Profiling Essential Genes in Human Mammary Cells by Multiplex RNAi Screening

Jose M. Silva, *et al.*

Science **319**, 617 (2008);
DOI: 10.1126/science.1149185

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of August 16, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/319/5863/617

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/319/5863/617/DC1

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

http://www.sciencemag.org/cgi/content/full/319/5863/617#related-content

This article **cites 31 articles**, 8 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/319/5863/617#otherarticles

This article has been **cited by** 54 article(s) on the ISI Web of Science.

This article has been **cited by** 17 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/319/5863/617#otherarticles

This article appears in the following **subject collections**:
Medicine, Diseases
http://www.sciencemag.org/cgi/collection/medicine

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2008 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

Profiling Essential Genes in Human Mammary Cells by Multiplex RNAi Screening

Jose M. Silva,[1] Krista Marran,[1] Joel S. Parker,[3] Javier Silva,[1] Michael Golding,[1]
Michael R. Schlabach,[2] Stephen J. Elledge,[2] Gregory J. Hannon,[1]* Kenneth Chang[1]

By virtue of their accumulated genetic alterations, tumor cells may acquire vulnerabilities that create opportunities for therapeutic intervention. We have devised a massively parallel strategy for screening short hairpin RNA (shRNA) collections for stable loss-of-function phenotypes. We assayed from 6000 to 20,000 shRNAs simultaneously to identify genes important for the proliferation and survival of five cell lines derived from human mammary tissue. Lethal shRNAs common to these cell lines targeted many known cell-cycle regulatory networks. Cell line–specific sensitivities to suppression of protein complexes and biological pathways also emerged, and these could be validated by RNA interference (RNAi) and pharmacologically. These studies establish a practical platform for genome-scale screening of complex phenotypes in mammalian cells and demonstrate that RNAi can be used to expose genotype-specific sensitivities.

The observation of genetic interactions is key to the definition of cellular networks. RNAi has enabled genetic approaches in both cultured mammalian cells (*1*–*5*) and intact animals (*6*–*9*). Large-scale screens of small interfering RNA (siRNA) (*10*–*12*) and shRNA collections (*5*, *13*–*16*) have generally adopted a one-by-one approach, interrogating phenotypes in a well-based format. This requires both considerable infrastructure and a substantial investment for each cell line to be screened. Alternatively, shRNA collections can be screened by assaying enrichment from pools, but this limits the range of phenotypes that can be addressed. Our focus was identifying essential genes or synthetically lethal genetic interactions through shRNAs that were selectively depleted from populations. This type of screen holds promise for the discovery of novel targets for cancer therapy and genetically validated combination therapies. Previously, one such screen was reported; however, this tested only ~500 shRNAs in a single pool (*17*). We therefore sought methods that allow multiplex analysis of phenotypic outputs on a genomic scale.

Pooled libraries drew from our previous collections wherein shRNAs are carried in a backbone derived from miR-30 (*18*). Combining RNA polymerase II promoters with miR-30–based shRNAs permits efficient suppression even with a single-copy integrant (*19*, *20*). Therefore, pooled shRNAs were transferred from pSM2 (*18*) to pLMP (*19*), wherein shRNA expression is driven from the murine stem cell virus long-terminal repeat promoter. Three different pools, containing ~6000, ~10,000, and ~20,000 shRNAs, were constructed to test screening at varying scales and levels of population complexity. Target cell populations were infected such that each cell contained, on average, a single integrated virus, and each individual shRNA occupied ~1000 cells. Three parallel infections generated biological replicate samples. Because our goal was to identify essential genes, genomic DNA was prepared from each replicate at three time points during a simple outgrowth assay (Fig. 1A).

Each shRNA cassette contains two unique identifiers: the shRNA itself and a random 60-nucleotide barcode. Barcode sequences were determined for the human shRNA library, and custom, multiplex format microarrays were prepared that contained both barcode and half-hairpin (HH) probes (*21*) (Fig. 1B). Proviral DNA fragments encompassing both shRNAs and barcodes were amplified from genomic DNA pools and hybridized to arrays in competition with a common reference.

We established a rigorous data analysis pipeline (*22*) for analyzing pooled shRNA screens. Correlations between biological replicates were high but diminished at later time points, whereas correlations between the reference channels remained unchanged (table S1). Overall, a gene was scored as a candidate if either its barcode or shRNA probe showed greater than 2-fold change with a false discovery rate (FDR) <10%.

We began with a pooled analysis of 6000 (6K) shRNAs in MCF-10A and MDA-MB-435. Although enriched gene sets varied considerably, similar numbers and largely overlapping gene sets showed depletion in both cell lines (tables S2 and S3). Among negatively selected shRNAs were many targeting regulators of the cell division cycle (*23*, *24*) (table S3). These included cyclins, cell division cycle (CDC) proteins, E2F family members, minichromosome maintenance deficient genes, proliferating cell nuclear antigen, and RNA polymerase II–associated genes. Additionally, the proteasome (15 of 25 subunits; $P = 5.61 \times 10^{-5}$) and anaphase-promoting complex/cyclosome (APC/C) (6 of 11 subunits; $P = 0.0139$) scored as being essential in both cell lines (table S3).

To validate candidates, we constructed a regulated shRNA vector, which linked shRNA and green fluorescent protein expression (fig. S1A). Inducible shRNAs against two APC/C subunits,

---

[1] Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [2] Department of Genetics, Center for Genetics and Genomics, Brigham and Women's Hospital, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA. [3] Expression Analysis, Inc., 4324 South Alston Avenue, Durham, NC 27713, USA.

*To whom correspondence should be addressed. E-mail: hannon@cshl.edu

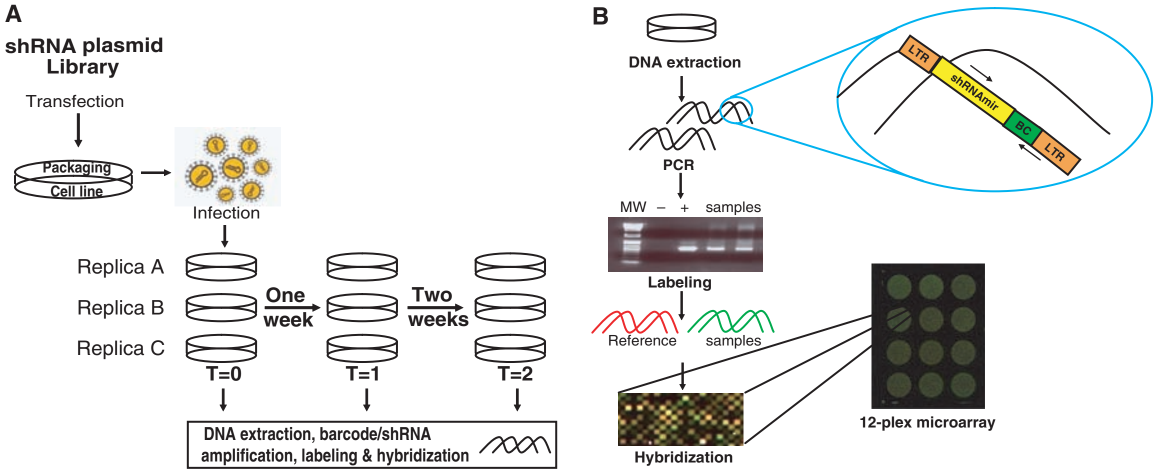
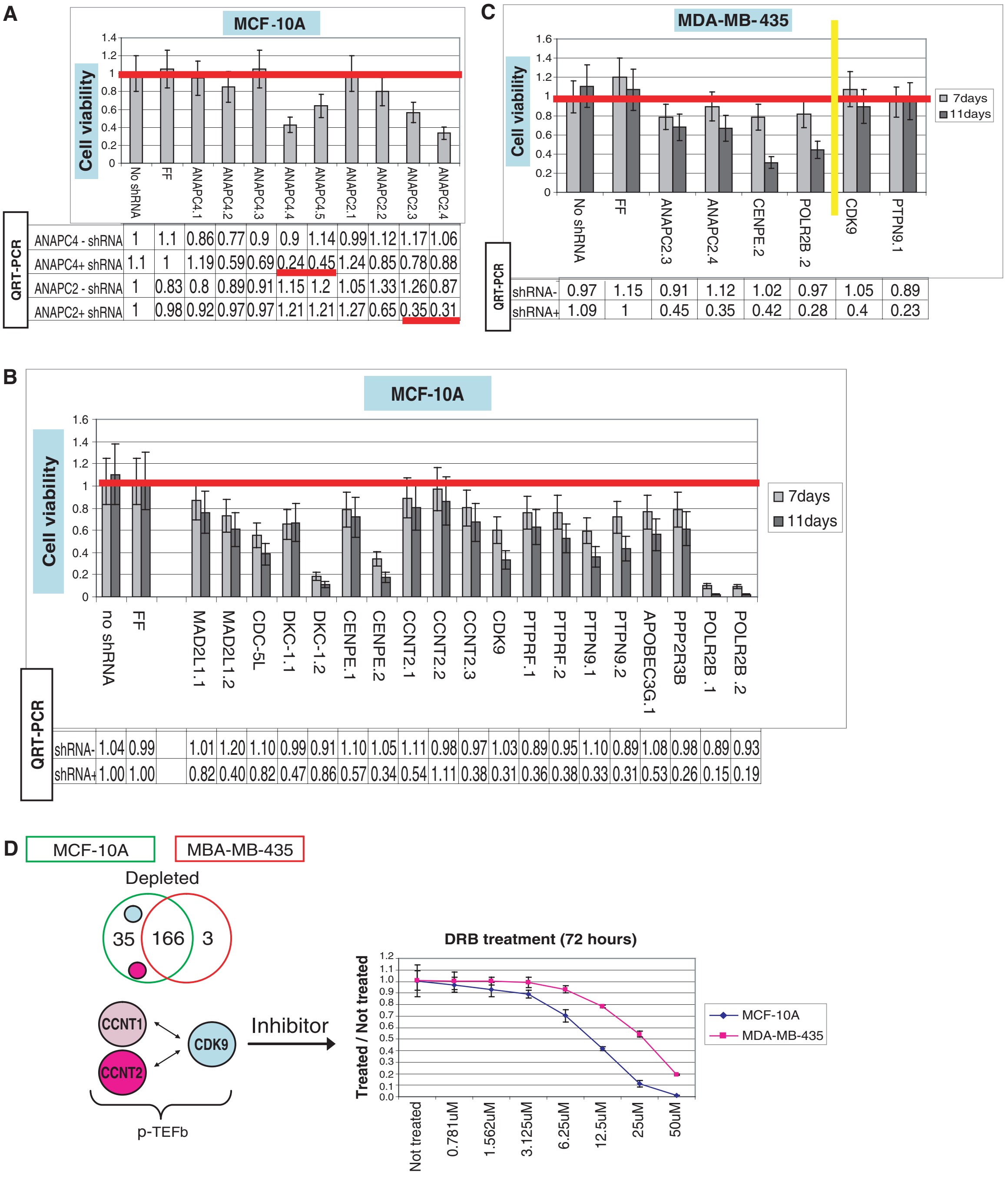


Fig. 1. Experimental approach. (**A**) shRNA plasmids were packaged into retroviruses in triplicate and introduced into replicate target cell populations at a multiplicity of ~0.3 to achieve ~1 integrant per cell. Over a 2-week culture period, time points were collected on day 2 or day 4 after infection and then once each week for 2 weeks. (**B**) The shRNA guide strand and the barcode region were amplified from genomic DNA from screening pools. Polymerase chain reaction (PCR) products were gel-purified, labeled, and hybridized to multiplex arrays in competition with a common reference.

ANAPC2 and 4, inhibited the growth of MCF-10A in a manner that correlated with mRNA knockdown (fig. S1A and Fig. 2A). Similarly, MDA-MB-435 was sensitive to ANAPC2 depletion (Fig. 2C). Nineteen additional MCF-10A lines were constructed with inducible shRNAs targeting 11 different candidates (Fig. 2B). Of these, 16 lines exhibited shRNA-dependent growth inhibition (30% to 95%), which correlated with mRNA knockdown in 14 cases. The exceptions were CDC-5L and DKC-1, where growth suppression could be due to off-target effects (Fig. 2B).

Among additional candidates were MAD2 and BUBR1, mitotic checkpoint proteins required for regulation of sister chromatid separation (*24*–*26*), and Kinesin-7/CENP-E, a component of the kinetochore (*27*). MAD2/MAD2L1 and Kinesin-7/CENP-E were validated as being essential in MCF-10A (table S3 and Fig. 2B). CENP-E depletion also inhibited growth in MDA-MB-435 (table S3 and Fig. 2C). Considered together, these studies showed that multiplex RNAi screens successfully identified essential components of cell growth and survival networks.

We also screened higher complexity populations containing 10,000 (10K) or 20,000 (20K) shRNAs. The 10K pool was introduced into MDA-MB-231, T-47D and ZR-75-1. The most complex pool (20K) was introduced into MCF-10A to allow direct comparison with the 6K screen. In all cases, cell numbers were scaled to maintain a representation of 1000 cells per shRNA. The quality of each screen was similar, with high correlations between biological replicates (table S4). We assessed the consistency of the MCF-10A screens by comparing depleted gene sets for the 6K and 20K pools. FDR thresholds were the same for both data sets (q < 0.1), but the fold-change criterion was relaxed from 2-fold to 1.5-fold for the 20K screen so that similar numbers of candidates were compared. A set of 172 genes ($P = 1.123 \times 10^{-9}$) overlapped in both data sets, despite some differences in the protocols



A

	No shRNA	FF	ANAPC4.1	ANAPC4.2	ANAPC4.3	ANAPC4.4	ANAPC4.5	ANAPC2.1	ANAPC2.2	ANAPC2.3	ANAPC2.4
ANAPC4 - shRNA	1	1.1	0.86	0.77	0.9	0.9	1.14	0.99	1.12	1.17	1.06
ANAPC4+ shRNA	1.1	1	1.19	0.59	0.69	0.24	0.45	1.24	0.85	0.78	0.88
ANAPC2 - shRNA	1	0.83	0.8	0.89	0.91	1.15	1.2	1.05	1.33	1.26	0.87
ANAPC2+ shRNA	1	0.98	0.92	0.97	0.97	1.21	1.21	1.27	0.65	0.35	0.31

C

	No shRNA	FF	ANAPC2.3	ANAPC2.4	CENPE.2	POLR2B .2	CDK9	PTPN9.1
shRNA-	0.97	1.15	0.91	1.12	1.02	0.97	1.05	0.89
shRNA+	1.09	1	0.45	0.35	0.42	0.28	0.4	0.23

B

	no shRNA	FF		MAD2L1.1	MAD2L1.2	CDC-5L	DKC-1.1	DKC-1.2	CENPE.1	CENPE.2	CCNT2.1	CCNT2.2	CCNT2.3	CDK9	PTPRF.1	PTPRF.2	PTPN9.1	PTPN9.2	APOBEC3G.1	PPP2R3B	POLR2B .1	POLR2B .2
shRNA-	1.04	0.99		1.01	1.20	1.10	0.99	0.91	1.10	1.05	1.11	0.98	0.97	1.03	0.89	0.95	1.10	0.89	1.08	0.98	0.89	0.93
shRNA+	1.00	1.00		0.82	0.40	0.82	0.47	0.86	0.57	0.34	0.54	1.11	0.38	0.31	0.36	0.38	0.33	0.31	0.53	0.26	0.15	0.19

Fig. 2. Validation of genes essential to multiple cell lines. Cell viability assays (bars) were performed on cell lines (MCF-10A or MDA-MB-435) expressing individual candidate shRNAs. Tables below the graphs show the level of target suppression, determined by quantitative real-time fluorescence PCR, with or without shRNA induction (indicated). An shRNA targeting luciferase (FF) and no shRNA serve as negative controls. (**A**) APC subunits ANAPC2 and ANAPC4 were suppressed by multiple hairpins in MCF-10A (1 to 5 for ANAPC4 and 1 to 4 for ANAPC2). Cell viability assays were carried out for 7 days after shRNA induction. (**B**) Nineteen additional inducible MCF-10A cell lines were generated to validate shRNAs that were depleted in the screen. Viability assays were carried out for 11 days. (**C**) Validation of shRNAs that were depleted (left of the yellow line) or not depleted (right of the yellow line) in MDA-MB-435 cells. Viability assays were carried out for 11 days. (**D**) The Venn diagram illustrates the 166 depleted genes that were common to both screens and the 35 and 3 genes specifically depleted from MCF-10A and MDA-MB-435, respectively. shRNAs targeting P-TEFb components CDK9 and cyclin T2 were both depleted specifically in MCF-10A. The graph shows a dose-response curve for growth inhibition of MCF-10A and MDA-MB-435 cells by DRB over a 72-hour period.

used to carry out each screen, and most of the validated targets from the 6K screen were found in the overlapping list of essential genes (tables S5 and S6). This suggests that a pool of ~20K shRNAs can be effectively screened.

We next sought to uncover cell line–specific genetic sensitivities that might reflect differences in the genetic constitutions of MCF-10A, MDA-MB-231, MDA-MB-435, and ZR-75.1. Initial comparisons focused on the 6K screens done with MCF-10A and MDA-MB-435. Filtering for shRNAs that had a low FDR ($q < 0.1$) and at least 2-fold depletion in MCF-10A but no more than 1.2-fold depletion in MDA-MB-435 yielded 35 genes (table S7). This compares to 166 genes that were important for growth in both cell lines and 3 genes that were differentially required in MDA-MB-435 (Fig. 2D and table S7). Among the candidates required in MCF-10A were two components of P-TEFb, CDK9 and cyclin T2 (*28*). We verified this differential sensitivity using both conditional shRNA expression (Fig. 2, B and C) and pharmacological inhibition (Fig. 2D). CDK9 is a DRB-sensitive kinase (*28*). Although DRB may also target other proteins, MCF-10A showed greater sensitivity to its effects than MDA-MB-435 cells (Fig. 2D).

We repeated the 6K screens of MCF-10A and MDA-MB-435 cells on the same array platform as the 10K screens of MDA-MB-231, T-47D, and ZR-75.1 cells and integrated the results (table S8). This was possible because more than 90% of the 6K shRNA set was contained within the 10K pool. Clustering of the resulting sensitivities (i.e., by fold-change, considering only shRNAs with $q < 0.1$) yielded a dendrogram wherein the more normal MCF-10A segregated from the other, more transformed lines (Fig. 3A). MDA-MB-435 also segregated, perhaps reflecting the observation that it is more related, by expression profiling, to melanoma than breast epithelia. Finally, the remaining lines separated into a group containing T-47D and ZR-75.1, both luminal tumor cell lines, and MDA-MB-231, a basal tumor cell line (*29*) (Fig. 3A).

Viewing this portrait of shRNA sensitivity in more detail revealed a number of pathways and complexes that were differentially required in MCF-10A. These included epidermal growth factor receptor (EGFR), an effect that could be reproduced pharmacologically using the EGFR inhibitor Tarceva (*30*) (Fig. 3B). DNA methyltransferases also scored either above or close to the threshold (table S8 and Fig. 3B). In accord with these results, MCF-10A cells showed a more than 50-fold greater sensitivity to 5-aza-deoxycytidine, a methyltrans-

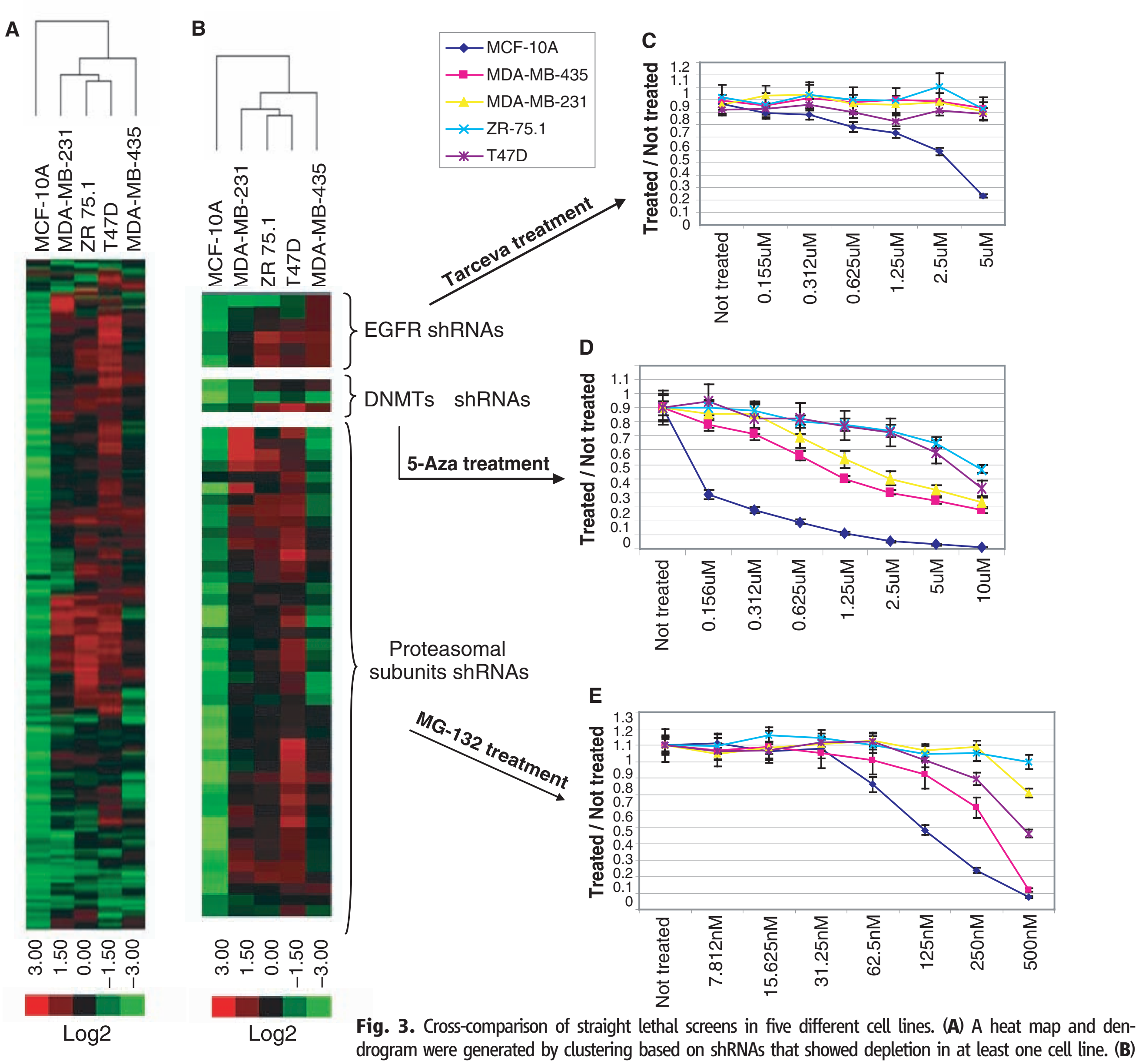


Fig. 3. Cross-comparison of straight lethal screens in five different cell lines. (**A**) A heat map and dendrogram were generated by clustering based on shRNAs that showed depletion in at least one cell line. (**B**) shRNAs targeting selected complexes or pathways were chosen from (A) to highlight responses to EGFR, methyltransferase, and proteasome lesions. Pharmacological inhibition of these pathways by treatment with Tarceva (**C**), 5-aza-deoxycytidine (**D**), or MG-132 (**E**) was used to validate the predictions of the shRNA screen.

ferase suicide substrate (*31*), than the other cell lines. As a final example, numerous proteasome subunits were preferentially depleted from MCF-10A (table S8 and Fig. 3B). These cells showed the greatest sensitivity to a proteasome inhibitor, MG-132 (*32*). Interestingly, MDA-MB-435 showed an intermediate level of sensitivity to the drug, and this was reflected precisely in their intermediate level of depletion of proteasomal shRNAs during the screen (table S8 and Fig. 3B).

We have validated a highly scalable approach for screening shRNA libraries. Although we used a phenotypic filter reflecting growth and survival, virtually any characteristic that allows separation of phenotypically distinct cells can be applied. We also validated the ability of functional shRNA screening to separate cell lines based on their genetic vulnerabilities in a manner that reflects their already defined characteristics (e.g., immortal versus tumor, basal versus luminal). Although one could attribute selective dependency to culture conditions in some cases, the overwhelming concordance of the shRNAs that affect proliferation and survival across these lines, many of which are cultured identically, strongly argues against this being a pervasive explanation. In all, this approach enables genome-wide screens for tumor-specific vulnerabilities to be carried out on large numbers of tumor lines. Moreover, it permits rational searches for lesions that synergize with existing therapeutics to produce a path toward genetically informed combination therapies.

References and Notes

1. P. J. Paddison, A. A. Caudy, G. J. Hannon, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1443 (2002).
2. S. M. Elbashir *et al.*, *Nature* **411**, 494 (2001).
3. L. Scherer, J. J. Rossi, *Curr. Pharm. Biotechnol.* **5**, 355 (2004).
4. A. Friedman, N. Perrimon, *Curr. Opin. Genet. Dev.* **14**, 470 (2004).
5. R. Bernards, T. R. Brummelkamp, R. L. Beijersbergen, *Nat. Methods* **3**, 701 (2006).
6. M. T. Hemann *et al.*, *Nat. Genet.* **33**, 396 (2003).
7. O. Snove Jr., J. J. Rossi, *Nat. Methods* **3**, 689 (2006).
8. M. H. Farah, *Curr. Drug Deliv.* **4**, 161 (2007).
9. A. de Fougerolles, H. P. Vornlocher, J. Maraganore, J. Lieberman, *Nat. Rev. Drug Discov.* **6**, 443 (2007).
10. P. Aza-Blanc *et al.*, *Mol. Cell* **12**, 627 (2003).
11. A. W. Whitehurst *et al.*, *Nature* **446**, 815 (2007).
12. F. Natt, *Curr. Opin. Mol. Ther.* **9**, 242 (2007).
13. P. J. Paddison *et al.*, *Nature* **428**, 427 (2004).
14. T. F. Westbrook *et al.*, *Cell* **121**, 837 (2005).
15. D. E. Root *et al.*, *Nat. Methods* **3**, 715 (2006).
16. C. Gazin *et al.*, *Nature* **449**, 1073 (2007).
17. V. N. Ngo *et al.*, *Nature* **441**, 106 (2006).
18. J. M. Silva *et al.*, *Nat. Genet.* **37**, 1281 (2005).
19. R. A. Dickins *et al.*, *Nat. Genet.* **37**, 1289 (2005).
20. F. Stegmeier, G. Hu, R. J. Rickles, G. J. Hannon, S. J. Elledge, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13212 (2005).
21. M. R. Schlabach *et al.*, *Science* **319**, 620 (2008).
22. Materials and methods are available as supporting material on *Science* Online.
23. K. I. Nakayama, K. Nakayama, *Nat. Rev. Cancer* **6**, 369 (2006).
24. J. M. Peters, *Nat. Rev. Mol. Cell Biol.* **7**, 644 (2006).
25. G. Fang, *Mol. Biol. Cell* **13**, 755 (2002).
26. G. J. Kops, D. R. Foltz, D. W. Cleveland, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8699 (2004).
27. T. J. Yen *et al.*, *Nature* **359**, 536 (1992).
28. J. Peng, Y. Zhu, J. T. Milton, D. H. Price, *Genes Dev.* **12**, 755 (1998).
29. R. M. Neve *et al.*, *Cancer Cell* **10**, 515 (2006).
30. J. D. Moyer *et al.*, *Cancer Res.* **57**, 4838 (1997).
31. L. Jackson-Grusby *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4681 (1997).
32. V. J. Palombella *et al.*, *Cell* **78**, 773 (1994).
33. We thank members of the Hannon and Lowe laboratories for helpful discussion and T. Moore of Open Biosystems for help and support. This work was supported by grants from NIH and the Department of Defense to G.J.H. and J.M.S. and by a kind gift from Kathryn W. Davis. G.J.H. and S.J.E. have a paid consulting relationship with Open Biosystems.

Supporting Online Material

www.sciencemag.org/cgi/content/full/319/5863/617/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S8
References

14 August 2007; accepted 20 December 2007
10.1126/science.1149185

Cancer Proliferation Gene Discovery Through Functional Genomics

Michael R. Schlabach,[1]* Ji Luo,[1]* Nicole L. Solimini,[1]* Guang Hu,[1]* Qikai Xu,[1] Mamie Z. Li,[1] Zhenming Zhao,[1] Agata Smogorzewska,[1,2] Mathew E. Sowa,[3] Xiaolu L. Ang,[3] Thomas F. Westbrook,[1] Anthony C. Liang,[1] Kenneth Chang,[4] Jennifer A. Hackett,[1] J. Wade Harper,[3] Gregory J. Hannon,[4] Stephen J. Elledge[1]†

Retroviral short hairpin RNA (shRNA)–mediated genetic screens in mammalian cells are powerful tools for discovering loss-of-function phenotypes. 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 cells and normal cells. We identified many shRNAs to be antiproliferative that target core cellular processes, such as the cell cycle and protein translation, in all cells examined. Moreover, 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.

We have recently generated 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 have also developed 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 (*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-nucleotide barcodes for pool deconvolution (*2*, *3*). To provide an alternative to these barcodes 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-nucleotide hairpin loop of our mir30-based platform and designed a polymerase chain reaction (PCR) strategy that amplifies only the 3′ half of the shRNA stem (Fig. 1B). As compared with full-hairpin sequences for microarray hybridization (*8*, *9*), HH barcodes entirely eliminate probe self-annealing during microarray hybridization (Fig. 1C and fig. S1, A and B), providing the critical dynamic range necessary for pool-based dropout screens. HH barcode signals are highly reproducible in replicate PCRs (R = 0.973, fig. S1A), highly specific (0.5% cross-reaction) (fig. S1C), and display a reasonable, although slightly compressed, dynamic range in mixing experiments with varied subpool inputs that are quantified by microarray hybridization (fig. S1, D and E). Taken together, these results indicate that HH barcodes are alternatives to the 60-nucleotide 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

---

[1] Howard Hughes Medical Institute and Department of Genetics, Center for Genetics and Genomics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [2] Department of Pathology, Massachusetts General Hospital (MGH), Boston, MA 02114, USA. [3] Department of Pathology, Harvard Medical School, Boston, MA 02115, USA. [4] Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: selledge@genetics.med.harvard.edu