Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1109363108">http://dx.doi.org/10.1073/pnas.1109363108</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Proceedings of the National Academy of Sciences (PNAS)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Dec 08 21:29:40 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/69006">http://hdl.handle.net/1721.1/69006</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer

Hiu Wing Cheung\textsuperscript{a,b,1}, Glenn S. Cowley\textsuperscript{b,1}, Barbara A. Wei\textsuperscript{b,1}, Jesse S. Boehm\textsuperscript{b,1}, Scott Rusin\textsuperscript{b}, Justine A. Scott\textsuperscript{b}, Alexandra East\textsuperscript{a}, Levi D. Ali\textsuperscript{b}, Patrick H. Lizotte\textsuperscript{c}, Terence C. Wong\textsuperscript{b}, Guozhi Jiang\textsuperscript{b}, Jessica Hisao\textsuperscript{a}, Craig H. Merem\textsuperscript{b,c}, Gad Getz\textsuperscript{b}, Jordi Barretina\textsuperscript{b,ab}, Shuba Gopal\textsuperscript{b}, Pablo Tamayo\textsuperscript{b}, Joshua Gould\textsuperscript{b}, Aviad Tsherniak\textsuperscript{b}, Nicolas Stransky\textsuperscript{b}, Biao Luo\textsuperscript{b}, Yin Ren\textsuperscript{d}, Ronny Drapkin\textsuperscript{e,f}, Sangeeta N. Bhatia\textsuperscript{b,d,g,hi}, Jill P. Mesirov\textsuperscript{b}, Levi A. Garraway\textsuperscript{a,b,cg}, Matthew Meyerson\textsuperscript{b,ce}, Eric S. Lander\textsuperscript{b,2}, David E. Root\textsuperscript{b,2}, and William C. Hahn\textsuperscript{a,b,cg,2}

\textsuperscript{a}Department of Medical Oncology, \textsuperscript{b}Center for Cancer Genome Discovery and \textsuperscript{c}Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02115; \textsuperscript{1}Broad Institute of Harvard and MIT, Cambridge, MA 02142; \textsuperscript{2}Harvard-MIT Division of Health Sciences and Technology and “Electrical Engineering and Computer Science, David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139; Departments of “Pathology and “Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115; and “Howard Hughes Medical Institute, Chevy Chase, MD 20815

Contributed by Eric S. Lander, June 13, 2011 (sent for review April 13, 2011)

A comprehensive understanding of the molecular vulnerabilities of every type of cancer will provide a powerful roadmap to guide therapeutic approaches. Efforts such as The Cancer Genome Atlas Project will identify genes with aberrant copy number, sequence, or expression in various cancer types, providing a survey of the genes that may have a causal role in cancer. A complementary approach is to perform systematic loss-of-function studies to identify essential genes in particular cancer cell types. We have begun a systematic effort, termed ProAccess, aimed at identifying genetic vulnerabilities across large numbers of cancer cell lines. Here, we report the assessment of the essentiality of 11,194 genes in 102 human cancer cell lines. We show that the integration of these functional data with information derived from surveying cancer genomes pinpoints known and previously undescribed lineage-specific dependencies across a wide spectrum of cancers. In particular, we found 54 genes that are specifically essential for the proliferation and viability of ovarian cancer cells and also amplified in primary tumors or differentially overexpressed in ovarian cancer cell lines. One such gene, PAX8, is focally amplified in 16% of high-grade serous ovarian cancers and expressed at higher levels in ovarian tumors. Suppression of PAX8 selectively induces apoptotic cell death of ovarian cancer cells. These results identify PAX8 as an ovarian lineage-specific dependency. More generally, these observations demonstrate that the integration of genome-scale functional and structural studies provides an efficient path to identify dependencies of specific cancer types on particular genes and pathways.

The application of whole-genome approaches to identify genetic alterations in cancer genomes is providing new insights into the spectrum of molecular events that occur in human tumors. Although in some cases this knowledge immediately illuminates a path toward clinical implementation, the long lists of genes with aberrant sequence, copy number, or expression in tumors already found make it clear that complementary information from systematic functional studies will be essential to obtain a comprehensive molecular understanding of cancer and to convert this knowledge into therapeutic strategies.

Most ovarian cancer patients present at an advanced stage with widely disseminated disease in the peritoneal cavity. Despite advances in surgery and chemotherapy, the majority of ovarian cancer patients re-present with relapsed and progressively chemotherapy-resistant and lethal disease. The Cancer Genome Atlas (TCGA) Project has characterized nearly 500 primary high-grade serous ovarian cancer genomes to identify recurrent genetic alterations in ovarian cancer (1).

A major feature of ovarian cancers is recurrent regions of copy-number alteration (1). A small number of these recurrent genomic events harbor known oncogenes and tumor suppressor genes, such as MYC, CCNE1, and RB (2). However, as in other cancers, the specific genes out of the ~1,800 genes targeted by recurrent amplification events remain undefined. Here, we report a genome-scale functional study to identify genes that are essential for the survival of 102 human cancer cell lines. The interrogation of a large number of cancer cell lines provides increased power to identify relationships between gene expression and dependence. To demonstrate the utility of integrating these data with results from cancer genome characterization, we focused on ovarian cancer and identified ovarian cancer lineage-specific dependencies.

Results

To identify genes essential for the proliferation and survival of human cancer cell lines, we performed genome-scale, pooled short hairpin RNA (shRNA) screens (3) in 102 cell lines (Fig. 1A and Fig. S1A), including 25 ovarian, 18 colon, 13 pancreatic, 9 esophageal, 8 non-small-cell lung cancer (NSCLC), and 6 glio-blastoma cancer cell lines (Table S1). Each cell line was infected in quadruplicate (Fig. S2A) with a pool of lentivirally delivered shRNAs, composed of 54,020 shRNAs targeting 11,194 genes, and propagated for at least 16 doublings. The abundance of shRNA sequences at the endpoint relative to the initial reference pool was measured by microarray hybridization. We developed a standardized analytical pipeline to assess the effects on proliferation induced by each shRNA (Figs. S1B and S2 B and C). Replicate infections were highly correlated (Fig. S2 D and E), and hierarchical clustering demonstrated that the replicates clustered tightly together (Fig. 1B). In all, we obtained 22 million individual measurements of shRNA effects on proliferation.


FREELY AVAILABLE ONLINE THROUGH THE PNAS OPEN ACCESS OPTION.

1H.W.C., G.S.C., B.A.W., and J.S.B. contributed equally to this work.

2To whom correspondence may be addressed. E-mail: lander@broadinstitute.org, droot@broadinstitute.org, or william_hahn@dfci.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1109363108/-/DCSupplemental.
and in lung adenocarcinoma (8), KRAS ranked highly in 2/3 analyses (23rd and 30th of 11,194; Fig. S4). PIK3CA itself strongly scored as a top differential in squamous cell carcinomas (10). To exhibit a similar dependence on mTOR, these observations suggest that cancer cell lines that harbor mutant BRAF exhibit a similar dependence on BRAF. Although initial reports suggest that clinical responses to BRAF inhibition in BRAF-mutant colon tumors are much less robust than those observed in BRAF-mutant melanomas (6), our observations indicate that BRAF is essential in colon cancer cell lines that express mutant BRAF (Fig. S4 A and B).

We next examined vulnerabilities in cell lines harboring PIK3CA mutations. PIK3CA itself strongly scored as a top differentially essential gene between PIK3CA mutant and PIK3CA wild-type cell lines in 2/3 gene-level analyses (Fig. 2 G and H). MTOR ranked highly in 2/3 analyses (23rd and 30th of 11,194; Fig. S4C); the top scoring MTOR-specific shRNAs strongly discriminated the PIK3CA mutant and wild-type classes (\(P = 6.0 \times 10^{-7}\); WoE; Fig. S4D). Indicating that cell lines that harbor PIK3CA mutations are also dependent on mTOR. These observations confirm prior work showing that mTOR plays an important role in PI3K signaling (7).

To assess how the number of cell lines analyzed affected these analyses, we repeated our WoE scoring of a set of shRNAs using data from smaller numbers of cell lines subsampled from the entire dataset. For the top scoring KRAS, BRAF, and PIK3CA shRNAs that were able to distinguish cell lines with mutant or wild-type alleles of these same respective genes, we found that the analysis of a smaller number of cell lines (<5) decreased the likelihood of discrimination between these two classes, whereas the comparison of groups composed of >10 cell lines greatly increased our ability to distinguish the two classes (Fig. 2 C, F, and J). We concluded that the analysis of a large number of cell lines overcomes the inherent heterogeneity of cell lines and reveals robust relationships between essentiality and particular cell features that persist across different genetic and epigenetic backgrounds. With this foundation, we undertook a preliminary exploration of what can be learned about genetic vulnerabilities of cancer cells with specific properties.

**Lineage-Specific Genetic Dependencies.** We hypothesized that a subset of genes showing enhanced dependencies in specific lineages would also be aberrantly activated in tumors due to amplification or overexpression. Recent studies have identified oncogenic transcription factors that are amplified, overexpressed, and essential in subsets of tumors from cancers of specific lineages, including NKX2-1 in lung adenocarcinoma (8), MITF in melanoma (9), and SOX2 in squamous cell carcinomas (10). To identify lineage-specific dependencies, we ranked shRNAs by their ability to discriminate cell lines of one lineage from cell lines from other lineages (Fig. 3A). We selected the top 150 genes (1.3% of those screened) based on ranking of the top-ranked shRNAs, the top 300 genes (2.7%) by the second-best ranking shRNAs, and the top 300 genes (2.7%) as assessed by the KS statistic (ref. 3; Table S3). Three categories of essential genes were considered for further analysis: (i) genes scoring by all three methods from individual lineage analyses; (ii) genes scoring by any method that also were amplified in primary tumors (essential and amplified; Table S3); and (iii) genes scoring by any method that also were differentially up-regulated in cell lines derived from that lineage (essential and overexpressed; Tables S3 and S4). An overview of these results from analyses performed across six cancer lineages is displayed in Table S5.

In colon cancer cell lines, we found KRAS, CTNNB1, and BRAF among 23 essential genes that scored by all three methods and each gene, (ii) WoE rank of the top two shRNAs targeting each gene, or (iii) a composite score of WoE ranks for all shRNAs for each gene using the Kolmogorov–Smirnov (KS) statistic (3). The KRAS and BRAF genes themselves were ranked highly by all three methods (Fig. 2 A and D). The top scoring KRAS- and BRAF-specific shRNAs significantly discriminated between the mutant and wild-type classes (\(P = 1.89 \times 10^{-3}\) and \(1.89 \times 10^{-4}\), respectively, WoE; Fig. 2 B and E).

Interestingly, because 7/10 BRAF-mutant cell lines were derived from nonmelanoma lineages (including 5 from colon cancer), these observations suggest that cancer cell lines that harbor mutant BRAF exhibit a similar dependence on BRAF. Although initial reports suggest that clinical responses to BRAF inhibition in BRAF-mutant colon tumors are much less robust than those observed in BRAF-mutant melanomas (6), our observations indicate that BRAF is essential in colon cancer cell lines that express mutant BRAF (Fig. S4 A and B).

To ensure that the pooled screening process accurately measured the activity of the shRNAs, we individually retested 350 shRNAs, chosen to sample the full range of array measurements obtained in a pooled screen in OVCAR-8 cells, by performing competition assays (Fig. 1C and Fig. S3). Specifically, we cloned these shRNAs into a modified pLKO.1 viral vector that coexpresses GFP and used these vectors to introduce shRNAs into ~50% of an OVCAR-8 cell population. We then monitored the proportion of GFP-expressing cells over time to measure the effects of each shRNA on proliferation. The percent depletion of the shRNAs in the individual pairwise competition tests was correlated to the log(fold depletion) of these shRNAs in the pooled screen \((R^2 = 0.58\); Fig. 1C and Table S2). We note that shRNAs that show the largest degree of depletion in the pooled screen exhibited more variability as would be expected at the limit of signal detection. These two correlated assessments were made at different time points, further confirming that they provide robust measures of the intrinsic proliferation effects of the individual shRNAs.

**Correlation of Genetic Dependency with Properties of Cell Lines.** We then sought to understand how the vulnerabilities of cancer cell lines relate to various properties, such as mutations in a specific gene, disruption of a specific pathway, or inclusion in a specific lineage. Because human cancer cell lines are genetically and epigenetically diverse, the analysis of a large number of cell lines ensures that the relationships between cell properties and the dependence of those cells on specific genes are not particular to one context. We tested whether this large dataset permits reliable inferences about the genetic vulnerability of cancers possessing specific properties. For each cell line classification, we used a class-discrimination feature selection method called the “weight of evidence” (WoE) statistic (4, 5) to rank shRNAs by their ability to distinguish the specified classes.

**Dependencies of Cell Lines with Oncogenic Mutations.** We first examined vulnerabilities of cell lines with KRAS or BRAF mutations. We defined “essential” genes by three complementary methods, including (i) WoE rank of the top shRNA targeting
found KRAS and IGF1R (11) among 35 essential and amplified genes. In pancreatic cancer cell lines, we identified KRAS among 23 essential genes that scored in all three analyses (Table S5). In NSCLC, we found NKX2-1 as the only essential gene that is both amplified and overexpressed and found CDK6 among 7 essential and amplified genes (8). These observations provide evidence that such integrative lineage analyses identify both known oncogenes and other relevant lineage-restricted dependencies.

In addition, we identified a number of particularly interesting, previously undescribed candidate lineage-specific dependencies. For example, we found MAP2K4, an activator of JNK and p38 (12), among 10 genes that showed selective essentiality and expression in NSCLC. We found MYB (13) and AXIN2 (14) among 9 genes that were essential and differentially expressed in colon cancer, and we identified SOX9 (15) among 18 genes in pancreatic cancer that emerged as lineage-specific dependencies nominated by all three gene-scoring methods.

To extend these observations, we selected the ovarian lineage for deeper analysis. Of the 582 genes (5.2%) nominated as candidates for enhanced dependency in ovarian cancer cell lines, we identified 22 essential genes that scored in all three analytical methods (Fig. 3B and Table S5) and found 5 essential and overexpressed genes (Tables S4 and S5). TCGA identified 1,825 genes residing on recurrently amplified regions in ovarian tumors, and we identified 50 amplified genes as also essential (Fig. 3B and Table S5). The set of amplified and essential genes included the known oncogene CCNE1 (16) and candidates including the FRS2 adaptor protein (17), the PRKCE protein kinase (18), RPTOR (19), and the PAX8 paired box transcription factor. Similarly to NKX2-1 in NSCLC, MITF in melanoma, and MYB in colon cancer, PAX8 was not only essential and amplified but also overexpressed in a lineage-specific manner.

Characterization of PAX8 Dependency in Ovarian Cancer. PAX8 was the only gene that was (i) identified as an essential gene in all three scoring methods, (ii) amplified in primary high-grade serous ovarian tumors, and (iii) differentially expressed in ovarian cancer cell lines (Table S5). Cell line subsampling analysis revealed that the large number of ovarian cell lines screened enabled the identification of this previously undescribed dependency (Fig. 3C).

PAX8 is a lineage-restricted transcription factor that plays an essential role in organogenesis of the Mullerian system (20), the thyroid, and the kidney (21). PAX8 was previously found to be

---

Fig. 2. Dependencies of cell lines with mutations in KRAS, BRAF, or PIK3CA. (A, D, and G) Distribution of shRNA ranks (x-axis) by the WoE scores (y-axis) for the class comparisons of KRAS mutant (mut) vs. KRAS wild-type (wt) cell lines (A), BRAF mutant vs. BRAF wt (D), and PIK3CA mutant vs. PIK3CA wt (G). shRNAs targeting KRAS, BRAF, and PIK3CA are highlighted in red, and their ranks are listed. Insets report the gene ranks of KRAS, BRAF, and PIK3CA for differential essentiality in the subset of cell lines with activating mutations in those respective genes. (B, E, and H) KRAS (B), BRAF (E), or PIK3CA (H) mutation status (mutant lines in green, wt lines in gray) correlates strongly with depletion of shRNAs targeting these genes. (Lower) Heatmaps report KRAS (B), BRAF (E), and PIK3CA-shRNA (H) fold depletion in each cell line. (C, F, and I) Subsets of the 102 cell lines were analyzed to assess convergence of the gene dependency results for the KRAS (C), BRAF (F), and PIK3CA (I) mutant vs. wt class comparison analyses as a function of the number of cell lines tested. Distributions of the scores in the top KRAS, BRAF, and PIK3CA hit shRNAs (given as the percentile of their depletion rankings, with a smaller percentage corresponding to more depleted, y axis) in the respective cell line class comparisons (using WoE) are shown for each of 100 trials, subsampling the indicated numbers of cell lines in each class (mutant and wt). The red bar indicates the median value for each group of subsamplings, boxes represent the 25th to 75th percentile of the data, and whiskers extend to the most extreme values of the group that are not considered outliers.
overexpressed in ovarian cancers (22) and implicated in follicular thyroid cancer development (23). We observed that PAX8 was the most differentially expressed gene when we compared ovarian cell lines to all of the other cancer cell lines (Fig. 4A). Furthermore, we found that PAX8 was amplified in 16% of primary ovarian tumors [log2(copy number ratio) > 0.3; n = 345] in a peak (2q13) that also contains PSD4 and LOC654433 (Fig. 4B).

We further examined the relationships among PAX8 amplification, PAX8 expression, and dependence on PAX8 in ovarian cancer cells. The PAX8-specific shRNA that scored 7th out of 54,020 shRNAs and the 2nd-ranking PAX8 shRNA both suppressed PAX8 (Fig. S4E). The sensitivity of cell lines to inhibition by the highest ranked PAX8-specific shRNA correlated with the level of PAX8 expression, based on comparison of cell lines with high vs. low PAX8 levels (P = 2.14 × 10−8; t test; Fig. 4C). Cell lines expressing high levels of PAX8 included the vast majority (21/25) of ovarian cancer cell lines as well as a renal and an endometrial cancer line. These observations suggested that the expression of PAX8 is selectively required for the proliferation/survival of cell lines expressing PAX8.

To confirm these findings, we introduced two distinct shRNAs targeting PAX8 into 17 cell lines. We found that suppression of PAX8 resulted in a >50% reduction in the viability in six of eight ovarian cancer cell lines (Fig. 4D), but failed to affect the proliferation of immortalized human ovarian surface epithelial cells (IOSE-T80; ref. 24) and eight other cell lines that did not express PAX8 (Fig. 4E). The six sensitive ovarian cell lines included three cell lines with amplification of 2q13 in which the PAX8 locus resides (Fig. S4F) and three additional cell lines that express higher levels of PAX8 protein compared with IOSE-T80 cells (Fig. 4D). Suppression of PAX8 induced apoptosis in these ovarian cancer cell lines (Fig. 4F). In contrast, the two cell lines (COV504 and OV-90) least sensitive to PAX8 suppression did not harbor the 2q13 amplification and expressed relatively low levels of PAX8 (Fig. 4D). These observations suggest that PAX8 represents a lineage-specific essential gene in a significant subset of ovarian cancer.

Discussion

The integrated analysis of functional dependencies and alterations in cancer genomes presented herein identified potential targets in ovarian, lung, colon, glioblastoma, pancreatic, and esophageal cancers. Among these candidate genes, we identified known oncogenes and lineage-specific dependencies as well as previously undescribed candidates, including the PAX8 transcription factor in ovarian cancer. Although shRNA screens performed in small numbers of cell lines have identified essential genes in specific contexts, the interrogation of genes across a large number of human cancer cell lines through Project Achilles provides a substantially more robust assessment of gene dependence and overcomes confounding effects due to the inherent heterogeneity of human cancer cell lines. These datasets will enable a wide range of analyses to connect particular cancer genotypes to dependencies.

As an initial approach, we elected to explore dependencies harbored by a majority of ovarian cancers. We pinpointed 5 genes displaying enhanced essentiality in ovarian cancer and differential overexpression in ovarian cell lines and 50 genes displaying enhanced essentiality in ovarian cancer and amplification in ovarian tumors. Further studies will be necessary to
extend the preliminary findings for these ovarian cancer dependencies. A single gene, PAX8, emerged from every one of these analyses. Our subsequent experiments confirmed PAX8 to be a lineage-specific survival gene that is essential for proliferation of ovarian cancer cells, highly expressed in ovarian cancer lines, and amplified in a substantial fraction of primary ovarian tumors.

PAX8 has been shown to play an essential role in the normal development of the thyroid gland (21) and female genital tract (20). The thyroid gland in Pax8-deficient mice is completely devoid of thyroid hormone-producing follicular cells and exhibits severe growth retardation within a week after birth (21). Thyroxine substitution enables Pax8-deficient mice to survive to adulthood (20). However, these mice are infertile because of the absence of uterus and vaginal openings (20), indicating that PAX8 is also essential for the development of the Mullerian duct. In the reproductive tract, PAX8 expression is restricted to secretory cells of the fallopian tube epithelium (22), which recent reports suggest represent a cell of origin for serous ovarian cancer (25). These findings suggest that PAX8 plays critical roles both in normal development of female genital tract and in high-grade serous ovarian tumors.

A number of genes involved in the differentiation programs for specific tissue lineages are amplified in cancers that arise from these tissues. For example, NKK2-1 (8), MITF (9), and SOX2 (10) are amplified and essential for the survival of significant subsets of NSCLC, melanoma, and squamous cell carcinomas, respectively. Our results show that PAX8 belongs to this distinct class of lineage-survival genes that are required for both normal development of specific tissues and for cancer cell proliferation/survival. Although different subtypes of ovarian cancer are likely to harbor distinct profiles of genetic alterations (2), our findings suggest that for most ovarian cancer cell lines, PAX8-driven transcription programs transiently active during normal development are coopted to maintain the malignant state.

Epithelial ovarian cancers have been classified into four major subtypes based on histology: serous, clear cell, endometrioid, and mucinous (2). These major subtypes show morphologic features that resemble those of the epithelia of the reproductive tract derived from the Mullerian duct (2). Previous studies using immunohistochemistry showed that 89–100% of serous, clear cell, and endometrioid subtypes and 8% of mucinous subtype express PAX8 (26). The majority of the ovarian cancer cell lines that we screened (20/25) belong to the serous subtype or had a mixed morphology. A larger collection of cancer cell lines will facilitate a deeper investigation of ovarian cancer subtypes.

Given the genetic heterogeneity of cancer, screening large numbers of cell lines is required to have sufficient statistical power to extract known and novel relationships and provide adequate representation of individual sublineage classes. Although the
classification of cell lines based on a single characteristic is unlikely to segregate cell lines into homogeneous groups, our observations indicate that by including many cell line representatives of each class of interest, it is still possible to discover underlying relationships between genotype or phenotype and essential genes. Integration of these data with information from the analyses of structural alterations in cancer genomes will further facilitate the systematic identification of genes critical to oncogenesis.

The approach described here can be extended to enable systematic interrogation of codependences beyond those analyzed here, including synthetic lethal relationships with activated oncogenes or inactivated tumor suppressor genes (27). Analyses of dependencies of cell lines with particular characteristics have the potential to discover novel targets, for example, by integrating the output of these types of screens with information emerging from the cataloging of mutations and other alterations in cancer genomes.

More generally, as large-scale efforts to characterize cancer genomes accelerate, these observations illustrate a path to functionally characterize the genes found to be altered in tumors and to identify the subset of such genes critical to cancer initiation and maintenance. To this end, we have made this dataset available (www.broadinstitute.org/igg) and will update the Project Achilles evidence data as additional data are obtained. Beyond the specific findings reported herein, we anticipate that this dataset will prove useful to identify correlations between genetic features and essential genes in human cancer cell lines.

Materials and Methods

Pooled shRNA Screening. Lentiviral pLKO.T-shRNA constructs were obtained from the RNAi Consortium, and the human 54K pool of 54,020 shRNA plasmids was assembled by combining 16 normalized subpools of ∼3400 shRNA plasmids. The list of 54,020 shRNAs can be found at http://www.broadinstitute.org/igg. Genome-scale pooled shRNA screens to identify genes essential for proliferation in 102 cancer cell lines were performed (3) using a lentivirally delivered pool of 54,020 shRNAs targeting 11,194 genes. The culture conditions for all cancer cell lines are listed in Table S1. Each cell line was infected in quadruplicate and propagated for at least 16 population doublings. The abundance of shRNA constructs relative to the initial DNA plasmid pool was measured by microarray hybridization (3) and analyzed by using a uniform pipeline. Detailed descriptions of each procedure can be found in SI Methods.

Data Processing, Class Comparison, and Gene Ranking. Raw .CEL files from custom Affymetrix barcode arrays were processed with a modified version of dChip software (3). The GenePattern modules shRNAscore and NormalizedCellLines were used to calculate the log fold change in shRNA abundances for each cell line at the conclusion of the screening relative to the initial plasmid DNA reference pool and to normalize these depletion values by using peak median absolute deviation normalization, a variation of Z score with median absolute deviation (3). Class definition files (.cls) were made by using the GenePattern module SubsetGetAndCls; definitions included cell line lineage (e.g., ovarian cancer, NSCLC, etc.) or genetic alterations (28, 29).

To compute the statistical evidence that a given shRNA contributes to the observed essentiality phenotype between two classes of interest, we used a WoE approach (4, 5). The Gene-E program (http://www.broadinstitute.org/cancer/software/GENE-E; ref. 3) was used to collapse shRNAscores to gene rankings by three complementary methods. These methods included (i) ranking genes by their highest shRNA depletion score, (ii) ranking genes based on the P value rank (correcting for different sets of shRNA targeting different genes) of their second best ranked shRNA, and (iii) ranking genes using a KS statistic in an approach similar to gene set enrichment analysis (5). The RNAi gene enrichment ranking (3) was developed for each procedure can be found in SI Methods. All data files, accessory files, and GenePattern modules can be found through the Integrative Genomics Portal (http://www.broadinstitute.org/igg).

Competition Assay. OVCAR-8 (5 × 10^6) cells were seeded into each well of a 96-well plate and spin-infected with 2 or 4 μL of lentiviruses (in duplicate) at 930 × g for 2 h at 30 °C in the presence of 4 μg/mL polybrene to transduce ∼50% of the cells. Cells were then trypsinized and replated into 24-plate wells. The percent of GFP+ cells at 3 and 7 d after infection was measured using BD LSR II flow cytometry system equipped with a high-throughput sampler (BD Biosciences). The fraction of GFP+ cells 7 d after infection relative to 3 d after infection was calculated. Data represent mean ± SD of duplicate infections.

ACKNOWLEDGMENTS. We thank M. Reich, A. Derr, and T. Liefield for assistance with analysis and data portal creation. This work was supported in part by National Institutes of Health Grants R33 CA128625, U54 CA112962, and RC2 CA148268, Novartis Pharmaceuticals, Richard Hunt, and the H. L. Snyder Foundation.


Cheung et al.

PNAS | July 26, 2011 | vol. 108 | no. 30 | 12377