Nucleic Acids Research, 2006, Vol. 34, Database issue ***D153–D157***
doi:10.1093/nar/gkj051

RNAi Codex: a portal/database for short-hairpin RNA (shRNA) gene-silencing constructs

A. Olson, N. Sheth, J. S. Lee, G. Hannon and R. Sachidanandam*

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA

Received August 11, 2005; Revised and Accepted October 4, 2005

ABSTRACT

Use of RNA interference (RNAi) in forward genetic screens is proliferating. Currently, short-interfering RNAs (siRNAs) and short-hairpin RNAs (shRNAs) are being used to silence genes to tease out functional information. It is becoming easier to harness RNAi to silence specific genes, owing to the development of libraries of readymade shRNA and siRNA gene-silencing constructs by using a variety of sources. RNAi Codex, which consists of a database of shRNA related information and an associated web-site, has been developed as a portal for publicly available shRNA resources and is accessible at http://codex.cshl.org. RNAi Codex currently holds data from the Hannon–Elledge shRNA library and allows the use of biologist-friendly gene names to access information on shRNA constructs that can silence the gene of interest. It is designed to hold user-contributed annotations and publications for each construct, as and when such data become available. We will describe features of RNAi Codex and explain the use of the tool.

INTRODUCTION

RNAi, or RNA interference, is the disruption of the expression of a gene by a double-stranded RNA (dsRNA), in which one strand is complementary (either perfectly or imperfectly) to a section of the gene's mRNA (1). A dsRNA can enter the cytoplasm, through the expression of a hairpin (or inverted repeats), through viral gene expression or through artificial

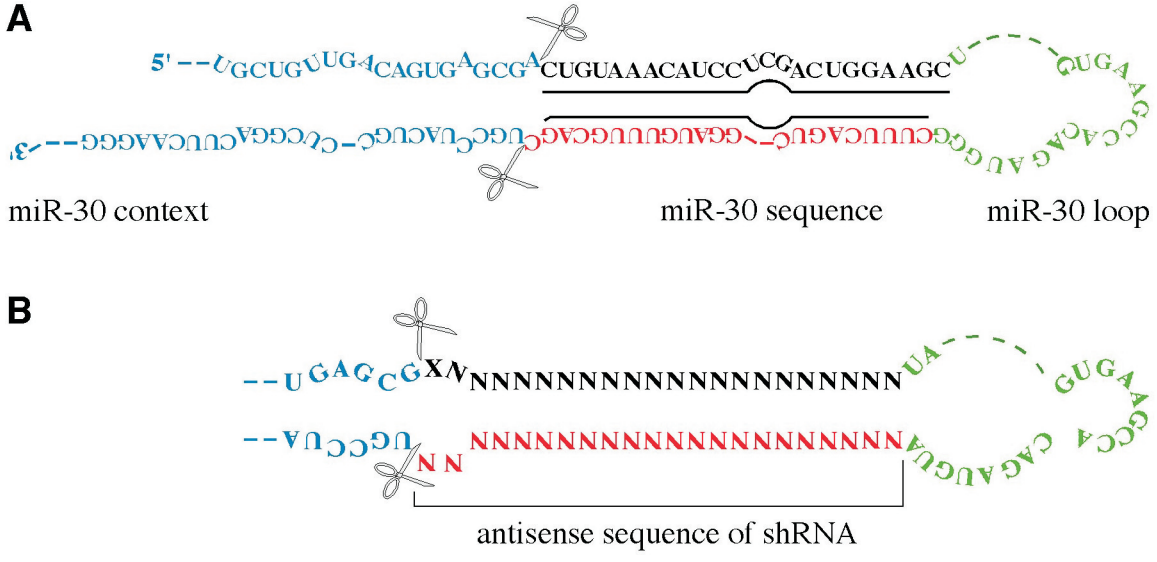


Figure 1. miR-30 based shRNA design (12). The figure shows the architecture of the constructs that are currently in RNAi Codex. The upper hairpin (**A**) is the primary transcript of the miR-30 miRNA. The sense and antisense strands are underlined. The lower hairpin (**B**) is the shRNA designed within the miR-30 context. The N's show the position of the sense and antisense strands on the hairpin. The figure has been adapted from the Open Biosystems' website (http://www.openbiosystems.com).

*To whom correspondence should be addressed. Tel: +1 516 367 8864; Fax: +1 516 367 8389; Email: sachidan@cshl.org

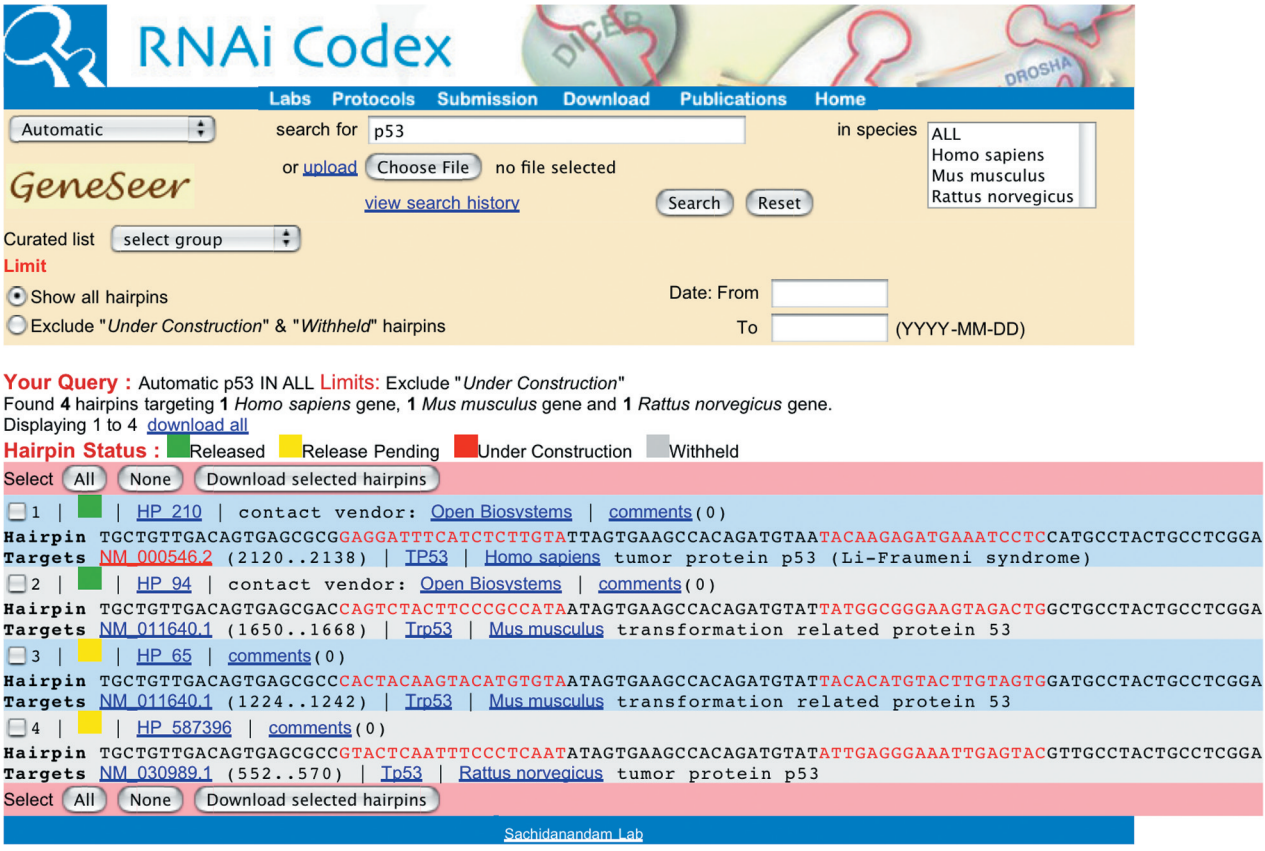


Figure 2. Results of a search for p53. This page of results shows all the hairpins in the database that target the p53 gene in the human, mouse and rat genomes. Color codes are used to show if the constructs are released and available (green), in the process of being released (yellow), under construction (red) or withdrawn (grey). For each hairpin, the actual sequence is shown, with the sequence of the sense and antisense strands highlighted in red. The name of the mRNA sequence is linked to resources at NCBI (http://www.ncbi.nlm.nih.gov). There is a direct link to the vendor's order page (Open Biosystems in the cases shown here), which can be used to purchase the hairpin. The *download all* link allows downloading all the search results into a csv file, which can be opened in spreadsheet programs. The user can also download specific hairpins by checking the check boxes and using the *Download selected hairpins* button. The *Comments* link shows user-supplied comments as well as publications that have referenced the construct. Clicking on the hairpin name takes the user to a view that is shown in Figure 3. The hairpin could be designed against a different target gene, it will appear in the results as long as it can target the gene of interest. The search bar in the top of the figure can be used for additional searches, which can be limited by conditions such as organisms, state of hairpins, and so on. Files containing search terms, which can be symbols, definitions, names (HUGO specified names) or GO Ids, can be uploaded to the website, to search for hairpins that target the relevant genes. The *search history* button can be used to retrieve old search results as well as combine results from two different searches using the logical operations AND, OR, NOT or XOR. The links on the top of the page take the user to protocols from the laboratories whose constructs are in the database.

constructs that enter the cell via the cell membrane. The disruption can take the form of mRNA degradation, translational repression or transcriptional repression through epigenetic modifications (2–5).

The introduction of large dsRNA into mammalian cells results in a general response (interferon or protein kinase PKR response) that leads to cell death (6). It was discovered that shorter dsRNA (<29 nt) can be used to bypass this response (7). Short-interfering RNAs (siRNAs) are short dsRNA with 2 nt 3′ overhangs and a 5′ phosphate group that mimic the product of Dicer activity. They can get incorporated directly into the RNAi silencing complex (RISC) resulting in silencing activity (8). This is a popular method of silencing genes in cells.

Another method of inducing RNAi is to insert hairpin constructs into the genome using vectors, which can then be stably expressed (9). The expressed hairpins are processed by Drosha and exported to the cytoplasm, where Dicer acts on them to create siRNAs, which then get incorporated into the RISC. These constructs are called short-hairpin RNAs (shRNAs) (9). shRNAs can also be chemically synthesized and introduced into the cytoplasm (10,11), but in this case it is important to mimic the product of Drosha, which has a 2 nt 3′ overhang. It is also possible to place the antisense strand in the context of a known microRNA (miRNA) hairpin. miRNAs are naturally occurring genes that play a role in switching genes on and off during development (2). The Hannon–Elledge library of shRNA constructs uses the context of the miR-30 miRNA, as shown in Figure 1 (12).

Both siRNAs and shRNAs allow gene silencing and operate through the same pathways. The design principles involved in both are similar, in terms of ensuring that the appropriate strand from the dsRNA gets incorporated in the RISC (13,14). Both can result in off-target effects, in which genes that share partial homology with either strand of the dsRNA get silenced (15,16). Unfortunately, it is difficult to make accurate quantitative predictions of these effects (17). Thus, annotating the shRNA constructs with functional information

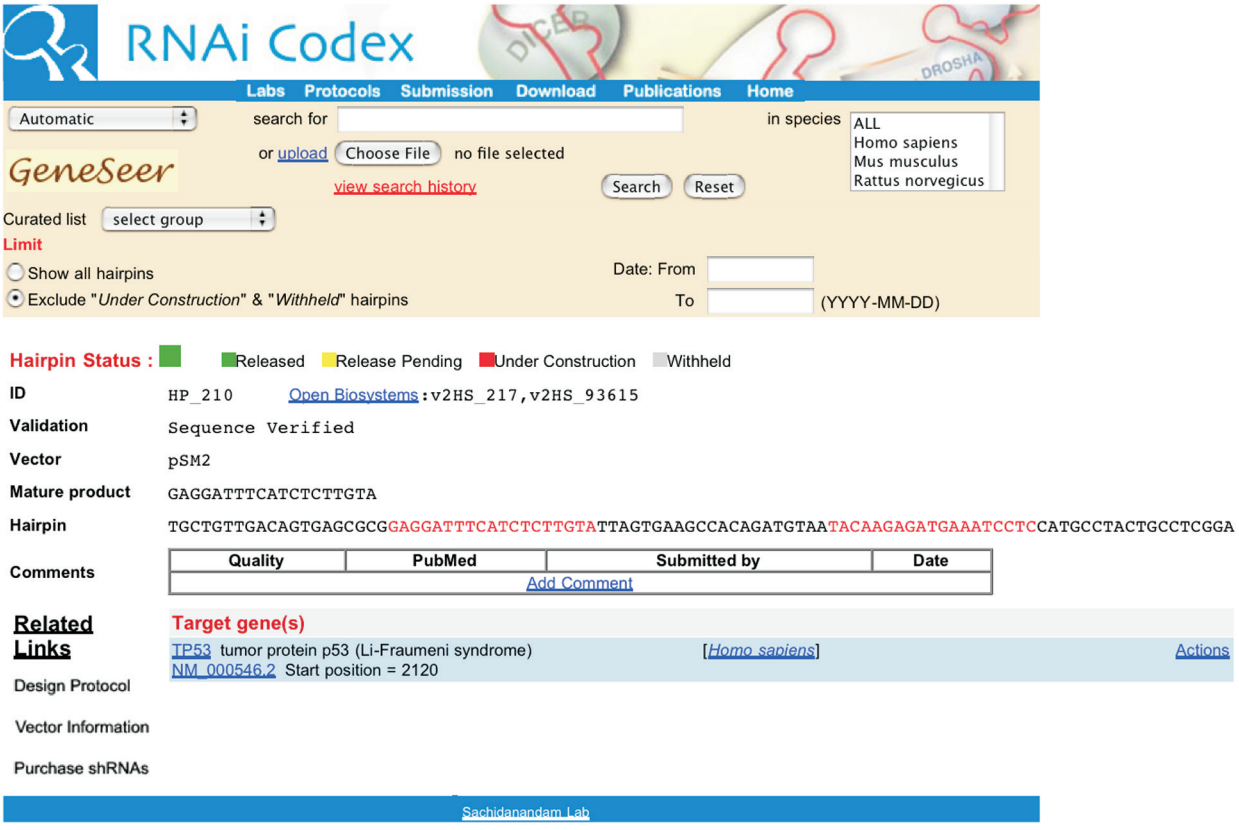


Figure 3. This page contains information on each hairpin. It can be reached from the page shown in Figure 2, by clicking on the hairpin link. This page shows comments and publications and other information regarding the hairpin construct. It also has links to protocols, vectors and to vendors. The *Actions* link pops-up a window that allows accessing other information such as homologs of the target gene in other species through the *find homologs* link and the *visualize* link allows visualizing the alignment of the constructs to the genomic region along with mRNA and expressed sequence tags in that region. The result of clicking on the *visualize* link is shown in Figure 4. A registered user can use the *Add Comment* link to annotate the hairpin with comments and publications. New users can register by clicking on the *Add Comment* link. Comments can only be selected from a controlled vocabulary so that it is machine-readable and allows statistical analysis of the dataset. Publications that reference a construct can also be added to the comments using uids from PubMed (http://www.ncbi.nlm.nih.gov). The controlled vocabulary will be expanded, based on user-feedback.

is useful as there is no reliable method that *a priori* predicts the performance of the shRNA construct under actual biological conditions.

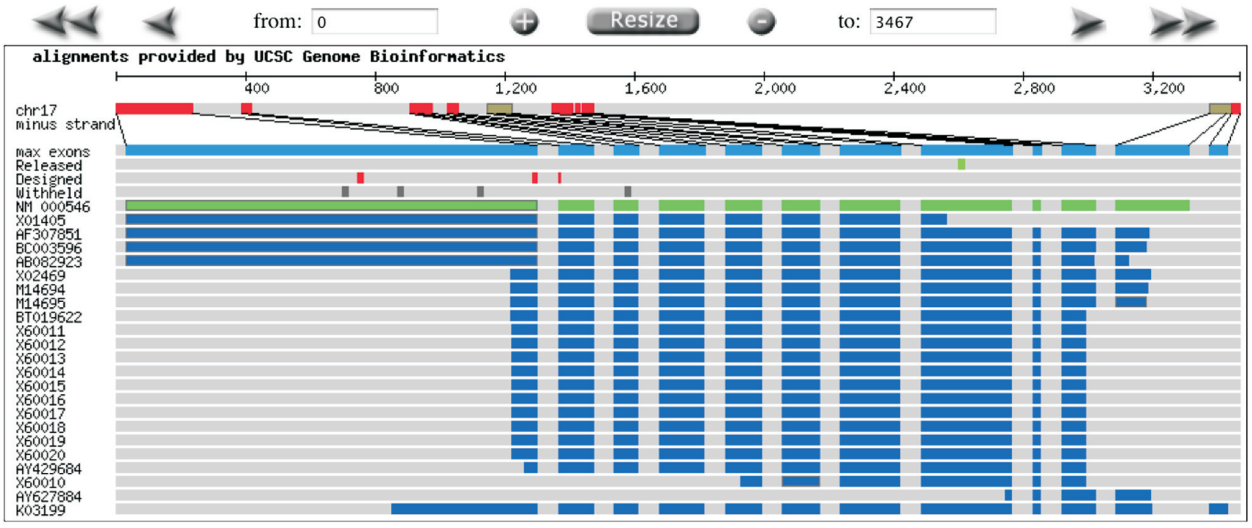
A central repository of shRNA constructs is essential since such a resource can act as a clearinghouse that can track results, identify patterns in shRNA performance and allow users to locate constructs from a variety of sources. RNAi Codex (http://codex.cshl.org) fulfills this role, though, at present, there is scant published information on the performance of specific shRNA constructs in the public domain. Our website and the associated database enable users to locate constructs from these libraries and purchase them from commercial vendors. We will explain our resource and give detailed instructions on the use of this tool.

MATERIALS AND METHODS

We built a database of shRNA constructs from the Elledge–Hannon collection (18). There are other collections (19), but these are not yet in the public domain. Each construct has associated with it several pieces of information such as the gene, the target sequence on the gene and the actual sequence of the construct. The database holds all this information. In addition, the database can also accept annotations of constructs by using a controlled vocabulary to log experiences from experiments as well as links to publications that reference the construct.

A problem with such databases is that it is difficult to locate appropriate constructs using names that might not be familiar to the database. To solve this problem, we previously built an extensive name translation service, GeneSeer (http://geneseer.cshl.org) (20), which allows the use of familiar names to identify corresponding silencing constructs. In addition, the system also allows searching for constructs using sequences. Functional groups of genes, such as *kinases*, *phosphatases*, *cancer-1000* and so on, have been annotated with the help of expert curators. This allows the creation of collections of shRNA constructs (mini-libraries) that can silence functional groups. The system is familiar with Gene Ontology (21) terms, user-friendly names such as p53 and names from other databases such as Swiss-Prot/TrEMBL (22) and HUGO (23).

The RNAi Codex checks the mature sequence of the hairpins against the gene of interest, and returns all the hairpins that can target the gene, even if they were initially designed against a different gene. This is a very useful feature since



highlighted regions correspond to features in track: Released, Withheld, NM_000546, Designed

GCACTGGCGT TCACCCCTCA GACACACAGG TGGCAGCAAA GTTTTATTGT AAAATAAGAG ATCGATATAA AAATGGGATA
TAAAAAGGGA GAAGGAGGGG AAGGGTGGGG TGAAAATGCA GATGTGCTTG CAGAATGTAA AAGATGTTGA CCCTTCCAGC
TGGACGTGGT GGCTCACAAT TGTAATCCCA GCACTCTGGG AGGCTGAGAC AGGTGGATCG CCTGAGCCCA GGAGTTTGAG
ACCAGCCTGG GCAACACTGT GAGACCCCAT CTCTACAAAA CATGCAAAAG TTGGCTGGCC ATGGTGGCAT GAACCTGTGG
TCCCAGCTAC TCCGGAGGCT GAGGCAGGAC TGCTCGAGCC GGGGAGGCAA AGGCTGCAGT AAGCCAAGAT CACGCCACTC
CACTCCAGCC TGGGCAACAA AGCGAGACCC AGTCTCAAAG AAAAAGAAAA AAAAAAAAAA AAAAGAAAAA AGAAATTGAC
CCTGAGCATA AAACAAGTCT TGGTGGATCC AGATCATCAT ATACAAGAGA TGAAATCCTC CAGGGTGTGG GATGGGGTGA
GATTTCCTTT TAGGTACTAA GGTTCACCAA GAGGTTGTCA GACAGGGTTT GGCTGGGCCA GCAGAGACTT GACAACTCCC
TCTACCTAAC CAGCTGCCCA ACTGTAGAAA CTACCAACCC ACCGACCAAC AGGGAGAGGG AACAAGCACC CTCAAGGGGG
TCAAGTTCTA GACCCCATGT AATAAAAGGT GGTTTCAAGG CCAGATGTAC ATTATTTCAT TAACCCTCAC AATGCACTCT
GTGAGGTAGG TGCAAATGCC AGCATTTCAC AGATATGGGC CTTGAAGTTA GAGAAAATTC AACAGTGAGG GACAGCTTCC
CTGGTTAGTA CGGTGAAGTG GGCCCCTACC TAGAATGTGG CTGATTGTAA ACTAACCCTT AACTGCAAGA ACATTTCTTA
CATCTCCCAA ACATCCCTCA CAGTAAAAAC CTTAAAATCT AAGCTGGTAT GTCCTACTCC CCATCCTCCT CCCCACAACA
AAACACCAGT GCAGGCCAAC TTGTTCAGTG GAGCCCCGGG ACAAAGCAAA TGGAAGTCCT GGGTGCTTCT GACGCACACC
TATTGCAAGC AAGGGTTCAA AGACCCAAAA CCCAAAATGG CAGGGGAGGG AGAGATGGGG GTGGGAGGCT GTCAGTGGGG
AACAAGAAGT GGAGAATGTC AGTCTGAGTC AGGCCCTTCT GTCTTGAACA TGAGTTTTTT ATGGCGGGAG GTAGACTGAC
CCTTTTTGGA CTTCAGGTGG CTGTAGGAGA CAGAAGCAGG GAGGAGAGAT GACATCTAGG GCCAGGAAGG GGCTGAGGTC
ACTCACCTGG AGTGAGCCCT GCTCCCCCCT GGCTCCTTCC CAGCCTGGGC ATCCTTGAGT TCCAAGGCCT CATTCAGCTC
TCGGAACATC TCGAAGCGCT CACGCCCACG GATCTGCAGC AACAGAGGAG GGGGAGAAGT AAGTATATAC ACTTGATAAG
AGGTCCCAAG ACTTAGTACC TGAAGGGTGA AATATTCTCC ATCCAGTGGT TTCTTCTTTG GCTGGGGAGA GGAGCTGGTG
TTGTTGGGCA GTGCTAGGAA AGAGGCAAGG AAAGGTGATA AAAGTGAATC CTCCACCGCT TCTTGTCCTG CTTGCTTACC
TCGCTTAGTG CTCCCTGGGG GCAGCTCGTG GTGAGGCTCC CCTTTCTTGC GGAGATTCTC TTCCTCTGTG CGCCGGTCTC
TCCCAGGACA GGCACAAACA CGCACCTCAA AGCTGTTCCG TCCCAGTAGA TTACCACTAC TCAGGATAGG AAAAGAGAAG
CAAGAGGCAG TACAGTGTGC AGGGTGGCAA GTGGCTCCTG ACCTGGAGTC TTCCAGTGTG ATGATGGTGA GGATGGGCCT
CCGGTTCATG CCGCCCATGC AGGAACTGTT ACACATGTAG TTGTAGTGGA TGGTGGTACA GTCAGAGCCA ACCTAGGAGA
TAACACAGGC CCAAGATGAG GCCAGTGCCC TCCCAGAGAC CCCAGTTGCA AACCAGACCT CAGGCGGCTC ATAGGGCACC
ACCACACTAT GTCGAAAAGT GTTTCTGTCA TCCAAATACT CCACACGCAA ATTTCCTTCC ACTCGGATAA GATGCTGAGG
AGGGGCCAGA CCTAAGAGCA ATCAGTGAGG AATCAGAGGC CTGGGGACCT GTCGTCTCTC CAGCCCCAGC TGCTCACCAT
CGCTATCTGA GCAGCGCTCA TGGTGGGGGC AGCGCCTCAC AACCTCCGTC ATGTGCTGTG ACTGCTTGTA GATGGCCATG
GCGCGGACGC GGGTGCCGGG CGGGGGTGTG GAATCAACCC ACAGCTGCAC AGGGCAGGTC TTGGCCAGTT GGCAAAACAT
CTTGTTGAGG GCAGGGGAGT ACTGTAGGAA GAGGAAGGAG ACAGAGTTGA AAGTCAGCAT GGAAGCCAGC CCCTCAGGGC
AACTGACCGT GCAAGTCACA GACTTGGCTG TCCCAGAATG CAAGAAGCCC AGACGGAAAC CGTAGCTGCC CTGGTAGGTT
TTCTGGGAAG GGACAGAAGA TGACAGGGGC CAGGAGGGGG CTGGTGCAGG GGCCGCCGGT GTAGGAGCTG CTGGTGCAGG
GGCCACGGGG GGAGCAGCCT CTGGCATTCT GGGAGCTTCA TCTGGACCTG GGTCTTCAGT GAACCATTGT TCAATATCGT
CCGGGGACAG CATCAAATCA TCCATTGCTT GGGACGGCAA GGGGGACTGT AGATGGGTGA AAAGAGCAGT CAGAGGACCA
GGCCAGGTCC CCAGCCCAAC CCTTGTCCTT ACCAGAACGT TGTTTTCAGG AAGTCTGAAA GACAAGAGCA GAAAGTCAGT
CCCATGGAAT TGGGCCTGCC CTTCCAATGG ATCCACTCAC AGTTTCCATA GGTCTGAAAA TGTTTCCTGA CTCAGAGGGG
GCTCGACGCT AGGATCTGAC TGCGGCTCCT CCATGGCAGT GACCCGGAAG GCAGTCTGGC TGCTGCAAGA GGAAAAGTGG
GGATCCAGCA TGAGACACAG GACTCATCAA GTTCAGTCAG GAGCTTACCC AATCCAGGGA AGCGTGTCAC CGTCGTGGAA
AGCACGCTCC CAGCCCGAAC GCAAAGTGTC CCCGGAGCCC AGCAGCTACC TGCTCCCTGG ACGGTGGCTC TAGACTTTTG
AGAAGCTCAA AACTTTTAGC GCCAGTCTTG AGCACATGGG AGGGGAAAAC CCCAATCCCA TCAACCCCTG CGAGGCTCCT
GGCACAAAGC TGGACAGTCG CCATGACAAG TAAGGGCAAG TAATCCGCCT GCCGGAGGAA GCAAAGGCCA CCCCTCTTGA
GTGTCTTGGG GACAGCTCTT TCCACCCCTG GAAGATGGAA ATAAACCTGC GTGTGGGTGG AGTGTTAGGA CCAACGGTTT
CCTAGGAGTA TGTGGTTTTG CTGTGTG

Figure 4. This figure shows the result of using the visualize link from the page shown in Figure 3. This is created using the program Light Weight Genome Viewer (lwgv) which can be downloaded from our website (http://lwgv.sourceforge.net). The top track is the genomic strand whereas the second track shows the exons in this region. The next three tracks show the hairpin designs from three categories (released, designed and withheld) and the bottom tracks show the alignments of expressed sequence tags to the genomic region. Below the tracks, the sequence of the region is shown, with the exons and the hairpins highlighted.

there are several constructs that can simultaneously silence genes from the mouse and human genomes.

We built a website that allows easy access to these resources and presents the results in a user-friendly manner. Figure 2 shows the results of a search conducted on RNAi Codex. Users can access all the data in the database (Figure 3). In addition, information from external sources is also shown, such as mapping information consisting of a view of the

mRNA and the position of the constructs on the mRNA (Figure 4). It is also possible to contact commercial vendors and purchase the shRNA constructs through the website (Figures 2 and 3).

The website also allows annotation of constructs, and addition of references to publications (Figure 3). This adds utility to the database since it is impossible for a single laboratory to verify the functional status of every construct or even a substantial portion of the constructs.

RESULTS

The database holds data for constructs targeting three organisms, the *Homo sapiens*, *Mus musculus* and *Rattus norvegicus* genomes. It currently holds 82 450 constructs targeting 31 039 human genes, 73 562 constructs targeting 30 381 mouse genes and 26 611 constructs targeting 15 410 rat genes. There are 6885 constructs that simultaneously target 5569 genes in both, the mouse and human genomes.

The RNAi Codex website (http://codex.cshl.org) acts as the primary means of accessing the database. Bulk downloads of data are also allowed from the website. We use three figures (Figures 2–4) to explain use of the system and the features available on the website. Figure 2 shows the result of searching RNAi Codex for constructs that silence the gene, p53, and also explains how to search for constructs. Figure 3 shows how information on each individual hairpin can be obtained. Figure 4 shows how the website allows visualization of the alignments of the antisense sequences with the mRNAs.

DISCUSSION

RNAi Codex is a unique resource for shRNA constructs, which will allow researchers worldwide access to information as well as allow purchase of the constructs. The underlying GeneSeer service (http://geneseer.cshl.org) (20) is undergoing constant improvement, which in turn improves functioning of the RNAi Codex, making searches easier and more accurate. We plan to incorporate data from other libraries that are coming online and also encourage user participation in annotating their experiences with specific constructs. Unfortunately, there is scant experimental data on shRNA constructs. We plan to use a human curator, along with user-contributions, to help in entering such data (and related publications), when they become available. The annotation of constructs will allow building models to predict the performance of shRNA constructs and help improve designs. In addition, RNAi Codex will also help track publications in the field.

ACKNOWLEDGEMENTS

We received generous support from the Cancer Center of CSHL and comments from numerous users. Xavier Roca helped to improve the manuscript. Funding to pay the Open Access publication charges for this article was provided by Cancer Center of CSHL.

Conflict of interest statement. None declared.

REFERENCES

1. Meister,G. and Tuschl,T. (2004) Mechanisms of gene silencing by double stranded RNA. *Nature*, **431**, 343–349.
2. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
3. Reinhart,B.J., Slack,F.J., Basson,M., Pasquinelli,A.E., Bettinger,J.C., Rougvie,A.E., Horvitz,H.R. and Ruvkun,G. (2000) The 21 nucleotide let 7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901–906.
4. Scott,R.J. and Spielman,M. (2004) Epigenetics: imprinting in plants and mammals—the same but different? *Curr. Biol.*, **14**, 201–203.
5. Lippman,Z. and Martienssen,R. (2004) The role of RNA interference in heterochromatic silencing. *Nature*, **431**, 364–370.
6. Gil,J. and Esteban,M. (2000) Induction of apoptosis by the dsRNA dependent protein kinase (PKR): mechanism of action. *Apoptosis*, **5**, 107–114.
7. Manche,L., Green,S.R., Schmedt,C. and Mathews,M.B. (1992) Interactions between double stranded RNA regulators and the protein kinase DAI. *Mol. Cell. Biol.*, **12**, 5238–5248.
8. Elbashir,S.M., Harborth,J., Lendeckel,W., Yalcin,A., Weber,K. and Tuschl,T. (2001) Duplexes of 21 nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, **411**, 494–498.
9. Paddison,P.J., Caudy,A.A., Bernstein,E., Hannon,G.J. and Conklin,D.S. (2002) Short hairpin RNAs (shRNAs) induce sequence specific silencing in mammalian cells. *Genes Dev.*, **16**, 948–958.
10. Siolas,D., Lerner,C., Burchard,J., Ge,W., Linsley,P.S., Paddison,P.J., Hannon,G.J. and Cleary,M.A. (2005) Synthetic shRNAs as potent RNAi triggers. *Nat. Biotechnol.*, **23**, 227–231.
11. Kim,D.H., Behlke,M.A., Rose,S.D., Chang,M.S., Choi,S. and Rossi,J.J. (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat. Biotechnol.*, **23**, 222–226.
12. Paddison,P.J., Cleary,M., Silva,J.M., Chang,K., Sheth,N., Sachidanandam,R. and Hannon,G.J. (2004) Cloning of short hairpin RNAs for gene knockdown in mammalian cells. *Nature Methods*, **1**, 163–167.
13. Khvorova,A., Reynolds,A. and Jayasena,S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209–216.
14. 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.
15. 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.
16. Jackson,A.L. and Linsley,P.S. (2004) Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.*, **20**, 521–524.
17. Sachidanandam,R. (2005) RNAi as a bioinformatics consumer. *Brief Bioinformatics*, **6**, 146–162.
18. Paddison,P.J., Silva,J.M., Conklin,D.S., Schlabach,M., Li,M., Aruleba,S., Balija,V., O'Shaughnessy,A., Gnoj,L., Scobie,K. *et al.* (2004) A resource for large scale RNAi based screens in mammals. *Nature*, **428**, 427–431.
19. Berns,K., Hijmans,E.M., Mullenders,J., Brummelkamp,T.R., Velds,A., Heimerikx,M., Kerkhoven,R.M., Madiredjo,M., Nijkamp,W., Weigelt,B. *et al.* (2004) A large scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature*, **428**, 431–437.
20. Olson,A.J., Tully,T. and Sachidanandam,R. (2005) GeneSeer: a sage for gene names and genomic resources. *BMC Genomics*, **6**, 134.
21. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.*, **25**, 25–29.
22. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, **31**, 365–370.
23. Wain,H.M., Lush,M.J., Ducluzeau,F., Khodiyar,V.K. and Povey,S. (2004) Genew: the Human Gene Nomenclature Database. *Nucleic Acids Res.*, **32**, 255–257.