Published in final edited form as:
Science. 2008 February 1; 319(5863): 620–624. doi:10.1126/science.1149200.

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], and **Stephen J. Elledge**[1,†]

[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

Abstract

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 lenti-viral 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

[†]To whom correspondence should be addressed. selledge@genetics.med.harvard.edu.
[*]These authors contributed equally to this work.

Supporting Online Material
www.sciencemag.org/cgi/content/full/319/5863/620/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
Tables S1 to S6
References
Data Sets S1 to S9

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 identified a distinct spectrum of mutations in these two cancer types (10,11). Also, the comparison between cancer cells and normal 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 kinase, phosphatase, ubiquitination pathway, and cancer-related genes (table S1). 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 a murine stem cell virus (MSCV)–derived 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 independent integrations per shRNA and with a multiplicity of infection of 1 to 2. Initial reference samples were collected 48 to 72 hours after infection. The remaining cells were puromycin-selected, propagated for several weeks, and collected again as the end samples. HH bar-codes 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. Correlations between initial samples across the triplicates and between the initial and end samples within each replica were high, indicating that the triplicates were highly reproducible and representation was well maintained throughout the experiment.

To identify shRNAs that consistently changed in abundance in each cell line, we analyzed data sets using a custom statistical package based on the Linear Models for Microarray data (Limma) method (13) for two-color cDNA microarray analysis (7). Whereas most shRNAs showed little changes in their abundance over time ($\log_2$ 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 three cancer cell lines from the normal HMECs, likely reflecting fundamental differences between cancer cells and normal cells (Fig. 2B). Furthermore, the two colon cancer cell lines were more similar to each other than to 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 HMECs showed statistically significant depletion (Fig. 2C and tables S3 to S6). The lists of antiproliferative shRNAs show significant overlap ($P < 1 \times 10^{-40}$), 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. 3, A and B). For example, shRNAs against multiple subunits of the anaphase promoting complex/cyclosome (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, $P = 3.2 \times 10^{-2}$), 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, $P = 2.4 \times 10^{-4}$; HMEC, $P = 8.6 \times 10^{-3}$) were identified (Fig. 3B). Several key proteins in the ubiquitination and sumoylation pathways, including most of the cullins, were also identified. Multiple shRNAs against the same gene scored in the screen, which suggests that their effects are unlikely due to off-target effects.

We next validated *EIF3S10* and *RBX1*: two genes that are essential for viability in all four cell lines. For each gene, we included shRNAs that scored in the screen as well as additional shRNA sequences present in our library (table S2). 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 antiproliferative phenotypes. Furthermore, the antiproliferative activity of the shRNAs correlated very well with the extent of target gene knockdown, as shown for *RBX1* (fig. S2A). Thus, these phenotypes are likely due to target gene knockdown rather than to off-target effects. This finding is consistent with a previous transfection-based screen with this library showing ~90% “on-target” efficiency (14).

In addition to the common set of shRNAs that impairs viability in all cell lines, we observed substantial numbers of genes that are selectively required for proliferation of each cell line (tables S3 to S6). These are particularly interesting because 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 (PP1), 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 and fig. S2B). This finding was corroborated with four additional PPP1R12A small interfering RNAs (siRNAs). These shRNAs and siRNAs resulted in comparable knockdown of PPP1R12A protein in both cell lines (fig. S2B), 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 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 and fig. S2C). These results suggest that 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 (tables S4 and S6). Not surprisingly, a much larger set of 695 genes is required by HMECs, likely reflecting the ability of normal cells to appropriately respond to various cellular stresses. Conversely, the relatively fewer genes that

are required by the cancer cells underscore 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*, which encodes the human homolog of MDM2 (the E3 ligase for p53) (Fig. 4D). HCC1954 and DLD-1 cells harbor inactivating mutations ($Tyr^{163}\rightarrow Cys^{163}$ and $Ser^{241}\rightarrow Phe^{241}$, respectively) in the *TP53* gene and are therefore insensitive to MDM2 knockdown. Multiple MDM2 shRNAs selectively impaired the viability of the p53 wild-type HMECs but not that of HCC1954 cells with mutant p53 (Fig. 4E and fig. S2D). 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 and fig. S2D).

Several genes appear to be selectively required by HCC1954 cells but not by HMECs (tables S4 and S6). Among these is the cell cycle regulator and spindle checkpoint kinase *BUB1* (Fig. 4G). We validated *BUB1* using both shRNAs and siRNAs to confirm that its knockdown is more detrimental to HCC1954 cells than to HMECs (Fig. 4H), despite similar levels of BUB1 protein reduction (fig. S2E). These results indicate that BUB1 is likely to play an integral role in supporting the oncogenic transformation of HCC1954 cells because they are more dependent on BUB1 function. One possible explanation for this enhanced dependency may be the near-tetraploid nature of the HCC1954 genome. As compared with the diploid HMECs, HCC1954 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 and an accompanying paper (19) demonstrate that highly parallel dropout screens that use complex pools of shRNAs can be achieved with the use of 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 and similar efforts, which focus on physical alterations of the cancer genome.

The most complex pool that we used contains 42,000 distinct shRNAs (fig. S3): an 80-fold increase in complexity as compared with that of previous dropout screens based on our designs (6). It is now conceivable for researchers to screen the entire human genome with ~3 shRNAs per gene using 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank A. L. Brass for the pMSCV-PM, pMSCV-PM-FF, and pMSCV-PM-mir30 vectors and for scientific advice; M. J. Solimini for help with data analysis; E. R. McDonald for scientific advice; T. Waldman and B. Vogelstein for the HCT116 and DLD-1 cell lines; and T. Moore from Open Biosystems for help with assembling library pools. G.H. is a fellow of the Helen Hay Whitney Foundation. X.L.A. is supported by a National Research Service Award fellowship, M.E.S. is supported by an American Cancer Society fellowship, and A.S. is supported by grant

T32CA09216 to the MGH Pathology Department. T.F.W. is a fellow of the Susan G. Komen Foundation and is supported by grant PDF0403175. This work is supported by grants from NIH and the U.S. Department of Defense to G.J.H., J.W.H., and S.J.E. G.J.H. has a paid consulting relationship with Open Biosystems.

References and Notes

1. Paddison PJ, et al. Nature 2004;428:427. [PubMed: 15042091]
2. Silva JM, et al. Nat Genet 2005;37:1281. [PubMed: 16200065]
3. Westbrook TF, et al. Cell 2005;121:837. [PubMed: 15960972]
4. Popov N, et al. Nat Cell Biol 2007;9:765. [PubMed: 17558397]
5. Kolfschoten IG, et al. Cell 2005;121:849. [PubMed: 15960973]
6. Ngo VN, et al. Nature 2006;441:106. [PubMed: 16572121]
7. Materials and methods are available as supporting material on *Science* Online.
8. Berns K, et al. Nature 2004;428:431. [PubMed: 15042092]
9. Brummelkamp TR, et al. Nat Chem Biol 2006;2:202. [PubMed: 16474381]
10. Sjoblom T, et al. Science 2006;314:268. [PubMed: 16959974]
11. Wood LD, et al. Science 2007;318:1108. [PubMed: 17932254]
12. Dickins RA, et al. Nat Genet 2005;37:1289. [PubMed: 16200064]
13. Smyth GK, Speed T. Methods 2003;31:265. [PubMed: 14597310]
14. Draviam VM, et al. Nat Cell Biol 2007;9:556. [PubMed: 17417629]
15. Ito M, Nakano T, Erdodi F, Hartshorne DJ. Mol Cell Biochem 2004;259:197. [PubMed: 15124925]
16. Jin H, Sperka T, Herrlich P, Morrison H. Nature 2006;442:576. [PubMed: 16885985]
17. Vassilev LT, et al. Science 2004;303:844. [PubMed: 14704432]
18. Solimini NL, Luo J, Elledge SJ. Cell 2007;130:986. [PubMed: 17889643]
19. Silva JM, et al. Science 2008;319:617. [PubMed: 18239125]

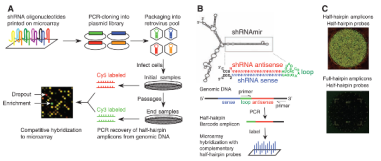

Fig. 1.

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

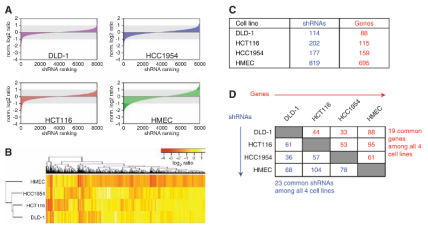

Fig. 2.

Pool-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 on the basis of their mean normalized $\log_2$ Cy3/Cy5 ratios. The shaded rectangle indicates the $\log_2$ ratio range within which an shRNA's abundance was considered unchanged. (**B**) Clustering of the four cell lines with the antiproliferative shRNAs identified in the screen. The color scale represents mean normalized $\log_2$ Cy3/Cy5 ratios of the probes. (**C**) Antiproliferative 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 antiproliferative 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).

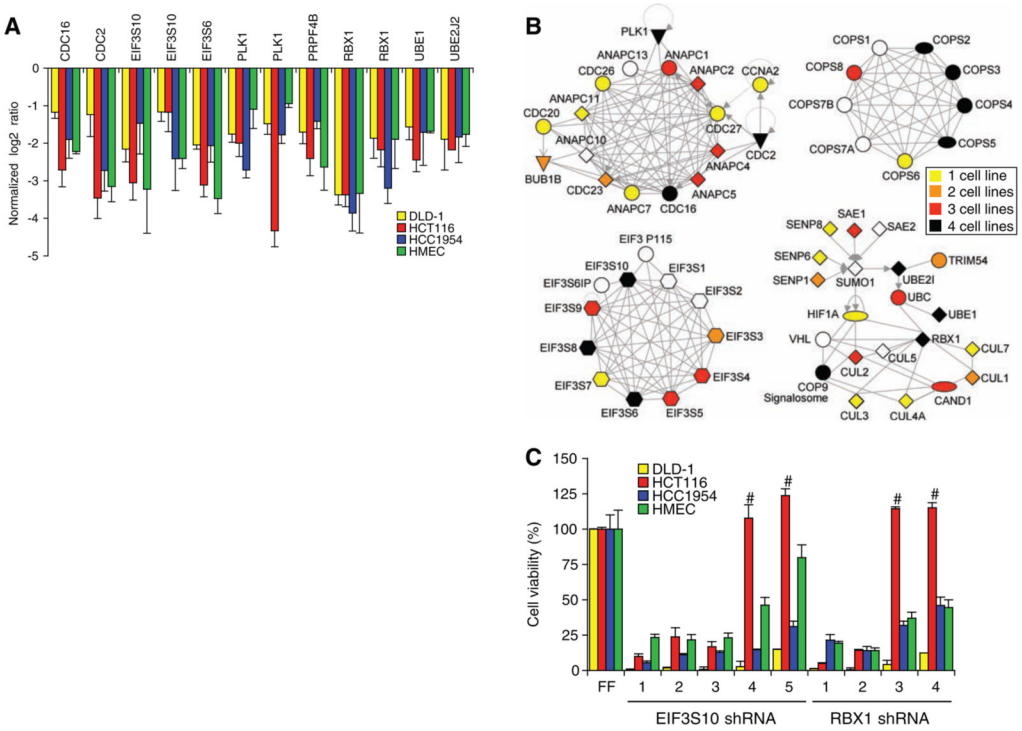

Fig. 3.

Genes commonly required for proliferation or survival of normal and cancer cells. Error bars represent SDs across triplicates. (**A**) Representative candidate shRNAs that reduce viability of all four cell lines. Multiple entries for the same gene indicate that multiple independent shRNAs have scored in the screen for that gene. For each shRNA, the normalized $\log_2$ Cy3/Cy5 ratio (i.e., its relative abundance in the end sample versus initial sample) is given. (**B**) Examples of core cellular modules required in all four cell lines. Shown are the APC/C (top left), the COP9 signalosome (top right), the eIF3 translation initiation complex (bottom left), and E3 ubiquitin ligases (bottom right). Color scheme indicates genes required for viability of one (yellow), two (orange), three (red), or all four (black) cell lines. (**C**) Validation of selected shRNAs from the screen that reduce viability of all four cell lines. Candidate shRNAs were individually packaged into viruses and infected into cells in 96-well plates in independent triplicates. Cell viability was measured at day 9 after infection. All viability reductions were significant ($P < 0.05$), except where indicated by the number sign. FF, negative control shRNA targeting firefly luciferase.

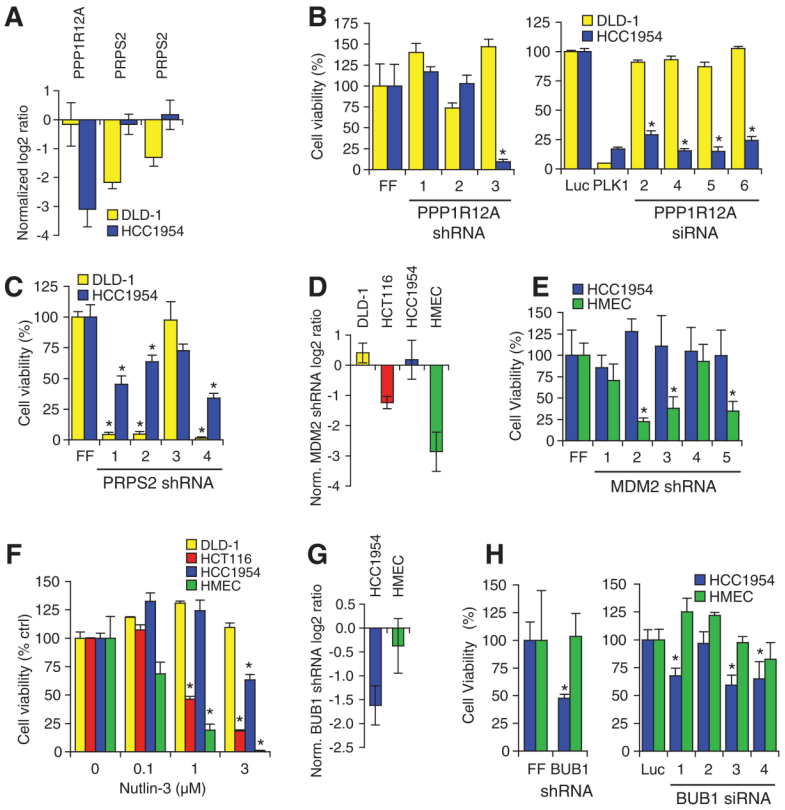

Fig. 4.

Genes selectively required for proliferation or survival of cancer cells. Error bars represent SDs across triplicates. (**A**) Identification of *PPP1R12A* (one shRNA) and *PRPS2* (two shRNAs) as two genes that are selectively required by HCC1954 or DLD-1 cells, respectively, in the screen. (**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 after infection (shRNA) or 4 days after transfection (siRNA) (*, $P < 0.05$). Luc, negative control siRNA targeting luciferase. PLK1, positive control siRNA targeting polo-like kinase 1. (**D**) Normalized $\log_2$ 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 after infection with retroviruses expressing five different MDM2 shRNAs (*, $P < 0.05$). (**F**) Differential sensitivity of the four cell lines to the MDM2 inhibitor nutlin-3. Cell viability is reflective of their p53 status (HMECs and HCT116 cells, p53 wild-type; HCC1954 and DLD-1 cells, p53 mutant). Cell viability was measured after 4 days of nutlin-3 treatment (*, $P < 0.05$). ctrl, control. (**G**) Normalized $\log_2$ ratios of a BUB1 shRNA from the screen. (**H**) Enhanced sensitivity of HCC1954 cells to BUB1 knockdown. Both shRNA (left) and siRNA (right) knockdown of BUB1 reduce HCC1954 cell viability but have no effect on HMEC viability. Cell viability was measured 4 or 9 days after transfection or infection, respectively (*, $P < 0.05$).