Data Mining on Survival Prediction after Chemotherapy for Diffuse Large-B-Cell Lymphoma and Genomics of Metastasis Cancer

Shen LU
Soft Challenge LLC
Little Rock, AR, USA
slu@softchallenge.net

Richard SEGALL
Arkansas State University at Jonesboro
Jonesboro, AR, USA
rsegall@astate.edu

Thomas HAHN
University of Arkansas at Little Rock
Little Rock, AR, USA
tfhahn@ualr.edu

ABSTRACT

This research pertains to the applications of data mining of microarray databases for large-B-cell Lymphoma and metastasis cancer, the latter of which little has been known about the genomic events that regulate the transformation of a tumor into a metastatic phenotype.

1. INTRODUCTION

Microarray technology has found its applications in recent years in many fields of life science. Generally speaking, all the data analysis behind these applications can be characterized into two major categories: (i.) discovery and (ii.) prediction. Discovery is to discover new knowledge, new genes involved in a pathway; prediction is to create predictive models to be used in such areas as toxicology and disease diagnosis. Fundamental to both discovery and prediction is the selection of genes that are differentially expressed (up or down) when comparing the samples of your interest to the control group.

Both discovery and prediction can help make diagnosis in the perspective of the lab research. Microarray analysis should be consistent with the clinical diagnosis. If both of them have the same conclusion, the diagnostic explanation can be accurate with a high probability; but on the other hand, if their conclusions conflict with each other, neither of them can be useful. In this paper, we use data mining techniques to build prediction models using microarray expression data. After that, we further check with the clinical gene signatures in order to find out if the significant genes that can be used to make prediction models for a particular disease, such as lymphoma, are in gene signature which is built based on clinical predictors, such as international prognostic index (IPI).

2. RELATED WORK

The authors Lu and Segall have performed many previous studies on applications of data mining to microarray databases as evidence by references Lu and Segall [((2011), [14]), ((2011), [16])] for application of statistical quality control of microarray gene expression, Lu et al. [((2013), [16]), ((2013), [17])] for comparison of data mining methods on microarray gene expression data on cancer, and Lu et al. ((2013), [18]) as a poster of preliminary research of this paper. Segall ((2006), [23]) ((was one of the first publications in the area of data mining of microarray databases for biotechnology. Segall [((2005), [24]), (2005), [25]) performed data mining of environmental factors on plants. Segall and Pierce [(2009)[26], (2009)[27]] discussed data mining of leukemia cell micro-arrays and Segall and Pierce [(2009)[28]) extended these using self-organized maps. Segall and Zhang ((2007)[29], (2006)[30], (2008)[31]) performed data mining for human lung cancer and other.

Wright et al. ((2003), [34]) used Bayes' rule to classify diffuse large B cell lymphoma (DLBCL) biopsy samples into two gene expression subgroups based on data obtained from spotted cDNA microarrays. They next used this predictor to discover these subgroups within a second set of DLBCL biopsies that had been profiled by using oligonucleotide microarrays. They identified the germinal center B-like (GCB) and activated B-cell like (ABC) DLBCL subgroups which have significantly different 5-yr survival rates after multiagent chemotherapy (62% vs. 26%; p=0.0051), in accordance with the analysis of other DLBCL cohorts.

Wright and Simon ((2003), [35]) proposed a model which can be used to draw gene variances from an inverse gamma distribution and estimate parameters afterwards. The motivation of their work is that DLBCL dataset has limited samples which makes estimation difficult since variance estimates made on a gene by gene basis will have few degree of freedom and the assumptions that all genes share equal variance is unlikely to be true. This model results in a test statistic that is a minor variation of those used in standard linear models and has more power than standard tests to pick up large changes in expression and does not increase the rate of false positives.

Ein-Dor et al. ((2005), [9]) performed research into the overlap genes of microarray expression data in order to find out whether the different results of the same genes are because of different technologies, or because of different patients and different types of analyses. They used a single method to experiment on a breast cancer microarray dataset. The result set of the genes are not unique which is strongly influenced by the subset of patients used for gene selection.

Colomo et al. ((2003), [6]) concluded that microarray gene expression profiling is associated with particular clinicopathological features but is not essential to predicting outcome in DLBCL patients.

Ross et al. ((2003), [22]) demonstrated that expression profiling of leukemic blasts can accurately identify all of the known prognostic subtypes. By analyzing the leukemic blasts microarray gene samples, the newly identified subtype discriminating genes are novel markers for those not identified in previous study. The newly selected genes are highly ranked as class discriminators that have not yet been used and should be used in clinical trials.

Hans et al. ((2004), [11]) divided diffuse large-B-cell lymphoma into prognostically important subgroups with germinal center B-cell like, activated b cell like and type 3 gene expression profiles using a cDNA microarray of the created tissue microarray blocks. They concluded that immunostains can be used to determine the GCB and non-GCB subtypes of DLBCL and predict survival similar to the cDNA microarray.

3. BACKGROUND

3.1 Microarray Profiling

For two-color microarray experiments, as shown in Figure 1, one must decide what the most appropriate comparison is to be made for each array of hybridization. The simplest comparisons can be separated into four general classes, such as direct comparison, reference design, balanced block design and loop design. In many ways, direct comparisons are the simplest conceptually; they are used when two distinct classes of experimental samples are to be compared, such as a treated sample and its untreated control. On each array, representatives of the two classes are paired and co-hybridized together such that the relative expression levels are measured directly on each array. The choice of appropriate pairing depends on the experimental question under study. For example, one can pair diseased and normal tissue from the same patient or randomly select animals from mutual and wild-type groups. The strategy to collect data for any given case is influenced by a wide range of factors, including the availability of samples, the quantity of RNA that can be obtained, the size of the study, and the logistical constraints in the laboratory.

For each gene, the process begins with defining an expression vector that represents its location in expression space. In this view of gene expression, each hybridization represents a separate distinct axis in space, and the log2(ratio) measured for that gene in that particular hybridization represents its geometric coordinate. In this way, expression data can be represented in m-dimensional expression space, where m is the number of hybridizations and where each gene expression vector is represented as a single point in that space. It should be noted that one could use a similar approach to representing each hybridization assay using a sample vector consisting of the expression values for each gene; these define a sample space whose dimension is equal to the number of genes assayed in each array.

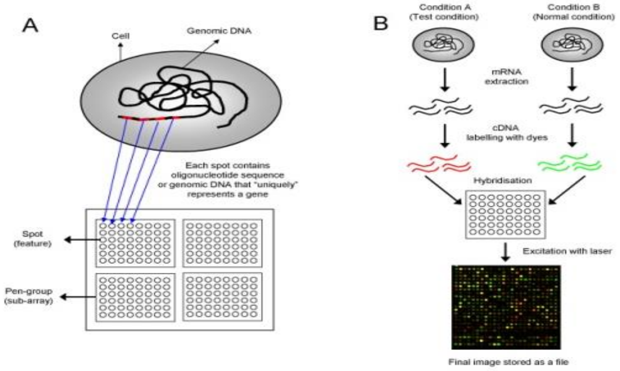

A
Cell
Genomic DNA
Each spot contains oligonucleotide sequence or genomic DNA that "uniquely" represents a gene
Spot (feature)
Pen-group (sub-array)
B
Condition A (Test condition)
Condition B (Normal condition)
mRNA extraction
cDNA labelling with dyes
Hybridisation
Excitation with laser
Final image stored as a file

Figure 1: Illustration of a microarray containing thousands of "spots" of genomic data [2]

3.2 Data Mining using Self-Organizing Maps (SOM) on Microarray Gene Expressions

We refer the reader to a complete discussion of Self-Organizing Maps (SOM) as was presented in our WMSCI 2012 paper Lu and Segall ((2012), [15]) and we are thus providing below a brief discussion.

Self-Organizing Maps (SOM) belong to competitive neural networks. Competitive learning is an adaptive process in which neurons in a neural network are sensitive to different input categories, sets of samples in a specific domain of the input space. ([1], [7], [8], [10], [12], [13], [19], [20], [21], [32])

According to Wikipedia ((2013)[33]), a self-organizing map consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional space. The procedure for placing a vector from data space onto the map is to find the node with the closest (smallest distance metric) weight vector to the data space vector.

A Self-Organizing Map consists of two layers as shown in figure 2. Suppose that we have a set of n-dimensional vectors. The first layer of SOMs is the input data which transfer to the second layer. The second layer has a number of neurons which are chosen arbitrarily and can be used to representing the feature space.

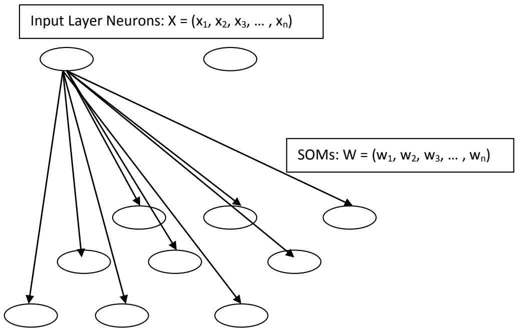

Input Layer Neurons: $X = (x_1, x_2, x_3, \ldots, x_n)$
SOMs: $W = (w_1, w_2, w_3, \ldots, w_n)$

Figure 2: SOMs Architecture

On the second layer, each neuron has the same dimension as the input neuron from the first layer. First of all, weights of the neurons on the second layer are set randomly. During the training process, they have their own weights vector and update those during the training process. When an input x arrives from the first layer to the second layer, the neuron that is best able to represent it wins the competition and is allowed to learn it even better. Moreover, not only the winning neuron but also its neighbors on the lattice are allowed to learn.

4. LYMPHOMA MICROARRAY GENE EXPRESSION PROFILE CLUSTERING

4.1 Background

After multi-agent chemotherapy, two subgroups of diffuse large-B-cell lymphoma had different outcomes. The germinal-center B-cell-like subgroup expressed genes that are characteristic of normal germinal-center B cell were associated with a good outcome. Whereas the activated B-cell-like subgroup expressed genes that are characteristic of activated blood B-cells were associated with a poor outcome. The international prognostic index (IPI) was generally used to stratify patients for therapeutic trials, but, its accuracy is not good enough.

In this paper, we explain how to check patients' genes with microarrays and analyze for genetic abnormalities; find patients with distinctive gene expression profiles; and construct molecular predictors by using genes. There were 160 patients in the training set and 80 patients in the test set. The following three gene expression subgroups were identified: (i.) germinal center B-cell-like, (ii.) activated B-cell-like, and (iii.) type 3 diffuse large-B-cell lymphoma, but only the germinal center B-cell-like subgroup contributed to the lymphoma. Seventeen genes were used to construct a predictor of the survival after chemotherapy. Patients of the germinal center B-cell-like subgroup had the highest survival rate. We compared the accuracy of this predictor with that of the international prognostic index. By using data mining methods to analyze microarray gene expression data, we can create predictors for the survival after chemotherapy.

4.2 Experiments

For hierarchical clustering, we used correlation as similarity measure. We did complete linkage clustering of the 74 significant genes which distinguished between germinal center B-cell lymphoma and activated B-cell lymphoma, as shown in figure 3 and did single linkage clustering of all genes as shown in figure 4. Output describing the meaning of the each node on the hierarchical structure of the 74 significant genes has also been generated.

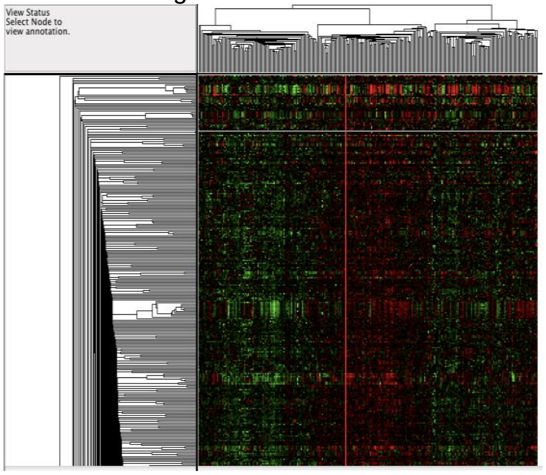

View Status
Select Node to
view annotation.

Figure 3: Visualization of 7399 genes from 275 patient cases

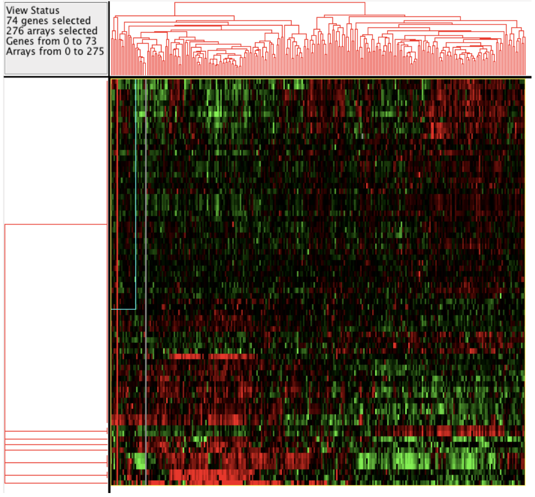

View Status
74 genes selected
276 arrays selected
Genes from 0 to 73
Arrays from 0 to 275

Figure 4: The hierarchical structure of the 74 significant genes which can distinguish germinal center B-cell lymphoma and activated B-cell lymphoma.

5. MICROARRAY GENE EXPRESSION DATA CLASSIFICATION

We used 240 patient cases and 522 significant genes chosen by using t-test (p< 0.01). Three data mining algorithms are tested, which are Naïve Bayesian model, Random Forest model and Self Organizing Map. The experimental results are listed below where TP=True Positive and FP=False Positive.

Table 1: Evaluation of Three Data Mining Models

	TP Rate	FP Rate	Precision	Recall	F-Meature	ROC Area	Class
Naive Bayesian	0.993	0	1	0.993	0.996	1	0
	1	0.007	0.99	1	0.995	0.996	1
Random Forest	1	0.01	0.993	1	0.996	1	1
	0.99	0	1	0.99	0.995	1	1
SOM	0.949	0.01	0.992	0.949	0.97	0.97	0
	0.99	0.051	0.935	0.99	0.962	0.97	1

Table 2: Statistics of Three Data Mining Models

Evaluation Measurement	Naive Bayesian	Random Forest	SOM
Correctly Classified Instances	99.58%	99.58%	96.67%
Incorrectly Classified Instances	0.42%	0.42%	3.33%
Kappa statistic	0.9915	0.9915	0.9323
Mean absolute error	0.0042	0.1255	0.0628
Root mean squared error	0.0645	0.1504	0.1772
Relative absolute error	0.85%	25.67%	12.85%
Root relative squared error	13.06%	30.42%	35.85%
Total Number of Instances	240	240	240

Figure 5: Comparison of the precision and recall on naïve Bayesian, random forest and SOM models

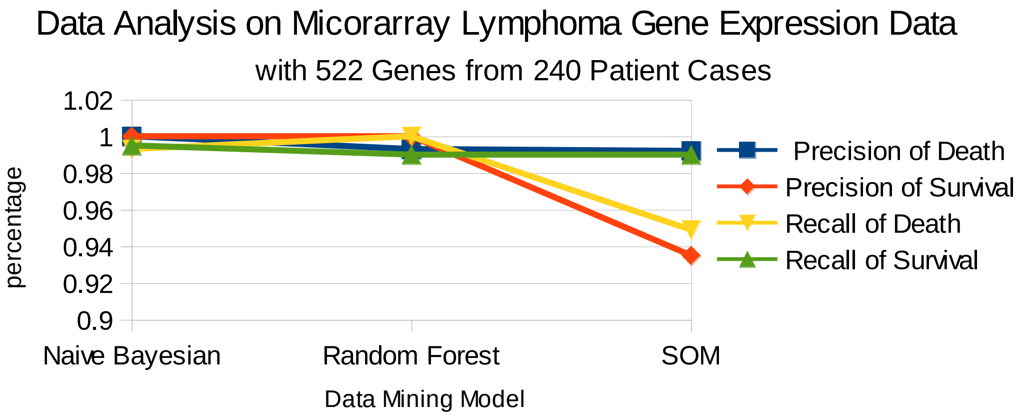


Table 3: Microarray Significant Features in Gene Signature

Gene Signature	Number of Genes	Number of Microarray Features	Percentage of Microarray Features in Signature	PValue
Germinal-Center-B	151	4	2.65%	0.01
Lymph-Node	357	13	3.64%	0.01
MHC-Class II	37	22	59.46%	0.01
Proliferation	1333	288	21.61%	0.01

With the microarray significant features which we used to make predicator, we checked with the gene signatures for germinal center B-cell signature, lynph-node signature, MHC-Class ll signature and proliferation signature, which we used to make predictions in clinical practice. We can see 59.46% MHC-Class ll signature characteristics are microarray significant features, 21.61% proliferation signature characteristics are microarray significant features, 3.64% lymph-Node signature characteristics are microarray significant features, and 2.65% germinal center B-cell signature characteristics are microarray significant features, which means microarray gene expression profiling and particular clinic pathological features are consistent. Therefore, there were 87.36% of microarray significant features in gene signatures and we can conclude that we can use microarray gene expression profiling alone to theoretically predict lymphoma.

6. DATA MINING OF MICROARRAY DATA FOR METASTASIS CANCER

6.1 Background

The data sets selected from the Broad Institute are two of those posted as available with unrestricted access as one of the web links posted on the web page of the Broad Institute Cancer Program Data Sets ((2008),[4]) and is that which is related to the “Genomic analysis of metastasis reveals an essential role for RhoC” research project of the Broad Institute. The selected data base for this research was used by Clark et al. ((2000),[5]) to illustrate the essential role of RhoC that is a member of thee Rho family of proteins that promote reorganization of the cytoskeleton and regulate cell shape, attachment, and motility. Figure 6 from Wikepedia ((2008),[33]) provides an illustration of RhoC also known as “Ras homolog gene family, member C”. According to Wikepedia ((2008),[33]), overexpression of this gene is associated with tumor cell proliferation and metastasis.

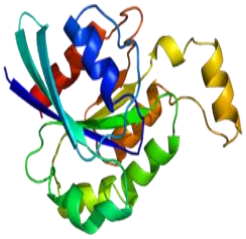


Figure 6: RhoC Genome [Source: Wikipedia ((2013),[33])

6.2 Experiments

The databases utilized for this research in the applications of data mining are those used in Clark et al. ((2000),[5]) as collected at affiliated sites of the Broad Institute ((2008),[3]). These data was collected from human A375 tumor cells, and successive metastases M1, M2 and M3 that were isolated, expanded in tissue culture, and re-introduced into host mice which exhibited more pulmonary metastases. That is M2 data is that collected from those injected with A375M1 cells, and M3 data is that collected from those injected with A375M2 cells. These constitute the first set of data for which data mining had been subjected.

The second data collection was for metastatic A375SM cells grown as a subcutaneous tumor to indicate that the expression of genes is truly intrinsic to the subjected metastatic cells. It was noted by the Broad Institute ((2008,[4]) that the tumor microenvironment may help to regulate the absolute level of gene expression.

The following figures were conducted using SAS Enterprise Miner version 5 using the data from Broad Institute ((2008),[4]) for A375 and A375SM tumor cells of metastasis cancer. Figure 7 and Figure 9 are the self-organized maps showing their frequency and normalized means, Figures 8 and 10 for the cluster proximities, and Tables 4 and 5 are the statistics from the SOM data mining. As can be seen from Figures 8 and 10, that the cluster proximities are generally much smaller of A375SM cells grown as a subcutaneous tumor. Figure 9 shows that the normalized means for A375SM are fewer but more intervals of frequency than those of Figure 6 for A375. Table 4 for A375 cells shows that the magnitude of the statistics are larger for those of the same segments of those of Table 5 for A375SM, indicating that the genes of the subcutaneous tumor are substantially and uniquely different from those of A375 cells of metastasis cancer.

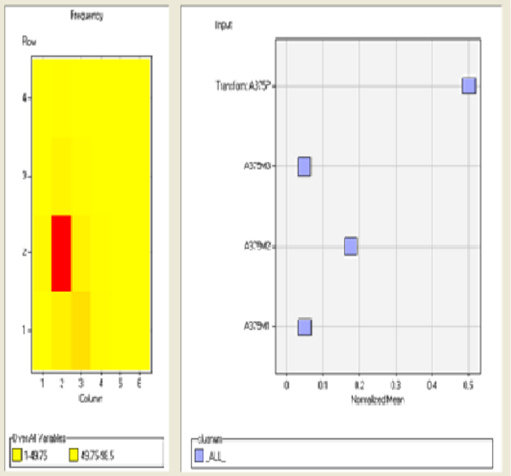


Figure 7: Self-Organized Map for A375 tumor cells

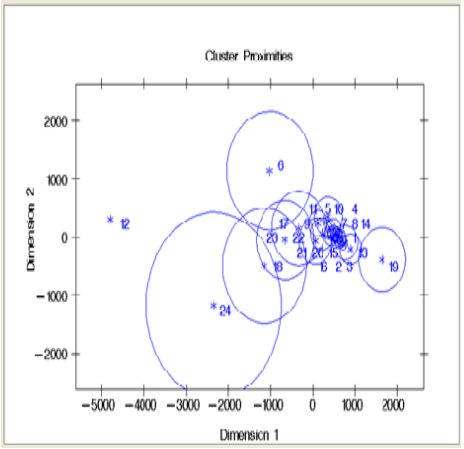
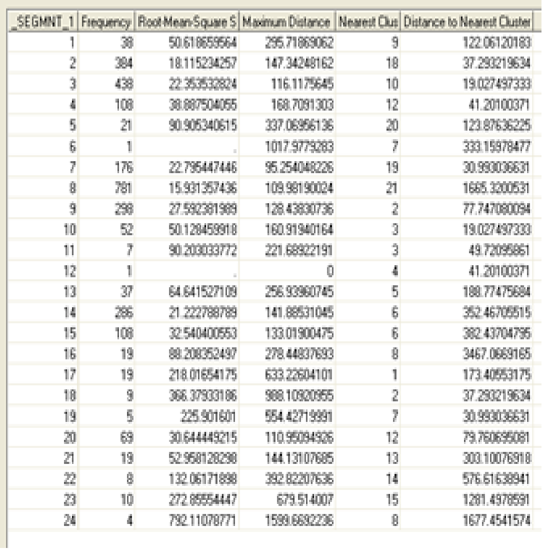


Figure 8: Clusters Proximities for A375 tumor cells



_SEGMNT_1	Frequency	Root-Mean-Square S	Maximum Distance	Nearest Clus	Distance to Nearest Cluster
1	38	50.618659564	295.71869062	9	122.06120183
2	384	18.115234257	147.34248162	18	37.293219634
3	438	22.353532824	116.1175645	10	19.027497333
4	108	38.887504055	168.7091303	12	41.20100371
5	21	90.905340615	337.06956136	20	123.87636225
6	1	.	1017.9779283	7	333.15978477
7	176	22.795447446	95.254048226	19	30.993036631
8	781	15.931357436	109.98190024	21	1665.3200531
9	298	27.592381989	128.43830736	2	77.747080094
10	52	50.128459918	160.91940164	3	19.027497333
11	7	90.203033772	221.68922191	3	49.72095861
12	1	.	0	4	41.20100371
13	37	64.641527109	256.93960745	5	188.77475684
14	286	21.222788789	141.88531045	6	352.46705515
15	108	32.540400553	133.01900475	6	382.43704795
16	19	88.208352497	278.44837693	8	3467.0669165
17	19	218.01654175	633.22604101	1	173.40553175
18	9	366.37933186	988.10920955	2	37.293219634
19	5	225.901601	554.42719991	7	30.993036631
20	69	30.644449215	110.95094926	12	79.760695081
21	19	52.958128298	144.13107685	13	303.10076918
22	8	132.06171898	392.82207636	14	576.61638941
23	10	272.85554447	679.514007	15	1281.4978591
24	4	792.11078771	1599.6692236	8	1677.4541574

Table 4: Statistics from SOM Data Mining for A375 Tumor Cells

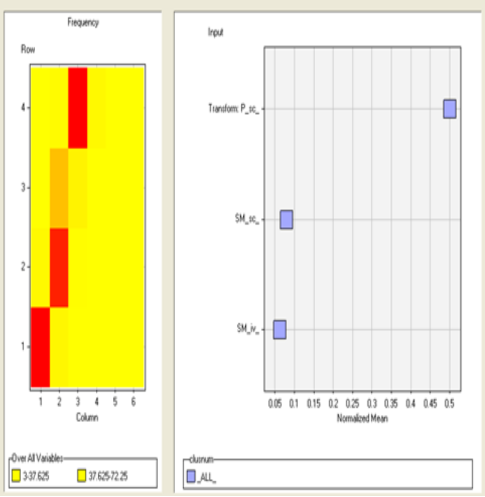


Figure 9: Self-Organizing Maps for A375SM tumor cells

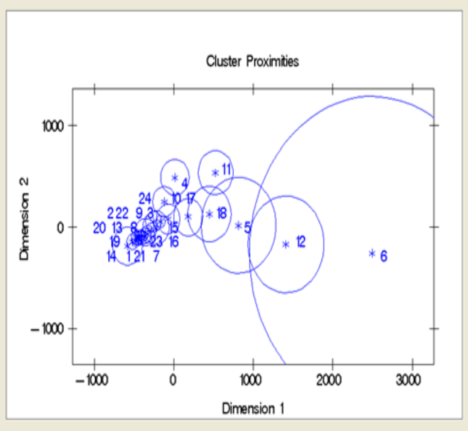
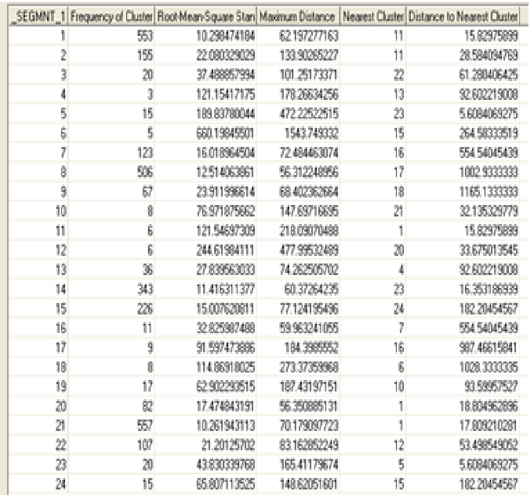


Figure 10: Cluster Proximities for A375SM Tumor Cells



_SEGMNT_1	Frequency of Cluster	Root-Mean-Square Stan	Maximum Distance	Nearest Cluster	Distance to Nearest Cluster
1	553	10.298474184	62.197277163	11	15.82975899
2	155	22.080329029	133.90265227	11	28.584094769
3	20	37.488857994	101.25173371	22	61.280406425
4	3	121.15417175	178.26634256	13	92.602219008
5	15	189.83780044	472.22522515	23	5.6084069275
6	5	660.19845501	1543.749332	15	264.58333519
7	123	16.018964504	72.484463074	16	554.54045439
8	506	12.514063861	56.312248956	17	1002.9333333
9	67	23.911996614	68.402362664	18	1165.1333333
10	8	76.971875662	147.69716695	21	32.135329779
11	6	121.54697309	218.09070488	1	15.82975899
12	6	244.61984111	477.99532489	20	33.675013545
13	36	27.839563033	74.262505702	4	92.602219008
14	343	11.416311377	60.37264235	23	16.353186939
15	226	15.007620811	77.124195496	24	182.20454567
16	11	32.825987488	59.963241055	7	554.54045439
17	9	91.597473886	184.3989552	16	987.46615841
18	8	114.86918025	273.37359968	6	1028.3333335
19	17	62.902293515	187.43197151	10	93.59957527
20	82	17.474843191	56.350885131	1	18.804962896
21	557	10.261943113	70.179097723	1	17.809210281
22	107	21.20125702	83.162852249	12	53.498549052
23	20	43.830339768	165.41179674	5	5.6084069275
24	15	65.807113525	148.62051601	15	182.20454567

Table 5: Statistics from SOM Data Mining of A375SM Tumor Cells

7. CONCLUSIONS

This paper discussed an open issue in Microarray gene expression application. For the lymphoma study two hierarchical structures of microarray gene expression data were built with 7399 genes and 74 significant genes, that visualized the characteristics of the microarray gene expression profiles. We used naïve Bayesian model, random forest model and self organizing maps (SOM) to predict lymphoma with microarray gene expression profile from 240 patients. The experimental results showed us that lymphoma can be predicted with microarray gene expression data by using naïve Bayesian, random forest and SOM algorithms. We further compared the difference between clinic pathological features and microarray features by using gene signatures for Germinal center B-cell like lymphoma, lymph-node lymphoma, proliferation lymphoma and MHC-Class II lymphoma. We can conclude that, since clinical features and microarray features are associated with each other, the predictions from both clinic pathological features and microarray gene features are consistent. For the metastasis cancer study we concluded that data mining of the microarrays using SOM was effective in distinguishing the uniqueness of the genes of the subcutaneous tumors.

REFERENCES

[1.] Amari, S-I., "Topographic organization of nerve fields". ***Bulletin of Mathematical Biology***. v42, n. 3, 1980, pp.339-364.

[2.] Babu, M.M., "An Introduction to Microarray Data Analysis", Chapter 11 in ***Computational Genomics*** by R. Grant, Editor, Horizon Press, UK, 2004, pp. 225-249, http://www.mrclmb.cam.ac.uk/genomes/madanm/microarray/chapter-final.pdf

[3.] Broad Institute (2008), http://www.broad.mit.edu/index.html

[4.] Broad Institute Cancer Program Data Sets (2008), www.broad.mit.edu/cancer/datasets.cgi

[5.] Clark, E. A., Golub, T. R., Lander, E.S., Hynes, R.O. (2000) **"**Genomic analysis of metastasis reveals an essential role for RhoC", ***Nature***, v. 406, n.6785, August 3, pp. 532-535, www.broad.mit.edu/cgi-bin/cancer/datasets.cgi

[6.] Colomo, L., A. Lopez-Guillermo, M. Perales, S. Rives, and A. Martinez. "Clinical impact of the differentiation profile assessed by immunophenotyping in patients with diffuse large B cell lymphoma". ***Blood***. 2003 101: 78-84. doi: 10.1 1182/blood-2002-04-1286

[7.] Didday, R. L., (1970) *The Simulation and Modeling of Distributed Information Processing in the Frog Visual System*. PhD thesis, Stanford University.

[8.] Didday, R. L., (1976) "A model of visuomotor mechanisms in the frog optic tectum". ***Mathematical Biosciences***, 30:169-180.

[9.] Ein-Dor, L. , I. Kela, G. Getz, D. Givol and E. Domany. "Outcome signature genes in breast cancer: is there a unique set?" ***Bioinformatics.*** Vol. 21, no. 2, 2005. pp. 171-178. doi: 10 1093/bioinformatics/bth469

[10.] Grossberg, S., "On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion system". ***Biological Cybernetics***. v21, 1976, pp. 145-159.

[11.] Hans, C.P., D. D. Weisenburger, T. C Greiner, R. D. Gascoyne, J. Delabie, G. Ott, H. K. Muller-Hermelink, E. Campo, R. M. Braziel, E. S. Jaffe, Z. Pan, P. Farinha, L. M. Smith, B. Falini, A. H. Banham, A. Rosenwald, L. M. Staudt, J. M. Connors, J. O. Armitage and W. C. Chan. "Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistocheistry using a tissue microarray". ***Blood.*** 2004 103:275-282. DOI:10. 1182/blood-2003-05-1545

[12.] Kohonen, T., K. Mkisara and T. Saramki, "Phonotopic maps insightful representation of phonological features for speech recognition". ***Proceedings of the Seventh International Conference on Pattern Recognition***, 1984, pp. 182-185.

[13.] Kohonen, T., "Self-organized formation of topologically correct feature maps". ***Biological Cybernetics***. v43. 1982, pp.59-69.

[14.] Lu, S. and R.S. Segall, "Statistical Quality Control of Microarray Gene Expression Data," ***Proceedings of 15th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2011***, Orlando, FL, July 19-22, 2011, pp. 206-211.

[15.] Lu, S. and R.S. Segall, "Multi-SOM: An Algorithm for High Dimensional, Small Size Datasets", ***Proceedings of 16th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2012***, Orlando, FL, July 17-20, 2012, p. 219-224.

[16.] Lu, S. and Segall, Richard S., "Statistical Quality Control of Microarray Gene Expression Data," ***Journal of Systems, Cybernetics and Informatics (JSCI),*** Vol.9, No. 7, 2011, pp. 63-68.

[17.] Lu, S., R.S. Segall, and T. Hahn, "Comparison of Data Mining Methods on Microarray Gene Expression Data on Cancer", Poster presented at the ***National Science Foundation (NSF) Bioinformatics Workshop to Foster Collaborative Research*****,** Little Rock, AR, March 3-5, 2013.

[18.] Lu, S., T. Hahn and R.S. Segall, "Data Mining on Survival Prediction after Chemotheraphy for Diffuse Large B-Cell Lymphoma", Poster presented at ***Joint Meeting of the Southern Section of the American Society of Plant Biologists (ASPB) and P3 Symposium*****,** Little Rock, AR, April 6-9, 2013.

[19.] Malsburg, C.v,d,, "Self-organization of orientation sensitive cells in the striate cortex". ***Kybernetik***. v14, 1973, pp.85-100.

[20.] Nass, M. and L. N. Cooper, A theory for the development of feature detecting cells in visual cortex. ***Biological Cybernetics***. v19, n. 1, 1975, pp. 1-18.

[21.] Pérez, R. , L. Glass and R. J. Shlaer, Development of specificity in cat visual cortex. ***Journal of Mathematical Biology***. v1. 1975, pp. 275-288.

[22.] Ross, M.E., X. Zhou, G. Song, S. A. Shurtleff, K. Girtman, W. Kent Williams, H-C Liu, R. Mhfouz, S. C. Raimondi, N. Lenny, A. Patel and J. R. Downing, "Classification of prediatric acute lymphomoblastic leukemia by gene expression profiling". ***Blood***. 2003 102: 2951-2959. doi: 10. 1182/blood-2003-01-0338

[23.] Segall, R. S. , "Data Mining of Microarray Databases for Biotechnology," ***Encyclopedia of Data Warehousing and Mining***, Edited by John Wang, Montclair State University, USA; Idea Group Inc., 2006, ISBN 1-59140-557-2, pp.734-739.

[24.] Segall, R. S., "Data Mining of Microarray Databases for the Analysis of Environmental Factors on Corn and Maize," ***2005 Conference of Applied Research in Information Technology,*** sponsored by Acxiom Laboratory for

Applied Research (ALAR), University of Central Arkansas (UCA), February 18, 2005.

[25.] Segall, R. S., “Data Mining of Microarray Databases for the Analysis of Environmental Factors on Plants Using Cluster Analysis and Predictive Regression”, Proceedings of the ***Thirty-sixth Annual Conference of the Southwest Decision Sciences Institute***, vol. 36, no. 1, March 3-5, 2005, Dallas, TX.

[26.] Segall, R. S. and R. M. Pierce, “Advanced Data Mining of Leukemia Cell Micro-Arrays”, ***Proceedings of 13th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2009***, Orlando, FL, July 10-13, 2009.

[27.] Segall, R. S. and R. M. Pierce, “Advanced Data Mining of Leukemia Cell Micro-Arrays”, ***Journal of Systemics, Cybernetics and Informatics (JSCI)***, Volume 7, Number 6, 2009, pp.60-66.

[28.] Segall, R. S., and R. M. Pierce, “Data Mining of Leukemia Cells using Self-Organized Maps,” ***Proceedings of 2009 Conference on Applied Research in Information Technology***, sponsored by Acxiom Laboratory of Applied Research (ALAR), University of Central Arkansas (UCA), Conway, AR, February 13, 2009, pp. 92-98.

[29.] Segall, R.S. and Zhang, Q., “Data Mining of Microarray Databases for Human Lung Cancer”, ***Proceedings of the Thirty-eighth Annual Conference of the Southwest Decision Sciences Institute***, vol. 38, no. 1, March 15-17, 2007, San Diego, CA.

[30.] Segall, R.S. and Zhang, Qingyu, “Data Visualization and Data Mining of Micro-Array Databases for Biotechnology”, ***Proceedings of the 2006 Conference of Applied Research in Information Technology***, sponsored by Acxiom Laboratory for Applied Research (ALAR), University of Central Arkansas, March 3, 2006, pp. 121-128.

[31.] Segall, R. S. and Q. Zhang, “Data Mining of Forest Cover and Human Lung Micro-array Databases for Bioinformatics”, ***Poster Session at Arkansas EPSCoR (Experimental Program to Simulate Competitive Research)P3 (Plant-Powered Production) Conference**,* Winthrop Rockefeller Institute of University of Arkansas System, Petit Jean Mountain, AR, August 20-22, 2008.

[32.] Swindale, N. V., “A model for the formation of ocular dominance stripes”, ***Proceedings of the Royal Society of London*, B**, 215, pp. 211-230, 1980.

[33.] Wikipedia (2013), RhoC, http://en.Wikipedia.org/wiki/RhoC

[34.] Wright, G., B. Tan, A. Rosenwald, E. H. Hurt, A. Wiestner, and L. M Staudt. “A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma”. ***PNAS***, August 19, 2003. vol.100, no, 17. www.pnas.org⍰cgi⍰doi⍰10.1073⍰pnas.1732008100

[35.] Wright, G.W. and R. M. Simon. “A random variance model for detection of differential gene expression in small microarray experiments”. ***Bioinformatics***. Vol. 19, no. 18, 2003. pp. 2448-2455. doi: 10. 1093/bioinformatics/btg345. ftp://linus.nci.nih.gov/pub/techreport/RVM_supplement.pdf

[36.] Zhang, Q. and R. S. Segall, “Data Mining of Forest Cover and Human Lung Micro-Array Databases with Four-Selected Software”, ***Proceedings of the 2007 Conference of Applied Research in Information Technology***, sponsored by Acxiom Laboratory for Applied Research (ALAR), University of Arkansas-Fayetteville, March 9, 2007.