Genome analysis

Determination of sample size in genome-scale RNAi screens

Xiaohua Douglas Zhang1,∗ and Joseph F. Heyse2

1Biometrics Research and 2BARDS, Merck Research Laboratories, West Point, PA 19486, USA

ABSTRACT

Motivation: For genome-scale RNAi research, it is critical to investigate sample size required for the achievement of reasonably low false negative rate (FNR) and false positive rate.

Results: The analysis in this article reveals that current design of sample size contributes to the occurrence of low signal-to-noise ratio in genome-scale RNAi projects. The analysis suggests that (i) an arrangement of 16 wells per plate is acceptable and an arrangement of 20–24 wells per plate is preferable for a negative control to be used for hit selection in a primary screen without replicates; (ii) in a confirmatory screen or a primary screen with replicates, a sample size of 3 is not large enough, and there is a large reduction in FNRs when sample size increases from 3 to 4. To search a tradeoff between benefit and cost, any sample size between 4 and 11 is a reasonable choice. If the main focus is the selection of siRNAs with strong effects, a sample size of 4 or 5 is a good choice. If we want to have enough power to detect siRNAs with moderate effects, sample size needs to be 8, 9, 10 or 11. These discoveries about sample size bring insight to the design of a genome-scale RNAi screen experiment.

Contact: Xiaohua_zhang@merck.com

1 INTRODUCTION

RNAi is a mechanism that inhibits gene expression with complementary nucleotide sequences of double-stranded RNA (Fire et al., 1998). The ‘functional’ component of RNAi is small interfering RNA (siRNA). RNAi offers an effective tool for silencing gene and has been seen as the third class of drug after small molecules and proteins. Genome-scale RNAi research relies on RNAi high-throughput screening (HTS) assays. Currently, an RNAi HTS experiment is usually conducted in 384-well plates in which the cells in a well are treated with a unique siRNA or a control although some experiments may use 96- or 1536-well plates. Hence, this article focuses on determining the sample size in experiments using 384-well plates. There are at least a positive control with specific knockdown effects and a negative control with unspecific knockdown effects due to siRNA induction of the innate cellular immunity in each plate. The measured response is usually the intensity emitted by labeled particles such as fluorescent dyes.

In RNAi HTS assays, one primary goal is to select siRNAs with a desired size of effect. The siRNA effect is represented by the magnitude of difference between the intensity of an siRNA and that of a negative reference in RNAi HTS experiments (Zhou et al., 2008). Recently, strictly standardized mean difference (SSMD) has been proposed to measure the magnitude of difference (Zhang, 2007b) and has been used to assess the size of siRNA effects (Lapan et al., 2008; Wiles et al., 2008; Zhang, 2007a, 2008a; Zhang et al., 2007; Zhou et al., 2008). Using SSMD, we may maintain a balanced control of both false positive rate and false negative rate (FNR) (Zhang, 2007a, 2008a; Zhang et al., 2007; Zhou et al., 2008), which relies on a reasonable sample size. In fact, it is critical to determine a sample size for the achievement of certain false negative level (FNL) and false positive level (FPL) in RNAi HTS assays.

Classical methods such as z-score method and t-test control the false positive rate of concluding no mean difference while actually there is any mean difference, even a tiny mean difference. However, in practice, what we are really interested in is whether an siRNA (or compound) has average inhibition/activation effects that are exactly the same as the negative reference. Instead, we are interested in the siRNAs or compounds with large effect sizes. Classical z-score method and t-test cannot effectively assess the size of siRNA effects whereas SSMD-based methods can [See Zhang et al. (2009) and Zhang (2007a, 2008a) for more detailed comparisons of SSMD methods and classical methods]; so we use SSMD-based methods to determine sample size in genome-scale RNAi screens.

The limitation of experimental time and cost usually does not allow a single experiment to have more than 200 384-well plates whereas 200 384-well plates is usually the minimal requirement for conducting a genome-wide screen with replicates. Therefore, currently a typical RNAi HTS project starts with a primary screen of about 20000 siRNAs most of which have no replicate. The siRNAs identified (called ‘hits’) in the primary screen are further investigated using one or more confirmatory screens in which each siRNA has replicates. A typical primary screen has 50–150 384-well plates and a typical confirmatory screen has 3–20 384-well plates. In the primary screen, we may use the negative control in a plate as a negative reference. The question is how many wells we should arrange for the negative control in a plate so that we can achieve a manageable false positive rate, while maintaining a reasonably low FNR. Currently, a negative control usually occupies 4, 8, 16, 20 or 24 wells per plate in a primary screen with 384-well plates. Are these numbers enough to obtain low false positive rate and FNR? A typical confirmatory screen usually has triplicates for each investigated siRNA. Are triplicates enough to achieve low false positive rate and FNR? In this article, we will investigate all these sample size issues in both primary and confirmatory screens.

2 METHODS AND RESULTS

2.1 Error rate calculation in RNAi HTS experiments

SSMD effectively measures the size of siRNA effects (Zhang, 2007a, 2008a; Zhang et al., 2007; Zhou et al., 2008). SSMD (denoted as β) is defined as

∗To whom correspondence should be addressed.
the ratio of mean to SD of the difference between the two groups, namely \( \beta = (\mu_1 - \mu_2)/\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \), where \( \mu_1, \mu_2, \sigma_1^2 \) and \( \sigma_2^2 \) are the means and variances in the two groups, respectively, and \( \sigma_{12} \) is the covariance between these two groups (Zhang, 2007b). If the two groups are independent then \( \beta = (\mu_1 - \mu_2)/\sqrt{\sigma_1^2 + \sigma_2^2} \). If the two independent groups have equal variance \( \sigma^2 \), then \( \beta = (\mu_1 - \mu_2)/\sqrt{2\sigma^2} \). SSMD can be interpreted as the ratio of mean to SD of the difference and it is linked to \( d' \)-probability, the probability that the difference is \( >0 \), through \( d' \)-probability = \( \Phi(\beta) \) under normality, where \( \Phi(\cdot) \) is the cumulative distribution function of the standard normal distribution (Zhang, 2008b). Because of these clear meanings of SSMD, we can construct meaningful and interpretable SSMD-based criteria for classifying the size of siRNA effects, namely \( |\beta| \geq 5 \) for ‘extremely strong’, \( 5 > |\beta| \geq 3 \) for ‘very strong’, \( 3 > |\beta| \geq 2 \) for ‘strong’, \( 2 > |\beta| \geq 1.645 \) for ‘fairly moderate’, \( 1.645 > |\beta| \geq 1 \) for ‘fairly weak’, \( 1 > |\beta| \geq 0.75 \) for ‘very weak’, \( 0.75 > |\beta| \geq 0.5 \) for ‘weak’, \( 0.5 > |\beta| \geq 0.25 \) for ‘very very weak’ and \( |\beta| \leq 0.25 \) for ‘extremely weak’ effects (Zhang, 2007a, 2009).

In a primary HTS experiment with no replicates for each siRNA, we use SSMD based on unpaired difference to assess the size of siRNA effects. For an unpaired difference in a primary screen, the estimate of SSMD is \( \hat{\beta} = (\overline{X}_1 - \overline{X}_0)/k \), where \( k = \sqrt{\frac{n_1 - 1}{n_1} + \frac{n_2 - 1}{n_2}} \). For a paired difference in a confirmatory screen, the estimate of SSMD is \( \hat{\beta} = (\bar{X}_1 - \bar{X}_0)/\sqrt{n_1 + n_2} \). where \( \bar{X}_1 \) and \( \bar{X}_0 \) are sample sizes, means and SDs of an siRNA and a negative reference in a plate, respectively (Zhang, 2007a). In a confirmatory HTS screen, there are several sets of source plates. Each set is unique and has replicates (usually triplicates), thus each siRNA has replicates. Because plate-to-plate variability is usually higher than within-plate variability, SSMD based on paired differences are often used to assess the size of siRNA effects. For a paired difference in a confirmatory screen, the estimate of SSMD is \( \hat{\beta} = (\bar{X}_1 - \bar{X}_0)/\sqrt{n_1 + n_2} \), where \( n, \bar{D} \) and \( SD_0 \) are sample size, sample mean and SD of the paired difference, respectively (Zhang, 2008a).

The SSMD estimate \( \hat{\beta} \) is proportional to a non-central \( t \)-distribution \( t(v, \beta) \) with \( v \) degrees of freedom and non-central parameter \( \beta \), namely \( \hat{\beta} = k \cdot t(v, \beta) \), where \( v = n - 2, b = \sqrt{\frac{n_1 - 1}{n_1} + \frac{n_2 - 1}{n_2}} + k = \sqrt{\frac{n_1 - 2}{n_1} + \frac{n_2 - 2}{n_2}} \) for a primary screen without replicates, and \( v = n - 1, b = \sqrt{n} \) and \( k = \frac{\sqrt{n_1 - 1}}{\sqrt{n_1}} \sqrt{\frac{n_2 - 1}{n_2}} \) for a confirmatory screen (Zhang, 2008a). Using the non-central \( t \)-distribution \( t(v, \beta) \), the restricted FPL (RFPL), in which the siRNAs with weak or no effects are selected as hits, and the FNL, in which the siRNAs with strong effects are not selected as hits, can be as follows (Zhang, 2007a, 2008a). In the unregulated direction, we may use the selection rule of declaring an siRNA as a hit if it has \( \hat{\beta} \geq \beta_0 \) as a hit and as a non-hit otherwise. In this selection process, given a larger value \( c_1 \), a smaller value \( c_2 \) of SSMD and a preset level \( \alpha \) of RFPL w.r.t. \( c_2 \), the formula to determine the cutoff \( \beta^* \) and then to calculate the corresponding FNL w.r.t. \( c_1 \) are

\[
\beta^* = kQ_{\beta_0, \alpha, c_1}(1 - \alpha)
\]

\[
FNL = F_{\beta_0, \alpha, c_1}(Q_{\beta_0, \alpha, c_1}(1 - \alpha))
\]

where \( F_{\beta_0, \alpha, c_1}(\cdot) \) is the cumulative distribution function of non-central \( t \)-distribution \( t(v, \beta_0) \), and \( Q_{\beta_0, \alpha, c_1}(1 - \alpha) \) is the \((1 - \alpha)\) quantile of \( t(v, \beta_0) \). From formula (1), the cutoff \( \beta^* \) of SSMD estimate \( \hat{\beta} \) depends on both \( c_2 \) (a population value of SSMD to indicate small effects) and its corresponding RFPL \( \alpha \). From formula (2), the FNL depends on both \( c_1 \) (a population value of SSMD to indicate large effects) and \( \alpha \). Based on both original and probability meanings of SSMD, the commonly used values of \( c_2 \) are 0.25 and the commonly used values of \( c_1 \) are 1.28, 1.645, 2, 3 and 5 (Zhang, 2007a, 2009).

### 2.2 Sample size consideration in primary screens

In primary screens, \( n_1 \) equals 1 for most sample siRNAs. The sample size \( n_2 \) in the reference group varies in different experiments. Therefore, the critical issue of sample size determination in a primary screen without replicates is to determine the number of replicates (i.e. number of wells per plate) in the negative reference group. An essential consideration for hit selection in a primary screen is the capacity available for confirmation screening or other investigations following the primary screen. In a typical primary screen, there are about 20000 siRNAs and the major goal is to select 300–800 siRNAs in one direction for followup research. If we control the RFPLs w.r.t. extremely weak or no effects to be 0.05, 0.025 and 0.01, respectively, for one direction, we would obtain 1000, 500 and 200 hits even if all the 20000 siRNAs have extremely weak or no effects. Clearly, a RFPL of 0.05 is too large; if a RFPL of 0.025 might be acceptable; and a RFPL of 0.01 is preferred. The FNLs (w.r.t. \( \beta = 1.28, 1.645, 2, 3 \) and 5; i.e. \( c_2 = 1.28, 1.645, 2, 3 \) and 5) under the control of RFPL > 0.025 and 0.01 w.r.t. \( \beta = 0 \) (i.e. \( c_2 = 0 \)) are shown in Figure 1A and B, respectively. The FNLs under the control of RFPL > 0.025 and 0.01 w.r.t. \( \beta = 0.25 \) (i.e. \( c_2 = 0.25 \)) are shown in Figure 1C and D, respectively.

If the RFPL w.r.t. \( c_2 = 0 \) is controlled to be 0.025, the curves of FNLs w.r.t. \( c_1 = 1.28, 1.645, 2, 3 \) go down relatively quickly when the sample size is \( <16 \) and becomes relatively flat when the sample size is \( >16 \) (black, red, green and blue solid lines in Figure 1A). Meanwhile, if the RFPL w.r.t. \( c_2 = 0 \) is controlled to be 0.01, the curves of FNLs w.r.t. \( c_1 = 1.28, 1.645, 2, 3 \) go down relatively flat when the sample size is \( <20 \) and becomes relatively flat when the sample size is \( >20 \) (Figure 1B).

Similarly, if the RFPL w.r.t. \( c_2 = 0.25 \) is controlled to be 0.025, the curves of FNLs w.r.t. \( c_1 = 1.28, 1.645, 2, 3 \) go down relatively quickly when the sample size is \( <20 \) and becomes relatively flat when the sample size is \( >20 \) (Figure 1D). A sample size of 4, 8, 16 and 24, respectively leads to an FNL of 0.785, 0.566, 0.08 and 0.056 w.r.t. \( c_1 = 3 \) when controlling RFPL = 0.01 w.r.t. \( c_2 = 0 \) (Figure 1B). A sample size of 24 also leads to an FNL of nearly 0.10 w.r.t. \( c_1 = 3 \) when controlling RFPL = 0.01 w.r.t. \( c_2 = 0.25 \) (Figure 1D). These analyses with the consideration of capacity in the followup investigation also support that a choice of 4 or 8 wells per plate is not enough to achieve an acceptable FNR; a choice of 16 wells per plate is acceptable and a choice of 20 to 24 wells per plate is preferable for the negative control.

![Fig. 1. Changes of error rates versus sample size by controlling RFPL for siRNAs with no effects (A and B) and with extremely weak effects (C and D) in a primary screen without replicates. All panels have the same legends as shown in (D).](image-url)
When we control RFPL = 0.05 w.r.t. c

Therefore, 4, 5 or 6 is the reasonably small sample size for detecting siRNAs with strong, fairly strong and moderate effects when controlling a manageable 384-well plates do not have a power large enough to detect siRNAs with extremely weak effects. Therefore, the primary screen without replicates in solid curves in Figure 1C) when controlling RFPL = 0.025 for siRNAs with extremely strong effects (light blue curve in Fig. 1). However, the FNL for the siRNAs with strong, fairly strong or moderate effects can be high. For example, even if 300 wells are occupied by a negative reference (which is equivalent to the situation where the majority of sample wells are used as a negative reference for hit selection), the FNLs for the siRNAs with strong, fairly strong and moderate effects are about 0.35, 0.53 and 0.71, respectively (black, green and red solid curves in Figure 1C) when controlling RFPL = 0.025 for siRNAs with extremely weak effects. Therefore, the primary screen without replicates in 384-well plates do not have a power large enough to detect siRNAs with strong, fairly strong and moderate effects when controlling a manageable false positive rate for the siRNAs with extremely weak or no effects.

2.3 Sample size consideration in confirmatory screens

In a confirmatory screen, the goal is to achieve a reasonably low FNR of missing siRNAs with large effects (i.e., to obtain a reasonably high power of selecting siRNAs with large effects) in the condition of controlling a preset low level (such as 0.05) of false positive rate of selecting siRNAs with extremely weak or no effects. Figure 2 displays the FNLs w.r.t. SSMD critical values of 1.28, 1.645, 2, 3 and 5, respectively, under the control of RFPL = 0.05 w.r.t. c2 = 0.25 (or c2 = 0), which shows that the larger the sample size, the smaller the FNL. When RFPL = 0.05 w.r.t. c2 = 0.25 (or c2 = 0), the FNL w.r.t. c1 = 5 is 0.019 (or 0.0007), the FNL w.r.t. c1 = 3 is 0.231 (or 0.069), and the FNL w.r.t. c1 = 2 is 0.505 (or 0.288) for a sample size of 3. Clearly, a sample size of 3 can achieve a reasonably high power for detecting siRNAs with extremely strong effects (light blue curve in Fig. 2), may achieve an acceptable power for detecting siRNAs with strong effects (blue curve), but cannot achieve an acceptable power for detecting siRNAs with strong, fairly strong or moderate effects (green, red and black solid curves).

When we control RFPL = 0.05 w.r.t. c2 = 0.25, in an experiment with sample size increasing from 3 to 4, 5 and 6 replicates, FNLs w.r.t. c1 = 2 reduces from 0.505 to 0.276, 0.135 and 0.061, respectively. The amount of reduction of FNL w.r.t. c1 = 2 is not large after 6 (green curve in Fig. 2A). When we control RFPL = 0.05 w.r.t. c3 = 0, in an experiment with sample size increasing from 3 to 4 and 5 replicates, FNLs w.r.t. c3 = 2 reduces from 0.288 to 0.092 and 0.025, respectively. The amount of reduction of FNL w.r.t. c3 = 2 is not large after a sample size of 5 (green curve in Fig. 2B). Therefore, 4, 5 or 6 is the reasonably small sample size for detecting siRNAs with strong effects. Similarly, 5, 6, 7 or 8 is the reasonably small sample size for detecting siRNAs with fairly strong effects (red solid curve); 8, 9, 10 or 11 is the reasonably small sample size for detecting siRNAs with moderate effects (black curve in Fig. 2).

3 DISCUSSION

So far, for a primary screen, we focus on exploring sample size issues in the setting of no replicates. Recently, there is an increasing trend that a primary screen is conducted with replicates (such as in duplicate or triplicate) (c.f. Brass et al., 2008; Ganesan et al., 2008). The FNR and restricted false positive rate (RFPR) in a primary screen with replicates can be calculated in the same way as in a confirmatory screen, and we can use Figure 2 to determine sample size. Figure 2 reveals that a sample size of 3 leads to too high FNR’s even when controlling FPR’s to be <0.05 w.r.t. siRNAs with small or no effects.

It has been noted that the signal-to-noise ratio is low in many RNAi HTS screens. There are many reasons contributing to this situation. One of them is the existence of unadjusted systematic experimental errors (Brideau et al., 2003; Makarenkov et al., 2007; Zhang, 2008b; Zhang et al., 2006, 2008a, b). The analysis in this article reveals another reason, namely the design of sample size in RNAi HTS projects. The siRNAs of the most interest are those with strong, fairly strong or moderate effects. The analysis of sample size reveals that, whether using a negative control or the majority of sample wells in a plate as a negative reference, a primary screen without replicates does not have a power large enough to detect these siRNAs of interest when controlling a manageable false positive rate for the siRNAs with extremely weak or no effects. The analysis also reveals that, in a confirmatory screen or a primary screen with replicates, a sample size of 3 or 2 is not large enough to achieve a reasonably low FNR for these siRNAs of interest, while controlling a low false positive rate for siRNAs with extremely weak or no effects. Therefore, the commonly used design of sample size, either in a primary or confirmatory screen, in most of the current RNAi HTS projects does not allow to achieve a power high enough to detect the siRNAs of interest that have strong, fairly strong or moderate effects.

Based on the analysis results in this article, we provide the following guidance to the design of a genome-scale RNAi screen experiment. In a primary screen using 384-well plates, an arrangement of 4 or 8 wells per plate is not enough to achieve an acceptably low FNR; an arrangement of 16 wells per plate is acceptable and an arrangement of 20 or 24 wells per plate is preferable for the negative control to be used as a negative reference for hit selection in a primary screen. In a confirmatory screen or a primary screen with replicates, a sample size of at least 4 is required for detecting siRNAs with strong, fairly strong or moderate effects. To search a tradeoff between benefit and cost, any sample size between 4 and 11 is a reasonable choice for selecting siRNAs with strong, fairly strong or moderate effects. If the main focus is the selection of siRNAs with strong effects, a sample size of 4 or 5 is a good choice. In the situation where the constraint in cost is not too strong, a sample size of 6, 7 or 8 is preferred especially when only one or two sets of source plates are investigated in a confirmatory screen. If we want to have enough power to detect siRNAs with moderate effects, sample size needs to be 8, 9, 10 or 11. Even though we focus on the 384-well plate layout, our sample size recommendation for the control wells can be readily applied to 1536-well plate layout although it may not be applied to 96-well
plate layout. The sample size recommendation for a confirmation screen or a primary screen with replicates can be applied to any plate layouts.

ACKNOWLEDGEMENTS
The authors thank Dr. D.J. Holder, Dr. K.A. Soper and Dr. R. Bain for their support in this research and also editors and two anonymous referees for their helpful comments.

Conflict of interest: X.D.Z. and J.F.H. are both employees of Merck Research Laboratories.

REFERENCES
Zhang, X.H.D. (2008a) Genome-wide screens for effective siRNAs through assessing the size of siRNA effects. BMC. Res. Notes, 1, 33.