Published online 31 August 2012

Nucleic Acids Research, 2013, Vol. 41, No. 1 ***e9***
doi:10.1093/nar/gks797

siSPOTR: a tool for designing highly specific and potent siRNAs for human and mouse

Ryan L. Boudreau[1], Ryan M. Spengler[2], Ray H. Hylock[3], Brandyn J. Kusenda[3], Heather A. Davis[3], David A. Eichmann[3] and Beverly L. Davidson[1,2,4,5,*]

[1]Department of Internal Medicine, [2]Program in Molecular & Cellular Biology, [3]Institute for Clinical & Translational Science, [4]Department of Molecular Physiology & Biophysics and [5]Department of Neurology, University of Iowa, Iowa City, IA 52242, USA

Received May 4, 2012; Revised June 30, 2012; Accepted July 30, 2012

ABSTRACT

RNA interference (RNAi) serves as a powerful and widely used gene silencing tool for basic biological research and is being developed as a therapeutic avenue to suppress disease-causing genes. However, the specificity and safety of RNAi strategies remains under scrutiny because small inhibitory RNAs (siRNAs) induce off-target silencing. Currently, the tools available for designing siRNAs are biased toward efficacy as opposed to specificity. Prior work from our laboratory and others' supports the potential to design highly specific siRNAs by limiting the promiscuity of their seed sequences (positions 2–8 of the small RNA), the primary determinant of off-targeting. Here, a bioinformatic approach to predict off-targeting potentials was established using publically available siRNA data from more than 50 microarray experiments. With this, we developed a specificity-focused siRNA design algorithm and accompanying online tool which, upon validation, identifies candidate sequences with minimal off-targeting potentials and potent silencing capacities. This tool offers researchers unique functionality and output compared with currently available siRNA design programs. Furthermore, this approach can greatly improve genome-wide RNAi libraries and, most notably, provides the only broadly applicable means to limit off-targeting from RNAi expression vectors.

INTRODUCTION

RNA interference (RNAi) is mediated by small RNAs (~21 nucleotides), which are loaded into the RNA-induced silencing complex (RISC), generating a functional complex capable of base-pairing with and repressing target transcripts (1). Scientists have devised strategies to co-opt the cellular RNAi machinery to silence virtually any gene of interest using small inhibitory RNAs (siRNAs), which may be chemically synthesized or expressed in the context of stem-loop RNAs [e.g. short-hairpin RNAs (shRNAs)]. RNAi tools are vital for functional genomics studies, which enrich our understanding of basic biological processes. In addition, RNAi-based therapeutics exhibit exciting potential to treat numerous human ailments by suppressing disease-associated genes (2). However, the utility of RNAi is appreciably limited by our ability to design siRNAs which are both potent and specific. There is considerable evidence supporting that siRNAs bind to and regulate unintended mRNAs, an effect known as off-target silencing (3–5). Although most siRNA design algorithms include BLAST to identify off-target transcripts with near-perfect complementarity, off-targeting primarily occurs when the seed region (nucleotides 2–8 of the small RNA) pairs with sequences within 3′-untranslated regions (UTRs) of unintended mRNAs thus inducing translational repression and transcript destabilization, similar to canonical micro-RNA based silencing (6–8). Notably, short stretches of complementarity—as little as 6 bp—may be sufficient to initiate off-target silencing (9) (Figure 1A).

Numerous reports support that seed-based off-targeting generates false positives in RNAi screens and dictates the toxicity potential of siRNAs (10–13). Anderson *et al.* reported that the extent of siRNA off-targeting correlates with the frequency of seed complements (hexamers) present in the 3′-UTRome (Figure 1B) (13). Upon evaluating subsets of siRNAs with differing off-targeting potential (low, medium and high; based on 3′-UTR hexamer distributions), the low subset had significantly diminished microarray off-target signatures and less adverse effects on cell viability as compared with the other subsets. These findings established the importance of considering seed complement hexamer frequencies as a key criterion for

*To whom correspondence should be addressed. Tel: +1 319 353 5573; Fax: +1 319 353 3372; Email: beverly-davidson@uiowa.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

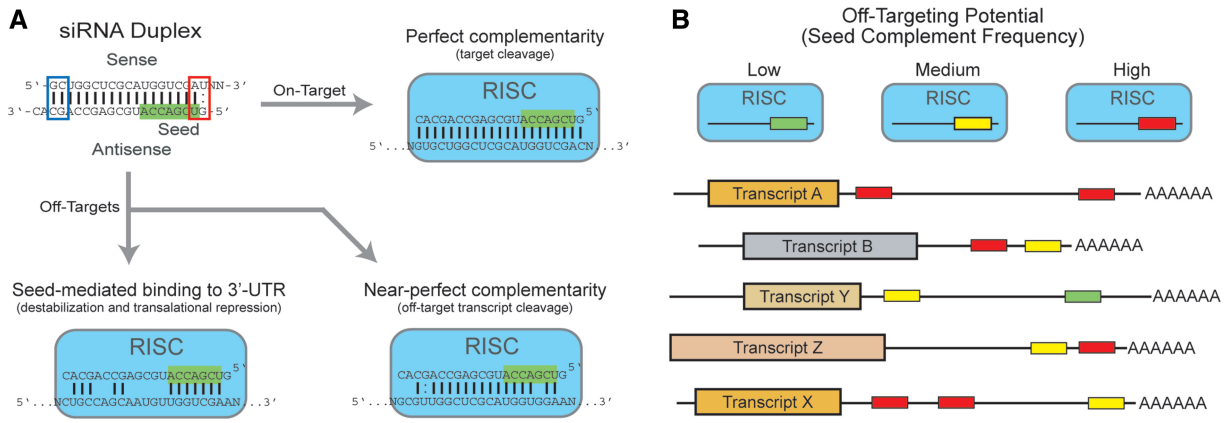

Figure 1. Diagram of on- and off-target silencing by siRNAs. (**A**) Cartoon depicting a siRNA duplex designed to exhibit proper strand-biasing [i.e. strong G-C (blue) and weak A/G-U (red) binding at the respective 5′ and 3′ ends of the sense strand] and contain a low off-targeting potential seed (green highlight). Upon loading into RISC, the antisense strand may direct on-target silencing (intended) and off-target silencing (unintended). (**B**) Schematic highlighting the relationship between the frequencies of seed complement binding sites in the 3′-UTRome and the off-targeting potential for siRNAs.

designing highly specific siRNAs, and some siRNA design algorithms have since incorporated seed-specificity guidelines (14–16). However, these algorithms remain strongly biased for silencing efficacy and because numerous potency-based filters are applied ahead of specificity guidelines, few candidate siRNAs with low off-targeting potential seeds emerge. This is reflected in recent literature and genome-wide RNAi libraries, where only 10% of siRNAs fall into the previously established low off-targeting range, per the Anderson *et al.* study (17,18). Although potency-based design is rational, only a fraction of the functional siRNAs for a given target transcript are predicted, and in many instances, highly functional siRNAs do not satisfy several design rules.

In recent work from our laboratory, we aimed to improve the safety profile of therapeutic RNAi by designing hairpin-based vectors containing siRNAs with low off-targeting potentials (17). We implemented a design scheme which focuses on seed specificity yet promotes efficacy. This approach proved successful in identifying therapeutic sequences which effectively silence target gene expression, induce minimal off-targeting and are well-tolerated in mouse and non-human primate brains (17,19). These promising results prompted us to extend the utility of this approach by developing a user-friendly tool to facilitate with the selection of low off-targeting potential siRNAs for broader application in therapeutic development and basic biological research. Here, we describe a specificity biased design algorithm which employs an improved means to score off-targeting potentials, and demonstrate its effectiveness and unique functionality in comparison with current publically available tools.

MATERIALS AND METHODS

Dataset and sequence retrieval

Pre-processed microarray datasets, annotations and sequences were obtained from previously published supplementary materials (20). This represents a compilation of microarray data from seven earlier reports describing gene expression changes in siRNA- or miRNA-treated HeLa cells. The relevant datasets, array IDs and corresponding sequence information are reproduced in Supplementary Table S1.

TargetScan 6.0 was used to determine the frequencies of seed complement binding sites (e.g. 6-mer, 7A1, 7m8 and 8-mer) for all possible 16 376 heptamers (corresponding to positions 2–8 of the small RNA) for each RefSeq 3′-UTR sequence (20). Human (GRCh37/hg19) and mouse (NCBI37/mm9) 3′-UTR sequences, and corresponding gene symbols and accession numbers were obtained from the UCSC Table Browser (http://genome.ucsc.edu/) using RefSeq annotations (21–25).

Formulating potential off-targeting score

Dataset selection

Expression data for endogenous micro-RNAs were excluded from the training and validation sets; several publications have suggested avoiding these seed sequences in RNAi sequence design (20,26). The GSE5814 dataset was also excluded, because 77 of the experiments tested siRNAs with the same seed sequence. Strand-biasing analyses were performed to determine whether sense or antisense strands induce detectable off-targeting in each experiment. Pairwise *t*-tests were performed comparing genes with at least one 7mer site ($\geq$1 8mer, 7M8 or 7A1) for either sense or antisense strand seed sequence, to those having no predicted 3′-UTR target site, including 6mer sites. Experiments exhibiting highly significant repression mediated by the sense strand (one-tailed; $P \leq 6E{-}5$), and little to no evidence for the antisense ($P > 0.05$) were removed from further analyses. Of the remaining studies, the Dharmacon2008 dataset qualitatively showed the most diversity in seed off-targeting potential, and it was set aside for downstream validation.

Establishing weighted probability of repression values and potential off-targeting score calculation

Following the dataset filtering described above, 53 microarray datasets from three independent studies

(Dharmacon2006, GSE5291 and GSE5769) were used as training data to establish potential off-targeting score (POTS). For each microarray dataset, transcripts with a single predicted 3′-UTR seed-binding site for either the sense or antisense strand of the given siRNA were considered. This was done to account for possible loading of the sense strand which may also mediate off-targeting. Transcripts with multiple target sites (8mer, 7M8, 7A1 or 6mer) for either strand were ignored so that the silencing potential for single sites for each site type could be determined. Background data for each microarray consisted of the remaining transcripts with no predicted 3′-UTR seed-binding sites for either siRNA strand. Transcripts containing seed-binding sites were parsed into groups based on seed site type, and cumulative distributions of gene expression values were generated for each transcript set.

Probability of repression (PR) values was calculated as a measure of the increased PR imparted by the presence of the single seed-binding sites, relative to background expectations. Statistical analyses were first performed on the datasets collectively to identify the log 2 fold-change value corresponding to the most significant divergence of repressive potentials across all site types. For this, these data were analyzed at discrete intervals (0.05 log 2 fold-change increments), comparing the mean differences in cumulative fractions (paired-samples *t*-test) for each site type set relative to the respective background values across all experiments. Fisher's method was used to summarize *P* values at each interval. The most significant interval (-0.3 log 2 $X^2 = 176.4$; $df = 8$; $P < 6E-34$) was used calculate PR values where,

$$\mathrm{PR}_{\text{site type}} = P(X_{\text{site type}_i} \leq -0.3) - P(X_{\text{nosite}_i} \leq -0.3).$$

These PR values were multiplied by seed-binding site frequencies (*N*) for each site type in the 3′-UTRome and summed to compute a weighted POTS using the following equation:

$$\mathrm{POTS} = N_{8\mathrm{mer}}\mathrm{PR}_{8\mathrm{mer}} + N_{7M8}\mathrm{PR}_{7M8} + N_{7A1}\mathrm{PR}_{7A1} + N_{6\mathrm{mer}}\mathrm{PR}_{6\mathrm{mer}}$$

To generate the final POTS used in the siSPOTR tool, PR values were calculated for both the validation and training datasets, and the median values served as the final PR value. Also, 8mer, 7M8, 7A1 and 6mer site counts for all 16 384 heptamers were calculated from Targetscan 6.0 (20) predictions based on human and mouse RefSeq-annotated 3′-UTRs.

Tissue-specific POTS analysis

Expression profiles from 177 human cell lines and tissues based on the U133A/GNF1H gene atlas were obtained from the BioGPS FTP site (http://biogps.org) (27,28). For each dataset, genes with median expression values of >100 for their corresponding probe sets were considered to be expressed. A tissue-specific POTS (tsPOTS) was calculated for each tissue, as described earlier, but limiting the 3′-UTRs to expressed genes when calculating site type frequencies. Spearman correlations were performed to evaluate variability in the rank-order of seed sequences by tsPOTS, as compared with POTS calculated based on all human 3′-UTRs.

Validating siSPOTR

Efficacy

The 2431 siRNAs in the Huesken Dataset were stepwise filtered according to the siSPOTR design scheme (i.e. strand-biasing, GC-content and POTS rank). For a comparison of efficacy, we used siDesign Center (Dharmacon), a highly utilized siRNA design tool which focuses primarily on potency. Target gene coding sequences were obtained using the Genbank Accessions provided in the Huesken siRNA Dataset and were used as input sequences into the siDesign Center tool for siRNA design using default settings. The top 10 hits by siDesign Center were considered the top candidates and were intersected with the Huesken siRNA dataset. Gene silencing efficacies for overlapping siRNAs were recorded and plotted.

Ranking off-targeting potential

To evaluate the ability of the PR values to estimate the relative extent of off-targeting, POTS values were calculated for the validation set (Supplementary Table S1), using the median value for each site type determined from the training set. Target site frequencies were calculated as described earlier, using human RefSeq 3′-UTR sequences for transcripts present on the array. POTS values were determined as the sum-product of the 8mer, 7M8, 7A1 and 6mer site frequencies and their respective PR values.

Cumulative distribution plots for gene expression values were generated by parsing the transcripts by site type with no limitation for transcripts with single sites. The number of down-regulated transcripts over background was calculated as described earlier, subtracting the background fraction at the same point. Seeds were ranked according to these values, and were compared with the rank-order of their estimated POTS values, using spearman rank correlations. Visual inspection of the correlation plot showed seven qualitatively distinct outliers in the right tail of the POTS distribution (red dots, Figure 5D). Spearman's rank correlation coefficients and *P* values were calculated with and without these samples included.

Suppression signatures

Microarray data for the validation datasets was processed on a per target gene basis (i.e. GAPDH, PPIB, and No Target groups) to discern off-targeting from gene expression changes resulting from on-target silencing. The microarray data for each group were evaluated to identify genes that were down-regulated by more than three standard deviations from the mean, across the datasets, for a given gene. These gene lists and accompanying gene expression values were imported into Partek Genomics Suite (Partek GS, St. Louis, MO) and used to perform hierarchical clustering by row (columns were ordered by increasing POTS) allowing visualization of the suppression signatures by heatmaps. Heatmaps

were partitioned to separate low POTS and high POTS siRNAs for each group. A qualitative assessment of suppression signature size was defined by the area of the broadest, dark blue regions for each lane and plotted on a common *x*-axis.

siRNA design tool comparison

We obtained RefSeq coding sequences for the 16 therapeutically relevant gene targets (Table 1). These sequences were used as input at each of the indicated siRNA tool websites [siDesign Center (Dharmacon, http://www.dharmacon.com/designcenter/DesignCenterPage.aspx), siRNA Target Finder (Genscript, https://www.genscript.com/ssl-bin/app/rnai), DSIR (Commissariat à l'Energie Atomique; France, http://biodev.cea.fr/DSIR/DSIR.html), and Applied Biosystems SVM siRNA Design Tool (http://www5.appliedbiosystems.com/tools/siDesign/) (14,26,29). These websites were selected for this comparison analysis because they are the select few of potency-based design tools that consider seed-based off-targeting. In each case, the optional parameters were adjusted to match our design scheme (e.g. 20–70% GC-content). At siDesign Center, output siRNAs for each of the 16 targets were sorted using by "Low Freq Seed" to identify candidates with low off-targeting potential among their top hits. For each target, up to 50 siRNAs were obtained for POTS analysis. At siRNA Target Finder, the Machine Learning option was used along with the Off-target filter (human, organ = house, seed size = 7, and Functional alignment option). Antiviral and Tradeoff options were deselected, and the output siRNAs (up to 10 per target gene) were used for POTS analysis. At DSIR, the default options were used and POTS for all candidates [ranging from 4 to 517 siRNAs per target gene (RTP801 and APOB, respectively)] were determined. For the Applied Biosystems siRNA Design Tool, sequences were uploaded and siRNAs obtained. For all siRNAs evaluated in these analyses, POTS were determined using positions 2–8 of the antisense strand.

Genome-wide shRNA coverage analysis and prospective library generation and comparison

The EMBOSS Splitter tool on the Galaxy web server (http://galaxyproject.org/) was used to generate a list of candidate siRNAs, for all human RefSeq 5′-UTR, CDS and 3′-UTR sequences using a 21-nt, 1-nt offset sliding window (30–32). Candidate siRNAs were filtered to promote antisense strand loading, retaining target sequences with the following pattern: $NN[G/C]_{3\text{-}4}N_{5\text{-}19}[A/T/C]_{20\text{-}21}$(14,33–35). Sequences falling outside of a 20–70% G/C content range were removed.

POTS values were obtained for the remaining sequences and were used to rank order candidate siRNAs for each transcript. Similar to previous publications and currently available RNAi libraries, candidates with near-perfect binding to other genes [0 or 1 mismatch across an 18-nt core (antisense positions 2-19)] were removed. For purposes of comparison with the RNAi Consortium human shRNA library (Broad Institute, MIT) (18) and coverage analysis, sequences corresponding to the 5′-UTR through the first 30-nt of the coding region were also removed. Candidate sites were grouped by Gene Symbol and duplicate values removed, noting sequences found in multiple transcript isoforms or with more than one site in the same transcript. A prospective shRNA library was generated by applying an additional filter to eliminate sequences with 'TTTT' or 'AAAA' motifs, allowing for compatibility with Pol-III expression-based systems. For each dataset, up to 10 candidates with the lowest POTS were included per gene.

For off-target comparison and coverage analysis with the RNAi Consortium shRNA library (one of the few with sequence information), POTS values were assigned based on position 2–8 of the reported antisense strand. POTS values were binned for each dataset for POTS distribution comparison. shRNA coverage analysis is reported based only on the genes included in the TRC dataset.

RESULTS

Low off-targeting siRNAs maintain potency

We first assessed whether siRNAs with low off-targeting potential have the capacity for potent silencing, because a diminished efficacy could explain their underrepresentation in the literature. Upon evaluation of 2431 randomly designed siRNAs described by Huesken *et al.* (henceforth referred to as the Huesken siRNA dataset) (36), we found that low off-targeting potential siRNAs (i.e. those having <2000 potential off-targets based on 3′-UTR seed complement hexamer distributions) exhibit comparable silencing efficiencies relative to the remaining sequences (~66 and 69% knockdown, respectively; Figure 2), with 1 in 4 siRNAs achieving $>80\%$ silencing, a commonly accepted threshold for potency. These results indicate that low off-targeting potential does not preclude siRNAs from being functional, suggesting that a siRNA design scheme weighted toward seed specificity would be capable of generating potent sequences.

Design of effective low off-targeting potential siRNAs

We thus developed a siRNA design algorithm termed siSPOTR (siRNA Seed Potential of Off-Target Reduction), which incorporates the most prominent determinants of siRNA efficacy while focusing mainly on seed specificity. For a given target sequence, all possible 21-mer siRNAs are filtered based on strand-loading and GC-content and then rank-ordered based on seed specificity.

Strand-biasing

First, siRNAs are selected to promote faithful loading of the antisense strand to mitigate potential off-targeting mediated by the sense strand. This is achieved using conventional siRNA design methodology based on duplex thermodynamic stability, with strong G–C binding at the 5′ end (2 bp) of the sense strand and weak A/G–U binding at the opposing end (2 bp; Figure 1A) (33,34), with target sites corresponding to $NN[G/C]_{3\text{-}4}N_{5\text{-}19}[A/T/C]_{20}[A/T/C/G]_{21}$. Notably, this differential stability represents the most significant attribute promoting siRNA efficacy, therefore encouraging potency in addition to specificity (i.e. preventing off-targeting from the sense strand)

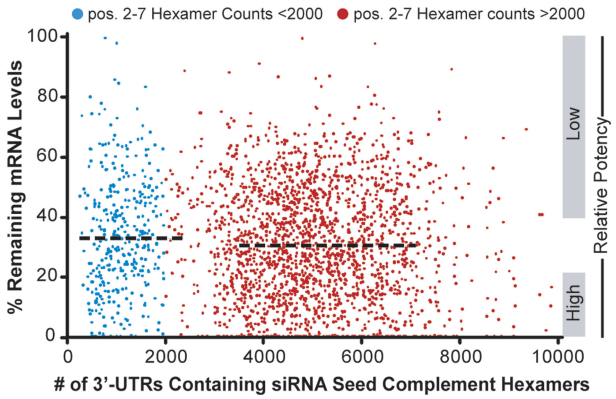

Figure 2. Effect of siRNA off-targeting potential on gene silencing capacity. A siRNA database composed of 2431 randomly designed siRNAs (targeting 31 unique mRNAs) and accompanying silencing data (36) was used to determine whether low off-targeting potential siRNAs (i.e. those having <2000 potential off-targets based on seed complement hexamer distributions in human RefSeq 3′-UTRs; blue) have similar capacities for gene silencing relative to the remaining 2068 siRNAs (mid-to-high off-targeting potentials; red). Approximately 1 in 4 of the low off-targeting potential siRNAs achieved >80% silencing (a commonly accepted threshold for potency), and overall their average efficiencies were comparable with the remaining siRNAs (~66 and 69% knockdown, respectively; dotted lines).

(14,35). To satisfy this criterion, weak G–U wobble pairing at the 3′ end of the target site can be introduced by converting cytosines into uridines. We allow sense strand modifications at position 20 and 21 (i.e. positions 1 and 2 of the antisense strand, respectively), while only permitting antisense strand modification at position 1. Previously published data support that the first position of the antisense strand does not influence targeting efficacy (37), and the ability to make these base conversions increases the number of potential target sites passing this strand-biasing filter.

GC-content

Next, putative target sequences are filtered based on GC-content, another strong determinant of siRNA potency (14,35). A range of 30–65% GC is considered optimal for identifying effective siRNAs and is generally used among potency-based siRNA design algorithms. To improve our yield of siRNAs with a potential for high specificity, we allow a broader range of 20–70% GC content. Our evaluation of the Huesken siRNA dataset supports that siRNAs within this range exhibit a suitable potential for efficient silencing of >80% (approximately 1 in 4 randomly designed siRNAs) (36) (data not shown).

Seed specificity

Finally, we rank candidate siRNAs by scoring seed specificity using a weighted system (POTS) that was formulated based on miRNA target recognition paradigms and siRNA off-targeting data derived from siRNA microarray studies (>50 unique siRNAs individually tested in Hela cells, Supplementary Table S1). Off-targeting among these datasets follows the well-characterized miRNA-based hierarchy of silencing potential based on seed site type (Figure 3A) (7); the presence of 8-mers within transcript 3′-UTRs confers a notably higher potential for down-regulation relative to the intermediate 7m8 and 7A1 sites, whereas 6-mer sites impart the least repressive potential over baseline transcripts (i.e. no sites). Statistical analyses performed on the datasets collectively revealed that the most significant divergence of the repressive potentials among all site types occurs at ≤ -0.3 log 2 fold-change ($P < 0.001$, Figure 3B). We next established a weighted PR (i.e. the likelihood for ≥ 0.3 log 2 fold-change down-regulation relative to baseline) for each site type by evaluating the siRNA experiments individually to control for the observed baseline variability among these datasets. The resulting PR values [8-mer (14.58%), 7m8 (7.68%), 7A1 (6.56%) and 6-mer (3.64%)] were calculated using the median values for each site type across the datasets. These PR values were then incorporated into the POTS formula which integrates both seed site type and frequency parameters. Previous reports have established that the potential for a miRNA to down-regulate a transcript depends not only on seed site types but also the frequencies of these sites within a target 3′-UTR (38–40). Grimson *et al.* reported that multiple miRNA seed sites in a single 3′-UTR primarily act in an independent and non-cooperative manner (e.g. two 8-mers impart twice the repressive potential relative to a single 8-mer). Our evaluation of the siRNA microarray experiments corroborated these results (data not shown), and thus, the POTS equation was formulated accordingly to provide an accurate estimation of off-targeting potentials.

$$\mathrm{POTS} = N_{8\mathrm{mer}}\mathrm{PR}_{8\mathrm{mer}} + N_{7\mathrm{M8}}\mathrm{PR}_{7\mathrm{M8}} + N_{7\mathrm{A1}}\mathrm{PR}_{7\mathrm{A1}} + N_{6\mathrm{mer}}\mathrm{PR}_{6\mathrm{mer}}$$

where N is the frequency of site in the 3′-UTRome, and PR is probability of repression.

We next calculated POTS for all possible 16 384 heptamers [note: heptamer sequences corresponding to positions 2–8 siRNAs/miRNAs determines all possible seed site type sequences (Figure 3A)] using transcriptome-wide human 3′-UTR sequences and observed a broad, non-uniform distribution of POTS, ranging from 5 to 5095 (Figure 3C). Not surprisingly, the highest scores were among heptamer sequences relevant to polyadenylation (e.g. AAAAAAA), whereas low POTS heptamers contain CpG dinucleotide motifs which are relatively rare within mammalian genomes. The POTS = 50 value is highlighted, representing an estimated but relevant cut-off which is employed henceforth for demonstrative purposes throughout this manuscript. This value is noteworthy because all 14 of the previously validated low off-targeting potential siRNAs tested by Anderson *et al.* have POTS < 50 (13). Furthermore, our evaluation of 750 siRNAs and accompanying *in vitro* cytotoxicity data support POTS < 50 as a conservative cut-off associated with an improved likelihood for tolerability (data not shown) (12). The siSPOTR specificity feature serves primarily to rank the off-targeting potential of siRNAs, and a firm cut-off for POTS values does not exist, much like

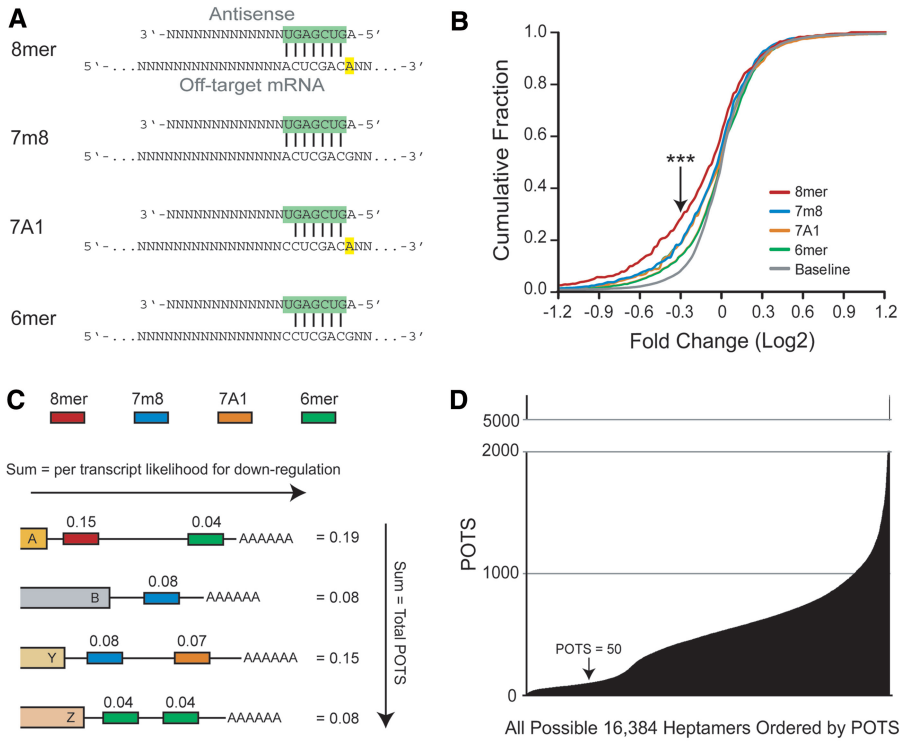

Figure 3. Formulation and distribution of POTS (potential off-targeting score). (**A**) Diagram illustrating the various seed site types. Seed sequences are highlighted in green. The adenosine corresponding to position 1 is highlighted in yellow and represents a defining feature for the 7A1 and 8mer binding site types. (**B**) The effect of seed site type on off-target silencing was determined using data from 54 microarray experiments testing unique siRNAs in HeLa cells. Cumulative distribution plots for gene expression values are shown for transcripts containing the relevant seed complement binding site types in their 3′-UTRs. Note: only transcripts containing singles sites for a given type and no other site types were considered. A shift to the left indicates an increased likelihood of being down-regulated relative to baseline transcripts (i.e. those lacking seed binding sites). ***Student *t* test indicated that the most significant divergence of the repressive potentials among these site types occurs at ≤ -0.3 log 2 fold-change ($P < 0.001$). (**C**) Schematic illustrating how POTS is calculated using seed site type frequency and PR values, shown above each respective site type. (**D**) The distribution of POTS scores—based on human 3′-UTR sequences—for all possible 16 384 heptamers is plotted. POTS < 50 is highlighted to indicate a relevant cut-off which is employed for purposes of this manuscript (refer to 'Results' section for further information regarding the relevance of this value).

for siRNA efficacy scores provided by potency-based siRNA design algorithms.

The importance of weighting seed site types is evident particularly in cases where seeds sharing the same core hexamer vary greatly in the number of genes containing the more potent 7- and 8-mer sites. For example, the seeds *CGCGAT*a and *CGCGAT*c each have 302 potential off-target transcripts (based on 3′-UTR hexamer counts) but respectively have 40 and 201 transcript 3′-UTRs with 7- or 8-mer sites. This 5-fold difference creates a considerable disparity in the off-targeting potentials of these seeds, resulting in a two-fold difference in their POTS values (Supplementary Tables S2 and S3). This illustrates the importance of considering position 8, which dictates the sequence of the most potent seed site types (i.e. 7m8 and 8mer). We calculated the mean site type frequencies for all possible heptamers binned by POTS values, revealing nearly a 5- to 10-fold reduction in the more potent site types for Low POTS heptamers, relative to those with medium-to-high POTS (e.g. for 8mers, mean values of ~45 compared with >350, respectively; Supplementary Table S4).

Finally, as means to further refine our prediction of off-targeting potentials, we considered the degree to which POTS is influenced by variations in gene expression changes across tissues. For this, transcriptional profiling data from 177 different human cell lines and tissues (BioGPS) were used to calculate tissue-specific POTS for all possible heptamers. Although gene expression patterns vary greatly across tissues, POTS ranks for each heptamer correlate strongly ($r^2 > 0.95$; Supplementary Figure S1). These data support that organism-wide application of POTS is suitable.

siSPOTR design example

We provide a step-wise example illustrating the use of siSPOTR for designing siRNAs targeting the human PPIB-coding sequence (CDS; Figure 4). The 648-nt target sequence is first divided up to produce all 631 possible 21-mer siRNA target sites, and the strand-biasing and GC-content filters described earlier are applied prior to determining POTS values for the resulting siRNAs. In this example, among the 113 PPIB-targeted siRNAs, which satisfy the strand-biasing and GC-content criteria, seven are represented in the siRNA validation datasets described later, allowing visualization of the measured off-targeting associated with their respective POTS values of 25, 29, 40, 407, 410, 510 and 560 (Figure 5).

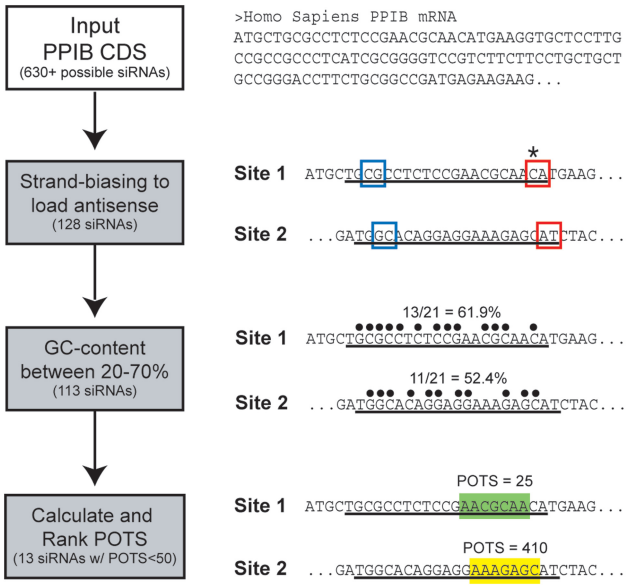

Figure 4. Workflow schematic for designing siRNAs targeting human PPIB using the siSPOTR algorithm. All possible 631 siRNAs targeting the human PPIB coding sequence (CDS) were filtered based on strand biasing [i.e. strong G–C (blue) and weak A/G–U (red) binding at the respective 5′ and 3′ ends of the sense strand] and GC-content, and the number of siRNAs passing each criteria are provided. Note: the asterisk denotes a cytosine base in the 3′ end of the target site; this base can be converted to a uridine to produce a weak G:U base-pairing in the resulting siRNA duplex. The heptamer seed sequence used for POTS determination is highlighted.

Validation of siSPOTR algorithm: efficacy and specificity

Efficacy

We gauged the capacity of siSPOTR to identify potent siRNA sequences among the siRNAs in the Huesken dataset (Figure 5A). The siRNAs satisfying the strand-biasing and GC-content criteria were rank ordered by POTS (low to high), yielding seven siRNAs with POTS < 50. Here, this relatively low number results from fewer sequences passing the strand-biasing filter, since the capacity for introducing duplex instability using G–U base-pairs, as described earlier, is not applicable to these pre-existing siRNAs. Surprisingly, these seven siRNAs each had >80% silencing efficacy, with a mean comparable with that of siRNAs within the database that were identified among the top hits generated by siDesign Center (Dharmacon), a widely used siRNA design website. Although siDesign Center yields more hits among this database, only two of these siRNAs has a POTS < 50. Indeed, siSPOTR identified five siRNAs not among the siDesign Center hits (Figure 5A, Venn diagram), highlighting the unique output potential of the siSPOTR algorithm.

Off-targeting potential

We next evaluated the predictive power of POTS to estimate the extent of off-target gene silencing observed among microarray experiments for 40 unique siRNAs targeting GAPDH, PPIB, or "No Target" (Supplementary Table S1). These 40 experiments were selected because the siRNAs encompass a broad range of POTS with relatively equal representation across low, medium and high scores. To improve our ability to discern sequence-specific off-targeting from gene expression changes associated with on-target silencing, the datasets were grouped by target gene prior to calculating differential gene expression and establishing "suppression signatures" for each siRNA. Furthermore, each of these 40 siRNAs exhibits >85% silencing efficacy, reducing the potential for detecting gene expression changes due to varying degrees of on-target silencing within groups. In support of the POTS approach, our analyses of these datasets reveals smaller sequence-specific "suppression signatures" among the low off-targeting potential siRNAs (POTS < 50), relative to siRNAs with higher POTS (Figure 5B). Notably, 13 of 28 higher POTS siRNAs produced greater "suppression signatures" than the largest one observed among the low POTS siRNAs (Figure 5C). It is important to note that our analyses and previously published data support that these "suppression signatures" consist of down-regulated transcripts that are enriched for 3′-UTR seed-binding motifs, suggesting that most are likely to be direct siRNA off-targets (6,41).

The prospect of using POTS to accurately rank off-targeting potentials among these 40 siRNAs was also assessed. Spearman rank correlation of the POTS scores and numbers of down-regulated off-targets observed for each siRNA indicated a positive correlation of modest significance (Figure 5D, dotted line, $P = 0.05$). As depicted by this plot, a few higher POTS siRNAs have low numbers of off-targets (red dots); however, none of the low POTS siRNAs showed high numbers of off-targets. Indeed, removing the overt outliers among the higher POTS siRNAs produces a highly significant correlation (solid line, $P < 1E{-}8$), providing further evidence that POTS is a reliable predictor of siRNA off-targeting potentials. These data, in conjunction with the efficacy validation, establish the robust capability of siSPOTR to identify highly specific and effective siRNAs.

Finally, we reasoned that training on more datasets (i.e. combining the training and validation sets described earlier) could generate a more accurate POTS for ranking siRNA off-targeting potentials. As expected, the Spearman rank correlation of POTS scores and numbers of down-regulated off-targets observed for each siRNA showed even greater significance (Supplementary Figure S2). These improved POTS values are used henceforth.

Comparison of siSPOTR with other algorithms

We subsequently compared the abilities of our design strategy and other publically available algorithms, particularly those which incorporate seed specificity parameters, to identify siRNAs with low off-targeting potential seeds (i.e. low POTS). The coding sequences of 16 therapeutically relevant genes (of varying sizes; comprising in total ~50 kb) were used as input, and the number of candidate siRNAs with POTS < 50 was determined for each algorithm. Our design scheme identified more low off-targeting potential siRNAs [at least four siRNAs (a typical starting number for initial efficacy screening) for

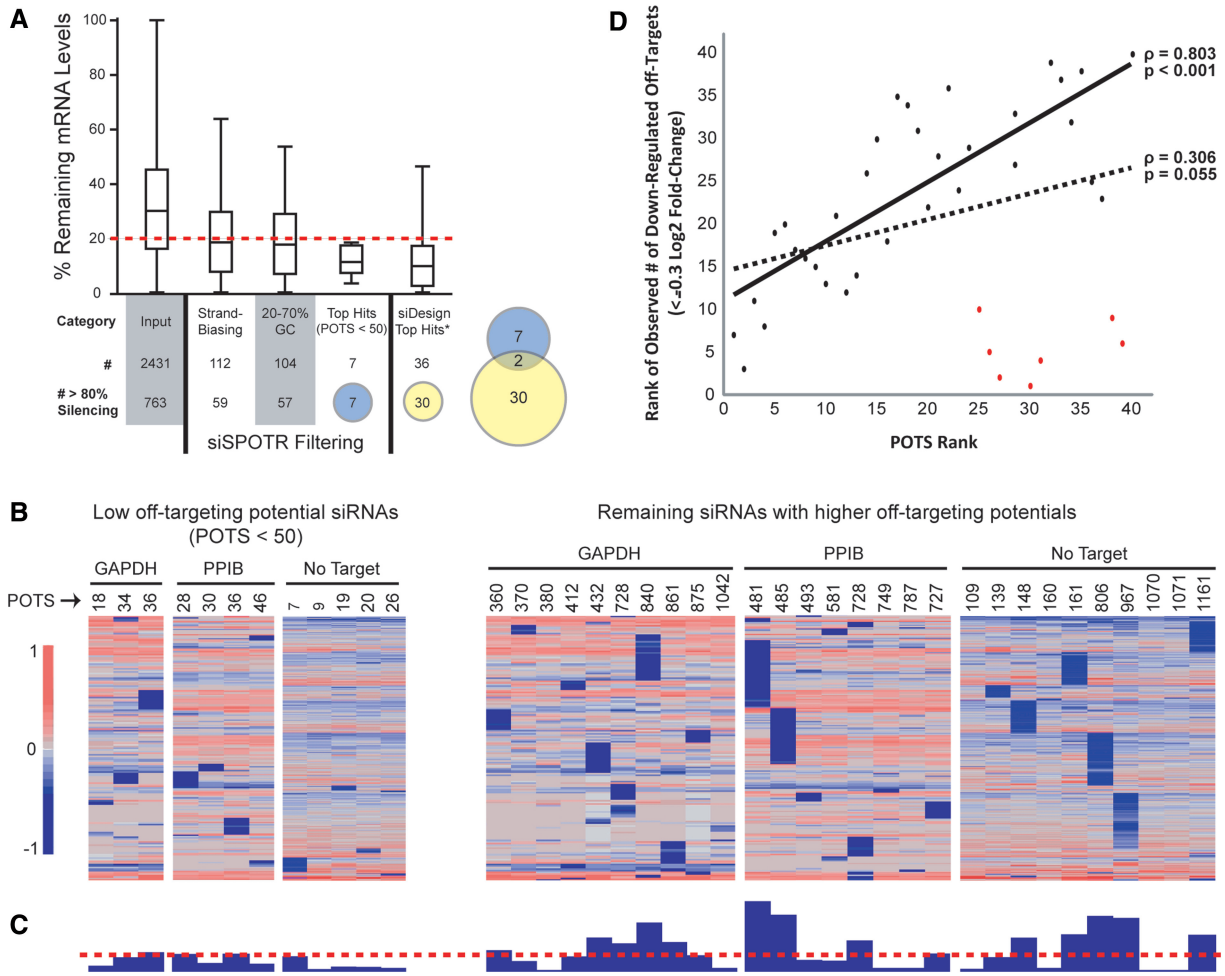

Figure 5. Validation of siSPOTR: efficacy and off-targeting. (**A**) siRNA efficacy was evaluated using a database of 2431 randomly designed siRNAs with accompanying silencing data. The number of siRNAs passing each stage of our stepwise filtering process is indicated along with the number of potent sequences among them (i.e. those with >80% silencing efficacy. *siDesign Center (Dharmacon) was used for comparison by inputting the relevant target gene sequences into the online tool ($N = 29$) and intersecting the top 10 hits for each gene with the 2431 siRNAs. The box and whiskers plot shows the max and min gene silencing values (whiskers) and the upper and lower quartiles (box). The accompanying Venn diagram shows that siSPOTR identified five unique and effective sequences not present among the siDesign Top Hits. (B-D) Microarray data from experiments testing 40 unique siRNAs were used to assess the reliability of POTS as an indicator for off-targeting potential. (**B**) Heatmaps representing sequence-specific gene "suppression signatures" unique to each siRNA were generated using hierarchical clustering of significantly down-regulated genes (>3 standard deviations from the mean) among the datasets on a per target gene basis (i.e. GAPDH, PPIB and No Target), and columns were ordered and parsed by POTS for each group. (**C**) A qualitative representation of "suppression signature" size (i.e. sum of dark blue regions) for each column is shown. The red dotted line marks the largest "suppression signature" among the siRNAs with POTS < 50. (**D**) Spearman rank correlation of the POTS scores and numbers of down-regulated off-targets (i.e. transcripts with 3′-UTRs containing 7- and 8-mer seed-binding sites and ≤ -0.3 log 2 fold-change) observed for each siRNA is plotted. Linear regression lines, including correlation coefficients and P values, for all data points (dotted line) and black dots (solid line) are provided. Red dots represent overt outliers.

all 16 of the input genes] relative to the other algorithms, which failed to generate at least four siRNAs with POTS < 50 for at least 8 of the 16 genes (Table 1). This observation emphasizes a considerable limitation of current siRNA design tools that are strongly biased towards potency, highlighting the unique functionality that siSPOTR provides to researchers seeking siRNAs with low off-targeting potentials.

Prospective applications to expressed RNAi and genome-wide RNAi libraries

The siSPOTR algorithm provides an attractive approach for limiting off-targeting from hairpin-based RNAi expression systems, which unlike siRNAs, are not amenable to chemical modifications that may reduce seed-based off-targeting (42–44). Recently, we published microarray data supporting that RNAi vectors expressing siRNAs with low off-targeting potentials (based on 3′-UTR hexamer frequencies) show reduced off-targeting relative to sequences with more promiscuous seeds (17). To ascertain whether POTS can be a reliable indicator of off-targeting from expressed RNAi, we evaluated the association of POTS with off-targeting for the expressed RNAi sequences tested in this previous study (eight constructs with POTS ranging from 11 to 653). Hierarchical clustering of differentially expressed genes ($N = 827$, $P < 0.0001$) among the

Table 1. Comparison of siRNA design tools

Gene	CDS (nt)	No. of siRNA candidates generated with POTS <50 [a]				
		siSPOTR	siDesign	Genscript	DSIR	AppBio
SNCA	423	4	0	0	0	0
SOD1	465	4	1	0	0	0
RTP801	699	19	5	1	0	0
TOR1a	999	14	3	6	6	1
SCA3	1 086	6	4	2	3	0
VEGF	1 239	22	4	4	1	2
MYC	1 365	31	7	2	4	3
BACE1	1 506	18	0	2	0	0
KRT6a	1 695	23	0	1	2	0
SCA1	2 448	42	2	1	3	1
SCA7	2 679	35	6	3	7	2
EGFR1	3 633	47	5	3	13	2
BCR-Abl	3 816	83	7	2	7	2
SCA2	3 942	42	2	2	13	1
HTT	9 435	82	3	N/A	8	N/A
APOB	13 692	66	1	N/A	14	N/A
Total	49 122	538	50	29	81	14
At least four siRNAs?		16 of 16	7 of 16	2 of 16	8 of 16	0 of 16

[a]POTS < 50 serves as a relevant cut-off for purposes of this manuscript (refer to 'Results' section for further information regarding the relevance of this value). N/A indicates that the online tool was unable to process transcripts of this length.

various RNAi sequences reveals that the clustering distance relative to the control (i.e. promoter-only vector) increases in agreement with rising POTS values (Supplementary Figure S3), supporting that low POTS RNAi sequences induce fewer gene expression changes as compared with sequences with higher POTS values. These data substantiate the utility of siSPOTR for improving the specificity of RNAi expression vectors.

Next, we investigated the feasibility of generating a genome-wide shRNA library using this algorithm. Genome-wide RNAi screens are broadly used to discover genes implicated in biological pathways and phenotypes; however, these screens can be plagued by off-target effects producing false leads (10,11). Although bioinformatic approaches show some practicality for distinguishing off-targets from bona fide targets (45,46), careful attention to sequence selection may greatly reduce off-targeting among libraries. There are currently several RNAi libraries available in synthetic siRNA or expressed forms (e.g. shRNAs). Here, we demonstrate the potential of our siRNA design scheme to generate genome-wide RNAi libraries with high specificity (based on POTS and BLAST, see methods). Our prospective shRNA library ("Low POTS") consists of 235 121 sequences (up to 10 shRNAs per target gene; $POTS_{median} = 37$) and provides at least 4 shRNAs with <50 POTS for more than 78% of all RefSeq mRNAs (Figure 6). These sequences have reduced (nearly 10-fold on average) off-targeting potential over those offered in a publically available shRNA library [178 265 sequences; $POTS_{median} = 322$; The RNAi Consortium (TRC)], which covers 0.70% of RefSeq mRNAs with at least four shRNAs having <50 POTS. A histogram of the POTS distributions for each of these libraries reveals an evident disparity, with >90% of the sequences having improved POTS relative to the TRC library which followed a near-random distribution mirroring POTS for all possible heptamers. For genome-wide siRNA design, the "low POTS" library coverage is even broader (data not shown), providing an additional means to enhance specificity in combination with chemical modifications to the seed (42–44).

siSPOTR online tool

Based on these observations, we developed an online tool employing the siSPOTR algorithm to assist users with designing RNAi sequences with low off-targeting potential for application in human and mouse (URL: sispotr.icts.uiowa.edu). The siSPOTR tool searches user-defined target sequences for siRNAs that pass strand-biasing and GC% filters and outputs candidate siRNAs rank-ordered by POTS from lowest to highest. For convenience, the sequences are ready-to-order with the necessary nucleotide substitutions made to promote proper strand-loading. In addition, DNA oligonucleotide sequences for generating corresponding shRNAs are supplied to assist users with generating RNAi expression vectors. The output also provides detailed off-targeting information for each siRNA including (i) the number of 3′-UTRs containing each seed site type, (ii) the putative off-target transcripts and (iii) counts of each seed site type on a per transcript basis. The siSPOTR tool also alerts the user if the siRNA seed sequence matches that of a known miRNA, as such an instance may confound experimental results given the regulatory roles miRNAs play in numerous biological processes and pathways. Furthermore, recognizing the ease of purchasing pre-validated siRNAs and shRNAs, we provide an accompanying online tool, which allows users to input siRNA sequences to obtain POTS values and the detailed off-targeting information described earlier. These tools will provide researchers with dependable means to minimize and evaluate off-targeting concerns associated with RNAi experiments.

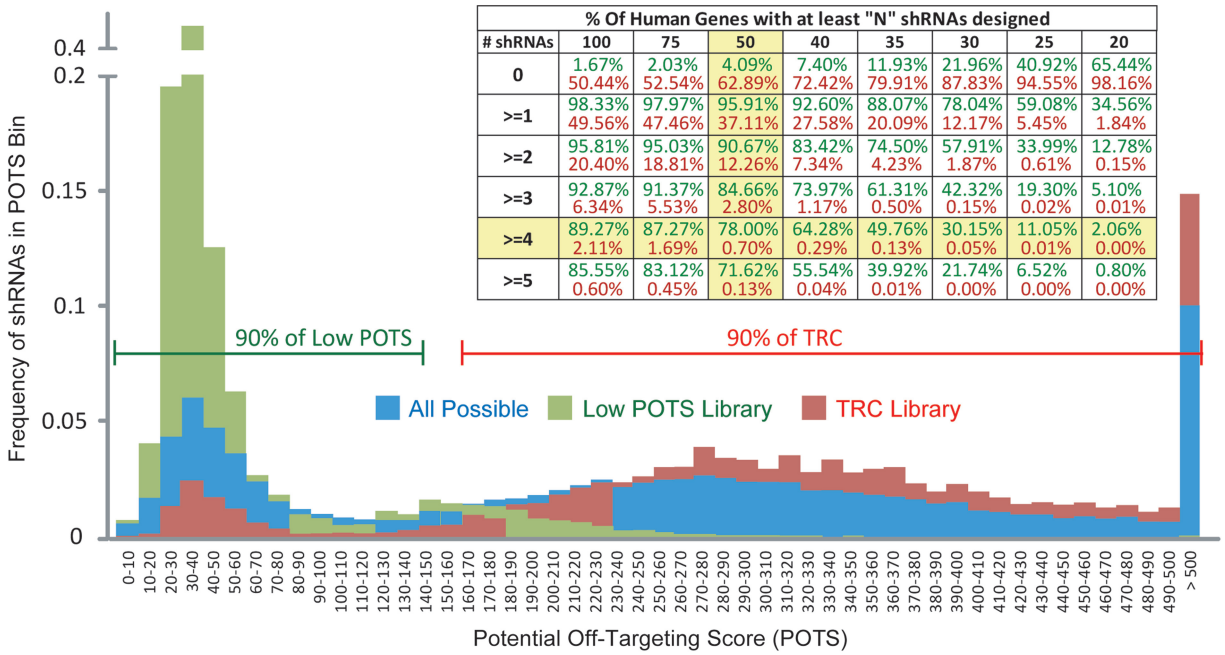

Figure 6. Comparison of off-targeting potentials among shRNA libraries. A histogram and complementing table presenting the POTS distributions and genome-wide coverage of shRNA library sequences are shown for our "Low POTS" library (green) and the TRC library (red). The POTS distribution of all possible heptamers (blue) serves as a reference. The range encompassing 90% of all sequences for each shRNA library is indicated. Yellow highlights intersect to emphasize the coverage disparities at a key point; POTS < 50 provides a conservative cut-off for low off-targeting potential, and at least four siRNAs are desired for a given gene when generating a library or performing initial efficacy screening.

DISCUSSION

Consideration of seed pairing stability

A recent report from the Bartel laboratory evaluated the impact of seed-pairing stability (SPS) and target abundance (TA; levels of potential binding sites in the cellular transcriptome) on seed-mediated silencing by small RNAs (miRNAs and siRNAs) (20). Their data support that seeds with weak SPS inherently have higher TA, and that both factors limit seed-based silencing potency, presumably from weaker binding and a dilution effect associated with the increased number of targets. In contrast to the siSPOTR approach, the authors propose that designing siRNAs with weak SPS and high TA seeds may minimize off-targeting potential. While the potency of such seeds may be low on average, the possibility of repressing considerably more off-targets exists. A comparison of the low POTS approach with the weak SPS strategy may be warranted. When accounting for repressive potentials in addition to the numbers of predicted off-targets, it is likely that siRNAs having weak SPS would consistently have higher numbers of off-targets expected to be down-regulated, relative to low POTS siRNAs. Even yet, SPS is worthy of consideration for siRNA design, and we have added SPS values to the siSPOTR output, so that users may avoid stronger SPS seeds among siRNAs with comparable POTS values.

The utility of siSPOTR

Off-target effects (e.g. false discovery rates and toxicity) pose a problem for gene silencing technologies, particularly for RNAi therapeutics, thus supporting the need for developing a user-friendly tool to assist researchers in designing siRNAs which are highly specific and efficacious. Here, and in prior work from our laboratory and others', we demonstrate that focusing on seed specificity in siRNA design may mitigate off-targeting by 5- to 10-fold, as supported by predictive analyses and transcriptional profiling data from RNAi studies (13,17). Unlike other siRNA design strategies, siSPOTR yields numerous candidate sequences with low off-targeting potentials, providing a broad and attractive approach towards alleviating off-target concerns. Other means to address off-targeting have been previously described. For example, in basic biological research, scientists may employ "same seed" controls (i.e. containing the same seed sequence as the experimental siRNA, but central mismatches to prevent silencing of the target of interest) to discern on-target versus off-target effects (17). Furthermore, research supports that off-targeting from synthetic siRNAs can be reduced by chemical modifications or using lower doses (26,42–44,47); however, specificity could be enhanced further by employing seeds with low POTS. By contrast, for expressed RNAi forms (e.g. shRNAs), our approach provides the only broadly applicable methodology to limit off-targeting potential. Although sequence-specific effects on hairpin expression, stability, and processing may also contribute to off-targeting potential, our data support that POTS values provide a good predictor of off-targeting for RNAi expression vectors. This is important particularly because dosing from RNAi expression vectors cannot be as readily

controlled, and shRNA-induced toxicities have been reported by several groups (48–51). Given the extensive use of RNAi expression systems in the laboratory and in therapeutic development, siSPOTR will serve as a valuable tool to the research community.

siSPOTR can easily be used in conjunction with other siRNA design algorithms (e.g. those weighted toward efficacy) to query their outputs for off-target potential. For instance, one can use Applied Biosystems' hyperfunctional (i.e. highly potent) siRNA design tool to identify hyperfunctional candidate sequences, which can subsequently be input into the siSPOTR tool to retrieve their POTS values (26). This combined approach aims to ascertain siRNAs with a highly desirable balance of potency and low off-targeting potential, providing an attractive means to identify therapeutic siRNAs for disease-relevant targets, particularly larger genes which have numerous low POTS siRNAs available (Table 1).

siSPOTR allows users to query the identities of predicted seed-based off-target transcripts as means to avoid potentially important cellular genes (e.g. those involved in cell cycle and viability). Off-target identity is an important contributor to the overall detrimental effects caused by disrupting gene networks, and the resulting tolerability for a given siRNA. However, declaring a predicted off-target to be important remains difficult due to a dependence on numerous variables [e.g. experimental system (i.e. cell type), duration and extent of knockdown, identities of other off-targets (e.g. a two-hit model), etc.]. Nevertheless, although researchers should consider the identities of predicted off-targets, it stands to reason that minimizing the off-targeting potential of the siRNA seed will inherently reduce the likelihood of unintentionally silencing important genes and further limit downstream events associated with cascading gene networks.

Finally, siSPOTR supports RNAi sequence design for human and mouse experimental systems; and moreover, all low POTS heptamers contain CpG motifs which are sparse throughout mammalian genomes. Furthermore, the ranking of heptamers by POTS for mouse and human reveals a significant correlation ($r^2 > 0.938$, plot not shown), suggesting that siSPOTR is likely applicable to other mammalian species.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4 and Supplementary Figures 1–3.

ACKNOWLEDGEMENTS

The authors thank the B.L.D and McCray laboratories, in addition to Edgardo Rodriguez and Yi Xing (University of Iowa) and Mark Behlke (Integrated DNA Technologies) for feedback, manuscript and website review and discussion, and members of the University of Iowa Research community for feedback on the siSPOTR tool.

FUNDING

National Institutes of Health [NS50210 and NS068280]; the National Center for Research Resources [UL1 RR024979]; the Hereditary Disease Foundation; the Roy J. Carver Trust; the Lori C. Sasser Fellowship (to R.L.B.). Funding for open access charge: NIH, the Roy J carver Trust.

Conflict of intertest statement. None declared.

REFERENCES

1. Provost,P., Dishart,D., Doucet,J., Frendewey,D., Samuelsson,B. and Radmark,O. (2002) Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.*, **21**, 5864–5874.
2. Davidson,B.L. and McCray,P.B. Jr (2011) Current prospects for RNA interference-based therapies. *Nat. Rev. Genet.*, **12**, 329–340.
3. Chi,J.T., Chang,H.Y., Wang,N.N., Chang,D.S., Dunphy,N. and Brown,P.O. (2003) Genomewide view of gene silencing by small interfering RNAs. *Proc. Natl. Acad. Sci. USA*, **100**, 6343–6346.
4. Semizarov,D., Frost,L., Sarthy,A., Kroeger,P., Halbert,D.N. and Fesik,S.W. (2003) Specificity of short interfering RNA determined through gene expression signatures. *Proc. Natl. Acad. Sci. USA*, **100**, 6347–6352.
5. Jackson,A.L., Bartz,S.R., Schelter,J., Kobayashi,S.V., Burchard,J., Mao,M., Li,B., Cavet,G. and Linsley,P.S. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.*, **21**, 635–637.
6. Jackson,A.L., Burchard,J., Schelter,J., Chau,B.N., Cleary,M., Lim,L. and Linsley,P.S. (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA*, **12**, 1179–1187.
7. Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
8. Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
9. Birmingham,A., Anderson,E.M., Reynolds,A., Ilsley-Tyree,D., Leake,D., Fedorov,Y., Baskerville,S., Maksimova,E., Robinson,K., Karpilow,J. *et al.* (2006) 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat. Methods*, **3**, 199–204.
10. Ma,Y., Creanga,A., Lum,L. and Beachy,P.A. (2006) Prevalence of off-target effects in Drosophila RNA interference screens. *Nature*, **443**, 359–363.
11. Schultz,N., Marenstein,D.R., De Angelis,D.A., Wang,W.Q., Nelander,S., Jacobsen,A., Marks,D.S., Massague,J. and Sander,C. (2011) Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-beta pathway and reveal microRNA regulation of TGFBR2. *Silence*, **2**, 3.
12. Fedorov,Y., Anderson,E.M., Birmingham,A., Reynolds,A., Karpilow,J., Robinson,K., Leake,D., Marshall,W.S. and Khvorova,A. (2006) Off-target effects by siRNA can induce toxic phenotype. *RNA*, **12**, 1188–1196.
13. Anderson,E.M., Birmingham,A., Baskerville,S., Reynolds,A., Maksimova,E., Leake,D., Fedorov,Y., Karpilow,J. and Khvorova,A. (2008) Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA*, **14**, 853–861.
14. Birmingham,A., Anderson,E., Sullivan,K., Reynolds,A., Boese,Q., Leake,D., Karpilow,J. and Khvorova,A. (2007) A protocol for designing siRNAs with high functionality and specificity. *Nat. Protoc.*, **2**, 2068–2078.
15. Naito,Y., Yamada,T., Ui-Tei,K., Morishita,S. and Saigo,K. (2004) siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.*, **32**, W124–129.

16. Jackson,A.L. and Linsley,P.S. (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nat. Rev. Drug Discov.*, **9**, 57–67.
17. Boudreau,R.L., Spengler,R.M. and Davidson,B.L. (2011) Rational design of therapeutic siRNAs: minimizing off-targeting potential to improve the safety of RNAi therapy for Huntington's disease. *Mol. Ther.*, **19**, 2169–2177.
18. Moffat,J., Grueneberg,D.A., Yang,X., Kim,S.Y., Kloepfer,A.M., Hinkle,G., Piqani,B., Eisenhaure,T.M., Luo,B., Grenier,J.K. *et al.* (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, **124**, 1283–1298.
19. McBride,J.L., Pitzer,M.R., Boudreau,R.L., Dufour,B., Hobbs,T., Ojeda,S.R. and Davidson,B.L. (2011) Preclinical safety of RNAi-mediated HTT suppression in the rhesus macaque as a potential therapy for Huntington's disease. *Mol. Ther.*, **19**, 2152–2162.
20. Garcia,D.M., Baek,D., Shin,C., Bell,G.W., Grimson,A. and Bartel,D.P. (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat. Struct. Mol. Biol.*, **18**, 1139–1146.
21. Karolchik,D., Hinrichs,A.S., Furey,T.S., Roskin,K.M., Sugnet,C.W., Haussler,D. and Kent,W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–496.
22. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
23. Fujita,P.A., Rhead,B., Zweig,A.S., Hinrichs,A.S., Karolchik,D., Cline,M.S., Goldman,M., Barber,G.P., Clawson,H., Coelho,A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–D882.
24. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
25. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2005) NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **33**, D501–D504.
26. Wang,X., Wang,X., Varma,R.K., Beauchamp,L., Magdaleno,S. and Sendera,T.J. (2009) Selection of hyperfunctional siRNAs with improved potency and specificity. *Nucleic Acids Res.*, **37**, e152.
27. Wu,C., Orozco,C., Boyer,J., Leglise,M., Goodale,J., Batalov,S., Hodge,C.L., Haase,J., Janes,J., Huss,J.W. 3rd *et al.* (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.*, **10**, R130.
28. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA*, **101**, 6062–6067.
29. Vert,J.P., Foveau,N., Lajaunie,C. and Vandenbrouck,Y. (2006) An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics*, **7**, 520.
30. Goecks,J., Nekrutenko,A. and Taylor,J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
31. Blankenberg,D., Von Kuster,G., Coraor,N., Ananda,G., Lazarus,R., Mangan,M., Nekrutenko,A. and Taylor,J. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.*, Chapter 19, Unit 19.10.1–21.
32. Giardine,B., Riemer,C., Hardison,R.C., Burhans,R., Elnitski,L., Shah,P., Zhang,Y., Blankenberg,D., Albert,I., Taylor,J. *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.
33. Schwarz,D.S., Hutvagner,G., Du,T., Xu,Z., Aronin,N. and Zamore,P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199–208.
34. Khvorova,A., Reynolds,A. and Jayasena,S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.
35. Matveeva,O., Nechipurenko,Y., Rossi,L., Moore,B., Saetrom,P., Ogurtsov,A.Y., Atkins,J.F. and Shabalina,S.A. (2007) Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. *Nucleic Acids Res.*, **35**, e63.
36. Huesken,D., Lange,J., Mickanin,C., Weiler,J., Asselbergs,F., Warner,J., Meloon,B., Engel,S., Rosenberg,A., Cohen,D. *et al.* (2005) Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.*, **23**, 995–1001.
37. Miller,V., Gouvion,C., Davidson,B. and Paulson,H. (2004) Targeting Alzheimer's disease genes with RNA interference: an efficient strategy for silencing mutant allele. *Nucleic Acids Res.*, **32**, 661–668.
38. Nielsen,C.B., Shomron,N., Sandberg,R., Hornstein,E., Kitzman,J. and Burge,C.B. (2007) Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA*, **13**, 1894–1910.
39. Doench,J.G. and Sharp,P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.
40. Grimson,A., Farh,K.K., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell.*, **27**, 91–105.
41. Burchard,J., Jackson,A.L., Malkov,V., Needham,R.H., Tan,Y., Bartz,S.R., Dai,H., Sachs,A.B. and Linsley,P.S. (2009) MicroRNA-like off-target transcript regulation by siRNAs is species specific. *RNA*, **15**, 308–315.
42. Jackson,A.L., Burchard,J., Leake,D., Reynolds,A., Schelter,J., Guo,J., Johnson,J.M., Lim,L., Karpilow,J., Nichols,K. *et al.* (2006) Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA*, **12**, 1197–1205.
43. Bramsen,J.B., Pakula,M.M., Hansen,T.B., Bus,C., Langkjaer,N., Odadzic,D., Smicius,R., Wengel,S.L., Chattopadhyaya,J., Engels,J.W. *et al.* (2010) A screen of chemical modifications identifies position-specific modification by UNA to most potently reduce siRNA off-target effects. *Nucleic Acids Res.*, **38**, 5761–5773.
44. Vaish,N., Chen,F., Seth,S., Fosnaugh,K., Liu,Y., Adami,R., Brown,T., Chen,Y., Harvie,P., Johns,R. *et al.* (2011) Improved specificity of gene silencing by siRNAs containing unlocked nucleobase analogs. *Nucleic Acids Res.*, **39**, 1823–1832.
45. Sigoillot,F.D., Lyman,S., Huckins,J.F., Adamson,B., Chung,E., Quattrochi,B. and King,R.W. (2012) A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nat. Methods*, **9**, 363–366.
46. Zhang,X.D., Santini,F., Lacson,R., Marine,S.D., Wu,Q., Benetti,L., Yang,R., McCampbell,A., Berger,J.P., Toolan,D.M. *et al.* (2011) cSSMD: assessing collective activity for addressing off-target effects in genome-scale RNA interference screens. *Bioinformatics*, **27**, 2775–2781.
47. Caffrey,D.R., Zhao,J., Song,Z., Schaffer,M.E., Haney,S.A., Subramanian,R.R., Seymour,A.B. and Hughes,J.D. (2011) siRNA off-target effects can be reduced at concentrations that match their individual potency. *PLoS One*, **6**, e21503.
48. Grimm,D., Streetz,K.L., Jopling,C.L., Storm,T.A., Pandey,K., Davis,C.R., Marion,P., Salazar,F. and Kay,M.A. (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature*, **441**, 537–541.
49. McBride,J.L., Boudreau,R.L., Harper,S.Q., Staber,P.D., Monteys,A.M., Martins,I., Gilmore,B.L., Burstein,H., Peluso,R.W., Polisky,B. *et al.* (2008) Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. *Proc. Natl. Acad. Sci. USA*, **105**, 5868–5873.
50. Boudreau,R.L., Martins,I. and Davidson,B.L. (2009) Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs in vitro and in vivo. *Mol. Ther.*, **17**, 169–175.
51. Martin,J.N., Wolken,N., Brown,T., Dauer,W.T., Ehrlich,M.E. and Gonzalez-Alegre,P. (2011) Lethal toxicity caused by expression of shRNA in the mouse striatum: implications for therapeutic design. *Gene Ther.*, **18**, 666–673.