Journal of Oncology Research
An open access journal
ISSN: 2637-6148
eSciRes
JOR-4-111
Research Article

ELAFT: An Ensemble-based Machine-learning Algorithm that Predicts Anti-cancer Drug Responses with High Accuracy

Lanka J[1], Housley SN[1,2], Benigno BB[1] and McDonald JF[1,2*]

[1]*Ovarian Cancer Institute, USA*

[2]*School of Biological Sciences, Georgia Institute of Technology, USA*

Abstract

The goal of precision cancer medicine is to optimize clinical efficacy by customizing therapies to individual patients. Therapies derived from genomic profiles of patient tumors are one such method. Machine-learning algorithms have emerged as useful tools to accomplish this goal, but predictive performance of single-model approaches remain inadequate. Here, we report the development of an ensemble-based machine-learning algorithm to predict clinical anti-cancer drug responses. Predictive models were developed for fifteen distinct cancer types incorporating data from 499 independent cell lines. Models were validated against a clinical dataset containing seven chemotherapeutic drugs, administered either singularly or in combination, to 23 ovarian cancer patients. We find an overall predictive accuracy of 91% with a precision of 89% and recall of 100%.

Keywords: Personalized medicine; Drug prediction; Machine learning; Ovarian cancer

Introduction

Precision cancer medicine is driven to match individual patients with treatment(s) that have the highest probability of decreasing tumor size or eradicating the patient's cancer [1]. Models designed to predict an optimal response rate are frequently derived from genomic profiles of patient tumors. These data include somatic alterations, such as, point mutations, deletions, amplifications, translocations, and chromosomal abnormalities that assist in matching a drug's capabilities to a specific molecular profile for a patient's benefit.

Despite the potential value, the complexity of genomic and drug sensitivity data poses challenges to identifying robust statistical relationships between genes and drugs. To overcome these impediments, a variety of computational approaches have been developed [2-12]. Among the most prominent, are machine learning (ML) algorithms that attempt to model regular, informative features of the data, affording a high-level abstraction of information. The majority of predictive modeling approaches rely on single algorithmic methods and each has distinct advantages and degeneracies typically resulting in accuracies of between 70-85%. One strategy to potentially improve the performance of predictive models is to combine multiple machine-learning methods (e.g. [13]). In the present study, we report on the development and validation of an ensemble-based machine-learning algorithm (ELAFT) that combines the strengths of a number of diverse ML approaches (*i.e.*, support vector machines, random forest classifier, K-nearest neighbor classifier, and logistic regression classifier) with a correlation-based feature selection approach (feature titration) to predict anti-cancer drug response with high accuracy (>90%).

Materials and Methods

National Cancer Institute (NCI) 60 Datasets

The National Cancer Institute panel of 60 human cancer cell lines (NCI-60) includes cell lines from 9 different types of cancers (lung, skin, renal, colon, ovarian, breast, leukemia, central nervous system (CNS), and prostrate) was used in our study (Figure 1a). The microarray gene expression data for these cell lines were obtained from the National Center for Biotechnology Gene Expression Omnibus [14] (Supplementary Table 1) and drug sensitivity profiles for seven drugs (carboplatin, paclitaxel, docetaxel, cisplatin, gefitinib, doxorubicin, and gemcitabine) were obtained from National Cancer Institute [15] (Supplementary Table 1, Figure 1b). The preprocessing of the gene expression data was performed as previously described [16]. The probe level expression data were used as features to build the predictive models.

The proteomics data for NCI-60 cell lines came from a recent study [17]. The dataset consisted of over 3,100 SwissProt proteotypic proteins and their systematic quantification of pathway activities obtained using pressure cycling technology and SWATH mass spectrometry (Supplementary Table 1). The quantifications of these proteins were used as features to build models.

The metabolomics data for NCI-60 were obtained from a recent study [18]. The data consisted of intracellular metabolic profiles that were used as features to build the model (Supplementary Table 1). The metabolomics dataset consisted of metabolic profiles for all the NCI-60 cancer types except for Leukemia. Thus, for assessments where the metabolomic dataset was involved, only 53 cell lines were used rather than 59 used in all gene expression and proteomics evaluations.

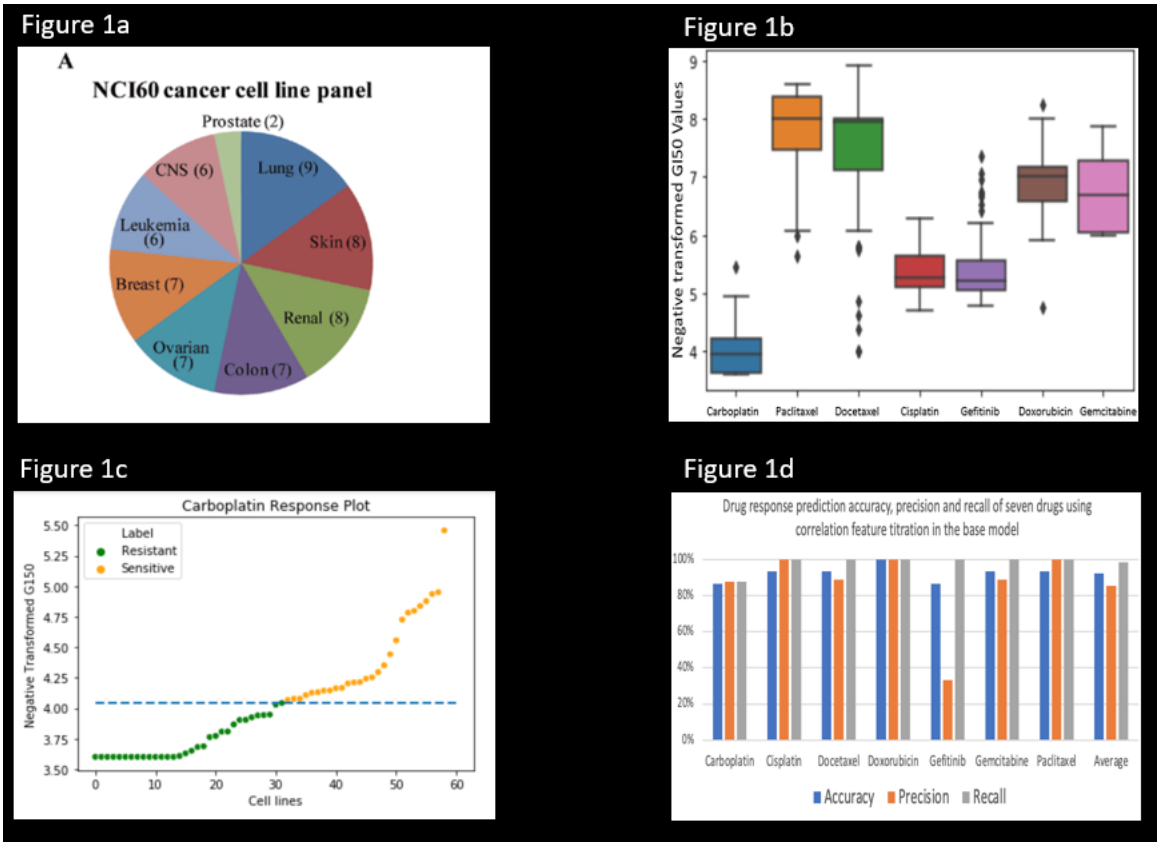

Figure (1a): National Cancer Institute (NCI)-60 Cell line panel. The panel consists of cell lines from nine cancer types (lung, skin, renal, colon, ovarian, breast, leukemia, central nervous system (CNS), and prostrate). The number of cell lines per each cancer type is indicated in the brackets ranging from two prostrate cancer cell lines to nine lung cancer cell lines; **(1b):** A box plot of the GI50 (Growth inhibitory 50%) negatively transformed response rate (y-axis) for each drug (x-axis) on NCI-60 cell lines; **(1c):** Graph showing carboplatin drug response curve on NCI-60 cell lines. The GI50 (Growth inhibitory 50%) negatively transformed response rate for the cell lines are shown on the y-axis. The plot shows the "two-class" labeling of the drug response. In this case, the cell lines with drug response above mean value are considered as "sensitive" to the drug (orange dots) and those below the mean are "resistant" to the drug (green dots). The mean value is shown by the horizontal blue dotted line; **(1d):** The prediction accuracy, precision and recall of ELAFT models on NCI-60 cell lines using gene-expression data. Models are generated for each of the seven drugs and the blue, orange and gray bars represent accuracy, precision and recall of each of these models applied on gene-expression dataset. The prediction accuracy levels of these models ranged from 86-100% with an average of 92%. Precision levels ranged from 33-100% with an average of 85.5% and recall ranged from 87-100% with an average of 98%.

Cancer Cell Line Encyclopedia (CCLE) and DepMap datasets

CCLE and DepMap datasets were downloaded from the DepMap database [19]. We used microarray datasets to allow adequate comparison to the ovarian cancer datasets that were analyzed later in the human patient studies (see below). Two datasets were obtained: one consisting of gene-expression data for 1037 cell lines from nineteen different cancer types, the other consisting of drug sensitivity profiles of 578 cell lines for eight drugs: carboplatin, paclitaxel, docetaxel, cisplatin, gefitinib, doxorubicin, gemcitabine and topotecan. For the gene expression dataset, DepMap provides RNA seq and microarray expression datasets. Not all the cell lines that had gene-expression data had drug sensitivity profiles and vice versa. Matching for both and filtering out cancer types that had less than ten cell lines resulted in 499 cell lines representing 15 different cancer types (skin, bladder, bone, brain, breast, colon, uterine, esophageal, gastric, head and neck, kidney, liver, lung, ovarian, pancreatic; Supplementary Table 2). The preprocessing of gene expression data was performed as previously described [16].

Ovarian Cancer patient data

Gene expression and drug response data for ovarian cancer (serous papillary) patients were obtained from the Ovarian Cancer Institute. Informed patient consents were obtained and approved under appropriate Georgia Institute of Technology Institutional Review Board protocol (H14337). Samples of primary tumors collected from 23 ovarian cancer patients at Northside Hospital (Atlanta) were snap frozen in liquid nitrogen within one minute of surgical removal and transferred to the lab for laser capture microdissection of cancer cells and subsequent microarray gene-expression analysis (Affymetrix, U133Plus 2.0 arrays, ThermoFisher

Scientific) as previously described [20,21]. These microarray datasets are available at National Center for Biotechnology Gene Expression Omnibus (GSE38666, GSE112798).

Gene- expression microarray (.CEL) files were normalized one by one against the DepMap gene-expression microarray data using standard quantile normalization [22,23] and using the mean of each probe. Patient responses to administered chemotherapies were monitored by measurement of CA-125 values prior to and after the treatment. Patients were characterized as responsive to treatments if their respective CA-125 values dropped below normal values (<21) (Supplementary Table 3)

Labeling of the drug response data

The drug response data (GI50 – Growth Inhibitory 50%) for seven drugs for NCI-60 cell lines and eight drugs for DepMap cell lines (Supplementary Tables 1 and 2) were transformed such that higher the number the more sensitive the cell line was to the drug administered. To assign a drug response label to the cell lines, three different methods were employed. The first method consisted of assigning two labels (sensitive and resistant, two-class classification). Cell lines with drug responses above the mean were designated as sensitive while those below the mean were designated as resistant (Supplementary Figure 1a, 1b). A second more stringent method again assigned two labels (sensitive and resistant) but cell lines displaying GI50 between 0.5 standard deviation above and below the mean were excluded from the training set but included in the test set (two-class separated labeling). The third method, a 3-class method, included the cell lines with GI50 values between the 0.50 standard deviation above and below the mean but labeled these as neutral class. All three labeling methods were separately evaluated to build and identify the optimized models (Supplementary Figure 2).

Feature selection for machine learning algorithm

Three different feature selection methods were evaluated. The first included all gene expression probes (27920 for NCI-60), or all the protein levels (3171 for NCI-60), or all the metabolic values (2182 for NCI-60) as features, A second method involved selecting features based on correlation and feature titration. A third method applied recursive feature elimination in the selection of a set of most informative features as previously described [24]. The correlation-based feature selection method was found to generate the highest accuracy and was utilized in the final method (Supplementary Figure 3).

Recursive Feature Elimination based feature selection was built as below

(1) *Initialization*

- Training cell lines: $X_0 = [x_1, x_2, \ldots, x_m]^T$
- Drug response class label: $y = [y_1, y_2, \ldots, y_m]^T$
- Current feature set: s = [1,2,…, n]
- Feature ranked list: *r* = [].

(2) *Feature ranking and training*

- Repeat the following steps until *s* = [].
- Create training data matrix with good features: $X = X_0(:, s)$.
- Train classifier
  - α = LR-train (X, y), KNN-train(X, y), RF-train(X, y),SVM-train(X, y)
- Compute feature weight: $w = \sum_k \alpha_k y_k x_k$.
- Compute feature ranking criteria: $c_i = (w_i)^2$.
- For all i find features with minimum rank: *f* = argmin(c)
- Update feature ranked list: r = [*s*(f), *r*].
- Remove features with smallest ranking criteria: *s* = *s*(1 : −1, *f* + 1 : length(s))

(3) *Feature selection and prediction:*

- Feature ranked List *r*.
- In each loop, the feature(s) with minimum ranking criteria (c_i) will be removed. The classifier (e.g., logistic regression, support vector machine, etc.) then retrains on the remaining features to obtain the new feature ranking. By training the classifier with these ranked feature subsets, evaluating these for prediction accuracy, retraining to get new feature ranked list, an optimal feature subset that optimizes prediction accuracy was obtained.

Correlation feature titration based feature selection was built as below

(1) *Initialization*

- Training cell lines: $X_0 = [x_1, x_2, \ldots, x_m]^T$
- Drug response class label: $y = [y_1, y_2, \ldots, y_m]^T$
- Current feature set: *s* = [*1,2,…, n*]
- Feature ranked list: *r* = [].

(2) *Feature ranking and training*

- Repeat the following steps until *s* = [].
- For all i, in *s* = [1,2,…, *n*]
  - Compute Pearson correlation coefficient
  - $r = \left[\frac{\mathrm{cov}(xy)}{\sigma_x \sigma_y}\right]^T$
  - Compute feature ranking criteria: $c_i = \mathrm{abs}(r_i)$
- Find features with maximum rank: f = argmax(c)
- Add features with max ranking criteria: *r* = *r*(1 : +1, f - 1 : length(*r*))
- Update feature ranked list: r = [*s*(*f*), *r*]
- Create training data matrix with good features: $X = X_0(:, r)$.
- Train classifier
  - α = LR-train (X, y), KNN-train (X, y), RF-train(X, y),SVM-train(X, y)

(3) *Feature selection and prediction:*

- Feature ranked List *r*.
- In each loop, the feature(s) with maximum ranking criteria (c_i) will be added. The classifier (e.g., logistic regression, support vector machine etc) then retrains on these features to obtain the new prediction accuracy. By training the classifier with these ranked feature subsets, evaluating these for prediction accuracy, an optimal feature subset that optimizes prediction accuracy was obtained.

Machine-Learning algorithm(s)

Five different machine learning algorithms, Support Vector Machines (SVM), Random Forest classifier (RF), K-Nearest Neighbor classifier (KNN), Logistic Regression classifier (LR) and ensemble (combined predictions of SVM, RF, KNN, and LR) classifier, were tested. The algorithms are implemented with default parameters in the python scikit-learn package (https://scikit-learn.org/stable/). Radial basis kernel with default settings was used for SVM. An ensemble classifier that was constructed as a collection of the above four classifiers working as a single classifier by taking the predicted probabilities of the above four classifiers and combining them to determine the final prediction for the test sample. Of these, the ensemble classifier provided more robust results (Supplementary Figure 4).

$$y = argmaxij{=}1mwj\ pij$$

- p is predicted probability of the classifier (SVM, RFC, KNN, LGR)
- wj is the weight assigned to the jth classifier, uniform weights are used
- i is class label (e.g., two class label, sensitive or resistant)

Validation Methods

Leave-one-out-cross validation (LOOCV) was used to validate model performance and has been described previously [25]. Briefly, LOOCV trains the model on all cell lines except one, and then tests it on the one left [25]. It then repeats this for all the cell lines finally generating accuracy, precision and recall metrics.

Success Metrics

Model accuracy, precision, and recall were used to evaluate performance. Accuracy is defined as the fraction of cancer cell lines the algorithm correctly predicted the drug response (sensitive or resistant) compared to their actual response observed in their drug sensitivity research studies. Precision is defined as the fraction of the cell lines the model accurately predicted as sensitive to the drug response over all cell lines it predicted as sensitive. Recall is defined as the fraction of the cell lines model accurately predicted as sensitive to drug response among all the cell lines that were sensitive (Figure 2).

$$\textbf{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \text{ or } \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

$$\textbf{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \text{ or } \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

$$\textbf{Accuracy} = \frac{\text{True Positive + True Negative}}{\text{Total}}$$

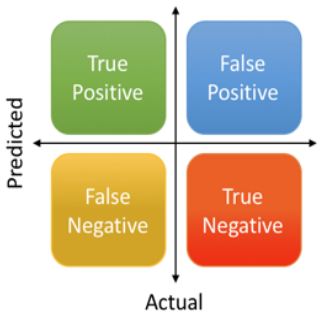

Figure 2: The success criteria of precision, recall and accuracy were employed to measure the performance of the ELAFT models built on NCI-60, DepMap and ovarian cancer datasets. The figure shows the formulas used for calculating these metrics. In each of the model assessments, the predicted responses were compared to the observed responses and assigned a prediction category of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) with TP being both observed and predicted response are responsive/sensitive, TN being both observed and predicted response are nonresponsive/resistant, FP being observed response is non responsive/resistant and predicted response is responsive/sensitive, FN being observed response is responsive/sensitive and predicted response is non responsive/resistant).

Final Design

The final design consisted of five steps: cancer type (the data used to build models), algorithms (the machine-learning algorithms evaluated), features (the features used in machine-learning algorithms), validation (the validation techniques employed) and the final step of the success metrics used for comparisons to obtain the optimized algorithm. At each step, different alternate methods (explained in the above sections) and the ones selected at each step through running multiple simulations (i.e., Cancer-specific dataset, ensemble algorithm, correlation based feature selection, LOOCV, accuracy/precision and recall) went into the final optimized algorithm (Figure 3a).

More than two million simulations were run to optimize the different components of the machine-learning algorithm to develop the final ELAFT model. ELAFT models were then used to predict the response of ovarian cancer patients (Figure

3b) based on their genetic profile to eight different drugs from which an effective drug from which that provided the optimal response is obtained together which patients response to each of the eight drugs.

Figure 3a

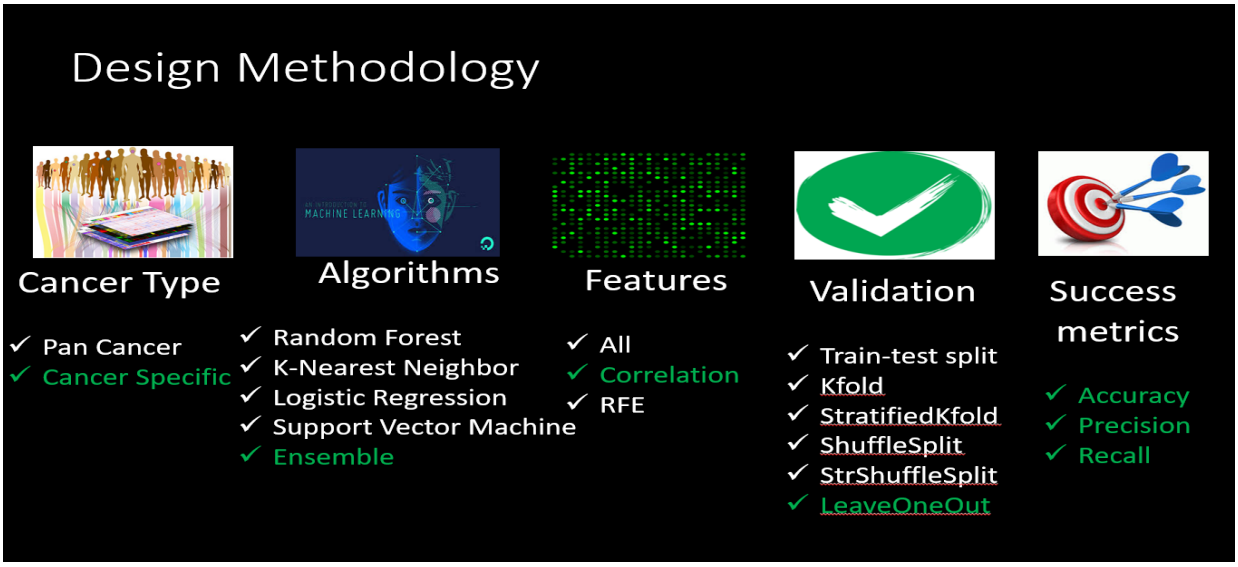

Figure 3b

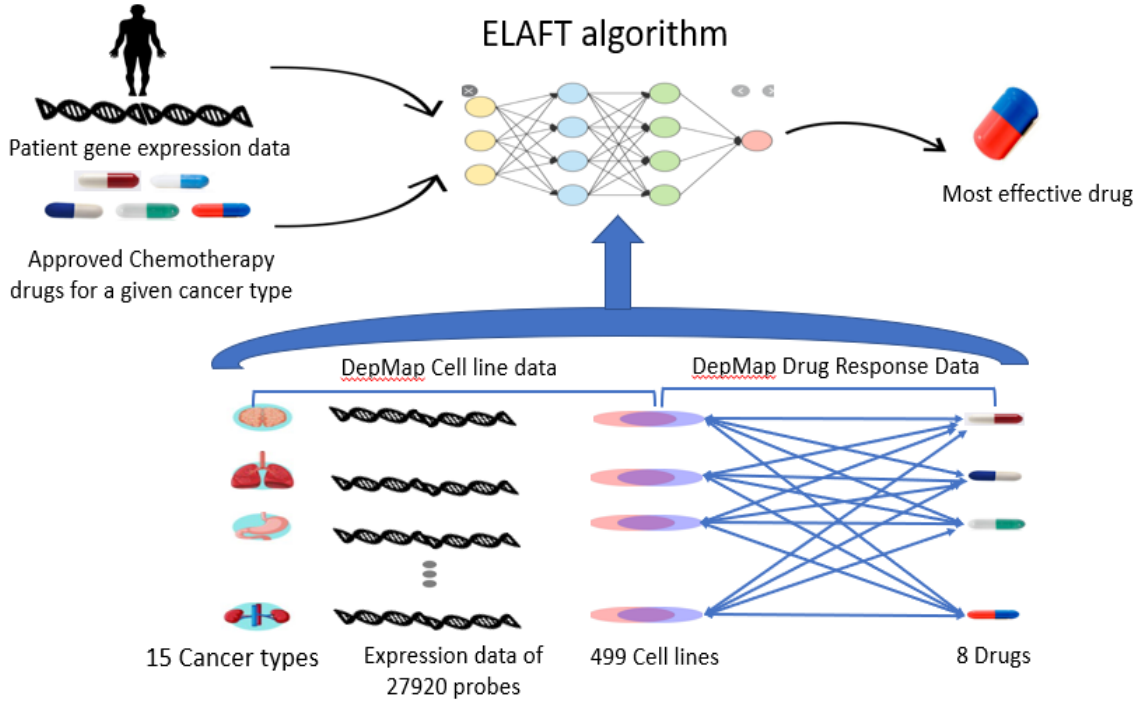

Figure (3a): Summary of methodology. The design consisted of five steps: Cancer type (the data used to build models), Algorithms (the machine learning algorithms evaluated), Features (the features used in building machine learning algorithms), Validation (the validation techniques employed) and the final step of the success metrics used to obtain the optimized algorithm. As can be seen in the figure, at each step different alternate methods were employed during the redesign and the final optimized algorithm is the one built based on the "green" text in each of the steps. More than two million simulations were run to optimize the different components of the machine-learning algorithm to develop the final algorithm; **(3b):** Overview of the ELAFT algorithm and how it enables personalized cancer medicine. The models for 15 cancer types are built using gene-expression data from 27920 probes from 499 DepMap cell lines combined with the response profiles of these cell lines to eight different chemotherapy drugs. These models are used to predict the response of a patient based on their genetic profile to eight different drugs from which an effective drug providing the optimal response is identified.

Results

ELAFT predicts cancer cell line response to drugs with high accuracy

Our initial models were built using gene expression and drug sensitivity profiles across nine cancer types (lung, skin, renal, colon, ovarian, breast, leukemia, prostate, and central nervous system (CNS)) represented in the National Cancer Institute-60 (NCI-60) panel of human cancer cell lines (Figure 1a).

In this initial study, predictive models were generated for each of seven drugs (carboplatin, cisplatin, paclitaxel, docetaxel, gemcitabine, doxorubicin, gefitinib). Cell lines with negatively transformed GI50 (Growth Inhibitory 50%) values (Figure 1b) above the mean were designated as drug sensitive while those below the mean, as drug resistant (Figure 1c). In all cases, cell lines used to build the models were distinct from those used in testing the models. Details on the development and optimization of the algorithm are presented in Materials & Methods.

We began by generating pan-cancer models utilizing gene-expression values and drug response profiles for all 59 cell lines comprising the NCI-60 panel. Individual models were built for each of the seven drugs and the predictive accuracies determined by leave-one-out-cross validation (LOOCV). Accuracies were found to vary across the seven drugs ranging from 100% for doxorubicin to 87% for carboplatin and gefitinib (Figure 1d), with an average accuracy across all drugs of 92%. On average, recall was higher (98%) than precision (86%) (Table 1).

Table 1: A table summarizing the results of ELAFT models on the gene-expression dataset on NCI-60 cell lines. Each row is a drug and the columns represent prediction performance levels of each of these models measured by "accuracy", "precision" and "recall" values. As can be seen an average accuracy of 92% is achieved.

Drug	Accuracy	Precision	Recall
Carboplatin	86.67%	87.50%	87.50%
Cisplatin	93.33%	100.00%	100.00%
Docetaxel	93.33%	88.89%	100.00%
Doxorubicin	100.00%	100.00%	100.00%
Gefitinib	86.67%	33.33%	100.00%
Gemcitabine	93.33%	88.89%	100.00%
Paclitaxel	93.33%	100.00%	100.00%
Average	92.38%	85.52%	98.21%

The predictive accuracy of the responsiveness of NCI-60 cells to drugs is dependent upon the type of data (proteomic, metabolomic, or gene expression) used in model building

Although most ML-based predictive models currently employ gene-expression (e.g., microarray, RNA-seq) data in model building, a variety of high-throughput technologies have and continue to be developed. Since, in addition to gene-expression data, both proteomic [17] and metabolic [18] profiling data are available for the NCI-60 cell lines, we were interested in exploring the relative predictive accuracies of models built using these alternate omic datasets. All model building and evaluation methods were as described using gene-expression data.

As was the case for models built using the gene-expression data, accuracies were found to vary across the seven drugs for models built with both the proteomic and metabolomic data (Figures 4a and 4b). Average accuracy remained highest for models built using the gene-expression data (92%) followed by the metabolic data (87%) and the proteomic data (82%) (Figure 4c). While Recall was, on average, slightly higher (78%) than Precision (76%) for models built using the proteomic data, the opposite was true for the metabolomic data where Precision was significantly higher (97%) than Recall (84%) (Figure 4d). Indeed, Precision was 100% for all drug models built using the metabolic data with the only exception being doxorubicin.

Tumor-specific ELAFT models predict drug response of cancer cell lines with high accuracy

While the NCI-60 dataset is useful in exploring the potential impact of different types of omic data on the predictive accuracy of ML-based models, its overall utility is limited due to the relatively small number of cell lines and cancer types represented in the panel. For this reason, we elected to further evaluate our model using the Cancer Cell

Line Encyclopedia's (CCLE) DepMap dataset comprised of 499 cell lines representing 15 cancer types (bladder, bone, brain, breast, colon/colorectal, endometrial/uterine, esophageal, gastric, head/neck, kidney, liver, lung, ovarian, pancreatic, skin). Because of the large number of cell line data available for each cancer, we were able to build individual predictive models for each of eight drugs and 15 cancers. Predictive accuracies range from 73% (gemcitabine for lung cancer) to 100% (paclitaxel and cisplatin for kidney cancer; gemcitabine for bone cancer; topotecan for esophageal cancer) with an overall average accuracy of 88% for all eight drugs across the 15 cancers (Table 2 and Supplementary Table 4).

Overall precision across all drugs and cancer types (92%) was slightly higher than recall (88%), with many cancers displaying precisions of 100% across several drugs. Recall values also varied across drugs and cancers ranging from 78% to 100% (Table 2 and Supplementary Table 4).

In contrast to the overall high predictive accuracies associated with the cancer-specific models, models built using the entire combined DepMap dataset (pan-cancer model), resulted in relatively low overall predictive accuracy (66%), precision (67%) and recall (68%) for all drugs (Table 3). These values are considerably lower than what we observed using the NCI-60 dataset (see above) possibly attributable to the increased diversity of cancer types comprising the DepMap dataset.

ELAFT models predict response of individual ovarian cancer (OC) patients to standard-of-care therapies with high accuracy

To assess the potential clinical utility of the ELAFT model, we utilized a human ovarian cancer dataset composed of gene-expression levels of 23 individual patient tumors combined with patient responses to chemotherapeutic treatments. We generated predicted responses to each of eight drugs previously employed in the treatment of OC using models developed using the DepMap dataset (see above). Standard-of-care chemotherapy for OC patients typically involves treatment with multiple drugs, most commonly, carboplatin and paclitaxel. In only one instance (Patient 992, Table 4) was a patient in our dataset treated with a single drug (Topotecan) allowing for straight forward testing of our prediction (patient predicted and observed to not respond to Topotecan treatment- True Negative (TN)).

Figure 4a

Drug	Accuracy	PPV_Precision	Sensitivity_Recall
Carboplatin	80.00%	100.00%	77.78%
Cisplatin	86.67%	100.00%	80.00%
Docetaxel	100.00%	100.00%	100.00%
Doxorubicin	80.00%	81.82%	90.00%
Gefitinib	86.67%	100.00%	50.00%
Gemcitabine	80.00%	100.00%	88.89%
Paclitaxel	93.33%	100.00%	100.00%
Average	86.67%	97.40%	83.81%

Figure 4b

Drug	Accuracy	PPV_Precision	Sensitivity_Recall
Carboplatin	78.57%	100.00%	85.71%
Cisplatin	78.57%	100.00%	62.50%
Docetaxel	78.57%	75.00%	100.00%
Doxorubicin	85.71%	85.71%	100.00%
Gefitinib	71.43%	0.00%	0.00%
Gemcitabine	85.71%	80.00%	100.00%
Paclitaxel	92.86%	90.00%	100.00%
Average	81.63%	75.82%	78.32%

Figure 4c

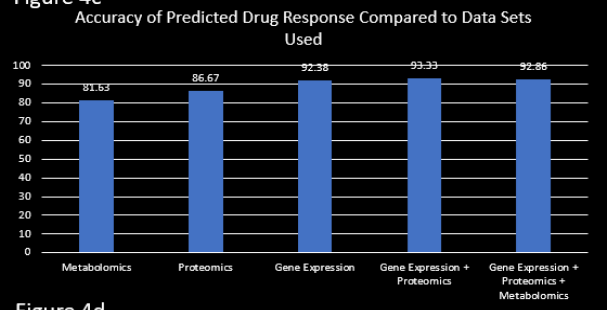

Figure 4d

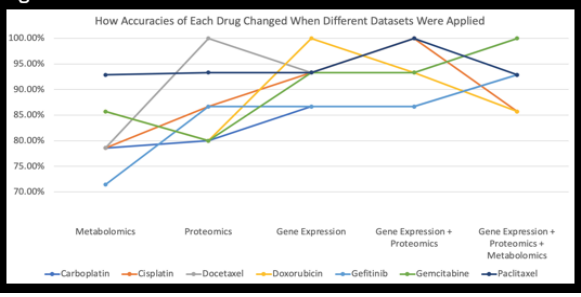

Figure (4): The results of ELAFT models on the proteomics dataset on NCI-60 cell lines. **(4a):** Each row is a drug and the columns represent prediction performance levels of each of these models measured by "accuracy", "precision" and "recall" values. The prediction accuracy levels of these models ranged from 80-100% with an average of 86.7%. Precision levels ranged from 82-100% with an average of 97.4% and recall ranged from 50-100% with an average of 83.8%; **(4b):** The results of ELAFT models on the metabolomics dataset on NCI-60 cell lines. Each row is a drug and the columns represent prediction performance levels of each of these models measured by "accuracy", "precision" and "recall" values. The prediction accuracy levels of these models ranged from 71-93% with an average of 81.6%. Precision levels ranged from 0-100% with an average of 75.8% and recall ranged from 0-100% with an average of 78.32%; **(4C):** Graph showing the prediction accuracy (%) comparisons of ELAFT models built using proteomics, metabolomics, gene expression, gene expression + proteomics, gene expression + proteomics + metabolomics as features in the models. Gene expression features provided better prediction accuracy (92%) followed by proteomics (87%) followed by metabolomics (82%). Combining gene expression and proteomics features together improved prediction accuracy marginally; **(4d):** Graph showing the prediction accuracy comparisons of ELAFT models for each of the drugs built using proteomics, metabolomics, gene expression, gene expression + proteomics, gene expression + proteomics + metabolomics as features in the models. Even at a individual drug level, Gene-expression features provided better prediction accuracy (e.g., carboplatin 87%) followed by proteomics (e.g., carboplatin 80%) followed by metabolomics (e.g., carboplatin 79%). Combining gene expression and proteomics features together improved prediction accuracy marginally.

Table 2: The results of ELAFT models on DepMap cell lines using the gene-expression dataset as features. These results are for paclitaxel drug response with models built with data from specific cancer cell line types (Individual cancer model). Each row is a cancer and the columns represent prediction performance levels of each of these models measured by "accuracy", "precision" and "recall" values. As the results show, the models built on specific cancer types are more than 20% accurate over models built on all cancer types (compare to results in Table 3). The prediction accuracy levels of these cancer specific type models ranged from 80-100% with an average of 88%. Precision levels ranged from 80-100% with an average of 92% and recall ranged from 78-100% with an average of 88%. The prediction performance metrics for other drugs are provided in Supplementary Table 4.

Cancer Type	Accuracy	Precision	Recall
Bladder Cancer	83.33	83.33	92.31
Bone Cancer	93.33	100.00	85.71
Brain Cancer	82.50	88.24	85.00
Breast Cancer	92.31	92.86	92.86
Colon/Colorectal Cancer	82.86	86.67	88.24
Endometrial/Uterine Cancer	95.45	100.00	90.91
Esophageal Cancer	86.96	90.00	81.82
Gastric Cancer	90.48	100.00	77.78
Head and Neck Cancer	88.89	85.71	92.31
Kidney Cancer	100.00	100.00	100.00
Liver Cancer	85.71	91.67	92.31
Lung Cancer	79.63	80.36	82.14
Ovarian Cancer	86.84	90.48	86.36
Pancreatic Cancer	86.49	93.33	83.33
Skin Cancer	88.37	93.33	84.21
Average	**88.21**	**91.73**	**87.69**

Table 3: The results of ELAFT models on DepMap cell lines using the gene-expression dataset as features. These results are for drug response with models built with data from all cancer cell line types (pan-cancer model). Each row is a drug and the columns represent prediction performance levels of each of these models measured by "accuracy", "precision" and "recall" values. As the results show, the models built on all cancer types is 20% less accurate than the models built on specific cancer types (compare to results in Table 2). The prediction accuracy levels dropped dramatically to an average of 66%, with precision and recall of 68%.

Drugs	Accuracy	Precision	Recall
Carboplatin	64.39%	65.04%	81.54%
Cisplatin	63.84%	63.99%	68.71%
Docetaxel	69.19%	71.96%	60.69%
Doxorubicin	62.55%	67.16%	39.76%
Gefitinib	73.62%	74.63%	83.01%
Gemcitabine	65.68%	64.42%	75.00%
Paclitaxel	63.28%	65.49%	59.63%
Topotecan	67.34%	67.49%	74.55%
Average	66.20%	67.50%	68.46%

In those cases where patients were observed to positively respond to the combination therapies, the prediction was scored as "true positive" (TP) if the patient was predicted to respond to at least one of the administered drugs (e.g., patients 242, 367 and 588). Conversely, in cases where patients were observed to not respond to the combination therapy, the prediction was scored as "false positive" (FP) if the patient was predicted to respond to at least one of the drugs (e.g., patients 286 and 1012). Instances where the patient is both predicted and observed not to respond to the combination therapy are scored as "true negative" (TN) (e.g., patients 272 and 545) while cases where the patient responded to the combination therapy but is predicted not to respond to either of the administered drugs the prediction was scored as "false negative" (FN), although no false negatives were found in our dataset. Overall predictive accuracy across the 23 patients was 91% with a precision of 89% and recall of 100%.

Table 4: Patient-by-patient analysis of ovarian cancer patients. Starting from the left and going right, the first column is the ID number assigned for each patient, the second column lists the drugs given to each patient (Carbo for carboplatin, Taxol for paclitaxel, Cis for cisplatin, GEM for gemcitabine), the third column is the observed response for each patient, as measured by CA-125 levels (R for responsive, NR for Nonresponsive) . The next eight columns are what the algorithm predicted the response would have been if the drug in that column was administered. The cells highlighted in blue are the predictions the algorithms made for the drugs that were administered for the patient. The column after the predictions for each drug is the final prediction that is based on the predicted response of the combination therapy given to the patients (cells in blue). This final predictive response is computed as responsive if at least one of the administered drugs is predicted as responsive, and called as nonresponsive otherwise. These final predicted responses were compared to the observed responses and given a prediction category of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) as shown in the final column(TP = both observed and predicted response are responsive, TN = both observed and predicted response are nonresponsive, FP = observed response is nonresponsive and predicted response is responsive, FN = observed response is responsive and predicted response is nonresponsive). There are only two incorrect predictions, both of which are false positives (red colored text in the final column, patients 286 and 1012).

Patient Number	Drug	OBSERVED	PREDICTED	PREDICTED	PREDICTED	PREDICTED	PREDICTED	PREDICTED	PREDICTED	PREDICTED	Final Pred based on Drug Administered	Prediction Category
		RESPONSE	Carboplatin	Paclitaxel	Cisplatin	Gemcitabine	Docetaxel	Doxorubicin	Gefitinib	Topotecan		
229	Carbo&GEM	R	R	R	NR	NR	R	R	R	R	R	TP
242	Carbo&Taxol	R	R	R	NR	NR	R	R	R	R	R	TP
272	Carbo&Taxol	NR	NR	NR	R	NR	R	NR	R	NR	NR	TN
286	Carbo&Taxol	NR	R	R	NR	NR	R	NR	R	NR	R	FP
317	Carbo&Taxol	R	R	R	NR	NR	R	NR	R	R	R	TP
336	Carbo&Taxol	R	R	R	NR	NR	R	R	R	R	R	TP
367	Carbo&Taxol	R	R	R	NR	NR	R	R	R	R	R	TP
413	Carbo&Taxol	R	R	R	NR	NR	R	R	R	R	R	TP
489	Carbo&Taxol	R	R	R	NR	NR	R	R	R	R	R	TP
528	Carbo&Taxol	R	R	R	NR	NR	R	R	R	R	R	TP
542	Carbo&Taxol	R	NR	R	NR	NR	R	NR	NR	R	R	TP
545	Carbo&Taxol	NR	NR	NR	NR	NR	R	NR	R	NR	NR	TN
588	Carbo&Taxol	R	NR	R	NR	NR	R	NR	R	R	R	TP
617	Carbo&Taxol	R	R	R	NR	NR	R	NR	R	R	R	TP
620	Carbo&Taxol	R	R	R	NR	NR	R	R	R	R	R	TP
813	Carbo/Cis/Taxol	R	R	R	NR	NR	R	NR	R	NR	R	TP
992	Topotecan	NR	NR	NR	NR	R	R	NR	R	NR	NR	TN
1012	Carbo & docetaxel	NR	NR	R	NR	NR	R	NR	R	NR	R	FP
1122	Carbo & Taxol	R	NR	R	NR	NR	R	NR	R	NR	R	TP
1129	Doxorubicin	NR	R	NR	NR	NR	R	NR	R	NR	NR	TN
1145	Carbo & Taxol	NR	NR	NR	NR	R	R	NR	R	NR	NR	TN
BJ1	Carbo & Taxol	R	NR	R	NR	R	R	NR	R	NR	R	TP
BJ4	Carbo&Taxol	R	NR	R	NR	R	R	NR	R	NR	R	TP

Discussion

It is now widely acknowledged that cancer is a highly complex disease and that, as a consequence, patients with even the same cancer type will often respond differently to the same therapeutic treatment [26]. A major goal of personalized cancer medicine is to accurately predict a patient's response to alternative therapeutic drugs/treatments based upon genomic profiles of tumor tissue biopsies thereby enabling selection of optimal therapeutic treatments for each individual patient [27].

Accurate predictions in science, including medical science, are generated in one of two ways [28]: If the cause and effect relationship(s) underlying a specific disease is known, then a specific diagnosis and treatment can be accurately predicted. While considerable progress in understanding the cause and effect relationships underlying some cancers has been made in recent years (e.g., [29,30], the extent of this understanding currently remains limited, thus far, resulting in relatively few examples of highly effective targeted gene therapies (e.g., [31]). A second way in which scientific predictions have and continue to be made is based on previously observed highly correlated trends. For example, many traditional chemotherapeutic drugs currently in use were originally identified through trial-and error screening and not based upon rational design [32].

Correlative-based predictions have experienced a resurgence in recent years with the development of machine learning (ML)-based methods to efficiently look for significant correlations embedded within large datasets. In the area of personalized medicine, ML-based approaches are being developed to potentially identify highly significant correlations between "omic" profiles (e.g., genomic, proteomic and/or metabolomic) of various types of cancers and to correlate these profiles with the therapeutic efficacy of drugs to kill the cancer cells or to significantly arrest their growth. Based upon identification of such correlations, models can be established to predict optimal drug treatments for individual cancer patients based on the "omic" profiles of their tumors without full knowledge of the underlying molecular mechanisms involved.

To date, a number of ML-based models have been developed for the prediction of optimal responses of cancer cell lines to drugs. A variety of computational approaches have been taken [e.g., regression methods (e.g., [33], random forests (e.g., [12]), modified rotation forest (e.g., [34]), support vector machines (e.g., [35]), deep learning and/or transfer learning (e.g., [36-39]) with reported predictive accuracies ranging from 70-85%. The predictive accuracy (>90%) of our ELAFT models compares favorably with these earlier studies, particularly in light of the fact that the ELAFT predictive models were developed and tested over fifteen

distinct cancer types incorporating data derived from 499 independent cell lines.

To evaluate the potential clinical accuracy of our ELAFT models to predict the responsiveness of cancer patients to various chemotherapeutic drugs, we employed a previously established dataset comprised of gene expression profiles (microarray) and drug responsiveness (based on observed changes in CA-125 levels) of ovarian cancer patients to treatments of seven chemotherapeutic drugs (carboplatin, cisplatin, paclitaxel, topotecan, gemcitabine, docetaxel, doxorubicin) administered either singularly or in combination to 23 ovarian cancer patients (Table 4). Predicted responses to each of these drugs were made based on ELAFT models developed using the DepMap dataset. Comparisons of these predictions with observed patient responses resulted in an overall predictive accuracy 91% with a precision of 89% and recall of 100%. While these results are highly encouraging, additional analyses of larger and more diversified cancer datasets are currently underway to fully evaluate the potential clinical utility of ELAFT as a useful tool in the treatment of cancer patients.

Acknowledgements

Dongjo Ban provided critical review of the manuscript. Funding was provided by the Ovarian Cancer Institute, Atlanta, GA, Northside Hospital (Atlanta) and The Deborah Nash Endowment Fund.

References

1. Letai A (2017) Functional precision cancer medicine - moving beyond pure genomics. Nature Med 23(9): 1028-1035.
2. Barretina J, Caponigro G, Stransky N, et al. (2012) The cancer cell line encyclopedia - Using preclinical models to predict anticancer drug sensitivity. Euro J Cancer 48: S5–S6.
3. Basu A, Bodycombe NE, Cheah JH, et al. (2013) An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell 154(5): 1151–1161.
4. Garnett MJ, Edelman EJ, Heidorn SJ, et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 483(7391): 570–575.
5. Gönen M, Margolin AA (2014) Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. Bioinfor 30(17): i556–i563.
6. Haider S, Rahman R, Ghosh S, et al. (2015) A copula based approach for design of multivariate random forests for drug sensitivity prediction. PLoS ONE 10(12): e0144490.
7. Huang C, Clayton EA, Matyunina LV, et al. (2018) Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. Sci Reports 8(1): 1-2.
8. Iorio F, Knijnenburg TA, Vis DJ, et al. (2016) A landscape of pharmacogenomic interactions in cancer. Cell 166(3): 740-754.
9. Menden MP, Iorio F, Garnett M, et al. (2013) Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. PLoS ONE8(4): e61318.
10. Rahman R, Matlock K, Ghosh S, et al. (2017a) Heterogeneity aware random forest for drug sensitivity prediction. Sci Reports 7(1).
11. Rahman R, Otridge J, Pal R (2017b) IntegratedMRF: random forest-based framework for integrating prediction from different data types. Bioinformatics 33(9): 1407-1410.
12. Riddick G, Song H, Ahn S, et al. (2010) Predicting in vitro drug sensitivity using Random Forests. Bioinformatics 27(2): 220-224.
13. Sagi O, Rokach L (2018) Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8(4): 1-2.
14. NCI-60 Gene Expression Data (2020).
15. NCI-60 Drug Response Data (2020).
16. Mezencev R, Matyunina LV, Jabbari N, et al. (2016) Snail-induced epithelial-to-mesenchymal transition of MCF-7 breast cancer cells: systems analysis of molecular changes and their effect on radiation and drug sensitivity. BMC Cancer 16: 236.
17. Guo T, Luna A, Rajapakse VN, et al. (2019) Quantitative proteome landscape of the NCI-60 cancer cell lines. IScience 21: 664-680.
18. Ortmayr K, Dubuis S, Zampieri M (2019) Metabolic profiling of cancer cells reveals genome-wide crosstalk between transcriptional regulators and metabolism. Nat Commun, 10(1).
19. DepMap: The Cancer Dependency Map Project at Broad Institute. (2020).
20. Lili LN, Matyunina LV, Walker L, et al. (2013) Molecular profiling supports the role of epithelial-to-mesenchymal transition (EMT) in ovarian cancer metastasis. J Ovarian Res 6(1): 49.
21. Bowen NJ, Walker LD, Matyunina LV, et al. (2009) Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. BMC Med Genomics 2: 71.
22. Amaratunga D, Cabrera J (2001) Analysis of data from viral DNA microchips. J Amer Stat Assoc 96(456): 1161-1170.
23. Bolstad BM, Irizarry RA, Astrand M, et al. (2003) A comparison of normalization methods for high-density oligonucleotide array data based on variance and bias. Bioinfo 19(2): 185-193.
24. Guyon I, Weston J, Barnhill S, et al. (2002) Gene selection for cancer classification using support vector machines. Machine Learning 46(1/3): 389-422.
25. Webb GI, Sammut C, Perlich C, et al. (2011) Leave-One-Out Cross-Validation. Encyclopedia of Machine Learning, Springer, Boston, MA, pp: 600–601.
26. Poudel P, Nyamundanda G, Patil Y, et al. (2019) Heterocellular gene signatures reveal luminal-A breast cancer heterogeneity and differential therapeutic responses. NPJ Breast Cancer 5(1): 1.
27. Bode AM, Dong Z (2017) Precision oncology- the future of personalized cancer medicine? NPJ Precis Oncol 1(1): 2.
28. McDonald JF (2018) Back to the future - The integration of big data with machine learning is re-establishing the importance of predictive correlations in ovarian cancer diagnostics and therapeutics. Gyn Onc 149(2): 230–231.

29. Chen YC, Peng D, Li D (2010) Molecular and cellular bases of chronic myeloid leukemia. Protein Cell 1(2): 124-132.
30. Peddi PF, Matthews EJ, Ma C (2011) Basis of triple negative breast cancer and implications for therapy. Int J Breast Cancer.
31. Lee YT, Tan YJ, Oon CE (2018) Molecular targeted therapy: treating cancer with specificity. Eur J Pharmacol 834: 188-196.
32. Rao AR, Motiwala HG, Karim OMA (2007) The discovery of prostate-specific antigen. BJU Int 101(1): 5-10.
33. Kurilov R, Haibe-Kains B, Brors B (2020) Assessment of modelling strategies for drug response prediction in cell lines and xenografts. Sci Rep 10(1): 2849.
34. Sharma A, Rani R (2020) Ensembled machine learning framework for drug sensitivity prediction. IET Syst Biol 14(1): 39-46.
35. Huang C, Mezencev R, McDonald JF, et al. (2017) Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. PLoS ONE, 12(10).
36. Baptista D, Ferreira PG, Rocha M (2020). Deep learning for drug response prediction in cancer. Brief Bioinfo 22(1): 360-379.
37. Manica M, Oskooei A, Born J, et al. (2019) Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. Mol Pharmaceutics 16(12): 4797-4806.
38. Rampášek L, Hidru D, Smirnov P, et al. (2019) Dr.VAE: improving drug response prediction via modeling of drug perturbation effects. Bioinformatics 35(19): 3743-3751.
39. Xia F, Shukla M, Brettin T, et al. (2018) Predicting tumor cell line response to drug pairs with deep learning. BMC Bioinformatics, 19(S18): 486.

***Corresponding author:** John F McDonald, School of Biological Sciences, Georgia Institute of Technology, 315 Ferst Drive, Atlanta, GA 30332, USA e-mail: john.mcdonald@biology.gatech.edu

Received date: May 03, 2021; **Accepted date:** July 09, 2021; **Published date:** August 16, 2021

Citation: Lanka J, Housley SN, Benigno BB, McDonald JF (2021) ELAFT: An Ensemble-based Machine-learning Algorithm that Predicts Anti-cancer Drug Responses with High Accuracy. *J Onc Res* 4(1): 111.