Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies

Ralf Kittler[1,7,8], Vineeth Surendranath[1,2,8], Anne-Kristin Heninger[1], Mikolaj Slabicki[1], Mirko Theis[1], Gabriele Putz[1], Kristin Franke[1], Antonio Caldarelli[1], Hannes Grabner[1,3], Karol Kozak[1,3], Jan Wagner[1,3], Effi Rees[4], Bernd Korn[4], Corina Frenzel[5], Christoph Sachse[5], Birte Sönnichsen[5], Jie Guo[6], Janell Schelter[6], Julja Burchard[6], Peter S Linsley[6], Aimee L Jackson[6], Bianca Habermann[1,2] & Frank Buchholz[1]

RNA interference (RNAi) has become an important technique for loss-of-gene-function studies in mammalian cells. To achieve reliable results in an RNAi experiment, efficient and specific silencing triggers are required. Here we present genome-wide data sets for the production of endoribonuclease-prepared short interfering RNAs (esiRNAs) for human, mouse and rat. We used an algorithm to predict the optimal region for esiRNA synthesis for every protein-coding gene of these three species. We created a database, RiDDLE, for retrieval of target sequences and primer information. To test this *in silico* resource experimentally, we generated 16,242 esiRNAs that can be used for RNAi screening in human cells. Comparative analyses with chemically synthesized siRNAs demonstrated a high silencing efficacy of esiRNAs and a 12-fold reduction of downregulated off-target transcripts as detected by microarray analysis. Hence, the presented esiRNA libraries offer an efficient, cost-effective and specific alternative to presently available mammalian RNAi resources.

The discovery that short dsRNAs trigger potent and specific mRNA degradation in mammalian cells through the evolutionarily conserved RNAi pathway[1] has revolutionized functional genomics[2]. For mammalian RNAi experiments, the two most commonly used approaches require chemically synthesized short interfering RNAs (siRNAs)[1] and vector-expressed short hairpin RNAs (shRNAs)[3,4]. Several libraries have recently been generated for the human and mouse genomes with these two technologies[5–8]. To ensure a high knockdown efficacy for each silencing trigger, the most recently released siRNA and shRNA libraries had been designed with algorithms that are based on thermodynamic and sequence-specific properties to predict efficient silencing molecules[9,10]. Although the use of these algorithms has improved the overall silencing efficacy, no prediction algorithm exists yet that can exclude off-target gene silencing, which has been recognized as a major challenge for the reliability of data generated in screens using siRNAs or shRNAs[11–16].

We and others have developed an alternative approach to generate silencing triggers for RNAi in mammalian cells. This technology is based on the enzymatic digestion of long dsRNA by bacterial RNase III or recombinant Dicer for the generation of esiRNA (reviewed in ref. 17).

We generated a first esiRNA library from cDNA clones that represented a random part of a transcript sequence. Although many of the esiRNAs produced from this source proved efficient in RNAi screening[18], we aimed to improve the silencing efficacy and specificity by using the optimal transcript fragment of each gene for esiRNA synthesis. In the present study, we identified the optimal esiRNA target regions for all protein-coding transcripts of the human, mouse and rat genomes *in silico*, and designed gene-specific primers to amplify these fragments from cDNA. Using this resource we generated an esiRNA library for the human genome and characterized several important aspects of these esiRNAs for RNAi screening, including the silencing effectiveness and the potential for off-target effects.

RESULTS

In silico prediction of optimized esiRNAs

Previously, we used cDNA clones as the DNA template with universal primers annealing to the vector backbone to amplify the cDNA insert[18]. Although this approach is convenient and effective, it has several disadvantages. Using the software program DEQOR[19] we observed that predicted efficient siRNAs are unequally distributed in most transcript sequences (data not shown), and that in most cases cloned cDNA fragments, which

[1]Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, D-01307 Dresden, Germany. [2]Scionics Computer Innovation, GmbH, Tatzberg 47-51, D-01307 Dresden, Germany. [3]Technology Development Studio, Pfotenhauerstrasse 108, D-01307 Dresden, Germany. [4]RZPD-Ressourcenzentrum für Genomforschung, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany. [5]Cenix BioScience GmbH, Tatzberg 47, Dresden 01307, Germany. [6]Rosetta Inpharmatics LLC, 12040 115th Avenue NE, Kirkland, Washington 98034, USA. [7]Present address: Department of Human Genetics, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to B.H. (habermann@mpi-cbg.de) or F.B. (buchholz@mpi-cbg.de).

RECEIVED 11 SEPTEMBER 2006; ACCEPTED 26 JANUARY 2007; PUBLISHED ONLINE 11 MARCH 2007; DOI:10.1038/NMETH1025

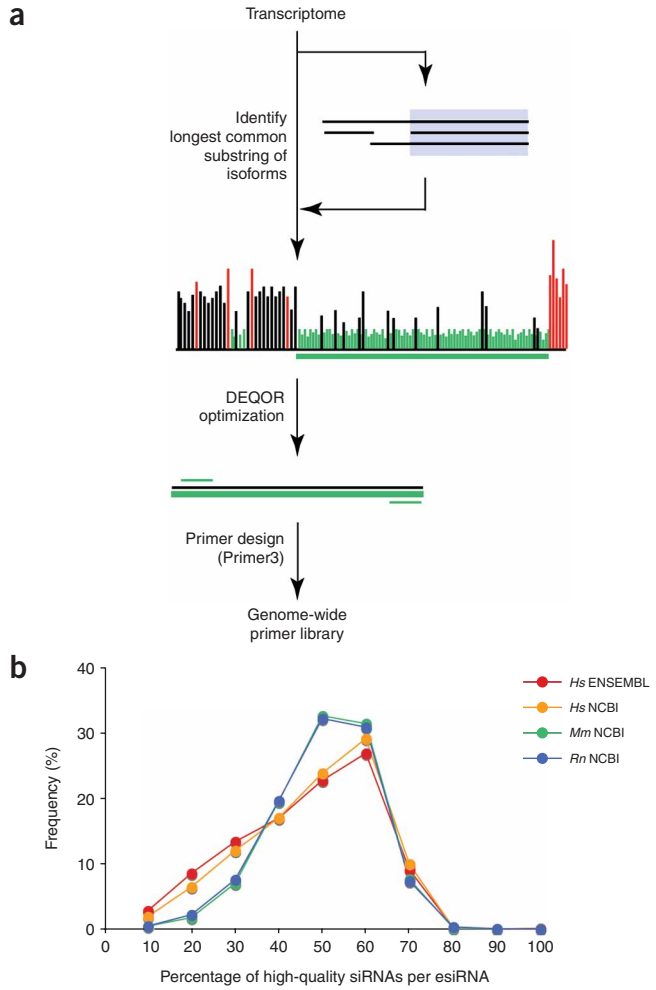

Figure 1 | *In silico* prediction of esiRNA template sequences. (**a**) *In silico* pipeline for optimized esiRNA template prediction. (**b**) Analysis of optimized esiRNAs per transcript or transcript variants for the human, mouse and rat genomes. The percentage of high-quality siRNAs per esiRNAs (DEQOR score ≤ 5) were counted for each esiRNA and plotted against the percentage of sequences. All four genomes show a nearly normal distribution and peak between 40% and 60% high-quality siRNAs per esiRNA. *Hs*, *Homo sapiens*; *Mm*, *Mus musculus*; *Rn*, *Rattus norvegicus*; ENSEMBL, human data set based on the ENSEMBL transcriptome prediction; NCBI, data sets based on the NCBI transcriptome predictions.

represent a random portion of a gene transcript, did not contain the best target region. In principle, the use of a cDNA fragment with the highest proportion of efficient siRNAs for esiRNA production should result in a better silencing potency than a fragment from the same gene that contains less potent siRNAs. For this reason, we developed a ready-to-use pipeline for predicting optimized esiRNA template regions of an entire transcriptome of any organism of choice (**Fig. 1a**). For each transcript, the pipeline identifies the optimal region for esiRNA-based gene-silencing according to state-of-the-art parameters of siRNA selection as implemented in the DEQOR software[19] and performs a primer prediction. For genes that encode multiple transcripts, representing different splice variants, the program automatically identifies the longest common region, which is subsequently used for DEQOR analysis. We applied this pipeline to human, mouse and rat genomes, and constructed genome-wide virtual esiRNA libraries for these three organisms. To obtain an overview of the *in silico* esiRNA prediction for all transcripts, we analyzed the percentage of high-quality siRNAs contained within each fragment (**Fig. 1b**). This analysis revealed a nearly normal distribution, with few fragments showing a very low or very high percentage of efficient siRNAs.

To make the esiRNA data sets available, we created the web-based database RiDDLE (**Fig. 2**). Primer pairs and the sequence of the selected esiRNA target region, as well as detailed information for each gene are readily accessible to the user. Thus, this database permits the retrieval of esiRNA primers for entire genomes, but also for single-gene experiments, gene families or orthologous genes in different species.

Generation of a genome-wide human esiRNA library

The *in silico* analysis of the human transcriptome delivered 19,565 primer pairs designed to amplify optimized template regions for esiRNA synthesis for most protein-coding genes (**Fig. 3a**). To perform the PCR reactions, we used a normalized cDNA library generated from mRNAs of 32 human tissues and 34 cell lines as the PCR template. Then we appended the primers with two different T7 promoter sequences that varied by a single 3′ nucleotide (guanine or cytosine) to obtain directionality in the resulting PCR product, which is helpful, for instance, for sequencing or subsequent cloning.

We obtained 16,287 PCR products (82%) that were of the predicted size; 621 products differed in size from the predicted amplicon length, and 556 products contained multiple PCR fragments (**Fig. 3b**). Notably, sequencing of 35 wrong-size PCR products revealed for 18 of the analyzed samples (51%) the amplification of a transcript of the desired gene that was alternatively spliced to the predicted fragment, whereas the remaining 17 samples represented unspecific products. Likewise, sequencing of 41 multiple-fragment PCR products identified in 20 cases (49%) a previously uncharacterized alternatively spliced transcript of the same gene in addition to the desired product. Hence, 50% of these esiRNAs also targeted the intended transcript.

We *in vitro* transcribed single-band PCR products of the predicted size and annealed them to generate long dsRNAs. Then we digested these dsRNAs with RNase III tagged with glutathione *S*-transferase to 18–25 bp, short dsRNAs and purified them. In total, we generated 16,242 high-quality esiRNAs with a yield of at least 20 μg that we adjusted to a normalized concentration and rearrayed in 384-well plates for large-scale knockdown studies.

Silencing efficacy of optimized esiRNAs

The *in silico* comparison of a random set of 3,173 optimized esiRNAs with their corresponding clone-based esiRNAs indicated a marked increase of the predicted silencing efficacy (**Fig. 4a**). DEQOR selection enhanced the average percentage of high-quality siRNAs per esiRNA from 35% for the clone-based to 43% for the DEQOR-optimized esiRNAs. To verify this improvement experimentally, we selected DEQOR-optimized esiRNAs for 15 genes and compared their knockdown efficacy to the clone-based esiRNA of the respective genes. For that purpose, we quantified the mRNA knockdown levels by quantitative reverse-transcriptase PCR (qRT-PCR) 36 h after transfection (**Fig. 4b**). Fourteen optimized esiRNAs

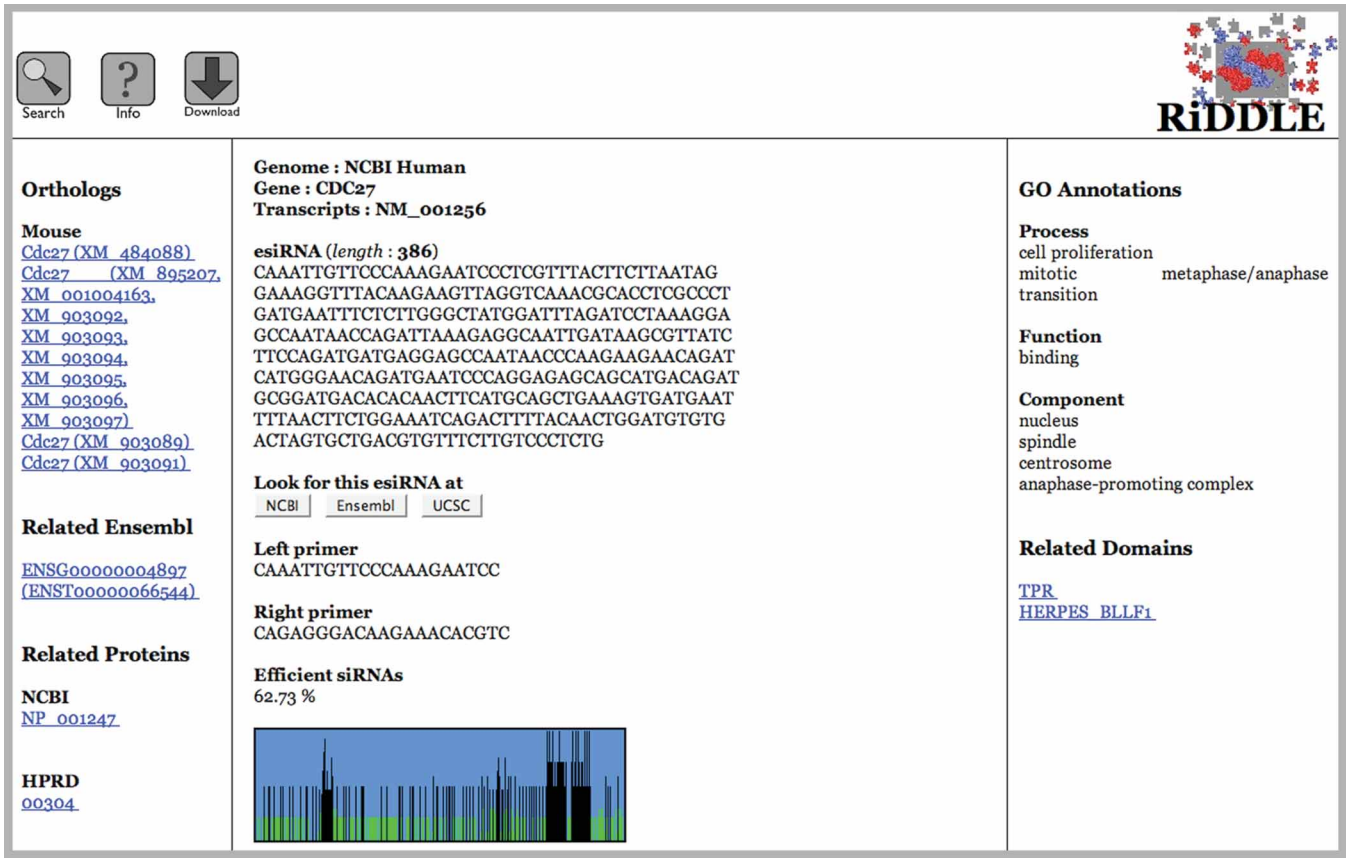

Figure 2 | Gene-centric information in the RiDDLE database. The gene *CDC27* is shown as an example. Gene-based information displayed includes the predicted esiRNA sequence, predicted left and right primer for target amplification and links to BLAST searches at NCBI, ENSEMBL and UCSC genome browser (center). Orthologous sequences in mouse and rat, if available, are shown and linked to the respective entries in the mouse and rat RiDDLE libraries (left). Gene Ontology (GO) annotations[29] are shown, as well as conserved domains found in the respective protein sequence, which are linked to the conserved domain database[30] (right).

and their corresponding clone-based esiRNAs caused a reduction in the amount of mRNA of the targeted transcript, with one qRT-PCR yielding inconclusive results (data not shown). The optimized esiRNAs showed an average transcript knockdown of 78.9%, whereas the clone-based esiRNAs resulted in an average knockdown of 67.9%. Seven (50%) optimized esiRNAs exhibited a significantly better knockdown efficacy than their corresponding clone-based esiRNAs. The seven clone-based esiRNAs, which were less efficient than the corresponding optimized esiRNAs, also exhibited a lower predicted silencing efficacy.

Next we compared the silencing efficacy of optimized esiRNAs and chemically synthesized siRNAs. For that purpose, we designed and synthesized 96 esiRNAs targeting 49 genes (typically two esiRNAs per gene) and obtained an equal number of corresponding siRNAs from a commercial supplier. The qRT-PCR analysis of knockdown 36 h after transfection (**Fig. 4c**) revealed overall a similar silencing efficacy for esiRNA and siRNA. We found that 76% of the esiRNAs and 71% of the siRNAs exhibited a knockdown efficacy of >70%. Notably, the mean knockdown efficacy of the two tested constructs per each transcript between the siRNAs and esiRNAs was highly correlated (**Supplementary Fig. 1** online), suggesting that the maximum knockdown of a gene that can be achieved by RNAi might be transcript-dependent.

Silencing specificity of optimized esiRNAs

To test the specificity of esiRNAs, we profiled transcript regulation by three siRNAs and three esiRNAs targeting each of two transcripts, *MAPK14* and *SMEK2* (also known as *KIAA1387*). We transfected the esiRNAs and siRNAs into HeLa cells and the colon-cancer cell line HCT116, and analyzed gene expression signatures by microarray profiling 12 and 24 h after transfection. We examined regulated signature transcripts for each siRNA and esiRNA to allow identification of all transcripts whose regulation was significantly different ($P < 0.01$) from the reference (mock-transfected RNA). In total, we analyzed 5–7 independent experiments for each interfering RNA (**Fig. 5a**). This analysis detected changes at the transcript level that were silenced by an average of 1.4-fold, demonstrating the sensitivity of this approach.

There was a marked difference in the number of transcripts consistently regulated by siRNAs and esiRNAs (**Fig. 5b**). siRNA signatures showed consistent downregulation of multiple transcripts (median = 25.5) across cell lines and time points. In every case, the downregulated signature was enriched for transcripts with 3′ untranslated region (UTR) sequence complementarity to the seed region hexamers of the siRNA guide strand (**Fig. 5c**), as has been shown previously for regulation of transcripts by siRNAs and miRNAs[14–16]. Thus, siRNA transcript silencing is sequence-specific rather than target-specific. In contrast, few transcripts other than

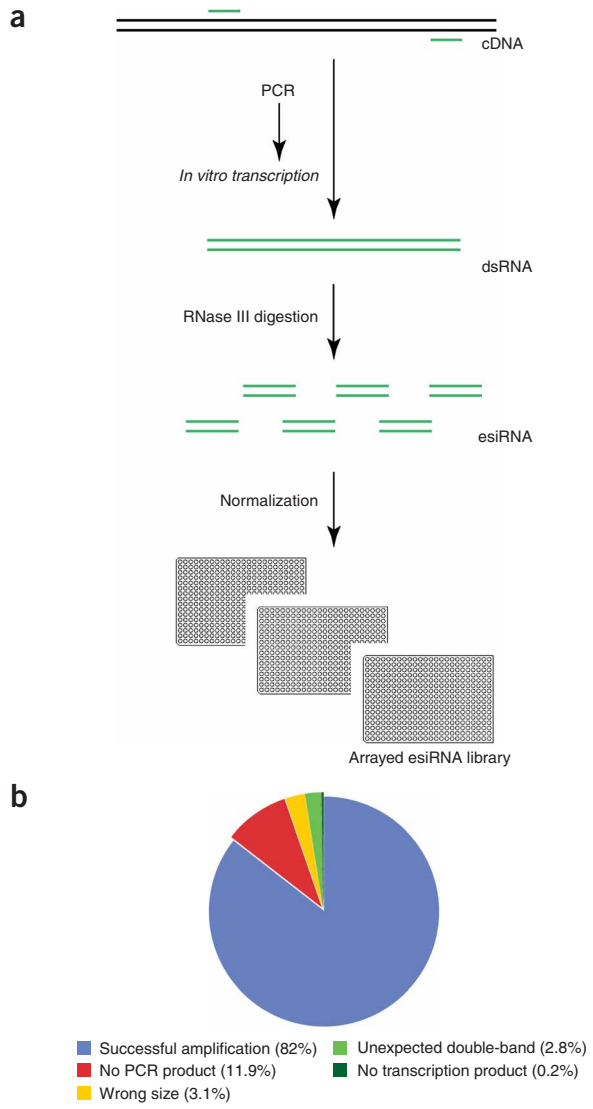

Figure 3 | Generation of a human genome-wide esiRNA library. (**a**) Flow-chart of esiRNA synthesis. The main steps of esiRNA synthesis are shown. A detailed protocol of this procedure is described in reference 23. (**b**) Statistics of library generation. Percentages of single-band PCR products with the expected size (successful amplification), complete PCR failures (no PCR product), unexpected products (wrong size, unexpected double-band) and *in vitro* transcription failures (no transcription product).

the intended target were downregulated by esiRNAs (median = 2). Neither siRNAs nor esiRNAs induced the expression of genes implicated in the interferon response pathway in the HeLa and HCT116 cells used in this study (**Supplementary Fig. 2** online). Notably, the esiRNAs designed to contain the predicted highest percentage of best silencers had the smallest signatures ($n = 0$), whereas the least specific esiRNAs had the lowest percentage of best silencers (**Fig. 5d**), indicating that the DEQOR score may be useful to estimate off-target effects.

To substantiate the improved specificity observed with esiRNAs, we targeted four additional genes with 2–3 siRNAs and 2 esiRNAs each, in duplicate. As before, there was a marked difference in the number of transcripts silenced by siRNAs and esiRNAs for these four genes. siRNAs silenced a median of 162 transcripts, whereas esiRNAs silenced a median of 12.5 transcripts (**Supplementary Fig. 3** online). Thus, the esiRNAs showed a 13-fold reduction in the number of unintended transcripts silenced in this comparison. Therefore, although one might assume that siRNA off-target effects are additive, the use of highly complex pools of siRNAs to the same target transcript markedly reduced the number of unintended transcripts silenced that are detectable by microarray analysis.

We wondered whether a reduction of off-target effects could also be achieved with low-complexity pools of siRNAs. For this purpose we profiled pools of increasing complexity of active siRNAs to *MAPK14*. We began with a pool of 3 siRNAs and gradually added more siRNAs up to a total of 12, maintaining the total concentration of siRNA. For comparison, we profiled a single siRNA at the same concentration it is represented in each pool. Reducing the concentration of the single siRNA *MAPK14-192* did not reduce either the number of unintended transcripts silenced or the magnitude of silencing (**Fig. 5e**). The signature of the single siRNA was highly enriched for transcripts with 3′ UTR sequences complementary to the siRNA seed region hexamers, and this enrichment was maintained regardless of the siRNA concentration. In contrast, both the number and magnitude of the off-target signatures were reduced by increasing the pool complexity. Furthermore, analysis of the pool signatures provided evidence of competition among siRNAs, as some siRNA seed regions were more highly represented in the signatures than others. When we added two more siRNAs to create a pool of three members, the seed region of siRNA-1 was less represented in the signature, which was now enriched for transcripts with 3′ UTR sequences complementary to the seed region of siRNA-2, *MAPK14-256*. As the pool size increased, the contribution of siRNA-1 continued to decrease to below significance. When we added more siRNAs to create pools of 9 or 12 members, the signature was now enriched for transcripts with 3′ UTR hexamers complementary to the seed region of siRNA-6, *MAPK14-154*. Thus, although this pool contains 12 siRNAs at equal concentration, the expression signature was dominated by one siRNA. In all cases, silencing of the intended target was maintained, demonstrating that the pools retain on-target efficacy while reducing the extent of off-target signatures. The siRNA pools, however, did not produce the same reduction in off-target silencing as the esiRNAs.

DISCUSSION

We provide here an open resource that allows cost-efficient and rapid synthesis of esiRNAs for nearly all human, mouse and rat genes. The published primer sequences can be used to synthesize single esiRNAs for small-scale experimentation or for the generation of large-scale libraries. For this purpose, we have recently developed a protocol that permits the rapid large-scale synthesis of esiRNAs in a standard-sized laboratory[20]. To facilitate the use of large esiRNA libraries, the human PCR products have been made available at the nonprofit German Genome Resource Center, and we also intend to make ready-to-transfect esiRNA libraries available through this resource provider. To date, we have completed an optimized esiRNA library for most human genes, and will generate in the near future a similar library for mouse and rat genomes. The ready-to-use pipeline we have established for the prediction of optimized esiRNA regions of an entire transcriptome furthermore permits the rapid generation of genome-wide libraries for any other organism of choice.

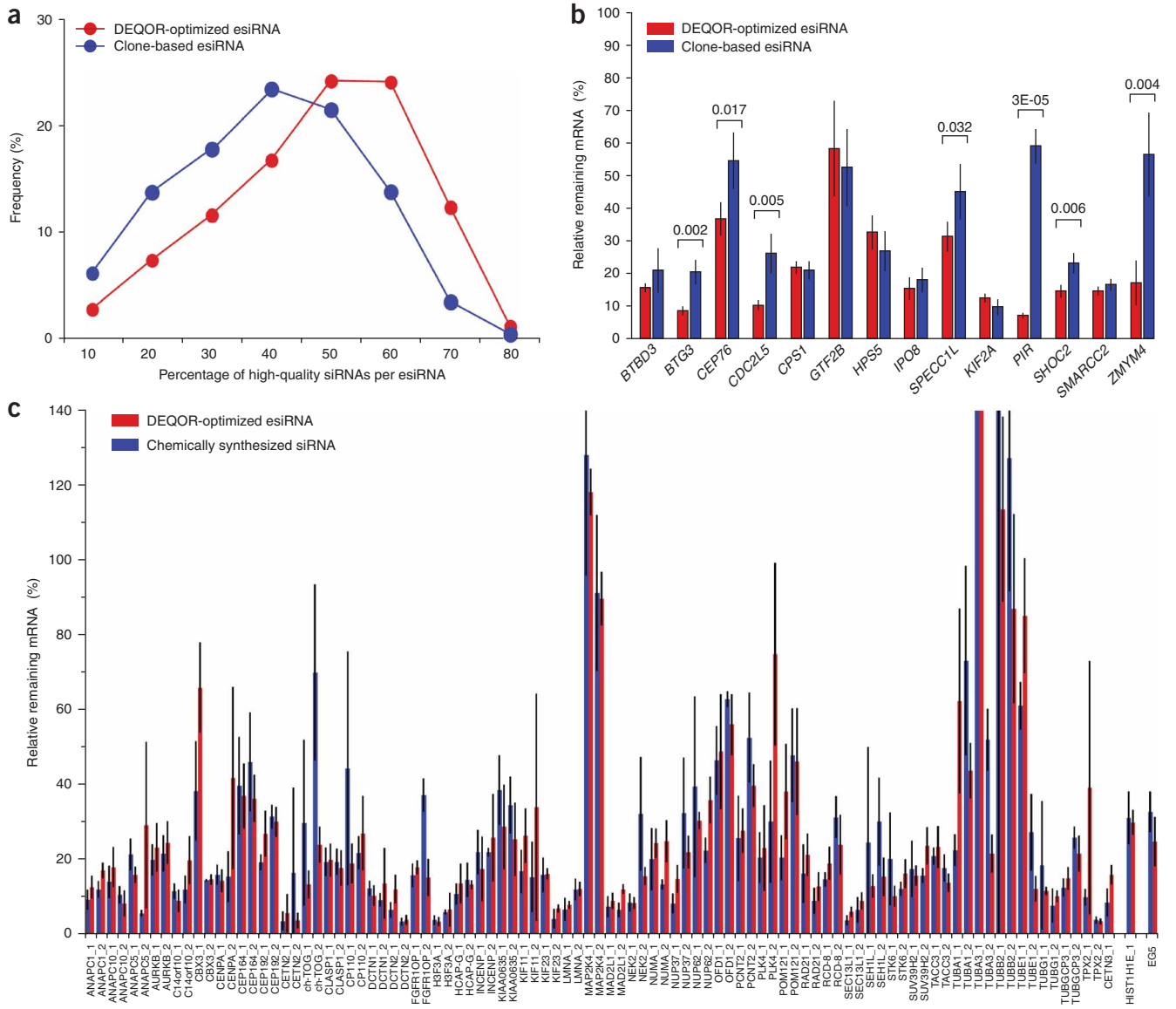

Figure 4 | Silencing efficacy of esiRNAs. (**a**) *In silico* comparison of predicted silencing efficacy of clone-based and DEQOR-optimized esiRNAs. Comparisons were performed for a random set of 3,173 genes by counting the number of high-quality siRNAs (DEQOR score ≤ 5) per esiRNA. The percentage of high-quality siRNAs per esiRNA was plotted against the sequence frequency. Note the shift to a higher percentage of high-quality siRNAs for the DEQOR-optimized esiRNAs. (**b**) Silencing efficacy of 14 optimized and clone-based esiRNAs measured by quantitative RT-PCR. Student's *t*-test *P* values are indicated above the bars for significant differences ($P \leq 0.05$). (**c**) Silencing efficacy of 96 optimized esiRNAs and chemically synthesized siRNAs measured by qRT-PCR. For **b** and **c**, HeLa cells were collected for RNA extraction 36 h after transfection. Expression levels were normalized against a negative control. Error bars indicate s.d., $n = 2$.

The approach we used for the generation of the optimized human esiRNA library allowed substantial improvements compared to the first clone-based library. As a key feature, the identification of common sequences for multiple transcripts of the same gene and the implementation of siRNA design criteria for the synthesis of optimized esiRNAs increased the efficacy of gene silencing. Additionally, the use of a well-defined size range of the PCR products allowed the streamlining of long dsRNA synthesis and digestion procedures, thereby reducing the number of drop-outs during esiRNA synthesis and doubling the yield of esiRNA obtained in the production process. We demonstrated in the present study that the use of normalized cDNA libraries as starting material for high-throughput generation of PCR products for the synthesis of long dsRNA is technically sound, as we obtained with one primer pair for each transcript 82% of the products we aimed to amplify.

All esiRNAs present in the reported library must target an existing polyadenylated transcript because their synthesis from cDNA relied on the presence of such a transcript in human cells. In contrast, siRNA and shRNA libraries are solely based on transcript sequences in databases, of which many are only *in silico* predicted and not experimentally verified. In this context, some of the primer pairs that did not produce a PCR fragment may reflect errors in gene predictions rather than PCR failure. Likewise, we

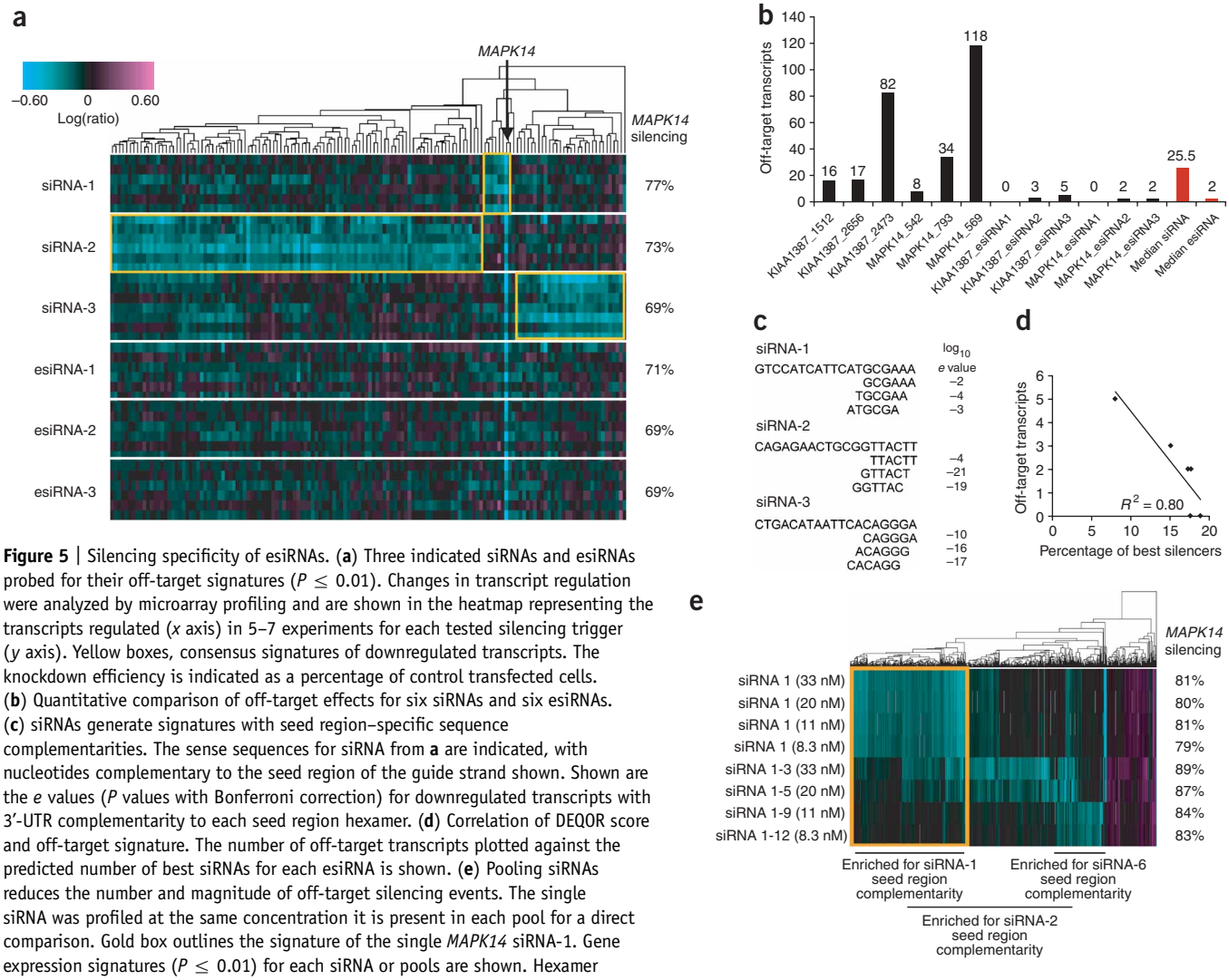

Figure 5 | Silencing specificity of esiRNAs. (**a**) Three indicated siRNAs and esiRNAs probed for their off-target signatures ($P \leq 0.01$). Changes in transcript regulation were analyzed by microarray profiling and are shown in the heatmap representing the transcripts regulated (*x* axis) in 5–7 experiments for each tested silencing trigger (*y* axis). Yellow boxes, consensus signatures of downregulated transcripts. The knockdown efficiency is indicated as a percentage of control transfected cells. (**b**) Quantitative comparison of off-target effects for six siRNAs and six esiRNAs. (**c**) siRNAs generate signatures with seed region–specific sequence complementarities. The sense sequences for siRNA from **a** are indicated, with nucleotides complementary to the seed region of the guide strand shown. Shown are the *e* values (*P* values with Bonferroni correction) for downregulated transcripts with 3′-UTR complementarity to each seed region hexamer. (**d**) Correlation of DEQOR score and off-target signature. The number of off-target transcripts plotted against the predicted number of best siRNAs for each esiRNA is shown. (**e**) Pooling siRNAs reduces the number and magnitude of off-target silencing events. The single siRNA was profiled at the same concentration it is present in each pool for a direct comparison. Gold box outlines the signature of the single *MAPK14* siRNA-1. Gene expression signatures ($P \leq 0.01$) for each siRNA or pools are shown. Hexamer enrichment was performed as in **a**. Heatmap as in **a**.

verified for 50% of the analyzed problematic PCR products (that is, multiple products and products of the wrong size) the correct gene identity, but found previously unknown exons or splice variants as the cause for unexpected fragments. To achieve full-genome coverage of the human library, we will use specific cDNA clones containing the target sequence as template to generate the missing PCR products. For the amplicons that cannot be obtained by this approach we will design an alternative primer pair.

To achieve a high saturation rate in genome-wide RNAi screens, a high silencing efficacy is desirable for each individual silencing trigger in the RNAi library. We demonstrated that a set of 96 optimized esiRNAs had a comparable silencing efficacy to a corresponding set of chemically synthesized siRNA that were designed with state-of-the-art prediction algorithms. This finding indicates that optimized esiRNAs have a similar silencing efficacy as predicted efficient siRNAs. Notably, there was a high correlation between the knockdown of siRNAs and esiRNAs for the same targeted transcripts. This suggests that the maximum knockdown effect that can be achieved with an efficient RNAi reagent is transcript-dependent. The underlying mechanism for this observation remains unclear, but previous studies have suggested a role of the secondary structure of the targeted transcript or proteins binding to specific mRNAs[21,22].

For large-scale RNAi studies the silencing specificity of the RNAi agent has become an important issue, because off-target effects can dramatically increase the false positive rate of a screen[14]. Recently, off-target effects, originating from perfect cross-matches[23], or cytosine-adenine-any base (CAN) repeats[24], have been identified to cause false positives in genome-wide RNAi screens in *Drosophila melanogaster* cells. The software program DEQOR minimizes the number of perfect cross-matches and repetitive elements in esiRNAs, because these sequences are given high penalty scores. Hence, our pipeline preferably selected regions that lack these sequences. Where cross-silencers could not be avoided in the esiRNA design process, the siRNAs with a perfect match to a different target are marked. Furthermore, esiRNAs with six or more CAN repeats are flagged in the RiDDLE database.

When comparing the global gene expression signatures of fourteen esiRNAs with fifteen siRNAs, we observed a marked reduction of off-target gene silencing for esiRNAs. A similar effect has been reported recently for dicer-prepared esiRNAs[25]. The increase in silencing specificity is likely achieved by the very low concentration

of individual siRNAs in a heterogeneous mixture, and/or by a competition of the different silencing triggers. More experiments are required to dissect these possibilities and clarify this process.

In this study we showed that a reduction of off-target effects can be gradually achieved by using pools with increasing numbers of chemically synthesized siRNAs. However, we found that even a pool of 12 siRNAs showed a larger off-target signature size than any of the esiRNAs. This indicates that a considerably larger number of individual siRNAs would have to be pooled to obtain a similar silencing specificity as obtained with esiRNAs.

The microarray experiments used to assess off-target effects in this study measure effects only at the transcript level. Because large changes at the mRNA level are more likely to achieve statistical significance in microarray studies than small changes, we might have missed some small changes at the transcript level. Off-target silencing causing small changes at the transcript level could potentially have large effects on the amounts of protein and could ultimately lead to a biological effect. If this were the case, then esiRNAs could produce more false phenotypes, despite the lower off-target signature detected using microarrays, owing to the large pool of siRNAs used. Although a first screen performed with the human esiRNA library produced a low rate of false positives (unpublished data), it remains to be seen whether a reduction of strong off-target silencing at the mRNA level with complex pools of siRNAs will generally translate into fewer false positives in RNAi screens.

METHODS

esiRNA library design. esiRNAs for human protein–coding genes of the ENSEMBL database were based on the December 2004 (V27) release; data sets for human, mouse and rat from the US National Center for Biotechnology Information (NCBI) were based on the April 2006 release. To exclude gene-prediction artifacts, we removed human ENSEMBL genes without expressed sequence tags and no homolog in mouse from the data set. We only considered sequences with more than 350 base pairs for esiRNA production. We identified longest common regions among splice variants of a gene with a modified version of suffix arrays[26] (see **Supplementary Methods** online). The resulting sequences of multiple isoforms, as well as full-length sequences of single-transcript genes were submitted to a stand-alone version of DEQOR[19]. We selected high-quality esiRNA regions of 400–600 bp based on the following criteria: (i) the region selected was within 3,000 bp of the protein-coding sequence upstream of the 3′ UTR, avoiding difficulties in cDNA amplification, and within 500 bp of the 3′ UTR, avoiding alternative polyadenylated transcripts; (ii) the region selected did not contain more than 5% siRNAs that match a transcript sequence of a different gene; (iii) the region selected had the highest percentage of high-quality siRNAs according to their DEQOR score (score ≤ 5). Selected regions were submitted to a stand-alone version of Primer3 (ref. 27) with default settings and a product size range of 400–600 bp. We step-wise reduced the length of regions that did not result in high-quality primers. We did not consider primer fragments of less than 220 bp for further analysis. For statistics on high-quality siRNAs per esiRNA and off-targets, we considered only the sequence fragments resulting from primer selection.

esiRNA synthesis. We generated the template for *in vitro* transcription by a two-step PCR approach (see **Supplementary Methods**) and synthesized esiRNAs in 96-well plates as previously described[20]. Primer sequences of the esiRNA templates used for the experimental analyses of silencing efficacy and specificity are available in **Supplementary Table 1** online.

siRNAs. We obtained the 96 siRNAs for the analysis of silencing efficacy from Ambion. All siRNAs for expression profiling were designed by Rosetta using an algorithm developed to increase silencing efficiency and minimize off-target silencing[11,15]. The Rosetta siRNA design algorithm is available online. siRNAs for specificity testing were synthesized by Sigma-Proligo. The siRNA sequences are available in **Supplementary Table 2** online.

Transfection. We added 30 nM siRNA or an equivalent amount of esiRNA in 15 µl OptiMEM (Invitrogen) to a mixture of 0.4 µl Oligofectamine (Invitrogen) and 4.6 µl OptiMEM in a 96-well tissue culture plate (Nunc). After 20 min of incubation at 18–23 °C we seeded 9,000 HeLa cells in 120 µl medium (DMEM, 11.5% FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin; Invitrogen) onto the transfection mix using a WellMate dispensing system (Matrix). We performed the transfections in duplicate (for the esiRNA and siRNA comparison) or triplicate (for the clone-based and DEQOR-optimized esiRNA comparison). For microarray analyses, we transfected cells in 6-well plates using Oligofectamine (HeLa) or Lipofectamine 2000 (HCT116) and 10 nM siRNA duplex or equivalent amount of esiRNA.

Quantitative RT-PCR. We lysed HeLa cells transfected in 96-well plates 36 h after transfection. We extracted total RNA using 96-well plate Invisorb kits (Invitek). cDNA was synthesized with High-Capacity cDNA reagent (ABI). We performed real-time qRT-PCR using Quantace SybrGreen qPCR mix (Quantance; see **Supplementary Table 3** online for primer sequences). We performed qRT-PCR on an ABI 7900HT machine, in duplicate for every well or cDNA, for both the target gene and for the 18*S* rRNA. We calculated relative remaining mRNA for target versus 18*S* rRNA levels, comparing samples versus two plate-specific controls (Ambion scrambled siRNA or an esiRNA targeting renilla luciferase). We performed quality control by evaluating melting curves; we excluded individual failed data points.

Microarray analysis. We hybridized RNA from siRNA-transfected cells or esiRNA-transfected cells against RNA from mock-transfected cells. We purified total RNA with the RNeasy kit (Qiagen) and processed it as described previously[28]. We performed ratio hybridizations with fluorescent label reversal to eliminate dye bias. Error models have been described previously[28]. Data shown are signature genes that display a difference in expression level ($P < 0.01$) relative to mock-transfected cells. No restrictions were placed on fold change in expression. We analyzed the data using Rosetta Resolver. We defined off-target signatures as transcripts showing significant regulation ($P \leq 0.01$) in both replicate transfections of the same RNAi agent and not significant in $>50\%$ of transfections targeting the gene.

Additional methods. Descriptions of the identification of the longest common substrings, the two-step PCR approach for the generation of templates for esiRNA production and the hexamer complementarity are available in **Supplementary Methods**.

URLs. RiDDLE database, http://cluster-12.mpi-cbg.de/cgi-bin/riddle/search; German Genome Resource Center, http://www.rzpd.de/products/esirna; Rosetta siRNA design algorithm, http://www.sigmaaldrich.com/rnai (pre-designed siRNAs).

Accession codes. Gene Expression Omnibus (GEO): GSE6807.

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

ACKNOWLEDGMENTS

We thank Eberhard Krausz (Technology Development Studio, Max Planck Institute for Molecular Cell Biology and Genetics) for providing robotics support, and the Rosetta Gene Expression Laboratory for microarray hybridizations. This work was supported by the EU grants "FunGenES" (LSHG-CT-2003-503494), "Mitocheck" (LSHG-CT-2004-503464), by BMBF grant PTJ-BIO/0313130, the NGFN2 grant SMP-RNAi (01GR0402) and the Max Planck Society.

AUTHOR CONTRIBUTIONS

R.K., A.K.H., M.S., M.T., G.P., K.F. and A.C. generated the human esiRNA library; V.S. and B.H. performed the *in silico* analyses; H.G. and J.W. performed automation; K.K. generated LIMS; E.R. and B.K. generated esiRNAs; C.F. performed QPCRs; C.S. and B.S. performed and analyzed QPCRs; J.G., J.S., J.B. and A.L.J. performed and analyzed microarray studies; P.S.L. analyzed microarray data; R.K. and F.B. designed and analyzed the experiments; R.K., V.S., A.L.J., B.H. and F.B. wrote the paper.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/naturemethods.

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

1. Elbashir, S.M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
2. Hannon, G.J. & Rossi, J.J. Unlocking the potential of the human genome with RNA interference. *Nature* **431**, 371–378 (2004).
3. Brummelkamp, T.R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
4. Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J. & Conklin, D.S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
5. Silva, J.M. *et al.* Second-generation shRNA libraries covering the mouse and human genomes. *Nat. Genet.* **37**, 1281–1288 (2005).
6. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
7. Root, D.E., Hacohen, N., Hahn, W.C., Lander, E.S. & Sabatini, D.M. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat. Methods* **3**, 715–719 (2006).
8. Bernards, R., Brummelkamp, T.R. & Beijersbergen, R.L. shRNA libraries and their use in cancer genetics. *Nat. Methods* **3**, 701–706 (2006).
9. Reynolds, A. *et al.* Rational siRNA design for RNA interference. *Nat. Biotechnol.* **22**, 326–330 (2004).
10. Pei, Y. & Tuschl, T. On the art of identifying effective and specific siRNAs. *Nat. Methods* **3**, 670–676 (2006).
11. Jackson, A.L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635–637 (2003).
12. Jackson, A.L. & Linsley, P.S. Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.* **20**, 521–524 (2004).
13. Scacheri, P.C. *et al.* Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl. Acad. Sci. USA* **101**, 1892–1897 (2004).
14. Lin, X. *et al.* siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res.* **33**, 4527–4535 (2005).
15. Jackson, A.L. *et al.* Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179–1187 (2006).
16. Birmingham, A. *et al.* 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods* **3**, 199–204 (2006).
17. Buchholz, F., Kittler, R., Slabicki, M. & Theis, M. Enzymatically prepared RNAi libraries. *Nat. Methods* **3**, 696–700 (2006).
18. Kittler, R. *et al.* An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature* **432**, 1036–1040 (2004).
19. Henschel, A., Buchholz, F. & Habermann, B. DEQOR: a web-based tool for the design and quality control of siRNAs. *Nucleic Acids Res.* **32**, W113–W120 (2004).
20. Kittler, R., Heninger, A.K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779–784 (2005).
21. Heale, B.S., Soifer, H.S., Bowers, C. & Rossi, J.J. siRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res.* **33**, e30 (2005).
22. Schubert, S., Grunweller, A., Erdmann, V.A. & Kurreck, J. Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. *J. Mol. Biol.* **348**, 883–893 (2005).
23. Kulkarni, M.M. *et al.* Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat. Methods* **3**, 833–838 (2006).
24. Ma, Y., Creanga, A., Lum, L. & Beachy, P.A. Prevalence of off-target effects in *Drosophila* RNA interference screens. *Nature* **443**, 359–363 (2006).
25. Myers, J.W. *et al.* Minimizing off-target effects by using diced siRNAs for RNA interference. *Journal of RNAi and Gene Silencing* **2**, 181–194 (2006).
26. Manber, U. & Myers, G. Suffix arrays: a new method for on-line string searches. *SIAM Journal on Computing* **22**, 935–948 (1993).
27. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
28. Hughes, T.R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19**, 342–347 (2001).
29. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
30. Marchler-Bauer, A. *et al.* CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* **30**, 281–283 (2002).