Expression profiling reveals off-target gene regulation by RNAi

Aimee L Jackson[1,2], Steven R Bartz[1,2], Janell Schelter[1], Sumire V Kobayashi[1], Julja Burchard[1], Mao Mao[1], Bin Li[1], Guy Cavet[1] & Peter S Linsley[1]

RNA interference is thought to require near-identity between the small interfering RNA (siRNA) and its cognate mRNA. Here, we used gene expression profiling to characterize the specificity of gene silencing by siRNAs in cultured human cells. Transcript profiles revealed siRNA-specific rather than target-specific signatures, including direct silencing of nontargeted genes containing as few as eleven contiguous nucleotides of identity to the siRNA. These results demonstrate that siRNAs may cross-react with targets of limited sequence similarity.

RNA interference (RNAi) is a potent method to suppress gene expression in mammalian cells, and has generated much excitement in the scientific community[1]. Specific gene silencing promises the potential to harness human genome data to elucidate gene function, identify drug targets and develop more specific therapeutics. Numerous reports in the literature describe the exquisite specificity of siRNAs, suggesting a requirement for near-perfect identity with the siRNA sequence[2–4]. However, most of the published analyses of siRNA-induced gene silencing have examined only one or a few genes in addition to the targeted gene, an approach not unlike 'looking for keys under the lamppost.' Cross-hybridization with transcripts containing partial identity to the siRNA sequence may elicit phenotypes reflecting silencing of unintended transcripts in addition to the target gene. We used expression profiling for an unbiased, genome-wide analysis of the efficacy and specificity of siRNA-induced silencing of two genes involved in signal transduction, the insulin-like growth factor receptor (*IGF1R*) and mitogen-activated protein kinase 1 (*MAPK14*, also known as *p38α*).

Using standard selection rules[5], we designed 16 siRNAs to target the coding region of *IGF1R* and 8 siRNAs to target *MAPK14* (see **Supplementary Table 1** online). After the siRNAs were transfected into HeLa cells, we compared the expression profiles resulting from silencing of the same target gene by different siRNAs (see **Supplementary Methods** online). Surprisingly, our analysis revealed few genes regulated in common by different siRNAs to the same target gene. Each of the 8 siRNA duplexes targeted to *MAPK14* produced a distinct expression pattern (**Fig. 1a**). Likewise, each of the 16 siRNA duplexes to *IGF1R* produced a unique expression pattern (**Fig. 1b**). Virtually identical gene expression patterns were observed in three independent experiments, demonstrating that gene regulation resulting from a particular siRNA was reproducible. Detailed analysis suggests that different siRNAs to the same target transcript elicit a small number of gene regulations in common (data not shown), but the vast majority of the transcript expression patterns were siRNA-specific rather than target-specific. The number and identity of altered transcripts did not correspond to the ability of the siRNA to silence the target gene. Target mRNA levels were correlated with protein levels and the extent of mRNA silencing measured by TaqMan was equivalent to that measured by array profiling. An siRNA targeted to luciferase reproducibly regulated the expression of several genes despite the lack of a homologous target in the human genome. Thus, we have observed patterns of gene regulation that are specific for the siRNA sequence used for silencing, rather than for the intended target.

We subsequently performed detailed concentration and kinetic analysis of MAPK14 protein and RNA knockdown by siRNA *MAPK14-1*. Although target gene silencing was detectable when the siRNA concentration was decreased 1,000-fold, off-target gene regulation was also detectable (**Fig. 1c**). Many of these genes showed nearly identical half-maximal responses with respect to siRNA concentration as *MAPK14* (~1nM). We were unable to titrate the off-target gene regulation from silencing of the intended target, indicating that off-target gene regulation is not simply an artifact of high siRNA concentration. We then analyzed temporal gene expression patterns. The MAPK14 protein demonstrated a half-life of approximately 40 h after siRNA transfection (see **Supplementary Fig. 1** online). In contrast, the *MAPK14* transcript demonstrated half-maximal degradation approximately 11 h post-transfection (see **Supplementary Fig. 1** online). Through expression profiling, we observed a surprising degree of gene regulation at early time points (6–12 h) well before any observable decrease in the MAPK14 protein (see **Supplementary Fig. 1** online). These gene expression changes therefore were unlikely to be secondary events resulting from loss of MAPK14 function. The expression signature could be divided into several temporally distinct groups of transcripts based on timing of half-maximal gene regulation (**Fig. 2a**). Group 1 contains a single transcript, the intended target *MAPK14*. Group 2 contains nine transcripts demonstrating similar kinetics of silencing to *MAPK14*, with half-maximal degradation at 7–13 h as determined by microarray. This same group of transcripts was downregulated with rapid kinetics in a separate experiment, demonstrating that these genes were reproducibly silenced by this siRNA. The rapid kinetics of transcript regulation suggests that these are direct transcript degradation events. This is in contrast to kinetic groups 3 and 4, for which half-maximal degradation occurs at approximately 40 h and therefore likely represents secondary gene expression changes.

None of the genes in kinetic group 2 (**Fig. 2b**) are known to function in the *MAPK14* pathway. Remarkably, all of these transcripts were found to contain regions of partial sequence identity to the siRNA duplex (**Fig. 2b**). Sequence alignment demonstrated that these genes could be divided into two subgroups. One subgroup contained a core of 14–15 nucleotides of similarity encompassing the central

[1]Rosetta Inpharmatics LLC, 12040 115th Avenue NE, Kirkland, Washington 98034, USA. [2]These authors contributed equally to this work. Correspondence should be addressed to A.L.J. (aimee_jackson@merck.com) or S.R.B. (steven_bartz@merck.com).

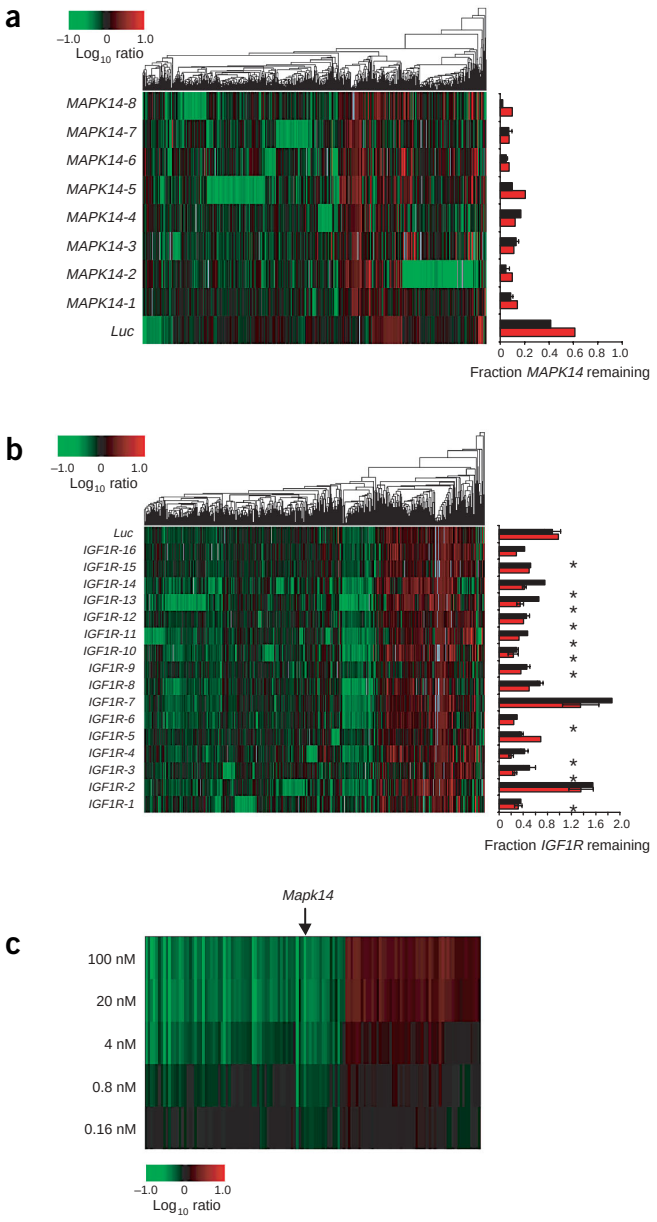

Figure 1 Expression profiling reveals siRNA sequence-specific and concentration-dependent gene expression patterns. (**a**) Eight different siRNA duplexes targeted to the *MAPK14* coding region were used for gene silencing in HeLa cells. *Luc*, siRNA targeted to luciferase. Green indicates decreased expression relative to mock transfection, red indicates increased expression. The bar graph represents the fraction of target protein (red) and RNA (black) remaining 48 h after siRNA transfection (see **Supplementary Methods** and ref. 8 for details on microarray analysis). (**b**) Sixteen different siRNA duplexes targeted to the *IGF1R* coding region were used for gene silencing. Asterisks indicate the *IGF1R* siRNA duplexes that reduce protein level by at least 60%. (**c**) HeLa cells were transfected with the indicated concentrations of *MAPK14-1* siRNA.

region of the siRNA sequence. The second subgroup contained a smaller core of similarity encompassing the nine nucleotides at the 3′ end of the siRNA sense-strand sequence. This is in contrast to transcripts in kinetic groups 3–5. Many of these genes lacked substantial sequence similarity to the siRNA, as determined by FASTA analysis, and those that did displayed only short stretches (<6–8 nucleotides) of similarity distributed randomly throughout the siRNA sequence (data not shown.) Thus, the bias for a core of sequence similarity encompassing the 3′ end of the siRNA is unique to the rapidly silenced transcripts. The same pattern of sequence similarity to genes silenced with rapid kinetics was also observed with two different siRNAs to *IGF1R* (data not shown.) Interestingly, for one of these siRNAs, the off-target gene silencing was directed by the antisense strand of the siRNA, whereas for the other siRNA the off-target gene silencing appeared to be directed by the sense strand. This suggests that both the sense and antisense strands of an siRNA duplex can contribute to transcript silencing.

On the basis of published reports that gene silencing was abolished by single nucleotide changes in the siRNA sequence[2,6], we would not have predicted that this limited degree of sequence similarity would be sufficient for transcript silencing. However, to test this possibility, we systematically substituted the nucleotide at each position of the siRNA sequence and determined the effect of the altered sequence on the expression signature. Although a comprehensive discussion of this analysis is beyond the scope of this report, representative results are presented in **Fig. 2c**. A single nucleotide substitution at position 4 reduced silencing of *MAPK14*, and abolished silencing of the off-target genes in subgroup 1 that contain similarity to *MAPK14* at this position. However, silencing was not abolished for the six off-target genes in subgroup 2 that do not contain similarity to *MAPK14* in this region, confirming that silencing of these off-target genes is independent of loss of *MAPK14* expression. A single nucleotide substitution at position 15 reduced *MAPK14* silencing, and abolished silencing of all nine off-target genes, presumably because all nine transcripts contain similarity to *MAPK14* in this region.

An siRNA duplex designed to target a different sequence in *MAPK14* (*MAPK14-2*) silenced *MAPK14* but not any of the group 2 transcripts, which are not similar to the new siRNA sequence (**Fig. 2c**). Finally, siRNAs for two off-target transcripts, *KPNB3* and *FLJ20291*, silenced the expression of *MAPK14* in addition to their intended targets (**Fig. 2d**). The *KPNB3* siRNA shares 14 contiguous nucleotides, and a total of 15 nucleotides, of identity with *MAPK14*. The *FLJ20291* siRNA shares only 11 contiguous nucleotides, and a total of 15 nucleotides, of identity with *MAPK14-1*. In conclusion, 15 nucleotides, and perhaps as few as 11 contiguous nucleotides, of sequence identity are sufficient to direct silencing of nontargeted transcripts and therefore, although RNA interference results in robust silencing of the desired target, off-target gene regulation can occur as a result of degradation of mRNA transcripts with partial identity to the siRNA sequence.

Although the early regulated genes show sequence similarity to the siRNA, not all possible transcripts with this level of sequence similarity were silenced. Despite repeated attempts, we have thus far been unable to identify patterns that could help predict off-target activity of siRNAs. Furthermore, the transcripts silenced with early kinetics are only a portion of the total signature. Although we cannot explain all of the off-target gene regulation, we believe that at least a portion of it can be explained by cross-hybridization to transcripts of similar sequence.

The finding of a core of sequence similarity to the 3′ end of the siRNA sense strand (the 5′ end of the antisense strand) invites speculation that this region of the siRNA plays an important role in transcript silencing. This is strikingly reminiscent of the finding that the 5′ ends of many microRNAs (miRNAs) contain seven nucleotides of sequence that are complementary to target regulatory motifs in the 3′ untranslated regions of mRNA[7]. Sequence complementarity to these regula-

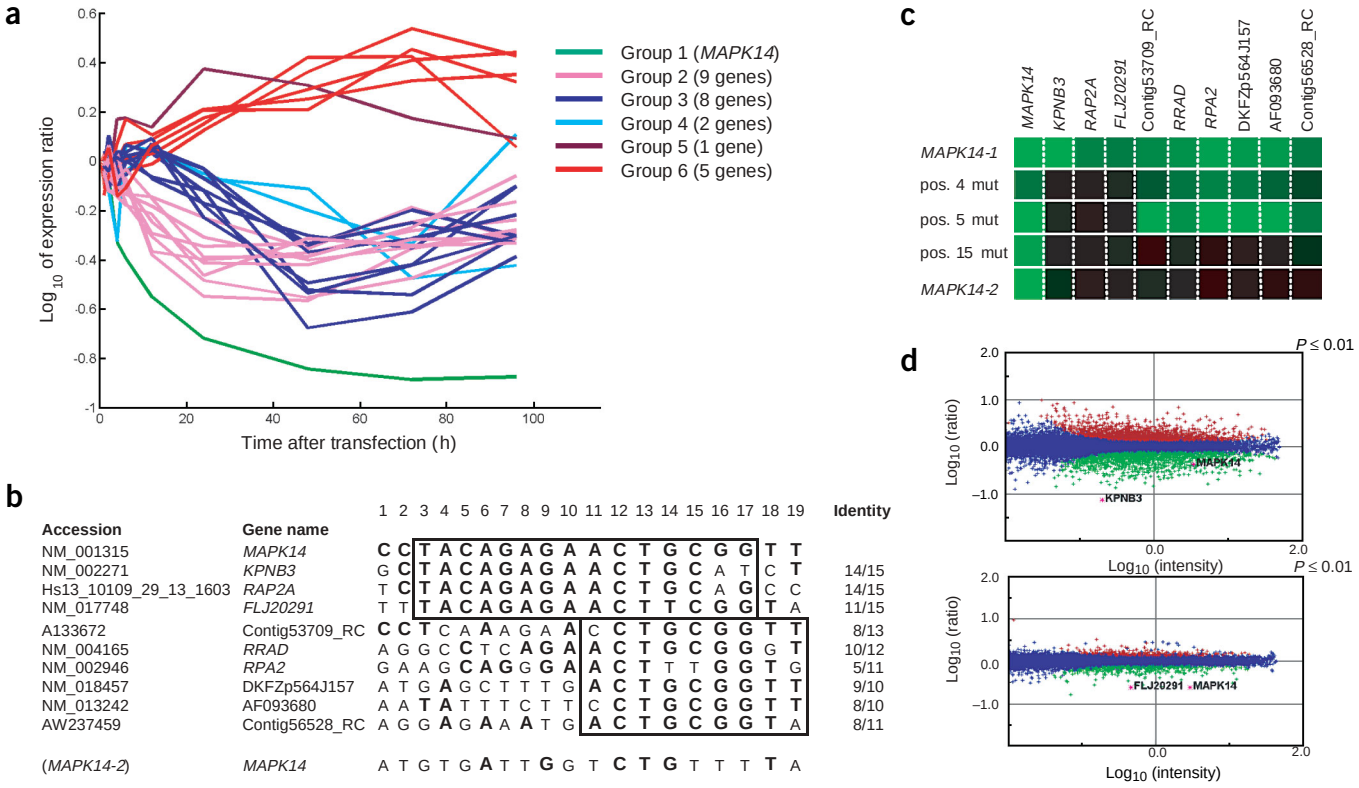

Figure 2 Kinetic and sequence analysis of *MAPK14* knockdown by RNAi. (**a**) RNA extracts were harvested at the indicated times after transfection of HeLa cells and processed for expression profiling. Transcript data from the microarray analysis were analyzed by trend plot to determine timing of half-maximal transcript degradation. Data are presented as $\log_{10}$ of expression ratio plotted as a function of time after transfection. (**b**) Sequence alignment of genes regulated with similar kinetics to *MAPK14*. Nucleotides with perfect identity to the *MAPK14* sequence are indicated in bold, mismatched nucleotides are indicated in small font. Accession numbers listed are from the NCBI GenBank database. Gene names are from LocusLink except for the EST contig designations, which are from www.pharp.org/est_assembly/human/gene_number_methods.html. The degree of sequence identity to *MAPK14-1* is indicated as the number of contiguous identical nucleotides /the total number of identical nucleotides. (**c**) HeLa cells were transfected with *MAPK14-1* siRNA, or *MAPK14-1* siRNA containing a single nucleotide substitution. (see **Supplementary Table 1** online for siRNA sequences). (**d**) Gene expression resulting from silencing by siRNAs to the off-target genes *KPNB3* and *FLJ20291*. Microarray data are plotted as $\log_{10}$ of expression ratio versus $\log_{10}$ of fluorescence intensity. Blue represents transcripts with unaltered expression. The targeted genes, as well as *MAPK14*, are indicated.

tory motifs was located exclusively at the 5′ ends of the miRNAs, suggesting a role in recognition or stabilization of the miRNA-mRNA interaction. The same may be true for the siRNA-mRNA interaction. Transcripts containing sequence similarity to this portion of the siRNA could potentially be collateral targets for silencing.

Given the small degree of similarity implicated in off-target gene regulation, it may be difficult to select an siRNA sequence that will be absolutely specific for the target of interest. Until siRNA design can be improved to convincingly reduce this off-target activity, incorporating multiple siRNA duplexes to silence a target gene of interest will increase the confidence with which an observed phenotype and expression pattern can be linked to target gene silencing. In this way, expression profiling in conjunction with gene silencing by RNAi will provide an effective means to identify and characterize gene function in cultured mammalian cells.

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

ACKNOWLEDGMENTS

We thank G. Schimmack, S. Kim, M. Imakura and J. McLarnan for expert technical assistance, M. Biery for assistance with computer graphics and the Rosetta Gene Expression Laboratory for microarray work.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 10 February; accepted 18 March 2003
Published online 18 May 2003; doi:10.1038/nbt831

1. Couzin, J. *Science* **298**, 2296–2297 (2002).
2. Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. *EMBO J.* **20**, 6877–6888 (2001).
3. Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P. & Sharp, P.A. *Genes Dev.* **13**, 3191–3197 (1999).
4. Hutvagner, G. & Zamore, P.D. *Science* **287**, 2056–2060 (2002).
5. Elbashir, S.M., Harborth, J., Weber, K. & Tuschl, T. *Methods* **26**, 199–213 (2002).
6. Martinez, L.A. *et al. Proc. Natl. Acad. Sci. USA* **99**, 14849–14854 (2002).
7. Lai, E.C. *Nat. Genetics* **30**, 363–364 (2002).
8. Hughes, T.R. *et al. Nat. Biotechnol.* **19**, 342–347 (2001).