Computational Biology and Chemistry 33 (2009) 421–428

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research article

On selecting mRNA isoform features for profiling prostate cancer

T. Murlidharan Nair[a,b,*]

[a] Department of Biology, Indiana University South Bend, 1700 Mishawaka Ave, South Bend, IN 46634, USA
[b] Department of Computer Science/Informatics, Indiana University South Bend, 1700 Mishawaka Ave, South Bend, IN 46634, USA

ARTICLE INFO

Article history:
Received 10 May 2009
Received in revised form
26 September 2009
Accepted 26 September 2009

Keywords:
mRNA isoforms
Splicing arrays
Multiple comparison
Prostate cancer

ABSTRACT

Alternative splicing of human pre-mRNA is a very common phenomenon and is a major contributor to proteome diversity. mRNA isoforms that arise as a result of alternative splicing also provide a more complete picture of the transcriptome as they reflect the additional processing a pre-mRNA undergoes before being translated into a functional product. It has been reported that molecular alterations of cells can occur as a result of the differential expression of mRNA isoforms, resulting in cancerous or normal tissue. Quantification of mRNA isoforms can thus be used as a better indicator in distinguishing a normal tissue from a cancerous tissue. In our earlier study we had used mRNA isoforms expression to identify biomarkers for prostate cancer (Li et. al, 2006. Cancer Res. 66 (8) 4079–4088). Here we have used statistical methods of multiple comparison and have developed a simple scoring scheme to extract isoform features. Further, we have rigorously analyzed the isoform expression data to understand the variability and heterogeneity associated with the expression levels between (i) prostate cancer cell lines and non-prostate cancer cell lines and (ii) normal prostate tissue and prostate cancer tissue. We found that there were several isoforms that showed significant difference in expression within the same class. We were also able to successfully identify isoforms with similar changes in expression levels, that when used as features for classification was able to provide robust class separation. The features selected using the multiple comparison methods had subsets that were common and disparate with those that were selected using statistical *t*-tests. This reveals the importance of selecting features using a combination of complementary methods.

1. Introduction

Cancer is a highly heterogeneous disease and one of the leading causes of death in both men and women alike. Over the years significant progress has been made in battling other diseases like heart disease and cerebrovascular disease, which has resulted in declined mortality rates. Death rate due to cancer has not changed significantly (http://www.cancer.org). Mortality rate due to prostate cancer, which has been the second leading cause of mortality in men have decreased, as a result of early detection based on prostate-specific antigen (PSA) screening (Crawford et al., 2001) and monitoring by digital rectal exam (Ilic et al., 2006). Elevated PSA levels are seen in another condition called benign prostate hyperplasia (BPH) (Schroder, 2005), thus making needle biopsy the only definitive diagnosis for prostate cancer. The heterogeneous nature of the disease with disparate molecular alterations makes early diagnosis a challenge.

Molecular alterations associated with prostate carcinomas range from epigenetic changes and genetic polymorphisms to alternative splicing and post-translational processes (Reynolds, 2008). Distinguishing carcinomas based on molecular alterations would involve the ability to accurately determine the molecular picture of the cell type. One approach to determining this is to use high-throughput technology using microarrays (Ramaswamy et al., 2001; Fournier et al., 2006; Ginestier et al., 2006) for profiling the levels of thousands of mRNA expressed in the cell in one single experiment. Most expression studies are based on measuring the levels of the unprocessed mRNA transcript. However, it is now a known fact that more than 60% of the human genes are alternatively spliced (Croft et al., 2000; Lander et al., 2001; Modrek et al., 2001). Alternatively spliced forms are capable of giving rise to unique proteins which contributes to the proteome diversity (Maniatis and Tasic, 2002). Conservation of sequence in alternative spliced forms is indicative of their functional importance (Chen and Zheng, 2008). Measuring the unprocessed mRNA transcript only provides an incomplete picture of the transcriptome. A probe detecting the level of the unprocessed transcript does not provide any information about the exons that was not queried specifically by it, even though it may be part of the same transcript. When non-specific probes are used, one often assumes that all exons that are part

* Correspondence address: Department of Biology, Indiana University South Bend, 1700 Mishawaka Ave, South Bend, IN 46634, USA. Tel.: +1 574 520 5068; fax: +1 574 520 4554.
E-mail address: mnair@iusb.edu.

1476-9271/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2009.09.006

of that pre-mRNA are equally expressed. This however is not the case as most pre-mRNAs undergo post-transcriptional processing. In an earlier study we profiled the expression of ~1500 mRNA isoforms from a panel of genes that have been previously associated with prostate and other cancers (Li et al., 2006). An mRNA isoform detection system known as the DASL assay (cDNA-mediated annealing, selection, and ligation) coupled with a universal array on fiber optic bundles was used to profile mRNA isoforms in a high-throughput fashion. The method has the advantage of handling partially degraded biological samples especially those derived from tissue blocks that have been formalin-fixed and paraffin embedded (Fan et al., 2004). One major challenge in analyzing microarray data especially from sources that are potentially heterogeneous is to identify with a reasonable certainty universal signatures of differential expression between two classes of samples being analyzed. The analysis of isoform expression data did reveal subset of isoforms that were uniquely expressed in prostate cancer cell lines and tissue but not in others. In this paper we rigorously analyzed the expression data to understand in detail the changes in isoform expression between prostate and non-prostate cancer cell lines, and between normal and prostate tumor tissue. We have used a simple scoring scheme using multiple comparisons to rank isoforms as most distinguishing between the two classes being compared. We also analyzed the changes in isoform expression levels within the same class with a view to illustrate the heterogeneity associated with the prostate carcinoma samples used in the study. Further, we also compared the ranking of all the features using the current method to those obtained using simple *t*-tests. The results revealed common and disparate features, underlining the importance of using complementary methods when selecting features.

2. Materials and methods

2.1. Data

The isoform expression data used in this study were obtained from earlier work by Li et al. (2006). The data consisted of isoform expression from cell lines and prostate tumor tissue. The cell lines for which expression data were collected included five prostate cancer cell lines viz. LNCap, LAPC4, RWPE2, PC3 and DU145, and twelve non-prostate cancer cell lines viz. colon cancer line (HT29, SW480, HCT116, LS174, Fet), breast cancer line (MCF7, MDA-MB-468), kidney cancer line (Caki-2), lung epidumoid carcinoma line (CALU1) and esophageal cancer lines (HCE-7, EC17 and TE3). The prostate tumor tissue from which expression data were collected consisted of formalin-fixed and paraffin-embedded tissue. The set contained 10 normal and 12 tumor samples.

2.2. Principal-component analysis and biplot-PCA

The structure of correlations within the mRNA isoform expression data was characterized using principal-component analysis and biplot-PCA. Biplot was used to represent the results of PCA (Gabriel, 1971). Biplot graphically displays a matrix $M=(m_{ij})$ of n rows and m columns, using row and column markers. The inner product of the markers represents the i,jth element of M. The rows

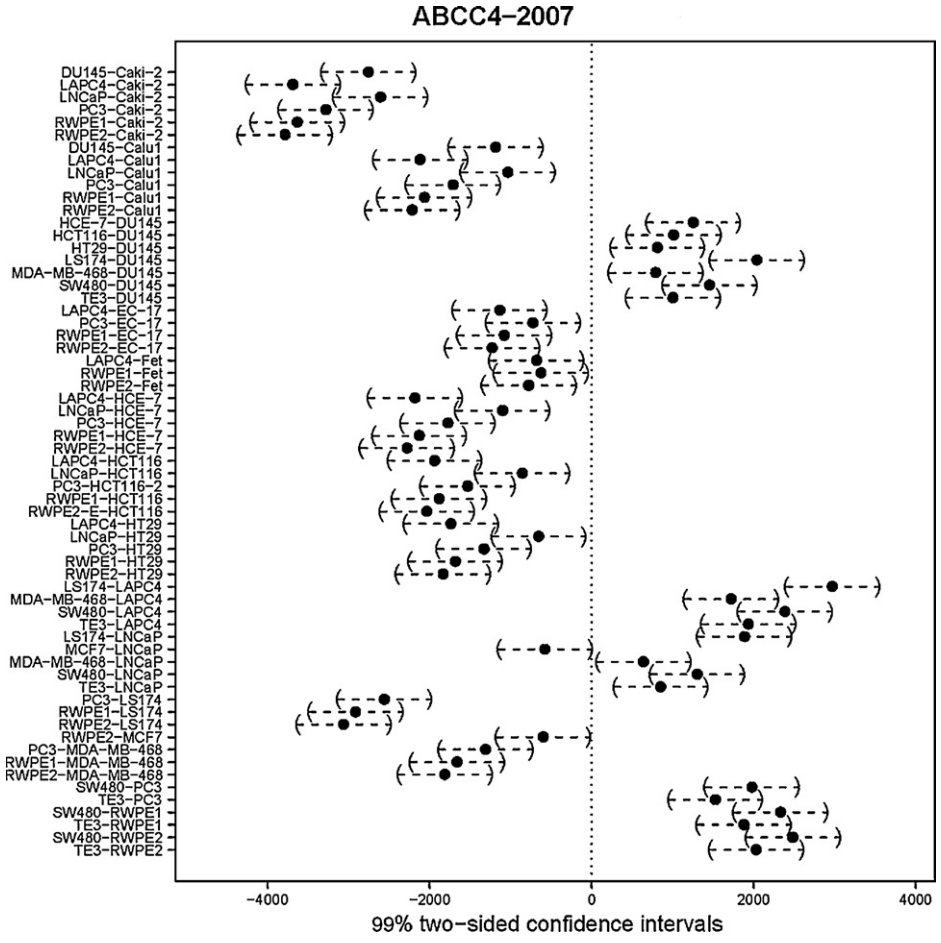

Fig. 1. Multiple comparison of the expression level of one of the mRNA isoforms of ATP-binding cassette, sub-family C, member 4 (ABCC4-2007) gene between the cell lines studied. Only comparisons that showed a significant difference in expression level are plotted. Similar exhaustive comparisons were done for all 1532 isoforms studied.

Table 1
Non-prostate cancer cell line vs. non-prostate cancer cell line. Percentage of isoforms that were found to be differentially expressed when different non-prostate cancer cell lines were compared with each other (within-class comparison).

	Caki-2	Calu1	EC-17	Fet	HCE-7	HCT116-2	HT29	LS174	MCF7	MDA-MB-468	SW480
Calu1	37.08										
EC-17	37.14	30.29									
Fet	39.10	37.21	35.44								
HCE-7	37.01	28.59	34.66	32.90							
HCT116-2	36.81	32.38	32.64	30.55	29.44						
HT29	43.86	40.08	39.62	36.10	37.27	38.64					
LS174	33.81	34.20	35.44	35.25	33.09	22.98	37.47				
MCF7	40.60	38.97	40.93	35.77	36.68	33.36	40.14	29.77			
MDA-MB-468	36.42	37.60	36.75	38.32	35.25	33.22	40.21	34.60	36.16		
SW480	43.15	39.75	38.45	40.34	37.99	36.42	38.12	30.29	35.64	38.71	
TE3	37.99	37.27	34.14	37.40	34.20	35.05	32.64	33.55	35.18	33.62	32.77

Table 2a
Prostate cancer cell line vs. non-prostate cancer cell line. Percentage of isoforms that were found to be differentially expressed when prostate cancer cell lines were compared with non-prostate cancer cell lines (between-class comparisons).

	DU145-E	LAPC4-E	LNCaP-E	PC3-E	RWPE1-E	RWPE2-E
Caki-2	56.20	59.79	61.10	56.72	56.79	59.07
Calu1	54.11	58.09	57.05	54.18	54.31	54.83
EC-17	53.46	56.98	55.87	51.83	52.74	53.26
Fet	54.24	52.94	54.37	54.96	54.63	53.66
HCE-7	50.46	55.16	54.44	50.00	51.57	51.96
HCT116-2	53.00	54.18	53.85	53.26	51.89	52.35
HT29	50.39	50.33	53.00	51.04	49.35	50.72
LS174	51.50	51.63	54.31	52.15	51.17	51.76
MCF7	54.44	53.52	54.05	55.29	53.98	54.83
MDA-MB-468	55.16	55.68	57.51	51.96	53.98	55.22
SW480	51.44	53.66	55.03	53.46	52.74	51.70
TE3	52.09	52.28	55.03	50.59	52.15	50.91

for the splicing array matrix correspond to the different isoforms of the genes being analyzed and the columns are the different cell lines or tissue samples. The row markers corresponding to the isoforms (not given in the figure) and the arrow or column markers represent the cell lines or tissue samples. The lengths of the arrows correspond to the variances among the different cell lines or tissue samples and the angle represents their correlation.

2.3. Multiple comparison scoring scheme

We used the multicomp package from the R statistical environment for our multiple comparisons (http://www.r-project.org). Tukey's method was used in our analysis. All possible comparisons within and between the classes were carried out. Making inferences about the differences between two population means by either using interval estimation or hypothesis testing approach is rather straightforward. However, when the comparison involves several means, approaches involving ANOVA *F*-test, only provides information whether the two means, being compared were significantly different. They do not give information on which means differ from which other means (Hochberg and Tamhane, 1987; Westfall, 1997). Multiple comparison procedure employed here circumvents this problem and provides detailed information about the differences in means. Exhaustive comparison of isoform expression levels was carried out within and between each class of samples. Each isoform was given a unit score for every significant difference it showed in a between-class or within class comparison. Sum of these scores was then used to rank the isoforms. A subset of top ranking isoforms was used as features for class separation using a clustering algorithm (Eisen et al., 1998).

Table 2b
Prostate cancer cell line vs. prostate cancer cell line. Percentage of isoforms that were found to be differentially expressed when different prostate cancer cell lines were compared with each other (within-class comparison).

	DU145-E	LAPC4-E	LNCaP-E	PC3-E	RWPE1-E
LAPC4-E	43.60				
LNCaP-E	44.26	39.88			
PC3-E	38.25	44.19	45.37		
RWPE1-E	43.02	46.93	47.45	41.97	
RWPE2-E	39.88	47.00	44.71	41.45	22.45

3. Results and discussion

3.1. Multiple comparison of prostate and non-prostate cancer cell lines

We carried out two sets of exhaustive comparisons. One involved the comparison between prostate and non-prostate cancer cell lines which we term as the between class comparison. The other comparison that we call within class comparison was done by comparing prostate cancer cell lines with prostate cancer cell lines and by comparing non-prostate cancer cell lines with non-prostate cell lines. Fig. 1 shows the comparison between different cell lines for one of the isoforms. Significant differences in expression levels are those that did not intersect the zero line. Only the cell lines that showed a significant difference in isoform expression levels have been plotted in the figure. In the interest of brevity, comparisons for only one isoform are shown. Table 1 shows the percentage of isoforms that showed differential expression between prostate cancer and non-prostate cancer cell lines. On average, about 53% of the isoforms showed significant differential expression between the two classes. With a view to understand heterogeneity and sample to sample variation, we analyzed the isoforms for differential expression within the same cell line class. Tables 2a and 2b give the percentage of isoforms that were differentially expressed within the same class. On average, about 35% of the isoforms were differentially expressed within non-prostate cancer cell lines and about 39% of the isoforms were differentially expressed within prostate cancer cell lines. Using the scoring scheme discussed earlier we selected a subset of one hundred top ranking isoforms as features and clustered the expression data for all the cell lines by hierarchical clustering. Fig. 2 gives the heat map and the results of clustering using the selected features. We were able to successfully separate the two classes and the clear difference in expression profiles is evident in the heat map diagram. The distinct differences in the expression levels of the isoforms associated with the two different classes is further emphasized when the features selected are subjected to biplot-principal-component-analysis (Gabriel, 1971). The analysis takes into account both negative and positive correlations in expression levels and the results are presented as a biplot graph in Fig. 3. The length of the arrow reflects the variance in the data for each sample. The variance in the expression seems to be comparable

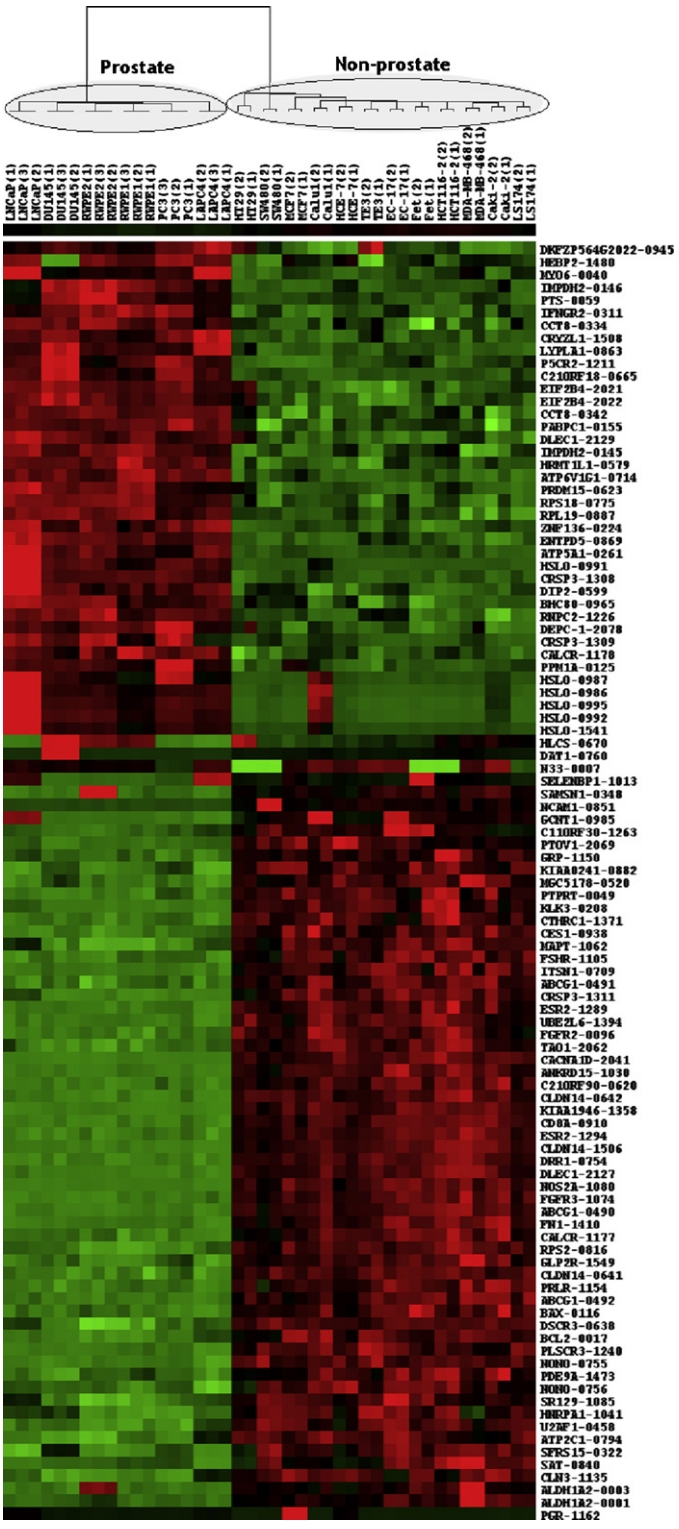

Fig. 2. Hierarchical clustering of the prostate cancer and non-prostate cancer cell lines using the top 100 features that were found to be most discriminatory based on the multiple comparison analysis ranking method.

in most cases. The biplot representation also displays the relationship among the samples. The angles between positively correlated vectors approach 0 degrees while those with negative correlations approach 180 degrees. Samples with no correlation are orthogonal. The samples that belong to the same class have distinctly similar projections of their vectors and the two classes show clear class separation.

3.2. Multiple comparison of isoform expression in prostate cancer tissue and normal tissue

We also carried out a similar exhaustive comparison for isoform expression data in normal and prostate cancer tissues. As we had done in the case of the cell line data, we carried out two sets of comparisons; in one we compared the expression of isoforms in samples belonging to two different classes which we term as between-class comparison, in the other we compared expression of isoforms in samples from the same class with each other in a within-class comparison. Fig. 4 shows the results of the multiple comparisons for one of the isoforms that showed significant difference in expression in the samples compared. Table 3 shows the percentage of isoforms that showed differential expression between tumor tissue and the normal tissue. The data in the table clearly shows the high variability in the number of isoforms that showed differential expression. The data should also be viewed keeping in mind the variability in the tumor content of the different samples that were analyzed. The tumor content varied from 40% to 95%. In order to understand the heterogeneity associated with the molecular alteration associated with prostate cancer, we analyzed the differential expression of isoforms within the different normal prostate tissue samples and within the prostate tumor tissue samples. Tables 4a and 4b show the results for the comparison within tumor and normal samples. The results clearly show that even in a within-class comparison there are molecular alterations that are distinct. The variations in the normal samples are much less as compared to the variations in the tumor samples. This clearly points to the heterogeneity associated with prostate tumors and the challenge in defining a single marker in identifying them. Even in the light of this variation, we were able to extract a subset of top ranking isoforms that showed significant differential expression in a majority of cases, and successfully use them as features for hierarchical clustering. The result of clustering the samples using the features extracted is shown in Fig. 5. It is noteworthy to point out that while we were able to successfully classify most of the samples, the tumor sample with the least tumor content (40%) was clustered among the normal sam-

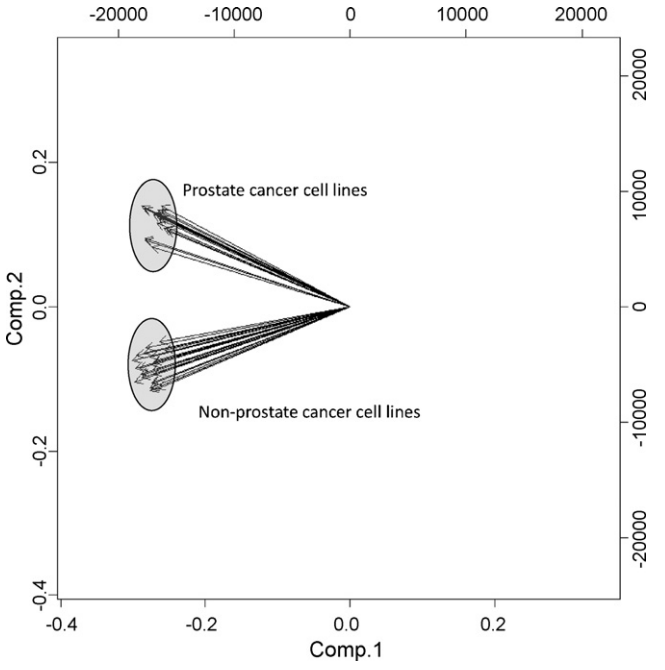

Fig. 3. Biplot showing the overall structure of the difference in isoform expression between prostate and non-prostate cancer lines. The length of each eigenvector is proportional to the variance in the expression for that cell line. The angle between eigenvectors represents correlations among different cell lines.

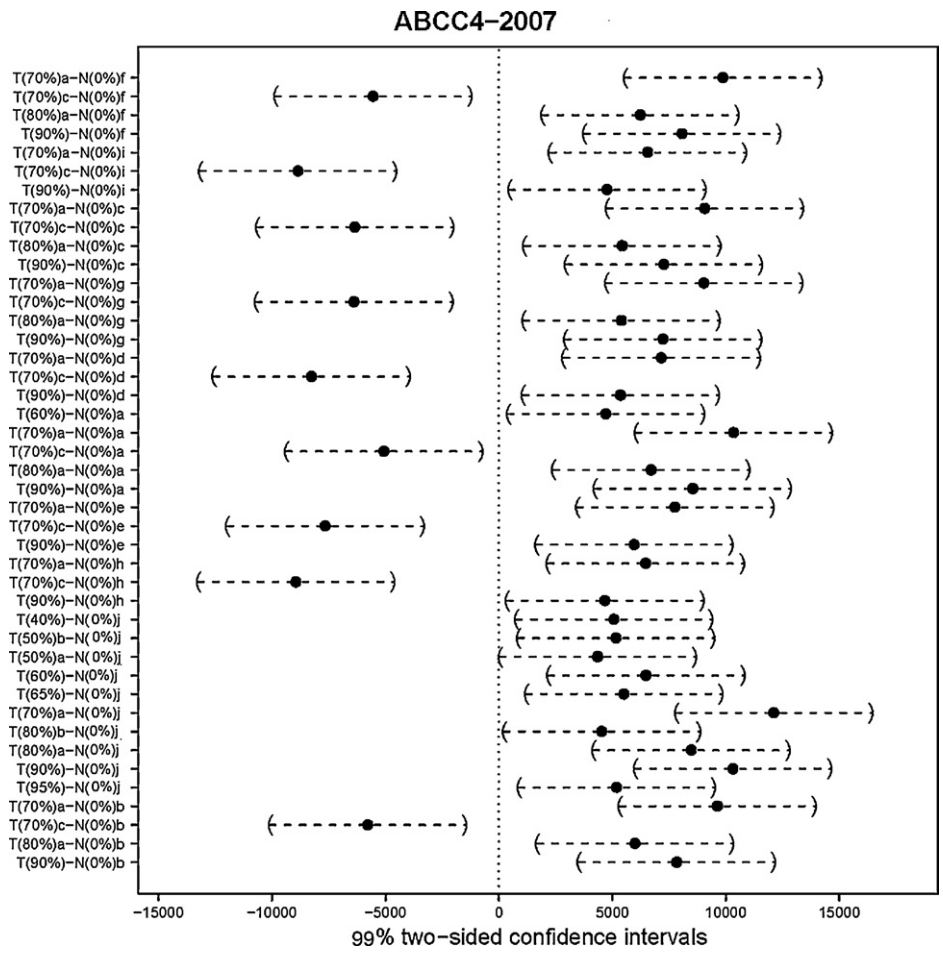

Fig. 4. Multiple comparison of the expression level of one of the mRNA isoforms of ATP-binding cassette, sub-family C, member 4 (ABCC4-2007) gene between prostate cancer tissue and normal prostate tissue. Only comparisons that showed a significant difference in expression levels between cancer tissue and normal prostate tissue are plotted. Similar exhaustive comparisons were done for all 1532 isoforms studied.

Table 3
Tumor tissue vs. normal tissue. Percentage of isoforms that were found to be differentially expressed when prostate tumor tissues were compared with normal prostate tissues (between-class comparison). The percentage of tumor content is indicated in the bracket.

	T(40%)	T(50%)a	T(50%)b	T(60%)	T(65%)	T(70%)a	T(70%)b	T(70%)c	T(80%)a	T(80%)b	T(90%)	T(95%)
N(0%)a	10.18	16.97	10.05	13.51	19.52	28.13	17.30	24.02	8.49	12.79	7.90	12.27
N(0%)b	4.50	10.51	9.14	8.16	19.65	26.57	15.60	25.39	10.90	9.46	11.16	15.60
N(0%)c	14.36	19.78	10.12	15.73	22.91	32.38	20.76	29.18	11.42	15.08	13.84	16.64
N(0%)d	6.27	15.01	5.94	8.75	19.84	28.52	14.69	24.15	10.31	8.68	9.99	14.43
N(0%)e	6.66	11.81	5.74	4.90	19.71	27.55	15.99	24.80	10.25	7.44	10.97	13.32
N(0%)f	6.01	12.47	9.33	10.51	22.72	27.68	17.49	27.35	13.45	9.66	12.01	16.25
N(0%)g	4.96	12.21	7.44	7.11	19.91	24.80	14.30	24.09	10.18	6.98	10.84	14.69
N(0%)h	6.72	9.66	8.94	6.92	19.97	25.59	16.64	24.48	12.99	9.14	12.99	15.93
N(0%)i	14.03	22.91	17.43	19.39	27.74	35.31	23.63	30.81	22.00	17.95	21.28	27.22
N(0%)j	9.40	13.58	12.60	13.38	22.32	29.11	18.02	27.81	13.84	12.73	14.49	18.80

Table 4a
Tumor tissue vs. tumor tissue. Percentage of isoforms that were found to be differentially expressed when different prostate tumor tissues were compared with each other (within-class comparison). The percentage of tumor content is indicated in the bracket.

	T(40%)	T(50%)b	T(50%)a	T(60%)	T(65%)	T(70%)a	T(70%)b	T(70%)c	T(80%)b	T(80%)a	T(90%)
T(50%)b	7.64										
T(50%)a	13.05	13.12									
T(60%)	9.92	6.59	9.66								
T(65%)	20.43	18.73	17.62	17.17							
T(70%)a	28.59	27.02	24.41	22.26	21.48						
T(70%)b	16.12	15.27	15.86	13.51	13.77	20.37					
T(70%)c	24.87	23.43	24.09	25.00	19.91	25.59	18.99				
T(80%)b	9.79	5.61	14.36	5.87	18.54	24.48	14.69	26.17			
T(80%)a	13.19	9.14	16.64	11.29	16.19	23.50	13.97	25.33	9.92		
T(90%)	12.27	10.44	16.97	11.62	15.54	22.85	15.21	22.32	9.20	7.77	
T(95%)	15.34	12.86	18.47	13.71	15.73	20.63	14.75	23.04	12.34	10.12	7.44

Table 4b
Normal tissue vs. normal tissue. Percentage of isoforms that were found to be differentially expressed when different normal prostate tissues were compared with each other (within-class comparison). Percentage of tumor content is indicated in the bracket.

	N(0%)a	N(0%)b	N(0%)c	N(0%)d	N(0%)e	N(0%)f	N(0%)g	N(0%)h	N(0%)i
N(0%)b	8.42								
N(0%)c	10.25	10.97							
N(0%)d	8.29	5.81	7.57						
N(0%)e	8.16	4.57	9.92	4.63					
N(0%)f	10.18	3.46	11.81	5.29	6.14				
N(0%)g	10.57	2.35	14.75	5.81	5.48	4.83			
N(0%)h	13.12	6.33	16.19	9.53	6.72	6.66	5.29		
N(0%)i	19.45	15.93	20.50	13.51	17.75	15.80	14.03	18.28	
N(0%)j	11.62	3.92	14.03	9.33	9.33	5.81	7.57	10.64	20.37

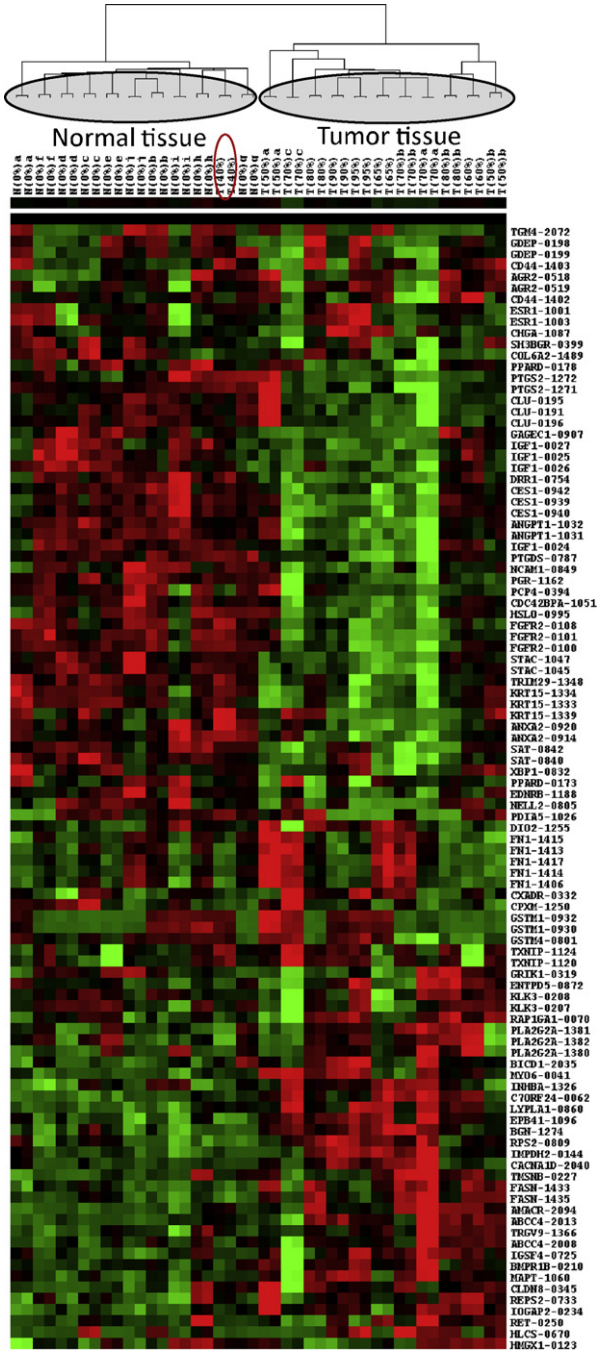

Fig. 5. Hierarchical clustering of prostate cancer tissue and normal tissue using the top 100 isoform features that were found to be most discriminatory based on the multiple comparison analysis ranking method.

ples. To further investigate the misclassification of this sample, all the samples were subjected to biplot-principal-component analysis (biplot-PCA). The results of the biplot-PCA are shown in Fig. 6. The results clearly show that the samples with lower tumor content are more similar to the normal samples. The tumor sample that was misclassified in the hierarchical clustering, appear in the biplot to be similar to the normal samples. It is also important to point out that the tumor sample with 50% tumor content is at the border of the two classes, but more close to the tumor class in the biplot.

3.3. Comparison of ranking of the features in prostate cancer cell line vs. prostate cancer tissue

One of the obvious questions that comes to mind is regarding the nature of concordance between the top ranking features. Are the features that distinguished prostate cancer cell lines from non-prostate cancer cell lines in agreement with those that distinguished the prostate cancer tissue from normal prostate tissue? Before delving into the details of the results, it is important to point out that the comparison of the cell lines involved comparing prostate cancer cell lines with non-prostate cancer cell lines and not non-cancer cell lines. The prostate cancer tissue on the other hand was compared with normal prostate tissue and not with other

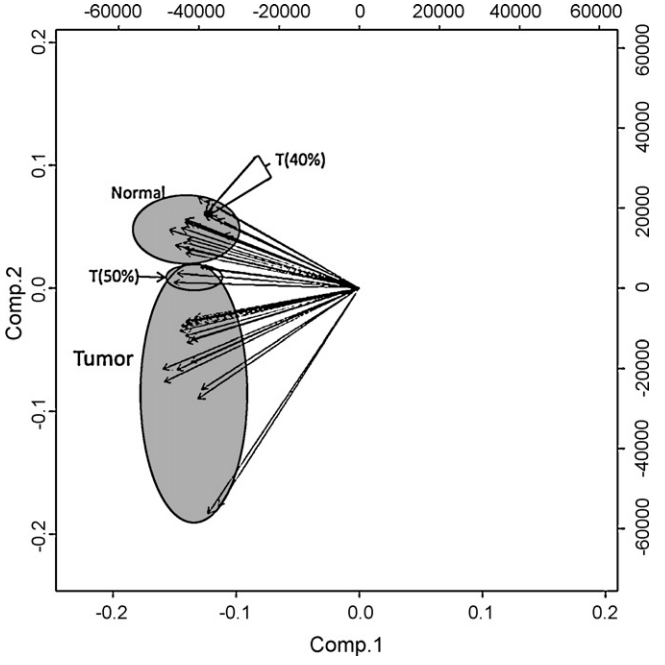

Fig. 6. Biplot showing the overall structure of the difference in isoform expression in tumor and normal prostate tissue. The length of each eigenvector is proportional to the variance in the expression of isoforms for that tissue sample. The angle between eigenvectors represents correlations among different tissue samples.

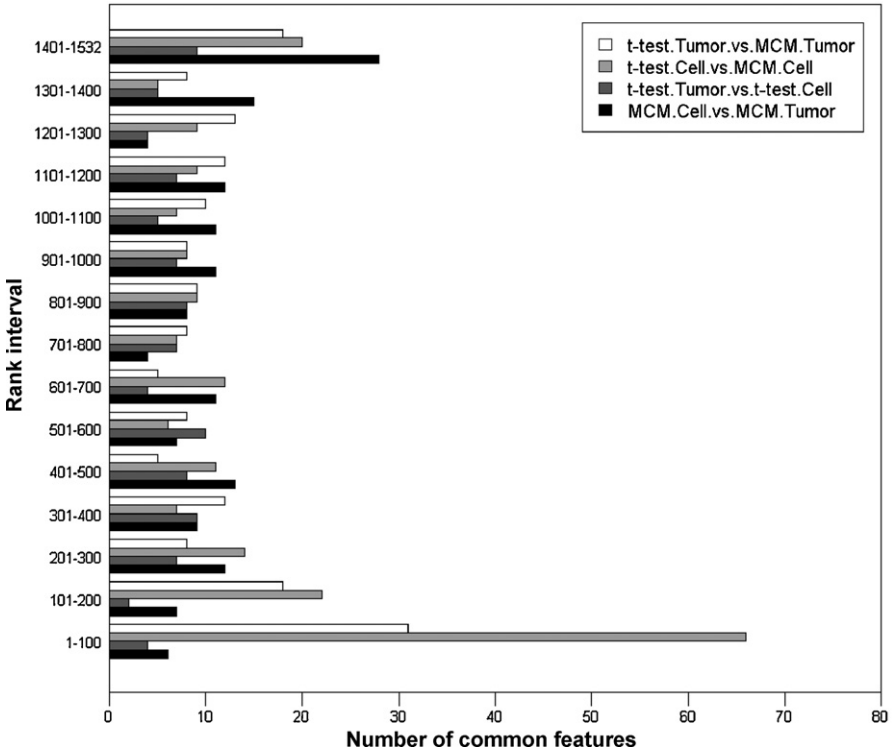

Fig. 7. Number of isoform features common at different rank intervals. Ranks were assigned to isoforms by *t*-tests or multiple comparison method of ranking (MCM).

non-prostate cancer tissues. This is an important fact to be kept in mind as the profile of the normal prostate tissue cannot be equated with that of non-prostate cancer cell lines. Further, there may be differences with cell lines itself when they are compared with tissue samples (Sandberg and Ernberg, 2005). The results of these comparisons should be viewed in the light of these arguments, but nevertheless it is important to understand the differences and similarities. The ranked features were divided at intervals of ranks of 100 and the intersection of features contained in those intervals was enumerated. Fig. 7 shows the number of features that were common at different rank intervals. The maximum common features were found at the intersection of first one hundred ranks of features for cell lines obtained using multiple comparison methods and the first one hundred ranks of cell line features obtained using *t*-tests. Within this interval there were 66 features that were common. In comparison only 31 features were common between the first one hundred ranks of the tumor tissue features obtained using multiple comparison methods and *t*-tests. There were less than ten common features when tumor tissue features were compared with cell line features within the same rank (1–100) using either method of feature extraction. The results point to some of the challenges in variable and feature selection that are encountered when profiling large data sets. Obtaining variable subsets of features that are capable of class separation presents a dilemma, as to which features are more discriminatory and biologically relevant. One approach to circumvent this is to use wrapper methods to assess subsets of variables according to their usefulness in class separation. Since, the biological problem we are dealing with is inherently complex, a computational wrapper approach might not capture the interdependence of features for their biological functionality. These need to be studied by experimentation.

4. Conclusion

We have used multiple comparison based approach to develop a simple scoring scheme to rigorously compare the expression levels of mRNA isoforms from prostate cancer cell lines and tissue samples. The method was able to capture features that were able to separate the classes using clustering algorithms. Exhaustive comparison of features within the same class in both cases (cell line and tissue) revealed that there were several isoforms that were differentially expressed. These comparisons revealed heterogeneity associated with cell lines and tissue samples. When the features obtained using the multiple comparison method was compared with those obtained using a *t*-tests, not many features were found at the intersection of similar ranks. The two seemingly disparate subsets however did provide robust class separation. The results point to the fact that when analyzing large data sets for features it is important to use complementary methods for feature extraction. Merely achieving class separation should not be used as a sole criterion for assigning biological functionality but only as an indicator of potential functional importance of the extracted features.

Acknowledgements

TMN would like to thank NSF DBI 0515986 and IUSB summer research for funding. TMN also thanks Prof Xiang-Dong Fu for his comments.

References

Chen, L., Zheng, S., 2008. Identify alternative splicing events based on position-specific evolutionary conservation. PLoS ONE 3, e2806.
Crawford, E.D., Miller, G.J., Labrie, F., Hirano, D., Batuello, J., Glode, L.M., 2001. Prostate cancer pathology, screening, and epidemiology. Rev. Urol. 3 (Suppl. 2), S2–S10.
Croft, L., Schandorff, S., Clark, F., Burrage, K., Arctander, P., Mattick, J.S., 2000. ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. Nat. Genet. 24, 340–341.
Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. 95, 14863–14868.
Fan, J.B., Yeakley, J.M., Bibikova, M., Chudin, E., Wickham, E., Chen, J., Doucet, D., Rigault, P., Zhang, B., Shen, R., McBride, C., Li, H.R., Fu, X.D., Oliphant, A., Barker, D.L., Chee, M.S., 2004. A versatile assay for high-throughput gene expression profiling on universal array matrices. Genome Res. 14, 878–885.

Fournier, M.V., Martin, K.J., Kenny, P.A., Xhaja, K., Bosch, I., Yaswen, P., Bissell, M.J., 2006. Gene expression signature in organized and growth-arrested mammary acini predicts good outcome in breast cancer. Cancer Res. 66, 7095–7102.

Gabriel, K.R., 1971. The biplot-graphics display of matrices with application to principal component analysis. Biometrika 58, 453–467.

Ginestier, C., Cervera, N., Finetti, P., Esteyries, S., Esterni, B., Adelaide, J., Xerri, L., Viens, P., Jacquemier, J., Charafe-Jauffret, E., Chaffanet, M., Birnbaum, D., Bertucci, F., 2006. Prognosis and gene expression profiling of 20q13-amplified breast cancers. Clin. Cancer Res. 12, 4533–4544.

Hochberg, Y., Tamhane, A.C., 1987. Multiple Comparison Procedures. Wiley, New York.

Ilic, D., O'Connor, D., Green, S., Wilt, T., 2006. Screening for prostate cancer. Cochrane Database Syst. Rev. 3, CD004720.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Li, H.R., Wang-Rodriguez, J., Nair, T.M., Yeakley, J.M., Kwon, Y.S., Bibikova, M., Zheng, C., Zhou, L., Zhang, K., Downs, T., Fu, X.D., Fan, J.B., 2006. Two-dimensional transcriptome profiling: identification of messenger RNA isoform signatures in prostate cancer from archived paraffin-embedded cancer specimens. Cancer Res. 66, 4079–4088.

Maniatis, T., Tasic, B., 2002. Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418, 236–243.

Modrek, B., Resch, A., Grasso, C., Lee, C., 2001. Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res. 29, 2850–2859.

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P., Poggio, T., Gerald, W., Loda, M., Lander, E.S., Golub, T.R., 2001. Multiclass cancer diagnosis using tumor gene expression signatures. Proc. Natl. Acad. Sci. U.S.A. 98, 15149–15154.

Reynolds, M.A., 2008. Molecular alterations in prostate cancer. Cancer Lett. 271, 13–24.

Sandberg, R., Ernberg, I., 2005. Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI). Proc. Natl. Acad. Sci. U.S.A. 102, 2052–2057.

Schroder, F.H., 2005. Detection of prostate cancer: the impact of the European randomized study of screening for prostate cancer (ERSPC). Can. J. Urol. 12 (Suppl. 1), 2–6; discussion 92–93.

Westfall, P., 1997. Multiple testing of general contrasts using logical constraints and correlations. J. Am. Stat. Assoc. 92, 299–306.