

RESEARCH

Open Access



Machine learning approach informs biology of cancer drug response

Eliot Y. Zhu^{1,2,3,4} and Adam J. Dupuy^{1,2*}

*Correspondence:
adam-dupuy@uiowa.edu

¹ Department of Anatomy and Cell Biology, The University of Iowa, Iowa City, IA, USA
Full list of author information is available at the end of the article

Abstract

Background: The mechanism of action for most cancer drugs is not clear. Large-scale pharmacogenomic cancer cell line datasets offer a rich resource to obtain this knowledge. Here, we present an analysis strategy for revealing biological pathways that contribute to drug response using publicly available pharmacogenomic cancer cell line datasets.

Methods: We present a custom machine-learning based approach for identifying biological pathways involved in cancer drug response. We test the utility of our approach with a pan-cancer analysis of ML210, an inhibitor of GPX4, and a melanoma-focused analysis of inhibitors of *BRAF*^{V600}. We apply our approach to reveal determinants of drug resistance to microtubule inhibitors.

Results: Our method implicated lipid metabolism and Rac1/cytoskeleton signaling in the context of ML210 and BRAF inhibitor response, respectively. These findings are consistent with current knowledge of how these drugs work. For microtubule inhibitors, our approach implicated Notch and Akt signaling as pathways that associated with response.

Conclusions: Our results demonstrate the utility of combining informed feature selection and machine learning algorithms in understanding cancer drug response.

Keywords: Drug resistance, Machine learning, Paclitaxel, Ovarian, PAX8, MECOM, AKT

Background

Drug resistance and off-target toxicity are two major obstacles for precision cancer treatment. Experimental approaches to understand these areas of research depend on the use of genetic screens or drug perturbation experiments paired with -omics profiling. However, such experiments require large commitments of resources including cell culture, genetic screening constructs, sequencing costs, and personnel. Analysis of publicly available pharmacogenomic datasets is a vastly less expensive option to understand the biology of cancer drugs. The difficulty with using *in silico* approaches is that meaningful signals may be weak and not easily detectable. Considering this challenge, Machine learning (ML) algorithms has become an increasingly popular strategy to build predictive models that utilize molecular patients of tumor or cancer cells to predict and understand patients' or cell lines' response to drugs [1–11].



Existing strategies for building drug response classifiers are incredibly diverse, utilizing various combinations of inputs, feature selection approaches, and algorithms. Here, we built a machine learning algorithm focused on informing the biological processes that drive cancer drug response. We do so by integrating prior knowledge of biological pathways and protein–protein interaction data. We tested our approach on two compounds: ML210, a selective covalent inhibitor of glutathione peroxidase 4 (GPX4) and the selective BRAF^{V600} inhibitors vemurafenib (VEM) and dabrafenib. We also used our approach to identify pathways that inform response to anti-tubulin drugs.

Methods

All analysis was performed in R using custom scripts. First, consider KEGG pathways belonging to Metabolism, Genetic Information Processing, Environmental Information Processing, and Cellular Processes. This list contains ~150 pathways. For each pathway, compute the pathway activity scores. The pathway activity score is defined as the t-score of the pathway activities across drug-sensitive and drug-resistant cell lines. Specifically, the pathway activity for pathway p , sample j , and gene i , is given by,

$$a_{pj} = \sum_{i=1}^k \frac{z_{ij}}{\sqrt{k}}$$

where z is the normalized gene expression. The number of genes to use for each pathway, or k , is determined using a greedy search strategy. That is, compute the t-score for each gene for a given pathway. Rank genes in increasing order if average t-scores are less than zero or in decreasing order, otherwise. Iterate over i until the maximum a_p is found. In other words, k is the smallest number that maximizes the t-score for a_p . See [12] for complete details on computing the pathway activity score.

For the BRAFi analysis, pathways are determined to be significant using a null distribution generated by permuting the cell line labels. For the ML210 and Paclitaxel analysis, pathways with pathway activity scores within the bottom or top 10th or 20th percentiles, respectively, were retained for further analysis. Significance thresholds were designed to return ~20% of the initial number of input KEGG pathways.

Next, take all genes from the pathways deemed significant. Bin these genes into mutually exclusive network modules. Genes are grouped together into mutually exclusive network modules through hierarchical clustering of the dissimilarity between genes. Dissimilarity is computed as 1 minus the standard topological overlap measure described in [13]. The adjacency matrix used to compute the topological overlap was derived using STRING protein–protein interactions [14]. Namely, we considered an edge to exist between two genes if they had a STRING combined score of ≥ 0.4 .

Then, determine the most informative genes in each module, separately, using Boruta, a random-forest-based feature selection algorithm, with default parameters [15]. Genes with a finalDecision of “Confirmed” was retained for further analysis.

Boruta determines variable importance by comparing the performance of an attribute relative to permuted versions of it within random forest classification.

Finally, take all the informative genes from the previous step and build a classifier using the support vector machine learning algorithm with recursive feature elimination (RFE). We used the implementation provided at <https://github.com/johncolby/SVM-RFE>. RFE involves running the SVM iteratively while removing the least informative feature at each iteration. The rank of the feature is inversely related to the iteration it was removed by the SVM algorithm. For our analysis, the rank for a feature is given as an average of a feature's rank across ten-fold cross validation for the ML210 and Paclitaxel analysis or leave-one-out-cross validation for the BRAFi analysis. The ranking of each feature determines the importance of the module it belongs to. The biological representation of each module was determined using Gene Ontology pathways enrichment analysis implemented by the limma R package [16].

To perform our machine learning analysis, we used RMA-normalized microarray gene expression from Genomics of Drug Sensitivity in Cancer (GDSC). We used ML210 and PTX drug response data from the Cancer Therapeutics Response Portal V2 (CTRP v2). We used VEM and Dabrafenib response data from GDSC. We used area under the curve (AUC) as the metric for drug response. The cutoff for ML210 resistance was set at an AUC of 9, which qualitatively separated two modes of the AUC distribution (Additional file 1: Figure S1). The cutoff for PTX resistance was set at 5 to distinguish the most sensitive cancers. The cutoff for drug response for BRAF inhibition was set at the 5th percentile of the AUC for VEM or Dabrafenib in the GDSC. Two BRAF inhibitors were used to compensate for missing data.

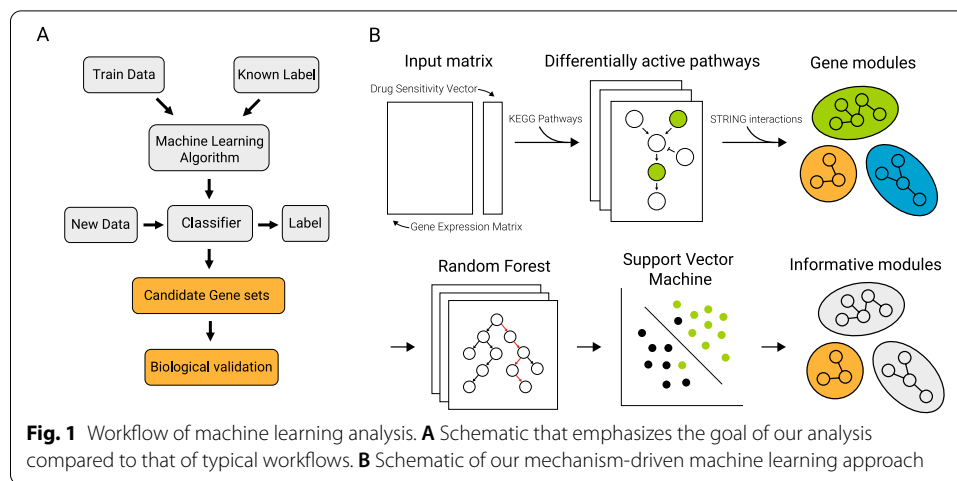
The *singscore* [17] R-package was used to compute the pathway enrichment scores for the 4-gene NOTCH3/PAX8 across ovarian cancer cell lines. The *biomaRt* R-package was used for data wrangling [18]. The *ggplot2* R-package was used for visualization [19].

For the t-test analysis, genes that had a Holm-Bonferroni corrected *p*-value of <0.1 were deemed as significant. Cell lines were labeled as sensitive or resistant to a drug of interest as described for each case study. Elastic net regression was performed using *glmnet* and *caret* R packages [20, 21]. AUCs for the respective drugs were regressed on the gene expression of the top 5000 most variably expressed genes. The optimal lambda was selected using ten-fold cross validation on models using different parameters determined by `tuneLength = 20`. Genes with non-zero coefficients were used for enrichment analysis.

Results

Design and conceptualization

We constructed a supervised learning algorithm to nominate biological processes that underlie cancer drug response. Our approach emphasizes prioritization of biologically meaningful features used for classification rather than predictive performance (Fig. 1). We trained our algorithm using only gene expression and drug sensitivity data. We opted to only use gene expression as this data type consistently performed the best as a standalone dataset in a metanalysis of the 44 machine learning algorithms submitted to the NCI-DREAM drug sensitivity prediction challenge [22]. We also favored gene



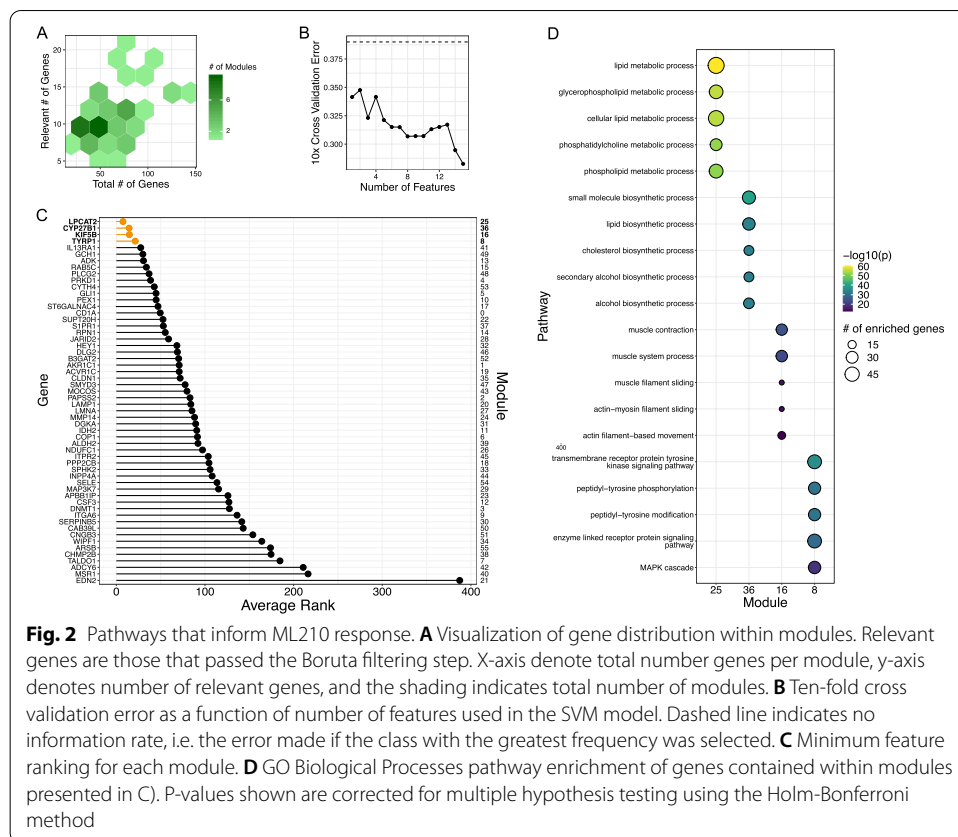
expression as it is known that transcriptomic diversity better explains phenotypic heterogeneity in some cancers, such as cutaneous melanoma [23].

Conceptually, our approach is based on the support vector machine learning algorithm combined with multiple layers of feature selection. Additionally, we use protein–protein interaction data to annotate important features with pathway-level information. Ultimately, our approach returns a ranked list of features, *i.e.* genes, that are grouped into mutually exclusive modules containing closely interacting genes. This strategy enables ranking of known biological processes like pathway enrichment analysis but requires much fewer informative, or differentially expressed, genes.

Case Study 1: Pathways that inform GPX4i sensitivity

ML210 was initially discovered in a high-throughput screening effort as an agent that was selective against HRAS-driven oncogenesis in fibroblasts [24]. However, ML210's mechanism of action was unknown at the time of its discovery. Later, it was found that ML210 kills cells via induction of ferroptosis through inhibition of GPX4 [25, 26]. We applied our approach on all cancer cell lines with gene expression and drug response data to ML210. Pathway activity feature selection returned pathways listed in Additional file 3: Table S1. This selection step retained 2439 genes. Boruta feature selection returned genes that enriched for GO Biological processes in Additional file 4: Table S2. This selection step retained 395 genes across 39 modules. Our method ranked lipid metabolism as the top pathway that determines sensitivity to ML210 (Fig. 2). This result is consistent with the knowledge that the balance of monounsaturated fatty acids (MUFAs) and polyunsaturated fatty acids (PUFAs) determines susceptibility to ferroptosis [27, 28]. As a negative control for the utility of our method, we performed enrichment analysis using genes determined to be significant using t-test or those retained by elastic net (Additional file 5: Table S3, Additional file 6: Table S4). The top results from our approach did not overlap with that of the standard analysis we tried.

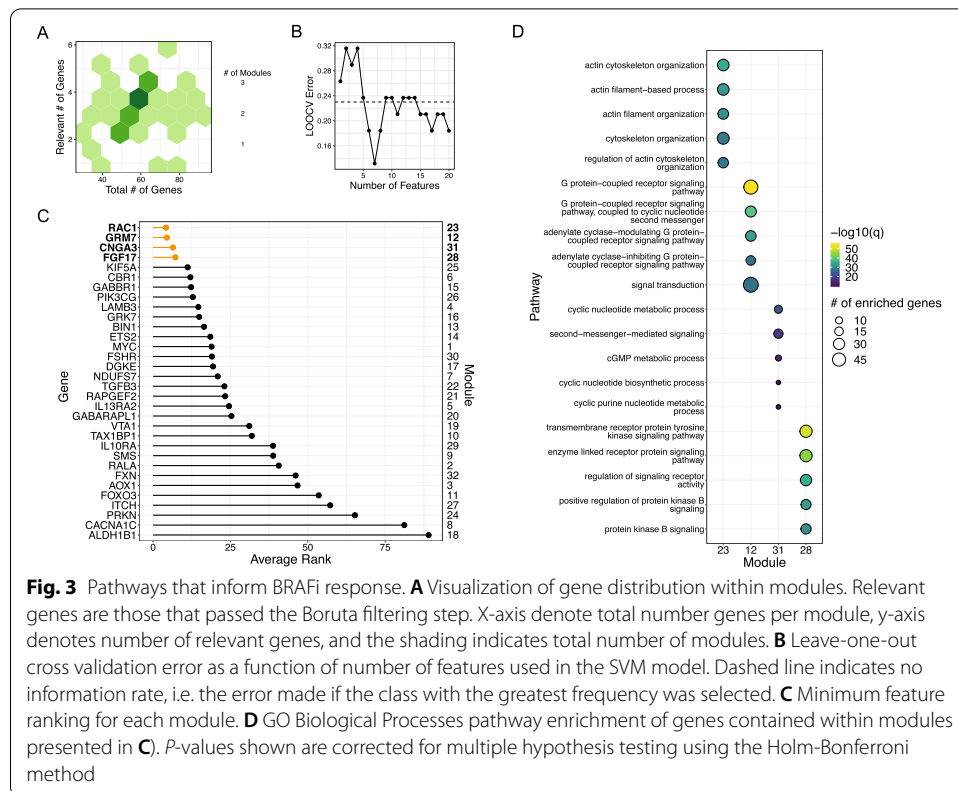
The first step of the proposed approach is to input a set of KEGG pathways. As described in the methods, we performed this analysis using KEGG pathways belonging to Metabolism, Genetic Information Processing, Environmental Information Processing, and Cellular Processes. To test what would happen if all pathways were included, we



repeated the analysis for ML210 using all human KEGG pathways (Additional file 2: Figure S2). Lipid metabolism and actin cytoskeleton pathways remained top candidates, but the other two top pathways changed.

Case Study 2: Pathways that inform BRAFi sensitivity

Next, we tested our approach on selective inhibitors of *BRAF*^{V600E}. We analyzed only cutaneous melanoma cell lines with the *BRAF*^{V600E} mutation, which is present in ~50% of this type of cancer. Even when this mutation is present, drug response to BRAF inhibitors is heterogenous with some melanomas more resistant to BRAF inhibition (BRAFi) than others. Pathway activity feature selection returned pathways listed in Additional file 7: Table S5. This selection step retained 3223 genes. Boruta feature selection returned genes that enriched for GO Biological processes in Additional file 8: Table S6. This selection step retained 169 genes across 36 modules. For the BRAF inhibitors, our method identified Rac1/cytoskeletal signaling as the most salient driver of drug resistance (Fig. 3). As a negative control for the utility of our method, we performed enrichment analysis using genes determined to be significant using t-test or those retained by elastic net (Additional file 9: Table S7, Additional file 10: Table S8). We found that Actin/cytoskeleton processes were highly ranked by our approach but not by the t-test nor elastic net. However, both our approach and the p-value strategy prioritized the “transmembrane receptor protein tyrosine kinase signaling pathway.” This finding is consistent

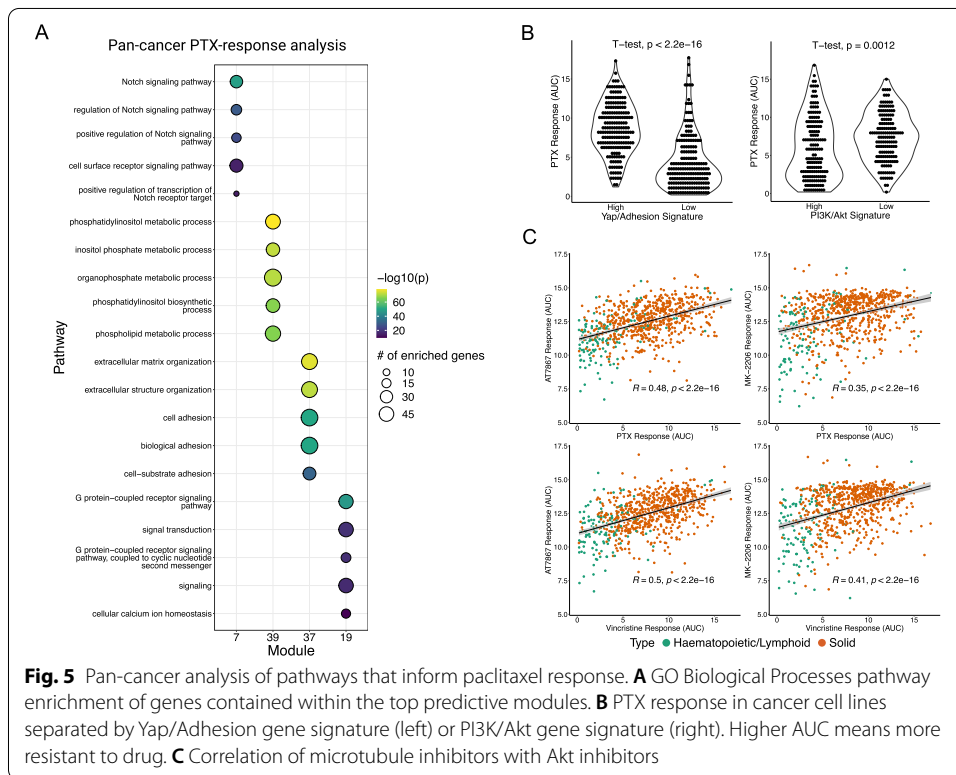
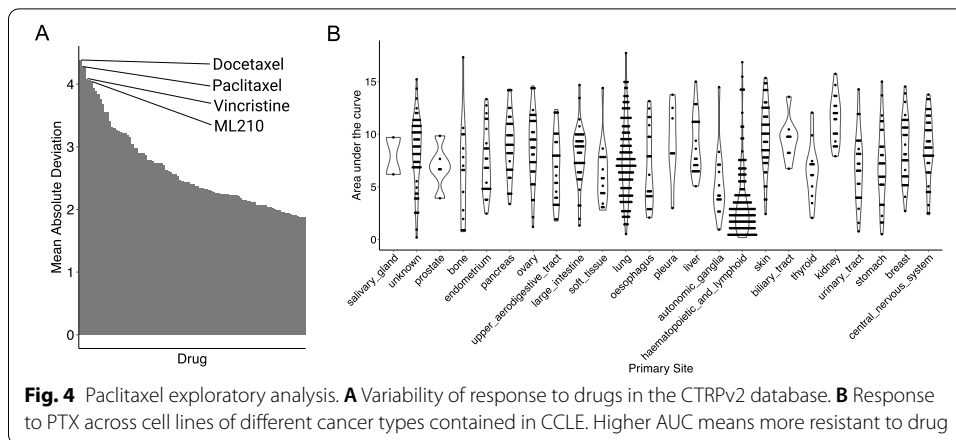


with other studies that report certain RTKs such as PDGFRB and CSF1R drive intrinsic drug resistance to BRAFi in *BRAF*^{V600} cutaneous melanoma [29, 30].

Case Study 3: Pathways that inform sensitivity to anti-tubulin drugs

For our last case study, we wondered if our approach could identify new insights for drugs where the mechanisms of response are less understood. We took an -omics approach and looked for drugs with heterogeneous response. To this end, we ranked drugs in CTRPv2 with respect to the mean absolute deviation of the AUC. In addition to ML210 discussed previously, three anti-tubulin drugs (paclitaxel (PTX), docetaxel, vincristine) were among those with the most variable response (Fig. 4A). Sensitivity to anti-tubulin drugs were highly correlated (Pearson correlation of 0.83 for paclitaxel and vincristine, 0.92 for paclitaxel and docetaxel, and 0.83 for vincristine and docetaxel), suggesting similar mechanisms of action. Pan-cancer analysis of response to paclitaxel shows that hematopoietic cancers are generally more sensitive to microtubule disruption. However, response within cancers of other sites, e.g. lung, ovary, was also heterogeneous (Fig. 4B).

In the era of precision oncology, anti-tubulin drugs are considered “non-targeted”, but unexpectedly we observed that the response to anti-tubulin drugs was highly disparate across different cancer cell lines. This suggests that there may be cancer cell intrinsic features that dictate sensitivity to these drugs. To explain this variation, we applied our analysis approach on PTX. Pathway activity feature selection retained 3232 genes. Boruta feature selection retained 822 genes across 49 modules. Pan-cancer analysis



suggested that Notch, Akt, and adhesion signaling may be involved in PTX-response (Fig. 5A). Notch signaling likely was used as a predictor because Notch is a critical driver of hematopoietic cancers, which happen to be generally sensitive to PTX-inhibition. As a negative control for the utility of our method, we performed enrichment analysis using genes determined to be significant using t-test or those retained by elastic net (Additional file 11: Table S9, Additional file 12: Table S10).

To confirm the relevancy of cell adhesion and Akt signaling, we computed previously published gene signatures for these pathways and tested whether the response to PTX was different between cell lines with high/low cell adhesion or Akt signaling signatures [31, 32]. Cell adhesion signaling is known to be regulated by Yap/TEADs, and in general,

cancer cells can be classified into Yap^{on} or Yap^{off} cancers [33]. Using a gene signature based on genes elevated in Yap^{on} cancers, we found that cancers with low Yap signature was more sensitive to PTX inhibition. Conversely, we found that cancer cell lines that had a high PI3K/AKT signature tended to more sensitive to PTX (Fig. 5B). As there are several targeted inhibitors of Akt, we further investigated the connection between PTX sensitivity and PI3K/AKT signaling by computing the correlation between PTX and Vincristine response with two different pan-Akt inhibitors (AT7867, MK2206). We observed statistically significant correlations between the response to microtubule and Akt inhibitors (Fig. 5C).

Since haemopoietic cancers have unique signaling features, *i.e.* Yap^{off} and Notch^{hi}, and contribute to a large percentage of PTX-sensitive samples, we performed the same analysis wherein we only used solid tumor cell lines. Pathway activity feature selection retained 2667 genes. Boruta feature selection retained 223 genes across 43 modules. Surprisingly, even when we excluded blood cancers, Notch signaling remained a predictor of response to PTX, along with Akt signaling (Fig. 6A). As a negative control for the utility of method we performed enrichment analysis using genes determined to be significant using t-test or those retained by elastic net (Additional file 13: Table S11, Additional file 14: S12). To confirm the connection between Notch and PTX response, we narrowed our focus on ovarian cancer, where PTX remains a standard of care. To get a general view of pathways associated with PTX-resistance, we identified genes expressed in ovarian cancer cell lines that were highly correlated with PTX response (Fig. 6B). Of note, one of these genes was MECOM. The locus at chromosome 3q21 contains MECOM and encodes the MDS1 and EVI1 proteins, under the control of two separate promoters. These proteins have been implicated in leukemia development [34–36].

Recently, it was shown that MECOM interacts with PAX8, a transcription factor that is an oncogene for ovarian and kidney cancers and can serve as an indicator of PAX8 transcriptional activity [37]. To determine the relationship with Notch signaling, we

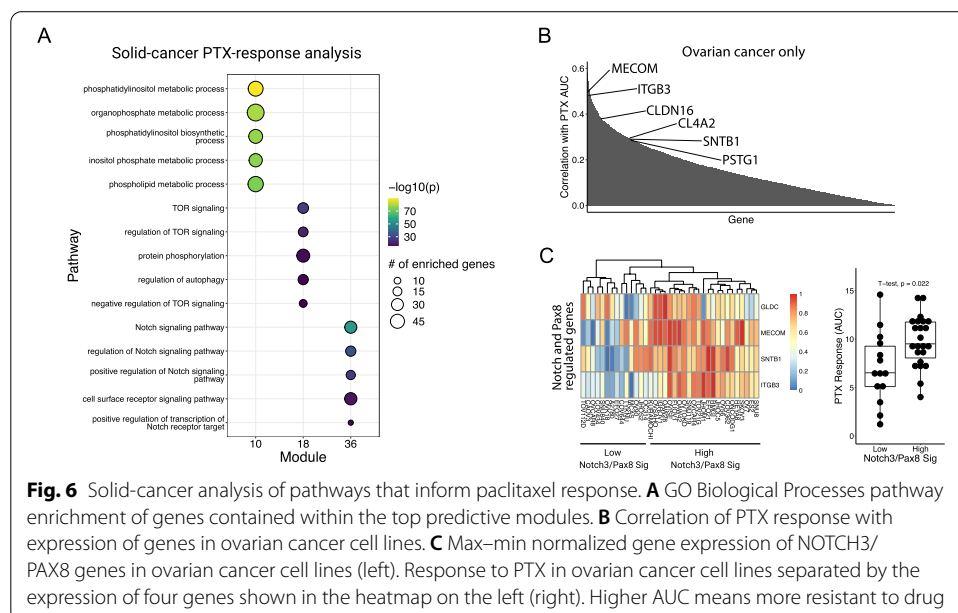


Fig. 6 Solid-cancer analysis of pathways that inform paclitaxel response. **A** GO Biological Processes pathway enrichment of genes contained within the top predictive modules. **B** Correlation of PTX response with expression of genes in ovarian cancer cell lines. **C** Max–min normalized gene expression of NOTCH3/PAX8 genes in ovarian cancer cell lines (left). Response to PTX in ovarian cancer cell lines separated by the expression of four genes shown in the heatmap on the left (right). Higher AUC means more resistant to drug

analyzed a published dataset where NOTCH3 was overexpressed in a murine ovarian surface epithelial cell line [38]. Interestingly, in this model, overexpression of NOTCH3 resulted in a four-fold increase in MECOM. In support of the connection between Notch and PAX8 signaling, we found that other genes positively regulated by NOTCH3 (> four-fold increase upon NOTCH3 overexpression), including NGLDC, SNTB1, and ITGB3, belonged to a the 29-gene PAX8 signature that was reduced upon PAX8 knockdown in multiple human ovarian cancer cell lines [37]. Profiling PTX response using a four-gene signature derived only from the NOTCH3 and PAX8 regulated genes, we observe that ovarian cell lines from CCLE with high NOTCH3/PAX8 transcriptional signature were more resistant to PTX (Fig. 6C). This observation suggests a previously unreported connection between drug resistance to PTX and NOTCH3/PAX8 signaling.

Discussion

Machine learning approaches for modeling cancer drug response have shown promise in predicting cancer drug sensitivity but may not inform biological processes that underlie response. Existing strategies used to reveal this information include pathway enrichment on highly weighted genes prior to the first hidden layer in a deep neural network, that obtained from models such as decision trees, or those with high Shapley values of deep neural networks [9, 39–41]. In this study, we extract biological meaning from a machine learning model by combining multiple layers of feature selection with a ranking process performed through the support vector machine. Furthermore, instead of using all available genes, we only utilize genes that fall within curated pathways and group such genes within interacting modules—sacrificing performance for interpretability. We demonstrate the utility of our approach with three test-cases. For each case, we also confirmed that standard analyses did not prioritize the same pathways that our approach did. Namely, we computed enriched pathways in genes that were differentially expressed between sensitive and drug resistant cell lines using the t-test. We also computed enriched pathways in genes, selected by elastic net, that could best model drug response.

Our knowledge-guided machine learning analysis nominated lipid metabolism as an important biological process that drove sensitivity to ML210. ML210 kills cancer via induction of ferroptosis through covalent interactions with its target, GPX4. Inhibition of GPX4 results in uncontrolled PUFA oxidation leading to ferroptosis [27]. However, there are clear biological determinants of ML210 sensitivity as some cancer cells are exquisitely sensitive while others are ambivalent towards it. Our approach correctly prioritized lipid metabolism as an important determinant of response to GPX4 inhibition. In general, cells with high PUFAs relative to MUFAs are more susceptible to GPX4 inhibition [27, 28]. This trend was also found in the Cancer Cell Line Encyclopedia metabolomics analysis, which demonstrated that the abundance of PUFAs was the most correlated with the genetic dependency on GPX4 [42]. Finally, it is known that some cell lines can protect themselves from lipid ROS by upregulating the lipid saturation pathway [43].

In the context of BRAF inhibition, our approach identified Rac1/cytoskeletal signaling as an important biological process underlying intrinsic drug resistance in cutaneous melanoma with oncogenic BRAF. Rac1 is a Rho family GTPase with diverse signaling properties including cytoskeletal regulation [44]. A mutated version of Rac1, RAC1^{P29S},

is a well-described driver of MAPK inhibitor resistance and metastasis in cutaneous melanoma [45–48]. Nevertheless, the Rac1 signaling axis can also drive resistance to MAPK inhibition [49, 50].

Our analysis of PTX-response suggests that inhibiting Akt-signaling may act synergistically with anti-tubulin drugs—additional analysis confirmed significant correlation between two anti-tubulin drugs and two selective Akt inhibitors. Co-targeting Akt and microtubules has been previously proposed [51–53]. Elevation of Akt signaling has also been shown to be positively correlated with PTX response in patients [54]. Here we provide -omics scale evidence that support this therapeutic strategy and the use of Akt pathway activation as a biomarker for PTX response. Our analysis also led us to a previously unreported connection between NOTCH3/PAX8 signaling and drug resistance to PTX.

Consistent with the finding that PAX8 is associated with PTX-resistance, patients with high PAX8 signature had worse overall survival [37]. Previous studies on PTX-resistance in ovarian cancer has implicated a critical role of cell adhesion in driving drug resistance and cell adhesion [55–57]. High expression of cell adhesion related genes was also identified using machine learning approaches to non-responders in patients. Consistent with these findings, both PAX8/MECOM and Notch regulated genes in ovarian epithelial cells enrich for cell-adhesion related pathways [37, 38]. Lastly, a deep learning algorithm developed by another research group also observed Notch signaling as an important predictor of PTX-response [39].

In summary, we developed a machine-learning approach to mine publicly available cancer pharmaco genomics data to generate hypothesis on biological pathways that underlie drug sensitivity. We tested our approach on inhibitors of GPX4, BRAF, and microtubules. Our approach revealed pathways that are consistent with existing knowledge on drug resistance to GPX4 and BRAF inhibition, and which were not detected by standard analysis methods. Furthermore, our PTX analysis informs future studies aimed to enhance the efficacy of anti-tubulin drugs.

Conclusions

We have developed a machine learning approach to inform the biology underlying cancer drug response. Our approach identified already known biological pathways that contribute to the drug response of ML210 and VEM/Dabrafenib. Our analysis also revealed a potentially novel connection between NOTCH3/PAX8 signaling and PTX drug resistance.

Abbreviations

VEM: Vemurafenib; PTX: Paclitaxel.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-022-04720-z>.

Additional file 1. Fig S1: Dotted line at AUC of 9 was the cutoff used to separate sensitive from resistant cancers.

Additional file 2. Fig S2: A) Minimum feature ranking for each module. **B)** GO Biological Processes pathway enrichment of genes contained within modules presented in **A**. P-values shown are corrected for multiple hypothesis testing using the Holm-Bonferroni method.

Additional file 3. Supplementary Table S1. KEGG pathways that passed pathway activity selection for ML210 analysis.

Additional file 4. Supplementary Table S2. Top 20 enriched GO Biological Processes of genes returned by Boruta. P-values shown are corrected for multiple hypothesis testing using the Holm-Bonferroni method.

Additional file 5. Supplementary Table S3. Top 20 Enriched GO Biological Processes of t-test derived genes for ML210 analysis. P-values shown are corrected for multiple hypothesis testing using the Holm-Bonferroni method.

Additional file 6. Supplementary Table S4. Top 20 Enriched GO Biological Processes enrichment of elastic-net derived genes for ML210 analysis. P-values shown are uncorrected for multiple hypothesis testing.

Additional file 7. Supplementary Table S5. KEGG pathways that passed pathway activity selection for BRAFi analysis.

Additional file 8. Supplementary Table S6. Top 20 GO Biological Processes enriched in important modules for BRAFi analysis. P-values shown are corrected for multiple hypothesis testing using the Holm-Bonferroni method.

Additional file 9. Supplementary Table S7. Top 20 GO Biological Processes enrichment of t-test derived genes for BRAFi analysis. P-values shown are uncorrected for multiple hypothesis testing.

Additional file 10. Supplementary Table S8. Top 20 GO Biological Processes enrichment of elastic-net derived genes for BRAFi analysis. P-values shown are corrected for multiple hypothesis testing using the Holm-Bonferroni method.

Additional file 11. Supplementary Table S9. Top 20 GO Biological Processes enrichment of t-test derived genes for PTX analysis. P-values shown are corrected for multiple hypothesis testing using the Holm-Bonferroni method.

Additional file 12. Supplementary Table S10. Top 20 GO Biological Processes enrichment of elastic-net derived genes for PTX analysis. P-values shown are uncorrected for multiple hypothesis testing.

Additional file 13. Supplementary Table S11. Top 20 GO Biological Processes enrichment of t-test derived genes for PTX analysis without blood cancers. P-values shown are uncorrected for multiple hypothesis testing.

Additional file 14. Supplementary Table S12. Top 20 GO Biological Processes enrichment of elastic-net derived genes for PTX analysis without blood cancers. P-values shown are uncorrected for multiple hypothesis testing.

Acknowledgements

We like to thank Michael Henry and Marion Vanneste for their helpful feedback.

Author contributions

Eliot Zhu and Adam Dupuy conceived the analysis. Eliot Zhu wrote the manuscript and performed the bioinformatic analysis. Both authors read and approved the final manuscript.

Funding

This study was funded by NIH NCI (F30 CA247102), the UIOWA Medical Scientist Training Program, (NIH NIGMS T32 GM007337), The Melanoma Research Foundation Medical Student Award, The American Skin Association Medical Student Award (E.Z.), The Iowa Department of Public Health Melanoma Research Award (A.D.) and the Holden Comprehensive Cancer Center, The University of Iowa (A.D.).

Availability of data and materials

The RMA normalized array gene expression matrix and vemurafenib and dabrafenib drug sensitivities were downloaded from GDSC [58]. Drug sensitivities for ML210 were downloaded from CTRP v2 [59–61]. Yap^{On} genes, namely those in PC1⁺, were obtained from [33]. Akt CMAP pathway signature genes were obtained from [31, 32]. Notch3 overexpression data was obtained from [38]. Gene expression of cancer cell lines for confirmatory analysis was obtained from the Cancer Cell Line Encyclopedia (CCLE) [62]. Lastly, code used to perform the analysis and generate the figures is accessible through Github (https://github.com/eyzhu/cancer_drug_ML_analysis). A guide will be provided to perform analysis on other drugs not assessed here. Requests for data from this study should be directed to Dr. Adam Dupuy (adam-dupuy@uiowa.edu).

Declarations

Ethics and consent to participate

Not applicable.

Consent to publication

Not applicable.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

¹Department of Anatomy and Cell Biology, The University of Iowa, Iowa City, IA, USA. ²Holden Comprehensive Cancer Center, The University of Iowa, Iowa City, IA, USA. ³Cancer Biology Graduate Program, The University of Iowa, Iowa City, IA, USA. ⁴The Medical Scientist Training Program, The University of Iowa, Iowa City, IA, USA.

Received: 15 November 2021 Accepted: 3 May 2022

Published online: 17 May 2022

References

- Dong Z, Zhang N, Li C, Wang H, Fang Y, Wang J, et al. Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer*. 2015;15:489.
- Dorman SN, Baranova K, Knoll JH, Urquhart BL, Mariani G, Carcangiu ML, et al. Genomic signatures for paclitaxel and gemcitabine resistance in breast cancer derived by machine learning. *Mol Oncol*. 2016;10(1):85–100.
- Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS ONE*. 2013;8(4): e61318.
- Daemen A, Griffith OL, Heiser LM, Wang NJ, Enache OM, Sanborn Z, et al. Modeling precision treatment of breast cancer. *Genome Biol*. 2013;14(10):R110.
- Chiu YC, Chen HH, Zhang T, Zhang S, Gorthi A, Wang LJ, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genomics*. 2019;12(Suppl 1):18.
- Gerdes H, Casado P, Dokal A, Hijazi M, Akhtar N, Osuntola R, et al. Drug ranking using machine learning systematically predicts the efficacy of anti-cancer drugs. *Nat Commun*. 2021;12(1):1850.
- Malik V, Kalakoti Y, Sundar D. Deep learning assisted multi-omics integration for survival and drug-response prediction in breast cancer. *BMC Genomics*. 2021;22(1):214.
- Zuo Z, Wang P, Chen X, Tian L, Ge H, Qian D. SWnet: a deep learning model for drug response prediction from cancer genomic signatures and compound chemical structures. *BMC Bioinform*. 2021;22(1):434.
- Liu Q, Hu Z, Jiang R, Zhou M. DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics*. 2020;36(Suppl_2):i911–i8.
- Kim Y, Bismeyer T, Zwart W, Wessels LFA, Vis DJ. Genomic data integration by WON-PARAFAC identifies interpretable factors for predicting drug-sensitivity in vivo. *Nat Commun*. 2019;10(1):5034.
- Cao X, Fan R, Zeng W. DeepDrug: a general graph-based deep learning framework for drug relation prediction. *bioRxiv*. 2020.
- Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward precise disease classification. *PLoS Comput Biol*. 2008;4(11): e1000217.
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13.
- Kursa MB, Rudnicki WR. Feature selection with the Boruta Package. *J Stat Softw*. 2010;36:1–13.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7): e47.
- Bhuva DD, Foroutan M, Xie Y, Lyu R, Cursons J, Davis MJ. Using singscore to predict mutation status in acute myeloid leukemia from transcriptomic signatures. *F1000Res*. 2019;8:776.
- Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. 2009;4(8):1184–91.
- Wickham H. *ggplot2: elegant graphics for data analysis*. Cham: Springer; 2016.
- Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(5).
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
- Costello JC, Heiser LM, Georgii E, Gonen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat Biotechnol*. 2014;32(12):1202–12.
- Hoek KS, Schlegel NC, Brafford P, Sucker A, Ugurel S, Kumar R, et al. Metastatic potential of melanomas defined by specific gene expression profiles with no BRAF signature. *Pigment Cell Res*. 2006;19(4):290–302.
- Weiwer M, Bittker JA, Lewis TA, Shimada K, Yang WS, MacPherson L, et al. Development of small-molecule probes that selectively kill cells induced to express mutant RAS. *Bioorg Med Chem Lett*. 2012;22(4):1822–6.
- Yang WS, SriRamaratnam R, Welsch ME, Shimada K, Skouta R, Viswanathan VS, et al. Regulation of ferroptotic cancer cell death by GPX4. *Cell*. 2014;156(1–2):317–31.
- Dixon SJ, Lemberg KM, Lamprecht MR, Skouta R, Zaitsev EM, Gleason CE, et al. Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell*. 2012;149(5):1060–72.
- Yang WS, Kim KJ, Gaschler MM, Patel M, Shchepinov MS, Stockwell BR. Peroxidation of polyunsaturated fatty acids by lipoxygenases drives ferroptosis. *Proc Natl Acad Sci USA*. 2016;113(34):E4966–75.
- Magtanong L, Ko PJ, To M, Cao JY, Forcina GC, Tarangelo A, et al. Exogenous monounsaturated fatty acids promote a ferroptosis-resistant cell state. *Cell Chem Biol*. 2019;26(3):420–32.
- Giricz O, Mo Y, Dahlman KB, Cotto-Rios XM, Vardabasso C, Nguyen H, et al. The RUNX1/IL-34/CSF-1R axis is an auto-crinally regulated modulator of resistance to BRAF-V600E inhibition in melanoma. *JCI Insight*. 2018;3(14).
- Nazarian R, Shi H, Wang Q, Kong X, Koya RC, Lee H, et al. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature*. 2010;468(7326):973–7.
- Creighton CJ, Fu X, Hennessy BT, Casa AJ, Zhang Y, Gonzalez-Angulo AM, et al. Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. *Breast Cancer Res*. 2010;12(3):R40.
- Zhang Y, Kwok-Shing Ng P, Kucherlapati M, Chen F, Liu Y, Tsang YH, et al. A pan-cancer proteogenomic atlas of PI3K/AKT/mTOR pathway alterations. *Cancer Cell*. 2017;31(6):820–32.

33. Pearson JD, Huang K, Pacal M, McCurdy SR, Lu S, Aubry A, et al. Binary pan-cancer classes with distinct vulnerabilities defined by pro- or anti-cancer YAP/TEAD activity. *Cancer Cell*. 2021;39(8):1115–34.
34. Kustikova O, Fehse B, Modlich U, Yang M, Dullmann J, Kamino K, et al. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science*. 2005;308(5725):1171–4.
35. Ottema S, Mulet-Lazaro R, Beverloo HB, Erpelinck C, van Herk S, van der Helm R, et al. Atypical 3q26/MECOM rearrangements genocopy inv(3)/t(3;3) in acute myeloid leukemia. *Blood*. 2020;136(2):224–34.
36. Fears S, Mathieu C, Zeleznik-Le N, Huang S, Rowley JD, Nucifora G. Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family. *Proc Natl Acad Sci USA*. 1996;93(4):1642–7.
37. Bleu M, Mermet-Meillon F, Apfel V, Barys L, Holzer L, Bachmann Salvy M, et al. PAX8 and MECOM are interaction partners driving ovarian cancer. *Nat Commun*. 2021;12(1):2442.
38. Price JC, Azizi E, Naiche LA, Parvani JG, Shukla P, Kim S, et al. Notch3 signaling promotes tumor cell adhesion and progression in a murine epithelial ovarian cancer model. *PLoS ONE*. 2020;15(6):e0233962.
39. Baptista D, Ferreira PG, Rocha M. Deep learning for drug response prediction in cancer. *Brief Bioinform*. 2021;22(1):360–79.
40. Tang YC, Gottlieb A. Explainable drug sensitivity prediction through cancer pathway enrichment. *Sci Rep*. 2021;11(1):3128.
41. Pham TH, Hagenbeek TJ, Lee HJ, Li J, Rose CM, Lin E, et al. Machine-learning and chemogenomics approach defines and predicts cross-talk of hippo and MAPK pathways. *Cancer Discov*. 2021;11(3):778–93.
42. Li H, Ning S, Ghandi M, Kryukov GV, Gopal S, Deik A, et al. The landscape of cancer cell line metabolism. *Nat Med*. 2019;25(5):850–60.
43. Talebi A, Dehairs J, Rambow F, Rogiers A, Nittner D, Derua R, et al. Sustained SREBP-1-dependent lipogenesis as a key mediator of resistance to BRAF-targeted therapy. *Nat Commun*. 2018;9(1):2500.
44. Marei H, Malliri A. Rac1 in human diseases: the therapeutic potential of targeting Rac1 signaling regulatory mechanisms. *Small GTPases*. 2017;8(3):139–63.
45. Davis MJ, Ha BH, Holman EC, Halaban R, Schlessinger J, Boggon TJ. RAC1P29S is a spontaneously activating cancer-associated GTPase. *Proc Natl Acad Sci U S A*. 2013;110(3):912–7.
46. Kawazu M, Ueno T, Kontani K, Ogita Y, Ando M, Fukumura K, et al. Transforming mutations of RAC guanosine triphosphatases in human cancers. *Proc Natl Acad Sci USA*. 2013;110(8):3029–34.
47. Watson IR, Li L, Cabeceiras PK, Mahdavi M, Gutschner T, Genovese G, et al. The RAC1 P29S hotspot mutation in melanoma confers resistance to pharmacological inhibition of RAF. *Cancer Res*. 2014;74(17):4845–52.
48. Mohan AS, Dean KM, Isogai T, Kasitinin SY, Murali VS, Roudot P, et al. Enhanced dendritic actin network formation in extended lamellipodia drives proliferation in growth-challenged Rac1(P29S) melanoma cells. *Dev Cell*. 2019;49(3):444–60.
49. Feddersen CR, Schillo JL, Varzavand A, Vaughn HR, Wadsworth LS, Voigt AP, et al. Src-dependent DBL family members drive resistance to vemurafenib in human melanoma. *Cancer Res*. 2019;79(19):5074–87.
50. Vanneste M, Feddersen CR, Varzavand A, Zhu EY, Foley T, Zhao L, et al. Functional genomic screening independently identifies CUL3 as a mediator of vemurafenib resistance via Src-Rac1 signaling axis. *Front Oncol*. 2020;10:442.
51. Wu YH, Huang YF, Chen CC, Huang CY, Chou CY. Comparing PI3K/Akt inhibitors used in ovarian cancer treatment. *Front Pharmacol*. 2020;11:206.
52. Kim SH, Juhn YS, Song YS. Akt involvement in paclitaxel chemoresistance of human ovarian cancer cells. *Ann N Y Acad Sci*. 2007;1095:82–9.
53. Lin YH, Chen BY, Lai WT, Wu SF, Guh JH, Cheng AL, et al. The Akt inhibitor MK-2206 enhances the cytotoxicity of paclitaxel (Taxol) and cisplatin in ovarian cancer cells. *Naunyn Schmiedebergs Arch Pharmacol*. 2015;388(1):19–31.
54. Yang SX, Costantino JP, Kim C, Mamounas EP, Nguyen D, Jeong JH, et al. Akt phosphorylation at Ser473 predicts benefit of paclitaxel chemotherapy in node-positive breast cancer. *J Clin Oncol*. 2010;28(18):2974–81.
55. Tumbarello DA, Temple J, Brenton JD. ss3 integrin modulates transforming growth factor beta induced (TGFβ) function and paclitaxel response in ovarian cancer cells. *Mol Cancer*. 2012;11:36.
56. Tumbarello DA, Andrews MR, Brenton JD. SPARC regulates transforming growth factor beta induced (TGFβ) extracellular matrix deposition and paclitaxel response in ovarian cancer cells. *PLoS ONE*. 2016;11(9):e0162698.
57. Ahmed AA, Mills AD, Ibrahim AE, Temple J, Blenkiron C, Vias M, et al. The extracellular matrix protein TGFβ1 induces microtubule stabilization and sensitizes ovarian cancers to paclitaxel. *Cancer Cell*. 2007;12(6):514–27.
58. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*. 2013;41(Database issue):D955–61.
59. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*. 2013;154(5):1151–61.
60. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov*. 2015;5(11):1210–23.
61. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol*. 2016;12(2):109–16.
62. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.