nature genetics

Second-generation shRNA libraries covering the mouse and human genomes

Jose M Silva[1,4], Mamie Z Li[2,4], Ken Chang[1,4], Wei Ge[3], Michael C Golding[1], Richard J Rickles[2], Despina Siolas[1], Guang Hu[2], Patrick J Paddison[1], Michael R Schlabach[2], Nihar Sheth[1], Jeff Bradshaw[3], Julia Burchard[3], Amit Kulkarni[3], Guy Cavet[3], Ravi Sachidanandam[1], W Richard McCombie[1], Michele A Cleary[3], Stephen J Elledge[2] & Gregory J Hannon[1]

Loss-of-function phenotypes often hold the key to understanding the connections and biological functions of biochemical pathways. We and others previously constructed libraries of short hairpin RNAs that allow systematic analysis of RNA interference–induced phenotypes in mammalian cells. Here we report the construction and validation of second-generation short hairpin RNA expression libraries designed using an increased knowledge of RNA interference biochemistry. These constructs include silencing triggers designed to mimic a natural microRNA primary transcript, and each target sequence was selected on the basis of thermodynamic criteria for optimal small RNA performance. Biochemical and phenotypic assays indicate that the new libraries are substantially improved over first-generation reagents. We generated large-scale-arrayed, sequence-verified libraries comprising more than 140,000 second-generation short hairpin RNA expression plasmids, covering a substantial fraction of all predicted genes in the human and mouse genomes. These libraries are available to the scientific community.

RNA interference (RNAi) has been exploited in organisms ranging from plants to fungi to animals for deciphering gene function through suppression of gene expression. Particularly in systems where targeted genetic manipulation is difficult or time consuming, RNAi has transformed the way in which gene function can be approached on a single-gene or genome-wide level[1–5].

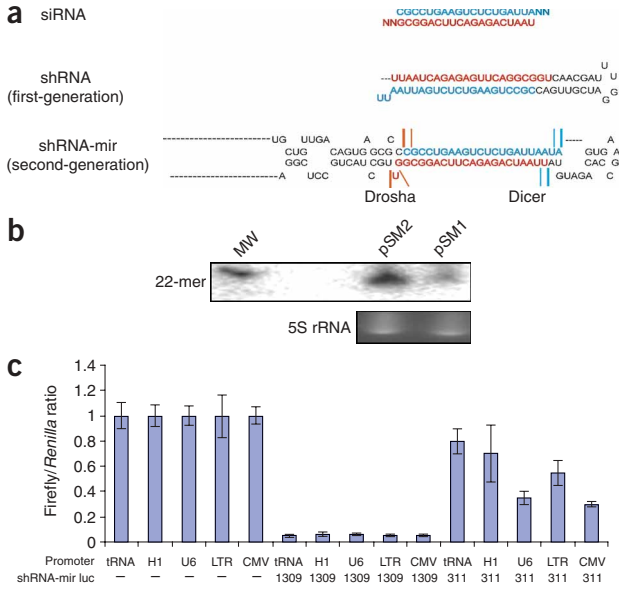

Figure 1 Design and structure of shRNA-mir cassettes. (**a**) The structures of several silencing triggers, including an siRNA, a portion of the shRNA precursor (as generated from our first-generation design in pSM1) and a segment of the shRNA-mir precursor produced by pSM2, are compared. The sequence of the target site (sense orientation) from firefly luciferase (luc1309, see **c**) is shown in blue (passenger strand), with the guide strand shown in red. For pSM2, mapped potential cleavage sites for Dicer and Drosha are indicated by blue and red lines, respectively. (**b**) Northern blotting was used to detect the mature small RNA produced after transfection of HEK 293 cells with shRNA and shRNA-mir cassettes expressed from pSM1 and pSM2, respectively, by the U6 snRNA promoter, but no substantial accumulation of pre-miRNA was observed. Transfection rates were normalized using a codelivered DsRed expression plasmid. (**c**) The ability of five different promoters (human tRNAval, human H1 RNA, human U6 snRNA, MSCV LTR and human CMV IE) to drive shRNA-mir expression and to silence luciferase in transient transfections was tested. Two different shRNAs were used, a highly efficient shRNA (luc1309) and a less efficient shRNA (luc311). In each case, the level of firefly luciferase was normalized to that of a nontargeted *Renilla* luciferase. Controls with empty vectors lacking a hairpin insert are also shown.

[1]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [2]Department of Genetics, Center for Genetics and Genomics, Brigham and Women's Hospital, Harvard Medical School, Room 158D, NRB, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [3]Rosetta Inpharmatics LLC, 401 Terry Ave. North, Seattle, Washington 98109, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to S.J.E. (selledge@genetics.med.harvard.edu) or G.J.H. (hannon@cshl.edu).

Received 15 June; accepted 18 August; published online 2 October 2005; doi:10.1038/ng1650

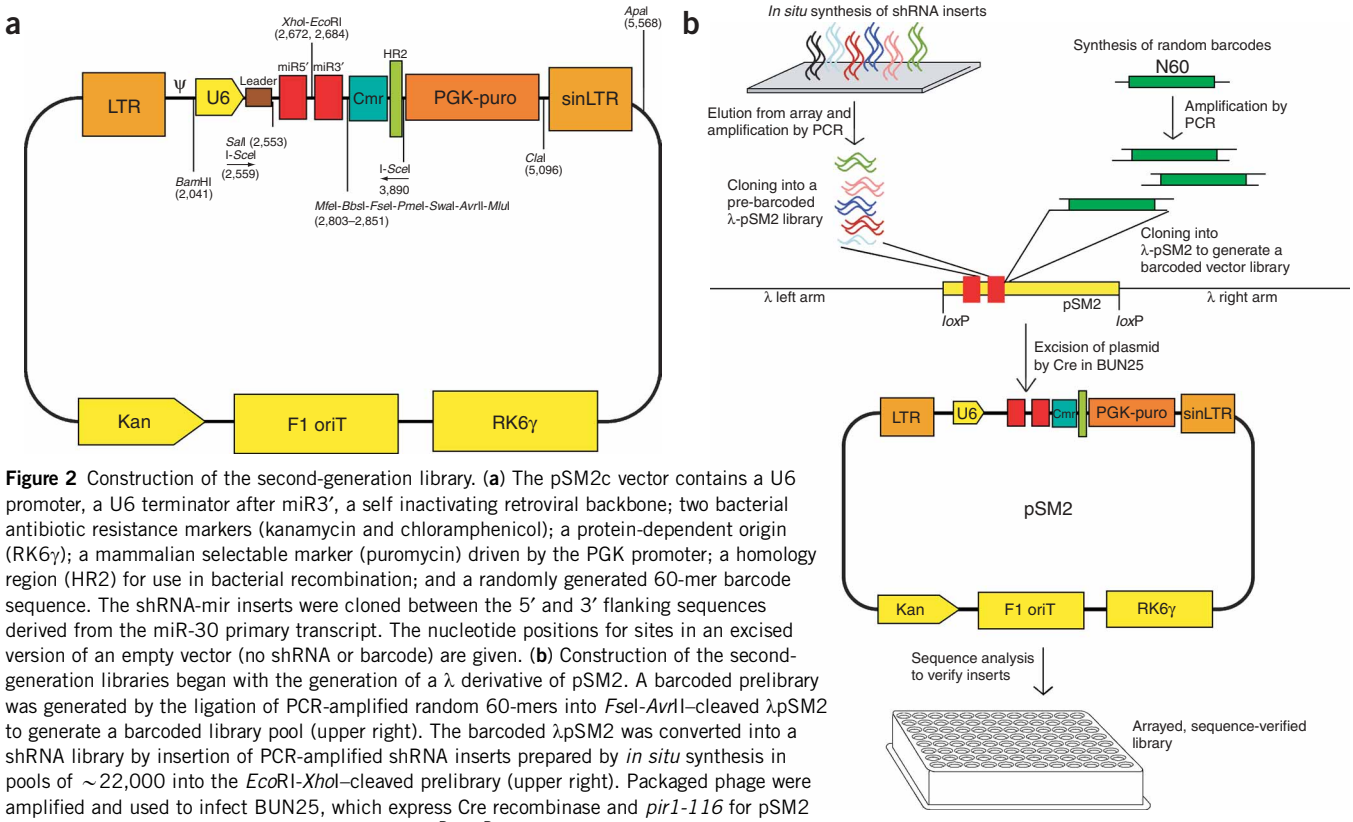

Figure 2 Construction of the second-generation library. (**a**) The pSM2c vector contains a U6 promoter, a U6 terminator after miR3′, a self inactivating retroviral backbone; two bacterial antibiotic resistance markers (kanamycin and chloramphenicol); a protein-dependent origin (RK6γ); a mammalian selectable marker (puromycin) driven by the PGK promoter; a homology region (HR2) for use in bacterial recombination; and a randomly generated 60-mer barcode sequence. The shRNA-mir inserts were cloned between the 5′ and 3′ flanking sequences derived from the miR-30 primary transcript. The nucleotide positions for sites in an excised version of an empty vector (no shRNA or barcode) are given. (**b**) Construction of the second-generation libraries began with the generation of a λ derivative of pSM2. A barcoded prelibrary was generated by the ligation of PCR-amplified random 60-mers into *Fse*I-*Avr*II–cleaved λpSM2 to generate a barcoded library pool (upper right). The barcoded λpSM2 was converted into a shRNA library by insertion of PCR-amplified shRNA inserts prepared by *in situ* synthesis in pools of ~22,000 into the *Eco*RI-*Xho*I–cleaved prelibrary (upper right). Packaged phage were amplified and used to infect BUN25, which express Cre recombinase and *pir1-116* for pSM2 replication. Each excision event gave rise to a Kan^R Cm^R colony. These were pooled and used for preparation of library DNA, which was transformed into BW F′DOT. Individual colonies were selected for sequence analysis.

The RNAi machinery can be programmed by exogenous or endogenous sources of double-stranded RNA. The most well-characterized source of endogenous triggers is microRNA (miRNA) genes[6,7]. It was initially assumed that miRNAs were transcribed from the genome as short hairpin RNAs[8] (shRNAs) that were directly processed by Dicer to yield the mature small RNAs that enter RISC[9–12]. Over the past year, however, a different picture has emerged. Numerous studies have shown that, in animals, miRNAs are transcribed by RNA polymerase II to generate long primary polyadenylated RNAs (pri-miRNAs)[13,14]. Through mechanisms not yet fully understood, the pri-miRNA is recognized and cleaved at a specific site by the nuclear Microprocessor complex[15–19]. This contains an RNase III family enzyme, Drosha, that cleaves the hairpin to produce a miRNA precursor (pre-miRNA) of ~70–90 nucleotides (nt) with a 2-nt 3′ overhang[15]. This distinctive structure signals transport of the pre-miRNA to the cytoplasm by a mechanism mediated by Exportin-5 (refs. 20,21). Only then is the pre-miRNA recognized by Dicer and cleaved to produce a mature miRNA. This probably involves recognition of the 2-nt 3′ overhang created by Drosha to focus Dicer cleavage at a single site ~22 nt from the end of the hairpin[22,23].

Mature miRNAs are superficially symmetrical, with 2-nt 3′ overhangs at each end generated by Drosha and Dicer, respectively. But the individual strands of the mature miRNA enter RISC in an unequal manner. As with small interfering RNAs (siRNAs), the thermodynamic asymmetry of the Dicer product is sensed such that the strand with the less stable 5′ end has a greater propensity to enter RISC and guide substrate selection[24,25]. This observation of thermodynamic asymmetry in small RNAs led to the development of rules for predicting effective siRNA sequences that have greatly improved the efficiency of those RNAs as genetic tools.

Several groups, including our own, previously described the design and construction of arrayed shRNA libraries that covered a fraction (approximately one-third) of human genes[26,27]. When these tools were developed, our knowledge of miRNA maturation was relatively incomplete. This led most groups to express a simple hairpin RNA that mimicked the pre-miRNA as an experimental trigger of the RNAi pathway.

Here we report the construction of a new generation of shRNA libraries (shRNA-mir) that uses our advancing understanding of miRNA biogenesis. As has previously proven successful in plants and animals, second-generation shRNA-mirs are modeled after endogenous miRNAs, specifically contained in the backbone of the primary miR-30 miRNA[28]. This natural configuration was up to 12 times more efficient than simple hairpin designs in producing the mature synthetic miRNAs. Additionally, we biochemically characterized the processing of these synthetic miRNAs, allowing us to predict the mature small RNA product(s) that will be generated from each vector. This has allowed us to select target sequences that maximize efficiency by directing preferential incorporation of the correct strand into RISC. Using these criteria, we produced and sequence-verified more than 140,000 shRNAs covering a substantial fraction of the predicted genes in the mouse and human genomes. We assayed the ability of a selected subset of shRNAs from the library to knock-down the expression of targeted genes by quantitative RT-PCR. We also tested this subset in a phenotypic assay and compared the performance of the first- and second-generation library designs. Overall, the shRNA-mir libraries

Table 1 Coverage by functional group

	Human		Mouse	
Group	Genes	Hairpins	Genes	Hairpins
Cancer	859	3,082	890	2,524
Cell cycle	531	2,166	482	1,552
Checkpoint	123	541	116	367
DNA repair	118	512	130	355
DNA replication	238	961	248	719
Enzymes	2,943	10,456	2,818	8,302
GPCRs	669	2,101	663	1,795
Kinases	618	2,648	575	2,250
Dual-specificity phosphatases	35	144	32	114
Tyrosine phosphatases	36	184	33	166
Phosphatases	206	765	187	628
Proteases	454	1,431	441	1,168
Proteolysis	302	1,458	270	858
Signal transduction	2,650	9,046	2,541	7,274
Protein trafficking	476	1,596	458	1,309
Transcription	820	2,865	767	2,209
Apoptosis	581	2,061	558	1,538

that we describe offer a convenient, flexible and effective tool for studying gene function in human cells. Additionally, for the first time to our knowledge, they extend the possibility of large-scale RNAi screens to mouse systems.

RESULTS

Design and construction of second-generation shRNA libraries

Expression of a simple, 29-bp hairpin from a U6 small nucleolar RNA (snRNA) promoter can induce effective suppression of target genes when delivered either transiently or stably from integrated constructs[26,29,30]. Longer hairpin structures are more effective inhibitors of gene expression than are shorter structures with stems of 19–21 nt. All these constructs, however, were designed to express a pre-miRNA hairpin, an intermediate in miRNA biogenesis, rather than a transcript that closely resembles a primary miRNA. Effective suppression could be achieved by redesigning an endogenous miRNA, miR-30, such that its targeting sequence was directed against a reporter gene[28]. We sought to compare directly the abundance of small RNAs produced from vectors with simple hairpin structures versus those that more closely resemble a natural miRNA. Because the efficient ectopic expression of endogenous miRNAs requires substantial flanking sequence[31], we developed a vector in which sequences from a remodeled miR30 were flanked by ~125 bases of 5′ and 3′ sequence derived from the primary transcript. Incorporation of appropriate cloning sites into this vector required altering only three positions in the precursor. We inserted this cassette into a vector equivalent to that in which we constructed our first-generation shRNA library (pSM1) and called the new shRNA vector pSM2. To distinguish the second-generation shRNAs from those in our first-generation library, we call them shRNA-mirs.

In order to use small RNA design rules to potentially enhance the efficacy of our shRNAs, we needed to understand how shRNA-mir was processed *in vivo*. To address this issue, we took advantage of existing studies of miR30 biogenesis that mapped its processing sites[32]. Using this information as a guide, we designed a series of constructs predicted to generate small RNAs targeting mouse p53, human PTEN and firefly luciferase. We transfected human 293 cells with pSM2 carrying each of these inserts and mapped the mature 3′ ends of the guide and passenger strands of p53 and PTEN shRNAs and the guide strand of the luciferase shRNA by random amplification of cDNA ends (RACE)-PCR (**Fig. 1a** and **Supplementary Fig. 1** online). Northern blotting indicated that maturation of shRNA-mirs produced 22-nt species (**Fig. 1b**); therefore, we were able to infer the 5′ end of each small RNA species. We considered the possibility of two processing sites at each end of the shRNA, because our analysis in the cases of p53 and PTEN shRNAs could not distinguish between processing at either of two terminal bases (**Fig. 1a**). In the case of luc1309, however, it was relatively obvious that the guide strand was cleaved in most cases (8 of 10) at the most 3′ indicated site. Two of ten sequences indicated that cleavage occurred 1 base 5′ of that site, perhaps reflecting a genuine heterogeneity in Drosha cleavage (RACE1, RACE3; **Supplementary Fig. 1**).

To test the performance of pSM2 relative to pSM1, we used both vectors to express a sequence targeting firefly luciferase. We inserted the sequence such that an identical mature small RNA would be generated from each construct after processing *in vivo* (**Supplementary Fig. 2** online). Of primary concern was the overall amount of mature small RNA generated from each construct, because dose-response experiments for shRNAs indicate that suppression correlates very well with the amount of RNA delivered[22], particularly at the relatively low doses expected to be achieved by expression from transfected or integrated constructs as compared with directly transfected synthetic RNAs. We transfected 293 cells with pSM1-luc and pSM2-luc, prepared RNA and assayed the processed small RNA by northern blotting. Cells transfected with pSM2-luc contained ~12 times more small RNA than did cells transfected with pSM1-luc (**Fig. 1b**).

Because primary miRNAs are transcribed mainly by RNA polymerase II (refs. 13,14), we wished to compare the performance of shRNA-mirs driven by various different promoters. We therefore cloned two different shRNA-mir cassettes targeting firefly luciferase downstream of three different RNA polymerase III promoters (tRNA-val[33], U6 (ref. 29) and H1 (ref. 34)) and two different RNA polymerase II promoters (MSCV-LTR and CMV[35]). These constructs were each prepared in a plasmid backbone that carried no other mammalian promoter. We transfected cells with each construct in combination with a homologous target expression plasmid encoding firefly luciferase and with a nontargeted reporter plasmid, encoding *Renilla* luciferase, as a means of normalization. We compared the performance of these plasmids in a four different cell lines, including two human, one mouse and one dog. We compared the ability of these constructs to suppress the luciferase target using a very efficient shRNA-mir (luc1309) and saw virtually no difference in the performance of the various promoters (**Fig. 1c**). But when we used a less efficient shRNA-mir (luc311), differences became apparent (**Fig. 1c**). In this, and numerous experiments with other shRNAs (data not shown), the U6 snRNA and CMV promoters gave the best and most consistent repression. On the basis of these results, we chose to retain the U6 snRNA promoter in our base library vector, pSM2. All our studies were carried out in transient assays. In situations in which constructs are stably integrated into the genome at single copy, different configurations of promoters and flanking sequences perform more efficiently than U6 (refs. 36,37). But we can also suppress gene expression by stable integration of pSM2 directly (**Supplementary Fig. 3** online).

On the basis of these tests, we constructed our second-generation shRNA library vector, pSM2 (**Fig. 2a**). The shRNA-mir expression cassette is carried in a self-inactivating murine stem cell virus. Expression of the small RNA is driven by the U6 snRNA promoter.

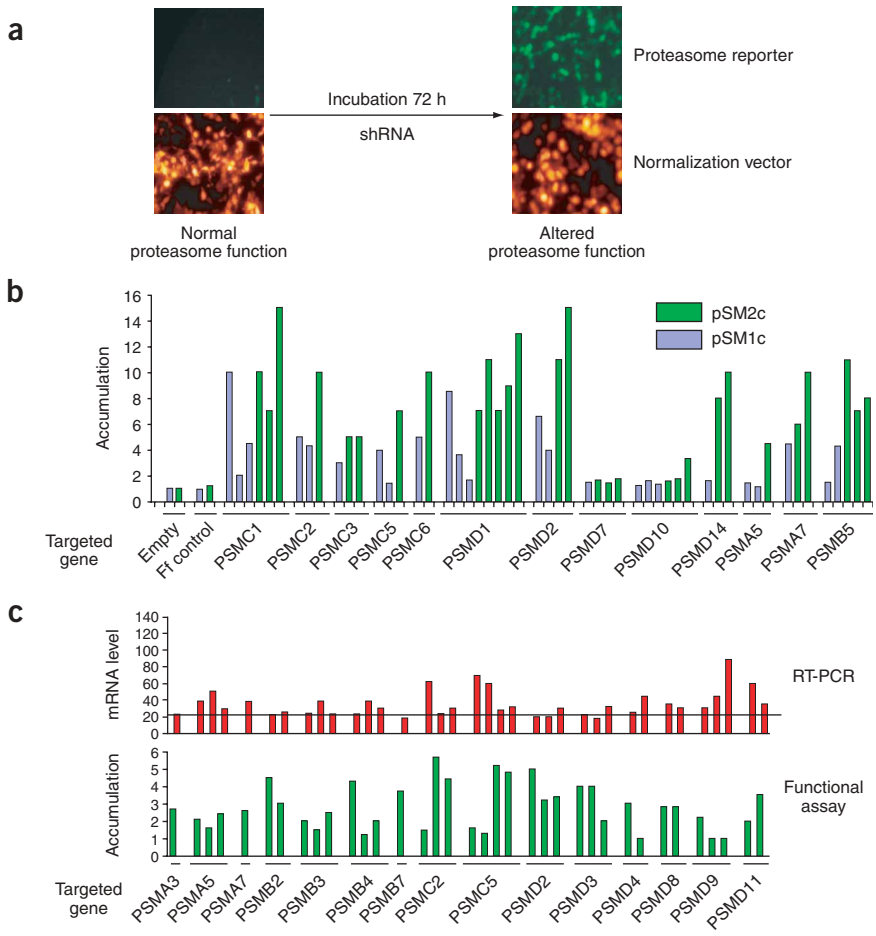

Figure 3 Validation of the second-generation library. (**a**) A schematic representation of the phenotypic assay for proteasome function. (**b**) Thirteen proteasome subunits were chosen because of their representation in both the first- and second-generation libraries (sequences are given in **Supplementary Table 1**). shRNA expression clones corresponding to each were assayed for activation of the proteasome reporter. Blue bars indicate first-generation clones; green bars indicate second-generation clones. In all cases, the activity of the proteasome reporter (green channel) was normalized for transfection using a DsRed expression plasmid. (**c**) In a separate study, 36 different proteasome shRNAs (sequences are given in **Supplementary Table 1**) were tested for their ability to suppress their target RNAs (upper panel). Quantitative RT-PCR was done 24 h after transfection of HeLa cells at an average efficiency of 80% as measured by a cotransfected normalization reporter (DsRed). The hypothetic maximum suppression, as calculated by transfection efficiency, is indicated by the black line. For comparison, functional assays for proteasome inhibition were done in parallel (lower panel).

As with the first-generation shRNAs, a U6 snRNA leader sequence lies between the promoter and the 5′ end of the miR-30 flanking region. We inserted synthetic oligonucleotides encoding shRNAs into *Xho*I and *Eco*RI sites in the miR-30 primary miRNA sequences. Immediately after the miR-30 cassette in each vector is a RNA polymerase III termination signal and a randomly generated 60-nt 'barcode' region to facilitate tracking of individual hairpin RNAs in complex populations. This feature is similar to that described for our first-generation RNAi library[26,30]. We designed the pSM2 vector such that inserts can be moved by an *in vivo* recombination strategy (MAGIC)[38]. One key difference between the first- and second-generation shRNA libraries is that the shRNA-mir cassette can be transferred without the constitutive U6 snRNA promoter. This allows the construction of mating recipients that contain inducible or tissue-specific promoters[36,37].

We designed six different shRNA-mir sequences for each of 34,711 known and predicted human genes and 32,628 mouse genes. In each case, we designed shRNAs such that the mature small RNA generated from each construct followed thermodynamic asymmetry rules that have been successfully applied for the design of siRNAs. In the approaches used to map the termini of the mature small RNAs generated from our vectors, we could not definitively distinguish between processing at two possible sites. Additionally, Dicer has been shown to generate some 3′ end heterogeneity in processing its substrates. Therefore, we chose sequences that gave similar thermodynamic profiles even if cleavage sites (**Fig. 1**) were shifted by a base at either end.

Construction of the library proceeded stochastically using a highly parallel *in situ* synthesis approach for oligonucleotide production (**Fig. 2b**). We synthesized groups of ~22,000 oligonucleotides, each containing a different shRNA-mir cassette, on glass-slide microarrays[39]. We eluted populations from the arrays and amplified them by PCR. To ensure efficient cloning, we inserted the pSM2 backbone into a lambda phage backbone such that it was flanked by *lox*P sites. λ-pSM2 contains unique *Xho*I and *Eco*RI sites for subcloning amplified hairpins and unique *Fse*I and *Avr*II sites for inserting barcode 60-mers. For λ-pSM2 barcodes, we first used a mixed library of random 60-nt sequences amplified with a primer set that included one primer with an *Fse*I site and one primer with a T7 promoter followed by the *Avr*II site. We 'lysogenized' amplified λ-pSM2 libraries containing barcodes into a strain that we constructed for this purpose, DH10βλ_{KP}, which has a wild-type *pir1* gene and the lambda repressor, *cI*, to allow λ-pSM2 to replicate as a 42-kb plasmid. We selected ~10^8 chloramphenicol- (Cm^R) and kanamycin-resistant (Kan^R) lysogens to serve as a barcoded library pool. We purified the barcoded λ-pSM2 libraries using cesium chloride gradient centrifugation; cleaved them with *Eco*RI and *Xho*I; ligated them to gel-purified, *Eco*RI-*Xho*I–cleaved, pooled shRNA-mir inserts from an individual chip; and packaged them. The average library size was ~5×10^7 recombinants per pool. To generate pSM2 library plasmid pools, we used the phage to infect an *Escherichia coli* strain that we constructed, BUN25, which expresses both Cre recombinase and the *pir1-116* gene, needed for high-copy RK6γ replication. We then transformed pooled plasmid libraries into a mating-competent host strain (BW F'DOT) and sequenced individual clones at random. Between 25% and 50% of clones had perfect inserts; we selected these and saved them as an arrayed set. We monitored accumulation of new clones from each pool dynamically; once a pool began to yield fewer unique clones per sequencing run, we halted sequencing and resynthesized the pool without the sequences that we had already obtained. We reiteratively

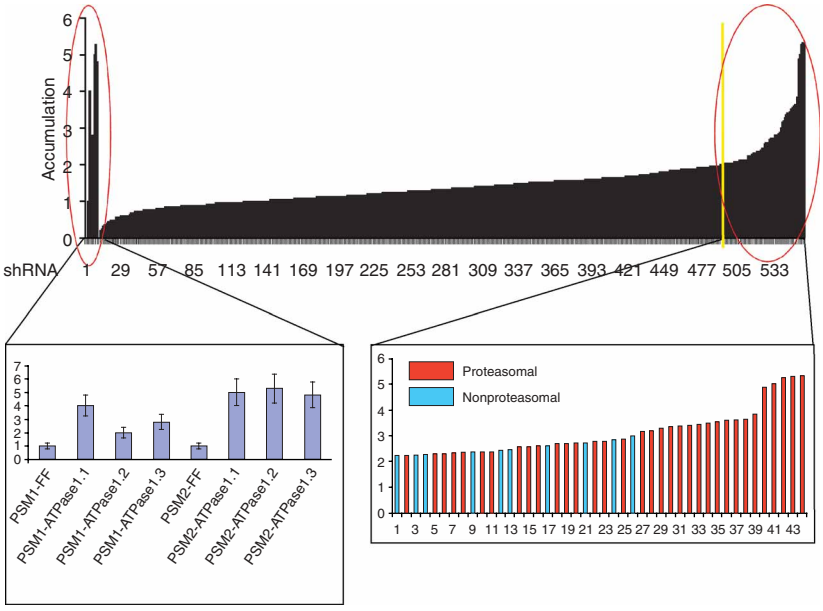

Figure 4 Performance of the second-generation library in a small-scale high-throughput screen. Forty-seven shRNAs targeting proteasome subunits were distributed among a series of 562 hairpins targeting human kinases (upper panel). The lower left panel shows the negative (FF) and the positive controls (ATPase1.1–ATPase1.3) from first- (pSM1) and second-generation (pSM2) libraries. The lower right panel shows the shRNAs that showed accumulation of the proteasome reporter over the cut-off (twofold or greater activation; yellow line). These are highly enriched for proteasomal shRNAs (red). Shown in blue are ten additional nonproteasomal shRNAs that also scored positive in the screen. Of these, five were also positive on a retest of individual clones. The sequences of the shRNAs, from left to right, for the lower right panel are given in **Supplementary Table 2**.

synthesized ~70 chips to maximize unique sequencing. Also, once we obtained three or more verified shRNAs for any given gene, we withdrew the remaining shRNAs targeting that gene from the population selected for resynthesis.

We have sequence-verified 79,805 shRNAs targeting 30,728 human genes and 67,676 shRNAs targeting 28,801 mouse genes so far. Coverage in selected functional groups is tabulated in **Table 1** for the mouse and human libraries. Our ultimate goal is to generate at least three shRNAs for each target locus. The full collection of mouse and human shRNAs, updated dynamically, can be accessed at RNAi Codex (see URL in Methods).

Validation of the second-generation shRNA libraries

To test the efficiency of the second-generation shRNA libraries, we used the same approach that we previously used to assess the performance of the first-generation reagents[26]. A green fluorescent protein (ZsGreen) reporter carrying the PEST domain of the mouse ornithine decarboxylase is normally degraded by the proteasome[40]. Therefore, cells with a destabilized ZsGreen expression plasmid show very low levels of fluorescence. Interference with proteasome function (*e.g.*, using a synthetic proteasome inhibitor) causes accumulation of the protein and a corresponding increase in fluorescence. The protein can also be stabilized by suppression of any gene required for proteasome function. Therefore, cotransfection of the reporters with an shRNA-mir expression plasmid can indicate whether a target protein is involved in the proteasome pathway (**Fig. 3a**). Using this assay as a primary test, we compared a series of shRNAs targeting proteasomal subunits that were obtained from either the first- or second-generation libraries (**Fig. 3b** and **Supplementary Table 1** online).

We chose 53 shRNAs targeting 13 different genes involved in proteasome function (**Fig. 3b**), 24 from the first-generation library and 29 from the second-generation library. We cotransfected cells with the reporter in combination with a plasmid encoding DsRed that allowed normalization of the transfections. The second-generation shRNAs performed substantially better than the first-generation shRNAs. Most of the plasmids derived from the second-generation library were as potent as the best shRNAs selected from a screen of the first-generation library.

To assess more thoroughly the performance of the second-generation libraries, we compared results from the proteasome assay for 36 shRNA-mir expression plasmids to suppression of target RNAs as measured by semiquantitative RT-PCR. We transfected HeLa cells with plasmids with ~80% efficiency, as measured by reference to a cotransfected reporter plasmid. Despite this incomplete transfection, all but 6 of the shRNA-mirs reduced the levels of their target RNAs by ~60% or more, and 13 of the 36 shRNA-mirs suppressed their targets by ~80%, the theoretical maximum (**Fig. 3c** and **Supplementary Table 1**). We obtained similar results with an additional 12 shRNAs that did not target proteasome subunits (data not shown). These studies also indicated that in some cases (*e.g.*, pSMB3), the functional assay did not show a large activation of the reporter despite substantial suppression of the targeted mRNA (**Fig. 3c**). Therefore, the functional assay slightly underestimated the efficacy of the library.

To test the performance of the library on a larger scale, we assayed a set of 515 kinase shRNAs that contained 47 hairpins directed to proteasome subunits using the phenotypic assay for proteasome function and a high-throughput protocol in 96-well plates (**Fig. 4**). In this context, 34 of 47 shRNAs targeting the proteasome scored positive (72%), compared with 10 shRNAs that had not previously been linked to proteasome function (1.9%). A secondary screen of those 44 potential positives from the primary screen showed positive signals from all 34 proteasomal shRNAs. Only 5 of the 10 nonproteasomal RNAs continued to activate the reporter, and none of these scored with more than one shRNA in the library (**Supplementary Table 2** online).

DISCUSSION

Since the discovery that an RNAi pathway was conserved in mammals, the exploitation of this silencing response as a genetic tool has evolved in concert with our deeper understanding of its biochemical mechanism. The initial applications of siRNAs as triggers of the silencing response required comprehension of the way in which Dicer processes long double-stranded RNA substrates in *Drosophila melanogaster*[41]. Similarly, studies of dicer-mutant *Caenorhabditis elegans* showed that endogenous loci could encode triggers of the RNAi machinery, and this led to the idea that such loci could be altered to target genes for experimental silencing[9–12].

Many strategies have been developed for producing miRNA-like triggers of the RNAi pathway. We mapped the processing sites on

precursor shRNA-mirs and can predict which small RNA is generated from each shRNA-mir expression vector. This enables us to apply siRNA design rules to shRNA-mir expression cassettes. A combination of increased small RNA production with better shRNA design yielded a pronounced increase in the performance of these silencing tools.

Guided by these design strategies, we constructed large libraries of sequence-verified shRNAs targeting most known and predicted genes in the human and mouse genomes. On average, each locus is currently covered by two shRNA-mirs, but our ultimate goal is to have three sequence-verified shRNA-mirs for each gene. The second-generation libraries resemble those that we previously reported in that they reside in flexible vectors that permit shuttling of shRNA-mir expression cassettes into virtually any desired expression vector using a bacterial mating strategy[38]. A unique feature of the second-generation library is that the expression cassette can be moved without also moving the constitutive U6 snRNA promoter also. This permits large-scale construction of secondary libraries under the control of tissue-specific and inducible promoters. Regulated expression of our library cassettes from RNA polymerase II promoters effectively suppresses gene expression both in cultured cells and in animals[36,37]. These recipient vectors can be used directly with any shRNA-mir encoded by the library described here.

The use of large-scale resources for suppressing gene expression by RNAi promises to revolutionize genetic approaches to biological problems in numerous model systems. The libraries described here should be useful in assessing the functions of individual genes and in genome-wide approaches. Strategies reported in accompanying papers[36,37] will permit large-scale application of these tools to screens that require long-term suppression of gene expression using single-copy integrants or inducible repression. Therefore, this coherent system of RNAi reagents can be used in both mouse and human experimental procedures.

METHODS

Construction of the lysogenic strain DH10βλ_{KP} and excision strain BUN25. Strain DH10βλ_{KP} [*mcrA* Δ*(mrr-hsdRMS-mcrBC)* ϕ*80 lacZΔM15* Δ*lacX74 deoR recA1 endA1 ara*Δ*139* Δ*(ara, leu)7697 galU galK* λ$^-$ *rpsL nupG tonA* λ*-pir1-npt*] containing λcI and the *pir1* gene was constructed to 'lysogenize' the λSM2 barcode library before introduction of the hairpin fragments. To generate this strain, we constructed λ_{KP} containing the *pir1* and Kan^R genes. To generate λ_{KP}, we amplified the *pir1* gene from BW23473 using primers MZL393 and MZL51 and cloned it into the pCR2.1 TOPO TA cloning vector. We excised the *pir1* gene from this clone on a *Bam*HI fragment and ligated it into *Bam*HI-cleaved pSE356, which contains an *npt* gene and a *Bam*HI restriction site flanked by two 1-kb λ DNA fragments[42] to generate pSE356pirWT. We recombined the *pir1* Kan^R fragment onto wild-type λ by amplifying λ on LE392/pSE356pirWT, collected the resulting phage and used them to infect DH10β. We infected 100 μl of DH10β cells with 10^6 plaque-forming units (PFUs) at 30 °C for 30 min in Luria broth plus 10 mM $MgSO_4$, diluted them with 900 μl of Luria broth, incubated them at 30 °C for 2 h with shaking and plated them on Luria broth containing 50 μg ml^{-1} kanamycin at 37 °C overnight to select λ_{KP} lysogens. We tested the ability of lysogens to lysogenize λ vectors containing R6Kγ origins of replication as extrachromosomal elements. We selected a strain capable of doing this and named it DH10βλ_{KP}.

Strain BUN25 [F' traD36 lacI^q Δ(lacZ)M15 proA$^+$B$^+$/e14$^-$(McrA$^-$) Δ(lac-proAB) thi gyrA96 (Nal^r) endA1 hsdR17 ($r_k^-m_k^+$) relA1 glnV44 λ-cre-npt umuC::pir116-Frt sbcDC-Frt] containing *pir1-116* and *cre* was constructed to allow the conversion of λSM2 shRNA libraries into pSM2 shRNA libraries. We generated a PCR fragment containing *pir1-116* using primers MZL393 and MZL51 (**Supplementary Table 3** online), cleaved it with *Bam*HI and ligated it into *Bam*HI-cleaved pUC18 to generate pML284. We isolated a fragment containing *Bst*BI-Frt-*cat*-Frt-*Nde*I (filled-in) from KD3 (ref. 43) and inserted it into the *Sma*I site of pML284 to generate pML334. We isolated an *Hpa*I fragment containing *UmuDC* from pSE117 and cloned it into pBluescript *Xho*I (filled in)-*Eco*RV to generate pML236. We eliminated one of the *Bam*HI site on pML236 by digesting it with *Pst*I-*Xba*I, filling in with T4 DNA polymerase and ligating. We isolated the Frt-*cat*-Frt-*pir116* sequence from pML334 as a *Kpn*I-*Sac*I (filled-in) fragment and ligated it into *Mlu*I-*Bam*HI (filled-in)-cleaved pML236-ΔBamHI to generate pML346. We integrated the 3.8-kb *Kpn*I-*Sac*I *UmuDC*-Frt-*cat*-Frt-*pir116-UmuC* fragment from pML346 into BNN132/pML104 by homologous recombination using the λ recombinase expressed from pML104 and confirmed it by colony PCR. We removed the *cat* gene by FLP-mediated excision *in vivo* using pCP20 (ref. 44), which expresses the FLP recombinase, to generate BUN24. We amplified a cassette that has Frt-*cat*-Frt flanked by 50 bp of homology to *sbcD* and 50 bp of homology to *sbcC* by primers MZL493 and MZL494 using KD3 as a template and used this cassette to replace *sbcD* and part of *sbcC* on BNN132 by homologous recombination. The deletions were confirmed by colony PCR. The strain was named BNN132sbcDC-Frt-cat-Frt. We then used a pair of outside primers (MZL495 and MZL496) that gave about 500 bp of homology to the upstream of *sbcD* and 500 bp of homology to the *sbcC* to amplify a PCR product from BNN132sbcDC-Frt-cat-Frt to recombine onto the *sbcDC* region of BUN24. The resulting strain was named BUN25 and is used to stabilize inverted repeats in *E. coli*[45].

Library vector construction. We inserted a pair of *lox*P-*Not*I-*lox*P duplexed oligos (MZL524 and MZL525) into the pSM2 *Bst*XI site to generate pSM2c-loxP. We inserted a second pair of duplexed oligos (MZL541 and MZL542), carrying the proper restriction sites for cloning barcodes into λSM2, into the *Bbs*I-*Mlu*I sites of pSM2c-loxP to create pML375. We digested λACT2 with *Not*I, gel-purified the λ arms and ligated them to *Not*I-digested pML375 to generate λSM2. We packaged the ligation mixture using MaxPlax lambda packaging extracts from Epicentre. We selected a λSM2 lysogen by infecting 200 μl of BW23473 cell (A_{600} = ~0.8) with 100 μl of λSM2 packaging mix in the presence of 10 mM $MgSO_4$ and 0.2% (w/v) maltose, incubated it at 30 °C for 30 min, added 900 μl of Luria broth, incubated it at 30 °C for 2 h with shaking to express the Cm^R marker and plated it on Luria broth containing 17 μg ml^{-1} of chloramphenicol at 30 °C overnight. The proper recombinants were confirmed by restriction analysis. Oligonucleotide sequences are given in **Supplementary Table 3**.

Barcode library construction. We amplified the 60-bp barcode oligos using barcode primers 1 and 2 (**Supplementary Table 3**). For PCR, we used 0.1 pmol of barcodes, 50 pmol of each primer, 25 nmol of each dNTP and 2.5 U of Taq DNA polymerase and the following conditions: 94 °C for 45 s; 13 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s; 72 °C for 10 min. We pooled ten PCR reactions together, purified them using a QIAquick PCR purification kit, digested them with *Bam*HI, *Eco*RI, *Xho*I and *Sal*I to remove these sites in the barcodes and gel–purified them. We digested the purified barcodes with *Fse*I and *Avr*II and then purified them using the QIAquick gel extraction kit. We mixed 2 μg of *Fse*I-*Avr*II–digested λSM2 ligated with 10 ng of *Fse*I-*Avr*II–digested barcodes with 1× ligation buffer and 0.5 μl of T4 DNA ligase in a 5-μl final volume and incubated it at 16 °C overnight. We packaged the ligation mixture and amplified it. The size of the λSM2-barcode library was 4.2×10^7. We used 20 ml of DH10βλ_{KP} cells (A_{600} = ~1) to lysogenize 2×10^9 of λSM2-barcode library as 42-kb plasmids. We mixed the cell and the phage in the presence of 10 mM $MgSO_4$ and 0.2% (w/v) maltose, incubated the mixture at 30 °C for 30 min and added 200 ml of Luria broth to recover at 30 °C for 1 h by shaking. We concentrated the mixture by centrifugation at 4,000 r.p.m. for 20 min, resuspended the pellet in 3 ml of Luria broth, plated the mixture on 10 large Luria broth plates with 17 μg ml^{-1} of chloramphenicol and incubated it at 30 °C overnight. We scraped the cells from plates and grew them in 3 l of terrific broth containing 17 μg ml^{-1} of chloramphenicol overnight. We prepared supercoiled λSM2-barcode library DNA using cesium chloride. The lysogenization efficiency was ~30%.

Oligonucleotide cleavage and PCR amplification. To collect oligonucleotides, we treated microarrays for 2 h with 2–3 ml of 35% NH_4OH solution (Fisher Scientific) at room temperature. We transferred the solution to 1.5-ml micro-centrifuge tubes and subjected it to speed vacuum drying at medium heat

(~55 °C) overnight. We resuspended the dried material in 200 µl of RNase- and DNase-free water and carried out PCR amplification in 50-µl reaction volumes using Invitrogen Platinum *Pfx* DNA Polymerase. To obtain a sufficient amount of PCR product, we did four 50-µl reactions for each sample. Each reaction contained 2× *Pfx* PCR amplification buffer, 0.3 mM of each dNTP, 1 mM $MgSO_4$, 0.3 µM of each primer, 0.5× PCR enhancer solution, 0.5 units of Platinum *Pfx* DNA Polymerase and 10 µl of template DNA. The primers used for amplification were 5′-mir30-PCR-xhoI-F and 3′-mir30-PCR-ecorI-R (sequences available on request). After an initial denaturation step of 94 °C for 5 min, DNA amplification occurred through 25 cycles of denaturing at 94 °C for 45 s and annealing and extension at 68 °C for 1 min and 15 s, followed by a final 7-min extension at 68 °C. We then combined the four reactions into one tube, cleaned up the PCR product using the QIAGEN MinElute PCR Purification Kit and eluted it in a total volume of 26 µl.

shRNA library construction. We digested the λSM2 barcode library and shRNA PCR products with *Eco*RI and *Xho*I overnight and gel-purified them. We set up ligations with shDNA oligos using 1.5 µg of *Xho*I-*Eco*RI–cleaved vector, 8–10 ng of *Xho*I-*Eco*RI–cleaved inserts generated from the PCR of shDNA oligos from the parallel microarray synthesis, 1 µl of 10× ligation buffer, 0.5 µl of T4 DNA ligase and water to a final volume of 10 µl. We incubated the ligation mixtures at 16 °C overnight and packaged them. We typically observed 30- to 90-fold stimulation of PFUs and 2×10^7 to 8×10^7 PFUs total for each library pool. We typically amplified 2×10^7 PFUs for each pool to generate a stock. To verify the ligation efficiency, we excised 10 µl of package mix by infecting 100 µl of BUN25 ($A_{600} = \sim 0.5$) and selected colonies on Luria broth with 30 µg ml^{-1} of chloramphenicol at 30 °C overnight. We carried out colony PCR using forward and reverse primers (**Supplementary Table 3**) and typically observed 85–95% correctly sized inserts, with some containing multiple inserts. To generate plasmid DNA from these libraries, we typically excised 5×10^7 PFUs through infection of BUN25 cells as described earlier. The cells were scraped from plates and grown in 2 l of Luria broth plus 13 g l^{-1} of circle growth at 37 °C for 7–8 h. We used the cesium chloride method to prepare DNA. DNAs were transformed into BW23474 F'DOT SbcC, and individual clones were sequenced using primer5′ (sequence available on request). Correct clones were individually rearrayed to form the final library.

Small RNA northern blots. We transfected 293 cells in 10-cm dishes at 60% confluency with 15 µg of shRNA plasmid DNA along with 5 µg of pDsRed-N1 (Clontech) using TransIT-LT1 (Mirus). At 48 h after transfection, we confirmed transfection efficiency by estimating the percentage of cells expressing DsRed (~80%) and then extracted and purified total RNA using Trizol. We carried out small RNA northern blots as described[46] using 30 µg of RNA per lane. For hairpins targeting EGFP at starting position 481 (**Fig. 1b**), northern probes were DNA oligos corresponding to the antisense strand (sequence available on request) of the mature RNA.

Proteasome assays. We grew bacteria cultures in 96-well plates for 36 h in GS96 medium (Bio101). We extracted plasmid DNA using Qiagen Ultrapure plasmids minipreps in a 96-well plate format. We determined DNA concentrations by mixing an aliquot of each sample with picogreen (Molecular Probes) and determining fluorescence on a Victor2 plate reader. We plated 1×10^5 HEK 293T cells per ml in 96-well optic plates (Corning). For the proteasome assay, we cotransfected cells in each well with 12.5 ng of pDsRed-N1 (Clontech), 5 ng of the Zsprosensor (Clontech) and 75 ng of each individual shRNA construct using 0.3 µl of LT-1 (Mirus) transfection reagent. After 24 h, we replaced the transfection medium. After 72 h, we removed the medium and replaced it with phosphate-buffered saline in order to read fluorescence. Fluorescence signals were read on a Victor2 plate reader. Signals in the green channel were normalized to transfection efficiency using customized scripts with fluorescence in the red channel serving as a normalization criterion. Cut-offs were assigned by using control shRNA transfections to determine the range for a negative outcome.

Plasmid transfections and mRNA quantification. We seeded 0.5×10^5 HeLa cells per well in 24-well plates and transfected them 24 h later with 1 µg per well of the appropriate plasmid. We delivered each plasmid to four wells using Lipofectamine 2000 (Invitrogen) in accordance with the manufacturer's protocols. We determined transfection efficiency by parallel transfection of a GFP-expressing plasmid and assayed the percentage of fluorescent cells by flow cytometry. For analysis of target gene mRNA knock-down, we collected cell lysates 24 h after transfection and prepared total RNA using RNeasy columns (Qiagen) in accordance with the manufacturer's protocols. We quantified mRNA by real-time PCR of reverse transcription products, using available Applied Biosystems TaqMan primer probe sets, and determined the percent mRNA remaining by comparison with mRNA levels from cells treated with transfection reagent alone.

URL. RNAi Codex is available at http://codex.cshl.edu/.

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

ACKNOWLEDGMENTS

We thank members of the laboratories of G.J.H., S.J.E. and S.W. Lowe for suggestions; J. Magnus for assistance with array PCRs; L. Nascimento, V. Balija, M. Kramer, T. Zutavern, S. Muller and B. Miller for assistance with sequencing; T. Moore for help with curation of the collection; and P. Linsley and S. Friend for their support of this project. This work was funded in part by awards from the Department of Defense Breast Cancer Research Program (G.J.H. and S.J.E.) and the US National Institutes of Health (G.J.H. and S.J.E.). S.J.E. and G.J.H. are investigators of the Howard Hughes Medical Institute.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

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

1. Nakayashiki, H. *et al.* RNA silencing as a tool for exploring gene function in ascomycete fungi. *Fungal Genet. Biol.* **42**, 275–283 (2005).
2. Tang, G. & Galili, G. Using RNAi to improve plant nutritional value: from mechanism to application. *Trends Biotechnol.* **22**, 463–469 (2004).
3. Dasgupta, R. & Perrimon, N. Using RNAi to catch *Drosophila* genes in a web of interactions: insights into cancer research. *Oncogene* **23**, 8359–8365 (2004).
4. Fraser, A. Towards full employment: using RNAi to find roles for the redundant. *Oncogene* **23**, 8346–8352 (2004).
5. Silva, J., Chang, K., Hannon, G.J. & Rivas, F.V. RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age. *Oncogene* **23**, 8401–8409 (2004).
6. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
7. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**, 522–531 (2004).
8. Reinhart, B.J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
9. Ketting, R.F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
10. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
11. Knight, S.W. & Bass, B.L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269–2271 (2001).
12. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834–838 (2001).
13. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
14. 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).
15. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
16. 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).
17. 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).
18. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).
19. Gregory, R.I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).
20. 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).

21. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
22. Siolas, D. *et al*. Synthetic shRNAs as potent RNAi triggers. *Nat. Biotechnol.* **23**, 227–231 (2005).
23. 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).
24. Schwarz, D.S. *et al*. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
25. Khvorova, A., Reynolds, A. & Jayasena, S.D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
26. Paddison, P.J. *et al*. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
27. Berns, K. *et al*. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
28. 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).
29. 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).
30. Westbrook, T.F. *et al*. A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**, 837–848 (2005).
31. Chen, C.Z., Li, L., Lodish, H.F. & Bartel, D.P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
32. Zeng, Y. & Cullen, B.R. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**, 112–123 (2003).
33. Kawasaki, H. & Taira, K. Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells. *Nucleic Acids Res.* **31**, 700–707 (2003).
34. Brummelkamp, T.R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
35. Zheng, L. *et al.* An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl. Acad. Sci. USA* **101**, 135–140 (2004).
36. Dickins, R.A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nat. Genet.,* advance online publication 2 October 2005 (doi:10.1038/ng1651).
37. 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).
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. Cleary, M.A. *et al.* Production of complex nucleic acid libraries using highly parallel *in situ* oligonucleotide synthesis. *Nat. Methods* **1**, 241–248 (2004).
40. Li, X. *et al.* Generation of destabilized green fluorescent protein as a transcription reporter. *J. Biol. Chem.* **273**, 34970–34975 (1998).
41. Carmell, M.A. & Hannon, G.J. RNase III enzymes and the initiation of gene silencing. *Nat. Struct. Mol. Biol.* **11**, 214–218 (2004).
42. Elledge, S.J. & Walker, G.C. Phasmid vectors for identification of genes by complementation of *Escherichia coli* mutants. *J. Bacteriol.* **162**, 777–783 (1985).
43. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).
44. Cherepanov, P.P. & Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9–14 (1995).
45. Chalker, A.F., Leach, D.R. & Lloyd, R.G. *Escherichia coli* sbcC mutants permit stable propagation of DNA replicons containing a long palindrome. *Gene* **71**, 201–205 (1988).
46. Caudy, A.A., Myers, M., Hannon, G.J. & Hammond, S.M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* **16**, 2491–2496 (2002).