Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs

Aly A Khan[1,2], Doron Betel[2], Martin L Miller[2,3], Chris Sander[2], Christina S Leslie[2,5] & Debora S Marks[4,5]

Transfection of small RNAs (such as small interfering RNAs (siRNAs) and microRNAs (miRNAs)) into cells typically lowers expression of many genes. Unexpectedly, increased expression of genes also occurs. We investigated whether this upregulation results from a saturation effect—that is, competition among the transfected small RNAs and the endogenous pool of miRNAs for the intracellular machinery that processes small RNAs. To test this hypothesis, we analyzed genome-wide transcript responses from 151 published transfection experiments in seven different human cell types. We show that targets of endogenous miRNAs are expressed at significantly higher levels after transfection, consistent with impaired effectiveness of endogenous miRNA repression. This effect exhibited concentration and temporal dependence. Notably, the profile of endogenous miRNAs can be largely inferred by correlating miRNA sites with gene expression changes after transfections. The competition and saturation effects have practical implications for miRNA target prediction, the design of siRNA and short hairpin RNA (shRNA) genomic screens and siRNA therapeutics.

Thousands of miRNAs (21- to 23-nucleotide single-stranded RNAs) have been identified in animals over the past 7 years[1,2]. Research on miRNAs has focused on their biochemical processing and mechanism of action[3], the scope of their regulatory programs and their differential expression profiles in development and disease[4]. Furthermore, various siRNA or miRNA (si/miRNA) constructs are widely used in functional genomics, and miRNA cellular/tissue profiles are measured in medical diagnostics[5]. Finally, si/miRNAs (and their inhibitors) are in clinical trials for use as medical therapeutics[6,7].

Contrary to expectations, however, some genes are strongly upregulated in si/miRNA transfections (**Supplementary Fig. 1** online). And although there have been encouraging successes in using of si/miRNAs in functional genomics and therapeutics, various unexpected effects have been reported, including a nonspecific immune response[8] and saturation of components of the shRNA or miRNA nuclear export machinery[9–11], such as exportin-5. It has been suggested that saturation-related effects can be avoided by using siRNAs[11] (because they do not rely on the nuclear export machinery) rather than shRNAs. Indeed, a recent prominent report specifically claimed that effective siRNAs targeting *APOB* and *F7* do not interfere with endogenous miRNA function[12]. However, use of siRNAs have not been problem-free, as scrambled siRNAs have been shown to cause dose-dependent upregulation of a target gene, *SREBF1*, in three different cell types[13], and an elegant report on combinatorial delivery of siRNAs in HEK293 cell lines demonstrated competition for RISC (RNA-induced silencing complex) machinery[14].

Here, we investigate the hypothesis that the unexplained si/miRNA-induced gene upregulation is due, at least partly, to a loss of function of endogenous miRNAs, as modeled in **Figure 1** and supported by previous reports[14]. In this model, transfected small RNAs compete with endogenous miRNAs for the RISC complex or other machinery further downstream than exportin-5 in the miRNA pathway, such as argonaute proteins or *TRBP* (also known as *TARBP2P*)[14–17]. Loss of available RISC through competition would be expected to relieve repression of the targets of endogenous miRNAs—an effect that may be observed as upregulation of target mRNAs and corresponding proteins. We reasoned that we should be able to detect this effect in gene expression profiles measured after si/miRNA perturbations. We also reasoned that this effect may be observable in the dose response and temporal dynamics of the misregulated off-target genes[18,19].

RESULTS

Endogenous miRNAs are upregulated after transfection

To test these hypotheses, we assembled data from published experiments in which small RNAs were transfected into cells in culture, which were then assayed using mRNA profiling or protein mass spectrometry (Online Methods and **Supplementary Fig. 2** online). In total, we gathered data from 151 experiments from seven different cell types, involving 29 different miRNAs (as well as 2 mutant and 2 chimeric miRNAs) and 42 unique siRNAs (**Table 1** and **Supplementary Tables 1** and **2** online). Notably, a large number of genes are upregulated in the si/miRNA experiments, rather than downregulated as would be expected (**Supplementary Fig. 1**).

To investigate whether predicted targets of cellular (endogenous) miRNAs respond to transfected si/miRNAs, we assessed global expression changes after si/miRNA transfection or miRNA inhibition (Online Methods). Briefly, we used available miRNA expression profiles[20–22] to define the ten most highly expressed endogenous

[1]Department of Computer Science, Columbia University, New York, New York, USA. [2]Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. [3]The Novo Nordisk Foundation Centre for Protein Research, University of Copenhagen, Copenhagen, Denmark. [4]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to C.S.L. (cleslie@cbio.mskcc.org) or D.S.M. (deboramarks@gmail.com).

Published online 24 May 2009; corrected after print 8 July 2009; doi:10.1038/nbt.1543

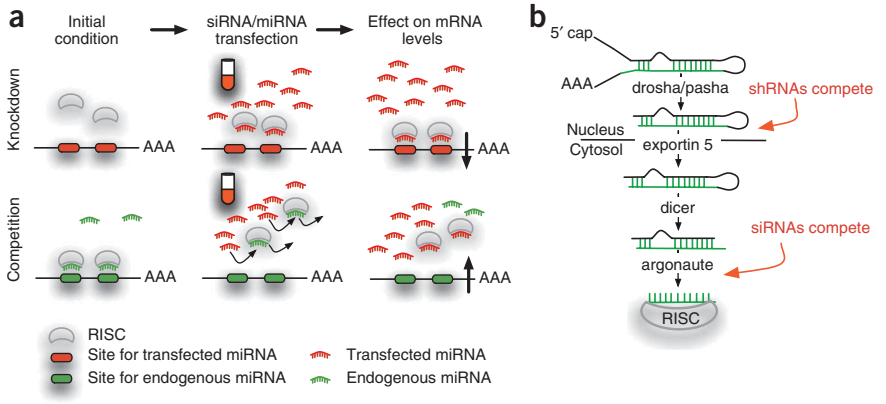

Figure 1 Schematic of the hypothesis that transfected si/miRNAs and the cell's endogenous miRNAs compete for RISC machinery. (**a**) Genes with sites (red) for the transfected small RNA (red) are downregulated after transfection. Genes with sites for endogenous miRNAs (green) may be upregulated after small RNA transfections. (**b**) Biogenesis of miRNAs. shRNAs and siRNAs enter the miRNA processing pathway at different points.

We also investigated whether the effect of miRNA transfection extends to protein levels by analyzing HeLa cell data from mass spectrometry experiments after miRNA transfection[24]. Significant changes in protein levels were observed (**Supplementary Table 2**). In particular, protein levels of target genes with sites for endogenous miRNAs and no sites for exogenous miRNAs (genes D – X) were upregulated when compared to baseline genes ($P < 1.3e-9$, pooled data). For example, when HeLa cells were transfected with the endogenously expressed let-7b, protein levels of genes with other endogenous target sites increased significantly relative to the baseline ($P < 8e-6$; **Fig. 2c**). Taken together, we find that global perturbations that follow miRNA transfection are consistent with a 'competition effect' in which the transfected miRNA competes with endogenous miRNAs for cellular protein machinery.

miRNAs in each cell type, which together make up 70–80% of the measured cellular miRNA content (**Supplementary Fig. 3** online). We then identified genes with predicted sites targeted by these ten 'endogenous' miRNAs (set D), genes with predicted sites for the 'exogenous' transfected si/miRNA (set X), and a 'baseline' set of genes with neither endogenous nor exogenous sites (set B). All miRNA predictions were based on conserved seed matches (see Online Methods). Differences in global expression changes between gene sets following perturbation were assessed for statistical significance by a one-sided Kolmogorov-Smirnov (KS) test (Online Methods).

When miR-124 is transfected into HeLa cells[23], genes with sites for HeLa-expressed (endogenous) miRNAs and without miR-124 sites (that is, genes in set D but not in set X, or D – X) are significantly upregulated compared to the baseline set ($P < 7.5e-34$; **Fig. 2a** and **Supplementary Table 2**). The magnitude of upregulation was even greater when we limited our analysis to genes with at least 2 endogenous sites and no sites for the transfected miRNA ($P < 2.2e-24$; **Fig. 2a**, blue line). Overall, we observed effects of this kind in 89% of the miRNA transfection experiments tested (using significance threshold $P < 0.05$, $n = 61$) (**Supplementary Table 2**).

To further investigate whether this is a general effect, we pooled all of the HeLa transfection experiments and repeated the analysis. We found that the set D – X was significantly upregulated in the pooled HeLa data ($P < 10^{-100}$; **Fig. 2b**). The same was also true for pooled data in A549, HCT116, HCT116 Dicer$^{-/-}$ and Tov21G cells, with $P < 10^{-10}$ for all cell types (**Supplementary Table 2**). Interestingly, in experiments in which an endogenous miRNA was overexpressed, the targets of other endogenously expressed miRNAs were also upregulated. For example, when HeLa cells were transfected with the endogenously expressed miRNAs miR-16 or let-7b, the set D – X was upregulated compared to the baseline set ($P < 5.6e-19$, $P < 6.1e-12$, respectively; **Supplementary Table 2** and **Supplementary Fig. 4** online).

As a positive control, to verify that transfected small RNAs affected their predicted targets, we compared changes in expression of the targets of the exogenous RNAs to the baseline gene set, both in individual transfection experiments and in sets of experiments grouped by cell type (**Table 1**). As expected, we found that expression of the target mRNAs was significantly downshifted compared to the baseline set ($P < 10^{-4}$ in all miRNA transfections; **Supplementary Table 2**).

siRNA transfections cause the same effect

To investigate whether siRNAs perturb the transcriptome similarly to miRNAs, we analyzed gene expression profiles of 42 siRNA transfections in HeLa cells[19,25,26]. We found that targets of endogenous miRNAs were significantly upregulated after transfection of an siRNA that targets *MAPK14* ($P < 7.4e-31$; **Fig. 2d**). Upregulation of targets of let-7 and

Table 1 Small RNA transfection data sets used in the analysis

Transfected miRNA	Cell type	References
miR-124, miR-1, miR-373, miR-124mut5-6, miR-124mut9-10, chimiR-1-124, chimiR-124-1	HeLa	23
miR-106b, miR-200a/b, miR-141miR-16, miR-15a/b, miR-103, miR-107, miR-192, miR-215, miR-17-5p, miR-20, let-7c, miR-195	HeLa, HCT116 HCT116 Dicer$^{-/-}$	30
miR-7, miR-9, miR-122a, miR-128a, miR-132, miR-133a, miR-142, miR-148, miR-181a	HeLa	36
miR-34a/b/c	HeLa,A549, TOV21G,HCT116 Dicer$^{-/-}$	21
miR-1, miR-155, let-7b, miR-30	HeLa	24
miR-181a, miR-124, miR-1	HeLa	35
miR-34a	HeLa	46
miR-124	HepG2	47
Transfected siRNA	**Cell type**	**Reference**
MAPK14-1pos4 mismatch, MAPK14-1pos5 mismatch, MAPK14-1pos15 mismatch, MAPK14-1*	HeLa	19
MPHOSPH1-2692, PIK3CA-2692, PRKCE-1295, VHL-2651, VHL-2652, SOS1-1582	HeLa	26
PIK3CB-6338, PLK-1319, PLK-772, PIK3CB-6340	HeLa	25
siAPOB-Hs1/Hs2/Hs3/Hs4	Huh7	18

This siRNA transfected at (1, 2, 4, 6, 12, 24, 48, 72, 96) hours and (0.16, 0.8, 4, 20, 100) nM.

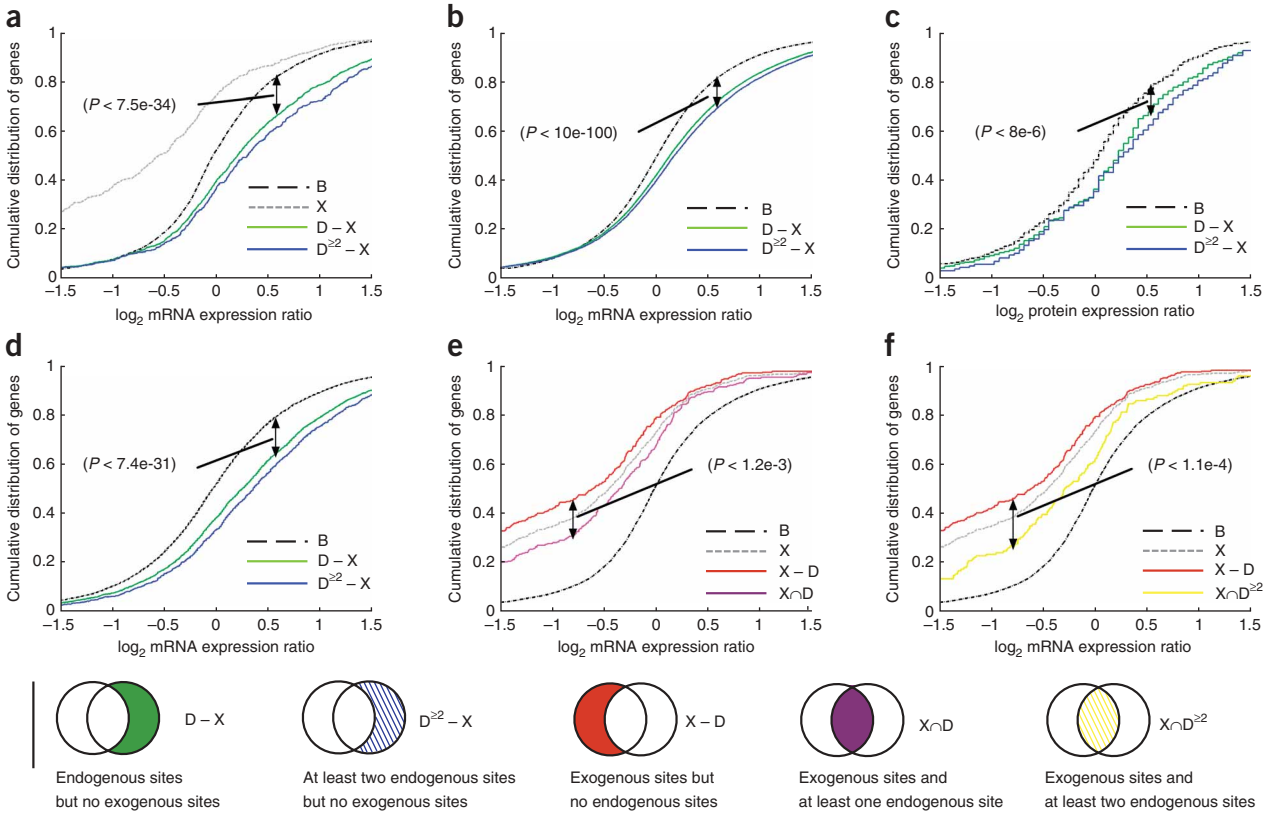

Figure 2 Genes with predicted target sites for endogenous miRNAs are significantly dysregulated after si/miRNA transfections. *P*-values shown are calculated by one-sided KS test as described in Online Methods. See bottom of figure for notation used to abbreviate gene sets (D – X, $D^{\geq 2}$ – X, X – D, X∩D and X∩$D^{\geq 2}$) tested for significant expression changes relative to a baseline genes (set B). (**a**) miR-124 transfection results in up-shift in gene expression for D – X. (**b**) Pooled data from 15 miRNA transfections into HeLa cells (miR-373, miR-124, miR-148b, miR-106b, miR-124, mut9-10, miR-1, miR-181, chimiR-124-1, chimiR-1-124, miR-16, mir-34a, mir-34b, miR-128a and miR-9). (**c**) Protein expression changes after let-7b transfection. (**d**) mRNA expression changes after MAPK14-siRNA is transfected into HeLa cells showing upregulation of genes with sites for endogenous miRNAs (green line)[19]. (**e**,**f**) Genes that contain sites for both endogenous miRNA and transfected small RNAs are less downregulated than if they contain only sites for transfected small RNAs. D, genes with target sites for endogenous miRNAs. X, genes with target sites for exogenously transfected small RNAs.

miR-15, two miRNAs highly expressed in HeLa cells, was especially notable (**Supplementary Table 3** online). Pooling the data from these siRNA experiments, we see a significant upward shift in the expression of genes with endogenous sites only relative to the baseline gene set ($P < 1 \times 10^{-100}$; **Supplementary Table 2**). Five different siRNAs designed to target *VHL, PRKCE, MPHOSPH1, SOS1* and *PIK3CA*[26], respectively. These siRNAs showed upregulation of similar sets of genes including *CCND1, DUSP4, DUSP5* and *ATF3* (**Supplementary Table 3**), despite having different direct targets. As each of these upregulated genes contains at least one site for an endogenous miRNA, this observation is consistent with upregulation as a consequence of the siRNA transfection, independent of the specific siRNA sequence.

Attenuated knockdown of targets with endogenous sites

To investigate whether the 'competition effect' might attenuate the strength of si/miRNA knockdown, we analyzed the expression of genes directly targeted by the transfected si/miRNA. Specifically, we partitioned the set of genes with sites for transfected miRNAs (set X) into two subsets—those without endogenous sites (labeled X – D) and those with endogenous sites (labeled X∩D). As a representative example of our results, after transfection of miR-16 into HeLa cells, predicted miR-16 target genes without endogenous sites (that is, X – D) were downregulated significantly more than targets with endogenous sites (that is, X∩D; $P <$ 1.2e-3, **Fig. 2e**). Limiting the analysis to genes with two or more sites for endogenous miRNAs resulted in an even more pronounced difference ($P <$ 1.1e-4, **Fig. 2f**). Pooling data across a panel of transfection experiments into HeLa cells also gave a significant result ($P <$ 3.6e-13, **Supplementary Table 2**). Taken together, these results suggest that competition with cellular machinery may attenuate the effectiveness of si/miRNA knockdown.

A quantitative model resolves the endogenous miRNA profile

To strengthen our analysis and predict the saturation effect on individual genes, we built a quantitative mathematical model of the changes in gene expression after si/miRNA transfection. This model can be used to predict which genes are likely to be upregulated or downregulated (off-target effects) after si/miRNA transfections. Considering each transfection into HeLa cells independently, we first fit a simple linear regression model (Online Methods) to predict the change in expression of genes based on the number of exogenous sites (n_X) and the number of endogenous sites (n_D) in their 3′ untranslated region (UTR) (**Fig. 3a**). In a large majority of experiments, the endogenous count n_D was found to be a significant variable for explaining expression changes (84 out of 109 experiments satisfying $P < 0.05$ by F statistic, **Supplementary Table 2**). As expected, the regression coefficient for the endogenous count was

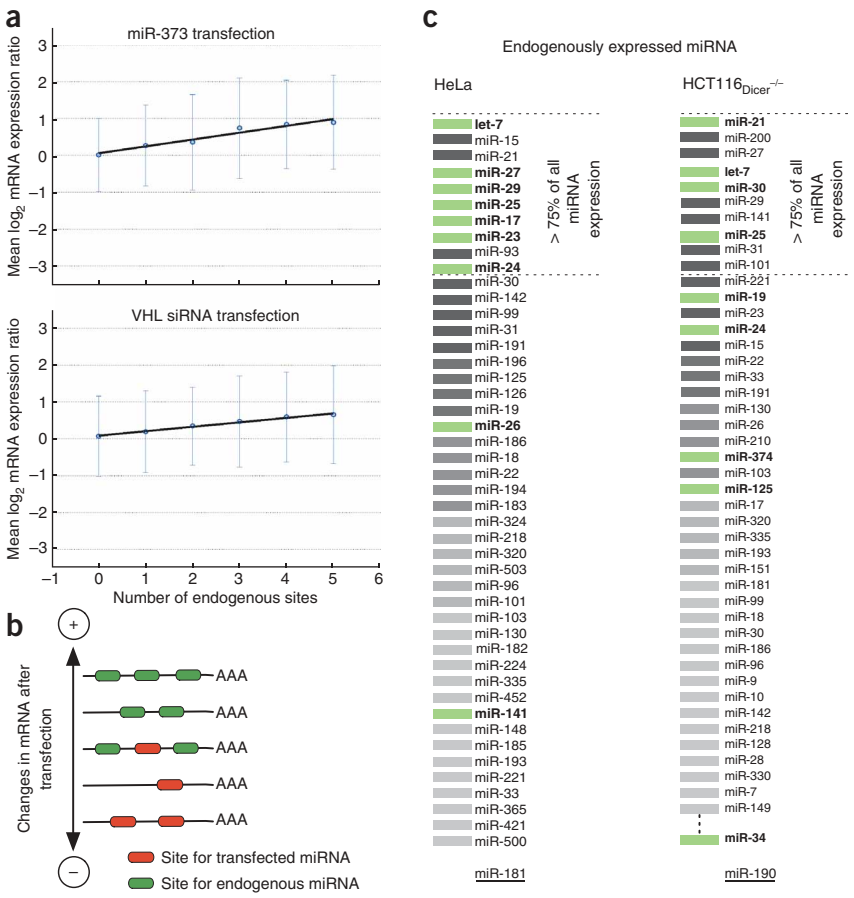

Figure 3 Quantitative model predicting expression change after transfection. (**a**) Linear regression fitting mean $\log_2$(expression change) to number of endogenous sites in genes having no exogenous sites, in miR-373 and VHL-siRNA experiments. One s.d. from mean depicted with error bars. (**b**) Cartoon of genes with different combinations of sites showing net effect on gene expression. (**c**) Endogenous miRNAs expressed in HeLa cells and HCT116 Dicer$^{-/-}$ (ranked in order of endogenous expression, with black to gray depicting decreasing relative expression). Green rectangles indicate the miRNAs that occurred most often with positive regression coefficient in stepwise regression models for >20 HeLa cell and 16 HCT116 Dicer$^{-/-}$ transfection experiments (Online Methods). These miRNAs can be interpreted as a predicted endogenous profile. Underlined miRNAs were predicted by regression analysis but are not known to be expressed endogenously.

always positive when significant, meaning that these sites correlate with upregulation, whereas the regression coefficient for the exogenous count was always negative. **Figure 3b** is a cartoon version of the expected effect on expression of a gene that contains different combinations of exogenous and endogenous sites.

We then refined the model to evaluate whether the presence of sites recognized by individual endogenous miRNAs could explain upregulation of targets in an experiment. More generally, we considered all human miRNA families as potential variables in the regression model and assessed whether sites of individual miRNAs accounted for expression changes in a transfection experiment. We ranked the importance of each miRNA by the number of experiments in which it was included in a forward stepwise regression model (Online Methods). Among the ten most frequently included miRNAs, we identified seven of the most highly expressed miRNAs in HeLa cells and four of the most highly expressed in HCT116 Dicer$^{-/-}$ cells, using no prior knowledge of the miRNAs expressed in those cell types (**Fig. 3c**). The top ranked miRNAs retrieved by this analysis, let-7 and miR-21, are thought to be the most highly expressed miRNA in HeLa and HCT116 Dicer$^{-/-}$ cells, respectively, therefore supporting a saturation model. Taken together, these results suggest that the endogenous miRNA profile in a cell can be largely determined from expression changes after transfection of small RNAs, which plausibly are due to competition for cellular resources.

The competition effect has a dose response

In a previous study[19], siRNA dose response was investigated by transfecting an siRNA targeting *MAPK14* into HeLa cells in a range of five doses, from 0.16 nM – 100 nM, followed by microarray profiling after 24 h. As described in the original publication, we observed that off-target genes (that is, genes other than *MAPK14* with sites in their 3′ UTR for the siRNA) were affected in a dose-responsive manner that mimicked the dose response of the main target (*MAPK14*) and that these off-target effects were not titrated away at lower transfection concentrations. However, we also found that many genes with sites for endogenous miRNAs follow a pattern of upregulation that mirrors the downregulation of off-targets (**Fig. 4a**). More specifically, a fivefold change in siRNA dose from 4 nM to 20 nM produced a twofold change in mean gene expression of the most responsive upregulated genes and the most responsive downregulated genes. The change in expression of both the endogenous target and off-target gene sets reaches near-maximal dose response at 20 nM. In summary, these siRNA saturation effects and off-target effects roughly scale with the dose response of the main target, at least for a significant fraction of genes in these sets, and we did not observe that they were titrated away at lower transfection concentrations.

Evidence for a transitory saturation effect

To measure the time dependence of the response of genes with sites for endogenous miRNAs, we examined published data[19] in which gene expression changes were monitored over a period of 96 h after transfection of an siRNA targeting *MAPK14*. Given our previous observations that genes putatively regulated by endogenous miRNAs were de-repressed after transfection, we expected that the temporal response of these genes would be similar to that of the intended siRNA target gene and the off-targets (that is, genes with nonconserved (NC) seed matches, X_{NC}; see Online Methods). We compared the mRNA changes of the putative off-target genes to *MAPK14* mRNA itself. Although the off-target genes followed a temporal downregulation pattern similar to *MAPK14* in the first 48 h, the expression level of the X_{NC} set of genes returned to near their original expression level by 92 h. In contrast, the intended target *MAPK14* had a gradually increasing downregulatory effect, with a half maximal effect at ~12 h and a sustained effect from 24–96 h (**Fig. 4b**).

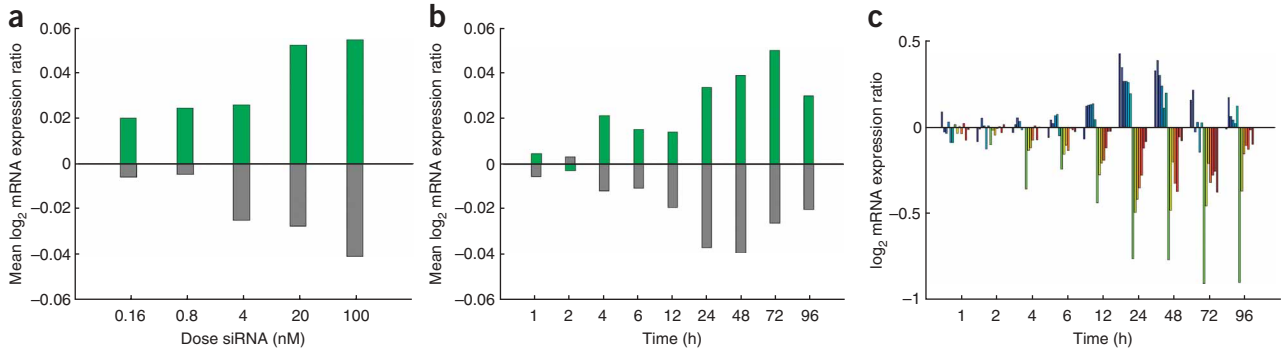

Figure 4 Competition effect shows dose-response and temporal dynamics proportional in magnitude, but opposite in direction, to targeted effect. (**a**) Effect of dose of siRNA transfection on mean $\log_2$(expression change) of X_{NC} (gray) and 90th percentile of $D^{\geq 2} - X_{NC}$ (green). (**b**) Mean $\log_2$(expression change) of X_{NC} (gray) and 90th percentile of $D^{\geq 2} - X_{NC}$ (green) versus time. (**c**) $\log_2$(expression change) of putative endogenously regulated genes (*SCML2, TNRC6, YOD1, CX3CL1, AKAP12* and *PGM2L1*), shown above zero, and *MAPK14* (light green) with a set of MAPK14-siRNA 'off-targets' (*MARK2*, *SLC35F3*, *HMGB3*, *FZD7*, *RPA2* and *IER5L*), shown below zero, over same time course as **b** (genes displayed in this order, from left to right).

We investigated the dynamics of a set of genes with at least two nonconserved endogenous sites (90th percentile for expression change, pooling all time points, ~1,000 genes), compared to a set of siRNA targeted genes (Online Methods). The genes in the endogenous set had maximal upregulation at 24–48 h with similar dynamics across the 92 h, consistent with being targets of endogenous miRNAs competing for components of the RISC (**Fig. 4b**). This similarity was also apparent in the expression patterns of the six most downregulated off-target genes and the six most upregulated genes in set $D - X_{NC}$ (**Fig. 4c**). The upregulated genes (*SCML2, TNRC6, YOD1, CX3CL1, AKAP12* and *PGM2L1*) each contain at least four sites for highly expressed endogenous miRNAs. The expression patterns of these genes are consistent with the model that they are targets of endogenous miRNAs competing for components of the RISC. As *TNRC6* is associated with *AGO2* (also known as *EIF2C2*) in RISC, this inadvertent TNRC6 upregulation may in turn affect the function of all microRNAs expressed in the cell[27].

We also investigated published experiments[18] that were designed to examine the off-target effects of four therapeutic siRNAs targeting the coronary artery disease target gene, *APOB*. We observed a significant saturation effect with all of the siRNAs ($P < 1e\text{-}8$ at 6 h, **Supplementary Table 2**), and we noticed that this effect reached its maximum at 6 h. Taken together, these investigations of the temporal dynamics of small RNA gene regulation after transfection show that the upregulatory effect mirrors the expected downregulatory effect, supporting the proposed competition model.

Cell cycle genes are upregulated after si/miRNA transfections

Dysregulation of endogenous miRNAs is known to contribute to tumorigenesis[28], and the experiments we analyzed were conducted in immortalized cell lines (e.g., HeLa cells). We were therefore not surprised to find a significant number of cell cycle, oncogene, and tumor suppressor genes (**Supplementary Fig. 5a** online) consistently upregulated across transfection experiments (**Supplementary Table 1**). For instance, known miRNA targets, including the oncogenes *HMGA2* (ref. 29), *CCND1* (refs. 30,31) and *DUSP2*, are upregulated after many different independent HeLa cell transfection experiments, including siRNA transfections. We also find that endogenous miRNA targets in HeLa cells are significantly enriched for cell cycle genes and oncogenes ($P < 2e\text{-}3$, $P < 4.5e\text{-}14$, respectively; see Online Methods and **Supplementary Fig. 5b** online). Taken together, these observations raise the possibility that cell cycle and oncogenes may be particularly susceptible to the proposed saturation effect.

miRNA inhibition may upregulate other endogenous targets

Finally, we examined published data[32] that measured expression changes after miRNA inhibition by 'antagomirs', which are chemically modified single-stranded RNA analogs that inhibit a specific target miRNA. Treatment of cells with antagomirs to miR-16 and miR-106b significantly upregulated genes that contained only endogenous sites ($P < 5e\text{-}16$ (D – X) and $P < 2e\text{-}30$ ($D^{\geq 2} - X$)), including *SSR3*, *PLSCR4* and *PTRF* (**Supplementary Table 3** online). Moreover, inhibition of miR-122 with locked nucleic acid (LNA) molecules[33] also produced significant upregulation of genes with sites for other endogenous miRNAs when compared with a saline transfection ($P < 2.5e\text{-}6$). Dose-dependent accumulation of a shifted heteroduplex band, implying that the LNA-antimiR binds stably to the miRNA, has been observed[33]. This finding is consistent with the hypothesis that the heteroduplex of miR-122::antimiR prevents the availability of free RISC machinery (**Supplementary Fig. 6** online), but clearly more experiments are needed to distinguish between the possible models and to assess the impact of the inhibition effect on the function of endogenous miRNAs.

DISCUSSION

We have shown that transfecting small RNAs affects the expression of genes predicted to be under endogenous miRNA regulation. This effect is observable at both the mRNA and protein levels. Moreover, this effect is observable in experiments that use siRNAs that target particular genes, and in experiments that use miRNA mimics and miRNA inhibitors designed to test the biological effects of miRNAs. Using a quantitative approach, we built a regression model that can identify many of the endogenous miRNAs expressed in a cell type simply from the changes in gene expression after small RNA transfections. The purpose of this approach is not to infer miRNA profiles *per se* but to provide independent evidence of the indirect perturbation of miRNA function. Finally, we used published data to show that the temporal dynamics and dose response of genes affected by the proposed competition effect follow the same patterns as those of the genes directly targeted by the transfection.

The most plausible model for these observations is saturation of the RISC complex (or other necessary small RNA processing or transport

machinery) and competition between the transfected small RNA and endogenous miRNAs for binding (**Fig. 1**). However, other models that may be consistent with the observed effect cannot be ruled out by our analyses. Although the precise mechanism of this competition effect remains to be established, the statistical significance of the observed shifts in transcript levels is clear, and the results of these analyses support the thesis that small RNA transfections unexpectedly and unintentionally (from the point of view of the investigators) disturb gene regulation by endogenous miRNA.

Our results have potentially important practical consequences for the use of siRNAs and shRNAs in functional genomics experiments. Although it is already known that siRNAs can produce unwanted off-target effects, such as unintended downregulation of mRNAs through a partial sequence match between the siRNA and target, the effects observed here are distinct and involve the de-repression of miRNA-regulated genes.

Our findings also have consequences for the development of miRNA target prediction methods. As measuring mRNA expression changes after si/miRNA perturbations is a standard way to validate miRNA target prediction methods[23,25,34], one should take the saturation effect into consideration. Despite concerted efforts, bioinformatic si/miRNA target prediction methods still greatly overpredict the number of targets by at least sevenfold[24,35–37]. Elegant work showing the dynamic (condition and cell-type dependent) regulation of UTR lengths[38] may explain some of these false positives, as shortening of UTRs may lead to loss of target sites, but is unlikely to explain all. The proposed competition effect may offer an explanation for false-positive target prediction in cases where UTRs have target sites for both the transfected and endogenous miRNAs (**Fig. 3b**). Moreover, as miRNAs may compete with each other for target sites in mRNAs, it may be important to consider RISC saturation in target prediction methodology.

Further, our results have consequences for the development of small RNA therapeutics, considered to hold substantial promise[39]. miRNA inhibitors, such as anti-miR-122, have been used to target cholesterol synthesis[40], hepatitis C virus[40,41] and herpes simplex virus[42]. Therapeutic siRNAs have also been designed for potential treatment of cancer, including in melanoma, against vascular endothelial growth factor VEGF-A/-C[43], and through anti-miR-21 in glioma[39,44,45]. Our work illustrates the potentially broad consequences of the perturbation of the cellular miRNA activity profile after introduction of si/miRNA inhibitors, and it suggests that these effects be considered quantitatively during development of small RNA therapies. Experiments that quantify the relative concentrations of protein machinery and small RNAs in a particular cellular context, as well as a fuller exploration of the kinetics of the various binding events involved in small RNA biogenesis and function, are clearly required. Our quantitative model implies a procedure for calibrating and potentially avoiding unwanted effects of the designed small RNA therapeutics.

Our work is subject to some limitations. In particular, this report does not attempt to resolve details of the mechanism behind the competition effect. The calculations of the effect, though carefully evaluated in statistical terms, are subject to the inaccuracies of miRNA target prediction, which entails both false positives and false negatives at the level of particular target genes. We therefore argue in terms of overall distributions, rather than attempting to quantify the involvement of individual target sites in transfection-mediated expression changes. In future work, it may be possible to identify quantitative criteria that determine the extent of the competition between exogenous and endogenous miRNAs and their effects on gene targeting. Quantitative detail will depend on knowing the concentration in the cell of the RISC complex and of other components of the small RNA machinery, the concentration of the transfected and endogenous miRNAs, the concentrations of the target mRNAs and the number of actual targets in the cell for a specific small RNA, as well as kinetic parameters such as the on and off rates of small RNAs in the RNA-protein complexes. Models that posit different concentration-dependent and kinetic scenarios could help focus the range of experiments needed to quantify these effects.

Finally, our results may have an important biological correlate, as the competition effect may have a role in normal biological or disease-related cellular processes, such as miRNA-dependent regulatory programs. For example, during both differentiation and disease processes such as cancer, miRNA profiles can change dramatically both in the identity of the dominant miRNAs and in total cellular miRNA concentration. Such changes, by means of competition for limited resources, may orchestrate observable changes in cellular regulatory programs with potential physiological consequences.

In summary, the observed statistically supported competition effect for small RNAs may point to new biological mechanisms and likely has important practical consequences for the use of small RNAs in functional genomics experiments, development of miRNA target and siRNA off-target prediction methods and development of small RNA therapeutics.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank N. Stroustrup, Y. Merbl, G. Altan-Bonnet, A. Arvey and N. Gauthier for useful discussions. This work was supported in part by National Institutes of Health (NIH) grant CA124380 and NIH grant CA121852.

AUTHOR CONTRIBUTIONS

A.A.K. performed the statistical and computational analysis and contributed to the manuscript. D.B. and M.L.M contributed to the computational analysis. C.S. contributed to discussions and the manuscript. C.S.L. designed the statistical and computational methods. D.S.M. conceived the idea for the project and contributed to the analysis. D.S.M. and C.S.L jointly supervised the research and wrote the manuscript.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Ruvkun, G. The perfect storm of tiny RNAs. *Nat. Med.* **14**, 1041–1045 (2008).
2. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
3. Filipowicz, W., Bhattacharyya, S.N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* **9**, 102–114 (2008).
4. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**, 522–531 (2004).
5. Tam, W. The emergent role of microRNAs in molecular diagnostics of cancer. *J. Mol. Diagn.* **10**, 411–414 (2008).
6. Rossi, J., Zamore, P. & Kay, M.A. Wandering eye for RNAi. *Nat. Med.* **14**, 611 (2008).
7. Leachman, S.A. *et al.* Therapeutic siRNAs for dominant genetic skin disorders including pachyonychia congenita. *J. Dermatol. Sci.* **51**, 151–157 (2008).
8. Robbins, M. *et al.* Misinterpreting the therapeutic effects of siRNA caused by immune stimulation. *Hum. Gene Ther.* **10**, 991–999 (2008).
9. Grimm, D. *et al.* Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**, 537–541 (2006).
10. Stewart, C.K., Li, J. & Golovan, S.P. Adverse effects induced by short hairpin RNA expression in porcine fetal fibroblasts. *Biochem. Biophys. Res. Commun.* **370**, 113–117 (2008).
11. McBride, J.L. *et al.* Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. *Proc. Natl. Acad. Sci. USA* **105**, 5868–5873 (2008).

12. John, M. *et al.* Effective RNAi-mediated gene silencing without interruption of the endogenous microRNA pathway. *Nature* **449**, 745–747 (2007).
13. Vankoningsloo, S. *et al.* Gene expression silencing with 'specific' small interfering RNA goes beyond specificity - a study of key parameters to take into account in the onset of small interfering RNA off-target effects. *FEBS J.* **275**, 2738–2753 (2008).
14. Castanotto, D. *et al.* Combinatorial delivery of small interfering RNAs reduces RNAi efficacy by selective incorporation into RISC. *Nucleic Acids Res.* **35**, 5154–5164 (2007).
15. Gruber, J. *et al.* RNAi of FACE1 protease results in growth inhibition of human cells expressing lamin A: implications for Hutchinson-Gilford progeria syndrome. *J. Cell Sci.* **118**, 689–696 (2005).
16. Wargelius, A., Ellingsen, S. & Fjose, A. Double-stranded RNA induces specific developmental defects in zebrafish embryos. *Biochem. Biophys. Res. Commun.* **263**, 156–161 (1999).
17. Oates, A.C., Bruce, A.E. & Ho, R.K. Too much interference: injection of double-stranded RNA has nonspecific effects in the zebrafish embryo. *Dev. Biol.* **224**, 20–28 (2000).
18. Burchard, J. *et al.* MicroRNA-like off-target transcript regulation by siRNAs is species specific. *RNA* **15**, 308–315 (2009).
19. Jackson, A.L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635–637 (2003).
20. Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
21. He, L. *et al.* A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007).
22. Cummins, J.M. *et al.* The colorectal microRNAome. *Proc. Natl. Acad. Sci. USA* **103**, 3687–3692 (2006).
23. Lim, L.P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
24. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).
25. Jackson, A.L. *et al.* Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179–1187 (2006).
26. Jackson, A.L. *et al.* Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA* **12**, 1197–1205 (2006).
27. Landthaler, M. *et al.* Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* **14**, 2580–2596 (2008).
28. Ventura, A. & Jacks, T. MicroRNAs and cancer: short RNAs go a long way. *Cell* **136**, 586–591 (2009).
29. Mayr, C., Hemann, M.T. & Bartel, D.P. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* **315**, 1576–1579 (2007).
30. Linsley, P.S. *et al.* Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol. Cell. Biol.* **27**, 2240–2252 (2007).
31. Bonci, D. *et al.* The miR-15a-miR-16–1 cluster controls prostate cancer by targeting multiple oncogenic activities. *Nat. Med.* **14**, 1271–1277 (2008).
32. Krutzfeldt, J. *et al.* Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685–689 (2005).
33. Elmen, J. *et al.* Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver. *Nucleic Acids Res.* **36**, 1153–1162 (2008).
34. Jackson, A.L. & Linsley, P.S. Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.* **20**, 521–524 (2004).
35. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
36. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
37. John, B. *et al.* Human MicroRNA targets. *PLoS Biol.* **2**, e363 (2004).
38. Sandberg, R. *et al.* Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).
39. Stenvang, J., Lindow, M. & Kauppinen, S. Targeting of microRNAs for therapeutics. *Biochem. Soc. Trans.* **36**, 1197–1200 (2008).
40. Fabani, M.M. & Gait, M.J. miR-122 targeting with LNA/2'-O-methyl oligonucleotide mixmers, peptide nucleic acids (PNA), and PNA-peptide conjugates. *RNA* **14**, 336–346 (2008).
41. Lupberger, J., Brino, L. & Baumert, T.F. RNAi: a powerful tool to unravel hepatitis C virus-host interactions within the infectious life cycle. *J. Hepatol.* **48**, 523–525 (2008).
42. Palliser, D. *et al.* An siRNA-based microbicide protects mice from lethal herpes simplex virus 2 infection. *Nature* **439**, 89–94 (2006).
43. Shibata, M.A. *et al.* Combination therapy with short interfering RNA vectors against VEGF-C and VEGF-A suppresses lymph node and lung metastasis in a mouse immunocompetent mammary cancer model. *Cancer Gene Ther.* **15**, 776–786 (2008).
44. Elmen, J. *et al.* LNA-mediated microRNA silencing in non-human primates. *Nature* **452**, 896–899 (2008).
45. Corsten, M.F. *et al.* MicroRNA-21 knockdown disrupts glioma growth in vivo and displays synergistic cytotoxicity with neural precursor cell delivered S-TRAIL in human gliomas. *Cancer Res.* **67**, 8994–9000 (2007).
46. Chang, T.C. *et al.* Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol. Cell* **26**, 745–752 (2007).
47. Wang, X. *et al.* Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res.* **34**, 1646–1652 (2006).

ONLINE METHODS

mRNA and protein experimental data sets. We collected data from four types of experiments: (i) transfection of a miRNA followed by mRNA profiling using microarrays[4,23,24,30,36,46]; (ii) transfection of an siRNA followed by mRNA profiling[18,19,26]; (iii) inhibition of miRNA followed by mRNA profiling[33]; and (iv) transfection of miRNA followed by protein profiling using mass spectrometry[24]. These four types of data sets of 150 experiments encompass seven different cell types, 20 different miRNAs and 40 different siRNAs (**Supplementary Table 2**). The synthetic transfected miRNAs are all commercially available siRNA or miRNA mimics (Dharmacon). Sequences of mimics can be found in the respective references. When possible, we used normalized microarray expression data as provided with the original publications. In all other cases, we used the 'affy' package in the 'R' software package to perform robust multi-array analysis (RMA) normalization of microarray probe-level data. For statistical analysis over multiple mRNA microarray profiling experiments, each experiment was independently centered using the mean $\log_2$(expression change) of genes lacking conserved endogenous or exogenous sites and normalized to have unit variance in $\log_2$(expression change) across all genes. This normalization results in a modified Z-transformation of the data, where genes with no exogenous or endogenous sites have mean 0. For the transfection experiments followed by mass spectrometry, we used normalized protein expression levels as provided by the authors of the original publication[24], **Supplementary Figure 2**.

Target prediction. We conducted four different types of miRNA target site searches using miRNA sequences grouped into families, and 3′ UTR alignment of five species. miRNAs were grouped into families as defined by identical nucleotides in positions 2–8. We searched for target sites for miRNA families in 3′ UTRs using four different types of seed matches: (i) 6-mers (position 2–7 and 3–8); (ii) 7-mers (position 2–8); (iii) 7-mer positions 2–7 m1A (the first nucleotide an A in the mRNA); and (iv) 8-mers (position 1–8). 7-mer positions 2–8 were selected for analysis because this choice gave the most significant *P*-values for downregulation of targets with sites for the transfected si/miRNA as compared to baseline genes based on a one-sided KS statistic (set X versus set B, as described below).

For target matches, we considered both nonconserved and conserved targets in human 3′ UTRs. 3′ UTR sequences for human (hg18), mouse (mm8), rat (rn4), dog (canFam2) and chicken (galGal2) were derived from RefSeq and the UCSC genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/multiz17way/). We used multiple genome alignments across the five species as derived by multiZ. The RefSeq annotation with the longest UTR mapped to a single gene was always used. To establish a conservation filter, we required that the 7-mer target site in human be present in at least three of the other four species, that is, exact matching in a 7-nucleotide window of the alignment in at least three other species, to be flagged as conserved. Restricting to conserved sites led to more significant *P*-values for downregulation of targets with exogenous sites as compared to baseline genes (one-sided KS statistic, set X versus set B, as defined below). We chose these stringent requirements so that our prediction method would be conservative and err on the side of underprediction rather than overprediction. However, we acknowledge that there are indeed functional siRNA and miRNA target sites that have mismatches, G:U wobbles in the 5′ end and are not conserved (refs. 48,49).

Endogenous miRNA expression. We used endogenous miRNA profiles from a compendium[20] for HeLa, A549, HepG2 and TOV21G, which provide relative miRNA expression levels from cloning and sequencing small RNA libraries. We used miRNA profiles from cloning and sequencing data[50] for HCT116 and HCT116 Dicer$^{-/-}$. For consistency across cell types, we took the top ten miRNAs with highest expression levels (clone counts), which corresponds to at least 75% of the miRNA content in each cell type, to be the set of endogenous miRNAs in our statistical analysis.

KS statistics. To compare the expression changes for two gene sets, we compared their distributions of Z-transformed $\log_2$(expression change) using a one-sided KS statistic, which assesses whether the distribution of expression changes for one set is significantly shifted downwards (downregulated) compared to the distribution for the other set. We chose the KS statistic to apply a uniform treatment of data despite the heterogeneity of the transfection experiments, which involve different cell types, different numbers of target genes with sites for the transfected si/miRNA, and different apparent transfection efficiencies. The KS statistic has the advantages that (i) it is nonparametric and hence does not rely on distributional assumptions about expression changes; (ii) it does not rely on arbitrary thresholds; and (iii) it measures significant shifts between the entire distributions rather than just comparing the tails. The KS statistic computes the maximum difference in value of the empirical cumulative distribution functions (cdfs):

$$\sup_x(F_1(x) - F_2(x))$$

where

$$F_j(x) = \frac{1}{n_j}\sum_{i=1}^{n_j} I_{X_i \le x}$$

is the empirical cdf for gene set $j = 1, 2$, based on n_j (Z-transformed) $\log_2$(expression change) values. We used the Matlab function kstest2 to calculate the KS test statistic and asymptotic *P*-value. Full KS test results are provided in **Supplementary Table 2**.

Notation. We use the following notation to describe sets of genes based on the number of sites for exogenous and endogenous miRNAs in their 3′ UTRs:

Nonconserved sites. Sets with subscript $_{NC}$ denote nonconserved sites have been used; subscript $_{NC}{}^{\geq 2}$ denotes two or more nonconserved sites

Endogenous sites. Sites for endogenous miRNAs, that is, miRNAs expressed in the cell.

Exogenous sites. Sites for exogenous si/miRNAs, that is, small RNAs introduced into the cell.

X ("*eXogenous*"). Set of genes containing at least one site for the exogenous (transfected) si/miRNA.

D ("*enDogenous*"). Set of genes containing at least one site for an miRNA endogenously expressed in the cell type.

B ("*Baseline*"). Set of genes containing neither exogenous nor endogenous sites.

$D - X$. Set of genes containing at least one endogenous site and no exogenous sites.

$X \cap D$. Set of genes containing at least one exogenous site and at least one endogenous site.

X-D. Set of genes containing at least one exogenous site and no endogenous sites.

$D^{\geq 2}$. Set of genes containing two or more endogenous sites.

$X \cap D^{\geq 2}$. Set of genes containing at least one exogenous site and at least two endogenous sites.

Regression analysis to model expression. We performed multiple linear regression to fit a linear model expressing the Z-transformed $\log_2$(expression change), denoted as *y*, in terms of the number of a gene's exogenous and endogenous target sites, denoted as n_X and n_D, respectively:

$$y = c_X n_X + c_D n_D + b$$

We use the Matlab regress function to fit the model and assess the significance of the fit as measured by the R^2 statistic. We used the *F* statistic, also computed by the regress function, to assess whether the linear model with two independent variables, n_X and n_D, significantly improves the fit over the simpler model: $y = c_X n_X + b$, given the number of sites for exogenous si/miRNAs a priori. All *P*-values from the *F* statistic across experiments are reported in **Supplementary Table 2**.

Forward stepwise regression analysis. As an extension to the linear model with two independent variables, we performed forward stepwise regression to fit the number of target sites for each of the (162) miRNA families to the Z-transformed $\log_2$(expression change) data. Starting again with the simpler model, $y = c_X n_X + b$, we incrementally added the number of target sites for the miRNA seed family with highest *F* statistic to the model. The procedure was continued until the *P*-value from the *F* statistic for the best remaining seed family failed to satisfy a significance threshold of $P < 0.05$. The final model can

doi:10.1038/nbt.1543

be viewed as a linear combination of the number exogenous target sites and the additive contribution of other miRNAs represented by their number of target sites n_i:

$$y = c_X n_X + \sum_i c_i n_i + b$$

Because we did not enforce a stringent significance criterion for including miRNA sites in the model, we do not expect every miRNA added to the model to be correct; however, miRNAs added consistently across different transfection experiments are likely to be significant. We repeated the forward stepwise regression for multiple experiments in HeLa and HCT116 Dicer$^{-/-}$ cells and computed the frequency of the most statistically significant additive factors with positive regression coefficient in the model for each cell type; we reported the ten most frequent of these miRNAs. All *P*-values from the *F* statistic across experiments are reported in **Supplementary Table 4** online.

Cell cycle and cancer genes. A list of expertly annotated genes for which mutations (both germline and somatic) have been causally implicated in cancer was obtained from the Cancer Genome Project (Cancer Gene Census catalog version 2008.12.16, http://www.sanger.ac.uk/genetics/CGP/Census)[51]. A list of genes that have consistently showed a periodic expression pattern during the cell cycle in several mRNA microarray studies was obtained from the Cyclebase database[52]. From these lists, we could match 312 and 651 genes to the mRNA data sets collected in this work, respectively. The gene sets were designated "oncogenes" and "cell cycle genes," respectively. To investigate if oncogenes or cell cycle genes were enriched for miRNA targets in HeLa cells compared to all genes we used Fisher's exact tests.

48. Vella, M.C., Reinert, K. & Slack, F.J. Architecture of a validated microRNA:target interaction. *Chem. Biol.* **11**, 1619–1623 (2004).
49. Didiano, D. & Hobert, O. Molecular architecture of a miRNA-regulated 3′ UTR. *RNA* **14**, 1297–1317 (2008).
50. Cummins, J.M. & Velculescu, V.E. Implications of micro-RNA profiling for cancer diagnosis. *Oncogene* **25**, 6220–6227 (2006).
51. Futreal, P.A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
52. Gauthier, N.P. *et al.* Cyclebase.org–a comprehensive multi-organism online database of cell-cycle experiments. *Nucleic Acids Res.* **36** Database issue, D854–D859 (2008).

Corrigendum: Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs

Aly A Khan, Doron Betel, Martin L Miller, Chris Sander, Christina S Leslie & Debora S Marks
Nat. Biotechnol. **6, 549–555 (2009); published online 24 May 2009; corrected after print 8 July 2009**

In the version of this article initially published, Figure 2f is not referenced in the figure legend and is referenced as Figure 2e in the main text. Also, on p.5, right col., para. 1, line 8, miR-21 should be miR-122. The errors have been corrected in the HTML and PDF versions of the article.

Erratum: Venture capital shifts strategies, startups suffer

Peter Mitchell
Nat. Biotechnol. **27, 103–104 (2009); published online 9 February 2009; corrected after print 8 July 2009**

In the version of this article initially published, "GSK Ventures" should have read "GlaxoSmithKline's SR One." The error has been corrected in the HTML and PDF versions of the article.

Erratum: New relief for gout

Jill U Adams
Nat. Biotechnol. **27, 309–311 (2009); published online 7 April 2009; corrected after print 8 July 2009**

In the version of this article initially published, the incidence of gout was incorrectly stated to be in the hundreds of millions worldwide and 300 million in the US (p. 309, para. 2). The incidence is known for industrialized countries, not worldwide. In the US, the number is 3 million. The last five lines of the paragraph should have read, "including about 1 in 100 adult men in industrialized countries (an estimated 3 million in the US according to the Centers for Disease Control)." The errors have been corrected in the HTML and PDF versions of the article.

Erratum: Biotech hirings and firings

Michael Francisco
Nat. Biotechnol. **27, 395, 2009; published online 7 April 2009; corrected after print 8 July 2009**

In the version of this article initially published, a company name was omitted from Table 2. GlaxoSmithKline should be listed in third place. The error has been corrected in the HTML and PDF versions of the article.

Erratum: Wyeth preemption case ruling sparks labeling confusion

Malorye Allison
Nat. Biotechnol. **27, 399–400 (2009); published online 8 May 2009; corrected after print 8 July 2009**

In the version of this article initially published, Phenergan is incorrectly mentioned in paragraph 2 as Merck's antinausea drug. Phenergan is made by Wyeth. The original version also states in paragraph 3 that Wyeth is located in Whitehouse Station, New Jersey. The company's correct location is Madison. The errors have been corrected in the HTML and PDF versions of the article.

Erratum: Academia and the company coin

Jim Kling
Nat. Biotechnol. **27, 411–414 (2009); published online 8 May 2009; corrected after print 8 July 2009**

In the version of this article initially published, on p. 411, left column, last paragraph, one of the researchers' names was incorrectly given as "Martin Feller." It should have read Martin Keller. The error has been corrected in the HTML and PDF versions of the article.