

# Journal of Biomolecular Screening

<http://jbx.sagepub.com/>

---

## The Use of Strictly Standardized Mean Difference for Hit Selection in Primary RNA Interference High-Throughput Screening Experiments

Xiaohua Douglas Zhang, Marc Ferrer, Amy S. Espeseth, Shane Douglas Marine, Erica M. Stec, Michael A. Crackower, Daniel J. Holder, Joseph F. Heyse and Berta Strulovici  
*J Biomol Screen* 2007 12: 497 originally published online 13 April 2007  
DOI: 10.1177/1087057107300646

The online version of this article can be found at:  
<http://jbx.sagepub.com/content/12/4/497>

---

Published by:



<http://www.sagepublications.com>

Additional services and information for *Journal of Biomolecular Screening* can be found at:

**Email Alerts:** <http://jbx.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jbx.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://jbx.sagepub.com/content/12/4/497.refs.html>

# The Use of Strictly Standardized Mean Difference for Hit Selection in Primary RNA Interference High-Throughput Screening Experiments

XIAOHUA DOUGLAS ZHANG,<sup>1</sup> MARC FERRER,<sup>2</sup> AMY S. ESPESETH,<sup>3</sup>  
SHANE DOUGLAS MARINE,<sup>2</sup> ERICA M. STEC,<sup>2</sup> MICHAEL A. CRACKOWER,<sup>4</sup>  
DANIEL J. HOLDER,<sup>1</sup> JOSEPH F. HEYSE,<sup>5</sup> and BERTA STRULOVICI<sup>2</sup>

RNA interference (RNAi) high-throughput screening (HTS) has been hailed as the 2nd genomics wave following the 1st genomics wave of gene expression microarrays and single-nucleotide polymorphism discovery platforms. Following an RNAi HTS, the authors are interested in identifying short interfering RNA (siRNA) hits with large inhibition/activation effects. For hit selection, the *z*-score method and its variants are commonly used in primary RNAi HTS experiments. Recently, strictly standardized mean difference (*SSMD*) has been proposed to measure the siRNA effect represented by the magnitude of difference between an siRNA and a negative reference group. The links between *SSMD* and *d*<sup>+</sup>-probability offer a clear interpretation of siRNA effects from a probability perspective. Hence, *SSMD* can be used as a ranking metric for hit selection. In this article, the authors investigated both the *SSMD*-based testing process and the use of *SSMD* as a ranking metric for hit selection in 2 primary siRNA HTS experiments. The analysis results showed that, as a ranking metric, *SSMD* was more stable and reliable than percentage inhibition and led to more robust hit selection results. Using the *SSMD*-based testing method, the false-negative rate can more readily be obtained. More important, the use of the *SSMD*-based method can result in a reduction in both the false-negative and false-positive rates. The applications presented in this article demonstrate that the *SSMD* method addresses scientific questions and fills scientific needs better than both percentage inhibition and the commonly used *z*-score method for hit selection. (*Journal of Biomolecular Screening* 2007;497-509)

**Key words:** strictly standardized mean difference, *d*<sup>+</sup>-probability, restricted false-positive rate, high-throughput screening, hit rate point

## INTRODUCTION

RNA INTERFERENCE (RNAi) HIGH-THROUGHPUT SCREENING (HTS) is broadly used in the identification of genes associated with specific biological phenotypes.<sup>1-6</sup> This technology has been hailed as the 2nd genomics wave following the 1st genomics wave of gene expression microarrays and single-nucleotide polymorphism discovery platform.<sup>7</sup> Following an RNAi HTS, we are interested in identifying short interfering RNAs (siRNAs) with large inhibition/activation effects compared to a negative reference. The size of this effect is represented by the

magnitude of difference between a tested siRNA and a negative reference group with no specific inhibition/activation effects. There are 2 main strategies of selecting hits with large effects. One is to use certain metric(s) to rank the siRNAs by their effects and then to select the largest number of potent siRNAs that is practical for confirmation and validation assays. The other strategy is to test whether an siRNA has effects strong enough to reach a pre-set specified effect. In this strategy, we need to control the false-negative and/or false-positive rates.

In the 1st strategy, currently 2 types of measures are commonly used to rank siRNA effects. One is mean difference, along with its variants such as signal-to-noise ratio and percentage inhibition; the other is *p*-value from either the *z*-score method or *t*-test of testing mean difference. The 1st measure cannot represent the magnitude of difference because it does not effectively capture the data variability.<sup>8</sup> When statistical significance is used, the *p*-value comes from testing the hypothesis of no mean difference between 2 groups. It addresses whether an siRNA has the same effect as the negative reference based on the sample observation. It is not designed to measure

<sup>1</sup>Biometrics Research, <sup>2</sup>Automated Biotechnology, <sup>3</sup>Molecular and Cellular Technology, and <sup>5</sup>BARDs, Merck Research Laboratories, West Point, Pennsylvania.  
<sup>4</sup>Biochemistry and Molecular Biology, Merck Frosst, Kirkland, Quebec, Canada.

Received Dec 15, 2006, and in revised form Jan 19, 2007. Accepted for publication Feb 1, 2007.

Journal of Biomolecular Screening 12(4); 2007  
DOI:10.1177/1087057107300646

how large the magnitude of difference is.<sup>9,10</sup> Thus, an siRNA effect that results in a low *p*-value may not cause a robust enough effect on the assay to indicate any meaningful biological association. Therefore, neither mean difference nor *p*-value can represent the magnitude of difference.

For a better metric to measure the magnitude of difference, Zhang<sup>11</sup> proposed strictly standardized mean difference (*SSMD*) and compared *SSMD* with *z*-factor and other quality control metrics in assessing assay quality. Unlike mean difference and percentage inhibition, *SSMD* is robust to both measurement unit and strength of positive controls; it takes into account data variability in both compared groups and has a probability interpretation.<sup>11,12</sup> Compared to *p*-value, *SSMD* directly measures how large the magnitude of difference is.<sup>11</sup> Here we evaluate *SSMD* as a ranking metric for hit selection in 2 in-house primary RNAi HTS experiments.

In the 2nd strategy, currently the mean  $\pm k$  SD method and its variants, such as the median  $\pm k$  MAD method, are commonly used in hit selection in primary HTS assays.<sup>13-19</sup> *SD* and *MAD* denote standard deviation and median absolute deviation, respectively. The mean  $\pm k$  SD method relies on the *z*-score of standard normal distribution  $N(0,1)$  and is thus also called the *z*-score method. In confirmatory HTS experiments, the *t*-test is popularly used. Both the *z*-score method and the *t*-test of testing no mean difference control the false-positive rate, in which we conclude that there exists (even a tiny) mean difference, but actually there is no mean difference. However, in practice, what we are really interested in is not whether an siRNA (or compound) has average inhibition/activation effects being the same as the negative reference. Instead, we are interested in the false-negative rate, in which the siRNAs or compounds with the large effects are not selected as hits, and the false-positive rate, in which the siRNAs or compounds with no or weak effects are selected as hits, in the process of hit selection. To address this question, Zhang<sup>12</sup> proposed an *SSMD*-based process with a flexible and balanced control of false positives and false negatives in hit selection. In this article, we evaluate the use of this process for hit selection in 2 in-house primary RNAi HTS experiments.

## METHODS

### **SSMD for measuring and ranking siRNA effects**

Consider 2 independent populations,  $P_1$  and  $P_2$ , and let  $D$  be the difference between the 2 populations. *SSMD* ( $\beta$ ) is defined as the ratio of the mean to the SD of  $D$ —namely,  $\beta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ , where  $\mu_1$ ,  $\mu_2$ ,  $\sigma_1$ , and  $\sigma_2$  are the population means and SD of  $P_1$  and  $P_2$ . *SSMD* measures the strength of the difference between 2 compared groups and is a good candidate for assessing differences between an siRNA and a negative reference group.<sup>11,12</sup> In addition, in cases where we are willing to postulate a distribution for  $D$ , the *SSMD* provides a metric that can be linked to the

probability (namely,  $d^+$ -probability) that a value in  $P_1$  is greater than a value in  $P_2$ .

Ideally, the negative reference group would be a set of known negative controls used in the experiment. If there are difficulties with the negative controls (e.g., they are small in number or have strong bias), then it is common to use some of the values from sample wells to estimate a negative reference distribution.<sup>14</sup> For instance, when a large proportion of the sample wells are expected to be negative, we might exclude the largest 1% values and the smallest 1% values and use the remaining 98% values in the sample wells as a negative reference group in each plate to calculate percentage inhibition, estimated *SSMD* value, and *z*-score in primary screens.<sup>12</sup>

Because *SSMD* is an unknown population parameter, we need to estimate it based on observed samples. Under the situation of equal variance in 2 compared groups, the uniformly minimal variance unbiased estimate ( $\hat{\beta}$ ) of *SSMD* ( $\beta$ ) is

$$\hat{\beta} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{2}{K} ((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2)}}, \quad K \approx n_1 + n_2 - 3.5 (n_1, n_2 \geq 2),$$

where  $n_1$ ,  $\bar{X}_1$ ,  $s_1^2$ , and  $n_2$ ,  $\bar{X}_2$ ,  $s_2^2$ , are the sample size, sample mean, and sample variance in the 2 compared groups, respectively. The estimated variance of  $\hat{\beta}$  is  $\hat{\sigma}_{\hat{\beta}}^2 \approx \frac{1}{2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$

when  $n_1 + n_2$  is large. The above formulas can be used to measure the magnitude of difference between a group of interest and a negative reference. In the primary HTS experiment with no replicates for most investigated siRNAs,  $n_1 = 1$ , so the formula to calculate the estimated *SSMD* ( $\hat{\beta}$ ) is reduced to

$$\hat{\beta} = \frac{x - \bar{X}_2}{\sqrt{\frac{2}{K} (n_2 - 1)s_2^2}}, \quad \text{where } K \approx n_2 - 2.5, \quad x \text{ is the observed}$$

value of an siRNA, and  $\bar{X}_2$ ,  $n_2$ , and  $s_2^2$  are the sample size, mean, and variance of the negative reference group in a plate, respectively. The estimated variance of  $\hat{\beta}$  is  $\hat{\sigma}_{\hat{\beta}}^2 \approx \frac{1}{2} \left( 1 + \frac{1}{n_2} \right)$  when  $n_2$  is large ( $n_2 \geq 50$ ). See Zhang<sup>12</sup> for details about the *SSMD* estimation and its variance.

### **SSMD-based hit selection**

For screens in which we want a fixed number of hits, we may simply rank the siRNAs based on *SSMD* and choose the siRNAs with the top ranks. For other screens, we may want to choose siRNAs based on a threshold with desired statistical characteristics. Although in many statistical hypothesis testing problems, it is common to pick a threshold based primarily on the false-positive rate, in screening experiments, it is often expedient to guard against false positives and false negatives. When we are interested in the siRNAs with large positive effects (namely, the siRNAs with  $SSMD \geq c$ , where  $c$  is a preset level of *SSMD*),

## Use of Strictly Standardized Mean Difference for Hit Selection

**Table 1.** Thresholds and Error Rates for 3 Different Testing Strategies

Strategy	Threshold ( $t$ )	RFPL	FNL
I: FNL = $\alpha$	$3 - 0.71 \times Z_\alpha$	$\Phi\left(Z_\alpha - \frac{3-c_2}{0.71}\right)$	$\alpha$
II: RFPL = FNL	$\frac{3+c_2}{2}$	$1 - \Phi\left(\frac{3-c_2}{1.42}\right)$	$1 - \Phi\left(\frac{3-c_2}{1.42}\right)$
III: Number of hits = $h$	Choose $t$ such that the number of siRNAs with $\hat{\beta} \geq t$ is $h$	$\Phi\left(\frac{c_2-t}{0.71}\right)$	$1 - \Phi\left(\frac{3-t}{0.71}\right)$

RFPL, restricted false-positive level; FNL, false-negative level; siRNA, short interfering RNA.

the false-negative rate (FNR) is the probability of concluding  $\beta < c$  when actually  $\beta \geq c$ , and the false-positive rate (FPR) is the probability of concluding  $\beta \geq c$  when actually  $\beta < c$ . In primary HTS screens, we usually do not want to miss siRNAs with strong positive effects (namely those with  $\beta \geq c$ ), but we may tolerate the false positives with fairly strong or fairly weak positive effects (namely, those with  $\beta > c_2$ , where  $0 \leq c_2 \leq c$ ). The rate of this type of false positives—namely,  $\Pr(\text{conclude } \beta \geq c \text{ given } \beta \leq c_2)$ —is called the restricted false-positive rate.

The decision rule for selecting siRNAs with large positive effects is<sup>12</sup> concluding  $\beta \geq c$  if  $\hat{\beta} \geq c - Z_\alpha \hat{\sigma}_{\hat{\beta}}$  and concluding  $\beta < c$  if  $\hat{\beta} < c - Z_\alpha \hat{\sigma}_{\hat{\beta}}$ , where  $\hat{\sigma}_{\hat{\beta}}$  is the estimated standard deviation of estimated SSMD, and  $Z_\alpha$  satisfies  $\Pr(Z \leq Z_\alpha) = 1 - \alpha$ , where  $Z$  is a standard normal distribution. That is, the threshold of  $\hat{\beta}$  is  $c - Z_\alpha \hat{\sigma}_{\hat{\beta}}$ . When the negative reference group works effectively and the sample size is large, the above decision rule can ensure that the FNR is no more than the preset level  $\alpha$ . The maximum of FNR of a decision rule is called the false-negative level (FNL). The maximum of the restricted false-positive rate of a decision rule is called the restricted false-positive level (RFPL). The RFPL of the above decision rule is equal to  $\Phi\left(Z_\alpha - \frac{c-c_2}{\hat{\sigma}_{\hat{\beta}}}\right)$ , where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution. In our experiments, we used the sample wells, after trimming off the values in the largest and smallest 1%, as the negative reference group. Accordingly, our  $n_1 = 1$ , and  $n_2$  ranged from 78 to 295, yielding  $\hat{\sigma}_{\hat{\beta}} \approx \sqrt{\frac{1}{2}\left(1 + \frac{1}{n_2}\right)} \approx \sqrt{\frac{1}{2}} \approx 0.71$ .

The relationship between FNL and RFPL of the decision rule is thus  $\text{RFPL} = \Phi\left(Z_{\text{FNL}} - \frac{c-c_2}{0.71}\right)$ . Based on the original meaning of SSMD and its link with  $d^+$ -probability,  $c = 3$  is commonly used, although occasionally, we use  $c = 2$ ; common choices of  $c_2$  are 0, 0.25, 0.5, 1, 1.645, and 2, depending on experimental needs.<sup>12</sup>

In this study, we applied the methods described above to 2 data sets: a hepatitis C virus siRNA HTS experiment and a mucin siRNA HTS experiment. We used 3 testing strategies—namely, controlling FNL (and/or RFPL), RFPL = FNL, and the fixed number of hits approaches. The thresholds and error rates

for these approaches are listed in **Table 1**. See Zhang<sup>12</sup> for the details about the SSMD-based approaches for hit selection.

### Percentage inhibition and z-score

Percentage inhibition is commonly used for data analysis in HTS experiments. In this article, percentage inhibition ( $y$ ) for a well in a plate is calculated using the following formula:

$$y = \frac{x - \bar{X}_2}{\bar{X}_+ - \bar{X}_2} \times 100, \text{ where } x \text{ is the raw value (in } -\log_2 \text{ scale) in}$$

a well in a plate,  $\bar{X}_2$  is the median of sample wells in the plate, and  $\bar{X}_+$  is the median of strong positive control wells in the plate. Percentage inhibition is usually used for ranking siRNA effects. The  $z$ -score is calculated using the following formula:

$$Z_{\text{obs}} = \frac{x - \bar{X}_2}{s_2}, \text{ where } x, \bar{X}_2, n_2, \text{ and } s_2^2 \text{ are the same as those}$$

in calculating  $\hat{\beta}$  in this article. The  $z$ -score can be used as both ranking siRNA effects and testing siRNA effects (namely, controlling the false-positive rate of concluding that an siRNA has a mean different from the mean of the negative reference group, but actually the 2 means are the same). The corresponding  $p$ -value (or false-positive rate) is  $2 \Pr(Z \geq |Z_{\text{obs}}|)$  when  $Z$  is a standard normal distribution. The false-positive rate is usually controlled to be 0.05 (which is equivalent to the mean  $\pm 1.96$  SD method, very close to mean  $\pm 2$  SD) or 0.0027 (which is equivalent to mean  $\pm 3$  SD) under the normality assumption. For a 1-sided test, the  $p$ -value is  $\Pr(Z \geq Z_{\text{obs}})$ .

### Experiments

In this article, we applied the SSMD-based methods to 2 primary RNAi HTS experiments: a primary hepatitis C virus (HCV) screen and a primary mucin screen. Both screens were carried out in a 384-well format, using oligos designed by Rosetta Inpharmatics (Seattle, WA). In both screens, the sample siRNAs are randomly placed within each type, and thus independence can be assumed. Note that the methods of hit selection identify the siRNA pools that affect the phenotype of

the cells being studied. Although the siRNA pools tested are designed to knock down mRNA levels of specific genes, at the point in the siRNA HTS experiment where hit selection is conducted, these selections are carried out in the absence of the confirmation of mRNA knockdown.

In the HCV screen,<sup>19</sup> a total of about 22,000 siRNAs genes were tested across 97 plates. The experiment was designed to identify host factors associated with HCV replication, using the HCV replicon assay system described in Zuck et al.<sup>1</sup> Huh-7 cells containing an HCV genotype 1b replicon with a β-lactamase reporter were transfected with siRNA pools targeting ~19,000 genes. Three days after siRNA transfection, the cells were stained for β-lactamase expression. siRNAs that decreased β-lactamase expression were considered to affect HCV replication. In this experiment, there were sample siRNAs (no replicates in the whole experiment); a strong positive control, labeled “Positive Control” (16 replicates per plate); and a negative control (16 replicates per plate).

In the mucin screen, a total of 22,368 genes were tested in pools of 3 siRNAs per well across 79 plates. The experiment was designed to identify genes involved in the secretion of mucin—more specifically, Muc5AC. A total of 20,000 H292 human lung epithelial cells per well were transfected with siRNA in reverse transfection mode. Twenty-four hours after siRNA transfection, a media change was carried out, and cells were induced with a Muc5AC induction cocktail. Forty-eight hours later, cell supernatant was harvested, and an enzyme-linked immunosorbent assay (ELISA) was carried out to measure the amount of Muc5AC secreted from H292 human lung epithelial cells into the cell supernatant. Genes interfering with either the generation or secretion of Muc5AC are considered hits. In this experiment, each sample siRNA has no replicates, and the positive control and the negative control both have 8 replicates per plate.

## RESULTS

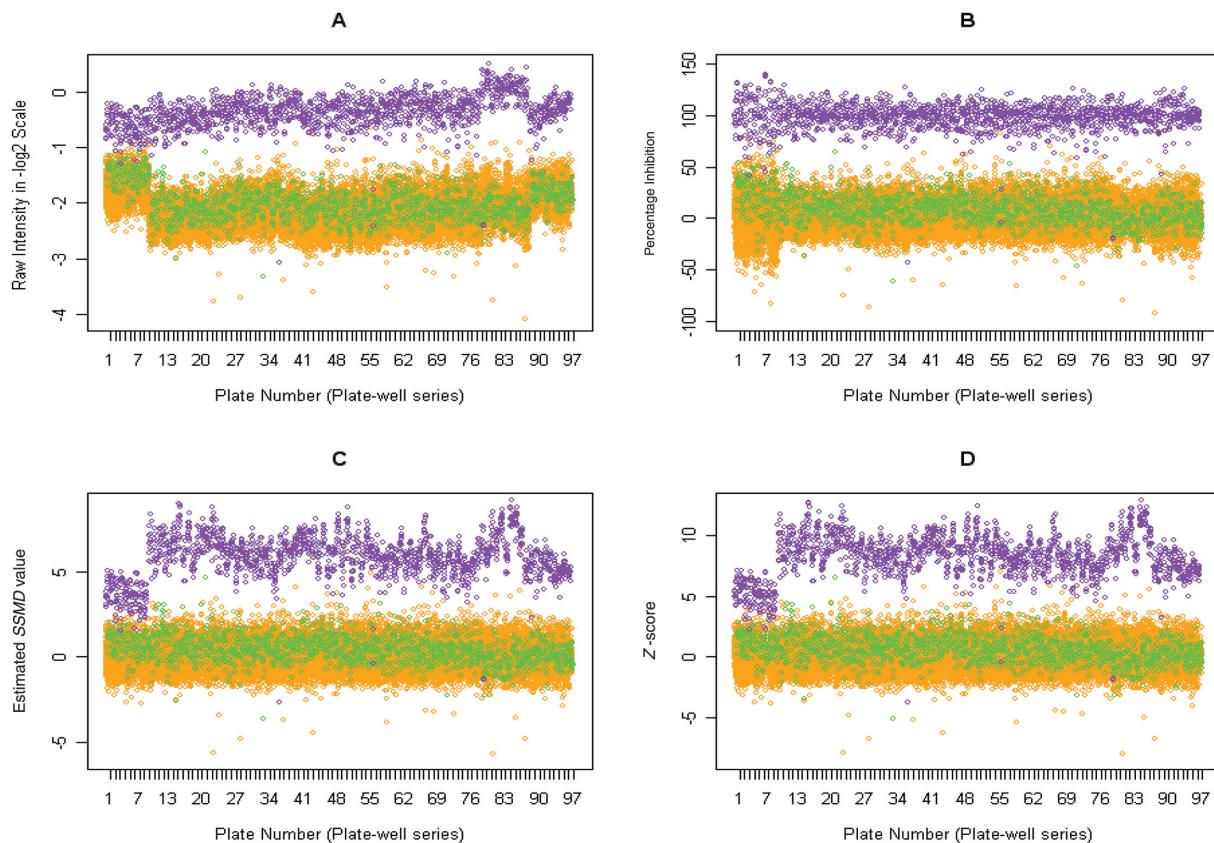
### **Hit selection using the ranking strategy**

Currently, percentage inhibition and z-score (or corresponding *p*-value) are commonly used for hit selection using the ranking strategy. A recently proposed parameter, *SSMD*, can be also used to rank siRNA effects with a better theoretical interpretation.<sup>12</sup> Here we investigated all these 3 measures in ranking siRNA effects in the 2 primary RNAi HTS experiments. Let us first focus on the HCV primary siRNA HTS experiment. The raw value, percentage inhibition, estimated *SSMD* value, and *z*-score for all the 97 plates in this experiment are displayed using plate-well series plots<sup>19</sup> in **Figure 1**.

**Figure 1A** shows the raw intensity (in  $-\log_2$  scale), where we can see that the positive control wells have apparently higher intensities in plates 79 to 88 than in the remaining plates. On the other hand, both negative control and sample wells in plates 1 to 9 have intensities apparently higher than those in the

remaining plates. The values of percentage inhibition in the sample wells are stretched in plates 1 to 9 and shrunken in plates 79 to 88 (**Fig. 1B**). Consequently, a value of 40 in percentage inhibition in plates 1 to 9 might be equivalent to a value of 20, not 40, in plates 79 to 88 and equivalent to a value of 30 in the remaining 78 normal plates.

By contrast, both *SSMD* and *z*-score in the sample wells are fairly stable, more stable than percentage inhibition, across plates (**Fig. 1C, D**), and they are not affected by the errors in the positive control wells. The *SSMD* value has the same meaning across plates, and so does the *z*-score. It is notable that the shape of *SSMD* values is the same as that of *z*-score values in **Figure 1**. Actually, for the siRNAs without any replicate, the *SSMD* value has a linear relationship to the corresponding *z*-score value—namely  $Z_{\text{obs}} = \sqrt{2}\hat{\beta}$ , if both are calculated using the same negative reference. Therefore, in terms of ranking siRNA effects, the *SSMD* method is equivalent to the *z*-score method, although the *z*-score value is  $\sqrt{2}$  times the estimated *SSMD* value. However, the *SSMD* method has a theoretically based criterion to classify the strength of siRNA effects. Based on the 1-2-3 rule,<sup>12</sup> using the estimated *SSMD* value to approximately represent the population value of *SSMD*, we identified 23 (and 113) siRNAs with strong inhibition effects, 254 (and 366) siRNAs with fairly strong inhibition effects, and 2668 (and 1795) siRNAs with moderate inhibition effects in the HCV (and mucin) screens. The method of using ranking strategy to select siRNAs with large inhibition effects is to choose a suitable number of siRNAs with the largest values of a given ranking metric. If, as an example, we decide to select 800 siRNAs from an HTS screen for further confirmatory experiment, the specific siRNAs selected will be highly dependent on the selection methods employed. **Figure 2A** shows a comparison between hit selection based on percentage inhibition relative to the negative reference (red points) and hit selection based on *SSMD* (or *z*-score, shown in blue). Selecting the 800 siRNAs with the largest percentage inhibitions (**Fig. 2A**, red) and graphing the number of hits on a plate-by-plate basis, it is apparent that there is a hill in plates 1 to 9 and a clear valley in plates 79 to 88. Among these 800 hits, there were 25.7 hits per plate in plates 1 to 9, 3.4 hits per plate in plates 79 to 88, and 6.86 hits per plate in the 78 normal plates on average. The number of hits per plate in plates 1 to 9 was about 4 times that in the normal plates. In contrast, if we select 800 siRNA hits by determining those with the largest *SSMD* (or *z*-score) (graphed by hit number per plate in blue, **Fig. 2A**), the number of hits per plate is fairly flat without any apparent hill or valley, as was observed when percentage inhibition was used as the criterion. Among these 800 hits with the largest *SSMD* (or *z*-score), there were 6.7 hits per plate in plates 1 to 9, 8.2 hits per plate in plates 79 to 88, and 8.4 hits per plate in the 78 normal plates on average. The number of hits per plate in plates 1 to 9 is nearly the same as (only slightly less than) plates 10 to 78 and 89 to 97, and the number of hits per plate in plates 79 to 88 is about the same as in plates 10 to 78 and 89 to 97.



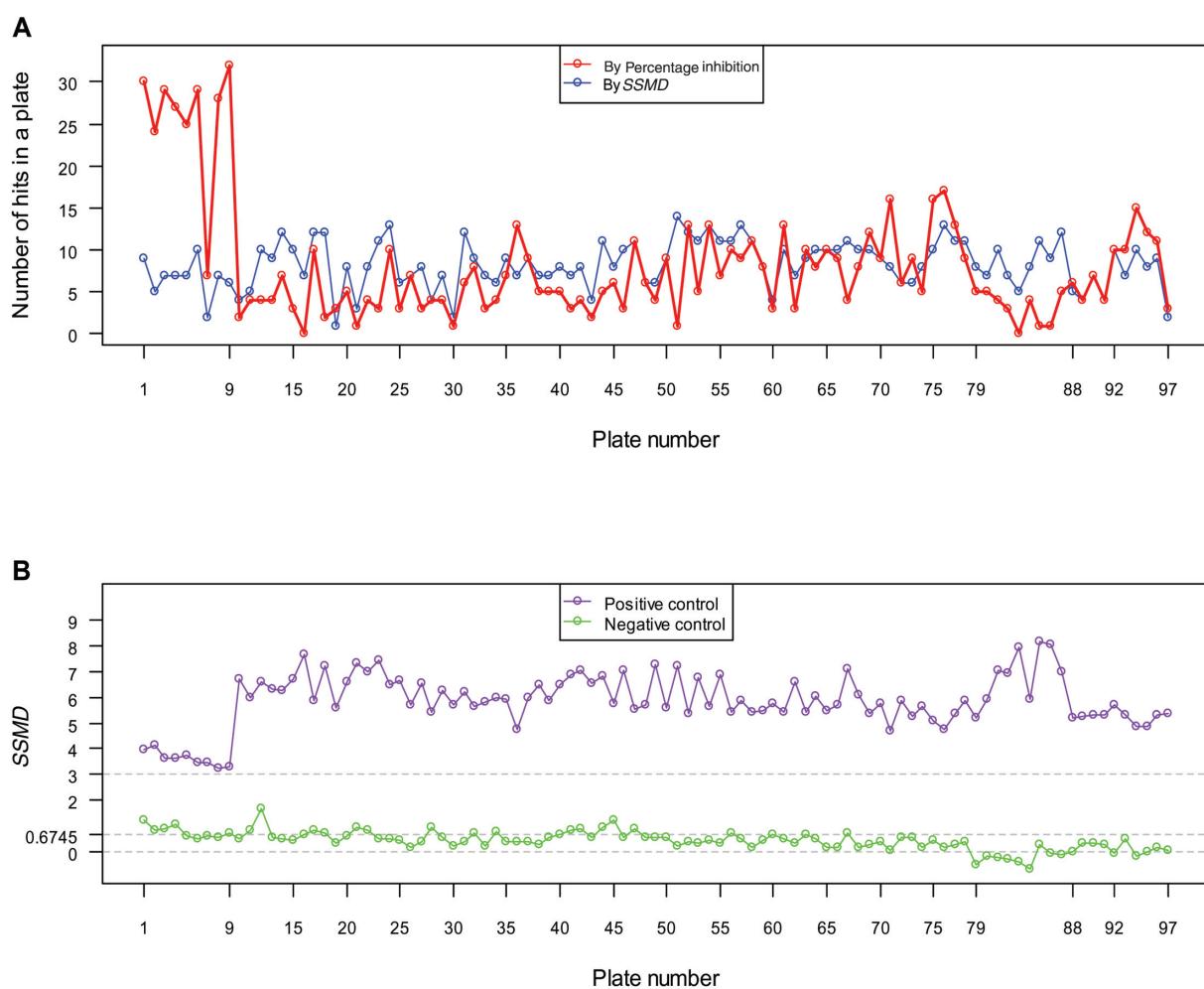
**FIG. 1.** Plate-well series plots to display percentage inhibition, strictly standardized mean difference (*SSMD*), and *z*-score for ranking short interfering RNA (siRNA) effects in the hepatitis C virus (HCV) siRNA primary experiment. Panels **A**, **B**, **C**, and **D** display raw intensity in  $-\log_2$  scale, percentage inhibition, estimated *SSMD* value, and *z*-score, respectively. In each panel, a point denotes a well in the plate, and the plate numbers are labeled in the *x*-axis; orange, green, and purple points represent the sample wells, negative control wells, and positive control wells, respectively.

**Figure 2B** displays the estimated *SSMD* value of positive controls (or negative controls) as a group in each plate. By comparing panels **A** and **B**, we can see that the hill (or valley) in the red line of panel **A** matched with the valley (or hill) in the purple line (for positive controls) of panel **B**, which suggests that the number of hits in a plate is strongly affected by the performance of positive controls when using percentage inhibition as a ranking metric. On the other hand, the number of hits in a plate is not affected by the performance of positive controls.

Now let us look at the mucin siRNA primary HTS experiment. **Figure 3A** shows the raw intensity (in  $-\log_2$  scale), where we can see that the positive control wells have apparently higher intensities in plates 29 to 41 than in the remaining plates. The raw intensities of sample wells in these plates are also higher than the sample wells in the other plates. Another apparent feature in these data is that, except for a few outliers, the positive controls are separated from the sample wells much better in plates 42 to 54 than in the other plates. As a result, the sample

siRNAs in plates 29 to 41 tend to have larger absolute value, whereas those in plates 42 to 54 tend to have smaller absolute value of percentage inhibition in either the positive or negative direction (**Fig. 3B**), which suggests the same value of percentage value may have different meanings depending on the quality of positive controls in different sets of plates. By contrast, both *SSMD* and *z*-score in the sample wells are fairly stable, more stable than percentage inhibition, across plates (**Fig. 3C, D**).

Among the 800 siRNAs with the largest percentage inhibition, the hit number in each plate is displayed by the red points and line in **Figure 4A**, which has a hill in plates 29 to 41 and a valley in plates 42 to 54. There were 14.8 hits per plate in plates 29 to 41, 4.5 hits per plate in plates 42 to 54, and 10.3 hits per plates in the 53 normal plates on average. The number of hits per plate in plates 29 to 41 (or plates 42 to 54) was about 1.5 times (or half of) that in the normal plates. Again, by comparing panels **A** and **B** of **Figure 4**, we can see that the hill (or valley) in the red line of panel **A** matched with the valley (or hill) in the purple line (for positive controls) of panel **B**.



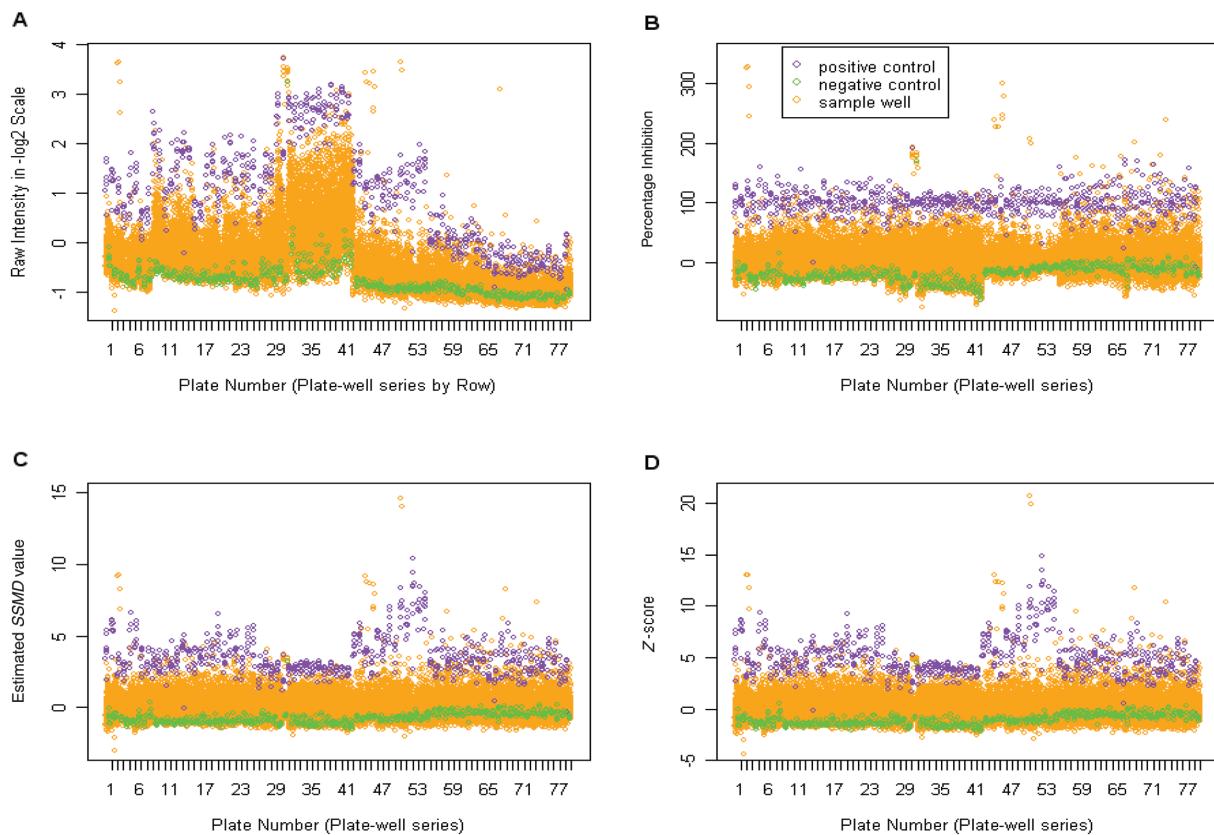
**FIG. 2.** The number of hits per plate using the ranking strategy (A) and the estimated strictly standardized mean difference (SSMD) values of positive (or negative) controls as a group (B) in the hepatitis C virus (HCV) short interfering RNA (siRNA) primary high-throughput screening (HTS) experiment. In panel A, the red (or blue) points display the number of hits per plate among the 800 siRNAs with the largest values of percentage inhibition (or SSMD). In panel B, the purple (or green) points denote the estimated SSMD values of positive (or negative) control wells as a group.

#### Hit selection using testing strategy

We applied the SSMD-based testing process<sup>12</sup> to select hits with large inhibition effects. To use this process for hit selection in primary RNAi HTS experiments, we should determine the threshold for  $\beta$ . As described in the Methods section, we determined the threshold using the 3 main approaches shown in Table 1. We applied these 3 approaches in both the mucin and HCV siRNA HTS experiments. The 1st approach focuses on the control of a low FNL for the siRNAs with strong positive effects ( $SSMD \geq 3$ ) and/or a low RFPL for the siRNAs with weak or no positive effects ( $SSMD \leq c_2$ ). When we applied this approach, we started with some common FNLs such as 0.01, 0.025, 0.10, and 0.15. The 2nd approach is the completely balanced error level

(CBEL) approach, in which we specify a  $c$  value and a  $c_2$  value first and then find the FNL value such that  $FNL = RFPL$  to obtain the threshold. When we applied this approach, we investigated the situations of  $c = 3$  and  $c_2 = 0, 0.25, 0.5, 1, 1.6449$ , and 2, respectively. The 3rd approach focuses on the number of siRNAs that we want to select for further research (such as for confirmatory screens). When we applied this approach, we considered 1000, 800, 600, 500, 400, 300, 200, and 100 siRNAs with the largest positive SSMD values (namely, with the strongest inhibition effects). The results of hit selection using these 3 approaches are displayed in Tables 2, 3, and 4, respectively.

In Tables 2 through 4, the empirical FNR and empirical FPR were calculated based on the positive and negative control wells, respectively. Namely, in the SSMD-based method, the empirical

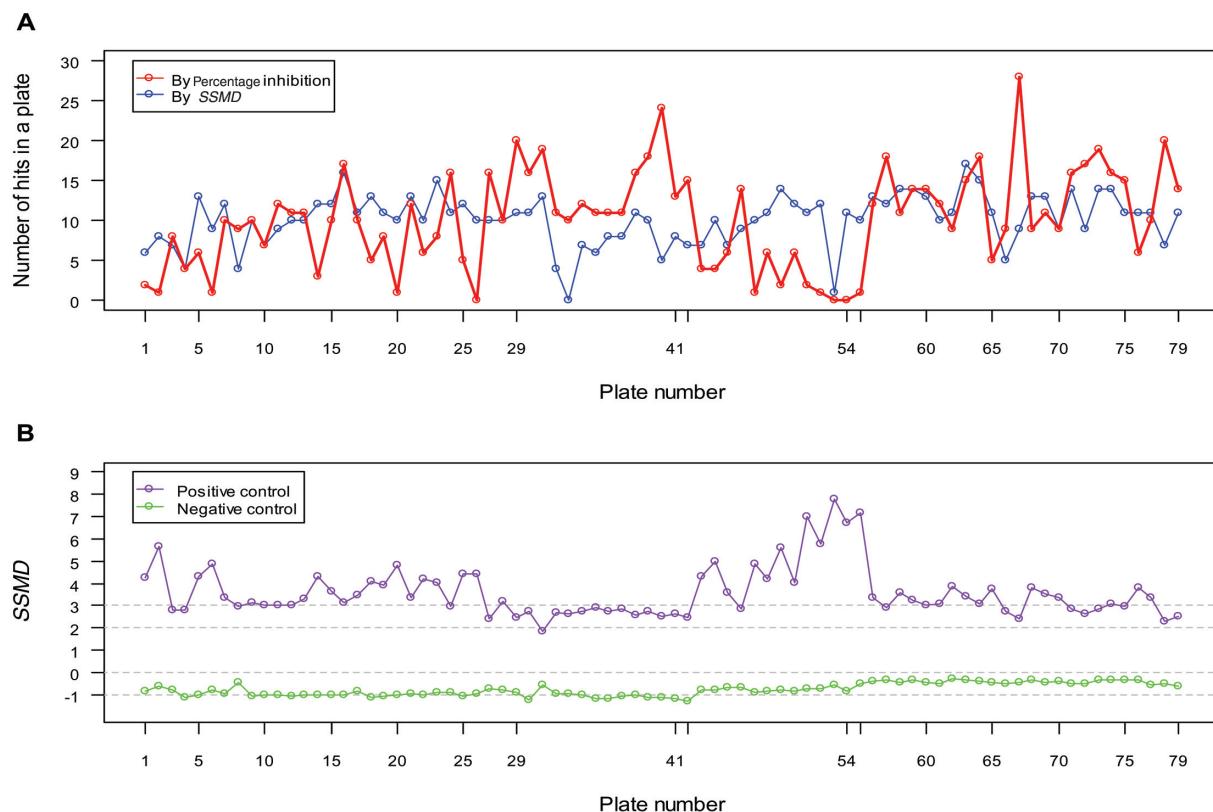


**FIG. 3.** Plate-well series plots to display percentage inhibition, strictly standardized mean difference (SSMD), and z-score for ranking short interfering RNA (siRNA) effects in the mucin siRNA primary experiment. Panels A, B, C, and D display raw intensity in  $-\log_2$  scale, percentage inhibition, estimated SSMD value, and z-score, respectively. In each panel, a point denotes a well in the plate, and the plate numbers are labeled in the x-axis; the orange, green, and purple points represent the sample wells, negative control wells, and positive control wells, respectively, as labeled in the legend of panel B. Note that 1 sample well is not shown in panels B to D. This sample well has a percentage inhibition of 897, an SSMD value of 20.77, and a z-score of 29.45 and is from plate 67.

FNR is the proportion of positive control wells with estimated SSMD values less than the threshold of  $\hat{\beta}$  for selecting hits; in the z-score method, the empirical FNR is the proportion of positive control wells with z-scores less than 3 (or 2). Similarly, in the SSMD-based method, the empirical FPR is the proportion of negative control wells with estimated SSMD values greater than or equal to the threshold; in the z-score method, the empirical FPR is the proportion of positive control wells with z-scores greater than or equal to 3 (or 2). Note that the mean difference between 2 groups cannot be used to represent the magnitude of difference. Therefore, it is nontrivial to calculate the theoretical FNL for the z-score method. The FPL in the z-score method has different meaning from the SSMD method. The z-score method takes any selected siRNA as a false positive when its mean is the same as or less than the mean of the negative reference, whereas the SSMD method takes a selected siRNA as a false positive when its effect (magnitude of difference) is not greater than a presetage level (defined by the value of  $c_2$ ).

From **Tables 2** to **4**, in the SSMD-based method, the empirical FNR is less than the corresponding controlled FNL, which suggests that the FNR was well controlled as we expected. As displayed in **Table 4**, when we selected siRNAs with the largest SSMD values, the FNL was higher, but the RFPL was lower in the mucin screen than in the HCV screen.

Given an FNL, the hit rate was between the RFPLs corresponding to  $c_2 = 0$  and  $c_2 = 1$ , respectively. We further explored the effect size ( $c_2$  value) under which the RFPL was equal to the hit rate given a controlled FNL. This  $c_2$  value is called hit rate point. **Figure 5** shows the search of the hit points in the cases of  $FNL = 0.05$  and  $0.15$  in the 2 screens, respectively. When the FNLs were  $0.05$  and  $0.15$ , the hit rate points were  $0.29$  and  $0.426$ , respectively, in the HCV screen, both of which were fairly small, suggesting that the selected hits mostly had weak inhibition effects. In the mucin screen, the hit rate points were  $0.464$  and  $0.712$ , respectively, both of which were nearly double the corresponding hit rate points in the HCV screen, which



**FIG. 4.** The number of hits per plate using ranking strategy (A) and the estimated strictly standardized mean difference (SSMD) values of positive (or negative) controls as a group (B) in the mucin short interfering RNA (siRNA) primary high-throughput screening (HTS) experiment. In panel A, the red (or blue) points display the number of hits per plate among the 800 siRNAs with the largest values of percentage inhibition (or SSMD). In panel B, the purple (or green) points denote the estimated SSMD values of positive (or negative) control wells as a group.

suggested that the selected hits in the mucin screen had inhibition effects stronger than those in the HCV screen under the same FNL. In each experiment, the hit rate points for other values of FNL are listed in the last row of the corresponding panel in **Tables 2, 3, and 4**.

We usually have 2 objectives in a primary RNAi HTS screen: screening out about 200 to 1200 candidate siRNAs for further confirmatory screens and selecting about 100 to 300 candidate siRNAs for preliminary gene pathway analysis. To select inhibition hits for further confirmatory screens, we focused on controlling the FNL. Based on **Table 2**, using the SSMD-based method, we controlled the FNL to be 0.025 and thus selected 767 (and 883) inhibition siRNAs in the HCV (and mucin) HTS experiment. To select inhibition hits for preliminary gene function and pathway analysis, we first adopted the ranking approach and selected 200 inhibition siRNAs in both the HCV and mucin HTS experiments. In this ranking approach, the FNL was 0.1099 in the HCV experiment, which was less than half of the FNL (i.e., 0.2620) in the mucin experiment (**Table 4**). To maintain FNL and RFPL to be the same in both experiments, we also adopted the CBEL approach and selected 127 (and 289) inhibition siRNAs in

the HCV (and mucin) HTS experiment based on the case of  $c = 3$  and  $c_2 = 1.6449$  (**Table 3**). Considering the balanced control of FNL and RFPL, the CBEL approach appeared to be more reasonable than the ranking approach for these 2 experiments.

## DISCUSSION

There are 2 main strategies for hit selection in RNAi HTS. One is to use metric(s), such as the signal-to-noise ratio, percentage inhibition, or  $p$ -value from either the  $z$ -score method or  $t$ -test of testing mean difference, to rank the siRNAs by potency and then to select a specified number of the strongest siRNAs for further investigation. The other strategy is to test whether an siRNA has effects strong enough to reach a preestablished cutoff point for hit selection—a strategy in which the false-negative and/or false-positive rates must be carefully controlled. Establishing cutoff points is commonly done using the mean  $\pm k$  SD method (or the  $z$ -score method) and its variants such as the median  $\pm k$  MAD method in hit selection in HTS assays.<sup>13-19</sup> Zhang<sup>11</sup> suggested the use of SSMD for measuring and ranking siRNA effects. Zhang<sup>12</sup> proposed an SSMD-based testing process

## Use of Strictly Standardized Mean Difference for Hit Selection

**Table 2.** Hit Selection Criteria, the Number and Rate of Selected Hits, Empirical FNR and FPR, and Hit Rate Point in the Application of Both the *SSMD*-Based Approach (with 5 FNLS) and the *z*-Score Method for Both the HCV and Mucin HTS Primary Screens

Method	<i>I SSMD-Based Method</i>					<i>II z-Score Method</i>	
	FNL Threshold of Estimated SSMD Value	0.01 $\hat{\beta} \geq 1.3483$	0.025 $\hat{\beta} \geq 1.6084$	0.05 $\hat{\beta} \geq 1.8322$	0.10 $\hat{\beta} \geq 2.0901$	0.15 $\hat{\beta} \geq 2.2641$	$z \geq 2$
<b>A</b>							
RFPL or FPL							
$c_2 = 0$	0.0288	0.0117	0.0049	0.0016	0.0007	0.025	0.00135
$c_2 = 0.5$	0.1161	0.0592	0.0303	0.0126	0.0065		
$c_2 = 1$	0.3119	0.1957	0.1206	0.0623	0.0375		
$c_2 = 2$	0.8207	0.7094	0.5934	0.4495	0.3549		
<b>B</b>							
HCV assay							
Hit number	1425	767	433	220	138	1224	204
Hit rate	0.0492	0.0265	0.0149	0.0076	0.0048	0.0422	0.0070
Empirical FNR	0.0026	0.0039	0.0052	0.0058	0.0071	0.0025	0.0064
Empirical FPR	0.1102	0.0599	0.0341	0.0187	0.0122	0.0921	0.0161
Hit rate point	0.1746	0.2342	0.2902	0.3662	0.4253	NA	NA
<b>C</b>							
Mucin assay							
Hit number	1325	883	605	412	323	1187	390
Hit rate	0.0590	0.0393	0.0269	0.0183	0.0144	0.0528	0.0174
Empirical FNR	0.0063	0.0174	0.0380	0.0854	0.1092	0.0079	0.0870
Empirical FPR	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016
Hit rate point	0.2383	0.3597	0.4635	0.6067	0.7115	NA	NA

*SSMD*, strictly standardized mean difference; HTS, high-throughput screening; FNL, false-negative level; RFPL, restricted false-positive level; FPL, false-positive level; HCV, hepatitis C virus; FPR, false-positive rate; NA, not applicable.

**Table 3.** Hit Selection Criteria, Number and Rate of Selected Hits, Empirical FNR and FPR, and Hit Rate Point in the Application of the *SSMD*-Based CBEL Approach for Both the HCV and Mucin HTS Primary Screens

$c_2$	0	0.25	0.5	1	1.645	2	3
FNL	0.0173	0.0264	0.0392	0.0795	0.17	0.2406	0.5
Threshold of Estimated SSMD Value	$\hat{\beta} \geq 1.5$	$\hat{\beta} \geq 1.625$	$\hat{\beta} \geq 1.75$	$\hat{\beta} \geq 2$	$\hat{\beta} \geq 2.3225$	$\hat{\beta} \geq 2.5$	$\hat{\beta} \geq 3$
<b>A</b>							
HCV screen							
Hit number	996	737	541	283	127	76	23
Hit rate	0.0344	0.0254	0.0187	0.0098	0.0044	0.0026	0.0008
Empirical FNR	0.0026	0.0039	0.0052	0.0052	0.0084	0.0110	0.0238
Empirical FPR	0.0728	0.0554	0.0451	0.0238	0.0116	0.0077	0.0039
Hit rate point	0.2074	0.2386	0.2720	0.3419	0.4615	0.5178	0.7576
<b>B</b>							
Mucin screen							
Hit number	1028	860	698	483	289	212	113
Hit rate	0.0458	0.0383	0.0311	0.0215	0.0129	0.0094	0.0050
Empirical FNR	0.0095	0.0174	0.0285	0.0665	0.1329	0.1899	0.3956
Empirical FPR	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016
Hit rate point	0.3017	0.3677	0.4260	0.5632	0.7391	0.8328	1.1727

*SSMD*, strictly standardized mean difference; CBEL, completely balanced error level; HTS, high-throughput screening; FNL, false-negative level; HCV, hepatitis C virus; FPR, false-positive rate; FNR, false-negative rate.

**Table 4.** The FNL, RFPL, and Hit Rate Point in the Application of the SSMD-Based Ranking Strategy for Hit Selection in Both the HCV and the Mucin HTS Primary Screens

Number of Hits Based on Ranking SSMD Estimates	1000	800	600	500	400	300	200	100
<b>A</b>								
HCV screen								
Threshold of estimated SSMD value	$\hat{\beta} \geq 1.4961$	$\hat{\beta} \geq 1.5872$	$\hat{\beta} \geq 1.7071$	$\hat{\beta} \geq 1.7780$	$\hat{\beta} \geq 1.8526$	$\hat{\beta} \geq 1.9661$	$\hat{\beta} \geq 2.1288$	$\hat{\beta} \geq 2.4171$
FNL	0.0171	0.0233	0.0343	0.0426	0.0530	0.0727	0.1099	0.2058
RFPL								
$c_2 = 0$	0.0175	0.0127	0.0081	0.0061	0.0045	0.0028	0.0014	0.0003
$c_2 = 0.25$	0.0396	0.0298	0.0201	0.0157	0.0120	0.0078	0.0041	0.0011
$c_2 = 0.5$	0.0803	0.0628	0.0445	0.0359	0.0284	0.0195	0.0109	0.0035
$c_2 = 1$	0.2424	0.2041	0.1596	0.1366	0.1149	0.0868	0.0559	0.0230
$c_2 = 1.645$	0.5830	0.5324	0.4651	0.4257	0.3850	0.3255	0.2478	0.1384
$c_2 = 2$	0.7611	0.7195	0.6600	0.6227	0.5822	0.5190	0.4280	0.2784
$c_2 = 3$	0.9829	0.9767	0.9657	0.9574	0.9470	0.9273	0.8901	0.7942
Hit rate point	0.2050	0.2259	0.2591	0.2769	0.2885	0.3236	0.3805	0.4990
<b>B</b>								
Mucin screen								
Threshold of estimated SSMD value	$\hat{\beta} \geq 1.5165$	$\hat{\beta} \geq 1.6617$	$\hat{\beta} \geq 1.8347$	$\hat{\beta} \geq 1.9672$	$\hat{\beta} \geq 2.0952$	$\hat{\beta} \geq 2.3009$	$\hat{\beta} \geq 2.5476$	$\hat{\beta} \geq 3.1487$
FNL	0.0183	0.0297	0.0504	0.0729	0.1013	0.1624	0.2620	0.5830
RFPL								
$c_2 = 0$	0.0163	0.0096	0.0049	0.0028	0.0016	0.0006	0.0002	0.0000
$c_2 = 0.25$	0.0372	0.0234	0.0128	0.0078	0.0047	0.0019	0.0006	0.0000
$c_2 = 0.5$	0.0761	0.0509	0.0301	0.0194	0.0123	0.0056	0.0020	0.0001
$c_2 = 1$	0.2335	0.1757	0.1199	0.0866	0.0615	0.0335	0.0146	0.0012
$c_2 = 1.645$	0.5718	0.4906	0.3946	0.3249	0.2630	0.1777	0.1018	0.0171
$c_2 = 2$	0.7521	0.6831	0.5920	0.5184	0.4467	0.3358	0.2203	0.0528
$c_2 = 3$	0.9817	0.9703	0.9496	0.9271	0.8987	0.8376	0.7380	0.4170
Hit rate point	0.3091	0.3808	0.4633	0.5407	0.6033	0.7278	0.8652	NA

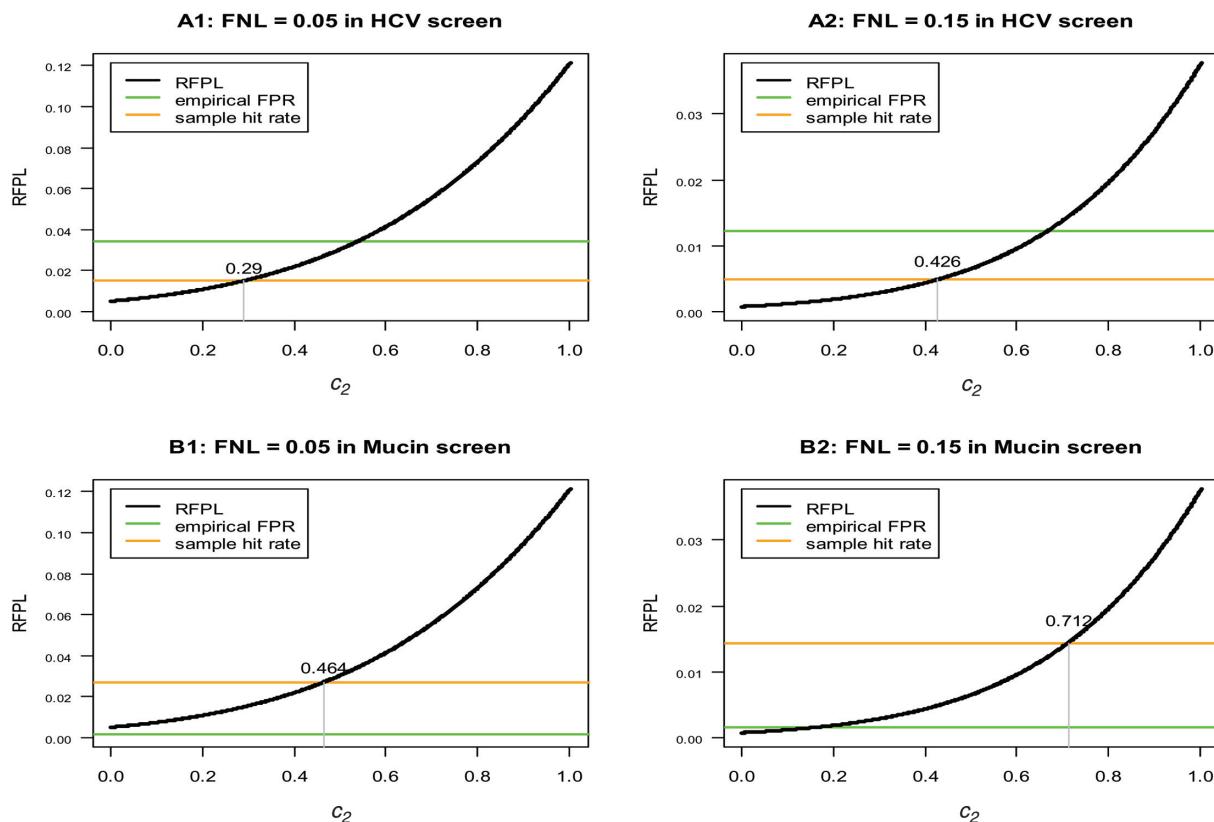
SSMD, strictly standardized mean difference; HTS, high-throughput screening; FNL, false-negative level; RFPL, restricted false-positive level; HCV, hepatitis C virus; NA, not applicable.

with a flexible and balanced control of both false negatives and false positives. In this article, through the application in 2 primary RNAi HTS experiments, we evaluated the use of *SSMD* for hit selection in both ranking and testing strategies and compared *SSMD* with the current commonly used ranking metrics and testing methods.

The use of percentage inhibition to rank siRNAs relies on the assumption that the same percentage inhibition values represent the same siRNA potency throughout the entire screen; that is, a value of 40 in percentage inhibition in 1 plate has the same meaning as a value of 40 in another plate. However, analyzing the data from 2 siRNA HTS experiments showed that this assumption did not hold for percentage inhibition even if we used the majority of sample wells as a negative reference. This is because percentage inhibition is strongly affected by the errors in not only the negative controls but also the positive controls. It has been observed early<sup>13</sup> that positive controls and negative controls in HTS are often unreliable due to positional effects or well-to-well variations in the assay. Consequently, selecting hits based on ranking siRNAs by percentage inhibition may be dominated by the poor performance of positive controls and/or negative controls.

As we saw in the HCV siRNA HTS experiment, based on percentage inhibition, the number of hits per plate in plates 1 to 9 was about 4 (or 7) times that in plates 10 to 78 and 89 to 97, and the number of hits per plate in plates 79 to 88 was about a half of that in plates 10 to 78 and 89 to 97. This observation might have made us to conclude that there were clusters of strong inhibition hits in plates 1 to 9 and many fewer hits in plates 79 to 88. However, by comparing this result with the behavior of positive controls and negative controls in Figure 2, we knew that this result was not caused by the different distribution of hits but by the irregular performance of positive and negative controls in these 2 sets of plates. We obtained similarly misleading results by ranking siRNAs by percentage inhibition for hit selection in the mucin screen.

When we used *SSMD* ranking for hit selection in the HCV siRNA HTS experiment, the number of hits per plate in plates 1 to 9 is nearly the same as (only slightly less than) in plates 10 to 78 and 89 to 97, and the number of hits per plate in plates 79 to 88 is about the same as in plates 10 to 78 and 89 to 97. These results, based on ranking *SSMD*, were matched with the observation obtained by both the values of sample wells and the comparison of the negative control wells and sample wells in



**FIG. 5.** Hit rate points, labeled by the numbers right above the intersections between the sample hit rate line and the restricted false-positive level (RFPL) curve in 2 cases in the hepatitis C virus (HCV) and mucin primary high-throughput screening (HTS) screens. Panels **A1** and **A2** display the results for the cases of  $c = 3$ , false-negative level (FNL) = 0.05 and  $c = 3$ , FNL = 0.15, respectively, in the HCV screen, and panels **B1** and **B2** display the results for the cases of  $c = 3$ , FNL = 0.05 and  $c = 3$ , FNL = 0.15, respectively, in the mucin screen. FPR, false-positive rate.

**Figure 1A.** Similarly, we obtained more sensible hit selection results using the ranking of SSMD than percentage inhibition in the mucin siRNA HTS experiment. Therefore, for hit selection using the ranking strategy, both SSMD and *z*-score work better than percentage inhibition in primary RNAi HTS experiments.

For the siRNAs without any replicate, as in most primary siRNA HTS experiments, the SSMD value has a linear relationship to the corresponding *z*-score value—namely,  $Z_{\text{obs}} = \sqrt{2}\hat{\beta}$ , if both are calculated using the same negative reference. Therefore, in terms of ranking siRNA effects, the SSMD method is equivalent to the *z*-score method in the primary screens. However, the *z*-score is aimed at addressing whether the mean of any values from an siRNA is the same as the mean of the values from the negative reference. It does not directly measure the strength of siRNA effects. The *z*-score is a function of both sample size and the strength of siRNA effects. Hence, the same *z*-score may be associated with different effects in different experiments when they have different sample sizes. By contrast, the SSMD value is not affected by the sample size and has consistent meanings in different experiments. The SSMD value has a clear

meaning (namely, the ratio of mean to SD of the difference) in measuring the strength of siRNA effects and has a clear probability meaning, denoted by *d*<sup>+</sup>-probability<sup>11,12</sup>; thus, it can be used to classify the strength of siRNA effects.

For hit selection using the testing strategy, we applied both the SSMD-based testing method and the *z*-score-based testing method in the 2 siRNA HTS primary screens. A good feature of the SSMD-based testing approaches is that, due to the ability of SSMD to measure the strength of siRNA effects, we can control both the false-negative rate, in which an siRNA with strong effect is not selected as a hit, and the restricted false-positive rate, in which an siRNA with weak or no effect is selected as a hit. To do so, we can specify a large value  $c$  and a small value  $c_2$  of SSMD and then maintain a flexible and balanced control of both the FNR, in which the siRNAs with  $\beta \geq c$  are not selected as hits, and the restricted false-positive rate, in which the siRNAs with  $\beta \leq c_2$  are selected as hits.

When scientists want a low FNL, an FNL of 0.01 or 0.025 may be a good choice. An SSMD-based decision rule with an FNL of 0.025 can ensure that the false-negative rate is no more

than 0.025. That is, if there are 200 siRNAs with  $\beta \geq 3$  in a screen, we should miss no more than 5 of them on average when we apply the *SSMD*-based decision rule with an FNL of 0.025 for hit selection. However, the false-positive rate may be up to 0.0117, even if no siRNAs have inhibition effects. That is, if all the 22,000 siRNAs have no inhibition effects, the *SSMD*-based decision rule with an FNL of 0.025 will pick up 257.4 siRNAs as hits on average. Using this decision rule, we selected 767 and 883 inhibition siRNAs as hits in the HCV screen and the mucin screen, respectively (panel I of **Table 3**), for further confirmatory screens. On the other hand, scientists may want a small number of hits with a low false-positive rate, such as in the selection of hits for preliminary gene function analysis. Based on the CBEL method, we selected 127 (and 289) inhibition siRNAs in the HCV (and mucin) HTS experiment. We also used the *SSMD* ranking method to select 200 inhibition siRNAs in both the HCV and mucin HTS experiments, which had an FNL of 0.1099 in the HCV experiment and 0.262 in the mucin experiment. Therefore, the *SSMD*-based method allows scientists to specify the strength of siRNA effects that they want to control and then to maintain a balanced control of both false negatives and false positives, depending on different scientific needs of siRNA HTS experiments.

By contrast, the *z*-score method only addresses whether an siRNA has the same mean as the negative reference. The *z*-score method hardly allows the specification of siRNA effects of interest due to the fact that the *z*-score is affected by both siRNA effects and sample size. It is nontrivial to obtain the false-negative rate in the *z*-score method while we can readily obtain the false-negative rate and false-positive rate in the *SSMD*-based testing method, as shown in **Tables 2** to **4**. For example, using a *z*-score  $\geq 2$ , there were 1224 (and 1187) inhibition siRNAs in the HCV (and the mucin) experiment (**Table 2**). We know that the false-positive rate of using a *z*-score  $\geq 2$  for hit selection was 0.025 under the normality assumption. However, it is nontrivial to calculate the false-negative rate in which the siRNAs with large effects are not selected as hits because we need to investigate the impact of both the siRNA effects and the sample size.

Therefore, the research in this article demonstrates in practice that the *SSMD* method addresses scientific questions and fills scientific needs better than both percentage inhibition and the *z*-score method. In these 2 primary RNAi HTS experiments, most of the sample siRNA wells (namely, the middle 98% values) were used as a negative reference to calculate *SSMD*. When the negative control wells work effectively, *SSMD* can also be calculated using the negative control wells as a negative reference. When *SSMD* is used as a ranking metric, the sample size may not be an issue. However, the *SSMD*-based testing methods are based on the condition of large sample size in the negative reference (namely,  $n_2 \geq 50$ ). Further research should be conducted to investigate the *SSMD*-based testing method when the sample sizes in both compared groups are small, such as in a confirmatory screen.

In this article, we applied the *SSMD* methods for hit selection in RNAi high-throughput screens. Besides RNAi screens, there are other screens, including enzyme, receptor, and cellular function assays. In an enzymatic assay, a signal is generated by processing a substrate into a product that is detected; in a receptor binding assay, the binding to a receptor is measured using a labeled ligand; and in a cellular function assay, a pathway is stimulated to lead to the transcription and expression of a reporter protein such as luciferase or  $\beta$ -lactamase. Screens generally measure changes in the signal generated when compounds (or siRNAs) are added, either by a reduction of the signal (e.g., enzyme inhibitors or receptor ligands) or an increase in signal (e.g., agonists in cellular functional assays). In fact, the HCV screen described in this article was first developed for the HTS of a compound collection,<sup>1</sup> producing similar data distribution, and the statistical methods used here could thus be applied for the analysis of data from the compound screens. Therefore, although the examples we provide describe the use of the *SSMD* methods for hit selection in RNAi-based high-throughput screens, these methods should be generally applicable to any assay where the endpoint is a difference in signal compared to a reference sample, including RNAi, enzyme, receptor, and cellular function assays.

## ACKNOWLEDGMENT

The authors thank Drs. Soper and Hazuda for their support in this research and the anonymous reviewers for their valuable comments.

## REFERENCES

1. 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.
2. MacKeigan JP, Murphy LO, Blenis J: Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. *Nat Cell Biol* 2005;7:591-600.
3. Nybakken K, Vokes SA, Lin TY, McMahon AP, Perrimon NA: Genome-wide RNA interference screen in *Drosophila melanogaster* cells for new components of the Hh signalling pathway. *Nat Genet* 2005;37:1323-1332.
4. Pelkmans L, Fava E, Grabner H, Hannus M, Habermann B, Krausz E, et al: Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* 2005;436:78-86.
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. Espeseth AS, Huang Q, Gates AT, 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.
7. Matson RS: *Applying Genomic and Proteomic Microarray Technology in Drug Discovery*. Boca Raton, FL: CRC Press, 2005.
8. Zhang J-H, 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.

## Use of Strictly Standardized Mean Difference for Hit Selection

9. Cohen J: The earth is round ( $p < 0.05$ ). *Am Psychol* 1994;49:997-1003.
10. Kirk R: Practical significance: a concept whose time has come. *Educ Psychol Meas* 1996;56:746-759.
11. Zhang XD: A pair of new statistical parameters for quality control in RNA interference high throughput screening assays. *Genomics*. In press.
12. Zhang XD: A new method with flexible and balanced controls of false negatives and false positives for hit selection in RNA interference high throughput screening assays. Submitted to *J Biomol Screen*.
13. 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.
14. 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.
15. Kevorkov D, Makarenkov V: Statistical analysis of systematic errors in high-throughput screening. *J Biomol Screen* 2005;10:557-567.
16. Malo N, Hanley JA, Cerqueira S, Pelletier J, Nadon R: Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006;24:167-175.
17. Zhang XD, Yang XC, Chung N, Gates AT, Stec E, Kunapuli P, et al: Statistical analysis in RNA interference high throughput screening experiments. Paper presented at The First RNAi Meeting, Cold Spring Harbor, NY, November 2005.
18. Zhang XD, Yang XC, Chung N, Gates AT, Stec E, Kunapuli P, et al: Exploring statistical methods for hit selection in RNA interference high throughput screening experiments. *Am Stat Assoc Proc* 2005;723-729.
19. Zhang XD, Yang XC, Chung N, Gates AT, Stec E, Kunapuli P, et al: Robust statistical methods for hit selection in RNA interference high throughput screening experiments. *Pharmacogenomics* 2006;7:299-309.

Address reprint requests to:

Xiaohua Douglas Zhang  
Merck Research Laboratories  
Biometrics Research, WP53B-120  
PO Box 4  
West Point, PA 19486

E-mail: xiaohua\_zhang@merck.com