AD_____

Award Number: DAMD17-03-1-0529

TITLE: Using RNA Interference to Reveal Genetic Vulnerabilities in Human Cancer Cells

PRINCIPAL INVESTIGATOR: Despina Siolas
Gregory Hannon, Ph.D.

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory
Cold Spring Harbor, NY 11724

REPORT DATE: July 2005

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20060315 029

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-07-2005	Annual Summary	1 Jul 2004 – 30 Jun 2005

4. TITLE AND SUBTITLE
Using RNA Interference to Reveal Genetic Vulnerabilities in Human Cancer Cells

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-03-1-0529

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Despina Siolas
Gregory Hannon, Ph.D.

E-mail: siolas@cshl.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cold Spring Harbor Laboratory
Cold Spring Harbor, NY 11724

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The sequencing of the human genome and recent advances in the application of RNA interference in both *in vitro* and *in vivo* mammalian systems have given scientists the ability to conduct in depth analyses of gene function. High throughput screens in cultured mammalian cells can be performed due to the advent of valuable genetic tools such as the Hannon-Elledge shRNA library. Concurrently, tremendous strides have been made in the field of cancer biology through the use of targeted therapeutics that enable a finer treatment of cancer cells. My goals are to augment our knowledge of the molecular basis of cancer and our ability to fight cancer through the use of the innovative genetic tool, RNA interference. My work studies the biology of shRNA processing and applying that knowledge to improve our short hairpin library. In addition, selection screens are being conducted targeting genes affecting apoptosis and growth arrest in cancer cells that would make suitable drug targets. This will be followed by comprehensive studies of the biological functions of these genes. These studies will provide great insight into the genes involved in predisposing normal and transformed breast cells to apoptosis and growth arrest.

15. SUBJECT TERMS
Cancer, RNAi, gene silencing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	42	

Table of Contents

Cover.....

SF 298.....2

Table of Contents.....3

Introduction.....4

Body.....6

Key Research Accomplishments.....8

Reportable Outcomes.....8

Conclusions.....9

References.....9

Appendices.....11

Introduction:

The completion of the Human Genome Project in 2003 unearthed the field of functional genomics as a new challenge to understanding humans at the molecular level. Efficient interpretation of the functions of human genes requires resources and strategies to be developed to enable large-scale investigations across entire genomes. Recent advances in the field of RNA interference have enabled researchers to conduct large-scale loss-of-function studies in mammals and address the trials presented by functional genomics.

Most eukaryotic cells harbor a natural response to double-stranded RNAs (dsRNA) that inhibits gene expression in a sequence-specific manner[1]. DsRNA silencing triggers are processed into small RNAs (siRNAs and miRNAs) that engage the RNA-induced silencing complex (RISC) to suppress expression of homologous targets. In cases in which the small RNA is perfectly complementary to the target, that RNA is cleaved and ultimately destroyed[1,2]. This pathway, known as 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[3-7].

In mammals, RNAi can be initiated in several ways. The most prevalent method of triggering RNAi is the delivery of one or more small interfering RNAs (siRNAs). SiRNAs are duplexes of ~21-22 nucleotides that bear two nucleotide 3' overhangs[1,8]. One strand (the guide strand) of the siRNA is incorporated into the effector complex of RNAi, the RNA-induced Silencing Complex, RISC and guides substrate selection via base pairing to its complementary target[1,2]. The RNAi machinery can also be programmed by endogenous sources of double-stranded RNA. The most well characterized source of endogenous triggers for the RNAi machinery are the microRNA genes[9,10]. Numerous studies have demonstrated that, in animals, miRNAs are transcribed to generate long primary polyadenylylated RNAs (pri-miRNAs)[11,12]. Through mechanisms not yet fully understood, the pri-microRNA is recognized and cleaved at a specific site by the nuclear Microprocessor complex[13-17] to produce a ~70-90 nucleotide microRNA precursor (pre-miRNA) which is exported to the cytoplasm [18,19]. Only then is the pre-miRNA recognized by Dicer and cleaved to produce a mature microRNA. This probably

involves recognition of the 2 nucleotide 3' overhang created by Drosha to focus Dicer cleavage at a single site ~22 nucleotides from the end of the hairpin[20].

Previously, several groups, including our own, described the design and construction of arrayed short hairpin RNA (shRNA) libraries that covered a fraction (~1/3) of human genes[21,22]. At the time when these tools were developed, our knowledge of microRNA maturation was relatively incomplete. This led most groups to the notion of expressing a simple hairpin RNA that mimicked the premiRNA. As our knowledge of the microRNA processing pathway and our understanding of strand preferences for RISC loading have grown, it seemed prudent to reevaluate whether the performance of encoded triggers of the RNAi pathway might be improved by remodeling a primary miRNA transcript to experimentally alter its targeting capability. Indeed such strategies have previously succeeded in both plants and animals[23,24].

My initial studies focused on the biology of miRNA processing which guided the new design of the Hannon-Elledge shRNA library. Concurrently, we have been testing shRNAs and developing screens that would allow us to target genes involved in apoptosis and growth arrest. Our approach is studying synthetic lethal interactions with p53 in cancer cells. This tumor suppressor protein induces apoptotic cell death in response to oncogenic stress. Loss of p53 function causes malignant progression through a mutation in the gene that encodes p53 or by defects in the signaling pathways that are upstream or downstream of p53[25]. The ultimate goal is to use RNAi as a genetic tool to find molecular vulnerabilities unique to breast cancer cells. These vulnerabilities are potential chemotherapeutic targets that can be exploited to kill cancer cells.

Body:

The biological studies described in my last update and our lab's ability to synthesize complex oligonucleotide populations using *in situ* microarray DNA synthesis contributed to the evolution of the new Hannon Elledge library which will enable a more effective genetic loss of function studies in mammalian systems. The result of this work is attached in the appendix, entitled "Second-Generation shRNA Libraries Covering the Mouse and Human Genomes." In addition, I am currently in the middle of conducting a screen searching for genes that would be synthetic lethal with p53. p53 inhibits tumor cell growth by evoking several responses to malignancy-associated stress signals including cell-cycle arrest, senescence, and apoptosis, with the option chosen being dependent on many factors that are both intrinsic and extrinsic to the cell[25]. p53 can also contribute to the repair of genotoxic damage, potentially allowing for the release of damaged cells back into the proliferating pool. Mutations in the p53 gene occur in about half of all human cancers, almost always resulting in the expression of a mutant p53 protein that has acquired transforming activity[25].

Synthetic lethality allows us to functionally define vulnerabilities in cancer cells because cancer arises from genetic lesions in somatic cells. Synthetic lethal interactions occur when mutations in two or more nonallellic genes synergize to kill cells. For example, a particular mutation may be tolerated when singly present in cells, but when combined may result in cell death. Thus, synthetic lethal interactions reveal situation in which cellular homeostasis is altered by a molecular lesion so that the action of another gene or pathway is required to compensate[26]. The fact that cancer cells arise from genetic alterations makes synthetic lethality ideally suited for identifying cellular targets required by cancer cells for viability.

For our initial screen, we have selected two cells lines that are as genotypically identical as possible except for their p53 status. We are using HCT116 and HCT116 p53 null colon cancer cell lines. While these cell lines are not breast cancer cell lines, they can serve as a basis for a more developed screen in MCF7 or MCF10A cell lines. We can work out the conditions for our screen in these genotypically and phenotypically well characterized cell lines and then expand into breast cancer cells. This would serve to verify our initial results and also expose genes that are responsible for apoptosis in combination with p53 loss across multiple cancers. We can also

compare genes that would cause the p53 null cancer cells to die versus cancer cells with wild type p53 .

My screen began with transfecting phoenix packaging cells with our new retroviral shRNA library constructed in a new format as described in the appendix. Four shRNA library subsets were used targeting human kinases, dual specificity phosphatases, protein tyrosine phosphatases, and a c600 control set that contains hairpins for proteasome subunits, cell proliferation genes and barcodes. Protein kinases are critical components of cellular signaling cascades that control cell proliferation and other responses to external stimuli. Kinases are attractive drug targets as their dysfunction can result in cancer. Viruses were pooled and both cell lines were infected separately such that each cell is targeted to carry on average one copy of the hairpin expression cassette. There are several advantages to this approach over transiently transfected screens: The knockdown effects can be monitored over extended periods, shRNA expression is more normalized, thereby facilitating the screening of cells in pools, and finally, this approach is very adaptable for high throughput studies[27].

Cell viability can be assayed in two ways. One will be using an fluorescent MTT dye reduction assay. Alternatively, genomic DNA will be extracted from cells at 0, 5, 10 and 15 days after hairpin selection with puromycin. Cells will be trypsinized, pooled and replated allowing them to grow to near confluency. This will entail a negative selection method where the fate of individual shRNAs in a complex population will be monitored by adopting a DNA-barcoding strategy. Each hairpin is linked to a 60mer DNA barcode that allows us to virtually count the number of cells that contain a specific shRNA cassette by looking at barcode representation in a cell population on a microarray. Genomic DNA extracted from each sample will be assayed for the presence of DNA barcodes by PCR. This DNA will be amplified using a primer containing a T7 RNA polymerase promoter sequence that allows for *in vitro* transcription of fluorescently labeled single-stranded RNA that will be subsequently hybridized to an custom microarray containing complements of these sequences. This technology will illustrate cell death as a loss of barcode representation in a population of cells. Comparison between the hybridization patterns of the different DNA samples over time allows the identification of shRNAs that are synthetically lethal with p53 and thus cause apoptosis or cell growth arrest.

Key Research Accomplishments:

- Assisting in the construction and validation of second-generation shRNA (shRNAmir) expression libraries that have been designed based on an increased knowledge of RNAi biochemistry. We have generated large-scale arrayed, sequence-verified libraries comprising more than 140,000 shRNAmir expression plasmids, covering a substantial fraction of all predicted genes in the human and mouse genomes.
- The initiation of a p53 synthetic lethal screen in cancer cells that can identify genes that are involved in apoptosis and growth arrest.

Reportable Outcomes:

Manuscripts:

Jose M. Silva, Mamie Z. Li, Ken Chang, Wei Ge, Michael C. Golding, Ricky Rickles, **Despina Siolas**, Guang Hu, Patrick J. Paddison, Michael R. Schlabach, Nihar Sheth, Jeff Bradshaw, Julia Burchard, Amit Kulkarni, Guy Cavet, Ravi Sachidanandam, W. Richard McCombie, Michele A. Cleary, Stephen J. Elledge, and Gregory J. Hannon. Second-Generation shRNA Libraries Covering the Mouse and Human Genomes. (submitted: *Nature Genetics* June 2005)

Siolas, D., Lerner C., Burchard J., Ge W., Linsley PS., Paddison PJ., Hannon GJ., Cleary MA. (2005) Synthetic shRNAs as potent RNAi triggers. *Nature Biotechnology*. 23(2):227-31

Presentations:

Minisymposium Talk:

Siolas, D., Lerner C., Burchard J., Ge W., Linsley PS., Hannon GJ., Cleary MA. (2005) Synthetic shRNAs as potent RNAi triggers. *American Association for Cancer Research Conference* Minisymposium Presentation, Anaheim, California, USA

Poster:

Siolas, D., Lerner C., Burchard J., Ge W., Linsley PS., Hannon GJ., Cleary MA. (2005) Synthetic shRNAs as potent RNAi triggers. *Era of Hope, Department of Defense Breast Cancer Research Program,* Philadelphia, PA USA

Awards:

Siolas, D. (2005) AACR-AstraZeneca Scholar-in-Training Award. American Association for Cancer Research, Philadelphia, PA, USA.

Conclusions:

The enhancement of our chemotherapeutic arsenal is crucial in our battle to conquer cancer. The shRNAmir libraries that I've helped develop provide a convenient, flexible and effective tool for studying gene function in human cells. Our new knowledge of hairpin processing enabled us to enhance the silencing capabilities of our hairpins by applying siRNA silencing guidelines to them. In this process, we were able to make more potent RNAi triggers that unexpectedly worked more efficiently than siRNAs with identical sequences. Libraries such as this allow the use of RNAi as a genetic tool to study cancer genes and identify the molecular pathways these genes affect.

I am currently using this library to study genetic vulnerabilities in cancer cells through screening for synthetic lethal combinations with p53 loss. The screen has been started in HCT 116 cells that have either wild type or no p53. This will allow us to work out the conditions of the screen and expand into MCF 7 and other breast cancer cell lines. Using shRNA libraries in cancer studies will expand our biological understanding of cancer and allowing us to enhance existing cancer therapies and develop new ones.

References:

1. Hannon, G.J. RNA interference. *Nature* **418**, 244-51. (2002).
2. Tomari, Y. & Zamore, P.D. Perspective: machines for RNAi. *Genes Dev***19**, 517-29 (2005).
3. Nakayashiki, H. et al. RNA silencing as a tool for exploring gene function in ascomycete fungi. *Fungal Genet Biol* **42**, 275-83 (2005).
4. Tang, G. & Galili, G. Using RNAi to improve plant nutritional value: from mechanism to application. *Trends Biotechnol* **22**, 463-9 (2004).
5. Dasgupta, R. & Perrimon, N. Using RNAi to catch Drosophila genes in a web of interactions: insights into cancer research. *Oncogene* **23**, 8359-65 (2004).
6. Fraser, A. Towards full employment: using RNAi to find roles for the redundant. *Oncogene* **23**, 8346-52 (2004).
7. 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**, 8401-9 (2004).
8. Meister, G. & Tuschl, T. Mechanisms of gene silencing by doublestranded RNA. *Nature* **431**, 343-9 (2004).

9. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97 (2004).
10. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**, 522-31 (2004).
11. Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**, 4051-60 (2004).
12. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957-66 (2004).
13. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9 (2003).
14. 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**, 231-5 (2004).
15. 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**, 2162-7 (2004).
16. Han, J. et al. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-27 (2004).
17. Gregory, R.I. et al. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235-40 (2004).
18. 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**, 3011-6 (2003). 21
19. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95-8 (2004).
20. Siolas, D. et al. Synthetic shRNAs as potent RNAi triggers. *Nat Biotechnol* **23**, 227-31 (2005).
21. Paddison, P.J. et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427-31 (2004).
22. Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431-7 (2004).
23. Zeng, Y., Wagner, E.J. & Cullen, B.R. Both natural and designed microRNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* **9**, 1327-33. (2002).
24. Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C. & Voinnet, O. In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. *Gene Dev* **18**, 2237-42 (2004).
25. Vousden KH, LU, X. Live or let die: the cell's response to p53. *Nat Rev Cancer.* 2002 Aug;2(8):594-604.
26. Paddison, PJ, Hannon, GJ RNA interference: the new somatic cell genetics? *Cancer Cell.* 2002 Jul;2(1):17-23.
27. Silva J, Chang K, Hannon GJ, Rivas FV. RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age. *Oncogene.* Nov 1 2004;23(51):8401-8409.

Appendix:

TECHNICAL REPORT

Second-generation shRNA libraries covering the mouse and human genomes

Jose M. Silva [1#], Mamie Z. Li [2#], Ken Chang [1#], Wei Ge [3], Michael C. Golding [1], Ricky 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*], and Gregory J. Hannon[1*]

[1] Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road
Cold Spring Harbor, NY 11724

[2] Department of Genetics
Center for Genetics and Genomics
Harvard Medical School
Room 158D, NRB
77 Avenue Louis Pasteur
Boston, MA 02115

[3] Rosetta Inpharmatics LLC,
A Wholly Owned Subsidiary of Merck & Co., Inc.
401 Terry Ave North
Seattle, Washington 98109, USA

*To whom correspondence should be addressed
#These authors contributed equally

Loss-of-function phenotypes often hold the key to understanding the connectivity and biological functions of biochemical pathways. We and others have previously constructed libraries of short hairpin RNAs (shRNAs) that allow systematic analysis of RNAi-induced phenotypes in mammalian cells [1,2]. Here we report the construction and validation of second-generation shRNA (shRNA^{mir}) expression libraries that have been designed based on an increased knowledge of RNAi biochemistry. In these constructs, silencing triggers have been designed to mimic a natural microRNA primary transcript, and each target sequence has been selected based on thermodynamic criteria for optimal small RNA performance. Biochemical and phenotypic assays have indicated that the new libraries are substantially improved compared to first-generation reagents. We have generated large-scale arrayed, sequence-verified libraries comprising more than 140,000 shRNA^{mir} expression plasmids, covering a substantial fraction of all predicted genes in the human and mouse genomes. These libraries are presently available to the scientific community.

Introduction

Most eukaryotic cells harbor a natural response to double-stranded RNAs (dsRNA) that inhibits gene expression in a sequence-specific manner [3]. DsRNA silencing triggers are processed into small RNAs (siRNAs and miRNAs) that engage the RNA-induced silencing complex (RISC) to suppress expression of homologous targets. In cases in which the small RNA is perfectly complementary to the target, that RNA is cleaved and ultimately destroyed [3,4]. This pathway, known as 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 [5-9].

In mammals, RNAi can be initiated in several ways. First, RNA molecules can be produced chemically [10] or enzymatically *in vitro* [11-14] and delivered to a cell. The most prevalent method of triggering RNAi is the delivery of one or more small interfering RNAs (siRNAs). SiRNAs are duplexes of ~21-22 nucleotides that bear two nucleotide 3' overhangs [3,15]. One strand (the guide strand) of the siRNA is incorporated into the effector complex of RNAi, the RNA-induced Silencing Complex, RISC, through the action of a RISC Loading Complex, RLC. Once in RISC, the siRNA guides substrate selection via base pairing to its complementary target [3,4].

At the heart of RISC is an Argonaute protein [16], which directly contacts the siRNA [17]. When the mRNA is engaged by this complex, the siRNA-mRNA interaction places the target in the correct alignment with the nuclease active site or "slicer" within the Argonaute PIWI domain, and the target is endonucleolytically cleaved [18-20].

The RNAi machinery can also be programmed by endogenous sources of double-stranded RNA. The most well characterized source of endogenous triggers for the RNAi machinery are the microRNA genes [21,22]. It was initially assumed that microRNAs were transcribed from the genome as short, hairpin RNAs [23] that were directly processed by Dicer to yield the mature small RNAs that enter RISC [24-27]. Over the past year, however, a different picture has emerged. Numerous studies have now demonstrated that, in animals, miRNAs are transcribed by RNA polymerase II to generate long primary polyadenylylated RNAs (pri-miRNAs) [28,29]. These primary transcripts probably adopt a complex secondary structure and fold, in the areas that harbor the mature microRNA sequences, into double-stranded RNA hairpins. Through mechanisms not yet fully understood, the pri-microRNA is recognized and cleaved at a specific site by the nuclear Microprocessor complex [30-34]. This contains an RNase III family enzyme, Drosha, that cleaves the hairpin to produce a ~70-90 nucleotide microRNA precursor (pre-miRNA) with a 2 nucleotide 3' overhang [30]. This distinctive structure signals transport of the pre-miRNA to the cytoplasm by a

mechanism mediated by Exportin-5 [35,36]. Only then is the pre-miRNA recognized by Dicer and cleaved to produce a mature microRNA. This probably involves recognition the 2 nucleotide 3' overhang created by Drosha to focus Dicer cleavage at a single site ~22 nucleotides from the end of the hairpin [17,37].

Mature miRNAs are superficially symmetrical, with 2 nucleotide 3' overhangs at each end having been generated by Drosha and Dicer, respectively. However, the individual strands of the mature miRNA enter RISC in an unequal manner. As with 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 [38,39]. This observation of thermodynamic asymmetry within 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.

Previously, several groups, including our own, described the design and construction of arrayed short hairpin RNA (shRNA) libraries that covered a fraction (~1/3) of human genes [1,2]. At the time when these tools were developed, our knowledge of microRNA maturation was relatively incomplete. This led most groups to the notion of expressing a simple hairpin RNA that mimicked the pre-miRNA. As our knowledge of the microRNA processing pathway and our understanding of strand preferences for RISC loading have grown, it seemed prudent to reevaluate whether the performance of encoded triggers of the RNAi pathway might be improved by remodeling a primary miRNA transcript to experimentally alter its targeting capability. Indeed such strategies have previously succeeded in both plants and animals [40,41].

Here we report the construction of a new generation of shRNA libraries (shRNA^{mir}) that takes into consideration our advancing understanding of microRNA biogenesis. In these constructs, the shRNA is harbored within the backbone of the primary mir-30 microRNA. This natural configuration proved to be up to 12 times more efficient in the production of the mature synthetic miRNAs than the previous design. Additionally, we have biochemically characterized processing of these synthetic microRNAs, allowing us to predict the mature small RNA product(s) that will be generated from each vector. This has allowed selection of target sequences that maximize efficiency by directing preferential incorporation of the correct strand into RISC. Using these criteria, we have produced and sequence-verified more than 140,000 shRNAs covering a substantial fraction of the predicted genes in the mouse and human genomes. We have assayed a selected subset of shRNAs from the library for their ability to knock-down the expression of targeted genes by quantitative RT-PCR. We have also tested this set in a phenotypic assay and compared the performance of the first- and second-generation library designs. Overall, the shRNA^{mir} libraries that we describe here provide a convenient, flexible and effective tool for studying gene function in human cells. Additionally, they, for the first time, extend the possibility of large-scale RNAi screens to mouse systems.

Results

Design and construction of second generation shRNA libraries

We have previously shown that expression of a simple, 29 basepair (bp) hairpin from a U6 snRNA promoter can induce effective suppression of target genes when delivered either transiently or stably from integrated constructs [1,42,43]. We also found that longer hairpin structures were more effective inhibitors of gene expression than were shorter structures with stems of 19-21 nucleotides. All of these constructs, however, were designed to express a pre-miRNA hairpin, an intermediate in microRNA biogenesis, rather than a transcript that closely resembles a primary microRNA. Cullen and colleagues had previously shown that effective suppression could be achieved by redesigning an endogenous microRNA, miR-30, such that its targeting sequence was directed against a reporter gene [40]. We sought to compare directly the abundance of small RNAs produced from vectors with simple hairpin structures to those that more closely resemble a natural microRNA. Since it had been previously shown that the efficient ectopic expression of endogenous microRNAs requires substantial flanking sequence [44], we developed a vector in which sequences from a remodeled miR30 are 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 3 positions in the precursor. This cassette was inserted into a vector equivalent to that in which we constructed our first-generation shRNA library (pSM1), with the new shRNA vector being designated, pSM2. To distinguish the second-generation shRNAs from those in our first-generation library, we have dubbed these shRNA^{mir}.

In order to enable the use of small RNA design rules to potentially enhance the efficacy of our shRNAs, it was necessary to understand how the shRNA^{mir} was processed *in vivo*. To address this issue, we took advantage of existing studies of miR30 biogenesis that mapped its processing sites [45]. Cullen and colleagues also mapped cleavage sites for a modified miR30 in which the mature microRNA sequences had been replaced by sequences targeting luciferase [45]. Using this information as a guide, we designed a series of constructs predicted to generate small RNAs targeting mouse p53, human PTEN and luciferase. To verify processing sites, 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 (we were unable to detect the passenger strand for this construct) by RACE-PCR (Fig. 1a; Supplementary Fig. 1). Since Northern Blotting indicated that maturation of shRNA^{mir}s produced 22 nucleotide species (Fig. 1b), we were able to infer the 5' end of each small RNA species. We consider the possibility of two processing sites at each end of the shRNA Since our analysis in the cases of p53 and PTEN shRNAs could not distinguish between processing at either of two terminal bases (Fig. 1a). However, in the

case of luc1309, the answer was relatively clear that the guide strand was cleaved in most cases (8/10) at the most 3' indicated site. Two out of ten sequences indicated cleavage 1 base 5' of that site, perhaps reflecting a genuine heterogeneity in Drosha cleavage (RACE1, RACE3; Supplementary Fig. 1). We therefore feel it most likely that cleavage is most prevalent at the sites indicated by the heavy red (drosha) and blue (dicer) lines, but our data is consistent with the possibility of some cleavage also occurring at the sites indicated by the lighter lines.

To test the performance of pSM2 in comparison to pSM1, we used both vectors to express a sequence targeting firefly luciferase. The sequence was inserted such that an identical mature small RNA would be generated from each construct following processing *in vivo* (Supplementary Fig. 2). Of primary concern was the overall amount of mature small RNA that would be generated from each construct. This was critical as dose-response experiments for shRNAs indicate that suppression correlates very well with the amount of RNA delivered [37], particularly at the relatively low doses that are expected to be achieved by expression from transfected or integrated constructs as compared to directly transfected synthetic RNAs. We transfected pSM1-luc and pSM2-luc into 293 cells, prepared RNA and assayed the processed small RNA by northern blotting. Cells transfected with pSM2-luc contained roughly 12-fold more of the small RNA than did cells transfected with pSM1-luc (Fig 1b).

As it is now clear that primary microRNAs are transcribed mainly by RNA polymerase II [28,29], we wished to compare the performance of shRNA^{mir}s driven by a variety of different promoters. We therefore cloned two different shRNA^{mir} cassettes targeting firefly luciferase downstream of three different RNA polymerase III promoters (tRNA-val [46], U6 [42] and H1 [47]) and two different RNA polymerase II promoters (MSCV-LTR and CMV [48]). These constructs were each prepared in a plasmid backbone that carried no other mammalian promoter. Each was transfected in combination with a homologous target expression plasmid encoding firefly luciferase and with a non-targeted 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 from human (HEK-293T, MBA-MD-231), one from mouse (NIH-3T3) and one from dog (MDCK). When the ability of these constructs to suppress the luciferase target was compared using a very efficient shRNA^{mir} (luc1309), we saw virtually no difference in the performance of the various promoters (Fig. 1c). However, when a less efficient shRNA^{mir} (luc311) was used, differences became apparent (Fig. 1c). In this, and numerous experiments with other shRNAs (not shown), the U6 snRNA and CMV promoters gave the best and most consistent repression. The MSCV LTR, tRNA val and H1 promoters worked less efficiently overall. Based upon these studies, we chose to retain the U6 snRNA promoter in our base library vector, pSM2. It is important to note that all of our studies have been 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 (see accompanying paper by Dickins et al., and Stegmeier et al., in press). However, we can also suppress gene expression by stable integration of pSM2 directly (Supplementary Figure 3).

Based upon these tests we constructed our second-generation shRNA library vector, pSM2, as shown in figure 2a. The shRNA^{mir} expression cassette is carried within a self-inactivating murine stem cell virus. Expression of the small RNA is driven by the U6 snRNA promoter. 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. Synthetic oligonucleotides encoding shRNAs are inserted into XhoI and EcoRI sites that lie within the miR-30 primary microRNA sequences. Immediately following the miR-30 cassette in each vector is a RNA polymerase III termination signal and a randomly generated 60 nucleotide 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 [1,43]. The pSM2 vector is also designed such that inserts can be moved by an *in vivo* recombination strategy (MAGIC) [49]. Key elements of this feature are the presence on the plasmid backbone of a protein-dependent origin of replication, RK6γ and a transfer origin (oriT) that is dependent upon a complementing locus in the host cells. To permit recombination into the recipient plasmid, the shRNA^{mir} cassette is flanked by I-SceI restriction sites which, when cut in the recipient strain, reveal homology regions for recombination into the recipient plasmid. One key difference between in the second-generation shRNA libraries is that the 5' homology region is the miR-30 flanking sequence itself rather than an artificial sequence. Thus, in the second generation libraries, the shRNA cassette is transferred without the U6 snRNA promoter. This allows the construction of mating recipients that contain inducible or tissue specific promoters (Stegmeier et al., in press). Finally, the pSM2 vector can be selected for integration into target cells using a puromycin selection marker.

Six different shRNA^{mir} sequences were designed for each of 34,711 different known and predicted human genes and 32,628 mouse genes. In each case, shRNAs were designed such that the mature small RNA generated from each construct followed thermodynamic asymmetry rules that have been successfully applied for the design of siRNAs. Based upon 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 were shifted by a base at either end (the cleavage positions indicated in Fig. 1).

Construction of the library proceeded stochastically using a highly parallel *in situ* synthesis approach for oligonucleotide production (Fig. 2b). Groups of ~22,000 oligonucleotides, each containing a different shRNA^{mir} cassette were

synthesized on glass-slide microarrays [50]. Populations were eluted from the arrays and amplified by PCR. In order to insure efficient cloning, the pSM2 backbone was inserted into a lambda phage backbone such that it was flanked by loxP sites. λ-pSM2 contains unique XhoI, EcoRI for subcloning amplified hairpins and unique FseI and AvrII sites for insertion of bar code 60mers. λ-pSM2 was first barcoded with a mixed library of random 60 nucleotide sequences amplified with a primer set which included one primer with an FseI site and one primer with a T7 promoter followed by the AvrII site. Amplified barcoded λ-pSM2 libraries were lysogenized into a strain 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. Approximately 10^8 Cm^RKm^R lysogens were selected and served as a bar coded library pool. Bar coded λ-pSM2 was CsCl purified, then cleaved with EcoRI and XhoI before being ligated to gel purified EcoR1-XhoI cleaved pooled $shRNA^{mir}$ inserts from an individual chip and packaged. Average library sizes were ~$5x10^7$ recombinants per pool. To generate pSM2 library plasmid pools, the phage were used to infect an *E. coli* strain we constructed, BUN25, that expresses both Cre recombinase and the *pir1-116* gene, needed for high copy RK6γ replication. Pooled plasmid libraries were then transformed into a mating competent host strain (BW F'DOT) and individual clones were sequenced at random. Clones with perfect inserts represented between 25 and 50% of the population, and these were selected and saved as an arrayed set. Accumulation of new clones from each pool was monitored dynamically and once a pool began to yield fewer unique clones per sequencing run, sequencing was halted and the pool was resynthesized without those sequences that had already been obtained. Approximately 70 chips were reiteratively synthesized to maximize unique sequencing. Also, once 3 or more verified shRNAs were obtained for any given gene, the remaining shRNAs targeting that gene were also withdrawn from population selected for resynthesis.

To date, we have sequence verified 79,805 shRNAs targeting 30,728 human genes and 67,676 shRNAs targeting 28,801 mouse genes. A tabulation of coverage within selected functional groups can be found in Table 1 for the mouse and human libraries. The ultimate goal is to generate 3 shRNAs for each target locus. Existing, sequence-verified shRNAs for human are listed in supplementary table 1, and verified mouse shRNAs are listed in supplementary table 2. The full collection, updated dynamically, can be accessed at http:\\codex.cshl.edu.

Validation of the second-generation shRNA libraries

To test the efficiency of the second-generation shRNA libraries, we took an approach that we had previously used to assess the performance of the first generation reagents [1]. A green fluorescent protein (ZsGreen) reporter harboring the PEST domain of the mouse ornithine decarboxylase is normally degraded by the proteasome [51]. Thus, cells harboring a destabilized ZsGreen expression plasmid show very low levels of fluorescence. Interference with proteasome

function, for example 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. Thus, co-transfection of the reporters with an shRNA^{mir} expression plasmid can reveal 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; Supplementary Table 3).

We chose a total of 53 shRNAs targeting 13 different genes that were known to be involved in proteasome function (Fig. 3b). 24 were from the first-generation library and 29 were from the second-generation library. These were co-transfected with the reporter in combination with a dsRED-encoding plasmid that allowed normalization of the transfections. It was immediately apparent that the second-generation shRNAs performed substantially better than the first-generation shRNAs. We noted that most of the plasmids derived from the second-generation library were as potent as the best shRNAs that had been selected from a screen of the first-generation library.

To gain a more detailed picture of 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 semi-quantitative RT-PCR. Plasmids were transfected into HeLa cells with approximately 80% efficiency, as measured by reference to a co-transfected reporter plasmid. Despite this incomplete transfection, all but 6 of the shRNA^{mir}s reduced the levels of their target RNAs by ~60% or more with 13/36 of the shRNA^{mir}s suppressing their targets by the theoretical maximum of ~80% (Fig. 3c, upper panel; Supplementary Table 3). Similar results were seen with an additional 12 shRNAs that did not target proteasome subunits (not shown). These studies were also illuminating, as they revealed that the functional assay in some cases, e.g., pSMB3, did not show a large activation of the reporter despite substantial suppression of the targeted mRNA (Fig. 3c, lower panel). Thus, the functional assay underestimated slightly 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 within it 47 hairpins directed to proteasome subunits using the phenotypic assay for proteasome function via a high-throughput protocol in 96-well plates (Fig. 4). In this context, 34/47 shRNAs targeting the proteasome scored as positives (72%) as compared to 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 again revealed positive signals from all 34 proteasomal shRNAs. However, only 5/10 of the non-proteasomal RNAs continued to activate the reporter, and none of these scored with more than one shRNA in the library (Supplementary Table 4).

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 dsRNA substrates in *Drosophila* [52]. Similarly, studies of dicer-mutant *C. elegans* demonstrated that endogenous loci could encode triggers of the RNAi machinery, and this led to the notion that such loci could be altered to target genes for experimental silencing [24-27]. At the time when the first such experiments were done, the nature of the primary microRNA transcript was unknown. Indeed, it was suspected that small RNAs were transcribed from the genome as short hairpin precursors, pre-miRNAs that were converted to mature small RNAs by Dicer cleavage. What has recently become clear is that pre-miRNAs are simply a processing intermediate, generated by cleavage of a longer primary transcript (pri-miRNA) by the Microprocessor [30-34].

Many strategies have been developed for producing miRNA-like triggers of the RNAi pathway. As we have mapped the processing sites on precursor shRNA^{mir}s, we can predict what 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 have constructed large libraries of sequence-verified shRNAs targeting the majority of the known and predicted genes in the human and mouse genomes. On average, each locus is covered by 2 shRNA^{mir}s presently; however, the ultimate goal is to have 3 sequence-verified shRNA^{mir}s for each gene. The second generation libraries resemble those that we have 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 [49]. A unique feature of the second generation library is that the expression cassette can be moved without the need to move also the constitutive U6 snRNA promoter. This permits large scale construction of secondary libraries under the control of tissue specific and inducible promoters. Indeed, regulated expression of our library cassettes from RNA polymerase II promoters has been shown to effectively suppress gene expression both in cultured cells and in animals (Dickins et al, see accompanying paper; Stegmeier et al., in press). These recipient vectors can be directly used with any shRNA^{mir} encoded by the library described herein.

The use of large-scale resources for suppressing gene expression via RNAi promises to revolutionize genetic approaches to biological problems in numerous model systems. For human cells, both siRNA and shRNA collections have previously been reported and are generally available to investigators to

probe a wide range of biological questions. The libraries described here should prove useful for assessing the functions of individual genes and for taking genome-wide approaches. Strategies reported in the accompanying paper and by Stegmeier and colleagues will permit large-scale application of these tools for screens which require long-term suppression of gene expression using single-copy integrants or inducible repression. Thus, we have produced coherent system of RNAi reagents with utility in both mouse and human experimental systems.

Methods

Construction of the lysogenic strain DH10βλ_{KP} and excision strain BUN25
DH10βλ_{KP} [*mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80 lacZΔM15 ΔlacX74 deoR recA1 endA1 araΔ139 Δ(ara, leu)7697 galU galK λ⁻ rpsL nupG tonA λ-pir1-npt*] is a strain containing λcI and the *pir1* gene that was constructed in order to lysogenize the λSM2-barcode library prior to introduction of the hairpin fragments. To generate this strain, λ_{KP} containing the *pir1* and Kn^R genes was constructed. To generate λ_{KP}, the *pir1* gene was amplified from BW23473 using primers MZL393 and MZL51, and cloned into the pCR2.1 TOPO TA cloning vector. The *pir1* gene was excised from the above clone on a BamH1 fragment and ligated into BamH1 cleaved pSE356, which contains an *npt* gene and a *Bam*HI restriction site flanked by two 1 kb λ DNA fragments [53] to generate pSE356pirWT. The pir1Km^R fragment was recombined onto wild type λ by amplifying λon LE392/pSE356pirWT and the resulting phage were collected and used to infect DH10β. 100 μl of DH10β cells were infected with 10^6 PFU at 30°C for 30 minutes in LB + 10 mM $MgSO_4$, diluted with 900 μl of LB incubated at 30°C for 2 h with shaking, and plated on LB containing 50 μg/ml kanamycin at 37°C overnight to select λ_{KP} lysogens. Lysogens were tested for the ability to lysogenize λ vectors containing R6Kγ origins of replication as extrachromosomal elements. A strain capable of doing this was selected and named DH10βλ_{KP}.

The 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*] strain containing *pir1-116* and *cre* was constructed to allow the conversion of λSM2 shRNA libraries into pSM2 shRNA libraries. A PCR fragment containing *pir1-116* gene was generated using primers MZL393 and MZL51 (see Supplementary table 4), cleaved with BamH1 and ligated into *Bam*HI-cleaved pUC18 to generate pML284. A fragment containing *Bst*BI-Frt-*cat*-Frt-*Nde*I (filled-in) was isolated from KD3 [54] and inserted into the *Sma*I site of pML284 to generate pML334. A *Hpa*I fragment containing *UmuDC* was isolated from pSE117 and cloned 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. The Frt-*cat*-Frt-*pir116* was isolated from pML334 as a *Kpn*I-*Sac*I (filled-in) fragment and ligated into *Mlu*I/*Bam*HI (filled-in) cleaved pML236-ΔBamHI to generate pML346. The 3.8 kb *Kpn*I-*Sac*I *UmuDC*-Frt-*cat*-Frt-*pir116-UmuC* fragment from pML346 was integrated into BNN132/pML104 by homologous recombination using the λ recombinase expressed from pML104, and confirmed by colony PCR. The *cat* gene was removed by FLP-mediated excision *in vivo* using pCP20 [55] which expresses the FLP recombinase to generate BUN24. A cassette that has Frt-*cat*-Frt flanked by 50 bp homology to *sbcD* and 50 bp homology to *sbcC* was amplified by primers MZL493/MZL494 and using KD3 as a template. This cassette was used to replace *sbcD* and part of *sbcC* on BNN132 by homologous recombination and the deletion were confirmed by colony PCR. The strain was

named BNN132sbcDC-Frt-cat-Frt. We then used a pair of outside primers (MZL495/MZL496) that gave about 500 bp homology regions to the upstream of *sbcD* and 500 bp homology regions to the *sbcC* to amplify a PCR product form 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 [56].

Library vector construction

A pair of loxP-*Not*I-loxP duplexed oligos (MZL524/MZL525) were inserted into the pSM2 *Bst*XI site to generate pSM2c-loxP. A second pair of duplexed oligos (MZL541/ MZL542), carrying the proper restriction sites for cloning barcodes into λSM2, were inserted into the *Bbs*I-*Mlu*I sites of pSM2c-loxP to create pML375. λACT2 was digested with *Not*I, and the λ arms were gel purified and ligated to *Not*I digested pML375 to generate λSM2. The ligation mixture was packaged using MaxPlax^{TM} 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 at 30°C for 30 minutes the added 900 μl of LB and incubated at 30°C for 2 hs with shaking to express the Cm^R marker, and plated on LB containing 17 μg/ml of chloramphenicol (Cm) at 30°C overnight. The proper recombinants were confirmed by restriction analysis. See Supplementary table 5 for sequences of referenced oligonucleotides.

Barcode library construction

The 60 base pair barcodes, gaagactaatgcggccggcca(n)$_{60}$gggccctatagtgagtcgtatta, were amplified using barcode primer 1 (aaattgcaatgaagactaatgcggccggcca) and barcode primer 2 (atatatggacgcgtcctaggtaatacgactcactatagggccc). The PCR conditions were: 0.1 pmol of barcodes, 50 pmol of each primer, 25 nmol of each dNTP, and 2.5 U of Taq DNA polymerase; 94°C for 45 seconds, (94°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds) x 13, 72°C for 10 minutes, 4°C forever. Ten PCR reactions were pooled together, purified using a QIAquick PCR purification kit, digested with *Bam*HI, *Eco*RI, *Xho*I and *Sal*I to remove these sites in the barcodes, and gel purified. The purified barcodes were digested with *Fse*I and *Avr*II and purified using the QIAquick gel extraction kit. Two micrograms of *Fse*I-*Avr*II digested λSM2 ligated with 10 ng of *Fse*I-*Avr*II digested barcodes with 1 x ligation buffer and 0.5 μl T4 DNA ligase in a 5 μl final volume at 16°C overnight. The ligation mixture was packaged and amplified. The size of the λSM2-barcode library was 4.2 x 10^7. We used 20 ml of DH10βλ_{KP} cells (A_{600} = ~1) to lysogenize 2 x 10^9 of λSM2-barcode library as 42 kb plasmids. The cell and the phage were mixed in the presence of 10 mM $MgSO_4$ and 0.2 % (w/v) maltose and incubated at 30°C for 30 minutes, added 200 ml of LB to recover at 30°C for 1 h by shaking. The mixture were concentrated by centrifugation at 4000 rpm for 20 minutes, resuspended in 3 ml of LB, plated on 10 large LB/Cm 17 μg/ml, and incubated at 30°C overnight. The cells were scraped from plates and grown in 3 L of TB

containing 17 μg/ml of Cm overnight. Supercoiled λSM2-barcode library DNA was prepared by cesium chloride. The lysogenization efficiency was approximately 30%.

Oligonucleotide Cleavage and PCR Amplification

To harvest 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 microcentrifuge tubes and subjected it to speed vacuum drying at medium heat (~55 °C) overnight. We resuspended the dried material in 200 μl of RNase/DNase free water and performed PCR amplification in 50μl reaction volumes using Invitrogen's Platinum® *Pfx* DNA Polymerase. To obtain a sufficient amount of PCR product, four 50 μl reactions were required for each sample. Each reaction contained 2X *Pfx* PCR amplification buffer, 0.3 mM of each dNTP, 1 mM $MgSO_4$, 0.3 uM of each primer, 0.5X 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 (5' CAGAGGCTCGAGAAGGTATATTGCTGTTGACAGTGAGCG 3') and 3'-mir30-PCR-ecorI-R (5' CGCGGCGAATTCCGAGGAGTAGGCA 3'). After an initial denaturation step of 94°C for 5 min, DNA amplification occurred through 25 cycles of denaturing at 94°C for 45 seconds and annealing and extension at 68° for 1 min and 15 sec followed by a final 7 minute extension at 68°. We then combined the four reactions into one tube, cleaned up the PCR product by use the QIAGEN® Minelute PCR Purification Kit and eluted in a total volume of 26 μl.

ShRNA library construction

The λSM2-barcode library and shRNA PCR products were digested with *Eco*RI and *Xho*I overnight and gel purified. Ligations with shDNA oligos were set up as following: 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 x ligation buffer, 0.5 μl of T4 DNA ligase, and water to 10 μl final. The ligation mixtures were incubated at 16°C for overnight and packaged. We typically observed 30- to 90-fold stimulation of plaque forming units (PFU) and 2 x 10^7 to 8 x 10^7 PFU total for each library pool. We typically amplify 2 x 10^7 PFU 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 = ~0.5) and selected colonies on LB/Cm 30 μg/ml 30°C overnight. Colony PCR was performed using forward (ggacgaaacaccgtgctcgc) and reverse primer (ttctgcgaagtgatcttccg) and 85 to 95 % correct sized inserts were typically observed with some containing multiple inserts. To generate plasmid DNA from these libraries, we typically excised 5 x 10^7 PFUs through infection of BUN25 cells as described earlier. The cells were scraped from plates and grown in 2 L of LB plus 13 g/l of circle growth 37°C for 7 to 8 h. Cesium chloride method was used to prepare DNA. DNAs were transformed into BW23474 F'DOT SbcC , and individual clones were sequenced using primer5' (TGTGGAAAGGACGAAACACC). Correct clones were individually rearrayed to form the final library.

Small RNA Northern Blots

293 cells were transfected in 10cm dishes at 60% confluency with 15 ug of shRNA plasmid DNA along with 5 ug of pDsRed-N1 (Clontech) using TransIT-LT1 (Mirus). 48hs post transfection, transfection efficiency was confirmed by estimating the percentage of cells expressing DsRed (~80%) and then total RNA was Trizol extracted and purified. Small RNA northern blots were carried out as described in [57] using 30 ug RNA/lane. For hairpins targeting EGFP at starting position 481 (Fig 1b), northern probes were DNA oligos corresponding to the anti-sense strand (ccggcatcaaggtgaacttcaa) of the mature RNA.

Proteasome assays

Bacteria cultures were grown in 96 well plates for 36 h in GS96 media (Bio101). Plasmid DNA was extracted using Quiagen Ultrapure plasmids minipreps in a 96 well plate format. DNA concentrations were determined by mixing an aliquot of each sample with picogreen (Molecular Probes) and determining fluorescence on a Victor2 plate reader. HEK 293T cells were plated in 96 well optic plates (Corning) at $1x10^5$ cells per ml. For the proteasome assay, 12.5 ng of the plasmid dsRed N-1 (Clontech), 5 ng of the Zsprosensor (Clontech) and 75 ng of each individual shRNA construct were cotransfected per well using 0.3 μl of LT-1 (Mirus) transfection reagent. After 24 hs the transfection media was replaced. After 72 h, media was removed and replaced with PBS 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 Quantitation

HeLa cells were seeded at 0.5×10^5 cells/well in 24-well plates and transfected 24 h later with 1 ug/well of the appropriate plasmid. Each plasmid was delivered to 4 wells by use of Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols. Transfection efficiency was determined by parallel transfection of a GFP-expressing plasmid and the percentage of fluorescent cells assayed by flow cytometry. For analysis of target gene mRNA knock down, cell lysates were collected 24 h after transfection, and total RNA was prepared by use of RNeasy columns (Qiagen) following the manufacturer's protocols. Messenger RNA quantitation was performed by Real-time PCR of reverse transcription products, using available Applied Biosystems TaqMan™ primer probe sets, and the percent mRNA remaining was determined by comparison with mRNA levels from cells treated with transfection reagent alone.

Acknowledgements

We thank members of the Hannon, Elledge, and Lowe labs for helpful suggestions. Jill Magnus assisted with array PCRs, and Lidia Nascimento, Vivek

Balija, Melissa Kramer, Theresa Zutavern, Stephanie Muller, and Beth Miller assisted with sequencing. Troy Moore helped with curation of the collection. We thank Peter Linsley and Steve Friend for their support of this project. This work was funded in part by awards from the DOD Breast Cancer Research Program (GJH and SJE) and the NIH (GJH and SJE). SJE and GJH are investigators of the Howard Hughes Medical Institute.

Table 1 Coverage by functional group

Group	Human Genes	Human Hairpins	Mouse Genes	Mouse Hairpins
Cancer	859	3082	890	2524
Cell cycle	531	2166	482	1552
Checkpoint	123	541	116	367
DNA repair	118	512	130	355
DNA replication	238	961	248	719
Enzymes	2943	10456	2818	8302
GPCRs	669	2101	663	1795
Kinases	618	2648	575	2250
Dual Specificity Phosphatases	35	144	32	114
Tryosine Phosphatases	36	184	33	166
Phosphotases	206	765	187	628
Proteases	454	1431	441	1168
Proteolysis	302	1458	270	858
Signal Transduction	2650	9046	2541	7274
Protein Trafficking	476	1596	458	1309
Transcription	820	2865	767	2209
Apoptosis	581	2061	558	1538

Figure Legends

Figure 1. Design and structure of shRNA^{mir} cassettes. (**a**) A comparison of the structures of several silencing triggers is shown. These include 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. 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-293T cells with shRNA and shRNA^{mir} cassettes expressed from pSM1 and pSM2, respectively, by the U6 snRNA promoter. In neither case was significant accumulation of pre-miRNA observed. Transfection rates were normalized using a co-delivered dsRED expression plasmid. (**c**) Five different promoters (human tRNAval, Human H1 RNA, Human U6 snRNA, MSCV LTR and Human CMV IE, as indicated) were tested for their ability to drive shRNA^{mir} expression and silence luciferase in transient transfections. Two different shRNAs were used, a highly efficient shRNA (luc1309) an a less efficient shRNA (luc311). In each case, the level of firefly luciferase was normalized to a non-targeted *Renilla* luciferase. Controls with empty vectors lacking a hairpin insert are also shown.

Figure 2. Construction of the second-generation library. (a) The pSM2c vector contains a U6 promoter, a U6 terminator following 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 using XhoI and EcoRI restriction sites. 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 lambda derivative of pSM2 that contained unique EcoR1, XhoI, FseI and AvrII sites, the latter two for insertion of bar codes. A bar coded pre-library was generated by the ligation of PCR amplified random 60 mers into Fse1-AvrII cleaved λpSM2 to generate a bar-coded library pool (upper right). The bar-codes λpSM2 was converted into a shRNA library by insertion of PCR amplified shRNA constructs prepared by *in situ* synthesis of inserts on a microarray in pools of ~22,000 into the EcoR1-XhoI cleaved pre-library (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 resistant colony. These were pooled and used for preparation of library DNA. This, in turn, was transformed into BW F'DOT and individual colonies were selected for sequence analysis.

Figure 3. Validation of the second-generation library. (a) A schematic representation of the phenotypic assay for proteasome function (see text) is shown. **(b)** Thirteen proteasome subunits were chosen because of their representation in both the first- and second-generation libraries (for sequences see Supplementary table 3). ShRNA expression clones corresponding to each were assayed for activation of the proteasome reporter. Blue bars indicate first-generation clones while green 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 (for sequences see Supplementary table 3) were tested for their ability to suppress their target RNAs (upper panel). QRT-PCRs were performed 24 h after transfection of HeLa cells at an average efficiency of 80% as measured by a co-transfected 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 performed in parallel (lower panel).

Figure 4. Performance of the second-generation library in a small-scale high-throughput screen. 47 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 positives controls (ATPase 1.1 to 1.3) from first (pSM1) and second (pSM2) libraries. The lower right panel shows the shRNAs that displayed accumulation of the proteasome reporter over the cut-off (2-fold or greater activation; yellow line). These are highly enriched for proteasome shRNAs (red). In blue are 10 additional non-proteasomal shRNAs that also scored positive in the screen. Of these, 5 were also positive on a retest of individual clones. The sequences of the shRNAs, in order from left to right, for the lower right panel are given in Supplementary Table 4.

Literature Cited

1. Paddison, P.J. et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427-31 (2004).
2. Berns, K. et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431-7 (2004).
3. Hannon, G.J. RNA interference. *Nature* **418**, 244-51. (2002).
4. Tomari, Y. & Zamore, P.D. Perspective: machines for RNAi. *Genes Dev* **19**, 517-29 (2005).
5. Nakayashiki, H. et al. RNA silencing as a tool for exploring gene function in ascomycete fungi. *Fungal Genet Biol* **42**, 275-83 (2005).
6. Tang, G. & Galili, G. Using RNAi to improve plant nutritional value: from mechanism to application. *Trends Biotechnol* **22**, 463-9 (2004).
7. Dasgupta, R. & Perrimon, N. Using RNAi to catch Drosophila genes in a web of interactions: insights into cancer research. *Oncogene* **23**, 8359-65 (2004).
8. Fraser, A. Towards full employment: using RNAi to find roles for the redundant. *Oncogene* **23**, 8346-52 (2004).
9. 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**, 8401-9 (2004).
10. Elbashir, S.M. et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-8. (2001).
11. Donze, O. & Picard, D. RNA interference in mammalian cells using siRNAs synthesized with T7 RNA polymerase. *Nucleic Acids Res* **30**, e46 (2002).
12. Kawasaki, H., Suyama, E., Iyo, M. & Taira, K. siRNAs generated by recombinant human Dicer induce specific and significant but target site-independent gene silencing in human cells. *Nucleic Acids Res* **31**, 981-7 (2003).
13. Kittler, R. et al. An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature* **432**, 1036-40 (2004).
14. Luo, B., Heard, A.D. & Lodish, H.F. Small interfering RNA production by enzymatic engineering of DNA (SPEED). *Proc Natl Acad Sci U S A* **101**, 5494-9 (2004).
15. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343-9 (2004).
16. 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**, 1146-50. (2001).
17. 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**, 1026-32 (2003).

18. 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**, 1434-7 (2004).
19. Liu, J. et al. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-41 (2004).
20. Parker, J.S., Roe, S.M. & Barford, D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *Embo J* **23**, 4727-37 (2004).
21. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97 (2004).
22. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**, 522-31 (2004).
23. Reinhart, B.J. et al. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-6. (2000).
24. 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**, 2654-9. (2001).
25. 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).
26. 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**, 2269-71. (2001).
27. 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**, 834-8. (2001).
28. Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**, 4051-60 (2004).
29. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* **10**, 1957-66 (2004).
30. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9 (2003).
31. 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**, 231-5 (2004).
32. 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**, 2162-7 (2004).
33. Han, J. et al. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**, 3016-27 (2004).
34. Gregory, R.I. et al. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235-40 (2004).
35. 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**, 3011-6 (2003).

36. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95-8 (2004).
37. Siolas, D. et al. Synthetic shRNAs as potent RNAi triggers. *Nat Biotechnol* **23**, 227-31 (2005).
38. Schwarz, D.S. et al. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208 (2003).
39. Khvorova, A., Reynolds, A. & Jayasena, S.D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209-16 (2003).
40. 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**, 1327-33. (2002).
41. Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C. & Voinnet, O. In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. *Genes Dev* **18**, 2237-42 (2004).
42. 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**, 948-58. (2002).
43. Westbrook, T.F. et al. A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**, 837-48 (2005).
44. Chen, C.Z., Li, L., Lodish, H.F. & Bartel, D.P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83-6 (2004).
45. Zeng, Y. & Cullen, B.R. Sequence requirements for micro RNA processing and function in human cells. *Rna* **9**, 112-23 (2003).
46. 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**, 700-7. (2003).
47. Brummelkamp, T.R., Bernards, R. & Agami, R. A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science* **21**, 21 (2002).
48. Zheng, L. et al. An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc Natl Acad Sci U S A* **101**, 135-40 (2004).
49. Li, M.Z. & Elledge, S.J. MAGIC, an in vivo genetic method for the rapid construction of recombinant DNA molecules. *Nat Genet* **37**, 311-9 (2005).
50. Cleary, M.A. et al. Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis. *Nat Methods* **1**, 241-8 (2004).
51. Li, X. et al. Generation of destabilized green fluorescent protein as a transcription reporter. *J Biol Chem* **273**, 34970-5 (1998).
52. Carmell, M.A. & Hannon, G.J. RNase III enzymes and the initiation of gene silencing. *Nat Struct Mol Biol* **11**, 214-8 (2004).
53. Elledge, S.J. & Walker, G.C. Phasmid vectors for identification of genes by complementation of Escherichia coli mutants. *J Bacteriol* **162**, 777-83 (1985).

54. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-5 (2000).
55. 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).
56. 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**, 201-5 (1988).
57. 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**, 2491-6. (2002).

Suplementary material

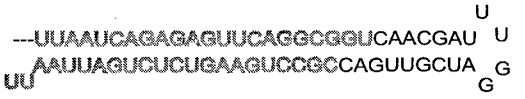
Supplementary Figures 1-3

Supplementary Tables 1-4

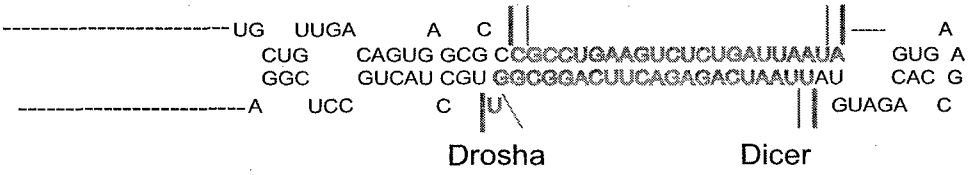
a

- siRNA

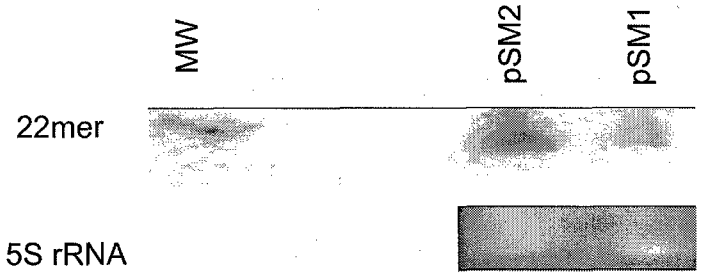
- shRNA (first-generation)



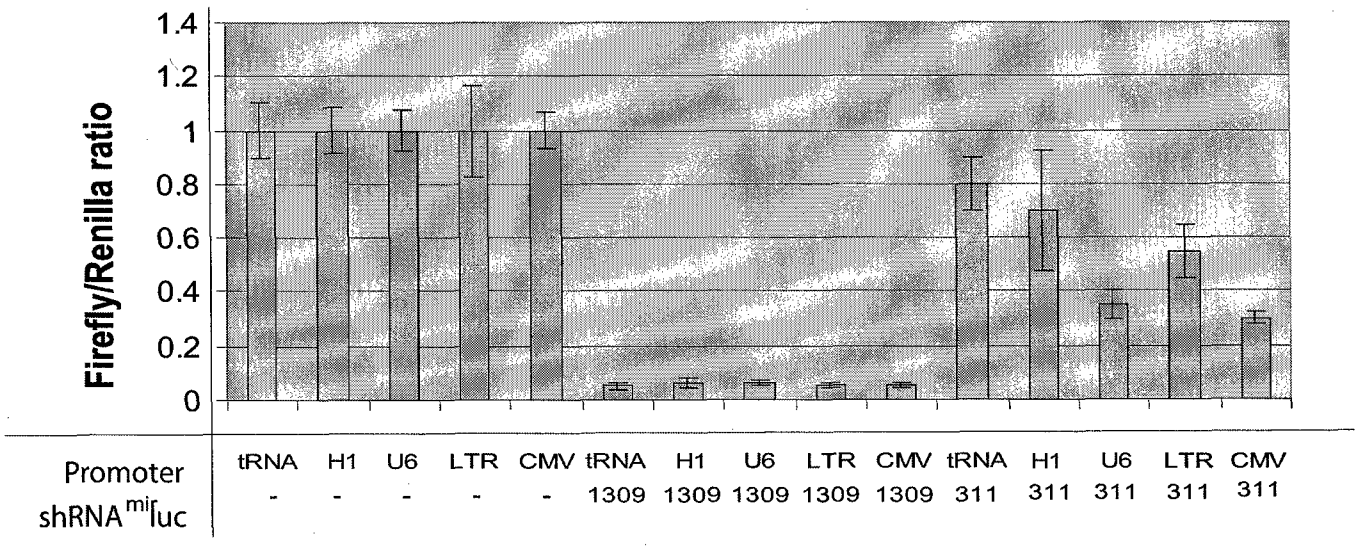
- shRNA^{mir} (second-generation)



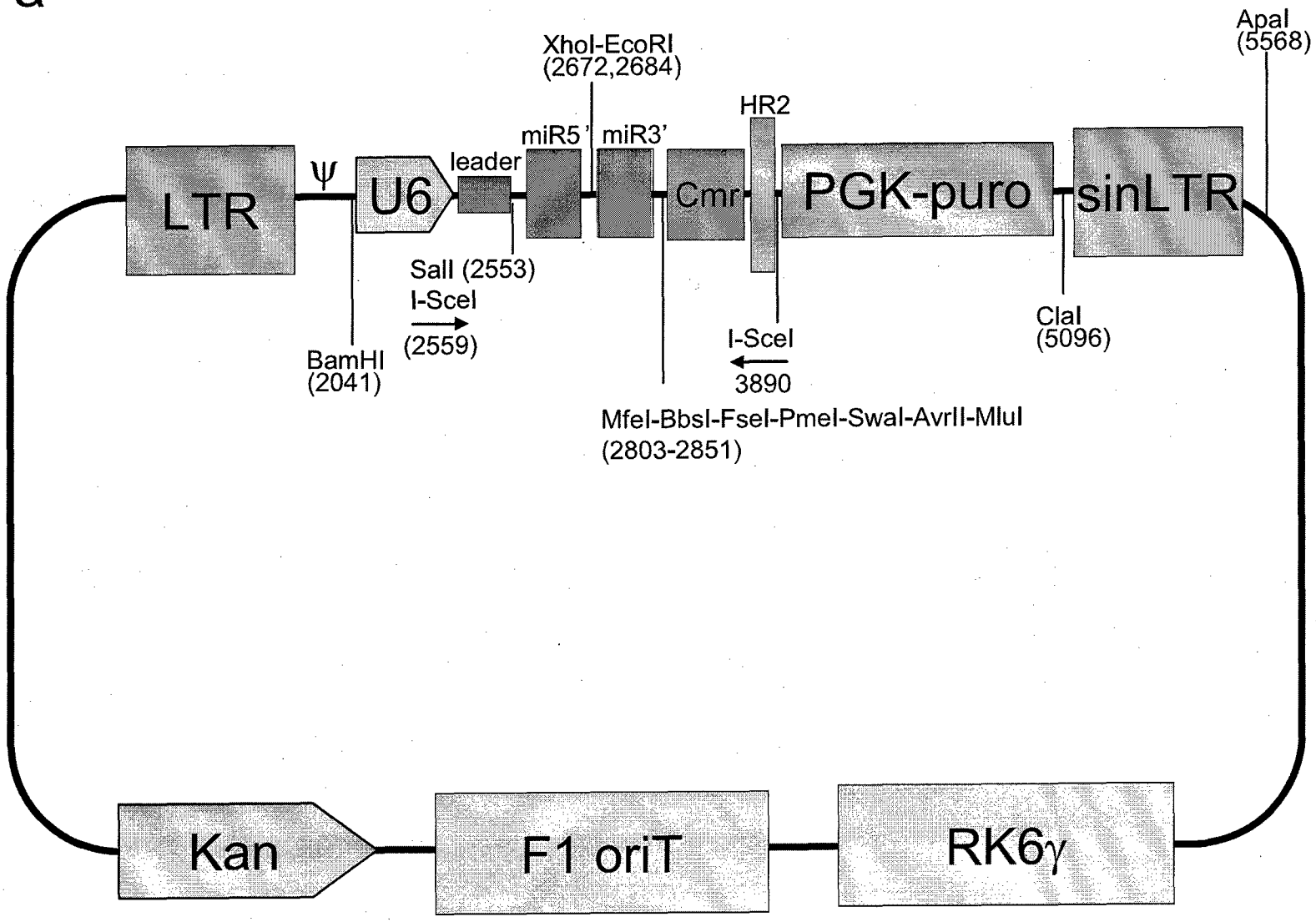
b



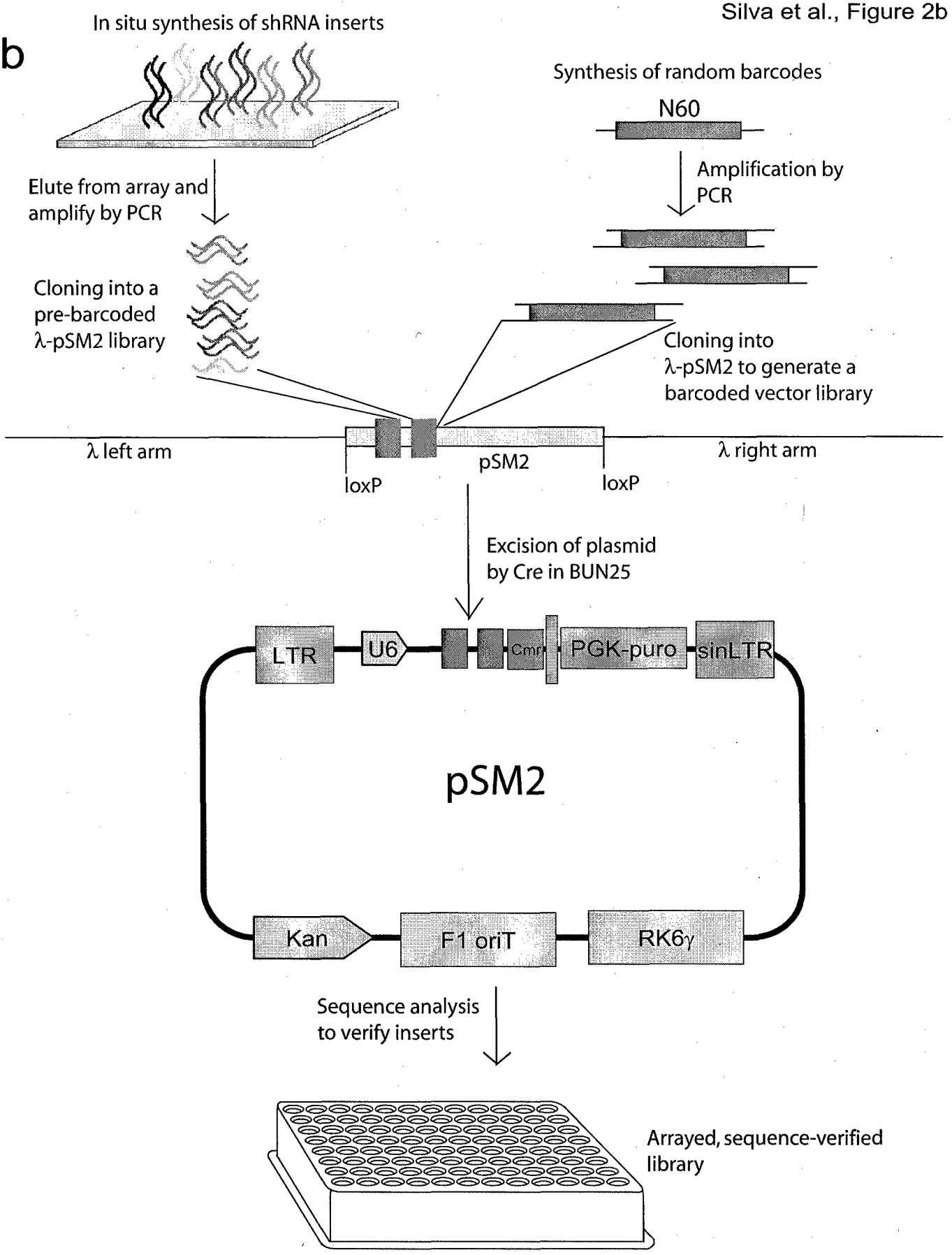
c

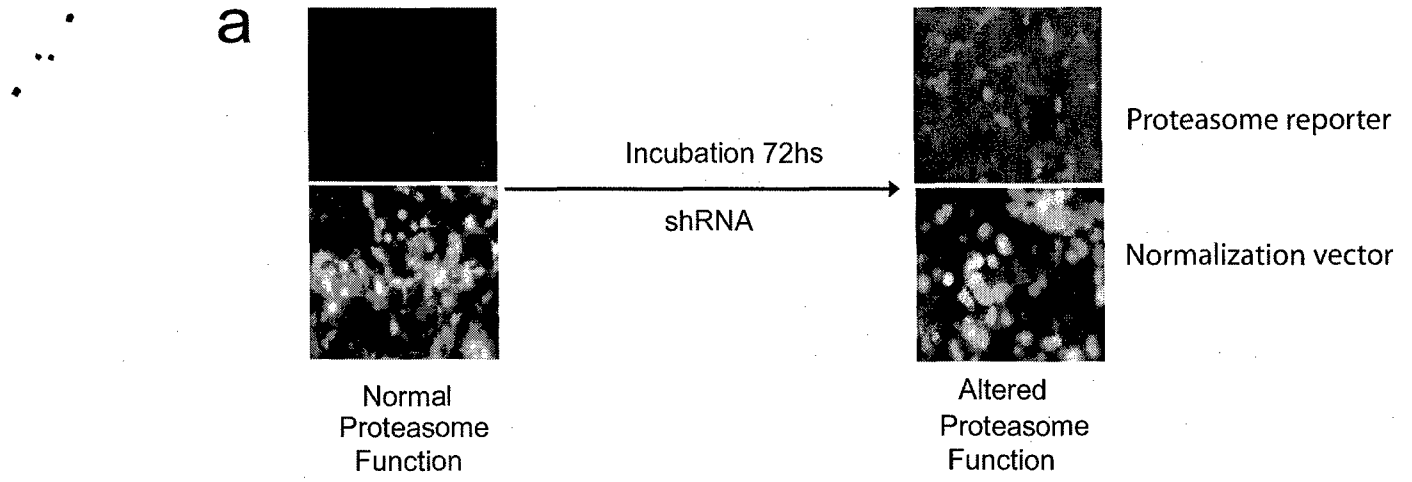


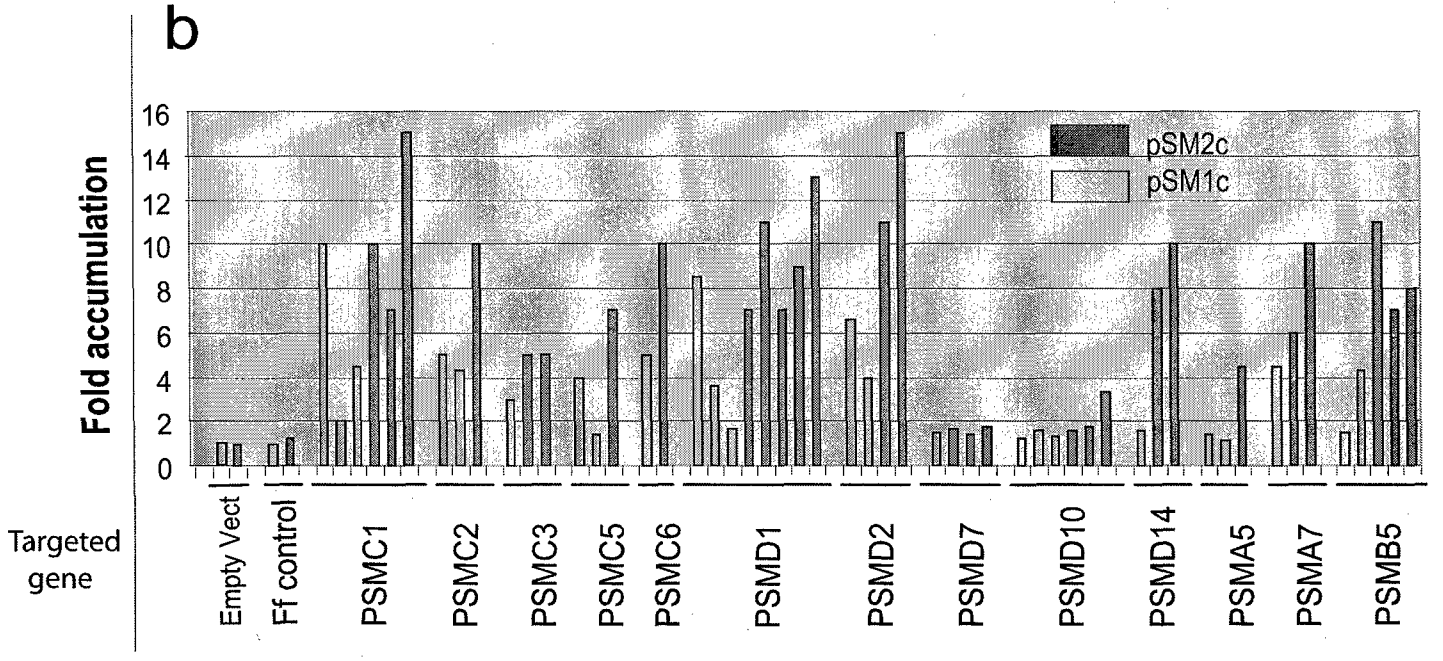
a

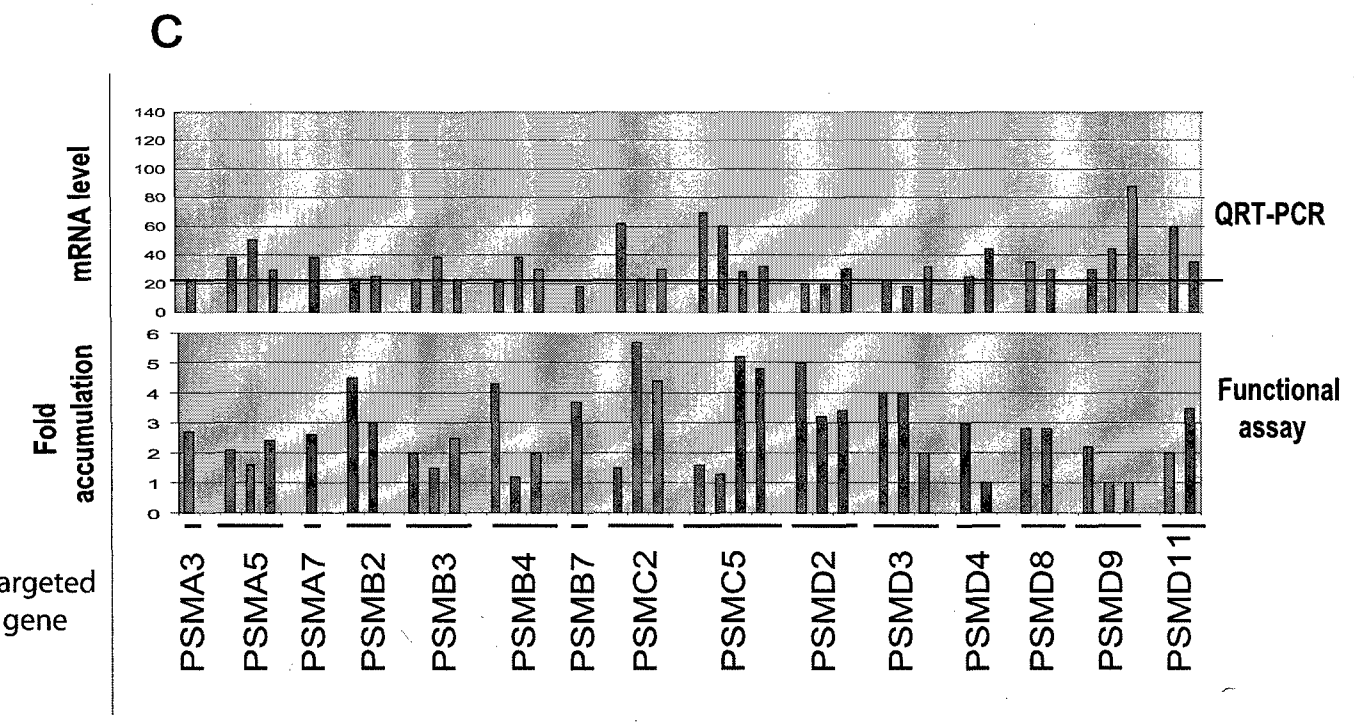

ApaI
(5568)
XhoI-EcoRI
(2672,2684)
HR2
miR5'
miR3'
LTR
ψ
U6
leader
Cmr
PGK-puro
sinLTR
SalI (2553)
I-SceI
(2559)
BamHI
(2041)
I-SceI
3890
ClaI
(5096)
MfeI-BbsI-FseI-PmeI-SwaI-AvrII-MluI
(2803-2851)
Kan
F1 oriT
RK6γ

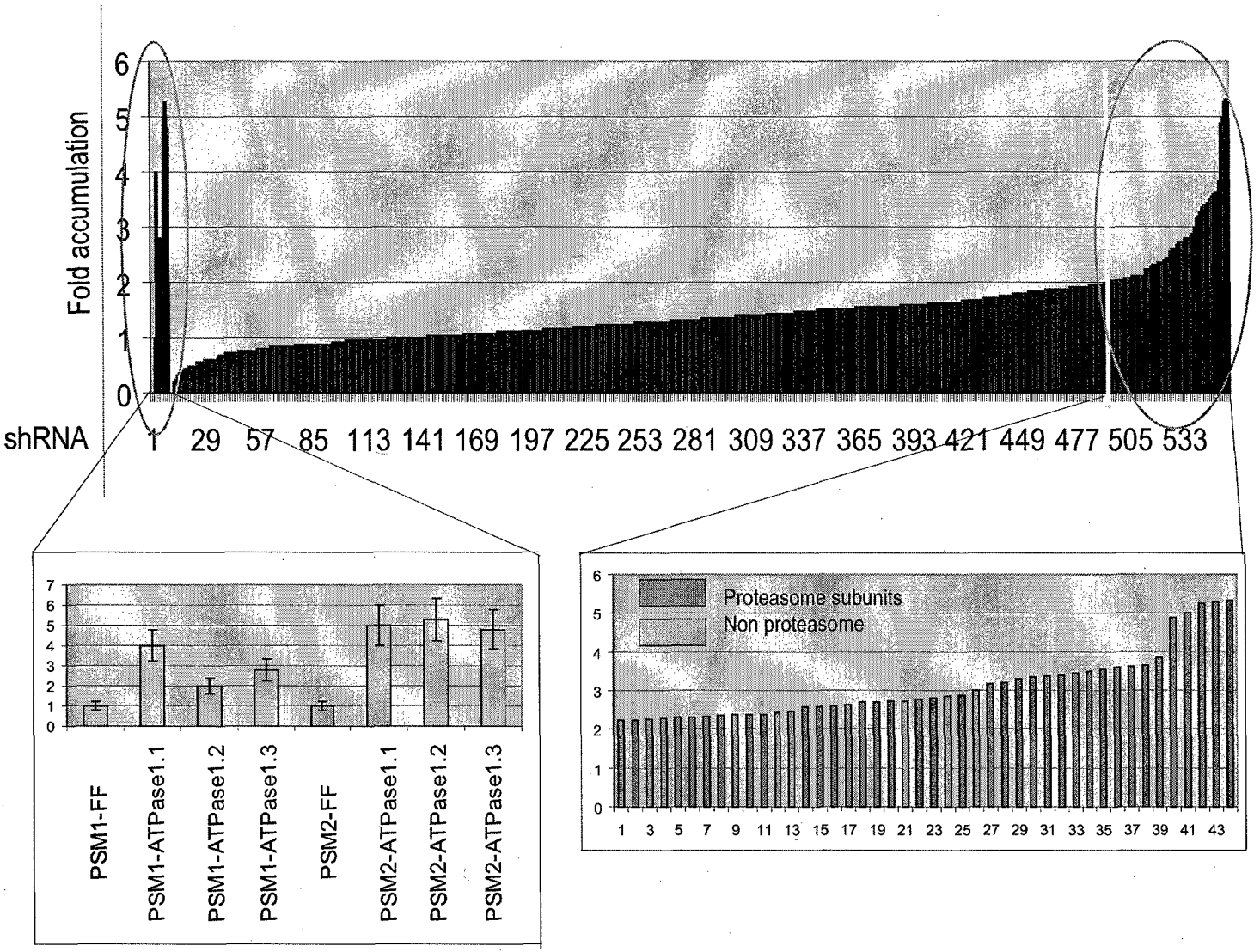
b


In situ synthesis of shRNA inserts
Synthesis of random barcodes
N60
Amplification by PCR
Elute from array and amplify by PCR
Cloning into a pre-barcoded λ-pSM2 library
Cloning into λ-pSM2 to generate a barcoded vector library
λ left arm
pSM2
λ right arm
loxP
loxP
Excision of plasmid by Cre in BUN25
LTR
U6
Cmr
PGK-puro
sinLTR
pSM2
Kan
F1 oriT
RK6γ
Sequence analysis to verify inserts
Arrayed, sequence-verified library


a
Incubation 72hs
shRNA
Proteasome reporter
Normalization vector
Normal Proteasome Function
Altered Proteasome Function


b
Fold accumulation
16
14
12
10
8
6
4
2
0
pSM2c
pSM1c
Targeted gene
Empty Vect
Ff control
PSMC1
PSMC2
PSMC3
PSMC5
PSMC6
PSMD1
PSMD2
PSMD7
PSMD10
PSMD14
PSMA5
PSMA7
PSMB5


c
mRNA level
140
120
100
80
60
40
20
0
QRT-PCR
Fold accumulation
6
5
4
3
2
1
0
Functional assay
Targeted gene
PSMA3
PSMA5
PSMA7
PSMB2
PSMB3
PSMB4
PSMB7
PSMC2
PSMC5
PSMD2
PSMD3
PSMD4
PSMD8
PSMD9
PSMD11


Fold accumulation
0
1
2
3
4
5
6
shRNA
1 29 57 85 113 141 169 197 225 253 281 309 337 365 393 421 449 477 505 533
PSM1-FF
PSM1-ATPase1.1
PSM1-ATPase1.2
PSM1-ATPase1.3
PSM2-FF
PSM2-ATPase1.1
PSM2-ATPase1.2
PSM2-ATPase1.3
Proteasome subunits
Non proteasome
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43

P53_1224 Guide Strand (Drosha site)

```
                    3'-TGTTCATGTACACATT-5'  RACE primer
               3'-AGGUGAUGUUCAUGUACACAUU-5' - predicted small RNA
5'-...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE1
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE2
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE3
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE4
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE5
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE6
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE7
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE8
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE9
   ...TTTTTTTTTTTTCCACTACAAGTACATGTGTAA... RACE10
```

P53_1224 Passenger Strand (Dicer site)

```
                    3'-TGTACATGAACATCAC-5'  RACE primer
               3'-AUAAUGUGUACAUGAACAUCAC-5' - predicted small RNA
5'-...TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE1
5'-...TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE2
5'-...TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE3
5'-...TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE4
5'-...TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE5
5'-... TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE6
5'-...TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE7
5'-...   TTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE8
5'-... TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE9
5'-...TTTTTTTTTTTTATTACACATGTACTTGTAGTG...-3' RACE10
```

Luc_1309 Guide Strand (Drosha site)

```
                    3'-ACTTCAGAGACTAATT-5'  RACE primer
               3'-UGGCGGACUUCAGAGACUAAUU-5' - predicted small RNA
5'-... TTTTTTTTTT.CCGCCTGAAGTCTCTGATTAA...-3' RACE1
5'-...TTTTTTTTTTTACCGCCTGAAGTCTCTGATTAA...-3' RACE2
5'-...TTTTTTTTTTT.CCGCCTGAAGTCTCTGATTAA...-3' RACE3
5'-...TTTTTTTTTTTTACCGCCTGAAGTCTCTGATTAA...-3' RACE4
5'-...TTTTTTTTTTTTACCGCCTGAAGTCTCTGATTAA...-3' RACE5
5'-...TTTTTTTTTTTTACCGCCTGAAGTCTCTGATTAA...-3' RACE6
5'-...  TTTTTTTTTACCGCCTGAAGTCTCTGATTAA...-3' RACE7
5'-...TTTTTTTTTTTTACCGCCTGAAGTCTCTGATTAA...-3' RACE8
5'-...TTTTTTTTTTTTACCGCCTGAAGTCTCTGATTAA...-3' RACE9
5'-...TTTTTTTTTTTTACCGCCTGAAGTCTCTGATTAA...3'- RACE10
```

PTEN_1137 Guide Strand (Drosha site)

```
                    3'-AAGGACGTCTTTCTCA-5' RACE primer
               3'-ACCCUAAAGGACGUCUUUCUGA-5' - predicted small RNA
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE1
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE2
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE3
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE4
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE5
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE6
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE7
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE8
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE9
5'-...TTTTTTTTTTTTGGGATTTCCTGCAGAAAGACT-3' RACE10
```

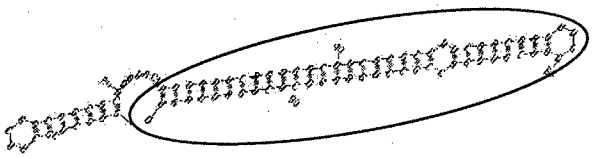
PETN_1137 Passenger Strand (Dicer site)

```
                    3'-AAAGACGTCCTTTAGG-5' RACE primer
               3'-AUUCAGAAAGACGUCCUUUAGG-5' - predicted small RNA
5'-... TTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE1
5'-... TTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE2
5'-...TTTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE3
5'-...  TTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE4
5'-...TTTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE5
5'-...TTTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE6
5'-...TTTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE7
5'-...TTTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE8
5'-...TTTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE9
5'-...TTTTTTTTTTTTAAGTCTTTCTGCAGGAAATCC-3' RACE10
```

Supplementary Figure 1. Mapping of Dicer and Drosha cleavage sites. To map cleavage sites for small RNAs generated by pSM2, we used 3' RACE. 293 cells were transfected with constructs corresponding to p53_1223, PTEN_1137 and luc_1309. Small RNAs were converted to cDNA after tailing with polyA polymerase and amplified using a specific primer (as indicated for each small RNA) and an anchored dT primer (according to the manufacturer's instructions, Roche). PCR products were cloned into the topo-TA vector (Invitrogen) and 10 clones were sequenced for each of the PTEN, p53 and luc guide strands and the PTEN and p53 passenger strands. Since the RNAs were A-tailed, the presence or absence of predicted terminal A residues was ambiguous, and they are therefore indicated in red.

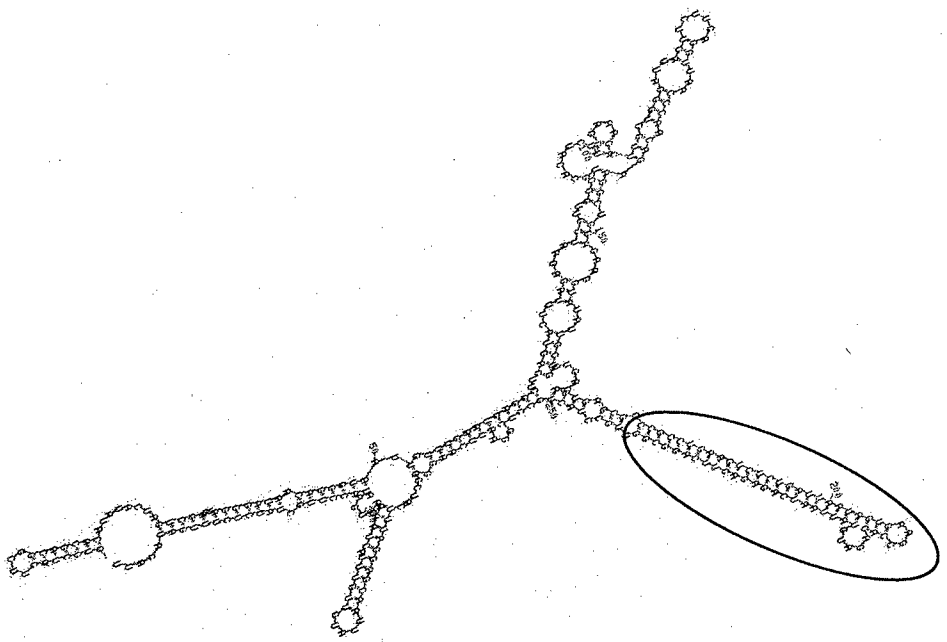
shRNA (first-generation)

gugcucgcuucggcagcacauauacuaUUAAUCAGAGACUUCAGGCGUUCAACGAU*uugg*
AUCGUUGAC**CGCCUGAAGUCUCUGAUUAA**UU



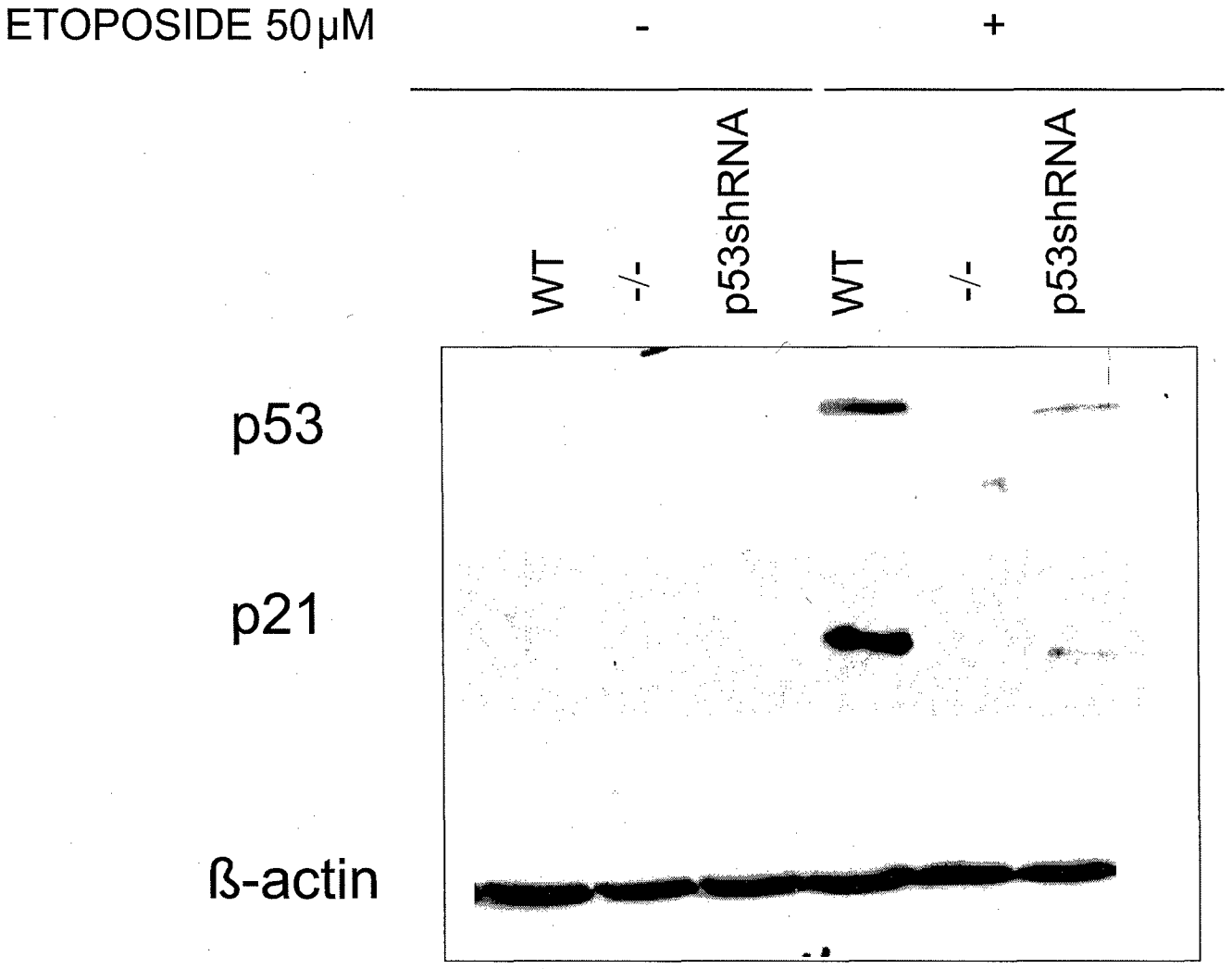
shRNA^{mir} (second-generation)

gugcucgcuucggcagcacauauacuagucgacuagggauaacaggguaauuguuugaaugaggcuucaguacuu
uacagaaucguugccugcacaucuuggaaacacuugcugggauuacuucuucagguuaacccaacagaaggcucga
gaaggguauauugcuguugacagugagcgcc**CGCCUGAAGUCUCUGAUUAA**UA*gugaagccaca*
*gaugua*UUAAUCAGAGACUUCAGGCGGUugccuacugccucggaauucaaggggcuacuuuagg
agcaauuaucuuguuuacuaaaacugaauaccuugcuaucucuuugauacauu



Underline : Leader sequence
CAPITAL: shRNA
CAPITAL BOLD: sense target sequence
Italic: shRNA loop

Supplementary Figure 2. The complete insert sequences for pSM1 and pSM2 containing a luciferase shRNA are shown along with their most stable potential secondary structures as predicted by RNA fold



Supplementary Figure 3. Stable suppression by pSM2. HCT116 cells were infected with pSM2_hsP53_2120 (v2HS_93615) and selected as a population for resistance to puromycin. To induce p53, populations were treated with 50 microM etoposide for 24 prior to lysis for Western blotting. For comparison HCT116-p53null cells were also examined. Lysates were examined for levels of p53, a p53 target, p21, and -actin as a control. Titers were approxmately 1x10^6/ml.