Journal of Biomolecular Screening
http://jbx.sagepub.com/

Illustration of SSMD, z Score, SSMD*, z* Score, and *t* Statistic for Hit Selection in RNAi High-Throughput Screens

Xiaohua Douglas Zhang

J Biomol Screen 2011 16: 775 originally published online 22 April 2011
DOI: 10.1177/1087057111405851

The online version of this article can be found at:
http://jbx.sagepub.com/content/16/7/775

Published by:
SAGE
http://www.sagepublications.com

Illustration of SSMD, z Score, SSMD*, z* Score, and *t* Statistic for Hit Selection in RNAi High-Throughput Screens

XIAOHUA DOUGLAS ZHANG

Hit selection is the ultimate goal in many high-throughput screens. Various analytic methods are available for this purpose. Some commonly used ones are z score, z* score, strictly standardized mean difference (SSMD), SSMD*, and *t* statistic. It is critical to know how to use them correctly because the misusage of them can readily produce misleading results. Here, the author presents basic concepts, elaborates their commonality and difference, describes some common misusage that people should avoid, and uses simulated simple examples to illustrate how to use them correctly. (*Journal of Biomolecular Screening* 2011;16:775-785)

Key words: SSMD, SSMD*, z score, z* score, *t* statistic, high-throughput screening

INTRODUCTION

THERE ARE MANY ANALYTIC METHODS FOR HIT SELECTION in high-throughput screens.[1] Among them, the commonly used ones are the z score, z* score, *t* statistic, strictly standardized mean difference (SSMD), and SSMD*.[2–10] They are suitable for hit selection in different situations. Thus, it is important to know how to use them for hit selection in high-throughput screens correctly. Otherwise, they may be misused. To avoid the misusage of these methods, in this article, I first present their concepts and calculations, elaborate their commonality and difference, and describe the situations where each metric is applicable. I use simulated examples to illustrate how to calculate and use them correctly for hit selection in RNAi high-throughput screens, to demonstrate the possible misusages of these metrics, and to show the misleading results and erroneous conclusions due to these misusages.

The analytic methods for hit selection in screens without replicates differ from those with replicates. For example, the z score method is suitable for screens without replicates, whereas the *t* statistic is suitable for screens with replicates. The calculation of SSMD for screens without replicates also differs from that for screens with replicates. Therefore, I discuss these methods separately: z score and SSMD for screens without replicates in the first section and *t* statistic and SSMD for screens with replicates in the second section.

Biometrics Research, Merck Research Laboratories, West Point, PA, USA.

Received Nov 29, 2010, and in revised form Jan 26, 2011. Accepted for publication Mar 9, 2011.

Journal of Biomolecular Screening 16(7); 2011
DOI: 10.1177/1087057111405851

Z SCORE AND SSMD IN PRIMARY SCREENS WITHOUT REPLICATES

z score and SSMD

The z score method is equivalent to the mean ± *k* SD method in which the siRNAs with measured value bigger than mean + *k* SD or smaller than mean – *k* SD are selected as hits, where *k* is a preset constant usually being 2 or 3. Note that SD represents the standard deviation. The z score method relies on the z score of the standard normal distribution $N(0,1)$ and hence the name. The formula for calculating the z score is shown in **Table 1**. *k* is also the cutoff (or critical value) of the z score. The choice of $k = 3$ comes from the well-known 3-sigma rule.[11]

SSMD is the mean of differences divided by the standard deviation of the differences between an siRNA and a negative reference.[10] When the data are preprocessed using log-transformation as we normally do, SSMD is the mean of log fold change divided by the standard deviation of log fold change with respect to a negative reference. In other words, SSMD is the average fold change (on the log scale) penalized by the variability of fold change (on the log scale). Thus, like z score and *t* statistic, SSMD captures both the mean and variability of siRNA (or small-molecule) effects. In a screen without replicates, we cannot directly calculate the variability of each siRNA. Thus, like z score, we have to assume that each siRNA has the same variability as a negative reference and then calculate the variability based on the negative reference and/or all the investigated siRNAs. On the basis of this assumption, we can calculate SSMD using the method-of-moment (MM) method or the uniformly minimal variance unbiased estimate (UMVUE) method[10] as shown in **Table 1**.

Table 1. Hit Selection Metrics and Their Calculation for Hit Selection in Primary Screens

Hit Selection Metric	*Calculation Formula*	
	Regular Version	*Robust Version*
Difference (i.e., fold change on the log scale)	Difference = $Y_i - \bar{Y}_N$	Difference* = $Y_i - \tilde{Y}_N$
z score	z score = $\frac{Y_i - \bar{Y}_N}{SD_N}$	z* score = $\frac{Y_i - \tilde{Y}_N}{MAD_N}$
Strictly standardized mean difference (SSMD) based on method-of-moment (MM) method	SSMD = $\frac{Y_i - \bar{Y}_N}{\sqrt{2}\ SD_N}$	SSMD* = $\frac{Y_i - \tilde{Y}_N}{\sqrt{2}\ MAD_N}$
SSMD based on uniformly minimal variance unbiased estimate (UMVUE)	SSMD = $\frac{Y_i - \bar{Y}_N}{\sqrt{\frac{2}{K}(n_N - 1)}\ SD_N}$, $K = 2 \cdot \left(\frac{\Gamma\left(\frac{n_N - 1}{2}\right)}{\Gamma\left(\frac{n_N - 2}{2}\right)} \right)^2 \approx n_N - 2.48$	SSMD* = $\frac{Y_i - \tilde{Y}_N}{\sqrt{\frac{2}{K}(n_N - 1)}\ MAD_N}$

Note: z score = $\sqrt{2}$ SSMD and z* score = $\sqrt{2}$ SSMD* when the MM method is used to estimate SSMD. Y_i denotes a measured activity value usually on the log scale of the ith investigated siRNA; n_N is the number of wells for a negative reference in a plate; $\bar{Y}_N$, $\tilde{Y}_N$, SD_N, and MAD_N denote the mean, median, standard deviation (SD), and median of absolute deviation (MAD) of measured values in a negative reference, respectively. The negative reference may consist of all sample wells in a plate. $\Gamma(\bullet)$ is a gamma function.

In screens without replicates, there is a linear relationship between z score and SSMD. That is, as shown in **Table 1**, z score = $\sqrt{2}$ SSMD when the MM method is used to estimate SSMD, and z score = $\sqrt{\frac{2}{K}(n_N - 1)}$ SSMD when the UMVUE method is used to estimate SSMD.[10] See **Table 1** for the definition of K and n_N. However, this relationship does not exist for a screen with replicates if the SSMD is calculated across replicates.

z score and SSMD**

In practice, the measured activity values are not normally distributed. Long-tailed or even skewed distributions may occur. Outliers appear frequently in high-throughput screening (HTS) data. True hits should behave differently from the siRNAs that do not have specific silencing effects. Because the majority of siRNAs in a primary HTS experiment do not have specific silencing effects, true hits (especially strong ones) should behave like outliers. To obtain estimates for the center and variability similar to mean and SD, but more robust to outliers and the violation of normal assumption, one common choice is median and MAD. MAD represents the median of absolute deviations, that is,

$$\text{MAD} = 1.4826\ \text{median}(|y_i - \text{median}(y)|).$$

The constant of 1.4826 is chosen so that MAD is equivalent to SD when the measured values are normally distributed. Similar to mean ± k SD, the method based on median and MAD is median ± k MAD, where k is often set to be 2 or 3. Similar to z score, we define *z* score* as in **Table 1**. Considering the robustness to outliers, similar to SSMD, we have SSMD* after replacing mean with median and replacing SD with MAD (**Table 1**).

In screens without replicates, the use of robust versions of estimators is strongly recommended because nearly all the analytic methods for screens without replicate have to be based on the strong assumption that all siRNAs in a plate or a screen have the same variability as a negative reference to estimate data variability. In many screens without replicates, all or most siRNAs are used to estimate mean, median, SD, and MAD. True hits with strong effects will strongly distort sample mean and SD but less on sample median and MAD. Consequently, SSMD* and z* score are more favorable than SSMD and z score for hit selection in primary screens without replicates.

Other considerations

One consideration in primary screens without replicates is whether z score, z* score, SSMD, and SSMD* should be calculated plate by plate (i.e., platewise) or on all plates (i.e., experimentwise). Considering the fact that large plate-to-plate variability commonly exists in HTS experiments, platewise calculation should be adopted in regular situations.[7,12] Another consideration is whether the sample wells or negative control wells should be used as the negative reference for the calculation

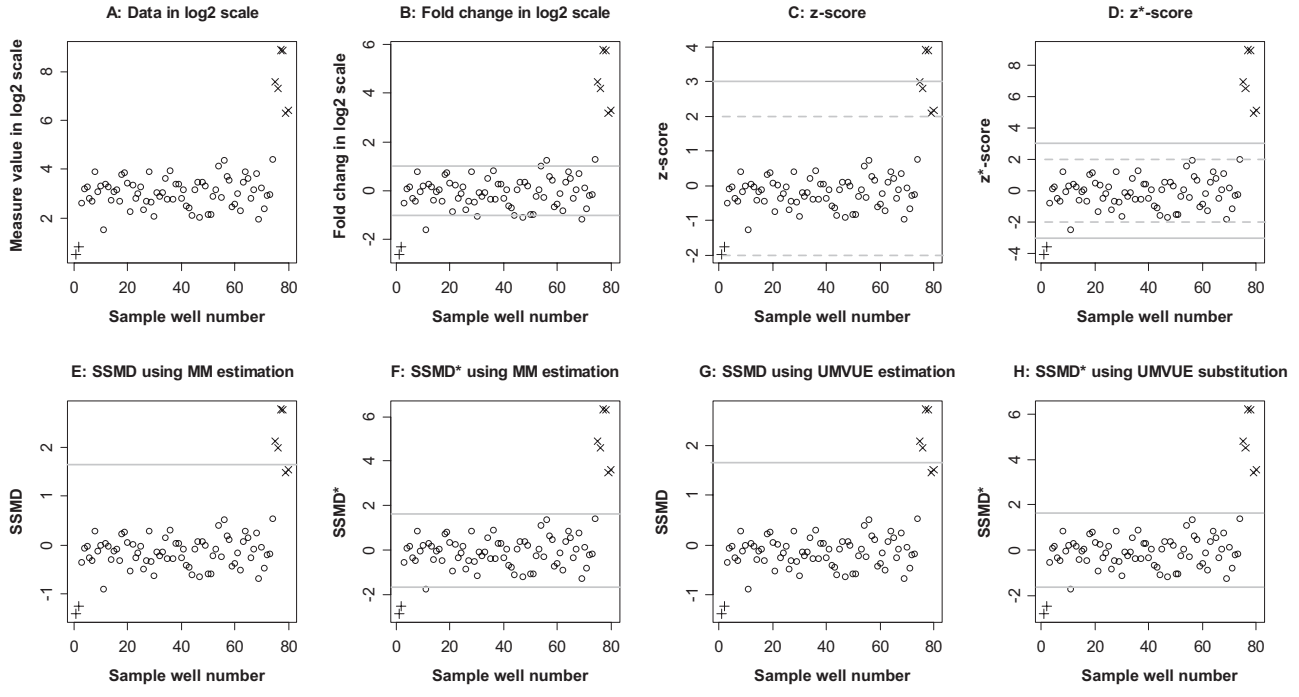

FIG. 1. A simulated data of sample wells in a 96-well plate: 2 siRNAs with strong downregulated effects ("+" points), 72 siRNAs with no or extremely effects (round points), and 6 siRNAs with very strong upregulated effects ("×" points). The solid gray lines denote $\log_2$-fold change = ±1 (i.e., twofold change in both directions) in panel **B**, z score = ±3 in panel **C**, z* score = ±3 in panel **D**, SSMD = ±1.645 in panels **E** and **G**, and SSMD* = ±1.645 in panels **F** and **H**, respectively. The dashed gray lines denote z score = ±2 in panel **C**, z* score = ±2 in panel **D**, respectively. SSMD, strictly standardized mean difference; UMVUE, uniformly minimal variance unbiased estimate.

of z score and SSMD. In most primary screens without replicates, the sample wells should be used as the negative reference.[7] This is because (1) most sample siRNAs have very small or no specific effects in the primary screens, (2) the number of wells for a negative control in a plate is usually not large, and (3) it is difficult to find reliable and effective negative controls in some RNAi HTS screens. On the other hand, if the siRNA are preselected to have certain effects or are arranged by certain biochemical features, it may be better to use the negative control wells as the negative reference.

Illustration of z score and SSMD* in primary screens without replicates*

To demonstrate the use of z score, z* score, SSMD, and SSMD* in primary screens without replicates, I simulate a 96-well plate that contains 2 siRNAs with strong downregulated effects ("+" points), 72 siRNAs with no or extremely weak effects (round points), and 6 with very strong upregulated effects ("×" points in **Fig. 1**). The data for all sample wells are listed in **Table 2**, which also lists the values of their mean, median, sample size, SD, and MAD. With these values, one can readily obtain the calculation of z score, z* score, SSMD, and SSMD* for each siRNA by plugging them into the formulas in **Table 1**. For example, siRNA 1 in **Table 1** has a measured value of 0.5. Its z* score and SSMD* are calculated as follows:

$$\text{z* score} = (0.5 - 3.11)/0.645 = -4.05,$$

$$\text{SSMD*} = \frac{0.5 - 3.11}{0.645\sqrt{\frac{2}{80-2.48}(80-1)}} = -2.84 \text{ based on the}$$

UMVUE estimate,

$$\text{SSMD*} = \frac{0.5 - 3.11}{0.645\sqrt{2}} = -2.86 \text{ based on the}$$

MM estimate.

The measured value on the $\log_2$ scale, calculated mean difference (i.e., average $\log_2$-fold change), z score, z* score, SSMD, and SSMD* are displayed in **Figure 1**. **Figure 1** reveals that the siRNAs with large effects behave like outliers as compared to the siRNAs with no or extremely weak effects; thus, they will inflate standard deviation and reduce the value of z score and SSMD in the regular (nonrobust) version when we use sample wells as a negative reference. Consequently, the use of regular z score and SSMD will miss the selection of some siRNAs with strong or

Table 2. Example Data Set for Demonstrating the Usage of Metrics for Hit Selection in One Plate in an RNAi Primary Screen without Replicates

Measured Value on the $\log_2$ Scale for Sample Wells	*Data Center*	*Variability*
0.50, 0.80, 2.61, 3.18, 3.28, 2.79, 2.68, 3.90, 3.08, 3.31,	Mean $\bar{Y}_N$: 3.321	SD_N: 1.419
1.52, 3.39, 3.28, 2.73, 3.07, 3.15, 2.69, 3.78, 3.85, 3.44,	Median $\tilde{Y}_N$: 3.110	MAD_N: 0.645
2.24, 3.36, 2.79, 3.00, 3.27, 2.35, 2.67, 3.88, 2.64, 2.08,	$n_N = 80$	
3.05, 2.88, 3.04, 3.64, 2.78, 3.92, 2.78, 3.40, 3.38, 2.80,		
3.14, 2.49, 2.40, 2.12, 3.16, 3.48, 2.01, 3.48, 3.32, 2.14,		
2.14, 2.87, 3.17, 4.13, 2.83, 4.38, 3.71, 3.56, 2.47, 2.56,		
2.98, 2.28, 3.48, 3.90, 3.62, 2.79, 3.17, 3.80, 1.96, 3.23,		
2.39, 2.91, 2.94, 4.41, 7.57, 7.29, 8.88, 8.86, 6.29, 6.42		

very strong effects (**Fig. 1C**,**E**,**G**). By contrast, the use of robust z* score and SSMD* will not have such an issue (**Fig. 1D**,**F**,**H**).

SSMD AND *T* STATISTICS IN SCREENS WITH REPLICATES

In general

In a confirmatory screen with replicate, the z score for an siRNA in a screen without replicates cannot incorporate the information across all its replicates correctly. Thus, people calculate *t* statistic or SSMD across the replicates. This is because the z score in the primary screen without replicates is based on a strong assumption that every siRNA has the same variability as a negative reference in the screens. The assumption cannot be held in most screens. However, without replicates, we cannot estimate data variability for an siRNA directly, and we have to estimate the variability based on this strong assumption in a screen without replicates. In a screen with replicates, we can directly estimate variability for each siRNA; consequently, we should use an SSMD or *t* statistic that does not rely on this strong assumption.

One issue with the use of the *t* statistic and associated *p* values is that they are affected by both sample size and effect size,[13] as also demonstrated in the formula for calculating the *t* statistic in the subsection "Paired SSMD and *t* statistic." People tend to think that a small *p* value indicates a large siRNA effect and that a large *p* value indicates a small or no siRNA effect. However, that is not true. An siRNA with a large *p* value may have a large effect. On the other hand, an siRNA with a small *p* value may have a small effect.[14] In RNAi screens, what we are really interested in is the size of siRNA effects. Thus, we need to separate effect size from the impact of sample size. The *t* statistic and associated *p* values come from testing for no mean difference and thus are not designed to measure the size of siRNA effects.

By contrast, SSMD directly assesses the size of siRNA effects.[15] SSMD has also been shown to be better than other commonly used effect sizes.[14] The population value of SSMD is comparable across experiments, and thus we can use the same cutoff for the population value of SSMD to measure the size of siRNA effects. A meaningful and interpretable SSMD-based criterion for classifying the size of siRNA effects is as follows: $|\text{SSMD}| \geq 5$ for extremely strong, $5 > |\text{SSMD}| \geq 3$ for very strong, $3 > |\text{SSMD}| \geq 2$ for strong, $2 > |\text{SSMD}| \geq 1.645$ for fairly strong, $1.645 > |\text{SSMD}| \geq 1.28$ for moderate, $1.28 > |\text{SSMD}| \geq 1$ for fairly moderate, $1 > |\text{SSMD}| \geq 0.75$ for fairly weak, $0.75 > |\text{SSMD}| \geq 0.5$ for weak, $0.5 > |\text{SSMD}| \geq 0.25$ for very weak, and $|\text{SSMD}| \leq 0.25$ for extremely weak effects.[15]

Both SSMD and the *t* statistic can be calculated on the paired and unpaired cases. For the paired case, a measured value for an siRNA is paired with a median value of a negative reference in the same plate. The mean and variability of the difference of all these pairs across all plates are used to calculate the *t* statistic and SSMD. For the unpaired case, all the measured values of an siRNA are formed as a group, and all the measured values of a negative reference in the whole screen are formed as another group. The means and variabilities of these two separate groups are used to calculate the *t* statistic and SSMD. Below, I describe the paired and unpaired cases separately.

Paired SSMD and t *statistic*

In confirmatory or primary HTS experiments with replicates, a *t* test for testing mean difference has been used for selecting hits. Because plate-to-plate variability is usually higher than within-plate variability, a paired *t* test is often used for hit selection. That is, for the *i*th siRNA with *n* replicates, we calculate the difference between the measured value (usually on the log scale) of the siRNA and the median value of a negative control in a plate, d_{ij} ($j = 1, \ldots, n$), then calculate the *t* statistic as $\frac{\bar{d}_i}{s_i/\sqrt{n}}$ (**Table 3A**), where $\bar{d}_i$ and s_i are the sample mean and

Table 3. Hit Selection Metrics and Their Calculation for Hit Selection in Confirmatory or Primary Screens with Replicates

A: Paired (or Matched) Method

Metrics	*Calculation Formula*	*Notation*
A1: Average difference	$\bar{d}_i$	$\bar{d}_i$ and s_i are respectively the sample mean and standard deviation of d_{ij}s where d_{ij} is the difference between the measured activity value (usually on the log scale) of the ith siRNA and the median value of a negative control in the jth plate. $\Gamma(\bullet)$ is a gamma function. s_0^2 is an adjustment factor. w_i and w_0 are weights with the constraint of $w_i + w_0 = 1$. A simple choice is to set $w_i = w_0 = 1/2$ and let s_0^2 be the median of all s_i^2s. More sophisticated estimates of w_i, w_0, and s_0^2 are also available.[18,19] MM: method of moment UMVUE: uniformly minimal variance unbiased estimate
A2: UMVUE estimate of SSMD	$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\sqrt{\frac{2}{n-1}}\frac{\bar{d}_i}{s_i}$	
A3: MM estimate of SSMD	$\frac{\bar{d}_i}{s_i}$	
A4: *t* statistic	$\frac{\bar{d}_i}{s_i/\sqrt{n}}$	
A5: UMVUE estimate of SSMD	$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\sqrt{\frac{2}{n-1}}\frac{\bar{d}_i}{\sqrt{w_i s_i^2 + w_0 s_0^2}}$	
A6: MM estimate of SSMD	$\frac{\bar{d}_i}{\sqrt{w_i s_i^2 + w_0 s_0^2}}$	
A7: *t* statistic	$\frac{\bar{d}_i}{\sqrt{(w_i s_i^2 + w_0 s_0^2)/n}}$	

B: Unpaired (or Unmatched) Method

Metrics	Calculation Formula	Notation
B1: Difference of means	$\bar{X}_i - \bar{X}_N$	n_i, $\bar{X}_i$, and s_i^2 are the number of replicates, sample mean, and variance of measured (or normalized) values in all replicates of the ith siRNA, respectively. n_N, $\bar{X}_N$, and s_N^2 are the number of wells, sample mean, and variance of the negative control wells of all plates, respectively. $K = 2\cdot\left(\frac{\Gamma\left(\frac{n_i+n_N-2}{2}\right)}{\Gamma\left(\frac{n_i+n_N-3}{2}\right)}\right)^2 \approx n_i + n_N - 3.48$
B2: SSMD estimate under unequal variance	$\frac{\bar{X}_i - \bar{X}_N}{\sqrt{s_i^2 + s_N^2}}$	
B3: *t* statistic under unequal variance	$\frac{\bar{X}_i - \bar{X}_N}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_N^2}{n_N}}}$	
B4: SSMD estimate under equal variance	$\frac{\bar{X}_i - \bar{X}_N}{\sqrt{\frac{2}{K}\left((n_i-1)s_i^2 + (n_N-1)s_N^2\right)}}$	
B5: *t* statistic under equal variance	$\frac{(\bar{X}_i - \bar{X}_N)}{\sqrt{\frac{2}{n_i+n_N-2}\left((n_i-1)s_i^2 + (n_N-1)s_N^2\right)\left(\frac{1}{n_i}+\frac{1}{n_N}\right)}}$	

SSMD, strictly standardized mean difference; MM, method-of-moment method; UMVUE, uniformly minimal variance unbiased estimate.

standard deviation of all the paired differences across all plates, respectively. We also calculate its corresponding *p* value based on the central *t* distribution (with $n-1$ degrees of freedom) that the above *t* statistic has under the null hypothesis of zero mean difference. The *t* statistic is affected by both sample size *n* and effect size (i.e., the ratio of population mean and standard deviation of the difference, estimated simply as $\frac{\bar{d}_i}{s_i}$).

Note that when one calculates the *p* value using the assumption that $\frac{\bar{d}_i}{s_i/\sqrt{n}}$ has a standard normal distribution, instead of a *t* distribution, the *t* statistic becomes a z score incorporating information across all replicates. Clearly, this z score differs from the z score described previously for primary screens without replicates. $\frac{\bar{d}_i}{s_i/\sqrt{n}}$ approximately has a standard normal distribution only when the number of replicates for each siRNA is very large (usually at least 20). Nearly no screens will afford such a large number of replicates. More important, theoretically $\frac{\bar{d}_i}{s_i/\sqrt{n}}$ should have a *t* distribution rather than a standard normal distribution, especially when the sample size is small. Thus, this z score is rarely used in HTS screens with replicates.

Similarly, as the paired *t* statistic, we can estimate SSMD based on paired (or matched) differences between an siRNA and the negative reference in a plate as follows. In a screen with *n* replicates, a simple estimate of SSMD is the MM estimate, $\text{SSMD} = \frac{\bar{d}_i}{s_i}$ for the *i*th siRNA. A better estimate, the UMVUE, of SSMD[10] is $\text{SSMD} = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\sqrt{\frac{2}{n-1}}\frac{\bar{d}_i}{s_i}$. The formulas for estimating SSMD using the paired method are shown in **Table 3A**.

For the paired *t* statistic or SSMD, the standard deviation is calculated based on the replicates of each individual siRNA. The number of replicates in an RNAi screen is usually small (typically 3–6). When only a small sample size is used for estimating each standard deviation, these estimates become highly variable. Considering that many siRNAs may have similar standard deviations, one may incorporate the information across siRNAs in the calculation of standard deviation for an individual siRNA to obtain a more stable estimate. A simple way to do so is to add a constant to the estimate of standard deviation.[16,17] A more common approach is to use a linear combination of a common factor and the individual sample variance—that is, $\sqrt{w_i s_i^2 + w_0 s_0^2}$, where s_0^2 is a common factor and $w_i + w_0 = 1$ (cf. Smyth[18]). A simple choice is to let s_0^2 be the median of all s_i^2s and to set $w_i = w_0 = \frac{1}{2}$. To derive w_i, w_0, and s_0^2 more sophisticatedly, Smyth[18] used a Bayesian method and Wright and Simon[19] used an inverse-gamma distribution. Malo et al.[12] explored the moderated *t* statistic corresponding to Wright and Simon's approach (termed RVM *t* statistic) for selecting hits in HTS experiments. We can use the similar idea to estimate SSMD by replacing s_i with $\sqrt{w_i s_i^2 + w_0 s_0^2}$, resulting in formulas A5 and A6 in **Table 3**.

Illustration of paired SSMD and t *statistic*

To demonstrate the use of the *t* statistic and SSMD in confirmatory screens with replicates, I simulate data for four replicate 384-well plates with focus on three siRNAs that have very strong upregulation, no effect, and moderate downregulation, respectively. Consequently, each siRNA has four replicates (i.e., $n = 4$). The data for the three siRNAs and negative control wells in the four plates are shown in **Table 4** and displayed in **Figure 2A**. The distributions of the simulated data for the negative control, siRNAs 1 to 3, are normal distributions $N(2.2, 0.55^2)$, $N(4.5, 0.5^2)$, $N(2.2, 0.7^2)$, and $N(1, 0.6^2)$ in plates 1 and 4; $N(1.2, 0.55^2)$, $N(3.5, 0.5^2)$, $N(1.2, 0.7^2)$, and $N(0, 0.6^2)$ in plate 2; and $N(5.2, 0.55^2)$, $N(7.5, 0.5^2)$, $N(5.2, 0.7^2)$, and $N(4, 0.6^2)$ in plate 3, respectively. Based on the means and standard deviations of these distributions, the population values (or true values) of SSMD are 3.094, 0, and −1.474 for the three siRNAs, respectively (**Tables 4** and **5B**).

Considering that the variability among plates is usually large, we should first calculate the difference between the measured value of an siRNA and the median of a negative control in each plate (as shown in **Table 4** and **Fig. 2B**) and then calculate average difference, SSMD, and/or *p* value from classical paired *t* test using the formula in **Table 3A**. For example, for siRNA 1 in **Table 5A**, the estimated SSMD is calculated as follows:

$$\bar{d}_1 = (2.73 + 1.80 + 2.055 + 2.015)/4 = 2.15$$

$$s_1 = \sqrt{\frac{(2.73-2.15)^2 + (1.80-2.15)^2 + (2.055-2.15)^2 + (2.015-2.15)^2}{4-1}} = 0.4025543.$$

$$t\,\text{statistic}_1 = \frac{\bar{d}_1}{s_1/\sqrt{4}} = \frac{2.15}{0.4025543/2} = 10.68.$$

$$\text{SSMD}_1 = \frac{\Gamma\left(\frac{4-1}{2}\right)}{\Gamma\left(\frac{4-2}{2}\right)}\sqrt{\frac{2}{4-1}}\frac{\bar{d}_1}{s_1} = \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma(1)}\sqrt{\frac{2}{3}}\frac{2.15}{0.4025543} = 3.865.$$

Similarly, we can obtain the results for other siRNAs. The results are shown in the rows marked with "Paired" in **Table 5A**. The results correctly indicate that there are significant mean differences for siRNA 1 and siRNA 3 but no significant mean difference for siRNA 2 at the significant level of 0.05. The estimated SSMD values (**Table 5A**) are also close to their true values (**Table 5B**) for the three siRNAs, and they correctly indicate that siRNAs 1, 2, and 3 have a very strong upregulated

Table 4. Example Data Set for Demonstrating the Usage of Metrics for Hit Selection in RNAi Confirmatory Screens with Replicates

	Sample Wells				
	siRNA 1	*siRNA 2*	*siRNA 3*	. . .	*Negative Control Wells (Raw Value on the Log2 Scale)*
Plate 1					
Raw value	4.85,	1.73,	1.40,	. . .	1.97, 2.45, 2.79, 3.06, 2.18, 1.45, 1.78, 1.61, 2.06, 2.57, 2.31, 1.42, 1.75, 1.79, 2.38, 2.49
Difference	2.73,	−0.39,	−0.72,	. . .	
Plate 2					
Raw value	7.18,	6.22,	3.54,	. . .	5.27, 5.33, 5.03, 5.48, 5.53, 4.82, 5.98, 6.12, 5.43, 5.10, 6.09, 5.27, 5.90, 4.88, 5.55, 4.79
Difference	1.80,	0.84,	−1.84,	. . .	
Plate 3					
Raw value	3.49,	0.77,	0.27,	. . .	0.75, 0.86, 2.05, 2.08, 1.73, 0.93, 2.23, 1.36, 0.59, 1.45, 1.93, 1.43, 0.71, 1.82, 1.44, 1.28
Difference	2.055,	−0.665,	−1.165,	. . .	
Plate 4					
Raw value	4.33,	2.43,	0.10,	. . .	2.97, 2.52, 1.45, 2.27, 1.90, 2.31, 3.64, 2.98, 1.89, 2.32, 2.03, 2.41, 1.65, 2.57, 2.91, 2.05
Difference	2.015,	0.115,	−2.215,	. . .	

effect, nearly zero effect, and a moderate downregulated effect, respectively.

Unpaired SSMD and t *statistic*

When there is no data variability between plates in a confirmatory or primary HTS experiment with replicates, one may also consider calculating SSMD and the *t* statistic using an unpaired (or unmatched) method. In this method, we pool all the replicates of an siRNA across plates to form one group and pool all the measured values of the negative control across plates as another group. Then we estimate SSMD for the magnitude of difference and apply a regular unpaired *t* test of testing for no mean difference between these two groups.[20]

The formula for calculating SSMD and the *t* statistic depends on whether each siRNA has the same variance as the negative control. Formulas B4 and B5 in **Table 3B** are respectively for calculating SSMD and the *t* statistic under the condition of equal variance (namely, the assumption that each siRNA has the same variance as the negative control), whereas formulas B2 and B3 are for calculating SSMD and the *t* statistic without such an assumption.

Possible usage of unpaired SSMD and t *statistic*

The unpaired SSMD and *t* statistic can only be applied to data with very small or no variability between plates. If there is large variability between plates in a screen, we need to normalize the data to remove it first and then apply the unpaired SSMD and/or *t* statistic to the normalized data. In fact, one mistake that has commonly been made for hit selection in screens with replicates is to directly apply the unpaired *t* statistic and/or SSMD to data with large between-plate variability without suitable normalization. If one does so, the estimated variability will be inflated, the *p* value will get larger, and SSMD will be underestimated.

Using siRNA 1 in **Figure 2** as an illustration, one may mistakenly pool the four replicates of siRNA 1 (the 4 separated black bars in **Fig. 2A**) together to form one group (the 4 closely located black bars in **Fig. 2C**), pool all gray bars in the four plates together to form another group (the 64 gray bars in **Fig. 2C**), and then calculate the regular *t* statistic, associated *p* value, and SSMD based on the black bars and gray bars in **Figure 2C** using the formulas in **Table 3B**. This will lead to a *p* value of 0.07 and an estimated SSMD of 0.921 (**Table 5C**), which suggests that there is no significant mean difference between siRNA 1 and the negative control and that siRNA 1 has only fairly weak upregulated effects. This result is clearly misleading. Similar misleading results can be reached for siRNA 3.

Figure 2A reveals large variability among the four plates in the screen. To apply the unpaired SSMD and *t* statistic, we need to normalize the data first. The simple normalized methods include

1. Platewise difference: each measured value minus the median of the negative reference wells in the same plate
2. Platewise z* score: each value minus the median and then is divided by the MAD of the negative reference wells in the same plate
3. Platewise percent activation/inhibition: each value minus the median of the negative reference wells and then is divided by the difference between the medians of a positive control and the negative reference in the same plate

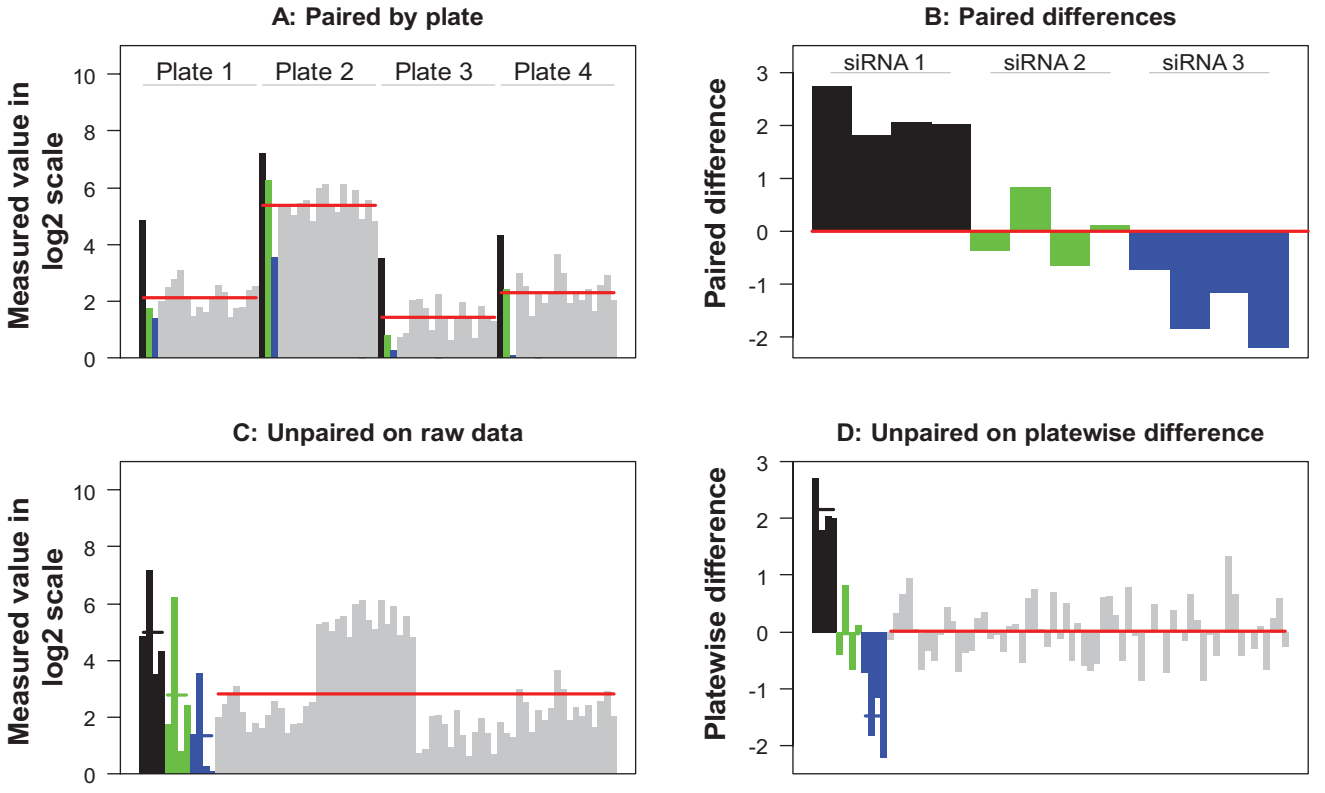

FIG. 2. A simulated confirmatory screen of four plates with focus on three siRNAs, siRNA 1 (black bars), siRNA 2 (green bars), and siRNA 3 (blue bars) and a negative control (gray bars). (**A**) Data arranged by plates, (**B**) paired (or matched) difference arranged by siRNAs, (**C**) comparison of raw data using unpaired method (i.e., comparing the pooled wells for an siRNA [as a group] with the pooled wells for the negative control [as another group]), and (**D**) comparison of platewise difference using unpaired method. In each panel, a red horizontal line denotes the baseline.

Table 5. Results for the Example in RNAi Confirmatory Screens with Replicates

Hit Selection Metrics		*siRNA 1*	*siRNA 2*	*siRNA 3*
A: Paired	Regular *p* value	0.0018	0.9445	0.021
	Estimated SSMD value	3.865	−0.027	−1.604
B: True SSMD value		3.094	0	−1.474
C: Unpaired on raw data	Regular *p* value	0.07	0.97	0.15
	Estimated SSMD value	0.921	−0.018	−0.664
D: Unpaired on difference	Regular *p* value	0.0009	0.906	0.019
	Estimated SSMD value	3.041	−0.060	−2.088

SSMD, strictly standardized mean difference.

For example, if we apply the unpaired SSMD and *t* statistic to the normalized data using platewise difference shown in the second row in each plate in **Table 4**, we will conduct the calculation based on **Figure 2D** instead of **Figure 2C**. Consequently, we will obtain results listed in **Table 5D**, which are close to those obtained using the paired SSMD and *t* statistic. The estimated SSMD values are also close to their true value and thus are fairly reasonable.

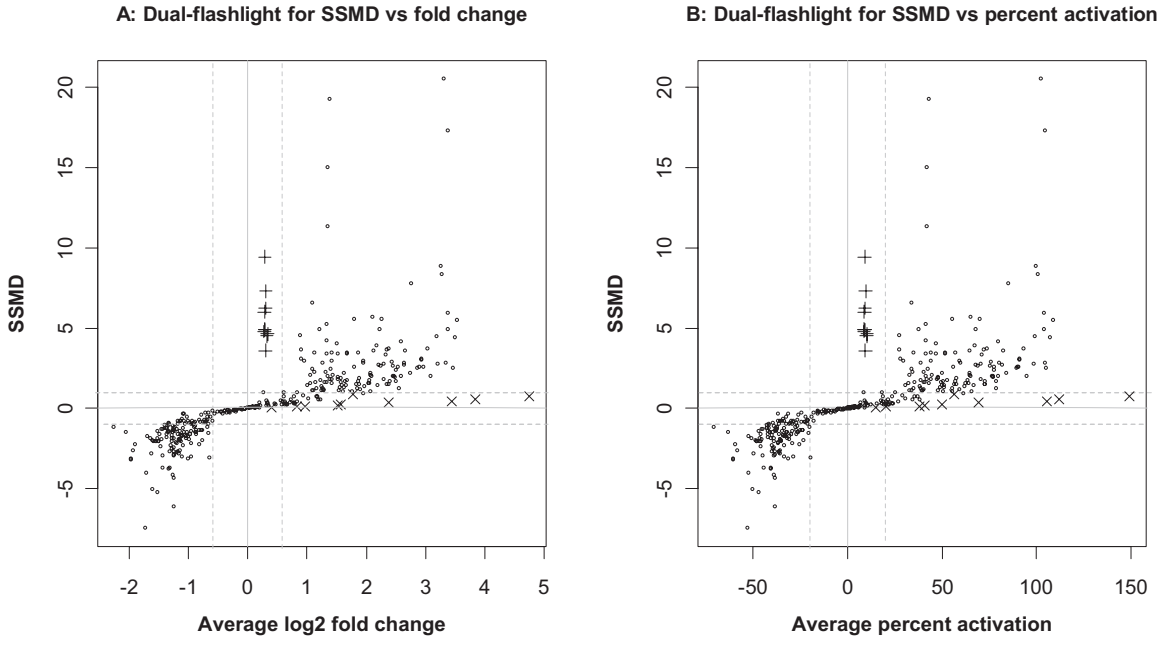

FIG. 3. Dual-flashlight plots for strictly standardized mean difference (SSMD) versus average fold change (**A**) and for SSMD versus percent activation (**B**). In either panel, the 10 "+" points come from a distribution with consistently very small fold change or percent activation; the 10 "×" points come from a distribution with both large mean fold change and large variability; the dashed gray horizontal lines denote SSMD = ±1. The vertical lines denote average fold change = 1.5 for upregulation and 1/1.5 for downregulation in panel **A** and average percent activation = ±20 in panel **B**.

Dual-flashlight plot

Similar to the classic *t* statistic, SSMD may overemphasize the siRNAs with very consistent but weak effects, as displayed in **Figure 3**. Considering that the mean deviation may have to have a reasonable size, we may use dual-flashlight plot[17] to overcome this issue. In a dual-flashlight plot, the *x*-axis represents a biologically more favored measure such as average fold change or average percent inhibition/activation, and the *y*-axis represents a statistically more favored measure such as SSMD, z score, or their variants. Therefore, using a dual-flashlight plot, scientists can select hits by simultaneously considering both biologically favored and statistically favored measures. The cutoff of average fold change or average percent inhibition/activation for hit selection is determined by the scientists who conduct the screens based on their biological knowledge. Meanwhile, the cutoff of estimated SSMD can be determined using false discovery and false nondiscovery rates as demonstrated by Zhang and colleagues.[21,22] The exact cutoff of estimated SSMD is experiment specific. However, in general, a cutoff between 1 and 1.645 (or between −1.645 and −1) can reach reasonably small false discovery and false nondiscovery rates in a screen with three or four replicates per siRNA.

Figure 3 demonstrates dual-flashlight plots for the simulated data partially shown in **Table 4** and **Figure 2**. Based on these dual-flashlight plots, if we consider both average fold change (or average percent activation) and SSMD and use the criterion of average fold change ≥1.5 (or average percent activation ≥20) and SSMD ≥1 for upregulation and average fold change ≤1/1.5 (or average percent activation ≤–20) and SSMD ≤–1 for downregulation, we should be able to exclude all the "+" points with consistently very small average fold change and all the "×" points with both large average fold change and large variability (i.e., very small SSMD).

Screens with duplicates only

For screens with duplicates only, there is not enough statistical power to detect siRNAs even with large effects.[23] For example, in the situation where we only have plates 1 and 2, the *p* values of testing for no mean difference for siRNAs 1, 2, and 3 are 0.129, 0.777 and 0.263, respectively, even if we use the paired *t* test. These *p* values would suggest that each of the three siRNAs does not have a significant mean difference from the negative control, which is not helpful. The UMVUE estimate does not exist for an siRNA with duplicates in the matched cases. Even worse is that people may mistakenly calculate SSMD and/or the *p* value for siRNAs with duplicates directly based on raw data similarly as in **Figure 2C**. If one does so, the *p* values for siRNAs 1, 2, and 3 are 0.289, 0.942, and 0.428, respectively;

the SSMDs for these three siRNAs are estimated to be 0.899, 0.079, and –0.521. All these results are biased and useless. As demonstrated in Zhang and Heyse,[23] for using SSMD in screens with replicates, it would be better to have four to seven replicates per siRNAs, and we need at least three replicates per siRNA.

On the other hand, the following three options may give some help for hit selection in a confirmatory screen with duplicates per siRNA: (1) use the MM estimate of paired SSMD as an approximation, (2) use unpaired SSMD based on normalized data such as platewise difference or platewise percent activation/inhibition, and (3) first calculate SSMD* or z* score (in **Table 1**) using the negative control as a negative reference in either plate and then count the frequency that the SSMD* or z* score is above a threshold for each siRNA in the two plates.

Using plates 1 and 2 as an example, in option 1, the MM estimated values of paired SSMD for siRNAs 1, 2, and 3 are 3.444, 0.259, and –1.616, respectively. The UMVUE estimated values of unpaired SSMD based on platewise differences for siRNAs 1, 2, and 3 are 3.161, 0.288, and –1.814, respectively. That is, siRNAs 1, 2, and 3 are assessed to have very strong upregulated effects, very weak upregulated effects, and moderate (by option 1) or fairly strong (by option 2) downregulated effects, respectively. Using option 3, the SSMD* values for siRNAs 1, 2, and 3 are 5.187, –0.741, and –1.368, respectively, in plate 1 and 3.854, 1.799, and –3.940, respectively, in plate 2. Thus, in option 3, siRNA 1 is assessed to have at least very strong upregulated effects in both plates, siRNA 2 is assessed to have fairly strong upregulated effects in plate 1 and weak downregulated effects in plate 2, and siRNA 3 is assessed to have at least moderate downregulated effects in both plates. Therefore, option 1 works the best in this example. On the other hand, this observation cannot be generalized because the estimation of a statistical parameter based on only two values is not reliable. As a reminder, the above methods provide possible analysis options for hit selection with duplicates that need to be assessed with care, but they cannot replace the need for a larger sample size when designing a screen.

GENERAL ANALYTIC STRATEGIES

The major role of z score and z* score is for hit selection in the primary screen without replicates, although they may also be used for data normalization. Because the plate-to-plate variability is usually large in high-throughput screens, z score and z* score should be calculated platewise. Because many true hits behave like outliers, z* score should generate more robust results and thus is generally better than z score. The *t* statistic for testing for no mean difference can be used for hit selection in screens with replicates. The *t* statistic is applicable in both the condition of equal variance and the condition of unequal variance, whereas z score and z* score are applicable only in the condition of equal variance. Because plate-to-plate variability is usually large in high-throughput screens, the paired *t* statistic is more appropriate than the unpaired *t* statistic for hit selection in high-throughput screens with replicates.

SSMD is applicable not only to screens with replicates but also to screens without replicates. For a screen without replicates, there is a simple linear relationship between z score and SSMD as well as between z* score and SSMD*, as shown in **Table 1**. For hit selection in a screen with replicates, similar to *t* statistic, paired SSMD (formula A2, A3, A5, or A6 in **Table 3A**) is more appropriate than unpaired SSMD (formula B2 or B4 in **Table 3B**). Because the variability for each siRNA or the control differs from each other, the SSMD in the condition of unequal variance (formula B2 in **Table 3B**) is more appropriate than that in the condition of equal variance (formula B4 in **Table 3B**) even if one wants to use unpaired SSMD for hit selection in a screen with replicates.

Sometimes, it may be easy to make a mistake by applying the formula of unpaired SSMD in the condition of equal variance (i.e., formula B4 in **Table 3B**) for a screen with duplicates. However, because the data variability in many siRNAs differs from the negative control in a screen, formula B2 in **Table 3B** may be used even if one wants to use unpaired SSMD. Otherwise, misleading results may be produced similarly as shown in Figure 4 in Barrows et al.[24] Similarly, the paired *t* statistic (or at least the unpaired *t* statistic under the condition of unequal variance) should be used if one wants to use the classical *t* test of testing for no mean difference for hit selection in high-throughput screens. It has been wrongly stated that z score = −3 is equivalent to SSMD = −3 (cf. Barrows et al.[24]), whereas actually z score = −3 is equivalent to SSMD = $-\frac{3}{\sqrt{2}} \approx -2.12$ when the MM method is used to estimate SSMD in a primary screen without replicates (**Table 1**). The illustration in this article should help scientists to use z score, z* score, SSMD, SSMD*, and *t* statistic correctly for their screens, thus avoiding similar misusages and misleading results.

For screens with duplicates only, there are some options of analytic methods that may be helpful. However, these options cannot replace the need for a larger sample size when designing a screen. For hit selection in screens with replicates, it would be better to have four to seven replicates per siRNA, and we should have at least three replicates per siRNA.[23] In addition, it is also a good strategy to use a dual-flashlight plot to select hits by simultaneously considering both SSMD and a biologically more favored measure (such as average fold change or average percent inhibition/activation; cf. **Fig. 3**). Finally, although I use RNAi screens as examples for illustration, all the points presented in this article are applicable to other screens, including small-molecule screens.

ACKNOWLEDGMENTS

The author thanks Drs. Soper and Heyse for their support in this research.

REFERENCES

1. Birmingham, A.; Selfors, L. M.; Forster, T.; Wrobel, D.; Kennedy, C. J.; Shanks, E.; Santoyo-Lopez, J.; Dunican, D. J.; Long, A.; Kelleher, D.; et al. Statistical Methods for Analysis of High-Throughput RNA Interference Screens. *Nat. Methods* **2009**, *6*, 569–575.
2. Banerjee, S.; Byrd, J. N.; Gianino, S. M.; Harpstrite, S. E.; Rodriguez, F. J.; Tuskan, R. G.; Reilly, K. M.; Piwnica-Worms, D. R.; Gutmann, D. H. The Neurofibromatosis Type 1 Tumor Suppressor Controls Cell Growth by Regulating Signal Transducer and Activator of Transcription-3 Activity In Vitro and In Vivo. *Cancer Res.* **2010**, *70*, 1356–1366.
3. Beller, M.; Sztalryd, C.; Southall, N.; Bell, M.; Jäckle, H.; Auld, D. S.; Oliver, B. COPI Complex Is a Regulator of Lipid Homeostasis. *PloS Biol.* **2008**, *6*, 2530–2549.
4. Kassner, P. D. Discovery of Novel Targets with High Throughput RNA Interference Screening. *Comb. Chem. High Throughput Screen.* **2008**, *11*, 175–184.
5. Lapan, P.; Zhang, J.; Pan, J.; Hill, A.; Haney, S. A. Single Cell Cytometry of Protein Function in RNAi Treated Cells and in Native Populations. *BMC Cell Biol.* **2008**, *9*, 43.
6. Zhou, H. L.; Xu, M.; Huang, Q.; Gates, A. T.; Zhang, X. H. D.; Castle, J. C.; Stec, E.; Ferrer, M.; Strulovici, B.; Hazuda, D. J.; et al. Genome-Scale RNAi Screen for Host Factors Required for HIV Replication. *Cell Host Microbe* **2008**, *4*, 495–504.
7. Zhang, X. H. D.; Yang, X. C.; Chung, N. J.; Gates, A.; Stec, E.; Kunapuli, P.; Holder, D. J.; Ferrer, M.; Espeseth, A. S. Robust Statistical Methods for Hit Selection in RNA Interference High-Throughput Screening Experiments. *Pharmacogenomics* **2006**, *7*, 299–309.
8. Mohr, S.; Bakal, C.; Perrimon, N. Genomic Screening with RNAi: Results and Challenges. *Annu. Rev. Biochem.* **2010**, *79*, 37–64.
9. Zhao, W. Q.; Santini, F.; Breese, R.; Ross, D.; Zhang, X. H. D.; Stone, D. J.; Ferrer, M.; Townsend, M.; Wolfe, A. L.; Seager, M. A.; et al. Inhibition of Calcineurin-mediated Endocytosis and alpha-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic Acid (AMPA) Receptors Prevents Amyloid beta Oligomer-induced Synaptic Disruption. *J. Biol. Chem.* **2010**, *285*, 7619–7632.
10. Zhang, X. H. D. 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.
11. Vysochanskij, D. F.; Petunin, Y. I. Justification of the 3-Sigma Rule for Unimodal Distribution. *Theory Probab. Math. Stat.* **1980**, *21*, 25–36.
12. Malo, N.; Hanley, J. A.; Carlile, G.; Liu, J.; Pelletier, J.; Thomas, D.; Nadon, R. Experimental Design and Statistical Methods for Improved Hit Detection in High-Throughput Screening. *J. Biomol. Screen.* **2010**, *15*, 990–999.
13. Cohen, J. The Earth Is Round (P-Less-Than .05). *Am. Psychol.* **1994**, *49*, 997–1003.
14. Zhang, X. H. D. Strictly Standardized Mean Difference, Standardized Mean Difference and Classical t-Test for the Comparison of Two Groups. *Stat. Biopharm. Res.* **2010**, *2*, 292–299.
15. Zhang, X. H. D. A Method for Effectively Comparing Gene Effects in Multiple Conditions in RNAi and Expression-Profiling Research. *Pharmacogenomics* **2009**, *10*, 345–358.
16. Tusher, V. G.; Tibshirani, R.; Chu, G. Significance Analysis of Microarrays Applied to the Ionizing Radiation Response. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5116–5121.
17. Zhang, X. H. D. Assessing the Size of Gene or RNAi Effects in Multifactor High-Throughput Experiments. *Pharmacogenomics* **2010**, *11*, 199–213.
18. Smyth, G. K. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, 1–28.
19. Wright, G. W.; Simon, R. M. A Random Variance Model for Detection of Differential Gene Expression in Small Microarray Experiments. *Bioinformatics* **2003**, *19*, 2448–2455.
20. Zhang, X. H. D. Genome-wide Screens for Effective siRNAs through Assessing the Size of siRNA Effects. *BMC Res. Notes* **2008**, *1*, 33.
21. Zhang, X. H. D. An Effective Method for Controlling False Discovery and False Non-discovery Rates in Genome-scale RNAi Screens. *J. Biomol. Screen.* **2010**, *15*, 1116–1122.
22. Zhang, X. H. D.; Lacson, R.; Yang, R.; Marine, S.; McCampbell, A.; Toolan, D.; Hare, T. R.; Kajdas, J.; Berger, J. P.; Holder, D. J.; et al. The Use of SSMD-based False Discovery and False Non-discovery Rates in Genome-scale RNAi Screens. *J. Biomol. Screen.* **2010**, *15*, 1123–1131.
23. Zhang, X. H. D.; Heyse, J. F. Determination of Sample Size in Genome-scale RNAi Screens. *Bioinformatics* **2009**, *25*, 841–844.
24. Barrows, N. J.; Sommer, G. L.; Garcia-Blanco, M. A.; Pearson, J. L. Factors Affecting Reproducibility between Genome-scale siRNA-based Screens. *J. Biomol. Screen.* **2010**, *15*, 735–747.

Address correspondence to:
Xiaohua Douglas Zhang
Biometrics Research, Merck Research Laboratories
WP53B-120, P.O. Box 4, West Point, PA 19486

E-mail: xiaohua_zhang@merck.com