

# Highly parallel identification of essential genes in cancer cells

Biao Luo<sup>a,1</sup>, Hiu Wing Cheung<sup>a,b,c,1</sup>, Aravind Subramanian<sup>a,1</sup>, Tanaz Sharifnia<sup>a</sup>, Michael Okamoto<sup>a</sup>, Xiaoping Yang<sup>a</sup>, Greg Hinkle<sup>a</sup>, Jesse S. Boehm<sup>a</sup>, Rameen Beroukhim<sup>a,b,c</sup>, Barbara A. Weir<sup>a,b,c</sup>, Craig Mermel<sup>a,b,c</sup>, David A. Barbie<sup>a,b,c</sup>, Tarif Awad<sup>d</sup>, Xiaochuan Zhou<sup>e</sup>, Tuyen Nguyen<sup>a</sup>, Bruno Piquani<sup>a</sup>, Cheng Li<sup>f</sup>, Todd R. Golub<sup>a,b,g,h</sup>, Matthew Meyerson<sup>a,b,c</sup>, Nir Hacohen<sup>a,b,i,2</sup>, William C. Hahn<sup>a,b,c,2</sup>, Eric S. Lander<sup>a,j,k,2,3</sup>, David M. Sabatini<sup>a,h,j,k,2</sup>, and David E. Root<sup>a,2,3</sup>

<sup>a</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142; <sup>b</sup>Department of Medical Oncology and Center for Cancer Genome Discovery and Departments of <sup>c</sup>Biostatistics and Computational Biology and <sup>d</sup>Pediatric Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; <sup>e</sup>Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142; <sup>f</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; <sup>g</sup>Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Boston, MA 02114; <sup>h</sup>Harvard Medical School, Boston, MA 02115; <sup>i</sup>Affymetrix Inc., 3380 Central Expressway, Santa Clara, CA 95051; <sup>j</sup>Atactic Technologies Inc., 2575 West Bellfort, Suite 270, Houston, TX 77054; and <sup>h</sup>Howard Hughes Medical Institute

Contributed by Eric S. Lander, October 27, 2008 (sent for review September 11, 2008)

**More complete knowledge of the molecular mechanisms underlying cancer will improve prevention, diagnosis and treatment.** Efforts such as The Cancer Genome Atlas are systematically characterizing the structural basis of cancer, by identifying the genomic mutations associated with each cancer type. A powerful complementary approach is to systematically characterize the functional basis of cancer, by identifying the genes essential for growth and related phenotypes in different cancer cells. Such information would be particularly valuable for identifying potential drug targets. Here, we report the development of an efficient, robust approach to perform genome-scale pooled shRNA screens for both positive and negative selection and its application to systematically identify cell essential genes in 12 cancer cell lines. By integrating these functional data with comprehensive genetic analyses of primary human tumors, we identified known and putative oncogenes such as *EGFR*, *KRAS*, *MYC*, *BCR-ABL*, *MYB*, *CRKL*, and *CDK4* that are essential for cancer cell proliferation and also altered in human cancers. We further used this approach to identify genes involved in the response of cancer cells to tumoricidal agents and found 4 genes required for the response of CML cells to imatinib treatment: *PTPN1*, *NF1*, *SMARCB1*, and *SMARCE1*, and 5 regulators of the response to FAS activation, *FAS*, *FADD*, *CASP8*, *ARID1A* and *CBX1*. Broad application of this highly parallel genetic screening strategy will not only facilitate the rapid identification of genes that drive the malignant state and its response to therapeutics but will also enable the discovery of genes that participate in any biological process.

oncogene | pooled library | RNAi | screen | shRNA

**A**lthough human cancers harbor hundreds of genetic alterations, only a subset of these alterations is likely to impact tumor initiation or maintenance. Furthermore, genes that are not altered at the genomic level may play essential roles in tumor development. Thus, to identify genes with important roles in cancer, systematic functional assessment of genes for their contribution to specific cancer phenotypes is complementary to structural characterization of the cancer genome. Integrating both structural and functional approaches will provide insight into therapeutic targets for treating cancer.

The recent development of RNAi libraries targeting the human and mouse genomes has enabled systematic genetic studies in mammalian cells by using arrayed and pooled screens (1–8). However, scaling up the application of this methodology to identify all essential genes across a diverse range of human cancers requires an integrated experimental and computational approach that is efficient, robust, and economical. Here, we describe the development and application of genome-scale high-throughput methods using our lentiviral RNAi library to systematically assess cancer

gene function and to integrate structural and functional approaches in the study of cancer.

## Results and Discussion

To apply RNA interference at genome scale, we developed a highly parallel “pooled screening” strategy that employs the previously described library created by The RNAi Consortium (TRC) (9, 10). The TRC library contains ≈170,000 lentivirally encoded short hairpin RNAs (shRNAs), with 5 or more independent shRNAs targeting each of 17,200 human genes, as well as an equivalent collection targeting each of 16,000 mouse genes. The pooled screening approach involves infecting cultured cells with a pool of shRNAs, allowing the cells to proliferate for a period, isolating the shRNA sequences from the resulting cells by PCR amplification, and measuring the relative abundance of the shRNAs (by cleaving the hairpins with a restriction enzyme and hybridizing them to a microarray complementary to the half-hairpin sequences) [Fig. 1*A*, *supporting information (SI) Fig. S1* and *SI Methods*]. In the experiments below, we used a sublibrary containing 45,000 shRNAs corresponding to ≈9,500 human genes (45k shRNA pool). We demonstrated that 4-fold changes in relative shRNA abundance are easily resolved (Fig. 1*B*) using this approach.

As an initial test of the system, we performed 2 positive-selection screens. The first screen was designed to identify genes whose inhibition renders T cells resistant to apoptosis induced by the activation of FAS, which functions in immune cell homeostasis (11). We infected Jurkat T cells with the 45k shRNA pool, so that the typical cell received 0.3 shRNAs [multiplicity of infection (MOI) = 0.3] and each shRNA was introduced to ≈200 independent cells. After selection to eliminate uninfected cells, the remaining cells were treated for 21 days with an activating anti-FAS antibody (12) at a dose sufficient to deplete the number of uninfected cells by a factor of ≈10<sup>5</sup>. To identify shRNAs that confer resistance, we

---

Author contributions: B.L., H.W.C., N.H., W.C.H., D.M.S., and D.E.R. designed research; B.L., H.W.C., T.S., M.O., X.Y., G.H., J.S.B., R.B., B.A.W., D.A.B., X.Z., T.N., and B.P. performed research; A.S., G.H., B.A.W., T.A., C.L., T.R.G., and M.M. contributed new reagents/analytic tools; B.L., H.W.C., A.S., J.S.B., R.B., B.A.W., C.M., M.M., N.H., W.C.H., E.S.L., D.M.S., and D.E.R. analyzed data; and B.L., H.W.C., A.S., T.S., J.S.B., N.H., W.C.H., E.S.L., D.M.S., and D.E.R. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

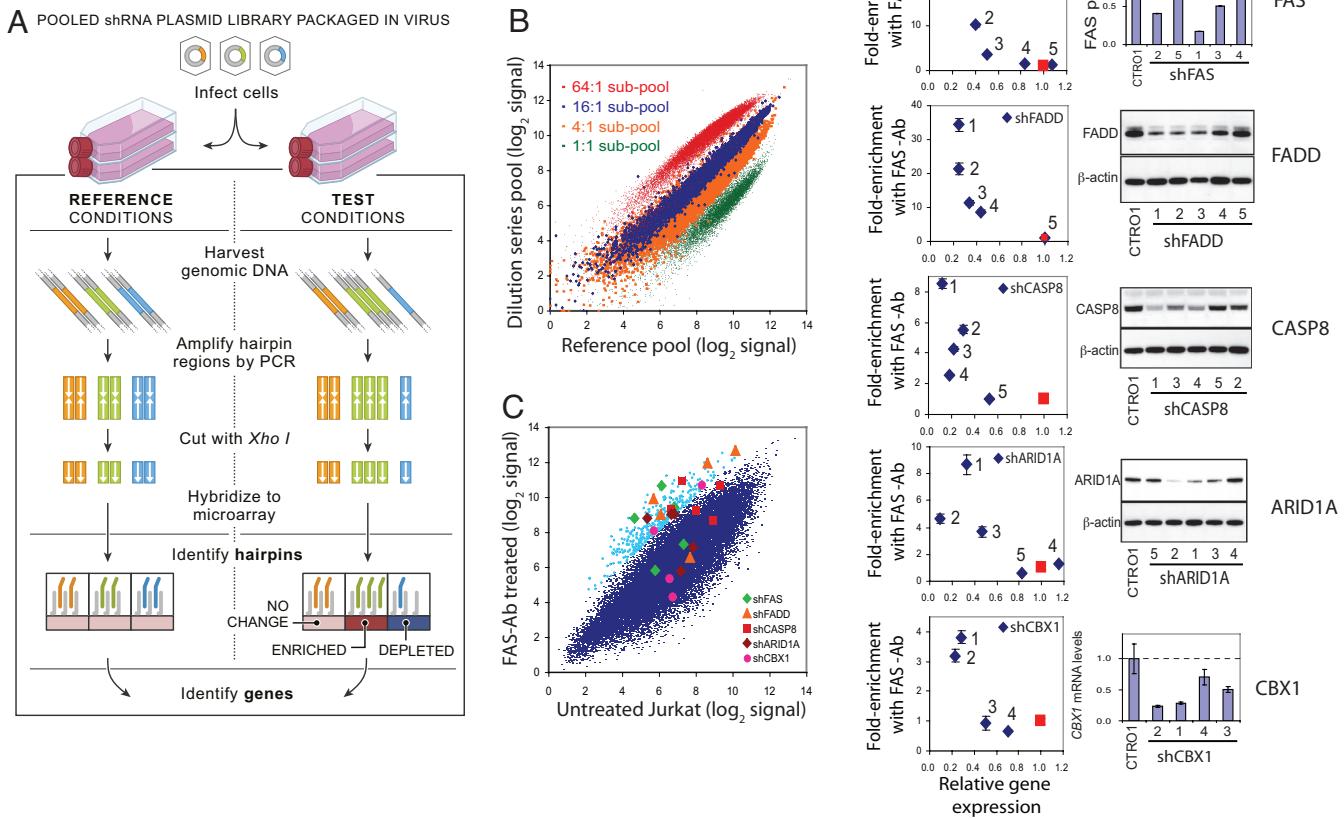
<sup>1</sup>B.L., H.W.C., and A.S. contributed equally to this work.

<sup>2</sup>N.H., W.C.H., E.S.L., D.M.S., and D.E.R. contributed equally to this work.

<sup>3</sup>To whom correspondence may be addressed. E-mail: lander@broad.mit.edu or droot@broad.mit.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0810485105/DCSupplemental](http://www.pnas.org/cgi/content/full/0810485105/DCSupplemental).

© 2008 by The National Academy of Sciences of the USA



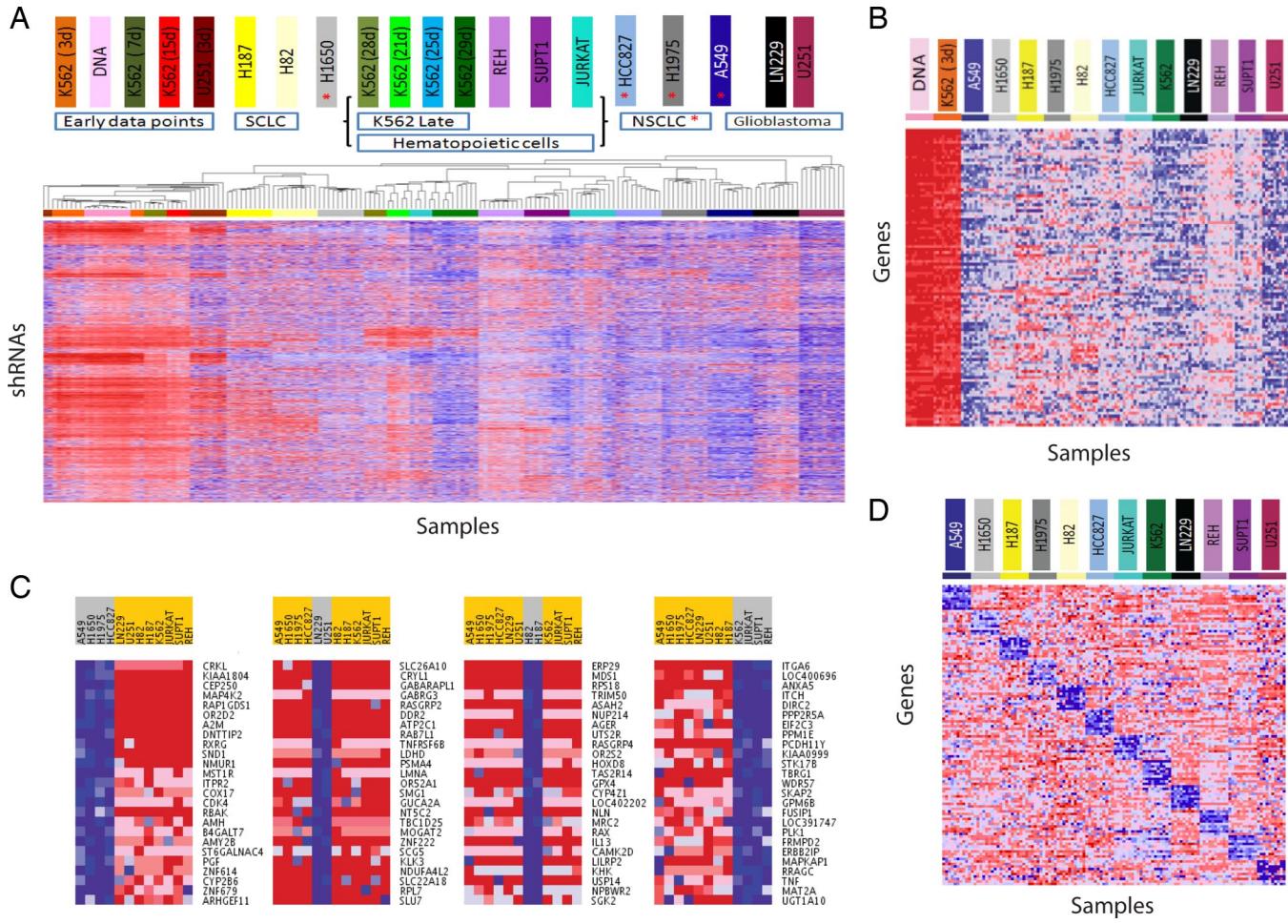
**Fig. 1.** Pooled RNAi screening strategy and performance using pools of 45,000 shRNA-expressing viruses. (A) Schematic of pooled shRNA library screens. (B) Performance evaluation of half-hairpin barcodes (hhbs) using pools containing known relative proportions of DNA. Two 45,000-shRNA pools were created by combining 4 subsets of the shRNA library plasmids (labeled in green, orange, blue, and red, each consisting of  $\approx 11,000$  different plasmids) in a 1:1:1:1 ratio of concentration for the “Reference pool” and in a 1:4:16:64 ratio for the “Dilution series pool.” To measure relative shRNA abundance in each pool, hhbs were hybridized to a custom Affymetrix barcode array. The observed separation of the 4 subsets of shRNAs according to their known relative proportions in the 2 pools illustrates the ability of hhbs to deconvolute the pooled shRNA library. (C) Primary screen results for genes required for FAS-induced apoptosis in Jurkat T cells. Cells were infected with the 45k pool viral library and cultured in the presence or absence of activating FAS antibody CH11 (FAS-Ab) for 3 weeks. Hybridization signals for hhbs amplified from the FAS-Ab treated group (average of 5 replicates) are plotted against those from the untreated group (average of 10 replicates). Array data for the 400 shRNAs (0.9% of pool) exhibiting highest enrichment in FAS-Ab treated group relative to untreated group are depicted in light blue. Array data for the shRNAs targeting the 5 hit genes are shown by distinct symbols. (D) Plot of target gene knockdown versus enrichment of shRNAs in FAS-treated samples for hit genes. FAS resistance was measured by relative proliferation rate of cells infected by individual candidate shRNA viruses (targeting FAS, FADD, CASP8, ARID1A, or CBX1) versus cells infected with a mixture of control shRNA viruses. Target gene suppression was measured by FACS (FAS), immunoblotting (FADD, CASP8, and ARID1A), or quantitative PCR (CBX1).

measured the overrepresentation of shRNAs in the surviving treated cells relative to untreated cells. To filter out shRNAs acting through off-target effects, we defined genes to be “hits” if at least 2 independent shRNAs against the gene were ranked in the top 0.9% of overrepresented shRNAs (Fig. 1C). There were 11 hits, of which 9 were confirmed by testing the shRNAs individually. We were able to reliably measure gene expression levels for 7 of these genes and found that 5 showed strong correlation between the level of resistance to FAS-induced apoptosis and the level of gene knockdown—confirming that the shRNA effect is “on-target” (Fig. 1D). The 5 genes include 3 with well-established roles in FAS-induced apoptosis (11, 13–15) (FAS, FADD, and CASP8) and 2 previously undescribed genes—ARID1A, a SWI/SNF chromatin remodeling complex component (16, 17), and CBX1, a chromatin silencing protein (18). For all 5 genes, the effective shRNAs also inhibited both FAS-induced CASPASE 8 cleavage and FAS-induced mitochondrial leakage (Fig. S2), indicating that, like the 3 known genes, the 2 previously undescribed genes act upstream of CASPASE 8 activation. The lack of downstream apoptosis genes among the hits could be due to false negative results (missed active

genes) or a true finding that stems from individual downstream genes not being absolutely required for apoptosis in these cells because of functional redundancy or the activation of compensatory processes.

The second positive selection screen sought to identify genes whose inhibition renders H82 small-cell lung cancer cells resistant to etoposide, a small molecule that alters the activities of topoisomerase IIA (TOPOIIA) (19, 20) and is used to treat small-cell lung and other cancers (21). By using a high dose of etoposide sufficient to eliminate unmodified H82 cells, 1 confirmed suppressor gene emerged: TOPOIIA itself (Fig. S3). Consistent with this observation, reduced TOPOIIA expression has been shown to confer etoposide resistance in SCLC lines (22). Together, these 2 positive selection screens demonstrate the utility of our approach in studying genes involved in cell viability.

We then turned to the more difficult challenge of identifying the genes that are essential for the proliferation of specific cancer cell lines, which involves the infection of cell lines with a pool of shRNAs and the identification of underrepresented shRNAs among surviv-



**Fig. 2.** Screens for essential genes in 12 cancer cell lines. The 45K pool viral library was used to infect 12 cancer cell lines in multiple replicates. Heat maps depict relative abundance of shRNAs, individually or combined by their gene target (red, high; blue, low). (A) Unsupervised hierarchical clustering of the hhb array data for 175 samples from screens of 12 cell lines (10 replicates per cell line for 4-week time points; 5 or 10 replicates for earlier time points, as noted), and the initial 45k shRNA DNA plasmid pool (10 replicates). The 10,117 shRNAs with the highest coefficient of variation in signal across all 175 samples ( $CV > 0.30$ ) were included in the clustering analysis. (B) Commonly essential genes. The average of “leading edge” shRNA signals for each of the top-100 commonly essential genes (requiring a minimum of 8 of 12 cell lines to contribute to the essentiality enrichment score) exhibits extensive depletion after 4 weeks. (C) Top cell lineage-specific essential genes for cell lines derived from: (i) 4 non-small-cell lung cancers, (ii) 2 glioblastomas, (iii) 2 small-cell lung cancers, and (iv) 4 leukemias. (D) Identification of cell line-specific essential genes based on relative shRNA depletion in 1 cell line versus the other 11 cell lines. Average signals for leading edge shRNAs for the top-10 specific essential genes for each cell line are displayed. *ABL1* and *BCR* are 1st and 5th best-scoring genes, respectively, in K562 cells.

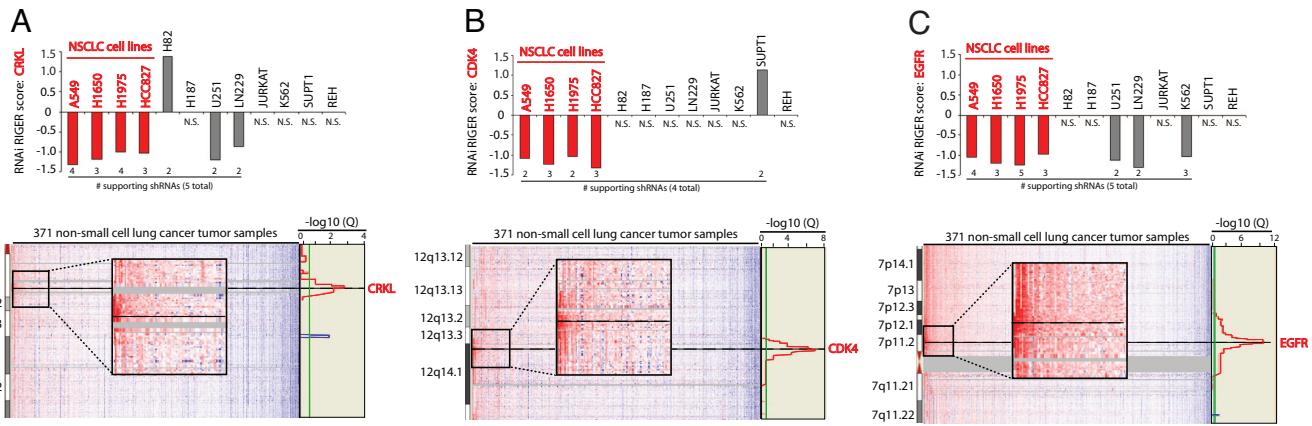
ing cells. These negative-selection screens require more precise quantification of shRNA abundance than positive-selection screens that seek to identify shRNAs that are dramatically overrepresented.

We performed negative-selection RNAi screens with the 45k shRNA pool in 12 cancer cell lines representing diverse cancer types, including small-cell lung cancer (H82, H187), non-small-cell lung cancer (A549, H1650, H1975, HCC827), glioblastoma (LN229, U251), CML (K562), and lymphocytic leukemia (Jurkat, SUPT1, REH). For each of the cell lines, we performed at least 10 independent infections and compared the abundance of each shRNA at  $\approx 28$  days after infection to the initial abundance in the DNA plasmid pool from which the lentiviral vectors were produced. For 2 of the cell lines (K562 (Fig. S4) and U251), we confirmed that the abundance at 3–4 days after infection was highly similar to the abundance in the plasmid pool, demonstrating that representation is preserved after viral packaging, viral transduction, and initial infection (Fig. 2A). In contrast, the representation is strikingly different at later time points (2–4 weeks), reflecting unequal survival of cells with different hairpins (Fig. S4 and Fig. 2A).

In total, we generated a database of 5.4 million measurements of the relative abundance of the 45,000 shRNAs across the 12 cell lines

and 10 replicates. Both unsupervised clustering and consensus clustering of these data clustered the replicates together, supporting the robustness of the results, and furthermore grouped the cell lines according to their developmental lineage (Fig. 2A and Fig. S5). To define genes as hits based on shRNA depletion data, we developed a statistic called an RNAi gene enrichment ranking (RIGER) score. Briefly, we examine the position of the 5 shRNAs targeting the gene in the full ranked list of the 45,000 shRNAs, assess whether the set is biased toward the top of the list based on a KS statistic, and calculate an enrichment score and gene ranking based on a permutation test (see *Materials and Methods*). The inclusion of all shRNAs targeting each gene increases the power of the screen, compensating for variation in gene suppression and off-target effects. We applied RIGER to each of the  $\approx 9,500$  genes, to identify the cancer-cell essential genes (Dataset S1).

The 12 cancer cell lines showed substantial correlation in their gene requirements for proliferation. For example, 530 genes ranked in the top 5% for essentiality in 5 or more cell lines, whereas only 2 genes would be expected if the cell lines were uncorrelated (Dataset S2). We identified “commonly essential” genes using a second application of RIGER to find genes enriched for essentiality



**Fig. 3.** Identification of known and putative oncogenes by integrating functional and structural genomics. RNAi RIGER scores for *CRKL* (A), *CDK4* (B), and *EGFR* (C) in each of the 12 cell lines relative to control and copy number changes in NSCLC tumors (26) at the loci encoding these genes. The number of shRNAs ranked in the leading edge of the RIGER analysis is noted. Two or more shRNAs for each gene were required to be in the RIGER leading edge to obtain a RIGER score for that gene; otherwise the RIGER result is labeled N.S. (no score). Significance of the observed copy number changes based on frequency and magnitude was calculated by using the GISTIC algorithm (41). False-discovery rates (red line,  $-\log_{10} Q$  values for amplification; blue line,  $-\log_{10} Q$  values for deletion; green line is 0.25 cutoff for significance) are depicted vertically along each chromosomal position.

among the 12 cell lines; we found 268 commonly essential genes with an FDR <25% (Fig. 2B and Dataset S3). Using gene-set enrichment analysis (GSEA) (23), we observed that the commonly essential genes showed a strong enrichment for certain molecular pathways including ribosomal proteins, mRNA processing and splicing, translation factors, and proteasome degradation (Dataset S4). For selected genes in these highly enriched pathways, we validated target specificity by comparing proliferation to target gene suppression for multiple shRNAs (Fig. S6).

In addition to these commonly essential genes, we identified “cell lineage-specific” essential genes (Fig. 2C and Dataset S5), which we defined as genes that exhibited a stronger phenotype in cell lines derived from a particular cancer type than in other cancer types. A total of 63 genes exhibited specific essentiality for the 4 non-small-cell lung cancer (NSCLC) cell lines that was significantly stronger than observed for randomly selected subsets of 4 cell lines ( $P < 0.05$ ). Similarly, 32 genes showed significant differential essentiality for the 4 lymphocytic and myelogenous leukemia lines. This type of analysis thus enables a systematic approach to identify an important class of genes that are differentially required for proliferation in a cancer-specific manner.

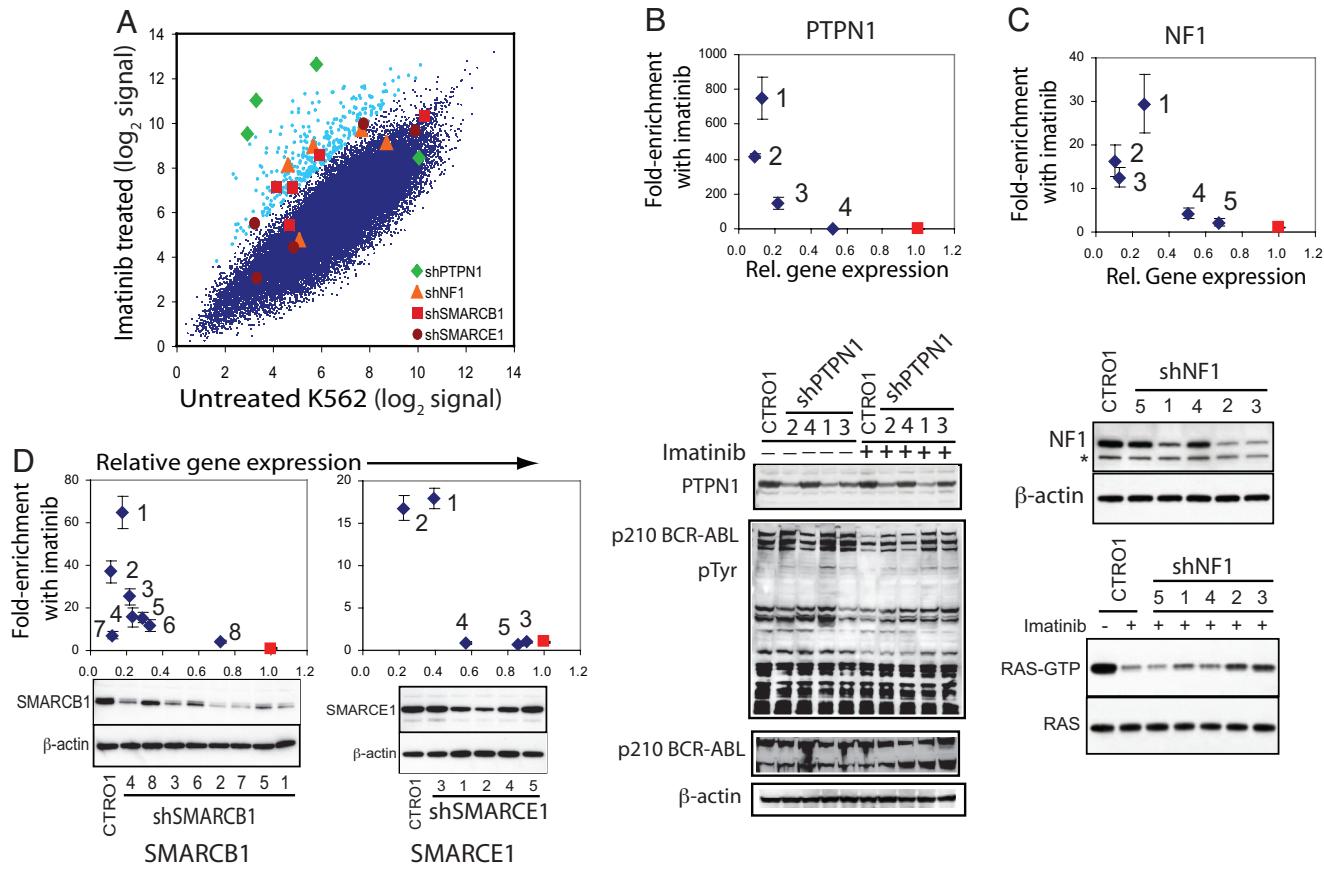
We also identified “cell line-specific” genes, which showed specific differential requirement in 1 cell line versus the other 11 (Fig. 2D, Fig. S7, and Dataset S6). Such genes can generate initial hypotheses about cancer-specific gene dependencies, but confirmation in additional cell lines would be required to define the cancer-specificity of these gene requirements. For example, chronic myelogenous leukemia is represented by only a single cell line (K562) among the 12. In K562 cells, we found that the 5 top-scoring genes included *ABL1* and *BCR* (ranked 1st and 5th, respectively, of  $\sim 9,500$  genes); these 2 genes are involved in the *BCR-ABL* translocation harbored by this cell line. We retested individually the 13 shRNAs against these genes, confirming that the inhibition is cell line-specific and that the level of inhibition is strongly correlated with the level of gene inhibition (Fig. S8). Thus, *ABL1* is readily identified as a selectively highly required gene in the K562 cell line, and in this positive-control case, we know that follow-up experiments would confirm this trait to be shared among CML cell lines, demonstrating the utility of this approach to identify bona fide oncogenes.

A particularly powerful way to characterize cancer cells may be to combine information about both structure (genomic mutation) and function (gene essentiality) to reveal oncogenes. Several recent studies have illustrated the ability to identify key cancer genes in this

manner (6, 24, 25). Indeed, when we searched for known oncogenes among the highest scoring genes in each cell line, we found several common oncogenes. For example, *KRAS*, *MYC*, and *MYB* were in the top 1% of essential genes in at least 1 cancer cell line (*KRAS* in LN229, A549, H1650, and H1975, and *MYC* and *MYB* in K562; Figs. S9 and S10). *KRAS* was found to be required in a *KRAS* mutant cell line, A549 (9th ranking gene, Dataset S1).

To extend this approach to genes resident in regions of copy number gain in human cancers, we intersected (i) the list of genes in regions of genomic amplification identified in a recent study of 371 NSCLC tumors (26) and (ii) the list of cell lineage-specific essential genes with strong preferential essentiality in the 4 NSCLC cell lines (Dataset S5). The top-scoring gene, *CRKL* ( $P = 0.010$ ; Fig. 3A), a member of an adapter protein family that activates the RAS signaling pathway (27); falls in one of the most significantly amplified regions in NSCLC 22q11.21, for which no oncogenes were previously known (26). We confirmed the essentiality of *CRKL* in A549 and H1975 NSCLC cells through: (i) a competitive cell survival experiment and (ii) experiments demonstrating that the level of *CRKL* knockdown was correlated with the level of growth inhibition (Fig. S11). The second-best-scoring gene was *CDK4* ( $P = 0.014$ ; Fig. 3B), which modulates the p16<sup>INK4a</sup>-cyclin D1-CDK4-RB growth regulatory pathway. This pathway is altered in the majority of NSCLCs, and high levels of *CDK4* are associated with tumor progression (26, 28, 29). The third-best-scoring gene was *EGFR* ( $P = 0.03$ ; Fig. 3C), a gene frequently amplified or mutated in NSCLC that has been successfully targeted by small-molecule inhibitors (26, 30–32). Although we screened only 4 NSCLC cell lines, the intersection of structural and functional data readily identifies 2 known oncogenes (*EGFR* and *CDK4*) and implicates an additional likely oncogene (*CRKL*) in human NSCLC. These observations suggest that the combination of large-scale structural and function data will accelerate the comprehensive identification of genes essential for the malignant state.

A further application of pooled shRNA screening is to perform suppressor and enhancer screens to identify genes that interact with known genes, pathways, and drugs. To test this approach, we screened for genes that modulate the response of CML cells to imatinib, a clinically approved inhibitor of *BCR-ABL* (33, 34). Such screens have the potential not only to identify genes that interact with *BCR-ABL* but also, importantly, highlight genes that may influence the development of imatinib resistance. We performed a positive-selection screen in which we infected K562 cells with the 45k pool, exposed these cells to a lethal dose of imatinib, and



**Fig. 4.** Screen for modifiers of the response to imatinib in K562 cells. K562 cells were infected with the 45k pool shRNA viral library and treated in the presence or absence of imatinib for 21 days (10 replicate infections for each group). (A) Averaged microarray hybridization signals for each shRNA in the imatinib-treated cell samples are plotted versus average hybridization signals for the untreated samples. The 400 shRNAs yielding the greatest resistance to imatinib are indicated in light blue. The shRNAs targeting 4 hit genes are labeled. (B-D) Knockdown validation of shRNAs conferring resistance to imatinib. The enrichment of shRNA-infected cells in response to imatinib was tested by coculturing GFP-labeled shRNA-infected cells with control cells for 3 weeks, followed by FACS analysis. Target gene knockdown by the shRNAs was determined by immunoblotting. (B) Cells infected with shPTPN1 were untreated or treated with imatinib, followed by immunoblotting for PTPN1, phosphotyrosine, ABL1 and  $\beta$ -actin. (C) Cells infected with shNF1 were treated with imatinib, followed by immunoprecipitation of GTP-bound RAS and immunoblotting for RAS. (D) Knockdown validation of shRNAs targeting *SMARCB1* and *SMARCE1*.

identified genes whose inhibition conferred survival (Fig. 4A). By using the same criteria as for the FAS-Ab modifier screen, 10 genes were identified as hits, 2 of which failed to be confirmed in tests of individual shRNAs. Target knockdown measurements were obtained for 7 of these genes, of which the shRNAs for 4: *PTPN1*, *NF1*, *SMARCB1*, and *SMARCE1* showed strong correlation between the level of resistance and the level of gene knockdown (Fig. 4B-D). One of these genes, *PTPN1*, has previously been reported to be a negative regulator of BCR-ABL signaling, because the expression of a dominantly interfering mutant of *PTPN1* rendered BCR-ABL-dependent cells resistant to imatinib (35-37). We found that shRNA-mediated inhibition of *PTPN1* leads to increased tyrosine phosphorylation of BCR-ABL in the presence or absence of imatinib (Fig. 4B Lower). Further confirming this finding, we also performed a separate screen to identify genes that permit cells to survive RNAi-mediated suppression of *BCR-ABL* and identified *PTPN1* as the top-scoring hit (Fig. S12). Among the other genes, *NF1* is a Ras GTPase that suppresses tumor formation by inhibiting ras activation (38, 39), and it is a tumor suppressor for both type 1 neurofibromatosis and juvenile myelomonocytic leukemia, a childhood leukemia with characteristics similar to CML (40). We found that shRNA-mediated inhibition of *NF1* partially restored levels of active RAS in imatinib-treated cells (Fig. 4C Bottom). Both *SMARCB1* and *SMARCE1* encode subunits of the SWI1/SNF5 matrix-associated actin-binding chromatin-remodeling complex

(17), and *SMARCB1* has been implicated as a tumor-suppressor gene in infantile malignant rhabdoid tumors and epithelioid sarcomas. These observations suggest a previously unrecognized role for this chromatin remodeling complex in imatinib-sensitivity of CML cells. Moreover, this screen suggests that this approach can be used to systematically identify genes and pathways that interact with a specific gene, pathway, or small-molecule perturbation.

## Conclusions

Extending the application of the experimental and analytical strategies described here to a much larger set of cancer cell lines will permit systematic discovery of genes involved in cancer cell proliferation and survival. The inclusion of 5 independent shRNAs targeting each human gene in this shRNA library provides power to discriminate specific from off-target effects in the primary screen and different levels of on-target knockdown, whereas the RIGER algorithm provides the means to rank genes based on these multiple shRNAs. Increasing the number of shRNAs available per gene and measuring the knockdown performance of each shRNA will further improve both the sensitivity of this approach to detect hit genes and the ability to discriminate against off-target effects. Although we have assessed only a single phenotype (proliferation) in a limited number of cell lines, this method may be applied to other phenotypes and cell types including more “normal” cultured

cells. We anticipate that systematic efforts to apply these approaches to study other cancer phenotypes will eventually lead to a more complete view of the Achilles' heels of different types of cancers. Our initial efforts suggest that such studies can be performed at a relatively modest cost, although they will require larger, validated shRNA libraries than we are currently generating, and an extensive collection of cell lines.

When combined with the increasingly complete structural analyses of cancer genomes by The Cancer Genome Atlas and other such efforts, the experimental and analytical strategies for pooled shRNA screens described herein provide a feasible strategy to systematically identify the key genes involved in cancer initiation, maintenance, and progression and likely targets for therapeutic intervention. Moreover, although we have used cancer cell proliferation to develop and validate these methodologies, the broad application of these approaches in other experimental contexts promises to provide insights into a wide range of biological phenotypes in mammalian cells.

## Materials and Methods

A genome-scale pooled shRNA library of 45,000 shRNAs in viral vectors (45k pool) was produced from the sequence-validated arrayed TRC shRNA library

and used for all of the screens reported here. The shRNA representation of the library was measured by using the half of the shRNA sequence as a molecular barcode (a "half-hairpin barcode", hhb), which was obtained by restriction enzyme digestion of PCR-amplified shRNA sequences from library-infected cells. The hhb representation was assessed by hybridizing the hhbs to a high-density Affymetrix custom microarray. The shRNA hhb hybridization data were preprocessed with modified Dchip software, and analyzed by using the RIGER algorithm. These computational analysis tools, dCHIP for RNAi and RIGER, are available online at [http://www.broad.mit.edu/rnai\\_analysis](http://www.broad.mit.edu/rnai_analysis). Detailed methods for all experiments are provided in *SI Methods*. For primers used in SYBR assays and TaqMan probes, see **Table S1**. For a key to the shRNA labels used in the figures, see **Dataset S7**. For analyses used to assess essential genes, see *SI Methods* and **Scheme S1**.

**ACKNOWLEDGMENTS.** We thank Preeti Gupta and Mike Mittman for assistance with Affymetrix microarray design; Jinyan Du, Casey Gates, Leigh Brody, Nathan Berkowitz, Neelum Khattak, Daniel Lam, Brian Wong, Jordi Barretina, and Cindy Nguyen for technical assistance and materials; and Pablo Tamayo, Preeti Gupta, Mike Mittman, Serena Silver, Jennifer Grenier, and Glenn Cowley for helpful discussions. This work is a project of The RNAi Consortium (TRC), financially supported by past and present consortium members: Academia Sinica, Bristol-Myers Squibb, Broad Institute, Eli Lilly, Novartis, Ontario Institute for Cancer Research, and Sigma-Aldrich. H.W.C. was supported by the Croucher Foundation.

- Schlabach MR, et al. (2008) Cancer proliferation gene discovery through functional genomics. *Science* 319:620–624.
- Silva JM, et al. (2008) Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science* 319:617–620.
- Brummelkamp TR, et al. (2006) An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. *Nat Chem Biol* 2:202–206.
- Berns K, et al. (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 428:431–437.
- Berns K, et al. (2007) A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* 12:395–402.
- Westbrook TF, et al. (2005) A genetic screen for candidate tumor suppressors identifies REST. *Cell* 121:837–848.
- Kolfschoten IG, et al. (2005) A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell* 121:849–858.
- Ngo VN, et al. (2006) A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* 441:106–110.
- Moffat J, et al. (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124:1283–1298.
- Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM (2006) Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat Methods* 3:715–719.
- Itoh N, et al. (1991) The propeptide encoded by the cDNA for human cell surface antigen Fas can mediate apoptosis. *Cell* 66:233–243.
- Yonehara S, Ishii A, Yonehara M (1989) A cell-killing monoclonal antibody (anti-Fas) to a cell surface antigen co-downregulated with the receptor of tumor necrosis factor. *J Exp Med* 169:1747–1756.
- Boldin MP, Goncharov TM, Goltsev YV, Wallach D (1996) Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell* 85:803–815.
- Chinnaiyan AM, O'Rourke K, Tewari M, Dixit VM (1995) FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. *Cell* 81:505–512.
- Muzio M, et al. (1996) FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* 85:817–827.
- Dallas PB, et al. (1998) p300/CREB binding protein-related protein p270 is a component of mammalian SWI/SNF complexes. *Mol Cell Biol* 18:3596–3603.
- Wang W, et al. (1996) Diversity and specialization of mammalian SWI/SNF complexes. *Genes Dev* 10:2117–2130.
- Furuta K, et al. (1997) Heterochromatin protein HP1Hsbeta (p25beta) and its localization with centromeres in mitosis. *Chromosoma* 106:11–19.
- Nitiss JL, Liu YX, Hsiung Y (1993) A temperature sensitive topoisomerase II allele confers temperature dependent drug resistance on amsacrine and etoposide: A genetic system for determining the targets of topoisomerase II inhibitors. *Cancer Res* 53:89–93.
- Osheroff N (1989) Effect of antineoplastic agents on the DNA cleavage/religation reaction of eukaryotic topoisomerase II: Inhibition of DNA religation by etoposide. *Biochemistry* 28:6157–6160.
- Hande KR (1998) Etoposide: Four decades of development of a topoisomerase II inhibitor. *Eur J Cancer* 34:1514–1521.
- Andoh T, et al. (1996) Reduced expression of DNA topoisomerase II confers resistance to etoposide (VP-16) in small cell lung cancer cell lines established from a refractory tumor of a patient and by in vitro selection. *Oncol Res* 8:229–238.
- Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545–15550.
- Boehm JS, et al. (2007) Integrative genomic approaches identify IKBKE as a breast cancer oncogene. *Cell* 129:1065–1079.
- Firestein R, et al. (2008) CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* 455:547–551.
- Weir BA, et al. (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450:893–898.
- Senechal K, Halpern J, Sawyers CL (1996) The CRKL adaptor protein transforms fibroblasts and functions in transformation by the BCR-ABL oncogene. *J Biol Chem* 271:23255–23261.
- Wikman H, et al. (2005) CDK4 is a probable target gene in a novel amplicon at 12q13.3-q14.1 in lung cancer. *Genes Chromosomes Cancer* 42:193–199.
- Zhao X, et al. (2005) Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res* 65:5561–5570.
- Lynch TJ, et al. (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 350:2129–2139.
- Paez JG, et al. (2004) EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* 304:1497–1500.
- Sordella R, Bell DV, Haber DA, Settleman J (2004) Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* 305:1163–1167.
- Deininger M, Buchdunger E, Druker BJ (2005) The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* 105:2640–2653.
- Wong S, Witte ON (2004) The BCR-ABL story: Bench to bedside and back. *Annu Rev Immunol* 22:247–306.
- Koyama N, et al. (2006) Inhibition of phosphotyrosine phosphatase 1B causes resistance in BCR-ABL-positive leukemia cells to the ABL kinase inhibitor ST1571. *Clin Cancer Res* 12:2025–2031.
- LaMontagne KR, Jr, Flint AJ, Franza BR, Jr, Pandergast AM, Tonks NK (1998) Protein tyrosine phosphatase 1B antagonizes signalling by oncoprotein tyrosine kinase p210 bcr-abl in vivo. *Mol Cell Biol* 18:2965–2975.
- LaMontagne KR, Jr, Hannon G, Tonks NK (1998) Protein tyrosine phosphatase PTP1B suppresses p210 bcr-abl-induced transformation of rat-1 fibroblasts and promotes differentiation of K562 cells. *Proc Natl Acad Sci USA* 95:14094–14099.
- Xu GF, et al. (1990) The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell* 62:599–608.
- Basu TN, et al. (1992) Aberrant regulation of ras proteins in malignant tumour cells from type 1 neurofibromatosis patients. *Nature* 356:713–715.
- Side LE, et al. (1998) Mutations of the NF1 gene in children with juvenile myelomonocytic leukemia without clinical evidence of neurofibromatosis, type 1. *Blood* 92:267–272.
- Beroukhim R, et al. (2007) Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc Natl Acad Sci USA* 104:20007–20012.

# Supporting Information

Luo et al. 10.1073/pnas.0810485105

## SI Methods

**Construction of Pooled shRNA Library.** The 45k pool of 45,182 shRNA plasmids from the TRC library was assembled from subpools made in 2 different ways: (i) For 20,053 (44%) of the shRNAs, first, equal amounts of normalized purified plasmid DNA were combined into 6 pools of  $\approx$ 3,300 plasmids. Each of these 3,300-plasmid pools was used to transform ElectroMAX DH5 $\alpha$ -E cells (Invitrogen) by electroporation, which were then plated onto 5 square bioassay dishes (24  $\times$  24 cm; Nunc). DNA was purified from the plated transformants using a HiSpeed Plasmid Maxi kit (Qiagen). (ii) For 25,129 (56%) of the shRNAs, the  $\approx$ 85 bacterial clones in each 96-well library plate were pooled and DNA was purified from these  $\approx$ 85-clone bacterial pools (Qiagen Qiaprep Spin Miniprep kit). These DNA preparations were combined in equal concentrations to form 7 pools of  $\approx$ 3,600 plasmids, which were then transformed, amplified and purified as in method (i). The 3,300- and 3,600-plasmid pools made by methods (i) and (ii) were then combined to create the 45k library DNA, which was used to transform bacteria. DNA purified from the plated transformants was used for virus production.

**Virus Pool Production, Infection, and Cell Propagation.** The 45k plasmid pool (50  $\mu$ g), along with 50  $\mu$ g of pCMV-dR8.91 and 10  $\mu$ g of pMD.G packaging plasmids (alternatively, the packaging plasmids pCMV-dR8.74psPAX2 and pMD2.G, respectively, function equally well for viral packaging with the pLKO.1 construct), was transfected into each of multiple T175 flasks of 293T cells. Virus with a titer of  $1 \times 10^7$  infectious units/ml from 48- and 72-h harvests after transfection was pooled, aliquoted, and stored at  $-80^{\circ}\text{C}$ .

To perform large-scale infections,  $3.6 \times 10^7$  target cells for each replicate were resuspended in 24 ml of culture medium containing 4  $\mu\text{g}/\text{ml}$  polybrene. The 45k library lentivirus was added in appropriate volume to achieve an MOI of 0.3. This mixture was split across a 12-well plate at 2 ml per well. A spin infection was performed by centrifugation at  $930 \times g$  for 2 h at  $30^{\circ}\text{C}$ .

For suspension cells, the 12 wells of each replicate plate were then pooled and centrifuged at  $335 \times g$  for 5 min. The supernatants were aspirated, and each pellet was resuspended in 200 ml of culture medium and added to a T175 flask. After 1 or 2 days, puromycin was added to the infected cells. Passaging was performing by transferring 30 ml of high-density cells into 200 ml of new medium containing puromycin. The remaining cells were centrifuged, resuspended in 0.5 ml PBS, and stored at  $-20^{\circ}\text{C}$  for genomic DNA purification.

For adherent cells, the supernatants of the 12 wells of each plate were aspirated after spin infection. Two milliliters of culture medium were added to each well and the cells cultured overnight. The next day, cells of each plate were trypsinized, pooled, and resuspended in 100 ml of growth medium for culture in 2 T175 flasks. One or 2 days after infection, puromycin was added to the infected cells. For every passage, 1/4 of the confluent cells were passaged into new flasks for continued culture, and 3/4 of the cells were harvested by centrifugation, resuspended in 0.5 ml of PBS, and stored at  $-20^{\circ}\text{C}$  for subsequent genomic DNA purification.

**Modifier Screens with Imatinib, Etoposide, and FAS-Induced Apoptosis.** The 45k library-infected cells were selected with puromycin. Five days after infection, half of the infected cells were treated

with perturbagens at each weekly passage, for 21 days; the other half was untreated at each passage. Final harvests of the infected cells were used for analysis. For the imatinib modifier screen, K562 cells were treated with 125 nM imatinib (Novartis). At this dose, cell numbers were depleted versus untreated control cells by 90% over the first 7-day passage, and by an additional  $>99\%$  in each of the following 2 weeks. For the etoposide screen, H82 cells were treated with 1  $\mu\text{g}/\text{ml}$  (1.7  $\mu\text{M}$ ) etoposide (Sigma-Aldrich). At the screening dose,  $\gg 99\%$  of all of the cells were killed within the first 7 days. One-half of the screening dose was sufficient to decrease cell numbers versus no-treatment control by  $>99\%$  during each 7-day passage. For the apoptosis screen, Jurkat cells were treated with 1.6 ng/ml activating FAS antibody CH11 (Upstate Biotechnology). At this dose, cell numbers were depleted by  $>99\%$  over the first 7 days. One-eighth of the screening dose produced a  $\approx 90\%$  reduction in cell number versus control during each 7-day passage. One-quarter of the screening dose produced a  $>99\%$  reduction in cell number versus control during each 7-day passage.

**Design of Affymetrix Half-Hairpin Barcode Microarray.** An Affymetrix microarray capable of interrogating 110,000 shRNAs was designed. Three probes targeted each 21-base subsequence of a 23-base target sequence that included the 21-base sense-strand sequence of each shRNA along with the base immediately flanking each side. The 3 probes targeting each shRNA were randomly distributed across the array.

**Purification of Genomic DNA from Harvested Cells.** Harvested cells were resuspended in PBS and lysed according to the QIAamp Blood Maxi Kit protocol (Qiagen). DNA was precipitated and purified with the QIAamp Maxi column. DNA was eluted by adding 500  $\mu\text{l}$  of Buffer AE to the membrane, incubating at  $25^{\circ}\text{C}$  for 5 min, and centrifuging at  $3,273 \times g$  for 2 min; then by adding an additional 200  $\mu\text{l}$  of Buffer AE to the membrane, incubating at  $4^{\circ}\text{C}$  for 16 h, and centrifuging at  $3,272 \times g$  for 2 min. The 2 eluates were pooled and stored at  $4^{\circ}\text{C}$ .

**Half-Hairpin Barcode Production.** The hairpin region of purified genomic DNA was amplified in PCR reactions containing 1  $\mu\text{M}$  biotinylated 5' primer [5'-BioAATGGACTATCATATGCT-TACCGTAACCTGAA-3'], 1  $\mu\text{M}$  3' primer [5'-TGTGGAT-GAATACTGCCATTGTCTCGAGGTC-3'], 200  $\mu\text{M}$  of each dNTP (TaKaRa), 1  $\times$  Ex Taq buffer (TaKaRa), 22.5 units of Ex TaqDNA polymerase (TaKaRa), and 30  $\mu\text{l}$  of genomic DNA template in a total reaction volume of 300  $\mu\text{l}$ . Thermal cycler PCR conditions consisted of heating samples to  $95^{\circ}\text{C}$  for 5 min; 35 cycles of  $94^{\circ}\text{C}$  for 30 sec,  $50^{\circ}\text{C}$  for 30 sec, and  $72^{\circ}\text{C}$  for 1 min; and  $72^{\circ}\text{C}$  for 10 min. Immediately after the first round of PCR amplification, reaction volumes were nearly doubled with the addition of another 270  $\mu\text{l}$  of PCR mixture, comprised as above except without DNA template. A second round of PCR amplification was performed by heating to  $95^{\circ}\text{C}$  for 7 min,  $55^{\circ}\text{C}$  for 2 min, and  $72^{\circ}\text{C}$  for 1 h.

Amplified hairpin DNA was digested into half-hairpins by adding 500 units of XhoI restriction enzyme and 600  $\mu\text{l}$  of 1  $\times$  NE Buffer 2 (New England Biolabs), followed by incubation at  $37^{\circ}\text{C}$  for 5 h to overnight. Digested DNA was purified by using a QIAquick PCR Purification kit (Qiagen) and eluted in 40  $\mu\text{l}$  of 0.2  $\times$  Buffer EB (Qiagen).

**Half-Hairpin Barcode Hybridization.** Half-hairpin targets combined with 1.5  $\mu$ M blocking primer [5'-GTCCTTCCACAA-GATATATAAAGCCAAGAAATCGAAATA-3'] were heated to 99 °C for 5 min, then 45 °C for 5 min. After incubation, this half-hairpin solution was added to a hybridization solution containing a final concentration of 0.1 $\times$  Fragmentation Buffer (Affymetrix), 150 pM Control Oligonucleotide B2 (Affymetrix), 1 $\times$  Eukaryotic Hybridization Controls (Affymetrix), 0.1 mg/ml Herring Sperm DNA (Promega), 0.5 mg/ml BSA, 1 $\times$  Hybridization Buffer (100 mM Mes, 1 M NaCl, 20 mM EDTA, 0.01% Tween-20), and 10% DMSO in a total volume of 300  $\mu$ l. This half-hairpin hybridization mixture was heated to 99 °C for 5 min, 45 °C for 5 min, and centrifuged for 5 min at 15,000  $\times$  g before hybridization to a TRCBCx520397F custom GeneChip microarray (Affymetrix) for 16 h at 40 °C.

GeneChip staining reagents were prepared according to the Affymetrix protocol for eukaryotic arrays. GeneChips were washed and stained by using the GeneChip Fluidics Station 450 (Affymetrix), following the FlexGE\_WS2v5 protocol with the temperature for Post Hyb Wash #2 with Stringent Wash Buffer B modified to 30 °C. GeneChips were scanned by using the GeneChip Scanner 3000 (Affymetrix).

**The 60-mer Barcode Hybridization.** The 752 reference pool hybridization targets were generated by combining the PCR products of reference pool DNA template PCR-amplified by using a biotinylated primer and dilution series pool DNA template PCR-amplified by using unlabeled primers as competitors. Dilution series pool hybridization targets were generated by combining the PCR products of dilution series pool DNA template PCR-amplified by using a biotinylated primer and reference pool DNA template PCR-amplified by using unlabeled primers as competitors. Each target mixture was heated to 95 °C for 10 min, then placed on ice for 10 sec before being added to a hybridization solution containing a final concentration of 0.1 $\times$  Fragmentation Buffer (Affymetrix), 150 pM Control Oligonucleotide B2 (Affymetrix), 1 $\times$  Eukaryotic Hybridization Controls (Affymetrix), 0.1 mg/ml Herring Sperm DNA (Promega), 0.5 mg/ml BSA, 1 $\times$  Hybridization Buffer (100 mM Mes, 1 M NaCl, 20 mM EDTA, 0.01% Tween-20), and 10% DMSO in a total volume of 200  $\mu$ l. Hybridization cocktails were heated to 99 °C for 5 min, 45 °C for 5 min, and centrifuged for 5 min at 15,000  $\times$  g before hybridization to TRC custom GeneChip microarrays (Affymetrix) for 16 h at 45 °C.

GeneChip SAPE Stain Solution was prepared according to the Affymetrix protocol for eukaryotic arrays. Antibody solution was substituted with 1 $\times$  Stain Buffer (100 mM Mes, 1 M NaCl, 0.05% Tween-20) and 2 mg/ml BSA. GeneChips were washed and stained by using the GeneChip Fluidics Station 450 (Affymetrix), following the EukGE\_WS2v5 protocol. GeneChips were scanned by using the GeneChip Scanner 3000 (Affymetrix).

**Preprocessing of Microarray Data with Modified dChip Software.** Half-hairpin barcode microarray data were analyzed by using a modified version of dChip software (1). After invariant set normalization was performed at the probe level, the average of the 3 perfect-match probes was used to represent the “shRNA signal.”

**Clustering Analysis.** We used hierarchical clustering to visualize the similarity between the cell lines screened. The hhb array hybridization data for 175 samples from 12 cell lines was filtered to remove constructs with low variation ( $CV < 0.3$ ) across the dataset. The 10,117 shRNAs with the highest coefficient of variation across all 175 samples were hierarchically clustered (2) by using the Pearson correlation metric.

To assess robustness of the observed clustering pattern in the 175-sample dataset, we used a resampling-based consensus

clustering approach (3). A total of 100 resamplings of the dataset were performed with 80% of the 175 samples included in each resampling. The consensus matrix counts the proportion of resamplings in which the 2 samples are clustered together.

**RNAi Gene Enrichment Ranking (RIGER).** To enrich for on-target genes in the primary screen, we developed a statistical approach that considers the phenotypic results for the multiple shRNAs targeting the same gene to determine RIGER. The inclusion of, on average, 5 shRNAs for each gene targeted by the TRC library greatly increases the power of the screen, mitigating inherent shRNA properties such as variable degree of gene suppression and off-target effects. RIGER is based on the GSEA methodology (4) and uses similar Kolmogorov-Smirnov (KS)-based statistics to calculate gene scores from a dataset of shRNA construct profiles. It considers the entire list of shRNAs, and thus does not depend on an arbitrary threshold; it is nonparametric, i.e., it does not assume any particular distribution such as a normal distribution; and it captures more information about the shRNA subset distribution than a mean or median (with contribution from all moments of the distribution). It also permits a weighting of the tails of the distribution in proportion to effect. The output of RIGER is a rank ordered list of genes, based on the depletion or enrichment of the shRNAs that target them.

**The RIGER Methodology.** The RIGER methodology proceeds through the following steps:

1. Feature selection: shRNAs are scored according to their differential effects between 2 classes, early time-point samples and late time-point samples. We used the signal-to-noise metric (5) to quantitate this differential effect.
2. Calculation of an Raw Enrichment Score: Enrichment scores are calculated in the same manner as for the GSEA method (4).
3. Calculation of a RIGER score: The raw ES values were normalized to account for variable numbers of shRNAs across different genes by dividing the raw ES by the directional mean of a size-matched null distribution generated by 100,000 random permutations of a hairpin set of the same size. Genes with insufficient support, i.e., lower than desired number of shRNAs in the “leading edge” of the subset distribution that contribute to the NES score, were filtered. The support requirement was set to 2 shRNAs.

**Description of Selected RIGER Output.** The output of the RIGER software includes a list of the genes sorted by their RIGER scores. The fields in the output are:

NAME: The Entrez Gene symbol of the targeted shRNA gene.  
#HAIRPINS: The number of shRNA constructs targeting that gene that were included in the current experiment.

ES: The enrichment score for the “shRNA construct set” calculated using the weighted-KS statistic. This is a measure of the degree to which these hairpins are overrepresented at the top or bottom of the ranked list of hairpins in the dataset.

RIGER\_SCORE: The normalized enrichment score of the “shRNA construct set” for a given gene. Positive scores indicate that the constructs are overall positively correlated with the phenotype (e.g., in a negative selection experiment that is testing early cells vs. late infection, these constructs would be lethal). Genes with insufficient support are set to have a score of 0.

RIGER\_RANK: The rank of the gene compared with all other genes with RIGER scores in the same direction. The rankings are computed separately for positive and negative RIGER scores.

**SUPPORT % (or #SUPPORT):** The percentage (or number) of shRNA constructs before the peak in the running enrichment score  $S$ . The larger the percentage, the more constructs contribute to the final enrichment score.

**HAIRPIN\_RANKS:** The specific ranks of the constructs for a gene in the rank-ordered construct list.

**HAIRPIN\_SIGNAL-TO-NOISE:** The specific signal-to-noise scores of the constructs for a gene in the rank-ordered construct list

**HAIRPIN\_FOLD\_CHANGE:** The specific fold changes of the constructs for a gene in the rank-ordered construct list.

**Computing the LateVsControl S/N Matrix from the shRNA signal.** Early time-point samples ( $n = 10$ ) and DNA control samples ( $n = 10$ ) were compared with end-point (4 week) samples from each of the 12 cell lines used in this study. Hairpin signals correlated with the early vs. late distinction for each cell line were identified by sorting all hairpins in the dataset according to their signal-to-noise statistic:  $(\text{MEDIAN}_{\text{class}0} - \text{MEDIAN}_{\text{class}1}) / (\text{STD}_{\text{class}0} + \text{STD}_{\text{class}1})$ , where MEDIAN and STD are the median and standard deviation of the array values.

**Essential Gene Analyses.** The analyses used to assess essential genes are described below and diagrammed in Scheme S1.

**Cell-Line Essential Gene Analysis.** After calculation of shRNA signal changes, the rank-ordered shRNA list obtained in the *LateVsControl S/N Matrix* was then processed by RIGER to find those genes with at least 2 shRNAs significantly overrepresented at the extremes.

**Commonly Essential Gene Analysis.** To find genes that were frequently lethal across multiple cell lines, we combined all 12 “cell-line essential” RIGER gene score lists into a single list of scores for 9,423 genes times 12 cell lines and resorted (by RIGER score). We then searched for genes consistently essential in this composite list by using a second application of RIGER to find genes that are overrepresented at the top of the list. We report those genes with “leading edge support” of at least 8 of 12 cell lines.

**Cell-Specific and Cell Lineage-Specific Essential Gene Analyses.** To identify genes that exhibit unusually high essentiality in some but not all cell lines, we applied the following approach.

We first standardized the normalized score for each shRNA according to:  $X_{ij} = (x_{ij} - \text{med}_j) / \text{mad}_j$ , where  $x_{ij}$  is the S/N for the  $i$ th shRNA in the  $j$ th cell line, med is median and mad is median absolute deviation, to obtain the ZMAD S/N matrix. This standardization put all hairpins on a normalized scale, and facilitated comparisons across hairpins. Next, the resulting rank ordered list obtained from sorting these normalized values was processed by RIGER. Thus the initial matrix of 44,961 shRNAs  $\times$  12 cell line scores was transformed into a 9,423 genes  $\times$  12 cell line specificity matrix of RIGER scores. We then analyzed this cell line specificity matrix by applying a class vector to find genes that correlate with a particular phenotype. To identify genes that are essential to the NSCLC cell lines, we applied a 4 (NSCLC) vs. 8 (non-NSCLC) class vector and scored correlated genes using signal-to-noise.

**P Value Calculations for CRKL, CDK4, EGFR.** We used a standard phenotype permutation test (5) to assess the statistical significance of the correlation between CRKL and the 4 NSCLC cell lines. Specifically, a null distribution of signal-to-noise scores was created from the 495 possible permutations of 4 vs. 8 cell lines. We computed the signal-to-noise score of CRKL for each random grouping. The observed score of CRKL (for the real data labels) was compared with the null scores to obtain the

nominal  $P$  values. A similar procedure was followed to assess the significance of the correlation between CDK4 or EGFR and the 4 NSCLC cell lines.

**High-Throughput Lentivirus Production for Validation.** Plasmid DNA for hairpins of interest was rearranged in 96-well plates, and high-titer shRNA-expressing lentiviruses were generated robotically by using the high-throughput method described previously (6). A pool of 85 control shRNAs (CTR01pool) targeting reporter genes (GFP, RFP, luciferase, and lacZ) was used to generate control lentiviruses.

**Cell Viability Assay with pLKO-GFP-shRNAs.** Hairpins of interest were recloned into a modified version of pLKO.1 coexpressing GFP to generate pLKO-GFP-shRNA plasmids. For infections, K562 or Jurkat cells were seeded at a density of  $1 \times 10^4$  cells per  $100 \mu\text{l}$  of media containing  $4 \mu\text{g}/\text{ml}$  polybrene into each well of 96-well plates (Costar 8795BC) using the MicroFill microplate dispenser (BioTek). Two or  $5 \mu\text{l}$  of lentivirus for Jurkat or K562 cells were added, respectively, to transduce  $\approx 50\%$  of the cells, and the plates were spun at  $930 \times g$  for 2 h at  $37^\circ\text{C}$ . After infection, media were aspirated gently, and cells were resuspended in  $200 \mu\text{l}$  of fresh media. Cells were passaged every 2–3 days to allow for optimal cell proliferation. Fractions of cell suspensions were taken for FACS analysis at different time points (3, 5, 7, 11, 13, and 20 days after infection) by using the BD FACSCalibur flow cytometry system equipped with a high-throughput sampler (BD Bioscience). Control infections were also performed on the same plate by using several dilutions (20, 10, 5, or  $2 \mu\text{l}$ ) of pLKO-GFP vector control virus. Cell viability was presented as the fold change between the fraction of  $\text{GFP}^+$  cells at 3 days and at 20 days after infection. Data represented the mean values from 3 separate infections in 2 experiments.

**Validating of Modifier Screen Proliferation Phenotype by Using a Coculture Assay.** Reference cell lines stably expressing GFP and the puromycin resistance gene were generated by transduction with pRRRL-PGK-GFP and pLKO-puro viruses, followed by both FACS sorting for GFP positivity and selection for puromycin resistance. Cell lines stably expressing an shRNA of interest or CTR01 pool were established by spin infection ( $930 \times g$  for 2 h at  $30^\circ\text{C}$ ) of  $1 \times 10^6$  cells per well of 24-well plates with lentiviruses, followed by selection with puromycin (at concentrations of  $1.5 \mu\text{g}/\text{ml}$  for K562,  $0.75 \mu\text{g}/\text{ml}$  for Jurkat, and  $2 \mu\text{g}/\text{ml}$  for H82 cells) for 5 days. Mixtures of shRNA-expressing cells and GFP-expressing reference cells were prepared and then were either untreated or treated with corresponding perturbagens; K562 cells were treated with  $0.125 \mu\text{M}$  imatinib, Jurkat cells were treated with  $1.6 \text{ ng}/\text{ml}$  of activating Fas antibody CH11, and H82 cells were treated with  $1 \mu\text{g}/\text{ml}$  etoposide. Cocultured cells were passaged weekly. After 3 weeks of coculture, the percentage of  $\text{GFP}^+$  cells was measured by FACS analysis using a BD FACSCalibur flow cytometry system. The ratios of shRNA-expressing cells to GFP-expressing reference cells were determined for both untreated and treated mixtures and normalized to CTR01. The fold change of this ratio between the 2 mixtures is reported as the fold of enrichment by the perturbagen. Data represent the mean values from 2 experiments.

**Key to shRNA Labels and Sequences.** Dataset S7 provides full information for shRNAs referenced in the figures and text.

**Lentiviral Infection for Gene Knockdown Validation.** K562, Jurkat, or H82 cells ( $1 \times 10^6$ ) were seeded into each well of 24-well plates and spin-infected with shRNA-expressing lentiviruses, along with  $4 \mu\text{g}/\text{ml}$  polybrene, by centrifugation at  $930 \times g$  for 2 h at  $30^\circ\text{C}$ . After infection, virus-containing media were aspirated gently and cells were resuspended in  $5 \text{ ml}$  of fresh media and

grown in 6-well plates. Twenty-four hours after infection, cells were selected in puromycin for 5 days. Mock infections without addition of virus were also treated with puromycin to ensure complete killing of uninfected cells. Cell pellets were collected by centrifugation at  $524 \times g$  rpm for 5 min for protein or total RNA extraction.

**Western Blotting.** Cell lysates were prepared by suspending cell pellets in lysis buffer [50 mM Tris-HCl (pH 8), 150 mM NaCl, 1% Nonidet P-40, 0.5% deoxycholate, and 0.1% SDS] containing Complete proteinase inhibitors (Roche) and phosphatase inhibitors (10 mM sodium fluoride and 5 mM sodium orthovanadate). Protein concentration was measured by using BCA Protein Assay kit (Pierce). An equal amount of protein (30  $\mu$ g) was separated by NuPAGE Novex Bis-Tris 4–12% gradient gels (Invitrogen) and then transferred onto a polyvinylidene difluoride membrane (Amersham) using a Bio-Rad electrophoretic tank blotting apparatus. The membrane was then incubated with primary antibodies for 1 h at room temperature against ABL1 (24–11; 1:200), BCR (N-20; 1:5000), CTDP1 (C-16; 1:250), MAPK6/ERK3 (I-15; 1:500), FADD (H-181; 1:2000), c-MYB (C-19; 1:2000), c-MYC (N-262; 1:2000), SF3B4 (C-20; 1:250), SMARCB1 (H-300; 1:2000), SMARCA4 (G7; 1:1000) and PTPN1 (H-135; 1:3000; Santa Cruz Biotechnology), and ARID1A (1:250; Abnova), CASPASE-8 (IC12; 1:500; Cell Signaling Technology), DNM1L (1:500; Novas), RAS (1:10000; Upstate), SMARCE1 (1:3000; Bethyl Laboratories, Inc.), USP39 (1:500; Novas) and NF-1 (1:40000; a kind gift from Karen Cichowski, Brigham and Women's Hospital and Harvard Medical School). After incubation with the appropriate horseradish peroxidase-linked secondary antibodies (Bio-Rad), signals were visualized by enhanced chemiluminescence plus Western blotting detection reagents (Amersham). Expression of  $\beta$ -actin was also assessed as an internal loading control by using a specific antibody (C-2; 1:4000; Santa Cruz Biotechnology). Images were scanned by CanoScan 8400F scanner (Canon), and intensities of bands were quantified by LabWorks image analysis software (UVP). After normalization to loading control, target gene knockdown was presented as the relative ratio to CTR01pool infections.

**Ras Activation Assay.** K562 cells ( $5 \times 10^6$ ) stably expressing CTR01pool or shNF1 hairpins were seeded in 20 ml of fresh media onto 100-mm culture dishes for 24 h before treatment with

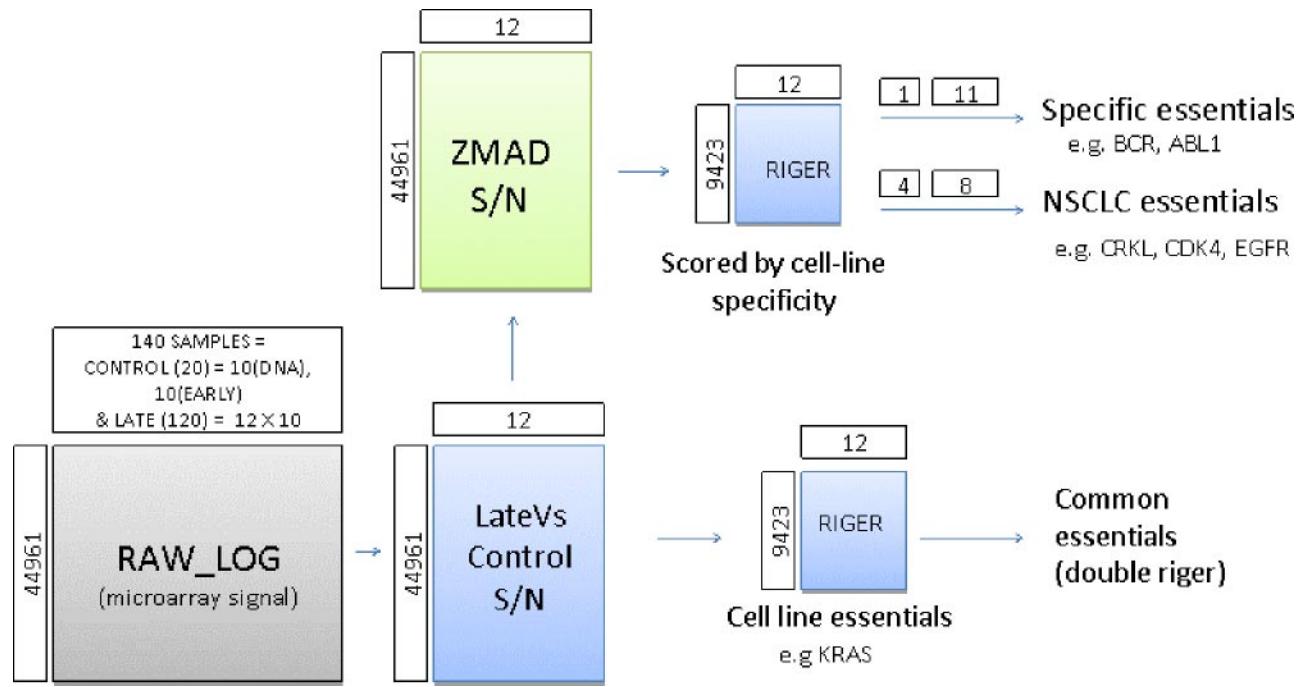
1  $\mu$ M imatinib or solvent for another 24 h. Cell lysis, immunoprecipitation of GTP-bound RAS, and immunoblotting for RAS were performed using a nonradioactive RAS Activation Assay kit (Upstate).

**Real-Time Quantitative Reverse-Transcription PCR.** Total RNA was extracted with TRI reagent (Molecular Research Center). Four micrograms of total RNA for each sample was used to synthesize the first-strand cDNA on 96-well plates by using Oligo(dT)<sub>20</sub>/random hexamer primer cocktails and SuperScript II reverse transcriptase (Invitrogen). Primers for SYBR assays and Taq-Man probes used in the study are listed in Table S1. Quantitative PCR reactions were performed by using the appropriate Universal PCR Master Mix (Applied Biosystems) and set up in 384-well plates by using the robotic MultiPROBE II HT Automated Liquid Handling System (PerkinElmer). Triplicate reactions for the gene of interest and the endogenous control, GAPDH gene, were performed separately on the same cDNA samples by using the ABI 7900HT real time PCR instrument (Applied Biosystems). The mean cycle threshold (C<sub>t</sub>) was used for the  $\Delta\Delta C_t$  analysis method (ABI User Bulletin #2), and target gene knockdown was presented as the relative ratio to the CTR01pool infections.

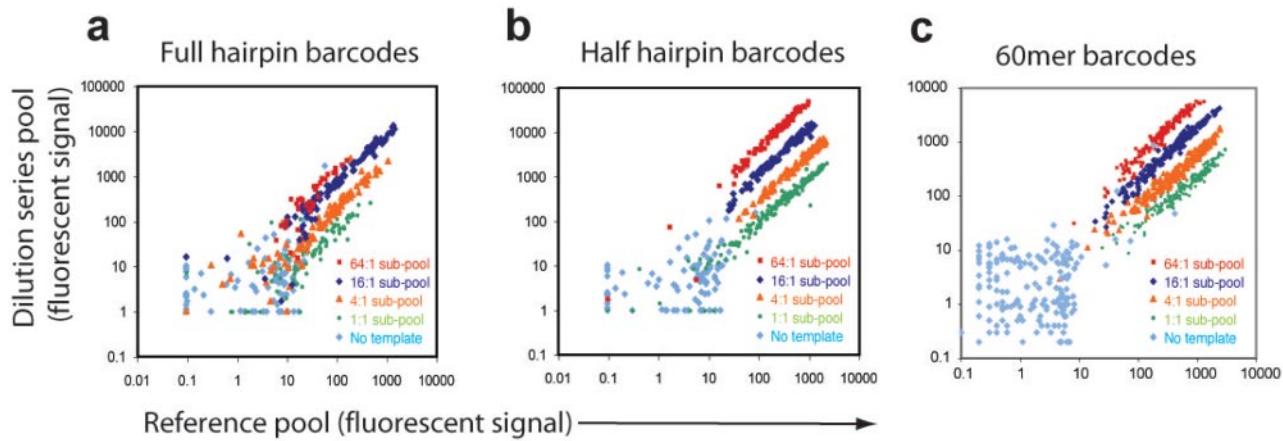
**Assessment of Mitochondrial Membrane Potential with DiOC<sub>6</sub> (3) in FAS-Induced Apoptosis.** Cell lines stably expressing an shRNA of interest or CTR01pool were treated with 1.6 ng/ml of activating FAS antibody CH11 for 2 days. Cells were incubated with 0.4 nM DiOC<sub>6</sub> (3) (3,3'-dihexyloxacarbocyanine iodide) (Invitrogen) at 37 °C for 15 min. The mitochondrial membrane staining by DiOC<sub>6</sub> (3) was measured by FACS.

**Assessment of Caspase-8 Cleavage by Immunoblotting.** Jurkat cells ( $1 \times 10^6$ ) stably expressing an shRNA of interest or CTR01pool were resuspended in 5 ml of fresh medium in each well of the 6-well plate and then treated with 1.6 ng/ml of activating FAS antibody CH11 for 18 h. Cell pellets were collected by centrifugation at  $524 \times g$  for 5 min and resuspended in 200  $\mu$ l of lysis buffer for protein extraction. After immunoblotting for Caspase-8, the images were scanned and the intensity of bands at 57-kDa and 43-kDa sizes was measured (which corresponds to full-length and cleaved Caspase-8, respectively) using LabWorks image analysis software (UVP). The ratios of cleaved-to-full-length Caspase-8 were calculated for shRNA-expressing cells and normalized to CTR01pool-expressing cells.

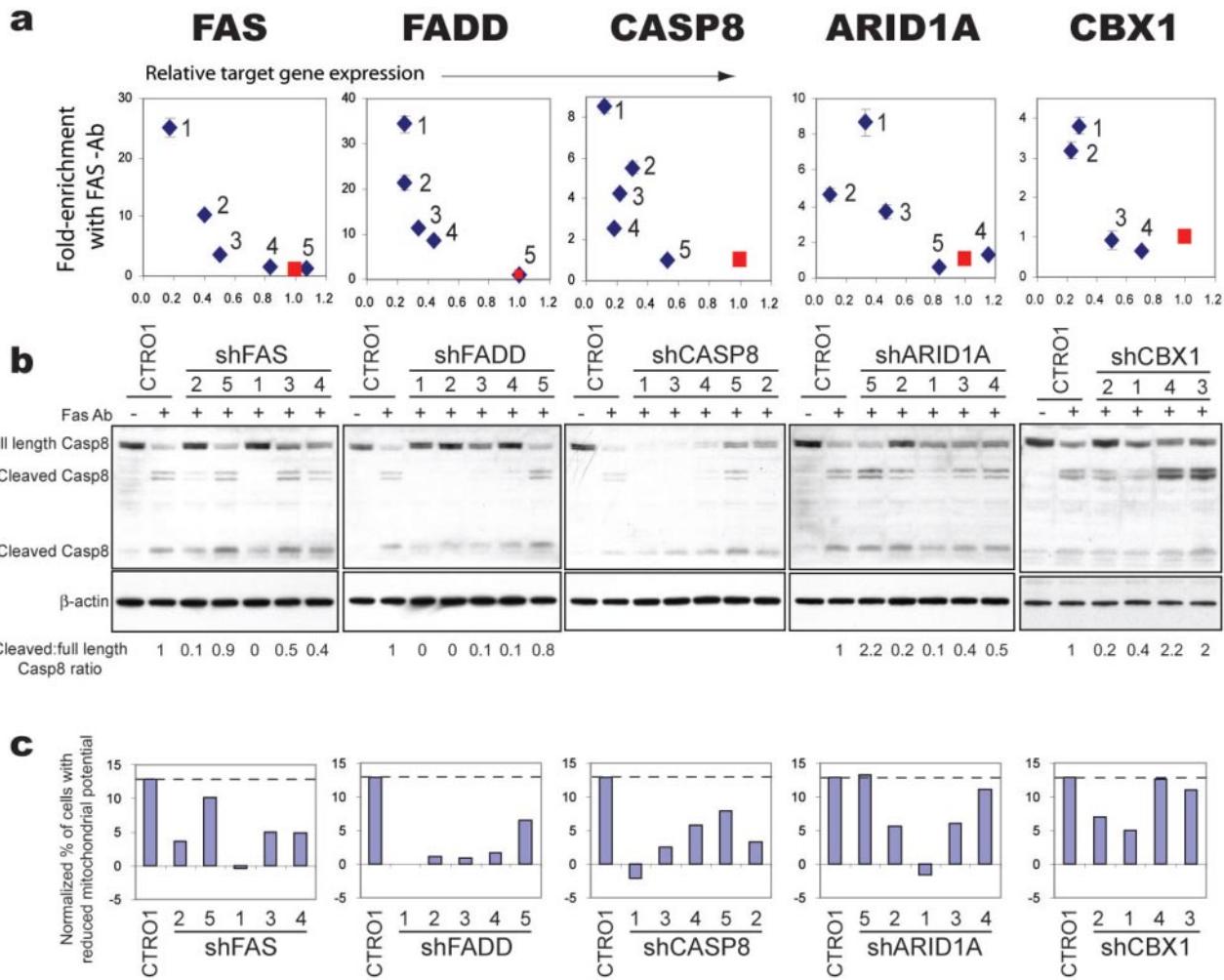
1. Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. *Genome Biol* 2:research0032.
2. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868.
3. Monti S, Tamayo P, Mesirov J, Golub T (2003) Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learn J* 52:91–118.
4. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545–15550.
5. Golub TR, et al. (1999) Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286:531–537.
6. Moffat J, et al. (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 124:1283–1298.



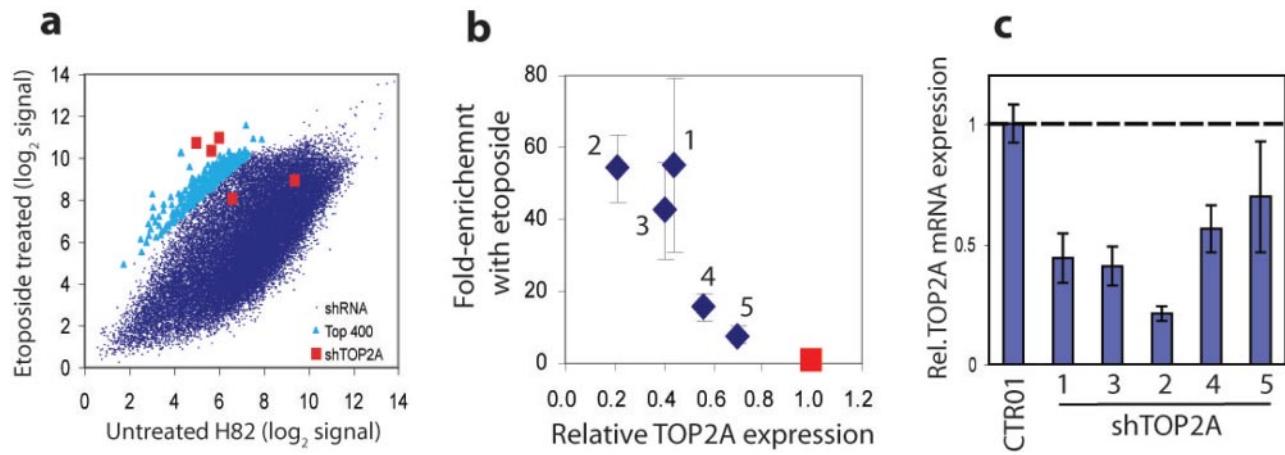
Scheme S1. Essential gene analysis diagram.



**Fig. S1.** Comparison of pool deconvolution performance using array hybridization for three barcode strategies: Full-hairpin barcodes (a), half-hairpin barcodes (b), and 60mer barcodes (c).

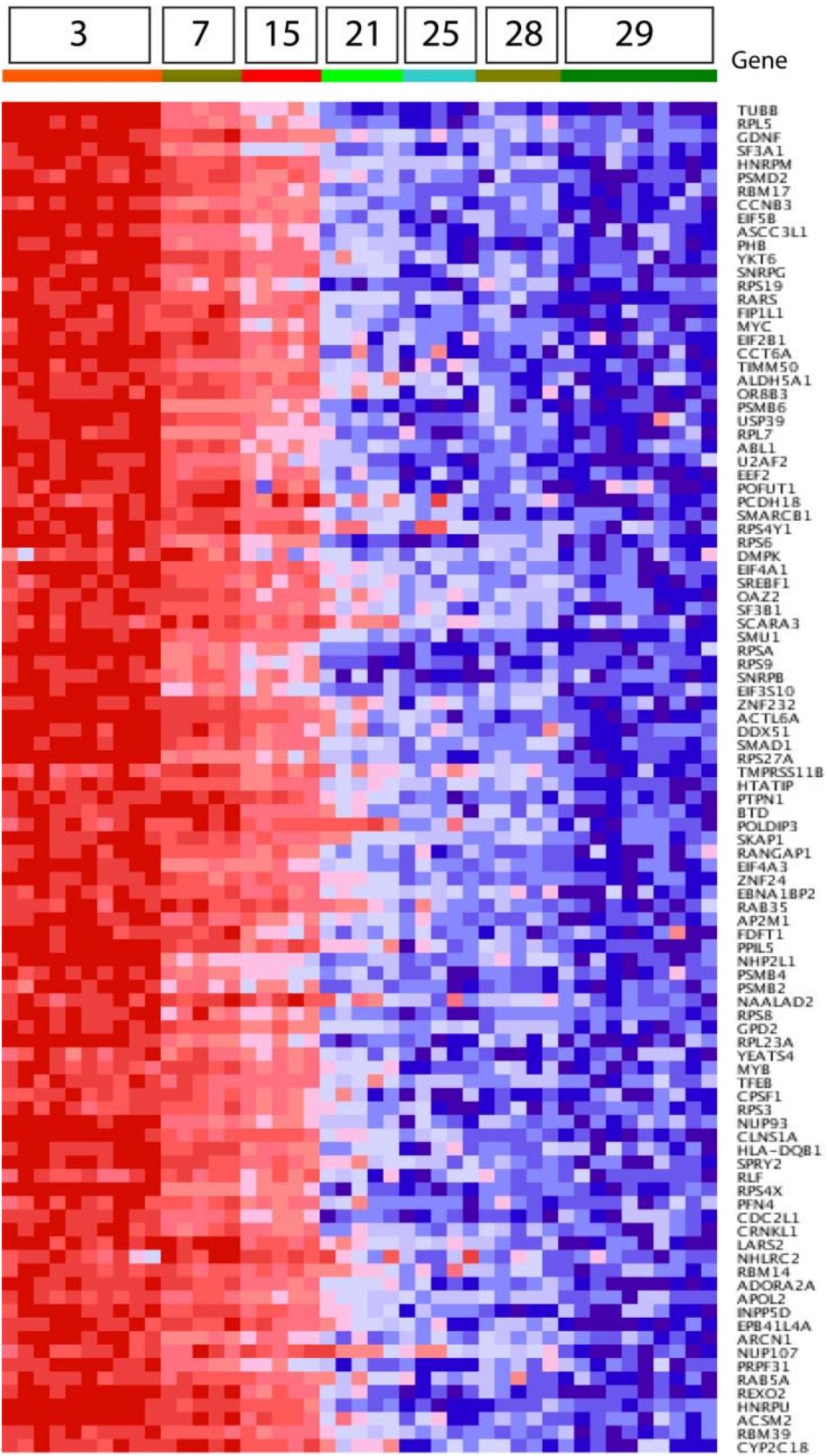


**Fig. S2.** Suppression of *FAS*, *FADD*, *CASP8*, *ARID1A*, or *CBX1* genes conferred resistance to FAS-Ab induced Caspase-8 cleavage and mitochondria leakage. (a) Plots of target gene knockdown for each shRNA to these hit genes versus the relative enrichment of these shRNAs in FAS-treated samples compared to control shRNAs (same plots as Fig. 1D Top). (b) Suppression of *FAS*, *FADD*, *CASP8*, *ARID1A*, and *CBX1* reduced the FAS-Ab induced Caspase-8 cleavage. Jurkat cells were infected with viruses carrying candidate or control shRNAs and were then treated with FAS-Ab for 18 h followed by immunoblotting for Caspase-8. The candidate shRNAs that conferred resistance to FAS-Ab showed decreased ratio of cleaved/full-length Caspase-8 compared to control-shRNA infected cells. (c) Suppression of *FAS*, *FADD*, *CASP8*, *ARID1A*, or *CBX1* inhibited FAS-Ab induced mitochondrial potential reduction. Cells infected with candidate or control shRNAs were treated with FAS-Ab for 48 h followed by FACS analysis.

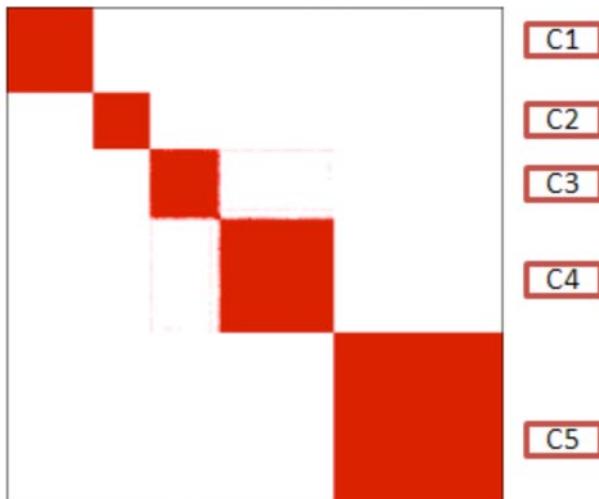


**Fig. S3.** Screen for genes that modulate the effect of etoposide on H82 small-cell lung cancer cells. H82 cells infected with the “45k pool” were grown for 3 weeks with weekly passage in the presence or absence of 1  $\mu\text{g}/\text{ml}$  etoposide. (a) Average probe set signals from 10 independent infections in each group, treated and untreated, are compared. The 400 shRNAs yielding the greatest resistance to etoposide are indicated in light blue. The shRNAs targeting TOP2A are indicated. (b) Relative fold-enrichment of cells infected with each individual TOP2A targeting shRNA treated with etoposide versus untreated. X-axis displays gene suppression measured by RT-PCR. (c) RT-PCR results for TOP2A suppression by the TOP2A-targeted shRNAs.

## Days after infection in K562



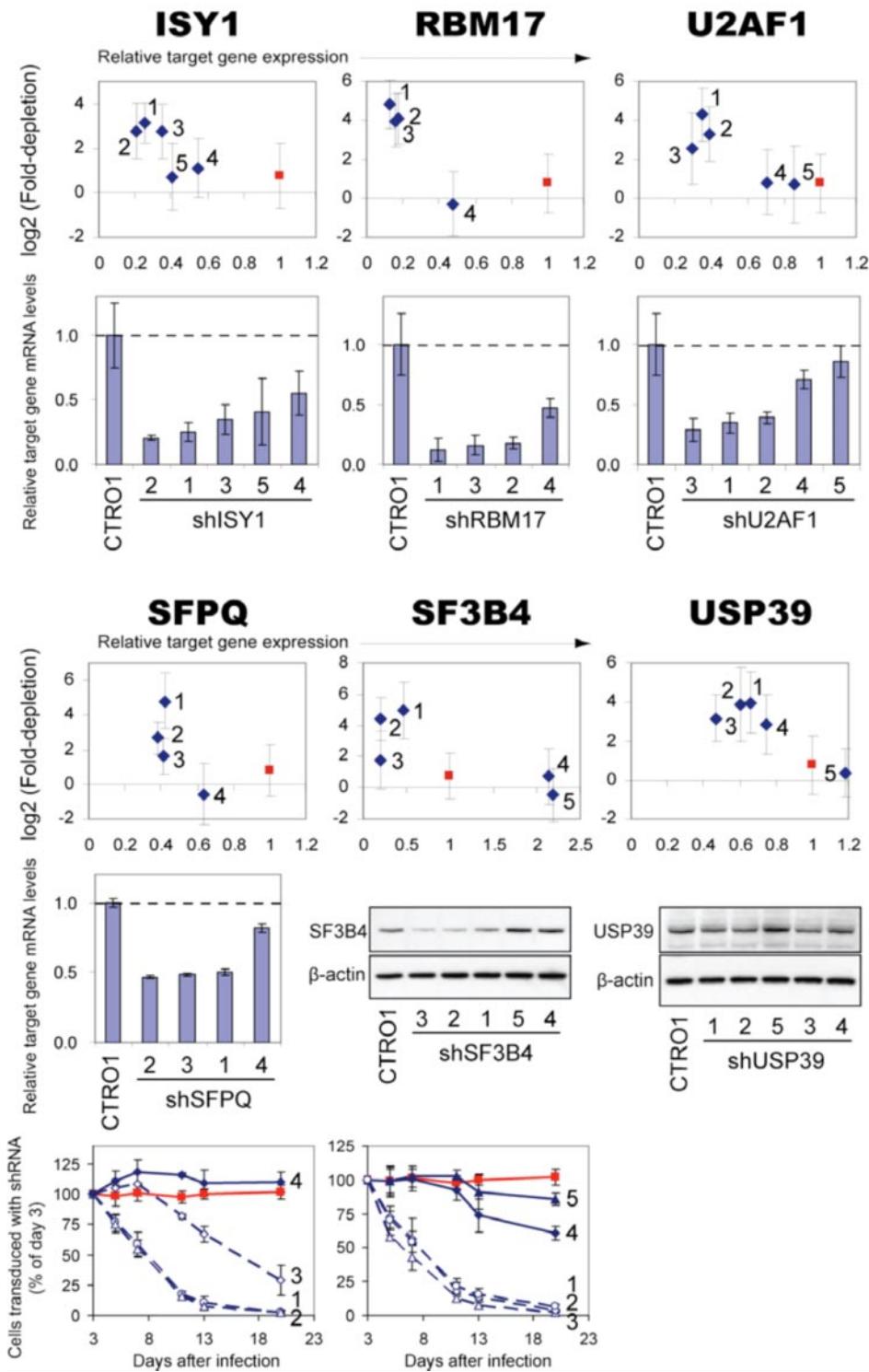
**Fig. S4.** Time course analysis for the top 100 essential genes in K562 cells. The heat map displays the average signal from the RIGER leading-edge shRNAs for each gene and illustrates the consistent depletion of these shRNAs over time.



C1	C2	C3	C4	C5	N	Cell-line	Lineage	Time Point
10					10	JURKAT	T-cell	Late (4 weeks)
10					10	REH	Pre-B ALL	Late (4 weeks)
10					10	SUPT1	T-ALL	Late (4 weeks)
			10	10	HCC827	Lung cancer	Late (4 weeks)	
			10	10	H1975	Lung cancer	Late (4 weeks)	
			10	10	H1650	Lung cancer	Late (4 weeks)	
			10	10	A549	Lung cancer	Late (4 weeks)	
10				10	H187	Lung cancer (SC)	Late (4 weeks)	
10				10	H82	Lung cancer (SC)	Late (4 weeks)	
			10	10	LN229	Glioblastoma	Late (4 weeks)	
			10	10	U251_LATE	Glioblastoma	Late (4 weeks)	
		10		10	U251_EARLY	Glioblastoma	Early (4d)	
		10		10	DNA		Early (3d)	
		10		10	K562_3D	CML	Early (3d)	
		5		5	K562_7D	CML	Intermediate (7d)	
		5		5	K562_15D	CML	Intermediate (15d)	
	5			5	K562_21D	CML	21d	
	5			5	K562_25D	CML	25d	
	5			5	K562_28D	CML	Late (28d)	
	10			10	K562_29D	CML	Late (29d)	
30	20	25	40	60	175	TOTAL		

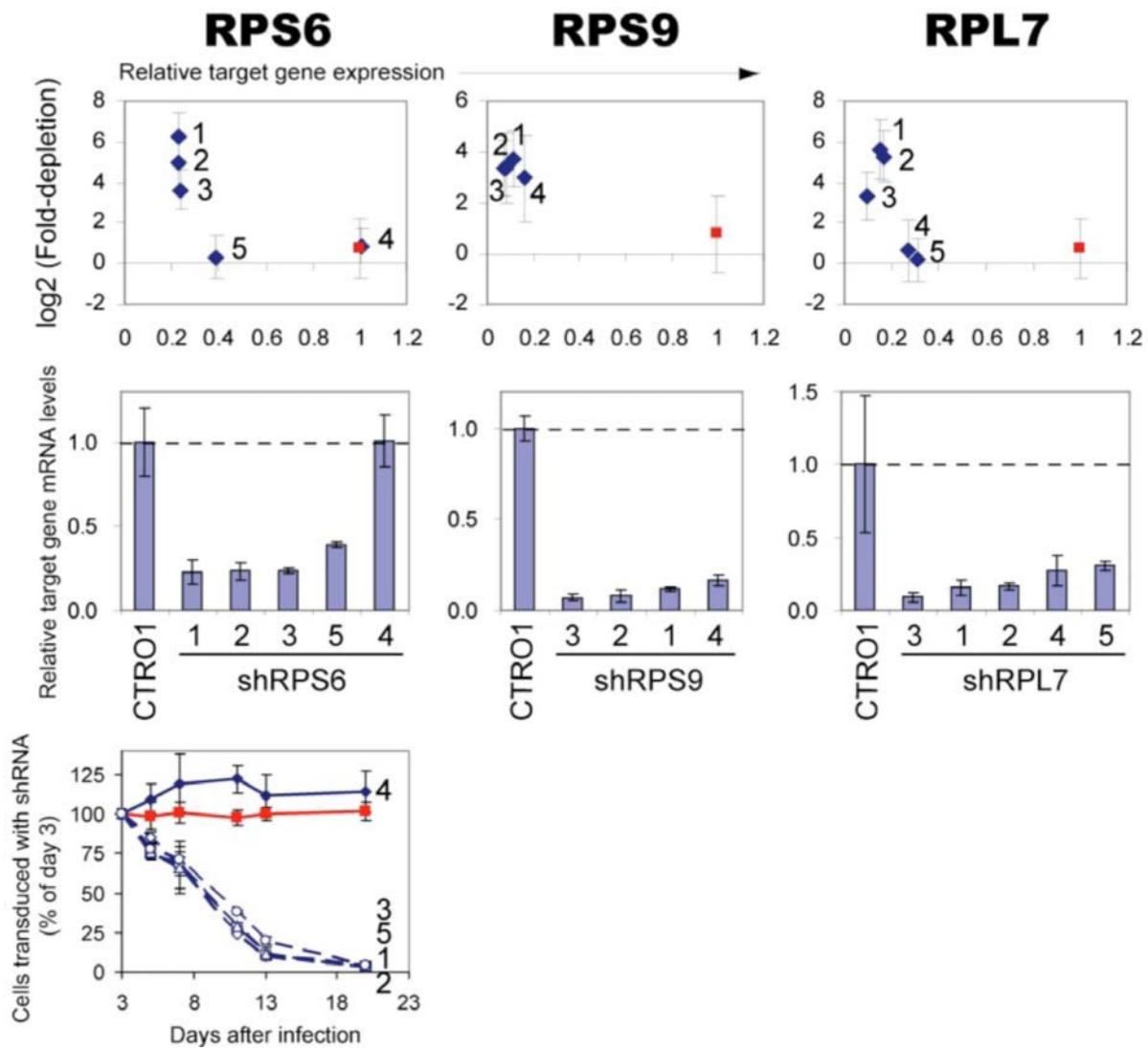
**Fig. S5.** Consensus clustering of cell lines by gene-essentiality screening data. Consensus clustering was performed on the shRNA hhb hybridization array data for the 12 cell lines and the pooled shRNA plasmids (5–10 replicates per cell line/time-point;  $n = 175$ ). A stable clustering into 5 classes (labeled C1–C5) was obtained. Red indicates that 2 samples always cluster together and white indicates that 2 samples never cluster together.

## a. RNA SPLICING PROTEINS



**Fig. S6.** Validation of target gene suppression for shRNAs targeting top-scoring commonly essential genes. Validation data are displayed for examples of essential genes involved in (a) mRNA splicing, (b) ribosomal function, (c) MYC signaling, (d) mRNA processing, (e) mRNA translation, and (f) various other biological processes. For these genes, depletion of shRNAs in the pooled screen for essentiality correlates with the degree of target gene knockdown, indicating a gene-specific effect. For each gene, the first plot (top to bottom) depicts the correlation between the fold-depletion of shRNAs over 4 weeks in K562 cells, as measured in the primary pooled screen, and target gene knockdown. The second plot depicts target gene suppression measured by immunoblotting or quantitative RT-PCR; for some genes, a third plot depicts time-resolved depletion of shRNA-infected cells using flow cytometry to determine percentage of LKO-GFP-shRNA infected cells in a mixed infected/uninfected cell population. Data for a LKO-GFP control infection are labeled in red.

## b. RIBOSOMAL PROTEINS



### C. MYC SIGNALING PROTEINS

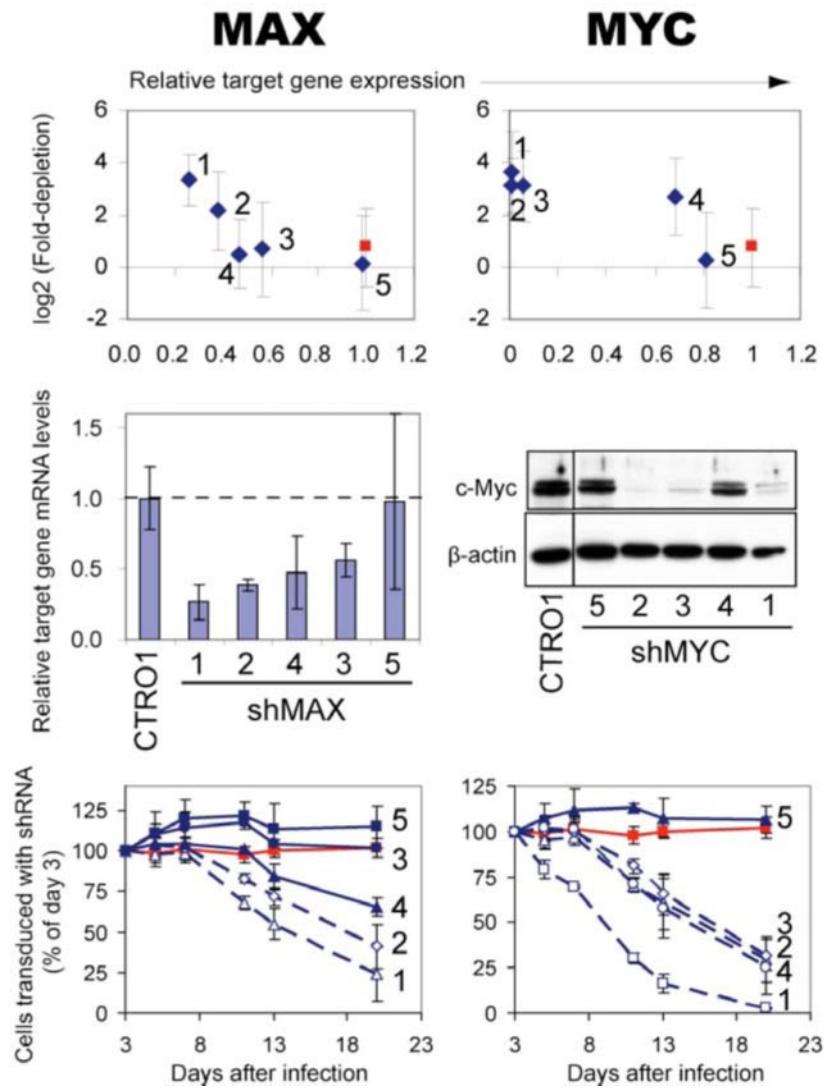
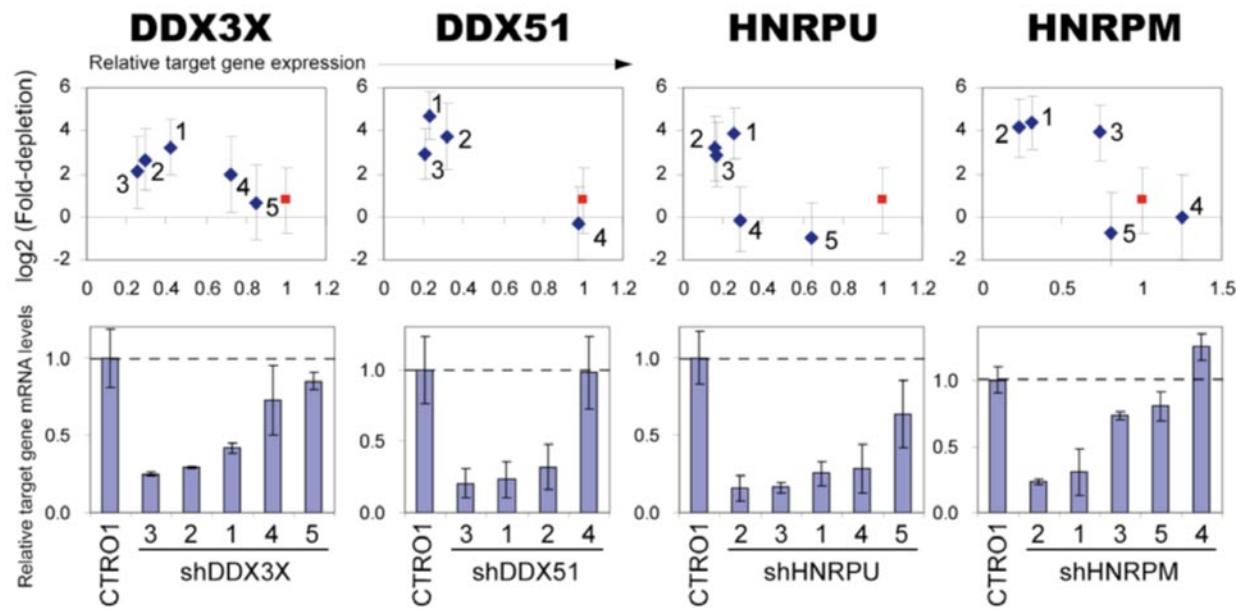


Fig. S6 continued.

## d. mRNA PROCESSING PROTEINS



## e. TRANSLATION PROTEINS

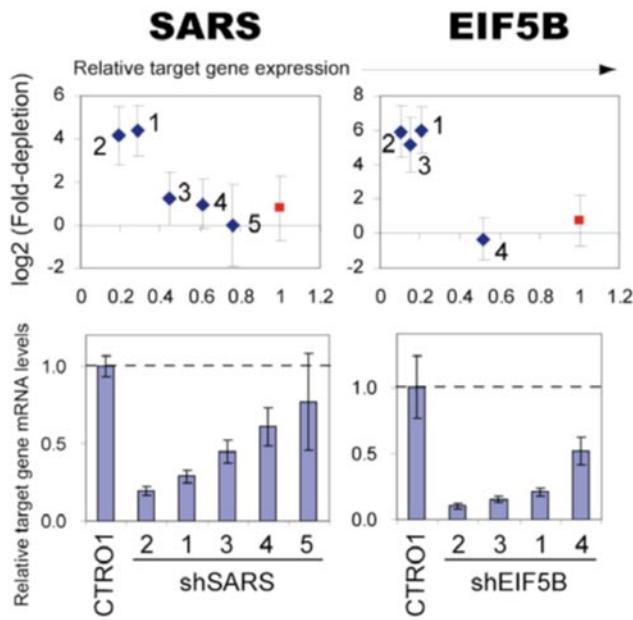


Fig. S6 continued.

**f. OTHERS**

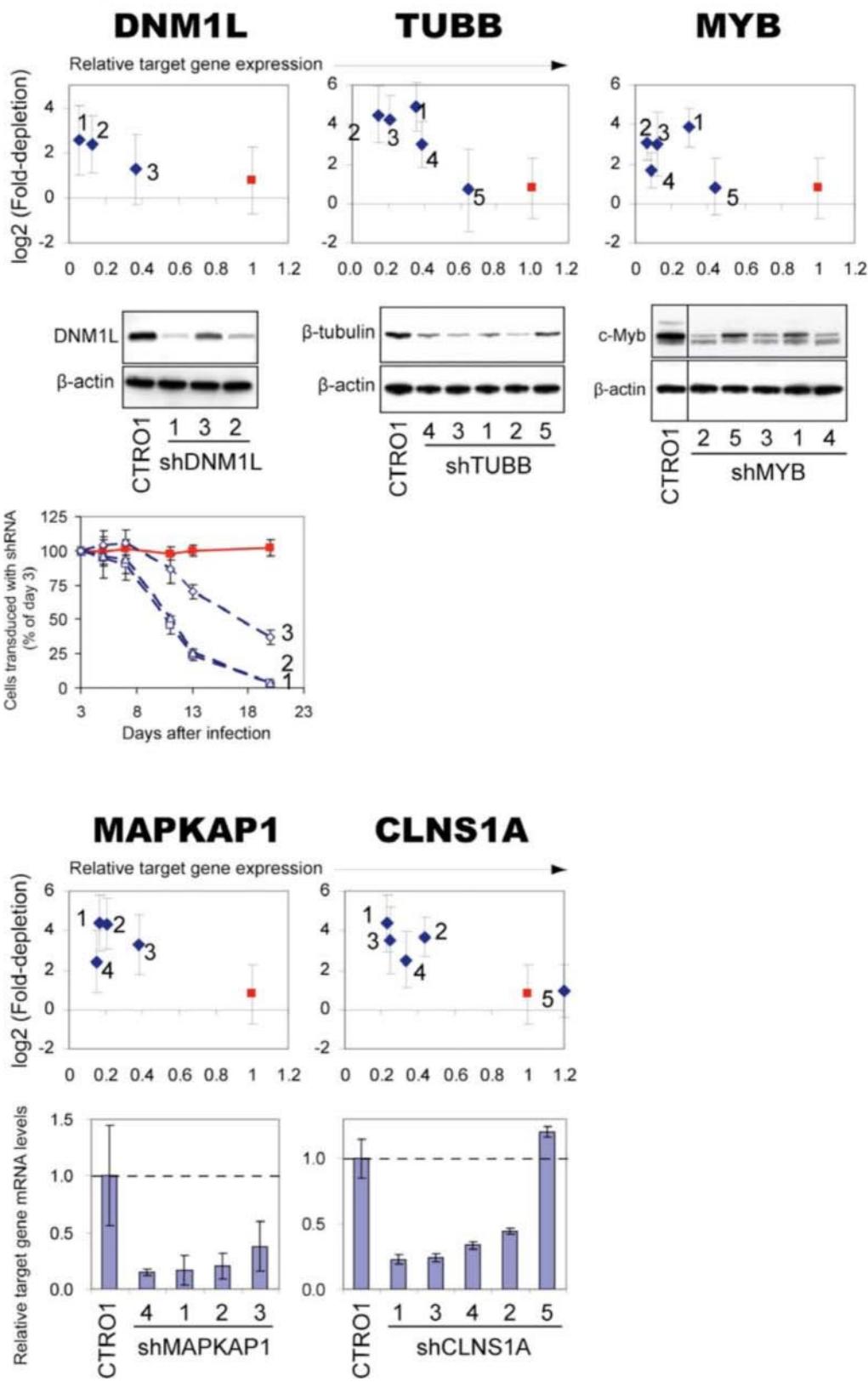
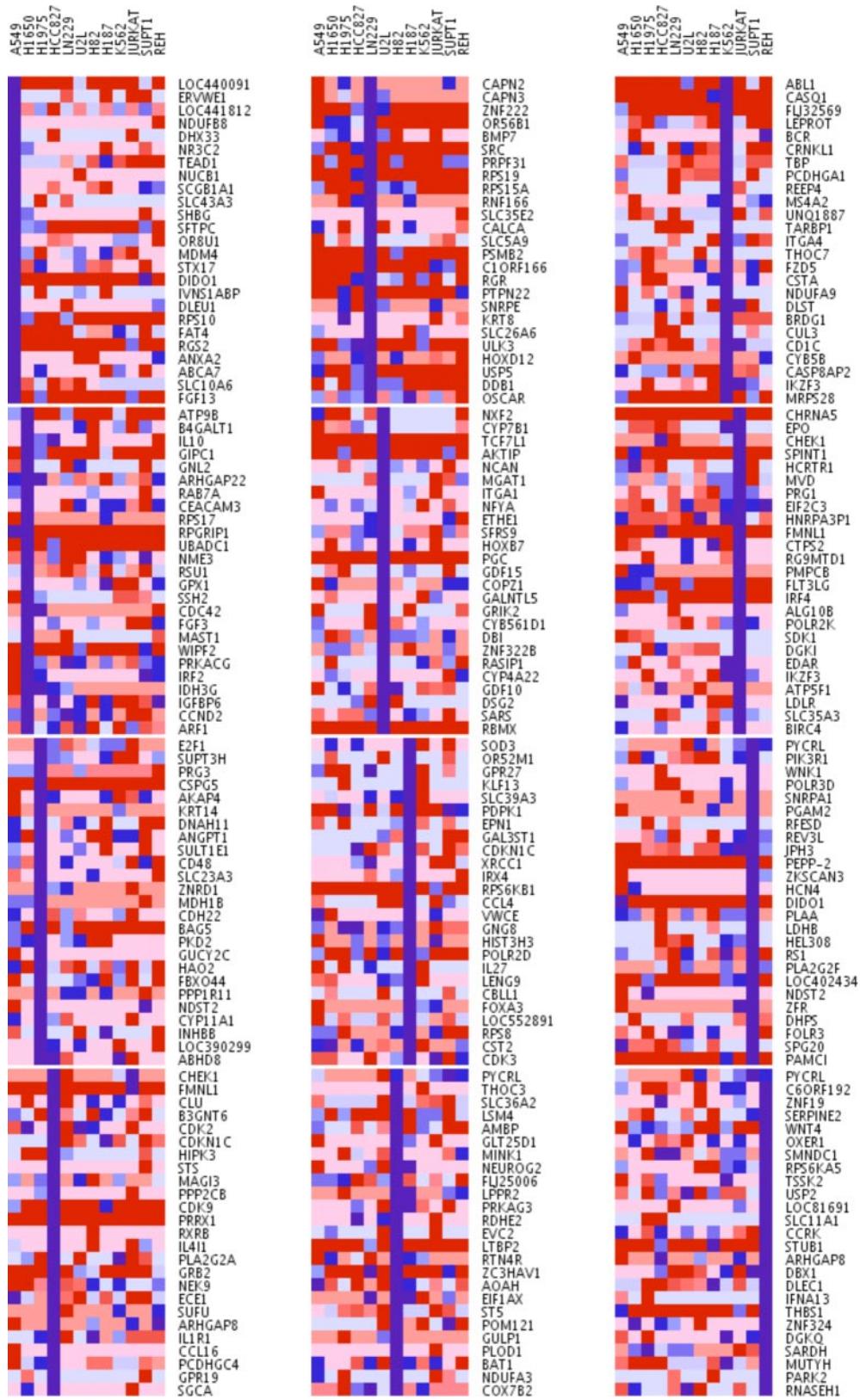
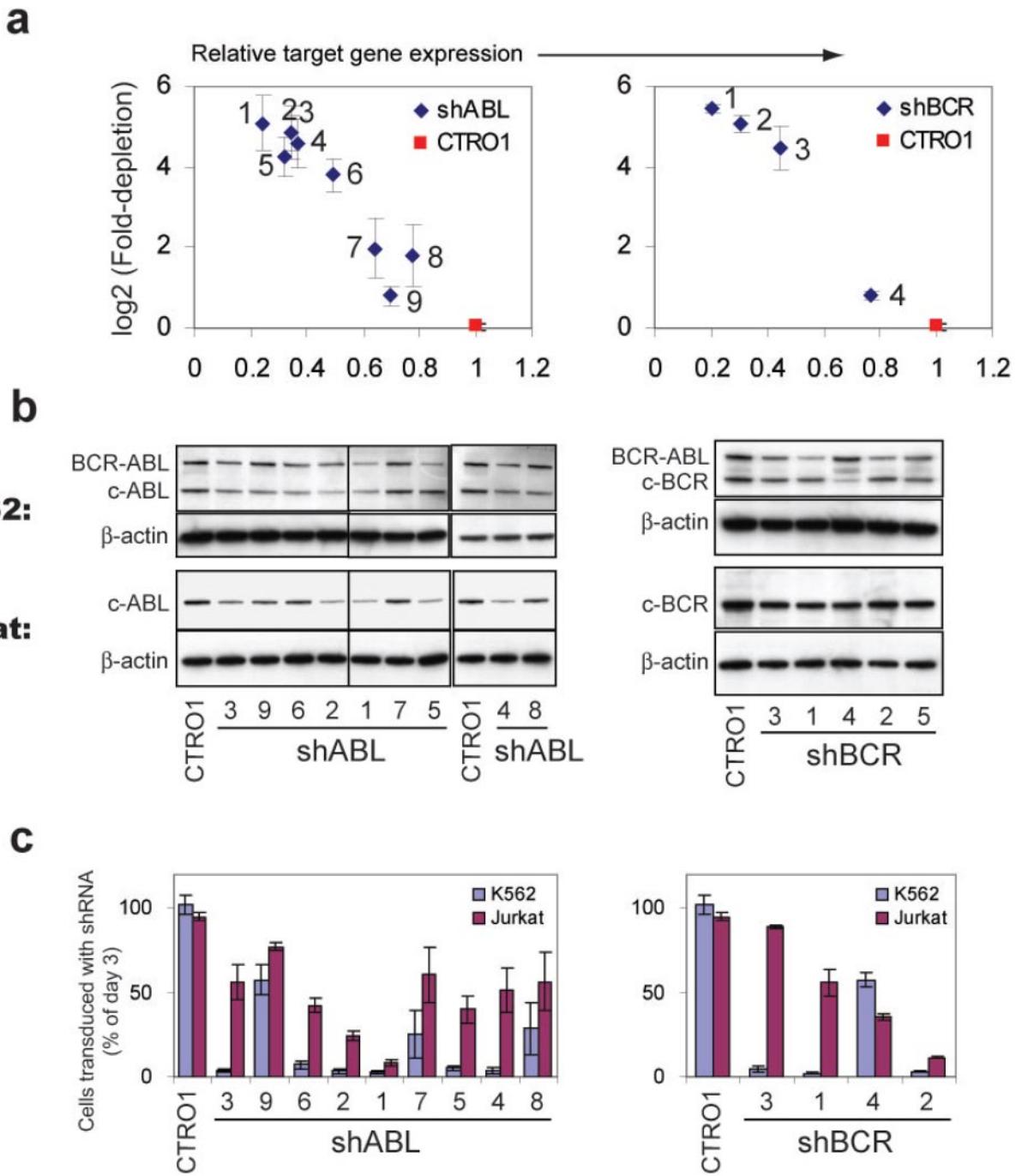


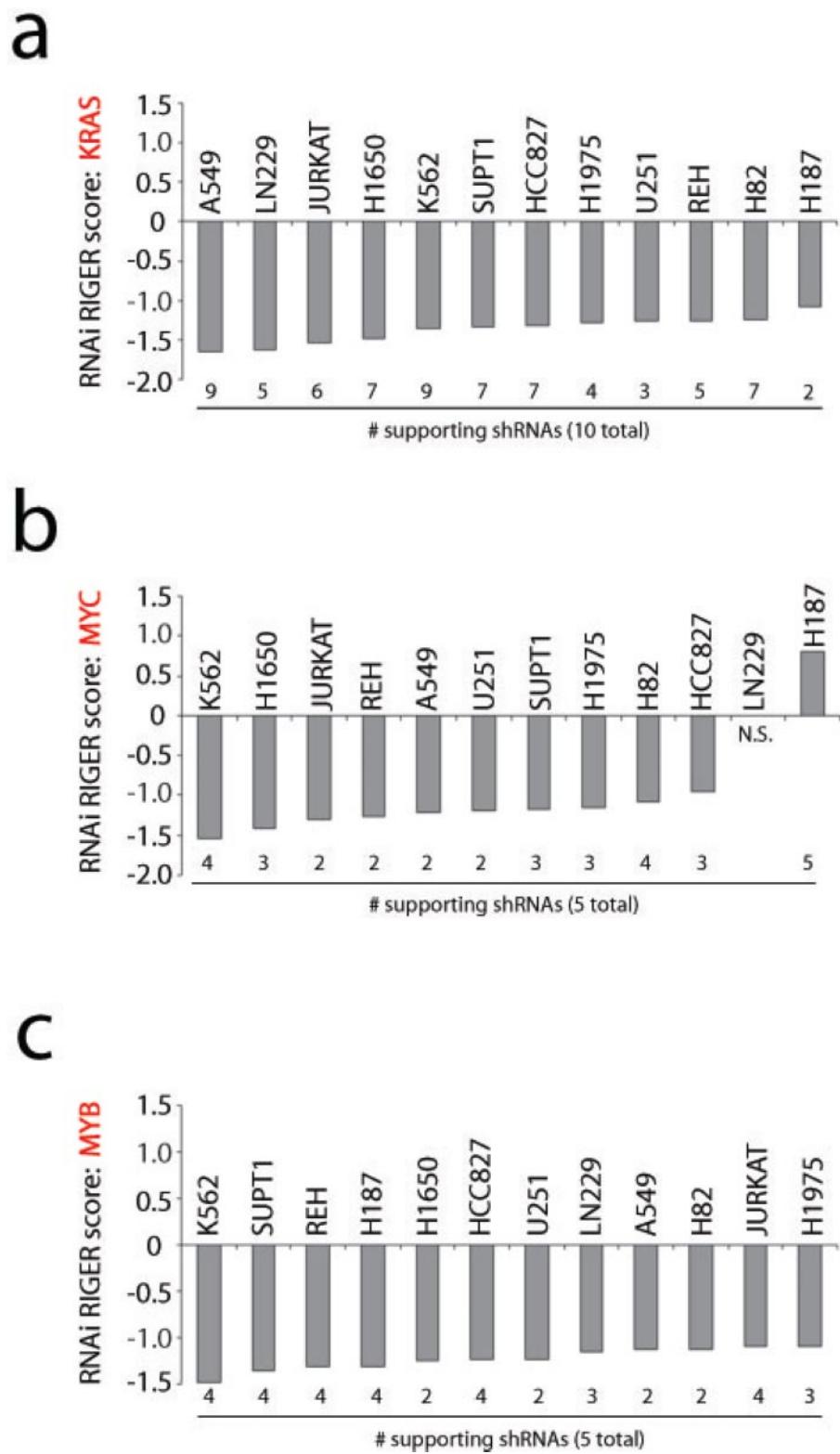
Fig. S6 continued.



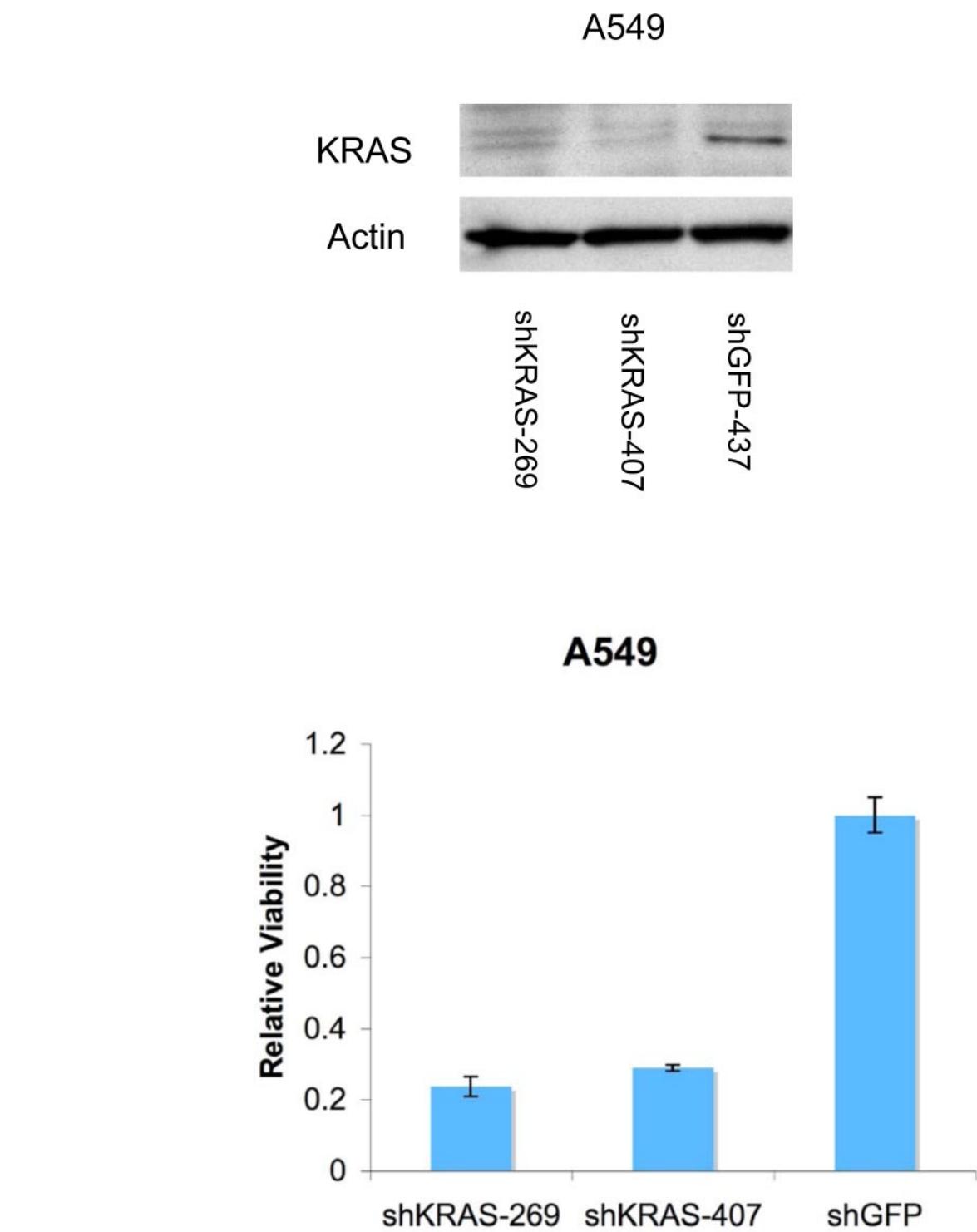
**Fig. S7.** Identification of cell line-specific essential genes based on relative shRNA depletion in one cell line versus the other 11 cell lines. Results for the top 25 specific essential genes for each cell line are displayed in the heat map.



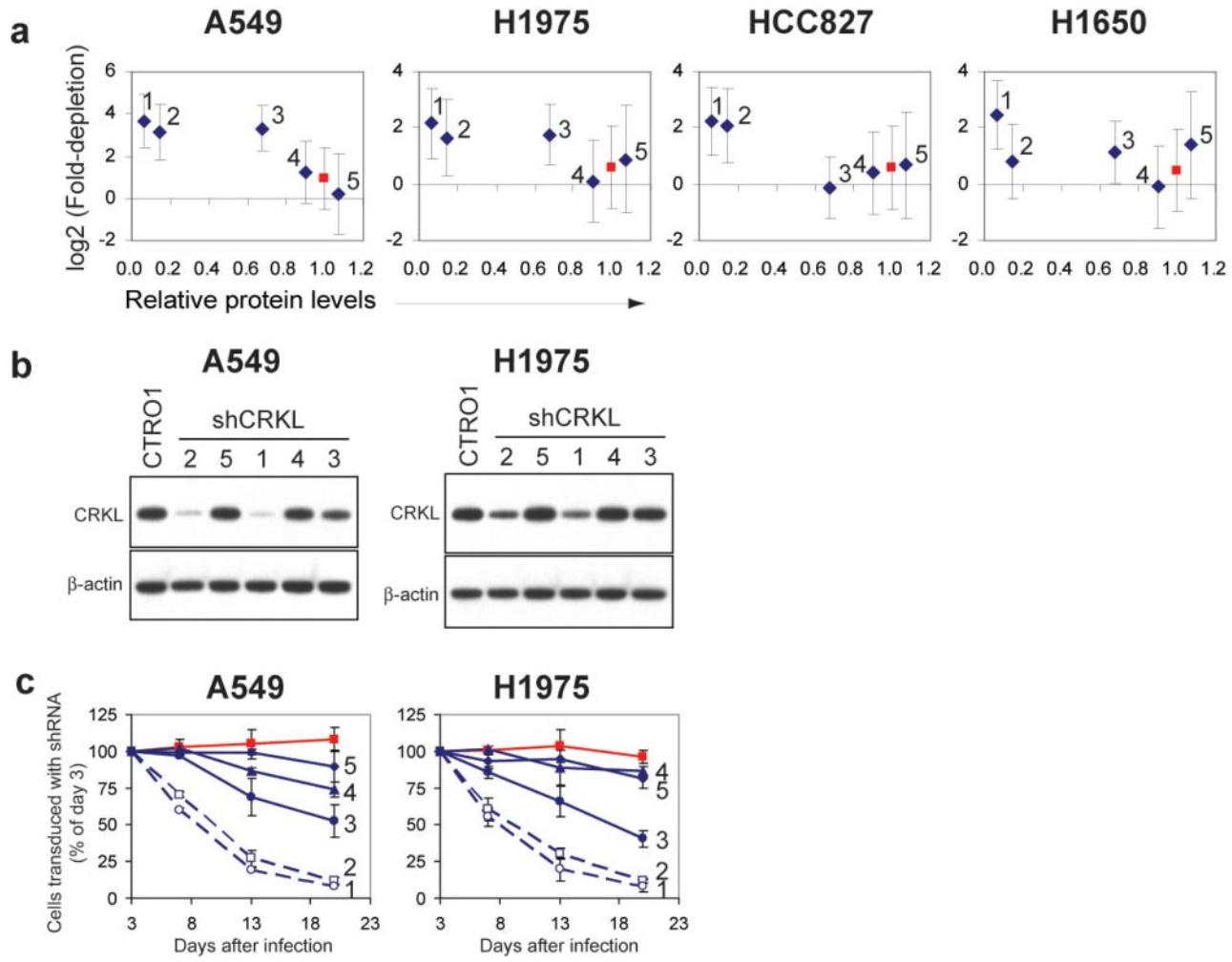
**Fig. S8.** Validation of target gene knockdown by shRNAs targeting *ABL1* (Left) and *BCR* (Right). (a) For shRNAs against *ABL1* and *BCR* in K562 cells, correlation between fold depletion of shRNA-virus-infected cells and BCR-ABL gene suppression. Fraction of LKO-GFP-shRNA-infected cells was measured at 3 and 20 days post-infection by FACS. (b) Immunoblot analysis of protein knockdown by shBCR and sh*ABL1* in K562 and Jurkat cells. (c) Differential anti-proliferative effect of sh*ABL1* and sh*BCR* in K562 versus Jurkat cells.



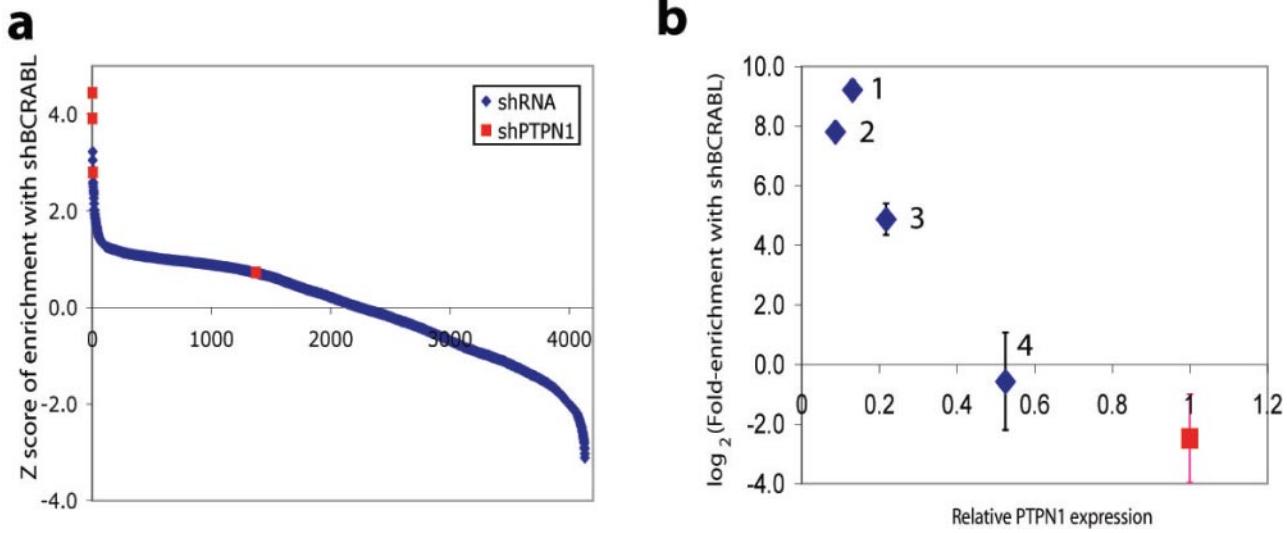
**Fig. S9.** KRAS, MYC, and MYB essentiality in 12 cell lines. Among known oncogenes, KRAS, MYC, ABL1 (see Fig. 2D), and MYB displayed the greatest proliferative requirement in one or more of the 12 screened cell lines. The normalized enrichment score for cell-line essentiality is provided for each cell line along with the number of shRNAs that contribute to the enrichment. Two or more shRNAs for each gene were required to be in the RIGER leading edge in order to obtain a RIGER score for that gene; otherwise the RIGER result is labeled. (N.S. (No Score).



**Fig. S10.** Two shRNAs that target KRAS. (a) shRNAs effectively suppress KRAS. (b) shRNAs exhibit a strong anti-proliferative effect in A549 cells that have an activating KRAS mutation.



**Fig. S11.** Validation of target gene knockdown by shRNAs targeting *CRKL*. (a) Anti-proliferative effects of the 5 shRNAs targeting *CRKL* in the 4 non-small lung cancer cell lines, as measured in the 45,000 shRNA primary pooled screen, plotted versus the level of protein knockdown by the same shRNAs in A549 cells. (b) Immunoblot analysis of protein knockdown by sh*CRKL* in A549 and H1975 cells. (c) Time-resolved depletion of sh*CRKL*-infected cells. Flow cytometry was used to determine percentage of LKO-GFP-sh*CRKL* infected cells in a mixed infected/uninfected cell population for A549 and H1975 cells. Data for a LKO-GFP control infection are labeled in red.



**Fig. S12.** Screen for modifiers of the response to shBCR-ABL. K562 cells were infected with a virus from a pool of 4,000 shRNAs from the TRC1 library that had been transferred into a MSCV-neo-shRNA expression vector. This shRNA screen was repeated on 6 cell populations. Two shABL1 and one shBCR-treated samples (the shBCR-ABL group) were compared to a sample treated with a mixture of control shRNAs targeting reporter genes and two duplicate cell populations that were not super-infected following the library treatment (the control group). (a) Enrichment of shPTPN1 by shBCR-ABL treatment in K562 cells. The enrichment Z-score for each shRNA (with versus without shBCR-ABL treatment) are plotted versus the Z-score rank; shRNAs targeting PTPN1 (ranks 1, 2, 6, and 1,372) are labeled as red squares. (b) Correlation between fold-enrichment (with versus without shBCRABL) versus PTPN1 suppression by shPTPN1s.

**Table S1. Primers for SYBR assays and TaqMan probes**

Gene	SYBR green forward primer	SYBR green reverse primer	TaqMan probe
<i>ISY1</i>	AAGTTGCCCTAAATGCTGG	CGCACCTGGCGTTTTTC	
<i>RBM17</i>	AGAGAGGAACGACAGAGACAG	TGTGATCGAGGTCTGAGTCC	
<i>U2AF1</i>			Hs00739599_m1
<i>SFPQ</i>			Hs00192574_m1
<i>RPS6</i>	TGGACGATGAACGCAAACCTC	TTCGGACCACATAACCCCTCC	
<i>RPS9</i>	GGAACTGCTGACGCTTGATGA	CCCAGGATGTAATCCAGCTCA	
<i>RPL7</i>	AAGCTGGCAACTTCTATGTACC	GGGCTCACTCCATTGATACCTC	
<i>MAX</i>			Hs00231142_m1
<i>DDX3X</i>	CAGGCAACAACGTCCCTCCA	AGTTTTCCAGACCCCTGTTGG	
<i>DDX51</i>	CAGGCCCTGCTTCGAGAG	GCCAGAGACTTCTGTCCCG	
<i>HNRPU</i>	ACCAGATGGAGCTAGGAGAGG	CTTCCTGAAACCCCTGATCGT	
<i>HNRPM</i>	TGCGGAAGTCTAAACAAGCA	GTAGCCATCACCTTTGCATTG	
<i>SARS</i>	ATGTGCTGAGTTCGATGACC	GCGTCACACTTCAGGATGG	
<i>EIF5B</i>	AGAAAACGGCGACTTGAACATAG	CCTGGAATCCGTACATTCTCT	
<i>MAPKAP1</i>	TGACCTGGACAGCACTTGG	TGCCCTTGAACTGTTCTCGC	
<i>CLNS1A</i>	GACAGGGGGACATCCCTACAT	TCCTGACCCAGCCATTATATAC	
<i>CBX1</i>	GCCGGAGCGGATTATTGGAG	GGAATGCCACGTCAGCCTT	
<i>TOP2A</i>	ACAAAGGTTGGGCACCAAG	CAGGCTGATAGCAGCATCATC	

**Other Supporting Information Files**

[Dataset S1 \(XLS\)](#)  
[Dataset S2 \(XLS\)](#)  
[Dataset S3 \(XLS\)](#)  
[Dataset S4 \(XLS\)](#)  
[Dataset S5 \(XLS\)](#)  
[Dataset S6 \(XLS\)](#)  
[Dataset S7 \(XLS\)](#)