Hit selection with false discovery rate control in genome-scale RNAi screens

Xiaohua Douglas Zhang[1,*], Pei Fen Kuan[2], Marc Ferrer[3], Xiaohua Shu[4], Yingxue C. Liu[1], Adam T. Gates[5], Priya Kunapuli[3], Erica M. Stec[3], Min Xu[5], Shane D. Marine[3], Daniel J. Holder[1], Berta Strulovici[3], Joseph F. Heyse[6] and Amy S. Espeseth[7]

[1]Biometrics Research, Merck Research Laboratories, West Point, PA 19486, [2]Department of Statistics, University of Wisconsin, Madison, WI 53707, [3]Automated Biotechnology, Merck Research Laboratories, North Wales, PA 19454, [4]Department of Statistics, Temple University, Philadelphia, PA 19101, [5]Antiviral Research, [6]BARDS and [7]RNA Therapeutics, Merck Research Laboratories, West Point, PA 19486, USA

ABSTRACT

RNA interference (RNAi) is a modality in which small double-stranded RNA molecules (siRNAs) designed to lead to the degradation of specific mRNAs are introduced into cells or organisms. siRNA libraries have been developed in which siRNAs targeting virtually every gene in the human genome are designed, synthesized and are presented for introduction into cells by transfection in a microtiter plate array. These siRNAs can then be transfected into cells using high-throughput screening (HTS) methodologies. The goal of RNAi HTS is to identify a set of siRNAs that inhibit or activate defined cellular phenotypes. The commonly used analysis methods including median $\pm k$MAD have issues about error rates in multiple hypothesis testing and plate-wise versus experiment-wise analysis. We propose a methodology based on a Bayesian framework to address these issues. Our approach allows for sharing of information across plates in a plate-wise analysis, which obviates the need for choosing either a plate-wise or experimental-wise analysis. The proposed approach incorporates information from reliable controls to achieve a higher power and a balance between the contribution from the samples and control wells. Our approach provides false discovery rate (FDR) control to address multiple testing issues and it is robust to outliers.

INTRODUCTION

RNA interference (RNAi) is a naturally occurring pathway for the regulation of gene expression in which small double-stranded RNAs (siRNAs) lead to the destruction of mRNAs with complementary nucleoside sequences (1). This pathway can be co-opted by experimentally introducing synthetic 19–21-mer double stranded RNAs designed to target specific mRNAs, thus knocking down the expression of the protein of interest (2). RNAi offers an effective way of silencing a gene and has been seen as the third class of drugs, after small molecules and proteins (3). The importance of RNAi was further recognized when the Nobel Prize in Medicine and Physiology was awarded to Drs Fire and Mello in 2006 for their research in this field. RNAi can be utilized on a genome-wide scale via high-throughput screening (HTS) technology that allows thousands of siRNAs to be tested in an unbiased manner to identify previously unknown genes involved in a biological pathway. The goal of RNAi HTS is to identify a set of siRNAs that inhibits or activates the cellular phenotype being evaluated. This process is called hit selection. One of the most fundamental challenges in HTS is to glean biological significance from mounds of data, which relies on the development and adoption of appropriate statistical designs and analytic methods for quality control and hit selection (4).

For hit selection, there are two major types of approaches: one is the use of analytic metrics to assess and rank the size of siRNA effects and the other is the use of hypothesis testing to control false positive and false negative rates (5). In the first approach, fold change, mean difference, percent activity, percent viability, percent inhibition and strictly standardized mean difference (SSMD) have already been proposed and explored in the literature (5–10). In the second approach, the most commonly used methods are the use of z-score or t-test for testing the null hypothesis that no difference exists between the means, i.e. mean $\pm k$SD or median $\pm k$MAD (median absolute deviation) (11–17). These methods usually control the false positive and false negative rates based on

*To whom correspondence should be addressed. Tel: +1 215 652 0522; Fax: +1 215 993 1835; Email: Xiaohua_zhang@merck.com

a single test. Given that a large number of siRNAs are tested in an HTS assay, the false positive rate will be inflated. Hence, one issue in these methods is the adjustment of error rates in multiple hypothesis testing.

Another issue arising from these classical approaches is whether to perform plate-wise or experiment-wise analysis. The plate-wise analysis can adjust for different systematic errors within each plate. Previously, we demonstrated that a plate-wise analysis has a higher probability of detecting true, modestly effective, hits than an experiment-wise analysis when there are not cluster(s) of true hits in any of the plates (17). However, plate-wise analysis may produce misleading results if a cluster of active siRNAs is located within a single plate. An experiment-wise analysis is not affected by the distribution of active siRNAs between plates; however, it cannot adjust for systematic errors within each plate. Current strategies require a decision on whether to use plate-wise or experiment-wise analysis. Finally, all the above methods of hit selection utilize information from only a negative reference. It remains unresolved whether a negative control or the majority of sample wells should be used as the negative reference to capture information on the variability (18).

In this article, we propose a methodology based on a Bayesian framework for hit selection to address the issues encountered in the classical approaches described above. Newton *et al.* (19) developed a Bayesian method to control false discovery rate (FDR) for analyzing microarray data. Like Newton *et al.*'s method, our proposed Bayesian method controls FDR via a direct posterior probability approach. In addition, to adapt the Bayesian FDR methods to solve issues of hit selection in genome-scale RNAi screens, our method has the following unique features. First, the proposed method incorporates information from reliable controls; thus it maintains a balance between the contributions from the sample and control wells. For screens employing effective positive and negative controls, an approach that utilizes this information will be more powerful in hit selection. Second, the proposed method implements robust statistical processes; thus it is robust to outliers. Third, the priors of the proposed method incorporate experiment-wise information whereas the likelihoods are constructed in a plate-wise base. Hence, this approach allows for sharing of information across plates in a plate-wise analysis and thus, eliminates the need to decide between a plate-wise or experiment-wise analysis. We illustrate the performance of our approach via both the case studies and simulations.

METHODS

An RNAi HTS experiment is usually conducted in 96-well, 384-well or 1536-well plates in which the cells in a well are treated with a unique siRNA or a control. A typical RNAi HTS project starts with a primary screen of about 20 000 to 50 000 siRNAs most of which have no replicate. The effective siRNAs identified (called 'hits') in the primary screen are further investigated using one or more confirmatory screens in which each siRNA has replicates. Currently, a typical primary screen may have 50–200 384-well plates and a typical confirmatory screen may have 3–20 384-well plates. The measured response is usually the intensity emitted by labeled particles such as fluorescent dyes. For simplicity, in this article, the term 'intensity of an siRNA' is used to refer to 'the measured intensity of a phenotype corresponding to the treatment of an siRNA', which may be light intensity, the intensity emitted by a dye or the ratio of intensity emitted by two dyes depending on individual experiments. Considering mathematical and computational convenience for modeling the observations with normal distribution, we typically log transform the raw intensity for statistical analysis. All experiments include a positive control with specific knockdown effects (usually occupying 4–16 wells) and a negative control with nonspecific knockdown effects (usually occupying 4–20 wells) in each plate.

Likelihood

Let $X_1, X_2, \ldots, X_K$ be the log-transformed intensities of sample siRNAs in a plate. Since each sample well consists of unique siRNAs targeting different genes, we assume that each well has its own distribution. Assume that the likelihood function is given by

$$X_i | \mu_i \sim N(\mu_i, \sigma^2), \qquad \qquad 1$$

where μ_i is the unknown mean for sample siRNA i. For the i-th siRNA, our hypotheses of interest are: H_0: siRNA $i \in \{\text{no effect}\}$, H_1: siRNA $i \in \{\text{activation effect}\}$ and H_2: siRNA $i \in \{\text{inhibition effect}\}$. Here, we construct two Bayesian models using hypothesis testing to identify hits.

Prior distributions

Let the prior distribution for μ_i be $\pi(\mu_i)$. For simplicity, we first construct the prior for μ_i of sample siRNA i as follows,

$$\mu_i \sim N(\theta_0, \tau^2), \qquad \qquad 2$$

where θ_0 and τ^2 are the center and variance of a negative reference, i.e. an siRNA with no specific effects. θ_0 is pre-estimated using the mean of sample siRNA wells excluding outliers in a plate. This model (i.e. Model 1) is based on the assumption that most of the sample siRNAs do not have an activation/inhibition effect. We could also use a uniform prior if no prior information is available for μ_i. However, this will be equivalent to the classical z-score method or mean $\pm k$SD approach. τ^2 in Model 1 is assumed to be known and is obtained in the same way as in our second model (i.e. Model 2) that is described below.

In cases where both effective positive (activation and inhibition) and negative controls are available and reliable, we propose using a prior that incorporates the information from these controls for selecting activating or inhibiting siRNAs, which leads to Model 2. Let the prior be a mixture distribution

$$\mu_i \sim p_0 \pi_0 + p_1 \pi_1 + p_2 \pi_2, \qquad \qquad 3$$

where p_0 is the prior proportion of nonhits, p_1 is the prior proportion of activation hits, p_2 is the prior proportion of inhibition hits and $p_0 + p_1 + p_2 = 1$. The densities π_0, π_1

and π_2 describe the fluctuations of μ_i within each hypothesis. We assume that $\pi_0 \sim N(\theta_0, \tau^2)$, $\pi_i \sim N(\theta_1, \tau^2)$ and $\pi_2 \sim N(\theta_2, \tau^2)$. Other choices of π_1's are possible but our case study suggests it is reasonable to model them as normal distributions for simplicity. The parameters θ_0, θ_1, θ_2, τ^2, σ^2, p_0, p_1, p_2, in the model are obtained using the methods provided in the following two paragraphs.

In a typical RNA HTS experiment, the negative control wells are from an siRNA with unspecific effects, e.g. luciferase in our case study for HIV. The observed intensities from the negative control wells in a plate can be assumed to be independent and identically distributed as $X_i|\mu_N \sim N(\mu_N, \sigma^2)$ where μ_N is the mean of the negative control. In many cases, the variability of the negative controls can be assumed to be the same across plates within each experiment. Therefore, we can pool the variability estimates across the plates to allow information sharing. This is also motivated by the fact that the number of negative controls in each plate is usually small; thus if we pool the information across the plates we can obtain a better estimate. To reduce the impact of outliers, we calculate the deviation of each well for a negative control from the median of the negative control in each plate and exclude outliers identified experiment wise using boxplot parameters. That is, let Y_{ij} be the intensity (in log scale) of the i-th negative control well and $\tilde{Y}_{\bullet j}$ be the median of intensities of negative control wells in plate j. Let $d_{ij} = Y_{ij} - \tilde{Y}_{\bullet j}$ and apply a boxplot rule to d_{ij} experiment wise to identify outliers for the negative control. Apply $\hat{\sigma}^2 = (\sum_i \sum_j (Y_{ij} - \bar{Y}_{\bullet j})^2)/(\sum_j n_j - A)$ to estimate σ^2 of the negative control, where $\bar{Y}_{\bullet j}$ and n_j are, respectively, the mean and number of negative control wells after excluding outliers in plate j and A is the total number of plates. Similar reasoning can be extended to other types of controls. The final estimated σ^2 can be obtained by pooling $\hat{\sigma}^2$'s from different controls.

The unconditional variance of X_i is $Var(X_i) = E[Var(X_i|\mu_i)] + Var[E(X_i|\mu_i)] = E[\sigma^2] + Var[\mu_i] = \sigma^2 + \tau^2 \sum p_m^2$ where $m = 0, 1, 2$. For Model 1, $p_0 = 1$ and $p_1 = p_2 = 0$; for Model 2, p_m's can be estimated using the EM algorithm similar to Newton *et al.* (19). A natural estimate for $Var(X)$ is the sample variance of the intensities of sample siRNAs $X_1, X_2, \ldots, X_K$ in a plate. Given $\hat{Var}(X)$ and $\hat{\sigma}^2$, an estimate for τ^2 is max $\{(\hat{Var}(X) - \hat{\sigma}^2)/\sum p_m^2, 0\}$ to avoid a negative variance estimate. θ_0, θ_1, and θ_2 are estimated plate wise from the mean of sample wells, activation control wells and inhibition control wells, respectively, after excluding the corresponding outliers in a plate. Note, outliers are excluded only in the process of estimating the parameters θ_0, θ_1, θ_2, σ^2 and τ^2, but are included for further analysis such as calculating their posteriors.

Posterior distributions

For our Model 1, using the likelihood given in Equation (1) and prior distribution Equation (2), the posterior distribution of μ_i is given by

$$\mu_i|X_i \sim N\left(\frac{\sigma^2\theta_0 + \tau^2 X_i}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right) \qquad \mathbf{4}$$

Our quantity of interest is μ_i in this model. From posterior distribution Equation (4), the Bayesian framework here allows for the inference on each sample siRNA incorporating both its own intensity X_i and the center of the negative reference θ_0, respectively, weighted by the variability of the sample siRNAs and of the negative control.

For our Model 2, using the likelihood function Equation (1) and prior distribution Equation (3), the posterior distribution of μ_i is given by

$$f(\mu_i|X_i) = \frac{f(X_i|\mu_i)\pi(\mu_i)}{\int f(X_i|\mu_i)\pi(\mu_i)d\mu_i} = \tilde{p}_0\tilde{f}_0 + \tilde{p}_1\tilde{f}_1 + \tilde{p}_2\tilde{f}_2 \qquad \mathbf{5}$$

where

$$\tilde{f}_m \sim N\left(\frac{\sigma^2\theta_m + \tau^2 X_i}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$$

and m = 0,1,2

$$\tilde{p}_m = \frac{p_m\phi\big((X_i - \theta_m)/\sqrt{\sigma^2 + \tau^2}\big)}{\sum_{k=0}^{2} p_k\phi\big((X_i - \mu_k)/\sqrt{\sigma^2 + \tau^2}\big)}$$

$\tilde{p}_m$'s are estimated using the EM algorithm similar to Newton *et al.* (19), and ϕ is the density function of standard normal distribution. From posterior distribution Equation (5), the Bayesian framework here allows for the inference on each sample siRNA incorporating both its own intensity X_i and the mixture center of the controls $\sum_{k=0}^{2} \tilde{p}_k\mu_k$, respectively, weighted by the variability of the sample siRNAs distribution and of the negative reference.

Bayesian hypothesis testing

We are interested in the hypotheses H_0: siRNA $i \in \{$no effect$\}$, H_1: siRNA $i \in \{$activation effect$\}$ and H_2: siRNA $i \in \{$inhibition effect$\}$. An HTS experiment typically consists of approximately 30 000 siRNAs of interest. In other words, we are conducting $\sim$30 000 hypothesis tests simultaneously. Given the large number of hypothesis tests, an effective way of controlling the error rate is by using the notion of FDR (19–21). We use the direct posterior probability approach to control FDR (19). One key in this approach is the calculation of the probability that an siRNA falls under each hypothesis given the data, namely $P(H_m|X_i)$.

Suppose large (small) values of μ_i relative to θ_0 are in favor of the presence of activation (inhibition) effect. Under the prior in Model 1, for the i-th sample siRNA, a decision rule based on priors can be set to be: siRNA $i \in H_0$ (i.e. no effect) if $|\mu_i - \theta_0| \leq a$; siRNA $i \in H_1$ (i.e. activation effect) if $\mu_i - \theta_0 > a$ and siRNA $i \in H_2$ (i.e. inhibition effect) if $\mu_i - \theta_0 < -a$, where a is the 0.95 quartile of the prior distribution of μ_i, i.e. $N(\theta_0, \tau^2)$. Using these priors and the posterior distribution in formula (4), we can obtain the posterior probabilities of $P(H_0|X_i)$, $P(H_1|X_i)$ and $P(H_2|X_i)$ (i.e. $P(|\mu_i - \theta_0| \leq a|X_i)$, $P(\mu_i - \theta_0 > a|X_i)$, and $P(\mu_i - \theta_0 \leq -a|X_i)$, respectively) under Model 1. Under Model 2, the posterior probabilities of $P(H_0|X_i)$, $P(H_1|X_i)$, and $P(H_2|X_i)$ are, respectively, $\tilde{p}_0$, $\tilde{p}_1$ and $\tilde{p}_2$ in the posterior distribution in formula (5).

Given $P(H_m|X_i)$, the FDR for selecting activation hits only can be calculated as follows: suppose our goal is to identify a list of J siRNAs for which H_1 is true. Let

$$\beta_i = 1 - P(H_1|X_i). \qquad 6$$

We rank the siRNAs according to increasing values of β_i and declare the first J siRNAs with values of β_i less than some bound κ to be placed in the list. Because β_i is the conditional probability that placing an siRNA on the list creates a type I error, the expected number of false discoveries, given the data, is

$$C(\kappa) = \sum_i \beta_i I[\beta_i \leq \kappa] \qquad 7$$

Therefore, the expected FDR, given the data, is $C(\kappa)/J$. Similarly, we can obtain the FDR for selecting inhibition hits only by replacing $\beta_i = 1 - P(H_1|X_i)$ with $\beta_i = 1 - P(H_2|X_i)$ and the FDR for selecting both activation and inhibition hits by replacing $\beta_i = 1 - P(H_1|X_i)$ with $\beta_i = 1 - P(H_1|X_i) - P(H_2|X_i)$.

With the control of FDR calculated using the above method, we may have two major strategies to select hits in RNAi HTS screens. One is to set up a prespecific FDR and search a list of J siRNAs with the smallest posterior probabilities under the null hypothesis, such that the expected FDR (namely $C(\kappa)/J$) given the data equals to the prespecific FDR value. The other is to fix the number (M) of hits to be selected for further analysis, for which we can rank the siRNAs according to increasing posterior probabilities under the null hypothesis and select the first M siRNAs as hits. The corresponding FDR is then the sum of their posterior probabilities under the null hypothesis divided by M. In addition, if we are interested in classifying each siRNA as either H_0 (no effect), H_1 (activation) or H_2 (inhibition), this can be achieved by declaring sample siRNA i to be H_k if $\tilde{p}_k = \max_m(\tilde{p}_m)$.

Simulations

We conducted two simulation studies to investigate the performance of our Bayesian approach. The first simulation explores the situation where each plate has about the same number of hits whereas the second simulation explores the situation in which some of the plates have an enriched number of hits in each experiment. In both studies, we simulated data for 100 384-well plates in one run, each plate having a negative control, an activation control, an inhibition control, true activation hits, true inhibition hits and nonhits. Specifically, in a plate there were 12 draws from normal distribution $N(\theta_0, \sigma^2)$ for the negative control, 10 draws from $N(\theta_0 + 4\sigma, \sigma^2)$ for an activation control, 10 draws from $N(\theta_0 - 4\sigma, \sigma^2)$ for an inhibition control, n_1 draws from $N(\theta_0 + h\sigma, \sigma^2)$ for activation hits, n_2 draws from $N(\theta_0 - h\sigma, \sigma^2)$ for inhibition hits and $384 - 32 - n_1 - n_2$ draws from $N(\theta_0, \tau^2)$ for nonhits. θ_0, τ and σ were sampled from three continuous uniform distributions U(4, 5), U(0.5, 0.6) and U(0.35, 0.45), respectively. The three continuous uniform distributions come from actual RNAi HTS experiments.

For the first simulation study, the number of true hits, n_1 and n_2, were both drawn from the discrete uniform

distribution U[0, 10] for each run of the 100 plates. For the second simulation study, in each run, we simulated E enriched plates and 100-E regular plates where E was a random draw from 1 to 9. n_1 and n_2 were both drawn from the same discrete uniform distribution, U[50, 70], in an enriched plate and from the discrete uniform distribution, U[0, 5], in a regular plate. In both studies, the data were simulated from various scenarios for the strength of true hits. The strength of a hit is indexed by h, the ratio of the deviation between the mean of a hit and θ_0 to the standard deviation of a control, i.e. $h = (\text{mean of a hit} - \theta_0)/\sigma$. Each scenario was repeated 200 times.

APPLICATIONS AND RESULTS

In the Methods Section, we proposed two Bayesian models for hit selection in RNAi HTS assays. In Model 1, we construct the prior using a negative reference, whereas in Model 2, we construct the prior based on a negative reference, an activation control and an inhibition control. Here we apply both Bayesian models to an HIV siRNA primary screen and a hepatitis C virus (HCV) siRNA primary screen.

Experiment 1: a genome-scale HIV siRNA screen

A HTS was conducted to identify cellular factors associated with HIV replication. In this screen, HeLa P4/R5 cells, which produce β-galactosidase upon infection with HIV (22), were transfected with siRNAs targeting 18 601 genes. Following transfection, the cells were infected with HXB2 HIV. To allow for the identification of host factors associated with early events, the level of HIV infection (as determined by β-galactosidase activity) was assessed at 48 h postinfection. The experiment consisted of 96 384-well plates. There were 304 sample siRNA wells, 8 negative control wells (Luciferase), 4 moderate inhibition control wells (CCNT), 16 strong inhibition control wells (cell no virus) and 4 activation control wells (TSG-101).

With either model, we have two strategies to select hits with FDR control: one is to fix the number of selected siRNAs and the other is to control a prespecific FDR. For the first strategy, we can obtain the FDR level corresponding to various preset numbers of selected siRNAs. For the second strategy, we can obtain the number of selected hits corresponding to various prespecific FDR levels. In either strategy, the FDR levels can be obtained for the selected siRNAs, which effectively addresses the multiplicity issue of hit selection in RNAi HTS assays. In contrast, in mean $\pm k$SD and median $\pm k$MAD, k is commonly chosen to be 2 or 3, which is based on a single test. Table 1 displays the results for both models under various preset number of selected hits or prespecific FDR levels.

Based on Table 1, using the ranking strategy in Model 1, if we select 100 siRNAs as hits, the FDR is only 0.039; on the other hand, if we select 1000 siRNAs as hits, the FDR is 0.228. Using the strategy of directly controlling FDR, if the prespecified FDR is 0.01, only 31 siRNAs are selected as hits; on the other hand, if the

Table 1. FDR level and number of selected hits in an HIV siRNA screen by Bayesian models

A: Bayesian model 1				B: Bayesian model 2							
				With moderate inhibition control				With strong inhibition control			
Fix list size		Fix FDR		Fix list size		Fix FDR		Fix list size		Fix FDR	
Size	FDR	FDR	Size	Size	FDR	FDR	Size	Size	FDR	FDR	Size
100	0.039	0.01	31	100	0.099	0.01	1	100	0.272	0.01	2
200	0.074	0.05	128	200	0.138	0.05	32	200	0.429	0.05	16
300	0.104	0.10	286	300	0.164	0.10	105	300	0.524	0.10	33
400	0.131	0.15	471	400	0.185	0.15	242	400	0.592	0.15	48
500	0.157	0.20	685	500	0.202	0.20	487	500	0.642	0.20	67
600	0.181	0.25	949	600	0.217	0.25	882	600	0.681	0.30	113
700	0.203	0.30	1266	700	0.230	0.30	1467	700	0.711	0.35	142
800	0.223	0.35	1642	800	0.241	0.35	2254	800	0.736	0.40	176
1000	0.228	0.40	2088	1000	0.262	0.40	3253	1000	0.776	0.50	270
Classify sample siRNAs				Classify sample siRNAs				Classify sample siRNAs			
FDR			0.312	FDR			0.366	FDR			0.282
# activation:			816	# activation:			119	# activation:			64
# inhibition:			536	# inhibition:			2439	# inhibition:			41
# no effect:			26 142	# no effect:			24 936	# no effect:			27 389

prespecified FDR is 0.30, 1266 siRNAs are selected as hits. One of the goals in a primary screen is usually to select a set of 500 to 1400 siRNAs with the lowest FDR for further investigation. With this consideration, it is reasonable to select hits by controlling the FDR to be somewhere between 0.25 and 0.35 depending on the capacity available for confirmation screening or other further investigations.

Model 1 does not use any information about the strength of the positive controls. Consequently, the hit selection results using Model 1 are not affected by the effectiveness of any positive controls in the experiment (Figure 1A and B). On the other hand, Model 2 does incorporate the strength of positive controls. Thus, hit selection results using Model 2 are highly affected by the positive controls adopted in the experiment. In the HIV RNAi screen, Model 2 identifies many more inhibition hits (i.e. 439 inhibition hits) using the moderate inhibition control compared to those (i.e. 27 inhibition hits) identified using the strong inhibition control (Figure 1C and E) when FDR = 0.20. This is because more sample siRNAs are similar to the moderate control than to the strong control.

Using our Bayesian models, we get an estimated probability of each siRNA being in each of the three hypotheses: H_0 (no effect), H_1 (activation) and H_2 (inhibition). According to the maximal posterior probability among these hypotheses, we can classify each sample siRNA into one of the three categories: no effect, activation and inhibition. In this classifying strategy, we can also obtain the corresponding FDR by summing up the posterior probabilities of all selected hits under the null hypothesis divided by the number of hits. The results are listed in the bottom rows of Table 1. Using the classifying strategy based on Model 1, we identified 816 siRNAs as activation hits and 536 siRNAs as inhibition hits with a corresponding FDR of 0.312 (Panel A of Table 1 and Figure 1B). As described earlier, classification results based on Model 2

are strongly affected by the strength of the positive controls. For example, there are 2439 siRNAs classified as inhibition hits with a FDR of 0.366 if the moderate inhibition control is used, whereas there are only 41 siRNAs classified as inhibition hits with a FDR of 0.282 if the strong inhibition control is used (Panel B of Table 1 and Figure 1D and F). This result hints that many siRNAs have moderate to weak inhibition effects whereas only a few siRNAs are strongly inhibitory.

When using the commonly used methods such as mean ± kSD and median ± kMAD, we also need to decide whether we should apply the methods experiment- or plate-wise. In the HIV screen, the between-plate variability is fairly large. Clearly, the data in plates 1–21 shift down while the data in plate 79 shift up compared with the data in the remaining plates. When applying median ± 3MAD experiment wise, the inhibition hits are mostly selected from Plates 1 to 21 and the activation hits are mostly selected from Plate 79 and plates 22 to 44 (Figure 2C and D). Clearly these results are dominated by the systematic errors and are unreliable. The plate-wise hit selection results are more reasonable (Figure 2A and B). That is one of the major reasons that median ± 3MAD is applied plate wise more often than experiment wise. On the other hand, an experiment-wise analysis is not affected by the distribution of active siRNAs between plates whereas a plate-wise analysis may produce misleading results when a cluster of active siRNAs is located within a single plate. Our Bayesian methods allow for sharing of information (experiment-wise) across plates in a plate-wise analysis and thus, avoid the need for a decision between plate-wise or experiment-wise analysis. The distribution of selected hits across the plates using our Bayesian methods also looks reasonable (Figure 1A–F).

In the methods of mean ± kSD and median ± kMAD, there is a need to decide whether a negative control or the
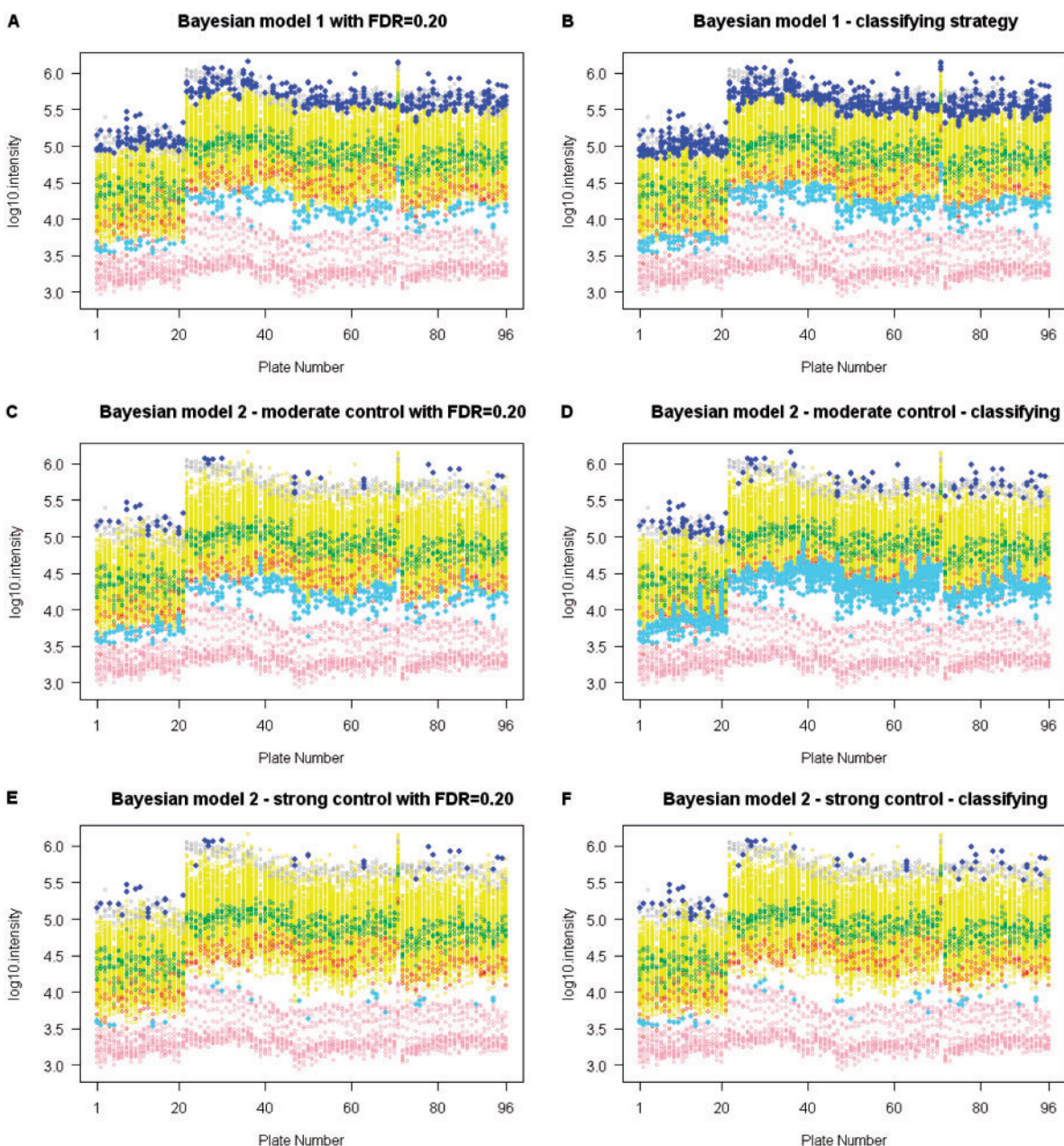
Figure 1. Hits identified by Bayesian methods with FDR control. In each panel, green, red, light pink, grey, yellow, blue and light blue points represent a negative control, moderate inhibition control, strong inhibition control, fairly strong activation control, nonhits, activation hits and inhibition hits, respectively. **A** and **B** show the result using priors based on the negative control only; **C** and **D** show the results using priors based on the negative, activation and moderate inhibition controls and **E** and **F** show the results using priors based on the negative, activation and strong inhibition controls.

majority of sample wells should be used as a negative reference to calculate the mean (or median) and SD (or MAD). Considering the fact that median $\pm$ kMAD is more robust to outliers, we concentrate on the comparison of median $\pm$ kMAD with our Bayesian methods. For median $\pm$ 3MAD, if using the negative control as a negative reference, we will select too many hits either plate wise (i.e. 8808 hits shown in Figure 2B) or experiment wise (i.e. 2138 hits shown in Figure 2D). On the other hand, if using the majority of sample wells as the negative reference, we will select much fewer hits (i.e. 106 hits shown in

Figure 2A and 29 hits shown in of Figure 2C). Our Bayesian models do not have such an issue in the determination of a reference because, from the posteriors shown in formulas (4) and (5), our Bayesian models naturally combine the information in both the negative control and sample wells. The numbers of selected hits using our Bayesian methods are reasonable (e.g. 685 hits with a FDR of 0.20 shown in Figure 1A and 1352 hits with a FDR of 0.312 shown in Figure 1B). These methods control FDR to adjust for multiplicity issues and provide each siRNA with more than a yes/no answer, but rather
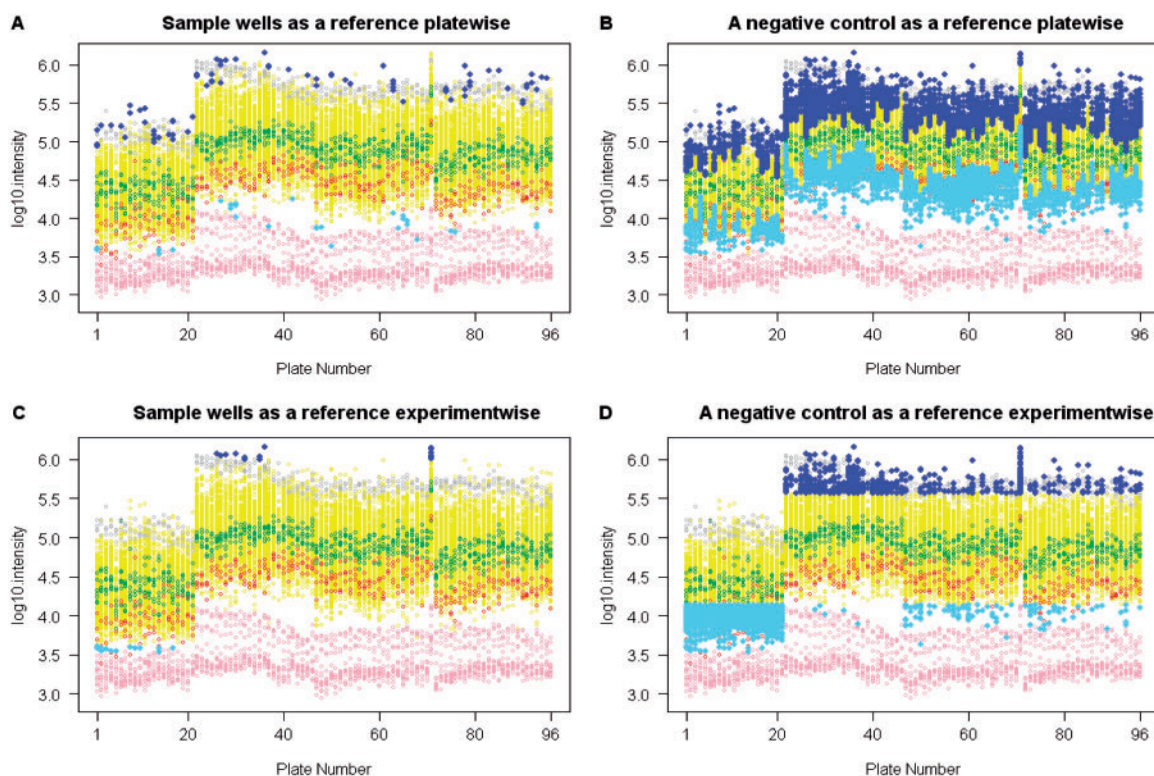
Figure 2. Hit selection results in the HIV screen, identified by median ± 3MAD plate wise (**A** and **B**) or experiment wise (**C** and **D**) and using the majority of sample wells (A and C) or a negative control (B and D) as a negative reference. In each panel, green, red, light pink, grey, yellow, blue and light blue points represent a negative control, moderate inhibition control, strong inhibition control, activation control, nonhits, activation hits and inhibition hits, respectively.

an estimate probability of being in each of three groups: no effect, activation and inhibition.

Experiment 2: a genome-scale HCV siRNA screen

A primary siRNA HTS experiment was conducted to identify host factors associated with HCV replication, using the HCV replicon assay system described in Zuck *et al.* (23). 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. The HTS was conducted using siRNAs in 97 plates, each with 384 wells (17). In this experiment, there were sample siRNAs (no replicates in the whole experiment), a strong positive control that targeted HCV (16 replicates per plate), a weaker positive control that targeted the HCV host factor, VAPA (8 replicates per plate) and a negative control (16 replicates per plate).

The results of hit selection using median ± kMAD experiment wise are dominated by the systematic shift in the first nine plates (Figure 3B). Plate-wise median ± kMAD tends to obtain more reasonable hits in this screen (Figure 3A). Using Model 1, we not only obtain more reasonable results (Figure 3C and D) but also know FDR corresponding to any group of selected hits (Table 2). Based on Model 1, more inhibition hits were

identified than activation hits given a prespecific FDR in the screen. For example, given a prespecific FDR of 0.20, we identified 290 inhibition hits and 66 activation hits (Panel A of Table 2 and Figure 3C). Using the classifying strategy of Model 1, we identify 591 inhibition hits and 150 activation hits, resulting in a FDR of 0.32 (Panel A of Table 2 and Figure 3D). To obtain a reasonable number of hits for further investigation while considering the control of FDR, we may fix the FDR to be somewhere between 0.25 and 0.40 for hit selection in this screen.

There was no control for identifying siRNAs that enhanced HCV replication in the screen. Hence, strictly speaking, the mixture model cannot be applied to this screen. However, one control appeared to act as an activation control (gray points in Figure 3). Therefore, we applied this control as an activation control and apply Model 2 to identify hits in the inhibition direction only. The numbers of selected hits corresponding to various prespecified FDRs are shown in Panel B of Table 2. Note, this siRNA was not designed to be an activation control and is not stable, even resulting in a median signal less than that of the negative reference in a few plates (Figure 3). If one uses Model 2 with this control to identify hits with activation effects, one would obtain misleading results: some selected activation hits have values less than the median of the negative control in some plates (Figure 3E and F). Hit selection using the classifying strategy of Model 2 is also strongly affected
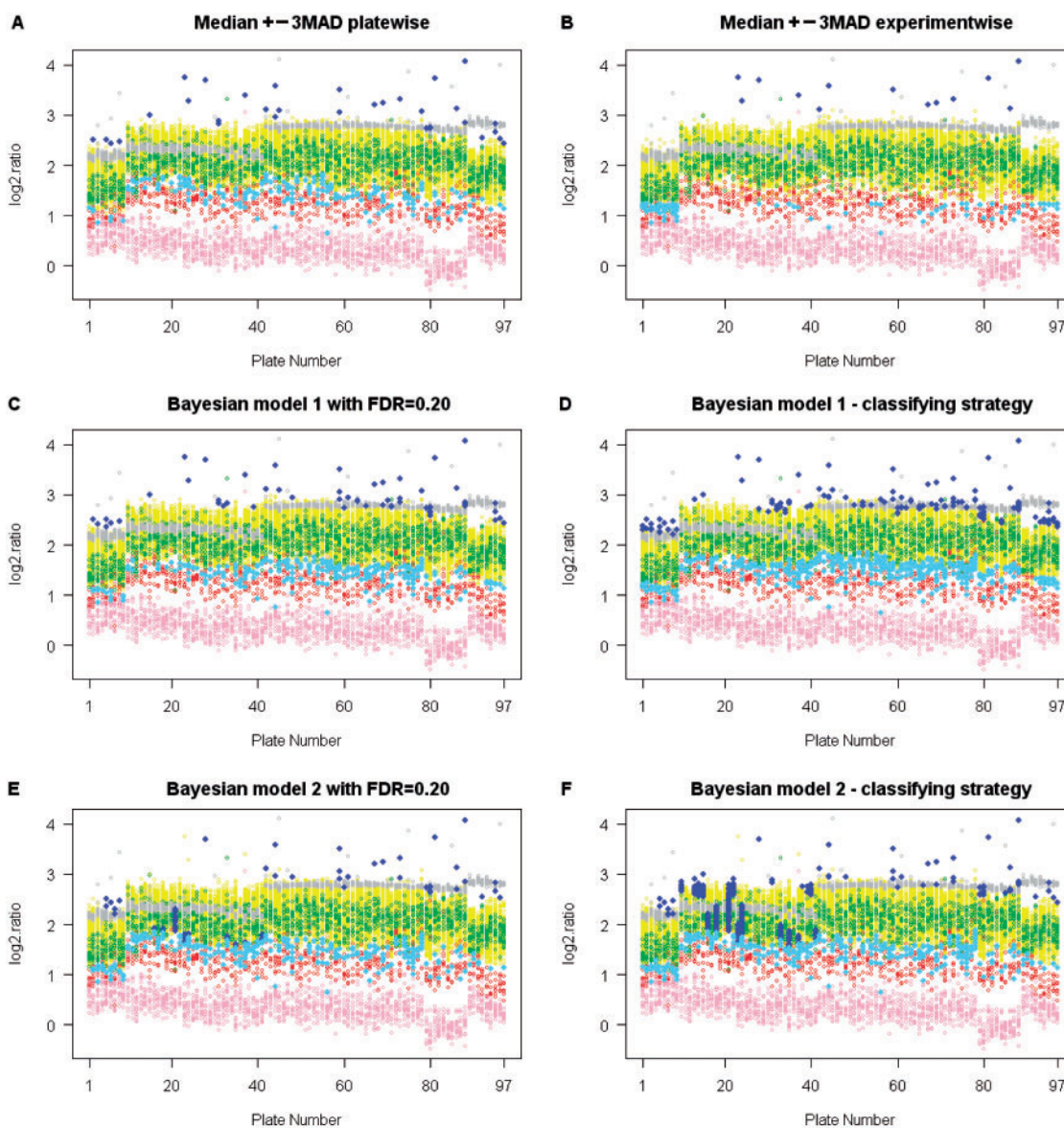
Figure 3. Hit selection results in the HCV screen, identified by median ± 3MAD plate wise (**A**) and experiment wise (**B**) using the majority of sample wells as a negative reference as well as by Bayesian Model 1 (**C** and **D**) and Bayesian Model 2 with the weaker inhibition control (**E** and **F**). In a panel, green, red, light pink, grey, yellow, blue and light blue points represent a negative control, weaker inhibition control, stronger inhibition control, pseudo activation control, nonhits, activation hits and inhibition hits, respectively.

by the quality of positive controls, and can thus generate similar misleading results (Panel B of Table 2 and Figure 3F).

ROC analysis of experiments

Receiver operating characteristic curve (ROC) analysis is commonly used to compare the performance of different methods in experiments in which both positive and negative controls are available and reliable. Here, we use ROC analysis to compare the performance of our proposed Bayesian approaches with plate-wise and experiment-wise median ± kMAD methods in both the HIV and HCV screens. The performance of each method is assessed via sensitivity (true positive rate) and specificity

(true negative rate). We used the negative control to define the true negative set in either screen. In the HIV screen, we used all three positive controls (i.e. activation, moderate inhibition and strong inhibition controls) to define the true positive set. In the HCV screen, we used the two inhibition controls to define the true positive set. The ROCs of five methods (i.e. plate-wise median ± kMAD, experiment-wise median ± kMAD, Bayesian Model 1, Model 2 using the weaker inhibition control and Model 2 using the stronger inhibition control) in the HIV and HCV screens are given in Figure 4A and B, respectively.

As evident in the ROC curves for the HIV screen in Figure 4A, Bayesian Model 1 has the highest power to detect the three positive controls given any false positive

Table 2. FDR level and number of selected hits using Bayesian models in an HCV siRNA screen

Selection strategy	A: Bayesian model 1 prior relying on the negative control			B: Bayesian model 2 mixture model					
				With weaker inhibition control			With stronger inhibition control		
	FDR	#INH	#ACT	FDR	#INH	#ACT	FDR	#INH	#ACT
Preset FDR	0.01	29	10	0.01	72	10	0.01	9	10
	0.05	74	17	0.05	157	17	0.05	13	18
	0.10	137	27	0.10	235	22	0.10	15	28
	0.15	210	41	0.15	295	41	0.15	16	44
	0.20	290	66	0.20	321	106	0.20	18	66
	0.25	395	96	0.25	352	195	0.25	19	108
	0.30	536	127	0.30	380	334	0.30	19	196
	0.35	685	190	0.35	395	579	0.35	19	372
	0.40	863	277	0.40	399	1067	0.40	21	801
	0.50	1366	574	0.50	528	4349	0.50	23	3968
Classify siRNAs	0.32	591	150	0.38	387	820	0.44	21	1589

Note: '#INH' and '#ACT' indicates the numbers of selected hits with inhibition and activation effects, respectively.
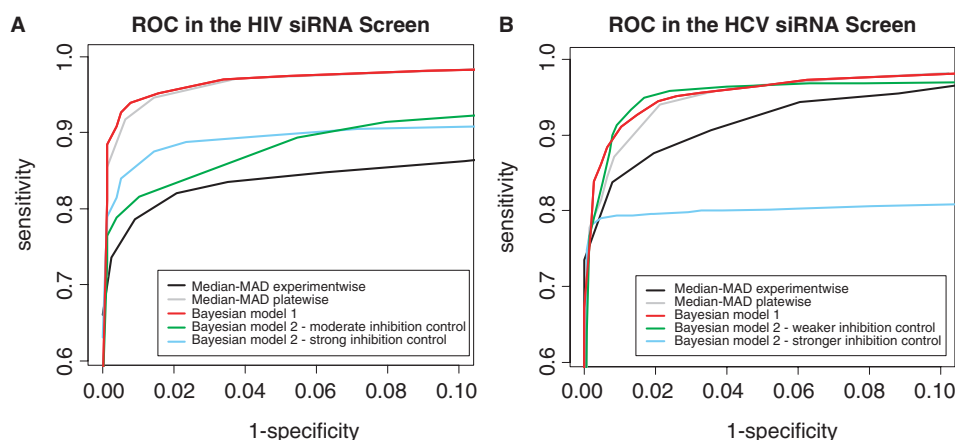

Figure 4. ROC curves for detecting all three positive controls in the HIV siRNA screen (**A**) and for detecting two inhibition controls in the HCV siRNA screen (**B**). The negative control in each screen is used to calculate the specificity in the screen.

rate in the range of interest (i.e. between 0 and 0.05); Bayesian Model 2 has powers higher than experiment-wise median $\pm k$MAD but lower than plate-wise median $\pm k$MAD and plate-wise median $\pm k$MAD performs slightly worse than Bayesian Model 1. As evident in the ROC curves for the HCV screen in Figure 4B, Bayesian Model 1 performed better than plate-wise median $\pm k$MAD. Bayesian Model 2 using the weaker inhibition control performs better than Model 1 when the false positive rate is in the range of 0 and 0.01 and worse than Model 1 when the false positive rate is in the range of 0.01 and 0.05. In both screens, experiment-wise median $\pm k$MAD always has a lower power than plate-wise median $\pm k$MAD given a false positive rate between 0 and 0.1. Bayesian Model 2 with the strong inhibition control did not perform well in either screen.

Simulation studies

The above results describing the performance of different methods on hit selection in siRNA HTSs are based on only two experiments. Thus, we conducted simulation studies as described in the Method Section to investigate the performance of our Bayesian approach in general. An exemplary run of simulated data in the first simulation study is shown in Figure 5A1.

We compared the performance of the Bayesian approach with the commonly used median $\pm k$MAD method via area under the receiver operating characteristic curve (AUROC). For each method, the sensitivity and specificity were computed at various target errors α. k was chosen as $\Phi^{-1}(1-\alpha/2)$ where Φ is the cumulative distribution function of the standard normal distribution. The sample wells with values greater than median $+ k$MAD or less than median $- k$MAD were declared to be hits. For the Bayesian approach, the target error α was the FDR. The negative control was used to calculate specificity and both activating and inhibiting true hits were used to calculate sensitivity. ROC curves in one run of 100 plates are shown in Figure 5A2. For each run, the AUROC was then computed as the area under a ROC curve. In real RNAi HTS experiments, the value of the approach is determined by its sensitivity in situations where the false positive rate is below 0.05 experiment wise,
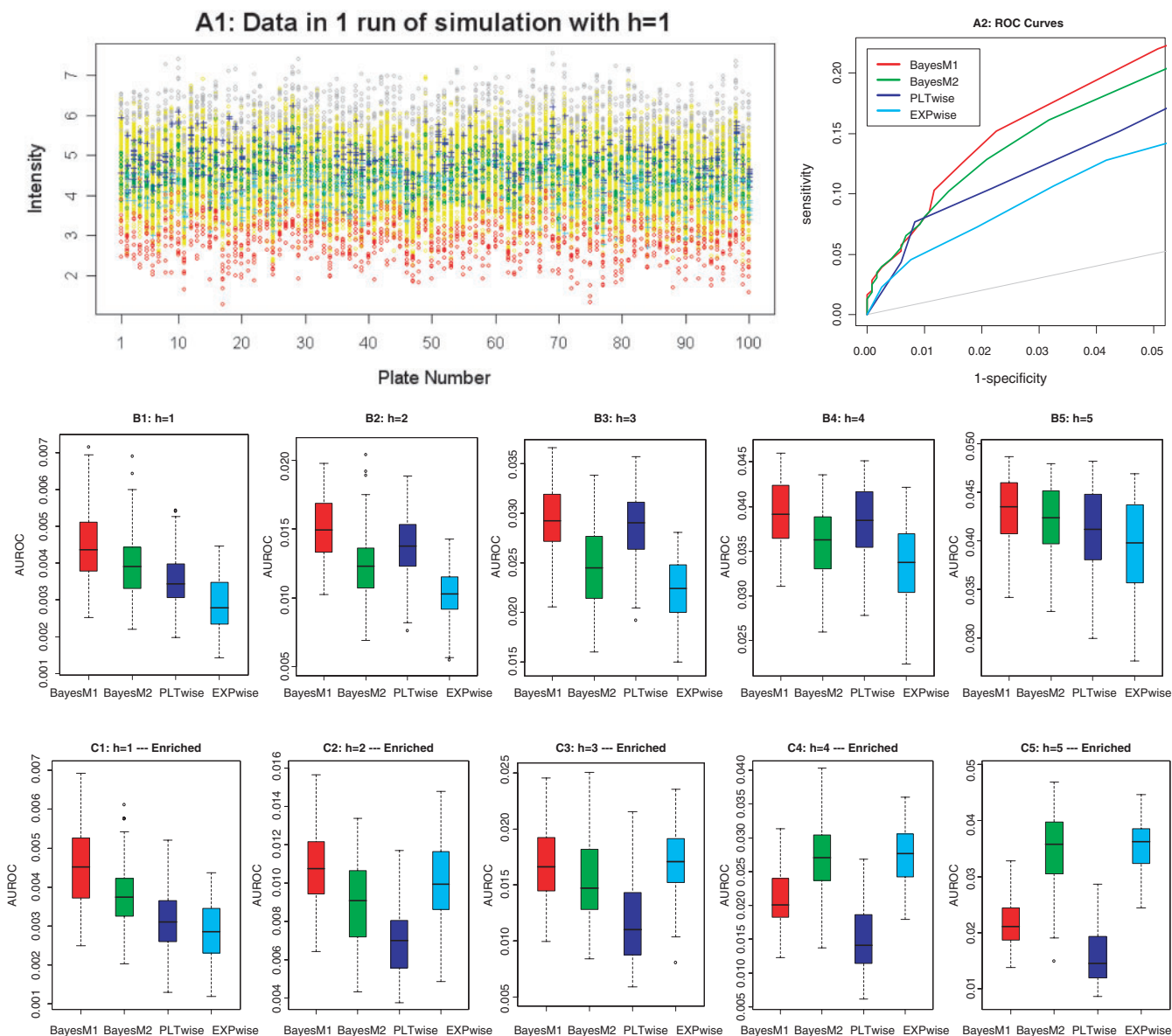
Figure 5. Simulated data and results. **A1** shows a simple realization of 100 plates in 1 run of the first simulation with $h = 1$ where $h = (\text{mean of a hit} - \theta_0)/\sigma$. In this panel, green, red, gray and yellow points represent a negative control, inhibition control, activation control and nonhits, respectively; blue and light blue crosses represent activation and inhibition hits, respectively. **A2** shows the ROC curves by applying four methods to the data in **A1**. **B1**−**B5** (and **C1**−**C5**) show AUROC's for the specificity between 1 and 0.95 in the first (and second) simulation study with $h = 1$, 2, 3, 4 and 5, respectively. In each **A1**, **B1**−**B5** and **C1**−**C5**, the Bayesian models 1–2, plate-wise and experiment-wise median $\pm\, k\text{MAD}$ are labeled as BayesM1 (red), BayesM2 (green), PLT wise (blue) and EXP wise (light blue), respectively.

which is equivalent to a specificity of above 0.95. Therefore, we calculated the AUROC in the situation where the specificity is between 0.95 and 1. When the specificity is above 0.95, a null method that randomly declares hits has a value of 0.00125 for AUROC and a perfect method has a value of 0.05 for AUROC. Figure 5B1–5 show the AUROC results in the first study and C1–5 show the AUROC results in the second study (Figure 5). Each panel summarizes the AUROC for 200 runs from different strengths of true hits.

From B1–5 of Figure 5 for the first study, Bayesian Model 1 overall has the largest AUROC in all scenarios, especially when the true hits have weak to moderate

effects. Employing plate-wise median $\pm\, k\text{MAD}$ tends to result in an AUROC slightly less than Bayesian Model 1. Using experiment-wise median $\pm\, k\text{MAD}$ resulted in the smallest AUROC of all three scenarios tested. Bayesian Model 2 resulted in an AUROC between Bayesian Model 1 and experiment-wise median $\pm\, k\text{MAD}$. Therefore, the first simulation study suggests that, in an experiment without any enriched plates, Bayesian Model 1 is the most effective method, especially with regard to detecting hits with weak or moderate effects; followed by the use of the plate-wise median $\pm\, k\text{MAD}$ method, Bayesian Model 2 and finally by the use of the experiment-wise median $\pm\, k\text{MAD}$ method.

DISCUSSION

We proposed a methodology for hit selection in genome-scale RNAi HTS based on a Bayesian framework. The proposed methodology aims to address issues arising from classical approaches including median $\pm k$MAD. One issue is the adjustment of error rate in multiple hypothesis testing. In this article, we construct Bayesian models to control FDR via the direct posterior probability approach (19). The second issue is the decision of whether a plate-wise or experiment-wise analysis should be used. The results in both HIV and HCV siRNA screens show that selected hits using an experiment-wise analysis can be dominated by systematic errors in a plate. A plate-wise analysis is more robust in the presence of systematic errors, but can result in false negatives in plates containing a cluster of true hits. Our proposed Bayesian methodology allows data to be shared across the plates of an experiment in a plate-wise analysis, thus obviating the need for choosing either plate-wise or experiment-wise analysis. The third issue is the use of controls: Should the negative control or the majority of sample wells be used as the negative reference when calculating the center and variability of the raw data? How should the information from positive and negative controls with varying degrees of efficacy be incorporated into the hit selection strategy? The choice of prior in our Bayesian methods enables us to incorporate information from various controls. From the posteriors in the proposed models, our Bayesian approach maintains a balance between the contribution from the sample wells and control wells. In addition, outliers are identified and excluded in the estimation of parameters in our Bayesian methods; thus our methods are robust to outliers. Finally, the Bayesian approach provides each siRNA with more than a yes/no answer, but rather an estimate probability of being in each of three groups: activation, inhibition and no effect.

We applied the Bayesian methods to two real genome-scale RNAi screens with different features. The HCV screen did not have a true activation control whereas the HIV screen did. The inhibition controls had data variability larger than the negative control in the HIV screen whereas the inhibition controls had data variability smaller than the negative control in the HCV screen. The strength of positive controls in the two screens also differed from each other. By using the Bayesian methods for hit selection we describe here, we were able to address effectively the issues of hit selection common to both screens and to obtain a reasonable pool of selected hits for both screens despite the differences between the two.

We focus on two Bayesian models for hit selection in RNAi HTS assays. In Model 1, we construct the prior using a negative reference, whereas in Model 2, we construct the prior based on a negative reference, an activation control and an inhibition controls. We apply both Bayesian models for hit selection in an HIV siRNA primary screen and an HCV siRNA primary screen. The applications show that the hit selection results using our proposed methods, especially Model 1, are more reasonable than those using classical methods (Figures 2–4). ROC analysis in the case studies shows that Model 1 is more powerful than the commonly used classical approaches. When the positive controls are moderately effective and of high quality, Model 2 is also powerful (Figure 4B). Simulation studies show that, in an experiment without any enriched plates, in general, Bayesian Model 1 has the best performance especially in detecting hits with weak or moderate effects; while Bayesian Model 2 performs better than the experiment-wise median $\pm k$MAD method and has a good performance in detecting hits with effects equal to or stronger than the positive controls (Figure 5B1–5). In an experiment with 1 to 9 enriched plates, in general, Bayesian Model 1 has the best performance in detecting hits with weak or moderate effects, while Bayesian Model 2 performs equivalently to classical experiment-wise methods in detecting hits with strong effects (Figure 5C1–5).

Bayesian Models 1 and 2 allow for three strategies to select hits: (i) fix the number of selected siRNAs, (ii) control a prespecific FDR or (iii) classify an siRNA based on its maximal posterior probability. In a typical primary siRNA HTS screen, one of the goals is usually to select a practical number (usually some number between 500 and 1500) of hits for further investigation. Therefore, there is a motivation to use the first strategy for hit selection in a primary screen. Our Bayesian methods allow selecting a fixed number of siRNAs with the lowest FDR. On the other hand, we cannot simply select a fixed number (say 1000) from every screen and ignore the error rate because different screens may have very different error rates. Thus, the second strategy may be more plausible especially when we want to control some fixed error rate in most screens. Considering both HIV and HCV screens discussed in this article, a reasonable choice of a fixed FDR is somewhere between 0.25 and 0.35, which allows us not only to fix a reasonable number of siRNAs for further analysis but also to control the FDR within reasonable range. The classifying strategy under Model 1 also works effectively in the two screens, which yields an FDR of around 0.30 and a reasonable number of selected hits (i.e. 1352 hits in the HIV screen and 741 hits in the HCV screen).

Model 1 does not use any information about the strength of positive controls. Consequently, the hit selection results using Model 1 are robust to the quality and strength of any positive controls used in the experiment. On the other hand, Model 2 incorporates the strength of positive controls. Thus, hit selection results using Model 2 are highly sensitive to the data quality and strength of positive controls adopted in the experiment. It is not uncommon that the strength of positive controls varies within or between experiments. It is also not uncommon that the data quality of positive controls is poor in some experiments. These facts make Model 1 more favorable than Model 2 for hit selection in RNAi HTS screens. On the other hand, the simulation studies show that Model 2 has a better performance in detecting hits with effects equal to or stronger than the positive controls when there are plates with enriched number of hits (Figure 5C1–5). If the data quality of positive controls is good and the positive controls are moderately or fairly strongly effective, Model 2 may be more

powerful (Figure 4B). Thus, Model 2 may help us to identify siRNAs with a desired potency if there are high quality and reliable controls with a similar effectiveness in the screen.

In our Bayesian methods, we preestimate some parameters in the priors, resulting in a fast and efficient algorithm. There are two reasons for the preestimation. One is to reduce computational time because it is critical to keep computational time as low as possible when analyzing large data sets from high-throughput screens. The second reason is to focus on selecting hits only in plates with good data quality, because hit selection results are often misleading in plates with poor quality (24–29). In the Bayesian models described in this article, we use the fact that the unconditional variance of an observed value should be larger than the expectation of its conditional variance (i.e. $Var(X_i) > E[Var(X_i|\mu_i)] = \sigma^2$) in a plate with good quality. In a plate with poor data quality, the pooled variance from different controls (i.e. estimate of σ^2) could be greater than the variability of sample wells (i.e. estimate of $Var(X_i)$) in a plate. In such a case, the estimate of τ^2 is zero (i.e. $\hat{\tau}^2 = 0$); consequently, the FDR is unavailable in that plate when using the above Bayesian models.

One strategy to handle the situation with $\hat{\tau}^2 = 0$ is the use of quality control methods described in (24–29) to assure all the plates have good quality during the experimental stage of a screen, thus preventing the occurrence of $\hat{\tau}^2 = 0$. Another strategy is to select hits in the plates with $\hat{\tau}^2 = 0$ using non-Bayesian methods such as those described in (5–7,10,13,17) and then to put a warning on all the hits from these plates. The third strategy is to build a Bayesian hierarchical model that puts priors on τ^2 and/or other preestimated parameters. For example, we may construct a hierarchical model as follows: $\pi(\tau) \propto \tau/(\tau^2 + \sigma^2)$, $\mu_i|\tau \sim N(\theta_0, \tau^2)$ and $X_i|\mu_i \sim N(\mu_i, \sigma^2)$ where θ_0 and σ^2 are preestimated as in Model 1. This leads to the posteriors:

$$\tau^2 + \sigma^2|X_i \sim \text{Inverse Gamma}\left(\frac{1}{2}, \frac{(X_i - \theta_0)^2}{2}\right)$$

and

$$\mu_i|\tau, X_i \sim N\left(\frac{X_i\tau^2 + \theta_0\sigma^2}{\tau^2 + \sigma^2}, \frac{\tau^2\sigma^2}{\tau^2 + \sigma^2}\right).$$

Based on these posteriors, we make inference on μ_i using direct sampling in the Monte Carlo simulation.

There are two major issues with hierarchical models: one is the long computational time and the other is that plates with good or poor quality are treated in the same way. For example, for the simple hierarchical model described above, we need to run direct sampling on each of 25 000 siRNAs tested. The number of iterations in the Monte Carlo simulation for each siRNA has to be large. Otherwise, due to sampling error, we may obtain incorrect results: i.e. a more effective siRNA may have less chance to be selected as a hit than a less effective siRNA in the same plate. Even if only 2000 iterations are used for each siRNA, this will result in a total of 50 000 000 iterations

when analyzing data from an HTS screen. Building more layers in a hierarchical model will result in even longer computational time, especially when there is at least one posterior requiring an indirect sampling in the Monte Carlo simulation (30). Therefore, in genome-wide RNAi primary screens, we do not recommend the use of Bayesian hierarchical models (especially those requiring the use of indirect sampling) although we may use them in confirmatory screens.

In summary, hit selection in genome-scale RNAi research is important for identifying targets for a new class of therapeutics for treating human diseases. The development of novel effective and powerful analytic methods for hit selection is critical to glean useful information from mounds of data. In this article, we propose a methodology based on a Bayesian framework. The case studies show that our methods effectively address multiple issues in classical methods including error rate issues in multiple hypothesis testing, experiment-wise versus plate-wise analysis and incorporation of information from multiple reliable controls. Both simulation and case studies show that the Bayesian Model 1 we describe is more powerful than classical methods in detecting siRNAs with weak and moderate effects. Our Bayesian Model 2 performs better than classical experiment-wise methods when there are no plates with enriched number of hits and performs equivalently to experiment-wise classical methods in detecting hits with strong effects when there are plates with enriched number of hits in an RNAi HTS experiment.

REFERENCES

1. Fire,A., Xu,S.Q., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, **391**, 806–811.
2. Hannon,G.J. and Zamore,P.D. (2003) Small RNAs, big biology: biochemical studies of RNA interference. In Hannon,G.J. (ed.), *a Guide to Gene Silencing*. Cold Spring Harbor Laboratory Press, New York, pp. 87–108.
3. Bumcrot,D., Manoharan,M., Koteliansky,V. and Sah,D.W.Y. (2006) RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nature Chemical Biology*, **2**, 711–719.
4. Eisenstein,M. (2006) Quality control. *Nature*, **442**, 1067–1070.
5. Zhang,X.D., Ferrer,M., Espeseth,A.S., Marine,S.D., Stec,E.M., Crackower,M.A., Holder,D.J., Heyse,J.F. and Strulovici,B. (2007)

The use of strictly standardized mean difference for hit selection in primary RNA interference high-throughput screening experiments. *J. Biomol. Screen.*, **12**, 497–509.

6. Malo,N., Hanley,J.A., Cerquozzi,S., Pelletier,J. and Nadon,R. (2006) Statistical practice in high-throughput screening data analysis. *Nat. Biotechnol.*, **24**, 167–175.

7. Zhang,X.D. (2007) A new method with flexible and balanced control of false negatives and false positives for hit selection in RNA interference high-throughput screening assays. *J. Biomol. Screen.*, **12**, 645–655.

8. Gou,D., Narasaraju,T., Chintagari,N.R., Jin,N., Wang,P.C. and Liu,L. (2004) Gene silencing in alveolar type II cells using cell-specific promoter in vitro and in vivo. *Nucleic Acids Res.*, **32**, e134.

9. Hsieh,A.C., Bo,R.H., Manola,J., Vazquez,F., Bare,O., Khvorova,A., Scaringe,S. and Sellers,W.R. (2004) A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.*, **32**, 893–901.

10. Zhang,X.D. (2008) Genome-wide screens for effective siRNAs through assessing the size of siRNA effects. *BMC Research Notes* **1**, 33.

11. Berns,K., Hijmans,E.M., Mullenders,J., Brummelkamp,T.R., Velds,A., Heimerikx,M., Kerkhoven,R.M., Madiredjo,M., Nijkamp,W., Weigelt,B. *et al.* (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature*, **428**, 431–437.

12. Boutros,M., Bras,L.P. and Huber,W. (2006) Analysis of cell-based RNAi screens. *Genome Biol.*, **7**, R66.

13. Chung,N.J., Zhang,X.D., Kreamer,A., Locco,L., Kuan,P.F., Bartz,S., Linsley,P.S., Ferrer,M. and Strulovici,B. (2008) Median absolute deviation to improve hit selection for genome-scale RNAi screens. *J. Biomol. Screen.*, **13**, 149–158.

14. Espeseth,A.S., Huang,Q., Gates,A., Xu,M., Yu,Y., Simon,A.J., Shi,X.P., Zhang,X., Hodor,P., Stone,D.J. *et al.* (2006) A genome wide analysis of ubiquitin ligases in APP processing identifies a novel regulator of BACE1 mRNA levels. *Mol. Cell. Neurosci.*, **33**, 227–235.

15. Haney,S.A. (2007) Increasing the robustness and validity of RNAi screens. *Pharmacogenomics*, **8**, 1037–1049.

16. Pelkmans,L., Fava,E., Grabner,H., Hannus,M., Habermann,B., Krausz,E. and Zerial,M. (2005) Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature*, **436**, 78–86.

17. Zhang,X.D., Yang,X.C., Chung,N.J., Gates,A., Stec,E., Kunapuli,P., Holder,D.J., Ferrer,M. and Espeseth,A.S. (2006) Robust statistical methods for hit selection in RNA interference high-throughput screening experiments. *Pharmacogenomics*, **7**, 299–309.

18. Gunter,B., Brideau,C., Pikounis,B. and Liaw,A. (2003) Statistical and graphical methods for quality control determination of high-throughput screening data. *J. Biomol. Screen.*, **8**, 624–633.

19. Newton,M.A., Noueiry,A., Sarkar,D. and Ahlquist,P. (2004) Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics*, **5**, 155–176.

20. Benjamini,Y. and Hochberg,Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*, **57**, 289–300.

21. Storey,J.D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **64**, 479–498.

22. Joyce,J.G., Hurni,W.M., Bogusky,M.J., Garsky,V.M., Liang,X.P., Citron,M.P., Danzeisen,R.C., Miller,M.D., Shiver,J.W. and Keller,P.M. (2002) Enhancement of alpha-helicity in the HIV-1 inhibitory peptide DP178 leads to an increased affinity for human monoclonal antibody 2F5 but does not elicit neutralizing responses in vitro - Implications for vaccine design. *J. Biol. Chem.*, **277**, 45811–45820.

23. Zuck,P., Murray,E.M., Stec,E., Grobler,J.A., Simon,A.J., Strulovici,B., Inglese,J., Flores,O.A. and Ferrer,M. (2004) A cell-based beta-lactamase reporter gene assay for the identification of inhibitors of hepatitis C virus replication. *Anal. Biochem.*, **334**, 344–355.

24. Zhang,X.D. (2007) A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics*, **89**, 552–561.

25. Iversen,P.W., Eastwood,B.J., Sittampalam,G.S. and Cox,K.L. (2006) A comparison of assay performance measures in screening assays: Signal window, Z' factor, and assay variability ratio. *J. Biomol. Screen.*, **11**, 247–252.

26. Sui,Y.X. and Wu,Z.J. (2007) Alternative statistical parameter for high-throughput screening assay quality assessment. *J. Biomol. Screen.*, **12**, 229–234.

27. Zhang,J.H., Chung,T.D.Y. and Oldenburg,K.R. (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen.*, **4**, 67–73.

28. Zhang,X.D. (2008) Novel analytic criteria and effective plate designs for quality control in genome-wide RNAi screens. *J. Biomol. Screen.*, **13**, 363–377.

29. Zhang,X.D., Espeseth,A.S., Johnson,E.N., Chin,J., Gates,A., Mitnaul,L.J., Marine,S.D., Tian,J., Stec,E.M., Kunapuli,P. *et al.* (2008) Integrating experimental and analytic approaches to improve data quality in genome-wide screens. *J. Biomol. Screen.*, **13**, 378–389.

30. Zhang,X.D., Roeder,K., Wallstrom,G. and Devlin,B. (2003) Integration of association statistics over genomic regions using Bayesian adaptive regression splines. *Hum. Genomics*, **1**, 20–29.