

RESEARCH

Open Access



# shRNA target prediction informed by comprehensive enquiry (SPICE): a supporting system for high-throughput screening of shRNA library

Kenta Kamatuka<sup>1</sup>, Masahiro Hattori<sup>2</sup> and Tomoyasu Sugiyama<sup>1\*</sup>

## Abstract

RNA interference (RNAi) screening is extensively used in the field of reverse genetics. RNAi libraries constructed using random oligonucleotides have made this technology affordable. However, the new methodology requires exploration of the RNAi target gene information after screening because the RNAi library includes non-natural sequences that are not found in genes. Here, we developed a web-based tool to support RNAi screening. The system performs short hairpin RNA (shRNA) target prediction that is informed by comprehensive enquiry (SPICE). SPICE automates several tasks that are laborious but indispensable to evaluate the shRNAs obtained by RNAi screening. SPICE has four main functions: (i) sequence identification of shRNA in the input sequence (the sequence might be obtained by sequencing clones in the RNAi library), (ii) searching the target genes in the database, (iii) demonstrating biological information obtained from the database, and (iv) preparation of search result files that can be utilized in a local personal computer (PC). Using this system, we demonstrated that genes targeted by random oligonucleotide-derived shRNAs were not different from those targeted by organism-specific shRNA. The system facilitates RNAi screening, which requires sequence analysis after screening. The SPICE web application is available at <http://www.spice.sugysun.org/>.

## 1 Introduction

Reverse genetics approaches, which enable the determination of gene function by analyzing loss-of-function in a phenotype, have been useful for investigating the role of genes in cells and organisms [1, 2]. Recent progress in whole genome sequencing and comprehensive expressed complementary DNA (cDNA) sequencing has enabled the use of systematic approaches to uncover the roles of genes that have been categorized as unknown function genes. Reverse genetics approaches, such as gene knock-out with homologous recombination and gene knock-down with antisense RNA, have been highly effective; however, they do not yield rapid results as gene silencing using double-stranded RNA (dsRNA) under RNA interference (RNAi). RNAi allows obtaining loss-of-function

phenotypes with high efficiency and specificity within a short period in a wide range of organisms.

Genome-wide reverse genetics performed by RNAi was first demonstrated in *Caenorhabditis elegans* to investigate genes involved in development [1]. Remarkably, the method required no laborious processes to obtain efficient induction of RNAi in the organism. For example, RNAi induction was demonstrated by feeding animals with *Escherichia coli* expressing dsRNA. The strategy also worked in other organisms such as planaria, a model animal for regeneration [3, 4]. However, the use of cDNA-derived dsRNA has been limited to invertebrates because long dsRNAs (>30 nucleotides (nt)) evoke interferon responses in vertebrates. This problem was solved by using small interfering RNA (siRNA) comprising ~21 nt, i.e., 19 bp with 2-nt 3' overhangs [5]. siRNAs could also be transformed from short hairpin RNA (shRNA) within transfected cells. The findings led to the development of siRNA-directed reverse genetics methods, which included

\* Correspondence: [tsugiyama@stf.teu.ac.jp](mailto:tsugiyama@stf.teu.ac.jp)

<sup>1</sup>Graduate School of Bionics, Tokyo University of Technology, 1401-1 Katakura-machi, Hachioji, Tokyo 192-0982, Japan

Full list of author information is available at the end of the article

RNAi library construction and screening systems [2, 6–8]. Methodological progress has also revealed that the efficiency of knockdown depends on sequence within each siRNA [9, 10]. Consequently, algorithms were developed to find efficient sequences from genome databases for RNAi and were utilized to design synthetic siRNA oligonucleotides. Associated web applications using these algorithms have facilitated the analysis of loss-of-function phenotypes [11–13].

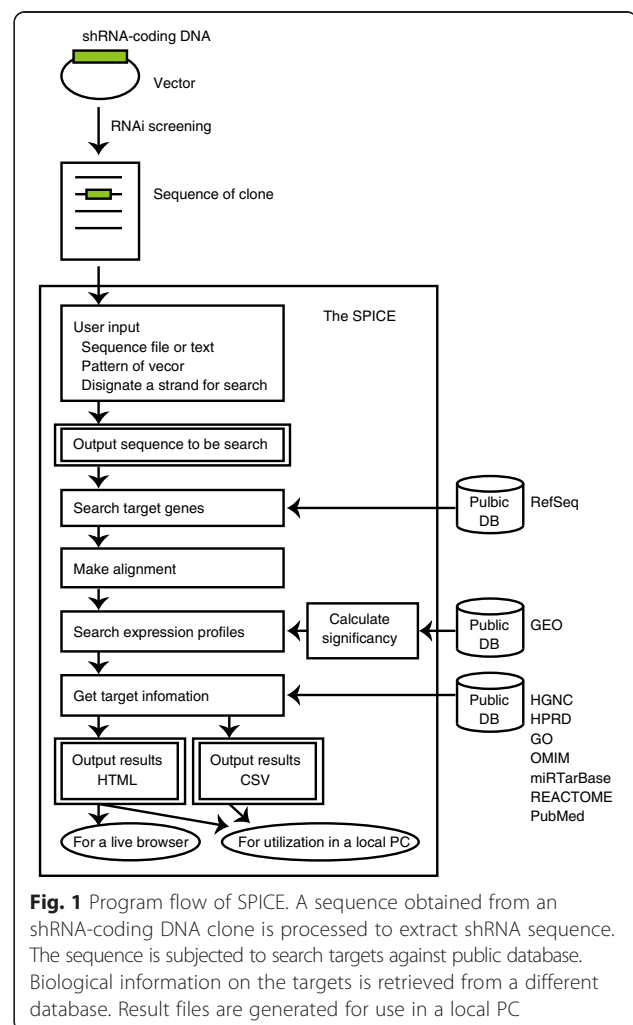
Because RNAi elicits a sequence-specific knockdown of gene expression, it is reasonable to associate the phenotype observed following siRNA-mediated knockdown with the biological functions of the target gene. Thus, most RNAi libraries were constructed using natural sequences specific to a known gene [14, 15] based on the original theory that siRNAs would specifically recognize the target mRNA without any mismatch between the target sequence and the guide strand of the siRNA. However, off-target silencing by siRNA occurs similar to that observed during silencing by micro RNA (miRNA) [16], suggesting non-assured specificity of siRNAs in the RNAi library. An siRNA would recognize a specific target gene, while also recognizing sequences of other genes with a few mismatches. Thus, several positions within a target gene might need to be further examined. On the other hand, some libraries were constructed using random oligonucleotides harboring artificial sequences that might include both specific and non-specific siRNAs to known genes [7, 17–20]. The main feature of these libraries is that every obtained shRNA needs to be subjected to sequence analysis to identify its target gene. If the siRNA sequence includes mismatches to a probable target gene, it would need further validation using additionally prepared shRNAs specific to the target gene. This is not as efficient as the other library; however, it offers the advantage that the library might have no bottleneck on the diversity of sequences because it was prepared using billions of siRNA sequences up to theoretical  $4^n$ , where  $n$  is the number of random oligonucleotides. In contrast to the general understanding of off-target silencing by siRNA [16], it is possible that the expression of a target gene with sequence mismatch can be specifically silenced by siRNAs. This might facilitate the use of reverse genetics methodology in genomics because construction of RNAi libraries is easy and inexpensive.

Sequence analysis of an obtained shRNA is an important process in the screening system for RNAi libraries generated from random oligonucleotides. Identification of a target gene might be simple for shRNAs carrying natural sequences. However, the identification might be difficult for siRNAs carrying non-specific sequences. A web application might reduce the laborious analysis of sequence databases. Although several bioinformatics tools are available in the public domain, utilizing each of

these tools separately for RNAi screening is not practically efficient. Here, we developed an automated web-based analysis and search tool, shRNA target prediction informed by comprehensive enquiry (SPICE), for investigating biological information about shRNA sequences. By integrating known bioinformatics tools [21–27] and additional processing of data for the efficient evaluation of sequence, SPICE displays target candidate genes with sequence alignment as well as information associated with each gene.

## 2 Web application

Our goal was to create a web application and provide a website that can support RNAi screening systems using random oligonucleotide RNAi libraries. To this end, the SPICE web application executes several tasks (Fig. 1): (i) identification of siRNA sequence region in vector harboring shRNA-encoding DNA, (ii) sequence alignment between passenger strand of the siRNA and human RefSeq DNA database, (iii) functional annotation of the siRNA target DNA using databases, (iv) calculation of



Gene Expression Omnibus (GEO) profile data to show significant microarray experiments in humans, and (v) preparation of downloadable summary files to support spreadsheet database construction in a local personal computer (PC). SPICE can mainly be utilized in RNAi screening using RNAi library that includes random oligonucleotide artificial sequences. It can also be used to investigate possible off-target candidates if the shRNA has a sequence specific to a gene. Case studies as examples of how to use SPICE are included in Additional file 1.

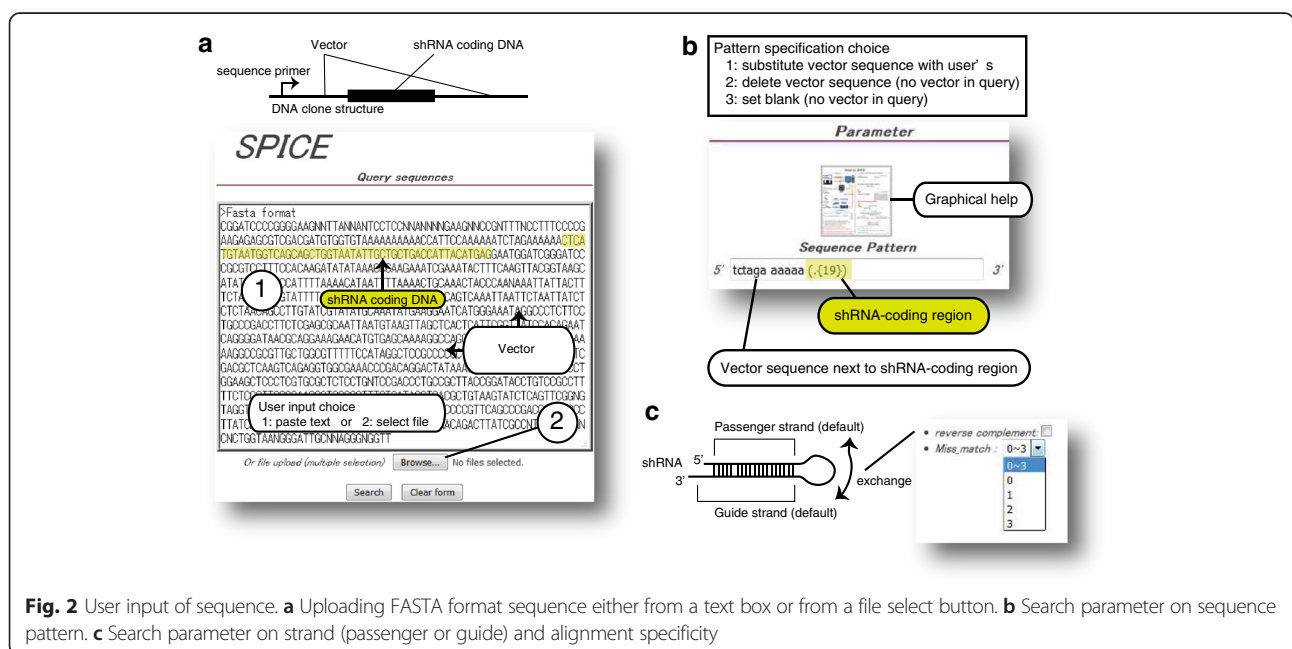
### 3 User input

In the first step, SPICE accepts either file upload or direct input by replacing the sample sequence that appears as default in a sequence box. For example, after obtaining an shRNA-coding DNA sequence from a sequencer, such as 3130xl Genetic Analyzer (Life Technologies/Applied Biosystems, Foster City, CA, USA), the deduced sequence in a FASTA file can be uploaded to the server using a file select button “Query sequences” (Fig. 2a). Alternatively, the FASTA sequence can be copied and pasted to the sequence box. The reverse direction of forward and reverse in the input sequence does not matter if the sequence is not modified by any other processing because shRNA-encoding DNA consists of inverted repeats [7]. Although the system supports vector sequence harboring shRNA-coding DNA, it should be noted that siRNA only sequence is acceptable by setting “Sequence parameter” either to “blank,” by which the whole input sequence will be used as query, or to “exactly (.{ 19 }),” by which the first 19 nt of input sequence will be queried (Fig. 2b). This is a pattern expression of input. The dot in the pattern means

any character such as A, T, G, and C except new line. The number between curly braces after the dot specifies the number of occurrences of the dot in the string. The parentheses group characters that were specified by the dot and curly braces. Thus, it is required to set “Sequence parameter” for the siRNA sequence within the input sequence by specifying vector sequence next to the siRNA sequence in the second step. For example, a default sample sequence **tatagaaaaa(.{ 19 }**) shows that an identical vector sequence **tatagaaaaa** is followed by a 19-base sequence of siRNA. The sample sequence **tatagaaaaa** might be replaced to other vector sequences of sufficient length. SPICE searches vector sequences and identifies an shRNA sequence in the input sequence. In the third step, additional options of “reverse complement” and “Miss\_match” can be specified (Fig. 2c). On checking “reverse complement,” the sequence will be queried as a guide strand of siRNA. Although the 5’ portion of shRNA might be a passenger strand in most cases [28], the probability is not 100 % [29]. Therefore, we made it possible to select whether the strand is a passenger or a guide. The “Miss\_match” option offers four kinds of search conditions, allowing indicated number of mismatches in alignment between query sequence and that in the database. The default “0–3” mismatch is searched in ascending order until hitting an alignment.

### 4 Identification of possible target genes by primary alignment of siRNA sequence and sequences in database

After execution, by clicking the search button, siRNA guide (antisense) and passenger (sense) strands are extracted



**Fig. 2** User input of sequence. **a** Uploading FASTA format sequence either from a text box or from a file select button. **b** Search parameter on sequence pattern. **c** Search parameter on strand (passenger or guide) and alignment specificity

from the sequence input. The strands are highlighted in the input sequence and listed with a number of target genes and mismatches, GC content, and a link (sequence name defined in “Query sequences”) to the detailed information window on the target genes (Fig. 3a). The information can be downloaded through “Download Result” for use in a local PC, as described in Fig. 3.

### 5 Prediction of siRNA targets

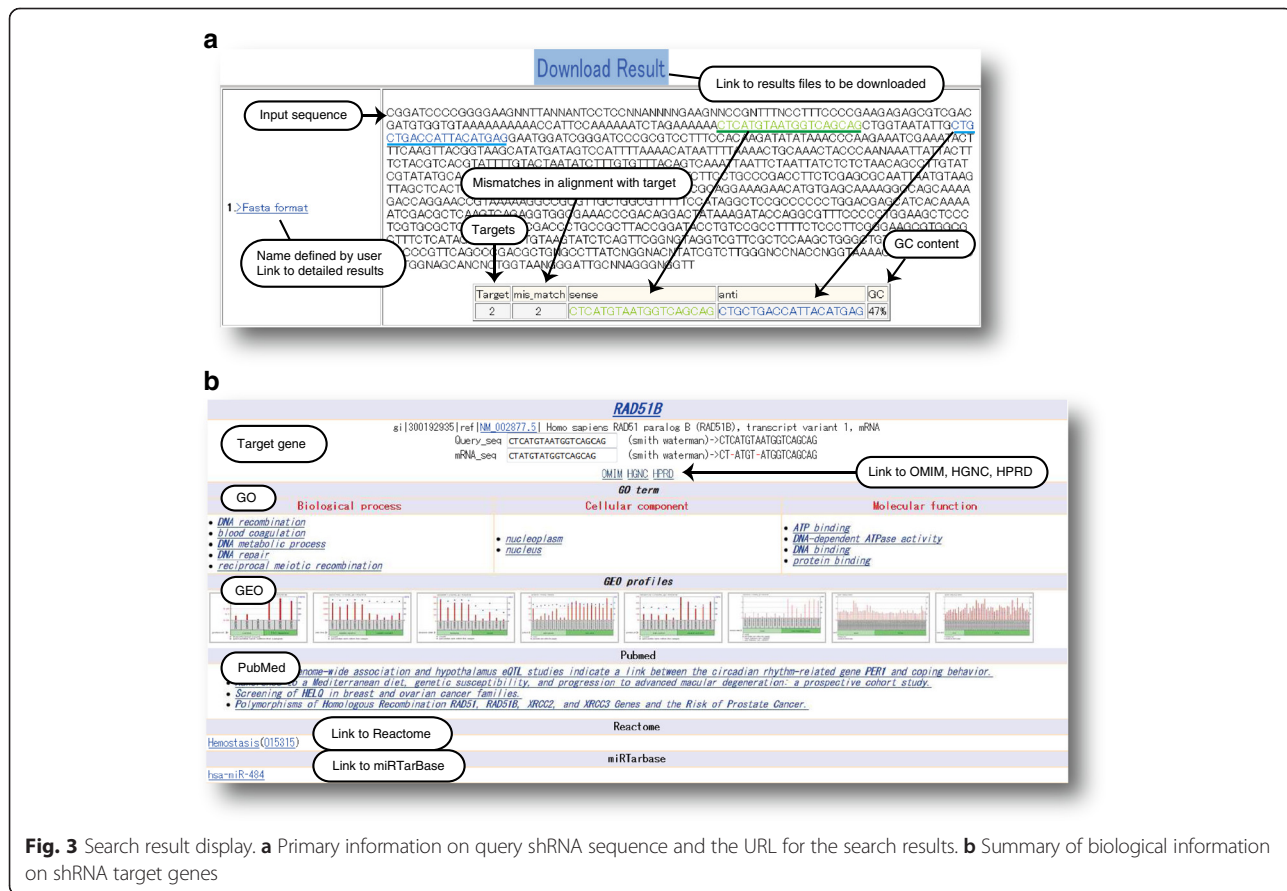
SPICE predicts target genes by performing GGGenome searches of siRNA sequences against sequences in the human RefSeq database using a parameter of mismatch [27]. GGGenome is an ultrafast search engine for nucleotide sequences and uses the Sedue software (Preferred Infrastructure, Japan) which is useful in handling short sequences. We limited human sequences to experimentally confirmed ones by using records both prefixed with “NM\_” and organisms “Homo sapiens.” SPICE selects and shows plus strands from the GGGenome search results because an input siRNA sequence is supposed to be a passenger strand (Fig. 3b). Next, alignment between the siRNA sequence and the selected strand is performed using the algorithm described by Smith and Waterman [30].

### 6 Displaying significant gene expression profiles

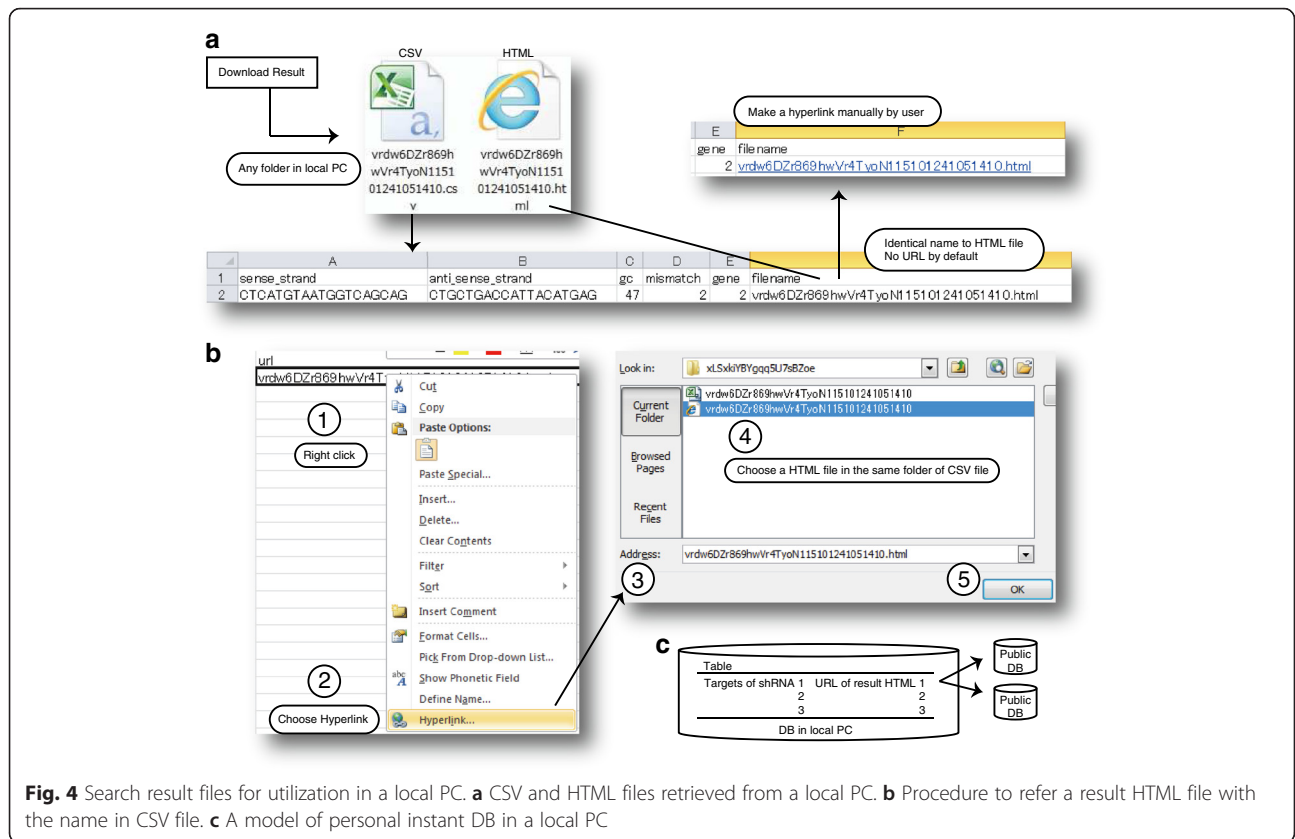
To display the expression of profiles of the predicted target genes, SPICE analyzes 361 kinds of selected DataSets of the GEO database [31]. Briefly, GEO contained 1335 kinds of human DataSets among 3413 kinds of whole DataSets. Then, 660 kinds of DataSets were extracted from human DataSets by searching descriptions that compared two experimental conditions with one experimental variable, which was indicated in the subset\_type descriptions. Next, 361 kinds of DataSets were chosen as DataSets having more than two samples in each condition. Marked differences between conditions in the expression of each gene in the selected GEO DataSets was previously evaluated using Welch’s *t* test ( $P < 0.01$ ). Therefore, the GEO profiles displayed in the box may exclusively list the novel expression of some subsets under the reported condition (Fig. 3b). Each cartoon of the GEO profiles has a URL for the original source data.

### 7 Links to other databases on siRNA targets

To obtain biological information about the siRNA targets, the name of the siRNA target was searched in each of the following databases: HUGO Gene Nomenclature Committee (HGNC) [32], Human Protein Reference Database (HPRD) [25], Gene Ontology (GO) [21], Online Mendelian



**Fig. 3** Search result display. **a** Primary information on query shRNA sequence and the URL for the search results. **b** Summary of biological information on shRNA target genes



**Fig. 4** Search result files for utilization in a local PC. **a** CSV and HTML files retrieved from a local PC. **b** Procedure to refer a result HTML file with the name in CSV file. **c** A model of personal instant DB in a local PC

Inheritance in Men (OMIM) [23], PubMed, miRTarBase [33], and REACTOME [34]. Links to these databases for each target are provided if there is any relation between the sequence in the database and the siRNA target (Fig. 3b).

### 8 Retrieval of search result files for use in a local PC

SPICE generates a downloadable compressed file (zip) that includes an HTML file showing the result and a

comma-separated value (CSV) table summarizing the siRNA profiles, e.g., sequence, GC contents, number of mismatches with the siRNA, and the name of the HTML file (Fig. 4). These files allow users to retain and utilize the search results in any directory/folder of a local PC. To prevent name redundancy of HTML files, the file is named by assembling 20 randomly chosen characters out of 62 different alphabets and numbers along with the time stamp. The HTML file will show results in any

**Table 1** Specificity of siRNA target prediction

shRNA (passenger)	Target	Reference	Target predicted by the SPICE <sup>a</sup>
GAUUUCCAAGAGGUUCU	<i>RPS6KA6</i>	Berns et al. [2]	<i>RPS6KA6</i> (1, 0)
UGGUUGAUGAGCCAAUGGA	<i>RPS6KA6</i>	Berns et al. [2]	<i>RPS6KA6</i> (1, 0)
GUACGGCCGUAGUCUCAAG	<i>HTATIP</i>	Berns et al. [2]	<i>KAT5</i> (1, 0): identical to <i>HTATIP</i>
CAAACGUCUGGAUGAAUGG	<i>HTATIP</i>	Berns et al. [2]	<i>KAT5</i> (1, 0): identical to <i>HTATIP</i>
GCCAUCCAGAUCCAGUUUC	<i>HDAC4</i>	Berns et al. [2]	<i>HDAC4</i> (1, 0)
GCAUGUGUUUCGCCUUGC	<i>HDAC4</i>	Berns et al. [2]	<i>HDAC4</i> (1, 0)
GCAGCAGAGGUGAUUCUGC	<i>CCT2</i>	Berns et al. [2]	<i>CCT2</i> (1, 0)
GAGAGGCGCUGUAGAGUUC	<i>CCT2</i>	Berns et al. [2]	<i>CCT2</i> (1, 0)
GAGUACAUCUGCCUUGCU	<i>KIAA0828</i>	Berns et al. [2]	<i>AHCYL2</i> (1, 0): identical to <i>KIAA0828</i>
GCUACAUCAAGAACUCAGC	<i>KIAA0828</i>	Berns et al. [2]	<i>AHCYL2</i> (1, 0): identical to <i>KIAA0828</i>
GAGACCTACCTCCGGATCA	<i>PLK1</i>	Jackson et al. [16]	<i>PLK1</i> (1, 0)

<sup>a</sup>Numbers in parentheses show (target genes, mismatches in alignment)

**Table 2** Number of target candidate genes against shRNA constructed using random oligonucleotides

Sequences of shRNA-coding DNA	Number of aligned genes with mismatches			
	0 <sup>a</sup>	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>
GGTAGCTCGTTAATTTAAT	0	0	1	3
GAATTCGCGGTACGGGGG	0	0	0	6
GAATGCAGCAGGAGGGAGG	0	1	8	278
GAATCGGACGCAGGCCTAT	0	0	2	53
GACTAGTCTACTTCCTTAG	0	0	0	31
GCGATCTCTACCACCTGGC	0	0	4	79
GCGTAATCGGGCGAATAAC	0	0	0	0
GATAGTGCACGTTGAAACA	0	0	1	38
GTAGGGGACCTAGAACAT	0	0	2	20
GACAAACCTGCCTAGTTA	0	0	0	17
GGATCCGAGTGCACCTAAA	0	0	0	28
GAAGTAATGCGCTCGATCC	0	0	1	13
GCCACGTATGGCCGTA CTG	0	0	0	10
GCACAATGTAAAAAGAATA	0	2	5	131
GAGCCAGTAAAAGAATATC	0	1	4	55
GTGCACTACTCGCAGTGGG	0	0	2	34
GTTCTAGACTTCAATAA	0	0	2	53
GACCTACCATGTAGCAAG	0	0	0	43
GAGAAGGCGTCTCCGAGA	0	0	4	82
GAGGAATGATATGGCCTTG	0	0	1	54
GGGGCAGTGGCAGGTGATC	0	1	7	140
GACCAGGGACATTAGAT	0	0	1	37
GATCTAAATATGGCAGGCA	0	0	0	44
GAAGACAGACCGGAGATC	0	0	7	112
GTCAGCCAGCAAAGAAGTG	0	0	7	115
GTGGACTAGTATCGATATA	0	0	1	10
GTGAACCAAGTTACTAGCT	0	0	0	39
GGCCCGGTTAGGAACACTT	0	0	0	7
GGCCAAGAAGATACACTAG	0	0	1	65
GGACAGGGCAACAATGTGG	0	0	1	44
GGAAGATATGCTCATTCCG	0	0	0	24
GGAAGTGGTAAGGACTCCA	0	0	3	65
GGAGTCAGTACAGATACGA	0	0	1	22
GCGTTAATTAATCAACATC	0	0	0	12
GGCGTGATCGTATGCACCA	0	0	0	13
GAGAAGTCCAGATCTTTCC	0	0	5	95
GGACCTGTCCCTAGAAGT	0	0	2	72
GAAAACGTGCAGACCCGAG	0	0	1	46
GCATTCCAAAGAAAAGAAT	0	0	7	167
GTCACATATCCTGGACACG	0	0	1	36
GTGTATAGTGAAGACAAT	0	0	0	48
GTGAGAACGTTGGAGAGCC	0	0	1	59

**Table 2** Number of target candidate genes against shRNA constructed using random oligonucleotides (Continued)

GTTTTGATCGGAGAATCCA	0	0	1	31
GCAAACATCCAGAAAGTAA	0	0	1	93
GTCCAGGGGTTGGGGATCG	0	0	1	34
GGGCGTAGGTCAAAGCATG	0	0	0	17
GGGCGGGGTAAAGAAACCA	0	0	0	17

Alignment between shRNA and human RefSeq was performed under conditions in which the indicated number of mismatches was allowed  
<sup>a</sup>Represents number of mismatched nucleotides

browser without searching again. Because text links to databases are active, original descriptions in the database can be referred. Remarkably, the resultant spreadsheet can be used as a front page of siRNA information by manually hyperlinking an HTML file name in the sheet onto a corresponding HTML file placed in the same folder. This can be easily accomplished using a basic function in a spreadsheet software package. For example, a hyperlink can be made using MS-EXCEL (Microsoft, CA, USA) as follows. (1) Locate the HTML file name in a table. (2) Choose a command “hyperlink...” from the “Insert” tab. (3) Choose the HTML file in the hyperlink insertion pop-up window. (4) Save the EXCEL file. The resultant table should have the HTML file name with the URL. This also facilitates the building of an instant database by combining multiple tables in a single file. Additionally, other comments on siRNA can be included by making new columns in the table.

**9 Evaluation of a web application**

The number of siRNA target candidates was compared between SPICE and other sequence search engines. A representative result for the shRNA GAUUAUCCAAAGAGGUUCU (passenger strand) targeting *RPS6KA6* gene [2] was used. SPICE showed only one target when executed without a mismatch. GGGenome search yielded five candidates including the target with no mismatch. The rest of the candidates were predicted genes that were indexed with “XM\_”, which indicate the sequence was predicted as gene by RefSeq. BLAST search showed 100 candidates including the target. Five genes had no mismatch in alignment. The rest of the candidates included one to six mismatches in alignment. Similar results were obtained for another shRNA UGGUUGAUGAGCCAAUGGA (passenger strand) targeting *RPS6KA6* gene [2]. Thus, all of the above applications listed the siRNA target. Of note, SPICE is sufficient for target prediction. Next, we investigated the specificity of the target prediction by using experimentally validated shRNA sequences (Table 1). SPICE showed the identical single target for each sequence, suggesting high specificity of target prediction.

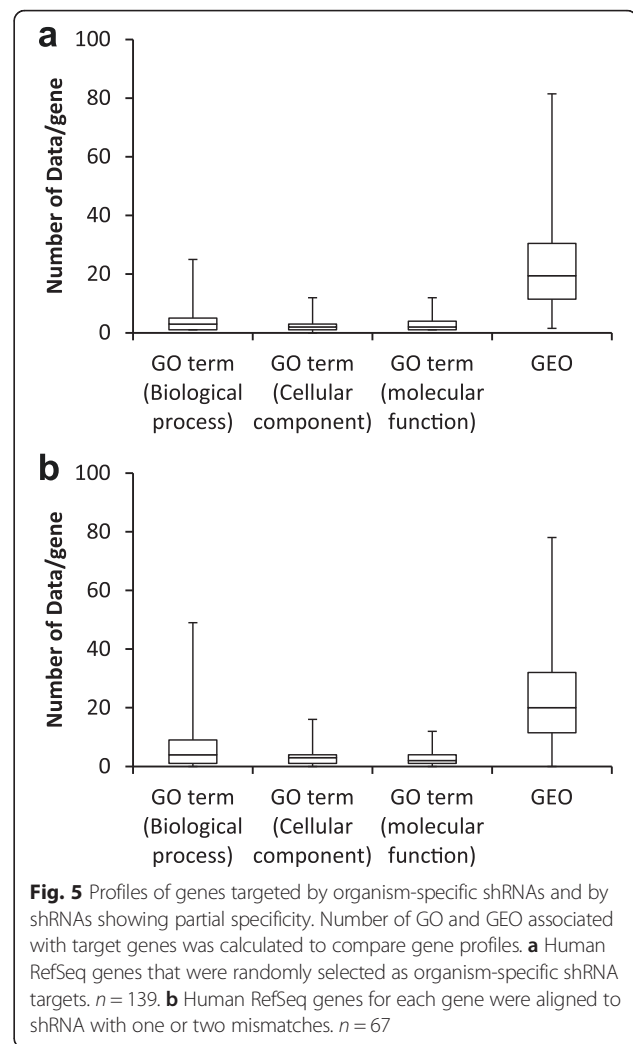
Next, we investigated gene expression profiles obtained using GEO. For example, there were 5245 kinds of GEO profiles on *RPS6KA6* gene in the current GEO database. The number was decreased by 3577 using a filter “Organism human.” Additionally, the number was decreased by 42 using a filter “Differential expression Up/down genes.” On the other hand, SPICE displayed 16 profiles that were confirmed manually using the original values in GEO profile data. We found that there was no overlap between the results, suggesting different sensitivities for the selections. Not surprisingly, the differential expression profile shown by SPICE might be only a part of the complete expression profile of the targeted gene.

Estimated time for receiving search results was 6 to 12 s per siRNA target gene. The time depends on how many targets an siRNA sequence has in the database. Because SPICE first searches targets with no mismatch and continues the process with mismatches until it finds a target, the number of target genes increased when searching with mismatches. It took approximately 10.5 min to search 76 target genes for an siRNA sequence.

### 10 Evaluation of random shRNA library using a web application

SPICE was developed for searching targets of shRNA obtained using random oligonucleotides. However, the characteristics of the shRNA sequence were not analyzed thoroughly. It is not clear how many shRNA clones from the RNAi library are sufficient to investigate all human genes. Therefore, we analyzed 47 clones obtained from an RNAi library constructed using random oligonucleotides (Table 2). Each sequence shows the DNA encoding the passenger strand of the shRNA. Interestingly, 19-nt sequences showed no perfect alignment with sequences in human RefSeq database (Table 2). By allowing a mismatch in the alignment, target genes increased from zero to four. Most of the sequences needed two to three mismatches to find targets. These results suggested that to obtain perfectly matched shRNA to any gene during RNAi screening, 47 times the number of shRNA clones against human genes might not be sufficient to cover all human genes.

Because most shRNA sequences in an RNAi library constructed using random oligonucleotides are not specific to the sequences in an organism, as described above, it is not assured that these shRNAs would target a series of genes as organism-specific shRNAs would. To investigate the similarity, we compared the profiles of genes targeted by organism-specific shRNAs with those of genes targeted by shRNAs derived from the RNAi library. We used 139 randomly selected human RefSeq sequences as representative targets of organism-specific shRNA. shRNAs from the RNAi library (Table 2) were



used as non-organism-specific shRNA. Genes in the human RefSeq database included one to two GO terms (median) (Fig. 5a). Approximately 18 GEO profiles were associated with a gene whose expression was significantly different among subsets. Comparably, shRNA targets shown in Table 2 showed a similar distribution (Fig. 5b). Thus, targets of shRNAs randomly derived from an RNAi library are not different from those of organism-specific shRNAs.

### 11 Conclusions

We have developed SPICE and provided the website for supporting RNAi screening systems using random oligonucleotide RNAi libraries. The SPICE web application can show siRNA target DNA with sequence alignment and the functional annotation. It also provides the downloadable summary files for database construction in local PC. SPICE can be used to facilitate sequence analysis of siRNAs carrying non-specific sequences to

natural sequences that will be obtained in RNAi screening.

## 12 Additional file

**Additional file 1: Search examples.** Execution using siRNA sequence or a sequence file form sequencer. (DOCX 464 kb)

### Competing interests

The authors declare that they have no competing interests.

### Compliance with ethical standard

No ethical approval was required for this work.

### Acknowledgements

This study was supported by the Takeda Science Foundation.

### Author details

<sup>1</sup>Graduate School of Bionics, Tokyo University of Technology, 1401-1 Katakura-machi, Hachioji, Tokyo 192-0982, Japan. <sup>2</sup>RNAi Inc., 4-1-4 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

Received: 1 September 2015 Accepted: 9 February 2016

Published online: 19 February 2016

### References

- RS Kamath, AG Fraser, Y Dong, G Poulin, R Durbin, M Gotta, A Kanapin, N Le Bot, S Moreno, M Sohrmann, DP Welchman, P Zipperlin, J Ahringer, Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**(6920), 231–237 (2003). doi:10.1038/nature01278
- K Berns, EM Hijmans, J Mullenders, TR Brummelkamp, A Velds, M Heimerikx, RM Kerkhoven, M Madiredjo, W Nijkamp, B Weigelt, R Agami, W Ge, G Cavet, PS Linsley, RL Beijersbergen, R Bernards, A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**(6981), 431–437 (2004). doi:10.1038/nature02371
- PW Reddien, AL Bermange, KJ Muffitt, JR Jennings, A Sanchez Alvarado, Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Developmental Cell* **8**(5), 635–649 (2005). doi:10.1016/j.devcel.2005.02.014
- Y Shiobara, C Harada, T Shiota, K Sakamoto, K Kita, S Tanaka, K Tabata, K Sekie, Y Yamamoto, T Sugiyama, Knockdown of the coenzyme Q synthesis gene *Smed-dlp1* affects planarian regeneration and tissue homeostasis. *Redox Biology* **6**, 599–606 (2015). doi:10.1016/j.redox.2015.10.004
- SM Elbashir, J Harborth, W Lendeckel, A Yalcin, K Weber, T Tuschl, Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**(6836), 494–498 (2001). doi:10.1038/35078107
- G Sen, TS Wehrman, JW Myers, HM Blau, Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nature Genetics* **36**(2), 183–189 (2004). doi:10.1038/ng1288
- Y Nishikawa, T Sugiyama, A shRNA library constructed through the generation of loop-stem-loop DNA. *The Journal of Gene Medicine* **12**(11), 927–933 (2010). doi:10.1002/jgm.1513
- RS Oh, WC Pan, A Yalcin, H Zhang, TR Guilarte, GS Hotamisligil, DC Christiani, Q Lu, Functional RNA interference (RNAi) screen identifies system A neutral amino acid transporter 2 (SNAT2) as a mediator of arsenic-induced endoplasmic reticulum stress. *Journal of Biological Chemistry* **287**(8), 6025–6034 (2012). doi:10.1074/jbc.M111.311217
- K Ui-Tei, Y Naito, F Takahashi, T Haraguchi, H Ohki-Hamazaki, A Juni, R Ueda, K Saigo, Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Research* **32**(3), 936–948 (2004). doi:10.1093/nar/gkh247
- SM Elbashir, J Harborth, K Weber, T Tuschl, Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* **26**(2), 199–213 (2002). doi:10.1016/S1046-2023(02)00023-3
- Z Arziman, T Horn, M Boutros, E-RNAi: a web application to design optimized RNAi constructs. *Nucleic Acids Research* **33**(Web Server issue), W582–588 (2005)
- Y Naito, J Yoshimura, S Morishita, K Ui-Tei, siDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics* **10**, 392 (2009). doi:10.1186/1471-2105-10-392
- L Li, X Lin, A Khvorova, SW Fesik, Y Shen, Defining the optimal parameters for hairpin-based knockdown constructs. *RNA* **13**(10), 1765–1774 (2007). doi:10.1261/rna.599107
- J Luo, MJ Emanuele, D Li, CJ Creighton, MR Schlabach, TF Westbrook, K-K Wong, SJ Elledge, A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**(5), 835–848 (2009). http://dx.doi.org/10.1016/j.cell.2009.05.006
- L Zhao, Y Pan, Y Gang, H Wang, H Jin, J Tie, L Xia, Y Zhang, L He, L Yao, T Qiao, T Li, Z Liu, D Fan, Identification of *GAS1* as an epirubicin resistance-related gene in human gastric cancer cells with a partially randomized small interfering RNA library. *Journal of Biological Chemistry* **284**(39), 26273–26285 (2009). doi:10.1074/jbc.M109.028068
- AL Jackson, J Burchard, J Schelter, BN Chau, M Cleary, L Lim, PS Linsley, Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. *RNA* **12**(7), 1179–1187 (2006)
- Q Guo, Y Kong, L Fu, T Yu, J Xu, W Chen, A randomized lentivirus shRNA library construction. *Biochemical and Biophysical Research Communications* **358**(1), 272–276 (2007). doi:10.1016/j.bbrc.2007.04.123
- Wu H, Dinh A, Mo YY (2007) Generation of shRNAs from randomized oligonucleotides. *Biol Proced Online* 9:9-17. doi:doi: 10.1251/bpo129
- Y Wang, YE Wang, MG Coticelli, RB Wilson, A random shRNA-encoding library for phenotypic selection and hit-optimization. *PLoS One* **3**(9), e3171 (2008). doi:10.1371/journal.pone.0003171
- M Nichols, RA Steinman, A recombinase-based palindrome generator capable of producing randomized shRNA libraries. *Journal of Biotechnology* **143**(2), 79–84 (2009). doi:10.1016/j.jbiotec.2009.06.010
- M Ashburner, CA Ball, JA Blake, D Botstein, H Butler, JM Cherry, AP Davis, K Dolinski, SS Dwight, JT Eppig, MA Harris, DP Hill, L Issel-Tarver, A Kasarskis, S Lewis, JC Matese, JE Richardson, M Ringwald, GM Rubin, G Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25**(1), 25–29 (2000). doi:10.1038/75556
- T Barrett, DB Troup, SE Wilhite, P Ledoux, C Evangelista, IF Kim, M Tomashevsky, KA Marshall, KH Phillippy, PM Sherman, RN Muetter, M Holko, O Ayanbule, A Yefanov, A Soboleva, NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Research* **39**(Database issue), D1005–1010 (2011). doi:10.1093/nar/gkq1184
- A Hamosh, AF Scott, JS Amberger, CA Bocchini, VA McKusick, Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research* **33**(Database issue), D514–517 (2005). doi:10.1093/nar/gki033
- SD Hsu, FM Lin, WY Wu, C Liang, WC Huang, WL Chan, WT Tsai, GZ Chen, CJ Lee, CM Chiu, CH Chien, MC Wu, CY Huang, AP Tsou, HD Huang, miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Research* **39**(Database issue), D163–169 (2011). doi:10.1093/nar/gkq1107
- TS Prasad, K Kandasamy, A Pandey, Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods in Molecular Biology* **577**, 67–79 (2009). doi:10.1007/978-1-60761-232-2\_6
- KA Gray, LC Daugherty, SM Gordon, RL Seal, MW Wright, EA Bruford, Genenames.org: the HGNC resources in 2013. *Nucleic Acids Research* **41**(Database issue), D545–552 (2013). doi:10.1093/nar/gks1066
- Y Naito, H Bono, GGRNA: an ultrafast, transcript-oriented search engine for genes and transcripts. *Nucleic Acids Research* **40**(Web Server issue), W592–596 (2012). doi:10.1093/nar/gks448
- A Khvorova, A Reynolds, SD Jayasena, Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**(2), 209–216 (2003). doi:10.1016/S0092-8674(03)00801-8
- JX Wei, J Yang, JF Sun, LT Jia, Y Zhang, HZ Zhang, X Li, YL Meng, LB Yao, AG Yang, Both strands of siRNA have potential to guide posttranscriptional gene silencing in mammalian cells. *PLoS One* **4**(4), e3382 (2009). doi:10.1371/journal.pone.0005382
- TF Smith, MS Waterman, Identification of common molecular subsequences. *Journal of Molecular Biology* **147**(1), 195–197 (1981). doi:10.1016/0022-2836(81)90087-5
- T Barrett, SE Wilhite, P Ledoux, C Evangelista, IF Kim, M Tomashevsky, KA Marshall, KH Phillippy, PM Sherman, M Holko, A Yefanov, H Lee, N Zhang, CL Robertson, N Serova, S Davis, A Soboleva, NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* **41**(Database issue), D991–995 (2013). doi:10.1093/nar/gks1193



32. KA Gray, B Yates, RL Seal, MW Wright, EA Bruford, Genenames.org: the HGNC resources in 2015. *Nucleic Acids Research* **43**(Database issue), D1079–1085 (2015). doi:10.1093/nar/gku1071
33. SD Hsu, YT Tseng, S Shrestha, YL Lin, A Khaleel, CH Chou, CF Chu, HY Huang, CM Lin, SY Ho, TY Jian, FM Lin, TH Chang, SL Weng, KW Liao, IE Liao, CC Liu, HD Huang, miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Research* **42**(Database issue), D78–85 (2014). doi:10.1093/nar/gkt1266
34. D Croft, AF Mundo, R Haw, M Milacic, J Weiser, G Wu, M Caudy, P Garapati, M Gillespie, MR Kamdar, B Jassal, S Jupe, L Matthews, B May, S Palatnik, K Rothfels, V Shamovsky, H Song, M Williams, E Birney, H Hermjakob, L Stein, P D'Eustachio, The Reactome pathway knowledgebase. *Nucleic Acids Research* **42**(1), D472–477 (2014). doi:10.1093/nar/gkt1102

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---