Oncogene (2004) 23, 8401–8409

www.nature.com/onc

RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age

Jose Silva[1], Kenneth Chang[1], Gregory J Hannon*[1] and Fabiola V Rivas[1]

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

Sequencing of complete genomes has provided researchers with a wealth of information to study genome organization, genetic instability, and polymorphisms, as well as a knowledge of all potentially expressed genes. The identification of all genes encoded in the human genome opens the door for large-scale systematic gene silencing using small interfering RNAs (siRNAs) and short hairpin RNAs (shRNAs). With the recent development of siRNA and shRNA expression libraries, the application of RNAi technology to assign function to cancer genes and to delineate molecular pathways in which these genes affect in normal and transformed cells, will contribute significantly to the knowledge necessary to develop new and also improve existing cancer therapy.
Oncogene (2004) **23,** 8401–8409. doi:10.1038/sj.onc.1208176

Keywords: RNAi; high throughput screening; functional genomics; cancer; apoptosis; synthetic lethality

Introduction

RNA interference (RNAi) is a conserved biological response discovered in the nematode *Caenorhabditis elegans*, as a response to double-stranded RNA (dsRNA). Initially demonstrated by Mello and coworkers, who showed that injection of long dsRNA into *C. elegans* led to sequence-specific degradation of the corresponding mRNAs, this silencing response has been subsequently found in other eukaryotes from yeast (*Neurospora crassa* and *Schizosaccharomyces pombe*) to mammals (Fire *et al*., 1998; Hannon, 2002; Montgomery, 2004). Although knowledge of the biological mechanism of RNAi has grown exponentially over the last few years, application of RNAi at the genome-wide level had to await the development of optimal techniques of delivery. These were pioneered in model organisms; for example, RNAi can be triggered by soaking *C. elegans* (Tabara *et al*., 1998) and *Drosophila* cells (Clemens *et al*., 2000) in a solution of dsRNA, or by feeding worms with *Escherichia coli* expressing gene-specific dsRNAs (Timmons and Fire, 1998). In mammalian cells, however, long dsRNAs (>30 nucleotides) elicit an antiviral interferon response (Minks *et al*., 1979; Manche *et al*., 1992). Thus, RNAi technology could not be applied to mammals until the discovery that short dsRNA duplexes, processed from long dsRNA into 21–28 cleavage fragments termed small interfering RNAs (siRNAs), were sufficient to trigger gene-specific silencing upon transfection into mammalian cells (Elbashir *et al*., 2001; Harborth *et al*., 2001). However, the silencing response to transfected siRNAs is transient, lasting from 3 to 7 days depending upon the rate of cell division making this approach unsuitable for analysis of long-term effects of silencing. The search for a more sustained silencing response has resulted in the development of an additional class of triggers, short hairpin RNAs (shRNAs), that can establish stable gene silencing by continuously supplying the RNAi trigger (see for review Paddison and Hannon, 2002). Researchers are now using this technology to understand biological mechanisms in both normal cells and in malignant ones, one of the major goals being to unravel the mysteries of transformation and to improve current cancer therapy.

Viewing cancer as a global epidemic of discrete afflictions is a confounding oversimplification. Each cancer is a unique disease arising from multiple genetic alterations, and the particular combination of genes mutated in any given patient probably determines the degree of malignancy and potential therapeutic vulnerabilities of that individual's cancer. Improved prevention, diagnosis, and treatment of cancer in patients, will require a detailed understanding of the specific molecular mechanisms that go away in specific cancers. This understanding must be derived both from an examination of the cancerous cell itself and from an investigation of the interactions between the cancer and its host. One of the ways in which these insights can be obtained is through functional genetic approaches in mammals.

Traditionally, functional genetic studies are divided into forward or reverse screens. In a typical forward genetic study, genes are mutated at random. The resulting changes in the phenotype of a cell or organism are then attributed to the mutated genes and, by inference, to their protein products. After identification of an abnormal phenotype, the mutations must be mapped, a process which is usually time-consuming and not easily applicable to mammalian systems. Conversely, reverse genetic approaches involve the disruption of a gene of interest, so as to determine its function and/

*Correspondence: GJ Hannon, Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA;
E-mail: hannon@cshl.edu

or involvement in a pathway. Classical reverse genetic approaches involve the creation of knockout cell lines or organisms, and can be expensive, time-consuming and unsuitable for genome-wide screens in mammals. Faster, simpler and cheaper alternatives of attenuating gene function in a sequence-specific manner have emerged in the form of antisense technology, ribozymes and more recently, RNAi (Figure 1).

To facilitate the use of RNAi in mammals, several groups have constructed first-generation RNAi libraries of shRNA expression vectors. For example, we have constructed a library comprising approximately 28 000 shRNA expression cassettes targeting 9610 human and 5563 mouse genes. The expression cassettes in our collections are sequence-verified and contained within multifunctional vectors that can be packaged into retroviruses, tracked in a mixed cell population by means of a random 60-mer DNA 'barcode', and shuttled into customized vectors through bacterial mating (Paddison *et al*., 2004). Bernards and colleagues have constructed a similar arrayed library with overlapping functionality (Berns *et al*., 2004). Given advances in our understanding of the RNAi mechanism over the last year, we have also constructed a second-generation library that uses improved expression cassettes and informatic tools for shRNA design. This library is presently available to academic investigators through several sources (e.g., Open Biosystems, MRC Geneservice). Given the availability of these powerful tools, there is a pressing need to discuss possible screening methodologies that are available for mammalian systems and how such approaches compare to those available in more traditional genetic models. Such a discussion is the purpose of this brief review.

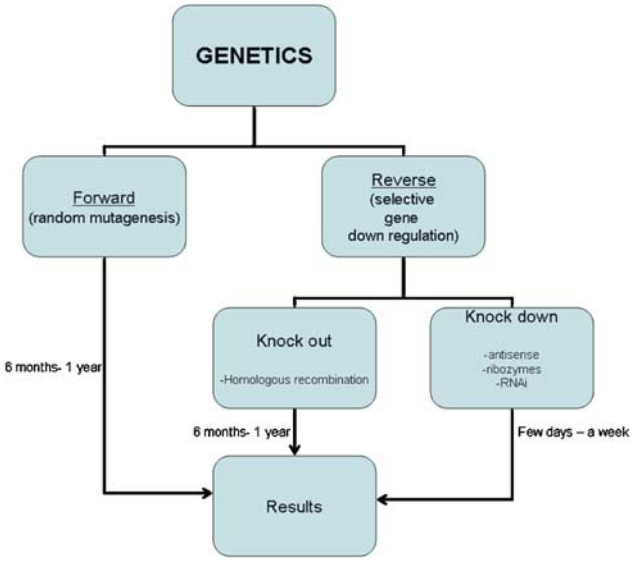

Figure 1 The diagram shows a comparison between forward and reverse genetics. A typical forward genetic study begins with the generation of random mutants in order to produce a specific phenotype. The next step is the identification of the genetic alteration that originated the phenotype. In mammalian cells, the time required to complete the study could be estimated between 6 months–1 year. Classical reverse genetic involves the selective downregulation of a gene function. Although knockout techniques produce the total abrogation of a protein by disruption of the two alleles of the genome, the time necessary to complete the process is also very long (6 months–1 year). On the other hand, new approaches that can achieve 90–95% of suppression of the gene expression like antisense, ribozymes or more recently RNAi are more simple and a much faster alternative

Genetic screens in mammalian cells

The ability to perform genetic screens has been popularized over the last 30 years primarily through the use of model organisms such as bacteria (Shuman and Silhavy, 2003), yeast (Forsburg, 2001), worms (Jorgensen and Mango, 2002), and flies (St Johnston, 2002). The demand for better tools to assign gene function has been made imperative by the advent of genomics, a field that within a few years has produced numerous monumental advances from the complete sequence of *S. cerevisiae* (Dujon *et al*., 2004) to a draft sequence of the human genome. (Celera Genomics Project, 2001; International Human Genome Sequencing Consortium, 2001). The availability of such copious sequence information has thrown into sharp relief the need for versatile technologies for decoding gene function. Nowhere has this been more apparent than in mammalian systems where only a few years ago the existing approaches to functional genomics offered few options. There have been isolated successes with overexpression screens (Michiels *et al*., 2002; Huang *et al*., 2004), insertional mutagenesis (Mikkers and Berns, 2003) and genome-wide two-hybrid studies (Chen and Han, 2000). However, approaches for routine, loss-of-function genetics on a large scale, particularly in cultured cells, were lacking.

In principle, cultured mammalian cells have many of the benefits associated with yeast as models for the study of eukaryotic cell genetics. Like yeast, they are amenable to gene transfer in mass or individually, a variety of selectable/detectable markers is available for establishing expression of heterologous genes, constructs can be maintained extrachromosomally (transient) or stably integrated via viral vectors, and in almost all cases, different genetic lines can be established from mammalian cell lines by cloning at limiting dilution. Additionally, mammalian cell lines present important advantages to yeast as models not only of cell genetics of higher eukaryotes, but also of biological process that have an impact at the organismal level. Accumulating data show that many human and murine genes are not represented in yeast, and there are no true counterparts to cellular processes such as transformation, apoptosis, tissue-specific differentiation and some signaling pathways in these lower eukaryotes (Aravind *et al*., 2001).

The technological shortfall in mammalian functional genomics has been, to a degree, met by recent advances toward the creation of routine RNAi-based tools for gene silencing in cultured cells and in animals. The initial finding that siRNAs and encoded shRNAs could trigger gene silencing in mammals has been extended by

the combination of RNAi with viral and episomal vectors that allow stable maintenance of silencing in mammals (Paddison and Hannon, 2003). This has reinvigorated and broadened interest in using mammalian cells for both forward and reverse genetic screens.

There are a number of possible approaches to screening large collections of silencing triggers for their effects on mammalian cells, and several of those will be discussed herein as examples of how RNAi libraries may be applied to cancer-relevant biological problems in mammalian cells. Three screening modalities will be discussed. First, individual siRNAs or shRNAs can be transfected and screened in multiwell format for activation or repression, of a reporter or activity in a cell-based or biochemical assay (Somma *et al*., 2002; Aza-Blanc *et al*., 2003; Brummelkamp *et al*., 2003; Boutros *et al*., 2004; Hsieh *et al*., 2004; Paddison *et al*., 2004). In this format, individual genes are transiently suppressed 'one-by-one' and analysis is carried out in a high throughput manner using a robotic platform. Second, cells can be infected with pools of shRNAs (at MOI of 1.0), followed by selection of individual colonies that can be scored phenotypically for alterations that result from the expression of a specific trigger (e.g., a morphological alteration). A variant of this procedure allows siRNAs or shRNAs to be delivered *in situ* via 'reverse transfection' with resulting phenotypes being scored in transfected microcolonies. Third, transduced pools of cells can be monitored in mass, via DNA 'barcodes' contained within the shRNAs, by high-density oligonucleotide microarrays for relative changes in shRNA representation following application of a selective stimulus (Paddison *et al*., 2004; Berns *et al*., 2004). A variant of these methods involves long-term selection for growth of colonies under specific conditions. This is defined as a genetic selection rather than a screen and will be discussed separately, below. Of course, the distinction between genetic selection and genetic screen can be subject to interpretation, particularly as modern techniques (e.g. barcode arrays) are mapped onto these longstanding definitions.

Transient screens (cell-based assays)

In these types of assays, cells are plated for high throughput studies, normally in 96- or 384-well plates, and each individual well is transfected with a different siRNA, shRNA or limited-complexity pool of triggers (Figure 2). The transfection of the RNAi triggers is usually mediated by lipid-based reagents that allow for high and reproducible transfection efficiencies on robotic platforms. An enormous variety of lipid-based transfection agents are already available in the market, and allow for transfection of siRNAs, plasmids, or both.

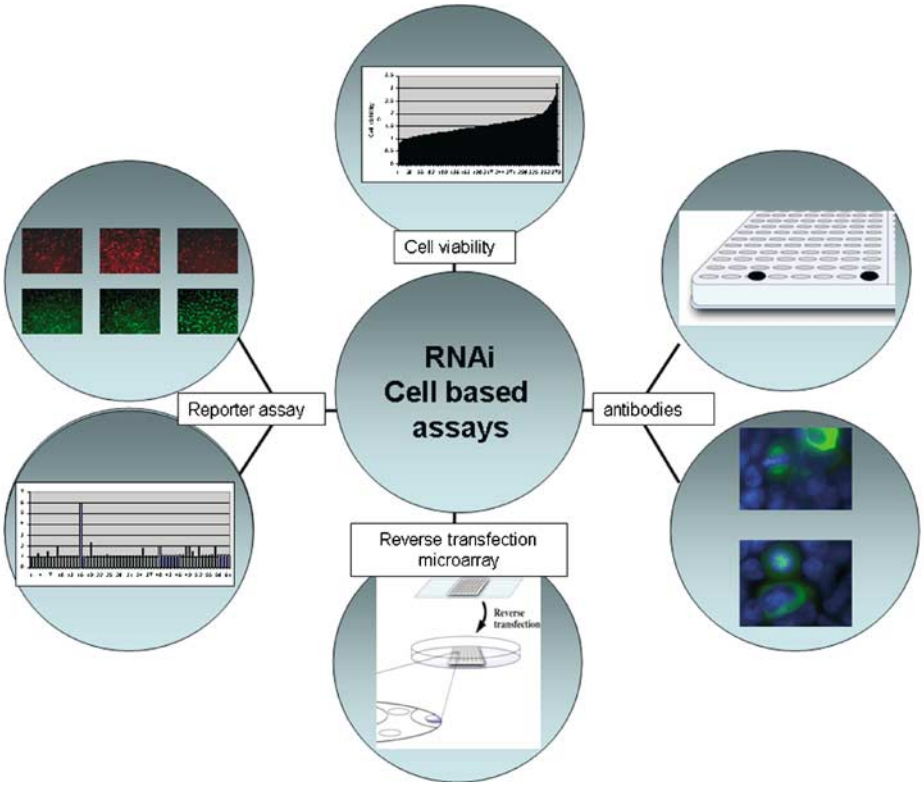

Figure 2 The graphic shows different possibilities for performing cell-based screens using RNAi. When combined with a simple cell viability assay it can identify essential genes. Fluorescent and luminescent reporters that are activated by a specific pathway offer the possibility to identify more detailed phenotypes. Additionally, classic immunofluorescence techniques increase the number of phenotypes that can be studied. The use of automated microscopy platforms allows the realization of high through put screens that can identify even light morphological abnormalities. Finally, the adaptation of these screens to a miniaturized format will simplify enormously the realization of these studies

Selection of the transfection carrier depends on both the cell type and the RNAi silencer being used, and their efficiency should be tested empirically with both positive and negative controls. In addition to classical methods of transfection, a novel approach named 'reverse transfection' has been recently described (Ziauddin and Sabatini, 2001). This technique is especially suitable for high throughput screens. A mixture of transfection reagent and nucleic acids is dispensed on a well of a tissue culture plate and dried for storage. The cells plated on the transfection layer incorporate the lipid–nucleic acid complexes, and after several hours, the effect of the transfection can be examined.

A promising new technology that greatly expands the potential throughput of RNAi-based screens adapts the reverse transfection method to a microarray-based platform. Mammalian cells are plated on a glass slide spotted in defined locations with transfection mixtures containing different RNAi triggers (Kumar *et al*., 2003; Mousses *et al*., 2003; Silva *et al*., 2004). Cells growing on the printed areas take up the nucleic acids, creating spots of localized transfection within a lawn of nontransfected cells. Although theoretically this technology shows a great potential for genome-wide analysis (thousand of knockdowns can be analysed on a slide), there are still technical limitations to be overcome. For instance, because of the small number of cells that are transfected (between 50 to a few hundred cells), variability and sensitivity are parameters that may compromise the results. Moreover, studies that require long incubation times to reveal a phenotype may not be amenable to this approach. Most importantly, there are thus far a limited number of cell lines that have been reported to have transfection efficiencies high enough to be used in this procedure, thus limiting its application.

Once the approach for performing an RNAi screen has been chosen and optimized, developing a phenotypic assay is the next step. There are an almost unlimited number of strategies to identify the phenotype of interest. Initial genome-wide screens have examined a simple phenotype: cell proliferation versus cell death. For instance, Aza-Blanc *et al*. (2003) described the application of an RNAi-based genetic screen in mammalian cells toward understanding the biology and mechanism of TRAIL-induced apoptosis. TRAIL is a TNF superfamily member that induces cytotoxicity in tumor cells when bound to its cognate receptors. Binding of TRAIL to specific death receptors (DR4 and DR5) induces apoptosis through recruitment of adaptor molecules, which results in the formation of the death-inducing signaling complex and the activation of downstream apoptotic pathways. To identify genes that modify cellular sensitivity to TRAIL-induced death, the authors screened a limited complexity siRNA library using HeLa cells in the presence or absence of TRAIL. After an incubation period, cell viability was measured by addition of a dye (Alamar Blue) that produces a fluorescent signal, which is reflective of the extent of cellular proliferation. Their screen, validated by the identification of known apoptotic and anti-apoptotic genes, also led to a functional linkage of genes like *DOBI*, *MIRSA*, GSK3α or *SRP72* to the TRAIL-mediated response. Although this study was limited to several hundred genes, it illustrates the potential of a genome-wide RNAi analysis in mammals aimed at identifying genes involved in cell growth and viability. A similar study has been carried out in Drosophila cells by Boutros *et al*. (2004), who identified 438 essential genes by simple quantification of cell number after transfection of individual dsRNAs to each of Drosophila's predicted 20 000 genes. Analogous assays can be designed by measuring parameters such as caspase activation, ATP content, or cell membrane permeability. In all cases mentioned, the effect of the RNAi trigger must extend to the vast majority of cells, so that loss of viability in a subpopulation is not masked by unaffected cells. This limits the approach to cell lines that are very easily transfected under the chosen cell culture conditions. The great advantage of this assay is its simplicity. However, such approaches are not illuminating regarding precisely which biological pathway is affected to elicit cell death.

More informative assays require that more specific phenotypes be examined. We and others have explored the use of reporter systems in which the activity of a protein or cellular pathway is monitored by easily detectable changes in expression (or activity) of a transgene (luciferase, fluorescent proteins, CAT, β-gal, etc…). For example, Brummelkamp *et al*. (2003) used this strategy to study the family of ubiquitin-specific proteases (deubiquinating enzymes or DUBs). Post-translational modification by conjugation of ubiquitin moieties plays a major role in the control of protein half-life and thus in their activity. Ubiquitin conjugating enzymes and DUBs mediate ubiquitination and deubiquitination, respectively, of cellular substrates. These families maintain the correct balance between how much protein is driven to degradation and how much protein is preserved. The authors designed a collection of RNA interference vectors to suppress 50 human DUBs, and searched for those relevant to the NF-κB pathway. They cotransfected an NF-κB-luciferase reporter gene and different DUB knockdown vectors into human cells, and measured the effect of DUB knockdown on tumor necrosis factor-α- (TNF-α) mediated activation of NF-κB. Importantly, they found that RNAi targeting of the cylindromatosis tumor suppressor gene *CYLD* enhanced the activation of an NF-κB reporter. Similarly, we have used a reporter approach to detect changes in the activity of the multisubunit 26S proteosome, the major nonlysosomal protease in eukaryotic cells, by using, as a readout, a green fluorescent protein genetically modified to be a target for degradation by the proteasome (Paddison *et al*., 2004). We tested an shRNA library of approximately 7000 constructs for the ability to block proteasome-mediated proteolysis, as reflected by the accumulation of the modified fluorescent protein. Our study revealed approximately 100 RNAi constructs that increased the accumulation of the reporter out of which 22 corresponded to 15 known proteasome subunits. This screen was conducted by

cotransfection of the reporter and shRNA expression vector into cell plated in 96-well plates. However, we have obtained similar results using a screening protocol in which RNAi was triggered *in situ* by reverse transfection on microarrays (Silva *et al*., 2004).

Although reporter constructs are a very convenient approach for cell-based assays, appropriate reagents are not always easily available to assay interesting phenotypes. In some cases, conventional techniques such as immunofluorescence (IF), have presented a feasible alternative. In a proof of concept experiment, Hsieh *et al*. (2004) used this approach to identify inhibitors of the phosphatidylinositol-3-kinase(PI3K)/Akt signaling pathway. Members of the PI3K family are characterized by their ability to phosphorylate the inositol ring 3′-OH group in inositol phospholipids, generating the second messenger phosphatidylinositol-3,4,5-triphosphate (PI-3,4,5-P(3)). This compound in turn recruits Akt to the inner cell membrane, where the kinase becomes phosphorylated and activated. Activated Akt modulates the function of numerous substrates involved in the regulation of cell survival, cell cycle progression, and cellular growth. Hsieh *et al*. tested functionally a set of siRNAs in a screen aimed at identifying negative regulators of the Akt phosphorylation. In this screen, upon siRNA transfection, modulation of the Akt phosphorylation was detected by IF staining with anti-phospho-Akt (detecting Akt phosphorylation at S473). As expected, the known Akt regulator PTEN scored positive in the screen, validating this approach for cell-based RNAi studies.

Some biological questions cannot be answered by examining changes in the intensity of a reporter or monitoring site-specific phosphorylation of a molecule. Instead, scoring a visible phenotype might be required to assay protein function. Such studies may be well suited for automated microscopy in which high-content images are automatically analysed by sophisticated image processing software to determine phenotype. The study of genes involved in different aspects of cytokinesis presents a clear example of such a process. As a proof-of-principle experiment, we knocked down the mitotic motor protein Eg5, as cytokinesis defects in cells where Eg5 function is inhibited are well established (Silva *et al*., 2004). Transfection mixtures contained a plasmid encoding an α-tubulin GFP fusion protein and individual shRNAs targeting Eg5. In this experiment, the GFP fusion protein identifies the cells that have been transfected, and also allows visualization of microtubules. Microscopic analysis of the transfected cells revealed a 'rosette' pattern characteristic of the cells displaying loss of kinesin Eg5. Similar results were obtained by IF staining with anti-α-tubulin antibodies. Using an analogous approach, Somma *et al*. (2002) present a nice example of molecular dissection of the cytokinesis pathway in *Drosophila* cells. Their phenotypic analysis identified genes required for different aspects of cytokinesis, such as central spindle formation, actin accumulation at the cell equator, contractile ring assembly or disassembly, and membrane behavior.

The aforementioned examples present a few of the limitless possibilities that can be explored by combining RNAi with assay systems that are limited only by the imagination of the investigator. However, many phenotypes require significant time to develop or assay or can only be probed by reintroduction of cells into a tissue setting.

Genetic screens and selections with stable populations

The need for examining phenotypes that develop over a time span that reaches from several days to several weeks has been addressed by the development of methods for stably integrating shRNA expression cassettes into the genome of target cells. In this mode, the shRNA expression library is most commonly packaged into retroviruses, transduced into cells, and stable integrants selected 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.

Once a population of cells that stably expresses shRNAs is produced, two alternative approaches can be undertaken for assaying the consequences of gene knockdown. Cells can be plated at low density to ascertain their phenotypic behavior through a positive selection. In this mode, only cells that have a specific characteristic such as the ability to proliferate under specific conditions (e.g. colony formation in soft agar, focus formation, insensitivity to growth inhibitory cytokines, etc.) will be selected. Following selection, individual colonies can be isolated and the identities of integrated shRNA cassettes determined by sequencing. In an alternative approach, the fate of shRNAs in a population of virally transduced cells can be monitored by adopting a DNA-barcoding strategy, which has been previously used in *S. cerevisiae* for following complex populations of mutants by DNA microarrays (Shoemaker *et al*., 1996).

The former approach has been well validated using both integrated and episomal cDNA libraries that cause ectopic expression of certain genes and with antisense RNA libraries that can inhibit gene expression. With shRNA libraries, a genetic selection was used by Berns and colleagues to investigate components of the p53 growth arrest pathway. These investigators used conditionally immortalized primary human fibroblasts that undergo a senescence program upon reassertion of the p53 and Rb tumor suppressor pathways. These cells were screened using a library of 23 703 shRNAs targeting 7914 different human genes (Berns *et al*., 2004) for constructs that allowed continued proliferation in face of reactivation of tumor suppressor function. Using colony formation assays, and screening 83 different populations of transduced cells, they identified six genes, all in the p53 pathway and including *p53* itself, which when suppressed, conferred resistance to both *p53*-dependent and $p19^{ARF}$-dependent prolifera-

tion arrest and also abolished DNA damage-induced G1 arrest by ionizing radiation.

Aside from successful applications in yeast and phage, the barcode strategy has not been formally validated as a mechanism to identify genes linked to specific phenotypes in mammalian cells. However, the potential power of this strategy has made it a tantalizing and irresistible possibility for applying complex RNAi libraries. There are several approaches to the use of molecular barcodes for tracking complex populations of RNAi triggers, and two have been exemplified in recent publications describing shRNA libraries. One possibility is to use the shRNA sequence itself as a barcode. Another is to link the shRNA sequence to an independent unique sequence within each vector. Irrespective of the barcoding strategy, the underlying concept is that by PCR amplification of integrated DNAs, one may essentially count (relatively speaking) the number of cells that contain a specific shRNA cassette. This is measured by hybridizing genomic PCR products containing the barcodes to custom microarrays that contain the complement of these sequences. By comparing barcode representations from cell populations treated in different ways, one may simultaneously assess the consequences of expressing a given individual shRNA on the cell's response to the treatment (Figure 3). If, for example, a particular shRNA provided resistance to a growth inhibitor stimulus, then the representation of its associated barcode should be increased after treatment. If a given shRNA sensitized a population to a specific stress, then the relative abundance if its barcode should diminish after the stress.

Despite the enormous promise of this approach, there are certain considerations that must be taken into

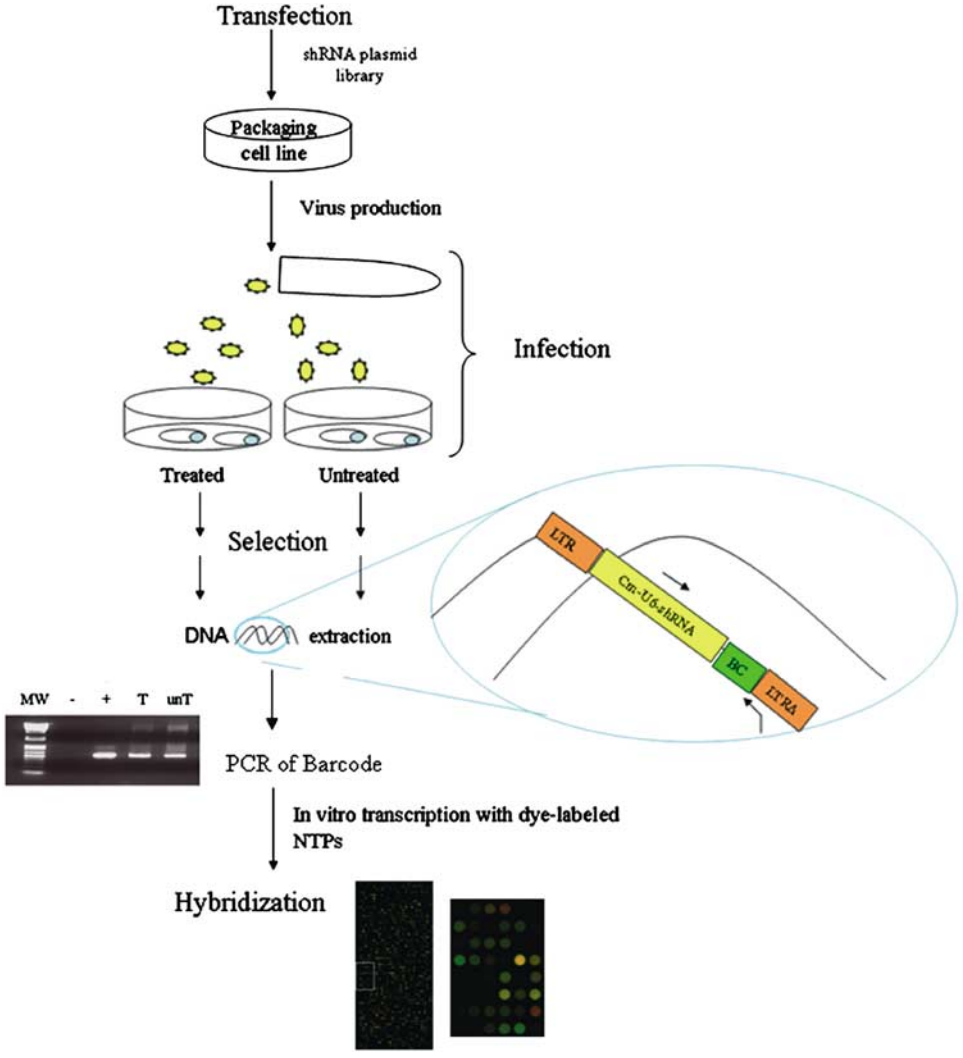

Figure 3 Sample protocol that illustrates a typical RNAi screen in a stably selected population. ShRNAs libraries are packaged into viruses, which are used to infect a target cell population. The populations are then divided into treatment groups, which are subjected to selection according to the experimental setting. After the desired time, DNA is extracted from each sample group and assayed for the presence of DNA barcodes by PCR. One of the PCR primers harbors a T7 RNA polymerase promoter sequence that allows for *in vitro* transcription of single-stranded RNA that is subsequently hybridized to an array containing the complement of the barcode sequences in the entire shRNA library. Comparison between the hybridization patterns of different treatment groups allows for the identification of shRNAs represented in each selected group

account when designing a barcode-based RNAi screen. The first is the potential noise associated with simultaneous analysis of complex populations by any given procedure. The second is the intrinsic noise produced by the chosen biological system.

Noise in the technical procedure used to measure barcode dynamics is easily defined and readily controlled. First, the PCR procedure used to amplify barcodes from the genome of infected populations must accurately reflect the relative proportions of individual members of the population. This is aided by the use of common primers for the amplification of each barcode sequence. Additionally, amplification should be kept in the linear range, which can be easily measured using a modification of a Q-PCR protocol. Remaining variables are applicable to any microarray procedures and include variation in labeling and hybridization. These are discussed in any of a number of published guides to microarray methodologies (for example, Churchill, 2002; Chuaqui *et al*., 2002; Forster *et al*., 2003). One source of potential problems, though not necessarily of noise, is the relative performance of individual probes on the arrays. At present, irrespective of the particular barcode strategy, none of the sequences being used for barcoding is specifically designed as a hybridization probe. While future advances may change this fact, it must be accepted that at present, some probes will simply perform poorly, giving weak signals irrespective of their frequency in the population. Thus far, we have demonstrated the validity of the technical approach to barcode amplification from the genomes of infected cells using populations with complexities reaching 20 000 individual elements.

Noise can also be introduced into the system from the biological components of the assay. Of principal concern in an outgrowth experiment is whether the genetic drift (Glass, 1954) of the population introduces random variation into the relative frequency of each shRNA clone irrespective of its specific biological function. Simulation studies suggest that such drift is most pronounced as complexity rises and population size falls. This leads to each member of the population being represented in relatively few cells. As the population proliferates and is stressed by propagation in culture, certain elements of the population may be lost. For example, most cell lines, split by trypsinization, do not have a 100% plating efficiency. Thus each passage creates a random sample of the total population. If each shRNA is present in only a few cells, there is a relatively high probability that some shRNAs will be lost at each sampling. A second problem is that cell lines are not completely uniform at the level of the individual cell. Thus, each cell within the cell line may have a different probability of doubling for each doubling of the population as a whole. This can also contribute to the loss or amplification of shRNAs over time. We find that both of these problems can be solved to a large degree by insuring that each shRNA in the population is represented in a large number of cells, as is predicted by population genetics simulations (Sachidanamdan *et al*., unpublished).

Given solutions of the aforementioned technical barriers, there are a number of ways in which barcode microarrays can be applied to RNAi-based screens. The simplest is to use barcode arrays rather than analysis of individual colonies to examine populations that have been positively selected following application of a stress. More complicated and fraught with difficulty is the use of barcodes to identify shRNAs that sensitize cells to a treatment or genetic lesion in a so-called synthetic lethal experiment.

Synthetic-lethal screens

A synthetic lethal interaction is defined by a situation in which two nonessential genes become essential when mutated in combination in the same cell (Basson *et al*., 1987; Bender and Pringle, 1991). Such an interaction implies that two genes lie in discrete pathways that normally compensate, each for the loss of the other. When both pathways are lost, death occurs because of a catastrophic loss of a specific biological process. In principle, a similar genetic interaction could also occur when multiple hypomorphic mutations in the same pathway reduce flux through that pathway to levels that are not tolerated by a cell. Carrying out synthetic lethal screens in genetically tractable models such as yeast or *C. elegans* is relatively straightforward. However, the development of this technology in mammalian cells is much less straightforward. In theory, genome-wide RNAi libraries offer a route toward this powerful method for detecting genetic interactions.

The ability to detect cell extinction events on a large-scale is key to realizing the use of RNAi in genome-wide screens for synthetic lethal interactions. There are several ways to accomplish this goal, including the use of highly parallel screens in which the effects of individual constructs are examined in multiwell dishes. In that mode, the successful identification of synthetic lethal interactions depends heavily on a combination of the timing of the event and achieving the appropriate cell plating density to enable measurement of growth inhibition. A more convenient way to approach the problem would be through the use of the DNA barcode strategy to follow loss of cell populations expressing certain shRNAs. In theory, the underlying methodology is straightforward. Cells of two discrete genotypes could be engineered with an shRNA library and then the two populations could be compared following some period of outgrowth. A slightly more sophisticated version would use a single engineered cell population carrying a conditional allele of the target gene. This could be generated either by conventional means or through the introduction into part of the population of a target-specific siRNA. Confounding such approaches is the incomplete expressivity of shRNA-mediated silencing. Even an effective shRNA will not silence its target to the same degree in all cells harboring the shRNA expression cassette. This is due both to position effects on the integrated cassettes and to the propensity to select for cells that have silenced or altered expression cassettes that encode deleterious genes. These problems effectively set a

background that might mitigate against detection of lethal interactions because of an insufficient power in the system to detect relatively minor shifts in populations. As an example, imagine that 70% of cells infected with a given shRNA virus show a synthetic lethality in combination with p53 loss. This would translate into the ability to see only an ~3-fold change in the representation of that shRNA on a barcode microarray, which is probably at the edge of what can be confidently detected. Therefore, our challenge in moving toward the goal of genome-wide synthetic lethal analysis is to increase both the quality of our RNAi reagents and the mechanisms to detect loss of such cell populations.

Conclusions

Over the past several years, we have made tremendous progress in our quest to harness the RNA interference pathway as a method for genetic manipulation of mammalian cells and animals. Last year, with the report of several efforts to construct genome-wide libraries of siRNAs and shRNAs for multiple mammalian models, we have entered a new era in which the application of powerful genetic approaches can be used both *in vivo* and *in vitro* to deepen our understanding of the cancer cell and to identify and validate potential new approaches to cancer therapy.

Acknowledgements
We thank JP O'Keefe for critical reading and suggestions. JS is supported by a postdoctoral fellowship from the US Army Prostate Cancer Research Program. FVR is a Jane Coffin Childs Memorial Cancer Research Fund fellow. GJH is supported by an Innovator Award from the US Army Breast Cancer Research Program and by grants from the National Institutes of Health.

References

Aravind L, Dixit VM and Koonin EV. (2001). *Science*, **291,** 1279–1284.
Aza-Blanc P, Cooper CL, Wagner K, Batalov S, Deveraux QL and Cooke MP. (2003). *Mol. Cell*, **12,** 627–637.
Basson ME, Moore RL, O'Rear J and Rine J. (1987). *Genetics*, **117,** 645–655.
Bender A and Pringle JR. (1991). *Mol. Cell Biol.*, **11,** 1295–1305.
Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL and Bernards R. (2004). *Nature*, **428,** 431–437.
Boutros M, Kiger AA, Armknecht S, Kerr K, Hild M, Koch B, Haas SA, Consortium HF, Paro R and Perrimon N. (2004). *Science*, **303,** 832–835.
Brummelkamp TR, Nijman SM, Dirac AM and Bernards R. (2003). *Nature*, **424,** 797–801.
Celera Genomics Project (2001). *Science*, **291,** 1304–1351.
Chen Z and Han M. (2000). *Bioessays*, **22,** 503–506.
Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangrea MA, Ahram M, Linehan WM, Knezevic V and Emmert-Buck MR. (2002). *Nat. Genet.*, **32** (Suppl)**,** 509–514.
Churchill GA. (2002). *Nat. Genet.*, **32** (Suppl)**,** 490–495.
Clemens JC, Worby CA, Simonson-Leff N, Muda M, Maehama T, Hemmings BA and Dixon JE. (2000). *Proc. Natl. Acad. Sci.*, **97,** 6499–6503.
Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisrame A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, Suleau A, Swennen D, Tekaia F, Wesolowski-Louvel M, Westhof E, Wirth B, Zeniou-Meyer M, Zivanovic I, Bolotin-Fukuhara M, Thierry A, Bouchier C, Caudron B, Scarpelli C, Gaillardin C, Weissenbach J, Wincker P and Souciet JL. (2004). *Nature*, **430,** 35–44.
Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K and Tuschl T. (2001). *Nature*, **411,** 494–498.
Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE and Mello CC. (1998). *Nature*, **391,** 806–811.
Forsburg SL. (2001). *Nat. Rev. Genet.*, **2,** 659–668.
Forster T, Roy D and Ghazal P. (2003). *J. Endocrinol.*, **178,** 195–204.
Glass B. (1954). *Adv. Genet.*, **6,** 95–139.
Hannon GJ. (2002). *Nature*, **418,** 244–251.
Harborth J, Elbashir SM, Bechert K, Tuschl T and Weber K. (2001). *J. Cell Sci.*, **114,** 4557–4565.
Hsieh AC, Bo R, Manola J, Vazquez F, Bare O, Khvorova A, Scaringe S and Sellers WR. (2004). *Nucleic Acids Res.*, **32,** 893–901.
Huang Q, Raya A, DeJesus P, Chao SH, Quon KC, Caldwell JS, Chanda SK, Izpisua-Belmonte JC and Schultz PG. (2004). *Proc. Natl. Acad. Sci.*, **101,** 3456–3461.
International Human Genome Sequencing Consortium (2001). *Nature*, **409,** 860–921.
Jorgensen EM and Mango SE. (2002). *Nat. Rev. Genet.*, **3,** 356–369.
Kumar R, Conklin DS and Mittal V. (2003). *Genome Res.*, **13,** 2333–2340.
Manche L, Green SR, Schmedt C and Mathews MB. (1992). *Mol. Cell Biol.*, **12,** 5238–5248.
Michiels F, van Es H, van Rompaey L, Merchiers P, Francken B, Pittois K, van der Schueren J, Brys R, Vandersmissen J, Beirinckx F, Herman S, Dokic K, Klaassen H, Narinx E, Hagers A, Laenen W, Piest I, Pavliska H, Rombout Y, Langemeijer E, Ma L, Schipper C, Raeymaeker MD, Schweicher S, Jans M, van Beeck K, Tsang IR, van de Stolpe O, Tomme P, Arts GJ and Donker J. (2002). *Nat. Biotechnol.*, **20,** 1154–1157.
Mikkers H and Berns A. (2003). *Adv. Cancer Res.*, **88,** 53–99.
Minks MA, West DK, Benvin S and Baglioni C. (1979). *J. Biol. Chem.*, **254,** 10180–10183.
Montgomery MK. (2004). *Methods Mol. Biol.*, **265,** 3–21.
Mousses S, Caplen NJ, Cornelison R, Weaver D, Basik M, Hautaniemi S, Elkahloun AG, Lotufo RA, Choudary A, Dougherty ER, Suh E and Kallioniemi O. (2003). *Genome Res.*, **13,** 2341–2347.
Paddison PJ and Hannon GJ. (2002). *Cancer Cell*, **2,** 17–23 . RNA interference: the new somatic cell genetics?.

Paddison PJ and Hannon GJ. (2003). *Curr. Opin. Mol. Ther.*, **5,** 217–224.
Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, Balija V, O'Shaughnessy A, Gnoj L, Scobie K, Chang K, Westbrook T, Cleary M, Sachidanandam R, McCombie WR, Elledge SJ and Hannon GJ. (2004). *Nature*, **428,** 427–431.
Shoemaker DD, Lashkari DA, Morris D, Mittmann M and Davis RW. (1996). *Nat. Genet.*, **14,** 450–456.
Shuman HA and Silhavy TJ. (2003). *Nat. Rev. Genet.*, **4,** 419–431.
Silva JM, Mizuno H, Brady A, Lucito R and Hannon GJ. (2004). *Proc. Natl. Acad. Sci.*, **101,** 6548–6552.
Somma MP, Fasulo B, Cenci G, Cundari E and Gatti M. (2002). *Mol. Biol. Cell*, **13,** 2448–2460.
St Johnston D. (2002). *Nat. Rev. Genet.*, **3,** 176–188.
Tabara H, Grishok A and Mello CC. (1998). *Science*, **282,** 430–431.
Timmons L and Fire A. (1998). *Nature*, **395,** 854.
Ziauddin J and Sabatini DM. (2001). *Nature*, **411,** 107–110.