4

Identifying siRNA-Induced Off-Targets by Microarray Analysis

Emily Anderson, Queta Boese, Anastasia Khvorova, and Jon Karpilow

Summary

RNA interference (RNAi) is an endogenous gene regulatory pathway that the research community has adopted to facilitate the creation of a functional map of the human genome. To achieve this, small interfering RNAs (siRNAs), short synthetic duplexes having complete homology to the intended target, are introduced into cells to silence gene expression via a posttranscriptional cleavage mechanism. While siRNAs can be designed to effectively knock down any target gene, recent studies have shown that these small molecules frequently trigger off-target effects. These unintended events can have a significant impact on experimental outcomes and subsequent data interpretation. As RNAi is envisioned to play a central role in developing a functional map of the human genome, the development of reliable protocols for identifying off-targeted genes is essential. This chapter focuses on the underlying features of siRNA-mediated off-targeting and the state-of-the-art methodology used to identify off-targeted genes via microarray-based gene expression analysis. Future adoption of standards in this field will allow a clean distinction between sequence-specific off-target gene regulation and other forms of gene modulation resulting from delivery effects and other events unrelated to the RNAi pathway.

Key Words: siRNA; RNAi; microarray; off-target effect; transfection.

1. Introduction

1.1. miRNA Targeting and siRNA Off-Targeting: The Caveats of Adopting Biological Pathways

Detailed studies of the endogenous substrates of the RNAi pathway (miRNA) **(Fig. 1)** have provided keen insights into the source of siRNA

From: *Methods in Molecular Biology, vol. 442: RNAi: Design and Application*
Edited by: S. Barik © Humana Press, Totowa, NJ

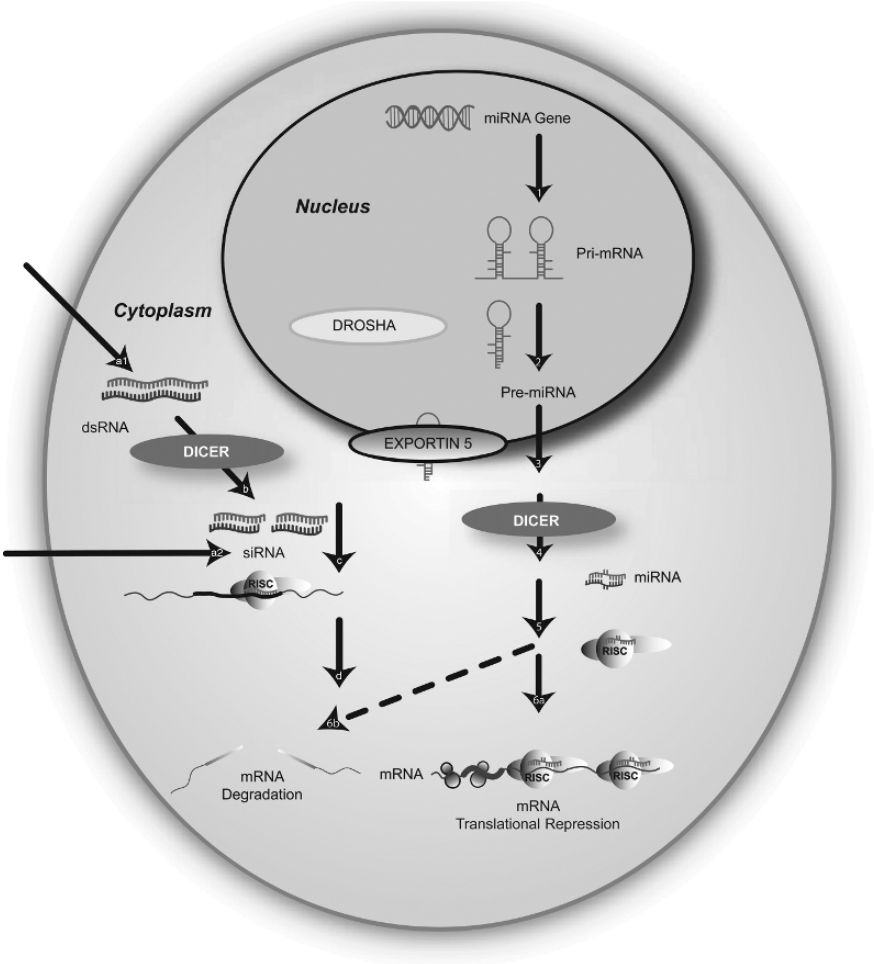

Fig. 1. The endogenous RNAi pathway depicting the key steps associated with the processing of and silencing by miRNAs and synthetic siRNAs. In the endogenous pathway: (1) miRNAs are transcribed forming complex stem-loop pri-miRNA structures; (2) Drosha processes these primary transcripts into pre-miRNAs in the nucleus; (3) pre-miRNAs are transported to the cytoplasm via Exportin 5; (4) Dicer further processes the pre-miRNAs into double-stranded miRNAs with characteristic mismatches and bulges; (5) RISC loads the active strand to form an activated complex and (6a) effect translational repression in cases of partial identity within the 3′-UTR of a given target or (6b) mRNA cleavage in cases of complete identity. mRNA regulation is also mediated by the introduction of long dsRNAs (a1) or siRNAs (a2) directly into the cytoplasm. In the case of long dsRNAs, these molecules are processed by DICER (b) into siRNAs that interact with RISC (c) and effect RNAi silencing via degradation of the mRNA target (d).

off-target effects. After undergoing a complex process of maturation (pri-miRNA → pre-miRNA → mature miRNA) that involves both nuclear and cytoplasmic RNase III endonucleases (Drosha and Dicer, respectively), the guide strand of miRNAs mediates translation attenuation of prospective targets through annealing of the miRNA seed region (positions 2–7) to complementary sites in the 3′ UTR of the target gene ***(1,2)***. These seemingly simple interactions allow modest (~2–4-fold) changes in the levels of gene expression and lead to the orchestrated suppression of dozens to hundreds of genes by a single miRNA.

While the primary mode of action of siRNAs is distinctly different from that of miRNAs (i.e., target cleavage vs. translation attenuation), there are several overlapping attributes that tie miRNA gene targeting with siRNA off-targeting. Like miRNAs, siRNA off-targeting induces modest changes in the expression of dozens to hundreds of messages. Moreover, studies by a number of groups have demonstrated a strong association between a transcript being an off-target and the presence of one or more siRNA seed complements in the 3′ UTR of off-targeted genes. Given the parallels among the number of targets, the level of gene modulation, and the apparent importance of 3′ UTR target sites for both molecules, the overall sentiment by the research community is that off-targeting by siRNA results from the promiscuous entry of these molecules into the miRNA pathway.

1.2. Techniques for Identifying Off-Targets

The mounting evidence that suggests the presence of off-targets has been derived from multiple sources. In an investigation of 10 separate siRNAs targeting the *MEN1* gene, Scacheri and colleagues used Western blot analysis to characterize the unexpected variation in the levels *p21* and *p53*, two genes considered to be indicators of overall changes in cellular physiology ***(3)***. As each siRNA altered expression of these proteins by differing amounts, the authors predicted that the siRNA affected the expression of targets other than *MEN1* and that these unintended targets disproportionately impacted *p21* and *p53* expression. In a different approach, Fedorov and coworkers employed a phenotypic assay that monitored cell viability to distinguish differences in siRNAs targeting a nonessential housekeeping gene, *PPIB* ***(4)***. Akin to Scacheri et al.'s findings, each duplex induced distinctly different levels of RNAi-dependent toxicity, thus lending further support to the notion that siRNAs modulate expression of genes other than the intended target.

While these and similar approaches provided evidence of siRNA off-target activity, the most definitive data that quantified off-targeted genes were first reported by Jackson et al. ***(5)***. Using microarray-based gene expression

profiling, the group monitored the genome-wide effects resulting from siRNA transfection and identified unique off-target signatures for multiple duplexes. Subsequent applications of microarray technologies have been responsible for the identification of (1) strong similarities between siRNA off-targeting and miRNA targeting mechanisms ***(6–8)***, (2) the importance of the 3′ UTR transcriptome seed complement frequency (SCF) in defining the extent of off-target signatures ***(9)***, and (3) strand-specific modification patterns that greatly enhance siRNA specificity ***(10)***. Thus, gene expression technologies have served to reveal the global effects of siRNAs and focus efforts toward improving and enhancing siRNA specificity.

1.3. Application of Microarrays to Off-Target Analysis

Successful application of microarray-based technologies to off-target studies requires attention to three principal phases: pre-experimental assay optimization, global gene expression profiling, and post-experimental data analysis **(Fig. 2)**. To facilitate the generation of reliable data sets, rigorous optimization of cell lines, plating conditions, delivery methods, and siRNA concentrations must be performed to minimize the confounding effects that can result from small alterations in experimental protocols. Once optimal conditions for potent silencing (>75% reduction of mRNA) without adverse viability effects (<10% cell death) are identified, the second phase, which includes technology selection, sample collection and preparation, hybridization, and data collection, can follow. Final data analysis requires the selection of software packages that are compatible with the chosen microarray technology and permit discrimination of off-targeted genes from genes whose expression changes are associated with delivery and/or downstream events that accompany targeted gene knockdown (*see* **Note 1**).

1.4. Global Gene Expression Profiling

1.4.1. Choice of Technology

A critical decision that must be addressed before performing any microarray-based off-target analysis involves the choice of array technologies. Current platforms include both cDNA and oligonucleotide-based arrays produced by a collection of commercial entities (e.g., Agilent, Affymetrix) that also provide optimized protocols for RNA purification, labeling, hybridization, and scanning. Though cross-platform concurrence studies have not been performed for off-target analysis, there is strong reproducibility among the various platforms for global gene profiling ***(28–32)***. Given this relative level of parity, issues associated with probe density, availability of organism-specific arrays,

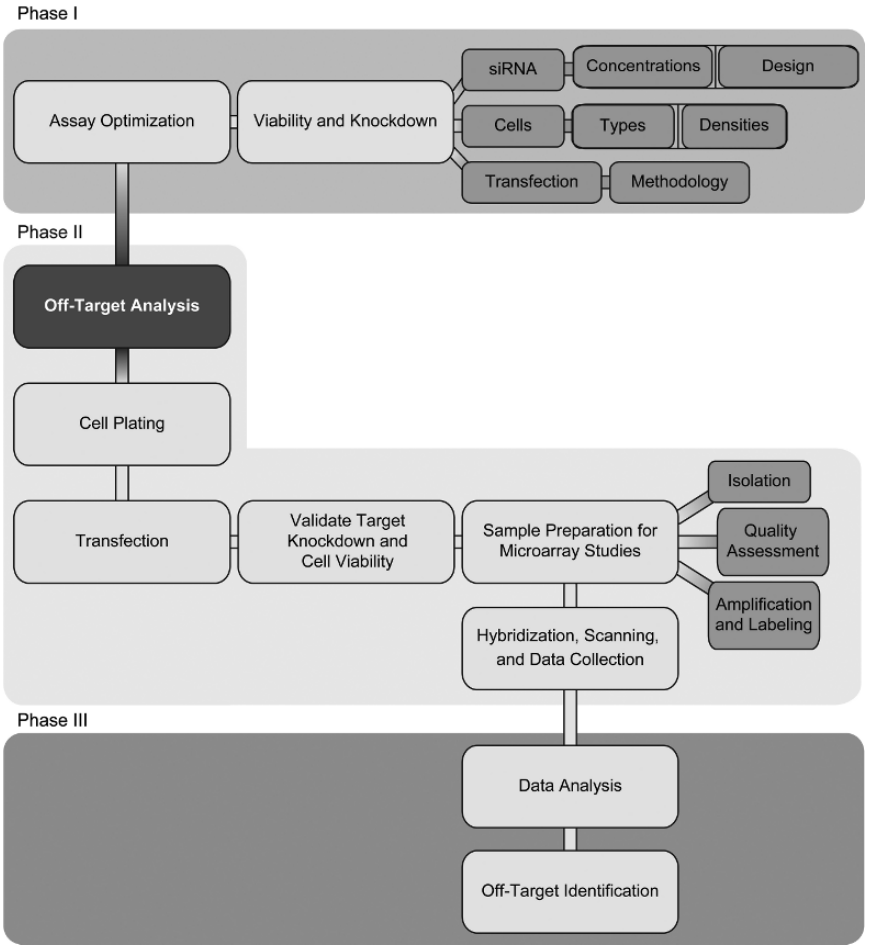

Fig. 2. General workflow for off-target analysis. Phase I assay optimization studies (involving the identification of acceptable siRNA concentrations and designs, cell types and densities, and transfection reagents and methodologies) are mandatory and precede all microarray-based studies. Experimental optimization is followed by transfections, RNA preparation, hybridization, and data collection (Phase II), and final data analysis, and off-target identification (Phase III).

up-to-date annotation, costs, ability to detect and distinguish variant splice forms, and available space for data storage are the critical areas of focus when deciding among technologies.

In our lab, we have adopted a dual-color spotted oligo microarray (Agilent Human 1A V2 arrays) that utilizes 60-mer *in situ* synthesized probes representing ~18,000 distinct gene IDs from the human transcriptome. The

two-dye system provides a high degree of specificity and sensitivity and allows for the incorporation of a reference sample with each array in a block reference design format ***(33,34)***. Total RNA preparation is performed using RNeasy kits (Qiagen), and RNA quality is assessed using the 2100 Bioanalyzer (Agilent). A Low Input RNA Fluorescent Linear Amplification Kit (Agilent) is used for RNA amplification/labeling and employs both Cy5™ and Cy3™ dyes for experimental and control samples, respectively. A NanoDrop spectrophotometer (ND-1000, NanoDrop) is used to assess the success of the amplification and labeling reactions.

Once a technology has been adopted, we typically perform a pilot study to verify compatibility between the final transfection conditions and the microarray technology selected. This can be achieved by hybridizing RNA from mock- and untransfected cells along with a self-self control array (e.g., mock-to-mock or untransfected-to-untransfected). These simple experiments allow one to determine if the transfection conditions optimized on the basis of gene knockdown and cell viability are themselves invasive at the more sensitive threshold of gene expression. Further expansion of the pilot studies can provide insights into the level of variability observed between biological replicates and permit protocol refinements to maximize identification of off-targeted genes (*see* **Notes 2–5**).

1.4.2. Data Analysis

Data processing and analysis depend on the platform used and the scope of the experiment. Several freeware gene expression analysis packages [such as Bioconductor (http://www.bioconductor.org/) and Cluster and Treeview from the Eisen lab (http://rana.lbl.gov/EisenSoftware.htm)] are available as well as more sophisticated commercial solutions [Spotfire DecisionSite for Functional Genomics and Spotfire DecisionSite for Microarray Analysis (Spotfire, Somerville, MA); Genespring (Agilent, Santa Clara, CA); Rosetta Resolver (Rosetta Biosoftware, Seattle, WA)]. Software selection depends on one's budgetary restrictions, available hardware, the intended depth of the analysis, and the flexibility and ease-of-use. At the very minimum, one must have the ability to process raw images (software usually supplied by the array manufacturer) and basic statistical analysis and graphing solutions to assess data quality. Among the RNAi research team at ThermoFisher Scientific (Dharmacon products), data sets are generally filtered in stages that take into consideration controls, flagged outliers, and measurements of absolute or relative signal intensity. Second-stage filtering often includes metrics that identify significant levels of expression change (by using an error model, a fold-change cutoff, or a combination of both). Relevant probes are then visualized via heatmap representation after being subjected to unsupervised

clustering methods (hierarchical clustering) to distinguish sample-specific gene expression changes (off-targets) as opposed to more general effects related to transfection or downstream gene modulation resulting from target-specific knockdown. Lists of gene names, accession numbers, and fold changes can subsequently be exported for further sequence or functional analysis.

2. Materials

2.1. Transfection

1. HyQ MEM-RS: 500 mL of HyClone (ThermoFisher Scientific, Waltham, MA, Cat#. SH3056401).
2. 96-Well plates *(2)* NUNC (ThermoFisher Scientific, Waltham, MA, Cat#. 12-565-383).
3. Deep Well Titer Plate (polystyrene): Nalge-Nunc (ThermoFisher Scientific, Waltham, MA, Cat#. 12-565-605).
4. Dharma*FECT*™ 1 siRNA Transfection Reagent (Dharmacon Products, Lafayette, CO, T-2001-03).
5. siRNA Controls:
   a. siGENOME® GAPD Control siRNA (ThermoFisher Scientific, Dharmacon Products, Lafayette, CO, Cat# D-001136-01): positive control for siRNA-mediated knockdown.
   b. Dharmacon® RISC-Free® Control siRNA (ThermoFisher Scientific, Dharmacon Products, Cat# D-001220-01): negative control for siRNA knockdown.
   c. PLK1 *SMART*pool reagent-positive control for transfection (ThermoFisher Scientific, Dharmacon Products, Cat# 003290): positive control for transfection.

2.2. Microarray Analysis

The various kits and specialized reagents are described in the corresponding sections in **Section 3**.

3. Methods

We provide a detailed description for performing off-target analysis using the Agilent microarray platform. To enhance the comprehension of this section, we present the protocol in the context of performing analysis on an siRNA targeting MAP2K1 (sense strand: 5′ GAGCAGATTTGAAGCAACT). For those in need of more general descriptions pertaining to the design and analysis of microarray experiments, a number of resources including *A Beginner's Manual*

to Microarrays by Eric M. Blalock (Kluwer Academic Press), *Microarray Methods and Applications: Nuts and Bolts* by Gary Hardiman (DNA Press), and *Microarray Methods and Protocols* by Robert Matson (CRC Press) are readily available. For additional information on instrumentation and software packages that are compatible with the Agilent platform described below, we suggest visiting the Agilent Web site at www.agilent.home.com.

For off-target analysis by microarray, duplicate plates (one for transfection validation and a second for microarray analysis) are prepared side-by-side. Following confirmation that the transfection was successful, the subsequent workflow includes (1) collection and labeling of samples, (2) hybridization, and (3) data collection (**Fig. 2**). Though the following protocol describes preparations for a single experiment, typically two or more biological replicates (different cell cultures) are performed to heighten confidence in identifying off-targeted genes.

3.1. Transfection

1. Cell plating: In order to have sufficient numbers of cells for (a) three 370 siRNA test samples, (b) a mock-treated (lipid alone) control, (c) untreated controls, and (d) transfection controls, generate the plate design described in **Fig. 3**. To achieve this, plate HeLa cells (strain CCL-2, ATCC, Manassas, VA) in 96-well plates at 10,000 cells per well 24 h prior to transfection and in the absence of antibiotics.
2. Preparation of siRNA for transfection: To prepare samples for transfection, aliquot 30 µL of RS-MEM into each well of a 96-well deep-dish plate. Subsequently add 2.5 µL of a 10 µM stock solution of the appropriate siRNA to each well. In a separate tube, combine 4 mL of HyQ MEM-RS with 60 µL of Dharma*FECT* 1. Gently mix the diluted transfection reagent and add 30 µL to the siRNA mixture in each well of the deep-well plate. Mix the resulting siRNA–lipid cocktail gently and allow to sit for 30 min in order for the liposome complexes to form. At the end of this incubation, add 190 µL of the appropriate cell culture media (without antibiotics) to each of the wells containing the siRNA–lipid cocktail.
3. Transfection: Remove the media from the 96-well plates containing the pre-plated cells and replace it with 100 µL of the siRNA–lipid–media cocktail. Return the plates to the incubator and culture the cells at 37 ºC (5% CO_2) overnight.

3.2. Preparation of Samples for Microarray Analysis and Verification of Transfection

1. Twenty-four hours after transfection, aspirate the wells designated for microarray analysis to remove the media. Lyse the cells using 70 µL of RLT buffer (Qiagen, Cat# 79216) with a 1:100 dilution of β-mercaptoethanol per the manufacturer's instructions. Pool twelve identical wells of lysate per transfection sample for RNA purification and store samples is not conducted on the same day.

Panel A - Master Template

	1	2	3	4	5	6	7	8	9	10	11	12
A	sample 1	sample 1	sample 1	sample 1	sample 1	sample 1	sample 1	sample 1	sample 1	sample 1	sample 1	sample 1
B	sample 2	sample 2	sample 2	sample 2	sample 2	sample 2	sample 2	sample 2	sample 2	sample 2	sample 2	sample 2
C	sample 3	sample 3	sample 3	sample 3	sample 3	sample 3	sample 3	sample 3	sample 3	sample 3	sample 3	sample 3
D	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated
E	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated	untreated
F	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	siCONTROL GAPDH	siCONTROL GAPDH	siCONTROL GAPDH
G	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	siCONTROL Risc Free	siCONTROL Risc Free	siCONTROL Risc Free
H	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	MOCK	PLK 1 SMART reagent	PLK 1 SMART reagent	PLK 1 SMART reagent

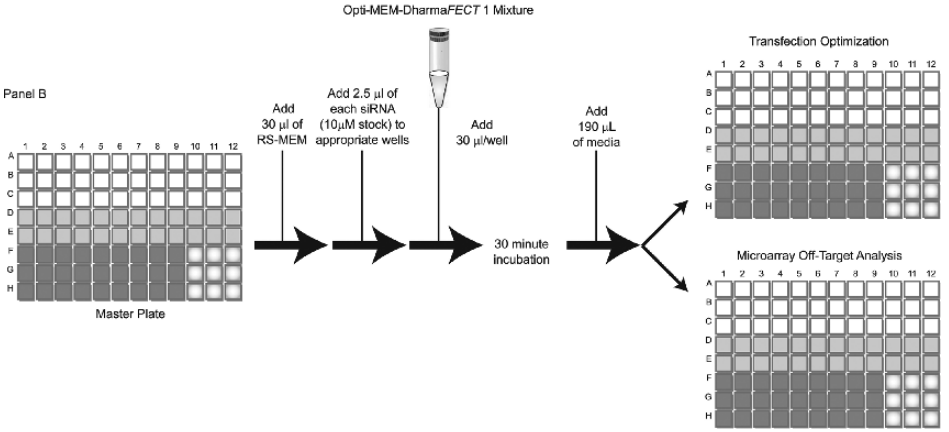

Fig. 3. An example 96-well plate map for microarray-based off-target analysis. Samples 1, 2, and 3 (rows A, B, C) represent wells with cells transfected with three different siRNA. "Untreated" and "Mock" treated wells are used as references on the Agilent platform (rows D–H). Dharmacon® siGENOME® GAPD Control siRNA, Dharmacon® RISC-Free® Control siRNA, and *PLK1* SMART pool Reagent occupy the wells of the lower right quadrant. These wells are used as positive and negative controls for knockdown and to control for transfection efficiency, respectively, in their appropriate assays. They are not collected for microarray analysis.

2. Wells dedicated to assessing transfection efficiency need to be evaluated for (a) culture viability using the alamarBlue™ assay (Biosource Intl., Camarillo, CA), and (b) siRNA-mediated knockdown of positive and negative control target genes using bDNA (Quantigene, Panomics, Fremont, CA). These two assays are compatible and can be run on a single plate (viability assay followed by knockdown efficiency assay). Transfections that pass both viability (90–95% viability) and knockdown thresholds (90% knockdown of a positive-control siRNA as normalized to a mock or untransfected control) are acceptable for microarray analysis.

3.3. Assessing Quantity and Quality of RNA for Microarray Analysis

1. Isolate total RNA from cell lysates using RNeasy Mini columns (Qiagen) with on-column DNase digestion per the manufacturer's protocol (http://www1.qiagen.com/literature/handbooks/PDF/RNAStabilizationAndPurification/FromAnimalAndPlantTissuesBacteriaYeastAndFungi/RNY_Mini/1035969_HB_BenchProtocol.pdf).
2. Assess total RNA yields using a NanoDrop (NanoDrop, Wilmington, DE). Samples are adjusted to 100 ng/μL (typical yields >100 ng/μL). Store samples at –80 ºC.
3. Assess the quality of the RNA preparation using a 2100 Bioanalyzer (Agilent) with the RNA 6000 Nano LabChip. Acceptable samples should have an RNA Integrity Number (RIN) of at least 9 and can be stored at –80 ºC.

3.4. RNA Amplification and Labeling

1. Amplify total RNA and label using the Low Input RNA Fluorescent Linear Amplification Kit (Agilent, Cat# 5184-3523) according to the manufacturer's protocol (http://www.chem.agilent.com/). Cy5™ CTP and Cy3™ CTP are obtained from Perkin Elmer (Wellesley, MA). Typically, 650 ng of total RNA is used per reaction. Experimental samples are labeled with Cy5™ while mock (lipid-treated) samples are labeled separately with Cy5™ and Cy3™. Untransfected samples are labeled with Cy3™. Remove excess dye from the labeled samples with Qiagen RNeasy Mini columns, implementing two 1-min. "dry" spins to remove excess ethanol (with a 180 degree rotation of the column in between each spin) before sample elution.
2. Quantify yields of labeled RNA: Yields of labeled samples are quantified with a NanoDrop instrument (NanoDrop). Typical yields are 300–500 ng/μL starting with 650 ng of total RNA and should be uniformly colored depending upon the fluorophore (pink for Cy3™, blue for Cy5™). Assess dye incorporation by measuring the respective dye wavelengths (Cy3™: 550 nm, Cy5™: 650 nm) with the "microarray" protocol on the NanoDrop instrument.

3.5. Hybridization and Scanning

1. Perform hybridizations to Human 1A (V2) microarray (Agilent, Cat# G4110B) according to the manufacturer's instructions using 750 ng of both Cy5™ and Cy3™ labeled sample per array. Hybridizations include:

    a. Each siRNA-transfected sample (Cy5™) is hybridized along with the mock-transfected (Cy3™-labeled) sample as a reference.
    b. A control array for gene expression effects of transfection is assembled by using mock-transfected (Cy5™) and untransfected (Cy3™) samples.
    c. A control array for assessing signal/expression levels in the absence of siRNA treatment as well as for determining dye-biased targets is assembled with mock-transfected (Cy5) and mock-transfected (Cy3) (a self-self array).

2. After a 17–20-hour hybridization at 60 ºC, wash arrays for 1 minute each in 6X and then 0.06X SSPE containing 0.025% N-lauroylsarcosine. Subsequently, submerge the arrays in acetonitrile and wash in nonaqueous drying and stabilization buffer (Agilent, Cat# 5185-5979).
3. Scan arrays using Agilent's model G2505B Microarray scanner and process the raw .tiff images according to the default protocols in Feature Extraction v 9.1 (Agilent).

3.6. Data Analysis

1. Data analysis is initiated in Microsoft Excel. Construct a flat text file of feature identifier columns (Feature Number, Row, Column, Control Type, Gene Name, Systematic Name, and Description if available) and select data columns for each sample (Log Ratio, PValue LogRatio, gProcessedSignal, rProcessedSignal, gIsWellAboveBG, rIsWellAboveBG). Record the processed data for saturated signals from the mock-mock array in separate columns [saturated green (gIsSaturated) and red signals (rIsSaturated), respectively] along with a column for the log of combined red and green signals (gProcessedSignal + rProcessedSignal) computed from the mock-mock array.
2. Conduct further analysis in Spotfire DecisionSite using the Functional Genomics Module. First, filter the data by control and low-signal/low-expression probes to eliminate sequences for which it would be difficult to accurately determine a twofold reduction in signal (Log Ratio –0.3). This is achieved by filtering probes for which the log (combined signal) from the mock-mock array is below 2.8, or approximately 630 intensity units. Assess the absolute range of processed signals among all of the arrays in a batch to make sure that each hybridization was successful. Remove probes with signals that saturate in either channel in the mock-mock array. Typically, by employing these procedures, the final probe set available for analysis is reduced from 21,073 to approximately 13,000–15,000 probes.
3. Remove probes with signals that do not change more than twofold in any sample (–0.3 < Log Ratio < 0.3). Plot the signals for the remaining probes on a heatmap after unsupervised hierarchical clustering (unweighted average, Euclidean distance, average value). Further comparison of siRNA-treated sample columns with mock-mock and mock-untransfected hybridizations will allow probes that appear to be differentially expressed in either of these control arrays or those that are inherently noisy (for instance, Log Ratio vs. mock of <–0.2 in three or more arrays in an experiment) to be discarded.
4. If a gene list is created and the numbers of genes representing off-targets are summed for each siRNA, duplicate probes should be filtered from the analysis. It should be noted that many genes are represented on the array as many as 10 times (either as replicate probes, different probes targeting the same gene, or probes to an older version of annotation to a given gene). Cross-comparison of results

obtained with these different probe sets can act as an internal control to confirm a particular gene is truly an off-target.

Using the protocols and siRNA targeting sequences described above, a clear off-target signature that includes 29 genes can be identified. An example of the identity of these targets and the relative level of downregulation by the MAPK siRNA are presented in **Table 1**.

Table 1
List of Genes That Are Typically Observed to Be Downregulated by Twofold or More (Log Ratio of < –0.3 or More) by the MAP2K1-Targeting siRNA Using the Experimental Conditions Described Here

Accession Number	Gene Name	Log Ratio
NM_139045	SMARCA2	–0.38
NM_013237	PX19	–0.36
NM_002960	S100A3	–0.41
NM_025075	NIF3L1BP1	–0.32
NM_014071	NCOA6	–0.39
NM_022149	MAGEF1	–0.33
NM_004120	GBP2	–0.32
NM_006756	TCEA1	–0.35
AB002370	KIAA0372	–0.33
NM_004176	SREBF1	–0.41
NM_004844	SH3BP5	–0.32
NM_001262	CDKN2C	–0.33
BC032468	DKFZp434D0215	–0.33
NM_016516	VPS54	–0.32
NM_012428	SDFR1	–0.33
NM_006628	ARPP-19	–0.37
NM_016542	MST4	–0.31
NM_015455	KIAA1194	–0.38
NM_002902	RCN2	–0.50
NM_032124	HDHD2	–0.33
NM_022740	HIPK2	–0.30
NM_003472	DEK	–0.43
NM_005120	TNRC11	–0.33
NM_013372	GREM1	–0.39
NM_003200	TCF3	–0.42
NM_031286	SH3BGRL3	–0.40
NM_138444	KCTD12	–0.40
NM_015134	M-RIP	–0.43
NM_012207	HNRPH3	–0.37

4. Notes

1. Assay optimization: As the goal of off-target analysis is to identify changes in gene expression profiles that are associated with the RNAi pathway, it is critical to minimize all other factors (RNAi-independent processes) that contribute to alterations in the global profile. High concentrations of lipid transfection reagents, and/or the presence of particular motifs, duplex lengths, or secondary structures have been shown to trigger the interferon response pathway and induce cytotoxicity. As these events have global repercussions on gene expression and are undetectable by monitoring target gene knockdown alone, assay optimization requires (at a minimum) the measurement of two interrelated parameters: target gene knockdown and cell viability.
2. Choice of technology: For assaying both of the above features, a wide array of technologies is available (branched DNA, quantitative PCR, CellTiter® Blue-, and MTT assays). Upon adopting a complementary pair, one can proceed to identify optimal conditions for multiple variables. These include selection of an appropriate cell line for the experimental model of interest and defining suitable plating densities, delivery methods, transfection times, and siRNA concentrations. Initial studies that employ highly functional siRNAs targeting nonessential genes [e.g., glyceraldehyde-3-phosphate dehydrogenase, *GAPDH* (NM_002046), or peptidylprolyl isomerase B/cyclophilin B, *PPIB* (NM_000942)] provide benchmarks against which general viability and target knockdown may be assessed.
3. Cell line: Recent unpublished microarray studies conducted in our laboratory have demonstrated that with few exceptions, off-targets observed in one cell type are consistently observed across unrelated cell types. For this reason, the selection of a cell type for off-target analysis is primarily driven by a limited set of criteria: ease of growth, transfectability, homogeneity, and resilience to manipulation. Though conceptually, both primary cells and immortal cell lines should be amenable to off-target studies, primary cells are generally avoided due to the potential presence of multiple cell types that require dissimilar transfection conditions. Instead, a diverse collection of cell lines are readily available from a number of institutions (e.g., the American Type Culture Collection, ATCC, Manassa, VA) and can be employed for off-target analysis. For most of our studies we use adherent cells such as HeLa, in part due to their robust growth in culture and the ease with which they can be transfected with lipid-based formulations like Dharma*FECT* 1-4 or comparable delivery reagents that have been developed for optimal delivery of small nucleic acids. Nonadherent cell lines can also be employed, but in many cases, these require alternative delivery methods that can (a) have deleterious effects on overall viability (electroporation) or (b) require significant investment in upfront reagent preparation (viral-based delivery).
4. Cell density: Once a cell line has been chosen, the plating density at which cells are transfected must be taken into consideration. In general, lipid-mediated transfection at low cell densities (e.g., 500–1,000 cells per well in a 96-well plate) often leads to a strong innate immune response and higher levels of cell death. In contrast, confluent monolayers tend to be refractory to this method of

transfection. As these changes can compromise off-target analysis, optimization studies typically include assessing a range of cell densities (e.g., 5,000, 10,000, 20,000, and 30,000 cells per well) that take into consideration culture format, cell size, cell growth rate, and sensitivity to crowding and transfection. It should be noted that while cell densities of 10–20,000 cells per well (in a 96-well format) are generally amenable to off-target profiling studies, minor alterations in transfection protocols (e.g., switching cell lines or lipid delivery formulations) demands a re-optimization of the cell density window.

5. Delivery methods: Lipid delivery reagents are currently the most common method of introducing synthetic siRNA into cells. In most cases, adherent cells are amenable to this mode of transfection, and appropriate conditions (formulation, concentration, transfection length) are readily identified. In contrast, nonadherent cell lines (e.g., THP-1, Jurkat cells) tend to be refractory to lipid-mediated transfection, and more specialized protocols (e.g., electroporation, nucleofection®, viral delivery, injection, etc.) often must be employed.

   Preferred lipid transfection reagents (a) efficiently deliver small nucleic acids, (b) are effective at low concentrations, (c) are nontoxic, and (d) have an extended shelf-life. Given this set of criteria, many of the commercially available lipid-delivery products that were originally designed for transfection of large DNA plasmids fall short as siRNA transfection reagents due to issues associated with cytotoxicity. These shortcomings have been addressed by a second generation of formulations [including Dharma*FECT*™ 1–4 (ThermoFisher Scientific, Dharmacon Products), siLentFect™ (BioRad), and HiPerFect™ (Qiagen)] that are compatible with small nucleic acid delivery. These new products are typically used at lower concentrations and exhibit fewer nonspecific, cytotoxic effects. Still, it is worth noting that no single reagent is universally applicable to all cell types. In a recent survey of five commercially available lipid transfection reagents in 20 different cell lines (using performance cutoffs of 80% silencing with less than 10% loss in cell viability), we observed that while certain cell lines were compatible with a wide array of reagents, others were more discriminatory (B. Robertson, personal communications). These findings suggest that cells respond differently to each lipid cocktail, a notion that is supported by microarray-based studies that identify both common and lipid-specific gene expression changes **(Fig. 4)**. For these reasons, when transfection conditions for a particular cell type have not been previously determined, a side-by-side comparison of several lipids should be made. In our laboratory, we typically perform an optimization matrix that tests 3 to 5 lipids at concentrations ranging from 0.025–1.6 μL per well (in a 96-well format) and regularly employ a highly functional control siRNA (>90% mRNA knockdown, 100 nM) targeting a nonessential gene (*GAPDH*, *PPIB*).

   In cases where cells are refractory to lipid-mediated delivery of siRNA, electroporation and DNA-based RNAi are often evaluated. Electroporation induces fewer nonspecific gene expression changes than lipid-mediated transfection ***(11)***, yet to minimize cytotoxicity this method still requires extensive optimization around parameters such as siRNA concentration, cell density, buffers, voltage, pulse

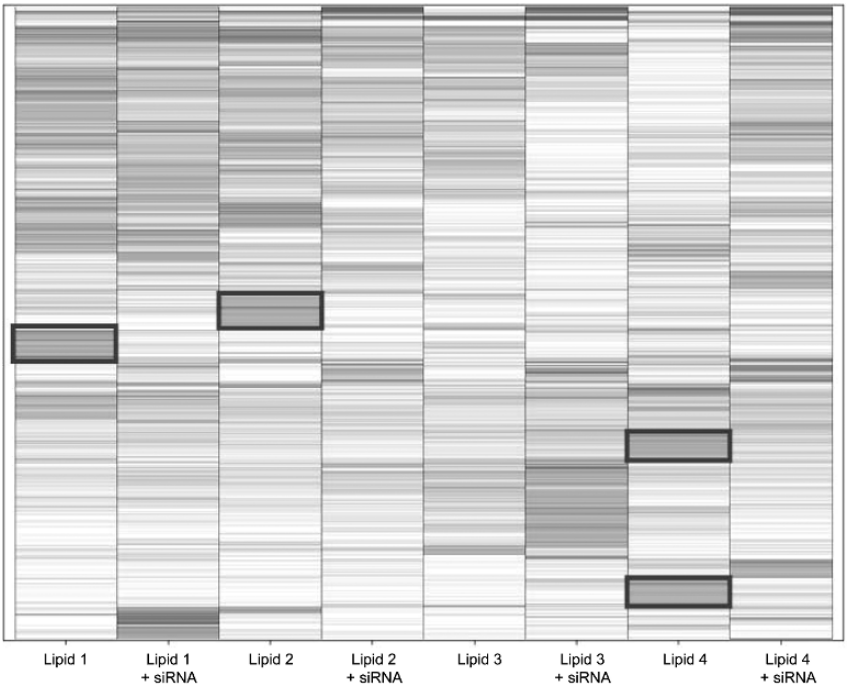

Fig. 4. Effects of lipid-based transfection on gene expression profiles. HeLa cells were transfected with four different lipid-based transfection in the presence and absence of a control siRNA (100 n*M*). As shown by the heatmaps for each treatment (normalized to mock or untransfected), lipid-based transfection induce both common and unique (boxes) changes in gene expression.

length and number, and interpulse interval length. Several manufacturers have developed specific buffers and preprogrammed protocols for cell types (e.g., amaxa Nucleofector® II), while alternative platforms provide extended flexibility by allowing for tailored protocols (e.g., Bio-Rad Gene Pulser XL).

Viral delivery systems offer another alternative for siRNA (or small hairpin RNA, shRNA) delivery and offer the added benefit of long-term silencing. Still, it is important to consider the potential shortcomings of these systems. First, work by Grimm and associates at Stanford ***(14)*** demonstrated that expression of shRNA from a viral construct can interfere with the processing of native RNAi substrates, miRNAs. Given this finding, it may be difficult to distinguish off-targets from changes in gene expression that result from perturbations in the endogenous miRNA processing mechanism. Second, because DNA-based RNAi provides long-term silencing, the profile changes observed in transduced cells will include both off-target effects as well as downstream effects resulting from silencing of the target gene. Differentiation between these two classes can be achieved

by simultaneously performing control studies with a second (equally functional) shRNA targeting the same gene, but this requires obvious additional costs and data analysis to correctly identify sequence-specific off-targets. Last, while there are multiple, commercially available custom viral-based silencing services (e.g., OpenBiosystems, Sigma-Aldrich), very little is known about how these molecules are processed. Studies by Vermeulen et al. demonstrated that Dicer processing of long siRNA (e.g., >23-mer duplexes) can often lead to multiple products ***(15)***. Given that the number and identity of most off-targets are determined by the 5′ terminal region of the guide strand (e.g., the seed region), the diverse population of molecules created by successive Drosha and Dicer processing of expressed hairpins could conceivably generate an expanded number of off-targeted genes.

6. siRNA attributes: Researchers planning to perform microarray-based off-target studies should also be conscious of the attributes associated with siRNA that contribute to the size, makeup, and intensity of the off-target signature. Studies by several groups have now demonstrated that siRNA off-targets are concentration-dependent ***(5)***, with off-target (and on-target) gene knockdown generally diminishing between 1–10 n*M* when lipid-mediated delivery technologies are employed. For these reasons, we typically perform off-target analysis at concentrations between 25–100 n*M*, which allows for clear identification of genes that are downregulated within 24 h after transfection. In addition to concentration, strand selection and nucleotide content play a significant role in the off-target signature. Off-target events can be induced by either strand of the siRNA duplex and both bioinformatic strategies and chemical modifications can be used to influence the contribution of either strand to the overall signature. Most siRNA rational design algorithms (which select functional sequences on the basis of thermodynamic and recognized nucleotide positional preferences) include differential end stability as a key element in siRNA selection to ensure RISC entry of the antisense strand. This process, which selects for disproportionate levels of stability at the duplex termini, can greatly minimize the sense strand off-target profiles and greatly simplify subsequent analysis. In addition, chemical modification patterns can also be incorporated into the sense and/or antisense strand to alter the off-target signature of each strand, respectively ***(10)***.

Specifically, modification patterns have now been identified that can (a) interfere with a strand's ability to enter RISC or (b) reduce the efficiency at which nonidentical pairings are permitted ***(10)***. While these techniques can help simplify subsequent off-target analysis, it is critical to note that in some instances, chemical modifications (as well as base-pair mismatches) can significantly alter off-target profiles. For example, in some cases, elimination of sense strand competition by chemical modification enhances the functionality of the antisense strand and amplifies the associated signature. In contrast, the addition of mismatches within the seed region (positions ~2–7) has been shown to shift

off-target profiles, further supporting bioinformatic studies that suggest siRNA off-targeting is similar to miRNA targeting ***(7)***.

Recent findings by our group have also shown that the number of off-targets generated by siRNA strongly correlates with the frequency at which the seed complement appears in the 3′ UTR transcriptome (referred to as the Seed Complement Frequency, SCF). siRNA with low SCF generally induce modest off-target profiles, while duplexes with medium-to-high SCF may exhibit more robust signatures. As the observed association has direct consequences on the rate at which false positives are generated in siRNA-mediated phenotypic screens, performing a preliminary bioinformatic analysis of the SCF for any siRNA can provide insights into the potential number and identity of off-targeted genes and aid in the selection of a duplex for future studies.

While concentration, strand selection, and seed frequency represent properties that can alter the off-target signature in an RNAi-dependent fashion, the length, structure, and sequence content of an siRNA can impact expression profiles through RNAi-independent mechanisms (*see* **Note 6**). Long siRNA (>23-mers) can induce a broad collection of IFN-response genes and toxicity at concentrations as low as 10 n*M **(25)***. In addition, particular sequence motifs and terminal end structures have likewise been shown to activate the innate immune response ***(26,27)***. Given these observations, we consistently utilize siRNA that are less than 23 bp in length and have been rationally designed to (a) bias antisense strand entry into RISC, (b) eliminate known sequences that induce the innate immune response, and (c) utilize seeds regions that have low SCFs.

References

1. Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* **11,** 1640–1647.
2. Rossi, J. J. (2005). RNAi and the P-body connection. *Nat. Cell Biol*. **7,** 643–644.
3. Scacheri, P. C., Rozenblatt-Rosen, O., Caplen, N. J., et al. (2004). Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl. Acad. Sci. USA* **101,** 1892–1897.
4. Fedorov, Y., Anderson, E. M., Birmingham, A., et al. (2006). Off-target effects by siRNA can induce toxic phenotype. *RNA* **12,** 1188–1196.
5. Jackson, A. L., Bartz, S. R., Schelter, J., et al. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol*. **21,** 635–637.
6. Lin, X., Ruan, X., Anderson, M. G., et al. (2005). siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res*. **33,** 4527–4535.

7. Jackson, A. L., Burchard, J., Schelter, J., et al. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12,** 1179–1187.
8. Birmingham, A., Anderson, E. M., Reynolds, A., et al. (2006). 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Meth.* **3,** 199–204.
9. Anderson, E. M., Birmingham, A., Baskerville, S., et al. (2008). Experimental Validation of the Importance of Seed Frequency to siRNA Specificity. RNA (in press).
10. Jackson, A. L., Burchard, J., Leake, D., et al. (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA* **12,** 1197–1205.
11. Fedorov, Y., King, A., Anderson. E., et al. (2005). Different delivery methods—Different expression profiles. *Nat. Meth.* **2**, 241.
12. Spagnou, S., Miller, A., and Keller, M. (2004). Lipidic carriers of siRNA: Differences in the formulation, cellular uptake, and delivery with plasmid DNA. *Biochemistry* **43,** 13348–13356.
13. Lv, H., Zhang, S., Wang, B., Cui, S., and Yan, J. (2006). Toxicity of cationic lipids and cationic polymers in gene delivery. *J. Control Release* **114,** 100–109.
14. Grimm, D., Streetz, K. L., Jopling, C. L., et al. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441,** 537–541.
15. Vermeulen, A., Behlen, L., Reynolds, A., et al. (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA* **11,** 1–9.
16. Harborth, J., Elbashir, S. M., Vandenburgh, K., et al. (2003). Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Nucleic Acid Drug Dev*. **13,** 83–105.
17. Khvorova, A., Reynolds A., and Jayasena, S. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216.
18. Chalk, A. M., Wahlestedt, C., and Sonnhammer, E. L. (2004). Improved and automated prediction of effective siRNA. *Biochem. Biophys. Res. Commun*. **319,** 264–274.
19. Ding, Y., Chan, C. Y., and Lawrence, C. E. (2004). Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res*. **32**(Suppl_2), W135–W141.
20. Henschel, A., Buchholz, F., and Habermann, B. (2004). DEQOR: A web-based tool for the design and quality control of siRNAs. *Nucleic Acids Res*. **32**(Suppl_2), W113–W120.
21. Naito, Y., Yamada, T., Ui-Tei, K., Morishita, S., and Saigo, K. (2004). siDirect: Highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res*. **32**(Suppl_2), W124–W129.
22. Pancoska, P., Moravek, Z., and Moll, U. M. (2004). Efficient RNA interference depends on global context of the target sequence: Quantitative analysis of silencing efficiency using Eulerian graph representation of siRNA. *Nucleic Acids Res*. **32,** 1469–1479.

23. Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W. S., and Khvorova, A. (2004). Rational siRNA design for RNA interference. *Nat. Biotechnol*. **22,** 326–330.
24. Schubert, S., Grunweller, A., Erdmann, V. A., and Kurreck, J. (2005). Local RNA target structure influences siRNA efficacy: Systematic analysis of intentionally designed binding regions. *J. Mol. Biol*. **348,** 883–893.
25. Reynolds, A., Anderson, E. M., Vermeulen, A., et al. (2006). Induction of the interferon response by siRNA is cell type—and duplex length—dependent. *RNA* **12,** 1–6.
26. Judge, A. D., Bola, G., Lee, A. C., and MacLachlan, I. (2006). Design of non-inflammatory synthetic siRNA mediating potent gene silencing *in vivo*. *Mol. Ther*. **13,** 494–505.
27. Marques, J. T., Devosse, T., Wang, D., et al. (2006). A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. *Nat. Biotechnol*. **24,** 559–565.
28. Bammler, T., Beyer, R. P., Bhattacharya, S., et al. (2005). Standardizing global gene expression analysis between laboratories and across platforms. *Nat. Meth.* **2,** 351–356.
29. Irizarry, R. A., Warren, D., Spencer, F., et al. (2005). Multiple-laboratory comparison of microarray platforms. *Nat. Meth*. **2,** 345–350.
30. Larkin, J. E., Frank, B. C., Gavras, H., Sultana, R., and Quackenbush, J. (2005). Independence and reproducibility across microarray platforms. *Nat. Meth*. **2,** 337–344.
31. MAQC Consortium: Shi, L., Reid, L. H., Jones, W. D., et al. (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol*. **24,** 1151–1161.
32. Patterson, T. A., Lobenhofer, E. K., Fulmer-Smentek, S. B., et al. (2006). Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat. Biotechnol*. **24**, 1140–1150.
33. Dobbin, K., and Simon, R. (2002). Comparison of microarray designs for class comparison and class discovery. *Bioinformatics* **18,** 1438–1445.
34. Dobbin, K., Shih, J. H., and Simon, R. (2003). Questions and answers on design of dual-label microarrays for identifying differentially expressed genes. *J. Natl. Cancer Inst*. **95,** 1362–1369.