Large-Scale Profiling of Kinase Dependencies in Cancer Cell Lines

Highlights

- Kinome-wide (714 gene) siRNA screens in 117 cell lines from ten cancer histotypes
- Integrating genotype data reveals cancer driver gene dependencies
- Integrating protein interaction data aids the interpretation of genetic dependencies
- Identified dependencies enable prediction of mutant cell line responses to drugs
Large-Scale Profiling of Kinase Dependencies in Cancer Cell Lines


The Breast Cancer Now Research Centre and CRUK Gene Function Laboratory, The Institute of Cancer Research, London SW3 6JB, UK

Systems Biology Ireland, University College Dublin, Dublin 4, Ireland

Royal Marsden Hospital, London SW3 6JJ, UK

UCL Cancer Institute, University College London, London WC1E 6DD, UK

Gustave Roussy Cancer Campus, 94805 Villejuif, France

UCSF Helen Diller Family Comprehensive Cancer Centre, San Francisco, CA 94158, USA

Cancer Research UK Cancer Therapeutics Unit, The Institute of Cancer Research, Sutton SM2 5NG, UK


Co-first author

Present address: UCSF Helen Diller Family Comprehensive Cancer Centre, San Francisco, CA 94158, USA

*Correspondence: alan.ashworth@ucsf.edu (A.A.), chris.lord@icr.ac.uk (C.J.L.)

http://dx.doi.org/10.1016/j.celrep.2016.02.023

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

SUMMARY

One approach to identifying cancer-specific vulnerabilities and therapeutic targets is to profile genetic dependencies in cancer cell lines. Here, we describe data from a series of siRNA screens that identify the kinase genetic dependencies in 117 cancer cell lines from ten cancer types. By integrating the siRNA screen data with molecular profiling data, including exome sequencing data, we show how vulnerabilities/genetic dependencies that are associated with mutations in specific cancer driver genes can be identified. By integrating additional data sets into this analysis, including protein–protein interaction data, we also demonstrate that the genetic dependencies associated with many cancer driver genes form dense connections on functional interaction networks. We demonstrate the utility of this resource by using it to predict the drug sensitivity of genetically or histologically defined subsets of tumor cell lines, including an increased sensitivity of osteosarcoma cell lines to FGFR inhibitors and SMAD4 mutant tumor cells to mitotic inhibitors.

INTRODUCTION

The phenotypic and genetic changes that occur during tumorigenesis alter the set of genes upon which cells are dependent. The best known example of this phenomenon of “genetic dependency” is oncogene addiction where tumor cells become dependent upon the activity of a single oncogene, which when inhibited leads to cancer cell death. Alternatively, tumor cells can become addicted to the activity of genes other than oncogenes, effects known as non-oncogene addictions (Luo et al., 2009), induced essential effects (Tischler et al., 2008), or synthetic lethal interactions (Kaelin, 2005). From a clinical perspective, identifying genetic dependencies in tumor cells could illuminate vulnerabilities that might be translated into therapeutic approaches to treat the disease. Examples of this approach include the development of drugs that target oncogene addiction effects, such as imatinib in the case of ABL addiction, and therapeutic approaches that exploit synthetic lethal effects, such as PARP inhibitors for BRCA-deficient cancers (Lord et al., 2015).

A number of groups have used high-throughput screening approaches such as RNAi or small molecule sensitivity screens to systematically identify genetic dependencies in tumor cell lines (Barrettina et al., 2012; Brough et al., 2011; Cowley et al., 2014; Garnett et al., 2012; Koh et al., 2012). A particular focus has been in dissecting genetic dependencies that involve kinases (Brough et al., 2011; Grueneberg et al., 2008), as these enzymes play key roles in a number of oncogenic processes (Greenman et al., 2007) and are pharmacologically tractable (Sakharkar and Sakharkar, 2007; Workman and Al-Lazikani, 2013; Zhang et al., 2009). Previously, we used high-throughput short interfering (si)RNA screening to identify the kinase dependencies in a panel of 20 breast cancer derived cell lines (Brough et al., 2011). Here, we describe as a resource an expansion of this approach, namely parallel siRNA screens targeting 714 genes in 117 genetically and histologically diverse tumor cell lines. Building on our previous work (Brough et al., 2011), we extend our analytical approach to describe how this data set may be used as a hypothesis-generating tool for identifying candidate therapeutic targets associated with specific tumor histotypes or mutations in cancer driver genes. We also illustrate how, by integrating this functional data with orthogonal data sources such as protein–protein interaction data sets, these genetic dependencies might be dissected mechanistically.
RESULTS

Kinase Genetic Dependencies Identified by Parallel siRNA Screens

We screened a panel of 136 tumor cell lines in triplicate in a plate-arrayed format using an siRNA library designed to target 714 genes (see Experimental Procedures and Figure 1A). The genes targeted with this library included 500 protein kinases (Manning et al., 2002), with the remaining targets comprising metabolic kinases (e.g., ATP-dependent 6-phosphofructokinases), lipid kinases (e.g., PIK3C2A), as well as proteins that lack kinase activity, but directly impact kinase signaling (e.g., the cyclin dependent kinase inhibitor CDKN1C). Cells were reverse transfected with siRNA and then cultured until cells reached 70% confluency (on average 4-7 days), at which point cell viability was assessed using the CellTiter-Glo assay. Following data processing and the application of quality control filters to ensure the reproducibility and high dynamic range of each screen (see Supplemental Information; Figure 1A; Table S1A), we retained 117 high quality screens for further analysis. The resulting resource (Table S1B) features tumor cell line models from ten different cancer types (breast, ovarian, lung, osteosarcoma, esophageal, pancreatic, head, and neck, cervical, CNS, and endometrioid; Figure 1B), and includes data for 69 lines not profiled in prior large-scale RNAi screens (Brough et al., 2011; Cowley et al., 2014; Koh et al., 2012).

To allow data to be compared between different cell lines, the viability data from each screen were standardized by the use of a robust Z score statistic (Table S1B). We considered candidate kinase genetic dependencies (KGDs) in the data set as those where the siRNA elicited $Z < -2$ effects. 76% of the kinases profiled in the screening library represented KGDs in at least one tumor cell model. Additionally, 53% and 26% represented KGDs in $\geq 5$ and $\geq 10$ cell lines, respectively (Figure 1C and Table S1C). On average, each tumor cell line model exhibited 51 KGDs. A set of six kinase-coding genes ($PLK1$, $AURKA$, $WE1$, $CHEK1$, $CDK11A$, and $GUCY2D$) represented KGDs in >70% of the cell lines screened and four of these ($PLK1$, $AURKA$, $WE1$, and $CHEK1$) are known to be involved in the mitotic cell-cycle checkpoint.

Candidate KGDs Associated with Tumor Histotypes

Using average linkage hierarchical clustering to cluster the siRNA Z score data (Figure 1D), we found that tumor cell lines frequently clustered according to tumor histotype. For example,
the majority of ovarian cancer cell lines formed a single cluster, as did those models derived from osteosarcomas (Figure 1D). Using median permutation (MP) tests on the Z score data, we found 151 KGDs associated with specific histotypes at a false discovery rate (FDR) of 0.1 (Table S1D). As expected, the number of KGDs associated with each histotype was correlated with the number of cell lines screened for that histotype (Spearman’s rho = 0.82), reflecting the increased statistical power resulting from a larger sample size. In breast cancer models, we found an increased requirement for ERBB3 and PIK3CA, members of the ERBB2 and PI3-kinase signaling pathways that are frequently dysregulated in this cancer histology (Miller et al., 2011). In contrast, models of osteosarcoma were more reliant upon genes involved in “skeletal system morphogenesis”, including PDGFRα, ACVR2B, TGFB2, DLG1, FGFR1, and FGFR2 (Su et al., 2008) (Gene Ontology enrichment p < 0.001 after correcting for multiple hypothesis testing, Berriz et al., 2009; Figures 2A and 2B). The FGFR1 and FGFR2 KGDs suggested that osteosarcoma models might be especially sensitive to small molecule FGFR inhibitors. Testing a set of 58 tumor cell lines for FGFR inhibitor sensitivity, we found AZD4547 (Gavine et al., 2012) and PD173074 (Bansal et al., 2003) to be more selective for osteosarcoma models (AZD4547, p = 7.6 \times 10^{-3}, PD173074 p = 3.9 \times 10^{-2}; Figures 2C and 2D; Table S1E) and
to have minimal effects in two non-tumor epithelial models (Figure S1). This osteosarcoma selective effect was independent of FGFR1 or FGFR2 amplification status and was also apparent when FGFR1 or FGFR2 amplified tumor cell lines were excluded from the analysis (AZD4572, p = 7.2 × 10^{-3} and PD173074, p = 4.3 × 10^{-2}; Figures 2C and 2D). Furthermore, the osteosarcoma selective nature of PD173074 was confirmed by a reanalysis of PD173074 sensitivity data derived from 660 tumor cell lines (Garnett et al., 2012) (Figure 2E; p = 1.4 × 10^{-3}). Taken together, these results suggested that FGFR inhibitors might show some utility in osteosarcoma, but that factors in addition to FGFR1 and FGFR2 amplification might explain drug sensitivity in this setting.

We also assessed the possibility that KGDs could be identified that were associated with specific subtypes of cancer. We, and others, have previously used RNAi data to identify KGDs associated with distinct breast cancer subtypes (Brough et al., 2011; Marcotte et al., 2012). To illustrate the utility of the expanded data set described here, we used MP tests to identify KGDs associated with the clear cell subtype of ovarian cancer (OCC).

We found three kinases (CAMK2N1, GRK2, and MAP3K9) to be KGDs in OCC models, compared to other ovarian cancer histologies such as serous ovarian cancer (Figure S2; Table S1F).

**Candidate KGDs Associated with Driver Gene Alterations**

By integrating the siRNA data with exome sequencing (Forbes et al., 2015) and copy number profiling data (Barretina et al., 2012), we identified KGDs associated with mutations in each of 200 candidate cancer driver genes (see Supplemental Information; Table S1G). We identified 4,247 putative dependencies associated with driver gene mutations (uncorrected MP test p ≤ 0.05; Table S1H). As the large number of tests performed using these 200 driver genes prohibited correction for multiple hypothesis testing, we focused our subsequent analysis on 21 key cancer driver genes (12 tumor suppressor genes and nine oncogenes; Futreal et al., 2004; Vogelstein et al., 2013) (Figure 3A) mutated in at least seven tumor cell lines in our panel. This identified 211 KGDs at an FDR of 0.5 (Table S1I) that could form the basis for subsequent validation experiments.

This approach reconfirmed the well-established ERBB2 oncogene addiction in models of breast cancer, but also established ERBB2 addiction/KGD in models of esophageal cancer (Figures 3B and 3C), where ERBB2 is recurrently amplified/over-expressed in 20% of tumors (Bang et al., 2010). This suggested that this particular genetic dependency was relatively independent of the underlying histotype. ERBB2 amplification was also associated with dependency upon other members of the ERBB2 signaling network including the ERBB2 binding partner ERBB3 (p = 2 × 10^{-3}), JAK2 (p = 1 × 10^{-3}), and the downstream effector of ERBB2, PIK3CA (p = 4 × 10^{-3}; Figure 3D). We found other KGDs associated with ERBB2 amplification, including a strong dependency upon the stress response kinase MEK3 (MAP2K3, p = 4 × 10^{-3}; Figure 3D) (Denjard et al., 1999) and PI3K1A (p = 2 × 10^{-5}), a kinase involved in inositol phosphate metabolism (Loijens and Anderson, 1996).

We assessed the possibility that some KGDs associated with cancer driver gene mutations might be private to or more profound in particular histotypes. For example, BRAF mutant melanomas are extremely sensitive to BRAF inhibition, whereas colorectal cancers with the same mutation show little response (Prahallad et al., 2012). We used a similar analysis as above to identify KGDs associated with driver gene mutations within particular histotypes and identified 943 KGDs (Tables S1J and S1K), compared to 211 in the prior analysis that combined all histotypes. Together, these 1,154 candidate dependencies could inform the design of subsequent validation studies. As an example of this, we selected for validation one of the KGDs associated with RB1 mutation in osteosarcoma (Kansara et al., 2014), DYRK1A (p = 6.8 × 10^{-3}; Figure S3A), a component of the DREAM complex (Sadasivam and DeCaprio, 2013) previously identified as a protein interaction partner of RB1 (Varjosalo et al., 2013). To confirm the dependency of RB1 null osteosarcoma models upon DYRK1A, we selected 14 osteosarcoma models and characterized these according to their RB1 mutation and protein expression status and established that multiple distinct DYRK1A siRNAs could replicate the RB1 selectivity observed in the initial screen as well as eliciting DYRK1A silencing (Figure S3). These results suggest that DYRK1A might represent a valid genetic dependency in RB1 defective osteosarcoma cells.

We also noted from our analysis of the siRNA data that some genetic dependencies associated with cancer driver gene mutations were observed independently in multiple histotypes. These included KGDs associated with ERBB2 amplification in breast and esophageal cancer models (e.g., ERBB2 p = 7.9 × 10^{-3} [breast] and p = 9.2 × 10^{-3} [esophageal] and MAP2K3 p = 3.3 × 10^{-2} [breast] and p = 4.4 × 10^{-3} [esophageal]; Figure S4A), but also a dependency upon the microtubule associated...
serine/threonine kinase MAST1 in CCND1 amplified breast or esophageal cancer models ($p = 1.1 \times 10^{-2}$ [breast] and $p = 1.3 \times 10^{-2}$ [esophageal]; Figure S3B). Likewise, a KGD upon Citron Rho-interacting kinase (CIT), a regulator of cytokinesis (Madaule et al., 1998) was also seen in CCND1 amplified breast or esophageal cancer models ($p = 2.0 \times 10^{-3}$ [breast], $p = 2.6 \times 10^{-3}$ [esophageal]; Figure 3E). In both osteosarcoma ($p = 1.4 \times 10^{-3}$) and lung cancer models ($p = 3.5 \times 10^{-2}$; Figure S1C), we identified an association between mutation/deletion of CDKN2A and dependency upon the cyclin dependent kinase gene CDK11A, which encodes a CDKN2A interacting protein (Varjosalo et al., 2013). In total, we identified 63 kinase dependencies associated with driver gene mutation status that were observed independently in more than one histotype (Table S1K).

**Integrating Data on Protein-Protein and Regulatory Interactions Facilitates the Interpretation of Genetic Dependencies**

The set of KGDs associated with cancer driver gene alterations can be used to frame testable hypotheses, such as “mutation in gene A drives dependency upon a gene B.” However, without further information, there are a number of potential mechanistic explanations for each genetic dependency. In model organisms, the problem of interpreting such dependencies has been addressed by integrating information from protein–protein (Beyer et al., 2007) and kinase-substrate interaction databases (Fiedler et al., 2009). To facilitate a mechanistic understanding of KGDs and to provide additional guidance for the design of subsequent experiments, we annotated our list of KGDs according to whether they involved known protein-protein interactions (Chatr-Aryamontri et al., 2015; Das and Yu, 2012), known kinase-substrate relationships (Lachmann and Ma’ayan, 2009), or known regulatory relationships (Cerami et al., 2011) between the driver gene and the identified dependency (see Experimental Procedures). Doing this, we found 113 KGDs involved pairs of genes with a previously reported functional relationship between the mutated driver gene and kinase target (Tables S1I and S1K).

For example, mutation/amplification of EGFR in lung cancer cell lines was associated with an increased dependency upon FES ($p = 3 \times 10^{-2}$; Figure 3F), previously identified as an EGFR binding partner (Jones et al., 2006). Similarly, in esophageal cancer models, we identified a significant association between mutation of the chromatin remodeling factor gene SMARCA4 and dependency upon the bromodomomain protein BRD4 ($p = 6 \times 10^{-3}$; Figure 3F), previously identified as a protein interaction partner of SMARCA4 (Rahman et al., 2011). Among the dependencies associated with a kinase-substrate interaction, we found that mutation of STK11 (LKB1) in ovarian cancer models was associated with an increased dependency upon MARK2 ($p = 2 \times 10^{-3}$; Figure 3G), an LKB1 substrate (Lizzano et al., 2004). Similarly, we found that MYC (cMYC) protein ($p = 1.2 \times 10^{-2}$; Figure 3G), which is known to phosphorylate and stabilize the cMYC protein (Sears et al., 2000). We also identified a series of dependencies between cancer driver genes and their transcriptional targets, the majority of which focused upon MYC. In lung cancer models, we found that MYC amplification was associated with an increased dependency upon CDKL5 ($5.6 \times 10^{-3}$; Figure 3H), a gene whose expression is regulated by MYC. Similarly, in esophageal models, we found MYC amplification to be associated with an increased dependency upon the MYC transcriptional target PRKCH (Zeller et al., 2006) ($p = 6.7 \times 10^{-3}$; Figure 3H).

For KGDs where a direct relationship between the driver gene and the kinase was not known, we used a simple information-flow type analysis to identify the shortest known molecular paths between driver gene and the kinase dependency (Tables S1I and S1K). For example, one of the strongest dependencies identified across all histotypes was between STK11 and SRP72 (Figure S4C). We found no evidence of a direct relationship between the two genes, but found that STK11 has been shown to regulate the expression of MYC (Nath-Sain and Marigani, 2009), which in turn has been shown to regulate SRP72 (Zeller et al., 2006), suggesting a putative path linking the driver gene and the kinase dependency. In esophageal cancer models, we found that ERBB2 amplification is associated with MASTL (Voets and Wolthuis, 2010) and NEK9 (Belham et al., 2003) KGDs (Figure 3I). We found no direct link between ERBB2 and either of these kinases, but both are CDK1 substrates and CDK1 itself is an ERBB2 substrate. In this instance, all members of the path (ERBB2/CDK1/NEK9/MASTL) were identified as ERBB2 dependencies. In total, 163 dependencies not supported by a direct link could be reached by adding one intermediate connection (e.g., CDK1 is an intermediate connection between ERBB2 and NEK9).

**Pathway and Network Level KGDs**

Work in model organisms has shown that a genetic mutation often results in an increased dependency on not just one gene, but multiple genes involved in a specific pathway or complex (Collins et al., 2007; Kelley and Ideker, 2005; Ryan et al., 2012). To explore the utility of this concept in interpreting our KGD data, we mapped the nominally significant KGDs ($p < 0.05$) identified for each cancer driver gene across all histotypes onto the high-confidence STRING functional interaction network (Franceschini et al., 2013) (see Experimental Procedures). For 11 of the 21 driver genes analyzed (KRAS, ERBB2, CCND1, PIK3CA, SMAD4, NOTCH2, ARID1A, NF1, FBXW7, MAP2K4, and RB1), we found that the dependencies associated with each driver gene were significantly more connected on the STRING interaction network than would be expected by chance (see Experimental Procedures; Figure S5; Table S1L). This suggested that these 11 driver genes might induce dependencies not just on individual genes, but on functional subnetworks. For two of these networks, we added known protein-protein and kinase-substrate interaction data to aid their interpretation. In the case of the network associated with ERBB2 amplification, this suggested that ERBB2 amplification might induce dependencies on direct binding partners and substrates of ERBB2 (JAK2, ERBB3, and PIK3CA), but also a network of genes involved in MAPK signaling (e.g., MAP2K3, MAP3K4, and MAP3K2) and inositol phosphate metabolism (including PIP5K1A, PIK3CA, and PIK3CD) (Figure 4A). Similarly, we found significantly more functional interactions among the kinases identified as dependencies associated with mutation of the tumor suppressor SMAD4, a member of the TGF-β pathway that...
**SMAD4** as the only driver gene significantly associated with sensitivity to A-443654, a pan-AKT inhibitor (Garraway and Lander, 2013). In addition to AKT1 and its substrates, we found a densely connected group of kinases that regulate the mitotic cell cycle in the **SMAD4** dependency network (Figure 4B), suggesting that **SMAD4** mutant tumor cell lines may have an increased sensitivity to perturbation of this process. To test this hypothesis, we analyzed a compendium of drug sensitivity profiles (Garrett et al., 2012) and found that **SMAD4** mutant cell lines have increased sensitivity to the Aurora Kinase inhibitor VX-680 (Harrington et al., 2004) \( (p = 4 \times 10^{-3}; \text{Figure 4C}) \). Furthermore, we found that **SMAD4** mutant cell lines also exhibited an increased sensitivity to the mitotic inhibitors paclitaxel \( (p = 8.3 \times 10^{-3}) \) and epothilone B \( (p = 3 \times 10^{-3}; \text{Figure 4C}) \), suggesting a general sensitivity to drugs that target the mitotic checkpoint.

We present the functional interaction networks for the dependencies associated with each driver gene in Figure S5 and Table S1. In addition to aiding the interpretation of dependencies, these subnetworks may be useful in alleviating some of the problems associated with false-positive effects in high-throughput genetic screens. Although there is a possibility of any given dependency being the result of off-target siRNA effects (Jackson and Linsley, 2010), the likelihood of an entire pathway being identified through off-target effects is likely to be much lower.

In the examples described above, we used the siRNA data to identify KGDs associated with defects in individual driver genes. Although there are hundreds of reported driver genes in cancer, some of these can be grouped into a small number of recurrently altered pathways (Garraway and Lander, 2013). Furthermore, it is possible that mutation in any member of such a pathway might have similar phenotypic effects. With this in mind, we considered whether we could identify candidate “pathway level” dependencies by grouping tumor cell lines according to mutations in any one of a set of driver genes belonging to the same pathway or complex. We obtained a previously curated list of pathways associated with driver gene mutations (Garraway and Lander, 2013) and manually updated this using literature information on well-established pathways (e.g., homologous recombination). For each pathway, tumor cell lines were grouped using a logical OR argument, i.e., if a cell line possessed a functional mutation of any gene member of the pathway then that cell line was considered mutated in that pathway. This resulted in a set of 15 pathway groupings (Table S1M) that were perturbed in at least seven tumor cell lines. Associating pathway mutations with KGDs was then performed in the same way as for individual genes using the MP test approach. This resulted in the identification of an additional 338 dependencies across all histotypes (Table S1N) and 748 histotype-specific dependencies (Table S1O).

As with individual driver genes, we found that the mutation of pathways was often associated with dependencies that were densely connected on the STRING functional interaction network. Indeed, using the dependencies identified across all histotypes, we found that nine of the 15 pathways (HR, PRC2, PI3K signaling, Cell Cycle Oncogenes, Cell Cycle Merged, TOR Signaling, MAPK Signaling, TGF B Signaling, and RAS/RAF Signaling) were associated with dependencies that were more functionally connected than would be expected by chance.

**Figure 4. Driver Gene KGDs and Functional Interaction Networks**

(A) Functional interaction network showing interactions between **ERBB2** amplification-associated KGDs. The nodes correspond to kinases that are identified as KGDs in **ERBB2** amplified cell lines. The nodes are scaled to indicate the significance of the KGD association \( p \) value. The blue edges correspond to experimentally determined protein-protein interactions, the pink arrows indicate the direction of experimentally determined kinase-substrate interactions, and the gray edges reflect high-confidence STRING functional interactions. Only KGDs that interact with at least one other **ERBB2** dependency are shown.

(B) Functional interaction network showing interactions among KGDs identified in **SMAD4** mutated cancer cell lines. Details as for **ERBB2** network in (A).

(C) Box plot showing AUC values of a panel of cell lines exposed to compounds targeting microtubules (paclitaxel and epothilone B) or Aurora Kinases (VX680) and classified into **SMAD4** mutant or wild-type groups. The top and bottom of the box represents the third and first quartiles and the box band represents the median (second quartile); whiskers extend to 1.5 times the interquartile distance from the box.

See also Figure S5 and Table S1L.

is frequently mutated or homozygously deleted in colorectal (Thiagalingam et al., 1996), pancreatic (Hahn et al., 1996), and esophageal (Dulak et al., 2013) cancers (Figure 4B). The integrated network we constructed from **SMAD4** KGDs revealed that AKT1 and a number of its substrates (**FGFR, MAP3K3, PIKFYVE, CHEK1**, and **WEE1**) were **SMAD4** mutation associated KGDs. Consistent with this, loss of **SMAD4** has been shown to be associated with increased AKT activation in colorectal and pancreatic tumor cell lines (Chen et al., 2014; Zhang et al., 2014). Furthermore, a recent large-scale drug screen identified **SMAD4** as the only driver gene significantly associated with sensitivity to A-443654, a pan-AKT inhibitor (Garrett et al., 2012). In addition to AKT1 and its substrates, we found a densely connected group of kinases that regulate the mitotic cell cycle in the **SMAD4** dependency network (Figure 4B), suggesting that **SMAD4** mutant tumor cell lines may have an increased sensitivity to perturbation of this process. To test this hypothesis, we analyzed a compendium of drug sensitivity profiles (Garrett et al., 2012) and found that **SMAD4** mutant cell lines have increased sensitivity to the Aurora Kinase inhibitor VX-680 (Harrington et al., 2004) \( (p = 4 \times 10^{-3}; \text{Figure 4C}) \). Furthermore, we found that **SMAD4** mutant cell lines also exhibited an increased sensitivity to the mitotic inhibitors paclitaxel \( (p = 8.3 \times 10^{-3}) \) and epothilone B \( (p = 3 \times 10^{-3}; \text{Figure 4C}) \), suggesting a general sensitivity to drugs that target the mitotic checkpoint.

We present the functional interaction networks for the dependencies associated with each driver gene in Figure S5 and Table S1. In addition to aiding the interpretation of dependencies, these subnetworks may be useful in alleviating some of the problems associated with false-positive effects in high-throughput genetic screens. Although there is a possibility of any given dependency being the result of off-target siRNA effects (Jackson and Linsley, 2010), the likelihood of an entire pathway being identified through off-target effects is likely to be much lower.

In the examples described above, we used the siRNA data to identify KGDs associated with defects in individual driver genes. Although there are hundreds of reported driver genes in cancer, some of these can be grouped into a small number of recurrently altered pathways (Garraway and Lander, 2013). Furthermore, it is possible that mutation in any member of such a pathway might have similar phenotypic effects. With this in mind, we considered whether we could identify candidate “pathway level” dependencies by grouping tumor cell lines according to mutations in any one of a set of driver genes belonging to the same pathway or complex. We obtained a previously curated list of pathways associated with driver gene mutations (Garraway and Lander, 2013) and manually updated this using literature information on well-established pathways (e.g., homologous recombination). For each pathway, tumor cell lines were grouped using a logical OR argument, i.e., if a cell line possessed a functional mutation of any gene member of the pathway then that cell line was considered mutated in that pathway. This resulted in a set of 15 pathway groupings (Table S1M) that were perturbed in at least seven tumor cell lines. Associating pathway mutations with KGDs was then performed in the same way as for individual genes using the MP test approach. This resulted in the identification of an additional 338 dependencies across all histotypes (Table S1N) and 748 histotype-specific dependencies (Table S1O).

As with individual driver genes, we found that the mutation of pathways was often associated with dependencies that were densely connected on the STRING functional interaction network. Indeed, using the dependencies identified across all histotypes, we found that nine of the 15 pathways (HR, PRC2, PI3K signaling, Cell Cycle Oncogenes, Cell Cycle Merged, TOR Signaling, MAPK Signaling, TGF B Signaling, and RAS/RAF Signaling) were associated with dependencies that were more functionally connected than would be expected by chance.

**Figure 4. Driver Gene KGDs and Functional Interaction Networks**

(A) Functional interaction network showing interactions between **ERBB2** amplification-associated KGDs. The nodes correspond to kinases that are identified as KGDs in **ERBB2** amplified cell lines. The nodes are scaled to indicate the significance of the KGD association \( p \) value. The blue edges correspond to experimentally determined protein-protein interactions, the pink arrows indicate the direction of experimentally determined kinase-substrate interactions, and the gray edges reflect high-confidence STRING functional interactions. Only KGDs that interact with at least one other **ERBB2** dependency are shown.

(B) Functional interaction network showing interactions among KGDs identified in **SMAD4** mutated cancer cell lines. Details as for **ERBB2** network in (A).

(C) Box plot showing AUC values of a panel of cell lines exposed to compounds targeting microtubules (paclitaxel and epothilone B) or Aurora Kinases (VX680) and classified into **SMAD4** mutant or wild-type groups. The top and bottom of the box represents the third and first quartiles and the box band represents the median (second quartile); whiskers extend to 1.5 times the interquartile distance from the box.

See also Figure S5 and Table S1L.
This suggested that mutation of one pathway may induce dependency on a second pathway, consistent with observations from yeast where it has been shown genetic dependencies can often be best explained as occurring between pairs of pathways (Kelley and Ideker, 2005). The dependency graphs associated with each pathway are presented in Figure S5.

In some instances, the association between a pathway and kinase siRNA had a predictive value no greater than the association with an individual member of the pathway. For example, alteration in the mTOR signaling pathway (mutation in TSC1 or TSC2 or STK11) was associated with an increased dependency upon the signal recognition particle SRP72 gene (rho = −0.40), but mutation of STK11 alone better explained the relative sensitivity of mutant and non-mutant cell lines in this regard (rho = −0.44). We therefore filtered these associations to identify 175 across-histotype (Table S1N) and 608 histotype-specific pathway dependencies where the pathway was a better predictor of dependency than any one individual gene (Table S1O). One example of a pathway dependency involved loss-of-function mutations in the genes encoding components of the SWI/SNF complex, mutated in ~20% of all human cancers. By grouping all tumor cell lines that had a loss-of-function mutation or homozygous deletion of any member of the SWI/SNF complex (including the genes ARID1A, SMARCA1, SMARCA4, ARID2, ARID1B, and PBRM1) and then carrying out MP tests on the siRNA data as before, we identified ten KGDs including TWF2 (Figure 5A), a gene encoding a protein that affects the stability of the actin cytoskeleton through interaction with G-actin (Pivovarova et al., 2013). Further dependencies were identified for this complex within specific histotypes including the uridine-cytidine kinase gene UCK2 (Van Rompay et al., 2001) in ovarian cancer models (Figure 5B) and the death-associated protein kinase gene DAPK1 in esophageal cancer models (Figure 5C).

We also investigated MAPK gene alterations (including RAS gene or BRAF mutations) as a pathway and found a CDK6 KGD (Figure 5D). The dependency of KRAS mutant tumor cell lines upon CDK6 was readily apparent (rho = −0.38), but was stronger when KRAS, HRAS, NRAS, or BRAF mutant tumor cell line...
models were combined as a group (rho = −0.43). CDKs, including CDK6, have been identified by a number of groups as putative non-oncogene addictions for KRAS mutant cancers (Barbie et al., 2009; Puyol et al., 2010). Our results suggest that CDK6 might be a non-oncogene addiction not just for KRAS mutant models, but also for cell lines with any one of a variety of MAPK activating mutations (NRAS, HRAS, and BRAF). We tested this hypothesis using published drug screening results for a CDK4/6 inhibitor (PD0332991) in 628 cell lines (Garnett et al., 2012) and found a significant association between mutation of the MAPK pathway and sensitivity to this inhibitor (p = 2.5 × 10^{-3}, Mann-Whitney U [MW U]-test). None of the individual members of this pathway showed as strong an association with this inhibitor (KRAS p = 1.9 × 10^{-1}, HRAS p = 5.8 × 10^{-2}, NRAS p = 1.7 × 10^{-2}, BRAF p = 2.8 × 10^{-2}, and MW U-test).

**DISCUSSION**

A key challenge in the study of cancer biology is to understand how driver mutations alter the cellular state to promote tumor progression and how these altered states may be exploited in the development of targeted therapeutic approaches to the disease (Yaffe, 2013). Here, we have used siRNA screening to quantitatively estimate the kinase requirements of tumor cell lines in an attempt to understand better the genetic dependencies present. By integrating our siRNA data with molecular and histotype classifications, we have identified KGDs associated with particular cancer histologies or the presence of particular driver gene mutations. By integrating the KGD data with additional sources of annotation, such as protein-protein interaction data, we have tried to exemplify how testable hypotheses can be framed to explain the associations between a biomarker, such as a driver gene mutation, and a kinase dependency. Our aim in providing this data and illustrating its potential utility is to present starting points for further work.

As with any large functional data set, it is important to point out where elements of the technology used might influence the interpretation of the data. In general, siRNA mediated gene silencing is transient, when compared to, for example, short hairpin (sh) RNA mediated RNAi. With this in mind, we used a relatively short cell culture period between transfection and cell viability assessment (4 to 7 day period). Nevertheless, we cannot predict whether longer-term cell culture or longer-term gene silencing might result in a somewhat different profile of genetic dependencies. Furthermore, we used an ATP-based assay of cell viability in the screens. Some modes of cell inhibition exist that might result in a somewhat different profile of genetic dependencies. For example, integrating protein-protein interaction data with functional data (Figure 5) was an approach pioneered in the study of yeast genetic interaction screens (Beyer et al., 2007; Kelley and Ideker, 2005). Here, we have integrated this type of data to help frame testable hypotheses relating to the observed dependencies. A more sophisticated level of protein-protein interaction data for human tumor cell lines (Krokan et al., 2015) will undoubtedly enhance our ability to understand genetic dependencies. Similarly the availability of phosphoproteomic data for the cell lines in our panel may facilitate a more mechanistic reconstruction of the signaling networks active in each cell line. A number of approaches (e.g., So et al., 2015; Terfve and Saez-Rodriguez, 2012) have been developed to integrate siRNA or small molecule perturbations with time-course phosphoproteomics data sets to reconstruct signaling networks. Currently phosphoproteomic data for cancer cell line panels are relatively limited (e.g., Casado et al., 2013; Creixell et al., 2015), but as the overlap of cell lines covered by these phosphoproteomic resources and our siRNA resource increases there will be opportunities for the development of further integrative modeling approaches. Similarly, the increased availability of protein expression data sets (e.g., Lawrence et al., 2015; Moghaddas Gholami et al., 2013) may provide further opportunities for the development of additional integrative approaches.

Finally, to make our resource as useful to the community as possible, we have made all of the data described in this manuscript available (https://cansar.icr.ac.uk/), alongside the computational scripts used to integrate data (https://github.com/GeneFunctionTeam/cell_line_functional_annotation).

**EXPERIMENTAL PROCEDURES**

**siRNA and Small Molecule Screening**

Cell lines were transfected with a plate-arrayed siRNA library targeting 714 kinases and kinase-related genes (Dharmacon SMARTpools). Positive control (siPLK1) and multiple negative controls (siCON1 and siCON2; Dharmacon, catalog numbers D-001210-01-20 and D-001206-14-20) and AllStar (QIAGEN, catalog number 1027281) were included on every plate. 20 breast cancer models were screened in a 96-well-plate format while the remaining cell lines were screened in a 384-well-plate format (Table S1A). All screens were performed in triplicate. Cell viability was estimated as cells reached 70% confluency (normally 4–7 days after transfection) using a CellTiter-Glo assay (Promega). Data processing and quality control was performed using the cellHTS R package (Boutros et al., 2008). Further details, including small molecule sensitivity testing, are provided in the Supplemental Information.

**Association Testing**

To identify associations between specific features (histotype or driver gene mutation) and sensitivity to specific siRNAs, a one-sided MP test was used. For each siRNA, we compared the observed difference between the median
Z score of the interest group and the median Z score of the “other” group to that expected based on random permutation. There were one million random samples that were created with the same sample sizes as the interest and other groups and the difference in the medians of the two groups calculated, allowing an empirically determined p value to be calculated. Correction for multiple testing was performed using the Benjamini and Hochberg FDR (Benjamini and Hochberg, 1995) and only those at an FDR of 50% are reported. For all small molecule association tests, we used a one-sided MW U test on area under the dose response curve values.

Data Access
All siRNA Z score data can be found in Table S1B and also at https://cansar.icr.ac.uk/.

Data Integration
Data from HINT (Das and Yu, 2012), BioGRID version 3.4.128 (Chatr-Aryamontri et al., 2015), and KEA protein-protein interaction databases were used (Lachmann and Ma’ayan, 2009). Kinase-substrate interactions were obtained from KEA (Lachmann and Ma’ayan, 2009), PhosphoSitePlus (Hornbeck et al., 2013) and Cheng et al., 2014. High confidence (combined score >0.7) functional interactions were obtained from the STRING database (Version 9.1; Franceschini et al., 2013). Gene expression relationships were obtained from Pathway Commons (Cerami et al., 2011). The shortest_path function in NetworkX (Hagberg et al., 2008) was also used. Further details are provided in the Supplemental Information.

ACCESSION NUMBERS
The accession number for the exome sequencing data of 11 ovarian cancer cell lines reported in this study is ENA: PRJEB9639.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, five figures, and one table and can be found with this article online at https://dx.doi.org/10.1016/j.celrep.2016.02.023.

AUTHOR CONTRIBUTIONS
C.J.L., A.A., C.J.R., and J.C. designed the experiments and wrote the manuscript. J.C. and C.J.R. performed statistical analyses and data integration. H.H., H.N.P., I.B., I.Y.C., J.F., R.B., R.M., R.R., S.C.-C., and S.P.-V. performed protein quantitation. H.N.P. performed RT quantitative (q) PCR. A.G., J.C., and W.W. processed the siRNA and small molecule inhibitor samples that were created with the same sample sizes as the interest and approved the final manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS
We thank Ulltán McDermott from the Wellcome Trust Sanger Institute for sharing DNA sequencing data prior to publication. This work was funded by Cancer Research UK (grant number C347/A8363), Breast Cancer Now, UCSF, and The EU FP7 project EurocanPlatform (grant number 280791). C.J.R. is a Sir Henry Wellcome Fellow jointly funded by Science Foundation Ireland, the Health Research Board, and the Wellcome Trust (grant number 103049/Z/13/Z) under the SFI-HRB-Wellcome Trust Biomedical Research Partnership. We acknowledge National Health Service (NHS) funding to the National Institute for Health Research (NIHR) Royal Marsden Hospital Biomedical Research Centre.

Received: July 31, 2015
Revised: November 7, 2015
Accepted: February 1, 2016
Published: March 3, 2016

REFERENCES


