

3' UTR seed matches, but not overall identity, are associated with RNAi off-targets

Amanda Birmingham¹, Emily M Anderson¹, Angela Reynolds¹, Diane Ilsley-Tyree², Devin Leake¹, Yuriy Fedorov¹, Scott Baskerville¹, Elena Maksimova¹, Kathryn Robinson¹, Jon Karpilow¹, William S Marshall¹ & Anastasia Khvorova¹

Off-target gene silencing can present a notable challenge in the interpretation of data from large-scale RNA interference (RNAi) screens. We performed a detailed analysis of off-targeted genes identified by expression profiling of human cells transfected with small interfering RNA (siRNA). Contrary to common assumption, analysis of the subsequent off-target gene database showed that overall identity makes little or no contribution to determining whether the expression of a particular gene will be affected by a given siRNA, except for near-perfect matches. Instead, off-targeting is associated with the presence of one or more perfect 3' untranslated region (UTR) matches with the hexamer or heptamer seed region (positions 2–7 or 2–8) of the antisense strand of the siRNA. These findings have strong implications for future siRNA design and the application of RNAi in high-throughput screening and therapeutic development.

RNAi is a nearly ubiquitous pathway that modulates gene expression by post-transcriptional mechanisms. For the pathway to be a useful tool in the research and therapeutic venues, the functional intermediate of RNAi, the siRNA, must be both potent and specific in its targeting of messenger RNA transcripts. Considerable efforts have been invested to identify the key thermodynamic and sequence parameters that promote strong siRNA gene knockdown^{1–3}. These studies have led to the development of multiple design algorithms that enhance the selection of highly functional duplexes^{4–6}.

Although the current rational design algorithms accurately predict target gene knockdown, less is known about the parameters that contribute to siRNA specificity. Unintended gene modulation can result from lipid delivery reagents and siRNA induction of the innate cellular immunity. Whereas these sources of nonspecific gene regulation can be minimized by selecting nontoxic lipid formulations and optimal siRNA structures, a third contributor to unintended gene knockdown is associated with off-targeting. Off-target gene knockdown⁷ is an RNAi-mediated event that results in modest, 1.5- to 4-fold changes in the expression of dozens to hundreds of genes. Off-target effects can be mediated by either strand of the siRNA and have been documented to occur when as few as 15 base pairs of complementarity exist between the siRNA

and target⁷. As off-targeting can induce measurable phenotypes⁸, it represents one of the largest impediments to phenotypic screening applications for RNAi.

Because early studies in RNAi established that single-base-pair mismatches could dramatically alter siRNA functionality^{9–11}, local alignment algorithms such as Basic Local Alignment Search Tool (BLAST)¹² had been adopted to predict and enhance siRNA specificity. Unfortunately, although BLAST is an exceptional tool for broad sequence alignments, it falls short in its ability to accurately predict small local alignments. This shortcoming is primarily the product of the minimal word-size restraints used by US National Center for Biotechnology Information (NCBI) nucleotide BLAST (blastn) implementation that preclude the alignment of short sequences with particular bulge or mismatch configurations⁴. In contrast, the Smith-Waterman local alignment algorithm¹³ is well suited for detailed alignment analysis of short sequences.

We have applied common and customized weighting parameters to the Smith-Waterman algorithm to investigate the contribution of overall identity to siRNA-mediated off-targeting. Application of the Smith-Waterman algorithm to a database of experimentally identified off-targets revealed that with the exception of cases of near-perfect identity (for example, 18/19, 18/20, 19/20 and others) none of the parameter sets tested accurately distinguish between off-targeted and untargeted genes. Instead, perfect matches between the hexamer or heptamer seed (positions 2–7 or 2–8 of the antisense strand) of an siRNA and the 3' UTR—but not the 5' UTR or open reading frame (ORF)—were associated with off-targeting. These findings intimate a strong mechanistic parallel between siRNA off-targeting and microRNA-mediated gene regulation.

RESULTS

Development of an experimental off-targets database

Microarray-based gene expression analysis has previously been used as a method for off-target identification⁷. For this reason, as a first step in identifying and testing key parameters associated with siRNA off-targeting, we generated a database of experimentally validated off-targeted genes from the expression signatures of HeLa cells transfected with one of twelve different siRNAs (100 nM)

¹Dharmacon Research, 2650 Crescent Drive, #100, Lafayette, Colorado 80026, USA. ²Agilent Technologies, 3500 Deer Creek Rd., MS 25U-7, Palo Alto, California 94034, USA. Correspondence should be addressed to A.K. (khvorova.a@dharmacon.com).

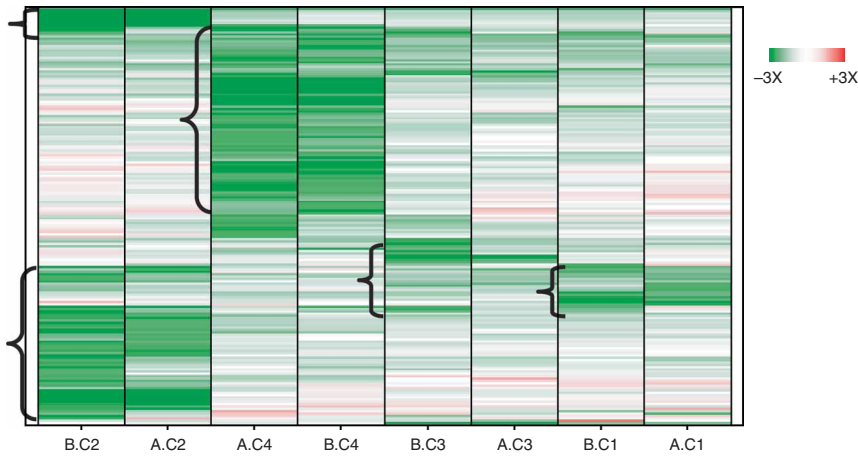


Figure 1 | Microarray analysis identifies off-targeted genes. Typical heatmap of HeLa cells transfected with four different *PPIB*-targeting siRNAs (C1, C2, C3 and C4). 'A' and 'B' represent biological replicates for transfection of each siRNA. Brackets highlight the clusters of sequence-specific off-targets of each siRNA. '+3X' and '-3X' represent fold changes in gene expression levels.

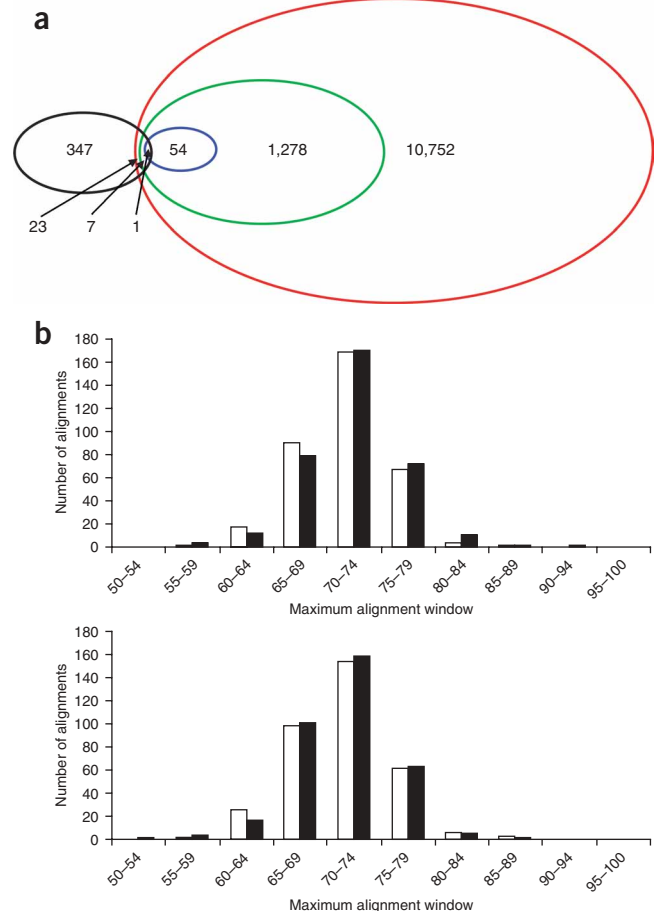
targeting three different genes (*PPIB*, *MAP2K1* and *GAPDH*). We transfected eleven rationally designed siRNA having a strong antisense strand bias toward RNA-induced silencing complex entry and one nonrationally designed siRNA into cells. Genes that were downregulated twofold or more by a given siRNA in one or more functional replicates, but were not modulated by other functionally equivalent siRNAs targeting the same gene were designated as off-targets. Expression signatures of cells transfected with the 12 siRNAs identified 347 off-targeted genes (Fig. 1 and Supplementary Table 1 online).

Comparison of *in silico* and experimental off-targets

Comparison of the validated off-target data set with *in silico* predicted off-targets showed that identity cutoffs do not accurately predict off-targeted genes. Using the Smith-Waterman alignment algorithm, the sense and antisense strands for each siRNA were aligned against the more than 20,000 genes represented on Agilent's Human 1A (V2) Oligo Microarray. The gene sequences that had $\geq 79\%$ identity with either the sense or antisense strands were designated as *in silico* predicted off-targets. We used common reward and penalty parameters (a match reward = 2, a mismatch penalty = -2 and a linear gap penalty = -3), and arbitrarily imposed

Figure 2 | Maximum sequence alignment does not accurately predict off-targeted gene regulation by RNAi. (a) Venn diagram shows overlap between 347 experimentally identified off-targets and *in silico* off-targets predicted by the Smith-Waterman alignment algorithm. Black, 347 experimentally validated off-targets for 12 separate siRNAs. Red, green and blue represent the number of off-targets predicted by Smith-Waterman using $\geq 79\%$ (for example, 15/19 or better), $\geq 84\%$ (for example, 16/19 or better) and $\geq 89\%$ (for example, 17/19 or better) identity filters, respectively. The numbers 23, 7 and 1 represent the number of genes that are common between the experimental and predicted groups at each of the identity filter levels ($\geq 79\%$, $\geq 84\%$ and $\geq 89\%$, respectively). (b) The sense (top) and antisense (bottom) sequences of each siRNA were aligned separately to the sequences of their corresponding 347 experimentally validated off-targets and a comparable number of control untargeted genes to identify the alignments with the maximum percent identity. The number of alignments in each identity window were then plotted for the off-targeted (black) and untargeted (white) populations.

a maximum cutoff of 1,000 alignments per siRNA. (Although multiple alignments between a given siRNA and mRNA were recorded, we carried out the analyses using only the best alignment between each pair). Surprisingly, the number of *in silico* predicted off-targets typically exceeded the number identified by microarray analysis by one to two orders of magnitude, regardless of whether alignments of one or both strands were included in the analysis (Fig. 2a and Supplementary Table 2 online). This results in a false positive rate of over 99% at the 79% identity cutoff. This number of predicted off-targets represented more than one third of the number of mRNAs in the human genome. Moreover, only 23 of the 347 experimentally validated off-targets were identified by *in silico* methods using this cutoff, which represents a false negative rate of approximately 93%. Higher cutoffs ($\geq 84\%$ and $\geq 89\%$) produced similarly poor overlap between experimental and *in silico* target predictions (7 and 1 commonly identified targets using the 84% and 89% identity filter, respectively), as well as gross mis-estimations of the number of off-targets (1,278 and 54, respectively). Figure 2b is a comparative histogram showing the distribution of best alignments of the 347 off-targeted



genes and a randomly chosen set of untargeted genes. Based on these observations, we concluded that overall sequence identity is a poor predictor of the number and identity of off-targeted genes.

Alignments are particularly sensitive to the weighting of matches, mismatches and gaps. With the long-term goal of creating a customized Smith-Waterman parameter set that can distinguish between off-targeted and untargeted populations, we synthesized individual siRNAs targeting human cyclophilin B (*PPIB*), firefly luciferase (*Ppyr\LUC*), and secreted alkaline phosphatase (*ALPPL2*)

in their native state or with one of three base pair mismatches at each of the 19 positions of the duplex (57 variants per siRNA). Next we performed a systematic single-mismatch analysis of siRNA functionality with the understanding that although such an analysis may provide insights into nucleotide pairing weights, it is not designed to take into account effects of multiple nonadjacent mismatches and/or secondary structure.

The results of *PPIB* and *Ppyr\LUC* studies (Fig. 3a–c) clearly show that the central region of the duplex (positions 9–12) is

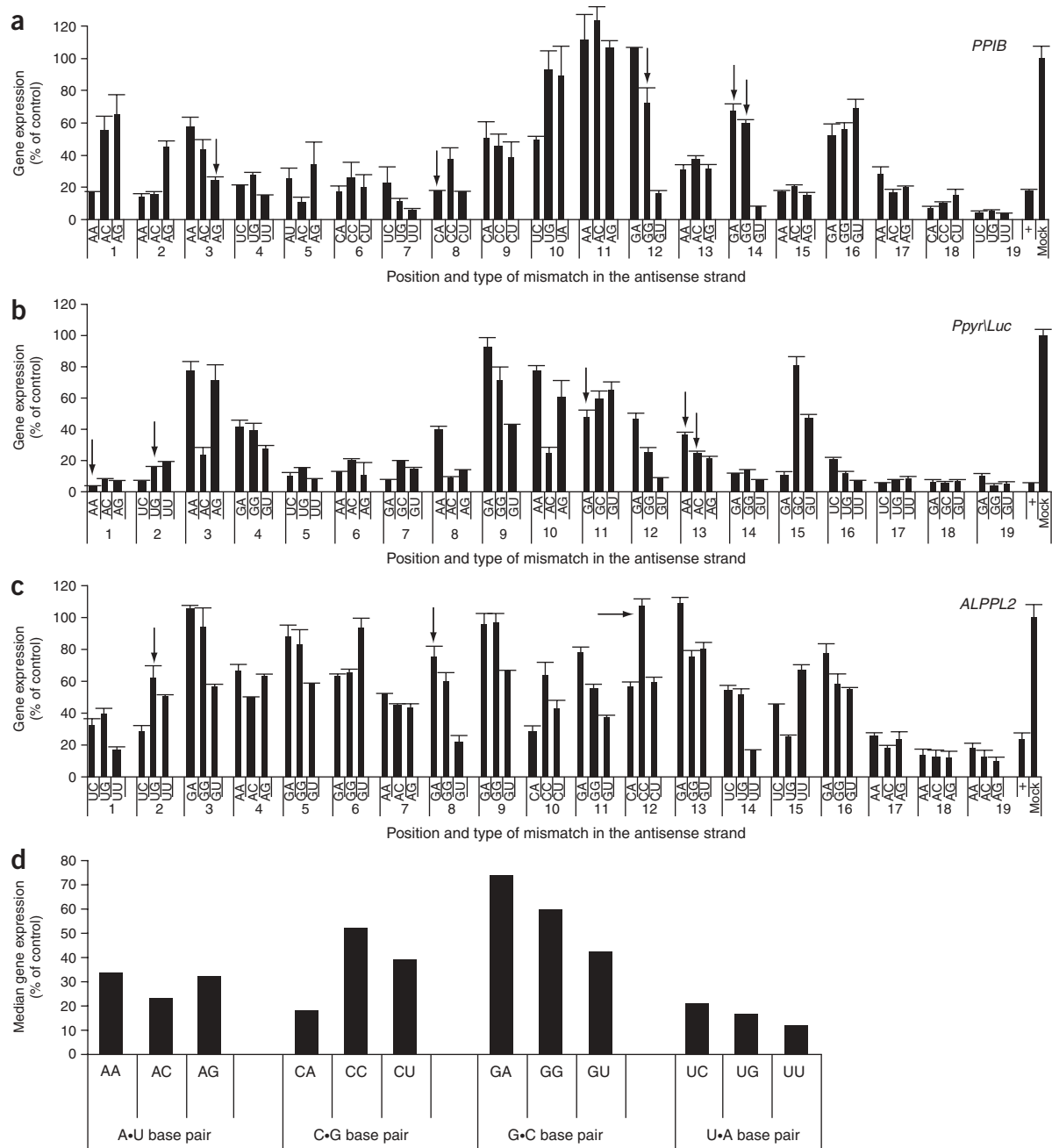


Figure 3 | Systematic single base pair-mismatch analysis of siRNA functionality. (a–c) Effects of single base pair mismatch in siRNAs targeting *PPIB* (a), *Ppyr\LUC* (b) and *ALPPL2* (c). Native forms of all three siRNAs induce > 90% gene knockdown. Position 1 refers to the 5'-most position of the antisense strand. The top base represents the antisense mutation, and the bottom base represents the mismatched sense strand nucleotide. 'Mock', lipid-treated cells; '+', native duplex. Arrows point to examples of positions that have equivalent bases with at least one other siRNA in the test group. Experiments were performed in triplicate. Error bars, s.d. from the mean. (d) Bar graph of overall impact of mismatch identity on siRNA function.

particularly sensitive to mismatches. In contrast, duplexes with mismatches at positions 18 and 19 exhibit consistent silencing, suggesting that the strength of base pairing in this region is less critical. Outside of positions 9–12 and 18–19, identical mismatches at any position can have widely disparate impacts on siRNA performance. For instance, while an AG mismatch at position 3 of the *PP1B* siRNA has little impact on overall duplex functionality, the same mismatch at the same position in the luciferase-targeting siRNA dramatically alters silencing efficiency. These findings suggest that with the exception of positions 18 and 19, the complete sequence has a role in determining the impact of mismatches, thus preventing the development of clear position-dependent mismatch criteria. Nonetheless, analysis of all mismatches in a position-independent manner identified a decided bias (Fig. 3d).

We incorporated the observed biases into 30 additional Smith-Waterman parameter sets (Supplementary Table 3 online) to test whether changes in the rewards and costs associated with matches and mismatches could improve the ability to predict off-targeted genes by overall alignment identity. As it is not feasible to test the entire parameter space of the Smith-Waterman algorithm, we chose parameter sets to represent possible biologically relevant interactions. We designed match and mismatch parameters that model various levels of helix stability in response to different pairings. As it is unclear how RNAi tolerates gaps, we included different gap penalties (both linear and affine) in the scoring matrices. The gap penalties simulate interactions allowing single-base and multi-base asymmetrical bulges in the helix. We analyzed two populations of siRNA-mRNA pairs (180 representing those with experimentally validated off-target interactions and 180 with no discernable off-target interactions) with each of the thirty unique scoring schemes. Analysis of off-targeted and untargeted populations using each of the modified parameter sets did not distinguish between the two data sets regardless of whether alignments for one or both strands were included. The finding that the distributions of maximum identity (in the best alignment) for each parameter set for off-targeted and untargeted populations are statistically indistinguishable ($P > 0.05$ after application of Bonferroni correction for multiple comparisons; Supplementary Fig. 1 online) supports the previous conclusion that overall sequence identity is a poor predictor of off-targeted genes.

Off-targeting and seed matches in 3' UTR

Recent studies on microRNA (miRNA)-mediated gene modulation have shown that complementary base pairing between the seed region and sequences in the 3' UTR of mRNA are associated with miRNA-mediated gene knockdown¹⁴. As siRNAs and miRNAs are believed to share some portion of the RNAi machinery, we investigated whether complementarity between the seed region of the siRNA and any region of the transcript was associated with off-targeting. To accomplish this, we scanned the 5' UTR, ORF and 3' UTR of 84 experimentally determined off-target genes for exact complementary matches to the antisense seed region (hexamer, positions 2–7 and heptamer, positions 2–8) of their respective siRNA. Then we compared this data set of siRNAs and their off-targeted genes to a control group (84 siRNA-mRNAs that shared no off-target interactions) to determine whether seed matches in any of the three regions correlated with off-targeting. For 5' UTR and ORF sequences, the frequency at which one or more hexamer seed matches were present in the experimental and control groups was

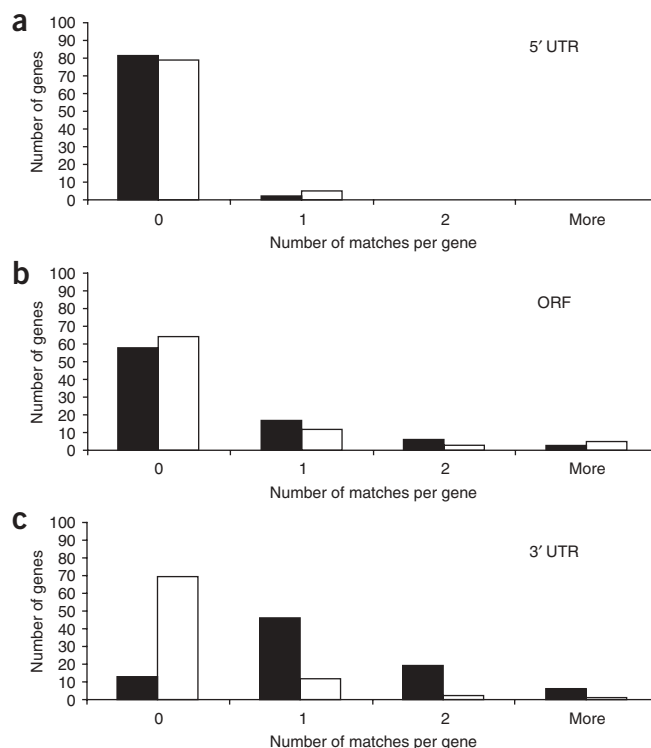


Figure 4 | Exact complementarity between the siRNA seed region and the 3' UTR (but not 5' UTR or ORF) distinguishes off-targeted from untargeted genes. (a–c) A search for complementarity between the siRNA antisense seed region (positions 2–7) and 5' UTRs (a), ORFs (b), and 3' UTRs (c) of off-targeted (84 genes, black bars) and untargeted (84 genes, white bars) genes was performed. The number of genes having one or more exact 3' UTR seed matches is almost fivefold greater in off-targeted genes than untargeted genes (82% of off-targeted genes versus 17% of untargeted genes).

statistically indistinguishable (at the $P > 0.05$ level using the χ^2 test for independence; frequencies were 2.3% and 5.9% for the 5' UTR, 30.9% and 23.8% for ORF sequences, respectively). In contrast, the incidence at which one or more hexamer matches were found in the 3' UTR of off-targets was nearly fivefold higher than that observed in the untargeted populations (84.5% in the experimental group, 17.8% in the control group; significant with $P < 0.001$; Fig. 4). The positive predictive value (defined as number of true positives / (number of true positives + number of false positives)) of the association between 3' UTR hexamer seed matches and off-targeted genes increased when multiple matches were required (for two or more 3' UTR matches: off-targeted genes = 29.76%, untargeted genes = 3.57%; Table 1). When four 3' UTR hexamer seed matches are present, we did not detect any false positives in this limited sample. As seed matches provide an enhancement over the predictive abilities of blastn and Smith-Waterman homology-based searches, we developed a web-based search tool (Supplementary Fig. 2 online) to allow identification of all possible off-targets based on 3' UTR hexamer seed matches for any given siRNA (<http://www.dharmacon.com/seedlocator/default.aspx>).

The frequency at which we observed heptamer seed matches in the 5' UTR, ORF and 3' UTR of experimental and control groups was similar to that documented for hexamers (heptamer frequency in experimental and control groups: 5' UTR: 0% and 1.2%; ORF:

Table 1 | Sensitivity, specificity and positive predictive power of siRNA hexamer and heptamer seed matches

Criteria	True positives	False positives	True negatives	False negatives	Sensitivity (%)	Specificity (%)	Positive predictive power (%)
At least 1 hexamer in 3' UTR	71	15	69	13	85	82	83
At least 2 hexamer in 3' UTR	25	3	81	59	30	96	89
At least 3 hexamer in 3' UTR	6	1	83	78	7	99	86
At least 4 hexamer in 3' UTR	4	0	84	80	5	100	100
At least 1 heptamer in 3' UTR	58	7	77	26	69	92	89
At least 2 heptamer in 3' UTR	8	0	84	76	10	100	100
At least 3 heptamer in 3' UTR	1	0	84	83	1	100	100
At least 4 heptamer in 3' UTR	0	0	84	84	0	0	NA

16.6% and 9.5%; 3' UTR: 69.1% and 8.3%) suggesting that the relevant seed region may consist of seven nucleotides (positions 2–8) rather than six (positions 2–7). As was observed with hexamer seed matches, increases in the number of 3' UTR heptamer seed matches were associated with improvements in the specificity of the association. The observed associations remained after 3' UTR length is controlled for by examining paired off-targeted and nontargeted control 3' UTRs with lengths equal to within thirty bases (**Supplementary Fig. 3** online), thus suggesting that 3' UTR-siRNA seed matches are an important parameter of off-targeting.

DISCUSSION

The work presented here demonstrates that with the exception of instances of near-perfect complementarity, the level of overall complementarity between an siRNA and any given mRNA is not associated with off-target identity. Although this approach is limited by its inability to assess the possible synergistic effects of multiple nonadjacent mismatches or secondary structures, these findings reveal that current protocols used to minimize off-target effects (for example, blastn and Smith-Waterman) have little merit aside from eliminating the most obvious off-targets (that is, sequences that have identical or near-identical target sites) and likely discard substantial numbers of functional siRNAs owing to unfounded specificity concerns.

Studies of the RNAi pathway have led to the identification of dozens of parameters that have a role in determining siRNA functionality. Individually, none of these parameters are sufficient to ensure siRNA functionality, yet when combined into a single algorithm, they provide a strong predictive function. The siRNA seed–3' UTR match is only one parameter in what is assumed to be an extremely complex phenomenon. Because the sheer number of genes that contain matches with any given siRNA seed region is very large in comparison to the number of actual off-targets for that siRNA, the value of the identified parameter (by itself) is limited. The identification of additional factors that have roles in off-targeting will likely lead to development of predictive algorithms that minimize off-target effects and enhance siRNA design.

METHODS

siRNA synthesis. We synthesized siRNA duplexes using 2' ACE chemistry¹⁵. The siRNA sequences are available in **Supplementary Table 1**.

Transfection. We obtained HeLa and HEK293 cells from ATCC. We grew cells at 37 °C in a humidified atmosphere with 5% CO₂

in Dulbecco's Modified Eagle medium (HyClone), 10% fetal bovine serum, and 2 nM L-glutamine. All propagation media were further supplemented with penicillin (100 U/ml) and streptomycin (100 µg/ml). For transfection experiments, we seeded cells at 1.0–2.0 × 10⁴ cells/well (in a 96-well plate) 24 h before the experiment (antibiotic-free medium). We transfected cells with siRNA (100 nM or 50 nM) using Lipofectamine 2000 (0.25 µl/well; Invitrogen). For targeting of *Ppyr\LUC* or *ALPPL2*, we performed cotransfections of plasmid (*Ppyr\LUC*: 70 ng/well pGL3 (Promega) or *ALPPL2*: 25 ng/well pCMV-ALPPL2) and siRNA using Lipofectamine 2000 at 0.5 µl/well in HEK293 cells plated at 2.5 × 10⁴ cells/well in a 96-well plate.

Gene knockdown and cell viability assay. Twenty-four hours after transfection, we assessed the knockdown of endogenous targets using a branched DNA assay (Genospectra). In all experiments, we used *GAPDH* as a reference. When *GAPDH* was the target gene, we used *PPIB* as a reference. We measured *Ppyr\LUC* knockdown (24 h post-transfection) with the Steady-Glo enzymatic assay (Promega). We measured *ALPPL2* (72 h post-transfection) knockdown using the Great EscAPE SEAP enzymatic assay (Clontech). To assess cellular viability, we added 25 µl of AlamarBlue reagent (Trek Diagnostic Systems) to each well, and incubated the cells for 2 h at 37 °C, 5% CO₂. Absorbance was read at 570 nm using a 600 nm subtraction. Transfections resulting in an optical density of ≥80% of the control were considered nontoxic.

Microarray experiments. For each sample, we amplified 1 µg of total RNA isolated from siRNA-treated HeLa cells (collected 24 h post-transfection), labeled the products with Cy5 (Cy-5 CTP; Perkin Elmer) using Agilent's Low Input RNA Fluorescent Linear Amplification Kit and hybridized against Cy3-labeled material derived from lipid-treated (control) samples. Hybridizations were performed using Human 1A (V2) Oligo Microarrays (Agilent; ~21,000 unique probes) according to a published protocol (<http://www.chem.agilent.com/scripts/literaturepdf.asp?iWHID=36866>). We washed the slides using 6× and 0.06× saline–sodium phosphate–EDTA (SSPE) buffers (Amresco; each with 0.025% N-lauroylsarcosine), dried them using Agilent's nonaqueous drying and stabilization solution, and scanned them on a Microarray Scanner (Agilent; model G2505B). The raw image was processed using Feature Extraction software (v7.5.1). Further analysis was performed using Spotfire Decision Site 7.2 software and the Spotfire Functional Genomics Module. We did not use outlier

flagging. Off-targets were identified as genes that were down-regulated by twofold or more (log ratio of more than -0.3) by a given siRNA in at least one experiment, but were not modulated by other functionally equivalent siRNA targeting the same gene. Biological replicate measurements denote replicate transfections of a given siRNA at 100 nM on two separate days. Functional replicate measurements can include siRNAs transfected at 50 and 100 nM in the same experiment, which generate an equivalent level of target knockdown and extent of off-target silencing. The list of off-targeted genes is available in **Supplementary Table 1**.

Computational analysis. We implemented the Smith-Waterman local alignment algorithm in C# and augmented it to extend alignments along the entire length of the shorter aligned sequence. The implementation also allowed the use of either uniform match rewards and mismatch costs or scoring matrices, and either linear or single affine gap costs.

The first stage of analysis used this implementation to align each strand of 12 siRNAs (including one nonrationally designed siRNA) against all GenBank mRNAs represented on the microarray chip. We archived the 1,000 highest-percent identity alignments (on either strand) for each siRNA; because of this cut-off, the alignment coverage at the 79% level was probably incomplete. We analyzed the archived alignments to determine their identity distributions and discover alignments with experimentally off-targeted mRNAs, using the validated dataset of 347 off-targets, including all accession numbers that were sequence-specifically downregulated by twofold or more in at least one functional replicate.

The parameter-testing work defined twelve scoring matrixes designed to reward complementarity rather than identity. Each scoring matrix was combined with at least one linear gap penalty (designed to allow only one gap at a time) and one single affine gap penalty (designed to allow multiple-gap runs) of varying weights to generate the 30 parameter sets described in **Supplementary Table 3**. We limited the dataset of experimental off-targets to include only those 180 that were sequence-specifically down-regulated by approximately twofold or more in two functional replicates for the 11 rationally designed siRNAs and had well-annotated coding sequences. We chose a control set at random from those mRNAs that were not substantially down-regulated by any of the test siRNAs; each siRNA was assigned as many controls as it had off-targets. For each parameter set, we used the Smith-Waterman algorithm to align each strand of the siRNAs with their off-targets' reversed mRNA (because of the complementary nature of the scoring matrices) and archived the best 20 alignments; we repeated the process for the control set. Analysis identified, for each siRNA-mRNA pair, the archived alignment (including both strands) containing the largest percentage of matched bases; this was termed the 'best-possible alignment'. We generated histograms showing distributions of the percent matches of these best possible alignments for each data set under each parameter set. Because all distributions except those for sets 29 and 30 were approximately normal, we subjected each off-target control distribution pair, except these two, to a two-tailed *t* test to determine whether their means were significantly

different. We subjected sets 29 and 30 to a χ^2 test for independence. We adjusted the results of all tests using the Bonferroni correction to account for multiple comparisons. We also conducted the analysis for each strand individually.

We performed the seed analysis using a stringent subset of the experimentally validated off-targets including only those 84 with well-annotated UTRs that were sequence-specifically down-regulated by at least twofold in two biological replicates for eight siRNAs measured in a single experiment; the control set was correspondingly narrowed. The analysis counted occurrences of exact substrings (identical to positions 13–18 inclusive, hexamer; and 12–18 inclusive, heptamer) of the siRNA sense strand to the 5' UTR, ORF and 3' UTRs of each off-target and control.

Additional information. A list of genes used in this study, with accession numbers, is available in a **Supplementary Note** online.

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

ACKNOWLEDGMENTS

We thank R. Knight for discussion and data analysis advice. We also thank the Dharmacon Production Team for synthesizing the siRNAs used in this work, and J. Kendall and A. O'Brien for providing assistance and direction in manuscript preparation.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

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

- Schwarz, D.S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
- Reynolds, A. *et al.* Rational siRNA design for RNA interference. *Nat. Biotechnol.* **22**, 326–330 (2004).
- Khvorova, A., Reynolds, A. & Jayasena, S. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
- Naito, Y., Yamada, T., Ui-Tei, K., Morishita, S. & Saigo, K. siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.* **32**, W124–W129 (2004).
- Jagla, B. *et al.* Sequence characteristics of functional siRNAs. *RNA* **11**, 864–872 (2005).
- Huesken, D. *et al.* Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.* **23**, 995–1001 (2005).
- Jackson, A.L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635–637 (2003).
- Lin, X. *et al.* siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res.* **33**, 4527–4535 (2005).
- Elbashir, S.M., Harborth, J., Weber, K. & Tuschl, T. Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* **26**, 199–213 (2002).
- Amarzguioui, M., Holen, T., Babaie, E. & Prydz, H. Tolerance for mutations and chemical modifications in a siRNA. *Nucleic Acids Res.* **31**, 589–595 (2003).
- Holen, T., Amarzguioui, M., Wiiger, M.T., Babaie, E. & Prydz, H. Positional effects of short interfering RNAs targeting the human coagulation trigger tissue factor. *Nucleic Acids Res.* **30**, 1757–1766 (2002).
- Altschul, S. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Smith, T.F. & Waterman, M.S. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981).
- Lim, L. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
- Scaringe, S.A. Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis. *Methods Enzymol.* **317**, 3–18 (2000).