Research article

Multivariate gene expression analysis reveals functional connectivity changes between normal/tumoral prostates

André Fujita*[1], Luciana Rodrigues Gomes[2], João Ricardo Sato[3], Rui Yamaguchi[1], Carlos Eduardo Thomaz[4], Mari Cleide Sogayar[2] and Satoru Miyano[1]

Address: [1]Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan, [2]Chemistry Institute, University of São Paulo, Av. Lineu Prestes, 748, São Paulo-SP, 05508-900, Brazil, [3]Mathematics, Computation and Cognition Center, Universidade Federal do ABC, Rua Santa Adélia, 166 – Santo André, 09210-170, Brazil and [4]Department of Electrical Engineering, Centro Universitário da FEI, Av. Humberto de Alencar Castelo Branco, 3972 – São Bernardo do Campo, 09850-901, Brazil

Email: André Fujita* - afujita@ims.u-tokyo.ac.jp ; Luciana Rodrigues Gomes - lugomes@iq.usp.br; João Ricardo Sato - jrsatobr@gmail.com; Rui Yamaguchi - ruiy@ims.u-tokyo.ac.jp; Carlos Eduardo Thomaz - cet@fei.edu.br; Mari Cleide Sogayar - mcsoga@iq.usp.br; Satoru Miyano - miyano@ims.u-tokyo.ac.jp

* Corresponding author

Abstract

Background: Prostate cancer is a leading cause of death in the male population, therefore, a comprehensive study about the genes and the molecular networks involved in the tumoral prostate process becomes necessary. In order to understand the biological process behind potential biomarkers, we have analyzed a set of 57 cDNA microarrays containing ~25,000 genes.

Results: Principal Component Analysis (PCA) combined with the Maximum-entropy Linear Discriminant Analysis (MLDA) were applied in order to identify genes with the most discriminative information between normal and tumoral prostatic tissues. Data analysis was carried out using three different approaches, namely: (i) differences in gene expression levels between normal and tumoral conditions from an univariate point of view; (ii) in a multivariate fashion using MLDA; and (iii) with a dependence network approach. Our results show that malignant transformation in the prostatic tissue is more related to functional connectivity changes in their dependence networks than to differential gene expression. The MYLK, KLK2, KLK3, HAN11, LTF, CSRP1 and TGM4 genes presented significant changes in their functional connectivity between normal and tumoral conditions and were also classified as the top seven most informative genes for the prostate cancer genesis process by our discriminant analysis. Moreover, among the identified genes we found classically known biomarkers and genes which are closely related to tumoral prostate, such as KLK3 and KLK2 and several other potential ones.

Conclusion: We have demonstrated that changes in functional connectivity may be implicit in the biological process which renders some genes more informative to discriminate between normal and tumoral conditions. Using the proposed method, namely, MLDA, in order to analyze the multivariate characteristic of genes, it was possible to capture the changes in dependence networks which are related to cell transformation.

Background

Cancer is one of the main public health problems in the United States and worldwide [1]. Among the diverse types of neoplasia, prostate cancer is the third most common cancer in the World [2], being ranked as the second leading cause of death in men, the first being lung cancer [1]. Its incidence and mortality varies in different parts of the World, being highest in Western countries, mainly among Africans [3].

With the widespread use of the prostate-specific antigen (PSA) test, more men are examined, and consequently, identification of patients with asymptomatic low-stage tumors has increased considerably [4,5]. Although the majority of prostate cancers is confined to the prostate gland, rarely affecting life expectancy, in about 30% of the cases, a specialized group of cells from the primary tumor mass may invade and colonize other distant tissues causing death, therefore, metastatic disease rather than the primary tumor itself is responsible for death, causing the prognosis to be directly related to the spread of the tumor. Unfortunately, the therapeutic approaches used nowadays against advanced stages of prostatic cancers are not effective [6]. Therefore, it is extremely important to understand the basic molecular biology involved in this disease in order to prevent the progression of the tumor [6]. However, the identification and analysis of these molecular mechanisms has been hampered by the heterogeneity and high molecular complexity of the process involved in the development of this disease.

In the last few years, several efforts have been made towards determining the genetic mechanisms involved in the development of this tumor [6,7]. A widely used approach in studying the development of several types of cancers has been the high-throughput gene expression microarray analysis, which has provided a wealth of information about tumor marker genes. Conventional methods of microarray data analysis have been systematically used to examine the differentially expressed genes [8], and molecular pathways [9] and discriminative methods have been used in order to identify biomarkers [10,11].

In general, discriminant studies focus only on the classification accuracy of the method and on a pre-step selection of the features (genes) which best classifies the samples [12]. This selection of features is often carried out by selecting a subgroup of the most differentially expressed genes [13] or in a multivariate fashion [12]. However, understanding of the structure responsible for regulation of these discriminative set of genes in prostatic cancer is required [14].

Many years of intensive research have demonstrated that signaling molecules are organized into complex biochemical networks. These signaling circuits are complicated systems consisting of multiple elements interacting in a multifarious fashion. Signaling networks are regulated both in time and space [15]; allow the cell to decide which cellular process (cell division, differentiation, transformation, or apoptosis) is the most appropriate response for each situation. Due to the high connectivity and complexity of these biological systems, small modifications in a few members ("hub" genes, i.e., highly functionally connected genes) of these biochemical networks are sufficient to perturb the whole system [16], consequently resulting in a change on the cell's phenotype [17]. Frequently, changes in the relative concentration of molecules, such as mRNAs and proteins, are the unique parameter analyzed in biological systems. However, the biomolecules' concentration is not the only important variable, but their compartmentalization and diffusion are also determinants of the cell's phenotype. Therefore, these approaches are reductionists in defining a good biomarker as the most differentially expressed gene or protein when comparing distinct cellular contexts.

Here, we report a cDNA microarray-based study in prostatic cancer aimed at understanding why some genes are good predictors in discriminating normal versus tumoral samples and others are not. We demonstrate that the discriminative information between normal and tumoral prostates is related to the change in functional connectivity between certain genes and not necessarily in their differential expression, as has often been assumed. Moreover, we present a systematic and straightforward approach based on MLDA (Maximum-entropy Linear Discriminant Analysis) to identify putative biomarkers in high dimensional data (when the number of features is greater than the number of observations), and a dependence network analysis in order to interpret sets of discriminative genes. This idea is illustrated in Figure 1.

Results

Simulation

The combination of PCA (Principal Component Analysis) + MLDA (Maximum-entropy Linear Discriminant Analysis) [18] was applied in a simulated data described in the Methods section in order to demonstrate that functional connectivity changes may be captured by the proposed approach. Figure 2 describes the weights in absolute values attributed by MLDA to each feature (artifically generated genes). The features are sorted in a decreasing order of weight. Red crosses represent the genes which have their functional connectivity altered between conditions 1 and 2. Blue crosses represent the genes which have their connectivities unaltered.

Samples classification

Applying the PCA combined with the MLDA approach to all ~25,000 genes available in our microarray dataset [19], it was possible to classify the samples with an accuracy of
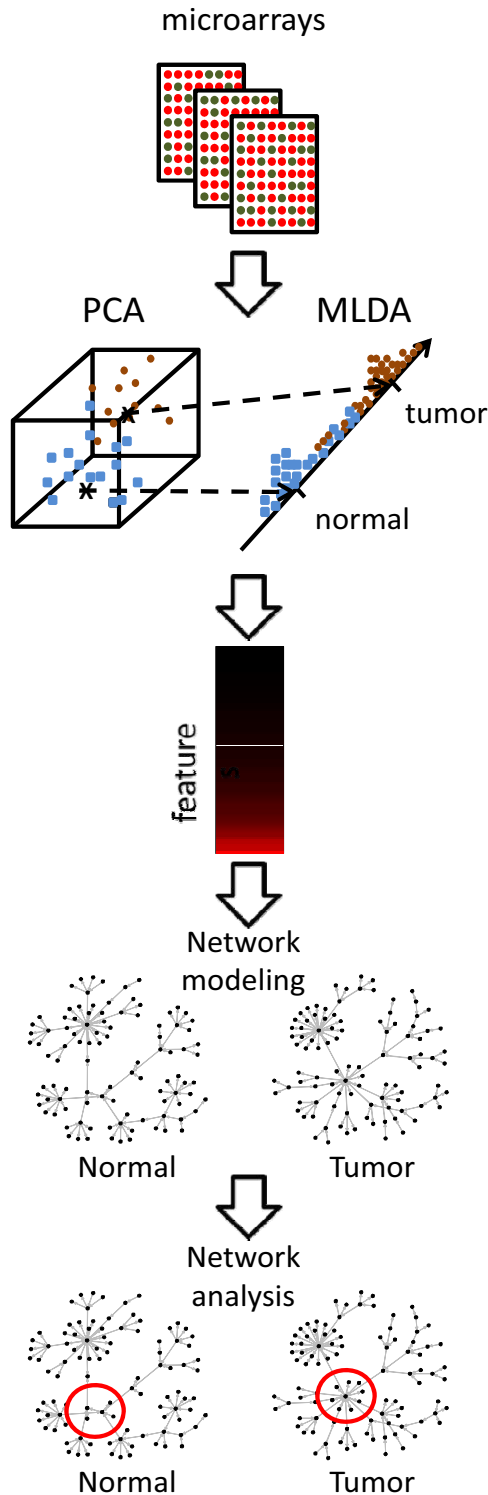
Figure 1
A pictorial scheme of the combination of PCA+MLDA and dependence network analysis for two populations (normal and tumoral prostatic tissues).
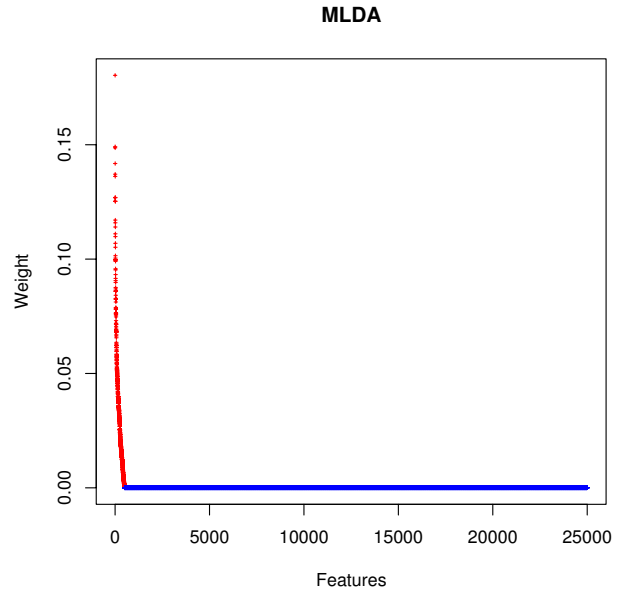
Figure 2
The discriminative weight of each simulated feature.
The features are sorted (in decreasing order) by the absolute value of the weight. Red crosses represent the 500 features that have their functional connectivities altered between conditions 1 and 2. Blue crosses represent the 24,500 features which have their functional connectivities unaltered.

96.5% (a misclassification of 2 out of 57 samples), using a leave-one-out cross validation.

Projection matrix ψ_{MLDA} analysis
The projection matrix ψ_{MLDA} contains the weights (degree of relationship between the gene and the normal/tumoral state) for each feature (gene). Figure 3 describes the weights in absolute values attributed by MLDA to each gene. The genes are sorted in a decreasing order of weight.

The most informative genes correlated to prostatic cancer
Table 1 illustrates the top 100 features identified as the most informative genes related to malignant transformation by the PCA+MLDA approach ranked in a decreasing order of weight values. This set of 100 most informative genes represents ~0.4% of the total number of genes available in the microarrays (~25,000 genes). Notice that these 100 genes have a MLDA weight different from zero, i.e., the 100th gene RPS28 has a MLDA weight (~0.035, Table 1) located before the convergence of the curve to zero (Figure 3, the horizontal red line indicates the 100th gene). In order to verify the stability and robustness of our results, 27 observations out of 32 from normal sample and 20 out of 25 from tumoral sample were randomly selected and the ψ_{MLDA} was re-calculated. This step was
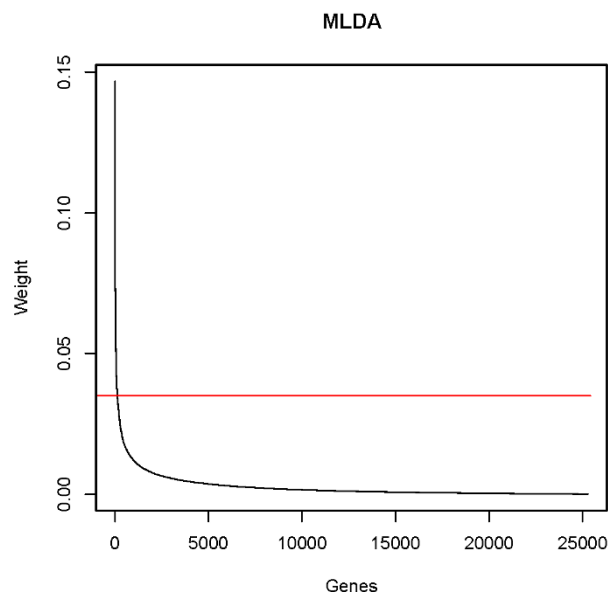
Figure 3
The discriminative weight of each gene. The genes are sorted (in decreasing order) by the absolute value of the weight. The horizontal red line indicates the 100th gene.

performed 100 times and the mean rank for each gene was obtained. About 80% of the originally obtained top 100 most discriminative genes were ranked as the top 100 most discriminative genes.

We have also manually annotated (which we believe be more accurate than automatic computer-based annotation, since it may be more efficient to capture semantic information from published articles) this set of 100 genes [see Table 1 and Additional file 1].

Putative differentially expressed genes
We have also searched for differentially expressed genes. About 25% of the genes listed in Table 1 do not present statistical evidence to be differentially expressed between normal and tumoral conditions.

Relevance networks
Both normal and tumoral relevance networks with the top 100 most informative genes were constructed, considering a false discovery rate of 5%, being illustrated in Figures 4 and 5, respectively. Nodes in red are the genes which have their functional connectivity (estimated using the non-parametric Hoeffding's D measure [20]) changed considerably between normal versus tumoral conditions, i.e., they become "hubs" (highly connected genes) [16] in tumoral prostates. "Hub" genes were maintained also

when relevance networks were constructed under different FDR thresholds (1, 5 and 10%).

Discussion
Firstly, the PCA+MLDA approach was applied to a simulated data set in order to illustrate that differences in connectivity may be behind the oncogenesis process. Sato *et al.* (2008) [21] have already demonstrated in another context (neuroscience) that the information contained in the connectivity may be useful to sample classification. The simulation was performed in a large scale multidimensional condition, where the relevant features (genes which have the connectivity changed) are only 2% (500 out of 25,000 genes). Interestinlgy, MLDA was able to correctly identify the discriminative features, represented by red crosses in Figure 2. Notice that the relevant features for discrimination do not present differential expression between conditions 1 and 2 (by construction).

In order to verify whether gene expression data contain the information to discriminate normal from tumoral prostatic samples, we have applied the PCA+MLDA approach to actual biological data, obtaining a high classification accuracy (96.5%) by the leave-one-out cross-validation. In this case, we have used all the principal components in order to avoid losing information. PCA is applied regarding computational cost and memory limitation. It is important to mention that the numerical results are identical in the absence of the PCA step [22]. Notice that MLDA does not require a pre-step feature selection, because it may also work for high dimensional data. Therefore, it was possible to include all of the 25,000 genes of the microarray dataset.

Since it was possible to verify that gene expression data retains information for classification, we analyzed the ψ_{MLDA} projection matrix which contains the weight values for each feature (gene). Notice that the majority of the genes shown in Figure 3 have weights near zero, and only a few genes actually have discriminative information (high weight).

By analyzing Table 1, it is possible to verify that most of the 100 informative genes had already been described in the literature as genes related to cancer (76 genes) and 45 genes had specifically been associated to prostate tumor. Interestingly, most of the other 24 genes do not have references describing their functionality. Therefore, they may be associated to cancer but have not been studied yet. The description of the 76 genes in the literature corroborates the results obtained by the PCA+MLDA method, indicating that these genes are informative to discriminate between normal and tumoral samples. The stability and robustnees of this result were verified by obtaining around 80% of the same top 100 genes when five obser-

Table 1: ψ_{MLDA}: the weights attributed by MLDA.

	Gene name	Official Full Name	ψ_{MLDA}	p-value (Wilcoxon)	References:
1	*MYLK	myosin light chain kinase	0.14672	0.00000	[24]
2	*KLK2	kallikrein-related peptidase 2	0.12512	0.01053	[49]
3	***KLK3**	kallikrein-related peptidase 3	0.12032	0.05625	[50]
4	HAN11	WD repeat domain 68	0.12019	0.00000	
5	*LTF	lactotransferrin	0.11594	0.00092	[39]
6	CSRP1	cysteine and glycine-rich protein 1	0.11355	0.00000	[51]
7	***TGM4**	transglutaminase 4 (prostate)	0.10452	0.06063	[42]
8	*ACTG2	actin gamma 2 smooth muscle enteric	0.09826	0.00000	[52]
9	MYL6	myosin light chain 6 alkali smooth muscle and non-muscle	0.09817	0.00045	[53]
10	*RDH11	retinol dehydrogenase 11 (all-trans/9-cis/11-cis)	0.09583	0.00018	[54]
11	*AZGP1	alpha-2-glycoprotein 1 zinc-binding	0.08817	0.00059	[55]
12	NPAL3	NIPA-like domain containing 3	0.08478	0.00008	
13	**PRO1073**	PRO1073 protein	0.08077	0.28733	
14	***FXYD3**	FXYD domain containing ion transport regulator 3	0.08024	0.05417	[56]
15	TPM2	tropomyosin 2 (beta)	0.07919	0.00001	[57]
16	CRYAB	crystallin alpha B	0.07560	0.00000	[58]
17	ACTA2	actin alpha 2 smooth muscle aorta	0.07372	0.01610	[59]
18	***RPS6**	ribosomal protein S6	0.07323	0.12130	[60]
19	TMEM130	transmembrane protein 130	0.07296	0.00005	
20	*ACPP	acid phosphatase prostate	0.07185	0.00037	[61]
21	*PCP4	Purkinje cell protein 4	0.07128	0.00000	[62]
22	*SYNPO2	synaptopodin 2	0.06943	0.00000	[63]
23	*SORBS1	sorbin and SH3 domain containing 1	0.06773	0.00000	[64]
24	*MSMB	microseminoprotein beta	0.06588	0.00076	[65]
25	ACTC	actin alpha cardiac muscle 1	0.06335	0.00001	
26	*TGFB3	transforming growth factor beta 3	0.06313	0.00000	[66]
27	***MALT1**	mucosa associated lymphoid tissue lymphoma translocation gene 1	0.06205	0.14208	[67]
28	ZNF532	zinc finger protein 532	0.06131	0.00000	
29	ANXA1	annexin A1	0.06119	0.00001	[68]
30	PALLD	palladin cytoskeletal associated protein	0.06116	0.00000	[69]
31	*MT2A	metallothionein 2A	0.06054	0.00141	[70]
32	ING5	inhibitor of growth family member 5	0.05872	0.93009	[71]
33	PGM5	phosphoglucomutase 5	0.05862	0.00000	
34	**SERPINA3**	serpin peptidase inhibitor clade A (alpha-1 antiproteinase antitrypsin) member 3	0.05828	0.19710	[72]
35	*KRT5	keratin 5 (epidermolysis bullosa simplex Dowling-Meara/Kobner/Weber-Cockayne types)	0.05699	0.00000	[73]
36	**RPL5**	ribosomal protein L5	0.05589	0.53873	[74]
37	*IGF1	insulin-like growth factor 1 (somatomedin C)	0.05549	0.00000	[75]
38	**ZNF92**	zinc finger protein 92 (HTF12)	0.05388	0.16056	
39	***FOLH1**	folate hydrolase (prostate-specific membrane antigen) 1	0.05361	0.08683	[76]
40	*CYR61	cysteine-rich angiogenic inducer 61	0.05318	0.00020	[77]
41	FHL1	four and a half LIM domains 1	0.05305	0.00000	[78]
42	*H19	H19 imprinted maternally expressed transcript	0.05221	0.00006	[79]
43	DMN	desmuslin	0.05219	0.00000	
44	NEFH	neurofilament heavy polypeptide 200 kDa	0.05186	0.00001	[80]
45	PPP1R12B	protein phosphatase 1 regulatory (inhibitor) subunit 12B	0.05149	0.00000	
46	ANTXR2	anthrax toxin receptor 2	0.05141	0.00002	[81]
47	MRLC2	myosin regulatory light chain MRLC2	0.05056	0.02204	[82]
48	C20orf103	chromosome 20 open reading frame 103	0.05055	0.00150	
49	UBA52	ubiquitin A-52 residue ribosomal protein fusion product 1	0.05033	0.00518	[83]
50	TRGV9	T cell receptor gamma variable 9	0.04983	0.00190	
51	*SPARC	secreted protein acidic cysteine-rich (osteonectin)	0.04969	0.00240	[84]
52	*AMACR	alpha-methylacyl-CoA racemase	0.04903	0.00011	[85]
53	**DNER**	delta/notch-like EGF repeat containing	0.04809	0.09301	[86]
54	PRNP	prion protein (p27-30)	0.04806	0.00000	[87]
55	PDK4	pyruvate dehydrogenase kinase isozyme 4	0.04751	0.00002	[88]
56	***APOD**	apolipoprotein D	0.04744	0.12931	[89]

Table 1: ψ_{MLDA}: the weights attributed by MLDA. *(Continued)*

|					
57	*HERPUD1	homocysteine-inducible endoplasmic reticulum stress-inducible ubiquitin-like domain member 1	0.04695	0.00001	[90]
58	FSTL1	follistatin-like 1	0.04692	0.00092	[91]
59	**HSPCB**	heat shock protein 90 kDa alpha (cytosolic) class B member 1	0.04663	0.08386	[92]
60	*GSTM2	glutathione S-transferase M2 (muscle)	0.04446	0.00000	[93]
61	*PTN	pleiotrophin	0.04440	0.00000	[94]
62	*ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)	0.04410	0.06528	[95]
63	*CTGF	connective tissue growth factor	0.04342	0.00004	[96]
64	*GUCY1A3	guanylate cyclase 1 soluble alpha 3	0.04303	0.05841	[97]
65	MT1F	metallothionein 1F	0.04303	0.00002	[98]
66	*TIMP3	TIMP metallopeptidase inhibitor 3	0.04225	0.00000	[99]
67	*LDHB	lactate dehydrogenase B	0.04217	0.00000	[100]
68	RNASE4	ribonuclease RNase A family 4	0.04167	0.00000	
69	ANPEP	alanyl aminopeptidase	0.04165	0.00002	[101]
70	*CAV1	caveolin 1 caveolae protein 22 kDa	0.04135	0.00000	[102]
71	TM9SF2	transmembrane 9 superfamily member 2	0.04122	0.01275	
72	*HSPB8	heat shock 22 kDa protein 8	0.04088	0.00000	[103]
73	TUBA1A	tubulin alpha 1a	0.04087	0.00018	
74	**PDLIM5**	PDZ and LIM domain 5	0.04077	0.32533	[104]
75	LPP	LIM domain containing preferred translocation partner in lipoma	0.04073	0.00003	[105]
76	**MAD2L1BP**	MAD2L1 binding protein	0.04051	0.62639	[106]
77	*ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif 1	0.04048	0.00011	[107]
78	***RHOA**	ras homolog gene family member A	0.04039	0.11368	[108]
79	*TXNIP	thioredoxin interacting protein	0.03995	0.00227	[109]
80	**OGDH**	oxoglutarate (alpha-ketoglutarate) dehydrogenase (lipoamide)	0.03974	0.07543	
81	**RPL35**	ribosomal protein L35	0.03971	0.17555	
82	*ANKH	ankylosis progressive homolog (mouse)	0.03856	0.00318	[110]
83	MPST	mercaptopyruvate sulfurtransferase	0.03856	0.00000	[111]
84	MORF4L2	mortality factor 4 like 2	0.03831	0.01337	[112]
85	CRISPLD2	cysteine-rich secretory protein LCCL domain containing 2	0.03799	0.00000	
86	*CD9	CD9 molecule	0.03787	0.00150	[113]
87	ALDH3A2	aldehyde dehydrogenase 3 family member A2	0.03696	0.00001	
88	SCN2B	sodium channel voltage-gated type II beta	0.03693	0.00024	[114]
89	*SPARCL1	SPARC-like 1 (mast9 hevin)	0.03693	0.00045	[115]
90	IGJ	immunoglobulin J polypeptide linker protein for immunoglobulin alpha and mu polypeptides	0.03683	0.00190	[116]
91	ZNF134	zinc finger protein 134	0.03670	0.00007	
92	**MRPL43**	mitochondrial ribosomal protein L43	0.03655	0.54934	
93	LOC152485	hypothetical protein LOC152485	0.03647	0.00000	
94	**CALM2**	calmodulin 2 (phosphorylase kinase delta)	0.03622	0.05417	[117]
95	COL9A2	collagen type IX alpha 2	0.03546	0.00141	
96	*PAGE4	P antigen family member 4 (prostate associated)	0.03541	0.00001	[118]
97	CALM1	calmodulin 1 (phosphorylase kinase delta)	0.03536	0.00098	[119]
98	*ACTB	actin beta	0.03508	0.01159	[120]
99	***AGR2**	anterior gradient homolog 2 (Xenopus laevis)	0.03498	0.56006	[121]
100	**RPS28**	ribosomal protein S28	0.03497	0.15578	

*: genes already described to be related to prostatic cancer. In bold are the genes which do not present statistical evidences to be differentially expressed between normal and tumoral conditions.

vations were excluded randomly from normal sample and five from tumoral sample in 100 re-calculations. For more details about annotation of the top 100 genes and the complete list of the ~25,000 genes, please see Additional file 2.

Comparing the weights obtained by MLDA and the differentially expressed genes, it is surprising that the most differentially expressed genes are not necessarily the most

discriminative ones. In other words, a multivariate combination of genes may be regulating the normal/tumoral state, i.e., the combination of genes may contain more information about normal/tumoral conditions than an univariate differentially expressed gene.

Since it is known that a complex network is involved in the regulation of several molecular processes, we further analyzed the dependence network involved in these puta-
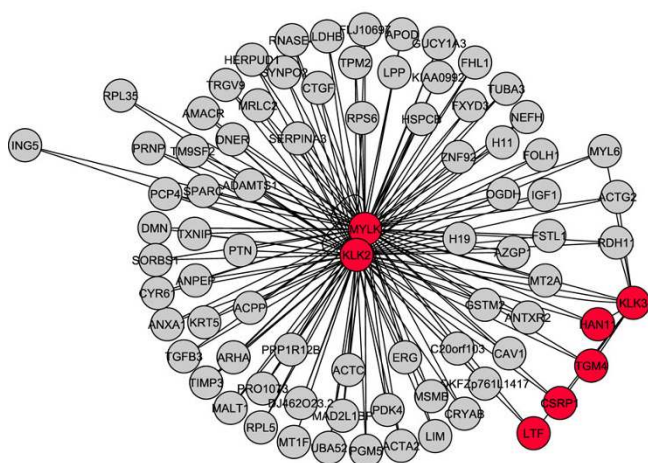
Figure 4
A normal prostate relevance network constructed with the top 100 most discriminative genes and FDR of 5%. Core genes are represented in red.

tive biomarkers in order to gain new insights. The analyis of Figures 4 and 5 indicate that exactly the top seven most discriminative genes described in Table 1 (MYLK, KLK2, KLK3, HAN11, LTF, CSRP1, TGM4) have considerably changed their functional connectivity between normal and tumoral conditions as illustrated by red nodes in Figures 4 and 5. These seven genes become "hubs" [16], i.e., highly connected genes in the tumoral condition, whereas in the normal condition, their connectivity was not different when compared to that of other genes. Furthermore, these seven genes maintained the position of the top seven most discriminative ones also when we have re-sampled the samples (the experiment which was performed in order to verify the stability and robustness of the top 100 genes). A Z-value summary table related to these seven genes is illustrated in Table 2. Z-values increase from normal to tumoral conditions, representing the changes in functional connectivities between these two conditions. The mean Z-values were calculated between the "hub" gene and the other 99 genes. In addition, in the list of the most discriminative features, there are genes which are more differentially expressed than these seven ones (lower p-value), however, their connectivity did not change. Krostka and Spang (2004) [17] have already suggested that differences in co-regulation between normal/disease states may be related to some pathologies. Moreover, Sato *et al*. (2008) [21] have reported that changes in networks connectivities may influence classification methods. These reports support our results showing that changes in functional connectivity may be closely related to the normal/tumoral states in prostate and that these changes in dependence may con-

tain an additional information when compared to differential gene expression.

Almost all top seven genes identified as the most discriminative features between normal and tumoral phenotypes had previously been described in the literature as being associated to cancer. The only gene that so far has not been correlated to cancer is HAN11, probably because little is known about this gene (only two articles were found in the literature describing this gene). Five of these top seven genes namely, MYLK, KLK2, KLK3, LTF and TGM4 had already been specifically related to prostate carcinoma (Table 1).

Myosin light chain kinase (MYLK) is one of them. This enzyme catalyzes the phosphorylation of a specific serine residue on the 20 kD light chain of myosin II (MCL20), consequently regulating the actin-myosin II interaction [23]. This reaction is responsible for smoothing muscle contraction/relaxation and organization of the cytoskeleton. Due to the central role played by the cytoskeleton in cell division and motility, it has been demonstrated that MYLK inhibition induces apoptosis in mammary prostate cancer cells and inhibits the growth of mammary and prostate tumors in rats and mice [24]. Furthermore, since MLC20 phosphorylation is necessary for cell motility [25,26], MYLK inhibition blocks cancer cell invasion and adhesion *in vitro*. As a result, some reports described the use of MYLK inhibitors as anti-cancer agents since they prevent cancer cells migration [27,28].

KLK3, also known as prostate specific antigen (PSA), is another gene which presents high functional connectivity in tumoral samples. PSA is a serine protease, secreted into seminal plasma, belonging to the human kallikrein gene family, being responsible for semen liquefaction. It is the first FDA (Food and Drug Administration)-approved tumor marker for cancer detection [29]. The prostatic gland volume affects the PSA level in serum, because it is produced and secreted by prostatic tissue [30,31]. However, increased levels of KLK3 are also observed in some patients with benign prostate hyperplasia. Therefore, elevated PSA concentration in patients' plasma may be indicative not only of prostate cancer, but, also of other prostatic pathologies. Consequently, the use of PSA as a cancer-specific marker is questioned.

Nowadays, 15 members of the kallikrein family (KLKs) are described in humans [32]. Among the KLKs, the highest homology is found between PSA and KLK2. In this case, the identity is 78% and 80% at the amino acid and DNA level, respectively [33]. KLK2 is another gene that presented functional connectivity changes between normal/tumoral conditions. The ratio of KLK2 to free PSA improves the discrimination of benign prostate hyperpla-
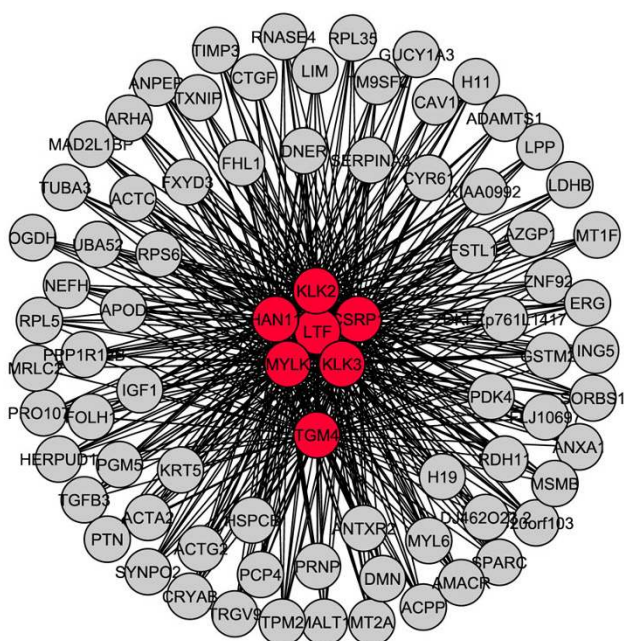
Figure 5
A tumoral prostate relevance network constructed with the top 100 most discriminative genes and FDR of 5%. Core genes are represented in red.

sia and prostate cancer patients [34]. In addition, it has already been described that KLK2 discriminates between high and low grade tumors [35]. There is evidence indicating that KLK2 is more closely correlated to the total volume and higher grade prostate cancers than PSA [36].

Identification of both of these classic biomarkers of prostate carcinomas (PSA and KLK2), in our list of the most informative genes, provides additional evidence to the hypothesis that functional connectivity changes and not only differential expression levels are highly correlated to normal/tumoral process.

Another gene classified as one of the most discriminative prostate cancer biomarkers, whose anti-tumorigenic role has already been described [37] is lactotransferrin (LTF). This non-heme iron-binding glycoprotein [38] is found in a variety of biological secretions, such as semen, as well as in several secretions derived from glandular epithelium cells, including the prostate. LTF mRNA and protein levels are downregulated in prostate cancer, with significant PSA recurrence associations, due to promoter silencing by hypermethylation [39]. It has been reported that bovine lactotransferrin significantly inhibits colon, esophagus, lung, bladder and liver cancers in rats [40]. Prostate cancer cells treated with LTF presented high apoptotic response,

growth arrest at G1 and reduced S phase, suggesting a role for specific cell cycle regulatory mechanisms in LTF-mediated cell growth inhibition [39].

CSRP1 (cysteine and glycine-rich protein 1) and TGM4 (human prostate-specific transglutaminase gene) are two other genes that become "hubs" [16] along tumoral development. The former belongs to the CSRP family, encoding a group of LIM domain proteins, which may be involved in regulatory processes which are important for development and cellular differentiation. Hirasawa and collaborators (2006) [41] suggest the use of CSRP as an important biomarker of hepatocellular carcinoma malignancy, because CSRP1 is inactivated in this model by aberrant methylation [41]. The latter, TGM4 was described as a candidate biomarker of region-specific epithelial identity in the prostate [42], being involved in the formation of stable protein-protein or protein-polyamide bounds [43].

Therefore, the literature supports the suggestion that these top seven genes (except for HAN11) may be considered as the most closely and informative prostate cancer biomarkers. Consequently, this suggests that the malignant transformation process in prostatic tissue is more correlated to functional connectivity changes in the gene dependence networks than differential gene expression itself.

Almost all of the 100 genes identified by PCA+MLDA are correlated to cancer, and, in many cases, to prostate cancer. Thus, TIMP3 and ADAMTS1 (Table 1) are genes classically correlated to invasion and the metastatic process, the main cancer attributes responsible for death.

Conclusion
In summary, our main goal using PCA+MLDA was not dimension reduction or verification of the classification accuracy, but to investigate the discriminative characteristics extracted from the whole microarray dataset and how one can interpret them, although this procedure may also be used for classification, yielding good results, as previously described.

We have demonstrated that changes in functional connectivity may underly the biological process which render some genes more informative to discriminate between normal and tumoral conditions. Using the proposed PCA+MLDA method in order to analyze the multivariate gene characteristic, it was possible to capture the changes in dependence networks which are related to cell transformation. Identification of seven genes (MYLK, KLK2, KLK3, HAN11, LTF, CSRP1, TGM4) which have their connectivity altered between normal/tumoral conditions may provide novel insights into specific targets against tumor progression.

Table 2: The seven "hub" genes.

Gene name	mean Z-value (normal)	Standard Error	mean Z-value (tumoral)	Standard Error
MYLK	1.138	0.107	2.464	0.177
KLK2	0.871	0.084	1.161	0.102
KLK3	1.070	0.100	0.953	0.073
HAN11	1.305	0.142	1.502	0.141
LTF	0.862	0.080	1.750	0.127
CSRPP1	1.254	0.139	1.601	0.157
TGM4	0.869	0.116	0.956	0.121

Mean Z-values obtained by Hoeffding's D measure and the corresponding standard errors.

Methods
Principal component analysis (PCA)
Principal component analysis is a dimension reduction technique used to reduce the high dimensional space (number of genes).

PCA is defined as linear transformations which maps the data to a new orthogonal coordinate system. These linear combinations are constructed so that the greatest variance by any projection lies on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

In other words, PCA summarizes the original features information by retaining characteristics of the dataset which most contribute to its variance.

For a gene expression data matrix $\mathbf{X}$ containing the genes in the columns and the observations in the rows (normalized to have zero mean and unit variance), the PCA transformation matrix ψ_{PCA} is given by

$$\psi_{\mathrm{PCA}} = eigenvectors(cov(\mathbf{X}^{\mathrm{T}})) \qquad (1)$$

where *cov* is the covariance matrix. In order to prevent losing any variance information, ψ_{PCA} is composed of all eigenvalues with non-zero eigenvectors. Here, PCA is used only to reduce computational and memory costs.

Maximum-entropy linear discriminant analysis (MLDA)
In gene expression data analysis, we usually have a large number of genes (features), but only a few number of observations, i.e., microarrays experiments.

A critical problem in applying conventional Linear Discriminant Analysis (LDA) to these types of data is the singularity and instability of the within-class scatter matrix calculated when the number of features approaches the number of available examples. In order to overcome this limitation, we applied the MLDA approach.

The MLDA method is concerned with the stabilization of pooled covariance matrix estimate $\mathbf{S_p}$. This covariance matrix $\mathbf{S_p}$ is constructed by selecting the largest dispersions regarding the $\mathbf{S_p}$ average eigenvalue. It is based on the maximum entropy covariance selection idea developed by Thomaz *et al* (2004) [18].

It is known that the estimated errors of small eigenvalues are greater than that of large eigenvalues. Therefore, Thomaz *et al.* (2007) [44] proposed to expand only the smaller and less reliable eigenvalues of $\mathbf{S_p}$, keeping most of the larger eigenvalues unchanged.

The algorithm may be described as follows:

1. Let the between-class scatter matrix $\mathbf{S_b}$ be defined as

$$\mathbf{S_b} = \sum_{i=1}^{g} n_i(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \qquad (2)$$

and the within-class scatter matrix $\mathbf{S_w}$ be defined as

$$\mathbf{S_w} = \sum_{i=1}^{g} (n_i - 1)\mathbf{S}_i = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)(\mathbf{x}_{i,j} - \bar{\mathbf{x}}_i)^T \qquad (3)$$

where $\mathbf{x}_{i,j}$ is the *m*-dimensional (*m*: number of genes) observation *j* from class $\prod_i$ (*i* = 1, 2, where 1 = normal and 2 = tumoral in our case) containing the gene expressions in the rows, n_i is the number of observations (microarrays) from class $\prod_i$, and *g* is the total number of classes (*g* = 2 in our case).

The vector $\bar{\mathbf{x}}_i$ is the unbiased sample mean and the matrix $\mathbf{S}_i$ is the sample covariance matrix of class $\prod_i$. The mean vector $\bar{\mathbf{x}}$ is calculated by

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{g} n_i\bar{\mathbf{x}}_i = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{n_i}\mathbf{x}_{i,j} \qquad (4)$$

where n is the total number of microarrays, i.e., $n = \sum_{j=1}^{g} n_j$.

2. Calculate the ψ eigenvectors and Λ eigenvalues of $\mathbf{S_p}$, where $\mathbf{S_p} = \mathbf{S_w}/[n - g]$.

3. Calculate $\bar{\lambda}$, i.e., the average eigenvalue

$$\bar{\lambda} = \frac{1}{m} \sum_{j=1}^{m} \lambda_j = \frac{trace(\mathbf{S_p})}{m} \qquad (5)$$

4. Construct the new matrix of eigenvalues based on the following largest dispersion criterion $\Lambda^* = diag\ [max(\lambda_i, \bar{\lambda}),..., max(\lambda_m, \bar{\lambda})]$

5. Construct the modified within-class scatter matrix $\mathbf{S_w^*}$

$$\mathbf{S_w^*} = \mathbf{S_p^*}(n - g) = (\psi \Lambda^* \psi^{T})(n - g) \qquad (6)$$

6. Finally, calculate the projection matrix ψ_{MLDA} which maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix (Fisher's criterion):

$$\psi_{MLDA} = eigenvector(\mathbf{S_w^{*-1}S_b}) \qquad (7)$$

The main advantage of MLDA is that it avoids both the singularity and instability of the within-class scatter matrix $\mathbf{S_w}$ when applied directly to gene expression data, which consists of a low number of observations and a high number of features.

The implemented R code is available in the Additional file 3.

Simulation
This simulation was designed in order to demonstrate that MLDA is capable to discriminate two different conditions and also to identify the intrinsic functional connectivity changes underlying the tumoral process. For this simulation, artificial gene expressions for 25,000 genes (features) were generated, based on the simulation illustrated in [21]. The 25,000 genes were divided in three sets A (250 genes), B (250 genes) and C (24,500 genes). For each gene, 30 observations representing "normal" condition and 30 observations representing "tumoral" conditions were generated. The model to investigate the situation where there are fuctional connectivity changes and there is no differences in gene expressions between conditions 1 and 2 were as follows:

$$\varphi^{(A)} = 1 + 0.3\varepsilon$$

$$\phi^{(B)} = \begin{cases} 1.3\phi^{(A)} + 0.3^{2} & \text{if condition 1.} \\ 0.9\phi^{(A)} + 0.3^{2} & \text{if condition 2.} \end{cases}$$

$$gene^{(A)} = \varphi^{A} + 0.3\theta_A$$

$$gene^{(B)} = \varphi^{B} + 0.5\theta_B$$

$$gene^{(C)} = \theta_C$$

where ε, , θ_A, θ_B and θ_C are independent Gaussian random variables with mean of zero and variance of one. This model considers two latent variables $\varphi^{(A)}$ and $\varphi^{(B)}$. Moreover, there is a functional relationship between A and B. Notice that there is no difference in means between A and B.

Differentially expressed genes
In order to identify putative differentially expressed genes, we have applied the non-parametric Wilcoxon test under a false discovery rate control (FDR) [45] of 5%. Wilcoxon procedure tests the median, therefore, it is more robust to outliers than the t-test (which tests the mean).

Relevance networks
Relevance networks [46] were constructed using the Hoeffding's D measure [20], a non-parametric association method (the R code is freely available in the Hmisc package at [47]), which is more robust to outliers than the Pearson's correlation. Pairwise correlations were measured and the false discovery rate (FDR) [45] was controlled to 1, 5 and 10%. "Hub" genes were determined by calculating the degree (the number of adjacent edges, i.e. functional connectivities) of each gene and selecting the highest ones.

Microarrays
We have analyzed the normal and tumoral prostate dataset publicly available at the Stanford MicroArray Database [48,19]. This dataset is composed of ~25,000 genes with 32 observations for normal state and 25 for tumoral condition.

Authors' contributions

AF has made substantial contributions to the conception, design and implementation of the study, and has also been responsible for drafting the manuscript. LRG has made substantial contributions to the biological interpretations, and has been responsible for drafting some parts of the manuscript. JRS has made substantial contributions to data analysis and applications of statistical concepts. RY, CET and MCS have discussed the results and critically revised the manuscript for important intellectual content.

SM has directed the work and has given the final approval of the version to be published.

Additional material

Additional file 1

Manual annotation. The manual annotation of the 100 genes described in Table 1.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-2-106-S1.doc]

Additional file 2

MLDA hyperplane weight. The MLDA hyperplane weight and the p-values (Wilcoxon test) for all the ~25,000 genes.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-2-106-S2.xls]

Additional file 3

R code. Implemented R code for MLDA.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-2-106-S3.r]

Acknowledgements

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ: **Cancer statistics.** *Cancer J Clin* 2008, **58:**71-96.
2. Parkin DM, Bray FI, Devesa SS: **Cancer burden in the year 2000. The global picture.** *Eur J Cancer* 2001, **37:**S4-S66.
3. Hsing AW, Tsao L, Devesa SS: **International trends and patterns in prostate cancer incidence and mortality.** *Int J Cancer* 2000, **85:**60-67.
4. Farkas A, Schneider D, Perrotti M, Cummings KB, Ward WS: **National trends in the epidemiology of prostate cancer, 1973 to 94: evidence for the effectiveness o prostate-specific antigen screening.** *Urology* 1998, **52:**444-448.
5. Han M, Partin AW, Piantadosi S, Epstein JI, Walsh PC: **Era specific biochemical recurrence-free survival following radical prostatectomy for clinically localized prostate cancer.** *J Urology* 2001, **166:**416-419.
6. Karan D, Lin MF, Hohansson SL, Batra SK: **Current status of the molecular genetics of human prostatic adenocarcinomas.** *Int J Cancer* 2003, **103(3):**285-293.
7. Reis EM, Nakaya H, Louro R, Canavez FC, Flatschart AVF, Almeida GT, Egidio CM, Paquola AC, Machado AA, Festa F, Yamamoto D, Alvarenga R, da Silva CC, Brito GC, Simon SD, Moreira-Filho CA, Leite KR, Camara-Lopes LH, Campos FS, Gimba E, Vignal GM, El-Dorry H, Sogayar MC, Barcinski MA, da Silva AM, Verjovski-Almeida S: **Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer.** *Oncogene* 2004, **23:**6684-6692.
8. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP, Lander ES, Loda M, Kantoff PW, Golub TR, Sellers WR: **Gene expression correlates of clinical prostate cancer behavior.** *Cancer cell* 2002, **1:**203-209.
9. Setlur SR, Royce TE, Sboner A, Mosquera JM, Demichelis F, Hofer MD, Mertz KD, Gerstein M, Rubin MA: **Integrative microarray analysis of pathways dysregulated in metastatic prostate cancer.** *Cancer Research* 2007, **67:**10296-10303.
10. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286:**531-537.
11. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V: **Molecular classification of cutaneous malignant melanoma by gene expression profiling.** *Nature* 2000, **406:**536-540.
12. Saeys Y, Inza I, Larrañaga P: **A review of feature selection techniques in bioinformatics.** *Bioinformatics* 2007, **23:**2507-2517.
13. Nguyen DV, Rocke DM: **Tumor classification by partial least squares using microarray gene expression data.** *Bioinformatics* 2002, **18:**39-50.
14. Hara T, Miyazaki H, Lee A, Tran CP, Reiter RE: **Androgen receptor and invasion in prostate cancer.** *Cancer Research* 2008, **68:**1128-1135.
15. Levchenko A: **Dynamical and integrative cell signaling challenges for the new biology.** *Biotechnol Bioeng* 2003, **30:**773-82.
16. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL: **The large-scale organization of metabolic networks.** *Nature* 2000, **407(6804):**651-654.
17. Krostka D, Spang R: **Finding disease specific alterations in the co-expression of genes.** *Bioinformatics* 2004, **20 Suppl 1:**i194-i199.
18. Thomaz CE, Gillies DF, Feitosa RQ: **A new covariance estimate for bayesian classifiers in biometric recognition.** *IEEE Transactions on circuits and systems for video technology* 2004, **14:**214-223.
19. Lapointe J, Li C, Higgins JP, Tijn M van de, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, Ekman P, DeMarzo AM, Tibshirani R, Botstein D, Brown PO, Brooks JD, Pollack JR: **Gene expression profiling identifies clinically relevant subtypes of prostate cancer.** *PNAS* 2004, **101:**811-816.
20. Hoeffding W: **A non-parametric test of independence.** *The Annals of Mathematical Statistics* 1948, **19:**546-557.
21. Sato JR, Mourão Miranda J, Amaro EJ, Morettin PA, Brammer MJ: **The impact of functional connectivity changes on support vector machines mapping of fMRI data.** *J Neurosci Methods* 2008, **172:**94-104.
22. Thomaz CE, Kitani EC, Gillies DF: **A Maximum Uncertainty LDA-based approach for limited sample size problems with application to face recognition.** *Journal of the Brazilian Computer Society* 2006, **12:**7-18.
23. Adelstein RS, Koonin EV, Altschul SF, Bork P: **Regulation of contractile proteins by phosphorylation.** *Journal of Clinical Investigation* 1983, **72:**1863-1866.
24. Gu LZ, Hu WY, Antic N, Mehta R, Turner JR, de Lanerolle P: **Inhibiting myosin light chain kinase retards the growth of mammary and prostate cancer cells.** *European Journal of Cancer* 2006, **42:**948-957.
25. Wilson AK, Gorgas G, Claypool WD, de Lanerolle P: **An increase or a decrease in myosin II phosphorylation inhibits macrophage motility.** *Journal of Cell Biology* 1991, **114:**277-283.
26. Klemke RL, Cai S, Giannini AL, Gallagher PJ, de Lanerolle P, Cheresh DA: **Regulation of cell motility by mitogen-activated protein kinase.** *Journal of Cell Biology* 1997, **137:**481-92.
27. Kaneko K, Satoh K, Masamune A, Satoh A, Shimosegawa T: **Myosin light chain kinase inhibitors can block invasion and adhesion of human pancreatic cancer cell lines.** *Pancreas* 2002, **24:**34-41.
28. Tohtong R, Phattarasakul K, Jiraviriyakul A, Sutthiphongchai T: **Dependence of metastatic cancer cell invasion on MLCK-catalyzed phosphorilation of myosin regulatory light chain.** *Prostate Cancer and Prostatic Diseases* 2003, **6:**212-216.
29. Stephan C, Jung K, Lein M, Diamandis EP: **PSA and other tissue kallikreins for prostate cancer detection.** *European Journal of Cancer* 2007, **43:**1918-1926.
30. Catalona WJ, Smith DS, Wolfert RL, Wang TJ, Rittenhouse HG, Ratliff TL, Nadler RB: **Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening.** *Journal of the American Medical Association* 1995, **274:**1214-1220.
31. Partin AW, Catalona WJ, Southwick PC, Subong EM, Gasior DW, Chan DW: **Analysis of percent free prostate-specific antigen**

(PSA) for prostate cancer detection: influence of total PSA, prostate volume, and age. *Urology* 1996, **48**:55-61.

32. Shaw JL, Diamandis EP: **Distribution of 15 human kallikreins in tissues and biological fluids.** *Clinical Chemistry* 2007, **53**:1423-1432.

33. Yousef GM, Diamandis EP: **The new human tissue kallikrein gene family: structure, function, and association to disease.** *Endocrine Reviews* 2001, **22**:184-204.

34. Kwiatkowski MK, Recker F, Piironen T, Pettersson K, Otto T, Wernli M, Tscholl R: **In prostatism patients the ratio of human glandular kallikrein to free PSA improves the discrimination between prostate cancer and benign hyperplasia within the diagnostic "gray zone" of total PSA 4 to 10 ng/mL.** *Urology* 1998, **52**:360-365.

35. Haese A, Becker C, Noldus J, Graefen M, Huland E, Huland H, Lilja H: **Human glandular kallikrein 2: a potential serum marker for predicting the organ confined versus non-organ confined growth of prostate cancer.** *Journal of Urology* 2000, **163**:1491-1497.

36. Haese A, Graefen M, Steuber T, Becker C, Noldus J, Erbersdobler A, Huland E, Huland H, Lilja H: **Total and Gleason grade 4/5 cancer volumes are major contributors of human kallikrein 2, whereas free prostate specific antigen is largely contributed by benign gland volume in serum from patients with prostate cancer or benign prostatic biopsies.** *Journal of Urology* 2003, **170**:2269-2273.

37. Brock JH: **The physiology of lactoferrin.** *Biochemistry and Cell Biology* 2000, **80**:1-6.

38. Teng CT: **Lactoferrin gene expression and regulation: An overview.** *Biochemistry and Cell Biology* 2002, **80**:7-16.

39. Shaheduzzaman S, Vishwanath A, Furusato B, Cullen J, Chen Y, Bañez L, Nau M, Ravindranath L, Kim KH, Mohammed A, Chen Y, Ehrich M, Srikantan V, Sesterhenn IA, McLeod DG, Vahey M, Petrovics G, Dobi A, Srivastava S: **Silencing of Lactotransferrin Expression by Methylation in Prostate Cancer Progression.** *Cancer Biology & Therapy* in press.

40. Tsuda H, Sekine K, Fujita K, Ligo M: **Cancer prevention by bovine lactoferrin and underlying mechanisms – A review of experimental and clinical studies.** *Biochemistry and Cell Biology* 2002, **80**:131-136.

41. Hirasawa Y, Arai M, Imazeki F, Tada M, Mikata R, Fukai K, Miyazaki M, Ochiai T, Saisho H, Yokosuka O: **Methylation status of genes upregulated by demethylating agent 5-aza-2'-deoxycytidine in hepatocellular carcionoma.** *Oncology* 2006, **71**:77-85.

42. Thielen JL, Volzing KG, Collier LS, Green LE, Largaespada DA, Marker PC: **Markers of prostate region-specific epithelial identity define anatomical locations in the mouse prostate that are molecularly similar to human prostate cancers.** *Differentiation* 2007, **75**:49-61.

43. Porta R, Esposito C, De Santis A, Fusco A, Iannone M, Metafora S: **Sperm maturation in human semen: role of transglutaminase-mediated reactions.** *Biology of Reproduction* 1986, **35**:965-970.

44. Thomaz CE, Duran FLS, Busatto GF, Gillies DF, Rueckert D: **Multivariate statistical differences of MRI samples of the human brain.** *Journal of mathematical imaging and vision* 2007, **29**:95-106.

45. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *Journal of the Royal Statistical Society Series B* 1995, **57**:289-300.

46. Butte A, Tamayo P, Slonim D, Golub TR, Kohane IS: **Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks.** *PNAS* 2000, **7**:12182-6.

47. **The R Project for Statistical Computing** [http://www.r-project.org/]

48. **Stanford MicroArray Database** [http://smd-www.stanford.edu/]

49. Lilja H, Ulmert D, Björk T, Becker C, Serio AM, Nilsson JA, Abrahamsson PA, Vickers AJ, Berglund G: **Long-term prediction of prostate cancer up to 25 years before diagnosis of prostate cancer using prostate kallikreins measured at age 44 to 50 years.** *Journal of Clinical Oncology* 2007, **25**:431-436.

50. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM: **Delineation of prognostic biomarkers in prostate cancer.** *Nature* 2001, **412**:822-826.

51. Miyasaka KY, Kida YS, Sato T, Minami M, Ogura T: **Csrp1 regulates dynamic cell movements of the mesendoderm and cardiac mesoderm through interactions with Dishevelled and Diversin.** *PNAS* 2007, **104**:11274-11279.

52. Untergasser G, Gander R, Lilg C, Lepperdinger G, Plas E, Berger P: **Profiling molecular targets of TGF-beta1 in prostate fibroblast-to-myofibroblast transdifferentiation.** *Mechanisms of ageing and development* 2005, **126**:59-69.

53. Li C, Kato M, Shiue L, Shively JE, Ares MJ, Lin RJ: **Cell type and culture condition-dependent alternative splicing in human breast cancer cells revealed by splicing-sensitive microarrays.** *Cancer Research* 2006, **66**:1990-1999.

54. Edwards S, Campbell C, Flohr P, Shipley J, Giddings I, Te-Poele R, Dodson A, Foster C, Clark J, Jhavar S, Kovacs G, Cooper CS: **Expression analysis onto microarrays of randomly selected cDNA clones highlights HOXB13 as a marker of human prostate cancer.** *British journal of cancer* 2005, **92**:376-381.

55. Bondar OP, Barnidge DR, Klee EW, Davis BJ, Klee GG: **LC-MS/MS quantification of Zn-alpha2 glycoprotein: a potential serum biomarker for prostate cancer.** *Clinical Chemistry* 2007, **53**:673-678.

56. Kayed H, Kleeff J, Kolb A, Ketterer K, Keleg S, Felix K, Giese T, Penzel R, Zentgraf H, Büchler MW, Korc M, Friess H: **FXYD3 is overexpressed in pancreatic ductal adenocarcinoma and influences pancreatic cancer cell growth.** *International journal of cancer* 2006, **118**:43-54.

57. Varga AE, Stourman NV, Zheng Q, Saña AF, Quan L, Li X, Sossey-Alaoui K, Bakin AV: **Silencing of the Tropomyosin-1 gene by DNA methylation alters tumor suppressor function of TGF-beta.** *Oncogene* 2005, **24(32)**:5034-5052.

58. Wittig R, Nessling M, Will RD, Mollenhauer J, Salowsky R, Münstermann E, Schick M, Helmbach H, Gschwendt B, Korn B, Kioschis P, Lichter P, Schadendorf D, Poustka A: **Candidate genes for cross-resistance against DNA-damaging drugs.** *Cancer Research* 2002, **62**:6698-6705.

59. Casey TM, Eneman J, Crocker A, White J, Tessitore J, Stanley M, Harlow S, Bunn JY, Weaver D, Muss H, Plaut K: **Cancer associated fibroblasts stimulated by transforming growth factor beta1 (TGF-beta1) increase invasion rate of tumor cells: a population study.** *Breast Cancer Res Treat* 2008, **110(1)**:39-49.

60. Martin PM, Aeder SE, Chrestensen CA, Sturgill TW, Hussaini IM: **Phorbol 12-myristate 13-acetate and serum synergize to promote rapamycin-insensitive cell proliferation via protein kinase C-eta.** *Oncogene* 2007, **26**:407-414.

61. Sharief FS, Mohler JL, Sharief Y, Li SS: **Expression of human prostatic acid phosphatase and prostate specific antigen genes in neoplastic and benign tissues.** *Biochem Mol Biol Int* 1994, **33(3)**:567-574.

62. Wei T, Geiser AG, Qian HR, Su C, Helvering LM, Kulkarini NH, Shou J, N'Cho M, Bryant HU, Onyia JE: **DNA microarray data integration by ortholog gene analysis reveals potential molecular mechanisms of estrogen-dependent growth of human uterine fibroids.** *BMC Women's Health* 2007, **7**:5.

63. Yu YP, Luo JH: **Myopodin-mediated suppression of prostate cancer cell migration involves interaction with zyxin.** *Cancer Research* 2006, **66**:7414-7419.

64. Vanaja DK, Ballman KV, Morlan BW, Cheville JC, Neumann RM, Lieber MM, Tindall DJ, Young CY: **PDLIM4 repression by hypermethylation as a potential biomarker for prostate cancer.** *Clinical cancer research* 2006, **12**:1128-1136.

65. Eeles RA, Z KJ, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL, Collaborators UGPCS, of Urological Surgeons' Section of Oncology BA, Collaborators UPS, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE, Easton DF: **Multiple newly identified loci associated with prostate cancer susceptibility.** *Nature Genetics* 2008, **40**:316-321.

66. Hisataki T, Itoh N, Suzuki K, Takahashi A, Masumori N, Tohse N, Ohmori Y, Yamada S, Tsukamoto T: **Modulation of phenotype of human prostatic stromal cells by transforming growth factor-betas.** *Prostate* 2004, **58**:174-182.

67. Li C, Hibino M, Komatsu H, Sakuma H, Sakakura T, Ueda R, Eimoto T, Inagaki H: **Primary mucosa-associated lymphoid tissue lymphoma of the prostate: Tumor relapse 7 years after local therapy.** *Pathology International* 2008, **58**:191-195.
68. Gianni-Barrera R, Gariboldi M, De Cecco L, Manenti G, Dragani TA: **Specific gene expression profiles distinguish among functional allelic variants of the mouse Pthlh gene in transfected human cancer cells.** *Oncogene* 2006, **25**:4501-4504.
69. Pogue-Geile KL, Chen R, Bronner MP, Crnogorac-Jurcevic T, Moyes KW, Dowen S, Otey CA, Crispin DA, George RD, Whitcomb DC, Brentnall TA: **Palladin mutation causes familial pancreatic cancer and suggests a new cancer mechanism.** *PLoS Medicine* 2006, **3**:e516.
70. Yamasaki M, Nomura T, Sato F, Mimata H: **Metallothionein is upregulated under hypoxia and promotes the survival of human prostate cancer cells.** *Oncology Reports* 2007, **18**:1145-1153.
71. Shiseki M, Nagashima M, Pedeux RM, Kitahama-Shiseki M, Miura K, Okamura S, Onogi H, Higashimoto Y, Appella E, Yokota J, Harris CC: **p29ING4 and p28ING5 bind to p53 and p300, and enhance p53 activity.** *Cancer Research* 2003, **63**:2373-2378.
72. Demeo DL, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, Rennard SI, Sandhaus RA, Stocks JM, Stoller JK, Strange C, Turino G, Silverman EK: **IL10 polymorphisms are associated with air flow obstruction in severe alpha1-antitrypsin deficiency.** *American Journal of Respiratory Cell and Molecular Biology* 2008, **38**:114-120.
73. Rumpold H, Heinrich E, Untergasser G, Hermann M, Pfister G, Plas E, Berger P: **Neuroendocrine differentiation of human prostatic primary epithelial cells in vitro.** *Prostate* 2002, **53**:101-108.
74. Lü B, Xu J, Zhu Y, Zhang H, Lai M: **Systemic analysis of the differential gene expression profile in a colonic adenoma-normal SSH library.** *Clinica Chimica Acta* 2007, **378**:42-47.
75. Johansson M, McKay JD, Stattin P, Canzian F, Boillot C, Wiklund F, Adami HO, Bälter K, Grönberg H, Kaaks R: **Comprehensive evaluation of genetic variation in the IGF1 gene and risk of prostate cancer.** *International Journal of Cancer* 2007, **120**:539-542.
76. Burger MJ, Tebay MA, Keith PA, Samaratunga HM, Clements J, Lavin MF, Gardiner RA: **Expression analysis of delta-catenin and prostate-specific membrane antigen: their potential as diagnostic markers for prostate cancer.** *International Journal of Cancer* 2002, **100**:228-237.
77. Pilarsky CP, Schmidt U, Eissrich C, Stade J, Froschermaier SE, Haase M, Faller G, Kirchner TW, Wirth MP: **Expression of the extracellular matrix signaling molecule Cyr61 is downregulated in prostate cancer.** *Prostate* 1998, **36**:85-91.
78. Shen Y, Jia Z, Nagele RG, Ichikawa H, Goldberg GS: **SRC uses Cas to suppress Fhl1 in order to promote nonanchored growth and migration of tumor cells.** *Cancer Research* 2006, **66**:1543-1552.
79. Berteaux N, Lottin S, Adriaenssens E, van Coppenolle F, Leroy X, Coll J, Dugimont T, Curgy JJ: **Hormonal regulation of H19 gene expression in prostate epithelial cells.** *Journal of Endocrinology* 2004, **183**:69-78.
80. Sanson M, Marineau C, Desmaze C, Lutchman M, Ruttledge M, Baron C, Narod S, Delattre O, Lenoir G, Thomas G, *et al.*: **Germline deletion in a neurofibromatosis type 2 kindred inactivates the NF₂ gene and a candidate meningioma locus.** *Human Molecular Genetics* 1993, **2**:1215-1220.
81. Rogers MS, Christensen KA, Birsner AE, Short SM, Wigelsworth DJ, Collier RJ, D'Amato RJ: **Mutant anthrax toxin B moiety (protective antigen) inhibits angiogenesis and tumor growth.** *Cancer Research* 2007, **67**:9980-9985.
82. Umeda D, Tachibana H, Yamada K: **Epigallocatechin-3-O-gallate disrupts stress fibers and the contractile ring by reducing myosin regulatory light chain phosphorylation mediated through the target molecule 67 kDa laminin receptor.** *Biochemical and Biophysical Research Communications* 2005, **333**:628-635.
83. Kanayama H, Tanaka K, Aki M, Kagawa S, Miyaji H, Satoh M, Okada F, Sato S, Shimbara N, Ichihara A: **Changes in expressions of proteasome and ubiquitin genes in human renal cancer cells.** *Cancer Research* 1991, **51**:6677-6685.
84. Hooi CF, Blancher C, Qiu W, Revet IM, Williams LH, Ciavarella ML, Anderson RL, Thompson EW, Connor A, Phillips WA, Campbell IG: **ST7-mediated suppression of tumorigenicity of prostate cancer cells is characterized by remodeling of the extracellular matrix.** *Oncogene* 2006, **25**:3924-3933.
85. Yemelyanov A, Czwornog J, Chebotaev D, Karseladze A, Kulevitch E, Yang X, Budunova I: **Tumor suppressor activity of glucocorticoid receptor in the prostate.** *Oncogene* 2007, **26**:1885-1896.
86. Kato K, Horiuchi S, Takahashi A, Ueoka Y, Arima T, Matsuda T, Kato H, Nishida J, Ji J, Nakabeppu Y, Wake N: **Contribution of estrogen receptor alpha to oncogenic K-Ras-mediated NIH3T3 cell transformation and its implication for escape from senescence by modulating the p53 pathway.** *Journal of Biological Chemistry* 2002, **277**:11217-11224.
87. Kaiser S, Park YK, Franklin JL, Halberg RB, Yu M, Jessen WJ, Freudenberg J, Chen X, Haigis K, Jegga AG, Kong S, Sakthivel B, Xu H, Reichling T, Azhar M, Boivin GP, Roberts RB, Bissahoyo AC, Gonzales F, Bloom GC, Eschrich S, Carter SL, Aronow JE, Kleimeyer J, Kleimeyer M, Ramaswamy V, Settle SH, Boone B, Levy S, Graff JM, Doetschman T, Groden J, Dove WF, Threadgill DW, Yeatman TJ, Coffey RJJ, Aronow BJ: **Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer.** *Genome Biology* 2007, **8**:R131.
88. Zhang Y, Ma K, Sadana P, Chowdhury F, Gaillard S, Wang F, McDonnell DP, Unterman TG, Elam MB, Park EA: **Estrogen-related receptors stimulate pyruvate dehydrogenase kinase isoform 4 gene expression.** *Journal of Biological Chemistry* 2006, **281**:39897-39906.
89. Ashida S, Nakagawa H, Katagiri T, Furihata M, Iizumi M, Anazawa Y, Tsunoda T, Takata R, Kasahara K, Miki T, Fujioka T, Shuin T, Nakamura Y: **Molecular features of the transition from prostatic intraepithelial neoplasia (PIN) to prostate cancer: genome-wide gene-expression profiles of prostate cancers and PINs.** *Cancer Research* 2004, **64**:5963-5972.
90. Segawa T, Nau ME, Xu LL, Chilukuri RN, Makarem M, Zhang W, Petrovics G, Sesterhenn IA, McLeod DG, Moul JW, Vahey M, Srivastava S: **Androgen-induced expression of endoplasmic reticulum (ER) stress response genes in prostate cancer cells.** *Oncogene* 2002, **21**:8749-8758.
91. Hodgson G, Hager JH, Volik S, Hariono S, Wernick M, Moore D, Nowak N, Albertson DG, Pinkel D, Collins C, Hanahan D, Gray JW: **Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas.** *Nature Genetics* 2001, **29**:459-464.
92. Chan CT, Paulmurugan R, Gheysens OS, Kim J, Chiosis G, Gambhir SS: **Molecular imaging of the efficacy of heat shock protein 90 inhibitors in living subjects.** *Cancer Research* 2008, **68**:216-226.
93. Ricci G, De Maria F, Antonini G, Turella P, Bullo A, Stella L, Filomeni G, Federici G, Caccuri AM: **7-Nitro-2,1,3-benzoxadiazole derivatives, a new class of suicide inhibitors for glutathione S-transferases. Mechanism of action of potential anticancer drugs.** *Journal of Biological Chemistry* 2005, **280**:26397-26405.
94. Yamashita S, Wakazono K, Nomoto T, Tsujino Y, Kuramoto T, Ushijima T: **Expression quantitative trait loci analysis of 13 genes in the rat prostate.** *Genetics* 2005, **171**:1231-1238.
95. Attard G, Clark J, Ambroisine L, Fisher G, Kovacs G, Flohr P, Berney D, Foster CS, Fletcher A, Gerald WL, Moller H, Reuter V, De Bono JS, Scardino P, Cuzick J, Cooper CS: **Duplication of the fusion of TMPRSS2 to ERG sequences identifies fatal human prostate cancer.** *Oncogene* 2008, **27**:253-263.
96. Yang F, Tuxhorn JA, Ressler SJ, McAlhany SJ, Dang TD, Rowley DR: **Stromal expression of connective tissue growth factor promotes angiogenesis and prostate cancer tumorigenesis.** *Cancer Research* 2005, **65**:8887-8895.
97. Dong Y, Zhang H, Gao AC, Marshall JR, Ip C: **Androgen receptor signaling intensity is a key factor in determining the sensitivity of prostate cancer cells to selenium inhibition of growth and cancer-specific biomarkers.** *Molecular Cancer Therapeutics* 2005, **4**:1047-1055.
98. Lee S, Bang S, Song K, Lee I: **Differential expression in normal-adenoma-carcinoma sequence suggests complex molecular carcinogenesis in colon.** *Oncology Reports* 2006, **16**:747-754.
99. Yegnasubramanian S, Kowalski J, Gonzalgo ML, Zahurak M, Piantadosi S, Walsh PC, Bova GS, De Marzo AM, Isaacs WB, Nelson WG: **Hypermethylation of CpG islands in primary and metastatic human prostate cancer.** *Cancer Research* 2004, **64**:1975-1986.
100. Leiblich A, Cross SS, Catto JW, Phillips JT, Leung HY, Hamdy FC, Rehman I: **Lactate dehydrogenase-B is silenced by promoter**

hypermethylation in human prostate cancer.** *Oncogene* 2006, **25:**2953-2960.

101. Wiese AH, Auer J, Lassmann S, Nährig J, Rosenberg R, Höfler H, Rüger R, Werner M: **Identification of gene signatures for invasive colorectal tumor cells.** *Cancer Detection and Prevention* 2007, **31:**282-295.

102. Karam JA, Lotan Y, Roehrborn CG, Ashfaq R, Karakiewicz PI, Shariat SF: **Caveolin-1 overexpression is associated with aggressive prostate cancer recurrence.** *Prostate* 2007, **67:**614-622.

103. Gober MD, Smith CC, Ueda K, Toretsky JA, Aurelian L: **Forced expression of the H11 heat shock protein can be regulated by DNA methylation and trigger apoptosis in human cells.** *Journal of Biological Chemistry* 2003, **278:**37600-37609.

104. Eeckhoute J, Carroll JS, Geistlinger TR, Torres-Arzayus MI, Brown M: **A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer.** *Genes & Development* 2006, **20:**2513-2526.

105. Crombez KR, Vanoirbeek EM, Ven WJ Van de, Petit MM: **Transactivation functions of the tumor-specific HMGA2/LPP fusion protein are augmented by wild-type HMGA2.** *Molecular Cancer Research* 2005, **3:**63-70.

106. Yun MY, Kim SB, Park S, Han CJ, Han YH, Yoon SH, Kim SH, Kim CM, Choi DW, Cho MH, Park GH, Lee KH: **Mutation analysis of p31comet gene, a negative regulator of Mad2, in human hepatocellular carcinoma.** *Experimental & molecular medicine* 2007, **39:**508-513.

107. Gustavsson H, Jennbacken K, Welén K, Damber JE: **Altered expression of genes regulating angiogenesis in experimental androgen-independent prostate cancer.** *Prostate* 2008, **68:**161-170.

108. Ghosh PM, Ghosh-Choudhury N, Moyer ML, Mott GE, Thomas CA, Foster BA, Greenberg NM, Kreisberg JI: **Role of RhoA activation in the growth and morphology of a murine prostate tumor cell line.** *Oncogene* 1999, **18:**4120-4130.

109. Xu W, Ngo L, Perez G, Dokmanovic M, Marks PA: **Intrinsic apoptotic and thioredoxin pathways in human prostate cancer cell response to histone deacetylase inhibitor.** *PNAS* 2006, **103:**15540-15545.

110. Coe BP, Henderson LJ, Garnis C, Tsao MS, Gazdar AF, Minna J, Lam S, Macaulay C, Lam WL: **High-resolution chromosome arm 5p array CGH analysis of small cell lung carcinoma cell lines.** *Genes Chromosomes & Cancer* 2005, **42:**308-313.

111. Wlodek L, Wróbel M, Czubak J: **Transamination and transsulphuration of L-cysteine in Ehrlich ascites tumor cells and mouse liver. The nonenzymatic reaction of L-cysteine with pyruvate.** *International journal of biochemistry* 1993, **25:**107-112.

112. Shadeo A, Chari R, Lonergan KM, Pusic A, Miller D, Ehlen T, van Niekerk D, Matisic J, Richards-Kortum R, Follen M, Guillaud M, Lam WL, Macaulay C: **Up regulation in gene expression of chromatin remodelling factors in cervical intraepithelial neoplasia.** *BMC Genomics* 2008, **9:**64.

113. Zhang XA, Lane WS, Charrin S, Rubinstein E, Liu L: **EWI2/PGRL associates with the metastasis suppressor KAI1/CD82 and inhibits the migration of prostate cancer cells.** *Cancer Research* 2003, **63:**2665-2674.

114. Pertin M, Ji RR, Berta T, Powell AJ, Karchewski L, Tate SN, Isom LL, Woolf CJ, Gilliard N, Spahn DR, Decosterd I: **Upregulation of the voltage-gated sodium channel beta2 subunit in neuropathic pain models: characterization of expression in injured and non-injured primary sensory neurons.** *Journal of Neuroscience* 2005, **25:**10970-10980.

115. Nelson PS, Plymate SR, Wang K, True LD, Ware JL, Gan L, Liu AY, Hood L: **Hevin, an antiadhesive extracellular matrix protein, is down-regulated in metastatic prostate adenocarcinoma.** *Cancer Research* 1998, **58:**232-236.

116. Yao R, Rich SA, Schneider E: **Validation of sixteen leukemia and lymphoma cell lines as controls for molecular gene rearrangement assays.** *Clinical Chemistry* 2002, **48:**1344-1351.

117. Rust R, Visser L, Leij J van der, Harms G, Blokzijl T, Deloulme JC, Vlies P van der, Kamps W, Kok K, Lim M, Poppema S, Berg A van den: **High expression of calcium-binding proteins, S100A10, S100A11 and CALM2 in anaplastic large cell lymphoma.** *British Journal of Haematology* 2005, **131:**596-608.

118. Sampson N, Untergasser G, Lilg C, Tadic L, Plas E, Berger P: **GAGEC1, a cancer/testis associated antigen family member, is a target of TGF-beta1 in age-related prostatic disease.** *Mechanisms of Ageing and Development* 2007, **128:**64-66.

119. Toutenhoofd SL, Foletti D, Wicki R, Rhyner JA, Garcia F, Tolon R, Strehler EE: **Characterization of the human CALM2 calmodulin gene and comparison of the transcriptional activity of CALM1, CALM2 and CALM3.** *Cell Calcium* 1998, **23:**323-338.

120. Chaib H, Cockrell EK, Rubin MA, Macoska JA: **Profiling and verification of gene expression patterns in normal and malignant human prostate tissues by cDNA microarray analysis.** *Neoplasia* 2001, **3:**43-52.

121. Wang Z, Hao Y, Lowe AW: **The adenocarcinoma-associated antigen, AGR2, promotes tumor growth, cell migration, and cellular transformation.** *Cancer Research* 2008, **68:**492-497.