SCIENTIFIC REPORTS

OPEN

Prediction of response to anti-cancer drugs becomes robust via network integration of molecular data

Marcela Franco[1], Ashwini Jeggari[2], Sylvain Peuget[1], Franziska Böttger[1,3], Galina Selivanova[1] & Andrey Alexeyenko [1,4]

Despite the widening range of high-throughput platforms and exponential growth of generated data volume, the validation of biomarkers discovered from large-scale data remains a challenging field. In order to tackle cancer heterogeneity and comply with the data dimensionality, a number of network and pathway approaches were invented but rarely systematically applied to this task. We propose a new method, called NEAmarker, for finding sensitive and robust biomarkers at the pathway level. scores from network enrichment analysis transform the original space of altered genes into a lower-dimensional space of pathways. These dimensions are then correlated with phenotype variables. The method was first tested using *in vitro* data from three anti-cancer drug screens and then on clinical data of The Cancer Genome Atlas. It proved superior to the single-gene and alternative enrichment analyses in terms of (1) universal applicability to different data types with a possibility of cross-platform integration, (2) consistency of the discovered correlates between independent drug screens, and (3) ability to explain differential survival of treated patients. Our new screen of anti-cancer compounds validated the performance of multivariate models of drug sensitivity. The previously proposed methods of enrichment analysis could achieve comparable levels of performance in certain tests. However, only our method could discover predictors of both *in vitro* response and patient survival given administration of the same drug.

The problem known as the "dimensionality curse"[1,2] - when a set of few (tens to hundreds) biomedical samples are described with a much larger number of molecular variables - undermines robustness of phenotype predictors. This was aggravated further when novel omics platforms expanded the variable space from thousands to nearly millions of potentially informative molecular features. In addition, profiling of cancer samples revealed that genomic alterations across tumors of the same type appear disparate and poorly overlapping[2]. As a result, variability between cancer samples is often higher than is assumed by the common parametric statistics[3]. Beyond a few success cases[4,5] molecular cancer signatures have been hard to corroborate in a novel, independent cohort. Across a number of meta-analyses, conclusions about practical applicability of the signatures range from entirely negative[6,7] to mixed or moderately positive[8]. The common understanding is that seemingly disparate individual events must be confluent to certain pathways that represent cancer hallmarks and pathways[9].

Modeling drug response *in vitro* was questioned by finding that molecular landscapes of cancer cell lines are be very different from those of original tumors[10]. A later, more comprehensive exploratory analysis demonstrated overall consistence of molecular aberrations between cell lines and primary tumors from matching cancer sites[9] – although these authors did not investigate the therapeutic relevance of discovered *in vitro* correlates. Haibe-Kains and co-authors published a discouraging comparison[11] between two large *in vitro* screens[12,13]. After that conclusion and the following polemics[14] and failures to replicate particular findings[15], the urgent need in cross-platform

[1]Department of Microbiology, Tumor and Cell Biology (MTC), Karolinska Institutet, Stockholm, Sweden. [2]Department of Cell and Molecular Biology, Karolinska Institutet, 171 77, Stockholm, Sweden. [3]Present address: OncoProteomics Laboratory, Department of Medical Oncology, VU University Medical Center, 1081HV, Amsterdam, The Netherlands. [4]National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Box 1031, 17121, Solna, Sweden. Correspondence and requests for materials should be addressed to A.A. (email: andrej.alekseenko@scilifelab.se)
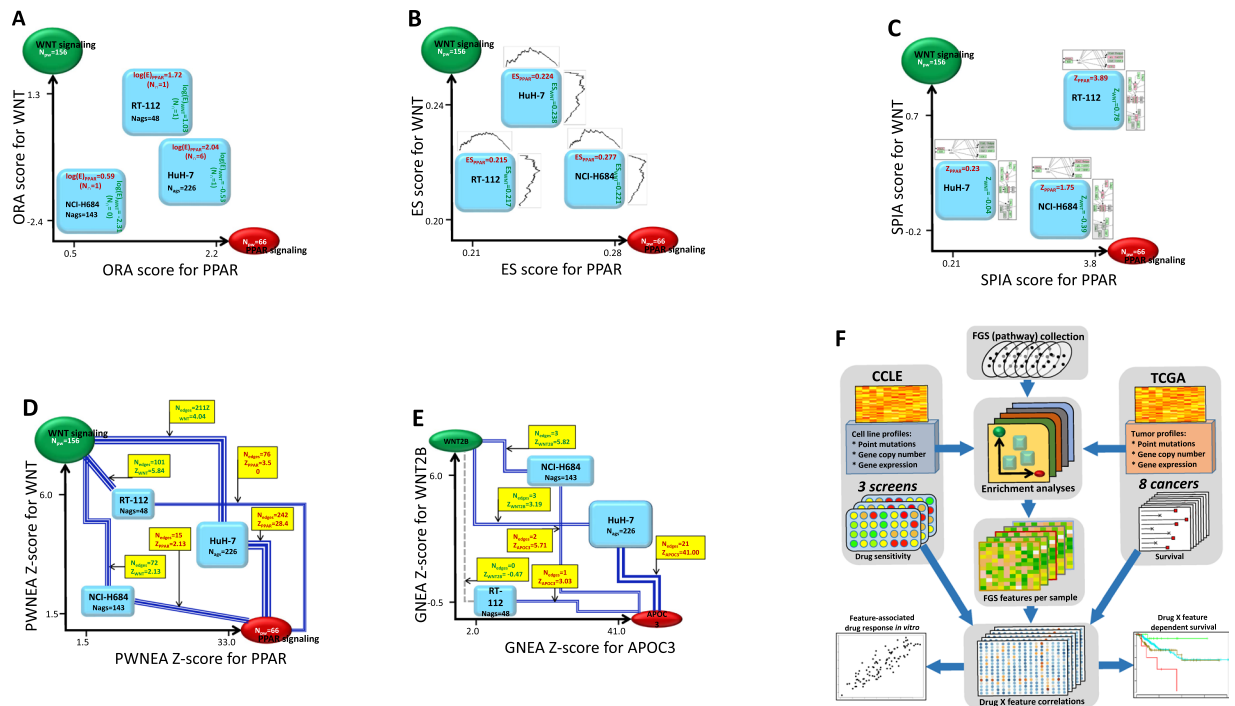
1

Figure 1. Rendering biological samples into pathway space with alternative enrichment methods. The placement of three cancer cell lines HuH-7, NCI-H684, and RT-112 in a 2-dimensional space of pathways 'PPAR signaling' and 'WNT signaling' (KEGG#03320 and KEGG#04310) (**A–D**) or, alternatively, in a space of two key genes from these pathways (**E**) was done by using cell line-specific altered gene sets, AGS, which originated from transcriptomics data and contained 226, 143, and 48 member genes, respectively (AGS of class `significant.affymetrix_ccle`). (**A**) ORA: enrichment of the three AGSs was analyzed against the two pathways (or, more generally, functional gene sets, FGS) using the overrepresentation analysis. The pathway enrichment scores were calculated from overlap between the gene sets. For clarity we here denote the pathway size N_{PW} which corresponds to N_{FGS} elsewhere in the article. Due to the relatively small gene sets sizes (N_{PW} and N_{AGS}), a noticeable ($N_{\cap} > 1$) and significant overlap was observed in just one out of six cases, which could limit the ORA sensitivity. (**B**) GSEA was calculated using the full ranked gene lists from each cell line sample[25]. (**C**) SPIA accounted for topological relationships of altered genes within the pathways. More weight was assigned to patterns of consistent up/down-regulation, i.e. where deregulated genes adjoined in regulatory cascades. Relatively disjoint regulatory events contributed with lower weights. The gene set submitted to SPIA can be of arbitrary size, up to full length, as in GSEA. The fold change values determine relative influence of the pathway genes. (**D**) NEA: the coordinates of the three AGSs in the space of two pathways were determined via network enrichment analysis. The NEA z-scores (on the axes and in yellow boxes) were calculated via network connectivity rates between corresponding AGS and FGS by taking into account the numbers of AGS-FGS links (N_{edges} in yellow boxes) and the node topology of the member genes (Fig. 1 and Methods). The summarized connections between AGSs and FGSs are shown by blue compound edges that represent multiple individual gene-gene edges in the global network ($N_{edges} \sim$ line width). Individual edges within AGSs and within FGSs are not used in the analysis. (**E**) GNEA: since the power of NEA to detect network enrichment was high, it was possible to apply NEA to the cell line AGSs versus individual gene network nodes WNT2B and APOC3 in the same way as it was done versus pathways in D. Even though the N_{edges} values were expectedly smaller than in B, four out of six Z-scores appeared rather high. (**F**) Analysis workflow: alternative enrichment analyses received as input CCLE (cell lines *in vitro*) and TCGA (clinical) omics datasets and created AGSxFGS matrices for the cancer cell lines and primary tumor samples. The *in vitro* screens and clinical follow-up observations provided information on anti-cancer drug sensitivity, which together with the AGSxFGS matrices was used for evaluation of drugXfeature correlations. The strength, consistence, and robustness of these correlations was used for the comparative assessment of the enrichment methods as presented in Table 2.

and clinical validation became even more apparent. It is dictated by both statistical and biological challenges, such as excessive data dimensionality, imperfect analytical tools, the heterogeneity of cancer genomes, and the downstream diversity of methylation and expression patterns[16]. Authors of one of the most up-to-date investigations still admitted that the ability of cancer cell drug screens "to inform development of new patient-matched therapies… remains to be proven"[17]. On the clinical side, oncologists expected reports on patient-specific alterations in the light of knowledge available from computerized support systems[18]. In our view, these challenges could be most systematically addressed by summarizing sparse, disparate events at the pathway level via the global interaction network.
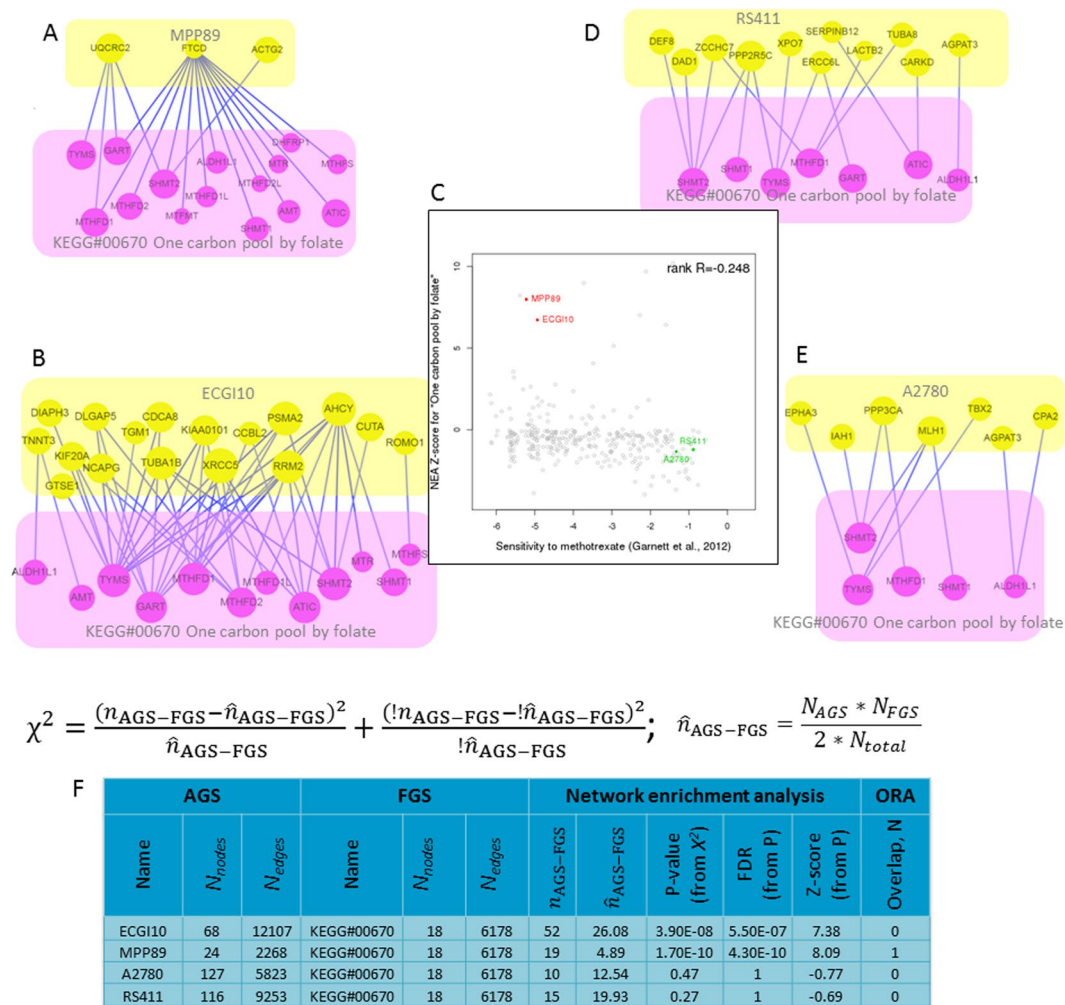
$$\chi^2 = \frac{(n_{AGS-FGS} - \hat{n}_{AGS-FGS})^2}{\hat{n}_{AGS-FGS}} + \frac{(!n_{AGS-FGS} - !\hat{n}_{AGS-FGS})^2}{!\hat{n}_{AGS-FGS}}; \quad \hat{n}_{AGS-FGS} = \frac{N_{AGS} * N_{FGS}}{2 * N_{total}}$$

	AGS			FGS			Network enrichment analysis					ORA
Name	N_{nodes}	N_{edges}	Name	N_{nodes}	N_{edges}	$n_{AGS-FGS}$	$\hat{n}_{AGS-FGS}$	P-value (from X^2)	FDR (from P)	Z-score (from P)		Overlap, N
ECGI10	68	12107	KEGG#00670	18	6178	52	26.08	3.90E-08	5.50E-07	7.38		0
MPP89	24	2268	KEGG#00670	18	6178	19	4.89	1.70E-10	4.30E-10	8.09		1
A2780	127	5823	KEGG#00670	18	6178	10	12.54	0.47	1	-0.77		0
RS411	116	9253	KEGG#00670	18	6178	15	19.93	0.27	1	-0.69		0

Figure 2. Network enrichment analysis of four cell line AGSs with differential response to methotrexate. While using AGSs of class significant.affymetrix_ccle, the response of cancer cell lines to methotrexate in CGP screen correlated with NEA scores (pane F) in regard to FGS "One carbon pool by folate" (pane C). The methotrexate-resistant cell lines MPP89 and ECGI10 (panes A and B) received higher NEA scores since the numbers of edges $n_{AGS-FGS}$ connecting them to the FGS significantly exceeded those expected by chance, $\hat{n}_{AGS-FGS}$ (52 vs 26.02 and 19 vs. 4.89, respectively; pane F). For comparison, the sensitive lines RS411 and A2780 (panes D and E) had fewer edges than expected (15 vs 19.93 and 10 vs. 12.54, respectively) and therefore received lower, negative scores. The table in F and the sub-networks in A, B, D, and E were created via the web-site for interactive NEA https://www.evinet.org/[76].

Adding omics data to clinical variables has demonstrated the potential for prediction of cancer disease outcome in a DREAM challenge[19]. One particularly winning strategy was to employ multigenic expression patterns. Such 'meta-genes'[20] were, despite the seemingly 'network-free' definition, nothing other than modules in a co-expression network, which allowed dimensionality reduction and a biological generalization. Another DREAM project revealed efficiency of summarizing gene expression in cancer cell lines over pathways[21].

Further, identifying patient sub-categories responsive to a treatment is more challenging than one-dimensional drug sensitivity or survival analyses. A practical method should profile individuals across the cohort, so that the profiles can be fit to clinical variables and covariates. Therefore, a crucial feature for biomarker discovery would be the ability to assign scores to individual samples rather than to derive feature-pathway associations from the whole data collection. In addition, further sample classification in a flow of new patients should not require re-running the analysis on the whole cohort, i.e. recalculating the data space, as is often the case.

In this work, we use acronym NEA to refer to a specific approach for network enrichment analysis, which ascends to the idea of accounting for the node degrees of individual genes[22]. Using that approach of significance estimation via comparing network connectivity to a null model, NEA[23,24] can characterize experimental and clinical samples with pathway scores by accounting for sample-specific gene set relationships in the global gene interaction network. The pathway-level output could be used in downstream analyses against arbitrary phenotype models. The ability to summarize rare alterations that cause the recurrent cancer phenotypes into pathway profiles provides higher statistical power, more information on the underlying biology, and robustness in phenotype

Method	Type of input data	Allows data type integration	Level of input (samples)	Network analysis	Level of output (features)
original data	Any	No	All genes	–	–
ORA	Any	Yes	Altered gene sets	–	Functional gene sets
ssGSEA	Expression	No	All genes	–	Functional gene sets
ZGSEA	Expression	No	All genes	–	Functional gene sets
SPIA	Expression	No	All genes	+	Functional gene sets
EGSEA.min	Expression	No	All genes	–	Functional gene sets
EGSEA.mean	Expression	No	All genes	–	Functional gene sets
PWNEA	Any	Yes	Altered gene sets	+	Functional gene sets
GNEA	Any	Yes	Altered gene sets	+	Network gene nodes

Table 1. Characteristic features of alternative input data types.

Step	What was evaluated	Measure	Scheme	Figure
1	Statistical power to detect correlation between omics-based features and sensitivity to anti-cancer drugs	Fraction of significant correlates per feature class	(a) Within 3 published *in vitro* screens; (b) within 8 TCGA clinical datasets	3
2	Validation across drug screens	Consistency of univariate correlates	Between 3 published *in vitro* screens	4
3	Sustainability of multivariate models between drug screens	Robustness under independent validation	From CTD *in vitro* screen to the novel ACT screen	5
4	Agreement between *in vitro* screens and clinical data	Applicability of same correlates in the both contexts	From 3 published *in vitro* screens to TCGA clinical datasets	6

Table 2. Steps of analysis using alternative methods from Table 1.

prediction. However, neither NEA nor alternative methods of pathway enrichment had been systematically applied to the task presented above: the discovery of biomarkers suitable for individual outcome prediction.

In the first section of Results, we provide a detailed explanation of method NEAmarker and an instructive example of its work, both in comparison with alternative methods. A representative set of such methods was selected by investigating a wide range of earlier proposed algorithms and approaches. In Methods (section "Alternative Methods of Pathway and/or Enrichment Analysis"), we discuss their principles, consider both applicability to biomarker discovery and software usability, and motivate our choice of methods presented in Fig. 1 and Table 1. Thereby performance of our method NEAmarker is measured in parallel with using original gene profiles and alternative enrichment methods: overrepresentation analysis (ORA), gene set enrichment analysis ssGSEA[25,26], signaling pathway impact analysis (SPIA[27]) and EGSEA[28] that integrates multiple enrichment analysis methods. The outline (Fig. 1F) and details of the comparative evaluation are reported in Results. More specifically, we run the alternative methods in order to: (1) assess content of relevant information in three published experimental *in vitro* drug screens[12,13,29] (dubbed CCLE, CGP, and CTD, respectively) as well as in eight TCGA clinical cancer cohorts, (2) investigate preservation of this content across drug screens, (3) perform a novel, small scale drug screen and demonstrate that the pathway-level multivariate models withstand the independent validation, and finally 4) validate the identified correlations in clinical treatment profiles from TCGA[30] (Table 2).

Results

Background. The main principle of NEA (Fig. 1D) can be understood via comparison to the simplest method for detecting enrichment called overrepresentation analysis (ORA)[31] (Fig. 1A). First, An experimental or clinical sample should be characterized by a set of altered genes (AGS), such as top ranking differentially expressed genes, or a set of somatic mutations, or a combination of these. The second component of the analysis is a collection of functional gene sets (FGS): pathways, ontology terms, or custom sets of biological importance. Enrichment scores of the AGSs can thus be used as the samples' coordinates in a lower-dimensional FGS space. In both ORA and NEA significance can be evaluated with appropriate statistical tests. In ORA, enrichment is measured by the number of genes shared between the FGS and AGS, normalized by the gene set sizes. Since NEA considers a third, network component - via counting *network edges that connect* any genes of AGS with any genes of FGS - it includes normalization by topological properties of the network nodes. Due to the presence of different interaction mechanisms in the global network, NEA does not expect FGS genes to be altered themselves and therefore is capable of detecting enrichment of e.g. transcriptomics-based AGS in a pathway that operates by other mechanisms, such as trans-membrane signaling, phosphorylation etc. NEA holds other key advantages, such as exceptionally high power to detect enrichment in a global network, given the latter is sufficiently dense. Hence, even smaller gene sets often connect to each other by multiple edges so that even an individual key network node can even appear as an ultimately reduced FGS. This gene-level network analysis, GNEA (Fig. 1E) provides a more focused alternative to the default analysis at the pathway level, PWNEA (Fig. 1D) and we therefore separately evaluated performance of PWNEA and GNEA in the present work.
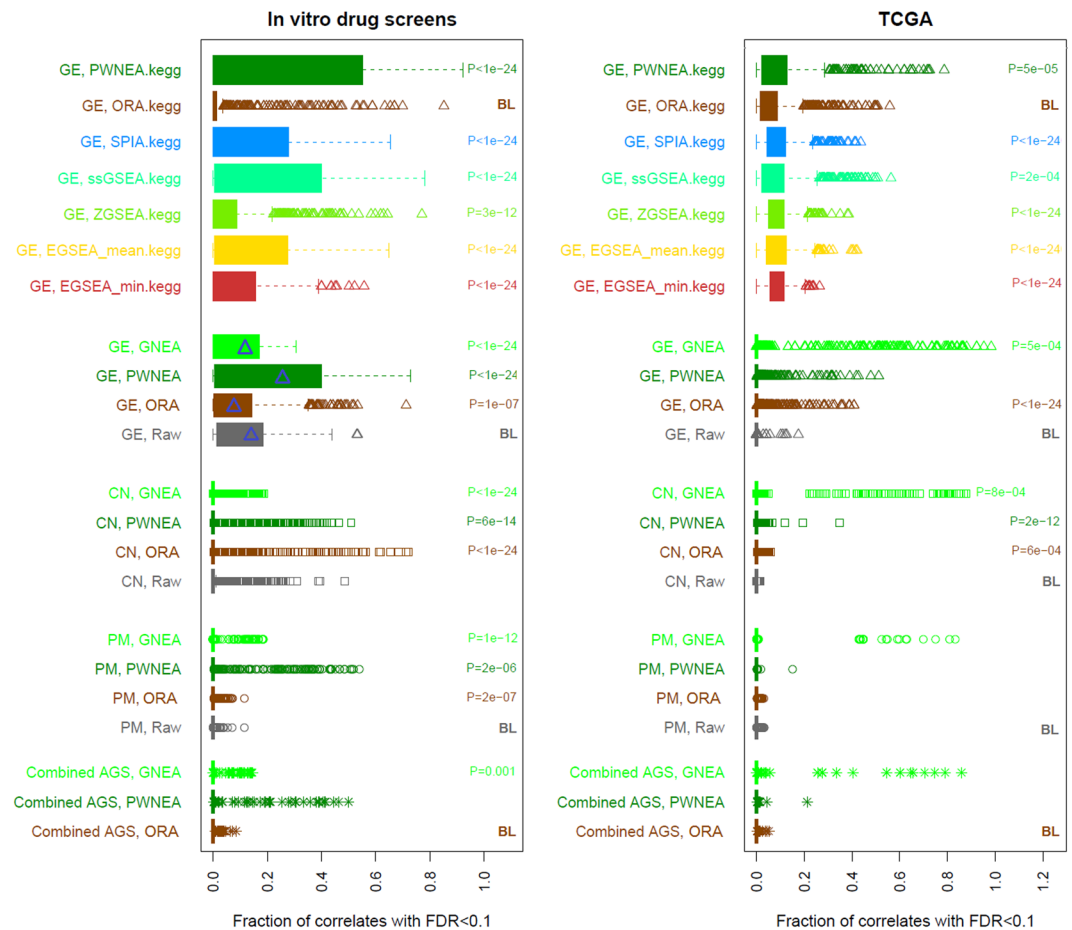
Figure 3. Comparison of the potential performance of different features, methods, and data types. The top seven boxplot rows (labeled ".kegg") present results obtained using the limited set of 197 KEGG pathways using gene expression data (label "GE"). Next, since ORA, PWNEA, and GNEA could accept any data type, the other boxplots present tests on the full set of 328 FGS (the respective PWNEA and ORA values for GE might differ, since the KGML gene sets were different from the core KEGG version). Each boxplot element combines correlation values of all features available for a given data type (point mutations, PM; copy number alterations, CN; and gene expression, GE) processed with each method. This was either done for the *in vitro* drugs response in the three screens (left pane; in total 365 tests of 320 distinct drugs) or for the survival of patients who had been administered the drugs (right pane; 42 drugs present both in a screen and in at least one of the eight TCGA cohorts). As an example, we calculated Spearman rank correlations between sensitivity of cell lines to drug RITA and transcriptomics features of these cell lines: either original Affymetrix (CCLE) gene expression profiles (18900 genes) or enrichment profiles of cell line specific AGSs of class top.400.affymetrix_ccle produced by GSEA (328 FGS features), pathway-level NEA (PWNEA; the same 328 features), and gene-level NEA (GNEA, in which 19027 nodes in the global network were treated as single-gene FGSs). The p-values of Spearman correlations between the features and drug sensitivity were then adjusted for multiple testing. The fractions of adjusted p-values below 0.1 became X-coordinates for the plot. The four examples (indicated by the blue triangle markers) gave fractions $1837/18900 = 0.097$; $23/328 = 0.070$; $78/328 = 0.236$; and $2090/19027 = 0.110$. The most detailed data can be found in Supplementary_table 5.FDR_rates_behind_Fig. 3.xlsx. The p-value next to boxplots are shown for categories producing significantly ($p < 0.001$ by Kolmogorov-Smirnov test) more fractions with FDR < 0.1 than the respective baseline category (labeled "BL"). The boxplot elements are defined in Methods.

Testing the multiple alternative methods implied different input, processing and output (Fig. 1, Table 1). Accordingly, our data analysis procedure included the method-specific steps for sample/patient characterization, enrichment analysis, and phenotype modeling. In order to maximally adapt GSEA to our applications, we tested two different ways of ranking gene lists, ssGSEA and ZGSEA (*Methods*) as well as two options of EGSEA, a method that combined multiple existing algorithms. In sections 3…5 of Results, we report the outcome of systematic analyses of the experimental datasets under these alternatives (Fig. 1F and Table 2).

We begin by introducing an example of data analysis and interpretation (Fig. 2). Using gene expression data, we observed a negative correlation between the PWNEA scores for pathway KEGG#00670 "One carbon pool by folate" for cancer cell line AGS features significant.affymetrix_ccle, on the one hand, and sensitivity to methotrexate on the other hand (Spearman rank $R = -0.248$; $p(H0) = 2.37e{-}06$). For the example, we focus on cell
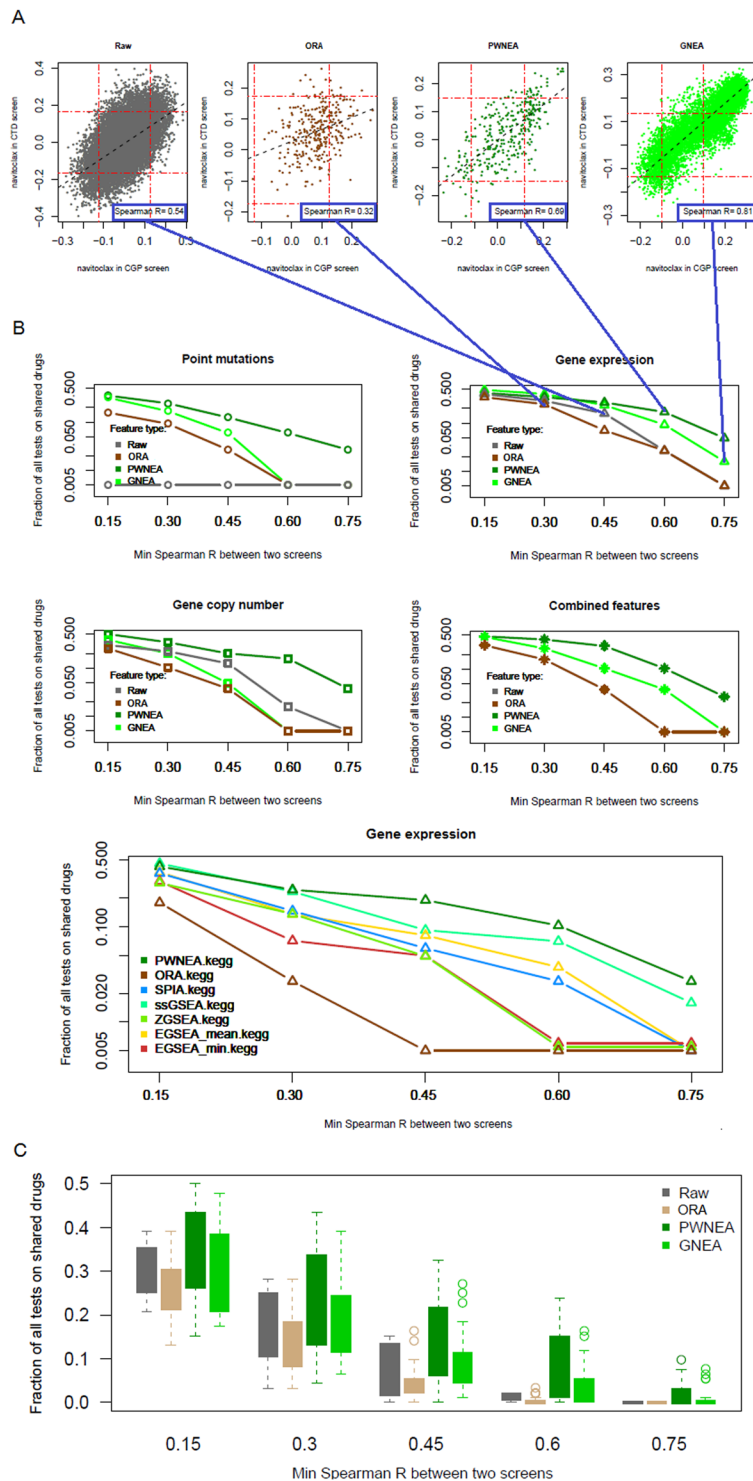
Figure 4. Consistency of drug-feature associations between drug screens. For each drug shared by any two *in vitro* drug screens (47 drugs in total), we calculated a rank correlation between drug-feature correlation coefficients in the two screens. (**A**) Agreement of drug-feature rank correlation coefficients between CGP and CTD screens of sensitivity to navitoclax using Affymetrix data as original gene expression values Affymetrix_CCLE (left pane) and AGS features of class significant.affymetrix_ccle profiled with ORA, PWNEA, and GNEA (other panes). The agreement in this case was worst while using ORA profiles (rank R = 0.32), whereas GNEA profiles performed best (rank R = 0.81). The red lines indicate FDR levels of 0.1 (false discovery rate by Benjamini-Hochberg method). The grey diagonal line is the linear regression fit. (**B**) Fractions of cases with rank correlation value above each of the five specified thresholds on example AGS classes. The features are grouped by type of profiling and by data type identically to Fig. 3. Four example values from A are mapped to the gene expression plot in B. In order to characterize sensitivity to each of the 47 drugs we used here, in parallel with the original gene profiles, AGS features of one class of each type: significant.filtered.exome.mini (PM),

significant.filtered.snp6.mini (CN), and significant.affymetrix_ccle (GE), and significant.filtered.combined. maxi ('combined'). The advantage of GNEA (with the exception of "Point mutations" and "Gene copy number") and PWNEA became apparent at the highest cutoffs R > 0.60 and R > 0.75. The lower pane presents testing the seven methods applicable to expression data only. This was done by using the same features for PWNEA and ORA as above and whole Affymetrix-CCLE matrices for the other methods. (**C**) Similarly to B, fractions of correlation values above each of the five specified thresholds were calculated for all feature classes and combined for all data types. For certain AGS feature classes, PWNEA and GNEA produced correlates highly conserved across screens (R > 0.6) in as many as 5–10% of cases.

lines which combined lowest sensitivity to methotrexate with highest PWNEA scores for KEGG#00670 (dubbed here Drug−/PW+) versus those possessing highest sensitivity and lowest PWNEA scores (Drug+/PW−) (ten cell lines in each set). Figure 2(A,B,D,E) displays the network connectivity of the FGS KEGG#00670 "One carbon pool by folate" with AGSs for two cell lines (MPP89, ECGI10) of group Drug−/PW+ and two cell lines (RS411, A2780) of group Drug+/PW−. The high score for MPP89 (Fig. 2A) was likely influenced by the network node of formimidoyltransferase cyclodeaminase FTCD, which provided 14 out of the 19 edges. FTCD was a member of four out of the ten AGS of the group Drug−/PW+. Methotrexate is a cytostatic drug that inhibits dihydrofolate reductase, thereby blocking synthesis of tetrahydrofolate, the downstream production of folic acid, and finally that of thymidine. Therefore an overexpressed FTCD, as an enzyme controlling the interconversion between formimidoyltetrahydrofolate and tetrahydrofolate[32], might rescue the thymidine production by supplying extra tetrahydrofolate[33].

For comparison, AGS of the other resistant cell line ECGI10 did not share any genes with the target pathway (Fig. 2B), although still received a higher NEA score. In this case, the drug resistance could potentially have been mediated by the DNA repair protein XRCC5 or by the adenosylhomocysteine hydrolase AHCY, which were earlier reported to be implicated in methotrexate resistance[34] and folate metabolism[35], respectively. Both these genes were strongly downregulated in ECGI10. Since AGS may include genes altered in either direction, higher NEA scores cannot be traditionally interpreted as 'pathway activation' but rather indicate 'pathway perturbation' in general. the pathway "One carbon pool by folate" was therefore unperturbed in the low-scoring cell lines A2780 and RS411, i.e. their AGS genes were not specifically connected to the pathway. For comparison, the alternative tested enrichment methods did not detect the association between "One carbon pool by folate" and methotrexate – in particular because no pathway genes correlated significantly with drug sensitivity at the gene expression, gene copy number, and point mutation levels.

Construction of sample-specific AGS.

In order to analyze data from the *in vitro* cancer cell screens and the primary tumor samples in the same manner, we constructed AGSs by following the same platform-specific approaches. Intuitively, having an AGS that is too big or too small could deteriorate specificity or sensitivity of NEA. Therefore, in order to prove that differences are not due to selecting AGS genes in a specific way, we tested and compared a number of options for AGS compilation. Mutation-based AGSs were created by first listing all point-mutated genes in each given sample (which might include hundreds and even thousands of passenger mutations) and then retaining only those with significant network enrichment against the rest of the set. This approach[36] had been proposed for distinguishing between driver and passenger mutations, so that the filtering should enrich AGSs in cancer driver genes. Next, AGSs from gene copy number and expression data included genes most deviating from the cohort means. This was achieved by using one of the three alternative algorithms (see Methods). Again, even such deviant gene sets could still be too large, e.g. due to listing copy number-alterations over extended chromosomal regions. In order to compact these, alternative AGS versions were derived by retaining only genes with significant network enrichment for signaling and cancer pathways or for the mutation-based AGS of the same sample, which reduced the AGS lists 3–10 fold. Finally and as an extra option, we tested combined AGSs, produced by concatenation of the platform-specific AGSs.

Statistical power to detect correlates of drug sensitivity.

The goal of this first, exploratory analysis was to compare the different methods and feature classes in their ability to explain the differential drug sensitivity. To this end, we counted how many features were significantly associated with a phenotype after adjusting for multiple testing. For example (Suppl. Fig. 1), we analyzed associations between point mutation profiles of cancer cell lines[37] and cell lines' sensitivity to each of the 203 anti-cancer drugs from Basu *et al.*[29]. The fraction of low p-values (e.g. $p(H_0) < 0.001$) in the total number of statistical tests did not exceed the level expected under "true null", i.e. in absence of any associations. Therefore, no genes received q-values (adjusted p-values)[38] below 0.05. On the contrary, the correlation analysis of gene expression[12] against the same drug sensitivity profiles discovered nearly 15,000 patterns of association between gene expression and drug sensitivity (out of in total $18,900 \times 203 = 3,836,700$ tests) with $p(H_0) < 0.001$. After the adjustment, more than 2500 of these gene-drug pairs remained significant at $q < 0.001$. These two examples demonstrate how dramatically the information content could vary depending on the feature type and data origin.

Applying this approach to the *in vitro* drug screen data, we evaluated all features of different types and classes. Respectively in TCGA data - again using all available features - we measured correlations of features with survival of patients who received one of the 42 frequently used drugs in any of the eight cohorts. We systematically and uniformly compared different feature types, i.e. original data from high-throughput platforms and NEA scores as well as classes within the types (e.g. transcriptomics data from Affymetrix vs. Agilent vs. RNA sequencing). Each case was tested on both relapse-free and overall survival and three follow-up intervals. We also analyzed the relative performance of different AGS classes (Suppl. Fig. 3).
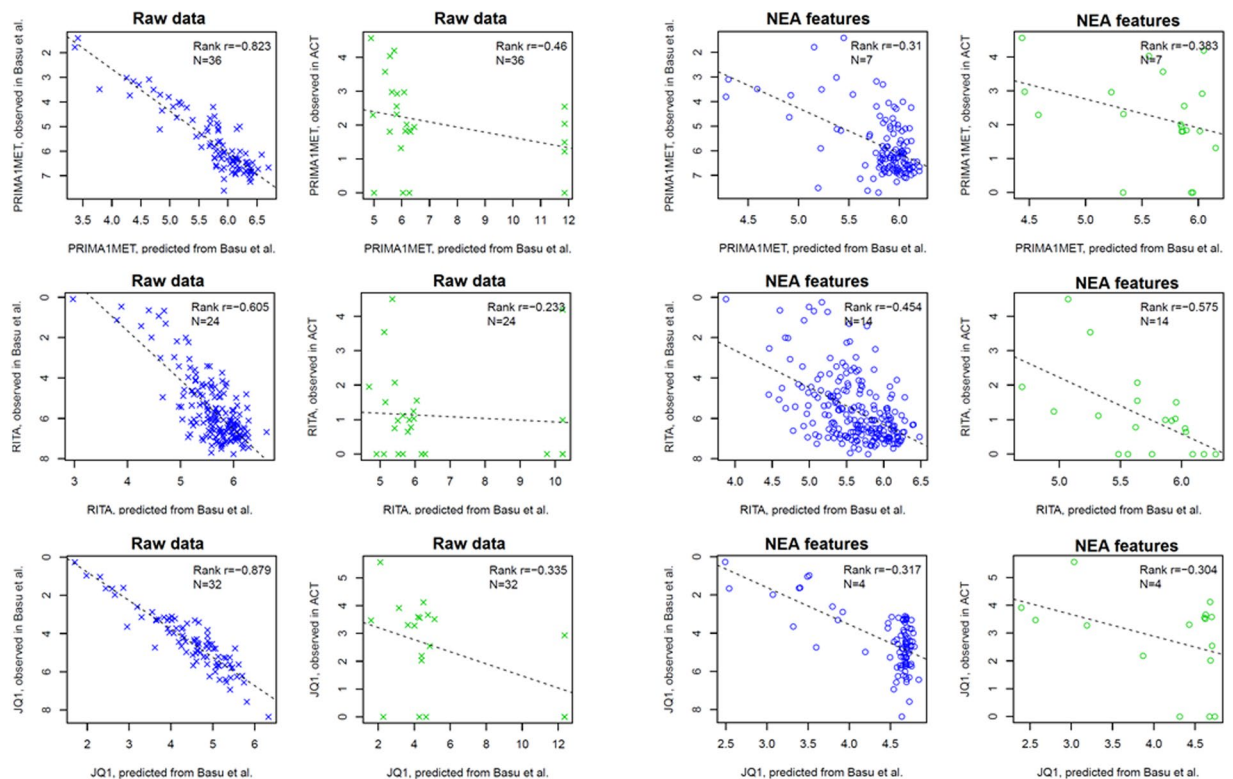
Figure 5. Predicted versus observed drug sensitivity across cancer cell lines in discovery versus validation screens. The predictive models for three compounds tested in the published CTD screen were validated in our ACT screen. Elastic net models were built under multiple cross-validation inside the training set (columns 1 and 3, blue) and then tested on non-overlapping sets of cell lines of the ACT screen (columns 2 and 4, green). Input variables were either original gene point mutation and expression profiles (columns 1 and 2, crosses) or PWNEA scores derived from these datasets for each cell line (columns 3 and 4, circles). Legends in each plot ('Rank r=') display the values of Spearman rank correlation between observed and predicted values and the number of non-zero terms in the model ('N='). Parameter alpha for the shown plots was set to 0.9. Since the drug sensitivity values from CTD screen were inverted compared to the other screens, the correlations are presented as negative values (note the inverted vertical scales for "observed in Basu *et al.*"). Detailed plots for models built under different alpha parameters are found in Supplementary Files glmnetModels.Basu_vs_new. raw.pdf and glmnetModels.Basu_vs_new.pwnea.pdf.

Overall, the NEA scores at both pathway level (PWNEA) and individual gene node level (GNEA) contained either approximately the same or larger amounts of information on drug sensitivity compared to the original gene profiles (Fig. 3; see full detailed results in Supplementary_table 5.FDR_rates_behind_Fig. 3.xlsx). In the drug screen data analysis, the ORA, PWNEA, and GNEA features performed apparently better than the respective original point mutation, gene copy number, and gene expression data. In the TCGA data analysis, the advantage of PWNEA and GNEA over both ORA and original gene profiles for particular drugs was even more pronounced, although not always overall significant. In the expression-based enrichment analyses limited to KEGG pathways (label ".kegg"), all the six tested methods significantly outperformed ORA, which served the baseline. In general, transcriptomics datasets more frequently manifested correlations with drug sensitivity than gene mutations (Suppl. Fig. 3). We also assessed relative performance of the different AGS classes. From each dataset with continuous values we created AGSs of fixed size (`top.200` and `top.400`) as well as sets of variable size where genes were included based on significance as referred to the cohort mean (`significant`) and, in addition to the latter, tested for network enrichment toward cancer gene sets (`significant.filtered.mini`) or any signaling pathways (`significant.filtered.maxi`). As illustrated in Suppl. Fig. 3, the different classes yielded variable results. We evaluated consistency and significance of these differences using the Kolmogorov-Smirnov test on the gene copy number and expression datasets for cell lines and TCGA samples (Suppl. Table 1). This evaluation, however, did not lead to an unequivocal conclusion. In the cell line datasets, the fixed size AGSs performed significantly better, while in the TCGA datasets the situation was rather opposite.

While the original features manifested considerable correlations in a number of classes, fractions of significant correlations were largely inferior when compared to NEA classes. Preliminarily, the different methods could be ranked by potential sensitivity in the following order: original gene profiles < [either ORA or EGSEA] < [either ssGSEA or SPIA] < [either PWNEA or GNEA]. However, we did not draw ultimate conclusions from the ranking. This exploratory analysis only informed us on the Type II error rates (i.e. statistical power to detect correlation). It suggested that multiple alternative methods and data types could appear – by following the state-of-the-art
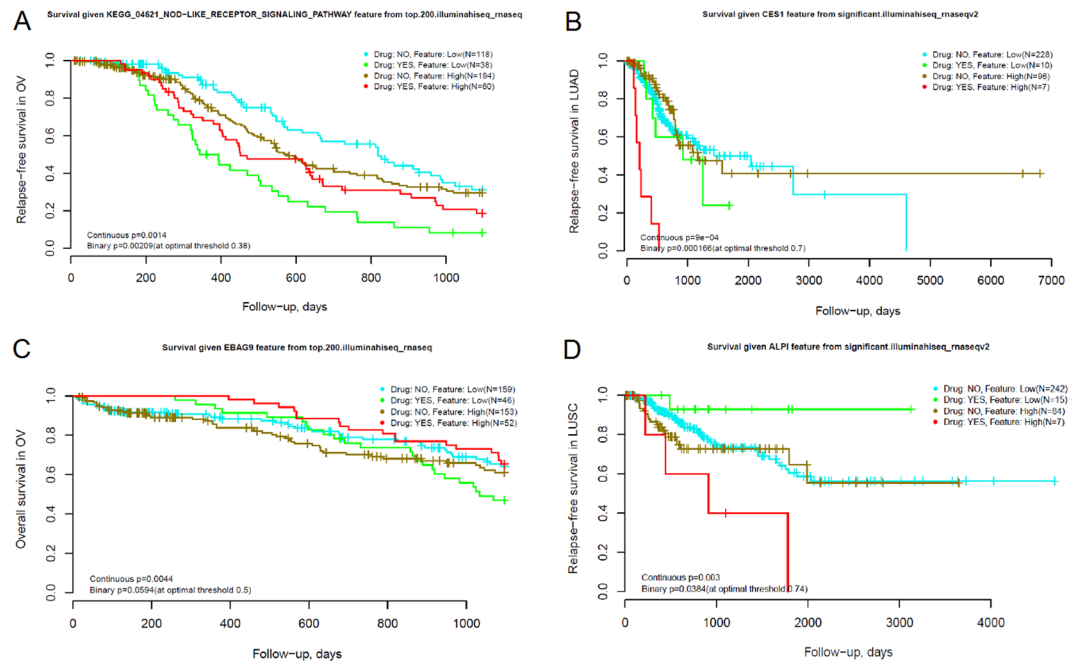
Figure 6. Clinical performance of NEA features discovered in drug screens. Each TCGA cohort was split into four categories by two factors: administration of the specific drug (as "treated/untreated") and a threshold for predictive feature (pathway or individual gene score, indicated in the plot header). While the primary feature evaluation, we calculated the factor interaction p-values without binarizing the cohort by the NEA score, i.e. in the continuous score space. Then for the visualization the binary classifications by the both factors were applied ("optimal threshold" value for the quantitative NEA feature). Therefore both continuous and binary p-values are indicated the legends. The plots present differential survival upon treatment with topotecan in ovarian carcinoma (**A,C**), gemcitabine in lung adenocarcinoma (**B**), and vinorelbine in lung squamous cell carcinoma (**D**).

approaches – potentially predictive of drug sensitivity. In order to evaluate robustness of these predictions we proceeded to the validation step as described below. We emphasize that only the AGS-based methods ORA, PWNEA, and GNEA enabled using discrete (gene point mutation and copy number) data and allowed integration over different platform types – while the other alternatives could be applied only to expression platforms.

Consistency of the discovered correlates between drug screens. In order to test reproducibility of the drug-feature associations in alternative experimental settings, we used data from three *in vitro* drug screens: CCLE[12], CGP[13], and CTD[29]. Similarly to an earlier presented comparison between CCLE and CGP screens[11], we found that the association values between drug sensitivity and original features only weakly agreed between the drug screens.

Albeit weak, these correlates were still significantly concordant across screens. Figure 4A gives examples of between-screen rank correlations while using original gene expression profiles as well as ORA, PWNEA, and GNEA features. When comparing results from CGP[13] and CTD[29] screens, the correlation values between Affymetrix expression data and sensitivity to navitoclax ranged from R = 0.31 (original gene profiles) to R = 0.81 (GNEA). More comprehensive analyses demonstrated (Fig. 4B,C) that applying PWNEA and GNEA considerably strengthened the concordance compared to the original gene profiles and ORA. the summarized ranking appeared as: [original gene profiles and ORA] < PWNEA < GNEA. In the tests using KEGG pathways with gene expression data, SPIA, ssGSEA, and EGSEA were somewhat inferior to PWNEA.

Next, we validated drug sensitivity profiles of three anti-cancer compounds, tested previously in the CTD screen - RITA, PRIMA-1[MET]/Apr-246, and JQ1 - in a new *in vitro* screen, named ACT (after "Advanced Cancer Therapies" centre at Karolinska Institutet). Activity of these compounds was re-tested in a panel of 20 cancer cell lines (the ACT set) for which gene expression and point mutation profiles data were available from the CCLE. Similarly to the results in Fig. 3, many of the both original gene profiles and NEA features showed significant correlation with drug sensitivity, which indicated a potential for creating multivariate prediction models.

As shown above, the original gene profiles were poorly preserved across drug screens. Therefore, we compared the CTD results with those from ACT screen in a more relevant multivariate approach using the "elastic net" method[39]. Starting from all available features, each model was finally reduced to a much smaller subset. Multivariate models are notoriously prone to over-fitting when the number of variables exceeds the number of samples. For this reason, validation on independent sets has become an essential requirement in such studies[40]. The CTD-based models were thus created using cell lines not found in the ACT screen. The comparison was also streamlined by using only the data from CCLE Affymetrix and point mutation datasets versus two respective

feature AGS classes `mutations.mgs` and `significant.affymetrix_ccle`. Using other classes produced similar results (data not shown).

Figure 5 demonstrates that by applying the same parameters for elastic net training, in each case it was possible to obtain a descriptive model from CTD drug screen data with a number (4…36) of non-zero terms and then substantiate the model (possibly with a poorer performance) using the ACT data in a smaller cell line set. For each model, we compared observed and predicted drug sensitivity values. The most important observation was that in all instances the signs of these correlations were consistent between CTD model and ACT validation, i.e. negative correlations in the training set remained negative upon validation.

Overall, the performance of the original profile models on the validation sets appeared comparable to that of PWNEA. However importantly, the former had much more freedom in model term selection since the initial feature space was around two orders of magnitude larger than that in PWNEA. Consequently, despite the rigorous cross-validation and feature selection implemented in the `glmnet` algorithm, using the original profiles generated more complex models (see the number of terms per model, N) which fit the training sets better. At the validation step however, the performance of the original data models significantly worsened - whereas the PWNEA-based models performed at the same level (all results obtained under variable parameters can be found in Supplementary Files glmnetModels.Basu_vs_new.raw.pdf and glmnetModels.Basu_vs_new.pwnea.pdf). This result essentially corroborated the previous conclusion about higher robustness conferred by NEA, compared to the usage of original gene profiles.

Agreement between *in vitro* screens and clinical data.

A more challenging task was to identify a conservation of associated features between the *in vitro* drug screens and clinical application of the same drugs. Any trustworthy setup of such an analysis would be very complex, so that even cross-validation and adjustment for multiple testing could not guarantee an unbiased probabilistic estimation. Thus, the final judgment should be made after a biologically independent *ad hoc* validation from the *in vitro* to the clinical domain. Even though the TCGA collection did not provide correctly balanced, randomized cohorts for estimation of relative risks, error rates etc., our task was simplified by only needing to compare the methods' performance. In the eight largest TCGA cohorts, we counted how many significant *in vitro*-detected features correlated with survival of patients who received same drug[30], (https://tcga-data.nci.nih.gov/docs/publications/tcga/; Suppl. Table 4)[41]. More specifically, molecular features of each class that significantly correlated with sensitivity to a drug in cell lines were required to also significantly correlate with patient survival in a TCGA cohort.Our survival analysis accounted for clinical covariates available from TCGA (Suppl. Table 4), which facilitated the 'net' effect estimation.

We matched correlates of same data types in CCLE and TCGA (possibly obtained using different omics platforms, e.g. Affymetrix microarray from CCLE could be matched to RNA-seq from TCGA etc.). Then we determined whether correlation p-values of individual features, in their turn, correlated between *in vitro* and TCGA data, i.e. if genes or FGSs with high (respectively low) correlation with drug response *in vitro* tended to correlate in the same manner with the patients' response. Due to the testing of alternative AGS classes, respective numbers of matching pairs in ORA, PWNEA, or GNEA were an order of magnitude higher than in raw data (column 2 in Table 3). Therefore we coupled this calculation with a significance test by randomly permuting feature and sample labels. Altogether, the permutation tests indicated that point mutation and copy number data had zero true discovery rates (TDR), i.e. their correlation p-values were preserved not more than expected by chance (see column 3 in Table 3). On the contrary, the TDR levels were substantial (0.02…0.805) for gene expression data and for AGSs processed with each of the enrichment analyses.

At the next step (remaining columns of Table 3) we calculated the numbers of significant cases that would also be practically usable, i.e. had both lower p-values (<0.001) and rank correlation values above 0.2. No such cases were identified in the gene expression data. ORA, PWNEA, and GNEA yielded 0.8%, 3.5%, and 5.9% of practically usable cases, respectively. Interestingly, most (56 out of 78) of the ORA cases were identified in the breast cancer cohort, whereas the preserved PWNEA and GNEA correlations distributed uniformly across all the TCGA cohorts (the prostate cancer cohort shared only one drug with one *in vitro* screen). Remarkably, the separate test using the 197 KGML KEGG pathways also demonstrated superiority of PWNEA over ORA, ZGSEA, EGSEA, ssGSEA, and SPIA - despite the reasonably good performance of the latter two on the *in vitro* datasets presented above. Thus at this crucial validation stage while translating drug sensitivity correlates between *in vitro* and clinical applications, robustness of the features increased in the following order: [point mutations and gene copy number changes] < [gene expression] < [ORA, ZGSEA, ESGSEA, ssGSEA, and SPIA] < PWNEA < GNEA.

This analysis proved significance of the produced correlates, and we reiterate that using the original data and the alternative enrichment methods did not seem efficient: although many transcriptomics profiles correlated with drug sensitivity, those patterns could not be traced back to the *in vitro* screens.

Most of the consistent NEA features were obtained for AGS based on gene expression data (Suppl. Table 2). They were identified for docetaxel, gemcitabine, and paclitaxel in BRCA (see the cancer cohort notation in Table 3); for dexamethasone, erlotinib, and topotecan in GBM; for gemcitabine in LUAD; and for gemcitabine, paclitaxel, tamoxifen, and topotecan in OV. While using gene copy number data, consistent PWNEA and GNEA features were found only for GBM (dexamethasone and topotecan). Consistent features that correlated with the response to cisplatin (LUSC) belonged to the combined, multi-platform types. One GNEA feature was based on somatic mutation analysis (gemcitabine in LUSC), although it did not match all the criteria. Below we present four promising findings predictive of survival in a TCGA cohort (Fig. 6).

The cancer emergence and progression were earlier linked to tissue inflammation through the NOD-like receptor signaling[42]. We found that the corresponding pathway score correlated with survival in ovarian carcinoma patients treated with topotecan (Fig. 6A).

Feature type		No. of available "feature X drug" tests	True discovery rate by permutation test, by requiring $p(H0) < 0.01$	True discovery rate by permutation test, by requiring $p(H0) < 0.01$								
				All TCGA cohorts (% of available correlates)	Bladder carcinoma, BLCA	Breast carcinoma, BRCA	Colon adeno carcinoma, COAD	Glioblastoma multiforme, GBM	Lung adeno carcinoma, LUAD	Lung squamous carcinoma, LUSC	Ovarian carcinoma, OV	Prostate adeno carcinoma, PRAD
Original gene profiles	Point mutations, PM	360	0	0	0	0	0	0	0	0	0	0
	Copy number alterations, CN	522	0	0	0	0	0	0	0	0	0	0
	Gene expression, GE	1080	0.149	0	0	0	0	0	0	0	0	0
Enrichment analysis, all platforms and FGS	ORA	9014	0.033	78 (0.8%)	0	56	0	3	3	8	8	0
	PWNEA	8822	0.146	305 (3.5%)	18	60	20	52	51	45	59	0
	GNEA	8630	0.805	505 (5.9%)	15	84	46	113	93	72	82	0
Enrichment analysis, GE on 197 KEGG pathway FGSs	ORA	5252	0.025	21 (0.3%)	0	5	0	2	4	2	8	0
	ZGSEA	1080	0.037	8 (0.7%)	0	2	0	0	1	0	5	0
	ssGSEA	1080	0.020	7 (0.6%)	0	7	0	0	0	0	0	0
	SPIA	1080	0.048	3 (0.3%)	0	0	0	3	0	0	0	0
	EGSEA.mean	1080	0.014	4 (0.3%)	0	1	0	2	0	0	1	0
	EGSEA.min	1080	0.006	0	0	0	0	0	0	0	0	0
	PWNEA	4988	0.241	364 (7.2%)	44	83	20	79	27	19	92	0

Table 3. Conservation of drug sensitivity correlates between the *in vitro* drug screens and clinical applications.

Carboxylesterases (CESs) are capable of hydrolyzing gemcitabine[43] - for instance, CES2 slows down hydrolysis of the gemcitabine pro-drug LY2334737[44]. We identified as many as 31 gene-wise NEA features which correlated with relapse-free survival in lung adenocarcinomas treated with gemcitabine. This list of network nodes from GNEA included CES1 (Fig. 6B), CES2, CES7, and a number of cytochromes with possible involvement in the catabolism of xenobiotics. Many of these genes were AGS members in both the gemcitabine-sensitive cell lines and in patients who responded to the gemcitabine treatment and – at the same time - were themselves part of KEGG pathways 00980 "Metabolism of xenobiotics by cytochrome p450", 00983 "Drug metabolism – other enzymes", and 00982 "Drug metabolism – cytochrome p450". Consequently, the ORA and PWNEA analyses detected enrichment of these pathways in the same patients. However the pathway scores correlated with response to gemcitabine neither in the CCLE and CTD screens nor in the LUAD cohort) and therefore would be useless as biomarkers. The gene expression profiles of carboxylesterases and cytochromes in cell lines and primary tumors did not correlate with gemcitabine response either. EBAG9 had been implicated previously in ovarian cancer progression[45], but it has not been shown to affect response to topotecan. Indeed, in the datasets of our study the expression of the gene itself correlated neither with cell line sensitivity to topotecan nor with patient survival. However, the GNEA features (Fig. 6C) for EBAG9 as a network node did correlate with sensitivity to topotecan *in vitro* (`top.200.affymetrix_ccle`; $p(H_0) = 4.2 \times 10^{-11}$) and with overall survival of OV patients (`top.200.illuminahiseq_rnaseq`; $p(H_0) = 4.4 \times 10^{-4}$ during the 3-year follow-up time while accounting for "clinical stage" as a covariate).

The intestinal-type alkaline phosphatase ALPI is known to modulate cancer cell differentiation[46] and cytoprotection[47,48]. In our analysis its GNEA feature was (Fig. 6D), in parallel with eleven others, negatively correlated with sensitivity to vinorelbine *in vitro* (`gnea.significant.affymetrix1`; $p(H_0) = 1.1 \times 10^{-07}$) and with relapse-free survival of OV patients ($p(H_0) = 0.003$).

This setup could not eliminate possible confounding effects from multi-drug treatment history and clinical factors that might determine administration of specific drugs. Nonetheless, the NEA scores apparently explained the differential sensitivity to anti-cancer drugs in a much more robust and efficient manner than the original data.

A visual inspection of the survival curves in Fig. 6 sheds light on usefulness of these tentative biomarkers in a clinical setting. As an example, in a 1-year survival perspective, relative risks (RR) would either increase (Fig. 6A,C) or decrease (Fig. 6B,D) given higher NEA scores of the patient samples. By using this fixed follow-up interval and the cohorts of limited size, the confidence intervals at the 95% level would be rather broad: $\ln(RR) = 0.405$ (95% CI: $[-0.07...0.88]$); $\ln(RR) = -2.061$ (95% CI: $[-3.99...-0.13]$); $\ln(RR) = 2.211$ (95% CI: $[-0.70...5.12]$); $\ln(RR) = -2.181$ (95% [CI: $-5.15...0.78$]) for Fig. 6A...D, respectively. The fractions of patients who might benefit from using these predictors could be estimated in terms of absolute risk reduction as 0.17, 0.62, 0.08, and 0.25. Inversely, the "number needed to treat", i.e. how many patients should be treated for one individual to benefit from the new test would have been 6.00, 1.60, 12.91, and 3.94, respectively[49]. However we note that additional responders could be detected by using multiple markers in parallel. As an example, beyond the "NOD–like receptor signaling pathway" at Fig. 6A, the response to topotecan in ovarian cancers similarly correlated with KEGG pathways "One carbon pool by folate" and "Bacterial invasion of epithelial cells" as well as with the GO term "Cytokine activity" (not shown). Predictions made with these markers would overlap only partially and therefore can complement each other. We presume that such discoveries should ultimately be evaluated by independent validation and careful clinical development. our combined analysis of independent cell screen and clinical results gave a first example of such validation.

Discussion

In our view, the advantages of our NEA approach are due to the following features of network-based data interpretation: (1) combining major types of molecular interactions in a biologically relevant way, (2) summarizing seemingly disparate molecular alterations at the level of pathways and processes, and (3) enabling lower-dimensional statistical analysis. In addition, a network context, with different types of evidence behind the edges provides better grounds for biological interpretation[36,50,51]. The poor performance of the individual gene analysis and alternative enrichment methods could be explained by the excessive dimensionality of the former and poorer sensitivity of the latter. In addition, the ability to use smaller and hence more specific AGS could have provided extra advantage of NEA over ORA and GSEA. On the other hand, NEA could also deteriorate on AGS of insufficient size when using sparser networks (around $10^4 \ldots 10^5$ edges) and networks with many missing nodes. These potential limitations were established earlier[36] and we tried to avoid them in the present work by using e.g. a denser network from data integration. We admit that a future, more comprehensive version of NEA might adopt advantages of the alternative enrichment methods by employing full gene lists (as in GSEA) and intra-pathway topology (as in SPIA). Indeed, at certain steps of our analysis these methods demonstrated performance comparable to that of NEA.

The individual molecular phenotypes of cell lines and tumors were characterized with AGSs compiled using a number of alternative methods. The analysis provided a primary comparison of their relative performance but – at the current stage – did not enable definite conclusions about performance of the different AGS classes. Indeed, AGS of fixed size (top.N) versus variable size (significant) compared differently in the cell lines versus the TCGA data (Suppl. Table 1). Further in the analysis of consistency *in vitro* versus clinical results, these classes were almost equally represented (Suppl. Table 2). We have also seen differences between different filtering approaches in AGSs of classes significant.mini and significant.maxi (Suppl. Fig. 2). Therefore an issue to be investigated further is the comparative performance and robustness of different feature classes, platforms etc. Importantly, multiple platforms' data can be integrated into combined AGSs. Although in our analysis such AGSs did not perform much better than platform-specific ones (most likely due to the domination of transcriptomics data), a more detailed evaluation should be done, including new platforms from TCGA and elsewhere, such as DNA methylation, protein phosphorylation etc. Given the diversity of carcinogenesis routes and the multiplicity of respective molecular mechanisms, combining platforms appears essential and most promising. Incorporation of approaches from sparse linear regression modeling, GSEA[26], SPIA[27], and PARADIGM[52] certainly represent promising ways in this direction. The statistical power of NEA was obviously far from full. As an example, there were 13 drugs for which the numbers of tested cell lines and patients treated in TCGA cohorts were sufficient for a significant estimation. For four drugs out of these 13, no reliable correlates could be found. One instructive example could be irinotecan, prescribed to 25 and 22 patients in COAD and GBM cohorts, respectively. The interesting feature of irinotecan is that its pharmacokinetic pathway involves the same enzymes as that of gemcitabine (Fig. 6B), namely CES1, CES2, CYP3A4, CYP3A5 and some others (https://en.wikipedia.org/wiki/Irinotecan#Interactive_pathway_map) – although the enzymes here work in an opposite direction: they activate irinotecan rather than degrade as they do to gemcitabine. Nonetheless, relevant GNEA scores might have been informative for response to irinotecan. The patients' response was sufficiently differential, too: while all the irinotecan-treated patients relapsed, the time to relapse varied from 78 to 1265 days. However, we did not observe almost any sensible correlation of the pathway genes neither as GNEA features nor as raw gene expression profiles. In regard of GNEA, this elucidated a lack of network linkage between the AGSs of responders (or non-responders) to the irinotecan pathway.

Further, our FGSs were created by third-party sources and never meant to be used in NEA. Thus, another step for NEA-based biomarker discovery would be the compilation of novel, specifically optimized FGSs. Ultimately, one could compile *de novo* pathways - similarly to the approach by Glaab *et al.*[53], but specifically informative of the drug response or disease prognosis. An example of such a functional set could be the presented above combination of the ten carboxylesterases and cytochromes.

Finally, given the low overlap of member genes between individual AGS, it is important to establish how AGS-level biomarker panels would practically summarize gene-level information and organize the accompanying statistical framework. Ways to compile and employ multi-platform AGSs, optimal FGS design, and construction of NEA-based biomarker panels should therefore become the topics of future studies.

Conclusions

We have presented method NEAmarker for using network enrichment scores in prediction of drug response and demonstrated its advantage compared to the conventional analyses of original gene profiles and alternative, previously presented enrichment methods. In the first place, NEAmarker allowed combining data from multiple omics platforms. Further, the NEA scores indicated higher statistical power to detect enrichment and were therefore more prone to manifest correlation with drug sensitivity. The higher robustness, anchored in the network context, enabled better preservation better between independent screens. Multivariate models using NEA scores were built from a lower-dimensional space, thereby proving more compact and, at the same time, robust when re-tested on novel data. Finally, corroborating *in vitro* phenotypes in corresponding clinical applications was possible by using our method but not by original profiles or alternative methods.

Methods

Drug screens. *Cell lines used in ACT screen.* In this analysis at ACT ("Advanced Cancer Therapies") centre, we used 20 cancer cell lines for which molecular data could be found in the CCLE Affymetrix set as well as in both CCLE and COSMIC point mutation sets: A375, HCT116, HDLM2, HT29, JVM2, K562, L428, MCF7, MDAMB231, MV411, NB4, PL21, RAJI, RKO, SJSA1, SKBR3, SKNAS, SW480, T47D, and U2OS. Eight of these cell lines had also been included in the CTD screen (A375, HCT116, HT29, MCF7, PL21, RKO, SW480, and

U2OS). In order to avoid overlap in the multivariate models, we excluded these eight cell lines while training the original models from the CTD data and only used them in the validation set.

Assay for cell proliferation used in ACT screen. Cell proliferation was estimated with the WST-1 assay (water soluble tetrazolium). Briefly, cells were incubated with each drug for 72 hours in a 96-well plate. At the end of this period, they were incubated with WST-1 reagent (Roche) for 2 hours. Absorbance at 450 nm was measured following the instructions from the manufacturer. The cell proliferation rate compared to that in the control was calculated.

For adherent cultures, cells were attached overnight before adding the compounds. For hematological malignancies, the compounds were added simultaneously with seeding cells. The initial cell density was chosen so as to avoid confluence at the end of the assay. Each compound was applied in six consecutive 3-fold dilutions. In RITA and Apr-246/PRIMA-1-met, the stocks were established at the concentration based on individually determined efficacy. Final concentrations were for RITA: 0.01, 0.04, 0.12, 0.37, 1.11, 3.33 μM and for Apr-246/PRIMA-1-met: 0.3, 1, 3, 9, 28, 83 μM. For JQ1 the cell lines HDLM2, HT29, MCF7, RAJI, RKO, SJSA1, SKBR3, and SW480 were tested using the final concentration range 1.66…0.007 μM in 1:3 serial dilutions. However later we found it necessary to raise the concentration by one order of magnitude, so that the final concentrations for the rest of the cell lines were 16.66, 5.55, 1.85, 0.61, 0.20 μM. Then we respectively adjusted IC50 values for the first group as if they were tested under the final concentrations. This was done by incrementing the initial-stock IC50 values of HDLM2, HT29, RKO, and SW480 by $log_3(10) \approx 2.09$. The cell lines MCF7, RAJI, SJSA1, SKBR3 did not show any sensitivity while using the initial stock (IC50 = 0), so that their IC50 values upon JQ1 treatment were declared missing.

IC50 was defined as the drug concentration inducing a 50% reduction in cell proliferation compared to the control. In the quantitative analysis, we used a universal scale for all the drugs where units 1…6 stood for dilution steps (1 = 1:300; 2 = 1:900; 3 = 1:2700; 4 = 1:8100; 5 = 1:24300 and 6 = 1:72900). Sensitivity to compounds was expressed in IC50 values varying from 0 (insensitive to compound) to 6 (fully sensitive to compound).

IC50 values and p-values of the model parameters were calculated using function `drm` from R package `drc`[48]. The model form (argument `fct`) was chosen as LL.4, where model parameters `Lowest` and `Highest` were fixed at cell proliferation rates 0% and 100%, respectively, while parameters slope and IC50 were left unfixed. The IC50 values are provided as Supplementary File IC50values.ACTscreen.xlsx. Missing values are presented as NA.

CCLE screen. Barretina *et al.*[12] analyzed cell line sensitivity to 24 drugs in 504 cell lines. These authors considered a range of numeric sensitivity metrics for their analysis and finally preferred 'normalized activity areas'. These original units were calculated as areas under compound response curves where higher values corresponded to higher sensitivity so that 0 stood for 'insensitive to compound' and 8 corresponded to 'full sensitivity'. Further, the activity area values were normalized for unequal luminescence in the assay. We rendered them normally distributed by log-transformation. Thus the values in our analysis range from −3.00 meaning 'insensitive to compound' to +2.31 meaning 'maximal sensitivity'.

CGP screen. Garnett *et al.*[13] analyzed 138 drugs in 714 cell lines. They used a combination of IC50 and the slope parameter to achieve the most complete description of responses. We decided to use the AUC as a single feature that reflects the both values. AUC was originally provided in the same table and ranged from 0% (fully sensitive) to 100% (insensitive). To approach the normal distribution, we transformed the values as $log(1 - AUC)$, so that now they ranged from −8.11 meaning 'insensitive to compound' to 0 meaning 'maximal sensitivity'.

CTD screen. The authors (Basu *et al.*)[29] mainly used areas under curve (AUC) for their quantitative analysis of 203 drugs in 242 cell lines. We reproduced this approach in our study. In completely insensitive cases, the full area under eight experimental points reached 8, whereas 0 stood for full sensitivity. Thus, the scale of this screen was inverted compared to the other screens, which was considered in all calculations.

Molecular data. *Gene expression.* The profiling was performed in CCLE study using Affymetrix GeneChip® Human Genome U133 Plus 2.0 Array and in CGP study by Affymetrix GeneChip® HT Human Genome U133 Array plate. The expression datasets were normalized as described when made public by the authors. Expression profiles in the CTD study were from CCLE. It has been shown earlier[11] that disagreement between CGP and CCLE could be attributed to the usage of different transcriptomics datasets only to a minor extent. We checked both the CCLE and CGP expression profiles and concluded that the latter provided poorer statistical power in regard to drug sensitivity as well as lower coverage of both genes (13891 unique mapped gene symbols vs. 18900 in CCLE) and cell lines (622 vs. 1034). For these reasons, we used the CCLE dataset in all the presented alternative analyses except NEA and ORA. AGS for the latter two were compiled from the both platforms (`affymetrix_ccle` and `affymetrix_cgp`). Expression values *x* of the downloaded datasets were transformed to $log_2(x)$.

Gene Copy Number. CCLE, CGP, and CTD all employed Affymetrix SNP 6.0 microarrays for gene copy number detection. We downloaded the CCLE dataset[12] for 994 cell lines. In addition, we downloaded COSMIC data[37] independently produced by the same platform and then post-processed in three different ways to provide total, absolute copy number per gene, number of copies of the minor allele, and a binary classification of gene copy number values into "gain" vs. "loss". All datasets were used as downloaded, without further processing or normalization.

Point mutations. CCLE provided point mutation data on sequencing of 1667 genes in 904 cell lines. In addition, we downloaded COSMIC data from exome sequencing of 1023 cell line genomes, which mapped to 19759 gene symbols. Mutation data from the both screens were used in the binary form, i.e. all specifying attributes were neglected.

Following the same approach, we employed TCGA data on somatic point mutations reported in MAF files. The column 'Variant_Classification' contained a number (more than 15) different codes, most frequent being Missense_Mutation, Nonsense_Mutation, and Silent. the latter constituted around 25% of the total number of somatic mutations reported in the eight cancers, while around half of such cases contained only mutations reported as silent. This fraction would not significantly affect the false positive and true discovery rates in any enrichment analysis. Furthermore, we found that in each cohort tens to hundreds of most frequently mutated genes (e.g. top 5% ranked by frequency per base pair length) had a significant rate of purely silent mutations (see Supplementary File Nmut_vs_frequency.4cohorts.pdf). Discarding such would exclude many potential cancer drivers – although of yet unknown mechanisms. We therefore included all mutation records present in the MAF files.

Alternative Methods of Pathway and/or Enrichment Analysis. We evaluated a number of existing multivariate, enrichment-based, and/or network analysis methods that could be potentially useful in the proposed analysis, accounting for their complexity, applicability to different experimental designs, and the ability to analyze individual samples rather than the whole cohort. Various statistical algorithms have been proposed to quantify functional relevance of pathways and other gene sets by accounting for gene network topology.

A number of methods can generate sparse regression models via network-based regularization, i.e. account for topological relations between potential predictors (typically gene expression variables). The regularization is based on certain assumptions, such as that e.g. term coefficients of neighbor nodes should be zeroes or non-zeroes simultaneously[54], that edge confidence weights should influence penalties on the model coefficients[55], or that there exists equivalence (or at least parallelism) between connectivity of nodes and covariance of model terms[56,57]. Advanced regularization of linear models in these methods often demonstrated promising efficiency[58]. However being very sophisticated, these models proved hard to tailor to novel, specific experimental designs. Notably, it was not feasible to include additional covariates or interaction terms which would be necessary for e.g. analyses similar to the one described in the present work - not even in the dedicated survival analysis method DegreeCox[59]. Using pathway membership information for summarizing cross-pathway linkage was proposed in[60] - however, adjusting its error rate model to other purposes has not been straightforward.

Technically, individual scores that estimate samples' uniqueness as compared to the rest of the collection can be obtained already from ORA, i.e. from the simplest analysis of dichotomous 2×2 tables applied to sample-specific gene sets[61–64], also called "class I" in the classification by Huang *et al.*[31]. For comparison, the most popular gene set enrichment analysis, GSEA[25] has been usually applied to finding pathway enrichment in gene lists pre-ranked by cohort-wise statistics. As an example, Haibe-Kains *et al.*[11] analyzed correlations between drug sensitivity and molecular features calculated on whole *in vitro* drug screens[12,13] which are among the datasets re-analyzed in this article. Those pathway enrichment scores represented correlates of drug sensitivity over the whole screened collections rather than characterized individual cell lines. Likewise, Iuliano and co-authors[65] matched molecular landscapes to survival in cancer sample cohorts in order to reverse-engineer relevant pathway and network structures. Thus, global methods often employ powerful, heavily optimized statistical techniques and are used for sample exploration or differential expression analysis[29,66] but cannot serve features for phenotype prediction in novel cell lines or tumors. An overview of network applications in cancer studies[52] showed that, indeed, most of the existing methods enabled exploratory analyses, discovery of driver genes and pathways as well as splitting a cohort into molecular subtypes, but did not characterize individual cases.

A number of hybrid approaches, such as SPIA[27] and iPAS[67] were also capable of calculating sample-specific pathway scores. However, their scores were based on gene expression values, which excluded the using of other data types. A genuinely integrative multi-omics method PARADIGM[68] (the program is currently distributed only via a company web portal), on the contrary, accounted for combinations of events in the chain DNA->mRNA->protein activity. As input, it required well characterized regulatory relationships – a complete set of which would rarely be available. Also, similarly to the former group of methods, it relied on comparing cancer to normal samples. Those dramatic alterations between the normal and cancer tissues encompassing thousands of genes would mask more fine-grained features that determine between-tumor heterogeneity, differences between sensitive and refractory cases etc. This requirement also precluded analyzing data where normal matches are missing, such as the widely used in our analysis cancer cell lines. Finally, EnrichNet[53] has been an algorithm closest in spirit to NEA: by using random walk with restart (hence not limited to 1-step network distances), it can trace AGS-FGS relationships via network paths. However it existed only in a single-AGS, web-based implementation and therefore was also not available for testing it the present analysis.

Even though individual enrichment scores can be correlated with phenotypes, they have still been rarely used in predictor models. In the case of ORA and GSEA, the major reason was that the enrichment is mostly detectable for large FGSs (hundreds to thousands genes), but such are unlikely to characterize functional differences between tumors - while compact, specific, and discriminatory gene sets tend to escape their limits of statistical power. Nonetheless, Drier *et al.*[69] have explored cancer cohorts with pathway-level sample scores derived from gene expression data in a quantitative way and found that certain sample clusters can be associated with patient survival. On the other hand, the network-based methods have been developed only recently and are therefore 'too young' to have been exploited fully. Above, we have also mentioned the network-based regularization of multiple regression models where inclusion of gene terms into the models was essentially coupled to their co-expression.

We finally decided to include in our testing, in parallel with PWNEA (pathway level NEA) and GNEA (gene node level NEA), the following methods:

(1) Using original gene profiles from respective omics platforms;
(2) ORA, over-representation analysis which was capable of working on exactly the same AGS and FGS as PWNEA;
(3) GSEA on full ranked gene lists, applying two alternative methods:

    a. ssGSEA, ranking by absolute gene expression value,
    b. ZGSEA, ranking by deviation of gene expression from the cohort mean;

(4) SPIA, measuring the pathway perturbation via known intra-pathway topology.
(5) EGSEA, that combined existing enrichment methods, using five out of the total twelve ones, which were capable of producing individual sample scores.

Using GSEA and SPIA was restricted to only transcriptomics data. SPIA, in addition, could only be run on pathways with known topology, which limited the set of available FGS to 197 KEGG pathways available in KGML format. This created an additional, specific line of testing on a limited collection of input data and FGSs for the methods ORA, ssGSEA, ZGSEA, SPIA, and PWNEA (see Figs 3,4 and Table 3).

Network Enrichment Analysis (NEA, PWNEA, and GNEA). *Network.* The network was based on the FunCoup method[50] with consecutive merging of five more resources as described and benchmarked previously[36]. The results of that benchmark indicated that FunCoup was superior to STRING (a method similar to FunCoup in terms of scale and the size of input data collection[70]), mostly due to the latter broadly using prokaryotic evidence and therefore less specific in cancer-related analyses. The second conclusion from the benchmark was that adding to the FunCoup network edges of curated databases significantly improved its performance. We therefore added the FunCoup-based network with functional links from KEGG[71], CORUM[72], and PhosphoSite[73], MSigDB transcription factor-related part[74]), and an own reverse-engineered network[36]. The resulting network thus combined a wide range of molecular mechanisms, functional relations, and metrics from high-throughput data sets: physical protein-protein interactions, membership in same protein complex, membership in the same pathway, correlation of mRNA profiles, correlation of protein abundance values, protein phosphorylation, coherence of GO annotations, concordance of upstream regulators (transcription factors and miRNAs), co-localization in same sub-cellular compartments, similarity of phylogenetic profiles etc. It contained 974,427 edges (links) between 19027 nodes (distinct human gene symbols).

Altered gene sets, AGS. Point mutation data (mutation gene sets):

- `mutations.mgs`: point-mutated genes that proved to be significantly NEA-enriched to either KEGG pathway set #05200 "Pathways in cancer" or to the full set of point-mutated genes annotated in the given genome (the approach described by Merid *et al.*[36]).
  Gene copy number and expression data:
- `top.200` and `top.400`: genes with copy number or mRNA expression value that in the given genome was among top 200 or top 400 most deviating from the gene's cohort mean using the one-sample Z-score. Each AGS thus had a fixed size, regardless of formal significance.
- `significant`: most deviating from the gene's cohort mean (same as above), but selected only if below the formal significance threshold (Benjamini-Hochberg[75] adjusted p-value < 0.05). These AGSs had variable sizes, depending on the significance criterion.
- `significant.filtered.mini`: members of the respective `significant` set had, in addition, to be also significantly NEA-enriched to either KEGG set #05200 "Pathways in cancer" or to `mutations.mgs` set of the same sample (whichever NEA score passed the significance threshold NEA FDR = 0.05).
- `significant.filtered.maxi`: members of the respective `significant` set were required to be significantly NEA-enriched to any of the signaling pathways (including all cancer ones) or to `mutations.mgs` set of the same sample.

Combined (multi-platform) AGS:

- `significant.filtered.combined.mini`: a merge (concatenation) of all sets of type `significant.filtered.mini`.
- `significant.filtered.combined.maxi`: a merge (concatenation) of all sets of type `significant.filtered.maxi`.

For convenience, AGS labels refer also to the platforms and sources, e.g. `top.200.cn_ccle`, `significant.filtered.maxi.affymetrix_ccle` etc.

FGS. The functional gene sets, FGSs, were AGS counterparts in the analysis. The main collection of 328 FGS was based on the KEGG pathways, the full collection of which was complemented with a number of separately published cancer pathways as well as specific GO terms corresponding to cancer-relevant signaling or hallmarks of cancer (around 70 cancer- and signaling-related gene sets from Reactome, Gene Ontology, WikiPathways and literature). Another approach was applied to enable compatibility with GSEA and SPIA. These methods were designed and are most suitable for analyzing expression data and, apart from that, SPIA was applicable only to pathways with well characterized intra-pathway topology. We therefore employed a special set of 197 KEGG pathways for which the topology was available in KGML files and tested on it SPIA, GSEA, ORA, and PWNEA

exclusively gene expression data (these results were separately labeled as ORA.kegg, SPIA.kegg, ssGSEA.kegg, ZGSEA.kegg, EGSEA.kegg, and PWNEA.kegg). The analysis on the FGS collection is referred to as pathway-level NEA (PWNEA).

In the other version of our analysis, called gene-wise NEA (GNEA), we treated each of the 19027 network nodes, regardless of their pathway or GO annotation, as a single-gene FGS.

Method

The major principles of NEA were described earlier[23]. In the current implementation, we evaluated enrichment of AGS versus FGS by the formula:

$$\chi^2 = \frac{(n_{\text{AGS–FGS}} - \hat{n}_{\text{AGS–FGS}})^2}{\hat{n}_{\text{AGS–FGS}}} + \frac{(!n_{\text{AGS–FGS}} - !\hat{n}_{\text{AGS–FGS}})^2}{!\hat{n}_{\text{AGS–FGS}}},$$

where $!n$ means "complement to n", i.e. all global network edges that did not belong to $N_{AGS\text{-}FGS}$. The number of links expected under true null, i.e. by chance, was determined by:

$$\hat{n}_{\text{AGS–FGS}} = \frac{N_{AGS} * N_{FGS}}{2 * N_{total}}$$

Node connectivity values (numbers of all edges for each given node) were pre-calculated by the algorithm in advance, given the input network. Then N_{AGS} and N_{FGS} reported the sums of connectivities of member nodes of AGS and FGS, respectively, and N_{total} was the number of edges in the whole network. Since it was desirable to provide normally distributed values for the downstream analyses (linear modeling, correlation, survival), we calculated p-values from the X^2 statistic $p(H_0) = f(X^2)$ using function `pchisq` available in R language and then re-calculated corresponding z-scores from the p-values as $Z = F(p(H_0))$ with function `qnorm`. Since X^2 is only defined on the non-negative domain, the z-scores were coerced negative in cases of depletion, i.e. when

$$\hat{n}_{\text{AGS–FGS}} > n_{\text{AGS–FGS}}.$$

An important feature of GNEA (gene-wise NEA) is that its enrichment estimates are, on average, based on fewer network edges compared to PWNEA, so that often $n_{\text{AGS–FGS}} = 0$. In such cases, the enrichment score is negative and the difference $n_{\text{AGS–FGS}} - \hat{n}_{\text{AGS–FGS}}$ reduces to $-\hat{n}_{\text{AGS–FGS}}$, which, in its turn, is a function of cumulative connectivity values N_{AGS} and N_{FGS}. In other words, lower NEA scores are then assigned to AGS-FGS pairs with more highly connected member nodes.

The steps of NEA described above can be performed with functions available in R package NEArender (https://cran.r-project.org/web/packages/NEArender/).

Signaling pathway impact analysis, SPIA.   The method by Tarca *et al.*[27] was implemented as an R package `SPIA`. The authors presented it as combination of two p-values: `pNDE` from common analysis of overrepresentation of differentially expressed genes in KEGG pathways and `pPERT` from a perturbation analysis by accounting for topological relations of the same genes within each KEGG pathway. Since the authors claimed that `pNDE` values are no different from p-values from the trivial ORA, we used the pure `pPERT` values from function `spia` (while the performance of ORA was evaluated separately). In order to get normally distributed values for our analyses, pPERT were transformed to *Z*-scores and signed according to the SPIA "Activated/inhibited" status as: `Z.spia = qnorm(pPERT/2, lower.tail = F)*ifelse(s1$Status=="Activated",1,−1)`.

Gene Set Enrichment Analysis, GSEA.   The R implementation of GSEA was downloaded from http://software.broadinstitute.org/gsea/msigdb/download_file.jsp?filePath=/resources/software/GSEA-P-R.1.0.zip (see also https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/R-GSEA_Readme). While GSEA possesses a sophisticated toolbox for significance estimation via permutation tests, we needed only the enrichment score and therefore calculated only the core ES values via function `GSEA.EnrichmentScore`. Normally, GSEA has been used for analyzing gene rankings from multi-sample analyses with replicates, such as a *t*-test of an experimental versus control group. The single-sample GSEA (so called ssGSEA) needed for our analysis was described by Barbie *et al.*[26]. They produced sample-specific lists by ranking genes by absolute expression values in each given sample. We implemented this analysis under acronym ssGSEA. However this approach might miss sample specificity. As an example, such ubiquitously expressed genes as GAPDH, RPS16, and RPS11 were found among the top 10 items in more than 90% of the CCLE cell line transcriptomes. For this reason, we additionally implemented and tested ranking genes in each sample by z-scores, i.e. by the standardized deviations from the genes' means across the whole cohort. Using this option, dubbed ZGSEA, was similar to mode `topnorm` for calculating AGS in function `samples2ags` of our package `NEArender`.

Overrepresentation analysis, ORA.   The overrepresentation analysis, ORA estimated the significance of overlap between AGS and FGS in 2×2 tables. We did it via Fisher's exact test using the function `gsea.render` in the R package `NEArender` described above. In order to get ORA values normally distributed, the "estimate" values from function `fisher.test` were augmented with a pseudo-score 0.1 and log-transformed.

EGSEA.   The ensemble of genes set enrichment analyses[28] can combine results from twelve individual algorithms which were previously presented by third parties. EGSEA then calculates collective gene set scores for unified enrichment estimation. Since the framework of our analysis required assigning enrichment scores to individual samples rather than to multi-sample contrasts such as "experiment vs. control", we excluded methods

that were present in EGSEA but only applicable to multi-sample contrasts, namely Camera, Roast, Fry, PADOG, CAGE, SAFE, and globaltest. We used the remaining suitable methods ORA, ZSCORE, GSVA, PLAGE, and ssGSEA. The correlation analysis and other models in our work required input in the form of enrichment scores rather than p-values. Therefore we could combine scores with two EGSEA methods which were not based on p-values, namely min rank and average rank.

Correlation between drug sensitivity and molecular features. In each of the four drug screens, we quantified correlation between the cell line sensitivity to each drug and each of the molecular features *F* according to a general model of the form:

$$S_d = \beta F + \varepsilon$$

where ε denotes residual, i.e. unexplained by feature *F*, variance. The features were either original gene profiles from the three platforms (point mutations screens, copy number arrays, and expression microarrays) or scores from GSEA, or scores from the two NEA modes, PWNEA and GNEA, i.e. pathway-level network enrichment scores and single-gene network enrichment scores, respectively. All data sets, except the point mutation set, contained continuous variables and were thus analyzed using Spearman rank correlation. The point mutation data were analyzed using a one-way ANOVA model with two levels of *F*: "any mutation" versus "wild type". P-values of both Spearman and ANOVA were adjusted by Benjamini and Hochberg method[75].

Elastic net models. Every tested model was built under 10-fold cross-validation using function `cv.glmnet` of R package glmnet (http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html) with the following parameters: `lambda.min.ratio` = 0.01 (the default) and `nlambda` = 25 (default was 100). Parameter `alpha` varied as {0.1; 0.3; 0.5; 0.9; 1.0}. The reported cross-validated mean error and the number of variables in the model corresponded to `lambda.1se`, i.e. largest value of lambda found within 1 standard error of the minimum lambda. The regression of observed on predicted values was plotted using `lambda.min`.

Drug sensitivity models in TCGA patients. We used the follow-up time profiles for which both status records "relapse/relapse-free" and "dead/alive" were available, which allowed creating "relapse-free survival" and "overall survival" variables. Depending on the cancer aggressiveness and chemotherapy type, different timeframes could become informative in the analysis of the eight TCGA cohorts. The follow-up timeframes were defined as 1/5th, 1/2nd, and full available (up to 18 years) intervals.

For the analysis reported in "Statistical power to detect correlates of drug sensitivity", we used 42 drugs which were applied to at least 10 patients in one of the eight cohorts. In Fig. 3 we report fractions of adjusted p-values (FDR) from this analysis calculated by Benjamini and Hochberg. For the analysis of "agreement between *in vitro* screen and clinical data" we only considered 14 of the compounds, which were found in the *in vitro* sets. The p-values from this analysis were Bonferroni-adjusted in the cross-comparisons between the *in vitro* and clinical results.

Matching significance of the drug-feature correlations that had been detected in the cell-line *in vitro* screens required accounting for multiple clinical variables. Such phenotype covariates as well as drug treatment data were obtained from TCGA as biotab files via

https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/tumor/*/bcr/biotab/clin/.

In order to measure and probabilistically estimate these effects, we fitted Cox proportional hazards regression models for every feature versus drug combination. Using all covariates available a cohort (such as "age at diagnosis", "year of diagnosis", "race", "gender", "ethnicity") could result in unrealistically complex models. We thus included only covariates most likely associated with the disease prognosis, such as tumor degree, pathological tumor stage, immunohistochemical statuses in BRCA, Gleason score in PRAD, Karnofsky score in GBM (Suppl. Table 4). Next, we reasoned that when the association "feature - drug response" truly exists, we should observe it specifically in the patients who did receive the drug in the given TCGA cohort. Our survival models of the form

$$\log\left(\frac{\lambda(t|C_1 \cdots C_k, D, F)}{\lambda_0(t)}\right) = \beta_1 C_1 + \cdots + \beta_k C_k + \beta_d D + \beta_f F + \beta_i D * F + \varepsilon$$

contained, apart from the covariates $C_1 \ldots C_k$ and the residual term ε, main effects "drug" *D* and "feature" *F* as well as the interaction term *D*F*. A significant main effect of a drug could be interpreted as patients' benefit in total and irrespective of the feature value, e.g. regardless of a gene mutation, or a gene expression, or a NEA-based pathway score. Conversely, a significant feature effect indicated that the feature correlated with survival directly, i.e. no matter if the drug was administered or not. Finally, significance of the interaction indicated efficacy of the drug specifically in patients with feature values either above or below a threshold, so that respective patterns could be explained by neither of the main effects. The interaction term was thus central for our purpose of detecting drug-feature correlations, whereas the significance of main effects of "feature" and "drug" was allowed although not required. As an example, a feature may or may not exhibit a significant correlation with survival in patients who did not receive the drug.

All survival analysis results were obtained using R package survival (https://doi.org/10.1007/978-1-4757-3294-8). In order to estimate significance of the model terms, we used function `coxph` with continuous feature vectors. However, for visualizing the survival curves (Fig. 6) each feature was binarized at a cutoff that yielded the lowest p-value for the interaction term. Apart from the interaction model, we also checked if the p-value and FDR distributions preserved their properties under a unifactorial model. To this end, sub-cohorts of respective drug-treated patients were included in the survival analysis with the single main factor "feature":

$$\log\left(\frac{\lambda(t|C_1 \dots C_k, D, F)}{\lambda_0(t)}\right) = \beta_f F + \varepsilon$$

Boxplots. We used the default parameters for function boxplot in R. The boxes contain data points within 25–75th percentile intervals (i.e. between quartiles Q1 and Q3). The maximal whisker length, MWL, is defined as 1.5 times the Q1–Q3 interquartile range (i.e. the box length). Whiskers can extend to either the MWL or the maximal available data point when the latter is below MWL. Markers thus correspond to data points that extend off the box by more than the MWL value.

Data Availability

The datasets generated for and analyzed during the current study are available from the corresponding author on reasonable request. The original drug screen data generated for this study (IC50 values) are included as one of the supplementary information files described below.

References

1. Bittner, L. R. Bellman, Adaptive Control Processes. A Guided Tour. XVI+ 255 S. Princeton, N. J., 1961. Princeton University Press. Preis geb. $ 6.50. *ZAMM - Z. Für Angew. Math. Mech.* **42**, 364–365 (1962).
2. Heng, H. H. *Debating Cancer: The Paradox in Cancer Research*. (World Scientific, 2015).
3. Lee, W. *et al.* Identifying and Assessing Interesting Subgroups in a Heterogeneous Population. *BioMed Res. Int.* **2015**, 462549 (2015).
4. Parker, J. S. *et al.* Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
5. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet Lond. Engl.* **365**, 671–679 (2005).
6. Roepman, P. *et al.* An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **15**, 284–290 (2009).
7. Subramanian, J. & Simon, R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J. Natl. Cancer Inst.* **102**, 464–474 (2010).
8. Waldron, L. *et al.* Comparative Meta-analysis of Prognostic Gene Signatures for Late-Stage Ovarian Cancer. *JNCI J. Natl. Cancer Inst.*, https://doi.org/10.1093/jnci/dju049 (2014).
9. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in. *Cancer. Cell* **166**, 740–754 (2016).
10. Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* **4** (2013).
11. Haibe-Kains, B. *et al.* Inconsistency in large pharmacogenomic studies. *Nature* **504**, 389–393 (2013).
12. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
13. Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
14. Stransky, N. *et al.* Pharmacogenomic agreement between two cancer cell line data sets. *Nature*, https://doi.org/10.1038/nature15736 (2015).
15. Vanden Heuvel, J. P. *et al.* Replication Study: Systematic identification of genomic markers of drug sensitivity in cancer cells. *eLife* **7** (2018).
16. Crystal, A. S. *et al.* Patient-derived models of acquired resistance can identify effective drug combinations for cancer. *Science* **346**, 1480–1486 (2014).
17. Seashore-Ludlow, B. *et al.* Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov.* **5**, 1210–1223 (2015).
18. Kurnit, K. C. *et al.* "Personalized Cancer Therapy": A Publicly Available Precision Oncology Resource. *Cancer Res.* **77**, e123–e126 (2017).
19. Margolin, A. A. *et al.* Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. *Sci. Transl. Med.* **5**, 181re1 (2013).
20. Cheng, W.-Y., Yang, T.-H. O. & Anastassiou, D. Biomolecular Events in Cancer Revealed by Attractor Metagenes. *PLoS Comput. Biol.* **9**, e1002920 (2013).
21. Costello, J. C. *et al.* A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32**, 1202–1212 (2014).
22. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
23. Alexeyenko, A. *et al.* Network enrichment analysis: extension of gene-set enrichment analysis to gene networks. *BMC Bioinformatics* **13**, 226 (2012).
24. Jeggari, A. & Alexeyenko, A. NEArender: an R package for functional interpretation of 'omics' data via network enrichment analysis. *BMC Bioinformatics* **18** (2017).
25. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
26. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
27. Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* **25**, 75–82 (2009).
28. Alhamdoosh, M. *et al.* Combining multiple tools outperforms individual methods in gene set enrichment analyses. *Bioinformatics* **33**, 414–424 (2017).
29. Basu, A. *et al.* An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* **154**, 1151–1161 (2013).
30. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
31. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
32. Tabor, H. & Wyngarden, L. The enzymatic formation of formiminotetrahydrofolic acid, 5,10-methenyltetrahydrofolic acid, and 10-formyltetrahydrofolic acid in the metabolism of formiminoglutamic acid. *J. Biol. Chem.* **234**, 1830–1846 (1959).
33. Rajagopalan, P. T. R. *et al.* Interaction of dihydrofolate reductase with methotrexate: Ensemble and single-molecule kinetics. *Proc. Natl. Acad. Sci. USA* **99**, 13481–13486 (2002).
34. Snijders, A. M. *et al.* Acquired genomic aberrations associated with methotrexate resistance vary with background genomic instability. *Genes. Chromosomes Cancer* **47**, 71–83 (2008).

35. Pogribny, I. P. *et al*. Epigenetically mediated inhibition of S-adenosylhomocysteine hydrolase and the associated dysregulation of 1-carbon metabolism in nonalcoholic steatohepatitis and hepatocellular carcinoma. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol*., https://doi.org/10.1096/fj.201700866R (2017).

36. Merid, S. K., Goranskaya, D. & Alexeyenko, A. Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis. *BMC Bioinformatics* **15**, 308 (2014).

37. Forbes, S. A. *et al*. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*., https://doi.org/10.1093/nar/gku1075 (2014).

38. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).

39. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw*. **33** (2010).

40. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Board on Health Care Services, Board on Health Sciences Policy & Institute of Medicine. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. (National Academies Press (US), 2012).

41. Liu, J. *et al*. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400–416.e11 (2018).

42. Saxena, M. & Yeretssian, G. NOD-Like Receptors: Master Regulators of Inflammation and Cancer. *Front. Immunol*. **5**, 327 (2014).

43. Williams, E. T. *et al*. Characterization of the Expression and Activity of Carboxylesterases 1 and 2 from the Beagle Dog, Cynomolgus Monkey, and Human. *Drug Metab. Dispos*. **39**, 2305–2313 (2011).

44. Pratt, S. E. *et al*. Human Carboxylesterase-2 Hydrolyzes the Prodrug of Gemcitabine (LY2334737) and Confers Prodrug Sensitivity to Cancer Cells. *Clin. Cancer Res*. **19**, 1159–1168 (2013).

45. Akahira, J.-I. *et al*. Expression of EBAG9/RCAS1 is associated with advanced disease in human epithelial ovarian cancer. *Br. J. Cancer* **90**, 2197–2202 (2004).

46. Shin, J. *et al*. The Intestinal Epithelial Cell Differentiation Marker Intestinal Alkaline Phosphatase (ALPi) Is Selectively Induced by Histone Deacetylase Inhibitors (HDACi) in Colon Cancer Cells in a Kruppel-like Factor 5 (KLF5)-dependent Manner. *J. Biol. Chem*. **289**, 25306–25316 (2014).

47. Giatromanolaki, A., Sivridis, E., Maltezos, E. & Koukourakis, M. I. Down-regulation of intestinal-type alkaline phosphatase in the tumor vasculature and stroma provides a strong basis for explaining amifostine selectivity. *Semin. Oncol*. **29**, 14–21 (2002).

48. Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-Response Analysis Using R. *PLOS ONE* **10**, e0146021 (2015).

49. Laupacis, A., Sackett, D. L. & Roberts, R. S. An Assessment of Clinically Useful Measures of the Consequences of Treatment. *N. Engl. J. Med*. **318**, 1728–1733 (1988).

50. Alexeyenko, A. & Sonnhammer, E. L. L. Global networks of functional coupling in eukaryotes from comprehensive data integration. *Genome Res*. **19**, 1107–1116 (2009).

51. Alexeyenko, A. *et al*. Comparative interactomics with Funcoup 2.0. *Nucleic Acids Res*. **40**, D821–D828 (2012).

52. Kim, Y.-A., Cho, D.-Y. & Przytycka, T. M. Understanding Genotype-Phenotype Effects in Cancer via Network Approaches. *PLOS Comput. Biol*. **12**, e1004747 (2016).

53. Glaab, E., Baudot, A., Krasnogor, N., Schneider, R. & Valencia, A. EnrichNet: network-based gene set enrichment analysis. *Bioinforma. Oxf. Engl*. **28**, i451–i457 (2012).

54. Kim, S., Pan, W. & Shen, X. Network-based penalized regression with application to genomic data. *Biometrics* **69**, 582–593 (2013).

55. Li, C. & Li, H. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics* **24**, 1175–1182 (2008).

56. Zhao, S. & Shojaie, A. A Significance Test for Graph-Constrained Estimation. *Biometrics* **72**, 484–493 (2016).

57. Dirmeier, S., Fuchs, C., Mueller, N. S. & Theis, F. J. netReg: network-regularized linear models for biological association studies. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btx677 (2017).

58. Shojaie, A. & Michailidis, G. Network enrichment analysis in complex experiments. *Stat. Appl. Genet. Mol. Biol*. **9**, Article22 (2010).

59. Veríssimo, A., Oliveira, A. L., Sagot, M.-F. & Vinga, S. DegreeCox – a network-based regularization method for survival analysis. *BMC Bioinformatics* **17**, 109–121 (2016).

60. Huttenhower, C. *et al*. Exploring the human genome with functional maps. *Genome Res*. **19**, 1093–1106 (2009).

61. Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. & Church, G. M. Systematic determination of genetic network architecture. *Nat. Genet*. **22**, 281–285 (1999).

62. Drăghici, S., Khatri, P., Martins, R. P., Ostermeier, G. C. & Krawetz, S. A. Global functional profiling of gene expression. *Genomics* **81**, 98–104 (2003).

63. Khatri, P., Draghici, S., Ostermeier, G. C. & Krawetz, S. A. Profiling Gene Expression Using Onto-Express. *Genomics* **79**, 266–270 (2002).

64. Robinson, M. D., Grigull, J., Mohammad, N. & Hughes, T. R. FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinformatics* **3**, 35 (2002).

65. Iuliano, A., Occhipinti, A., Angelini, C., De Feis, I. & Lió, P. Applications of Network-based Survival Analysis Methods for Pathways Detection in Cancer. in *Computational Intelligence Methods for Bioinformatics and Biostatistics* (eds di Serio, C., Liò, P., Nonis, A. & Tagliaferri, R.) 8623, 76–88 (Springer International Publishing, 2015).

66. Yu, X., Zeng, T. & Li, G. Integrative enrichment analysis: a new computational method to detect dysregulated pathways in heterogeneous samples. *BMC Genomics* **16** (2015).

67. Ahn, T., Lee, E., Huh, N. & Park, T. Personalized identification of altered pathways in cancer using accumulated normal tissue data. *Bioinforma. Oxf. Engl*. **30**, i422–429 (2014).

68. Vaske, C. J. *et al*. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, i237–i245 (2010).

69. Drier, Y., Sheffer, M. & Domany, E. Pathway-based personalized analysis of cancer. *Proc. Natl. Acad. Sci*. **110**, 6388–6393 (2013).

70. von Mering, C. *et al*. STRING 7–recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res*. **35**, D358–362 (2007).

71. Kanehisa, M., Goto, S., Kawashima, S. & Nakaya, A. The KEGG databases at GenomeNet. *Nucleic Acids Res*. **30**, 42–46 (2002).

72. Ruepp, A. *et al*. CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res*. **36**, D646–D650 (2007).

73. Hornbeck, P. V. *et al*. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res*. **40**, D261–D270 (2012).

74. Liberzon, A. *et al*. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. **1**, 417–425 (2015).

75. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to MultipleTesting. *J. R. Stat. Soc. Ser. B Methodol*. **57**, 289–300 (1995).

76. Jeggari, A. *et al*. EviNet: a web platform for network enrichment analysis with flexible definition of gene sets. *Nucleic Acids Res*., https://doi.org/10.1093/nar/gky485

Acknowledgements

Author Contributions

M.F. performed the drug screen and participated in the data analysis. A.J. participated in the data analysis. S.P. and F.B. participated in the cell line analysis. G.S. designed and supervised the drug screen, the cell line and data analyses, participated in writing the manuscript. A.A. designed the study, analyzed the data and wrote the manuscript. All authors read, edited, and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-39019-2.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.