

## Two Novel Nonparametric Methods for Cancer Diagnosis through Microarray Analysis

Jesús A. Rodríguez\*; Luis Rivero\*; Matilde L. Sánchez-Peña\*; Clara E. Isaza\*†; Mauricio Cabrera-Ríos\*

**Diagnosing cancer using microarray analysis to study differential gene expression has been a recent focus of intense research. Although several very sophisticated analysis tools have been developed with this aim in mind, it still remains a challenge to keep these methods free of parametric adjustments as well as maintain their transparency for the final user. Nonparametric methods in general have been associated with these last two characteristics, thus becoming attractive tools for microarray analysis in cancer research. In particular, diagnosing cancer via microarray analysis is an exercise whereby tissue is characterized according to its differential gene expression levels. In this manuscript, two novel nonparametric methods for cancer diagnosis using microarray data are described and their performance assessed against a baseline approach that utilizes the Mann-Whitney test for median differences. Both methods show promising results in terms of their potential use in making diagnoses. [P R Health Sci J 2010;3:305-311]**

*Key words: Microarray analysis, Nonparametric statistics, Cancer Diagnosis Tools*

Cancer consists of a series of diseases characterized by the uncontrolled growth and dispersion of abnormal cells. Cancer has assumed a global importance in the field of public health due to its associated mortality rates (1) and economic impact. Cancer diagnosis has traditionally been carried out through morphologic characterization, although in recent years there has been an increasing interest in supporting this method with genetic profiling (2). Microarray experiments, which first appeared in 1995 (3), can be used to quantify the relative expression of tens of thousands of genes in a simultaneous manner, thereby providing a convenient way to determine genetic profiles. These experiments usually generate large amounts of information, which are stored in databases and then moved to electronic repositories. Several studies have used microarrays to characterize gene expression (4-6).

One important task when analyzing microarray data is that of determining which genes changed their expressions significantly from one state to another, for example, from tissues in a cancerous state to tissues in a healthy state. In general, the procedure in which such a task is undertaken is known as gene filtering and has been extensively explored due to its potential for recognizing a reduced number of genes, which recognition can offer a shortcut to illness diagnosis, prognosis, and treatment (4-5, 7-21). Gene filtering has been explored through a variety of techniques based on normal distribution, such as the 2 sample t-test (17), ANOVA (22), and the Welch t-test (19), among others. Some authors stress the fact that gene expression data do not follow a normal distribution (18, 21, 23), proposing the

use of nonparametric statistical tests such as the Mann-Whitney (MW) test (18), also known as the Wilcoxon test.

Genes selected through a filtering procedure can be used for many purposes. Of particular interest to this study was defining a classifier to determine whether a given tissue belongs to a particular category (i.e. cancer or healthy) through measuring the relative expressions of the selected genes. Thus, the interest was on developing a cancer diagnosis that is based on classification.

As a precedent, our research group has previously proposed a strategy (based on the Wilcoxon test) to carry out gene filtering and tissue classification (24-25), aiming first for simplicity rather than performance. In this study, using this initial strategy as a baseline, classification performance was targeted through the development of two new methods. The first method employed the Wilcoxon test for gene filtering and classification; however, this revised method introduced a gene-set selection step right after filtering to enhance classification performance. The second method capitalized on this new structure, and used the Nemenyi-Damico-Wolfe (NDW) multiple comparison nonparametric test as a distinctive enhancement strategy. For brevity, the descriptions of the Wilcoxon and the NDW tests have been omitted here but

\*Department of Industrial Engineering University of Puerto Rico at Mayagüez; †Department of Biology Universidad Autónoma de Nuevo León, México

Address correspondence to: Mauricio Cabrera-Ríos, PO Box 9000, ININ –UPRM, Mayagüez, PR 00681-9000. Tel: (787) 832-4040, ext. 3240 • E-mail: mauricio.cabrera1@upr.edu

can be readily perused in a textbook on nonparametric statistical methods, such as that of Hollander and Wolfe (26).

The structure of this paper is as follows: In the next section, microarray databases are described in general, along with the details of the proposed methods. The computational setting is then discussed in the ensuing section, followed by an assessment of the classification performance of the proposed methods vs. that of the baseline approach. Finally, conclusions are drawn and future plans are described.

### Methods

As explained previously, microarray experiments usually generate large amounts of information, which are stored in databases and then moved to electronic repositories. The focus in this work is on the general case in which the database of interest is laid out in a matrix form, with  $n$  rows representing genes and columns containing  $l$  healthy tissues (class 1) and  $m$  carcinogenic tissues (class 2). At each intersection, the expression level of a particular gene (row) on a specific tissue (column) is available.

For the purposes of the performance assessment proposed in this work, it was necessary to extract tissues from both of the classes defined previously to create a reference set (R). The remaining tissues made up a validation set (V). R and V were, then, mutually exclusive. Tissues in R were used to carry out all of the necessary computations for each method; tissues in V were used to assess the classification performance of each method with previously untried material.

The baseline diagnostic strategy (24-25) consisted of three stages: gene filtering, class separability, and classification. An important modification was the introduction of an additional stage (called the *selection* stage), which included the identification of a set of genes with the best separability properties for the two classes of interest. The term “validation” is used here to refer to the separability stage to avoid any confusion with the selection stage. Figure 1 shows a schematic description of the general strategy that was followed according to the methods proposed here. A more detailed description of each stage follows.

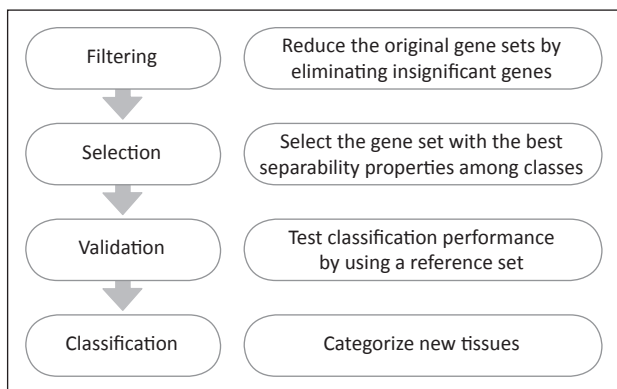


Figure 1. General scheme of the proposed methods

### Filtering

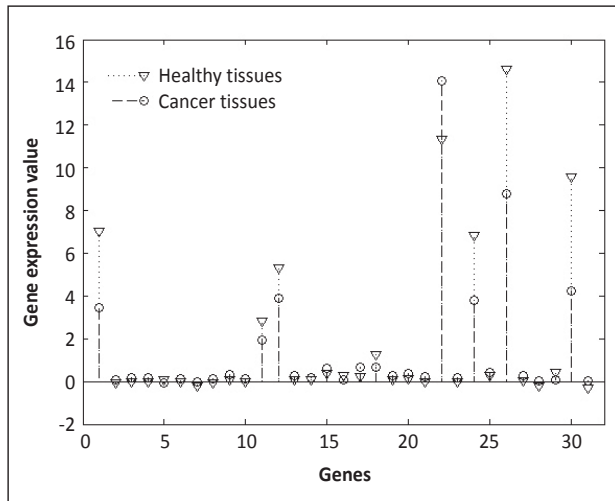
In this stage, the objective was to reduce the original gene set to include only those genes that showed significantly different levels of expression in the two classes (healthy, cancer). To this end, a two-sided Wilcoxon test for the difference of medians was applied to each gene (using a significance value of  $\alpha$ ). The reduction of the number of genes was then reflected in the number of rows for all sets, R and V.

A filtering stage (used to search for cancer biomarkers) is also common in microarray analysis, in which a rather high discrimination rate is important and, therefore, a single constant threshold value for  $\alpha$  is not desirable. Several modifications to vary  $\alpha$ , including the well-known Bonferroni correction, can be used in those cases (27). Many authors have also used the False Discovery Rate (FDR) correction (21, 23, 28-32), which was originally defined by Benjamini and Hochberg (33) as the expected proportion of errors among the rejected hypotheses. In terms of gene selection, FDR represents the expected number of genes identified as being significant without actually being important for the analysis. However, when classification is the objective—as was the case in this study—a more liberal approach to gene filtering can be adopted so as to favor generalization capabilities and therefore better classification performance. A larger set of genes for classification purposes was pursued in this study; thus, no correction was used for gene filtering.

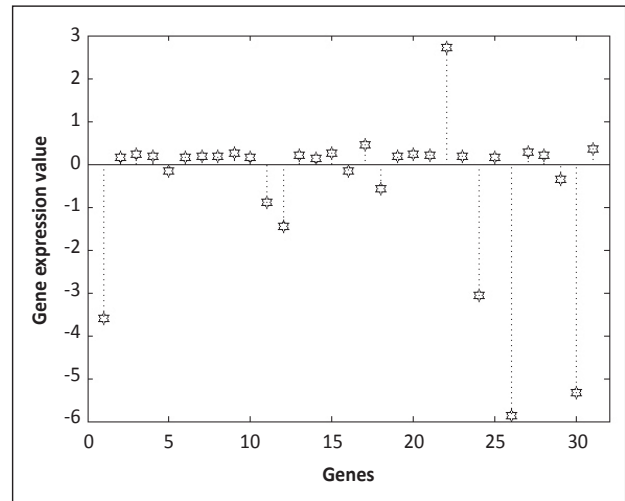
### Selection

Even though a reduced set of significant genes was obtained from the filtering stage, the medians of healthy tissues may differ from one gene to another in their relative position with respect to the medians of cancer tissues. In other words, some genes were overexpressed and others were underexpressed in one condition with respect to the other condition. This situation is shown graphically in Figures 2 and 3. Referring to the case presented by these figures, the median expression of all of the cancer tissues (0.28395) can be seen to be larger than the median expression of all of the healthy tissues (0.1131); however, in the sample of 31 genes shown below, several instances show the median expression of the cancer tissues as being below that of the healthy tissues. A clear example of this can be seen in gene 26 in both figures. It has been empirically noted that this inconsistent behavior severely hampers classification performance.

In order to deal with this issue, it was proposed that the significant genes (from the filtering stage) be further categorized as overexpressed or underexpressed. A gene (previously deemed significant) with a median expression across the cancer tissues being larger than the median expression across the healthy tissues was considered to be overexpressed. Conversely, a gene with a larger median expression across the healthy tissues was deemed underexpressed. Two distinct groups of significantly expressed genes resulted from this process (*overexpressed and underexpressed*).



**Figure 2.** Medians of tissues from a filtered expression matrix for a sample of 31 genes.



**Figure 3.** Difference of medians of cancer tissues - healthy tissues from a filtered expression matrix for a sample of 31 genes.

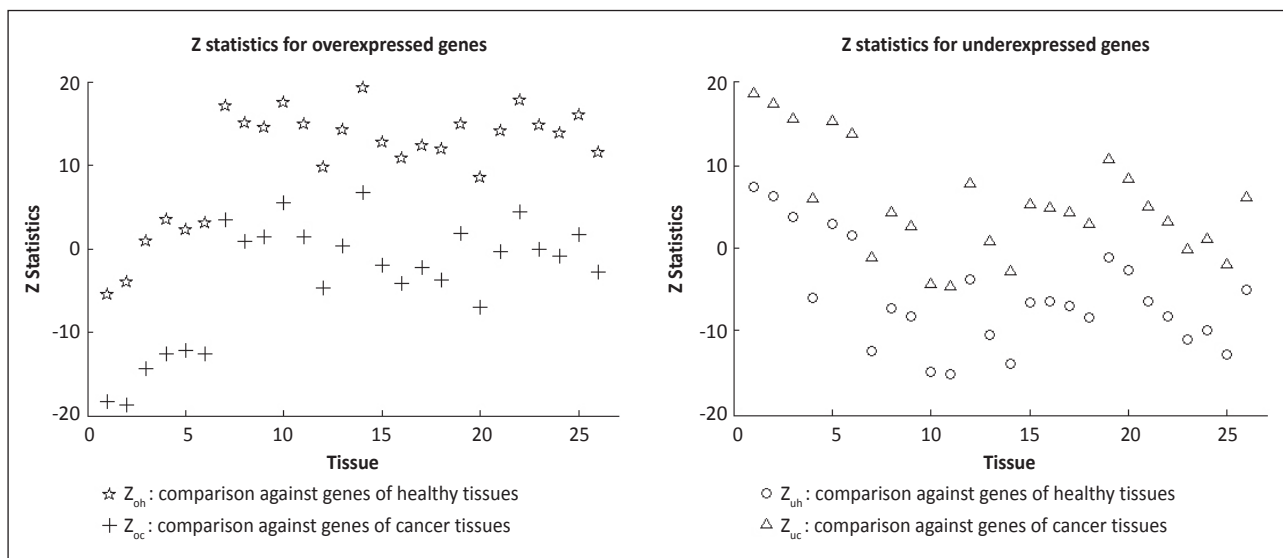
This selection process provided the main difference between the strategies that were proposed in this study (as explained in the next section). It should be remembered at this point that the baseline strategy did not include a selection step.

**Method 1. Selection with the Wilcoxon test**

Using the two-sided Wilcoxon test, each tissue of the reference set R characterized in the filtered genes can be compared to (1) the set of tissues belonging to the healthy class and to (2) that of the cancer class. If the *overexpressed* or *underexpressed* category is now taken into consideration, four quantities estimating the difference between medians can be computed as follows: Let the difference values be represented

by the statistic  $Z_{ijk}$  for the  $k^{th}$  tissue in the reference set being compared against the  $j^{th}$  class (*Healthy* or *Cancer*), using the significant gene set of the  $i^{th}$  category (*Overexpressed* or *Underexpressed*). Utilizing this notation for the 4<sup>th</sup> tissue in a reference set (by way of an example) will result in the following associated values:  $Z_{OH4}$ ,  $Z_{OC4}$ ,  $Z_{UH4}$ , and  $Z_{UC4}$ . In order to provide competitive classification capability in the posterior analysis stages, the four different statistics are now compared and a search for the category showing the best separability between classes is undertaken.

Figure 4 shows the results—using a database first described in Wong et al. (4) in which 43 cervix tissue samples were subjected to microarray experiments—obtained for the 26



**Figure 4.** Z values corresponding to 26 tissues. The first 6 tissues belong to the healthy class and the rest to the cancer class. Z values were plotted separately for overexpressed (left scatter plot) and for underexpressed genes (right scatter plot).

tissues contained in a reference set (6 tissues from the *healthy* class and 20 from the *cancer* class). Overall, the category with the greatest distance between the z-values of both classes and with the lowest within-class variability should be selected. From Figure 4, it can be appreciated that z-values from an overexpressed category (with its two associated classes) show the most competitive characteristics. The standardized mean of the chosen category can be seen in Figure 5.

One way to formalize this comparison is through the use of a larger-is-better index that includes variability within each class and separability between classes. For each category  $i \in \{O, U\}$ —where O means *overexpressed* and U means *underexpressed*—the index will be defined as follows:

$$Index_i = \frac{Abs \left( \frac{\sum_{k \in H} Z_{iHk} - \sum_{k \in C} Z_{iCk}}{|H| - |C|} \right)}{\sqrt{\sigma_{iH}^2 + \sigma_{iC}^2}}$$

where |H| and |C| are the number of healthy and cancer tissues in the reference set, respectively;  $\sigma_{iH}^2$  and  $\sigma_{iC}^2$  are the sample variances of the Z-statistics pertaining to the healthy and cancer classes, respectively; and Abs is the absolute value function.

**Method 2. Selection using Multiple Comparison via the Nemenyi-Damico-Wolfe (NDW) test**

As in the previous method, the starting point was the data in the reference set organized by classes (H for *healthy* and C for *cancer*) and categories (U for *underexpressed* and O for *overexpressed*). For each category, two Nemenyi-Damico-Wolfe tests were carried out, one comparing against the tissues in the healthy class and another one comparing against the tissues in the cancer class. The NDW multiple comparison test was used to determine the median differences of multiple samples vs. a control group preserving a family-wise error rate. This is a one-sided test that is based on the calculation of joint rankings; thus, it has a nonparametric nature. Because of its one-directionality, care must be exercised regarding how the hypotheses are set and, therefore, how the Z statistic must be computed. Since two tests are required for each of the matrices, a first computation on the differences of average joint rankings according to the category of the group under analysis must be carried out utilizing the formulae found in Table 1. In this table, the Rs stand for average joint rankings, and a dot substituting for a subindex indicates the term over which the average is computed. The undotted subindices indicate the groups involved in the joint ranking.

In order to obtain the statistics ( $Z_{ijk}$ ) of interest, the definition of  $T_{ijk}$  from table 1 needs to be used, resulting in the following expression:

$$Z_{ijk} = \frac{T_{ijk}}{\sqrt{\left( \frac{1}{n_{ij}} + \frac{1}{n_k} \right) \left( \frac{N_{ij}^2 + N_{ij}}{12} \right)}}$$

where  $n_{ij}$  and  $n_{ik}$  are the number of measurements of gene expressions in category  $i$ , for the control group in class  $j$ , and for the  $k_{th}$  tissue, respectively.  $N_{ij}$  is the number of observations in the complete sample; that is,

$$N_{ij} = n_{ij} + \sum_{k \in R} n_{ik}$$

The use of a larger-is-better index (as defined in the previous section) aided in the selection of the most adequate category for classification purposes.

**Table 1.** Determining the direction of the hypotheses to be used for multiple comparisons when analyzing tissue  $k$ . In this table, the Rs stand for average joint rankings, and a dot substituting for a subindex indicates the term over which the average is computed. The undotted subindices indicate the groups involved in the joint ranking.

		Category	
		Overexpressed	Underexpressed
		Cancer median > Healthy median	Healthy median > Cancer median
Class of the control group	Cancer	$T_{OCk} = R_{\cdot Ok} - R_{\cdot OC}$	$T_{UCk} = R_{\cdot Uk} - R_{\cdot UC}$
	Healthy	$T_{OHk} = R_{\cdot Ok} - R_{\cdot OH}$	$T_{UHk} = R_{\cdot Uk} - R_{\cdot UC}$

**2.3 Validation**

In order to estimate the expected performance of the methods, tissues of the validation set (V) were used to “simulate” the classification of new tissues. The control group was compounded by the tissues of the reference set (R). Only those genes belonging to the category chosen in the previous step (whether overexpressed or underexpressed) were used in this test.

For each tissue of the validation set, a Z statistic was calculated following either method 1 (Wilcoxon test) or method 2 (Nemenyi-Damico-Wolfe test). Z statistics obtained in this step were used to assign tissues of the validation set to the healthy or cancer class, according to their posterior probability when compared against the Z statistics distribution of the tissues of the reference set.

**2.4 Classification**

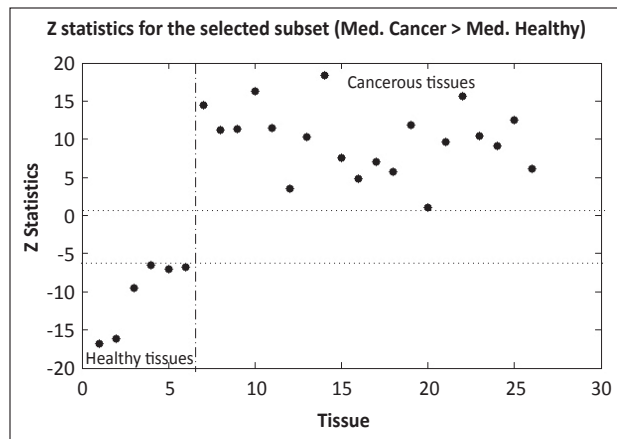
Finally, the resulting classification was compared with the actual state of each tissue to assess performance on key indicators.

**3. Computational Setting**

The methods proposed here are relatively straightforward and were implemented using functions available from the MATLAB Statistics Toolbox™ (The MathWorks, Inc., Natick, Mass.). The authors can be contacted for a copy of the code.

Both methods proposed here were tested against the baseline strategy discussed in Isaza-Brando et al. (24), which study used the Wilcoxon test. In a previous evaluation, the baseline strategy was deemed more competitive at handling a low





**Figure 5.** Mean Z statistic of the category with the best separability (with overexpressed genes)

number of replicates, unequal variances, and nonnormality than several normality-based approaches (24); therefore, the latter approaches were not considered for this evaluation.

Three factors at different levels were considered for this validation study: (i) Database (cervix cancer (4), colon cancer (5), pancreatic cancer (6)), (ii) Significance value  $\alpha$  (0.10, 0.05, 0.01, 0.005), and (iii) Proportion of tissues in the reference set and in the validation set (80/20 split, 70/30 split, 60/40 split).

The characteristics of the databases used in this study can be seen in Table 2. Also, once a proportion was chosen for the reference and validation sets, a random selection of tissues for validation purposes was carried out so as not to bias the results. Finally, if the split did not yield an integer, the result was rounded up to the next integer.

**Table 2.** Characteristics of the databases used in the validation study.

Cancer Type	Number of Healthy Tissues	Number of Cancer Tissues	Number of genes	Source
Cervix	8	35	10,692	(4)
Colon	18	18	7,457	(5)
Pancreas	5	31	12,687	(6)

## Results

Table 3 summarizes the classification percentages of the cancer tissues correctly categorized across all databases. The 80/20 split is shown here since, as expected, the more information contained in the reference set, the better the classification percentages obtained.

From Table 3 it is clear that both of the proposed new methods outperformed the baseline strategy, while maintaining similar levels of performance between them. Due to this similarity of performance, their superior classification performance when compared to the baseline strategy is attributed to the addition of a gene-set selection phase.

**Table 3.** Comparative results for the proposed methods vs. the baseline strategy using the previously mentioned 80/20 split. Classification percentage of cancer tissues correctly categorized across all databases.

		Reference set (80 % of data)	Validation set (20% of data)
<b>Baseline Strategy</b>	Average	59.8%	56.2%
	Median	57.7%	53.6%
	Min.	48.3%	25.0%
	Max.	76.9%	85.7%
<b>Method 1</b>	Average	95.2%	92.0%
	Median	100.0%	100.0%
	Min.	50.0%	50.0%
	Max.	100.0%	100.0%
<b>Method 2</b>	Average	93.3%	91.1%
	Median	100.0%	100.0%
	Min.	50.0%	50.0%
	Max.	100.0%	100.0%

Table 4 shows the sensitivity and specificity values resulting from the proposed methods across all databases. *Sensitivity* estimates the ability to detect true positives (correctly detect cancer tissues), and *specificity*, the ability to detect true negatives (correctly detect healthy tissues). In general, high sensitivity values in a diagnostic test—though ideal—often come at the expense of specificity. This holds true for the diagnostic methods developed in this study, as can be corroborated in Table 4.

Table 5 shows the results of systematically varying the significance value,  $\alpha$ . As can be appreciated, reducing the significance value improves the classification performance. This last result should, however, be considered carefully when choosing  $\alpha$ , since too low a value could significantly reduce the number of genes after the filtering stage and unnecessarily weaken the large-sample approximations used throughout the proposed methods. The authors recommend a value not below 0.01 for classification purposes, based on the empirical evidence presented.

The results presented here provide evidence that nonparametrics are ideal for classification purposes. Pending, however, is the application of the methods described herein to unknown tissues so that the effectiveness of said methods might be more rigorously assessed. In addition, it is not the intention of this manuscript to imply that microarray data analysis can alone provide an accurate diagnosis of cancer; a great deal of information must be considered before arriving at such a conclusion.

## Conclusions and Future Studies

In this study, the issue of diagnosing cancer via tissue classification, using microarray-generated data, was approached. In particular, two nonparametric methods were introduced that capitalized on a gene-set selection stage to dramatically improve the rate of correctly classified cancer tissues. The methods performed satisfactorily without requiring complex computational approaches and by keeping user-defined parameters at a minimum.

**Table 4.** Sensitivity and specificity values for the proposed diagnostic methods, across all databases, using an 80/20 split.

			Method 1				Method 2			
			Mean	Median	Max.	Min.	Mean	Median	Max.	Min.
Across all databases	Sensitivity	Reference	99%	100%	100%	71%	100%	100%	100%	95%
		Validation	97%	100%	100%	50%	100%	100%	100%	100%
	Specificity	Reference	79%	100%	100%	0%	74%	100%	100%	0%
		Validation	69%	100%	100%	0%	67%	100%	100%	0%

**Table 5.** Correct classification percentages for cancer tissues in the verification set for different alpha values and using the 80/20 split.

Significance Value, $\alpha$	Baseline Strategy	Method 1	Method 2
0.1	47.2%	86.3%	88.1%
0.05	56.0%	88.1%	88.1%
0.01	60.9%	96.8%	91.3%
0.005	60.9%	96.8%	96.8%

The robustness of the methods was demonstrated through the use of different databases with several combinations of tissues in the reference and validation sets.

Future studies in this line of research will include the application of the proposed methods to the classification of truly unknown tissues with the purpose of establishing a more rigorous assessment. One probable next step will be to link analyzed tissues to available patient information while simultaneously considering current knowledge regarding cancer biomarker genes so that diagnostic capabilities might be strengthened.

### Resumen

El diagnóstico de cáncer es un área de investigación muy activa dentro del análisis de microarreglos. Esencialmente, para poder diagnosticar se busca identificar un patrón de variación de la expresión diferencial genética que se pueda asociar al cáncer de tal manera que sea distinguible de otros estados de salud. Aunque se han desarrollado herramientas de análisis muy sofisticadas para identificar estos patrones, queda aún camino por recorrer para tener métodos cuyo funcionamiento sea completamente transparente para los usuarios finales y que no requieran un ajuste excesivo de parámetros de funcionamiento. Los métodos no paramétricos han sido tradicionalmente asociados con estas dos últimas características, por tal razón, se han convertido en herramientas cuyo uso es atractivo para el análisis de microarreglos en el estudio del cáncer. En particular, el diagnóstico de cáncer a través de análisis de microarreglos se puede entender como un problema de clasificación de tejidos cuando el tejido bajo análisis está caracterizado

por sus niveles diferenciales de expresión genética. En este artículo se describen dos métodos no paramétricos para el diagnóstico de cáncer basado en datos de microarreglos. El desempeño de clasificación de los métodos propuestos se comparan con una estrategia de base que utiliza la prueba

Mann Whitney para diferencias de medianas en dos poblaciones. Ambos métodos propuestos muestran resultados promisorios para su uso en diagnóstico de cáncer.

### Acknowledgements

This study was made possible thanks to UPRM-BioSEI grant 330103080301, awarded to M. Cabrera-Rios at UPRM; C.E. Isaza acknowledges the support she received from this grant as a collaborator. For their assistantships, J. Rodríguez and M. Sánchez thank the Department of Industrial Engineering at UPRM; the support of the UPRM-BioSEI grant was also much appreciated. L. Rivero is grateful for the support received from NSF REU Grant 0851879 (PI: V. Cesani).

### References

- Danaei G, Vander Hoorn S, Lopez AD, et al. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet*. 2005 Nov 19;366(9499):1784-1793.
- Golub TR, Slonim DK, Tamayo P, et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science*. 1999;286(5439):531-537.
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995 Oct 20;270(5235):467-470.
- Wong YF, Selvanayagam ZE, Wei N, et al. Expression Genomics of Cervical Cancer : molecular classification and prediction of radiotherapy response by DNA microarray. *Clinical Cancer Res*. 2003;9(15): 5486-5492.
- Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional Gene Expression Profiles of Colorectal Adenoma, Adenocarcinoma, and Normal Tissue Examined by Oligonucleotide Arrays. *Cancer Res*. 2001;61(7): 3124-3130.
- Iacobuzio-Donahue CA, Maitra A, Olsen M, et al. Exploration of Global Gene Expression Patterns in Pancreatic Adenocarcinoma Using cDNA Microarrays. *Am J Pathol*. 2003;162(4):1151-1162.
- Gollub J, Ball CA, Binkley G, et al. The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res*. 2003;31(1):94-96.
- Shyamsundar R, Kim Y, Higgins J, et al. A DNA microarray survey of gene expression in normal human tissues. *Genome Biol*. 2005;6(3):R22.
- Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet*. 2000 Mar;24(3): 236-244.
- Lee JW, Lee JB, Park M, Song SH. An extensive comparison of recent classification tools applied to microarray data. *Comput Stat Data An*. 2005 Apr 1;48(4):869-885.

11. Welsh JB, Zarrinkar PP, Sapinoso LM, et al. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc Natl Acad Sci U S A*. 2001;98(3):1176-1181.
12. Alon U, Barkai N, Notterman DA, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A*. 1999;96(12):6745-6750.
13. Khan J, Wei JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med*. 2001 Jun;7(6):673-679.
14. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A*. 2001;98(24):13784-13789.
15. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000 Feb 3;403(6769):503-511.
16. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet*. 2000 Mar;24(3):227-235.
17. Dudoit S, Yang YH, Callow MJ, Speed TP. Statistical Methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat Sin*. 2002;12:111-139.
18. Troyanskaya OG, Garber ME, Brown PO, et al. Nonparametric methods for identifying differentially expressed genes in microarray data. *Bioinformatics*. 2002 Nov 1;18(11):1454-1461.
19. Pan W. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics*. 2002 Apr 1;18(4):546-554.
20. Townsend J, Hartl D. Bayesian analysis of gene expression levels: Statistical quantification of relative mRNA level across multiple strains or treatments. *Genome Biol*. 2002;3(12):RESEARCH0071.
21. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98(9):5116-5121.
22. Kerr MK, Afshari CA, Bennett L, et al. Statistical Analysis of a Gene Expression Microarray Experiment with Replication. *Stat Sin*. 2002;12:203-217.
23. Zhang S. A comprehensive evaluation of SAM, the SAM R-package and a simple modification to improve its performance. *BMC Bioinformatics*. 2007;8(1):230.
24. Isaza-Brando C, Uribe Mastache L, Perez Vicente HA, et al. Cancer Diagnosis through Microarray Analysis using the Mann-Whitney statistical test. In: *Proceedings of the Industrial Engineering Research Conference*. Miami, FL: Institute of Industrial Engineers; 2009:763-740.
25. Perez Vicente H, Uribe Mastache L, Cabrera-Ríos M, Isaza-Brando C. Diagnóstico de cáncer a partir de datos de microarreglos. In: *Memorias del VI Congreso Internacional en Innovación y Desarrollo Tecnológico*. Cuernavaca, Morelos, México; 2008.
26. Hollander M, Wolfe DA. *Nonparametric Statistical Methods*. 2nd ed. New York, NY: Wiley-Interscience, John Wiley & Sons, Inc.; 1999.
27. Draghici S. *Data Analysis Tools for DNA Microarrays*. 1st ed. London, England: Chapman & Hall; 2003.
28. Efron B, Tibshirani R. Empirical bayes methods and false discovery rates for microarrays. *Genet Epidemiol*. 2002 Jun;23(1):70-86.
29. Efron B, Storey JD, Tibshirani R. Microarrays empirical bayes methods, and false discovery rates [Report]. Stanford, Ca: Stanford University; 2001 [cited 2009 Nov 6]. Available at: <http://www-stat.stanford.edu>
30. Efron B, Tibshirani R. On testing the significance of sets of genes. *Ann Appl Statistics*. 1(1):107-129.
31. Subramanian A, Kuehn H, Gould J, et al. GSEA-P: A desktop application for Gene Set Enrichment Analysis. *Bioinformatics*. 2007 Dec 1;23(23):3251-3253.
32. Cui X, Hwang JTG, Qiu J, Blades NJ, Churchill GA. Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics*. 2005 Jan;6(1):59-75.
33. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met*. 1995;57(1):289-300.