



Determining juvenile cancer types from gene expression using gene contribution and differential analysis

Eric S. Hald^{1†}, Ryan J. Stoner^{1†} and Derrick K. Rollins^{1,2*}

*Correspondence: drollins@iastate.edu



CrossMark

← Click for updates

[†]These authors contributed equally to this work.

¹Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa 50011, USA.

²Department of Statistics, Iowa State University, Ames, Iowa 50011, USA.

Abstract

Background: The usefulness of DNA microarrays is limited to the efficacy of methods of gene importance analysis available, which could have far-reaching implications in diagnosis, discovery, and treatment of the genetic nature of diseases. This article applies a powerful differential method (DM) based on principle component analysis (PCA) to a vast DNA microarray data set containing data from 88 samples of juvenile small round blue cell tumors.

Methods: Using this DM for ranking the most critical genes from microarray data, the top 25 genes in the resulting rank-ordered list were associated with four functional categories: directly cancer-related, protein synthesis, cell cycle control, and neurological function.

Conclusions: The strength of the DM is demonstrated in the ability to tie these functions to previous cancer research. The results show the method's ability to differentiate between different types of similar juvenile cancers and that the method could also be useful in exploring the genetic nature of various diseases.

Keywords: Data mining, microarrays, gene expression data, principle component analysis, bioinformatics

Introduction

With the advent of the human genome project and countless other research projects devoted to gene identification and function, several methods of analyzing the resulting vast arrays of data have arisen [1-3]. Some of the most useful gene data has been found to be in DNA microarray data sets. Microarray data from various organisms have been used to suggest the function as well as the contribution of a multitude of genes [1,2]. In order to process these data sets, several methods have been developed and proposed as accurate means by which to divulge useful information about genes from the vast DNA microarray data sets. Methods ranging from clustering to neural networks to PCA have made for both intriguing and varied results in determining gene expression from DNA microarrays [1-4]. These methods all have their own unique pros and cons relating to their use in analysing data sets.

We sought out microarray data with high variability to demonstrate the ability of the Rollins and Teh [4] differential method (DM) to handle variability in data sets. This search yielded a DNA microarray developed for Khan et al., [2]. The microarray contains data taken from a study of 84 different samples, or *arrays*, composed of four different types of childhood cancers classified as small round blue cell tumors: neuroblastoma (NB), non-Hodgkin lymphoma (BL), rhabdomyosarcoma (RMS), and Ewing sarcoma (EWS) [3]. Five of the samples included

in the microarray were non-cancerous in nature and made up of various normal human tissues. Various methods [2,3] for analyzing this data set have been applied in the hopes of determining the genetic differences between these similar childhood cancers. The realization of the genetic differences between these similar cancers has potential for both diagnosis and treatment. Microarray analysis provides for fascinating possibilities in the realm of gene expression, the focus of the application of the method to this data set.

In this study we employ the Rollins and Teh [4] differential method (DM), an extension of the method developed in Rollins et al., [1], to a DNA microarray from Khan et al., [2]. We hypothesize that this application will showcase the powerful capabilities of the comprehensive method to span the differences found in various sets of DNA microarray data. The DM allows the experimenter to choose between two types of contribution plots, known as eigengene (EG) or eigenassay (EA). This method can also process very large sets of data, discarding the need for data reductions that may unintentionally eliminate important data. In this article, we explain the methodologies behind the DM and apply it to the complete, unabridged Khan data set. Next, we use the resulting gene data to make comparisons to results obtained in previous studies. Functions of the genes found to be differentially expressed will be stated and an analysis of these functions will be shown

to give a greater understanding of the power of the DM.

Methods

The expression data derived from the DNA microarray is defined as X , an m by n matrix. Along the rows of X there are m genes expressed, and along the columns n assays are expressed. This means that each cell of the matrix is the expression level of the i^{th} gene in the j^{th} assay and is given as x_{ij} . Also, *variables* for X , or any matrix, are along the columns, and *measurements* for X , or any matrix, are along the rows. This means that for X the assays are *variables* while genes are *measurements*. An exception is the transposed matrix of X , or X^T , where the *variables* are expressed along the rows and *measurements* are expressed along the columns. The DM is applied to the microarray gene expression data, and since the DM is based on PCA, the microarray expression data matrix orthogonally transforms into a new coordinate data system. This space is used to maximize composite variability [1-4]. In short, the new coordinate system has the first axis point towards the greatest spread in data and the second axis point orthogonal to the first axis. This is defined as the PC loading matrix. Eigenvectors, the principal components (PCs), are found from the correlation matrix, or the covariance matrix. The first PC has the largest eigenvalue; thus, it is ranked first. We derive the EG contribution plot, an n by n loading matrix, from X and the EA contribution plot, an m by n loading matrix, from X^T [4].

Eigengene contribution plot approach

After applying PCA to X , the values within the EG scores matrix, an m by n pseudo data matrix, must be found. From there, the EG differential gene contribution between two groups can be determined. Before finding the elements of the scores matrix, however, X must standardize into the matrix Z . Each element of Z is defined as z_{ij} and must follow the equation:

$$z_{ij} = (x_{ij} - \bar{x}_j) / s_j$$

where the sample mean and sample standard deviation of the data column j are \bar{x}_j and s_j , respectively. To find the elements of S^{EG} , the EG scores matrix, the equations

$$\begin{aligned} s_{ij}^{EG} &= l_{1j}^{EG} z_{i1} + l_{2j}^{EG} z_{i2} + \dots + l_{nj}^{EG} z_{in} = \sum_{k=1}^n l_{kj}^{EG} z_{ik} \\ &= g_{ij1}^{EG} + g_{ij2}^{EG} + \dots + g_{ijn}^{EG} = \sum_{k=1}^n g_{ijk}^{EG}; \quad (1) \end{aligned}$$

$$i = 1, \dots, m; j = 1, \dots, n; k = 1, \dots, n$$

apply where s_{ij}^{EG} is the score for the i^{th} gene using the j^{th} vector of EG loadings, l_{ij}^{EG} is the i^{th} loading for the j^{th} EG vector, and g_{ijk}^{EG} is the contribution for the i^{th} gene, on the k^{th} assay from the j^{th} EG loading vector. Next, two groups are chosen so that

$A =$ Group A with n_A assay members and $B =$ Group B with n_B assay members. No members should be in common between both groups such that

$$2 \leq n_A + n_B \leq n \quad (2)$$

Before giving the EG differential contribution it is necessary to give the mean contribution for the i^{th} gene from the j^{th} EG loading vector from Group A and Group B, respectively. The equations are given as

$$g_{ij}^{-EG_A} = \frac{1}{n_A} \sum_{\text{over } k'} l_{k'j}^{EG} z_{ik'} = \frac{1}{n_A} \sum_{\text{over } k'} g_{ijk'}^{EG} \quad (3)$$

$$g_{ij}^{-EG_B} = \frac{1}{n_B} \sum_{\text{over } k''} l_{k''j}^{EG} z_{ik''} = \frac{1}{n_B} \sum_{\text{over } k''} g_{ijk''}^{EG} \quad (4)$$

where the assay members are k' and k'' in Groups A and B, respectively. Now we can find the EG differential gene contribution between both groups for the i^{th} gene from the j^{th} EG loading vector by the equation,

$$dg_{ij}^{-EG} = g_{ij}^{-EG_A} - g_{ij}^{-EG_B} \quad (5)$$

The EG differential gene contribution will also be known as $T_{diff_j}^{EG}$ [4].

Eigenassay contribution plot approach

The EA contribution plot approach applies PCA to X^T ; thus, creating an m by n loading matrix that will be used to find the values within the EA scores matrix, an n by n pseudo data matrix. From there the EA differential gene contribution between two groups can be determined. Unlike the EG contribution plot approach, X^T does not need to be standardized. This allows us to solve for the elements within S^{EA} , the EA scores matrix, with the following equations:

$$\begin{aligned} S_{ij}^{EA} &= l_{1j}^{EA} x_{1i} + l_{2j}^{EA} x_{2i} + \dots + l_{mj}^{EA} x_{mi} \\ &= \sum_{p=1}^m l_{pj}^{EA} x_{pi} = \sum_{p=1}^m g_{ijp}^{EA}; \quad (6) \end{aligned}$$

$$i = 1, \dots, n; j = 1, \dots, n; p = 1, \dots, m$$

The score for the i^{th} assay using the j^{th} vector of EA loadings is S_{ij}^{EA} . l_{ij}^{EA} is the i^{th} loading for the j^{th} EA vector, and g_{ijp}^{EA} is the contribution for the p^{th} gene on the i^{th} assay from the j^{th} EA loading vector. Next, we need to choose two groups so that

A = Group A with n_A assay members and B = Group B with n_B assay members. One group should not contain any members in common with the other such that it follows Eq. 2. Again the mean contribution must be found, but this time it is for the p^{th} gene from the j^{th} EA loading vector for both groups. The equations are given as:

$$g_{jp}^{-EA_A} = \frac{l_{pj}^{EA}}{n_A} \sum_{\text{over } i'} x_{pi'} = \frac{1}{n_A} \sum_{\text{over } i'} g_{i'jp}^{EA} \quad (7)$$

$$g_{jp}^{-EA_B} = \frac{l_{pj}^{EA}}{n_B} \sum_{\text{over } i''} x_{pi''} = \frac{1}{n_B} \sum_{\text{over } i''} g_{i''jp}^{EA} \quad (8)$$

where the assay members are i' and i'' in Groups A and B, respectively. Now we can find the EA differential gene contribution between both groups for the p^{th} gene from the j^{th} EA loading vector by the equation,

$$dg_{jp}^{-EA} = g_{jp}^{-EA_A} - g_{jp}^{-EA_B} \quad (9)$$

The EA differential gene contribution will also be known as $T_{diff,p}^{EA}$ [4].

Results and discussion

Application of the DM

Our study involves the application of the DM developed in Rollins and Teh [4]. The DM was created to determine a ranked order set of genes that express most differently in two or more groups of assays. Initial testing of the DM involved the use of simulated and real gene data for two groups of assays, mutant and wild type mice. In simulated data study two hundred of the 40,000 simulated mouse genes were purposely differentiated between the two groups of mice. The DM was able to distinguish these 200 genes based on the differential between their gene contributions [4]. With the success of the DM when applied in this work, this is the first application to real cancer data that we are calling the Khan data set. For the Khan data set, there are essentially five different groups of assays—the four different cancer types and the non-cancerous samples. In applying the DM, we show that it may have far-reaching possibilities in a relatively quick and thorough analysis of DNA microarray data, specifically data sets with discernible groups of assays.

We applied the DM to the unfiltered complete data set, which included all 6,567 genes. We discovered that the EG loading plots yielded more interesting groupings for analysis than their EA score plot counterparts. For the top ten PCs, PC 5 yielded a strikingly similar and more refined grouping than that of PC 6 in the DM analysis. Figure 1 shows this grouping of 11 assays, visibly separated from the rest of the 88 assays. Biologically interpretable grouping is desired when using

the DM. After finding such a grouping, the contribution associated with the 11 distinguished assays and the differential contribution associated with the remaining 77 assays are found. Using the knowledge of the original data set, we found that PC 5 separates all of the BL samples from the rest of the samples. Thus, we applied the DM to find the differential gene expression between samples with BL to those without BL. After finding the individual contributions of both groups of assays, the differential contribution, $T_{diff,p}^{EG}$, is found through simple subtraction of the linear combination contribution values for each gene. Figure 2 shows rank ordered $T_{diff,p}^{EG}$ values, from highest to lowest, plotted against the gene index. The genes of interest are those with the highest values, as these correspond to a high degree of differential expression between the two

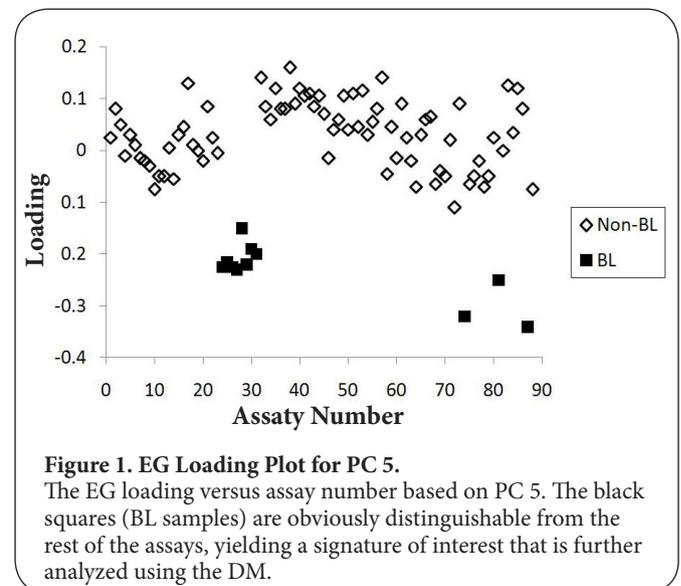


Figure 1. EG Loading Plot for PC 5.
 The EG loading versus assay number based on PC 5. The black squares (BL samples) are obviously distinguishable from the rest of the assays, yielding a signature of interest that is further analyzed using the DM.

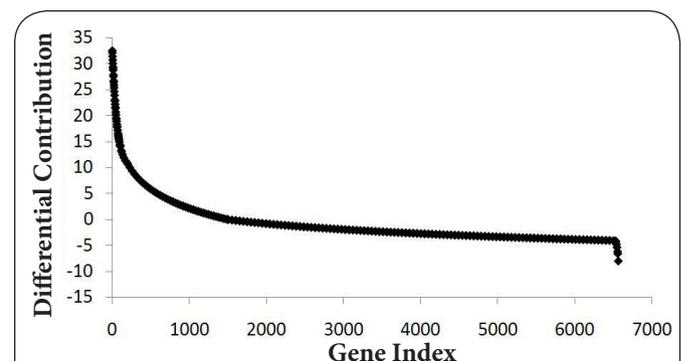


Figure 2. Rank-ordered Differential Contributions for BL Assays versus non-BL Assays.
 Rank-ordered differential contribution, from equation 5, plotted against gene index. $T_{diff,p}^{EG}$ EGs ranked in decreasing order. Differential is between BL assays and non-BL assays. Genes of importance are found in the top left of the graph and are more spread out than the majority of the genes.

groups of assays. We chose to further analyze and compare the top 25 genes of interest to investigate the effectiveness of the DM. **Table 1** lists the top 25 genes determined by the DM, including their abbreviations and commonly known functions. Next, we utilize these genes and their functions are utilized to determine the success of the DM based upon past studies and the genes' possible contributions to the cancers being studied.

Top genes analysis

In analyzing and comparing the top genes determined by the DM, we did two types. First, we compared our top genes to those found in previous studies by other methods. Second, we related the functions of the top genes to the cancers being studied. The two previous studies that use the same microarray data, Khan et al., [2] and Pal et al., [3], include lists of genes that supported the results of their methods. In directly comparing our top 25 list of genes, none of them can be found

in either of the corresponding important genes lists in those two studies. In fact, when we increase our list of top genes to 100, only two genes, number 66 and number 77, can be found to match the top 96 genes found by Khan et al., [2] with none occurring in the seven important genes found by Pal et al., [3]. While these results may seem discouraging, further analysis of the two previous studies yields prior existence of such discrepancies within these two older methods as well as possible short comings in their filtering techniques. Of the seven genes found to be important using relational fuzzy clustering in Pal et al., [3], only four may be found in the top 96 genes reported in Khan et al., [2]. This discovery reveals a common disparity found when using different methods to analyze the vast DNA microarray data sets. In addition, seven of our top 100 genes were found to be absent from the analysis involving the previous studies. These genes were subjectively filtered out of the data set before any analysis on the data was completed. This, coupled with the pervasive nature of

Table 1. Top 25 genes based on differential between gene contributions between BL and non-BL samples.

Gene Symbol	Gene Function
<i>RPL36A</i>	role in tumor cell proliferation
<i>TUBA1A</i>	gene expression in morphologically differentiated neurologic cells
<i>RPLP0</i>	protein coding; inhibits translation
<i>EEF2</i>	protein synthesis
<i>STMN1</i>	destabilization of microtubules; role in cell cycle
<i>HNRPA1</i>	pre-mRNA processing; myometrial protein in parturition
<i>SPEG</i>	growth regulation in differentiation of vascular smooth muscle
<i>TERT</i>	cellular senescence; maintenance of telomere ends; repressed in postnatal cells
<i>RPL6</i>	gastric cancer resistance; binds to C domain of human T-cell leukemia virus
<i>RPS5</i>	variable expression in colorectal cancers
<i>EEF1G</i>	regulation of assembly of multisubunit complexes; enzymatic delivery to ribosomes
unknown	Microtubules
<i>GABRB1</i>	mediates fastest inhibitory synaptic transmission in central nervous system
<i>GNAL</i>	activation of adenylyl cyclase; mediation of hormones/neurotransmitters
<i>HMGBI</i>	unwinding of double-stranded DNA; binding of preferentially single-stranded DNA
<i>GAPDH</i>	catalysis of energy-yielding step in carbohydrate metabolism
<i>HNRPC</i>	pre-mRNA processing; mRNA metabolism and transport
<i>SRc25a3</i>	transporter for energy-linked phosphate anion transport
<i>BTF3</i>	required for transcriptional initiation
<i>RPS15A</i>	S8P family; 40S subunit
unknown	Unknown
<i>RPS4X</i>	encodes for S4; possible role in Turner's syndrome
<i>EIF4A1</i>	required for mRNA binding to ribosome; involvement in cap regulation
<i>ACY1</i>	expression reduced or undetectable in small-cell lung cancer cell lines and tumors
<i>UBA52</i>	targets cellular proteins for degradation; maintenance of chromatin structure; regulation of gene expression and stress response

Listing of top 25 genes found to have the greatest differential based on gene contribution analysis. Genes were rank-ordered by these differential values. Genes in bold designate genes possibly linked to cancer according to the *Atlas of Genetics and Cytogenetics in Oncology and Haematology* [13].

disagreement in analysis techniques, allows one to have confidence in the DM's continued application in analyzing DNA microarray data.

To support the list of top contributing genes found by the DM, we next discuss these genes' functions. In determining our genes' contributions to the cancers being considered, we reviewed the current gene data using "Gene" [14]. The resulting gene functions generally fell into one of four different categories: specific links to cancer, protein synthesis, cell cycle regulation, and neurological association. These four functional categories will be used to show the biological importance of the genes chosen by the DM.

Some of the top 25 genes have generalized or specific relations to cancer. In addition to being expressly identified by their contributions to cancers, some additional genes have also been considered to have possible links to cancer. Of the top 25 genes, 16 have been considered to be associated with some type of cancer according to the *Atlas of Genetics and Cytogenetics in Oncology and Haematology* [13]. This means that over three-fifths of the genes found to be important by the DM have already been considered to have a possible association with the development of cancer. With the Human Genome Project being completed, it can only be expected that this number will increase as researchers uncover more of the mysteries involved with genes and gene function. Some genes have been directly named after the role they play in cancer. The top gene in our analysis, *RPL36A*, has a vital role in tumor cell proliferation. It encodes for the protein L36a and has been shown to be over-expressed in hepatocellular carcinoma. The means by which *RPL36A* becomes over-expressed is unknown, but one particular result of its over-expression is applicable to our study. Previous studies of the gene have linked its over-expression to accelerating the cell cycle [5]. The cell cycle plays a vital role in all cancers and their proliferation, as cells must complete multiple cycles in the process of proliferating and metastasizing. The DM picked out this gene because it showed differential expression in the BL samples as opposed to the other samples. *RPL36A* may be over-expressed to a much greater or less degree in BL than in any of the other cancers. The ninth gene in our analysis, *RPL6*, is associated with gastric cancer resistance. The gene encodes for the protein L6, and its high expression in tumors increases the growth rate of the cells. Counteracting the proliferation with down regulation may suppress the transition from the G₁ phase to the S phase; thus impeding the cell cycle and proliferation. Also, *RPL6* expression has ties to prognosis and can function as a biomarker for gastric cancer [12]. These findings may pave the way for more in-depth research regarding *RPL36A* and *RPL6* involvement in small round blue cell tumors. Such research would greatly aid in showing the power and potential of the DM.

The largest portion of the top 25 list of genes was genes linked to protein synthesis. Proteins are the building blocks of all life and are also essential to cancer cells as they proliferate.

Protein synthesis is a complicated process, but the relation of these genes to the process has several varied consequences when considering a link to cancer. For example, *EEF2*, number four in our list, involves a kinase that has been shown to possibly promote cancer cell survival, particularly in glioblastoma cells [6]. In order for cancer to proliferate, it must have a plethora of proteins for use in cell division and growth. Another of the top 25 genes, *ACY1*, has low expression levels in small-cell lung cancer. It tends to be a putative tumor suppressor. This is because, without the gene deacetylating amino acids, protein synthesis is limited. The suppressor role, however, is not the case for neuroblastoma according to Long et al., [9]. In neuroblastoma, *ACY1* expression can be clearly seen but it is still unclear if the gene has further roles in other cancerous tumors.

mRNA is essential in the formation of these proteins, and whenever problems in mRNA transcription or translation occur, necessary proteins cannot be made. A number of the top 25 genes found by the DM are genes that have been found to inhibit transcription (*HMGB1*, *BTF3*) or translation (*RPLP0*, *EEF1G*) and to govern the metabolism and transport of the mRNA (*HNRPC*). *HMGB1* in particular also hinders the abilities of most treatments for cancer cells as stated by Tang et al., [11]. The gene increases tumor cell autophagy, which allows resistance to a treatment. For tumor cells to induce apoptosis and decrease autophagy, *HMGB1* must be repressed. Further research regarding the relationship of *HMGB1* and autophagy will be of use. These effects can vary greatly based upon the expression of the associated genes and may be key in understanding the significance of the genes found by the DM.

Another functional grouping that contains some of the top 25 genes involve mechanisms that govern the cell cycle. A couple of genes involve the control of microtubules. Microtubules are essential in mitosis of cells during cell division, a key component in the cell cycle and in cell proliferation. *STMN1* and *Homo sapiens* clone 24703 beta-tubulin mRNA, complete cds (unknown gene symbol), have both been linked to microtubule stabilization. Microtubules must be stabilized for a cell to enter into mitosis. *STMN1* regulates this stability through depolymerisation of microtubules during the interphase of a cell. When highly-expressed, it corresponds to microtubule destabilization; which, in the case of cancer, would restrict and retard cell proliferation. Interestingly enough, however, *STMN1* over-expression supports tumor-cell invasion for hepatocellular carcinoma according to Hsieh et al., [10]. Only when *STMN1* was silenced did the tumor-cell invasion become suppressed. The high expression destabilizes the cell microtubules, but there is polyploidy formation as well [10]. This explains the apparent contradiction. Future research will be able to determine the effects and provide more information about *STMN1* expression on other types of cancer.

An additional gene affecting the cell cycle is the telomerase *TERT*, number eight on the list. *TERT* maintains telomere ends and is repressed in postnatal cells. This repression is applicable

to the juvenile cancers being analyzed. Its expression is most important in the developmental stage of humans; thus it has an increased likelihood to be linked to the juvenile diseases being studied. In addition, a prior study has linked the over-expression of *TERT* to samples of BL [7]. Both of these genes provide strong support for the DM, as it ranked *STMN1* and *TERT* very highly in importance.

The fourth and final category for the top 25 genes includes genes of neurological importance. Genes that affect neurological pathways and functions are especially important due to the nature of the four cancers being studied. NB and EWS have important neurological effects, whereas BL does not [8]. *TUBA1A*, *GNAL*, and *GABRB1* are all important to neurological cells and functions. While they have not been directly linked to these cancers at this time, their level of expression could play a significant role in the proliferation of NB and EWS tumors due to these cancers' neurological ties.

Additional differential analysis

Due to knowledge of the DNA microarray data set contents, we conducted additional differential analysis for BL against samples for each type of cancer and the five non-related tissues. Although the order of the genes changed for each additional analysis, the top genes remained relatively constant. This provides support for the effectiveness of the PCA analysis, the first part of the DM process. The PC found to have a signature of interest separated the BL samples from all of the rest. Differential analysis of these two groups yields the same basic results as if the BL group was analyzed with any one particular group of cancer. This allows for confidence in the DM when applying it to data sets of unknown nature, which would be the case in any diagnostic application.

Conclusions

This article enhances the application for the powerful PCA-based differential method developed by Rollins and Teh [4]. It establishes a foundation for the possible use of the DM in analyzing DNA microarray data with the goal of disease diagnosis or discovery. The DM is able to handle all of the DNA microarray data without the need of any subjective filtering methods for dimensional reduction. This makes it a very powerful method, with a great advantage over other methods such as neural networks or clustering. The DM selects assay-specific signatures based upon loading and score plots, giving the user the capability to selectively determine the signature of most interest. It is unique in its use of using rank-ordered differential gene contribution to establish a list of genes that contribute most to the chosen assay-specific signature.

This article presented the application of the DM to a DNA microarray including 83 samples from four types of juvenile cancers, NB, BL, RMS, and EWS, all forms of small round blue cell tumors. Five non-related tissues were also included in the microarray. PCA on the complete data set of 6567 genes

established a signature that completely separated BL samples from the rest. Application of the DM yielded a rank-ordered list of all 6567 genes, listed in order of their importance to the signature. The top 25 genes from this list were closely analyzed to find as many links as possible to the cancers within the data. Top genes were found to be included in one of four categories: directly cancer-related, protein synthesis, cell cycle control, and neurological effects. Gene functions as well as findings from various studies of the genes yielded information that promotes a general relation of the top genes to the juvenile cancers for which the DM found them to be important. These findings, coupled with further research regarding the nature of the top genes found by the DM, supports the strength DM in differentiating between two different groups of assays. It lays the foundation for future applications in disease diagnosis and provides exciting possibilities for discovery of unknown genetic causes for several diseases.

List of abbreviations

PCA: principal component analysis
DM: differential method
NB: neuroblastoma
BL: non-Hodgkin lymphoma
RMS: rhabdomyosarcoma
EWS: Ewing sarcoma
EG: eigengene
EA: eigenassay
PC: principal component

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Authors' contributions	ESH	RJS	DKR
Research concept and design	✓	--	✓
Collection and/or assembly of data	✓	--	--
Data analysis and interpretation	✓	✓	✓
Writing the article	✓	✓	✓
Critical revision of the article	--	--	✓
Final approval of article	--	--	✓
Statistical analysis	✓	--	✓

Acknowledgement

We would like to thank Dr. Javed Khan for making available his data set. We also thank Ailing Teh for her aid in using the PCA methods and Amy Roggendorf for assisting with the final draft. This material is based upon work supported by the National Science Foundation under Grant No. EEC 0552584.

Publication history

Editor: Jimmy Efir, East Carolina University, USA.
Received: 05-Feb-2014 Final Revised: 06-Mar-2014
Accepted: 13-Mar-2014 Published: 02-Apr-2014

References

1. Rollins DK, Zhai D, Joe AL, Guidarelli JW, Murarka A and Gonzalez R. A novel data mining method to identify assay-specific signatures in functional genomic studies. *BMC Bioinformatics*. 2006; 7:377. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)

2. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C and Meltzer PS. **Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks.** *Nat Med.* 2001; **7**:673-9. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
3. Pal NR, Aguan K, Sharma A and Amari S. **Discovering biomarkers from gene expression data for predicting cancer subgroups using neural networks and relational fuzzy clustering.** *BMC Bioinformatics.* 2007; **8**:5. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
4. Rollins DK and Teh A. **An extended data mining method for identifying differentially expressed assay-specific signatures in functional genomic studies.** *BioData Min.* 2010; **3**:11. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
5. Kim JH, You KR, Kim IH, Cho BH, Kim CY and Kim DG. **Over-expression of the ribosomal protein L36a gene is associated with cellular proliferation in hepatocellular carcinoma.** *Hepatology.* 2004; **39**:129-38. | [Article](#) | [PubMed](#)
6. Wu H, Yang JM, Jin S, Zhang H and Hait WN. **Elongation factor-2 kinase regulates autophagy in human glioblastoma cells.** *Cancer Res.* 2006; **66**:3015-23. | [Article](#) | [PubMed](#)
7. Ladetto M, Ricca I, Compagno M, Ferrero D, Rocci A, Dell'Aquila M, Omede P, Francese R, De Marco F, Vallet S, Santo L, Drandi D, Astolfi M, Boccadoro M and Tarella C. **Telomerase inhibitor BIBR1532 induces telomere shortening and growth inhibition in extra-germinal center (GC)-derived lymphoid cell lines.** *Blood.* 2004; **104**:11, 925A. | [Article](#)
8. Kleinschmidt-DeMasters BK, Lovell MA, Donson AM, Wilkinson CC, Madden JR, Addo-Yobo SO, Lillehei KO and Foreman NK. **Molecular array analyses of 51 pediatric tumors shows overlap between malignant intracranial ectomesenchymoma and MPNST but not medulloblastoma or atypical teratoid rhabdoid tumor.** *Acta Neuropathol.* 2007; **113**:695-703. | [Article](#) | [PubMed](#)
9. Long PM, Stradecki HM, Minturn JE, Wesley UV and Jaworski DM. **Differential aminoacylase expression in neuroblastoma.** *Int J Cancer.* 2011; **129**:1322-30. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
10. Hsieh SY, Huang SF, Yu MC, Yeh TS, Chen TC, Lin YJ, Chang CJ, Sung CM, Lee YL and Hsu CY. **Stathmin1 overexpression associated with polyploidy, tumor-cell invasion, early recurrence, and poor prognosis in human hepatoma.** *Mol Carcinog.* 2010; **49**:476-87. | [Article](#) | [PubMed](#)
11. Tang D, Kang R, Cheh CW, Livesey KM, Liang X, Schapiro NE, Benschop R, Sparvero LJ, Amoscato AA, Tracey KJ, Zeh HJ and Lotze MT. **HMGB1 release and redox regulates autophagy and apoptosis in cancer cells.** *Oncogene.* 2010; **29**:5299-310. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
12. Wu Q, Gou Y, Wang Q, Jin H, Cui L, Zhang Y, He L, Wang J, Nie Y, Shi Y and Fan D. **Downregulation of RPL6 by siRNA inhibits proliferation and cell cycle progression of human gastric cancer cell lines.** *PLoS One.* 2011; **6**:e26401. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
13. **Atlas of Genetics and Cytogenetics in Oncology and Haematology.** | [Website](#)
14. **Gene.** | [Website](#)

Citation:

Hald ES, Stoner RJ and Rollins DK. **Determining juvenile cancer types from gene expression using gene contribution and differential analysis.**

J Med Stat Informs. 2014; **2**:2.

<http://dx.doi.org/10.7243/2053-7662-2-2>