

PROGNOSTICATE THE DRUGS FOR MULTIPLE MYELOMA PATIENTS BY USING GENE EXPRESSION TECHNIQUE WITH POLYCLONAL AND MONOCLONAL SAMPLES

R. Hemalatha and T.P. Abhinayaa Devi

Department of Computer Science, Tiruppur Kumaran College for Women, India

Abstract

A major protest in cancer treatment is predicting the clinical response to anti-cancer drugs for each individual patient. For complex diseases such as Myeloma, characterized by high inter-patient variance, the implementation of precision medicine approaches is dependent upon understanding the pathological processes at the molecular level. Myeloma is one of the horrible diseases in the world claiming plurality of lives. Accurately predicting drug responses to Myeloma is a most important problem preventing oncologists' efforts to ensemble the most powerful drugs to treat Myeloma, which is a root goal in precision medicine. It entails the design of therapies that are matched for each individual patient. In this article, it considers a review of approaches that have been proposed to tackle the drug sensitivity prediction problem especially with respect to the personalized Cancer therapy. There are a total of 44 drug sensitivity prediction algorithms. In that the gene expression microarrays consistently provided the best predictive power of the individual profiling data sets; however, performance was increased by including multiple, independent data sets. The proposed algorithm surpassed Bayesian Multitask Multiple Kernel Learning (BMMKL) classification which currently represent the state-of-the-art in drug-response prediction and finally passed the gene expression data to Cytoscape for visualization.

Keywords:

Drug Sensitivity Prediction, Gene Expression Microarrays, Prediction Algorithms

1. INTRODUCTION

The human body is made up of cells, tissues, muscles and veins. The cells which grows in an uncontrolled manner or the cells which shows some changes in the body are termed to be Cancerous Cells. The cancer cells may be in the form of masses or lumps, which are medically termed as tumour as shown in Fig.1.

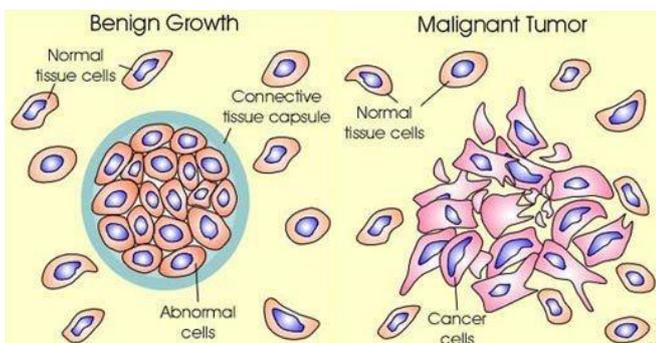


Fig.1. Normal cells vs Cells affected by Cancer

There are more than 200 different types of cancer. A primary tumour is the name where the cancer starts. Cancer sometimes

spread to other parts of the body and this is called a secondary tumour or a metastasis.

Cancers are grouped into five main categories.

- Carcinoma: This type of cancer begins in the skin or in the tissues that line or cover internal organs.
- Sarcoma: This type of cancer begins in the connective or supportive tissues.
- Blood cancer: This type of cancer begins in the blood cells and spreads through blood.
- Lymphoma and Myeloma: This type of cancer begin in the cells of the immune system.
- Brain and Spinal cord cancer: This type of cancer are known as the central nervous system cancer [1].

1.1 MYELOMA DISEASE

Cancer begins when cells in the body start to grow out of control. Multiple myeloma is a cancer formed by malignant plasma cells (Fig.2). Normal plasma cells are found in the bone marrow and are an important part of the immune system.



Fig.2. Myeloma Cancer in Bone marrow

The immune system is made up of several types of cells that work together to fight infections and other diseases. Lymphocytes (lymph cells) are the main cell type of the immune system. The major types of lymphocytes are T cells and B cells.

When B cells respond to an infection, they sophisticated and change into plasma cells. Plasma cells make the antibodies (also called immunoglobulins) that help the body attack and kill germs. Lymphocytes are in many areas of the body, such as lymph nodes, the bone marrow, the intestines, and the bloodstream. Plasma cells, however, are mainly found in the bone marrow. Bone marrow is the soft tissue inside some hollow bones. In addition to

plasma cells, normal bone marrow has cells that make the different normal blood cells.

When plasma cells become cancerous and grow out of control, they can produce a tumor called a plasmacytoma. These tumors generally develop in a bone, but they are also rarely constructed in other tissues. If someone has only a single plasma cell tumor, the disease is called an isolated (or solitary) plasmacytoma. If someone has more than one plasmacytoma, they have multiple myeloma as shown in Fig.3.

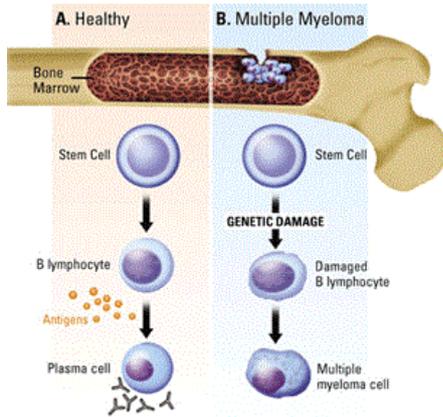


Fig.3. Normal vs Multiple Myeloma Bone Marrow

1.1.1 Symptoms:

The symptoms of multiple myeloma are Bone problems, Anemia, Extreme thirst, leading to drinking a lot of fluids, Dehydration, Abdominal (belly) pain, kidney failure and hyperviscosity.

1.1.2 Treating Myeloma:

The medicines used to treat the different types of Myeloma may vary but treatments can include:

- Chemotherapy: Medicines that destroy the cancer cells.
- Bone marrow or stem cell transplants: To replace the patient's bone marrow with healthy bone marrow.
- Blood and platelet transfusions: To replace red blood cells and platelets.
- Antibiotic: To fight infections as the patient's immune system is not working well [2].

1.2 DRUG SENSITIVITY

In recent years, the study of predictive modeling of drug sensitivity has received a boost due to the ever-growing interest in precision medicine and the availability of large-scale pharmacogenomics datasets. Drug sensitivity prediction is an integral part of personalized medicine that refers to therapy tailored to an individual patient, rather than a one-size-fits-all approach designed for an average patient [1].

Personalized therapy in the 20th century was primarily based on advanced visualization approaches, such as X-ray or Magnetic Resonance Imaging (MRI) scans along with multiple pathology tests.

The ability to measure genomic features on an individual basis in the last two decades opened up numerous possibilities for personalized therapy. The genomic characterizations provide

substantially more detailed information on a patient with a genetic disease as compared to phenotypic observations or non-molecular tests.

The National Cancer Institute (NCI) defines personalized medicine as "A form of medicine that uses information about a person's genes, proteins, and environment to prevent, diagnose, and treat disease" [35].

An underlying theme of personalized medicine is the incorporation of patient's genomic information besides other characteristics in therapeutic decisions. This review primarily considers various algorithms that have been proposed to decipher and obtain relevant information from genomic and drug sensitivity characterization sources. These data sources are used to design models that can predict: (1) the sensitivity of a new tumor culture or cell line to an existing drug or combinations of them or (2) the sensitivity of an existing tumor culture to a new drug or to a drug combination.

It must be noted that cell lines are the most commonly-used approach to study cancer biology and test various cancer treatments. Cell lines usually contain cells of one type and can be either genetically identical or diverse. These cell lines are easy to grow in a laboratory, and multiple research groups can have access to the same cell line for corroborating research findings and to build a body of knowledge based on a specific genetic type of cell. Cancer cell lines are expected to reflect the properties (such as genotypic and phenotypic characteristics, mutation status, gene expression and drug sensitivity) of the original cancer type from which they were cultured [2], [3].

However, the laboratory environment along with the absence of other regular cell types within the cell line restricts the ability of a cell line itself to mimic in vitro cancer growth, as well as the efficacy of anti-cancer drugs [4].

The accuracy and precision of drug sensitivity modelling is highly dependent on associated factors, like data types, feature selection, model selection and model validation. Therefore, this paper structured the review in a manner such that we can discuss each one of these individual steps. In terms of the models used to predict drug sensitivity, this paper considers model based on genomic characterization alone, along with models based on both genomic and functional characterizations.

2. LITERATURE REVIEW

Kareem states that the white blood cell nuclei can be detected by using some image arithmetic operation and some image threshold operations [3].

Abdullah [29] proposes a technique for the Acute Myelogenous Leukaemia (AML) detection. By using the combination between the contrast enhancement WBC can be detected. Partial contrast and bright stretching are used as the contrast enhancement technique. To produce a better result, HSI color space technique is used instead of the RGB color model. The end result states that after the usage of the enhancement technique, the reduction of the interference on the background images is found to a great extent and extraction of the morphological images from the blood samples is allowed efficiently.

Scotti [4] used a technique, which is based on the removal of background and white blood cell segmentation clustering. Some

low-pass filters and threshold operations were used for removing the background and clustering the WBC segmentation.

Ghosh [5] deals with finding out the accurate threshold for the leukocyte segmentation. The fuzzy divergence technique and some functions such as Gamma, Gaussian and Cauchy techniques are used in this paper. The paper deals with the advantage and the disadvantage of using this technique. The advantage is that the nucleus segmentation is performed well by using this technique. Whereas, the disadvantage is that the cytoplasm extraction is not performed well.

Akinola and Oyabugbe [30] designed to determine how data mining classification algorithm perform with increase in input data sizes. They used three data mining classification algorithms Decision Tree, Multi-Layer Perceptron (MLP) Neural Network and Naïve Bayes were subjected to varying simulated data sizes. The time taken by the algorithms for trainings and accuracies of their classifications were analyzed for the different data sizes.

Sharma [31] focuses on the application of various data mining classification techniques using different machine learning tools such as WEKA and Rapid miner over the public healthcare dataset for analyzing the health care system. The percentage of accuracy of every applied data mining classification technique is used as a standard for performance measure. The best technique for particular data set is chosen based on highest accuracy.

In [34], Sinha differentiates five types of leucocytes. K-means algorithm is used for the differentiation. The author says that some other clustering model algorithms can also be used for the differentiation purpose.

3. MATERIALS AND METHODS

3.1 BAYESIAN MULTITASK MULTIPLE KERNEL LEARNING (BMMKL)

The most commonly-used Bayesian Multitask Multiple Kernel Learning (BMMKL) method for regression modelling consists of support vector regression. BMMKL utilizes kernelized regression to represent the relationships between features and Bayesian inference for learning the model.

Support Vector Regression (SVR) is based on the statistical learning theory or Vapnik Chervonenkis (VC) theory developed by Vladimir Vapnik and Alexey Chervonenkis during 1960s [6]-[8].

To explain SV regression, consider $x(i, j)$ and $y(i)$ where $(i = 1, 2, \dots, n; j = 1, 2, \dots, M)$ to denote the training predictor features and output response samples, respectively. In terms of SVR, the goal is to find a function $f(x(i, 1:M))$ that has at most deviation ϵ from actual values $y(i)$ [9], and a characteristic of the function is minimized. Thus, the deviation of the function approximator f from the actual values is being limited while minimizing some characteristic of the approximator. For explanation purposes, consider the following linear function:

$$f(x(i, 1:M)) = \sum_{j=1}^M w(j)x(i, j) + b \quad (1)$$

Here, the function characteristic will be the norm of w . The problem can be posed as an optimization problem as follows: minimize $0.5\|w\|^2$

$$\text{Subjected to } \left| y_i - \sum_{j=1}^M w(j)x(i, j) + b \right| \leq \epsilon$$

To extend the SV algorithm to the nonlinear scenario, kernels are usually employed. Some commonly-used kernels are [9]:

Homogenous polynomial kernels k with $p \in N$ and:

$$K(x, x') = \langle x, x' \rangle^p \quad (2)$$

Inhomogenous polynomial kernels k with $p \in N, c > 0$ and:

$$K(x, x') = (\langle x, x' \rangle + c)^p \quad (3)$$

Radial basis kernels:

$$K(x, x') = \frac{e^{-\|x-x'\|^2}}{2} \quad (4)$$

Support vector machines can handle non-linearity based on the use of kernels and tend to have better predictive performance compared to regularized regression approaches [10].

3.2 SUPPORT VECTOR MACHINE

It is set of related supervised learning method used for regression and classification. They belong to a family of generalized linear classification. A special property of Support Vector Machine is, Support Vector Machine simultaneously minimize the empirical classification error and maximize the geometric margin. So Support Vector Machine called Maximum Margin Classifiers. Support Vector Machine is based on the Structural risk Minimization (SRM). Support Vector Machine map input vector to an advanced dimensional space where a maximal unravelling hyper plane is constructed.

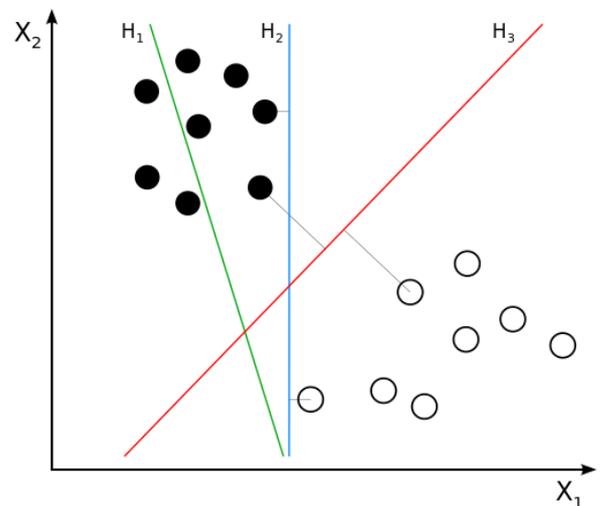


Fig.4. SVM

The two parallel frenzied planes are constructed on each side of the hyper plane that separate the data. The separating hyper plane is the hyper planes that increases the distance between the two parallel hyper planes as shown in Fig.4. An expectation is made that the larger the distance between these parallel hyper planes or margin, the improved and the generalization error of the classifier will be more.

3.3 MYELOMA CANCER DATASET DESCRIPTION

The Myeloma cancer dataset consists of 38 samples: 11 samples of Acute Lymphoblastic Leukemia (ALL) and 27 samples of Acute Myeloid Leukemia (AML). The source of the gene expression measurement is captured from 16 peripheral blood samples and 22 bone marrow samples. Gene expression levels in these 38 samples are measure by applying high density oligonucleotide microarrays.

Relation: GeneLeukemia

No.	1: SNUM	2: CLASS	3: BM_PB	4: TE_IF_ALL	5: FAB_IF_AML	6: Year	7: Gender	8: pct_Blasts	9: Treatment_Response	10: PS	11: Source	12: D49950	13: D63880
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	
35	848	ALL	BM	B-cell	199..	F				0.84	DFCI	80.0	368.0
36	847	ALL	BM	B-cell	198..	M				0.81	DFCI	191.0	741.0
37	848	ALL	BM	B-cell	199..	F				0.94	DFCI	51.0	579.0
38	849	ALL	BM	B-cell		M				0.84	DFCI	128.0	391.0
39	855	ALL	BM	B-cell		F		Success		0.73	St-Jude	74.0	449.0
40	856	ALL	BM	B-cell		F		Success		0.84	St-Jude	20.0	494.0
41	859	ALL	BM			F		Failure		0.68	St-Jude	205.0	891.0
42	857	ALL	PB	T-cell	199..	M		Failure		0.16	DFCI	53.0	72.0
43	858	ALL	PB	B-cell	199..	M		Success		0.8	DFCI	145.0	734.0
44	859	ALL	PB	B-cell	199..	M		Failure		0.85	DFCI	84.0	545.0
45	870	ALL	PB	B-cell	199..	F		Success		0.73	DFCI	20.0	440.0
46	871	ALL	PB	B-cell	199..	F		Success		0.93	DFCI	105.0	113.0
47	872	ALL	PB	B-cell	199..	F		Success		0.77	DFCI	146.0	78.0
48	828	AML	BM	B-cell	M2		79.0	Failure		0.44	CALGB	297.0	279.0
49	829	AML	BM	M2		M	34.0	Failure		0.74	CALGB	190.0	277.0
50	830	AML	BM	B-cell	M5		93.0	Failure		0.8	CALGB	326.0	136.0
51	831	AML	BM	M4		M	77.0	Failure		0.61	CALGB	197.0	321.0
52	832	AML	BM	B-cell	M1		86.0	Failure		0.47	CALGB	259.0	283.0
53	833	AML	BM	T-cell	M2		70.0	Failure		0.89	CALGB	100.0	151.0
54	834	AML	BM	M2		M	77.0	Success		0.64	CALGB	355.0	226.0
55	835	AML	BM	M1			67.0	Success		0.21	CALGB	295.0	20.0
56	836	AML	BM	M5			76.0	Success		0.94	CALGB	283.0	410.0
57	837	AML	BM	M2			44.0	Success		0.95	CALGB	311.0	185.0
58	838	AML	BM	M1			80.0	Success		0.73	CALGB	70.0	118.0
59	850	AML	BM	M4			93.0	Failure		0.97	CALGB	346.0	164.0
60	851	AML	BM	M2			57.0	Failure		1.0	CALGB	211.0	209.0
61	852	AML	PB	M4			86.0	Success		0.61	CALGB	109.0	131.0
62	853	AML	BM	M2			76.0	Success		0.89	CALGB	106.0	94.0

Fig.5. Dataset of Multiple Myeloma Patient

Each sample contains 7129 gene expression levels.

Number of Cases: 38

Number of Attributes: 7129 plus the class attribute

Number of Missing attributes: None

The domain of the attributes is set to real as the gene expression may fall under real values.

4. PROPOSED FRAMEWORK

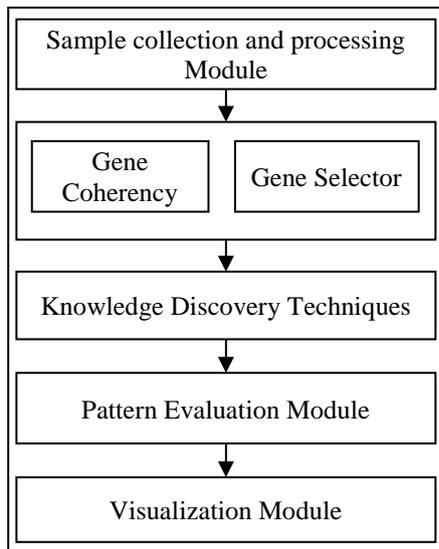


Fig.6. Proposed Framework

The Fig.7(a), In addition to the six profiling data sets, three different categories of data views were compiled using prior biological knowledge, yielding in total 22 genomic views of each cell line. The Fig.7(b) has Bayesian multitask MKL combines

nonlinear regression, multiview learning, multitask learning and Bayesian inference.

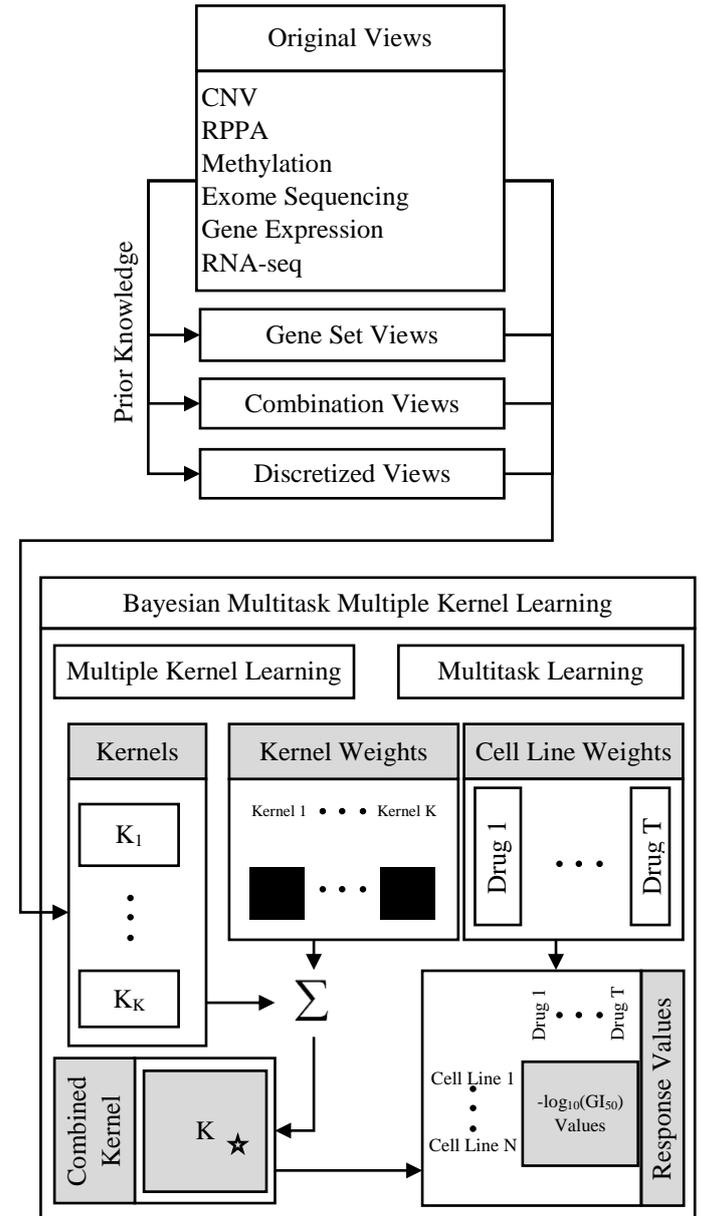


Fig.7. Proposed Methodology

Nonlinear regression: response values were computed not directly from the input features but from kernels, which define similarity measures between cell lines. Each of the K data views was converted into an $N \times N$ kernel matrix K_k ($k = 1, \dots, K$), where N is the number of training cell lines. Specifically, the Gaussian kernel was used for real-valued data, and the Jaccard similarity coefficient for binary-valued data. Multiview learning: a combined kernel matrix K^* was constructed as a weighted sum of the view-specific kernel matrices K_k , $k = 1, \dots, K$.

The kernel weights were obtained by multiple kernel learning. Multitask learning: training was performed for all drugs simultaneously, sharing the kernel weights across drugs but allowing for drug-specific regression parameters, which for each drug consisted of a weight vector for the training cell lines and an intercept term.

Bayesian inference: the model parameters were assumed to be random variables that follow specific probability distributions. Instead of learning point estimates for model parameters, the parameters of these distributions were learned using a variational approximation scheme.

5. PREDICTION PERFORMANCE

In this paper, the BMMKL is implemented by using the algorithmic code provided in [11]. Based on the parameters used in [11], this paper has considered 200 iterations and gamma prior values (both a and b) of one. Subspace dimensionality has been considered to be 20, and the standard deviation of hidden representations and weight parameters are selected to be the defaults 0.1 and one, respectively.

Table.1. Coefficients between Actual and Predicted Drug Sensitivity

Drug Name	Correlation Coefficients (BMMKL)
AZD0530	0.2245
Erlotinib	0.4100
Lapatinib	0.4641
Crizotinib	0.2647

The Table.1 shows the predictive performance in the form of correlation coefficients between actual and predicted drug sensitivity (AUC) for BMMKL.

6. BIOLOGICAL ANALYSIS

The standard procedure used to compare drug sensitivity prediction algorithms consists of evaluating a fitness measure, such as a difference measure of a norm between experimental responses and predicted values or a similarity measure of the correlation coefficient between the vector of predicted and actual experimentally-derived sensitivities.

Other than evaluating the performance of a generated model using existing drug sensitivity data based on new experiments can be conducted to validate hypotheses generated from the estimated models. For example, [12] used in vitro DIPG myeloma cancer cell lines to validate model hypotheses of synergistic drug pairs. The commonly-used experimental approaches included the use of

- In-vitro cancer cell lines*: Cell lines are the most commonly-used approach to study cancer biology and test various cancer treatments. Some commonly-used cancer cell line databases are NCI60 [13], [14], the Cancer Cell Line Encyclopedia [15], the Genomics of Drug Sensitivity in Cancer [16] and the Cancer Therapeutics Response Portal (CTRP) [17];
- In vitro primary tumor cultures*: Primary tumor cultures are the cell cultures established from the tumor biopsy of the patient and reflect the surroundings in the original tissue. As compared to cell lines, primary cell cultures are more heterogeneous and also contain a small fraction of normal stromal cells, which can be an advantage since tumor response is often influenced by an interplay of the tumor cells and the non-malignant cells in the tumor microenvironment;

- In vitro three-dimensional cell cultures*: 3D cell cultures attempt to provide an environment that is similar to what is available to in vivo tissues [18], including 3D scaffolding, nutrient-rich interstitial fluid, the exchange of growth factors and other biological effectors;
- In vitro Genetically-Engineered Mice Models (GEMM)*: GEMM consists of mice whose genes have been modified to reflect genetic aberrations observed in a specific human tumor [19] [20]. GEMMs [21] have played a vital role in understanding tumor proliferation, as they effectively capture the tumor microenvironment along with the human tumor phenotype. However, the timelines to generate these models are longer, along with the expenses involved;
- In vivo xenograft mice*: Xenograft mice are created by transplanting human tumor cells into the organ type in which the tumor originated, in immunocompromised mice that accept human cells [22]. This approach has shown high drug efficacy predictive capability with respect to human clinical activity [23]–[25], but can fail to replicate the immune system of the human cancer and, thus, unable to model the lymphocyte-mediated response to the tumor [24].

7. RESULTS AND DISCUSSION

The predictive performance evaluation based on existing drug sensitivity data or new validation experiments can consider the biological analysis of the variables deemed significant by the generated predictive model. An approach to do that is to consider prior biological knowledge of Protein-Protein (PP) interactions or other forms like protein-DNA, protein-reaction, reaction-compound and protein-mRNA, among others, to evaluate the observed number of such interactions in the model of selected variables. This procedure can be conducted using platforms, such as the Search Tool for the Retrieval of Interacting Genes (STRING) [32] or Cytoscape [33].

To illustrate the biological relevance of features selected by a predictive model, we will consider the NCI-DREAM drug sensitivity prediction challenge where five different data types of gene expression (microarray data), methylation, RNA sequencing, RPPA and SNP6 are considered to generate individual bayesian multitask multiple kernel learning models for each subtype that are then combined together to generate the integrated prediction [48]. For our current analysis, we used the bayesian multitask multiple kernel variables. Each individual set was used as input to the STRING [36] database to analyze the Protein-Protein Interactions (PPI). The enrichment p-values for each of the five PPI networks are as follows: gene expression (0.508), methylation (0.840), RNA-Seq (0.059), RPPA (0) and SNP6 (0.319).

The lowest p-value for RPPA shows that the important features obtained from the RPPA-generated bayesian multitask multiple kernel are significantly enriched as compared to random sub-networks. We passed the RPPA and gene expression output networks to Cytoscape to use the CyTargetLinker [27] application.

CyTargetLinker extends the output network from STRING using Regulatory Interaction Networks (RegINs), which are local database files, to extend the network with the possible

interactions: drugs, genes, mRNA, additional transcription factors and other targets. The RegIN files can be obtained from the CyTargetLinker website [37].

The DrugBank database is used to identify drug interactions containing 7759 compounds (drugs), some FDA approved and some experimental.

DrugBank [38] combines comprehensive pharmacological and pharmaceutical drug data along with the corresponding targets, sequences, structure and pathway information.

The Fig.8 shows the interactions among the top 50 features selected by gene expression and plotted using the visualization platform Cytoscape. Only nine out of these 50 (denoted by the red circles) had known interactions. The connections shown by the violet squares are the drug interactions obtained from the DrugBank database.

On the other hand, when we observe the known prior connections among the top 50 features selected using the RPPA-based model, a significantly higher number of connections were identified, as shown in Fig.9. For RPPA-based top features, there were 41 proteins found (red circles) to interact with 434 drugs (violet squares).

These forms of visualization displaying different types of interactions provide insights upon the biological relevance of the selected variables. The predictive performance of a model is one aspect of the goodness of fit of a model, whereas the biological relevance of the selected features is an equally important aspect of a predictive model.

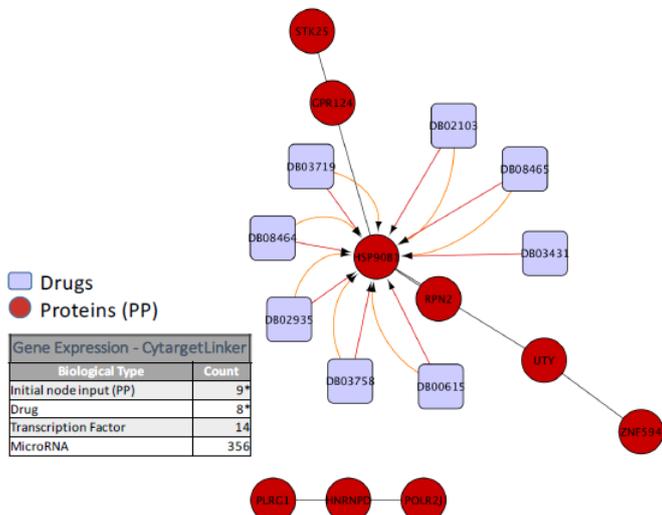


Fig.8. Cytoscape Visualization

In Fig.8, protein and drug interactions among top features selected in gene expression data. The top 50 genes selected from gene expression data are passed to the STRING database to find protein-protein interaction. Then, this new network is passed to Cytoscape using the CyTargetLinker App to extend the Protein-Protein (PP) interactions in a 7759 drug database (DrugBank); only the elements marked with an * are displayed in this network.

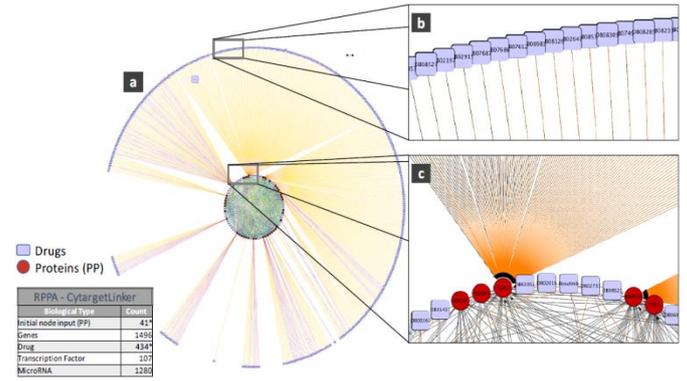


Fig.9.(a) Protein and drug interactions among top features selected in the RPPA data; (b) Outer ring showing details of the drug connection with CDK2; (c) Inner ring detail showing the CDK2 drug connectivity network from both the inner and outer rings

In Fig.9.(a) Protein and drug interactions among top features selected in the RPPA data; (b) Outer ring showing details of the drug connection with CDK2; (c) Inner ring detail showing the CDK2 drug connectivity network from both the inner and outer rings. Images obtained from Cytoscape using the CyTargetLinker App (version 3.0.1). Only the elements marked with an * are displayed in this network.

8. CONCLUSION AND FUTURE SCOPE

The prediction accuracy of the current state of the art techniques for drug sensitivity prediction is still limited and often based on a single snapshot of the underlying system. The prediction accuracy can potentially be increased by incorporating diverse genomic characterization datasets, including DNA mutations, gene expression and protein expression along with epigenomic characterizations.

Advances in this area are necessary for delivering on the promise of precision medicine. Moreover, as larger and more complex omic data sets become available, the greater the expectations and opportunities for generating novel applications. Such systems are also likely to be based on new investigations about the interplay of multiple types of molecular data and their capacity for predicting drug responses. These models will progressively incorporate information encoded at various levels of biological control (e.g. tumor epigenetics), resolution analyses (e.g. multiple biopsies per tumor) and the tumor micro-environment. Also, as new advances in cancer immunotherapy materialize, there will be additional opportunities for developing prediction models tailored to this type of treatments or their combination with small molecule drugs.

To bring these models closer to the clinic, it is important to require high levels of expertise and involvement from bioinformaticians and clinically oriented researchers during all key development phases. Similarly, as new opportunities and expectations arise, there will be further needs for powerful, interpretable and reproducible computational methods in different cancer domains, especially those with limited treatment options for patients. This will be accompanied by a higher demand for researchers who can design, adapt and evaluate these models, as well as for clinicians well trained to interpret results and to

facilitate the creation of stronger bridges between the lab and the clinic.

In future, this work can be combined with genetic algorithm or Particle Swarm Optimization (PSO) technique which is used to solve a problem by iteratively trying to improve drug sensitivity with regard to a given measure of quality.

REFERENCE

- [1] Cancer Research UK, Available at: <http://cancerresearchuk.org>
- [2] Cell Division and Cancer, Available at: <http://www.abpischools.org.uk – cancer cells>
- [3] H.T. Madhloom, S.A. Kareem, H. Ariffin, A.A. Zaidan, H.O. Alanazi and B.B. Zaidan, “An Automated White Blood Cell Localization and Segmentation using Image Arithmetic and Automatic Threshold”, *Journal of Applied Sciences*, Vol. 10, No. 11, pp. 959-966, 2010.
- [4] F. Scotti, “Robust Segmentation and Measurements of White Cells in Blood Microscopic Images”, *Proceedings of IEEE Instrumentation and Measurement Technology Conference*, pp. 43-48, 2006.
- [5] M. Ghosh, D. Das, C. Chakraborty and A.K. Ray, “Automated Leukocyte Recognition using Fuzzy Divergence”, *Micron*, Vol. 41, No. 7, pp. 840-846, 2010.
- [6] V. Vapnik and A. Lerner, “Pattern Recognition using Generalized Portrait Method”, *Automation Remote Control*, Vol. 24, pp. 774-780, 1963.
- [7] V. Vapnik and A. Chervonenkis, “A Note on One Class of Perceptrons”, *Automation Remote Control*, Vol. 25, pp. 735-739, 1964.
- [8] V.N. Vapnik, “*Statistical Learning Theory*”, Wiley, 1998.
- [9] A.J. Smola and B. Scholkopf, “A Tutorial on Support Vector Regression”, *Statistics and Computing*, Vol. 14, pp. 199-222, 2004.
- [10] V. Cherkassky and Y. Ma, “Comparison of Model Selection for Regression”, *Neural Computation*, Vol. 15, pp. 1691-1714, 2003.
- [11] M. Gonen and A.A. Margolin, “Drug Susceptibility Prediction against a Panel of Drugs using Kernelized Bayesian Multitask Learning”, *Bioinformatics*, Vol. 30, pp. 566-563, 2014.
- [12] C.S. Grasso et al., “Functionally-defined Therapeutic Targets in Diffuse Intrinsic Pontine Glioma”, *Nature Medicine*, Vol. 21, No. 6, pp. 555-559, 2015.
- [13] R.H. Shoemaker, “The NCI60 Human Tumour Cell Line Anticancer Drug Screen”, *Nature Review Cancer*, Vol. 6, No. 10, pp. 813-823, 2006.
- [14] R. Pal et al., “Boolean Relationships among Genes Responsive to Ionizing Radiation in the NCI 60 ACDS”, *Bioinformatics*, Vol. 21, No. 8, pp. 1542-1549, 2005.
- [15] J. Barretina et al., “The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity”, *Nature*, Vol. 483, pp. 603-607, 2012.
- [16] W.E.A. Yang, “Genomics of Drug Sensitivity in Cancer (GDSC): A Resource for Therapeutic Biomarker Discovery in Cancer Cells”, *Nucleic Acids Research*, Vol. 41, pp. 955-961, 2013.
- [17] A. Basu et al., “An Interactive Resource to Identify Cancer Genetic and Lineage Dependencies Targeted by Small Molecules”, *Cell*, Vol. 154, No. 5, pp. 1151-1161, 2013.
- [18] J.B. Kim, “Three-Dimensional in Vitro Tissue Culture Models of Breast Cancer-A Review”, *Breast Cancer Research and Treatment*, Vol. 85, No. 3, pp. 281-291, 2004.
- [19] J.W. Gordon et al., “Genetic Transformation of Mouse Embryos by Microinjection of Purified DNA”, *Proceedings of National Academy of Science of the United States of America*, Vol. 77, No. 12, pp. 7380-7384, 1980.
- [20] J.C. Walrath, “Genetically Engineered Mouse Models in Cancer Research”, *Advances in Cancer Research*, Vol. 106, pp. 113-164, 2010.
- [21] A. Richmond and Y. Su, “Mouse Xenograft Models vs. GEM Models for Human Cancer Therapeutics”, *Disease Models and Mechanism*, Vol. 1, No. 2-3, pp. 78-82, 2008.
- [22] R.S. Kerbel, “Human Tumor Xenografts as Predictive Preclinical Models for Anticancer Drug Activity in Humans: better than commonly Perceived-but they can be Improved”, *Cancer Biology and Therapy*, Vol. 2, pp. 134-139, 2003.
- [23] J.I. Johnson et al., “Relationships between Drug Activity in NCI Preclinical in Vitro and in Vivo Models and Early Clinical Trials”, *British Journal of Cancer*, Vol. 84, pp. 1424-1431, 2001.
- [24] C.C. Scholz, “Correlation of Drug Response in Patients and in the Clonogenic Assay with Solid Human Tumour Xenografts”, *European Journal of Cancer*, Vol. 26, No. 8, pp. 901-905, 1990.
- [25] C. Khanna et al., “The Dog as a Cancer Model”, *Nature Biotechnology*, Vol. 24, No. 9, pp. 1065-1066, 2006.
- [26] B. Haibe-Kains, “Inconsistency in Large Pharmacogenomic Studies”, *Nature*, Vol. 504, pp. 389-393, 2013.
- [27] J.C. Costello, et al., “A Community Effort to Assess and Improve Drug Sensitivity Prediction Algorithms”, *Nature Biotechnology*, Vol. 32, No. 12, pp. 1202-1212, 2014.
- [28] Q. Wan and R. Pal, “An Ensemble based Top Performing Approach for NCI-DREAM Drug Sensitivity Prediction Challenge”, *PLoS ONE*, Vol. 9, No. 6, pp. 1-7, 2014.
- [29] A.N. Aimi Salihah, M.Y. Mashor and Azian Azamimi Abdullah, “Improving Blast Segmentation of Cute Myelogenous Leukaemia (AML) Images using Bright Stretching Technique”, *Proceedings of International Post Graduate Conference on Engineering*, pp. 1-5, 2010.
- [30] S. Olalekan Akinola and O. Jephthar Oyabugbe, “Accuracies and Training Times of Data Mining Classification Algorithms: An Empirical Comparative Study”, *Journal of Software Engineering and Applications*, Vol. 8, pp. 470-477, 2015
- [31] Tanvi Sharma, Anand Sharma and Vibhakar Mansotra, “Performance Analysis of Data Mining Classification Techniques on Public Health Care Data”, *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4, No. 6, pp. 1-5, 2016.
- [32] D. Szklarczyk et al., “STRING v10: Protein-Protein Interaction Networks, Integrated over the Tree of Life”, *Nucleic Acids Research*, Vol. 43, pp. 447-452, 2015.
- [33] P. Shannon et al., “A Software Environment for Integrated Models of Biomolecular Interaction Networks”, *Genome Research*, Vol. 13, pp. 2498-2504, 2003.

- [34] N. Sinha and A.G. Ramakrishnan, "Automation of Differential Blood Count", Proceedings of IEEE International Conference on Convergent Technologies for the Asia-Pacific region, pp. 371-375, 2003.
- [35] National Cancer Institute, "NCI Dictionary of Cancer Terms", Available at: [https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid= 561717](https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=561717)
- [36] String, Available at: <http://string-db.org/>
- [37] CyTargetLinker, Available at: <http://projects.bigcat.unimaas.nl/cytargetlinker/>
- [38] Drugbank, Available at: <http://www.drugbank.ca/>