Application of MUTIC to the exploration of gene expression data in prostate cancer

B. Goertzel[1], C. Pennachin[1], L.S. Coelho[1] and M.A. Mudado[1,2]

[1]Biomind LLC, Rockville, MD, USA
[2]Departamento de Bioquímica e Imunologia, UFMG, Belo Horizonte, MG, Brasil
Corresponding author: L.S. Coelho
E-mail: lucio@biomind.com

ABSTRACT. We show here an example of the application of a novel method, MUTIC (model utilization-based clustering), used for identifying complex interactions between genes or gene categories based on gene expression data. The method deals with binary categorical data which consist of a set of gene expression profiles divided into two biologically meaningful categories. It does not require data from multiple time points. Gene expression profiles are represented by feature vectors whose component features are either gene expression values, or averaged expression values corresponding to gene ontology or protein information resource categories. A supervised learning algorithm (genetic programming) is used to learn an ensemble of classification models distinguishing the two categories based on the feature vectors corresponding to their members. Each feature is associated with a "model utilization vector", which has an entry for each high-quality classification model found, indicating whether or not the feature was used in that

model. These utilization vectors are then clustered using a variant of hierarchical clustering called Omniclust. The result is a set of model utilization-based clusters, in which features are gathered together if they are often considered together by classification models - which may be because they are co-expressed, or may be for subtler reasons involving multi-gene interactions. The MUTIC method is illustrated here by applying it to a dataset regarding gene expression in prostate cancer and control samples. Compared to traditional expression-based clustering, MUTIC yields clusters that have higher mathematical quality (in the sense of homogeneity and separation) and that also yield novel insights into the underlying biological processes.

INTRODUCTION

A variety of methods for analyzing gene expression data has arisen in recent years, including but not limited to: identifying which genes are maximally differentiated between two categories; clustering genes based on co-expression across multiple samples or multiple experiments (Eisen et al., 1998; Spellman et al., 1998; Ben-Dor et al., 1999; Tamayo et al., 1999; Sharan and Shamir, 2000; Sharan et al., 2001; Dopazo and Azuaje, 2005); using supervised categorization algorithms to learn rules distinguishing two or more categories of gene expression profiles from each other (Golub et al., 1999; Brown et al., 2000; Dudoit et al., 2002; Guyon et al., 2002; Cho et al., 2004), and inference of genetic interaction networks from gene expression time series data (Markowetz and Spang, 2003; Vert and Kanehisa, 2003; Markowetz, 2004; Nachman et al., 2004; Sohler et al., 2004). These methods serve various purposes, such as induction of diagnostic models, qualitative understanding of the biological phenomena underlying a dataset, and identification of specific actors (e.g., genes, proteins) that may be involved in a certain biological phenomenon. In this paper, we use model utilization-based clustering (MUTIC), a novel method for gene expression data analysis, whose goal is to identify the interactions among genes, proteins and biological processes that are most relevant to the phenotypic distinction underlying a given binary categorization of gene expression profiles.

Clustering is the most common tool for interaction identification. By determining which genes or gene categories have expression-value profiles that cluster together across multiple samples or multiple experiments, one gets a picture of which genes are "associated" with each other. However, these associations do not usually have a clear interpretation, as co-expression can occur for a variety of reasons. Furthermore, many types of interactions are in principle not identifiable by directly clustering gene expression values. For instance, one will not recognize ternary interactions wherein, say, C is only highly expressed when both A and B are highly expressed together.

The technique used here, MUTIC, is oriented toward capturing interactions that ordinary expression-based clustering misses. The end result of MUTIC looks superficially similar to that of traditional gene expression clustering: one obtains a set of clusters (of genes or gene categories), where the elements of a cluster are hypothesized to have a significant interrelationship. What is novel is that these clusters are not determined based on co-expression but via a more involved analysis. The semantics of the clusters is different: MUTIC clusters represent genes or gene categories that are usefully considered in combination when formulating classification rules distinguishing one category of gene expression profiles from another. The elements of such a cluster may or may not be co-expressed across the set of gene expression profiles under analysis.

The MUTIC method is described in detail elsewhere (Goertzel et al., 2006). Here, we discuss its application to a dataset regarding human gene expression, drawn from prostate tumor and control cells. In the context of that dataset, we review a number of potentially interesting biological interactions that the new method finds but traditional expression-based clustering misses.

Experimental setup

In this section we describe the particularities of the application of MUTIC reported here, encompassing the dataset and data preprocessing as well as the parameters used by the several MUTIC phases properly told.

Test dataset

The prostate tumor dataset used for validating the method is composed of a train dataset containing 102 samples (52 prostate tumor tissue samples, or cases, and 50 controls corresponding to normal tissue samples) and a test dataset containing 34 samples (25 cases, 9 controls). The train dataset has been reported by Singh et al. (2002), while the test dataset was reported in another experiment (Welsh et al., 2001). The use of this train-test pair in a classification experiment was in turn reported by Tan and Gilbert (2003). This reference indeed points to the URL where the datasets (as they were used here) are available: http://sdmc.lit.org.sg/GEDatasets/Datasets.html#Prostate. In both train and test datasets, each sample was characterized by the gene expression of 12,600 features. Nevertheless, most of those features showed suspicious near-zero expression values in all samples. Those null-valued features in either train or test sets were removed and as a result only 1704 genes were effectively used throughout the analysis.

This dataset then underwent our process of enhancement, which consists in adding synthetic features corresponding to the GO (gene ontology) and PIR (proteing information resource) categories. Each synthetic feature corresponding to a given GO or PIR category contains for each sample in the dataset the average expression value of all genes present in the dataset and under that category. The enhancement process added 2430 GO-related and 644 PIR-related features.

Experimental and analytical setup

The prostate tumor dataset was used to produce Utility Profiles, by running a large number of differently configured genetic programming-based categorization processes to create a diverse classification model ensemble. The execution of the genetic programming algorithm

was done using the Biomind ArrayGenius Software, available at http://ondemand.biomind.com:8090. In particular, we used the metatasking capability of ArrayGenius: the software, upon receiving the dataset as input, was instructed to run 1000 genetic programming (GP; Koza, 1992) classification tasks with parameters selected randomly within specified ranges. Ranges used for parameter variation are detailed below (parameters not mentioned were left at their ArrayGenius default values):

- All combinations of use of direct and categorial features (see section above on feature vector enhancement) were allowed.
- Each GP task used only the top d most differentiated (among categories) features in the dataset, where d is a randomly chosen number between 10 and 1000.
- In terms of GP-specific parameters, fitness function was varied across all available alternatives.

Utility Profiles were then built using the results of the thousand categorization tests produced by the means above. Recalling the basics of MUTIC, a utility profile of a given feature is the vector composed of the utility or importance values of that feature in each one of the 1000 tests. Utility in this context is simply the frequency of utilization of the feature across the classifiers in an ensemble generated by a given task.

The Utility Profiles produced in this manner were then used as inputs for Omniclust clustering. This produces a set of feature clusters (where each feature can be a gene or a gene category, as described above). These clusters may have a more general semantics than clusters formed directly from gene expression vectors using standard methods. In these utilization-based clusters, features are gathered together if classification models habitually find it useful to consider them together.

For the sake of comparison, we also used Omniclust to perform clustering of the prostate cancer (PC) dataset in the traditional way, by simply clustering the feature vectors associated with the gene expression profiles. As usual in the MUTIC approach, only the clusters at the first level of the dendrogram produced by Omniclust were analyzed both quantitative and qualitatively.

The purpose of the quantitative comparison is to show that the clusters obtained using MUTIC are of high quality in a purely mathematical sense. The purpose of the qualitative analysis is to look for novel biological insights that MUTIC may have uncovered. The results obtained with both approaches are detailed in the next section.

RESULTS

Quantitative comparison

Clustering is a qualitative data analysis method; there are no robust, commonly accepted, objective metrics for comparing different clustering algorithms to each other. Dopazo and Azuaje (2005) gave a comprehensive overview of contemporary clustering methods and a review of methods for comparing them to each other.

Choosing a variant of a standard technique, we measured the quality of a clustering as the product homogeneity x separation. Homogeneity is calculated as *1/(1+A)* where A is the average of the distances of all members of the cluster to their nearest cluster-mates. Separation is simply the minimum distance from any given member of the cluster to elements outside the

cluster. These particular definitions of separation and homogeneity were used in order to minimize the influence of the size of the cluster on its quality. (As we have observed empirically, using more traditional definitions of separation and homogeneity, e.g., defining homogeneity as the average of all similarities among all members of a cluster, causes small clusters to habitually display a better quality than larger ones, which is an undesirable bias.)

If one straightforwardly compares MUTIC to plain expression-based clustering, according to this cluster quality metric, one finds that MUTIC produces dramatically clearer clusters, with roughly 10 times greater quality. This comparison, however, is somewhat unfair to the standard method, because the separation values are bound to be larger for MUTIC simply because it involves fewer features (only the ones that have nonzero model usage). Thus, to make a fairer comparison, we also tried standard expression-based clustering using a smaller set of features: only the *N* most-differentiated features, where differentiation was measured using the same categories used for supervised categorization, and *N = 1000* was chosen as the same number of features having nontrivial Utility Profiles. (*N = 1000* was a consequence of the feature selection policy used in the GP classification experiments, explained above.) The results of this comparison are shown in Table 1. As we see, MUTIC still comes out far ahead here, with a cluster quality still around one order of magnitude higher.

Table 1. Clustering qualities.

Clustering	Quality of 1st cluster	Quality of 20th cluster
Utilization-based clustering	0.6226	0.5423
Expression-based clustering	0.0550	0.0083
Expression-based clustering, 1000 most differentiated features	0.0948	0.0111
Expression-based clustering, 1000 most differentiated features, using Average sparseness policy	0.3695	0.1695
Expression-based clustering, 1000 most differentiated features, using Median sparseness policy	0.4149	0.1819
Expression-based clustering, 1000 most differentiated features, using Custom sparseness policy	0.3691	0.2683

Another possible source of unfair comparisons could be the sparse nature of utility-based vectors as compared to gene expression vectors. In order to detect a potential unfair advantage based on sparseness, we applied three different sparseness policies to the gene expression vectors:

- Average Policy: all values in a given feature vector below the average of those values were set to zero.
- Median Policy: all values in a given feature vector below the median of those values were set to zero.
- Custom Policy: in a generalization of the Median Policy, in this one, all the lowest P % values in a given feature vector are set to zero. P was chosen as the average sparseness ratio (number of zero-ed dimensions over the total number of dimensions) in the utility-based data.

Using any one of these three sparseness policies raises the quality of the expression-based clustering to the same order of magnitude as the utility-based clustering. Nevertheless, even the highest quality value (achieved using the Median Policy) is roughly 2/3 of the quality obtained for utility-based clustering. Also, the quality differences between the 1st and 20th ranked clusters indicate a sharper decline of quality as rank increases when clustering sparsified expression vectors as opposed to utility vectors. It appears, therefore, that only part of the high quality of the utility-based clusters is explained by the sparseness of the utility vectors, as shown by Goertzel et al. (2006).

We emphasize that our cluster quality assessment method was in no way engineered to favor the utilization-based clusters; and nor was the Omniclust method devised specifically to showcase utilization-based clustering. In fact, it was devised for standard expression-based clustering and is used for this purpose within the Biomind ArrayGenius product. The essential result is that the clusters found via utilization-based clustering are drastically more clear and distinct than what traditional expression-based clustering yields.

Qualitative comparison

Below are the analyses of the top 5 quality clusters from the PC dataset.
Cluster #1 includes the features:
- 960_g_at   Cell division cycle 42 (CDC42)
- GO:0005220   Inositol 1,4,5-triphosphate-sensitive calcium-release channel activity (INS3P)
- NM_004651   *Homo sapiens* ubiquitin specific protease 11 (USP11)
- GO:0004868   Serpin
- GO:0007422   Peripheral nervous system development (PNSD)
- GO:0019887   Protein kinase regulator activity
- SF002466   Glycophorin
- GO:0019867   Outer membrane
- GO:0003700   Transcription factor activity
- NM_014015   *Homo sapiens* dexamethasone-induced transcript (DEXI)
- XM_032901   *Homo sapiens* KIAA0226 gene product (KIAA0226)

This cluster displays various features with direct and indirect relation to PC, the majority has roles in signal transduction that lead to cell differentiation and multiplication (CDC42, INS3P, USP11, protein kinase regulator activity, transcription factor activity, DEXI). CDC42 is directly related to cell cycle and PC (Erlich et al., 2006). Active BRCA2 leads to cell growth and proliferation in PC, and USP11 seems to inactivate BRCA2 ubiquitination in breast cancer (Schoenfeld et al., 2004; Moro et al., 2006). Its role in PC modulation with BRCA2 should be studied further. Protein kinase regulator and transcription factor activities together with DEXI are related to cell transduction pathways, both already described having roles in PC (Zerbini et al., 2005; Clark et al., 2005; Shimada et al., 2006). DEXI is known to suppress PC tumor angiogenesis (Yano et al., 2006). Some other features, such as serpin and PNSD, have distinct roles in PC. The first, is a family of proteins known to be bound to 70-90% circulating PSA, the prostate-specific antigen, a known PC marker (Stephan et al., 2000; Kuvibidila and Rayford, 2006). Peripheral nervous systems are known to regulate prostate growth and function. There have been studies linking

neuropeptide receptors from peripheral nervous system (muscarinic) to PC growth modulation (Ventura et al., 2002). Some other features have unkown roles in PC (KIAA0226, glycophorin and outer membrane). The roles of these features in PC should be studied more carefully.

Cluster #2 includes the features:

- 1280_i_at   Serine/threonine kinase (STK)
- GO:0015629   Actin cytoskeleton
- NM_003374   *Homo sapiens* voltage-dependent anion channel 1 (VDAC1)
- GO:0005635   Nuclear membrane
- GO:0005778   Peroxisomal membrane
- GO:0006560   Proline metabolism
- NM_016621   *Homo sapiens* BRAF35/HDAC2 complex (BHC80)
- NM_013231   *Homo sapiens* fibronectin leucine-rich transmembrane protein 2 (FLRT2).
- GO:0008435   Anticoagulant activity (AAc)

This cluster provides many features related to PC, especially to cell growth and proliferation (STK, BHC80 and AAc). STK is known to regulate PC (Schrantz et al., 2004; Clark et al., 2005). BHC80 complex is related to cell cycle regulation, and a spliced form of BRAF35, BRAF25, is known to be down regulated in PC (Wang et al., 2002). Also, mutations in histone deacetylase 2 were found to be related to human cancers (Ropero et al., 2006). The coagulation system has a major role in PC (Kohli et al., 2002). Although also related to the coagulation system, some enzymes and enzyme precursors are involved in cancer metastasis, especially the plasmin/plasminogen system (Duffy and Duggan, 2004). Plasminogen activator inhibitor type 1 has roles in cancer dissemination and is more expressed in PC tissue (Chorostowska-Wynimko et al., 2004; Riddick et al., 2005). VDAC1 does not have described roles in PC but is known to regulate apoptosis (Abu-Hamad et al., 2006). Does it have any regulation role in PC? Other features from this cluster (nuclear membrane, actin cytoskeleton) appear to have roles in cellular morphology of tumor cells. There have been reports on nuclear membrane alterations in PC (Fischer et al., 2004). Actin cytoskeleton reorganization leads to cell differentiation and migration in cancers, and it seems that human PC cells have testosterone receptors which when stimulated cause modifications in actin cytoskeleton (Kampa et al., 2002). Also, a serine/threonine kinase (first feature of the cluster), named PAK6, seems to be correlated to actin reorganization and is expressed differently among PC cell lines (Schrantz et al., 2004). Some receptors, such as the PPAR, in peroxisomal membrane have shown a relation to PC (Leibowitz and Kantoff, 2003). Other features do not show any clear relation to PC (FLRT2 and proline metabolism).

Cluster #3 includes the features:

- GO:0006493   O-linked glycosylation
- GO:0007338   Fertilization (sensu Metazoa)
- GO:0045595   Regulation of cell differentiation
- NM_000853   *Homo sapiens* glutathione S-transferase theta 1 (GSTT1)
- GO:0012505   Endomembrane system
- NM_014874   *Homo sapiens* mitofusin 2, nuclear gene encoding mitochondrial protein (MFN2)
- GO:0016049   Cell growth

This cluster has interesting features related to PC. Cell growth and regulation of cell differentiation (GO:0016049, GO:0045595) are features unarguably related to most kinds of human cancers. Some antigens present in PC tumors are glucosylated and have O-linked (GO:0006493) and N-linked oligosaccharide chains (Holmes et al., 1996). Also, the tissue factor protein (initiator of coagulation) is known to bind to plasminogen with N- and O-linked oligosaccharide chains (Gonzalez-Gronow et al., 2002). The endomembrane system (GO:0012505), composed of intracellular membranous trafficking system, as Golgi, ER, vesicles, etc., was found to have roles in PC as some particular vesicles are found in PC cells, called prostasomes (Llorente et al., 2004). Also, the prostate-specific membrane antigen is directed to the plasma membrane in appropriate post-Golgi vesicles, with dependence upon N-glycosylation of the protein and help from microtubules (Christiansen et al., 2005). GSTT1 is known to be mutated in some kinds of PC (Dong, 2006; Yang et al., 2006). There was no evidence of direct relation of MFN2, a protein known to regulate membrane fusion in mitochondria, to PC, but mitochondria are known to be of great importance in PC (Costello et al., 2005). More studies are required to prove MFN2 relations to PC. Fertilization (GO:0007338), characterized as the union of male and female gametes, is an unusual feature present in this cluster and its relations to PC can be extrapolated. Fertilization is probably altered in PC patients, as it has been shown that sperm motility and maturation is enhanced in some kinds of PC (Yeung et al., 1997; Wang et al., 2001).

Cluster #4 includes the features:
- GO:0005216   Ion channel activity
- GO:0006171   cAMP biosynthesis
- SF002282   Cytoskeletal keratin
- GO:0007397   Histogenesis and organogenesis

All features present in this cluster have connections to PC. Histogenesis and organogenesis (GO:0007397) are terms related to the formation of tumor tissues and cancers by definition. There is a cytokeratin (SF002282) highly expressed in many cancers, including PC (Egland et al., 2006). Activities of some ion channels (GO:0005216), such as potassium and calcium channels (alpha-types), have been related to PC cell proliferation (Skryma et al., 1997; Mariot et al., 2002; Van Coppenolle et al., 2004). PC cells initially depend on circulating androgens. These hormones activate signal transduction pathways via G protein receptors with cAMP production (GO:0006171) via adenylate cyclase. Some hormone agonists and antagonists seem to decrease the metastatic progression of PC (Dondi et al., 2006).

Cluster #5 includes the features:
- 1175_s_at   Cytochrome P450, family 2, subfamily C, polypeptide 8 (CYP2C8)
- GO:0006330   Single-stranded DNA binding
- GO:0042578   Phosphoric ester hydrolase activity
- NM_005419   *Homo sapiens* signal transducer and activator of transcription 2113 kDa (STAT2)
- NM_018651   *Homo sapiens* zinc finger protein 167 (ZNF167)

Although too general, the single-stranded DNA binding feature (GO:0006330) has a member protein with direct roles in PC cell survival and proliferation, as well as in other cancers: human telomerase (Folini et al., 2005). The enzyme, phosphodiesterase 4, responsible for cAMP breakdown (a phosphoric ester hydrolase activity - GO:0042578) has been shown to be

hypomethylated in PC cells (Ho et al., 2006). Some STAT proteins are found to be activated in PC, but there is no direct relation for STAT2 (Ni et al., 2002). Is this molecule also involved? The other features have no direct evidence of relation to PC (CYP2C8, ZNF167), and more specific tests are needed in order to determine their relations to PC.

DISCUSSION

We applied a novel analytical method, model utilization-based clustering or MUTIC, to a gene expression dataset pertinent to the genetics of PC.

The method corroborated previous results (Goertzel et al., 2006) as it produced clusters with high significance, biologically and mathematically. As shown in the qualitative comparison section, it was able to stress inter-gene and inter-process interactions that could not be identified via standard expression-based clustering analysis.

From the biological analysis of the clusters, 70% of the features from the 5 better quality clusters analyzed have some level of relation to PC (confirmed by the literature). The 30% remaining features need a more detailed study to confirm their relations to PC. From this set, many features have unknown function (e.g., KIAA0226, from cluster 1). With 'wet lab' experiments, those features could be linked to PC and/or have its function discovered or more detailed.

In conclusion, the results presented here reinforce the potential of this method to aggregate genes and gene categories that are in fact relevant to the biological phenomenon under study. MUTIC could also be used as a good tool for biologists to study gene products with still unclear functions as well as to help expand and extrapolate the existing biological pathway and ontology databases (e.g., KEGG, Reactome, Biocarta, GO, etc.) with the novel gene linkage it produces.

REFERENCES

Abu-Hamad S, Sivan S and Shoshan-Barmatz V (2006). The expression level of the voltage-dependent anion channel controls life and death of the cell. *Proc. Natl. Acad. Sci. USA* 103: 5787-5792.

Ben-Dor A, Shamir R and Yakhini Z (1999). Clustering gene expression patterns. *J. Comput. Biol.* 6: 281-297.

Brown MP, Grundy WN, Lin D, Cristianini N, et al. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA* 97: 262-267.

Cho JH, Lee D, Park JH and Lee IB (2004). Gene selection and classification from microarray data using kernel machine. *FEBS Lett.* 571: 93-98.

Chorostowska-Wynimko J, Skrzypczak-Jankun E and Jankun J (2004). Plasminogen activator inhibitor type-1: its structure, biological activity and role in tumorigenesis. *Int. J. Mol. Med.* 13: 759-766.

Christiansen JJ, Rajasekaran SA, Inge L, Cheng L, et al. (2005). N-glycosylation and microtubule integrity are involved in apical targeting of prostate-specific membrane antigen: implications for immunotherapy. *Mol. Cancer Ther.* 4: 704-714.

Clark DE, Errington TM, Smith JA, Frierson HF Jr, et al. (2005). The serine/threonine protein kinase, p90 ribosomal S6 kinase, is an important regulator of prostate cancer cell proliferation. *Cancer Res.* 65: 3108-3116.

Costello LC, Franklin RB and Feng P (2005). Mitochondrial function, zinc, and intermediary metabolism relationships in normal prostate and prostate cancer. *Mitochondrion* 5: 143-153.

Dondi D, Festuccia C, Piccolella M, Bologna M, et al. (2006). GnRH agonists and antagonists decrease the metastatic progression of human prostate cancer cell lines by inhibiting the plasminogen activator system. *Oncol. Rep.* 15: 393-400.

Dong JT (2006). Prevalent mutations in prostate cancer. *J. Cell Biochem.* 97: 433-447.

Dopazo J and Azuaje F (2005). Data analysis and visualization in genomics and proteomics. John Wiley and Sons, Chichester, West Sussex, Hoboken.

Dudoit S, Fridlyand J and Speed T (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* 97: 77-87.

Duffy MJ and Duggan C (2004). The urokinase plasminogen activator system: a rich source of tumour markers for the individualised management of patients with cancer. *Clin. Biochem.* 37: 541-548.

Egland KA, Liu XF, Squires S, Nagata S, et al. (2006). High expression of a cytokeratin-associated protein in many cancers. *Proc. Natl. Acad. Sci. USA* 103: 5929-5934.

Eisen MB, Spellman PT, Brown PO and Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95: 14863-14868.

Erlich S, Tal-Or P, Liebling R, Blum R, et al. (2006). Ras inhibition results in growth arrest and death of androgen-dependent and androgen-independent prostate cancer cells. *Biochem. Pharmacol.* 72: 427-436.

Fischer AH, Bardarov S Jr and Jiang Z (2004). Molecular aspects of diagnostic nucleolar and nuclear envelope changes in prostate cancer. *J. Cell Biochem.* 91: 170-184.

Folini M, Brambilla C, Villa R, Gandellini P, et al. (2005). Antisense oligonucleotide-mediated inhibition of hTERT, but not hTERC, induces rapid cell growth decline and apoptosis in the absence of telomere shortening in human prostate cancer cells. *Eur. J. Cancer* 41: 624-634.

Goertzel B, Pennachin C, de Souza Coelho L and Mudado M (2006). Identifying complex biological interactions based on categorical gene expression data. In: Proceedings of the 2006 IEEE Congress on Evolutionary Computation, Vancouver, 1434-1441.

Golub TR, Slonim DK, Tamayo P, Huard C, et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286: 531-537.

Gonzalez-Gronow M, Gawdi G and Pizzo SV (2002). Tissue factor is the receptor for plasminogen type 1 on 1-LN human prostate cancer cells. *Blood* 99: 4562-4567.

Guyon I, Weston J, Barnhill S and Vapnik V (2002). Gene selection for cancer classification using support vector machines. *Machine Learning* 46: 389-422.

Holmes EH, Greene TG, Tino WT, Boynton AL, et al. (1996). Analysis of glycosylation of prostate-specific membrane antigen derived from LNCaP cells, prostatic carcinoma tumors, and serum from prostate cancer patients. *Prostate* 7: 25-29.

Kampa M, Papakonstanti EA, Hatzoglou A, Stathopoulos EN, et al. (2002). The human prostate cancer cell line LNCaP bears functional membrane testosterone receptors that increase PSA secretion and modify actin cytoskeleton. *FASEB J.* 16: 1429-1431.

Kohli M, Fink LM, Spencer HJ and Zent CS (2002). Advanced prostate cancer activates coagulation: a controlled study of activation markers of coagulation in ambulatory patients with localized and advanced prostate cancer. *Blood Coagul. Fibrinolysis* 13: 1-5.

Koza JR (1992). Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge.

Kuvibidila S and Rayford W (2006). Correlation between serum prostate-specific antigen and alpha-1-antitrypsin in men without and with prostate cancer. *J. Lab. Clin. Med.* 147: 174-181.

Leibowitz SB and Kantoff PW (2003). Differentiating agents and the treatment of prostate cancer: vitamin D3 and peroxisome proliferator-activated receptor gamma ligands. *Semin. Oncol.* 30: 698-708.

Llorente A, de Marco MC and Alonso MA (2004). Caveolin-1 and MAL are located on prostasomes secreted by the prostate cancer PC-3 cell line. *J. Cell Sci.* 117: 5343-5351.

Mariot P, Vanoverberghe K, Lalevee N, Rossier MF, et al. (2002). Overexpression of an alpha 1H (Cav3.2) T-type calcium channel during neuroendocrine differentiation of human prostate cancer cells. *J. Biol. Chem.* 277: 10824-10833.

Markowetz F (2004). A bibliography on learning causal networks of gene interactions. http://www.molgen.mpg.de/~markowet/docs/network-bib.pdf or <http://www.molgen.mpg.de/%7Emarkowet/docs/network-bib.pdf>. Accessed September 20, 2007.

Markowetz F and Spang R (2003). Reconstructing gene regulation networks from passive observations and active interventions. In: 7th Ann. Intl. Conf. Res. Comput. Molec. Biol. (Panel).

Moro L, Arbini AA, Marra E and Greco M (2006). Up-regulation of Skp2 after prostate cancer cell adhesion to basement membranes results in BRCA2 degradation and cell proliferation. *J. Biol. Chem.* 281: 22100-22107.

Nachman I, Regev A and Friedman N (2004). Inferring quantitative models of regulatory networks from expression data. *Bioinformatics* 20: i248-i256.

Ni Z, Lou W, Lee SO, Dhir R, et al. (2002). Selective activation of members of the signal transducers and activators of transcription family in prostate carcinoma. *J. Urol.* 167: 1859-1862.

Riddick AC, Shukla CJ, Pennington CJ, Bass R, et al. (2005). Identification of degradome components associated with prostate cancer progression by expression analysis of human prostatic tissues. *Br. J. Cancer* 92: 2171-2180.

Ropero S, Fraga MF, Ballestar E, Hamelin R, et al. (2006). A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition. *Nat. Genet.* 38: 566-569.

Schoenfeld AR, Apgar S, Dolios G, Wang R, et al. (2004). BRCA2 is ubiquitinated *in vivo* and interacts with USP11, a deubiquitinating enzyme that exhibits prosurvival function in the cellular response to DNA damage. *Mol. Cell Biol.* 24: 7444-7455.

Schrantz N, da Silva CJ, Fowler B, Ge Q, et al. (2004). Mechanism of p21-activated kinase 6-mediated inhibition of androgen receptor signaling. *J. Biol. Chem.* 279: 1922-1931.

Sharan R and Shamir R (2000). CLICK: a clustering algorithm with applications to gene expression analysis. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 8: 307-316.

Sharan R, Elkon R and Shamir R (2001). Cluster analysis and its applications to gene expression data. In: Ernst Schering Workshop on Bioinformatics and Genome Analysis, Springer Verlag, Berlin.

Shimada K, Nakamura M, Ishida E and Konishi N (2006). Molecular roles of MAP kinases and FADD phosphorylation in prostate cancer. *Histol. Histopathol.* 21: 415-422.

Singh D, Febbo PG, Ross K, Jackson DG, et al. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1: 203-209.

Skryma RN, Prevarskaya NB, Dufy-Barbe L, Odessa MF, et al. (1997). Potassium conductance in the androgen-sensitive prostate cancer cell line, LNCaP: involvement in cell proliferation. *Prostate* 33: 112-122.

Sohler F, Hanisch D and Zimmer R (2004). New methods for joint analysis of biological networks and expression data. *Bioinformatics* 20: 1517-1521.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, et al. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9: 3273-3297.

Stephan C, Jung K, Lein M, Sinha P, et al. (2000). Molecular forms of prostate-specific antigen and human kallikrein 2 as promising tools for early diagnosis of prostate cancer. *Cancer Epidemiol. Biomarkers Prev.* 9: 1133-1147.

Tamayo P, Slonim D, Mesirov J, Zhu Q, et al. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* 96: 2907-2912.

Tan AC and Gilbert D (2003). Ensemble machine learning on gene expression data for cancer classification. *Appl. Bioinformatics* 2: S75-S83.

Van Coppenolle F, Skryma R, Ouadid-Ahidouch H, Slomianny C, et al. (2004). Prolactin stimulates cell proliferation through a long form of prolactin receptor and K+ channel activation. *Biochem. J.* 377: 569-578.

Ventura S, Pennefather J and Mitchelson F (2002). Cholinergic innervation and function in the prostate gland. *Pharmacol. Ther.* 94: 93-112.

Vert JP and Kanehisa M (2003). Extracting active pathways from gene expression data. *Bioinformatics* 19: ii238-ii244.

Wang C, McCarty IM, Balazs L, Li Y, et al. (2002). Immunohistological detection of BRAF25 in human prostate tumor and cancer specimens. *Biochem. Biophys. Res. Commun.* 295: 136-141.

Wang J, Lundqvist M, Carlsson L, Nilsson O, et al. (2001). Prostasome-like granules from the PC-3 prostate cancer cell line increase the motility of washed human spermatozoa and adhere to the sperm. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 96: 88-97.

Welsh JB, Sapinoso LM, Su AI, Kern SG, et al. (2001). Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* 61: 5974-5978.

Yang J, Wu HF, Zhang W, Gu M, et al. (2006). Polymorphisms of metabolic enzyme genes, living habits and prostate cancer susceptibility. *Front. Biosci.* 11: 2052-2060.

Yano A, Fujii Y, Iwai A, Kawakami S, et al. (2006). Glucocorticoids suppress tumor angiogenesis and *in vivo* growth of prostate cancer cells. *Clin. Cancer Res.* 12: 3003-3009.

Yeung CH, Perez-Sanchez F, Soler C, Poser D, et al. (1997). Maturation of human spermatozoa (from selected epididymides of prostatic carcinoma patients) with respect to their morphology and ability to undergo the acrosome reaction. *Hum. Reprod. Update* 3: 205-213.

Zerbini LF, Wang Y, Correa RG, Cho JY, et al. (2005). Blockage of NF-kappaB induces serine 15 phosphorylation of mutant p53 by JNK kinase in prostate cancer cells. *Cell Cycle* 4: 1247-1253.