The Use of SSMD-Based False Discovery and False Nondiscovery Rates in Genome-Scale RNAi Screens

XIAOHUA DOUGLAS ZHANG, RAUL LACSON, RUOJING YANG, SHANE D. MARINE, ALEX McCAMPBELL, DAWN M. TOOLAN, TIM R. HARE, JOOLEN KAJDAS, JOEL P. BERGER, DANIEL J. HOLDER, JOSEPH F. HEYSE, and MARC FERRER

In genome-scale RNA interference (RNAi) screens, it is critical to control false positives and false negatives statistically. Traditional statistical methods for controlling false discovery and false nondiscovery rates are inappropriate for hit selection in RNAi screens because the major goal in RNAi screens is to control both the proportion of short interfering RNAs (siRNAs) with a small effect among selected hits and the proportion of siRNAs with a large effect among declared nonhits. An effective method based on strictly standardized mean difference (SSMD) has been proposed for statistically controlling false discovery rate (FDR) and false nondiscovery rate (FNDR) appropriate for RNAi screens. In this article, the authors explore the utility of the SSMD-based method for hit selection in RNAi screens. As demonstrated in 2 genome-scale RNAi screens, the SSMD-based method addresses the unmet need of controlling for the proportion of siRNAs with a small effect among selected hits, as well as controlling for the proportion of siRNAs with a large effect among declared nonhits. Furthermore, the SSMD-based method results in reasonably low FDR and FNDR for selecting inhibition or activation hits. This method works effectively and should have a broad utility for hit selection in RNAi screens with replicates. (Journal of Biomolecular Screening 2010:1123-1131)

Key words: false discovery rate, false nondiscovery rate, q*-value, p*-value, RNAi, high-throughput screening

INTRODUCTION

G

ENOME-SCALE RNA INTERFERENCE (RNAi) SCREENS have been widely used to investigate gene functions and to discover drug targets. The success of RNAi screens relies on the selection of short interfering RNAs (siRNAs) with a desired size of inhibition/activation effects. For hit selection in genome-scale RNAi screens, it is critical to control false positives and false negatives. Traditional definitions of false positives and false negatives are inappropriate for hit selection in RNAi screens because they come from testing zero average effects, whereas many siRNAs may have very small but real nonzero average effects. To address for deficiencies in the application of traditional approaches to the identification of false positives and false negatives in RNAi screening, new definitions have been proposed: false positives are defined to be the siRNAs with a small effect among selected hits, and false negatives are defined to be the siRNAs with a large effect among declared nonhits. Based on the proposed definitions, a new statistical method based on strictly standardized mean difference (SSMD) has been put forth for controlling false discovery and false nondiscovery rates from a theoretical basis.

In this article, we explore how to apply the newly proposed SSMD-based method for hit selection using actual genome-scale RNAi screening data. A side-by-side comparison is made between results using the SSMD-based method and those using the traditional methods. The false positives in an experiment are controlled by the false-positive rate (FPR) and/or false discovery rate (FDR). The false negatives are controlled by the false-negative rate (FNR) and/or false nondiscovery rate (FNDR). As pointed out by Storey and Tibshirani, FPR and FDR are often mistakenly equated, but their difference is actually very important. So are FNR and FNDR. Zhang proposes p*-value and q*-value to control FNR and FNDR, respectively, and provides the calculation of SSMD-based p*-value, p*-value, q*-value, and q*-value from a theoretical basis. In this article, we examine actual genome-scale RNAi data to illustrate the difference between FDR and FPR as well as that between FNDR and FNR. In particular, we demonstrate how to use SSMD-based q*-value
and $q^*$-value to select inhibition hits in a primary siRNA screen for diabetes drug targets as well as how to select activation hits in a confirmatory screen for neurological disease drug targets.

**METHODS AND MATERIALS**

*An siRNA primary screen to identify novel targets for diabetes*

A genome-wide RNAi screen was conducted to identify modulators of glucose output using a multiplexed array gene expression assay. A 7000-gene siRNA library representing the druggable genome was tested using pools of 3 siRNAs for each target in 4 replicates. A human hepatoma cell line (PLC PRF 5) was transfected with siRNA and treated with dexamethasone, cAMP, and a suboptimal dose of insulin. After 48 h, a 4-gene quantitative nucleic acid protection assay (qNPa; High Throughput Genomics, Tucson, AZ) in a 384-well format was used to measure the gene expression of beta-actin, glucose-6-phosphatase (G6PC), and pyruvate dehydrogenase kines 4 (PDK4) (in addition to a negative control gene). siRNAs that modulate the gene expression of these readouts were analyzed and tested for their ability to modulate gluconeogenesis and insulin sensitivity. The major goal in this screen was to select siRNAs with inhibition effects to identify potential diabetes drug targets.

*An siRNA confirmatory screen for neurological disease targets*

Following a primary screen without replicates in which a total of 23,653 pools of 3 siRNAs were tested, 960 pools of siRNAs were selected for further investigation in a confirmatory screen. Most of these 960 siRNAs showed activation activity in the primary screen. This confirmatory siRNA screen was carried out to identify genes involved in the regulation of ABCA1 levels by enzyme-linked immunosorbent assay (ELISA). siRNAs that substantially altered total ABCA1 levels in H4 cells were considered hits. The 960 pools of siRNAs were arranged in 3 source plates; 4 experimental plates were generated from each source plate. The goal in the above confirmatory screen was to select activation hits to identify potential neurological disease targets.

**SSMD-based methods controlling false discovery and false nondiscovery rates**

In RNAi screens with replicates, there are several sets of source plates. Each set is unique and is repeated $n$ times, and thus each siRNA has $n$ replicates. Because plate-to-plate variability is usually much higher than within-plate variability, we calculated the difference (i.e., $D$) between the measured intensity of an siRNA and average intensity of a negative control in the plate where this siRNA is located. Then we calculated the average difference (i.e., $D$) of all replicates for this siRNA—namely, $D = \frac{1}{n} \sum_{i=1}^{n} (D_{i} - \bar{D})$ and the corresponding sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_{i} - \bar{D})^2$. Let $\beta$ denote SSMD. Then an unbiased estimate of SSMD is $\hat{\beta} = \frac{k}{\sqrt{n-1} s_0}$, where $k = 2 \left( \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} \right)$ and $\Gamma(\cdot)$ is a gamma function. The classical $t$-statistic for testing no mean difference here (i.e., paired $t$-test) is $t = \sqrt{n} \frac{\bar{D}}{s_0}$. One difference between the classical $t$-statistic and the SSMD estimate is the latter is more robust to sample size than the former. See Zhang23 for more details on how these methods are developed.
RESULTS

A diabetes target identification siRNA primary screen

Based on traditional methods of testing $H_0$: mean difference $\geq 0$ for selecting inhibition hits, the $p$-value, $q$-value, $p^*$-value, and $q^*$-value for each siRNA in the diabetes siRNA target identification screen are represented by black, red, blue, and green points, respectively, in Figure 1. In general, one controls FPR (or FDR) by establishing a specified level—say, FPR = 0.05 (or, say, FDR = 0.2)—in hit selection. That is, select all the siRNAs with an FPR-associated $p$-value less than 0.05 (or with an FDR-associated $q$-value less than 0.2) as hits. If we do so, the resulting FNR and FNDR are above 0.9 and 0.85, respectively. Therefore, using the traditional concepts of false positives and false negatives, both of these latter values are undesirably high (Fig. 1A). The reason is that there are many siRNAs with small but real effects among nonhits, and all these siRNAs will be treated as false negatives. In reality, we can tolerate the inclusion of some siRNAs with small or even moderate inhibition effects in the list of nonhits. Thus, the false negatives should be defined as the siRNAs with a large inhibition effect rather than those with a very small inhibition effect. This concept of false negatives requires a metric to assess effect size.

The $t$-value from traditional $t$-tests is dominated by sample size: the larger the sample size, the larger the absolute value of $t$-value; thus, it cannot be used to assess effect size. Mean difference (or, equivalently, average fold change) is a statistical parameter that is robust to sample size: its estimated value approaches its true value when sample size gets larger. However, because mean difference does not contain information about data variability (variance), the same value of mean difference may correspond to different $p$-values even when the sample size is the same for all siRNAs. For example, an siRNA with an average fold change of about 1.5 in the inhibition direction may have an FPR ranging from 0 to 0.2 (black points in Fig. 1B), an FNR from 0.8 to 1 (blue points in Fig. 1B), an FDR from 0 to 0.3 (red points in Fig. 1C), and an FNR from 0.75 to 0.8 (green points in Fig. 1C). Meanwhile, an siRNA with a $p$-value of about 0.05 may have an average fold change ranging from 1.1 to 3 (black points in Fig. 1B), and an siRNA with a $q$-value of about 0.1 may have an average fold change ranging from 1.1 to 3 (red points in Fig. 1C). Thus, neither $t$-value nor average fold change can be used to assess effect sizes.

SSMD is the ratio of mean to standard deviation of the difference between an siRNA and a negative reference group. Thus, SSMD is robust to sample size similar to mean difference and contains information about data variability similar to $t$-value. There is a meaningful and interpretable SSMD-based criterion for classifying the size of siRNA effects. Thus, we can use SSMD to define false negatives and false positives as shown in Zhang. That is, for selecting inhibition hits, the false negatives are defined as the siRNAs with true SSMD values of less than $\beta_1$, where $\beta_1$ is a negative value such as −2 or −3 among declared nonhits. The false positives are defined as the siRNAs with true SSMD values greater than $\beta_2$, where $\beta_2$ is a negative value such as 0 or −0.25 among selected hits. Based on the calculation of the corresponding FPR, FNR, FDR, and FNDR provided in Zhang, we can obtain a $q$-value, $q^*$-value,
When using SSMD for selecting inhibition hits in RNAi screens, we normally use the following decision rule: an siRNA is selected as a hit if it has an estimated SSMD value less than or equal to a critical value; it is considered a nonhit otherwise. The application of this decision rule relies on the determination of a critical value that is normally achieved through the consideration of FDR and FNDR. Table 1 shows the FDRs and FPRs with respect to extremely weak inhibition effects or no inhibition (i.e., \( \beta_2 = -0.25 \)), as well as the FNDRs and FNRs with respect to very strong inhibition effects or stronger inhibition (i.e., \( \beta_2 = -3 \)).

![Diagram](image_url)

**Table 1.** FDR and FNDR for Selecting Inhibition Hits among 6613 siRNAs under Investigation in a Primary siRNA Screen with 4 Replicates for Diabetes

<table>
<thead>
<tr>
<th>Potential Critical Value of SSMD</th>
<th>( \beta_2 = -0.25 )</th>
<th>( \beta_2 = -3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FDR</td>
<td>FPR</td>
</tr>
<tr>
<td>–0.5</td>
<td>0.4075</td>
<td>0.2492</td>
</tr>
<tr>
<td>–0.75</td>
<td>0.2717</td>
<td>0.1348</td>
</tr>
<tr>
<td>–1</td>
<td>0.1773</td>
<td>0.0762</td>
</tr>
<tr>
<td>–1.28</td>
<td>0.1185</td>
<td>0.0432</td>
</tr>
<tr>
<td>–1.645</td>
<td>0.0819</td>
<td>0.0230</td>
</tr>
<tr>
<td>–2</td>
<td>0.0665</td>
<td>0.0136</td>
</tr>
<tr>
<td>–3</td>
<td>0.0506</td>
<td>0.0044</td>
</tr>
</tbody>
</table>

FDR, false discovery rate; FNDR, false nondiscovery rate; FPR, false-positive rate; FNR, false-negative rate; SSMD, strictly standardized mean difference.

**FIG. 2.** The \( q \)-values, \( q^* \)-values, \( p \)-values, and \( p^* \)-values with respect to (shortened as “wrt” in the figure legends) various true values of SSMD (A, B) in a primary diabetes siRNA target identification screen with 4 replicates.
effects (i.e., $\beta_1 = -3$) for 7 potential critical values of $-0.5$, $-0.75$, $-1$, $-1.28$, $-1.645$, $-2$, and $-3$ in the diabetes siRNA screen.

The error rates in Table 1 can be used to demonstrate the difference between FDR and FPR, as well as between FNDR and FNR. For example, the use of the critical value of $-1.28$ leads to the selection of 1371 inhibition hits. The corresponding FDR with respect to extremely weak or no inhibition effects (i.e., $\beta_2 = -0.25$) is 0.1185, which indicates that among the 1371 selected hits, on average 162 (1371×0.1185) have extremely weak inhibition effects or no inhibition. By contrast, the critical value of $-1.28$ leads to an FPR of 0.0432. This FPR indicates that, assuming all the 6613 siRNAs included in the experiment have extremely weak or no inhibition effects, the critical value of $-1.28$ will lead to the selection of 286 (6613×0.0432) inhibition hits on average due to chance alone. Clearly, the FPR does not have meaning directly related to the 1371 selected inhibition hits, especially taking into consideration the insupportable assumption that all the 6613 siRNAs included in the experiment have extremely weak effects or no inhibition effects.

Similarly, the critical value of $-1.28$ leads to an FNDR of 0.0143 with respect to very strong inhibition effects or stronger inhibition effects, as well as an FNR of 0.0609 with respect to extremely weak effects or no inhibition. The FNDR of 0.0143 indicates that among the 5242 (6613 − 1371) declared nonhits for inhibition, on average 75 (5242×0.0143) have very strong effects or stronger effects. By contrast, the FNR of 0.0609 indicates that, assuming all the 6613 siRNAs included in the experiment have a very strong inhibition effect or stronger inhibition effects, the critical value of $-1.28$ will lead to the identification of 403 (6613×0.0609) siRNAs as nonhits on average due to chance alone. Again, the FNR does not answer the question of interest about the proportion of siRNAs with very strong effects or stronger effects among the declared nonhits, whereas FNDR does. In addition, the assumption that all the 6613 siRNAs included in the experiment have very strong inhibition effects or stronger inhibition effects is not supported. Therefore, FDR and FNDR can address important questions of interest in the process of hit selection in RNAi screens better than FPR and FNR. Consequently, one should focus on the control of FDR and FNDR through $q$-value and $q^*$-value rather than the control of FPR and FNR through $p$-value and $p^*$-value in genome-scale RNAi screens.

From Table 1, the use of the critical value of $-0.5$ leads to the selection of 2482 inhibition hits. The corresponding FDR with respect to extremely weak effects or no inhibition (i.e., $\beta_2 = -0.25$) is 40.75%, and the corresponding FNDR with respect to very strong inhibition effects or stronger inhibition effects (i.e., $\beta_1 = -3$) is 8.2×10^{-4}. That is, among the 2482 selected hits, on average 1011 have extremely weak inhibition effects or no inhibition; among the 4131 nonhits, on average none have very strong inhibition effects or stronger inhibition effects. Clearly, the number of false positives resulting from the critical value of $-0.5$ is too high. On the other hand, the use of the critical value of $-3$ leads to the selection of 296 inhibition hits. The corresponding FDR with respect to extremely weak inhibition effects or no inhibition is 5.06%, and the corresponding FNDR with respect to strong inhibition effects or stronger inhibition effects is 12.14%. That is, among the 296 selected hits, on average 15 have extremely weak inhibition effects or no inhibition; among the 6317 nonhits, on average 767 have very strong inhibition effects or stronger inhibition effects. Clearly, the number of false negatives resulting from the critical value of $-3$ is too high. Following this logic, we can determine that the critical values of $-1$, $-1.28$, and $-1.645$ all lead to reasonable FDRs and FNDRs.

Besides FDR and FNDR, another consideration that needs to be taken into account when planning a primary RNAi screen is the capacity to carry out follow-up confirmatory assays on leads from the primary screen. Confirmatory capacity might typically range from 300 to 2000 siRNAs. With this consideration, we use the critical value of $-1.645$ for selecting inhibition hits, which leads to the selection of 976 inhibition hits. From Table 1, on average 80 of these 976 selected hits (i.e., 8.19%) have extremely weak inhibition effects or no inhibition. Meanwhile, on average 225 of the 5637 nonhits for inhibition have very strong inhibition effects or stronger inhibition effects. The value of $-1.645$ also has a reasonable probabilistic meaning. That is, if the estimated SSMD value equals the true SSMD value, the probability that one value randomly generated from an siRNA is greater than another value from the negative reference (i.e., $d^*$-probability) is 0.05 under normal assumption.

Similarly, we use the following decision rule to select activation hits: an siRNA is selected as a hit if it has an estimated SSMD value greater than or equal to a specified critical value; it is considered a nonhit otherwise. Based on the calculation provided in Zhang,23 we obtain the FDRs and FNDRs for 7 potential critical values in the activation direction, as shown in Table 2. From Table 2, the use of the critical value of 1.645 for selecting activation hits leads to the selection of 905 activation hits. The corresponding FDR with respect to extremely weak activation effects or no activation (i.e., $\beta_2 = 0.25$) is 17.75%, and the FNDR with respect to very strong activation effects or stronger activation effects (i.e., $\beta_1 = 3$) is 2.07%. That is, among the 905 selected activation hits, on average 161 have extremely weak activation effects or no activation; among the 5708 nonhits for activation, on average 118 have very strong activation effects or stronger activation effects.

An siRNA confirmatory screen for neurological disease targets

In the confirmatory screen for neurological disease targets described in the Materials and Methods section, the goal is to
Table 2. FDR and FNDR for Selecting Activation Hits among 6613 siRNAs under Investigation in a Primary siRNA Screen for Diabetes Drug Targets

<table>
<thead>
<tr>
<th>Potential Critical Value of SSMD</th>
<th>FDR</th>
<th></th>
<th>FNDR</th>
<th></th>
<th>Number of Selected Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_2 = 0$</td>
<td>$\beta_2 = 0.25$</td>
<td>$\beta_1 = 2$</td>
<td>$\beta_1 = 3$</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.3212</td>
<td>0.5999</td>
<td>0.0030</td>
<td>$4 \times 10^{-6}$</td>
<td>2512</td>
</tr>
<tr>
<td>0.75</td>
<td>0.1960</td>
<td>0.4477</td>
<td>0.0169</td>
<td>0.00017</td>
<td>2081</td>
</tr>
<tr>
<td>1</td>
<td>0.1351</td>
<td>0.3224</td>
<td>0.0458</td>
<td>0.0016</td>
<td>1643</td>
</tr>
<tr>
<td>1.28</td>
<td>0.0967</td>
<td>0.2381</td>
<td>0.0857</td>
<td>0.0071</td>
<td>1257</td>
</tr>
<tr>
<td>1.645</td>
<td>0.0705</td>
<td>0.1775</td>
<td>0.1310</td>
<td>0.0207</td>
<td>905</td>
</tr>
<tr>
<td>2</td>
<td>0.0575</td>
<td>0.1466</td>
<td>0.1616</td>
<td>0.0361</td>
<td>653</td>
</tr>
<tr>
<td>3</td>
<td>0.0411</td>
<td>0.1065</td>
<td>0.2016</td>
<td>0.0684</td>
<td>289</td>
</tr>
</tbody>
</table>

FDR, false discovery rate; FNDR, false nondiscovery rate; SSMD, strictly standardized mean difference.

select activation hits. Accordingly, we need to search for a critical value via the following decision rule for selecting activation hits: an siRNA is selected as an activation hit if it has an estimated SSMD value greater than or equal to a specified critical value; it is declared a nonhit otherwise. The determination of this critical value mainly relies on the control of FDR and FNDR through $q$-value and $q^a$-value in Figure 3A.

From Figure 3A, a critical value between 0.75 and 1.645 may generate an acceptably low FDR with respect to extremely weak activation effects or no activation (i.e., $\beta_2 = 0.25$; red points), an acceptably low FDR with respect to no activation (i.e., $\beta_2 = 0$; black points), an acceptably low FNDR with respect to strong activation effects or stronger activation effects (i.e., $\beta_1 = 2$; green points), and an acceptably low FNDR with respect to very strong activation effects or stronger activation effects (i.e., $\beta_2 = 3$; blue points).

To determine the exact critical value for selecting activation hits, we need to know detailed information about the error rates and the number of selected hits generated by the decision rule in association with some potential critical values of 0.5, 0.75, 1, 1.28, 1.645, 2, and 3, which are shown in Table 3. From Table 3, the use of the critical value of 0.5 leads to the selection of 595 activation hits. The corresponding FDR with respect to extremely weak activation effects or no activation is 23.45%. That is, among the 595 selected hits, on average 140 have extremely weak activation effects or no activation. Clearly, the number of false positives resulting from the critical value of 0.5 is too high. On the other hand, the use of the critical value of 2 leads to the selection of 374 activation hits and a corresponding FNDR of 23.45% with respect to strong activation effects or stronger activation effects. That is, among the 586 declared nonhits, on average 137 have very strong activation effects or stronger activation effects. Thus, the number of false negatives resulting from the critical value of 2 is too high. Following this logic, the use of the critical value of 1.645 leads to the selection of 427 activation hits, and we would on average miss 77 siRNAs with very strong activation effects or stronger activation effects, which is still assessed as slightly too many.

The use of the critical value of 1.28 leads to the selection of 470 activation hits, an FDR of 0.0531 with respect to extremely weak activation effects, or no activation, as well as an FNDR of 0.0539 with respect to very strong activation effects or stronger activation effects (Table 3). That is, if we declare an siRNA an activation hit if it has an estimated SSMD value greater than or equal to 1.28 and a nonhit otherwise, then we will select 470 siRNAs as activation hits and declare the remaining 490 siRNAs as nonhits. Among the 470 selected as hits, on average 25 siRNAs have extremely weak activation effects or no activation; among the 490 declared nonhits, on average 26 siRNAs have very strong activation effects or stronger activation effects. Therefore, considering the associated reasonable false discovery rates and false nondiscovery rates, we choose to adopt the decision rule with the critical value of 1.28 for selecting activation hits in this screen. The value of 1.28 also has a reasonable probabilistic meaning: if the estimated SSMD value equals the true SSMD value, the probability that one value randomly generated from an siRNA is greater than another value from the negative reference is 0.90 under normal assumption. The 470 selected hits should be investigated in further analysis such as pathway analysis and deconvolution screening.

One point in the new method of controlling false discovery rates and false nondiscovery rates in hit selection in genome-scale RNAi screens is that we control the FDR with respect to extremely weak effects or no effect (i.e., with respect to $\beta_2 = 0.25$ for activation), whereas we control FNDR with respect to very strong effects or stronger effects (i.e., with respect to $\beta_1 = 3$ for activation), not with respect to extremely weak effects or no effect (i.e., with respect to $\beta_1 = 0.25$ for activation). This is important as illustrated below. For each critical value shown in Table 3, the corresponding FNDRs with respect to $\beta_1 = 0.25$ are...
all very high, ranging from 68.17% to 73.74%. That is, it is not feasible to achieve a low FNDR with respect to extremely weak effects or no effect. More important, since we observe that about 76% of the siRNAs have activation effects in the experiment, most of which would be expected to have a large or medium effect, it is meaningless to control FNDR with respect to extremely weak effects or no effect. Similarly, we can see that it is not feasible and arguably meaningless to attempt to control FNDR with respect to siRNAs that have no effect, that is, $\beta_1 = 0$ (Table 3).

### Table 3. FDR and FNDR for Selecting Activation Hits among 960 siRNAs under Investigation in a Confirmatory siRNA Screen with 4 Replicates for Diabetes

<table>
<thead>
<tr>
<th>Critical Value of SSMD</th>
<th>$\beta_1 = 0$</th>
<th>$\beta_1 = 0.25$</th>
<th>$\beta_1 = 2$</th>
<th>$\beta_1 = 3$</th>
<th>$\beta_1 = 0$</th>
<th>$\beta_1 = 0.25$</th>
<th>Number of Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.1172</td>
<td>0.2345</td>
<td>0.0116</td>
<td>0.00003</td>
<td>0.7377</td>
<td>0.6817</td>
<td>595</td>
</tr>
<tr>
<td>0.75</td>
<td>0.0625</td>
<td>0.1422</td>
<td>0.0647</td>
<td>0.00126</td>
<td>0.7514</td>
<td>0.7114</td>
<td>556</td>
</tr>
<tr>
<td>1</td>
<td>0.0340</td>
<td>0.0858</td>
<td>0.1698</td>
<td>0.01252</td>
<td>0.7573</td>
<td>0.7248</td>
<td>513</td>
</tr>
<tr>
<td>1.28</td>
<td>0.0216</td>
<td>0.0531</td>
<td>0.2851</td>
<td>0.0539</td>
<td>0.7602</td>
<td>0.7316</td>
<td>470</td>
</tr>
<tr>
<td>1.645</td>
<td>0.0125</td>
<td>0.0313</td>
<td>0.3886</td>
<td>0.1451</td>
<td>0.7618</td>
<td>0.7356</td>
<td>427</td>
</tr>
<tr>
<td>2</td>
<td>0.0082</td>
<td>0.0210</td>
<td>0.4479</td>
<td>0.2345</td>
<td>0.7626</td>
<td>0.7374</td>
<td>374</td>
</tr>
</tbody>
</table>

FDR, false discovery rate; FNDR, false nondiscovery rate; SSMD, strictly standardized mean difference.
When we use SSMD for controlling false discovery and false nondiscovery rates, we use one value such as 0 or 0.25 for controlling FDR and another value such as 3 or 2 for controlling FNDR. In context of mean fold change as a statistical parameter, it might be argued that one would want to consider the use of 2 values of mean fold change as well—for example, the use of a value of 1 or 1.2 of mean fold change for controlling FDR and another value of 2 for controlling FNDR. The corresponding formulas for calculating FDR and FNDR based on mean fold change in this adapted approach have been derived and provided by Zhang.23 The major issue with this approach is that there is no theoretical framework to set up the 2 constants for mean fold change. Another consideration is that the siRNAs with the same or similar observed value of mean fold change may have very different q*-values and very different q*-values (Fig. 3B). For example, the siRNAs with a value of about 1.5 for the estimated mean fold change may have an FDR ranging from 0 to 0.38 with respect to a true mean fold change of 1.2, as well as an FNDR ranging from 0 to 0.10, with respect to a true mean fold change of 2 (the gray vertical line in Fig. 3B).

DISCUSSION

One important goal in genome-scale RNAi screens is to select siRNAs with a specified magnitude of inhibition effect or activation effect, which usually requires statistical control of false positives and false negatives. As demonstrated in the real RNAi screen shown in Figure 1, traditional methods of controlling a low FPR or FDR lead to an unacceptably high FNR and FNDR. The traditional methods are inappropriate for hit selection in RNAi screens when the goal is to control for the proportion of siRNAs with a small effect among selected hits and to control for the proportion of siRNAs with a large effect among declared nonhits. To address for this need, a new method of controlling false positives and false negatives has been developed.23 In this article, we illustrate how to use this newly developed SSMD-based method to select inhibition hits in a primary siRNA screen for diabetes targets and to select activation hits in a confirmatory siRNA screen for neurological disease targets. We also demonstrate that this SSMD-based method results in reasonably low false discovery and false nondiscovery rates with respect to the specified effect sizes in genome-scale RNAi screens.

The differences between FDR and FPR and between FNR and FNDR are demonstrated in the diabetes target example shown in Figure 2 and Table 1. The FPR does not answer the question of interest about the proportion of siRNAs with extremely weak effects or no effects among the declared hits, whereas FDR does. Similarly, the FNR does not answer the question of interest about the proportion of siRNAs with very strong effects or stronger effects among the declared nonhits, whereas FNDR does. In addition, FPR relies on the assumption that all 6613 siRNAs included in the diabetes target RNAi experiment have extremely weak or no inhibition effects. FNR relies on the assumption that all 6613 siRNAs included in the experiment have very strong inhibition effects or stronger inhibition effects. Both assumptions are insupportable. By contrast, FDR and FNDR do not rely on either of these assumptions. Consequently, whenever possible, we should focus on the control of FDR and FNDR through q*-value and q*-value rather than the control of FPR and FNR through p-value and p*-value in genome-scale RNAi screens.

It should be noted that these newly developed methods result in not only a low FDR with respect to extremely weak effects or no effect but also a low FNDR with respect to very strong effects or stronger effects (not FNDR with respect to no effect, as in traditional methods). As demonstrated in Table 3, it is arguably meaningless and infeasible to control FNDR with respect to the no-effect case in a genome-scale RNAi screen (and especially so in a confirmatory screen). As guidance, the SSMD-based p-value, q-value, p*-value, and q*-value that are part of the proposed method require sample sizes (number of replicates) to be at least 3 and appear to work best when sample size is 4 or more.28

Traditional methods of controlling false positives and false negatives are largely based on testing for no mean difference (or, equivalently, average fold change being one), whereas the newly proposed methods are based on a different statistical parameter, SSMD. One may adapt the traditional method of testing mean difference to select hits in a similar fashion as in the SSMD-based method. That is, use one small value of true mean fold change to indicate a small effect and another larger value of true mean fold change to indicate a large effect and then control FDR with respect to the smaller value and FNDR with respect to the larger value of true fold change. The major issue with this approach is that, unlike SSMD, mean difference cannot effectively measure the size of siRNA effects because it cannot capture data variability. As described in Zhang,22 there exist theoretically based thresholds (such as −5, −3, −1, −.75, −.5, −.25, etc.) of the population values of SSMD for classifying siRNA effects. In contrast, there is no theoretical framework to suggest thresholds of average fold change for classifying siRNA effects. Another potential issue is that, as demonstrated in Figures 1C and 3B, the same value of observed mean fold change may correspond to different p-value and p*-value with respect to a single true value and with the same sample size, which is an undesirable feature of this approach to the process of controlling false positives and false negatives.

In summary, the traditional methods of controlling FDR and FNDR are inappropriate for hit selection in RNAi screens. A new method based on SSMD of controlling for false discovery and false nondiscovery rates has been proposed.23 In this article, we explore the utility of this SSMD-based method in
controlling FDR and FNDR for hit selection in RNAi screens. As demonstrated in 2 genome-scale RNAi screens, this SSMD-based method works effectively for hit selection and also addresses the unmet need of controlling the proportion of siRNAs with a small effect among selected hits and controlling the proportion of siRNAs with a large effect among declared nonhits.

ACKNOWLEDGMENT

The authors thank Drs. Soper, Bain, and Cleary for their support in this research, as well as Erica Stec and Francesca Santini for their discussions.

Conflict of interest statement. All the authors are employees of Merck Research Laboratories.

REFERENCES


Address correspondence to:
Xiaohua Douglas Zhang
Biometrics Research, WP53B-120
Merck Research Laboratories
West Point, PA 19486

E-mail: xiaohua_zhang@merck.com