CANCER DISCOVERY



Essential Gene Profiles in Breast, Pancreatic, and Ovarian Cancer Cells

Richard Marcotte, Kevin R. Brown, Fernando Suarez, et al.

Cancer Discovery 2012;2:172-189. Published OnlineFirst December 29, 2011.

Updated Version	Access the most recent version of this article at: doi:10.1158/2159-8290.CD-11-0224
Supplementary Material	Access the most recent supplemental material at: http://cancerdiscovery.aacrjournals.org/content/suppl/2011/12/29/2159-8290.CD-11-0224. DC1.html

Cited Articles	This article cites 78 articles, 24 of which you can access for free at: http://cancerdiscovery.aacrjournals.org/content/2/2/172.full.html#ref-list-1
----------------	---

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

Essential Gene Profiles in Breast, Pancreatic, and Ovarian Cancer Cells

Richard Marcotte⁴, Kevin R. Brown¹, Fernando Suarez⁴, Azin Sayad¹, Konstantina Karamboulas¹, Paul M. Krzyzanowski⁴, Fabrice Sircoulomb⁴, Mauricio Medrano^{3,4}, Yaroslav Fedyshyn¹, Judice L.Y. Koh¹, Dewald van Dyk¹, Bohdana Fedyshyn¹, Marianna Luhova¹, Glauber C. Brito¹, Franco J. Vizeacoumar¹, Frederick S. Vizeacoumar⁵, Alessandro Datti^{5,7}, Dahlia Kasimer¹, Alla Buzina¹, Patricia Mero¹, Christine Misquitta¹, Josee Normand⁴, Maliha Haider⁴, Troy Ketela¹, Jeffrey L. Wrana^{2,5}, Robert Rottapel^{3,4,6}, Benjamin G. Neel^{3,4}, and Jason Moffat^{1,2}

ABSTRACT

Genomic analyses are yielding a host of new information on the multiple genetic abnormalities associated with specific types of cancer. A comprehensive description of cancer-associated genetic abnormalities can improve our ability to classify tumors into clinically relevant subgroups and, on occasion, identify mutant genes that drive the cancer phenotype ("drivers"). More often, though, the functional significance of cancer-associated mutations is difficult to discern. Genome-wide pooled short hairpin RNA (shRNA) screens enable global identification of the genes essential for cancer cell survival and proliferation, providing a "functional genomic" map of human cancer to complement genomic studies. Using a lentiviral shRNA library targeting -16,000 genes and a newly developed, dynamic scoring approach, we identified essential gene profiles in 72 breast, pancreatic, and ovarian cancer cell lines. Integrating our results with current and future genomic data should facilitate the systematic identification of drivers, unanticipated synthetic lethal relationships, and functional vulnerabilities of these tumor types.

SIGNIFICANCE: This study presents a resource of genome-scale, pooled shRNA screens for 72 breast, pancreatic, and ovarian cancer cell lines that will serve as a functional complement to genomics data, facilitate construction of essential gene profiles, help uncover synthetic lethal relationships, and identify uncharacterized genetic vulnerabilities in these tumor types. Cancer Discovery; 2(2); 172-89. © 2011 AACR.

INTRODUCTION

Recent technological advances have revolutionized our understanding of cancer genetics. Transcriptional profiling, copy number variation, and deep sequencing have revised the classification of many tumors into molecular subtypes that provide improved prognostic information compared with conventional clinical and histopathologic classification schemes (1, 2). Yet often, these subtypes provide little functional information about the molecular events that drive cancer cell behavior. Genome-wide sequencing studies have identified hundreds to thousands of mutations in individual tumors (3-6), yet it frequently is difficult to know which of these are essential for pathogenesis (i.e., "drivers"), as opposed to "passenger" mutations. Even when a driver oncogene (e.g., KRAS, MYC) or tumor suppressor gene (e.g., TP53, BRCA1/2) is known, these can be poor targets for drug development. In addition, unanticipated gene/pathway dependencies

R. Marcotte, K.R. Brown, and F. Suarez contributed equally to this article.

doi: 10.1158/2159-8290.CD-11-0224

© 2011 American Association for Cancer Research.

("synthetic lethality") can arise as a consequence of the genetic abnormalities in cancer cells, as illustrated by the sensitivity of BRCA1/2-mutant breast cancer cells to PARP inhibitors (7, 8). The systematic identification of such synthetic lethal relationships might suggest new drug targets (9). Comparison of the genetic abnormalities and functional vulnerabilities of cancer cells should help researchers identify new drivers and provide insight into the complex systems biology of cancer.

RNA interference technology has enabled genome-wide loss-of-function screens in mammalian cells. Most screens have used siRNAs, usually arrayed in multiwell plates. Arrayed screens have focused mainly on specific gene families, such as kinases, phosphatases, or selected candidate genes, and have yielded new insights into mechanisms of cancer cell signal transduction, cell division, and cell death (10, 11). Cell proliferation assays in multiwell plates are usually constrained to a few population doublings, and gene "knockdowns" in these conditions typically last for at most a week. Therefore, siRNA screens are, by nature, transient, and might underestimate the roles of long-lived proteins on a given phenotype. Moreover, given their cost and the need for extensive automation to interrogate most of the genome, siRNA screens usually are performed on only limited numbers of cell lines and might fail to capture the genetic heterogeneity in cancer. These properties make it difficult to use arrayed screening approaches to construct extensive functional genomic maps of cancer cells.

The more recent development of large retroviral- or lentiviral-based short hairpin RNA (shRNA) libraries facilitates genome-wide screening of cultured cancer cells in a pooled format (12–14), providing a potential solution to the limitations of arrayed screens. Cells are infected with these libraries at a low multiplicity of infection (MOI) and allowed to proliferate for 3 to 4 weeks, after which shRNAs that have been selectively depleted (referred to as "dropouts") or enriched are identified

Authors' Affiliations: ¹Donnelly Centre and Banting & Best Department of Medical Research, Departments of ²Molecular Genetics and ³Medical Biophysics, University of Toronto; ⁴Campbell Family Cancer Research Institute, Ontario Cancer Institute, Princess Margaret Hospital, University Health Network; ⁵Samuel Lunenfeld Research Institute; ⁶Division of Rheumatology, Department of Medicine, St. Michael's Hospital, Toronto, Canada; ⁷Department of Experimental Medicine and Biochemical Sciences, University of Perugia, Perugia, Italy

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Authors: Jason Moffat, Donnelly Centre, 160 College Street, Room 802, University of Toronto, Toronto, Canada M5S 3E1. Phone: 416-978-0336; Fax: 416-978-8287; E-mail: j.moffat@utoronto.ca. Robert Rottapel. E-mail: rottapel@uhnresearch.ca; Benjamin G. Neel. E-mail: bneel@uhnresearch.ca

Marcotte et al.

on custom microarrays or by deep sequencing. Pooled screens have been used to define genes necessary for cancer cell proliferation/survival in cell culture (12–14), genes that enhance or interfere with the action of specific oncogenes (15), or genes that enhance the effects of antineoplastic drugs, suggesting potential new combination therapies (16, 17).

Most large-scale pooled shRNA screens have surveyed cancer cell lines representing multiple histotypes but usually with few representatives of any one tumor type, or they have focused on cell lines from different histotypes bearing the same genetic abnormality (e.g., KRAS mutations) (15, 18). As an initial step toward a more comprehensive understanding of the vulnerabilities of breast cancer (BrCa), pancreatic ductal adenocarcinoma (PDAC), and high-grade serous ovarian carcinoma (HGS-OvCa), we performed near genome-wide pooled shRNA screens on 72 cancer cell lines and established a unique informatics approach to monitor the dynamic evolution of cancer cell populations. We chose breast cancer because the extensive genomic information and subtype classification schemes that exist for this tumor type facilitate integrated genomic/functional genomic analysis. Ongoing genomic efforts should provide similar information for PDAC and HGS-OvCa, but we focused on these malignancies primarily because they typically are detected at an advanced stage, their prognosis remains dismal, and there is therefore an urgent need to define new therapeutic targets. Our large functional genomic data set can be used in conjunction with orthogonal efforts to map the structural variation within cancer genomes, such as The Cancer Genome Atlas (TCGA) or the International Cancer Genome Consortium (19), to accelerate the identification of drivers. Initial analysis revealed only partial overlap between genomic and functional genomic classifications of cancer and uncovered novel, unanticipated, cancer cell-specific dependencies in these 3 major types of cancer, some of which could be amenable to targeted therapies.

RESULTS

Classifying shRNA Activity across a Compendium of Pooled shRNA Screens

To catalogue essential genes across a defined set of cancer types, we performed genome-wide pooled screens using a library of 78,432 shRNAs targeting 16,056 unique Refseq genes ("80K" library, Supplementary Table S1), developed by The RNAi Consortium (TRC) (20–22). A total of 72 cancer cell lines were screened: 29 breast, 28 pancreatic, and 15 ovarian (Fig. 1A and Supplementary Table S2). Each line was screened in triplicate, and at least 3 time points were assessed for overall shRNA abundance during population outgrowth. The screens were highly reproducible between replicate biological populations for all of the cell lines (R_{avg} (BrCa) = 0.9; R_{avg} (PDAC) = 0.92; R_{avg} (HGS-OvCa) = 0.87). The result was a data set containing more than 50 million data points from more than 200 independent cell populations.

Current scoring algorithms for shRNA and siRNA screens assess dropouts at only a single time point. We reasoned that adding additional time points would provide a detailed history of individual shRNA performance, allow us to model shRNA kinetics during population outgrowth, and increase our confidence in the essentiality score derived for each gene. We also developed a set of heuristics to classify shRNAs as fast, continuous, or slow dropouts, based on the rate at which an shRNA disappeared from the bulk population of cells during the screen (see Methods and Supplementary Table S3). Examples of these profiles are shown at the right of Figure 1A. Using heuristics designed to identify the most potent shRNAs in the fast, continuous, or slow classes resulted in the classification of ~2% of the shRNAs in the library into one of these categories, with 40% fast, 30% continuous, and 30% slow dropouts. These classification criteria largely restricted hairpins to a single class. Moreover, dropout behavior largely appeared to be characteristic of the gene targeted by the hairpin rather than the shRNA itself: within any cell line, a given gene almost always fell into a single dropout class (Fig. 1B). Altering our heuristics would allow us to classify more hairpins but would result in greater overlap between dropout classes and also lead to classification of less potent hairpins. On average, ~0.4% of shRNAs were enriched in any given cell line; due to our shRNA bar code detection procedures (see Supplementary Table S4 and Methods), this is almost certainly a substantial underestimate of the true number of enriched hairpins (see Discussion).

To explore the hypothesis that classes of shRNAs are related to functional categories, we compared the biological functions of the gene targets [as assessed by gene ontology (GO) categories] of hairpins classified only as fast or continuous dropouts with those classified only as slow dropouts. Fast and continuous dropout shRNAs target genes enriched for the proteasome (e.g., PSMA1, PSMB2), ribosome (e.g., RPS17), splicing machinery (e.g., SNRPD2, SF3B2, AQR, HNRNPC, THOC1), metabolism of proteins (e.g., ARCN1, COPZ1), transcription (e.g., POLR2D, POLR2E, PABPN1), and translation (e.g., EIF3B), all of which are highly conserved housekeeping functions (Fig. 1C and Supplementary Fig. S1). Conversely, shRNAs classified as slow dropouts target genes that are enriched for regulation of protein phosphorylation, signaling, signal transduction, and kinase activity (e.g., PTPRG, EPGN, PPP1R3B, RNF128, SERPINA3, CSNK2B, ST3GAL3, UBOX5, TNKS2). These results suggest that classifying hairpins on the basis of dropout rate reveals different functional properties of the genes that they target, providing further evidence that hairpin behavior usually reflects the properties of the underlying target gene, rather than the specific hairpin per se.

Validation of Fast and Slow Dropouts Defined by Hairpin Class

To assess further the validity of our results, we performed secondary screens using an orthogonal, siRNA-based assay. Fifty genes (Supplementary Table S5) available in the Dharmacon SMARTpool siRNA library were selected for further analysis. All of the shRNAs for each gene chosen fell into a single dropout class in our original screen, representing either the fast (n = 17) or slow (n = 33) categories. The chosen siRNAs were transfected into MDA-MB-231, MIA PaCa-2, and KP-4 cell lines. After 7 days, cells were enumerated and compared with a mock-transfected population. Most (80%–93%) of the fast dropout and 29% to 38% of the slow dropout genes inhibited growth significantly (Supplementary Table S5B; adj. P < 0.05, Wilcoxon rank sum test) in this assay (Fig. 1D

RESEARCH ARTICLE



Figure 1. Outline of procedure for time-course shRNA screening. **A**, schematic representing steps involved in shRNA functional screening. **B**, hairpins were classified based on heuristic rules (see Supplementary Table S3). The proportion of genes falling into each category is shown as a stacked-bar chart. The pink, white, and cyan bar segments detail genes that are targeted by a single hairpin class, while the red and blue segments indicate the sparse overlap between classes. Stacked bars do not reach a value of 1.0 as the minimal overlap observed between fast and slow, and all 3 classes are omitted. The full classification results are shown in Supplementary Table S3. **C**, enrichment plot of G0 "biological process" terms and MSigDB pathways within classified hairpins. Darker blue colors indicate more significant *P* values corresponding to the enrichment. The full set of enriched terms is shown in Supplementary Figure S1. **D**, validation of selected shRNA hits by siRNA. Growth inhibition due to target gene knockdown was determined relative to mock transfected controls in MDA-MB-231, MIA PaCa-2, and KP-4 cell lines. Red dashed lines indicate the mean of 60 replicate mock transfections, while the blue dashed lines indicate mean ± 3 *SD*.

AC American Association for Cancer Research

Marcotte et al.

and Supplementary Fig. S2). These findings again indicate that the shRNA kinetics detected by our pooled shRNA screens reflect the properties of specific genes (i.e., fast dropouts are more likely than slow dropouts to score in this short-term assay), rather than the quality/knockdown efficiency of the shRNA reagents targeted against a given gene.

Conversion of shRNA Class Behavior into an Activity Score

To convert time-course information from dropout screens into an individual value for each shRNA, we developed the "shARP" (shRNA Activity Rank Profile) score. The shARP score assigns a value to each hairpin by calculating the average slope between the measured microarray expression intensity at each time point and the intensity at time zero (i.e., T_0), producing a weighted average of the fold-change across time and accounting for the growth rate of the cell line (see Methods). By integrating information across the time course, shARP can discriminate among "fast," "continuous," and "slow" hairpins: in general, slow dropouts have less negative shARP scores than fast or continuous dropouts (Supplementary Fig. S3A).

Because each gene represented in the TRC library is targeted by an average of 5 hairpins, shARP scores for the different shRNAs must be converted into a gene-level score to uncover the behavior of specific genes in a screen. To this end, we defined the GARP (Gene Activity Ranking Profile) score (see Methods) as the average of the 2 lowest shARP scores. We then compared the performance of the GARP score with 2 previously developed scoring metrics, RNAi gene enrichment ranking (RIGER) (12) and redundant siRNA activity (RSA) (23), for their respective ability to rank the genes in our large panel of screens. These methods differ in how the shRNA sets for a given gene are treated. GARP, RIGER "Weighted sum" (RIGER_WS), and RIGER "Second-best hairpin" (RIGER_SB) consider only the 2 best hairpins or the second-best single hairpin, respectively. By contrast, RIGER "Kolmogorov-Smirnov" (RIGER_KS) and RSA consider the behavior of the complete set of hairpins against a given gene.

In the absence of a "gold standard" set of essential human genes, we benchmarked the 5 scoring approaches using 2 largely nonoverlapping reference sets likely to be enriched for essential genes (see Methods): highly conserved genes from 8 diverse species, including *Arabidopsis thaliana, Bos taurus, Caenorhabditis elegans, Canis familiaris, Macaca mulatta, Mus musculus, Rattus norvegicus,* and *Saccharomyces cerevisiae* (24); and housekeeping genes (25). If, as seems reasonable, one assumes that housekeeping and ortholog gene sets are enriched for essential genes, then the intersection between the top 500 genes from each scoring method and each of the reference gene sets indicated that GARP outperforms the other scoring approaches in representative breast (HCC1187), ovarian (OVCAR5), and pancreatic (HPAF-II) cancer cell lines (Fig. 2A).

To examine the performance of each scoring metric more thoroughly, we determined the intersection of scored genes and reference sets across our entire panel of 72 cell lines, then computed the area under the curve (AUC) for each overlap. The cumulative distribution of AUCs for each scoring method and reference set is shown in Figure 2B, which depicts the fraction of scored gene sets with an AUC greater

than a sliding threshold on the x-axis. Again, GARP performed better overall than the other scoring methods, when considering either the conserved gene reference set or the housekeeping gene reference set. Although RSA and RIGER_ KS consider all shRNAs targeting each gene, RSA consistently produced higher AUCs, presumably because of its different underlying statistical approach. More important, the genes that drive the performance of GARP in each of the ortholog and housekeeping gene sets are largely nonoverlapping (Fig. 2C). By examining the top 500 hits as defined by each method, we found that GARP identifies a set of genes that are common to RSA and RIGER_KS, as well as a unique subset of genes that are distinct from the overlap between RSA and RIGER_KS in HCC1187, OVCAR5, and HPAF-II cell lines (Fig. 2D). In general, the unique subsets of genes identified by GARP in the 72 cell lines account for its superior ability to identify genes in the housekeeping or highly conserved reference sets.

Correlating Gene Activity Scores with Target Expression Levels

All RNAi screens are susceptible to off-target effects (11, 26). Although it is difficult to quantify such effects from screening data alone, if off-target effects are common, one might expect hairpins directed against nonexpressed genes to "score" as frequently as those targeting expressed genes. We performed RNA-seq on 7 ovarian cancer cell lines and determined the fraction of nonexpressed genes that were adjudged "essential" by GARP and the other published scoring metrics (Fig. 2E). Regardless of the scoring method, the vast majority of "hits" reflected genes that were expressed, defined as "reads per kilobase of exon model per million mapped reads" (RPKM) greater than zero. However, compared with the other scoring systems, GARP identified fewer nonexpressed genes as "hits."

We also used receiver operating characteristic curves to compare the relative ability of each score to "call" essential transcripts that also were expressed. Notably, the AUCs computed by GARP ($\mu = 61.8 \pm 5.6\%$) were higher than those computed by RSA ($\mu = 56.1 \pm 3.5\%$; P = 0.013), RIGER_KS ($\mu = 56.4 \pm 4.3\%$; P = 0.006), and RIGER_SB ($\mu = 55.3 \pm 4.2\%$; P = 0.016), using a paired *t* test with unequal variances. The difference between GARP and RIGER_WS ($\mu = 59.6 \pm 5.4\%$; P = 0.079) was not significant, although GARP still outperformed RIGER_WS in 5 of 7 (71%) cell lines. Taken together, these results suggest that GARP performs at least as well as, and by many metrics superior to, previous scoring systems for shRNA screens (see Discussion).

A Snapshot of Gene Essentiality across Three Major Tumor Types

We looked for essential genes across all of the cell lines. We identified 297 genes with significant GARP scores ($P \le 0.05$) in at least 50% of the 72 screens (Supplementary Table S6). These "general essential" genes were enriched in housekeeping functions involving the ribosome (P = 2.4e-50; Fisher's exact test), proteasome (P = 1.4e-11; Fisher's exact test), spliceosome (P = 7.7e-28; Fisher's exact test), DNA replication (P = 2.5e-3; Fisher's exact test), protein metabolism

Discovering Cancer Dependencies with Essential Gene Profiles



Figure 2. Comparison of shRNA scoring approaches. **A**, overlap of genes ranked by GARP, RIGER (RIGER_KS, RIGER_SB, and RIGER_WS), and RSA with 2 data sets enriched in essential genes: "housekeeping (HK) genes" defined by gene expression (see Methods), and human genes conserved in 8 eukaryotic species. Representative results for the breast cancer line HCC1187 (left), ovarian cancer cell line OVCAR5 (center), and pancreatic cell line HPAF-II (right) are shown. Overlaps are included for 10 samples of randomly selected genes (black lines, mean ± *SD*). **B**, cumulative distribution plots of AUC for each overlap. At each threshold on the x-axis, the fraction of cell lines with higher AUCs is indicated for each scoring method; left, overlap with highly conserved orthologs; right, overlap with housekeeping genes. **C**, stacked-bar chart showing proportion of genes from overlap of top 500 GARP-scored genes and reference sets (HK and orthologs) unique to HK set, unique to ortholog set, or common to both. **D**, Venn diagrams of top 500 ranked genes for each scoring method in cell lines HCC1187, OVCAR5, and HPAF-II. **E**, GARP shows the lowest rate of targeting nonexpressed transcripts. "Nonexpression rates" were determined by RNA-seq for genes ranked by GARP, RSA, and RIGER in 7 ovarian shRNA screens. The proportion of nonexpressed genes with RPKM = 0 was calculated for the top *N* ranked genes within each cell line, and the results were averaged for each scoring metric (see Methods).

AAC American Association for Cancer Research

FEBRUARY 2012 CANCER DISCOVERY | 177

Marcotte et al.

(P = 1e-20; Fisher's exact test), or mRNA processing (P = 1.3e-12; Fisher's exact test) (Fig. 3A), consistent with observations in the fast hairpin dropout class (Fig. 1C). Not surprisingly, general essential genes displayed potent inhibitory effects on cell growth; all general essentials were "fast dropouts" in at least one cell line and behaved as "fast dropouts" in an average of 14



Figure 3. Identification of general and tissue-specific essential genes. A, representation of selection criteria underlying definition of the 297 general essential genes. The histogram at right shows the number of cell lines in which each gene is found essential. The "Wordle" (inset) depicts the most frequently occurring KEGG pathways among the general essential genes, in which the most frequently occurring terms are illustrated in larger type. **B**, heat map of differentially essential genes in the 3 tumor types. Here, 155 genes were identified as specific to breast cell lines, 66 to ovarian cell lines, and 187 to pancreatic cell lines. The complete list of genes is available in Supplementary Table S8. **C**, range of zGARP scores for 6 tissue-specific genes. *P* values comparing the range of scores were determined using the Kruskal-Wallis test. cell lines (Supplementary Fig. S3B). Notably, there was significant overlap (P < 1.6e-109; Fisher's exact text) between general essentials identified in our study and in a recent report (18) that also used pooled shRNA screening but tested a smaller (~54,000 vs. ~78,000) set of shRNAs (Supplementary Fig. S4), and an in-depth cell line-by-cell line comparison showed significant overlap in 11 of 15 cell lines in common between the screens (see Supplementary Materials and Supplementary Table S7).

Next, we wished to determine whether the 72 cell lines could be classified on the basis of "functional genomics" (i.e., their relative sensitivity to shRNAs in the 80K library), and if so, how such groups would compare with their respective histotypes. Of the genes with a significant normalized GARP score (P < 0.05) in at least 2 cell lines (~5,508 genes), we identified the 10% most variable across all of the lines. These genes (n = 551) were subjected to unsupervised complete linkage hierarchical clustering using the Pearson correlation, which divided the cell lines into 2 major groups (Supplementary Fig. S5). Notably, cell lines derived from the same tissue did not cluster into the same groups, although one of the major clusters was markedly enriched (P = 8.7e-5) for breast cancer cell lines: 23 of 29 breast cancer cell lines appeared in this cluster, which contained a total of 36 cell lines. Accordingly, this cluster included genes that we classified as breast-specific (FOXA1, CDK4, SNW1, ATP5B, CLYBL, NDUFS7, CAD, INTS10,) or luminal/HER2-specific (TFAPC2, AF3GL2) in subsequent analyses (see below). Interestingly, even though all of the PDAC lines contained a KRAS mutation, these lines scattered throughout the 2 major clusters, with 10 PDAC lines in the breast-enriched set and 18 in the other major cluster (see Discussion).

We also performed a supervised analysis of the 72 cell line data set to identify 2 other types of essential genes: (1) tissuespecific and (2) subtype-specific essentials, the latter of which can be classified further into essential genes found in a subset of cell lines from different tumor types or essentials found in selected cell lines within a single tumor type. Nonparametric statistical testing identified 66 ovarian-, 187 pancreatic-, and 155 breast-specific essential genes, distinguishable by a significantly lower normalized GARP score (see Methods) in that tumor type (Wilcoxon rank sum test, P < 0.01) relative to the other 2 tumor types. A heat map of these tissue-specific essential genes is shown in Figure 3B. We also examined each list of tissue-specific essentials for gene set enrichment using the Molecular Signatures Database (27). The breast cancer-specific essentials were enriched for a broad array of functions [e.g., cell cycle (q = 7.1e-10), ubiquitin-mediated proteolysis (q = 1.1e-3), spliceosome (q = 9.3e-4), oxidative phosphorylation (q = 4.4e-3), snRNP assembly (q = 2.1e-3), and pyrimidine metabolism (q = 3.5e-4)]. Pancreatic tissuespecific essential genes were enriched for signaling in the immune system (q = 2e-3), the ETS pathway (q = 2.2e-2), and the LAIR pathway (q = 2.2e-2) (Supplementary Fig. S6).

More detailed examination of the tissue-specific essentials revealed that the pancreatic cancer-specific essentials included *KRAS* and the cell-cycle regulator *CDK6* (Fig. 3C and Supplementary Table S8). *KRAS* is mutated in most pancreatic tumors and promotes ETS-mediated transcription (28), which could account for the aforementioned enrichment for *ETS* pathway genes. Most pancreatic cancer lines fail to

RESEARCH ARTICLE

express the CDK inhibitor *CDKN2A* (data not shown), which might sensitize them to a reduction in CDK activity. Novel pancreatic cancer-specific essential genes included *DONSON* (Fig. 3C), a centrosomal protein involved in DNA-damage response signaling and genomic integrity (29); the histone acetyltransferase *MYST3* (30); and the transcriptional repressor *SSX4* (Supplementary Table S8), although the latter gene has been implicated in other tumor types [non-small cell lung carcinoma (NSCLC), endometrial, cervical] (31, 32). Also of note are *PRKAA1* (encoding the catalytic subunit of PKA) and *TRAF6*, which, along with *TLR4* (another hit in the screen), is involved in autophagy and adaptive and innate immune responses (33) (Supplementary Table S8).

Ovarian cancer-specific essentials of note included the cyclin-dependent kinase *CDK12* (P = 0.015), which narrowly missed our *P*-value cutoff of P < 0.01, is involved in RNA splicing and was recently found to be mutated somatically in 3% of HGS-OvCa cases (34). Two other ovarian cancerspecific essentials, *PIM1* and *CARD11*, have not been linked to HGS-OvCa but are involved in promoting cell-cycle progression and in NF-κB activation, respectively (35, 36), both of which are broadly implicated in oncogenesis.

Breast cancer-specific essentials included known mammary oncogenes such as *AKT1*, a downstream effector of another mammary oncogene, *PIK3CA* (37); and *CDK4*, which encodes a binding partner of cyclin D1; the latter is amplified in a substantial percentage of breast cancers (38). Other breast cancer-specific essentials reportedly are overexpressed in breast cancer, including *ERH*, *GRN*, and *KDM1A* (39–41); coactivate estrogen receptor α (ER α) (*SNW1*); or are involved in tamoxifen resistance (*ABCC2*) (17, 42, 43). *FOXA1* knockdown was particularly deleterious to ER⁺ breast cancer cell lines (see "supervised analysis of breast cancer cell lines" below, and Fig. 3C), consistent with its well-established role in ER action (44, 45), and its prognostic significance (46) in breast cancer.

Functional Screening Results Partially Recapitulate Breast Cancer Subtypes

Breast tumors can be classified by transcriptional profiling into multiple subtypes with different prognostic significance (2, 47). By contrast, unsupervised analysis clusters breast cancer cell lines into 2 major subgroups, basal and luminal/ HER2. The basal cluster can be divided further (by unsupervised clustering) into Basal A and Basal B, with the Basal A subgroup most reminiscent of classic triple-negative breast tumors and Basal B enriched for the recently described claudin-low tumor subtype (48). The luminal/HER2 cluster can be subdivided further on the basis of HER2 amplification.

We wished to determine whether breast cancer cell lines also could be classified on the basis of functional genomics, and if so, how such groups would compare with genomic subclasses. Of the genes with a significant normalized GARP score (P < 0.05) in at least 2 breast cancer cell lines (~3,500 genes), we identified the 10% that varied the most across all of the lines. Subjecting these genes (n = 348) to unsupervised hierarchical clustering using Pearson correlation and complete linkage clustering divided the cell lines into 2 major groups. A heat map representation of the genes (n = 41; Supplementary Table S9) that most significantly discriminated between these 2 major clusters [t test; false discovery rate (FDR) < 0.1, Benjamini-Hochberg correction] is shown in Figure 4A. Notably, the MCF-7 and KPL-1 cell lines, which are derived from the same tumor, clustered most closely in this analysis. Moreover, the 2 functional genomic clusters corresponded almost exactly with the basal and luminal/HER2 subtypes defined by expression profiling of the same cell lines. The single outlier, Sk-Br-3, is classified as luminal/HER2 by microarray analysis but behaved like a basal cell line in our functional genomic screen. This unsupervised clustering approach robustly distinguished basal and luminal/HER2⁺ cell lines even when we used the top 50% most variable genes for analysis (Supplementary Fig. S7). By contrast, unsupervised analysis of our functional genomic data did not segregate the breast cancer cell lines further into, for example, Basal A and Basal B, or HER2 versus luminal.

We also carried out a supervised analysis on the same data set, searching for the genes that best distinguished basal from luminal/HER2 cell lines (Fig. 4B). Remarkably, all 26 of the genes that met our significance criteria (*t* test; FDR < 0.1) were identified by unsupervised clustering (Fig. 4A) as well and included well-known determinants of the luminal/ HER2 subtype, such as *ESR1*, *FOXA1*, *SPDEF*, and *TFAP2C*. Although known drivers of the HER2 subtype (e.g., HER2 and HER3) did not achieve scores significant enough to appear in the unsupervised or supervised clustering panels, they clearly were more essential to the HER2 subtype (Fig. 4C). Finally, we interrogated the TCGA database for the levels of expression of the 41 genes in breast cancer subtypes. Overall, expression did not correlate with functional subtype specificity, except for the few examples depicted in Figure 4D.

Identification of Putative Oncogenic Drivers through Integrative Analyses

Genome-scale functional screens can be coupled with genomic information to identify potential cancer driver genes. To facilitate such analyses, we developed a representation ("query plot") that permits simultaneous visualization of copy number information from publicly available data (see Methods) and GARP scores from functional screens. Each query plot displays amplification data (from tumors and cell lines) for each gene (as a percentage of total tumors and cell lines) as downwardly projecting bars, and the percentage of cell lines in which the same gene scored as essential (P < 0.05; GARP) as upwardly projecting bars (for details, see Methods).

Initial query plots were generated for known oncogenes, such as KRAS, FOXA1, and ERBB2. Notably, compared with the other tumor types, KRAS was particularly essential in pancreatic cancer cell lines (Fig. 5A), consistent with the known role of KRAS mutations in PDAC. Although FOXA1 is amplified in breast, ovarian, and pancreatic cancer cell lines, it was essential only in luminal/HER⁺ breast cancer cells (Fig. 5B). As expected, ERBB2 is clearly the focal point of the ERBB2 amplicon in primary tumors and cell lines and was essential in a subset of breast cancer cell lines (Fig. 5C). The ERBB2 locus also was amplified in a subset of PDAC and HGS-OvCa cell lines and tumors but was essential only in selected PDAC cell lines. Scanning across the ERBB2 amplicon, we see general essential genes (e.g., RPL19, PSMD3), as well as CDK12, recently implicated as a driver in HGS-OvCa (34). Taken together (also see the previous discussion of ovarian cancer-specific essentials), these data suggest that CDK12 may be the key driver within the ERBB2 amplicon in HGS-OvCa.

AC American Association for Cancer Research

FEBRUARY 2012 CANCER DISCOVERY | 179

Marcotte et al.



Figure 4. Subtype classification of breast functional screening results. **A**, unsupervised hierarchical clustering of functional screening data classifies the breast cancer cell lines into the known breast cancer subtypes. Seventeen genes corresponded to the luminal/HER2⁺ subtypes, and 24 were specific to the basal subtype. **B**. supervised clustering of screening results according to known breast cancer subtypes. **C**, normalized zGARP scores for genes displaying differential essentiality in specific breast subtypes. Only MYO3B and ESR1 had signals strong enough to be included in the unsupervised analysis at the FDR threshold used. The Kruskal-Wallis test was used to determine P values comparing the range of scores. **D**, the normalized gene expression level of 4 genes identified as luminal/HER2⁺ specific (*FOXA1, SPDEF*) or basal subtype-specific (*CENPO, NLN*) was obtained from the TCGA data portal. The expression data included 42 basal subtype, 63 HER2⁺ subtype, and 223 luminal subtype tumor samples. The Kruskal-Wallis test was used to determine P values comparing the range of expression values.

To survey the landscape of potential drivers for breast, pancreatic, and ovarian cancer, we examined genes with measurable amplification in multiple cell lines that also scored as essential (P < 0.05; GARP) in multiple cell lines of a given tumor type. Several general essential genes identified by GARP are involved in cell signaling and represent putative therapeutic targets. For example, the gene encoding the receptor tyrosine kinase DDR1 hit in 49 of

72 cell lines (P < 0.05; GARP). To explore further the role of *DDR1*, 3 shRNAs from the TRC library, as well as 3 hairpins from an alternative shRNA library (see Methods), were used for secondary assays in the breast cancer cell lines Cal51, MCF7, Sk-Br-3, BT-20, HCC1954, and HCC38. As positive controls for cell killing, we used 2 shRNAs that were lethal in almost all cell lines, targeting small nuclear ribonucleoprotein D1 polypeptide (*SNRPD1*) and the 26S



RESEARCH ARTICLE



proteasomal subunit non-ATPase 1 (*PSMD1*), respectively (Supplementary Table S6). All 6 *DDR1*-directed hairpins efficiently lowered *DDR1* transcript levels and inhibited proliferation in the 6 breast cancer cell lines (Fig. 6B, C). *DDR1* depletion had similar effects on the 3 pancreatic cancer cell lines tested, KP-4, Panc08.13, and Panc10.05 (Fig. 6D), confirming that it is essential for cancer cell proliferation. Notably, the most efficacious *DDR1* shRNAs were as effective at inhibiting proliferation as the *SNRPD1*and *PSMD1*-positive controls.

We selected for further analysis an additional 5 genes (*SKAP1, PRUNE, EIF3H, EPS8,* and *ITGAV*) that mapped within amplicons in at least 1 of the 3 tumor types (Supplementary Fig. S8) and scored as essential in our screens. SKAP1, which scored in multiple breast (13), PDAC (13), and HGS-OvCa (7) cell lines (total of 33), is an SRC kinase-associated phosphoprotein thought to function only in T cells (49). We confirmed

that SKAP1 is expressed in breast cancer cell lines and found that SKAP1 knockdown significantly reduced proliferation in HCC1954 and MCF7 breast cancer cells (Fig. 7A-C). PRUNE encodes a phosphodiesterase reportedly involved in cell migration and was essential in 25 of our cell lines (PDAC: 14; BrCa: 7; HGS-OvCa: 4) (50). PRUNE knockdown (by either of 2 shRNAs) significantly reduced the proliferation of Sk-Br-3 and MDA-MB-436 cells (Fig. 7D-F). Knockdown of the translation initiation factor EIF3H, which scored as essential in 8 BrCa, 5 PDAC, and 4 HGS-OvCa cell lines, by any of 3 independent shRNAs, significantly reduced cell proliferation in the HGS-OvCa cell lines tested (OVCAR5, OVCAR8, and A2780) (Fig. 7G-I). Furthermore, knocking down EPS8, which encodes an adaptor protein involved in endocytosis and was essential in 28 of our cell lines (PDAC: 12; BrCa: 10; HGS-OvCa: 6), also significantly reduced proliferation in certain PDAC-derived cancer cell lines (Fig. 7J-L). Lastly, ITGAV, which

Marcotte et al.



Figure 6. Confirmation of essential role of *DDR1* in breast and pancreatic cell lines. **A**, as described in Figure 5, but for *DDR1*. **B**, relative number of cells remaining 6 days postinfection with 3 different *DDR1* shRNAs. *SNRPD1* and *PSMD1* shRNA were used as positive controls, because these were the 2 most potent shRNAs in pooled screening. Control shRNA is the average of a GFP, a LacZ, and a Luciferase shRNA. *DDR1* shRNA viability/ cell number is depicted relative to control. Experiments were performed in triplicate. *P* value was derived using a one-tailed Student *t* test. (shDDR1-83: TRCN0000121083; shDDR1-86: TRCN0000121086; shDDR1-63: TRCN0000121163). Small graph inset, percentage transcript remaining determined by qPCR relative to GFP shRNA control. **C**, relative number of cells remaining 9 days postinfection with 3 different pGIPZ *DDR1* shRNAs. Negative control was a nonsilencing shRNA. Small graph inset, percentage transcript remaining determined by qPCR relative to GFP shRNA. Small graph inset, percentage transcript remaining determined by qPCR relative to GFP shRNA. Small graph inset, percentage transcript remaining determined by qPCR relative to GFP shRNA. Small graph inset, percentage transcript remaining determined by qPCR relative to GFP shRNA. Small graph inset, percentage transcript remaining determined by qPCR relative to GFP shRNA. Small graph inset, percentage transcript remaining determined by qPCR relative to GFP shRNA.

encodes integrin α V, a component of the vitronectin receptor, was essential in 21 of our cell lines (PDAC: 8; BrCa 7; HGS-OvCa: 6). Accordingly, knockdown of *ITGAV* significantly inhibited the proliferation of Panc10.05, SU86.86, and Panc03.27 cells (Fig. 7M–O).

Finally, we tested whether the effects of knocking down a gene identified in our screens could be rescued by re-expressing an shRNA-resistant form of that gene. We developed a competition assay using PL45 cells expressing (GFP⁺) or not expressing (GFP⁻) mouse *Itgav* (Fig. 7P), then monitored the effect on proliferation of an shRNA targeting the 3' untranslated region of human

ITGAV. Indeed, cells expressing *Itgav* were resistant to the effects of the human shRNA, confirming that the inhibitory effect of *ITGAV* knockout on proliferation was specific (Fig. 7Q, R).

DISCUSSION

Functional genomic approaches can complement genomic analyses, providing a more comprehensive view of cancer cell biology and suggesting new therapeutic strategies. Earlier work using large-scale pooled shRNA screens to investigate gene essentiality in cancer cell lines surveyed multiple histotypes

RESEARCH ARTICLE

but examined relatively few examples of each type (18) or sought synthetic lethal relationships with the same oncogene expressed in different types of cancer (15). Our study represents an extensive functional genetic survey of 3 major cancers (BrCa, PDAC, and HGS-OvCa), establishes a new metric for scoring shRNA dropout screens, uncovers complexity in the relationship between genomic and functional genomic classification schemes, and reveals unexpected gene dependencies and new potential therapeutic targets in these malignancies.

Previous studies quantified hairpin dropout at a single (usually fairly long) time point. Dynamic measurements provide additional power for tracking shRNA abundance in a given population of cells, and we found that this helps to group shRNAs into functional classes that reflect the intrinsic properties of their corresponding target genes (Fig. 1C). For example, shRNAs that drop out early are more likely to target housekeeping genes. We developed a new scoring approach (shARP) that captures these dynamic properties, as well as a gene-level metric, GARP. More important, if one considers "housekeeping" or "highly conserved" gene subsets (which are largely nonoverlapping; see Fig. 2C) as reasonable surrogates for essentiality, GARP outperforms previous scoring metrics in its ability to "call" general essential genes (Fig. 2A, B).

Analysis of our 72 screens allowed us to define 3 types of essential genes: (i) general essentials, (ii) tissue-specific essentials, and (iii) subtype-specific essentials. General essentials are, as expected, enriched for highly conserved, housekeeping functions, such as those associated with transcription, translation, splicing, and the proteasome (Fig. 3A, Supplementary Table S6). Moreover, the general essentials identified in our screen showed considerable (although not complete) overlap with those defined in an earlier study (18) (Supplementary Fig. S4). Such similarity in experimental results across a large number of experiments in different institutions argues for the robustness of dropout screening methodology and strongly suggests that the genes identified as essential in all of these screens are, in fact, generally required for cancer cell proliferation/survival. Several possible explanations likely contribute to the lack of complete overlap in general essentials defined in previous work and our study. For example, we observed, retrospectively, that calculating fold-changes in shRNA bar code abundance for different endpoints within the same screen yields rank-ordered gene lists that do not overlap perfectly. Our screening approach, which employs multiple time points, helps to buffer some of the noise introduced by endpoint assays, obviates the need to terminate a screen at a precise number of generations, and permits comparison of shRNA dropout profiles across multiple cell lines. Thus, we believe that dynamic shRNA profiles provide a more robust metric than fold-change measurements for ranking shRNA dropouts and, hence, essential genes.

Although we are confident that genes that score as general essentials in our screen (and particularly those that are also classified as general essentials by the other scoring metrics) are, in fact, required for cancer cell proliferation, inherent limitations of shRNA screening methodology make it likely that we are underestimating the actual number of general essential genes. Indeed, we have noted that some genes with GARP P > 0.05 are, in fact, required for proliferation in secondary assays. Because biologically significant results can be obtained for hairpins that lie outside the statistically significant range, future users of our resource should not *a priori* exclude from further investigation genes with GARP

scores lying outside a stringent statistical cutoff. Furthermore, it should be noted that the shRNA detection strategy that we used in our screens precludes efficient identification of genes whose depletion enhances cell proliferation: in order to detect dropouts, hybridizations are carried out with a standard amount of excess probe, which limits the detection of "enhancing" shRNAs.

Interestingly, some genes that qualified as general essentials in our screen do not serve obvious housekeeping roles. Perhaps the most intriguing is DDR1, which encodes a receptor tyrosine kinase (RTK) that binds to collagens (51). Secondary screens using shRNAs from 2 independent libraries confirmed that DDR1 is required in the 6 breast and 4 pancreatic cancer lines tested. In addition, several previous studies reported increased DDR1 expression in various human tumors, including breast, ovarian, lung, esophageal, and pediatric brain cancers (reviewed in ref. 52), and DDR1 knockdown suppresses tumorigenicity in HCT116 cells (53). Furthermore, the DDR1 locus is amplified in a significant number of breast, ovarian, and pancreatic tumors and cell lines (Fig. 6A). Notably, although they are smaller than control, Ddr1 knockout mice are viable (51), arguing for a selective requirement in malignant cells. These data suggest that DDR1 might be an attractive target for antineoplastic drug development, although it will be important to confirm its requirement for tumor maintenance, not just for cancer cell proliferation in tissue culture. Recently, DDR2, which shares substantial sequence similarity with DDR1 and also encodes a collagen-responsive RTK, was found to be mutated and required in about 5% of squamous cell carcinomas of the lung (54). We found that DDR2 also was essential in 36 of our cancer cell lines. Taken together, these results raise the possibility that both DDRs play a broader role in oncogenesis than previously appreciated.

Supervised analysis of our screening results identified tissuespecific genes, which are potentially responsible for enrichment in different biological processes and may reveal emergent properties of cancer cells from tissues of different origin. However, these tissue-specific genes did not drive (unsupervised) clustering of the lines into histotype-specific groups. Notably, one major cluster was enriched in breast cancer cell lines, including 13 of 14 lines of luminal/HER2 subtype, and this grouping was driven mainly by breast cancer-specific genes and luminal/ HER2-specific genes. Most HGS-OvCa and PDAC cell lines segregated into the same group, even though breast and ovarian cancer share some oncogenic drivers (BRCA1/2, HER2, HER3) and, unlike PDAC, usually lack KRAS mutations. Moreover, while nearly all of the PDAC lines tested have a KRAS mutation, these lines did not cosegregate by unsupervised clustering analysis. These data indicate that additional genetic events in PDAC modulate the gene essentiality landscape imposed by KRAS mutations. Furthermore, they suggest that, in contrast to previous reports (15, 55-57), it might be difficult to identify a universal set of "KRAS" synthetic lethal genes; we were not able to uncover a set of essential genes specific to mutant KRAS. Rather, there might be context-dependent KRAS synthetic lethality imposed by the cell of origin of the malignancy and/or its other underlying genetic abnormalities.

Because the breast cancer cell lines we studied have been analyzed by expression profiling (and copy number variation analysis), we could explore the relationship between genomics and functional genomics in these lines. Remarkably, applying an unsupervised clustering algorithm to these screening data resulted in clustering of the breast cancer cell lines into functional subsets

American Association for Cancer Research

FEBRUARY 2012 CANCER DISCOVERY | 183





that are essentially identical to the classic luminal/HER2 and basal subgroups (58, 59). Reassuringly, several of the genes responsible for the clustering of the luminal/HER2 cell lines are well-known drivers of these subtypes (e.g., ESR1, FOXA1, SPDEF). The genes whose knockdown preferentially impaired basal breast cancer cell proliferation are less well studied in this subtype or in breast cancer in general. However, they include genes encoding proteins involved in DNA repair (POLE) and with antiapoptotic function (XIAP). Our functional genomic classification does not separate the luminal/HER2 or basal subgroups further (i.e., into HER2-specific, luminal-specific, or Basal A- or Basal B-specific) by unsupervised clustering. For the luminal/HER2 breast cancer lines, this is not surprising, because HER2 and luminal subgroups also are not distinguishable by expression profiling (59). Basal A and Basal B cell lines, however, can be discerned by transcriptional differences (59). Failure to separate these subgroups by functional

genomic clustering could reflect insufficient numbers of cell lines in our screen or biological nuances that are not completely captured at the mRNA expression level. Notably, genes specific for each subclass (i.e., luminal, HER2, Basal A, Basal B) can be identified by supervised methods (Fig. 4C); such genes, if validated, could yield new insights into subgroup-specific biological differences and suggest subgroup-specific therapeutic targets.

Recently, PDAC and HGS-OvCa were classified into subgroups based on expression differences (34, 60). However, unsupervised clustering did not reveal subgroup-specific essential gene maps for our pancreatic or ovarian cancer cell lines. Conceivably, the number of cell lines that we screened did not provide sufficient predictive power. Alternatively, our cell lines might not adequately represent the range of transcriptional subclasses seen in tumors. It also is possible that the transcriptional subclasses themselves are not predictive





Figure 7 continued. L. Western blot of EPS8 knockdown in pancreatic cell lines. **M**, **N**, same as (A, B) for *ITGAV* in pancreatic cell lines. **O**, Western blot showing effects of *ITGAV* knockdown in pancreatic cell lines. **P**, experimental scheme for *ITGAV* rescue experiment. **Q** flow cytometry results for *ITGAV* rescue experiment. **R**, quantification of results in (Q). *P<0.05; **P<0.01; ***P<0.001.

of gene sets that are essential for viability. For example, our PDAC collection contains cell lines that conform to both the classic and quasi-mesenchymal pancreatic subtypes (60), yet these lines did not segregate from each other by unsupervised functional genomic clustering. Additional genomic data and further analyses are needed to explore the relationship between functional genetic screening data and genomic data to identify better prognostic and therapeutic factors for pancreatic and ovarian cancers.

Finally, by combining copy number information with the results of our dropout screens, we identified—and confirmed—several unexpected vulnerabilities in breast, pancreatic, and ovarian cancer cell lines. For example, *SKAP1* encodes an adaptor that is generally thought to function only in T cells, where it reportedly modulates T-cell antigen receptor–induced activation of the Ras-ERK-AP1 pathway (49, 61), as well as integrin clustering and adhesion (62). Recently, *SKAP1* was identified in genomewide association studies as a susceptibility locus for ovarian (63)

and prostate cancer (64). Although this could indicate a role for SKAP1 in immune surveillance, our data suggest that these alleles might have cell-autonomous effects on tumorigenesis. We identified PRUNE as a gene that was preferentially essential in breast and pancreatic cancer cell lines (P < 0.028) and confirmed this finding in several breast cell lines (Fig. 7). PRUNE encodes a phosphodiesterase belonging to the DHH superfamily, which reportedly binds NM23-H1 to promote metastasis (65). PRUNE also binds to glycogen synthase kinase and reportedly regulates cell migration by modulating focal adhesions (50). Moreover, amplification and overexpression of PRUNE reportedly correlates with advanced breast carcinomas (66, 67). EIF3H encodes a translation initiation factor and is amplified and overexpressed in a variety of cancer types including colon (68), NSCLC (69), prostate (70), breast (71), and HCC (72). Notably, several other members of the eIF-3 translation initiation complex scored as essential in our screens: EIF3A (67 lines), EIF3B (70 lines), EIF3C (66 lines), EIF3D (65 lines), EIF3G (58 lines), and EIF3I (57 lines).

EPS8 is an epidermal growth factor receptor (EGFR) substrate that mediates RAC1 activation and trafficking of EGFR in a RAB5-dependent manner (73). EPS8 also has been implicated in cell migration and ovarian cancer metastasis (74), and *EPS8* overexpression has been observed in PDAC (75) and oral squamous cell carcinoma (76). Finally, *ITGAV* encodes integrin α chain V, a component of the vitronectin receptor, which has been associated with multiple different cancers (77).

Taken together, our results have interesting implications for the systems biology of cancer. Our finding that functional genomic and genomic classification schemes yield only partially overlapping results implies that functional genomic studies reveal nuances in cancer cell biology that have not been captured by existing genomic analyses. By (re)-analyzing genomic data in cancer cells grouped by similar gene essentiality profiles, it is likely that new drivers and synthetic lethal relationships will emerge. Future exploitation of our functional genomic resource will require validation of the multiple types of essential genes that we have identified, along with integrative analysis of functional genomic and detailed genomic information from these cell lines.

METHODS

Cell Lines

A full description of each cell line used, where the cell lines were obtained, and the method and date of cell line authentication are detailed in Supplementary Table S2.

shRNA Dropout Screens

Each cell line was grown to a population size of $\sim 2 \times 10^8$ cells in the requisite medium (see Supplementary Table S2). Cells were washed with warm PBS, trypsinized, resuspended in warm medium, and counted. An aliquot of the 80k human shRNA lentivirus pool (22) and either polybrene (4-8 g/mL) or protamine sulfate (5 g/mL) were added such that an MOI of 0.3 to 0.4 would be achieved (determined by prior testing of each cell line). Cell-lentivirus mixtures were plated into 15-cm-diameter culture dishes and incubated at 37°C with 5% CO₂. Twenty-four hours postinfection, the medium was replaced with fresh medium containing puromycin (puromycin concentration determined by prior tests on each cell line), and cells were incubated for an additional 48 hours. Culture dishes were washed with warm PBS to remove dead cells, and surviving cells were collected by trypsinization and resuspended in warm medium. Cell populations were quantified, aliquots of 2×10^7 cells were removed and pelleted by centrifugation, and 3 replicate populations of 2 \times 10⁷ cells were plated at appropriate density into 15-cm-diameter culture dishes. When replicate populations reached 80% to 90% confluency, cells from each replicate were collected by trypsinization and mixed (cells from different replicates were not comingled). From these mixtures, 2 aliquots of 2×10^7 cells were removed, pelleted, and frozen down, while one aliquot of 2×10^7 cells was replated for further growth. This step was repeated until a minimum of 6 to 8 doublings for each replicate cell population was obtained. Genomic DNA was prepared from cell pellets using the QIAmp Blood Maxi kit (Qiagen #51194). Genomic DNA was precipitated using ethanol and sodium chloride and resuspended at 400 ng/L in 10 mmol/L Tris-HCl, pH 7.5. shRNA populations from cell lines were amplified via PCR and prepared and applied to GMAP arrays as described previously (22).

Identification of Hairpin Classes

Hairpins were segregated into classes using rules based on a Boolean combination of features that describe the hairpin behavior through the time course. The rules involved features describing the rate of dropout over the first and second time intervals (i.e., the slope of the expression intensity change between time points), the ratio of the dropout rates, the fold-change at the end point relative to T_0 (FC₂), and the initial expression intensity at T_0 . The classes were as follows: K–fast dropouts, which lead to rapid hairpin depletion; S–slow dropouts, displaying a lag prior to hairpin depletion; E–enhancers, or hairpins that show an increase in abundance over time; and C–continuous dropouts, or hairpins that show a constant rate of depletion over time. These rules are summarized in Supplementary Table S3.

Scoring of shRNA Screens

To incorporate measurements from multiple time points in an shRNA screen, we developed the shARP score, as shown in the following equation:

$$shARP = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(\frac{\Delta y_i}{\Delta x_i} \right)$$

where *n* is the number of time points, Δy is the change in expression intensity at t_i relative to t_0 , and Δx is the number of doublings for the cell line at t_i relative to t_0 . The shARP scores are determined for each of the 78,432 hairpins in the library. They are then used to calculate the GARP score by averaging the 2 lowest shARP scores. A significance value is assigned to each GARP score through bootstrapping, in which the shARP scores are randomly permuted 1,000 times, GARP scores are recomputed, and a *P* value is determined by the frequency with which the actual GARP score is lower than the permuted GARP scores. To facilitate comparisons between screens, GARP scores were *Z*-score normalized.

In addition to the GARP score, we applied 2 previously published shRNA scoring methods to rank the normalized shARP scores from our screens. First, we applied RSA (23) using the R package provided by the original authors. Briefly, all shARP scores were normalized using a robust Z-score before applying RSA iteratively to each screen, setting the UB parameter to 0, the LB parameter to -3, and using the Entrez Gene ID as the unique identifier for each gene. Second, we ranked the normalized shARP scores by RIGER (12), as implemented in the GENE-E software package. Genes were ranked by each of the 3 RIGER methods to collapse hairpins to genes: RIGER_WS, RIGER_SB, and RIGER_KS. To avoid RIGER scoring dropouts and enhancers separately, the shARP score distribution was shifted below zero prior to applying Kolmogorov-Smirnov scoring by subtracting the maximum value from each distribution.

Benchmarking Scoring Methods

To benchmark our scoring method against existing approaches such as RIGER and RSA, the top-ranked genes from each screen were overlapped against 2 data sets likely enriched in essential genes. First, housekeeping genes are genes universally expressed to maintain cellular function: the more tissues in which a gene is expressed, the higher the likelihood that it will be essential (78). Second, highly conserved orthologs are genes that are shared among species, which have a higher likelihood of being essential (79, 80). Housekeeping genes (n = 1,722) were identified as genes expressed in at least 73 of 79 tissues in a human expression compendium (25). Highly conserved orthologs (n = 1,617) were identified as human genes with orthologs in 8 different species (A. thaliana, B. taurus, C. elegans, C. familiaris, M. mulatta, M. musculus, R. norvegicus, and S. cerevisiae), as determined by InParanoid (24). There are 315 genes in common between the housekeeping genes and conserved orthologs.

The top 500 genes ranked by each scoring method in each screen were selected and overlapped with the reference sets,

RESEARCH ARTICLE

starting with the top 5 genes up to the top 500 in 5 gene steps. The overlap was plotted, and the AUC was calculated for each overlapping gene set.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

R. Marcotte, K.R. Brown, T. Ketela, R. Rottapel, B.G. Neel, and J. Moffat designed the research. R. Marcotte, M. Haider, Y. Fedyshyn, M. Luhova, B. Fedyshyn, P. Mero, M. Medrano, K. Karamboulas, D. van Dyke, F.J. Vizeacoumar, F.S. Vizeacoumar, A. Datti, J. l. Wrana, T. Ketela, C. Misquitta, D. Kasimer, and J. Normand contributed to and/or performed experiments. R. Marcotte, K.R. Brown, F. Sircoulomb, A. Sayad, J.L.Y. Koh, P.M. Krzyzanowski, F. Suarez, G.C. Brito, T. Ketela, R. Rottapel, B.G. Neel, and J. Moffat analyzed the data. R. Rottapel, B.G. Neel, and J. Moffat supervised the research. R. Marcotte, K.R. Brown, B.G. Neel, and J. Moffat wrote the manuscript.

Acknowledgments

The authors thank Corey Nislow and Guri Giaever for use of equipment; Francisco Real, Ann Marie Mes-Masson, Patricia Tonin, Graham Fletcher, Gordon Mills, and Robert Bast for cell lines; and Luigi Naldini for plasmids.

Grant Support

This work was supported by the Ontario Institute for Cancer Research and Terry Fox Research Institute through the Selective Therapies Program (R. Rottapel, B.G. Neel, J.L. Wrana, and J. Moffat), the Canadian Institutes for Health Research (J.L. Wrana; grant number PRG-82679, B.G. Neel, and J. Moffat), the Canadian Foundation for Innovation (J. Moffat), and the Ontario Research Fund (B.G. Neel, J. Moffat, and R. Rottapel). Work in the Neel and Rottapel laboratories is supported in part by the Ontario Ministry of Health and Long Term Care and the Princess Margaret Hospital Foundation. J. Moffat holds a CIHR New Investigator Award. B.G. Neel is a Canada Research Chair, Tier 1 and the recipient of the Premier of Ontario's Summit Award. R. Marcotte is the recipient of a postdoctoral fellowship from the Canadian Breast Cancer Foundation.

Received September 6, 2011; revised December 15, 2011; accepted December 16, 2011; published OnlineFirst December 29, 2011.

REFERENCES

- 1. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529–41.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98:10869–74.
- Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature 2010;467:1109–13.
- Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature 2009;462:1005–10.
- Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 2010;463:184–90.
- Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010;463:191–6.

- 7. Ashworth A. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. J Clin Oncol 2008;26:3785–90.
- Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 2005;434:917–21.
- Ashworth A, Lord CJ, Reis-Filho JS. Genetic interactions in cancer progression and treatment. Cell 2011;145:30–8.
- 10. Moffat J, Sabatini DM. Building mammalian signalling pathways with RNAi screens. Nat Rev Mol Cell Biol 2006;7:177–87.
- 11. Sigoillot FD, King RW. Vigilance and validation: keys to success in RNAi screening. ACS Chem Biol 2011;6:47-60.
- Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, et al. Highly parallel identification of essential genes in cancer cells. Proc Natl Acad Sci U S A. 2008;105:20380–5.
- Schlabach MR, Luo J, Solimini NL, Hu G, Xu Q, Li MZ, et al. Cancer proliferation gene discovery through functional genomics. Science 2008;319:620–4.
- Silva JM, Marran K, Parker JS, Silva J, Golding M, Schlabach MR, et al. Profiling essential genes in human mammary cells by multiplex RNAi screening. Science 2008;319:617–20.
- Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, Westbrook TF, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell 2009;137:835–48.
- Bivona TG, Hieronymus H, Parker J, Chang K, Taron M, Rosell R, et al. FAS and NF-kappaB signalling modulate dependence of lung cancers on mutant EGFR. Nature 2011;471:523–6.
- Mendes-Pereira AM, Sims D, Dexter T, Fenwick K, Assiotis I, Kozarewa I, et al. Breast Cancer Special Feature: genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. Proc Natl Acad Sci U S A 2011 Apr 11. [Epub ahead of print].
- Cheung HW, Cowley GS, Weir BA, Boehm JS, Rusin S, Scott JA, et al. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc Natl Acad Sci U S A 2011;108:12372–7.
- Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, et al. International network of cancer genome projects. Nature 2010;464:993-8.
- Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, et al. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell 2006;124:1283–98.
- 21. Root DD, Wang K. High-affinity actin-binding nebulin fragments influence the actoS1 complex. Biochemistry 2001;40:1171–86.
- 22. Ketela T, Heisler LE, Brown KR, Ammar R, Kasimer D, Surendra A, et al. A comprehensive platform for highly multiplexed mammalian functional genetic screens. BMC Genomics 2011;12:213.
- Konig R, Chiang CY, Tu BP, Yan SF, DeJesus PD, Romero A, et al. A probability-based approach for the analysis of large-scale RNAi screens. Nature Methods 2007;4:847–9.
- O'Brien KP, Remm M, Sonnhammer EL. Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res 2005;33:D476-80.
- 25. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 2004;101:6062–7.
- Moffat J, Reiling JH, Sabatini DM. Off-target effects associated with long dsRNAs in Drosophila RNAi screens. Trends Pharmacol Sci 2007;28:149–51.
- 27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
- Zhang JS, Koenig A, Harrison A, Ugolkov AV, Fernandez-Zapico ME, Couch FJ, et al. Mutant K-Ras increases GSK-3β gene expression via an ETS-p300 transcriptional complex in pancreatic cancer. Oncogene 2011;30:3705–15.
- Fuchs F, Pau G, Kranz D, Sklyar O, Budjan C, Steinbrink S, et al. Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. Mol Syst Biol 2010;6:370.

AC American Association for Cancer Research

FEBRUARY 2012 CANCER DISCOVERY | 187

- Katsumoto T, Yoshida N, Kitabayashi I. Roles of the histone acetyltransferase monocytic leukemia zinc finger protein in normal and malignant hematopoiesis. Cancer Sci 2008;99:1523–7.
- Hasegawa K, Koizumi F, Noguchi Y, Hongo A, Mizutani Y, Kodama J, et al. SSX expression in gynecological cancers and antibody response in patients. Cancer Immun 2004;4:16.
- 32. Gure AO, Chua R, Williamson B, Gonen M, Ferrera CA, Gnjatic S, et al. Cancer-testis genes are coordinately expressed and are markers of poor outcome in non-small cell lung cancer. Clin Cancer Res 2005;11:8055–62.
- Shi CS, Kehrl JH. Traf6 and A20 differentially regulate TLR4-induced autophagy by affecting the ubiquitination of Beclin 1. Autophagy 2010;6:986–7.
- Bell D, Berchuck A, Birrer M, Chien J, Cramer DW, Dao F, et al. Integrated genomic analyses of ovarian carcinoma. Nature 2011;474:609–15.
- Bachmann M, Moroy T. The serine/threonine kinase Pim-1. Int J Biochem Cell Biol 2005;37:726–30.
- Lamason RL, McCully RR, Lew SM, Pomerantz JL. Oncogenic CARD11 mutations induce hyperactive signaling by disrupting autoinhibition by the PKC-responsive inhibitory domain. Biochemistry 2010;49:8240–50.
- Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science 2007;318:1108–13.
- Ormandy CJ, Musgrove EA, Hui R, Daly RJ, Sutherland RL. Cyclin D1, EMS1 and 11q13 amplification in breast cancer. Breast Cancer Res Treat 2003;78:323–35.
- Zafrakas M, Losen I, Knuchel R, Dahl E. Enhancer of the rudimentary gene homologue (ERH) expression pattern in sporadic human breast cancer and normal breast tissue. BMC Cancer 2008;8:145.
- Serrero G. Autocrine growth factor revisited: PC-cell-derived growth factor (progranulin), a critical player in breast cancer tumorigenesis. Biochem Biophys Res Commun 2003;308:409–13.
- 41. Lim S, Janzer A, Becker A, Zimmer A, Schule R, Buettner R, et al. Lysine-specific demethylase 1 (LSD1) is highly expressed in ERnegative breast cancers and a biomarker predicting aggressive biology. Carcinogenesis 2010;31:512–20.
- 42. Lu S, Becker KA, Hagen MJ, Yan H, Roberts AL, Mathews LA, et al. Transcriptional responses to estrogen and progesterone in mammary gland identify networks regulating p53 activity. Endocrinology 2008;149:4809–20.
- 43. Kiyotani K, Mushiroda T, Imamura CK, Hosono N, Tsunoda T, Kubo M, et al. Significant effect of polymorphisms in CYP2D6 and ABCC2 on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. J Clin Oncol 2010;28:1287–93.
- Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat Genet 2011;43:27–33.
- Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineagespecific transcription. Cell 2008;132:958–70.
- Badve S, Turbin D, Thorat MA, Morimiya A, Nielsen TO, Perou CM, et al. FOXA1 expression in breast cancer–correlation with luminal subtype A and survival. Clin Cancer Res 2007;13:4415–21.
- 47. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
- Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 2010;12:R68.
- Kosco KA, Cerignoli F, Williams S, Abraham RT, Mustelin T. SKAP55 modulates T cell antigen receptor-induced activation of the Ras-Erk-AP1 pathway by binding RasGRP1. Mol Immunol 2008;45:510–22.
- Kobayashi T, Hino S, Oue N, Asahara T, Zollo M, Yasui W, et al. Glycogen synthase kinase 3 and h-prune regulate cell migration by modulating focal adhesions. Mol Cell Biol 2006;26:898–911.

- Vogel WF, Aszodi A, Alves F, Pawson T. Discoidin domain receptor 1 tyrosine kinase has an essential role in mammary gland development. Mol Cell Biol 2001;21:2906–17.
- Vogel WF, Abdulhussein R, Ford CE. Sensing extracellular matrix: an update on discoidin domain receptor function. Cell Signal 2006;18:1108–16.
- 53. Kim HG, Hwang SY, Aaronson SA, Mandinova A, Lee SW. DDR1 receptor tyrosine kinase promotes prosurvival pathway through Notch1 activation. J Biol Chem 2011;286:17672–81.
- Hammerman PS, Sos ML, Ramos AH, Xu C, Dutt A, Zhou W, et al. Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. Cancer Discovery 2011;1:78–89.
- 55. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462:108–12.
- Scholl C, Frohling S, Dunn IF, Schinzel AC, Barbie DA, Kim SY, et al. Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. Cell 2009;137:821–34.
- 57. Singh A, Settleman J. Oncogenic K-ras "addiction" and synthetic lethality. Cell Cycle 2009;8:2676-7.
- Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, et al. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. PLoS One 2009;4:e6146.
- Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 2006;10:515–27.
- Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011;17:500–3.
- Lehmann R, Meyer J, Schuemann M, Krause E, Freund C. A novel S3S-TAP-tag for the isolation of T-cell interaction partners of adhesion and degranulation promoting adaptor protein. Proteomics 2009;9:5288–95.
- 62. Wang H, Liu H, Lu Y, Lovatt M, Wei B, Rudd CE. Functional defects of SKAP-55-deficient T cells identify a regulatory role for the adaptor in LFA-1 adhesion. Mol Cell Biol 2007;27:6863–75.
- Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat Genet 2010;42:874–9.
- 64. Huang CN, Huang SP, Pao JB, Chang TY, Lan YH, Lu TL, et al. Genetic polymorphisms in androgen receptor-binding sites predict survival in prostate cancer patients receiving androgen-deprivation therapy. Ann Oncol 2011 Jun 6. [Epub ahead of print].
- D'Angelo A, Garzia L, Andre A, Carotenuto P, Aglio V, Guardiola O, et al. Prune cAMP phosphodiesterase binds nm23-H1 and promotes cancer metastasis. Cancer Cell 2004;5:137–49.
- 66. Zollo M, Andre A, Cossu A, Sini MC, D'Angelo A, Marino N, et al. Overexpression of h-prune in breast cancer is correlated with advanced disease status. Clin Cancer Res 2005;11:199–205.
- 67. Forus A, D'Angelo A, Henriksen J, Merla G, Maelandsmo GM, Florenes VA, et al. Amplification and overexpression of PRUNE in human sarcomas and breast carcinomas—a possible mechanism for altering the nm23-H1 activity. Oncogene 2001;20:6881–90.
- Pittman AM, Naranjo S, Jalava SE, Twiss P, Ma Y, Olver B, et al. Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. PLoS Genet 2010;6.
- 69. Cappuzzo F, Varella-Garcia M, Rossi E, Gajapathy S, Valente M, Drabkin H, et al. MYC and EIF3H coamplification significantly improve response and survival of non-small cell lung cancer patients (NSCLC) treated with gefitinib. J Thorac Oncol 2009;4:472–8.
- Saramaki O, Willi N, Bratt O, Gasser TC, Koivisto P, Nupponen NN, et al. Amplification of EIF3S3 gene is associated with advanced stage in prostate cancer. Am J Pathol 2001;159:2089–94.
- 71. Savinainen KJ, Linja MJ, Saramaki OR, Tammela TL, Chang GT, Brinkmann AO, et al. Expression and copy number analysis of

RESEARCH ARTICLE

TRPS1, EIF3S3 and MYC genes in breast and prostate cancer. Br J Cancer 2004;90:1041-6.

- Okamoto H, Yasui K, Zhao C, Arii S, Inazawa J. PTK2 and EIF3S3 genes may be amplification targets at 8q23-q24 and are associated with large hepatocellular carcinomas. Hepatology 2003;38:1242–9.
- 73. Lanzetti L, Rybin V, Malabarba MG, Christoforidis S, Scita G, Zerial M, et al. The Eps8 protein coordinates EGF receptor signalling through Rac and trafficking through Rab5. Nature 2000;408:374–7.
- 74. Chen H, Wu X, Pan ZK, Huang S. Integrity of SOS1/EPS8/ABI1 tri-complex determines ovarian cancer metastasis. Cancer Res 2010;70:9979-90.
- 75. Welsch T, Endlich K, Giese T, Buchler MW, Schmidt J. Eps8 is increased in pancreatic cancer and required for dynamic actin-based cell protrusions and intercellular cytoskeletal organization. Cancer Lett 2007;255:205–18.
- 76. Yap LF, Jenei V, Robinson CM, Moutasim K, Benn TM, Threadgold SP, et al. Upregulation of Eps8 in oral squamous cell carcinoma promotes cell migration and invasion through integrin-dependent Rac1 activation. Oncogene 2009;28:2524–34.
- 77. Cooper CR, Chay CH, Pienta KJ. The role of alpha(v)beta(3) in prostate cancer progression. Neoplasia 2002;4:191-4.
- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci U S A 2007; 104:8685-90.
- Amsterdam A, Nissen RM, Sun Z, Swindell EC, Farrington S, Hopkins N. Identification of 315 genes essential for early zebrafish development. Proc Natl Acad Sci U S A 2004;101:12792–7.
- Doyle MA, Gasser RB, Woodcroft BJ, Hall RS, Ralph SA. Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. BMC Genomics 2010;11:222.

AC American Association for Cancer Research