AD_____

Award Number: W81XWH-041-0197

TITLE: Genomic Approaches for Detection and Treatment of Breast Cancer

PRINCIPAL INVESTIGATOR: Stephen J. Elledge, Ph.D.

CONTRACTING ORGANIZATION: Brigham & Women's Hospital
Boston, MA 02115

REPORT DATE: July 2007

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
31-07-2007	Annual	1 JUL 2006 - 30 JUN 2007

4. TITLE AND SUBTITLE
Genomic Approaches for Detection and Treatment of Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-041-0197

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Stephen J. Elledge, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

Email:

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

Brigham & Women's Hospital
Boston, MA 02115

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

None provided.

15. SUBJECT TERMS
None provided.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			
U	U	U	UU	20	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Introduction

A key part of our research plan has been the development and use of retroviral vectors expressing RNA interference RNAs to identify human genes involved in causing or restraining cancer. In our first progress reports we described our efforts to develop shRNA libraries and showed they could be used to identify tumor suppressors. Ultimately our goal is to screen of complex pools of shRNA expressing retroviruses each marked with a bar code that allows the results of the screen to be read out by microarray hybridization. We demonstrated this could be accomplished in enrichment screens for shRNAs that caused cellular transformation and growth in soft agar. However, a key goal has been to identify shRNAs that debilitate or kill cancer cells. In order for this to be possible in complex pools, it is imperative that each vector knock down its target with high penetrance. We have successfully achieved this level of knockdown and can now see particular shRNA expressing viruses drop out of complex pools.

Body

Identification of cancer-specific lethal genes.

Retroviral shRNA-mediated genetic screens in mammalian cells are powerful tools for discovering loss-of-function phenotypes. We have been working on the generation of shRNA libraries for the express purpose of performing screens to kill cancer cell. Here we describe a highly parallel multiplex methodology for screening large pools of shRNAs using half-hairpin barcodes for microarray deconvolution. We carried out dropout screens for shRNAs that affect cell proliferation and viability in cancer and normal cells. We identified many shRNAs to be anti-proliferative that target core cellular processes such as the cell cycle and protein translation in all cells examined. More importantly, we identified genes that are selectively required for proliferation and survival in different cell lines. Our platform enables rapid and cost-effective genome-wide screens to identify cancer proliferation and survival genes for target discovery. Such efforts are complementary to the Cancer Genome Atlas and provide an alternative functional view of cancer cells.

In the last progress report, we described the generation of barcoded, microRNA-based shRNA libraries targeting the entire human genome that can be expressed efficiently from retroviral or lentiviral vectors in a variety of cell types for stable gene knockdown (*1, 2*). Furthermore, we also described a method of screening complex pools of shRNAs using barcodes coupled with microarray deconvolution to take advantage of the highly parallel format, low cost, and flexibility in assay design of this approach (*2, 3*). Although barcodes are not essential for enrichment screens (positive selection) (*3-5*), they are critical for dropout screens (negative selection) such as those designed to identify cell lethal or drug sensitive shRNAs which we have proposed to do for this Innovator award (*6*). Hairpins that are depleted over time can be identified through the competitive hybridization of barcodes derived from the shRNA population before and after selection to a microarray (Fig. 1A).

We previously described the use of 60-mer barcodes for pool deconvolution (*2, 3*). To provide an alternative to these bar-codes that enables a more rapid construction and

screening of shRNA libraries, we have developed a methodology called half-hairpin (HH) barcoding for deconvoluting pooled shRNAs (*7*). We took advantage of the large 19-nt hairpin loop of our mir30-based platform and designed a PCR strategy that amplifies only the 3'-half of the shRNA stem (Fig. 1B). Compared to using full hairpin sequences for microarray hybridization (*8, 9*), HH barcodes entirely eliminate probe self-
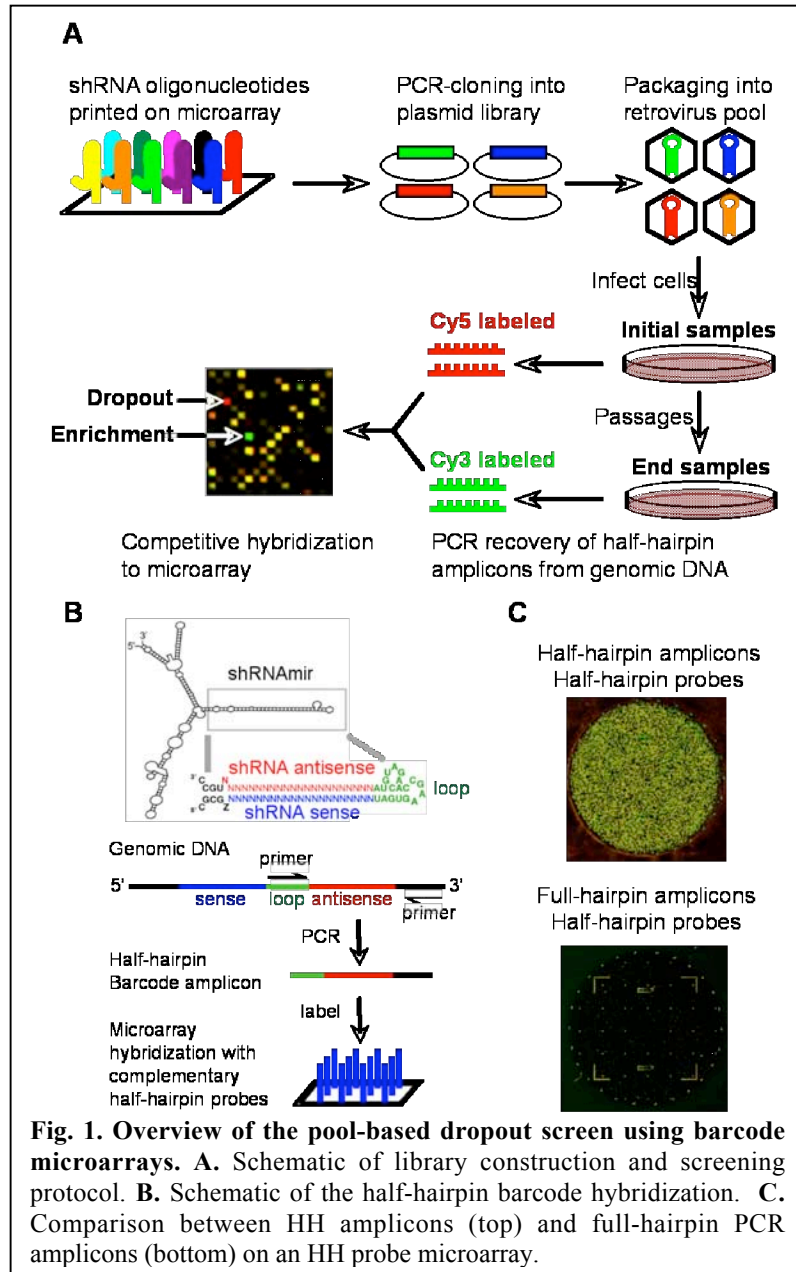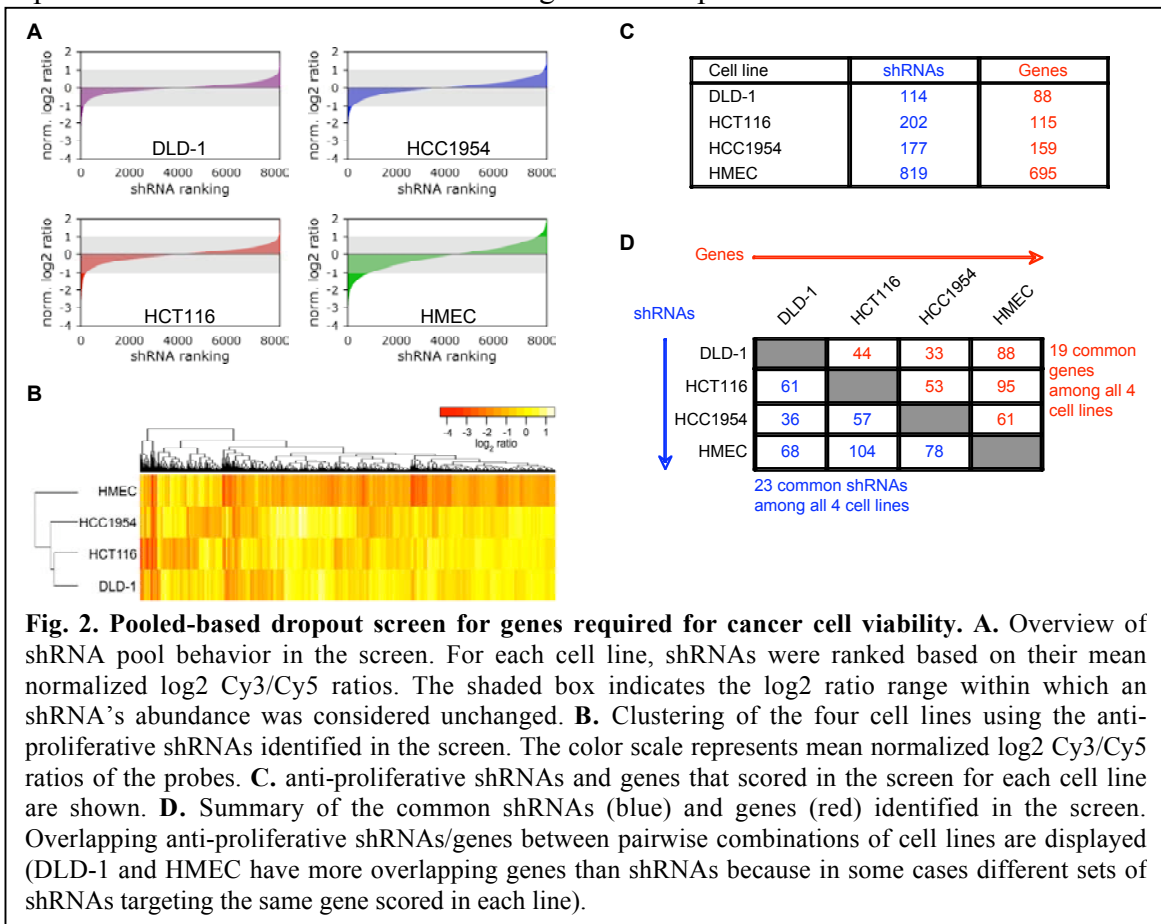

Fig. 1. Overview of the pool-based dropout screen using barcode microarrays. A. Schematic of library construction and screening protocol. **B.** Schematic of the half-hairpin barcode hybridization. **C.** Comparison between HH amplicons (top) and full-hairpin PCR amplicons (bottom) on an HH probe microarray.

annealing, providing the dynamic range necessary for pool-based dropout screens. HH barcode signals are highly reproducible in replicate PCRs (*R*=0.973), highly specific (0.5% cross reaction), and display reasonable dynamic range in mixing experiments where sub-pool inputs are varied in competing hybridization experiments. Taken together, these results indicate HH barcodes are alternatives to the 60-mer barcodes originally designed into our library. Our central goal is to develop the means to rapidly perform dropout screens to systematically identify genes required for cancer cell proliferation and survival that could represent new drug targets. We used our screening platform to interrogate human DLD-1 and HCT116 colon cancer cells, human HCC1954 breast cancer

cells and normal human mammary epithelial cells (HMECs). We compared colon and breast cancer cells, two types of cancers with distinct origins, to maximize our ability to identify common and cancer-specific growth regulatory pathways. Recent large-scale efforts have identified a distinct spectrum of mutations in these two cancer types (*10, 11*). Also, the comparison between cancer and normal mammary epithelial cells should reveal potential growth and survival adaptations specific to cancer cells. We constructed a

highly complex pool of 8203 distinct shRNAs targeting 2924 genes consisting of annotated kinases, phosphatases, ubiquitination pathway and cancer-related genes. We chose these genes because they are central regulators of signaling pathways that should provide a rich source of phenotypic perturbation. These shRNAs were placed in an MSCV- retroviral vector (*12*), MSCV-PM, that functions efficiently at single copy.

We screened each cell line in independent triplicates (*7*). Cells were infected with an average representation of 1000 per shRNA and a multiplicity of infection (MOI) of 1-2. Initial reference samples were collected 48–72 hours post-infection. The remaining cells were puromycin-selected, propagated for several weeks and collected again as the end samples. HH barcodes were PCR-recovered from genomic DNA, labeled with Cy5 and Cy3 dyes, respectively, and hybridized to a HH barcode microarray (Fig. 1A). The Cy3/Cy5 signal ratio of each probe reports the change in relative abundance of a particular shRNA between the beginning and the end of the experiment. Both correlations between initial samples across the triplicates and between the initial and end samples within each replica were high, indicating the triplicates were highly reproducible and representation was well-maintained throughout the experiment.

Fig. 2. Pooled-based dropout screen for genes required for cancer cell viability. A. Overview of shRNA pool behavior in the screen. For each cell line, shRNAs were ranked based on their mean normalized log2 Cy3/Cy5 ratios. The shaded box indicates the log2 ratio range within which an shRNA's abundance was considered unchanged. **B.** Clustering of the four cell lines using the anti-proliferative shRNAs identified in the screen. The color scale represents mean normalized log2 Cy3/Cy5 ratios of the probes. **C.** anti-proliferative shRNAs and genes that scored in the screen for each cell line are shown. **D.** Summary of the common shRNAs (blue) and genes (red) identified in the screen. Overlapping anti-proliferative shRNAs/genes between pairwise combinations of cell lines are displayed (DLD-1 and HMEC have more overlapping genes than shRNAs because in some cases different sets of shRNAs targeting the same gene scored in each line).

To identify shRNAs that consistently changed abundance in each cell line, datasets were analyzed using a custom statistical package based on the LIMMA method (*13*) for 2-color cDNA microarray analysis (*7*). Whereas the majority of shRNAs show little changes in their abundance over time (log2 ratio between –1 and 1), a small fraction of

shRNAs showed depletion (Fig. 2A). Based on their shRNA dropout signatures, unsupervised hierarchical clustering segregated the 3 cancer cell lines from the normal
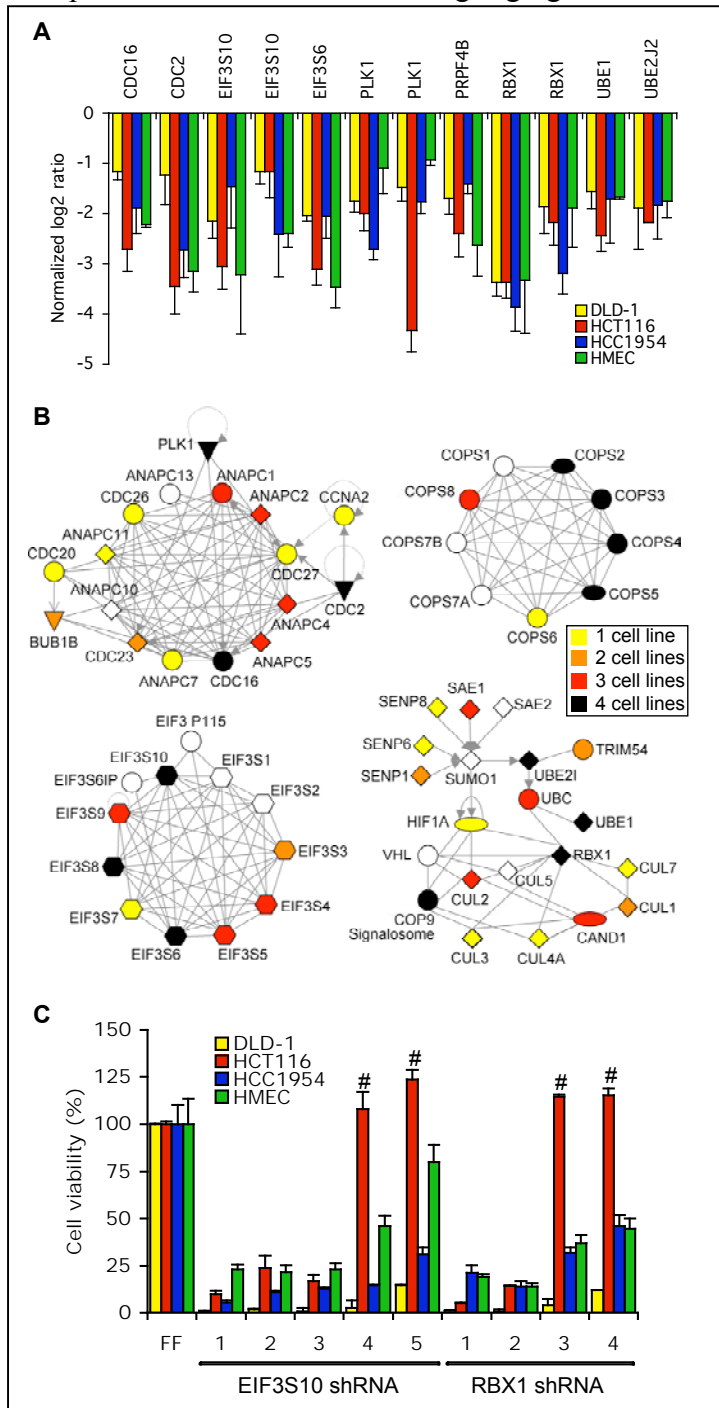

Fig. 3. Genes commonly required for proliferation of cancer cells. A. Representative candidate shRNAs that reduce viability of all four cell lines. For each shRNA, the normalized log2 Cy3/Cy5 ratio (i.e. its relative abundance in the end sample vs. initial sample) is given. B. Core cellular modules required in all cell lines. Shown are the APC the COP9 signalosome, the eIF3 translation initiation complex and E3 ubiquitin ligases Color indicates genes required for viability of 1 (yellow), 2 (orange) 3 (red), or 4 (black) lines. C. Validation of shRNAs from the screen that reduce viability of four lines. Candidate Cell viability was measured at day 9 post-infection. All viability reductions were significant ($p < 0.05$), except where indicated by #. FF, negative control shRNA targeting firefly luciferase.

HMECs, likely reflecting fundamental differences between cancer and normal cells (Fig. 2B). Furthermore, the two colon cancer cell lines were more similar to each other than the breast cancer line, reflecting the differences in their tissues of origin and paths to tumorigenesis. Overall we found 114 shRNAs (1.4%) representing 88 genes (3.0%) in DLD-1 cells, 202 shRNAs (2.5%) representing 115 genes (3.9%) in HCT116 cells, 177 shRNAs (2.2%) representing 159 genes (5.4%) in HCC1954 cells, and 819 shRNAs (10.0%) representing 695 genes (23.8%) in HMEC cells showed statistically significant depletion (Fig. 2C). The lists of anti-proliferative shRNAs show significant overlap ($p < 10^{-40}$), with 23 shRNAs and 19 genes scoring in all four lines (Fig. 2D). As expected, our screen recovered components of core cellular modules essential for all cell lines (Fig. 3A and 3B). For example, shRNAs against multiple subunits of the anaphase promoting complex (APC/C) (DLD-1

$p = 9.65 \times 10^{-5}$; HCT116 $p = 2.99 \times 10^{-9}$; HCC1954 $p = 1.41 \times 10^{-5}$; HMEC $p = 5.80 \times 10^{-6}$), the COP9 signalosome (DLD-1 $p = 2.48 \times 10^{-6}$; HCT116 $p = 9.34 \times 10^{-6}$; HCC1954 $p = 4.54 \times 10^{-5}$; HMEC $p = 0.032$), and the eukaryotic translation initiation factor 3 (eIF3) complex (DLD-1 $p = 1.42 \times 10^{-5}$; HCT116 $p = 7.98 \times 10^{-8}$; HCC1954 $p = 0.00024$; HMEC $p = 0.0086$) were identified (Fig. 3B). A number of key proteins in the ubiquitination and sumoylation pathways, including most of the cullins, were also identified. Importantly, multiple shRNAs against the same gene scored, suggesting their effects are specific.

We next validated *EIF3S10* and *RBX1*, two genes essential for viability in all 4 cell lines. For each gene, we included shRNAs that scored in the screen as well as additional shRNA sequences present in our library. Cells were infected with individual retroviral shRNAs and cell viabilities were assessed (Fig. 3C). For each gene, all of the shRNAs that scored in the screen and many additional shRNAs gave anti-proliferative phenotypes. Furthermore, the anti-proliferative activity of the shRNAs correlated very well with the extent of target gene knockdown, as shown for *RBX1*. Thus, these phenotypes are due to target gene knockdown rather than off-targets. This is consistent with a previous screen with this library showing ~90% "on target" efficiency (*14*).

In addition to the common set of shRNAs that impair viability in all cell lines, we observed a substantial number of genes that are selectively required for proliferation of each cell line. These are particularly interesting as they may reflect differences in the underlying oncogenic context, and therefore represent potential cancer-selective drug targets. We validated the gene *PPP1R12A*, which encodes a regulatory subunit of protein phosphatase 1, for its selective requirement in HCC1954 but not DLD-1 cells (Fig. 4A). The PPP1R12A shRNA that gave the greatest depletion (shRNA 3) showed the strongest effect on HCC1954 cells but only marginally affected DLD-1 viability (Fig. 4B). This was corroborated with 4 additional PPP1R12A siRNAs. Importantly, these shRNAs and siRNAs resulted in comparable knockdown of PPP1R12A protein in both cell lines, indicating that the selective requirement for *PPP1R12A* by HCC1954 cells is not due to different degrees of protein knockdown. PPP1R12A has been shown to target protein phosphatase 1 (PP1) isoforms to several substrates including myosin and merlin (*15, 16*). Thus, PP1 activity reduction by PPP1R12A knockdown may lead to increased phosphorylation of key proteins that disrupt the viability of HCC1954 cells. Conversely, *PRPS2*, which encodes phosphoribosyl pyrophosphate synthetase 2, an enzyme involved in nucleoside metabolism, is more selectively required by DLD-1 than HCC1954 cells (Fig. 4C). These results suggest distinct, genetic context-dependent vulnerabilities exist between these tumor cell lines.

Comparison between HCC1954 cells and normal HMECs also revealed a distinct subset of genes selectively required by each cell line. Not surprisingly, a much larger set of 695 genes are required by HMECs, likely reflecting the ability of normal cells to appropriately respond to various cellular stresses. Conversely, the relatively fewer genes required by the cancer cells underscores their ability to evade and overcome growth-inhibitory cues. Among the genes identified as essential for HMECs and HCT116 cells, but not DLD-1 or HCC1954 cells, is *HDM2* encoding the human homolog of MDM2, the E3 ligase for p53 (Fig. 4D). HCC1954 and DLD-1 cells harbor inactivating mutations (Y163C and S241F, respectively) in the *TP53* gene and are therefore insensitive to MDM2 knockdown. Multiple MDM2 shRNAs selectively impaired the viability of the p53 wildtype HMECs, but not HCC1954 cells with mutant p53
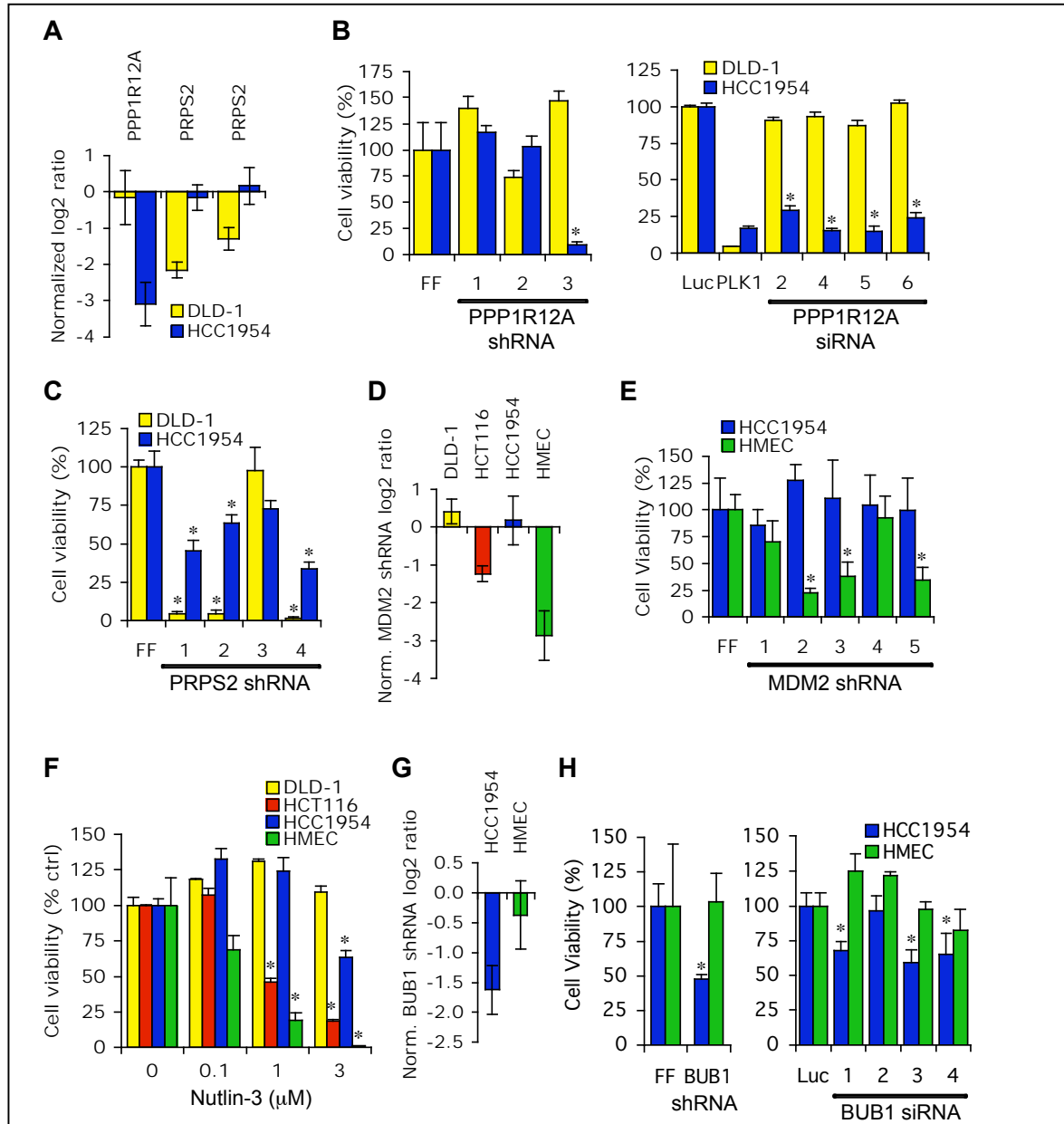
Fig. 4. Genes selectively required for proliferation or survival of cancer cells.
A. Identification of *PPP1R12A* (1 shRNA) and *PRPS2* (2 shRNAs) as two genes that are selectively required by HCC1954 or DLD-1 cells, respectively, in the screen. All error bars represent standard deviations across triplicates**. B. and C.** Validation of *PPP1R12A* (**B**) and *PRPS2* (**C**) as selectively required for viability of HCC1954 or DLD-1 cells, respectively. Cells were either infected with individual retroviral shRNAs or transfected with individual siRNAs in triplicates. Cell viability was measured at 9 days post-infection (shRNA) or 4 days post-transfection (siRNA). (* $p < 0.05$). FF, negative control shRNA targeting firefly luciferase. Luc, negative control siRNA targeting luciferase. PLK1, positive control siRNA targeting polo-like kinase 1.**D.** Normalized log2 ratios of an MDM2 shRNA in the screen. **E.** shRNA knockdown of MDM2 selectively impairs the viability of HMECs. Cell viability was measured 9 days post-infection with retroviruses expressing 5 different MDM2 shRNAs (* $p < 0.05$).**F.** Differential sensitivity of the 4 cell lines to the MDM2 inhibitor nutlin-3. Cell viability is reflective of their p53 status (HMECs and HCT116 cells, p53 wildtype; HCC1954 and DLD-1 cells, p53 mutant). Cell viability was measured after 4 days of nutlin-3 treatment (* $p < 0.05$).

p53 (Fig. 4E). Furthermore, we were able to pharmacologically validate this finding by interfering with MDM2 function using the inhibitor nutlin-3 (*17*) and recapitulating the sensitivity of these cells to MDM2 inactivation (Fig. 4F).

Importantly, a number of genes appear to be selectively required by HCC1954 cells but not HMECs. Among these is the cell cycle regulator and spindle checkpoint kinase *BUB1* (Fig. 4G). We validated *BUB1* using both shRNA and siRNAs to confirm its knockdown is more detrimental to HCC1954 cells than HMECs (Fig. 4H), despite similar levels of BUB1 protein reduction. These results indicate that BUB1 is likely to play an integral role in supporting the oncogenic transformation of HCC1954 cells as they are more dependent on BUB1 function. One possible explanation for this enhanced dependency may be the near-tetraploid nature of the HCC1954 genome. Compared to the diploid HMECs, HCC9154 cells may rely more heavily on the spindle checkpoint to maintain genomic stability. Such a dependency is an example of "non-oncogene addiction" where cancer cells come to be highly dependent for growth and survival on the functions of genes that are themselves not oncogenes (*18*).

Our study demonstrates that highly parallel dropout screens using complex pools of shRNAs can be achieved using HH barcodes in combination with highly-penetrant vectors. Our ability to identify anti-proliferative shRNAs specific to particular cell lines indicates that different cancer cells have distinct growth and survival requirements that cluster with cancer type. Targeting such key vulnerabilities is an attractive approach for cancer-selective therapeutics. The functional genetic approach demonstrated here presents an alternative and complementary effort to sequencing-based approaches such as the Cancer Genome Atlas (TCGA) and similar efforts, which focus on physical alterations of the cancer genome.

The most complex pool we employed contains 42,000 distinct shRNAs, an 80-fold increase in complexity compared to previous dropout screens based on our designs. It is now conceivable to screen the entire human genome with ~3 shRNAs per gene with a pool of ~100,000 shRNAs in ~100 million cells. Thus, a large number of cancer and normal cell lines can be rapidly screened in this manner, through what we hope will become a Genetic Cancer Genome Project, with the goal of generating "cancer lethality signatures" for different cancer types and thus identifying cancer type-specific lethal genes representing potential drug targets. We propose to now take this screen genome wide for additional breast cancer lines.

A shRNA synthetic lethal screen with the Ras oncogene

As described above, the past few years, we have developed an shRNA library targeting the entire human genome that incorporate several advances in mammalian RNAi technologies. Our shRNA library is delivered using either retroviral or lentiviral vectors and expressed in a microRNA context from a strong Pol II promoter to achieve strong, stable gene knockdown. Importantly, we have also developed barcode technologies that utilize either a dedicated 60-nt barcode or the anti-sense strand of the shRNA itself as barcode to track the abundance of individual shRNA in complex pools by microarray hybridization. This latter feature enables us to carry out massively parallel screens of tens of thousands of shRNAs simultaneously in complex pools. Previously we have demonstrated that we could carry out enrichment screens to identify novel tumor suppressors using our library. More recently, we have demonstrated that we can carry out depletion screens using our shRNA library and barcode technology to look for genes specifically required for cancer cell survival. This latter study is particularly exciting not just because depletion screens are technologically more difficult to achieve than enrichment screens, its greater implication is that we are now positioned to directly probe cancer cell vulnerabilities and identify potential drug targets for cancer therapeutics on a genome scale.

Building on these successes, we designed a synthetic lethal screen aiming to identify genes whose knockdown constitutes synthetic lethality with the Ras oncogene, the most frequently mutated oncogene in human cancers. While several well-characterized downstream effector pathways (such as the MAPK, PI3K and RalGDS pathways) have been implicated in mediating the effect of the Ras oncogene, the network of genes supporting the mutant Ras phenotype have not been systematically characterized. Clinically, Ras mutation is associated with poor prognosis with little options for targeted therapies. Farnesyl transferase inhibitors have met with only limited success whereas inhibitors targeting Ras effectors such as MEK and PI3K are yet to prove their efficacies. A major hurdle in the field has been the slow pace of discovering, through traditional genetic and biochemical means, genes that critically support the Ras oncogene in tumors and therefore could serve as new drug targets. Our approach should identify such genes in an unbiased fashion, and therefore overcome this bottleneck in target identification.

The detailed design of the Ras synthetic lethal (RSL) screen using our barcode technology is illustrated in Figure 5. Cells are infected with a pool of ~13,000 distinct retroviral shRNAs and an initial (reference) sample is taken at day-2 post infection before the knockdown phenotype has occurred. Cells are then propagated for an additional number of doublings before the end sample is taken. The shRNA barcodes are PCR-recovered from genomic DNA of the initial and end samples, labeled with Cy5 and Cy3 fluorescent dyes, respectively, and competitively hybridized to a custom microarray containing the corresponding probes. If an shRNA enhances cell proliferation, it will enrich in the pool over time and show a Cy3:Cy5 ratio >1. Conversely, shRNAs that are toxic will "drop out" in the pool over time and show a Cy3:Cy5 ratio <1. To discover genes that are synthetically lethal with Ras, two isogenic cell lines that differ only by the presence of a mutant Ras allele are screened in parallel with the same pool to identify shRNAs that selectively drop out in the Ras mutant cells but not the Ras wild type cells

(see below).

We chose to conduct the RSL screen in the DLD-1 cancer cell line for the following reasons. First, DLD-1 cells were derived from a tumor with an endogenous, activating mutation in K-Ras (K-RasG13D). Thus, unlike immortalized cell lines engineered to overexpress mutant Ras, all the cellular genetic context supporting Ras-driven tumorigenesis in vivo should be present in these cells. Second, the K-RasG13D allele in DLD-1 cells has been knocked out to derive Ras-wild type (Ras-WT) clones (Figure 6A). Importantly, phenotypic comparison of this Ras-mutant (Ras-Mut) and Ras-WT isogenic pair show that Ras-WT cells exhibit reduced MAPK signaling, decreased cell proliferation and abrogated anchorage-independent growth (not shown), indicating that Ras mutation is functionally important for the proliferation and transformation of DLD-1 cells. Third, our retroviral shRNA library can infect the DLD-1 isogenic pair with high efficiency and achieve strong gene knockdown at the level of single integrant (not shown).

So far, we have carried out a proof-of-principle screen using a library of 8011 shRNAs targeting annotated kinases, phosphatases, ubiquitin ligase/proteasome pathway genes and genes previously implicated in cancer. The DLD-1 isogenic cell lines were infected in parallel in independent triplicates and the relative abundance of each shRNA at the beginning (population doubling 0, or PD0) and the end of the experiment (PD16) were compared by barcode microarray hybridization (Figure 5). As expected, most shRNAs show little change in their relative representation over time (less than 2-fold in either direction, Figure 6B). A small number of shRNAs dropped out in both cell lines (Figure 6C). As expected, among these "straight lethal" shRNAs are those that target essential cellular machineries such as the eukaryotic translation initiation factor-3 complex and the anaphase promoting complex. Importantly, a number of candidate shRNAs show synthetic lethality with the Ras-Mut cells (Figure 6), thus demonstrating the feasibility of this approach.

To begin validating the candidate genes from the pilot screen, we have developed a relatively high-throughput competition assay in 96-well plate format (Figure 7A). Candidate RSL shRNAs are individually packaged as retroviruses and infected into a mixture of DLD-1 Ras-Mut and Ras-WT cells. The Ras-WT cells are labeled with green fluorescent protein (EGFP) so the two populations can be distinguished by FACS analysis. A RSL shRNA will selectively impair the viability of the Ras-Mut cells and therefore deplete the fraction of Ras-Mut cells in the mixture. From the pilot screen, we found 22 shRNAs against 15 genes to preferentially kill Ras-Mut cells over Ras-WT cells by the competition assay (not shown), thus providing a first-level validation for the screen. Interestingly, an shRNA targeting polo-like kinase 1 (PLK1) scored as synthetically lethal (Figure 7B, 7C). Although PLK1 is an essential mitotic kinase, a recently developed PLK1-specific inhibitor, BI-2536, shows excellent in vivo anti-tumor activity yet is well tolerated in animal models. This suggests that tumor and normal cells have differential sensitivity to PLK1 inhibition. Indeed, in the competition assay we were able to identify a dose range of BI-2536 where it preferentially killed Ras-Mut cells (Figure 7D). This example demonstrates that our screen will discovery drug targets where inhibitors already exist in various stages of development, and therefore illustrate the power of our approach in rapidly translating basic research into drug discovery.

We have several strategies in place for in-depth validation of candidate RSL genes

to exclude shRNA off-target effects and to elucidate their mechanisms in supporting the oncogenic activity of mutant Ras. We believe we will soon be able to expand this screen to the genome scale with our entire shRNA library. We expect to identify many genes that have not been previously identified as being required to support the Ras oncogene.
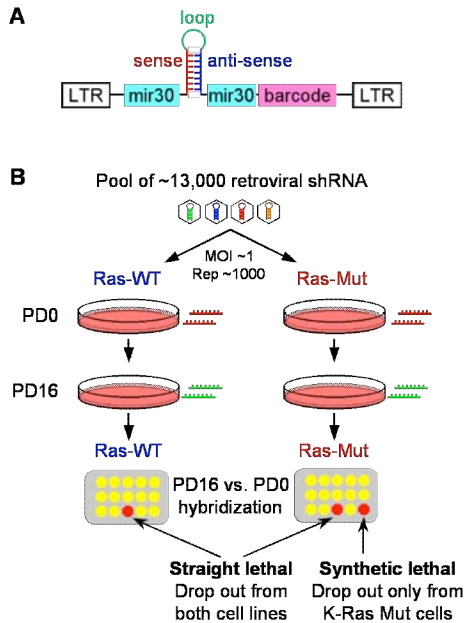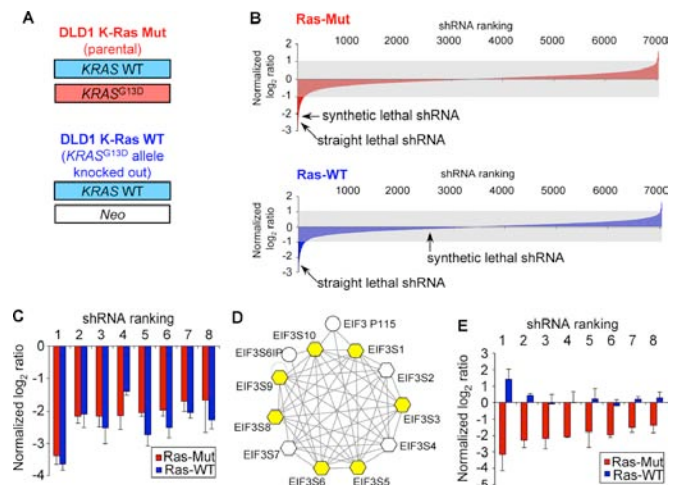

Figure 5. Design of a Ras synthetic lethal screen using barcoded shRNA retroviral libraries. A. The Elledge-Hannon mir30 retroviral shRNA vector. B. Schematics of the Ras synthetic lethal screen. An isogenic pair of DLD-1 cell lines, Ras-WT and Ras-Mut, are screened in parallel with a pool of retroviral shRNAs. Cells are infected with a multiplicity of infection (MOI) of 1 (i.e. one virus per cell) and a representation (Rep) of 1000 (i.e. 1000 independent integrants for each shRNA species). Initial and end samples are harvested at day-2 post infection (PD0) and after 16 population doublings (PD16), respectively. For each cell line, the shRNA barcodes are PCR-recovered from PD0 and PD16 genomic DNA samples, labeled with Cy5 and Cy3 dyes, respectively, and competitively hybridized to a microarray containing the corresponding probes. The Cy3/Cy5 ratio for a probe indicates the change in relative representation of that shRNA in the population over time ($\log_2 Cy3/Cy5 < 0$, depleted; $\log_2 Cy3/Cy5 \approx 0$, unchanged; $\log_2 Cy3/Cy5 > 0$, enriched).

Figure 6. A pilot Ras synthetic lethal screen. A. The isogenic DLD-1 cell lines used for the screen. The parental line (Ras-Mut) contains a wild type and a G13D mutant K-Ras allele. The Ras-WT derivative has the G13D mutant allele knocked out and replaced with a neomycin cassette. B. Representative results from the pilot screen with 8011 shRNAs. For each cell line, the normalized $\log_2$ ratios of Cy3/Cy5 signals (i.e. PD16 vs. PD0) for the shRNAs are calculated and rank-ordered. A negative $\log_2$ ratio indicates the shRNA has dropped out over time. Most shRNAs show little change in their representations over time ($\log_2$ ratios between −1 and 1, shaded in gray box), whereas a small number of shRNAs show significant depletion ($\log_2 < -1$) or enrichment ($\log_2 > 1$). The behavior of an expected "straight lethal" shRNA (toxic to both cell lines) and a "synthetic lethal" shRNA (selectively toxic to Ras-Mut cells) are indicated. C. Examples of straight lethal shRNAs from the screen ranked by log2 ratio. D. shRNAs against many subunits of the eukaryotic translation initiation factor 3 (EIF3) complex were recovered as straight lethals (shaded yellow) from the pilot screen. E. Examples of synthetic lethal shRNAs from the screen. Error bars represent S.D. across independent triplicates.
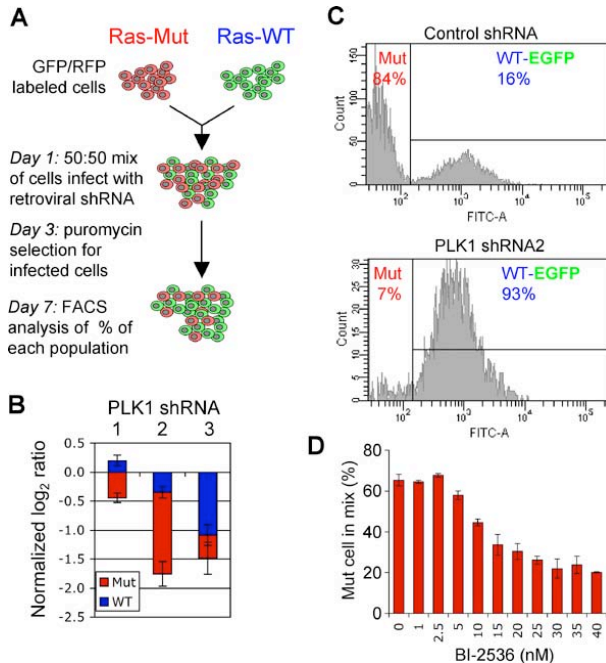
Figure 7. Validation of polo-like kinase 1 (PLK1) as a candidate RSL gene. A. Schematics of the competition assay. Ras-Mut or Ras-WT cells are labeled with either green or red fluorescent proteins and mixed together prior to infection with either candidate shRNA or control shRNA. This mixture is analyzed by FACS on day-7 to determine the relative fraction of Mut and WT cells. B. The $\log_2$ ratios of 3 shRNAs against PLK1 in the pilot screen. One hairpin, PLK1-shRNA2, shows particularly strong synthetic lethality with the Ras-Mut cells. C. Validation of the PLK1-shRNA2 by competition assay using unlabeled Ras-Mut cells and EGFP-labeled Ras-WT cells. As Ras-Mut cells have a proliferation advantage, they constitute the majority of the mixed population by day 7 in the control shRNA infected sample. In PLK1-shRNA2 infected sample, however, Ras-WT cells dominate, indicating selective depletion of Ras-Mut cells. D. The PLK1-selective inhibitor, BI-2536, selectively depletes Ras-Mut cells in a dose-dependent manner in the competition assay. 50:50 mixtures of Ras-Mut and Ras-WT cells were treated with various concentrations of BI-2536 (x-axis) for 5 days, the % of Ras-Mut cells in the mixture were then analyzed by FACS (y-axis). Error bars represent S.D. across triplicates.

Investigation of Auto-Antibodies as Breast Cancer Biomarkers

To identify auto-antibodies that could be used as biomarkers of breast cancer, we have generated a phage display library of coding fragments encompassing all open reading frames of the human genome. We designed approximately 467,000 overlapping oligonucleotides to cover 23,959 known protein-coding sequences. Nineteen pools spanning all open reading frames of the genome were constructed consisting of 22,000 oligonucleotides per pool, with each oligonucleotide encoding 36 codons and a 21-nucleotide region of overlap between each consecutive fragment. Furthermore, three pools encoding the N-terminal and C-terminal fragments of all open reading frames of the genome were constructed consisting of 18,000 oligonucleotides per pool, with each oligonucleotide encoding the first or last 24 codons of each open reading frame. These twenty-two HsORF pools (Homo *sapiens* open reading frame pools) were cloned into the T7FNS2 vector, a derivative of the T7-based vector T7Select 10-3b (Novagen).

The T7Select 10-3b system of lytic phage display is a mid-copy vector that displays between 5-15 copies on the surface of the T7 capsid. The natural T7 10 capsid protein is produced in two forms (10A and 10B), where 10B is produced by a translational frameshift of 10A and is approximately 10% of the protein that comprises the capsid as we described in last years report. The vector generates only the 10B version of the capsid protein, and the coding sequences for the displayed peptides are cloned into a multiple cloning site to produce C-terminal fusions of the peptides of interest with 10B. Phage are amplified on a bacterial host that carries an ampicillin-resistant plasmid expressing additional 10A capsid protein from a T7 promoter. We modified the T7Select

14

10-3b vector to generate T7FNS2 by inserting a sequencing encoding a FLAG epitope in the Not1 and Xho I sites to allow for selection of in-frame peptides.

Cloning of the HsORF pools into T7FNS2 resulted in a C-terminal fusion of the ORF fragments with the T7 10B capsid protein, followed by a C-terminal FLAG epitope. We have successfully cloned and validated all 22 pools to generate a phage display library encompassing all open reading frames. During the cloning of the library, we took several quality controls measures to ensure a highly accurate library with less than 5% of the phage resulting from vector re-ligation or cloning of multiple inserts. By sequence analysis of phage from several pools, approximately 55% of the phage inserts were entirely correct, 19% of the inserts had 1 or more mutation, but were still in frame, and 12% of the inserts had insertions or deletions resulting in frameshifts. These mutations, insertions, and deletions occurred during the oligonucleotide synthesis. Furthermore, approximately 74% of the library encodes in-frame phage, a much higher percentage of in-frame phage than libraries previously generated from cDNA sources using conventional methods.

This library of phage-displayed peptides will be screened for binding to antibodies present in patient sera. We have collaborated with the Breast SPORE blood bank at the Dana-Farber Cancer Institute and have obtained 250 serum samples collected from breast cancer patients. Furthermore, we have collaborated with Phil de Jager at the Harvard-Partners Center for Genetics and Genomics and have obtained 200 serum samples collected from healthy volunteers as well as 200 serum samples collected from patients with multiple sclerosis to serve as negative controls.

We plan to immunoprecipitate the IgG from pools of patient sera using protein G beads and then immunoprecipitate specific peptide-displaying phage using the IgG-bound beads. We will PCR the ORF fragment insert from the immunoprecipitated phage and use this amplicon for pool deconvolution. We plan to use either microarray technology or massively parallel Solexa sequencing for pool deconvolution. Currently, we are optimizing the phage immunoprecipitations and pool deconvolution aspects of the screen. Taken together, these technologies will be used to determine the peptide specificity for the entire antibody repertoire of a given pool of serum samples. By comparing the immunoreactive peptide profiles of serum samples from multiple breast cancer patients and excluding peptides observed in sera from healthy volunteers, we aim to identify tumor-associated biomarkers.

Key research accomplishments

1. Demonstration that our new version 2 libraries are capable of being screened in large pools to identify hairpins that are toxic to cancer and normal cells. This is a key finding essential to accomplishing the goals of this grant.
2. Development of a new barcoding method that works with our version 2 libraries that gives us an ability to quickly screen libraries before their 60-mer barcodes are sequenced/
3. Identification of genes that are selectively toxic to cancer cells.
4. Identification of several genes that appear to be synthetically lethal with ras.
5. Pharmacological validation of Plk1 as a synthetic lethal with Kras.
6. Generation of genome wide phage display libraries that cover the entire coding capacity of the human genome for the auto-antibody profiling.

With respect to the Statement of Work, we have accomplished many of the goals of year 1, 2 and 3 shown below.

Stament of Work. Year 1.

Task 1 **(Months 1-12)**
In the first year we anticipate beginning to work out the conditions for using the bar coding method to follow retroviruses containing hairpins as mixtures in complex libraries. We now have a library of 22,000 hairpins covering about 8,000 genes. We will be performing exploratory screens and optimizations to test the quality of the barcoding method. We must have this method working well to carryout the synthetic lethal screens.
We accomplished this goal in two ways. The first is we performed a bar code screen for potential tumor suppressors and identified several genes described in our first report and in Westbrook et al, 2005). Secondly, we have improved our vectors to allow single copy knockdown as described in Stegmeier et al. 2005. This was absolutely essential for the bar coding experiments we have proposed to kill cancer cells.
Task 2 **(Months 1-24 and possibly longer, an ongoing effort)**
We will continue to expand the library during this period to encompass more genes. This will be done in collaboration with Dr. Greg Hannon.
We have accomplished this goal by the generation of a second generation library in the mir30 context as described in Silva et al., 2005. This covers 140,00 human and mouse shRNAs as was described in last years report. We have also developed new and better knockdown vectors to allow us to knock down genes with greater penetrance. Right now we feel we have nearly genome-wide coverage and are working on a new library which if successful wil be a much better and more trustworthy library.
Task 3 **(Months 6-24)**
We also will begin the process of analyzing the human genome for coding sequences to set up the bio-informatics analysis to generate a list of sequences we wish to express to look for auto-antibodies. We should begin synthesizing oligo nucleotides to cover human genes.
We have designed oligonucleotides to cover the human genome. We are through cloning them in phage display vectors. We are characterizing the libraries and trying to figure

out how best to screen them. We ran into the problem that screening them by microarray ran into cross hybridization problems which we are addressing bioinformatically.

Year 2.

Task **4 (Months 13-24)**
In this period we plan to begin to carryout screen to look for genes which when knocked down by shRNA will interfere with the growth of cells containing defined mutations that lead to breast cancer. We will start with known tumor suppressors such as loss of p53 and Rb. We will use the barcoding methods. We may also screen for genes that sensitize cells to killing by gamma IR.
We initially tried PTEN mutants but were unable to find synthetic lethals. We have now successfully started with Kras and identified a few reproducible genes in a pilot experiment that are selectively toxic with Kras mutant cells.
Task 5 **(Months 18-36)**
We will begin to synthesize shRNA clones corresponding to the mouse genome.
We have completed this and now have about 100,000 sequence verified shRNAs in our version 2 library already.
Task 6 **(Months 12-24) We will expand the library of short coding regions for the autoantibody project and work out conditions to express these protein fragments in bacteria in a high through-put fashion.**
We have made the libraries and are working on developing methods to analyze the results.s.

Year 3.

Task 7 **(Months 24-36)**
We will continue to screen for synthetic lethals with tumor causing mutations relevant to breast cancer. In addition, by this time we will be retesting the synthetic lethal positives from the initial screens performed in year two.
We are in the process of performing straight lethal and synthetic lethal experiments with ras.

Task 8 **(Months 24-36) We plan to work out the conditions for placing the proteins expressing short segments of human proteins for the auto-antibody screening project onto glass slides for screening purposes.**
We have abandoned this aim in that we switched our approach to a phage display library which does not require glass slide. We just made our first comprehensive library in a T7 display vector.
Task 9 **(Months 24-36) We will continue to characterize the mouse shRNA library.**
We are characterizing the mouse library. It is being transferred into our best knockdown vector.

Years 4 and 5.

These years are listed together as they will be consumed with executing the long-term goals of the Tasks outlined in years 1 through 3.

Task 10 (Months 36-60)
We will begin to screen human sera for autoantibodies against our arrays of human protein fragments. We will work out these methods and attempt to begin a higher through-put analysis to determine if common epitopes are eliciting a response in breast cancer patients.
We have made the libraries and obtained the sera samples and hope to have data for the next report.

Task 11 (Months 36-60) **We will infect mice with retroviral libraries and screen for tumor suppressors in the breast and possibly other tissues.**
We have not gotten to the point where we can do this aim as we are consumed with finding the cancer the lethals.

Task 12 (Months 36-60) **We will be examining the genes we have found in various screens using standard molecular biological approaches to understand their roles in control of the responses we screened for in previous tasks.**
We are doing this with some of our tumor suppressor hits and some potential oncogenes we have found. We are following up on the cancer-specific lethals as well as ras synthetic lethals.

Reportable outcomes

Solimini, N.L., Luo, J., Elledge, S.J. (2007) Non-oncogene addiction and the stress phenotype of cancer cells. *Cell*. 130:986-8.

Schlabach, M.R., Luo, J., Solimini, N.L., Hu, G., Xu, Q., Li, M.Z., Zhao, Z., Smogorzewska, A., Sowa, M.E., Ang, X.L., Westbrook, T.F., Liang, A., Chang, K., Hackett, J.A., Harper, W.J., Hannon, G.J., and Elledge, S.J. (2008) Cancer Proliferation Gene Discovery Through Functional Genomics. (2008) *Science* 319:620-4.

Silva, J.M., Marran, K., Parker, J.S., Silva, J., Golding. M., Schlabach, M.R., Elledge, S.J., Hannon, G.J., Chang, K. (2008) Profiling essential genes in human mammary cells by multiplex RNAi screening. (2008) *Science* 319:617-20.

Conclusions

Progress on barcode screening for essential genes

It is clear from our current studies that we have overcome the main problem with performing bar code screens which is getting sufficiently good knockdown from single copy vectors and being able to reproducibly measure their abundance in complex pools by microarray hybridization. This has solved the main hurdle we had when we started this analysis. We have used a pool of shRNAs against kinases, ubiquitin-related proteins and the Cancer 1000 to look for genes toxic to cancer but not normal cells. We find that it is much easier to both kill and increase the proliferation rates of normal cells relative to cancer lines but we are finding genes that are specifically toxic to cancer lines. We have also begun to find genes that are synthetically lethal with Kras mutations. We hope to start cMyc synthetic lethals in the next year.

Screens for Tumor suppressors using the RNAi library and for Oncogenes using the ORFeome library

This ongoing effort should be completed in the next year. We ran into trouble with our cell transformation assay. Apparently the supplier of our specialized media for HMECs switched some of their components and nothing worked. We worked hard for 6 months and have finally overcome that problem which is important for both the shRNA screens as well as the overproduction screens for oncogenes

In addition we are performing the same screens with retroviral ORFeome libraries which are the equivalent to normalized full length cDNA libraries. We have identified a few genes in a pilot screen and we plan to take this genome-wide now that we have solved our transformation assay problems. One gene we are following up is PVRL4/Nectin-4. It potently transforms HMECs and is overproduced in 62% of ductal carcinomas.

Investigation of Auto-Antibodies as Breast Cancer Biomarkers

We are still in early stages of this project but we now have the libraries in hand and the patient samples. We have run into read-out problems on the microarrays because we are forced to use the sequences that are coded as opposed to optimized barcodes. We could read them out by highly parallel sequencing but it is too expensive to do it for 200 normals and 200 breast cancer patients so we are trying to optimize the probes on the microarray to limit cross hybes and we will see how that works..

References

1.   P. J. Paddison *et al.*, *Nature* **428**, 427 (2004).
2.   J. M. Silva *et al.*, *Nat. Genet.* **37**, 1281 (2005).
3.   T. F. Westbrook *et al.*, *Cell* **121**, 837 (2005).
4.   N. Popov *et al.*, *Nat. Cell Biol.* **9**, 765 (2007).
5.   I. G. Kolfschoten *et al.*, *Cell* **121**, 849 (2005).
6.   V. N. Ngo *et al.*, *Nature* **441**, 106 (2006).
7.   L. D. Wood *et al.*, *Science* **318**, 1108 (2007).
8.   K. Berns *et al.*, *Nature* **428**, 431 (2004).
9.   T. R. Brummelkamp *et al.*, *Nat. Chem. Biol.* **2**, 202 (2006).
10. T. Sjoblom *et al.*, *Science* **314**, 268 (2006).
11. L. D. Wood *et al.*, *Science* **318**, 1108 (2007).
12. R. A. Dickins *et al.*, *Nat. Genet.* **37**, 1289 (2005).
13. G. K. Smyth, T. Speed, *Methods* **31**, 265 (2003).
14. V. M. Draviam *et al.*, *Nat. Cell Biol.* **9**, 556 (2007).
15. M. Ito, T. Nakano, F. Erdodi, D. J. Hartshorne, *Mol. Cell. Biochem.* **259**, 197 (2004).
16. H. Jin, T. Sperka, P. Herrlich, H. Morrison, *Nature* **442**, 576 (2006).
17. L. T. Vassilev *et al.*, *Science* **303**, 844 (2004).
18. N. L. Solimini, J. Luo, S. J. Elledge, *Cell* **130**, 986 (2007).