Lessons from Nature: microRNA-based shRNA libraries

Kenneth Chang[1], Stephen J Elledge[2] & Gregory J Hannon[1]

Loss-of-function genetics has proven essential for interrogating the functions of genes and for probing their roles within the complex circuitry of biological pathways. In many systems, technologies allowing the use of such approaches were lacking before the discovery of RNA interference (RNAi). We have constructed first-generation short hairpin RNA (shRNA) libraries modeled after precursor microRNAs (miRNAs) and second-generation libraries modeled after primary miRNA transcripts (the Hannon-Elledge libraries). These libraries were arrayed, sequence-verified, and cover a substantial portion of all known and predicted genes in the human and mouse genomes. Comparison of first- and second-generation libraries indicates that RNAi triggers that enter the RNAi pathway through a more natural route yield more effective silencing. These large-scale resources are functionally versatile, as they can be used in transient and stable studies, and for constitutive or inducible silencing. Library cassettes can be easily shuttled into vectors that contain different promoters and/or that provide different modes of viral delivery.

RNAi is an evolutionarily conserved, sequence-specific gene-silencing mechanism that is induced by dsRNA. Each dsRNA silencing trigger is processed into 21–25 bp dsRNAs called small interfering RNAs (siRNAs)[1–3] by Dicer, a ribonuclease III family (RNase III) enzyme[4]. The resulting small RNAs enter the RNA-induced silencing complex (RISC), which uses a single-stranded version of the small RNA as a guide to substrate selection[1,3,5,6]. Perfect complementarity between the substrate and the small RNA leads to target-RNA cleavage by an RNAse H–like active site within an Argonaute protein that forms the core of RISC[7,8]. For each small RNA, the two strands of the Dicer product are treated differently. The strand with the less stable duplex at its 5′ end is incorporated preferentially into RISC[9,10]. This realization has led to rational designs of effective siRNA sequences and yielded substantial improvements in both the efficiency and reliability of RNAi.

miRNAs are a class of endogenous dsRNAs that exert their effects through the RNAi pathway. More than a decade elapsed between the discovery of the first miRNA, *Caenorhabditis elegans lin-4* (refs. 11,12) and the achievement of at least a superficial understanding of the biosynthesis, processing and mode of action of this class of noncoding regulatory RNAs[13,14]. miRNAs are transcribed by RNA polymerase (pol) II as long primary polyadenylated transcripts (pri-miRNAs)[15,16]. Through mechanisms that have yet to be discovered, the pri-miRNA is recognized and cleaved at a specific processing site by the RNase III enzyme, Drosha, in the context of the Microprocessor complex, to produce an miRNA precursor (pre-miRNA) of approximately 70–90 nucleotides (nt)[17–22]. The pre-miRNA has a 2-nt 3′ overhang at one end[17]. This distinctive structure is recognized by the Exportin-5–Ran-GTP heterodimer, and the pre-miRNAs is shuttled to the cytoplasm[23,24]. Only then is the miRNA precursor recognized and processed by Dicer into a mature miRNA, using the 3′ overhang as a guide for site-specific cleavage at the second processing site[25,26].

[1]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, Howard Hughes Medical Institute, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [2]Department of Genetics, Center for Genetics and Genomics, Howard Hughes Medical Institute, Harvard Medical School, Room 158D, NRB, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. Correspondence should be addressed to hannon@cshl.org

PUBLISHED ONLINE 23 AUGUST 2006; DOI:10.1038/NMETH923

We have build upon our ever-deepening understanding of the biology of the RNAi and miRNA pathways to construct large-scale libraries of artificial miRNAs that cover the majority of genes in the mouse and human genomes. Libraries covering the rat genome are currently under construction (G.J.H., S.J.E. and R. Gibbs; unpublished data). Initial designs, including that of our first-generation library, were based upon a simple hairpin structure that mimicked an intermediate in the miRNA maturation pathway[27,28], the pre-miRNA; however, more recent and more effective designs, such as that adopted in our second-generation library, are based upon modified primary miRNA transcripts[29,30]. In this review, we discuss the construction and features of our second-generation library with occasional reference to the first-generation library.

miRNA-based shRNA libraries

A miR-30–based shRNA expression cassette. The discovery of endogenous triggers of RNAi suggested that RNAi might be induced in mammalian cells by synthetic genes that express mimics of these naturally occurring regulatory molecules. Several groups have tested this approach by expressing artificial miRNAs in the form of shRNAs with stems of varying lengths (19–29 nt) and loops of 4–15 nt[25,27,28,31,32]. Artificial miRNAs can be expressed from both pol II and pol III promoters, resulting in a varying extent of silencing.

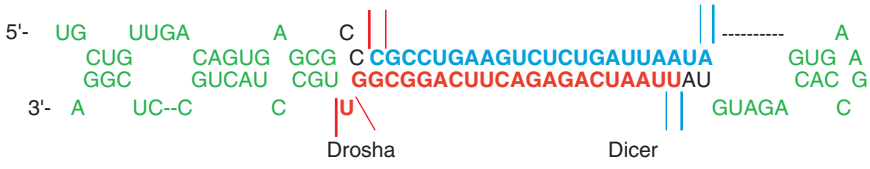
Presently, there is no clear consensus as to the most effective manner to present synthetic miRNAs to the RNAi pathway. After extensively testing different design strategies, however, our results indicate that expression of the artificial miRNA in the context of a natural miRNA primary transcript provides the highest levels of mature miRNA in RISC and generally effective silencing. These structures can be transcribed by either pol II or pol III promoters, although in the context of an integrated genomic copy, the former provided the most consistent results. This permits the use of standard expression strategies for artificial miRNAs, and allows access to all inducible and tissue-specific expression systems that have been developed for protein expression. Given previous reports of artificial miRNAs based upon miR-30 (refs. 15–17,29,30,33,34) and our own attempts to express artificial miRNA from several precursors, we chose to model our library design upon miR-30 (**Fig. 1**). Previous studies have indicated that the presence of sequences flanking the native miRNA is essential for its efficient processing. Therefore we designed each mir-30–based shRNA (shRNA^{mir}) in a carrier such that it is flanked by ~125 bases 5′ and 3′ of the pre–miR-30 sequence.

Vector backbone. We constructed a second-generation library shRNA vector, pSM2 (**Fig. 2**), in which the shRNA^{mir} expression cassette can be packaged in a self-inactivating murine stem cell virus (MSCV). As with our first-generation shRNA library vector, pSM1 (**Fig. 2**), the expression of the small RNA is driven by the U6 promoter along with the U6 snRNA leader sequence. In the second-generation library, the U6 snRNA leader sequence lies between the promoter and the 5′ end of the miR-30 flanking region, and a pol III termination signal is inserted immediately after the miR-30 cassette. This arrangement was determined empirically in transient transfection experiments to give effective silencing.

To facilitate the tracking of shRNAs in complex pools, a randomly generated 60 nt DNA 'bar-code' region was created downstream of the miR-30 cassette such that individual hairpin RNAs are identified through their hybridization to custom bar codes on oligonucleotide microarrays. One of the key design features of pSM2, unlike its predecessor, is the ability to swap library inserts (miR-30 cassette, bar code and the microbial selection marker) into recipient plasmids containing other promoters (for example, pol II promoters for tissue-specific or inducible expression) by mating-assisted, genetically integrated cloning (MAGIC; see **Box 1** and **Fig. 3**)[28,30].

a

shRNA^{mir} (second-generation)



b

5'-miR30PCRxhoI-F (forward primer):

5'-CAGAAGGCTCGAGAAGGTATATGCTGTTGACAGTGAGCG-3'

3'-miR30PCREcoRIF (reverse primer):

5'-CTAAAGTAGCCCCTTGAATTCCGAGGCAGTAGGCA-3'

DNA template:

5'-TGCTGTTGACAGTGAGCGCCCGCCTGAAGTCTCTGATTAATAGTGAAGCCACAGATGTA-3'

5'-TTAATCAGAGACTTCAGGCGGTTGCCTACTGCCTCGGA-3'

Figure 1 | Structure of shRNA^{mir}. (**a**) A segment of the shRNA^{mir} precursor produced by pSM2. The sequence of the target site (sense orientation) from firefly luciferase (*luc1309*) is shown in blue (passenger strand) with the guide strand shown in red. For pSM2, the mapped potential cleavage sites for Dicer and Drosha are indicated by blue and red lines, respectively. Bases in green are those that remain as in miR-30 miRNA. (**b**) Primer and DNA template for shRNA amplification. The universal primers shown here contain *Xho*I and *Eco*RI sites to allow cloning into *Xho*I and *Eco*RI sites of miR-30 cassette of pSM2. Underlined sequences in each universal primer are regions complementary to respective portions (underlined) of the DNA template. The DNA oligonucleotide was designed for *luc1309* shRNA^{mir} amplification. Sense or passenger strand is in blue and antisense or guide strand is in red.

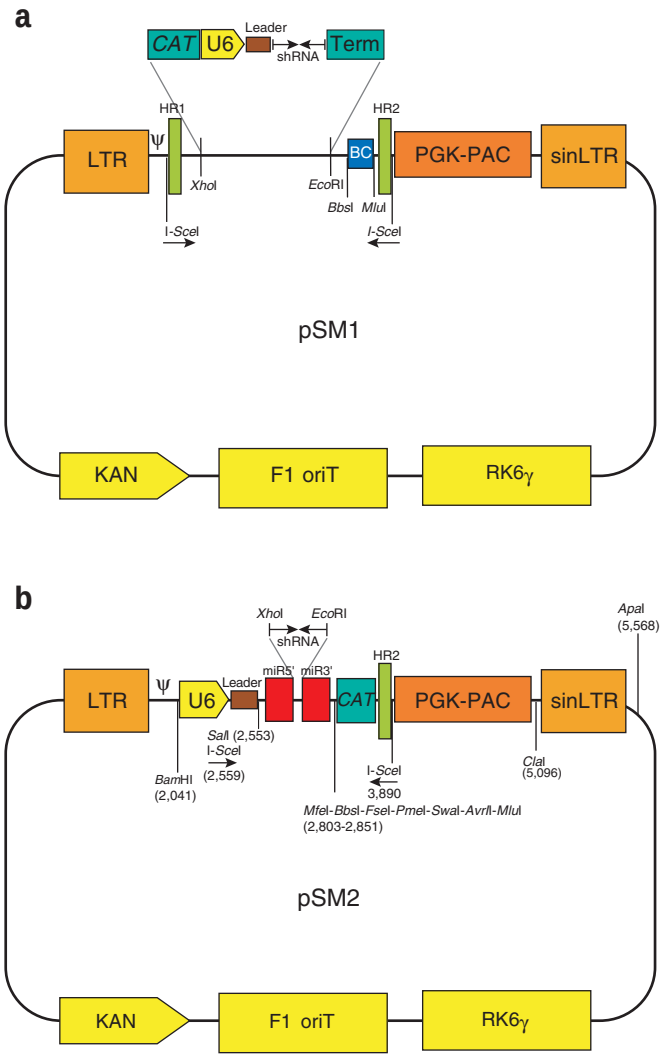


Figure 2 | Schematics of first- and second-generation library vectors. (**a**) pSM1 vector contains a U6 promoter with a 27-nt leader sequence; a U6 terminator (Term); a self-inactivating retroviral backbone; Ψ, the retroviral packaging signal; bacterial antibiotic-resistance markers kanamycin (KAN) and chloramphenicol (*CAT*); a protein-dependent origin (*RK6γ*) that requires the transacting factor, π, encoded by the relaxed copy-number gene allele, *pir1-116*; a mammalian selectable marker, puromycin, driven by the PGK promoter (PGK-*PAC*); homology regions (HR1 and HR2) for use in bacterial recombination; and a randomly generated 60-mer bar-code sequence (BC) inserted between *Bbs*I and *Mlu*I restriction sites. Upon amplification, shRNA cassettes are inserted into *Xho*I and *Eco*RI cloning sites. F1 is the origin of replication for production of single-stranded DNA; oriT is the *cis*-acting origin of transfer for gene transfer using the MAGIC system (**b**) pSM2 vector contains a U6 promoter with a leader sequence, a U6 terminator following mir-30 3′ a self-inactivating retroviral backbone; bacterial antibiotic-resistance markers kanamycin and chloramphenicol; a protein-dependent origin (*RK6γ*); a mammalian selectable marker, puromycin, driven by the PGK promoter; a homology region (HR2) for use in bacterial recombination and a randomly generated 60-mer bar-code sequence cloned between *Fse*I and *Avr*II sites. The shRNA^{mir} inserts were cloned between the 5′ and 3′ flanking sequences derived from the mir-30 primary transcript using *Xho*I and *Eco*RI restriction sites. The nucleotide positions for sites in an excised version of an empty vector (no shRNA or bar code) are given. LTR and sinLTR denote the long terminal repeat of the self-inactivating retroviral backbone and Ψ indicates the retroviral packaging signal.

Selection of shRNA sequences. The artificial miRNAs used in our second-generation library are designed based upon a proprietary algorithm developed by Rosetta Inpharmatics (Ge *et al.*, unpublished data). In general, these designs were reverse-engineered based upon tests of several thousand siRNAs. Overall, the designs follow the thermodynamic rules suggested by analysis of endogenous miRNAs and by studies of siRNA incorporation into RISC. However, additional positional biases are also introduced by the algorithm. Second-generation libraries were designed to include six different shRNA^{mir} constructs to each known and predicted gene of the human (34,711 genes) and mouse (32,628 genes) genomes.

Library construction. DNA oligonucleotide templates for each second-generation silencing trigger consist of common 5′ and 3′ miR-30 regions flanking a 22-nt sense and a 22-nt antisense target sequence that is separated by the 15-nt miR-30 loop (**Fig. 1a**). These shRNA^{mir} templates are designed to be amplified by universal primers (**Fig. 1b**). The amplicon encoding each artificial miRNA is inserted between *Xho*I and *Eco*RI restriction sites that are flanked by pri–miR-30 sequences.

The cloning strategies used for creating first- and second-generation libraries were notably different. Rather than using conventional methods for the generation of large and complex libraries of defined nucleic acids, we used a more cost-effective approach by synthesizing oligonucleotides on printed microarrays. Though there are several available technological approaches to synthesizing DNA microarrays, we chose and validated inkjet synthesis (Agilent Technologies, Inc.) as it provided the most accurate production of the long oligonucleotides (96 nt) required for our application. PCR amplification of ammonium hydroxide–cleaved templates yielded an average of ~25% to >60% of perfect shRNAs, using a combination of thermostable proofreading polymerases and PCR-enhancing agents. Oligonucleotides corresponding to more than 32,000 known and predicted genes in human (>195,000 oligonucleotides) and mouse (>187,000 oligonucleotides) genomes were synthesized on more than 21 glass-slide microarrays, each capable of carrying 22,000–24,000 oligonucleotide probes per array.

To maximize cloning efficiency, we inserted the pSM2 backbone into lambda phage flanked by *loxP* sites. This was first used to create a high-complexity library of bar-coded cosmids comprising ~10^8 independent clones. This pool then served as the vector population for library construction. shRNA^{mir} inserts were amplified from oligonucleotides that were cleaved from microarrays in pools of ~22,000. These were inserted between the *Xho*I and *Eco*RI sites in bar-coded λ-pSM2. More that 10^6 independent clones were routinely obtained for each pool. Phage populations containing the shRNA^{mir} inserts were used to infect BNN132, a bacterial strain that expresses Cre recombinase. Exision of the shRNA^{mir} plasmid from the phage occurred with ~50% efficiency (comparison of plaque-forming versus colony-forming units). Each pool was then entered into a customized, high-throughput sequencing pipeline.

As with the first-generation library, all pSM2 clones were sequence-verified and stored individually in a multiwell format. Depending upon the individual array synthesis, between 25 and 50% of all clones contained the correct shRNA^{mir} sequence.

BOX 1 THE MAGIC SYSTEM

MAGIC is a gene-transfer system designed to work exclusively *in vivo*[38]. Although it was originally designed for transfer of open reading frames, we adapted this system for use in our shRNA system. For use with MAGIC, the pSM2 vector was designed with a conditional origin of replication (R6Kγ) and a transfer system for the fertility factor F′. In the presence of F′ with a functional *tra* operon, the transfer origin can mobilize pSM2 for exchange from one strain into a second strain. Additionally, the pSM2 mir-30–bar code–*CAT* (chloramphenicol acetyltransferase) region is flanked by I-*Sce*I sites.

The recipient vector, whether it is a standard plasmid or a lentiviral vector, must also contain two I-*Sce*I sites flanking a negative selectable marker, the gene encoding PheS294G. Outside of the I-*Sce*I sites in the recipient lies homology to the 5′ of mir-30 on one side and homology to the sequences just after the chloramphenicol acetyltransferase gene on the other side. The *pheS* gene encodes a mutant phenylalanine tRNA synthetase, which can charge $tRNA^{Phe}$ with phenylalanine or chloro-phenylalanine (Cl-Phe) if the latter is supplied in the medium. Cl-Phe is toxic to cells if incorporated into proteins and is not used by the wild-type PheS protein. Thus, the gene encoding mutant PheS is a dominant, negative selectable marker.

Recombinational shRNA gene transfer occurs when pSM2 is mated into the strain containing the recipient vector of interest, typically a vector containing a β-lactamase gene (conferring ampicillin resistance) and a nonconditional pUC origin. In the recipient strain the I-*Sce*I restriction enzyme is induced, and cleaves both pSM2 and the recipient vector. The resulting linear molecules can be recircularized by homologous recombination between the liberated mir-30–*CAT* fragment and the recipient vector from which the gene encoding PheS294G was excised by I-*Sce*I. Only a correct recombination event will give rise to a replication-competent vector that contains the ampicillin resistance marker plus the chloramphenicol resistance marker, and that lacks *pheS*.

It should be noted that the pSM2 plasmid has a conditional origin of replication. Thus, it cannot replicate in the recipient strain and therefore any cell that becomes chloramphenicol resistant must have transferred the resistance cassette to the nonconditional origin on the ampicillin-resistant recipient plasmid to create the desired recombinant. This is illustrated in **Figure 3**. The advantage of this system is that whole libraries of shRNAs and their bar codes can be transferred into different vector backbones with different markers, different promoters or in combination with other genes without the need for conventional cloning, lending great versatility to the existing libraries.

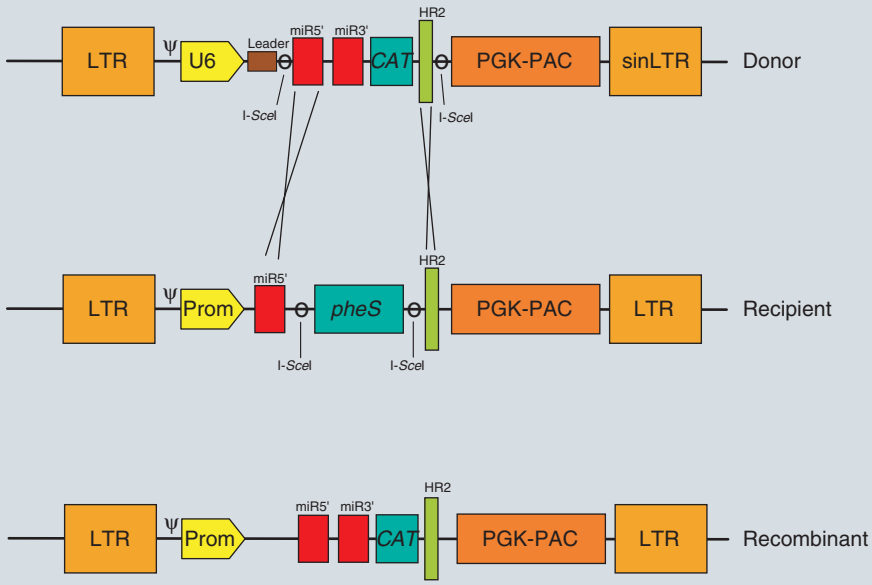


Figure 3 | Mating scheme for pSM2 library clones. The $shRNA^{mir}$ cassette from the library expression plasmid (donor) is transferred into a recipient with wild-type MSCV LTRs, an alternative promoter and ampicillin-resistance marker (not shown). The recombinant plasmid is recovered by selecting for clones that exhibit chloramphenicol and ampicillin resistance.

Sequencing progress was monitored dynamically for each pool, and sequencing was halted when accumulation of new clones dropped to an unacceptable rate. At this point, sequences that were already identified were withdrawn from the library, and arrays were synthesized to produce new $shRNA^{mir}$ oligonucleotide populations for cloning and sequencing. Each time the number of sequence-verified $shRNA^{mir}$ for a given gene exceeded three, the remaining oligonucleotides were also withdrawn from the synthesis queue. Iterative synthesis of more than 70 custom microarrays has thus far yielded 87,283 verified human $shRNA^{mir}$ clones and 76,896 verified mouse $shRNA^{mir}$ clones (see **Box 2**). Summaries of coverage for specific functional groups are given in **Table 1** and **Supplementary Table 1** online. All second-generation shRNA clones are made available to the research community through Open Biosystems Inc. as soon as they are sequence-verified and pass Open Biosystems quality control.

Validation of $shRNA^{mir}$ designs

Our initial studies indicated that longer hairpin structures (~29-nt stems) were more effective triggers than those containing shorter stems of 19–21 nt[25]. All of these structures, however, were designed around the pre-miRNA, which is an intermediate in miRNA biogenesis. When advances in our basic understanding of miRNA biology revealed in initial upstream processing step, we sought to compare the relative potency of artificial miRNAs designed around expression of the pre-miRNA to those modeled on the pri-miRNA transcript. Direct comparison of small RNA production by northern analysis showed that a substantially higher amount of small RNAs (22 nt)

are produced with shRNA^{mir} designs (modeled on the pri-miRNA transcript) as compared to simple hairpin designs modeled on the pre-miRNA. Overall, approximately 12 times more small RNA was incorporated into RISC when the same mature sequence was produced from pSM2 as compared to pSM1.

To successfully apply siRNA design rules to shRNAs, it was crucial to predict precisely the mature small RNA that is generated from pSM2 by Dicer and Drosha processing. Initial designs were based upon studies of Dicer and Drosha processing sites in native miR-30, and on extensive biochemical experimentation with processing of artificial miRNAs both *in vitro* and *in vivo*. We verified assumptions inherent in shRNA^{mir} design by directly mapping the 3′ ends of several small RNAs generated from shRNA^{mir} vectors. We were able to infer the 5′ end of mature small RNAs by counting 22 nt from mapped 3′ ends of both the guide and passenger strands of shRNA^{mir} constructs (**Fig. 1a**). For some synthetic hairpins that have been tested, we could not determine the cleavage site definitively as our studies indicated that mature products could vary by a single base at their 3′ ends (**Fig. 1a**). Perhaps this reflects a genuine heterogeneity in Drosha cleavage that also has been observed with Dicer substrate processing *in vitro* and for mature miRNAs *in vivo*. Thus, for target-sequence selection for the design of shRNA^{mir}, we chose target sequences that gave similar thermodynamic profiles, even if processing sites were translocated by a base.

Library validation

To test the performance of the first- and second-generation shRNA libraries in a biological context, we chose an assay for which we could expect to recover numerous shRNAs targeting a known biological process. We focused on an assay that could report, in a quantitative manner, proper function of the 26*S* proteasome. The proteasome is a major nonlysosomal protease in eukaryotic cells, and its function can be detected in an assay in which the mouse ornithine decarboxylase (MODC), a well-known target of the proteasome, containing a PEST domain, directs degradation without the need for ubiquitination[35]. A green fluorescent protein (Zs green) is fused to the MODC degron such that any disruption of proteasome function can be detected by the accumulation of fluorescence. Thus we sought to identify shRNAs that could compromise proteasome function. Our goal was to examine the performance of both versions of our shRNA libraries using the same approach.

A collection of approximately 7,000 first-generation library plasmids, targeting 4,873 genes, was individually transfected into HEK293T cells along with the MODC reporter and a transfection control (DsRed). We recovered close to 100 RNAi constructs that resulted in the increase of Zs green fluorescence above background. Of these, 22 corresponded to 15 proteasome subunits, and all scored positively in a secondary screen involving only these 22 hairpins and 33 nonscoring proteasome hairpins from the first round. An additional 14 shRNAs targeting proteasome directly also scored in a more focused assay. In all, the positively scoring hairpins targeted mostly the 19*S* base subunits including five out of five ATPases and the two largest nonATPases. We found that targeting of the 19*S* lid and the 20*S* core produced lower hit rates. All positively scoring hairpins showed measurable suppression by western blotting and fluorescence microscopy[28].

To compare the performance of the second-generation library to that of the first, we chose 53 shRNAs (24 first-generation and 29 second-generation hairpins), targeting 13 different genes involved in proteasome function, and tested their individual ability to effect accumulation of Zs green in transient assays. We found that the majority of second-generation shRNA^{mir} were as potent as the best hairpins identified in the screen of the first generation library. There was a good correlation between disruption of proteasome function and reduction in mRNA levels as measured by quantitative reverse-transcriptase PCR. This indicated that, on average, the second-generation shRNA^{mir} performed substantially better than first-generation shRNAs modeled on pre-miRNAs. To test the second-generation library in the context of a screen, a set of 515 human kinase shRNA^{mir} containing 47 shRNA^{mir} directed against proteasome subunits was assayed for effects on proteasome function. We found that 34 of the 47 hairpins scored positive, and ten other shRNA^{mir} that were not previously linked to proteasome function also scored positive. In a second screen with these 44 candidates, all but five shRNAs (nonproteasomal) scored positive upon retesting[30].

Vectors and shRNA expression

It is now known that miRNAs are predominantly transcribed by pol II. Direct comparison of various pol III promoters (U6, H1 and tRNA-val) and pol II promoters (CMV and MSCV–long terminal repeat, MSCV-LTR) with a highly potent shRNA^{mir} in transient assays showed little difference in suppression. With a less efficient hairpin, however, the CMV and U6 promoters demonstrated the highest level of suppression, but both performed equally well. Based upon these results, we chose to retain the U6 promoter for the expression of library shRNA^{mir}. For long-term suppression, pSM2 can efficiently produce stable populations by random integration of

Table 1 | A sampling of second-generation library genes by functional groups

Functional category	Human genes in the category	Human shRNA clones	Mouse genes in the category	Mouse shRNA clones
Apoptosis[a]	587	2,525	584	1,834
Cancer-relevant[b]	871	3,793	937	3,005
Cell cycle	536	2,646	505	1,815
Checkpoint	124	650	121	434
DNA repair	118	620	139	413
DNA replication[c]	240	1,168	258	841
Enzymes	3,010	12,928	2,934	9,703
GPCR[d]	690	2,632	716	2,132
Kinases[e]	625	3,163	579	2,440
Phosphatases[f]	210	933	198	725
Proteases[d]	470	1,761	467	1,384
Proteolysis[g]	305	1,778	286	1,021
Signal transduction[d]	2,724	11,250	2,690	8,660
Trafficking	488	1,993	483	1,560
Transcription[d]	841	3,576	799	2,604

[a]Curated by Y. Luzebnik & colleagues (CSHL). [b]Broadly defined set of cancer genes curated by CSHL colleagues. [c]Curated by B. Stillman & P. Paddison (CSHL). [d]Based on GO Database annotation. [e]Based on Manning and colleagues (2002). [f]Curated by N. Tonks & colleagues (CSHL). [g]Curated by W. Harper (Harvard University).

BOX 2 HANNON-ELLEDGE LIBRARY AT A GLANCE

Library vector

Vector description. Second generation: pSM2 (see **Fig. 2**; ref. 30).
Viral vector backbone: self-inactivating retroviral vector.
Plasmid pSM2 is related to pSM1 (first-generation library vector; **Fig. 2**; ref. 27).
Major modifications: (i) in pSM2 the shRNA is embedded in an miR-30 cassette, (ii) pSM2 contains elements to swap the miR-30 expression cassette (with the 60-nt bar code and microbial selection gene but without the U6 promoter) into recipient plasmids containing other promoters by mating-assisted bacterial recombinational cloning.
Additional features: in both pSM1 and pSM2, the shRNA expression cassette is under the control of the U6 pol III promoter. The vectors also contain unique 60-nt bar-code sequences.

Library stability. Under the recommended handling and growth conditions, the second-generation library clones are very stable and more stable than first-generation clones. Upon sequence verification, second-generation library plasmids are recovered and retransformed into a low recombination bacteria host.
We evaluated the recombination rate in a random sampling of clones by monitoring DNA quality by electrophoresis and observed recombination in ~5% of clones (see http://www.openbiosystems.com).

Infectivity. A variety of human and mouse cell lines have been tested with pSM2 virus: 293T, 293, MCF10A, NIH3T3, IMR90 and derivatives, as well as a variety of cancer cell lines.

Silencing. No data available.

Library production

Vector DNA production yield. Both first and second-generation vectors contain the π protein–dependent origin of replication (*RK6γ*). Yield is typically less than with pUC origin plasmids. Yield can be increased by using *pir1* strains and/or growing in rich medium. There are other variables that influence DNA yield, of note is the method of DNA extraction and reagent quality.
If grown in normal medium (LB) yield is 2–3 μg/ml (after 16–18 h at 37 °C), whereas in rich medium (GS96 or 2XYT), the yield is 5–7 μg/ml.

Viral titer. Production conditions: Typically, 12 μg of library DNA along with helper plasmids (4 μg of packaging helper plasmid, and 8 μg of pVSV.G expressing the pseudotying envelope protein) are transfected into packaging cells (LinX) plated on a 10-cm dish (cell confluence of 50–70%), using lipid carriers. Two virus harvests (filtration-clarified culture supernatant) are collected at 48 h and 72 h after transfection. The two harvests can be pooled if necessary but for most applications, virus from a single harvest is sufficient.
Viral titer is determined by infection of human (HEK293) and mouse (NIH3T3) cells by counting of antibiotic-resistant transduced colonies. These cells are highly infectable and thus serve as a standard for comparing infectivity with other cells. Typical virus titers for the second-generation library are ~1–1.3 × 10^6 I.U./ml for both HEH293 and NIH3T3 cells.

Library coverage

As of June 2006, the second-generation library contains shRNAs directed against:
32,216 human genes (87,283 shRNAs)
30,629 mouse genes (76,896 shRNAs)

Gene coverage distribution. Average: 2–3 shRNAs per target gene.
For human genes:
18% of target genes (5,800 genes) covered by exactly 1 shRNA
26% of target genes (8,362 genes) covered by 2 shRNAs
34% of target genes (11,056 genes) covered by 3 shRNAs
22% of target genes (6,998 genes) covered by ≥4 shRNAs
For mouse genes:
19% of target genes (5,740 genes) covered by exactly 1 shRNA
31% of target genes (9,359 genes) covered by 2 shRNAs
36% of target genes (10,983 genes) covered by 3 shRNAs
15% of target genes (4,547 genes) covered by ≥4 shRNAs
Region of transcripts targeted: open reading frame.
List of target genes: available online (http://codex.cshl.edu/ and http://www.openbiosystems.com).

shRNA sequences. The shRNA were designed using a proprietary algorithm of Rosetta Inpharmatics.
A list of shRNA sequences is available online (http://codex.cshl.edu/ and http://www.openbiosystems.com).

Functional tests

Sequence verification. Sequencing success rate is 35–55% (sequencing of one strand). In the second-generation library, 87,283 and 76,896 shRNA clones against human and mouse genes, respectively, were sequence-verified. In the first-generation library, 28,659 and 9,119 shRNA clones against human and mouse genes, respectively, were sequence-verified.

Validation data. Validation is only available for a small proportion of clones as a large-scale effort to systematically evaluate each individual shRNA for either collection is not currently planned. Rather, validation of candidates from functional screens have been and will continue to be the means in which library clones will be validated for the time being.
1. First- and second-generation library clones were validated in a proteosome function screen[28,30].
2. Genetic screen for candidate tumor suppressor identified REST (first-generation library)[31].
3. Probing tumor phenotypes using stable and regulated synthetic miRNA precursors (second-generation library)[37].
4. Identification and validation of oncogenes in liver cancer[39].
5. Validation in context of other vector[36,37].

Nonspecific off-target effect. In a test of several second-generation vectors, no induction of the interferon response was observed, as monitored by EIF2-α.

Distribution

Glycerol stocks (arrayed in 96-well format) and plasmid DNA of individual clones and pools of clones are available from Open Biosystems.
Sets of individual clones in glycerol stock are also available grouped by functions of targets (see **Table 1**).

library cassettes into target genomes[36]. Suppression of gene expression at single copy facilitates screening of shRNAs in pools. This is true for both positive selections and for negative selections using bar-code hybridization for tracking depleted clones. Initial experiments aimed at generating gene knockdown using single-copy integrants of pSM2 were highly variable using the U6 promoter. We therefore began to explore other promoters to drive miR-30 expression. CMV–miR-30 which had been shown by others and ourselves to work in transient assays, did not fare well at single copy in the cell lines tested. However, extensive testing led to the development of two strategies that allowed much higher-penetrance knockdown at single copy. We found that including a substantial spacer sequence, *GFP* in the initial case, between CMV and miR-30 greatly increased the ability of shRNA^{mir} to knock down gene expression[36]. The underlying biochemical reason why the spacer leads to increased knockdown efficiency is not presently clear. We also found that the non–self inactivating (non-SIN) version of pSM2, MSCV, was also highly efficient at single copy gene knockdown when miR-30 was driven from the MSCV-LTR promoter[37]. This is likely due both to the broad spectrum of cell types in which the MSCV-LTR is active and to the 1.8-kilobase packaging signal serving as a *de facto* spacer sequence. These findings have led to the development of much more efficient vectors for expression of shRNA^{mir} libraries that should allow routine, high-efficiency knockdown of gene expression and efficient bar code–based screening.

Understanding the interplay between the spacer region and the shRNA^{mir} in combination with pol II promoters allowed us to take advantage of existing conditional-promoter technologies to generate conditional knockdown vectors. We have developed lentiviral vectors with both constitutive and conditional promoters based on the various Tet repressor–regulated systems[36]. These are the PRIME vectors, which have been demonstrated to allow efficient regulated gene knockdown even at single copy. The PRIME series also has additional spacer fragments in place of *GFP* to allow their use with screens that already use *GFP* as a reporter. Similar vectors based upon conditional promoters inserted into MSCV-based SIN vectors have also been reported and validated.

Bar-code versus arrayed screening formats

The second-generation Hannon-Elledge libraries are presently available in arrayed formats suitable for high-throughput DNA preparation, plasmid transfection, virus production and well-by-well screening approaches. The use of such methods, however, is beyond the budgets of all but the most well-funded laboratories. At the core of our philosophical approach toward the generation of the tools described herein was to democratize the use of RNAi as a genome-wide tool. Therefore, we have been developing strategies that take advantage of pooled collections of sequence-verified shRNAs. Of course, the use of such pools in positive-selection schemes is easy to envision. But also, we have developed methods that permit the discovery of shRNAs that become either enriched or depleted from a population in response to a specific stress using bar-code microarrays to measure relative abundance of individual constructs within a mixed population. Current protocols allow the simultaneous screening of up to 10,000 individual constructs using arrays that report for approximately 85% of all members of the population. The behavior of constructs is consistent with biological replicates in outgrowth experiments displaying correlation coefficients in the range of 0.85–0.95 (unpublished data). We therefore feel that in the near future, the use of highly parallel tools for forward genetics in mammalian cells will become accessible to a substantial fraction of the community.

Outlook

The development of RNAi-based tools for investigating gene function has always progressed hand-in-hand with advances in our basic understanding of the RNAi pathway. The development of second-generation shRNA^{mir} vectors provides another confirmation of this maxim whereby increased understanding of miRNA biogenesis has led to the design of more efficient vector strategies for transient or stable RNAi. We have generated comprehensive shRNA^{mir} libraries targeting the majority of genes in the mouse and human genomes and have made these readily available to the scientific community. The major characteristics of the shRNAs developed by our groups are summarized in **Box 2**.

Looking toward the future, it is likely that new insights into the function of RISC and its associated RNAs will allow the production of shRNA libraries with even greater efficacy toward desired targets and reduced nonspecific or off-target effects. But even with present designs, the creation of a broad spectrum of different shRNA expression vectors and focused and genome-wide shRNA sets in those vectors will facilitate the application of these tools to a range of biological problems. With respect to our own Hannon-Elledge libraries, the production of arrayed collections of lentiviral vectors, to complement existing MSCV collections, is already well underway. Additionally, screening pools harbored in a range of constitutive and inducible viral vectors should be available within the near future for use in both positive-selection and bar-code format screens.

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

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

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

1. Zamore, P.D., Tuschl, T., Sharp, P.A. & Bartel, D.P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).
2. Hamilton, A.J.B. D.C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
3. Hammond, S.M., Bernstein, E., Beach, D. & Hannon, G.J. An RNA-directed nuclease mediates posttranscriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
4. Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
5. Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R. & Hannon, G.J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
6. Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P. & Sharp, P.A. Targeted mRNA degradation by double-stranded RNA *in vitro*. *Genes Dev.* **13**, 3191–3197 (1999).
7. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
8. Song, J-J., Smith, S.K., Hannon, G.J. & Joshua-Tor, L. Crystal structure of argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
9. Schwarz, D.S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
10. Khvorova, A., Reynolds, A. & Jayasena, S.D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).

11. Lee, R.C., Feinbaum, R.L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
12. Wightman, B., Burglin, T.R., Gatto, J., Arasu, P. & Ruvkun, G. Negative regulatory sequences in the *lin-14* 3′-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. *Genes Dev.* **5**, 1813–1824 (1993).
13. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
14. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**, 522–531 (2004).
15. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
16. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957–1966 (2004).
17. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
18. Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. & Hannon, G.J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235 (2004).
19. Landthaler, M., Yalcin, A. & Tuschl, T. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr. Biol.* **14**, 2162–2167 (2004).
20. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).
21. Gregory, R.I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).
22. Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).
23. Yi, R., Qin, Y., Macara, I.G. & Cullen, B.R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).
24. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
25. Siolas, D. *et al.* Synthetic shRNAs as potent RNAi triggers. *Nat. Biotechnol.* **23**, 227–231 (2005).
26. Song, J.J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat. Struct. Biol.* **10**, 1026–1032 (2003).
27. Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J. & Conklin, D.S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
28. Paddison, P.J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
29. Zeng, Y., Wagner, E.J. & Cullen, B.R. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
30. Silva, J.M. *et al.* Second-generation shRNA libraries covering the mouse and human genomes. *Nat. Genet.* **37**, 1281–1288 (2005).
31. Westbrook, T.F. *et al.* A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**, 837–848 (2005).
32. Kim, D. *et al.* Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat Biotechnol.* **23**, 222–226 (2005).
33. Chen, C.Z., Li, L., Lodish, H.F. & Bartel, D.P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
34. Zeng, Y. & Cullen, B.R. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**, 112–123 (2003).
35. Ghoda, L., Sidney, D., Macrae, M. & Coffino, P. Structural elements of ornithine decarboxylase required for intracellular degradation and polyamine-dependent regulation. *Mol. Cell. Biol.* **12**, 2178–2185 (1992).
36. Stegmeier, F., Hu, G., Rickles, R.J., Hannon, G.J. & Elledge, S.J. A lentiviral microRNA-based system for single copy Pol II regulated RNAi in mammalian cells. *Proc. Natl. Acad. Sci. USA* **102**, 13212–13217 (2005).
37. Dickins, R.A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat. Genet.* **37**, 1289–1295 (2005).
38. Li, M.Z. & Elledge, S.J. MAGIC, an *in vivo* genetic method for the rapid construction of recombinant DNA molecules. *Nat. Genet.* **37**, 311–319 (2005).
39. Zender, L. *et al.* Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125**, 1253–1267 (2006).