

Systematic identification of genomic markers of drug sensitivity in cancer cells

Mathew J. Garnett^{1*}, Elena J. Edelman^{2*}, Sonja J. Heidorn^{1*}, Chris D. Greenman^{1†}, Anahita Dastur², King Wai Lau¹, Patricia Greninger², I. Richard Thompson¹, Xi Luo², Jorge Soares¹, Qingsong Liu^{3,4}, Francesco Iorio^{1,5}, Didier Surdez⁶, Li Chen², Randy J. Milano², Graham R. Bignell¹, Ah T. Tam², Helen Davies¹, Jesse A. Stevenson², Syd Barthorpe¹, Stephen R. Lutz², Fiona Kogera¹, Karl Lawrence¹, Anne McLaren-Douglas¹, Xenia Mitropoulos², Tatiana Mironenko¹, Helen Thi², Laura Richardson¹, Wenjun Zhou^{3,4}, Frances Jewitt¹, Tinghu Zhang^{3,4}, Patrick O'Brien¹, Jessica L. Boisvert², Stacey Price¹, Wooyoung Hur^{3,4}, Wanjuan Yang¹, Xianming Deng^{3,4}, Adam Butler¹, Hwan Geun Choi^{3,4}, Jae Won Chang^{3,4}, Jose Baselga², Ivan Stamenkovic⁷, Jeffrey A. Engelman², Sreenath V. Sharma^{2†}, Olivier Delattre⁶, Julio Saez-Rodriguez⁵, Nathanael S. Gray^{3,4}, Jeffrey Settleman², P. Andrew Futreal¹, Daniel A. Haber^{2,8}, Michael R. Stratton¹, Sridhar Ramaswamy², Ultan McDermott¹ & Cyril H. Benes²

Clinical responses to anticancer therapies are often restricted to a subset of patients. In some cases, mutated cancer genes are potent biomarkers for responses to targeted agents. Here, to uncover new biomarkers of sensitivity and resistance to cancer therapeutics, we screened a panel of several hundred cancer cell lines—which represent much of the tissue-type and genetic diversity of human cancers—with 130 drugs under clinical and preclinical investigation. In aggregate, we found that mutated cancer genes were associated with cellular response to most currently available cancer drugs. Classic oncogene addiction paradigms were modified by additional tissue-specific or expression biomarkers, and some frequently mutated genes were associated with sensitivity to a broad range of therapeutic agents. Unexpected relationships were revealed, including the marked sensitivity of Ewing's sarcoma cells harbouring the *EWS* (also known as *EWSR1*)-*FLI1* gene translocation to poly(ADP-ribose) polymerase (PARP) inhibitors. By linking drug activity to the functional complexity of cancer genomes, systematic pharmacogenomic profiling in cancer cell lines provides a powerful biomarker discovery platform to guide rational cancer therapeutic strategies.

There is compelling evidence that the likelihood of a patient's cancer responding to treatment can be strongly influenced by alterations in the cancer genome. For example, the use of drugs to selectively target the protein product of the *BCR-ABL* translocation in chronic myeloid leukaemia (CML) has revolutionized the treatment of this disease, with five-year survival rates of 90% in treated patients¹. Although targeting of specific genetic changes in defined patient subsets has been successful, a poorly explained range of responses to appropriately selected therapies is often still observed in patients^{2,3}. Moreover, a large number of cancer drugs have not been linked to specific genomic alterations that could be used as biomarkers to specify their selective therapeutic effectiveness⁴. As drug pipelines generate new classes of compounds, systematic methods to identify predictive biomarkers during their early development could have a profound effect on the design, cost and ultimate success of new cancer drug development.

The NCI60 cell line panel and associated drug screens pioneered the approach of using cancer cell lines to link drug sensitivity with genotype data^{5,6}. Cancer cell lines have subsequently been used to identify rare drug-sensitizing genotypes, including mutant *EGFR*, *BRAF* and the *MLL4-ALK* translocation, which are highly predictive of clinical responses^{2,3,7}. Here, we report the results of a large-scale screen of human cancer cell lines, incorporating detailed genomic and gene expression analysis, to identify systematically drug-sensitivity biomarkers to a broad range of cancer drugs.

Therapeutic biomarker discovery

To capture the high degree of genomic diversity in cancer and to identify rare mutant subsets with altered drug sensitivity, we assembled 639 human tumour cell lines, representing the spectrum of common and rare types of adult and childhood cancers of epithelial, mesenchymal and haematopoietic origin (Fig. 1a and Supplementary Data 1). Cell lines were subjected to sequencing of the full coding exons of 64 commonly mutated cancer genes, genome-wide analysis of copy number gain and loss using Affymetrix SNP6.0 microarrays, and expression profiling of 14,500 genes using Affymetrix HT-U133A microarrays. The presence of seven commonly rearranged cancer genes and of microsatellite instability (MSI) was also investigated. The 130 drugs selected for screening covered a wide range of targets and processes implicated in cancer biology (Fig. 1b and Supplementary Data 2). They encompassed both targeted agents ($n = 114$) and cytotoxic chemotherapeutics ($n = 13$), including approved drugs used in clinical practice ($n = 31$), drugs in development undergoing studies in clinical trials ($n = 47$), and experimental tool compounds ($n = 52$). To gain insight into drug-to-drug variation, we included multiple drugs designed against well-credentialed targets (Fig. 1b). The effect of 72 h of drug treatment on cell viability was used to derive a multi-parameter description of drug sensitivity, including the half-maximal inhibitory concentration (IC_{50}), and the slope of the dose-response curve (Supplementary Fig. 1). In total, we assayed 48,178 drug-cell-line

¹Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. ²Massachusetts General Hospital Cancer Center, Harvard Medical School, Charlestown, Massachusetts 02129, USA.

³Department of Cancer Biology, Dana Farber Cancer Institute, 44 Binney Street, Boston Massachusetts 02115, USA. ⁴Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Avenue, Boston, Massachusetts 02115, USA. ⁵EMBL-EBI, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. ⁶Laboratoire de génétique et biologie des cancers, Institut Curie, 75248 Paris, Cedex 05, France. ⁷Division of Experimental Pathology, Institute of Pathology, Centre Hospitalier Universitaire Vaudois (CHUV), 1005 Lausanne, Switzerland. ⁸Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. [†]Present addresses: Department of Computing, University of East Anglia, Norwich NR4 7TJ, UK (C.D.G.); The Genome Analysis Centre, Norwich Research Park, Norwich NR4 7UH, UK (C.D.G.); Oncology Drug Discovery, Novartis Institutes for Biomedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA (S.V.S.).

*These authors contributed equally to this work.

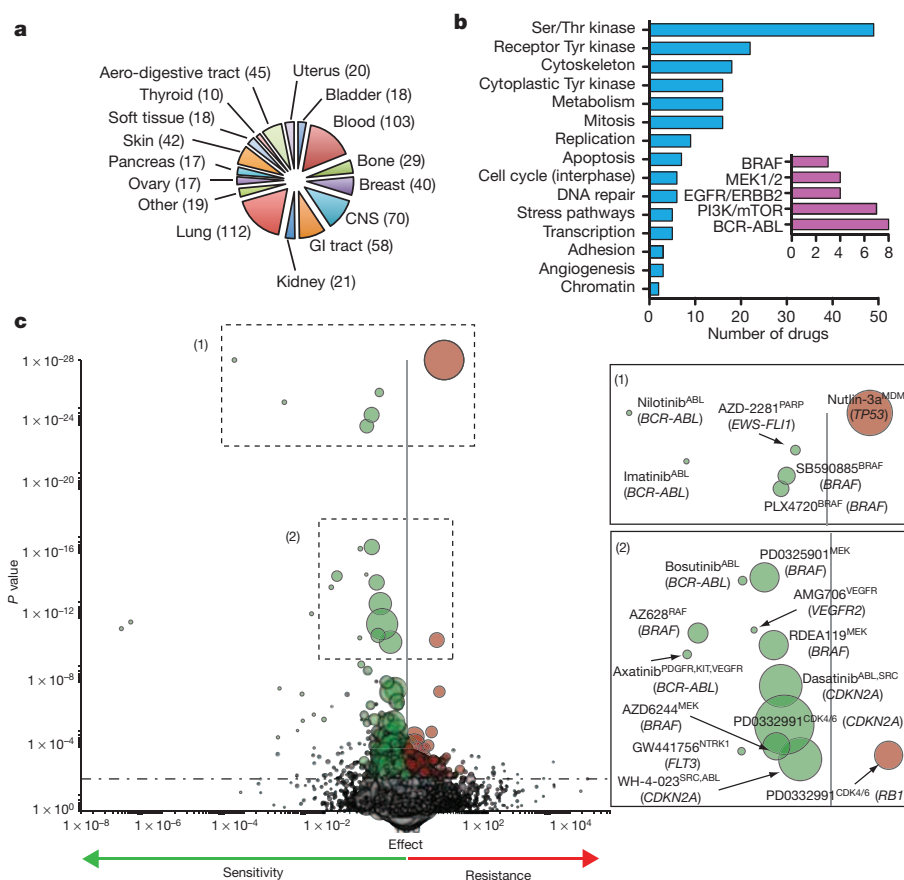


Figure 1 | A systematic screen in cancer cell lines identifies therapeutic biomarkers. **a**, The number of tumour-derived cell lines used for screening classified according to tissue type ($n = 639$ in total). CNS, central nervous system; GI, gastrointestinal. **b**, The panel of 130 screening drugs classified according to their therapeutic targets, primary effector pathways, and cellular functions. A single drug may be included in multiple categories. The inset indicates the number of drugs screened against a selection of prototype cancer targets. **c**, A volcano plot representation of MANOVA results showing the magnitude (effect, x -axis) and significance (P value, inverted y -axis) of all

drug-gene associations. Each circle represents a single drug-gene interaction and the size is proportional to the number of mutant cell lines screened (range 1–334). The horizontal dashed line indicates the threshold of statistical significance (0.2 false discovery rate, $P < 0.0099$). Insets (1) and (2) are magnified views of selected highly significant associations; the drug name, therapeutically relevant target(s) (in superscript) and cancer gene (in brackets) are given for each. The P values for nilotinib^{ABL} (BCR-ABL), $P = 2.54 \times 10^{-65}$, and nutlin-3a^{MDM2} (TP53), $P = 2.78 \times 10^{-37}$, have been capped at 1×10^{-28} in this representation.

combinations with a range of 275–507 cell lines screened per drug (mean = 368 cell lines per drug; Supplementary Data 2 and Supplementary Fig. 2). Clustering of compounds across cell lines based on IC_{50} values indicated that drugs with overlapping specificity were highly correlated, supporting the selectivity of the biological effects observed in the data set (Supplementary Fig. 3 and Supplementary Tables 1 and 2).

Tumours from a particular tissue frequently have a shared set of somatic mutations. To gain insight into how this might relate to drug sensitivity, we performed an analysis to identify associations between cancer tissue type and drug sensitivity based on IC_{50} values. As expected, in some instances tumour-type-specific sensitivity may be explained by the prevalence of cancer gene mutations (for example, breast cancer sensitivity to inhibitors of the phosphatidylinositol 3-kinase (PI3K) pathway that is commonly altered in this tumour type; Supplementary Data 3 and 4). In other cases, however, our current understanding of cancer genomes could not explain the observed associations. For instance, renal cell carcinoma (RCC) cells were sensitive to five SRC inhibitors (for example, AZD0530, $P < 1 \times 10^{-4}$, $n = 9$ RCC and 294 non-RCC cell lines)⁸, glioma cells were sensitive to a ROCK inhibitor (GSK269962A, $P < 1 \times 10^{-6}$, $n = 23$ glioma and 266 non-glioma cell lines)⁹. This analysis also identified therapeutic associations already used in the clinic with incompletely understood molecular bases, such as the sensitivity of myeloma cells to lenalidomide ($P < 1 \times 10^{-5}$, $n = 3$ myeloma and

455 non-myeloma cell lines)¹⁰. For most drugs, however, sensitive cell lines were scattered across multiple cancer types.

Cancer genes are drug-sensitivity biomarkers

Single gene mutations are increasingly being adopted as clinical biomarkers for the optimal application of cancer therapeutics. To identify associations between individual mutated cancer genes and drug sensitivity across the cell line panel, we used a multivariate analyses of variance (MANOVA) incorporating the IC_{50} value and slope of the dose-response curve. This analysis revealed a large number of individual gene-drug associations, a subset of which (448/9,039, 5%) were highly significant and are discussed here (Fig. 1c and Supplementary Data 5). Interestingly, most of the cancer genes analysed, including those that are not known to be direct targets of the drugs tested, were associated with either sensitivity or resistance to at least one drug in our panel (65/69, 94%) (Supplementary Fig. 4). Similarly, sensitivity to most drugs tested was associated with a mutation in at least one cancer gene (118/130, 91%). Thus, diverse cancer gene mutations are implicated as markers of sensitivity or resistance to a broad range of anticancer drugs, indicating that genomic biomarkers could inform the therapeutic selectivity of many cancer drugs.

The mutated cancer genes most clearly associated with drug sensitivity are oncogenes that are direct targets of the relevant drug. For example, the BCR-ABL rearrangement conferred sensitivity to multiple ABL inhibitors (for example, $P = 2.54 \times 10^{-65}$ for nilotinib;

Fig. 1c and Supplementary Fig. 5)¹, several of which are approved for CML treatment. Similarly, *BRAF* mutation was associated with sensitivity to BRAF and inhibitors of MEK1 and MEK2 (for example, $P = 1.25 \times 10^{-24}$ for PLX4720; Figs 1c and 2a–c)³, including a structural analogue of vemurafenib, which in clinical trials has extended the survival of *BRAF*-mutation-positive melanoma patients. Additionally, *ERBB2* (also known as *HER2*) amplification was associated with sensitivity to EGFR-family inhibitors including lapatinib ($P < 1 \times 10^{-7}$; Fig. 2d)¹¹, which is licensed for the treatment of HER2-positive breast cancer. We were also able to detect known associations between *EGFR*, *FLT3* and *PIK3CA* mutations and drugs that target the products of these genes (Supplementary Data 5)^{12,13}. A number of associations were driven by marked responses in small subsets of outlier cell lines. For example, two *FGFR2*-mutated cell lines were exquisitely sensitive to the FGFR inhibitor PD-173074 (Fig. 2e; $P < 1 \times 10^{-5}$)^{14,15}, confirming the need for large panels of cell lines to capture low-frequency drug-sensitizing genotypes.

We also found associations between the presence of inactivating mutations in tumour suppressor genes and several drugs, which in some instances provide insight into the interplay between tumour suppressors and the cellular machinery in mediating drug sensitivity.

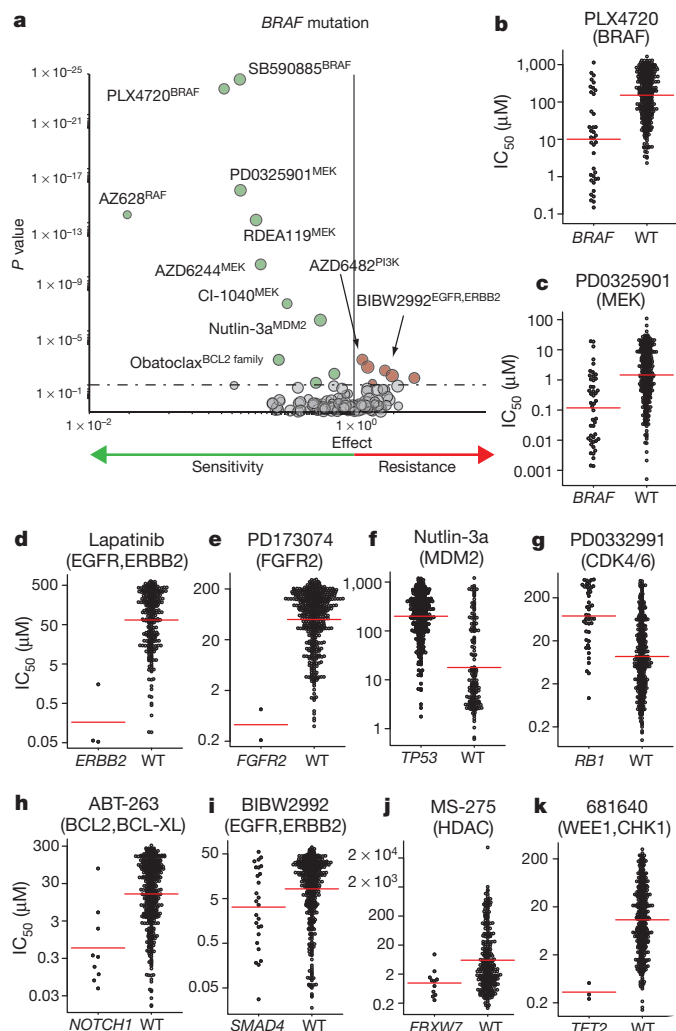


Figure 2 | Biomarkers of drug sensitivity and resistance. a, Gene-specific volcano plot of drug sensitivity associated with *BRAF* mutations in cancer cell lines (range 22–54). b–k, Scatter plots of cell line IC_{50} (μM) values from selected drug–gene associations. IC_{50} values are on a log scale comparing mutated (gene symbol given) or non-mutated (wild type (WT)) cell lines. Each circle represents the IC_{50} of one cell line and the red bar is the geometric mean. The drug name is indicated above each plot and therapeutic drug target(s) are bracketed.

For example, mutation of *TP53*, an important regulator of apoptosis and cell cycle arrest in response to cellular stress, confers resistance to nutlin-3a ($P < 1 \times 10^{-36}$), an inhibitor of the MDM2 E3 ligase that negatively regulates p53 protein levels (Fig. 2f)¹⁶. Similarly, mutational inactivation of *RBI1*, a key repressor of cell cycle progression in normal cells, confers resistance to PD-0332991 ($P < 1 \times 10^{-10}$), an inhibitor of the upstream cyclin-dependent kinases (CDKs) 4 and 6, which drive cell cycle progression by inhibiting pRb through phosphorylation (Fig. 2g)¹⁷. Conversely, mutational inactivation of *CDKN2A*, encoding the CDK inhibitory protein p16, was associated with sensitivity to PD-0332991 ($P < 1 \times 10^{-11}$; Supplementary Data 5)¹⁷, presumably because *CDKN2A*-mutated cells have an enhanced requirement for signalling through the CDK4/6–pRb signalling pathway.

In other instances genomic associations appear related to enrichment of mutations in a particular tissue type. The association of the *BRAF* and *NRAS* mutations with sensitivity to obatoclax mesylate, a pro-apoptotic drug that targets BCL2 family anti-apoptotic proteins (BCL2, BCL-XL (also known as BCL2L1) and MCL1), probably results from the enrichment of these mutations in melanoma, as drug sensitivity among melanoma cell lines was not correlated with the presence or absence of these mutations (Supplementary Fig. 6). The tissue-specific effect of obatoclax may be related to inhibition of the melanoma survival mediator MCL1 (ref. 18), because sensitivity of melanoma lines to ABT-263, another BCL2 inhibitor that does not target MCL1, was not correlated with *BRAF* or *NRAS* mutation. Moreover, an ABT-263-insensitive melanoma cell line can be sensitized to this drug by short interfering (si)RNA-mediated depletion of MCL1 (Supplementary Fig. 7).

The genomic associations identified for 13 clinically approved cytotoxic chemotherapeutics in our panel were generally less significant than for targeted drugs, indicating that single gene biomarkers may be less informative for this class of drugs with broad action across many cancers (Supplementary Figs 8 and 9). Intriguingly, we did not find general associations between targeted or cytotoxic drug-sensitivity patterns and mutations in *TP53*. It may be that functional inactivation of p53, through mutations or abrogation of signalling pathways that regulate its activity, is an almost universal feature of cancer cell lines and thus differential drug sensitivity between mutant and non-mutant cell lines is not observed¹⁹.

Several other novel gene–drug associations were identified that cannot be readily explained on the basis of our current knowledge of signalling pathways and may reflect unappreciated biological relationships. Mutation of *NOTCH1* was associated with sensitivity to ABT-263 ($P < 1 \times 10^{-9}$; Fig. 2h and Supplementary Fig. 10), perhaps due to decreased expression of BCL2 family members in *NOTCH1*-mutant cell lines (Supplementary Fig. 11). Amplification of *CCND1* (cyclin D1) or loss of *SMAD4* were associated with sensitivity to multiple EGFR-family inhibitors including lapatinib and BIBW2992; and for *SMAD4* this correlated with elevated *EGFR* gene expression (Fig. 2i and Supplementary Fig. 12). Inactivation of *STK11* (also known as *LKB1*; $P < 0.01$), thought to relieve repression of mTOR, was associated with sensitivity to the HSP90 inhibitor 17-AAG. Additionally, loss of *FBXW7* was associated with sensitivity to the histone deacetylase (HDAC) inhibitor MS-275 ($P < 1 \times 10^{-5}$; Fig. 2j), and *TET2* loss was associated with sensitivity to the WEE1 and CHK1 (also known as CHEK1) inhibitor 681640 ($P < 1 \times 10^{-4}$; Fig. 2k). These associations, and others presented here (Supplementary Data 5), represent candidate biomarkers of drug sensitivity and may ultimately be useful for the deployment of targeted therapies in cancer.

Complex genomic correlates of drug sensitivity

In most instances sensitivity of cancer cells to drugs is likely to depend on a multiplicity of genomic and epigenomic variables. Indeed, single gene–drug associations were only rarely able to explain the range of drug sensitivities observed across cell lines for any given drug (Fig. 2). We thus applied elastic net regression²⁰, a penalized linear modelling

technique, to identify cooperative interactions among multiple genes and transcripts across the genome and defined response signatures for each drug. Elastic net regression identified 26,938 feature–drug associations (Supplementary Data 6) from which 534 associations corresponding to 69 different drugs were highly significant (defined as $-2.95 > \text{effect} (e) > 2.79$ and frequency $(f) > 0.76$; Fig. 3a and Supplementary Fig. 13 and Supplementary Data 7).

In many instances transcriptional features showed correlations with drug sensitivity that were equal to or stronger than those observed with gene mutation (Fig. 3a and Supplementary Table 3). For example, although sensitivity to the EGFR and ERBB2 inhibitor lapatinib correlated with *ERBB2* expression and mutation, the strongest correlate for this drug was actually expression of the matrix metalloproteinase *MMP28* ($e = -29.28, f = 1$) (Fig. 3a). Notably, for most drugs, including

those with clear linkage to cancer gene mutations, elastic net modelling identified multi-feature signatures of drug sensitivity. For example, together with *BRAF* mutation, sensitivity to RAF or MEK1 and MEK2 inhibitors was recurrently associated with 67 features. These features included expression of *SPRY2*, *DUSP4* and *DUSP6*, which are known regulators of MAPK signalling (Supplementary Fig. 14)^{21,22}. Interestingly, expression of 8 genes identified as markers of sensitivity to the MEK inhibitor AZD6244 significantly overlapped with an 18-gene signature of sensitivity to this drug (hypergeometric test of the overlap significance: $P = 3 \times 10^{-9}$)²³. In some cases, elastic net modelling identified complex patterns of biomarkers corresponding to distinct subsets of sensitive cancer cell lines. Thus, sensitivity to dasatinib, an inhibitor of multiple kinases including ABL and SRC, correlated with both *BCR-ABL* translocation and with multi-gene transcriptional signatures that were expressed in sensitive cell lines lacking that gene translocation (Fig. 3b).

Elastic net modelling also identified transcriptional correlates of sensitivity for drugs without a known sensitizing mutational event. This included expression of LAG3, which correlated with sensitivity to the SGK inhibitor GSK-650394 ($e = -29.9, f = 1$) and the correlation between expression of the NADPH dehydrogenase family member *NQO1* with sensitivity to the HSP90 inhibitor 17-AAG ($e = -22.21, f = 1$; Fig. 3c). Consistent with these findings, *NQO1* was previously shown to metabolize 17-AAG into a more potent inhibitor of HSP90 (ref. 24).

A small number of features were recurrently associated with altered sensitivity to drugs from different classes, indicating that they may be broadly involved in mediating drug sensitivity by diverse mechanisms such as drug efflux (for example, *ABCB1*; Supplementary Data 6). To give further insight into this data set, we mapped the elastic net drug signatures onto the target of the drugs (Supplementary Data 7) and onto 457 known cancer genes (<http://www.sanger.ac.uk/genetics/CGP/Census/>; Supplementary Data 8). Collectively, these observations illustrate that in many instances multi-feature genomic signatures incorporating markers related to mutations, tissue lineage, cellular differentiation states and cellular pathways have the potential to expand and refine our current understanding of drug sensitivity.

EWS-FLI1 is a PARP inhibitor biomarker

We identified a highly significant association between the *EWS-FLI1* rearrangement that is characteristic of Ewing's sarcoma tumours and sensitivity to olaparib (AZD2281), an inhibitor of PARP ($P = 1.03 \times 10^{-26}$; Figs 1c and 4a). Screening of a structurally distinct PARP inhibitor, AG-014699, across a large panel of cell lines confirmed the sensitivity of Ewing's sarcoma cell lines (geometric mean IC_{50} for *EWS-FLI1* = 4.7 μM versus 64 μM for wild type, $P < 0.0001$ Mann–Whitney test, $n = 291$ cell lines; Fig. 4a and Supplementary Fig. 15 and Supplementary Data 9). Cells from Ewing's sarcoma were more sensitive to olaparib ($P = 2.84 \times 10^{-11}$) than cells from other tumour types, including sarcomas of bone and soft tissue (Supplementary Fig. 15 and Supplementary Data 3). PARP inhibitors have activity in *BRCA1*- and *BRCA2*-mutant cancers due to the defects in homologous recombination present in these tumours and their consequent reliance on alternative DNA damage repair pathways that are targeted by these inhibitors²⁵. A comparison of olaparib and AG-014699 sensitivity in a panel of cell lines using a 6-day viability assay and colony formation experiments confirmed the marked sensitivity of Ewing's sarcoma to PARP inhibitors, an effect that was comparable to that observed in *BRCA*-deficient cells (Fig. 4b, c and Supplementary Figs 16 and 17)²⁶. Furthermore, treatment with olaparib selectively induced apoptosis in Ewing's sarcoma compared to control cells (Fig. 4d). Unlike in Ewing's sarcoma, we did not observe an association between *BRCA1* and *BRCA2* mutations and sensitivity to PARP inhibitors in the 3-day screening format, which is probably due to a requirement for several rounds of division in these cells to accumulate toxic levels of DNA damage.

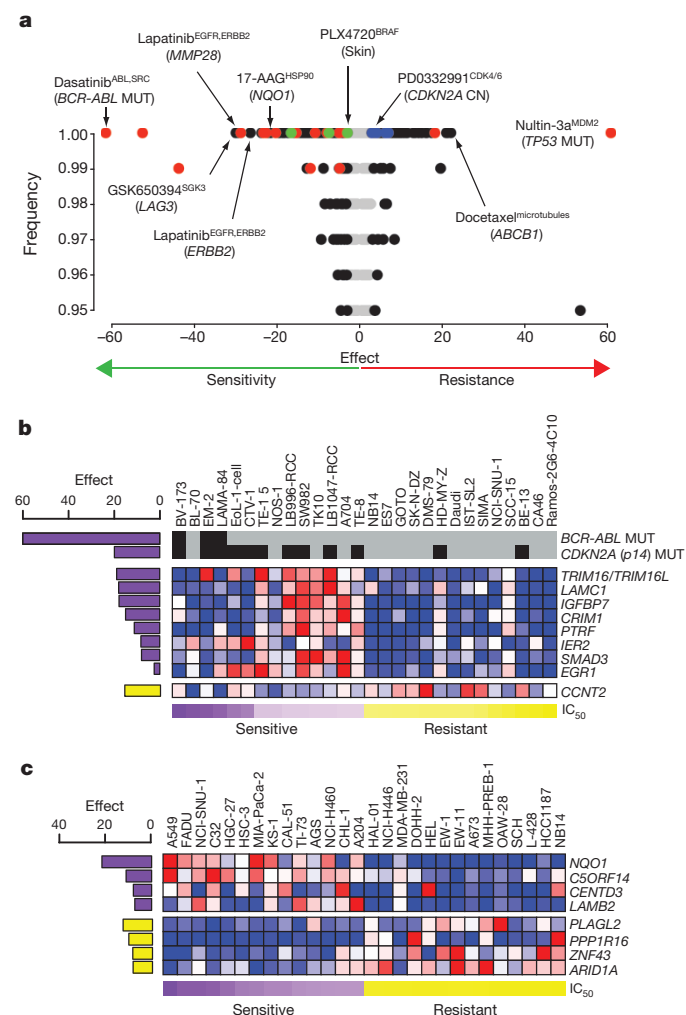


Figure 3 | Multi-feature genomic signatures of drug response. **a**, The top drug–feature associations identified by the elastic net are plotted for their frequency and effect size. Associations are coloured black for expression features, red for mutations, blue for copy number, and green for tissue. MUT, mutation. CN, copy number. **b**, **c**, Heatmaps of highly significant elastic net features associated with response to dasatinib (inhibitor of SRC and ABL) (**b**) and 17-AAG (HSP90 inhibitor) (**c**) for the 14 most sensitive (purple) and resistant (yellow) cell lines. For each cell line mutation features are at the top of the heatmap shown in black (present) or grey (absent), followed by expression features (blue corresponds to lower expression, red to higher expression). To the left of each feature is a bar indicating the absolute value of the effect size. Bars in purple are negative effects, indicating features associated with sensitivity, and bars in yellow are positive effects, indicating features associated with resistance. The natural log IC_{50} values are represented at the bottom. For clarity, only the top four features associated with sensitivity and resistance to 17-AAG are shown.

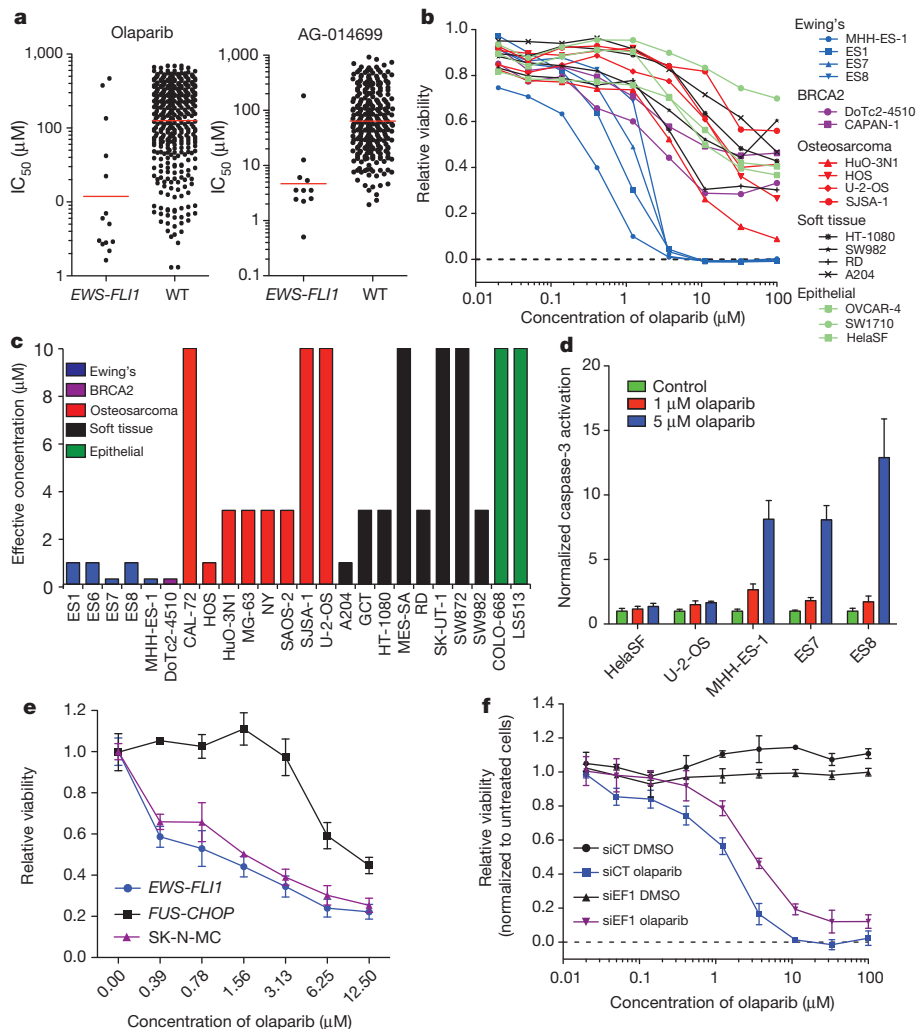


Figure 4 | Ewing's sarcoma cell lines are sensitive to PARP inhibition. **a**, The IC₅₀ values of wild-type and *EWS-FLI1*-translocation-positive cell lines to olaparib and AG-014699. **b**, Dose-response curves to olaparib after 6 days of constant drug exposure. Cell lines are classified according to tissue subtype. **c**, Colony formation assays were performed for 7–21 days over a range of olaparib concentrations (0.1, 0.32, 1, 3.2 or 10 μ M) and the concentration at which the number of colonies is reduced >90% for each cell line is indicated. **d**, Olaparib induced apoptosis in Ewing's sarcoma cell lines after 72 h treatment. **e**, Sensitivity to olaparib of *EWS-FLI1* and *FUS-CHOP* transformed mouse mesenchymal cells compared to the SK-N-MC cell line (which harbours the *EWS-FLI1* fusion). **f**, Sensitivity to olaparib of A673 cells transiently transfected with (siEF1) and without (siCT) *EWS-FLI1*-specific siRNA. All error bars are s.d. from triplicate measurements except for **b** where error bars have been removed for clarity.

To assess whether the sensitivity to PARP inhibitors is due to the presence of the *EWS-FLI1* rearrangement or intrinsic to the mesenchymal precursor cell type from which Ewing's sarcoma arises, we compared the sensitivity to olaparib of mouse mesenchymal cells transformed with either *EWS-FLI1* or the related liposarcoma-associated translocation *FUS-CHOP* (also known as *DDIT3*)^{27,28}. *EWS-FLI1*-transformed cells showed sensitivity comparable to human Ewing's sarcoma cells, whereas the *FUS-CHOP*-transformed cells were relatively resistant (Fig. 4e and Supplementary Fig. 18). Moreover, expression of *EWS-FLI1* in NIH3T3 cells conferred increased sensitivity to olaparib (Supplementary Fig. 19), whereas olaparib sensitivity was partially reverted by transiently depleting *EWS-FLI1* from Ewing's sarcoma cells (Fig. 4f and Supplementary Fig. 20). Higher *FLI1* expression was also correlated with sensitivity to olaparib even when considering only non-Ewing's sarcoma cell lines ($r = -0.32$ between IC₅₀ and *FLI1* expression, $n = 413$, $P = 1.68 \times 10^{-11}$), suggesting that the sensitivity to olaparib of Ewing's sarcoma lines might be related to *EWS-FLI1* transcriptional activity.

Mutations in *BRCA1* or *BRCA2* are not present in these Ewing's sarcoma lines (Supplementary Data 1), and we have observed no evidence to indicate that the DNA damage response is defective in Ewing's sarcoma cells (data not shown). However, for reasons that are currently unclear, the *EWS-FLI1* translocation was associated with sensitivity to cytotoxic drugs, including DNA damaging agents such as camptothecin ($P < 1 \times 10^{-5}$), cisplatin ($P < 1 \times 10^{-4}$) and mitomycin-C ($P < 0.001$) (Supplementary Figs 21 and 22). Together with the report of olaparib sensitivity in prostate cancer cell lines

bearing the translocation between *ERG*, which like *FLI1* is a member of the *ETS* gene family, and *TMPRSS2*²⁹, our data support the potential utility of *ETS* gene fusions as biomarkers of sensitivity to PARP inhibitors. Notably, however, unlike the effect reported in prostate cancer, we observe potent cell death with olaparib treatment alone in Ewing's sarcoma cells. These observations raise the possibility that PARP inhibitors would be useful in the treatment of Ewing's sarcoma, a tumour of children and young adults with a 15% five-year survival rate in patients with metastatic disease or relapse after chemotherapy³⁰.

Discussion

High-throughput cancer cell line screening for drug-sensitivity patterns provides a strategy to identify appropriate cancer subtypes and biomarkers that may guide the early-phase clinical trials of multiple novel compounds under development. The validity of this approach is supported by its effective identification of drug-genotype associations that have already been established clinically, and it sets the stage for clinical testing of novel therapeutic biomarkers, such as the association between the *EWS-FLI1* translocation in Ewing's sarcoma cells and sensitivity to PARP inhibitors. The Supplementary Data accompanying this report, as well as the ongoing public web resource from future screenings (Genomics of Drug Sensitivity in Cancer; <http://www.cancerRxgene.org>), will hopefully enhance the discovery and validation of additional predictive cancer biomarkers.

Although the large number of cell lines screened facilitates representation of rare cancer genotypes and mitigates against the effects of individual samples, the data set presented here is limited both by the number

of available genotypes, as well as the number of targets interrogated by currently available drugs. Despite the apparent utility of using tumour-derived cell lines grown in two-dimensional culture, it is likely that experimental models that better mimic the tumour environment would in some instances further improve our understanding of drug sensitivity and provide additional insights. Nonetheless, we can discern an initial landscape of drug sensitivity patterns across a broad set of different cancer types and genomic backgrounds.

The identification of 'outliers' that are exquisitely sensitive to a drug as a result of a specific genetic abnormality within a targeted pathway remains the most compelling vision for targeted cancer therapies. *BCR-ABL*-positive CML, *BRAF*-mutant melanoma and *EGFR*-mutant-positive lung cancer and drugs that target the protein products of these genes are now well-established associations. The observation of PARP-inhibitor sensitivity by *EWS-FLI1*-positive Ewing's sarcoma cell lines points to the likelihood of new potent gene-drug associations as novel chemical and genomic space are explored. Even in the absence of outlier effects, pharmacogenomic profiling reveals a wealth of biomarkers that may prove useful for patient stratification. Although further work is needed to assess their potential clinical utility, in some instances these biomarkers may help explain heterogeneity in clinical responsiveness even among preselected patient populations.

This work, as well as an accompanying report³¹, provides a systematic and extensive view of the genomics underlying the sensitivity of human cancer cell lines to the diverse array of cancer drugs currently in use or under development. The emergent picture is of a complex network of biological factors that affect sensitivity to the majority of cancer drugs. This underscores both the challenge of identifying preselected patient populations for targeted therapies, as well as the opportunity to improve existing therapies and find new therapeutic avenues by identifying more predictive biomarkers.

METHODS SUMMARY

Cells were treated with nine concentrations (twofold dilutions) of drug for 72 h before measuring cell number relative to controls. A MANOVA was used to examine how drug IC₅₀ and slope values associate with tissue type, the mutation status of 64 cancer genes (including gene amplifications and homozygous deletions), rearrangements and MSI. The elastic net regression used the same genomic data sets as the MANOVA and also incorporated additional copy number data from a total of 426 cancer genes, transcriptional profiles and tissue types to identify features associated with drug response as measured by cell line IC₅₀.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 July 2011; accepted 2 March 2012.

1. Druker, B. J. *et al.* Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N. Engl. J. Med.* **355**, 2408–2417 (2006).
2. Kwak, E. L. *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **363**, 1693–1703 (2010).
3. Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with *BRAF* V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
4. McDermott, U. & Settleman, J. Personalized cancer therapy with selective kinase inhibitors: an emerging paradigm in medical oncology. *J. Clin. Oncol.* **27**, 5650–5659 (2009).
5. Shoemaker, R. H. *et al.* Development of human tumor cell line panels for use in disease-oriented drug screening. *Prog. Clin. Biol. Res.* **276**, 265–286 (1988).
6. Weinstein, J. N. *et al.* An information-intensive approach to the molecular pharmacology of cancer. *Science* **275**, 343–349 (1997).
7. McDermott, U. *et al.* Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc. Natl Acad. Sci. USA* **104**, 19936–19941 (2007).
8. Suwaki, N. *et al.* A HIF-regulated VHL-PTP1B-Src signaling axis identifies a therapeutic target in renal cell carcinoma. *Sci. Transl. Med.* **3**, 85ra47 (2011).
9. Deng, L. *et al.* Rho-kinase inhibitor, fasudil, suppresses glioblastoma cell line progression *in vitro* and *in vivo*. *Cancer Biol. Ther.* **9**, 875–884 (2010).
10. Weber, D. M. *et al.* Lenalidomide plus dexamethasone for relapsed multiple myeloma in North America. *N. Engl. J. Med.* **357**, 2133–2142 (2007).
11. Konecny, G. E. *et al.* Activity of the dual kinase inhibitor lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res.* **66**, 1630–1639 (2006).

12. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
13. O'Farrell, A. M. *et al.* SU11248 is a novel FLT3 tyrosine kinase inhibitor with potent activity *in vitro* and *in vivo*. *Blood* **101**, 3597–3605 (2003).
14. Kunii, K. *et al.* *FGFR2*-amplified gastric cancer cell lines require *FGFR2* and *ErbB3* signaling for growth and survival. *Cancer Res.* **68**, 2340–2348 (2008).
15. Byron, S. A. *et al.* Inhibition of activated fibroblast growth factor receptor 2 in endometrial cancer cells induces cell death despite PTEN abrogation. *Cancer Res.* **68**, 6902–6907 (2008).
16. Vassilev, L. T. *et al.* *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
17. Konecny, G. E. *et al.* Expression of p16 and retinoblastoma determines response to CDK4/6 inhibition in ovarian cancer. *Clin. Cancer Res.* **17**, 1591–1602 (2011).
18. Boisvert-Adamo, K., Longmate, W., Abel, E. V. & Apolin, A. E. Mcl-1 is required for melanoma cell resistance to anoikis. *Mol. Cancer Res.* **7**, 549–556 (2009).
19. Brown, C. J., Lain, S., Verma, C. S., Fersht, A. R. & Lane, D. P. Awakening guardian angels: drugging the p53 pathway. *Nature Rev. Cancer* **9**, 862–873 (2009).
20. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* **67**, 301–320 (2005).
21. Hanafusa, H., Torii, S., Yasunaga, T. & Nishida, E. Sprouty1 and Sprouty2 provide a control mechanism for the Ras/MAPK signalling pathway. *Nature Cell Biol.* **4**, 850–858 (2002).
22. Patterson, K. I., Brummer, T., O'Brien, P. M. & Daly, R. J. Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochem. J.* **418**, 475–489 (2009).
23. Dry, J. R. *et al.* Transcriptional pathway signatures predict MEK addiction and response to selumetinib (AZD6244). *Cancer Res.* **70**, 2264–2273 (2010).
24. Kelland, L. R., Sharp, S. Y., Rogers, P. M., Myers, T. G. & Workman, P. DT-Diaphorase expression and tumor cell sensitivity to 17-allylamino, 17-demethoxygeldanamycin, an inhibitor of heat shock protein 90. *J. Natl Cancer Inst.* **91**, 1940–1949 (1999).
25. Fong, P. C. *et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* **361**, 123–134 (2009).
26. McCabe, N. *et al.* BRCA2-deficient CAPAN-1 cells are extremely sensitive to the inhibition of poly(ADP-ribose) polymerase: an issue of potency. *Cancer Biol. Ther.* **4**, 934–936 (2005).
27. Riggi, N. *et al.* Development of Ewing's sarcoma from primary bone marrow-derived mesenchymal progenitor cells. *Cancer Res.* **65**, 11459–11468 (2005).
28. Riggi, N. *et al.* Expression of the FUS-CHOP fusion protein in primary mesenchymal progenitor cells gives rise to a model of myxoid liposarcoma. *Cancer Res.* **66**, 7016–7023 (2006).
29. Brenner, J. C. *et al.* Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. *Cancer Cell* **19**, 664–678 (2011).
30. Balamuth, N. J. & Womer, R. B. Ewing's sarcoma. *Lancet Oncol.* **11**, 184–192 (2010).
31. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* <http://dx.doi.org/10.1038/nature11003> (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Lo Grasso of the Scripps Research Institute for providing the inhibitor JNK9L. This work was supported by a grant from the Wellcome Trust (086357; M.R.S., P.A.F., J.S., D.A.H.) and by grants from the National Institutes of Health (P41GM079575-02 to N.S.G. and 1U54HG006097-01 to N.S.G. and D.A.H.). S.R. is supported by a Physician-Scientist Early Career Award from the Howard Hughes Medical Institute. U.M. is supported by a Cancer Research UK Clinician Scientist Fellowship.

Author Contributions M.J.G., C.H.B., U.M. and S.V.S. supervised data collection. M.J.G., C.H.B., U.M. and S.R. supervised data analysis. C.D.G. and K.W.L. conceived and wrote the curve-fitting algorithm and performed the MANOVA; E.J.E. and S.R. performed elastic net analysis and analysed the data. P.G., I.R.T. and J.S. developed and managed screening databases with assistance from A.B. and W.Y.; S.J.H. performed most of the Ewing's sarcoma related studies with contributions from D.S., A.D., X.L., F.K. and L.C., with I.S. and O.D. providing critical reagents; R.J.M., A.T.T., J.A.S., S.B., S.R.L., K.L., A.M.-D., J.L.B., X.M., T.M., H.T., L.R., F.J. and P.O'B. performed cell line screening experiments. S.P. performed *MCL1* siRNA experiments. Q.L.W.Z., T.Z., W.H., X.D., H.G.C. and J.W.C. synthesized screening compounds, and N.S.G. provided guidance on their selection and use; F.I. and J.S.-R. performed compound activity clustering; G.R.B. and H.D. performed cell line genotyping and genetic analysis; J.A.E. and J.B. provided guidance regarding clinical relevance of the work; M.J.G. and C.H.B. wrote the manuscript with major contributions from S.R., S.J.H. and U.M.; M.R.S., D.A.H., J.Se. and P.A.F. conceived the study, analysed the data and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to U.M. (um1@sanger.ac.uk) or C.B. (cbenes@partners.org).

METHODS

Cell lines. All cell lines were sourced from commercial vendors. Cells were grown in RPMI or DMEM/F12 medium supplemented with 5% FBS and penicillin/streptavidin, and maintained at 37 °C in a humidified atmosphere at 5% CO₂. Cell lines were propagated in these two media to minimize the potential effect of varying the media on sensitivity to therapeutic compounds in our assay, and to facilitate high-throughput screening. To exclude cross-contaminated or synonymous lines, a panel of 92 SNPs was profiled for each cell line (Sequenom) and a pair-wise comparison score was calculated. In addition, we performed short tandem repeat (STR) analysis (AmpFISTR Identifier, Applied Biosystems) and matched this to an existing STR profile generated by the providing repository. More information on the cell lines screened, including their SNP and STR profiles is available on the Genomics of Drug Sensitivity in Cancer website (<http://www.cancerRxgene.org>).

Screening drugs. Compounds were from academic collaborators or commercial vendors. We have provided a full description for each compound including its name, source, PubChem and/or ChEMBL accessions, screening concentration as well as therapeutically relevant molecular target(s) (Supplementary Data 2). Compounds were stored as 10 mM aliquots at –80 °C, and were subjected to a maximum of five freeze–thaw cycles. The range of concentrations selected for each compound was based on *in vitro* data of concentrations inhibiting relevant kinase activity and cell viability.

Cell viability assays. Cells were seeded in either 96-well or 384-well microplates in medium supplemented with 5% FBS and penicillin/streptavidin. The optimal cell number for each cell line was determined to ensure that each was in growth phase at the end of the assay (~70% confluency). Adherent cell lines were plated 1 day before treatment with a 9-point twofold dilution series of each compound using liquid handling robotics, and assayed at a 72-h time point. Cells were fixed in 4% formaldehyde for 30 min and then stained with 1 μM of the fluorescent nucleic acid stain Syto60 (Invitrogen) for 1 h. Suspension cell lines were treated with compound immediately following plating, incubated for 72 h, and then stained with 55 μg ml^{–1} resazurin (Sigma) prepared in glutathione-free media for 4 h. Quantification of fluorescent signal intensity was performed using a fluorescent plate reader at excitation and emission wavelengths of 630/695 nm for Syto60, and 535/595 nm for resazurin. All screening plates were subjected to stringent quality control measures and a Z-factor score comparing negative and positive control wells was calculated across all screening plates (median = 0.70, upper quartile = 0.86, lower quartile = 0.47, *n* = 4,857 plates). Drug screening was performed at two sites using matched cell line collections (the site of drug screening and plate format for each IC₅₀ are described in Supplementary Data 2 and 10). As a control the drug camptothecin (drugs number 1003 and 195) was screened at both sites. The IC₅₀ values were highly correlated (*r*² = 0.3244, slope = 1.030, *n* = 252 cell lines) and these drugs were nearest neighbours in our cluster analysis of drugs (Supplementary Table 1). Measurements of cell viability during 6-day assays using threefold dilution series of olaparib were performed in 96-well plates using Cell Titer Blue (Promega) according to the manufacturer's instructions. A673 and NIH3T3 cells were also exposed to a threefold dilution series of olaparib and cell viability was measured after 72 h of drug exposure using Cell Titer 96 Aqueous One Solution Cell (Promega) according to the manufacturer's instructions. Measurements of cellular apoptosis were performed using Apo-ONE caspase assay (Promega) following manufacturer's instructions. The A673 Ewing's sarcoma cell line was transiently transfected with allstars non-targeting siRNA control (Qiagen; siCT) or an siRNA targeting the *EWS-FLI1* translocation³² (siEF1, 5'-GGCAGCAGAACCCUUCUACG) and treated with olaparib immediately after siRNA transfection. Transient knock-down of *MCL1* in HT-144 and A549 cells was performed using the following siRNA sequences: oligonucleotide 1, 5'-CTGGTTTGGCATATCTAATAA; oligonucleotide 2, 5'-CCCGCCGAATTCATTAATTTA; oligonucleotide 3, 5'-AAGGGTTAGGACCAACTACAA; oligonucleotide 4, 5'-CCCTAGCAACC TAGCCAGAAA) and controls were mock transfected.

Colony formation assays. Cells were plated at low density into 35-mm cell-culture plates and the following day treated with the indicated drug concentration or vehicle control (DMSO). The medium was changed and cells re-drugged every 3–4 days. When sufficient colonies were visible, typically after 7–21 days, cells were washed once in PBS before fixing in ice-cold methanol for 30 min while shaking. Methanol was aspirated and Giemsa stain added at a dilution of 1:20 overnight while shaking. The following day cells were rinsed in distilled water and air dried.

Genomic and transcriptional characterization of cancer cell lines. A total of 64 of the most frequently mutated cancer genes were sequenced to base-pair resolution across all coding exons for each gene by capillary sequencing in our panel of human cancer cell lines, which formed the basis for the cell lines chosen for this drug screen. The presence of seven of the most commonly rearranged cancer

genes (for example, *BCR-ABL*, *MLL-AFF1* and *EWS-FLI1*) was determined across the drug screen cell line panel by the design of breakpoint-specific sequence primers that enabled the detection of the rearrangement following capillary sequencing. Analysis of MSI was carried out according to the guidelines set down by The International Workshop on Microsatellite Instability and RER Phenotypes in Cancer Detection and Familial Predisposition³³. Samples were screened using the markers BAT25, BAT26, D5S346, D2S123 and D17S250 and were characterized as MSI if two or more markers showed instability. Total integral copy number values across the footprints of the cancer genes were determined from Affymetrix SNP6.0 microarray data using the PICNIC algorithm to predict copy number segments in each of the cell lines³⁴. For a gene to be classified as amplified, the entire coding sequence must be contained in one contiguous segment defined by PICNIC, and have a total copy number of eight or more. A copy number of eight was chosen to ensure that the majority of segments above this threshold consisted of focal amplifications more likely to have been selectively amplified. Deletions must occur within a single contiguous segment with copy number zero. For gene expression analysis, RNA was extracted from each cell line using a standard Trizol protocol and hybridized to the HT-HGU133A Affymetrix whole genome array. Normalized gene expression intensities were generated using the Robust Multi-Array Average (RMA) algorithm³⁵. The genomic data used for this analysis are provided in the Supplementary Data and transcriptional data are available from ArrayExpress (accession number E-MTAB-783). A complete description of the characterization of our cancer cell line collection is also available from the Cancer Genome Project webpages (<http://www.sanger.ac.uk/genetics/CGP/CellLines/>).

Curve fitting of drug sensitivity data. Dose–response curves were fitted to the fluorescence signal intensities. We required a method that would first allow us to model the heteroscedasticity in the luminescence data, and second allow us to incorporate prior knowledge of response, especially at drug concentrations at which the data are less informative. A bespoke Bayesian sigmoid model was thus implemented to facilitate this, yielding a full description of the uncertainties in the data, and allowing reasonable interpretation of predicted response at concentrations outside the tested range. Response curves were fitted to the fluorescence signal intensities using a Bayesian sigmoid model. Drug–response data consisted of 16 (96-well format) or 42 (384-well) drug-free positive controls, 8 (96-well) or 32 (384-well) negative (no cells) controls and drug–response points for nine half-fold concentrations. Technical replicate intensity responses were averaged. Generalized sigmoidal response curves are fitted as follows.

$$\text{Intensity } x_{lc} \text{ is assumed to have the mean value } x_{lc} = I_{\min} + \frac{I_{\max} - I_{\min}}{(1 + e^{\beta(lc - \alpha)})^f}.$$

Parameters I_{\max} and I_{\min} are the mean intensities of the positive and negative controls, α and β are scale and gradient responses, f is a shape parameter, lc denotes the log concentration. We assume that the intensity x_{lc} has variance $\text{Var}(x_{lc}) = BE(x_{lc})$, where B represents a noise parameter. We assume that x_{lc} has a gamma distribution:

$$P(x_{lc}) = \frac{B^{E(x_{lc})/B}}{\Gamma(E(x_{lc})/B)} x_{lc}^{E(x_{lc})/B - 1} e^{-Bx_{lc}}$$

Positive and negative controls x_{\max} and x_{\min} are also gamma distributed:

$$P(x_{\max}) = \frac{B^{I_{\max}/B}}{\Gamma(I_{\max}/B)} x_{\max}^{I_{\max}/B - 1} e^{-Bx_{\max}}$$

$$P(x_{\min}) = \frac{B^{I_{\min}/B}}{\Gamma(I_{\min}/B)} x_{\min}^{I_{\min}/B - 1} e^{-Bx_{\min}}$$

The concentration giving percentage response p is given by:

$$IC_p = \alpha + \frac{1}{\beta} \log\left(\left(\frac{100}{100-p}\right)^{\frac{1}{f}} - 1\right)$$

We use Markov Chain Monte Carlo simulations to obtain mean posterior parameter estimates. Parameters were initialized from maximum likelihood estimates and 100,000 iterations obtained. The final 10,000 values were used for subsequent inference. The IC₅₀ has a normal prior with 95% probability mass covering a range from 1,000-fold below the minimum concentration tested to 1,000-fold higher than the maximum tested concentration. We assume uninformative priors on the remaining parameters. Response curves are plotted using mean posterior values of IC_p for p ranging between 0% and 100%. Confidence intervals for IC_p are obtained from the associated posterior. Results were manually curated as part of quality control.

MANOVA. A fixed effects MANOVA was used to correlate response with genomics. An $n \times 2$ dose–response matrix consisting of IC₅₀ and slope parameter β for n cell lines was constructed for each drug. A linear (no interaction terms)

model explained these observables with factors including tissue type, the mutation status of cancer genes, chromosomal re-arrangements, and MSI status. Size effects and significances were obtained. A gene was defined as mutated if it fulfilled any of these criteria: a coding sequence variant in the cancer gene, a total copy number of 0 (homozygous deletion) or more than 7 (amplification). Only those genes with >1 mutated cell lines in the panel were included for analysis (65 cancer genes and 3 chromosomal re-arrangements in total). The effect measures the relative difference in the mean IC₅₀ from the wild-type to mutant group (for example, an effect of 0.1 or 10 indicates a ~10-fold decrease or increase in drug concentration, respectively). A Benjamini–Hochberg multiple testing correction threshold with false discovery rate of 20% was used to identify a candidate list of significant associations for the purposes of this paper.

Elastic net analysis. We take the previously described approach of the elastic net²⁰, a multivariate variable selection technique with a penalization approach. Genomic data including mutation status of 64 cancer genes, rearrangements, continuous copy number data from 426 genes causally implicated in cancer (<http://www.sanger.ac.uk/genetics/CGP/Census/>), and genome-wide transcriptional profiles, as well as tissue type, were used as input variables (Supplementary Data 1 and 11). The elastic net was used to select which of these features were associated with drug response as measured by IC₅₀ across the cell line panel.

Let X be a $n \times p$ matrix of input features (where p is the number of features and n is the number of cell lines) and y be a vector of drug sensitivities of length n . For any non-negative λ_1 and λ_2

$$L(\lambda_1, \lambda_2, \beta) = |y - X\beta|^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \quad (1)$$

Let $\hat{\beta}$ be the naive elastic net estimator. Then $\hat{\beta} = \arg \min_{\beta} \{L(\lambda_1, \lambda_2, \beta)\}$. A scaling factor of $(1 + \lambda_2)$ is added to the naive elastic net to prevent double shrinking.

$$\hat{\beta}(\text{elasticnet}) = (1 + \lambda_2) \hat{\beta}(\text{naiveelasticnet}) \quad (2)$$

To determine the optimal λ_1 and λ_2 , we let $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$. Then $(1 - \alpha) \sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j|$ is the elastic net penalty. Tenfold cross validation is performed to optimize λ_1 and λ_2 in equation (2), denoted as $\hat{\lambda}_1$ and $\hat{\lambda}_2$, respectively. To find the variables that are associated with drug response, $\hat{\lambda}_1$, $\hat{\lambda}_2$, X and y are inputted into equation (2) to solve for vector β . The variables with non-zero β values are determined to be features associated with drug response.

This procedure was repeated 100 times for each drug to assess the stability of the features when applying the tenfold cross validation procedure. For each of the 100 runs, a feature list was built for the drug comprised of genes, transcripts, and tissues with weights assigned to each. The final signature of markers for a drug consisted of all features that appear in any of the 100 runs along with the statistics on the frequency that the feature appears and the average weight given to that feature over the 100 runs.

The weights (β) were used to assess effect sizes of features in a drug's marker signature. The effect size of a feature was calculated by multiplying the feature's weight by its standard deviation across the cell line panel. The effect size is therefore a normalization of the feature's weight to account for the different scales used to measure the different genomic features. Features with higher stability of correlation in cross-validation (f) were considered of the highest confidence of truly being associated with drug response. The most significant features associated with drug response are those with both large frequency and effect size. Highly significant associations are defined as those with an effect size that was ± 2 s.d. from the mean, $-2.95 > e > 2.79$ and a frequency (f) of 2 s.d. above the mean, $f > 0.76$.

Drug clustering based on IC₅₀ values. The data set was filtered to remove cell lines for which less than 50% of drugs in the panel were tested, resulting in 400 cell lines remaining. Missing data points were inputted by using a nearest-neighbour approach based on Pearson correlation (PC) scores between IC₅₀ values. We also applied a step to reduce effects inherently present in the data due to the integration of data that were obtained from two different screening sites and which may be amplified by the nearest-neighbour missing point imputation. To do this, we performed an ANOVA for each cell line to assess the difference in the average IC₅₀ values for drugs at the two screening sites. Cell lines with a P value < 0.01 according to this test were filtered out. Of the original data set, 265 cell lines (41.47%) were used for cluster analysis.

As done for missing point imputation, we measured the extent of similarity among drug sensitivity/resistance profiles with PC scores pair-wisely by considering for each drug its pattern of IC₅₀ values across the cell lines (log values were considered). Then we used these similarity scores as input to the affinity-propagation clustering (APC) algorithm in a recursive manner³⁶. The APC algorithm is based on a 'passing-message-between-datapoints' strategy, it requires as input a pair-wise similarity matrix and gives as the output a set of disjointed clusters. It also indicates, for each of the computed clusters, a datapoint called the 'cluster exemplar' (that is, the 'cluster centroid'): the datapoint that best interpolates all the other datapoints in its cluster. The APC algorithm requirement consists in a square matrix containing similarity scores between each pair of datapoints and a set of probabilities, one for each datapoint to be elected as the exemplar of its clusters. Implicitly, these probabilities determine the number of clusters to be computed. If they are uniformly distributed then the APC algorithm from the data automatically determines the number of clusters. In the first step of our clustering strategy, we applied the APC algorithm to the whole set of drugs in the screening by assuming the probability of each drug to be elected as a cluster exemplar uniformly distributed (so the number of clusters was automatically determined by the APC algorithm). Particularly, we set the input parameter from which this probability is computed equal to the median value of all the pair-wise similarities. The algorithm provided 22 clusters. In each of these clusters, a drug was indicated as the cluster exemplar. The intra-cluster similarity score (odds ratio) of a community C containing n drugs is computed as the average similarity (that is, correlation between IC₅₀ patterns) between all the possible couple of drugs belonging to C (total average correlation) divided by the expected average similarity between all the possible couple of drugs in a randomly selected set of n drugs. We derived a network representation of the clustered data by considering each drug as a network node and by adding a link between a couple of nodes corresponding to drugs in the same cluster. This generated the 'network communities' indicated with different colours in Supplementary Fig. 3. We then, recursively, clustered again the cluster exemplars with the APC algorithm, to obtain 'communities of communities' (or 'rich clubs') and added links to the network as explained above. This procedure was recursively re-applied over cluster exemplars until convergence (no datapoints were clustered together) ending up into a hierarchical network depicted in Supplementary Fig. 3 and as described in ref. 37. The network visualization in Supplementary Fig. 3 has been obtained by using Cytoscape³⁸.

The inter-cluster similarity score for each couple of communities A and B was computed as follows:

$$\rho(A, B) = \frac{\sum_{X \in A} \sum_{Y \in B} \rho_{X,Y}}{|A||B|}$$

Where $\rho_{X,Y}$ is the Pearson correlation coefficient between the patterns of IC₅₀ values of drug X and drug Y .

Empirical P values were computed for each of these scores by permutation test: for a number $n = 10,000$ of trials and each $\rho(A, B)$, we sampled two random sets of drugs A^* and B^* (containing $|A|$ drugs and $|B|$ drugs, respectively) from the whole set of 131 drugs and we computed $\rho(A^*, B^*)$. Then we computed the empirical P value for a given $\rho(A, B) \geq 0$ (respectively ≤ 0) as the number of times that $\rho(A^*, B^*)$ was greater (resp. less) than or equal to $\rho(A, B)$ across the n permutation trials, divided by n .

32. Prieur, A., Tirode, F., Cohen, P. & Delattre, O. EWS/FLI-1 silencing and gene profiling of Ewing cells reveal downstream oncogenic pathways and a crucial role for repression of insulin-like growth factor binding protein 3. *Mol. Cell. Biol.* **24**, 7275–7283 (2004).
33. Boland, C. R. et al. A National Cancer Institute Workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248–5257 (1998).
34. Greenman, C. D. et al. PICNIC: an algorithm to predict absolute allelic copy number variation with microarray cancer data. *Biostatistics* **11**, 164–175 (2010).
35. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
36. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
37. Iorio, F. et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl Acad. Sci. USA* **107**, 14621–14626 (2010).
38. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).