A resource for large-scale RNA-interference-based screens in mammals

Patrick J. Paddison[1]*, Jose M. Silva[1]*, Douglas S. Conklin[1]*†, Mike Schlabach[2]†, Mamie Li[2], Shola Aruleba[1], Vivekanand Balija[1], Andy O'Shaughnessy[1], Lidia Gnoj[1], Kim Scobie[1], Kenneth Chang[1], Thomas Westbrook[2]†, Michele Cleary[3], Ravi Sachidanandam[1], W. Richard McCombie[1], Stephen J. Elledge[2]† & Gregory J. Hannon[1]

[1]*Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA*
[2]*Department of Biochemistry, Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA*
[3]*Rosetta Inpharmatics, 12040 115th Avenue NE, Kirkland, Washington 98034, USA*

* These authors contributed equally to this work
† Present addresses: Department of Biomedical Sciences, Center for Functional Genomics, University at Albany, East Campus, B342A, One University Place, Rensselaer, New York 12144-2345, USA (D.S.C.); Department of Genetics, Harvard Partners Center for Genetics and Genomics, Harvard Medical School Room 158D, NRB, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA (M.S., T.W. and S.J.E.)

Gene silencing by RNA interference (RNAi) in mammalian cells using small interfering RNAs (siRNAs) and short hairpin RNAs (shRNAs) has become a valuable genetic tool[1–10]. Here, we report the construction and application of a shRNA expression library targeting 9,610 human and 5,563 mouse genes. This library is presently composed of about 28,000 sequence-verified shRNA expression cassettes contained within multi-functional vectors, which permit shRNA cassettes to be packaged in retroviruses, tracked in mixed cell populations by means of DNA 'bar codes', and shuttled to customized vectors by bacterial mating. In order to validate the library, we used a genetic screen designed to report defects in human proteasome function. Our results suggest that our large-scale RNAi library can be used in specific, genetic applications in mammals, and will become a valuable resource for gene analysis and discovery.

In invertebrates, RNAi has been harnessed as a powerful genetic tool to reduce endogenous gene expression through programming organisms or cells with homologous double-stranded (ds)RNA (reviewed in ref. 1). In *Caenorhabditis elegans*, this approach has been used in large-scale, genome-wide screens[2–5]. With the advent of RNAi in mammals and the refinement of techniques to trigger gene silencing, expression from any human or mouse transcript can, in principle, be inhibited using siRNAs or shRNAs.

To facilitate the use of RNAi as a genetic tool in mammals, we have constructed a large-scale library of RNAi-inducing shRNA expression vectors targeting human and mouse genes. We began by testing a number of variables that might affect shRNA performance. We previously showed that shRNAs containing 29 nucleotides of dsRNA and simple loop structures are effective silencing triggers when expressed from U6 small nuclear RNA promoters[6,7]. Other published accounts of shRNA-mediated inhibition include differences in promoter choice (reviewed in ref. 8), orientation of sense and antisense strands, length of RNA duplex (19–29 nucleotides), loop structure, and the addition of a 27-nucleotide U6 leader sequence[9]. After comparing each of these variables with a series of eight shRNAs targeting firefly luciferase or green fluorescent protein (GFP), we found that RNA polymerase III promoters were largely interchangeable (Supplementary Fig. 1a, b), that 29-nucleotide hairpins were more effective than shorter shRNAs (Supplementary Fig. 1a and data not shown), that changes in loop structure had minimal effects (data not shown), and that the addition of the U6 leader sequence had a positive effect (Supplementary Fig. 1c). The finalized hairpin design for this library is presented in Supplementary Fig. 2a. It contains a 27-nucleotide U6 leader sequence, followed by 29 base pairs (bp) of dsRNA and a 4-nucleotide loop. Recent studies have suggested that each shRNA is processed from its stem end by Dicer through a single cleavage event (D. Siolas and G.J.H., unpublished data). Thus, the combination of Drosha and Dicer processing is predicted to generate a precisely defined siRNA from each 29-nucleotide shRNA expression vector.

shRNAs were designed covering approximately 10,000 human and 5,000 mouse genes with between three and nine constructs each (Supplementary Fig. 2b). Only coding sequences were targeted, and each shRNA was chosen such that it contained >3 mismatches to any other gene. Where possible, shRNAs had sequence identity to the mouse orthologue of the targeted gene. In pilot experiments,

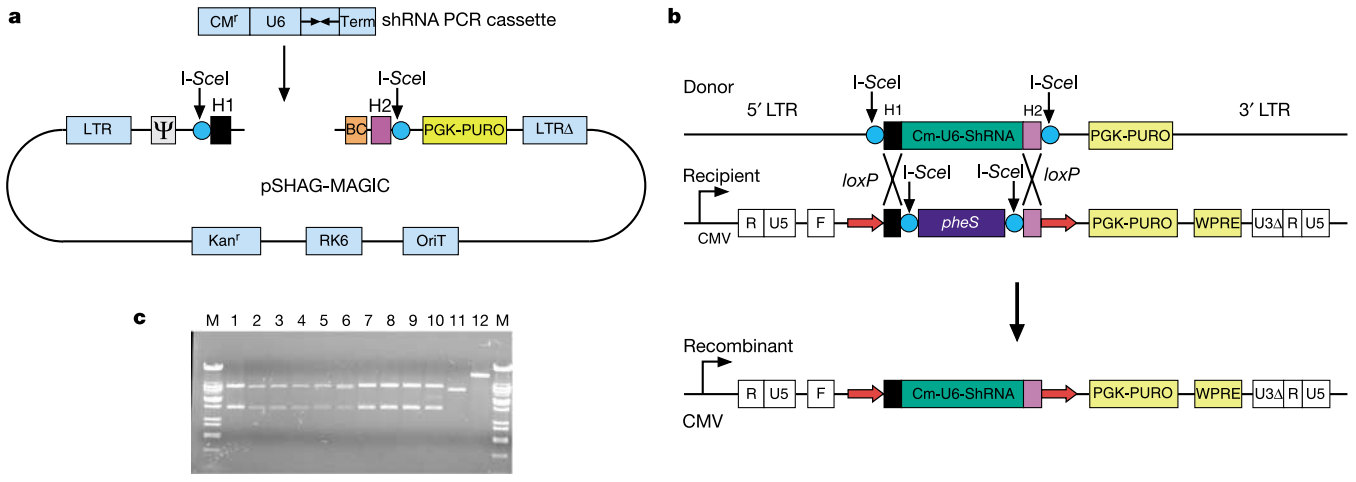

Figure 1 pSHAG-MAGIC shRNA cassette movement strategy. **a**, Map of the pSHAG-MAGIC vector. CMr, chloramphenicol-resistance gene; Kanr, kanamycin-resistance gene; LTR, long terminal repeat; OriT, origin of transfer. **b**, A diagrammatic representation of DNA exchanges occurring once the pSHAG-MAGIC donor vector has been transferred to cells containing a recipient vector by mating. In this case pLenti-LoxP (M.L. and S.J.E., unpublished data) is the recipient vector. WPRE, woodchuck hepatitis B post-transcriptional responsive element. **c**, Plasmids from ten independent colonies (post-mating) were digested with *Nde*I and the digestion products were separated on a 0.5% agarose gel (lanes 1–10). The parental plasmids, pSHAG-MAGIC (lane 11) and pLenti-LoxP (lane 12), each contain a single *Nde*I site, and on cleavage generate a 7.0- or 10.6-kilobase (kb) fragment, respectively. The pSHAG-MAGIC vector contains a unique *Nde*I site in the U6 promoter, which is transferred into the recipient vector. The correct mating product generates two fragments of 3.5 kb and 7.4 kb. M, marker.

between 25–75% of cloned shRNAs contained significant mutations, which arose during chemical synthesis. As a result we sequence-verified each shRNA in the library (see Supplementary Fig. 2b).

Our current sequence-verified library has a representation of 9,610 human and 5,563 mouse genes. This corresponds to 28,659 shRNAs targeting human genes and 9,119 shRNAs targeting mouse genes. Examples of library coverage for selected functional categories of human genes are presented in Supplementary Table 1. The most thoroughly represented functional groups in our current library are kinases and phosphatases, in which approximately 85% of all known kinases and phosphatases are targeted by three or more hairpins. Most other functional classes contain between 30–60% coverage with three or more hairpins, and >80% of genes in functional categories listed in Supplementary Table 1 are targeted by at least one sequence-verified hairpin.

The shRNA expression library has been constructed in a vector that contains a number of convenient design features (pSHAG-MAGIC, Fig. 1a). The vector is capable of producing self-inactivating murine-stem-cell-virus (MSCV) particles in commonly available retroviral packaging lines. We have also incorporated a new *in vivo* subcloning technology. This method is called 'Mating-Assisted Genetically Integrated Cloning' (MAGIC) (M.L. and S.J.E., manuscript in preparation). The MAGIC system consists of a donor vector (the library vector), in which the fragment of interest is flanked by two different 50-bp homology regions, H1 and H2, which in turn are flanked with linked I-*Sce*I sites. The donor vector also includes an F′ origin and a conditional origin of replication (RK6γ; Fig. 1a, b). The recipient vector, which also contains I-*Sce*I-linked H1 and H2 sites surrounding a negative selectable marker (*pheS*), resides in a bacterial strain that contains an inducible I-*Sce*I gene (Fig. 1b). After transfer of the donor vector into the recipient host by bacterial mating, I-*Sce*I cleaves both donor and recipient vectors, and these breaks are healed by homologous recombination via the H1 and H2 sequences. Selection against the unrecombined recipient containing *pheS* and I-*Sce*I sites and for the capture of the appropriate insert (chloramphenicol resistance) gives essentially 100% recovery of the desired plasmid. To test MAGIC transfer of shRNA clones, we developed a lentivirus recipient vector based on the FUW vector[10]. Using a test shRNA clone in the mating, we observed a mating efficiency of $>6 \times 10^6$ clones per ml of mixed bacteria. Restriction analysis showed that 10 out of 10 recombinants have the expected structure (Fig. 1c).

The shRNA library was designed to function for both genetic selections and screens. For selections conferring a growth advantage, the library can be used in a pooled fashion. Genetic screens (for example, for lethal events) can be carried out in a 96-well format; however, this method is labour intensive and time consuming. To facilitate such screens, we have adopted a DNA bar-coding strategy

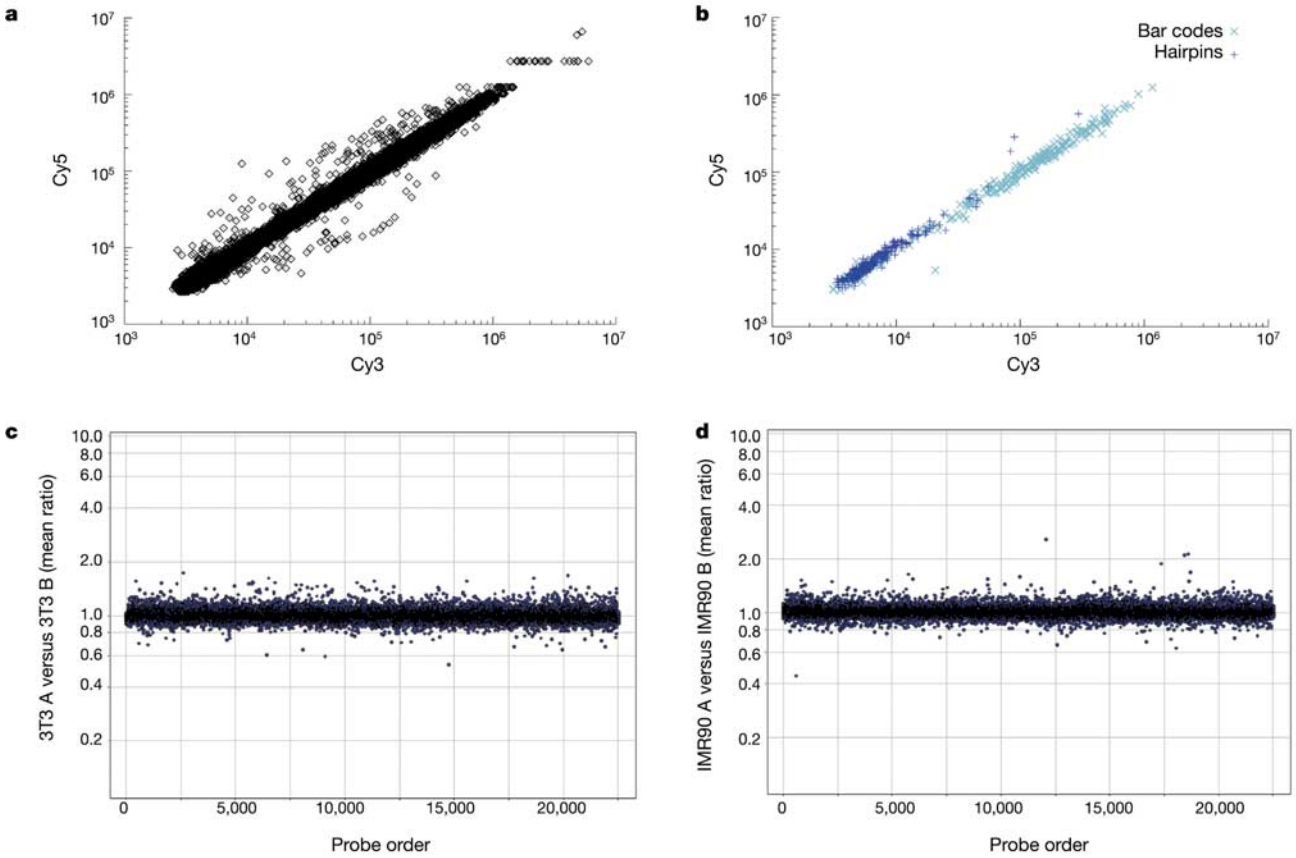

Figure 2 Microarray analysis of pSHAG-MAGIC library bar codes. **a**, Self–self hybridization of pSHAG-MAGIC library bar codes obtained from approximately 15,000 library plasmids prepared from *E. coli*. The DNA microarray (Agilent Technologies) is composed of 20,241 complementary 60-nuclotide oligonucleotides out of a total of 22,575 elements on the array, including controls; the remaining represent various positive and negative controls. Cy-labelled cRNA for library bar codes was generated and competitively hybridized as described in the Supplementary Methods. **b**, An analysis of a subset of 255 60-nucleotide bar codes from **a** versus 255 60-nucleotide shRNA probes. Each 60-nucleotide shRNA probe contains a direct repeat of the 29-nucleotide gene-targeting sequence with a 2-nucleotide spacer. **c**, Microarray analysis of pSHAG-MAGIC bar codes from transduced NIH3 T3 cells. Two pools of 2×10^7 cells (A and B) were infected separately and harvested 48 h after infection. Bar codes were processed and competitively hybridized as described in the Methods. The *x* axis is an index of all the elements (22,575), bar codes, hairpins and controls on the array. The *y* axis shows the mean measured ratios from two experiments (forward and reverse colour orientation) of 3T3 pool A hybridized against 3T3 pool B, plotted on a log scale with the numbers indicating fold change. Most probes, with the exception of a few, scatter around 1.0 (no change) with a range between 0.5 (twofold decrease) and 2.0 (twofold increase). **d**, Microarray analysis of pSHAG-MAGIC bar codes in transduced normal human diploid fibroblasts (IMR90) cells carried out as in **c**.

that has been used previously in *Saccharomyces cerevisiae* deletion collections, to follow individual mutants in complex populations via microarray analysis[11–13]. We have linked a unique 60-nuclotide DNA bar code to each shRNA vector to allow us to follow the fate of shRNAs in populations of virally transduced cells. To monitor relative frequencies of individual shRNA species, we designed microarrays (Agilent Technologies) containing 20,241 60-nucleotide DNAs complementary to each shRNA bar code.

Sequence analysis of library clones revealed that a fraction (about 4%) contained orphan shRNAs without bar codes. We therefore sought to determine whether we could use the shRNA sequence as a bar code. We chose 255 non-orphan shRNAs and deposited on arrays both associated bar codes and oligonucleotides designed have 60 hybridize to the shRNA itself. Such oligonucleotides were designed nucleotides in order to contain direct repeats of the 29-nucleotide antisense sequence. Cy-labelled antisense RNA was generated from about 15,000 library constructs such that it contained the reverse complement of the bar code and the entire shRNA. A linear log–log plot of a self hybridization revealed consistent cRNA production and hybridization from the plasmid library (Fig. 2a). However, hairpins gave substantially lower hybridization signals, essentially at background levels, relative to the corresponding bar codes (Fig. 2b), indicating that using hairpins as probes is not an optimal bar-coding strategy under these conditions.

We next asked whether bar codes could be used to report the representation of individual shRNAs in transduced mammalian cells. Normal human diploid fibroblasts (IMR90) or NIH3T3 cells

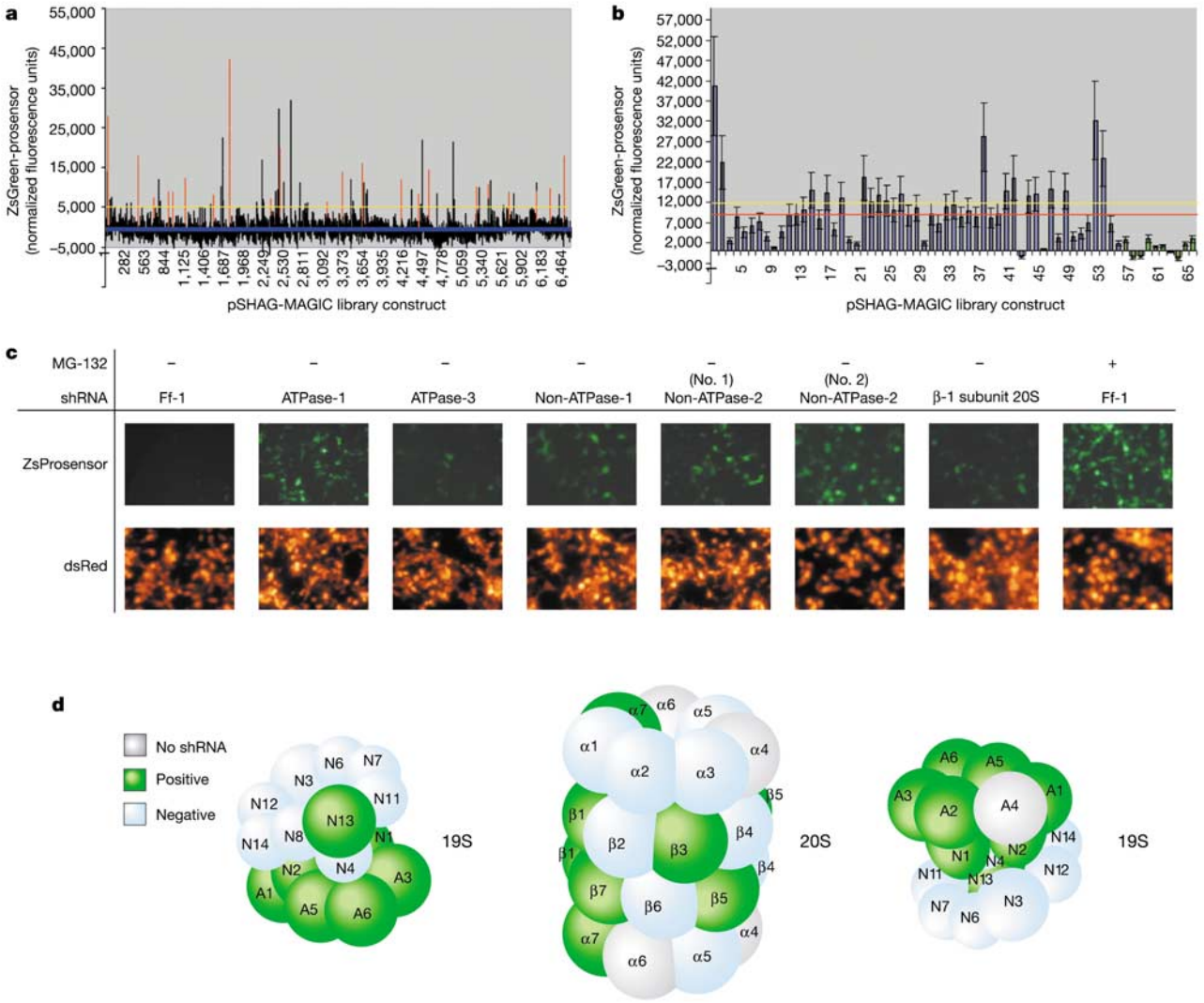

Figure 3 A reverse genetic screen for defects in human proteasome function. **a**, A graph of relative ZsGreen fluorescence for all pSHAG-MAGIC clones transfected. Red vertical bars indicate 22 positively scoring proteasome shRNAs corresponding to 15 known proteasome subunits, whereas black bars indicate library shRNAs. The yellow horizontal bar indicates a cut-off that was set based on control experiments. **b**, A replicate transfection experiment as in **a** using a set of 22 positively scoring proteasome hairpins and 33 that failed to score (blue bars), along with non-proteasome hairpin controls (green bars). These experiments were carried out in triplicate. All 22 of the proteasome hairpins with a positive score from the original screen were re-tested and also achieved a positive score. None of the 33 non-scoring proteasome hairpins from the original screen managed to match the score of the 22 positively scoring proteasome shRNAs from the first round (yellow line). However, in this experiment 36 of 55 scored well above (~2 s.d.) the mean background (red line). **c**, Fluorescence microscopy images showing representative results for individual proteasome hairpins as carried out in **a** and **b**. A shRNA targeting firefly luciferase (Ff-1) and small molecule proteasome inhibitor, MG-132 (Sigma), were used as negative and positive controls, respectively. **d**, A diagrammatic representation of the 26S proteasome colour-coded according to pSHAG-MAGIC library hits. Subunits coloured green had strong positive shRNA hits from the library in the primary screen. Grey subunits were not represented by any shRNAs in the approximately 7,000 tested. Subunits coloured blue were represented by at least 1 shRNA but did not score in the screen. For nomenclature see Supplementary Table 2.

were infected with retroviruses derived from the library population that was used for control hybridizations. To test for experimental variability, two populations of 2×10^7 cells (denoted A and B) were infected independently at a multiplicity of infection of approximately 1. Examination of colour-reversal experiments and plots comparing the relative intensity of the bar-code signals in these populations (Figs 2c, d) indicated that all steps of the procedure were highly reproducible, with the ratio of intensities varying by less than twofold in all but a few isolated cases for each cell line.

Finally, we sought to test the performance of the shRNA library in a biological context. We focused on an assay for which we could predict the recovery of a substantial number of shRNAs targeting genes in a known biological pathway. The 26S proteasome is the major non-lysosomal protease in eukaryotic cells. To search the library for shRNAs that compromise proteasome function, we used a reporter assay in which a fluorescent protein is coupled to a well-characterized degradation signal. The mouse ornithine decarboxylase (MODC) gene contains a PEST sequence that directs proteasomal degradation without the need for ubiquitination[14]. Roughly 7,000 shRNA expression plasmids, corresponding to approximately one-quarter of the complete library, were individually co-transfected with two expression constructs. The first encoded a *Zoanthus* green fluorescent (ZsGreen)–MODC degron fusion. In cases in which shRNAs compromised proteasome function, ZsGreen–MODC was expected to accumulate, giving a detectable signal. The second plasmid encoded *Discosoma* red fluorescent protein (DsRed). This permitted normalization of signals, which controlled for transfection efficiency.

An analysis of 6,712 shRNAs targeting 4,873 genes revealed approximately 100 RNAi constructs that increased the accumulation of ZsGreen–MODC (Fig. 3a). Twenty-two of these corresponded to 15 known proteasome subunits. As a secondary test, these 22 putative proteasome-positive shRNAs, and an additional 33 shRNAs that targeted proteasome subunits but which had not scored in the original screen, were selected from the population. These were assayed in replicate transfections for the ability to increase the stability of the unstable fluorescent protein. Again, the 22 shRNAs scored positively, whereas the 33 proteasome shRNAs not detected in the initial screen scored less well. However, an additional 14 shRNAs did score above background in the focused assay (Fig. 3b, c).

Notably, many of the positive shRNAs targeted 19S base subunits, including five out of five ATPases and the two largest non-ATPases (1 and 2) (Fig. 3c, d; see also Supplementary Table 2). Compared with the 19S base, targeting subunits of the 19S lid or the 20S core produced a lower hit rate. However, two out of three of the most important catalytic components, located in subunits β1 (peptidyl-glutamyl hydrolysing or caspase-like) and β5 (chymotrypsin-like)[15,16], achieved a positive score in our assay (Fig. 3). The other two β-subunits that we identified are involved in important *cis* contacts between neighbouring subunits (β3) and in *trans* contacts between the two β-rings (β7)[17,18]. In all cases tested, activation of the reporter was accompanied by a reduction in the expression of targeted proteasomal components (Fig. 4a and data not shown).

The c-Myc oncoprotein is a target for ubiquitin-mediated degradation by the proteasome[19]. As was observed for ZsGreen–MODC degron and bulk-ubiquitinated protein (not shown), c-Myc levels specifically increased upon transfection of cells with shRNAs that targeted proteasomal subunits (Fig. 4b).

Here, we report progress towards the construction of a genome-wide library of RNAi-inducing constructs for use in mammalian cells. At present the resource targets approximately 10,000 human and 5,000 murine genes, and it continues to expand. To examine the performance of the library, we have tested about one-quarter of its constituent clones (roughly 7,000 shRNA expression vectors) individually for the ability to inhibit the degradation of a direct proteasome target. Nearly 50% of the shRNAs in this collection that were expected to target proteosomal proteins were recovered as positives. The availability of this resource to the research community will open the door to the use of RNAi in mammalian systems as a large-scale tool for biological discovery. □

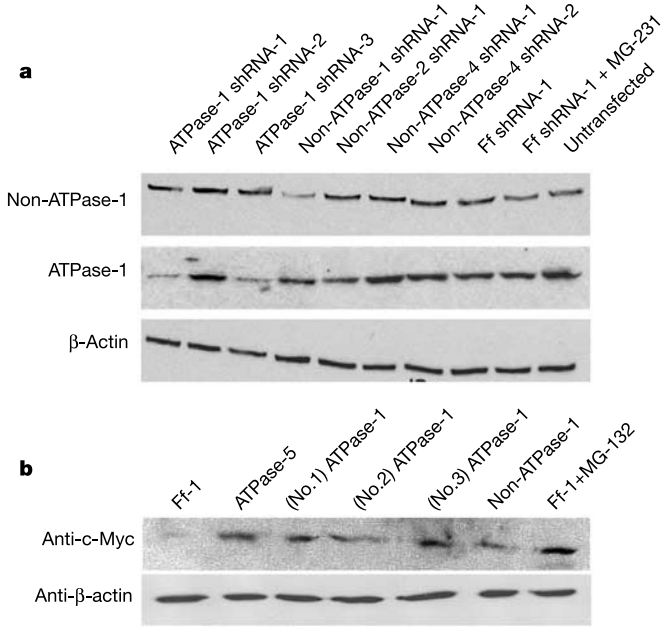

Figure 4 Further validation of selected pSHAG-MAGIC proteasome hairpins. **a**, Western blot showing specific suppression of the ATPase-1 and non-ATPase-1 of the 19S regulatory subunit in transiently transfected HEK293 cells. Cells were transfected with shRNAs as indicated. Knockdown of protein levels for shRNA-1 and -3 against ATPase-1 proteasome correlated with the severity of the relative scores in the pZsGreen–MODC accumulation assay. The lane labelled untransfected indicates the control lane where cells were not transfected. **b**, A western blot showing increased steady-state levels of endogenous c-Myc in HEK293 cells transiently transfected with library shRNAs as indicated. c-Myc is normally degraded by ubiquitin-mediated proteolysis in these cells[19].

Methods

A detailed protocol for the design and creation of shRNA polymerase chain reaction (PCR) cassettes is available at http://www.cshl.edu/public/SCIENCE/hannon.html. Oligonucleotides were purchased from Sigma-Genosys. The pSHAG-MAGIC shRNA library was cloned in the MAGIC-competent *Escherichia coli* strain BW23474F′DOT (M.L. and S.J.E., manuscript in preparation). The full sequence of the pSHAG-MAGIC vector and details of the mating protocol are available for download at the shRNA library website (http://www.cshl.edu/public/SCIENCE/hannon.html). Bar-code microarrays were synthesized by Agilent Technologies and were analysed as described in the Supplementary Methods.

Plasmids were prepared as described in the Supplementary Methods. Transfections were carried out in 96-well plates using plasmid mixtures described in Supplementary Methods. 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 16 independent control shRNA transfections to determine the range for a negative outcome.

Received 16 November 2003; accepted 26 January 2004; doi:10.1038/nature02370.

1. Hannon, G. J. RNA interference. *Nature* **418,** 244–251 (2002).
2. Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299,** 2039–2045 (2003).
3. Lee, S. S. *et al.* A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. *Nature Genet.* **33,** 40–48 (2003).
4. Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408,** 331–336 (2000).
5. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408,** 325–330 (2000).
6. 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).
7. Hemann, M. T. *et al.* An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes *in vivo*. *Nature Genet.* **33,** 396–400 (2003).
8. Paddison, P. J. & Hannon, G. J. siRNAs and shRNAs: skeleton keys to the human genome. *Curr. Opin. Mol. Ther.* **5,** 217–224 (2003).

9. Paul, C. P., Good, P. D., Winer, I. & Engelke, D. R. Effective expression of small interfering RNA in human cells. *Nature Biotechnol.* **20,** 505–508 (2002).
10. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295,** 868–872 (2002).
11. Birrell, G. W., Giaever, G., Chu, A. M., Davis, R. W. & Brown, J. M. A genome-wide screen in *Saccharomyces cerevisiae* for genes affecting UV radiation sensitivity. *Proc. Natl Acad. Sci. USA* **98,** 12608–12613 (2001).
12. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).
13. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).
14. Ghoda, L., Sidney, D., Macrae, M. & Coffino, P. Structural elements of ornithine decarboxylase required for intracellular degradation and polyamine-dependent regulation. *Mol. Cell. Biol.* **12,** 2178–2185 (1992).
15. Chen, P. & Hochstrasser, M. Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* **86,** 961–972 (1996).
16. Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U. & Wolf, D. H. The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* **272,** 25200–25209 (1997).
17. Bochtler, M., Ditzel, L., Groll, M., Hartmann, C. & Huber, R. The proteasome. *Annu. Rev. Biophys. Biomol. Struct.* **28,** 295–317 (1999).
18. Coux, O. An interaction map of proteasome subunits. *Biochem. Soc. Trans.* **31,** 465–469 (2003).
19. Kim, S. Y., Herbst, A., Tworkowski, K. A., Salghetti, S. E. & Tansey, W. P. Skp2 regulates Myc protein stability and activity. *Mol. Cell* **11,** 1177–1188 (2003).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank T. Moore and B. Simmons from Open Biosystems for their help in organizing and rearraying the library, and colleagues at CSHL and elsewhere (as indicated in Supplementary Table 1) as well as J. LaBaer and C. Perou for curating gene lists. G. Katari and J. Faith helped with bioinformatic analysis and shRNA choice, and members of the Lowe laboratory (CSHL) provided advice on vector optimization. This work was supported by an Innovator Award from the US Army Breast Cancer Research Program (G.J.H.), a contract from the National Cancer Institute (G.J.H.), grants from the NIH (G.J.H., W.R.M., S.J.E.) and the US Army Breast Cancer Research Program (G.J.H., D.S.C.), the Howard Hughes Medical Institute (S.J.E.), and by generous support from Oncogene Sciences and Merck. P.J.P. is an Arnold and Mabel Beckman Fellow of the Watson School of Biological Sciences and is supported by a predoctoral fellowship from the US Army Breast Cancer Research Program. J.M.S. is supported by a postdoctoral fellowship from the US Army Prostate Cancer Research Program. S.J.E. is an Investigator of the Howard Hughes Medical Institute. G.J.H. is a Rita Allen Foundation Fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.org) or S.J.E. (selledge@genetics.med.harvard.edu).

A large-scale RNAi screen in human cells identifies new components of the p53 pathway

Katrien Berns[1]*, E. Marielle Hijmans[1]*, Jasper Mullenders[1], Thijn R. Brummelkamp[1], Arno Velds[1], Mike Heimerikx[1], Ron M. Kerkhoven[1], Mandy Madiredjo[1], Wouter Nijkamp[1], Britta Weigelt[2], Reuven Agami[3], Wei Ge[4], Guy Cavet[4], Peter S. Linsley[4], Roderick L. Beijersbergen[1] & René Bernards[1]

[1]*Division of Molecular Carcinogenesis and Center for Biomedical Genetics,* [2]*Division of Experimental Therapy, and* [3]*Division of Tumor Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*
[4]*Rosetta Inpharmatics, Inc., 12040 115th Avenue NE, Kirkland, Washington 98034, USA*

* These authors contributed equally to this work

RNA interference (RNAi) is a powerful new tool with which to perform loss-of-function genetic screens in lower organisms and can greatly facilitate the identification of components of cellular signalling pathways[1–3]. In mammalian cells, such screens have been hampered by a lack of suitable tools that can be used on a large scale. We and others have recently developed expression vectors to direct the synthesis of short hairpin RNAs (shRNAs) that act as short interfering RNA (siRNA)-like molecules to stably suppress gene expression[4,5]. Here we report the construction of a set of retroviral vectors encoding 23,742 distinct shRNAs, which target 7,914 different human genes for suppression. We use this RNAi library in human cells to identify one known and five new modulators of *p53*-dependent proliferation arrest. Suppression of these genes confers resistance to both *p53*-dependent and $p19^{ARF}$-dependent proliferation arrest, and abolishes a DNA-damage-induced G1 cell-cycle arrest. Furthermore, we describe siRNA bar-code screens to rapidly identify individual siRNA vectors associated with a specific phenotype. These new tools will greatly facilitate large-scale loss-of-function genetic screens in mammalian cells.

RNAi is a defence mechanism triggered by double-stranded (ds)RNAs to protect cells from parasitic nucleic acids. The dsRNAs are processed into siRNAs, which target homologous RNAs for destruction[6]. In mammalian cells, an RNAi response can be triggered by 21-base-pair siRNAs, which can cause strong, but transient, inhibition of gene expression[7]. By contrast, vector-expressed shRNAs can suppress gene expression over prolonged periods[4,5]. In the current study, we create a large set of vectors and use them to search for components of the p53 tumour-suppressor pathway. This pathway is crucial for genome integrity as it transmits both anti-proliferative and pro-apoptotic signals in response to a variety of stress signals[8].

To construct a human RNAi library (the 'NKi library'), we selected 7,914 human genes for shRNA-mediated reduction in expression, known as knockdown. This collection of genes includes components of major cellular pathways, including the cell cycle, transcription regulation, stress signalling, signal transduction and important biological processes such as biosynthesis, proteolysis and metabolism. In addition, genes implicated in cancer and other diseases are included in the library (see Supplementary Table 1). To increase the likelihood of obtaining a significant inhibition of gene expression, we constructed three different shRNA vectors against each gene (23,742 vectors in total; Fig. 1a and Supplementary Fig. 1). The oligonucleotides specifying the shRNAs were annealed and cloned in a high-throughput fashion into pRetroSuper (pRS), a retroviral vector that contains the shRNA expression cassette[9]. Using a pool of three knockdown vectors against a single gene, we obtain on average 70% inhibition of expression for approximately 70% of the genes in the library (see ref. 10; data not shown). The vector-based shRNA library can be used for functional genetic screens in both short-term and long-term assays using DNA transfection or retroviral transduction.

To validate the RNAi library, we developed a cell system to screen for bypass of p53-dependent proliferation arrest. We generated primary human BJ fibroblasts, which ectopically express the murine ecotropic receptor, the telomerase catalytic subunit (TERT) and a temperature-sensitive allele of SV40 large T antigen (tsLT), yielding BJ-TERT-tsLT cells. As can be seen in Fig. 1b, these cells proliferate when grown at 32 °C, the temperature at which tsLT binds and inactivates both retinoblastoma protein (pRB) and p53, but enter into a synchronous proliferation arrest after a shift to 39 °C, at which tsLT is inactive. To determine whether this proliferation arrest is p53 dependent, we infected BJ-TERT-tsLT cells with pRS–p53 (which targets *p53* for suppression) at 32 °C, and shifted them to 39 °C after two days. Figure 1b shows that knockdown of *p53* allowed temperature-shift-induced proliferation arrest to be bypassed. Knockdown of the RB pathway component $p16^{INK4A}$ alone did not allow growth arrest to be bypassed, but simultaneous suppression of both $p16^{INK4A}$ and *p53* yielded a further increase in escape from growth arrest compared with knockdown of *p53* alone (Fig. 1b). Thus, the conditional proliferation arrest in BJ fibroblasts depends primarily on *p53*.

We isolated polyclonal plasmid DNA from each of the 83 pools of