Finding Cancer-Related Gene Combinations Using a Molecular Evolutionary Algorithm

Chan-Hoon Park[a,*] Soo-Jin Kim[b,*] Sun Kim[a] Dong-Yeon Cho[a] Byoung-Tak Zhang[a,b]
[a]School of Computer Science and Engineering, [b]*Graduate Program in Bioinformatics*
Seoul National University
Seoul 151-742, Korea
{chpark, sjkim, skim, dycho, btzhang} @bi.snu.ac.kr

***Abstract*—High-throughput data such as microarrays make it possible to investigate the molecular-level mechanism of cancer more efficiently. Computational methods boost the microarray analysis by managing large and complex data systematically. However, combinatorial interactions among genes have not been considered as a unit of the analysis since previous methods mainly focus on a whole gene or a single isolated gene. Here, we introduce a molecular evolutionary algorithm called probabilistic library model (PLM). In the PLM, library elements are generated from gene combinations. An evolutionary procedure is adopted to learn the probabilistic distribution of training samples. We apply the PLM to prostate cancer microarray data. The experimental results show that the PLM classifiers perform better than conventional methods such as neural networks and decision trees in accuracy. We also examine the evolved library to find cancer-related gene combinations.**

Keywords-component; Microarrays; Probabilistic library model; Prostate cancer classification

I. INTRODUCTION

Revealing the mechanism of cancer is a research area of constant interest. The nature of cancer, a disease with many variations such as multiple heterogeneous genetic and epigenetic changes, makes it difficult to study the cancer apart from its molecular basis [1]. As high-throughput data such as microarrays becomes easy to access, the cancer researches tend to be more systematic based on computational methods for managing large and complex data [2]. The gene expressions on the microarray show large and complex parallel interactions among genes and these interactions are valuable to discover the mechanisms of cancer development. Thus, computational approaches are required to analyze the data and collect cancer-related genes, while removing massive redundant information [6-16]. For this purpose, the correlations between genes and samples are calculated in general. They provide a list of important genes in cancer classification [6-8], whereas the effects of gene combinations are ignored. The gene combinations are important since they can provide some clues to find the cancer-related pathways among complex networks of genes. In addition, there are some studies that these combinations make synergic effects especially in cancer therapy [26, 27].

Machine learning techniques are alternative ways to analyze the microarrays [9-14]. They have shown good performances in the cancer classification tasks. We can also employ them to select relevant genes to cancer [9, 12-14]. However, previous studies do not consider the gene-gene interactions. Even though some approaches utilize the gene combinations by mapping into the high-dimensional space, they cannot obtain human-interpretable solutions.

In this paper, we propose a simple, but effective method to discover significant gene pairs for cancer classification. The proposed method is based on the probabilistic library model (PLM) [15, 16]. The PLM is motivated from molecular computing [19, 20], which only use simple operations such as selection and amplification. In the PLM classifier, gene combinations are selected to build a library, and an evolutionary learning is performed to obtain the joint probability distribution of given examples. Since the combinatorial effects among genes are explicitly considered in the PLM framework, one can get a list of meaningful decision rules after the learning process. The strong point of PLM framework is the interpretability of the results. The decision rules are represented as combinations of genes. The importance of each combination is immediately interpretable based on their fitness values or weights in the PLM classifiers. It can be compared to the decision trees. However, the PLM can search the problem space more flexibly using massive small probabilistic combinations of features with simple operations.

We apply the proposed method to the prostate cancer classification [17, 18]. The experimental results show that our method outperforms conventional machine learning algorithms in accuracy. By examining the gene pairs of the library, we could find the cancer-related genes with several candidates of co-regulated genes on the pathway.

The paper is organized as follows. In Section 2, the PLM classifiers are explained. Section 3 describes the evolutionary learning method to find the optimal library. In Section 4, the experimental results are described. Section 5 draws the conclusion and further research.

II. PROBABILISTIC LIBRARY MODEL

The PLM is a probabilistic framework proposed by Zhang and Jang [15]. For PLM classifiers, a training set D of K label is represented as follows:

*: Both authors have equally contributed to this work.

$$D = \{(\mathrm{x}_i, y_i)\}_{i=1}^{K} \quad (1)$$

$$\mathrm{x}_i = (x_{i1}, x_{i2}, ..., x_{in}) \in \{0,1\}^n \quad (2)$$

$$y_i \in \{0,1\}, \quad (3)$$

where x_i is a set of features from a sample i and y_i is its class.

A library element in the PLM is a decision rule or an individual which is a conjunction of binary variables x_i and a class label y_i. The number of variables in an individual is defined as the *order* of the individual. For example, the individual z = (x, y) = ($x_1 = 1, x_2 = 1, x_5 = 0, y = 1$), is a conjunction of three variables with class 1. Thus, z is the individual with the *order* of 3.

The library includes multiple copies of the individuals and the number of copies means the importance of the individual in a whole library. The learning procedure updates the number of copies to accurately discriminate training examples.

Substantially, the library expresses the joint probability $P(X,Y)$ of the input pattern X and the output class Y. Since each individual has a class label, the decision on a query can be made based on the individuals. Given an example, the class is determined by matching it against each individuals and taking the majority class. This ensemble approach naturally makes use of huge number of individuals to make decision robust.

The classification of the PLM can be explained to the conditional probability of each class on the input. Given input x, the class y* is decided by computing the conditional probability of each class, and then selecting the class which has the highest conditional probability, as follows:

$$y^* = \arg\max_{Y \in \{0,1\}} P(Y \mid \mathrm{x}) \quad (4)$$

$$= \arg\max_{Y \in \{0,1\}} \frac{P(Y, \mathrm{x})}{P(\mathrm{x})}. \quad (5)$$

Here, $P(Y,\mathrm{x})=P(Y|\mathrm{x})P(\mathrm{x})$, and Y represents the candidate classes.

To accomplish Equation (4), PLM classifiers initialize the library with nth *order* individuals and evolve their distribution. That is, the empirical probability distribution $P(X,Y)$ can be represented by a set of point estimators that constitute the library L of individuals:

$$P(X,Y) \approx \frac{1}{|L|} \sum_{i=1}^{|L|} f_i^{(n)}(X_1, X_2, ..., X_n, Y), \quad (6)$$

where $f_i^{(n)}(X_1, X_2, ..., X_n, Y)$ is the ith individual of *order* n and $|L|$ is the size of library. By increasing the $|L|$, the approximation can be more arbitrarily accurate. More theoretical background can be found in [15].

1. Let the library L represent the current empirical distribution $P(X, Y)$.

2. Given an input x,

3. Classify x using L as follows:

3.1 Extract all individuals matching with x into M.

3.2. Separate the individuals from M according to their classes:

- Extract the individuals with label Y=0 into M^0.

- Extract the individuals with label Y=1 into M^1.

3.3. Compute $y^* = \arg\max_{Y \in \{0,1\}} c(Y \mid \mathrm{x}) / |M|$

Figure 1. The procedure of PLM decision making.

The procedure of decision making in PLM classifiers is summarized in Figure 1. Given an input x, all individuals matching with x is extracted from the library L. The extraction is implemented by matching x with each and every library element of the library. The class decision of x is made by selecting the class which has the higher number of elements.

In step 3.1, the count c(x) of x in M approximates the evidence which is the probability of observing the matching individuals:

$$c(\mathrm{x})/|L| = |M|/|L| \approx P(\mathrm{x}). \quad (7)$$

Step 3.2 computes the frequencies $c(Y|\mathrm{x})$ of each class. These are an approximation of a posteriori probabilities which is the conditional probabilities given the example:

$$c(Y \mid \mathrm{x})/|M| = |M^Y|/|M| \approx P(Y \mid \mathrm{x}). \quad (8)$$

Thus, the procedure computes the maximum a posteriori (MAP) criterion:

$$y^* = \arg\max_{Y \in \{0,1\}} c(Y \mid \mathrm{x}) / |M|$$

$$= \arg\max_{Y \in \{0,1\}} c(Y \mid \mathrm{x})$$

$$\approx \arg\max_{Y \in \{0,1\}} P(Y \mid \mathrm{x}). \quad (9)$$

which validates the class decision in Equation (4).

The PLM has some advantages caused by molecular computing. The massive parallelism of the molecular operations can be easily transformed to the parallel computing. In particular, the PLM naturally finds potential relations by explicitly considering the pair of features as a unit. This nature directly provides the human-interpretable rules, which is similar to the branch of rules in decision trees. However, the PLM considers the probabilistic computation of individuals in decision making while decision trees make a decision deterministically.

III. Evolutionary Learning of the PLM

The evolutionary learning of PLM can be summarized as a procedure of finding a proper probability distribution of the library which best fits in the training examples using a gradient descent [16]. As a result of the learning procedure, we can adjust the number of copies of each individual to reduce the magnitude of the classification error.

An evolutionary procedure is applied after initializing the PLM classifier. Given a training input pattern x_i and its class y_i, the total quantity c for a class Y is computed as follows:

$$c(Y \mid \mathrm{x}_\mathrm{i}) = \sum_{j=1}^{|L|} c_j I_{z_j=(\mathrm{x},Y)} \tag{10}$$

$$\text{where } I_{z_j=(\mathrm{x},Y)} \begin{cases} 1 & \text{if } z_j = (\mathrm{x}, Y) \\ 0 & \text{otherwise} \end{cases}, \tag{11}$$

where c_j is the number of copies for the individual z_j. With a weight vector w, the conditional probability given the input x_i is represented as follows:

$$P(Y \mid \mathrm{x}_\mathrm{i}) \approx \frac{c(Y \mid \mathrm{x}_i)}{|M|} = \sum_{j=1}^{|L|} w_j I_{z_j=(\mathrm{x},Y)} \tag{12}$$

$$\text{where } w_j = c_j / |M|. \tag{13}$$

An error e_i of a training example x_i and class Y is given by

$$e_i = P^*(Y \mid \mathrm{x}_i) - P(Y \mid \mathrm{x}_i), \tag{14}$$

where $P^*(Y \mid \mathrm{x}_i) = I_{y_i=Y} \in \{0,1\}$ is the target probability for the training example x_i. From error of each training example, the error function of a whole training dataset D is defined as follows:

1. Let the library L represent the current empirical distribution *P(X, Y)*.
2. Given a training example (x, y),
3. Classify x using *L* as described in Figure 1. Let the result of classification y^*.
4. Update *L* if $y^* \neq y$.
   - $L_n \leftarrow L_{n-1} + \Delta c(\mathrm{x}, y)$
5. Normalize *L*.
6. Goto step 2 unless the termination condition is met.

Figure 2. The procedure of evolutionary learning of the PLM.

$$E(\mathrm{w}) = \frac{1}{2}\sum_{i \in D} e_i^2. \tag{15}$$

The gradient descent is applied to minimize the error function E(w), i.e.

$$w_j \leftarrow w_j + \Delta w_j, \tag{16}$$

$$\text{where } \Delta w_j = -\eta \frac{\partial E}{\partial w_j}, \tag{17}$$

and η operates as a learning rate which determines the amount of update in each generation. The update rule for gradient search is defined as follows:

$$\Delta w_j = \eta \sum_{i \in D} (P^*(Y \mid \mathrm{x}_i) - P(Y \mid \mathrm{x}_i)) I_{z_{ji}=(\mathrm{x},Y)}. \tag{18}$$

The rule can be modified to perform weight updates for each training example by adopting a stochastic gradient descent, which is to approximate the gradient descent by updating w incrementally, following the calculation of the error for each example. The modified rule is given by

$$\Delta w_j = \eta (P^*(Y \mid \mathrm{x}) - P(Y \mid \mathrm{x})) I_{z_{j=(\mathrm{x},Y)}} \tag{19}$$

$$\approx \Delta c_j. \tag{20}$$

where $P^*(Y \mid \mathrm{x})$ and $P(Y \mid \mathrm{x})$ are the target value and the system output. From Equation (20), the update is implemented by

TABLE I. PERFORMANCE COMPARISON.

Algorithms	Accuracy (%)
PLM	88.24
K-nearest neighbors	85.29
Neural Networks	79.41
Decision Trees	79.41
Support Vector Machine	79.41
Bayesian Networks	73.53

controlling the number of copies c_j with a certain amount of value, Δc_j. Hence, learning procedure becomes a task of adjusting the number of individuals towards minimizing the incorrect output.

We update the library by increasing the count of correctly matched individuals, $c(x,y)$, in case of misclassifying the x. The update rule is defined as follows:

$$L \leftarrow L + \Delta L, \tag{21}$$

$$\text{where } \Delta L = \Delta c(x, y). \tag{22}$$

Figure 2 describes the evolutionary process to adjust the library elements of the PLM classifier based on the stochastic gradient descent. As a new training example (x,y) is given, the individuals matching with x is extracted from the library. If class y^* of x which is predicted by the classifier is correct, no action is performed. If y^* is incorrect, the library is modified according to Equation (21) and (22) to reduce the error of the PLM classifier. When the update occurs, the number of copies of each element is normalized to keep the initial library size.

IV. EXPERIMENTAL RESULTS

We apply the PLM classifier to prostate cancer microarray data obtained from [17, 18]. The training data set contains 52 tumor samples and 50 normal samples with 12,600 genes expression profiles [18] which are measured by the Affymetrix HG-U95Av2 chip. To concentrate on the analysis of the cancer mechanism, cancer-related genes are extracted from the Cancer Gene Census [21]. Hence, the number of genes is reduced from 12,600 to 225. For the PLM classifier, we binarize each expression pattern by setting all elements to '1' if the values are greater than the average expression level of each sample, '0' otherwise. An independent set of test data (25 tumor and 9 normal samples) from [17] is also prepared in the same way.

We use the second order uniform PLM to classify the prostate cancer microarray data. Thus, all individuals in the library are composed of two genes and a class of sample. The library is initialized by randomly selected genes from training examples with the probability of 0.5. Unless the training examples are selected, the individuals are sampled from random examples. The individuals are set to 50,000, and the number of duplicates is initialized to 1,000. The learning rate η is important because it control the adaptability and stability of the library. The larger η causes the library updated more rapidly. In experiment, η is set to 0.005.

TABLE II. HIGH-RANKED GENE COMBINATIONS RELATED TO CANCER. THE GENES ON THE PTEN DEPENDENT PATHWAY AND WNT-SIGNALING PATHWAY ARE MARKED BY SYMBOL * AND †, RESPECTIVELY.

Gene Combinations					
	a	***b***		***a***	***b***
1	CTNNB1*	IGH@	12	BCL3	EIF4A2
2	NPM1	ITGB1†	13	MYH11	SHC1
3	IGH@	MYH11	14	CTNNB1*	EIF4A2
4	BCL3	LASP1	15	CTNNB1*	ILK†
5	EIF4A2	MAPK3†	16	NACA	ILK†
6	CEBPA	COX6C	17	CCND1*	FOXO3A†
7	EIF4A2	IGH@	18	LASP1	ITGB1
8	MYH9	ILK†	19	BCL3	CTNNB1*
9	MAF	MYH9	20	FOXO3A†	MAF
10	BCL3	PDPK1†	21	ILK†	PDPK1†
11	HDAC1*	ILK†	22	CTNNB1*	MAF

TABLE III. BIOLOGICAL PROCESS ENRICHED IN HIGHLY WEIGHTED GENES. OVERREPRESENTED TERMS WERE CHOSEN BY HYPERGEOMETRIC TESTING AND MULTIPLE TESTING ADJUSTMENTS USING THE FALSE DISCOVERY RATE (FDR). *ADJUSTED *P*-VALUE BY FDR.

GO ID	Biological Process	*p-value	Genes
GO:0048469	Cell maturation	1.57E-3	HDAC1, ITGB1, CCND1, MAPK3, CTNNB1, MYH11, LASP1, MYH9, BCL3, MAF, PDPK1, EIF4A2, CEBPA, COX6A, FOXO3A, ILK
GO:0021700	Developmental maturation	1.60E-3	
GO:0030099	Myeloid cell differentiation	1.60E-3	
GO:0030097	Hemopoiesis	1.96E-3	
GO:0048534	Hemopoietic or lymphoid organ development	2.18E-3	
GO:0002520	Immune system Development	6.14E-3	
GO:0045944	Positive regulation of transcription from RNA polymerase II promoter	6.15E-3	

A. Prostate Cancer Classification

Table I summarizes the classification accuracy of the PLM and the conventional machine learning methods. Given the test dataset, the PLM has 88.24% classification accuracy which is the best performance of the applied methods. With the small random combinations of genes in the PLM, we can obtain bet-

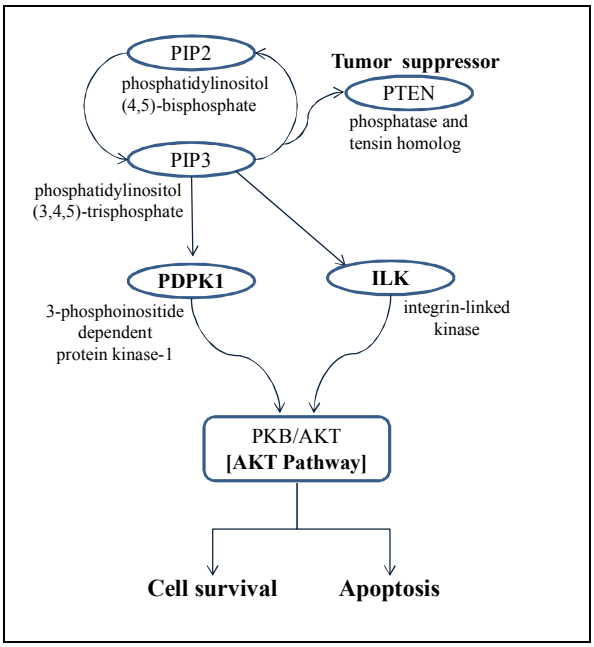

Figure 3. A simple illustration of AKT/PTEN pathway. It is known that the pathway is related to prostate cancer and it is found that PDPK1 and ILK are significant genes for the cancer by our method.

ter classification performance than other methods which use whole genes as input features. The PLM outperforms the similar rule generating or structure inferring methods such as decision trees and Bayesian networks. This means that the PLM can be used as a candidate method to build a reliable system both in classification of prostate cancers and discovery of valuable genes.

We can further improve the performance by increasing the size of library. Moreover, the higher-order library can be made with more computational resources. It allows us to control the library more elaborately for the better classification performance and complex interaction analysis.

B. Finding Prostate Cancer-Related Genes

In the microarray data analysis, the PLM can be used to find cancer-related genes and gene combinations. Table II shows the highly weighted gene combinations obtained from the 10 repeated experiments. The library elements are sorted according to their weights and the 22 highest combinations are presented.

The activation status of cell signaling pathways controls cell fate and deregulation of these pathways demonstrates carcinogenesis. The PTEN/Akt pathway and Wnt signaling pathway are reported to relate to prostate cancer. Thus, prostate cancer is affected by genes on PTEN/Akt pathway and Wnt signaling pathway. The Wnt signaling pathway dysfunction is an important component of prostatic tumorigenesis. By changing the activity of Wnt signaling pathway, prostate cancer cells upset the normal balance between formation and destruction of tumor cells [22]. The PTEN/Akt signaling cascades also play critical roles in the transmission of signals from growth factor receptors to regulate gene expression and prevent apoptosis [23]. Components of this pathway are mutated or abnormally expressed in prostate cancer. Table II shows the high-ranked gene combination list found by our method. Here, many genes are located in the Wnt or PTEN/Akt pathway.

To examine the meaning of selected gene combinations, we compare the functional correlations between genes in Table II with those extracted from the Gene Ontology (GO) term [25]. The GO has become a standard to validate the functional coherence of genes. In the GO, gene functions are organized into three hierarchical trees. These trees are parallel to each other and stand for biological process, molecular function and cellular component, respectively. A node of tree represents a functional annotation of genes. If the genes in Table II are closely related, they might reflect their functional relevance in a specific biological context. We examine significant terms with p-value < 0.01. The results are shown in Table III. Among genes in Table 2, 16 genes (HDAC1, ITGB1, CCND1, MAPK3, CTNNB1, MYH11, LASP1, MYH9, BCL3, MAF, PDPK1, EIF4A2, CEBPA, COX6A, FOXO3A and ILK) are annotated in a significant level. These genes belong to specific functional categories which are related to cell and development maturation, hemopoiesis, cell differentiation and regulation of transcription.

Figure 3 shows the PTEN/Akt pathway known to be related to the prostate cancer. The abnormal condition of PTEN/Akt pathway can induce apoptosis instead of cell survival [24]. In Table II, we find a gene combination, PDPK1 and ILK, located in the pathway. Therefore, it suggests that these gene combinations are significantly related to the prostate cancer.

V. CONCLUSION

We propose an effective method to discover significant gene pairs for cancer classification. The proposed method is based on the probabilistic library model (PLM) which is motivated from molecular computing. Here, gene combinations are selected to build a library, and an evolutionary learning is performed to obtain the joint probability distribution of given samples. Since the combinatorial effects among genes are explicitly considered in the PLM framework, a set of meaningful decision rules can be found after the learning process.

We apply the proposed method to the prostate cancer classification. The empirical evidences support that our method outperforms conventional machine learning algorithms in accuracy. By examining the gene pairs of the library, we also find the cancer-related genes with several candidates of co-regulated genes on the pathway.

One important future research is to investigate the effect of various library orders, especially higher-orders. These complex library elements take the combinations of more than three genes into account. We expect that they can contribute to the discovery and understanding of the complicated relationship among the cancer-related genes.

ACKNOWLEDGMENT

This work was supported by KOSEF through the National Research Laboratory Program (No. M1040000 0349-06J0000-34910) and MOCIE through the Molecular Evolutionary Com-

puting (MEC) Project. C.-H. Park and B.-T. Zhang were also supported by the Ministry of Education and Human Resources Development under the BK21 Program. The ICT at Seoul National University provided research facilities for this study.

REFERENCES

[1] G. Russo, C. Zegar, and A. Giordano, "Advantages and limitations of microarray technology in human cancer," Oncogene, 2003, 22(42), pp. 6497-6507.

[2] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray," Science, 1995, 270 (5235), pp. 467-470.

[3] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J.P. Mesirov, T. Poggio, W. Gerald, M. Loda, E.S. Lander, and T.R. Golub, "Multiclass cancer diagnosis using tumor gene expression signatures," PNAS, 2001, 98(26), pp. 15149-15154.

[4] P.S. Mischel, T.F. Cloughesy, S.F. Nelson, "DNA-microarray analysis of brain cancer: molecular classification for therapy," Nature Review Neuroscience, 2004, 5(10), pp. 782-792.

[5] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring," Science, 1999, 286(5439), pp. 531-537.

[6] A. Cromer, A. Carles, R. Millon, G. Ganguli, F. Chalmel, F. Lemaire, J. Young, D. Dembélé, C. Thibault, D. Muller, O. Poch, J. Abecassis and Bohdan Wasylyk, "Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis," Oncogene, 2004, 23(14), pp. 2484-2498.

[7] P. Warnat, R. Eils, and B. Brors, "Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes," BMC Bioinformatics, 2005, 6(265).

[8] P.T.C. Wan, M.J. Garnett, S.M. Roe, S. Lee, D. Niculescu-Duvaz, V.M. Good, C.G. Project, C.M. Jones, C.J. Marshall, C.J. Springer, D. Barford and R. Marais, "Mechanism of Activation of the RAF-ERK Signaling Pathway by Oncogenic Mutations of B-RAF," Cell, 2004, 116(6), pp. 855-867.

[9] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines," Machine Learning, 2002, 46(1), pp. 389-422.

[10] A. Statnikov, C.F. Aliferis, I. Tsamardinos, D. Hardin and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," Bioinformatics, 2005, 21(5), pp. 631-643.

[11] Y. Wang, I.V. Tetko, M.A. Hall, E. Frank, A. Facius, K.F.X. Mayer, H.W. Mewes, "Gene selection from microarray data for cancer classification−a machine learning approach," Computational Biology and Chemistry, 2005, 29, pp. 37-46.

[12] L. Xu, A.C. Tan, D.Q. Naiman, D. Geman and R.L. Winslow, "Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data," Bioinformatics, 2005, 21(20), pp.3905-3911.

[13] D.R. Rhodes, A.M. Chinnaiyan, "Integrative analysis of the cancer transcriptome," Nature Genetics, 2005, 37, pp. S31-S37.

[14] D.R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, and A.M. Chinnaiyan , "Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression," PNAS, 2004, 101(25), pp. 9309-9314.

[15] B.T. Zhang, H.Y. Jang, "Molecular programming: evolving genetic programs in a test tube," The Genetic and Evolutionary Computation Conference (GECCO 2005), 2005, 2, pp.1761-1768.

[16] S. Kim, M.O. Heo, and B.T. Zhang, "Text classifiers evolved on a simulated DNA computer," IEEE Congress on Evolutionary Computation (CEC 2006), 2006, pp. 9196-9202.

[17] J.B. Welsh, L.M. Sapinoso, A.I. Su, S.G. Kern, J. Wang-Rodriguez, C.A. Moskaluk, H.F. Frierson, Jr. and G.M. Hampton, "Analysis of Gene Expression Identifies Candidate Markers and Pharmacological Targets in Prostate Cancer," Cancer Research, 2001, 61, pp. 5974-5978.

[18] D. Singh, P.G. Febbo, K. Ross, D.G. Jackson, J. Manola, C. Ladd, P. Tamayo, A.A. Renshaw, A.V. D'Amico, J.P. Richie, E.S. Lander, M. Loda, P.W. Kantoff, T.R. Golub, and W.R. Sellers, "Gene expression correlates of clinical prostate cancer behavior," Cancer Cell, 2002, 1, pp. 203–209.

[19] M.H. Garzon, K.V. Bobba, and A. Neel, "Efficiency and reliability of semantic retrieval in DNA-based memories," DNA Computing 9, LNCS 2943, 2004, pp. 157–169.

[20] R.S.Braich, N. Chelyapov, C. Johnson, P.W.K. Rothemund, and L. Adleman, "Solution of a 20-variable 3-SAT problem on a DNA computer," Science, 2002, 296(5567), pp. 499–502.

[21] P. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman and M. Stratton, "A CENSUS OF HUMAN CANCER GENES," Nature Reviews Cancer, 2004, 4, pp. 177-183.

[22] G. Yardy and S. Brewster, "Wnt signalling and prostate cancer," Prostate Cancer and Prostatic Diseases, 2005, 8, pp. 119–126.

[23] J. Paez and W. Sellers, "PI3K/PTEN/AKT pathway. A critical mediator of oncogenic signaling," Cancer Treat Res. 2003, 115, pp. 145 167.

[24] J. McCubrey, L. Steelman, S. Abrams, L, J. Lee, F. Chang, F. Bertrand, P. Navolanic, D. Terrian, R. Franklin, A. D'Assoro, J. Salisbury, M. Mazzarino, F. Stivala, and M. Libra, "Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance," Adv Enzyme Regul., 2006, 46, pp. 249-279.

[25] F. Al-Shahrour, R. Díaz-Uriarte and J. Dopazo, "FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes," Bioinformatics, 2004, 20(4), pp. 578-580.

[26] D.C. Yu, Y. Chen, J. Dilley, Y. Li, M. Embry, H. Zhang, N. Nguyen, P. Amin, J. Oh and D.R. Henderson, "Antitumor Synergy of CV787, a Prostate Cancer-specific Adenovirus, and Paclitaxel and Docetaxel," Cancer Research, 2001, 61, pp. 517-525.

[27] D. Bocangel, M. Zheng, A. Mhashilkar, Y. Liu, R. Ramesh, K.K. Hunt and S. Chada, "Combinatorial synergy induced by adenoviral-mediated mda-7 and Herceptin in Her-2+ breast cancer cells," Cancer Gene Therapy, 2006, 13, pp. 958-968.