\frac{2}{K}((n_P-1)s_P^2 + (n_N-1)s_N^2)}} \\ &\approx \frac{\bar{X}_P - \bar{X}_N}{\sqrt{\frac{2}{n_P+n_N-3.5}((n_P-1)s_P^2 + (n_N-1)s_N^2)}} \end{aligned} \quad (3)$$

when $n_P, n_N, \geq 2$, where n_P and n_N are sample sizes of the positive control and the negative reference respectively and $K = 2 \cdot \left(\frac{\Gamma\left(\frac{n_P+n_N-2}{2}\right)}{\Gamma\left(\frac{n_P+n_N-3}{2}\right)}\right)^2 \approx n_P + n_N - 3.5$.

SSMD estimate in equation 2 looks similar to the *t*-statistic in situation of unequal variance,

$$t\text{-statistic} = \frac{\bar{X}_P - \bar{X}_N}{\sqrt{s_P^2/n_P + s_N^2/n_N}}. \quad (4)$$

SSMD estimate in formula (3) looks similar to the *t*-statistic in situation of equal variance,

$$t\text{-statistic} = \frac{\bar{X}_P - \bar{X}_N}{\sqrt{\frac{1}{n_P+n_N-2}((n_P-1)s_P^2 + (n_N-1)s_N^2)} \cdot \sqrt{\frac{1}{n_P} + \frac{1}{n_N}}}. \quad (5)$$

However, there is a major difference between *t*-statistic and SSMD. That is, in situations of either equal or unequal variance, when n_P and/or n_N increase, *t*-statistics increase and corresponding *p*-values decrease whereas SSMD estimates tend to be closer to its population value; as n_P and n_N go to infinity, *t*-statistics approach infinity and corresponding *p*-values approach zero whereas SSMD estimates approach its population value.[23] In other words, a *t*-statistic is a function of both sample size of controls and degree of differentiation between 2 controls and is thus highly affected by sample size of controls whereas the values of an SSMD estimate fall around SSMD population value. This difference makes *t*-statistic a poor metric and makes SSMD a good metric for assessing data quality.

For example, no matter how poor an assay one has, one can increase *t*-statistic to reach a very large value (or reduce the corresponding *p*-value to reach a very small value) simply by increasing sample size of controls in a plate. By contrast, increasing sample size of controls will make an SSMD estimate approach to its population value more closely and will not surely increase value of the SSMD estimate. The population value of SSMD reflects only the degree of differentiation between 2 controls and is not affected by sample size of controls. Z factor should have a similar property to SSMD in terms of sample size impacts, which may be 1 major reason why Z factor has been widely used to assess quality in HTS assays whereas *t*-statistic has not. Statistical power for testing null hypotheses has been proposed for a QC metric.[22] However, power is a more complicated term than is *t*-statistic and its corresponding *p*-value. Power is further affected by type I error besides by effect size and sample size of controls. Thus, *t*-statistic, *p*-value, and power are all most suitable for null hypothesis testing, but are not suitable for measuring effect size (see a book[30,31] edited by Harlow, Mulaik, and Steiger for serious criticisms). Therefore, *t*-statistic, *p*-value, and power can hardly serve as QC metrics in RNAi HTS assays especially given the consideration that many siRNAs may have tiny effects on measured response.

Increasing sample size of controls should affect the precision but not the overall value of a good QC metric. On the other hand, increasing the sample size of each sample siRNA or compound in a plate should increase assay quality when hit selection is correspondingly based on the mean or median of replicates of each sample siRNA or compound, not the value of individual

sample wells. Assuming the replicate of every sample siRNA (or compound) in a plate is r, the calculation corresponding to equations 1, 2, and 3 are respectively,

$$Z \text{ factor}_{\text{r}} = 1 - \frac{3(s_P + s_N)}{|\bar{X}_P - \bar{X}_N|}\frac{1}{\sqrt{r}}, \tag{6}$$

$$\text{SSMD}_{\text{r}} = \frac{\bar{X}_P - \bar{X}_N}{\sqrt{s_P^2 + s_N^2}}\sqrt{r}, \tag{7}$$

and

$$\begin{aligned}\text{SSMD}_{\text{r}} &= \frac{\bar{X}_P - \bar{X}_N}{\sqrt{\frac{2}{K}((n_P-1)s_P^2 + (n_N-1)s_N^2)}}\sqrt{r} \\ &\approx \frac{\bar{X}_P - \bar{X}_N}{\sqrt{\frac{2}{n_P+n_N-3.5}((n_P-1)s_P^2 + (n_N-1)s_N^2)}}\sqrt{r}\end{aligned} \tag{8}$$

where n_P and n_N are sample sizes. One reminder is that equations 7 and 8 are adopted only in situations where the process of hit selection is first to calculate the mean or median of replicates of each sample siRNA or compound in each plate, and then to select hits based on the summarized value of each sample siRNA or compound, not the value of individual sample wells.

SSMD-based QC criteria

A Z-factor-based criterion for QC (namely, Z factor = 1 for "ideal," 1 > Z factor ≥ 0.5 for "excellent," 0.5 > Z factor > 0 for "doable," Z factor = 0 for "yes/no type," and Z factor < 0 for "screen essentially impossible")[19] has widely been used to measure the quality of data in a plate.[11,13-19] However, this criterion does not consider the fact that positive controls in different experiments may have different sizes of effects. As a result, applying the same criterion to positive controls with different sizes of effects will lead to misleading results. For example, in a simulation study where the assay quality in each simulated experiment is ideal: data variability consisting of only biological variability but no assay variability, the positive control in experiment A has stronger inhibition effects than the positive control in experiment B (panels **A1** and **B1** of **Fig. 1**). If the above Z-factor-based QC criterion is used in both experiments, 24 plates are "screen essentially impossible," 76 "doable" in experiment A (panel **A3** of **Fig. 1**), and 100 "screen essentially impossible" in experiment B (panel **B3** of **Fig. 1**). Because, theoretically, the positive controls in the 2 experiments have different sizes of inhibition effects, we should use different QC thresholds in these 2 experiments. Otherwise, assays with good quality to distinguish moderate or fairly strong positive controls may be judged as having poor quality (such as in experiment B). This is important especially given the consideration that a moderate or fairly strong positive control is usually more instructive and relevant to the hits of interest than a very strong positive control.

In addition, it is common that 2 or more positive controls are adopted in a single experiment such as in experiment C (**Fig. 1**). Applying the same Z-factor-based QC criterion to both controls leads to inconsistent results: 6 "excellent" and 93 "doable" by 1 positive control represented by purple points and 5 "doable" and 95 "screen essentially impossible" by the other control represented by red points (panel **C3** of **Fig. 1**). Considering the fact that different positive controls may have different sizes of effects, we may think about constructing different criteria for positive controls with different sizes of effects based on Z factor. However, considering the fact that SSMD has a better probabilistic and statistical base than Z factor, I construct detailed QC criteria based on SSMD instead of Z factor. These criteria take into account the size and direction of effects of positive controls in an HTS assay.

An siRNA with its population value of SSMD between 1 and 2, between 2 and 3, or between 3 and 4.7 has a moderate, fairly strong, or strong effect, respectively.[24] In most RNAi HTS experiments, the positive controls have moderate, fairly strong, or strong effects. For a fairly strong positive control, if the population value of SSMD were known, we would declare a plate to have good quality if the SSMD population value is between 2 and 3. In reality, the population value of SSMD is usually unknown and we need to use the estimated value of SSMD to approximately represent its population value. Considering the variability in SSMD estimation, for a plate with a fairly strong positive control whose value is theoretically greater than the negative control, it is reasonable to judge the plate to have excellent quality if the estimated SSMD value is above 3, good quality if between 2 and 3, inferior quality if between 1 and 2, and poor quality if below 1. Thus, we have criterion Ib of **Table 1**. Following this idea, I construct 8 SSMD-based QC criteria, criteria Ia, Ib, Ic, Id, IIa, IIb, IIc, and IId, which take into account the size and direction of effects of positive controls in an HTS assay (**Table 1**).

The thresholds in the SSMD-based criteria have a theoretical basis and a probabilistic interpretation. The rationale for these thresholds is as follows. The SSMD values of 0.5, 1, 2, and 3 have clear meanings: the size of mean difference being 0.5×, 1×, 2×, and 3× the standard deviation of the difference respectively. The d^+-probability associated with SSMD of an siRNA is the probability that a value from this siRNA is greater than a value from a negative reference. Based on the relationship between SSMD and d^+-probability, the SSMD values of 1, 2, and 3 (or –1, –2 and –3) also indicate that the minimums (or maximums) of the corresponding d^+-probabilities are respectively about 0.5, 0.95, and 0.975 (or 0.5, 0.05, 0.025) in the situation where the difference has a symmetric unimodal distribution with finite variance. Similarly, the SSMD value of 4.7 (or –4.7) indicates that the minimums (or maximums) of the

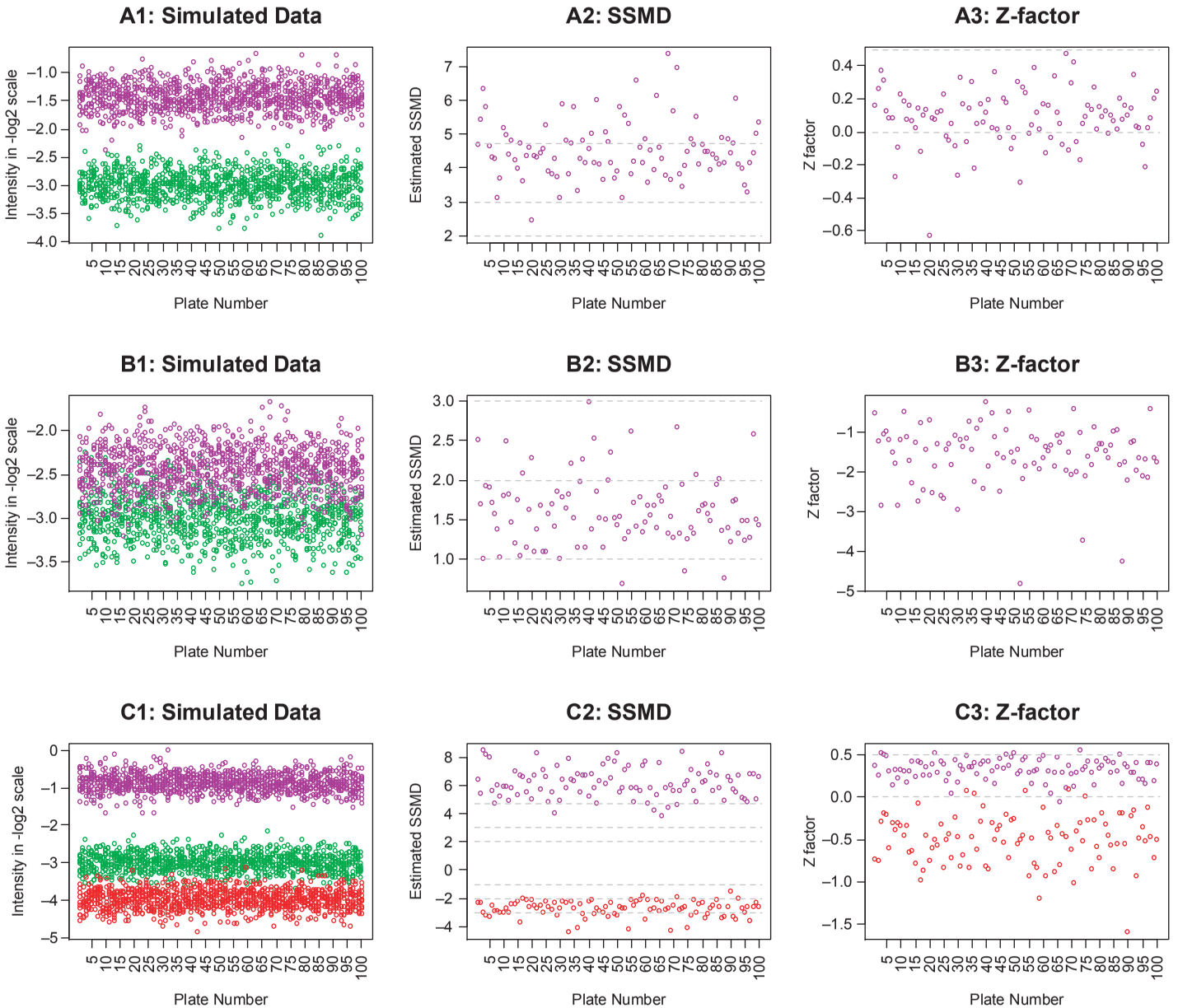

FIG. 1. Data, strictly standardized mean difference (SSMD), and Z factor in 3 simulated experiments, A, B, and C, in which the positive controls have different effect sizes: a very strong inhibition control (purple points) in experiment A, a fairly strong inhibition control (purple points) in experiment B, a strong inhibition control (purple points) and a moderate activation control (red points) in experiment C. In each simulated experiment, there are 100 plates each with 10 replicates for each positive or negative control. The data for each control in each experiment are generated from a normal distribution with standard deviation of 0.25, namely, $N(\mu, 0.25^2)$, where $\mu = -3$ for each negative control in the 3 experiments, and $\mu = -1.44, -2.43, -0.88, -3.99$ for the strong positive control in experiment A, moderate positive control in experiment B, very strong positive control and fairly strong positive control in experiment C, respectively.

Table 1. SSMD-Based QC Criteria in RNAi HTS Assays Taking into Account Effect Size of a Positive Control in an Assay under Two Situations

Situation I The intensity of a positive control is theoretically greater than that of a negative reference.				
Quality Types	*Criterion Ia: For a Moderate Control*	*Criterion Ib: For a Fairly Strong Control*	*Criterion Ic: For a Strong Control*	*Criterion Id: For a Very Strong Control*
Excellent	$\hat{\beta} \geq 2$	$\hat{\beta} \geq 3$	$\hat{\beta} \geq 4.7$	$\hat{\beta} \geq 6.67$
Good	$2 > \hat{\beta} \geq 1$	$3 > \hat{\beta} \geq 2$	$4.7 > \hat{\beta} \geq 3$	$6.67 > \hat{\beta} \geq 4.7$
Inferior	$1 > \hat{\beta} \geq 0.5$	$2 > \hat{\beta} \geq 1$	$3 > \hat{\beta} \geq 2$	$4.7 > \hat{\beta} \geq 3$
Poor	$\hat{\beta} < 0.5$	$\hat{\beta} < 1$	$\hat{\beta} < 2$	$\hat{\beta} < 3$
Situation II The intensity of a positive control is theoretically less than that of a negative reference.				
Quality Types	*Criterion IIa: For a Moderate Control*	*Criterion IIb: For a Fairly Strong Control*	*Criterion IIc: For a Strong Control*	*Criterion IId: For a Very Strong Control*
Excellent	$\hat{\beta} \leq -2$	$\hat{\beta} \leq -3$	$\hat{\beta} \leq -4.7$	$\hat{\beta} \leq -6.67$
Good	$-2 < \hat{\beta} \leq -1$	$-3 < \hat{\beta} \leq -2$	$-4.7 < \hat{\beta} \leq -3$	$-6.67 < \hat{\beta} \leq -4.7$
Inferior	$-1 < \hat{\beta} \leq -0.5$	$-2 < \hat{\beta} \leq -1$	$-3 < \hat{\beta} \leq -2$	$-4.7 < \hat{\beta} \leq -3$
Poor	$\hat{\beta} > -0.5$	$\hat{\beta} > -1$	$\hat{\beta} > -2$	$\hat{\beta} > -3$

Note: $\hat{\beta}$ denotes estimated strictly standardized mean difference (SSMD) value. In practice, we usually know whether the measured intensity of a positive control is theoretically larger than a negative reference. If effect size of a positive control is known biologically, adopt the corresponding criterion based on this table; otherwise, use the strategies in **Table 3** for choosing which SSMD-based criterion to be used in an RNAi high-throughput screening (HTS) experiment. QC, quality control; RNAi, RNA interference.

corresponding d^+-probabilities are 0.99 (or 0.01) respectively. In the situation where the difference has a normal distribution, SSMD = 0.5, 1, 2, 3, and 4.7 corresponds to d^+-probability = 0.69, 0.84, 0.97725, 0.99865, and 0.9999987. The SSMD-based criteria consider various distributions and are thus robust to different symmetric distributions.

Strategies for using analytic metrics for quality control

To use the SSMD-based QC criteria for judging whether a plate passes or fails QC in a screen with only 1 positive control, the following strategy may be adopted: a plate passes QC if it has excellent or good quality; it fails QC if it has inferior or poor quality. When multiple positive controls are used in an HTS screen, we may need to evaluate data quality based on 2 or more positive controls. In many cases, to pass QC in a plate, we may need both positive controls to pass QC in that plate, especially in the experiments with the objective of selecting both the siRNAs with strong effects and those with moderate or fairly strong effects. The strategies for the adoption of SSMD-based criteria for QC in an HTS experiment are summarized in **Table 2**.

As demonstrated in the previous section, for the simulation study, the usage of the popularly used Z-factor-based QC criterion leads to misleading QC results in experiments where a weaker positive control is adopted; it also leads to inconsistent results in experiments where 2 positive controls have different sizes of effects. Now, based on the SSMD-based QC criteria and related strategies, criteria Ic and Ia are applied to the strong positive control in experiment A and the moderate positive control in experiment B respectively, which produces the following sensible QC evaluation results: 32 excellent, 67 good, and 1 inferior plates in experiment A, and 14 excellent, 83 good, and 3 inferior plates in experiment B. Criteria Id and IIb are applied to the very strong positive control and the fairly strong positive control in experiment C respectively, which yields the following evaluation results: 30 excellent, 64 good, 6 inferior plates by the very strong positive control and 30 excellent, 67 good, 3 inferior plates by the fairly strong positive control. All the QC results obtained using SSMD-based QC criteria are much more reasonable than those obtained using the Z-factor-based QC criterion.

To apply SSMD-based QC criteria listed in **Table 1**, we need to know the size and direction of effects of positive controls in an HTS assay. In practice, we usually know whether the

Table 2. Strategies for Quality Controls in an HTS Experiment with One or Two Positive Controls

Strategy 1: Using 1 positive control,
a plate passes QC if it has good or excellent quality;
a plate fails QC if it has poor or inferior quality.
Strategy 2: Using 2 positive controls,
a plate passes QC if it has good or excellent quality in both positive controls;
a plate fails QC if it has inferior or poor quality in both positive controls;
depending on experimental need and cost, a plate may pass or fail QC if it has good or excellent quality in 1 positive control and inferior or poor quality in the other positive control.

Note: HTS, high-throughput screening; QC, quality control.

Table 3. Strategies for Using SSMD-Based Criteria for QC in an HTS Experiment in Situations where There Is Not Enough Information about Effect Sizes of Positive Controls

- For the RNAi HTS assays in which cell viability is the measured response, criterion Id or IId should be adopted for the controls without cells (namely, the wells with no cells added) or background controls.
- If the difference is not normally distributed especially when it is highly skewed, QC criteria Id and IId may be used even for a strong or fairly strong positive control.
- In a viral assay in which the amount of viruses in host cells is the interest, criterion Ic or IIc is usually used and criterion Id or IId is occasionally used for the positive control consisting of siRNA from the virus.
- If there is not enough information about 1 or more positive controls each consisting of an siRNA,
  - adopt criterion Ic or IIc when there is only 1 positive control in an experiment;
  - adopt criterion Ic or IIc for the stronger positive control and criterion Ib or IIb for the weaker positive control when there are 2 positive control in an experiment.
- In compound HTS assays, positive controls usually have very strong effects and thus usually criterion Id or IId (and occasionally criteria Ic and IIc) should be adopted in compound HTS assays.

Note: SSMD, strictly standardized mean difference; QC, quality control; HTS, high-throughput screening, RNAi, RNA interference.

measured intensity of a positive control is theoretically greater than a negative reference. It is more difficult to obtain the information about the sizes of positive controls. In practice, most positive controls adopted in RNAi HTS experiments have moderate, fairly strong, or strong effects. Therefore, the QC criteria Ia, Ib, Ic, IIa, IIb, and IIc listed in **Table 1** work for most RNAi HTS experiments. In some cases, a good positive biological control is unavailable for experimenters and some very strong positive controls must be used. For example, in some experiments where the goal is to screen out siRNAs capable of inhibiting cancer cell growth, the wells with no cells are used as a positive control because no better positive biological controls are available. Because these positive control wells have no cells added at all, the size of inhibition effect in this positive control is very strong. In such a case, we may use QC criterion Id or IId. The detailed strategies for adopting an SSMD-based QC criterion for a positive control in an HTS experiment in situations where there is not enough information about the sizes of positive controls are described in **Table 3**.

EFFECTIVE PLATE DESIGNS

Construction of plate designs

Currently, RNAi HTS screens are often conducted in 384-well plates (e.g., designs A1, B1, and C1–C5 in **Fig. 2**). Hence we focus on design of wells in 384-well plates. In RNAi HTS experiments, there commonly exist systematic errors of measurement that have either a linear or a bowl-shaped relationship to well position. For example, the data in **Figure 3** have systematic errors with a linear relationship to column number (i.e., lower values in left columns and higher values in right columns). The data in **Figure 4** have systematic errors with a bowl-shaped relationship to well position (i.e., lower values in the middle and higher values in the edge).

A typical format for current plate designs is that multiple types of controls are arranged in the 4 edge columns and sample siRNAs are arranged in the remaining columns (e.g., designs A1 and B1 in **Fig. 2**). Design A1 or similar designs are commonly used in current RNAi HTS experiments. However, these designs cannot effectively identify and adjust for systematic errors, especially in confirmatory screens where we cannot use the majority of sample wells to display potential systematic errors, because the sample siRNAs are preselected to have inhibition or activation effects. For example, the linear or bowl-shaped pattern in the systematic errors is not apparent based on the negative control data in panel **A2** of **Figures 3** and **4**. Because the negative control wells only occupy 1 column, they cannot be used to adjust for the systematic errors. On the other hand, design B1 displays and adjusts for potential systematic errors more effectively than design A1 because, in design B1, we may use the negative control alone to identify systematic errors linearly linked to row/column numbers. For example, the overall values of negative control wells in the 2nd columns are lower than in the 23rd columns, which indicates the existence of systematic errors linearly linked to column number (green points in panel **B2** of **Fig. 3**). Using either a linear model or a smoothing method based on the negative control, we remove the effect of the systematic errors (panel **B3** of **Fig. 3**).

Both designs A1 and B1 do not allow for the control wells to be arranged in the middle columns. Thus, the data of negative control wells in those designs cannot effectively display the systematic errors in the middle columns such as the bowl-shaped pattern (panels **A2** and **B2** of **Fig. 4**) and thus cannot help to remove/reduce impact of these systematic errors (panels **A3** and **B3** of **Fig. 4**). To display and adjust for systematic errors with various patterns, we need plate designs with controls arranged in the middle of a plate.

Considering the fact that robots can now readily arrange controls anywhere in a plate, I propose designs C1–C5 (**Fig. 2**). The arrangement of the negative control in design C1 is the same as in design C2, which can effectively display various patterns of systematic errors especially for those with a sharp change in the edge row/column. Design C3 is a variant of design C1 or C2 that allows for 1 more column to accommodate other controls. Similar variants can be obtained to allow for 4 edge columns to accommodate other controls. Designs C4 and C5 may not work as effectively as designs C1–C3 especially when there is a sharp change in the edge; however, fewer wells are needed for the negative control (20 and 16 respectively in designs C4 and C5, compared to 24 in designs C1–C2). For designs C4 and C5, positive controls can be arranged in the edge similar to design C1 or in the middle similar to design C2. Designs C1–C5 help to display sys-

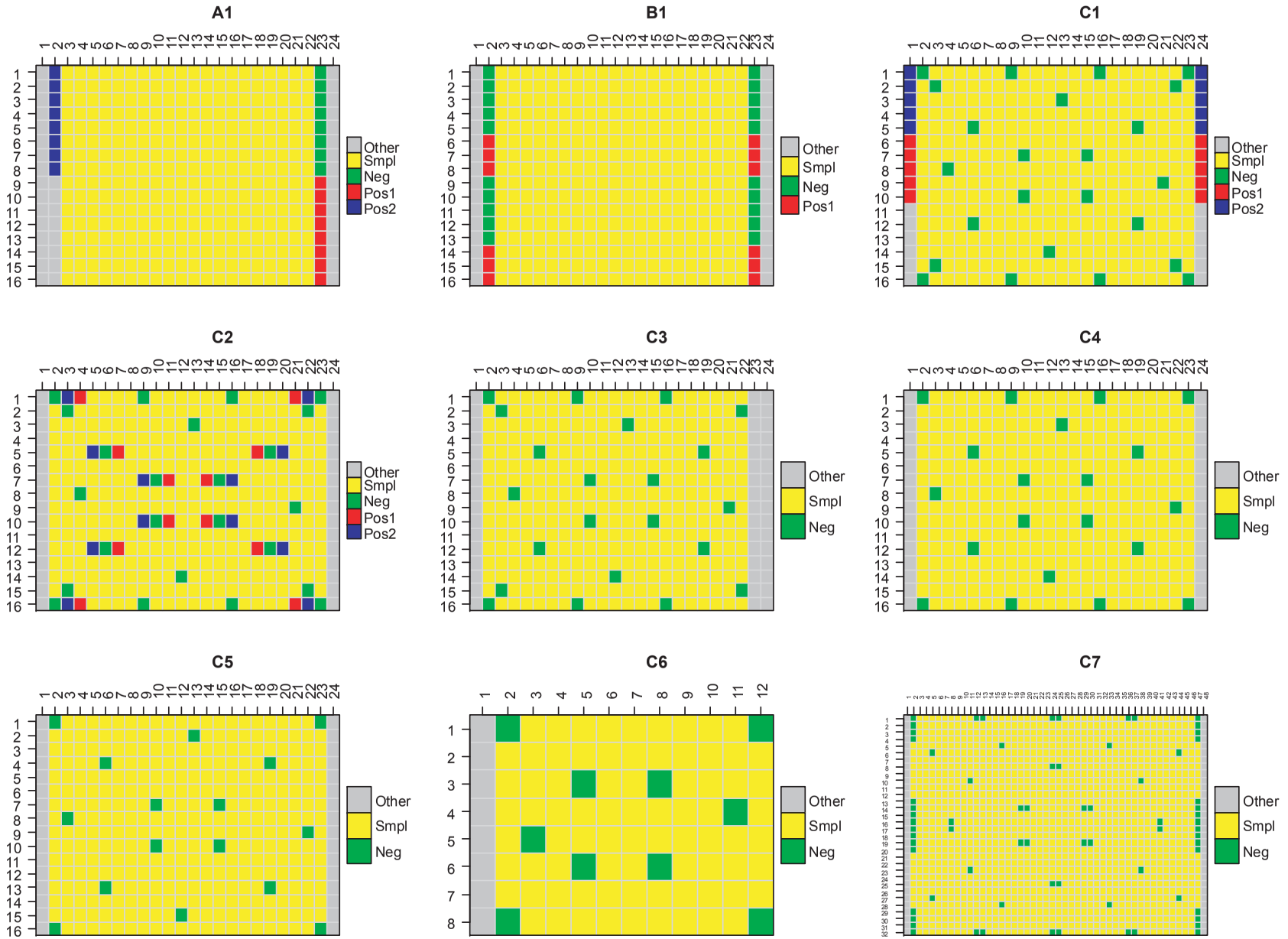

FIG. 2. Plate designs in a 384-well (panels **A1**, **B1**, and **C1–C5**), 96-well (panel **C6**), or 1536-well plate (panel **C7**). Designs A1 and B1 (shown in panels **A1** and **B1**, respectively) are for situations where controls can only be arranged in the 4 edge columns, and designs C1–C7 (shown in panels **C1–C7**, respectively) are for situations where controls are allowed to be arranged anywhere in a plate. The colors represent types of wells as shown in the legend to the right of each panel: green = negative control, red = 1st positive control, blue = 2nd positive control, yellow = sample siRNAs, gray = other controls.

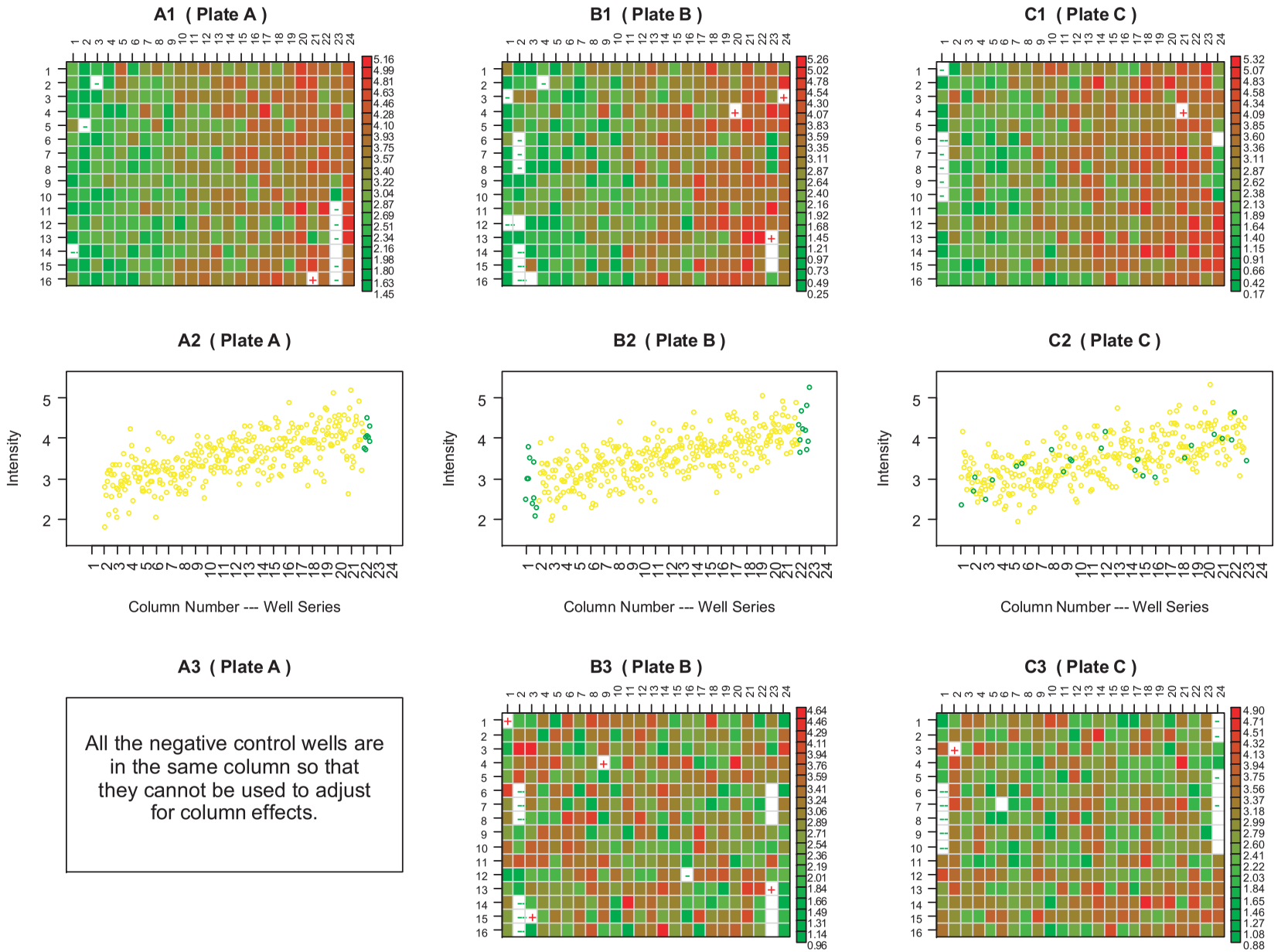

FIG. 3. Data with systematic errors having a linear relationship to column number in 3 plates extracted from RNA interference high-throughput screening experiments. Panels **A1**, **B1**, and **C1** show the measured intensities (in log10 scale) in all wells in a plate from 3 experiments that have plate designs A1, B1, and C1 (**Fig. 2**), respectively. Panels **A2**, **B2**, and **C2** display intensities using well-series plots. Panels **A3**, **B3**, and **C3** display the adjusted data using the negative control wells. In panels **A1**, **B1**, **B3**, **C1**, and **C3**, a green "–" in a white well indicates that the value in that well is lower than the value denoted by the strongest green color and a red "+" indicates that the value in the well is higher than the value denoted by the strongest red color. In panels **A2**, **B2**, and **C2**, the green and yellow points denote the intensities in the negative control and sample siRNA wells, respectively.

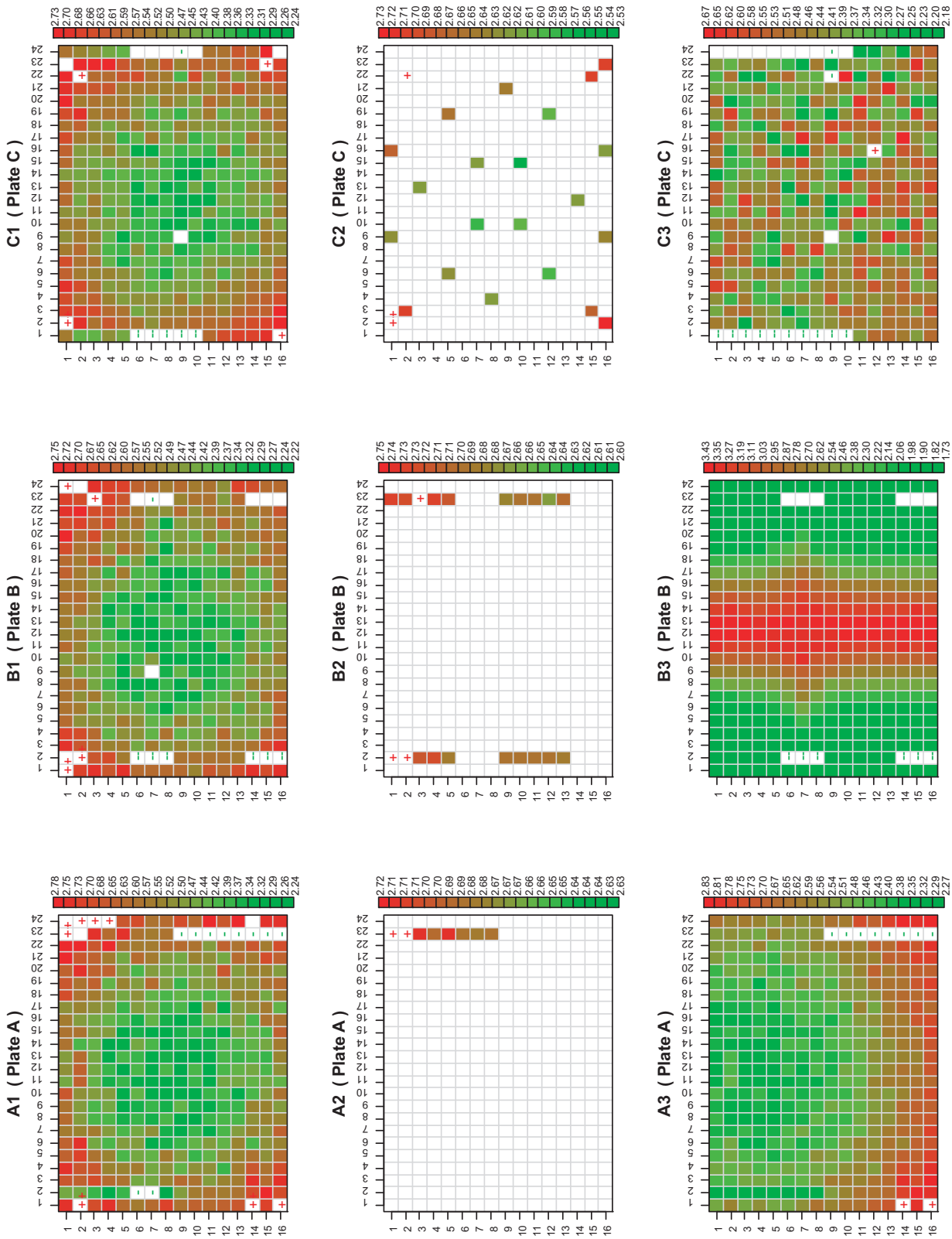

FIG. 4. Data with systematic errors having bowl-shaped relationship to well position in 3 plates extracted from RNA interference high-throughput screening experiments. Panels **A1**, **B1**, and **C1** show the measured intensities (in log10 scale) in all wells in a plate from 3 experiments that have plate designs A1, B1, and C1 (**Fig. 2**), respectively. Panels **A2**, **B2**, and **C2** display intensities of the negative control. Panels **A3**, **B3**, and **C3** display the data adjusted using the negative control wells. In each panel, a green "–" in a white well indicates that the value in that well is lower than the value denoted by the strongest green color and a red "+" indicates that the value in the well is higher than the value denoted by the strongest red color.

tematic errors with various patterns (panel **C2** of **Figs. 3** and **4**) and to remove/reduce impact of systematic errors (panel **C3** of **Figs. 3** and **4**).

Guidelines for adopting plate designs in HTS experiments

In primary screens, we may assume that the majority of sample siRNA wells have no effects. Thus, it is reasonable to use the majority of sample wells to adjust for systematic errors of measurement. However, even if we do so, plate designs with controls arranged only in edge columns cannot identify the plates with enriched hits. Considering these facts, I propose the following guideline for choosing plate designs in primary screens: if possible, adopt 1 of designs C1, C3–C5, and their variants especially in the situation where sample siRNAs are not randomly arranged in the screen; otherwise, adopt design B1 or a similar design. To adopt designs C1, C3–C5, or their variants in primary screens, it may be necessary to work with vendors of genomic libraries so that a control such as luciferase or polo-like kinase-1 can be arranged in the green wells of a plate during the process of generating a genomic library.

In confirmatory screens, it is infeasible to use sample wells to adjust for systematic errors because the sample siRNAs are preselected to have inhibition or activation effects. Considering this fact, I propose the following guideline for choosing plate designs in confirmatory screens: adopt 1 of designs C1–C5 and their variants whenever possible; adopt design B1 or a similar design, and avoid design A1 or a similar design, if the negative control cannot be arranged in the middle of a plate.

Choosing 1 design from designs C1–C5 or their variants relies on the ability to arrange positive control wells in the middle of a plate and the tradeoff between the ability of the chosen design to display systematic errors and the number of negative control wells. If the positive control cannot be arranged in the middle of a plate or if the arrangement may greatly affect the measured intensities of their neighbor wells, choose a design with positive controls only in the edge columns such as design C1; otherwise, choose a design with positive controls in the middle of a plate such as design C2. If more wells are needed in the edge for other types of controls, choose design C3 or similar designs. If more wells are needed for sample siRNAs, choose design C4, C5, or their variants.

The above plate designs and guidelines are constructed for experiments with a 384-well plate, which is summarized in **Table 4**. Similar plate designs and guidelines can be constructed for experiments with a 96-well or 1536-well plate. For example, design C6 is a design for a 96-well plate and design C7 is for a 1536-well plate. In a 1536-well plate, the intensity of a well with a strong effect is more likely to affect the measured intensity of its neighbor wells; thus in design C7, we arrange many negative control wells in the 2nd and 47th columns to establish buffering borders. When the impact of edge wells on the neighbor wells is strong, the negative control wells in these 2 columns are not used for smoothing but are used for the comparison with positive controls in the edge columns. For experiments conducted in 1536-well plates, only 1 primary screen with replicates may be needed with no follow-up confirmatory screens. In such a case, to adopt design C7 or similar designs, it may be necessary to work with vendors of genomic libraries to arrange a control such as luciferase in the green wells during the process of generating a genomic library.

Table 4. Guidelines for the Adoption of Plate Designs Conducted in 384-Well Plates

Guideline 1: for adopting plate designs in primary screens
- If possible, adopt 1 of designs C1, C3–C5, and their variants especially in the situation where sample siRNAs are not randomly arranged in the screen; otherwise, adopt design B1 or a similar design.
- To adopt designs C1, C3–C5, or their variants in primary screens, it may be necessary to work with vendors of genomic libraries so that a control such as luciferase or polo-like kinase-1 can be arranged in the middle of a plate during the process of generating a genomic library.

Guideline 2: for adopting plate designs in confirmatory screens
- Adopt 1 of designs C1–C5 and their variants whenever possible.
- Adopt design B1 or a similar design and avoid design A1 and similar designs in the situation where the negative control cannot be arranged in the middle of a plate.

Guideline 3: for making a choice among designs C1–C5
- If the positive control cannot be arranged in the middle of a plate or if the arrangement may greatly affect the measured intensities of their neighbor wells, choose a design with positive controls only in the edge columns such as design C1; otherwise, choose a design with positive controls in the middle of a plate such as design C2.
- If more wells are needed in the edge for other types of controls, choose design C3 or similar designs. If more wells are needed for sample siRNAs, choose design C4, C5, or their variants.

DISCUSSION

One common setback that scientists encounter is that they invest a vast amount of time and effort to derive conclusions from their experiments, but only to realize that these conclusions are not valid due to poor experimental design or poor data quality. This can be attributed to the lack of appropriate designs and effective metrics for QC in the very beginning. The plate designs proposed in this study display/adjust effectively for systematic errors. The novel SSMD-based QC criteria serve as a gatekeeper for good quality assays. SSMD combines data variabilities of a positive control and a negative reference in a theoretically more acceptable way than Z factor. The SSMD estimate and the relationship between SSMD and d^+-probability that are proved in this study provide a solid probabilistic and statistical basis for the SSMD-based criteria.

Although this study proposes the use of different criteria for controls with different effect sizes in RNAi HTS experiments especially in situations where there is no replicate for the majority of sample siRNAs in a plate, there exist disputes about

whether a single criterion or multiple criteria should be used. The adoption of multiple criteria for controls with different effect sizes (such as the SSMD-based QC criteria in **Table 1**) takes into account the fact that different positive controls may have different effect sizes; however, it makes it complicated to apply them in experiments in which the sizes of positive controls are unknown. To make the application easier, the strategies for using multiple criteria for different controls are proposed in **Table 3**.

A single criterion (such as the popularly used Z-factor criterion) is simple to apply in experiments; however, it cannot take into account the fact that different positive controls may have different effect sizes and lead to inconsistent QC results in experiments with 2 or more positive controls with different effect sizes. In RNAi HTS experiments with no replicate for sample siRNAs, it is important to take into account the sizes of positive controls because (1) a moderate or fairly strong positive control is usually more instructive and relevant to the hits of interest than a strong positive control in RNAi HTS assays, and (2) it has strongly been recommended that HTS assays incorporate as many controls as possible (http://nsrb.med.harvard.edu/assaydev.html) and different controls have different effect sizes. Controls in compound screens are straightforward whereas siRNA controls in RNAi screens are usually neither straightforward nor as strong as the positive controls in compound HTS. For example, some RNAi screens in which cell viability is measured use only the background wells as a positive control whereas others may use a weak positive control. Applying the same QC criterion to these 2 positive controls will lead to misleading QC results for detecting siRNAs with moderate or fairly strong effects: judging the screens with strong positive controls as good quality assays even if they have poor quality, and judging the screens with weak positive controls as poor quality assays even if they have good quality for detecting hits of interest that have moderate or fairly strong effects.

In addition, it is well known that data variability or noise comes from 2 sources: measurement variability and biological variability (c.f., Stone et al[28]). SSMD and Z factor take into account both the mean difference and the data variability. When an siRNA has weak effect, its biological variability is usually large and the ratio of mean difference to biological variability is small. In assays with no replicates, increasing assay quality reduces measurement variability but not biological variability and thus may not lead to a large value of SSMD or Z factor even in an excellent assay with tiny measurement variability.

On the other hand, if one strongly prefers the use of a single QC criterion instead of multiple QC criteria for positive controls with different effect sizes, criterion Id or IId can be adopted because criterion Id or IId has a stronger probabilistic and statistical base than the commonly used Z-factor criterion. Another reason for using criterion Id or IId is that $|SSMD| \geq 4.7$ leads to QC results more similar to QC results of using Z factor ≥ 0.5 than those of using Z factor > 0. This may have practical meanings because it has been empirically proposed that an assay should pass QC if Z factor ≥ 0.5, instead of Z factor > 0, in compound HTS experiments (http://nsrb.med.harvard.edu/assaydev.html). In addition, in the situation where it is feasible to arrange 2 or more replicates per sample siRNA (or compounds) in a plate, a sensible QC strategy is to use a single QC criterion such as criterion Id or IId for different positive controls so that we can search the number of replicates to reach a similar power for the positive control that have the weakest effect in each HTS experiment. Of course, this strategy may substantially increase the cost of a genome-scale screen.

Table 5. A Strategy to Implement both Analytic Criteria and Effective Plate Designs to Improve Data Quality in RNAi HTS Experiments

1. Whenever possible, adopt effective plate design in the very beginning of the experiment following the guidelines in **Table 2**.
2. Check for systematic errors of measurement using a negative reference;
   - If systematic errors exist, try to adjust for them
   - If obvious systematic errors exist and cannot be adjusted, QC results reached
   - using any analytic metric may not be reliable.
3. Adopt SSMD-based criteria for QC if systematic errors do not exist or have been adjusted;
4. Examine the plates failing QC to investigate potential causes.
   - Check whether the failure is only caused by the contaminated positive controls
     - If yes, the plate may be used for hit selection;
     - Otherwise, redo the plate when possible
   - Check whether the failure is only caused by 1 or 2 extreme outliers in the controls
     - If yes and if the plate passes QC after removing the outliers, the plate may be used for hit selection;
     - Otherwise, redo the plate when possible

Note: RNAi, RNA interference; HTS, high-throughput screening; QC, quality control; SSMD, strictly standardized mean difference.

The novel SSMD-based criteria and effective plate designs provide tools to achieve high quality RNAi HTS data. For example, if the systematic errors shown in **Figure 3** have not been identified and adjusted appropriately, hit selection will be dominated by the column effects; thus the majority of selected hits with low values come from left columns and the majority of selected hits with high values come from right columns. These selected hits are useless because they are selected based not on their siRNA effects but on their positions in a plate. Good plate designs (e.g., designs B1 and C1) help to display and adjust for the column effects (panels **B3** and **C3** of **Fig. 2**). Therefore, we should always plan our experimental designs (including plate designs) appropriately in the beginning and adopt effective analytic methods to avoid the unpleasant situation in high-throughput biotechnologies, as pointed out by John Quackenbush,[12] in which "people tend to go out blindly and do experiments, then go back and try to analyse them and figure out what the question is afterwards."

To implement both novel criteria and effective plate designs to improve data quality in RNAi HTS experiments, I suggest

the use of the strategy displayed in **Table 5**. Following this strategy, we may display systematic errors and then adjust for them in the situation where we can adopt effective plate designs in the very beginning. In the situation where we may not control the plate design, the strategy may still help us to identify plates with bad quality and redo them or exclude them for further data analysis. In either situation, following the strategy will help to obtain high quality data.

Rigorous statistical analysis may become the key to good science in high-throughput biotechnologies including microarrays.[12] In this study, I concentrate on developing analytic methods and plate designs based on a theoretical foundation as well as based on simulation studies with ideal conditions. To avoid generating another ad-hoc method, these methods have not been applied in any complete data sets. In a follow-up study,[32] Zhang and colleagues explore the implementation of the proposed SSMD-based QC methods, effective plate design, and the selection of effective controls in multiple real RNAi HTS experiments. The SSMD-based QC criteria and effective plate designs should be generally applicable to any assay where the endpoint is a difference in signal compared to a reference sample, including enzyme, receptor, and cellular function assays in addition to RNAi-based high-throughput screens. It may be noted that small molecule HTS experiments tend to have very strong or strong positive controls. Hence, usually criterion Id or IId (and occasionally criterion Ic or IIc) is used in small molecule HTS experiments.

In conclusion, effective analytic methods and appropriate plate designs are critical for obtaining high quality data in RNAi HTS assays. Based on both original and probabilistic meanings of SSMD, I develop novel SSMD-based criteria for QC from a solid probabilistic and statistical base. Considering the fact that robots can now readily arrange controls anywhere in a plate, plate designs with controls arranged across a plate are proposed to effectively display and adjust for systematic errors of measurement; and guidelines are proposed for using these designs in primary and confirmatory screens. The novel SSMD-based QC criteria and the proposed plate designs and guidelines greatly help in obtaining high quality data in genome-wide RNAi screens and other HTS screens.

ACKNOWLEDGMENTS

The author thanks Drs. Daniel Holder, Keith Soper, and Joseph Heyse for their support in this research, Pei Fen Kuan for proofreading, and Editor Campbell and two anonymous reviewers for their helpful comments.

APPENDIX
SSMD and its estimate

SSMD has recently been proposed for measuring the magnitude of difference between 2 populations.[23] Let random variables P_1 and P_2 denote 2 populations of interest. SSMD (denoted as β) is defined as the ratio of mean to standard deviation of the difference between the 2 populations, namely, $\beta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}}$, where μ_1, μ_2, σ_1^2, σ_2^2 are the means and variances in the 2 populations, respectively, and σ_{12} is the covariance between these 2 populations. If the 2 populations are independent, $\beta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$. If the 2 independent populations have equal variances (namely, $\sigma_1^2 = \sigma_2^2 = \sigma^2$), then $\beta = \frac{\mu_1 - \mu_2}{\sqrt{2\sigma^2}}$.

SSMD defined above is an unknown population parameter that needs to be estimated from observed samples. Suppose we have 1 sample (with sample size n_1, sample mean $\bar{X}_1$, and sample standard deviation s_1) from population P_1 and another independent sample (with sample size n_2, sample mean $\bar{X}_2$, and sample standard deviation s_2) from population P_2. Let $N = n_1 + n_2$. Zhang[23] derived the maximum-likelihood estimate and method-of-moment estimate of SSMD. The method-of-moment estimate of SSMD is $\hat{\beta} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2 + s_2^2}}$. When the 2 groups have equal variances, we can prove a better estimate of SSMD, the uniformly minimal variance unbiased estimate (UMVUE), as follows. It is well known that the following properties hold in the situation where P_1 and P_2 have normal distributions and equal variance: $\bar{X}_1$, $\bar{X}_2$, s_1^2, s_2^2 are all independent with each other; $\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{\sigma^2}$ is distributed to $\chi^2(N-2)$; and $(\bar{X}_1, \bar{X}_2, s_1^2, s_2^2)$ is a complete sufficient statistic of (μ_1, μ_2, σ^2). Based on these properties,

$$\mathrm{E}\left(\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2}{K}(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}\right) = \sqrt{\frac{K}{2}} \cdot \mathrm{E}(\bar{X}_1 - \bar{X}_2) \cdot \mathrm{E}\left(\frac{1}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}\right) = \frac{\mu_1 - \mu_2}{\sqrt{2\sigma^2}} = \beta,$$

where $K = 2 \cdot \left(\frac{\Gamma\left(\frac{N-2}{2}\right)}{\Gamma\left(\frac{N-3}{2}\right)}\right)^2 \approx N - 3.5$ and $n_1, n_2 \geq 2$.

Set $\hat{\beta} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2}{K}\left((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right)}}$. Then $\hat{\beta}$ is an unbiased estimate of β and is a function of complete sufficient statistic $(\bar{X}_1, \bar{X}_2, s_1^2, s_2^2)$; thus $\hat{\beta}$ is a UMVUE of β.

Relationship between SSMD and d^+-probability

The probability of the difference D being positive is called positive difference probability, which is denoted by d^+-probability.[23-25] A general relationship between SSMD and d^+-probability is d^+-probability = $\Pr(D > 0) = \Pr(Z_D > -\beta) = 1 - \Pr(Z_D \leq -\beta) = 1 - F_Z(-\beta)$, where Z_D is the standardized difference (i.e., $z_D = \frac{D - \mu_D}{\sigma_D}$) and $F_Z(\cdot)$ is the cumulative distribution function of Z_D.

If D has a unimodal distribution with finite variance $\sigma_D^2 > 0$, we have

$$\text{Link I:} \begin{cases} d^+\text{-probability} \geq 1 - \frac{4}{9\beta^2}, & \text{for } \beta \geq \sqrt{\frac{8}{3}} \\ d^+\text{-probability} \geq \frac{4}{3} - \frac{4}{3\beta^2}, & \text{for } 1 < \beta \leq \sqrt{\frac{8}{3}} \\ d^+\text{-probability} \leq \frac{4}{9\beta^2}, & \text{for } \beta \leq -\sqrt{\frac{8}{3}} \\ d^+\text{-probability} \leq \frac{4}{3\beta^2} - \frac{1}{3}, & \text{for } -1 \geq \beta \geq -\sqrt{\frac{8}{3}}. \end{cases}$$

If D has a symmetric unimodal distribution with finite variance $\sigma_D^2 > 0$, we have

$$\text{Link II: } \begin{cases} d^+\text{- probability} \geq 1 - \frac{2}{9\beta^2}, & \text{for } \beta \geq \sqrt{\frac{8}{3}} \\ d^+\text{- probability} \geq \frac{7}{6} - \frac{2}{3\beta^2}, & \text{for } 1 < \beta \leq \sqrt{\frac{8}{3}} \\ d^+\text{- probability} \leq \frac{2}{9\beta^2}, & \text{for } \beta \leq -\sqrt{\frac{8}{3}} \\ d^+\text{- probability} \leq \frac{2}{3\beta^2} - \frac{1}{6}, & \text{for } -1 \geq \beta \geq -\sqrt{\frac{8}{3}}. \end{cases}$$

If D has a normal distribution, we have Link III: d^+-probability = $\Phi(\beta)$ where $\Phi(\cdot)$ is a cumulative distribution function of the standard normal distribution.

Link III holds because d^+-probability = $\Pr(Z_D > -\beta) = \Pr(Z_D < -\beta) = \Phi(\beta)$ when D is normally distributed. To prove Links I and II, we need to use Vysochanskii-Petunin inequality[33]: For all $k > 0$, the following inequality holds for an arbitrary random variable X having a unimodal distribution and finite variance $\sigma^2 > 0$,

$$\begin{cases} \Pr(|X - \mu_X| \geq k\sigma) \leq \frac{4}{9k^2}, & \text{for } k \geq \sqrt{\frac{8}{3}} \\ \Pr(|X - \mu_X| \geq k\sigma) \leq \frac{4}{3k^2} - \frac{1}{3}, & \text{for } k \leq \sqrt{\frac{8}{3}}. \end{cases}$$

The well-known 3-sigma rule is a case of Vysochanskii-Petunin inequality with $k = 3$. Considering the value of any probability is between 0 and 1, we require $k \geq 1$.

When $\beta > 0$, $\Pr(D \leq 0) = \Pr(D - \mu_D \leq -\beta\sigma_D) \leq \Pr(|D - \mu_D| \geq \beta\sigma_D)$. In the situation where D has a unimodal distribution with finite variance $\sigma_D^2 > 0$, applying Vysochanskii-Petunin inequality with $k = \beta$ to variable D, we get

$$\begin{cases} \Pr(|D - \mu_D| \geq \beta\sigma_D) \leq \frac{4}{9\beta^2}, & \text{for } \beta \geq \sqrt{\frac{8}{3}} \\ \Pr(|D - \mu_D| \geq \beta\sigma_D) \leq \frac{4}{3\beta^2} - \frac{1}{3}, & \text{for } 1 \leq \beta \leq \sqrt{\frac{8}{3}}. \end{cases}$$

Considering d^+-probability = $\Pr(D > 0) = 1 - \Pr(D \leq 0)$, we then have

$$\begin{cases} d^+\text{- probability} \geq 1 - \frac{4}{9\beta^2}, & \text{for } \beta \geq \sqrt{\frac{8}{3}} \\ d^+\text{- probability} \geq \frac{4}{3} - \frac{4}{3\beta^2}, & \text{for } 1 < \beta \leq \sqrt{\frac{8}{3}}. \end{cases}$$

Similarly, when $\beta < 0$, $\Pr(D > 0) = \Pr(D - \mu_D > -\beta\sigma_D) \leq \Pr(|D - \mu_D| \geq (-\beta)\sigma_D)$. Applying Vysochanskii-Petunin inequality with $k = -\beta$ to variable D, we get

$$\begin{cases} d^+\text{- probability} \leq \frac{4}{9\beta^2}, & \text{for } \beta \leq -\sqrt{\frac{8}{3}} \\ d^+\text{- probability} \leq \frac{4}{3\beta^2} - \frac{1}{3}, & \text{for } -1 \geq \beta \geq -\sqrt{\frac{8}{3}}. \end{cases}$$

Thus, we prove Link I.

In the situation where D has a symmetric unimodal distribution with finite variance $\sigma_D^2 > 0$, $\Pr(D \leq 0) = \Pr(D - \mu_D \leq -\beta\sigma_D) = \frac{1}{2}\Pr(|D - \mu_D| \geq \beta\sigma_D)$ when $\beta > 0$, and $\Pr(D > 0) = \Pr(D - \mu_D > -\beta\sigma_D) = \frac{1}{2}\Pr(|D - \mu_D| > (-\beta)\sigma_D) \leq \frac{1}{2}\Pr(|D - \mu_D| \geq (-\beta)\sigma_D)$ when $\beta < 0$. Using the above relationships and Vysochanskii-Petunin inequality, we can readily prove Link II similarly in the way we proved Link I.

REFERENCES

1. Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE, Mello CC: Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 1998;391:806-811.
2. Hannon GJ, Zamore PD: Small RNAs, big biology: biochemical studies of RNA interference. In Hannon GJ (ed): *A Guide to Gene Silencing*. New York: Cold Spring Harbor Laboratory Press, 2003:87-108.
3. Bumcrot D, Manoharan M, Koteliansky V, Sah DWY: RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nat Chem Biol* 2006;2:711-719.
4. Mahanthappa N: Translating RNA interference into therapies for human disease. *Pharmacogenomics* 2005;6:879-883.
5. Bard F, Casano L, Mallabiabarrena A, Wallace E, Saito K, Kitayama H, et al: Functional genomics reveals genes involved in protein secretion and Golgi organization. *Nature* 2006;439:604-607.
6. Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, et al: A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 2004;428:431-437.
7. DasGupta R, Kaykas A, Moon RT, Perrimon N: Functional genomic analysis of the Wnt-Wingless signaling pathway. *Science* 2005;308:826-833.
8. Espeseth AS, Huang Q, Gates A, Xu M, Yu Y, Simon AJ, et al: A genome wide analysis of ubiquitin ligases in APP processing identifies a novel regulator of BACE1 mRNA levels. *Mol Cell Neurosci* 2006;33:227-235.
9. Majercak J, Ray WJ, Espeseth A, Simon A, Shi XP, Wolffe C, et al: LRRTM3 promotes processing of amyloid-precursor protein by BACE1 and is a positional candidate gene for late-onset Alzheimer's disease. *Proc Natl Acad Sci U S A* 2006;103:17967-17972.
10. Zhao HF, L'Abbe D, Jolicoeur N, Wu MQ, Li Z, Yu ZB, et al: High-throughput screening of effective siRNAs from RNAi libraries delivered via bacterial invasion. *Nat Methods* 2005;2:967-973.
11. Zuck P, Murray EM, Stec E, Grobler JA, Simon AJ, Strulovici B, et al: A cell-based beta-lactamase reporter gene assay for the identification of inhibitors of hepatitis C virus replication. *Anal Biochem* 2004;334:344-355.
12. Eisenstein M: Quality control. *Nature* 2006;442:1067-1070.
13. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R: Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006;24:167-175.
14. Zhang XD, Yang XC, Chung NJ, Gates A, Stec E, Kunapuli P, et al: Robust statistical methods for hit selection in RNA interference high-throughput screening experiments. *Pharmacogenomics* 2006;7:299-309.
15. Boutros M, Bras LP, Huber W: Analysis of cell-based RNAi screens. *Genome Biol* 2006;7:R66.
16. Park HJ, Partridge E, Cheung P, Pawling J, Donovan R, Wrana JL, et al: Chemical enhancers of cytokine signaling that suppress microfilament turnover and tumor cell growth. *Cancer Res* 2006;66:3558-3566.
17. Riester D, Wirsching F, Salinas G, Keller M, Gebinoga M, Kamphausen S, et al: Thrombin inhibitors identified by computer-assisted multiparameter design. *Proc Natl Acad Sci U S A* 2005;102:8597-8602.
18. Tian H, Cao LC, Tan YP, Williams S, Chen LL, Matray T, et al: Multiplex mRNA assay using electrophoretic tags for high-throughput gene expression analysis. *Nucleic Acids Res* 2004;32:e126.
19. Zhang JH, Chung TDY, Oldenburg KR: A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 1999;4:67-73.
20. Eastwood BJ, Farmen MW, Iversen PW, Craft TJ, Smallwood JK, Garbison KE, et al: The minimum significant ratio: a statistical parameter to characterize the reproducibility of potency estimates from concentration-response assays and estimation by replicate-experiment studies. *J Biomol Screen* 2006;11:253-261.
21. Iversen PW, Eastwood BJ, Sittampalam GS, Cox KL: A comparison of assay performance measures in screening assays: signal window, Z′ factor, and assay variability ratio. *J Biomol Screen* 2006;11:247-252.
22. Sui YX, Wu ZJ: Alternative statistical parameter for high-throughput screening assay quality assessment. *J Biomol Screen* 2007;12:229-234.
23. Zhang XD: A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics* 2007;89:552-561.

24. Zhang XD: A new method with flexible and balanced control of false negatives and false positives for hit selection in RNA interference high-throughput screening assays. *J Biomol Screen* 2007;12:645-655.
25. Zhang XD, Ferrer M, Espeseth AS, Marine SD, Stec EM, Crackower MA, et al: The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. *J Biomol Screen* 2007;12:497-509.
26. Brideau C, Gunter B, Pikounis B, Liaw A: Improved statistical methods for hit selection in high-throughput screening. *J Biomol Screen* 2003;8:634-647.
27. Gunter B, Brideau C, Pikounis B, Liaw A: Statistical and graphical methods for quality control determination of high-throughput screening data. *J Biomol Screen* 2003;8:624-633.
28. Stone DJ, Marine S, Majercak J, Ray WJ, Espeseth A, Simon A, et al: High-throughput screening by RNA interference—control of two distinct types of variance. *Cell Cycle* 2007;6:898-901.
29. Ramadan N, Flockhart I, Booker M, Perrimon N., Mathey-Prevot B: Design and implementation of high-throughput RNAi screens in cultured Drosophila cells. *Nature Protocol* 2007;2:2245-2263.
30. Harlow LL, Mulaik SA, Steiger JH: *What If There Were No Significance Tests?* Mahwah, NJ: Lawrence Erlbaum Associates, 1997.
31. Zhang XD: New concept of contrast in statistical analysis. In: *The 2007 American Statistical Association Proceedings*, meeting proceedings of the Joint Statistical Meetings, Salt Lake City, Utah, July 29 to August 2, 2007. Alexandria, VA: American Statistical Association, 2007:637-648.
32. Zhang XD, Espeseth A, Johnson E, Chin J, Gates A, Mitnaul L, et al: Integrating experimental and analytic approaches to improve data quality in genome-wide RNAi screens. *J Biomol Screen* 2008;378-389.
33. Vysochanskij DF, Petunin YI: Justification of the 3-sigma rule for unimodal distribution. *Theory of Probability and Mathematical Statistics* 1980;21:25-36.

Address correspondence to:
Xiaohua Douglas Zhang, Ph.D.
Biometrics Research, WP53B-120
Merck Research Laboratories
770 Sumneytown Pike
West Point, PA 19486

E-mail: xiaohua_zhang@merck.com