

RESEARCH ARTICLE

Open Access

# Key regulators in prostate cancer identified by co-expression module analysis

Junfeng Jiang<sup>1,2†</sup>, Peilin Jia<sup>2,3†</sup>, Zhongming Zhao<sup>2,3,4,5\*</sup> and Bairong Shen<sup>1\*</sup>

## Abstract

**Background:** Prostate cancer (PrCa) is the most commonly diagnosed cancer in men in the world. Despite the fact that a large number of its genes have been investigated, its etiology remains poorly understood. Furthermore, most PrCa candidate genes have not been rigorously replicated, and the methods by which they biologically function in PrCa remain largely unknown.

**Results:** Aiming to identify key players in the complex prostate cancer system, we reconstructed PrCa co-expressed modules within functional gene sets defined by the Gene Ontology (GO) annotation (biological process, GO\_BP). We primarily identified 118 GO\_BP terms that were well-preserved between two independent gene expression datasets and a consequent 55 conserved co-expression modules within them. Five modules were then found to be significantly enriched with PrCa candidate genes collected from expression Quantitative Trait Loci (eQTL), somatic copy number alteration (SCNA), somatic mutation data, or prognostic analyses. Specifically, two transcription factors (TFs) (NFAT and SP1) and three microRNAs (hsa-miR-19a, hsa-miR-15a, and hsa-miR-200b) regulating these five candidate modules were found to be critical to the development of PrCa.

**Conclusions:** Collectively, our results indicated that genes with similar functions may play important roles in disease through co-expression, and modules with different functions could be regulated by similar genetic components, such as TFs and microRNAs, in a synergistic manner.

**Keywords:** Prostate cancer, Co-expression, Gene Ontology, Module, Transcription factor, MicroRNA

## Background

Prostate cancer (PrCa) is the sixth leading cause of cancer-related deaths of men in the world [1] and the second leading cause in the United States [2]. Due to the high risk of metastasis, it has become fundamentally important to uncover the underlying mechanisms of PrCa. Factors such as age, ethnicity, family history, heritability, diet, lifestyle, environment, and androgens have long been recognized as contributors to the risk of PrCa [3-5]. As demonstrated by twin studies, PrCa's genetic component is estimated to be as high as 42-57% [6,7].

To elucidate the underlying pathophysiology and molecular mechanisms of PrCa, numerous genetic and genomic studies have been conducted, including gene expression

profiling [8-12], expression Quantitative Trait Loci (eQTL) mapping [13-15], somatic copy number alteration (SCNA) identification [16], gene mutation detection [17], prognostic gene discovery [18], microRNA (miRNA) expression profiling [14,19], and transcription factor (TF) enrichment [20], among others. The gene expressions profiled by microarray technology have been a major strategy to detect mRNA abundance. Traditional, single, and gene-based strategies have been widely applied for gene expression analyses, but they suffered from limitations, such as multiple testing burdens [21], small numbers of differentially expressed genes [22], lack of interactions/regulations among genes [23], or low replication rates [24].

Alternatively, gene co-expression module analysis attempts to study combined effects by identifying groups of genes that are coordinately expressed [21,25-27]. For instance, Horvath and colleagues have developed a widely used algorithm, the Weighed Gene Co-expression Network Analysis (WGCNA) [28], to search for co-expression modules. The R package WGCNA implements a suite of tools

\* Correspondence: zhongming.zhao@vanderbilt.edu; bairong.shen@suda.edu.cn

†Equal contributors

<sup>2</sup>Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN 37203, USA

<sup>1</sup>Center for Systems Biology, Soochow University, Jiangsu, China

Full list of author information is available at the end of the article

for network construction, module detection, module significance examination, module preservation computation, and hub gene query, among many others [29-31].

To our knowledge, no co-expression module has been constructed for the identification of key regulators in PrCa until now. Moreover, traditional co-expression studies that start from whole human genes on a chip (or top differentially expressed genes) often result in very large modules (e.g., >1000 genes). Although functional assessments, such as GO enrichment, the functional gene/SNP enrichment test, and hub gene analysis could help to explore the functions of modules, such interpretation typically results in noisy results (e.g., a lot list of GO terms or genes). In this study, we developed a framework for gene co-expression module construction in PrCa using the WGCNA approach and augmented by Gene Ontology [32] biological process (GO\_BP) annotations. We argued that although GO\_BP terms are broadly defined for each functional group, there may be subsets of genes in a biological process (GO\_BP term) that are coordinately expressed, e.g., in a disease-associated fashion. For example, different co-expression modules may underlie different diseases, although they all execute the same biological functions as defined by GO\_BP terms. To this end, we developed a systematic framework (Figure 1) to search for co-expression modules within each GO\_BP term and demonstrated it in PrCa. We primarily found 118 preserved GO\_BP terms in two PrCa datasets and constructed 55 co-expression modules. We then assessed these modules for their enrichment of PrCa candidate genes collected by eQTLs, SCNA, somatic mutation data, or prognostic studies using the hypergeometric test. As a result, 5 modules were identified as significantly associated with PrCa, and several TFs and miRNAs were found to be potential key regulators of these candidate modules.

## Methods

### Processing and analysis of microarray gene expression data

Two microarray gene expression datasets were downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). To ensure the data quality, we searched for studies that have a well-defined phenotypic description for PrCa, have a sample size around 100 or more, and have preferably been measured using similar platforms in order to obtain a high proportion of overlapping genes. We included both case and control samples to determine disease-specific signals, as similarly done in Chen *et al.* [21]. Two datasets that fulfilled these criteria were downloaded for our further analyses (Table 1). The first dataset (GEO accession ID: GSE17951 [33]), which was used as the training dataset in our work, was generated using the Affymetrix U133Plus2 array on prostate tissue samples

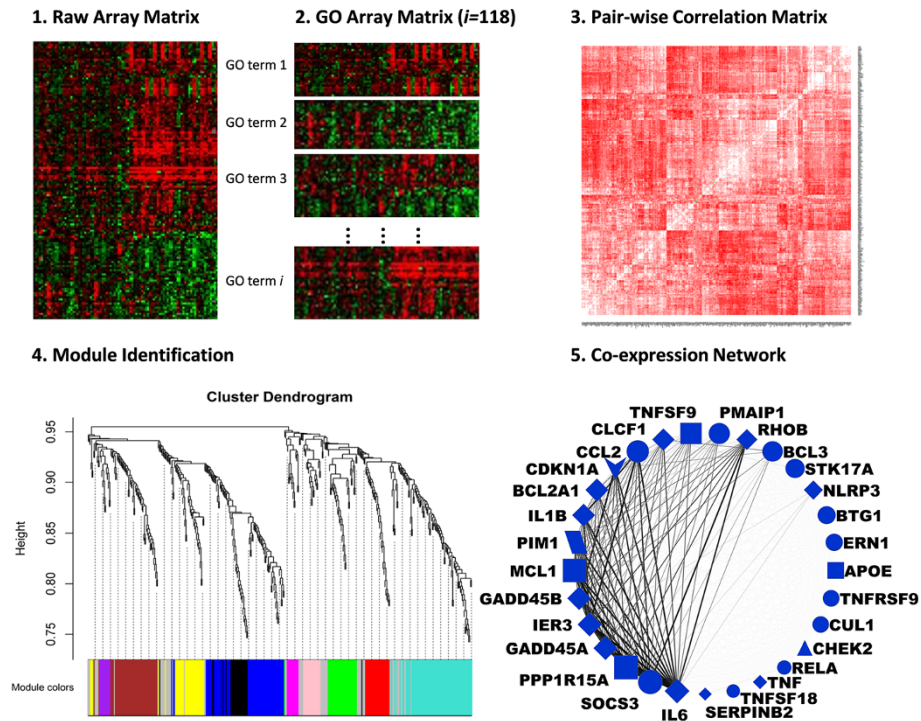
from 109 cancer patients and 45 control samples [33] (Table 1). The second dataset (GEO accession ID: GSE6956 [10]), which was used as the testing dataset, was collected using the Affymetrix Human Genome U133A 2.0 array for 69 fresh-frozen prostate tumors and 20 control samples from surrounding normal prostate tissue [10]. For each dataset, we performed the following quality control steps. First, outlier samples were detected and removed. We calculated the inter-array correlation (IAC) based on Pearson's correlation coefficient for tumor and control samples, respectively. We excluded the samples with low mean IAC and/or those that emerged as a clear outlier by the hierarchical clustering approach [34]. Second, samples were quantile-normalized after  $\log_2$  transformation [35]. Probes with missing expression values in more than 30% samples were removed from further analyses. For each gene, we selected the probe with the highest intensity to represent the expression level of the gene. This resulted in 21,049 genes involved in 82 tumor and 40 control samples in the training dataset and 13,211 genes in 60 tumor and 19 normal samples in the testing dataset, with 13,211 genes shared by the two datasets.

### Highly-preserved GO\_BP terms

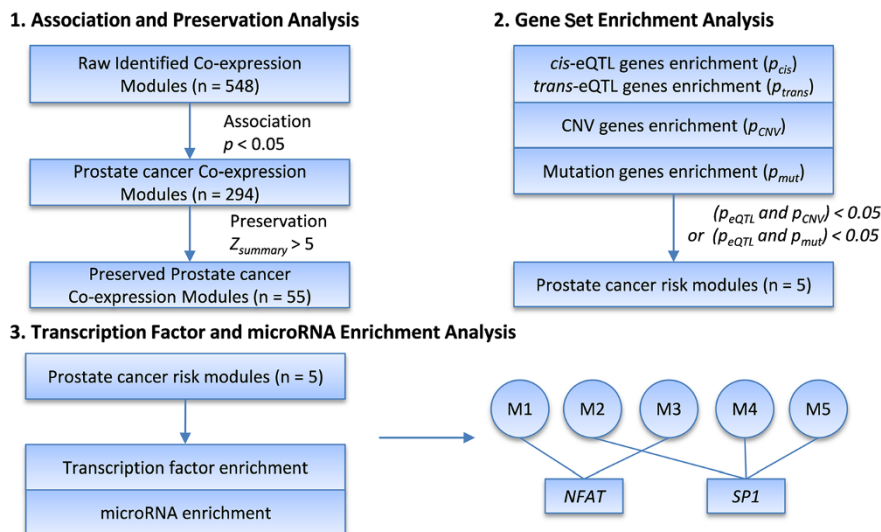
The GO database provides three annotation categories (domains): Molecular Function (MF), Biological Process (BP), and Cellular Component (CC). In this study, we focused on the BP category, based on the notion that genes that participate in the same biological processes tend to be expressed coordinately [36]. We downloaded the GO\_BP gene sets from the Molecular Signatures Database (MSigDB) [37], including 825 GO\_BP terms in the C5 category. To avoid too broadly or too narrowly defined GO\_BP terms, we only kept the terms with 50 to 500 measured genes. As a result, 226 GO\_BP terms were eligible to build the gene expression matrices for the following analysis.

Before the construction of the modules, we first evaluated the importance of each term associated with PrCa by calculating a preservation score. The preservation score aims to assess the level of preservation between the training and testing datasets for a gene set (i.e., a GO\_BP term) and is typically measured based on both density and connectivity patterns among the genes [29]. The parameter  $Z_{\text{summary}}$  [29] implemented in WGCNA [28] was employed to compute the preservation score. In general, a value of  $Z_{\text{summary}} < 2$  indicates no evidence of preservation between the training and the testing datasets,  $2 < Z_{\text{summary}} < 10$  implies weak to moderate evidence, and  $Z_{\text{summary}} > 10$  indicates strong evidence. At this stage, we aim to perform pre-selection of GO\_BP terms that are suitable for the following co-expression clustering analysis. To this end, we chose a moderate threshold of preservation,  $Z_{\text{summary}} = 5$  [29], to select

### A Co-expression Network Construction



### B Module Significance Enrichment



**Figure 1 Overview of workflow.** (A) GO-based gene co-expression network construction. It has five steps: (1) Raw microarray data processing and analysis. (2) GO term expression data matrix-building. (3) Pairwise correlation analysis of genes in GO term across different samples. (4) Expression profile clustering and module identification using WGCNA. (5) Visualization of co-expression modules by Cytoscape. (B) Framework of module significance analyses. The details are provided in Methods.

GO\_BP terms and denote them as preserved GO\_BP terms in both training and testing datasets.

#### Application of WGCNA in PrCa

Given that genes within a GO\_BP term are well-defined with similar biological functions, we asked whether they

tend to co-express in a specific disease. Thus, we performed gene co-expression analysis for each preserved GO\_BP term instead of all genes on the chip. We took the gene expression matrix for each GO\_BP term as the input and applied WGCNA to detect co-expression modules. Gene co-expression correlation was measured

**Table 1 Summarization of PrCa microarray gene expression datasets used in the study**

GEO accession ID	Type	Before QC <sup>a</sup>		After QC		# Genes
		# Tumors	# Controls	# Tumors	# Controls	
GSE17951	Training	109	45	82	40	21049
GSE6956	Testing	69	20	60	19	13211

<sup>a</sup>QC: quality control.

by Pearson's correlation coefficients. In this step, we built a pairwise co-expression matrix. This GO\_BP term-based matrix was then utilized to construct an initial gene co-expression network by the *blockwiseModules* function in WGCNA. Notably, the initial co-expression network constructed and based only on Pearson's correlation coefficients was not always a scale-free network. Rather, to obtain a scale-free network, a weighted adjacency matrix needs to be constructed using a selected power determined through a soft-thresholding approach in WGCNA.

Co-expression modules were then defined by a robust dynamic hierarchical tree cut algorithm using the measurement of dissimilarity (i.e., 1-topological overlap matrix) [26,38]. To ensure a suitable number of genes for next-step analysis, we set the minimum module size as 10. The adjacent modules were merged based on the parameter of cutHeight, i.e., modules with a minimum cutHeight at 0.25 were merged. Principle component analysis (PCA) of the expression matrix for each module was then performed. We denoted the first principal component (PC) as the module eigengene and used it to represent the overall expression profile of the module [39]. For each gene, we computed a module membership (kME) based on the correlation between the gene expression and the module eigengene. Those genes with a lower membership ( $kME \leq 0.3$ ) were removed from the module and assigned to the grey module.

To validate whether the identified modules were associated with PrCa, we conducted a two-step evaluation procedure. First, for each module, we adopted the module eigengene to assess its trait association (denoted as  $p_{cor}$ ) based on Pearson's correlation coefficients. We used the false positive rate (FDR) for multiple testing correction [40]. Second, for PrCa-associated modules, we further evaluated the module preservation in the testing dataset. Since the size of our identified module was generally less than 100, we defined a module to be preserved if it has a  $Z_{summary}(\text{module}) > 5$ .

#### Enrichment test

Four types of large-scale, PrCa-associated genetic/genomic data, eQTL genes, recurrent SCNAs, somatic mutations, and prognostic genes were collected for the enrichment test of the identified modules. The eQTL genes were collected from the online eQTL database, SeeQTL ([http://www.bios.unc.edu/research/genomic\\_software/seeQTL/](http://www.bios.unc.edu/research/genomic_software/seeQTL/)), which re-

analyzed nine independent HapMap studies in lymphoblastoid cell lines (LCLs); performed a consensus meta-analysis with comprehensive quality control, population stratification control, and FDR control; and provided the q-value as the significance measurement [41-47]. We retrieved a total of 8652 genes regulated by *cis*-eQTL (7071 genes) or *trans*-eQTL (4140 genes) from the SeeQTL database. Here, *cis*-eQTL represents the cases where the regulated genes are located within 1 Mb of the regulatory SNP, while *trans*-eQTL indicates associations for more distant eQTL. We denoted them as *cis*-eQTL and *trans*-eQTL gene sets, respectively.

A list of PrCa susceptibility genes located in SCNA regions was downloaded from the cBio data portal [48]. These SCNA regions were obtained using GISTIC2 (q-value <0.1) [49], based on The Cancer Genome Atlas (TCGA) prostate adenocarcinoma data (<https://tcga-data.nci.nih.gov/tcga/>). In sum, we retrieved 567 unique genes and denoted them as the SCNA gene set.

Genes with somatic mutations were collected from two sources. First, we manually collected 47 mutated genes from the Human Prostate Gene Database (PGDB) [17]. Second, 107 significantly mutated genes identified in the TCGA prostate adenocarcinoma (<https://tcga-data.nci.nih.gov/tcga/>) samples were retrieved from the cBio portal. As a result, we obtained 149 unique genes and denoted them as the mutant gene set.

We downloaded the RNASeqV2 and clinical data for prostate adenocarcinoma from the TCGA portal (<https://tcga-data.nci.nih.gov/tcga/>). The Univariate Cox model was applied to define the prognostic genes [18]. FDR was applied for multiple testing correction of the raw Wald  $p$  values. Finally, we obtained 737 genes with  $FDR < 0.05$ .

Gene set enrichment analysis for PrCa-associated modules was performed using the hypergeometric test. Multiple testing correction was controlled by the FDR method. The module was taken as a candidate if it significantly enriched with any of the two gene sets among eQTL, SCNA, mutation, and prognostic genes with  $FDR < 0.05$ . We further performed enrichment tests of the candidate module genes with TF and miRNA using the online tool WebGestalt [50]. The resultant  $p$  values were corrected for multiple testing using the FDR method. We set the significance level for FDR at 0.01 and the minimum number of genes for a category at two. For simplification, the top 5 enriched TFs or miRNAs were collected for further analyses.

## Results

### The identified GO\_BP-based co-expression modules

GO\_BP terms categorize genes that function in the same or similar biological processes. Hence, genes in the same GO\_BP term could be expected to have coordinated expression patterns. In our study, among the 226 GO\_BP terms that fulfilled our query criteria (size between 50 and 500), 118 had a preservation score of  $Z_{\text{summary}}(\text{GO\_BP}) > 5$  and were considered well preserved between the training and the testing datasets. For each of these 118 GO\_BP terms, we built a weighted co-expression network using the entire training dataset of 82 prostate tumor samples and 40 control samples (see Methods). Highly co-expressed genes within a term were then clustered into modules, each labeled with a specific color (Figure 1A4). This resulted in 548 modules in total. To examine their association with prostate cancer, we calculated the correlation between the module expression profiles (represented by the module eigengene) and the PrCa disease status (represented by a vector of 1 for case and 0 for control). With FDR  $< 0.05$ , we identified 294 of the 548 modules that showed statistically significant association with PrCa in the training dataset. To further evaluate the association, we calculated the preservation score,  $Z_{\text{summary}}$ , for each single module against the testing dataset. 55 of the 294 modules were preserved with  $Z_{\text{summary}}(\text{module}) > 5$ . As shown in Additional file 1: Table S1, these 55 modules belong to 35 GO\_BP terms. Many of these terms have been reported to be associated with cancer development, such as “biosynthetic process” [18], “cell-cell signaling” [51,52], “inflammatory response” [53], “response to stress” [54], “post translational protein modification” [55], “immune system process” [56], “phosphorylation” [57], “regulation of apoptosis” [58], and “regulation of cell proliferation” [59] in many cancer types, including PrCa. Although expected, these results confirmed that the identified modules are important to PrCa, and the method for the analysis is reasonable. In addition, the successful detection indicates that there are indeed subsets of genes within each single GO\_BP term that are co-expressed and associated with PrCa. This further verified the rationale of our procedure to examine co-expression patterns in each GO\_BP term, rather than in the whole gene set on chip, and proved its ability to discover disease-associated genes and modules.

### Characterization of the identified PrCa associated modules

Our enrichment test of the 55 preserved PrCa-associated modules showed that six modules were significantly enriched with PrCa candidate genes, such as eQTL, SCNA, or mutant genes (FDR adjusted  $p < 0.05$ ) (Additional file 1: Table S1). Their functions are annotated as “response to stress (labeled in green in Figure 1A4)” ( $p_{\text{cis-eQTL}} = 0.017$ ,  $p_{\text{trans-eQTL}} = 3.16 \times 10^{-3}$ ,  $p_{\text{mutation}} = 1.37 \times 10^{-3}$ ),

“cellular localization (turquoise)” ( $p_{\text{cis-eQTL}} = 7.30 \times 10^{-3}$ ,  $p_{\text{SCNA}} = 0.024$ ,  $p_{\text{prog}} = 1.62 \times 10^{-3}$ ), “protein localization (brown)” ( $p_{\text{cis-eQTL}} = 0.034$ ,  $p_{\text{SCNA}} = 0.039$ ,  $p_{\text{mutation}} = 6.84 \times 10^{-4}$ ), “regulation of apoptosis (green)” ( $p_{\text{cis-eQTL}} = 8.93 \times 10^{-3}$ ,  $p_{\text{SCNA}} = 0.040$ ), “regulation of apoptosis (red)” ( $p_{\text{cis-eQTL}} = 0.027$ ,  $p_{\text{trans-eQTL}} = 6.59 \times 10^{-3}$ ,  $p_{\text{mutation}} = 6.39 \times 10^{-5}$ ), and “apoptosis go (black)” ( $p_{\text{cis-eQTL}} = 8.93 \times 10^{-3}$ ,  $p_{\text{trans-eQTL}} = 1.08 \times 10^{-3}$ ,  $p_{\text{mutation}} = 6.41 \times 10^{-5}$ ). We calculated the pairwise similarities between the six modules using Fisher’s exact test. As a result, “regulation of apoptosis (red)” and “apoptosis go (black)” were found with a large proportion of overlapping genes ( $p = 3.7 \times 10^{-41}$ ). We therefore merged these two modules and referred to the resultant module as M1. Other modules were listed as M2-M5, as summarized in Table 2.

For these identified PrCa-associated modules, we recalculated the Module Membership (kME) of each gene by its correlation with the module eigengene (Additional file 2: Table S2). In particular, we presented the five modules in Figure 2, in which nodes were ranked by their kME values and edge thickness reflected the correlations’ coefficients. Node shapes represented different gene functions, as shown in Additional file 2: Table S2. Genes with higher kME values were highly interconnected in modules M1, M3, and M4, indicating a strong co-expression pattern in PrCa (Figure 2). In modules M2 and M5, relatively moderate connections were observed. This is likely due to a weak PrCa association (M2,  $p_{\text{cor}} = 0.015$ ) or a low preservation score (M5,  $Z_{\text{summary}} = 5.2$ ). The biological functions of modules M1 and M2 are associated with the GO\_BP term “apoptosis.” Representative genes include *IL6* [60], *SOCS3* [61], *GADD45A* [62], *PIMI1* [63], *IL1B* [64], *CDKN1A* [65], *CCL2* [66], *PMAIP1* (also known as *NOXA*) [67], and *RHOB* [68] in module M1, and *DHCR24* [69], *BNIP3* [70], and *IGF1R* [71] in module M2. For modules M3, M4, and M5, although the corresponding GO\_BP terms were not directly related to cancer, we found PrCa-relevant genes in these three modules, including *BTG2* [72], *FOS* (also known as *c-Fos*) [73], and *CXCR4* [74] in module M3; *ARFGAP3* [75] and *CDH1* [76] in module M4; and *SMAD3* [77] and *MXII* [78] in module M5.

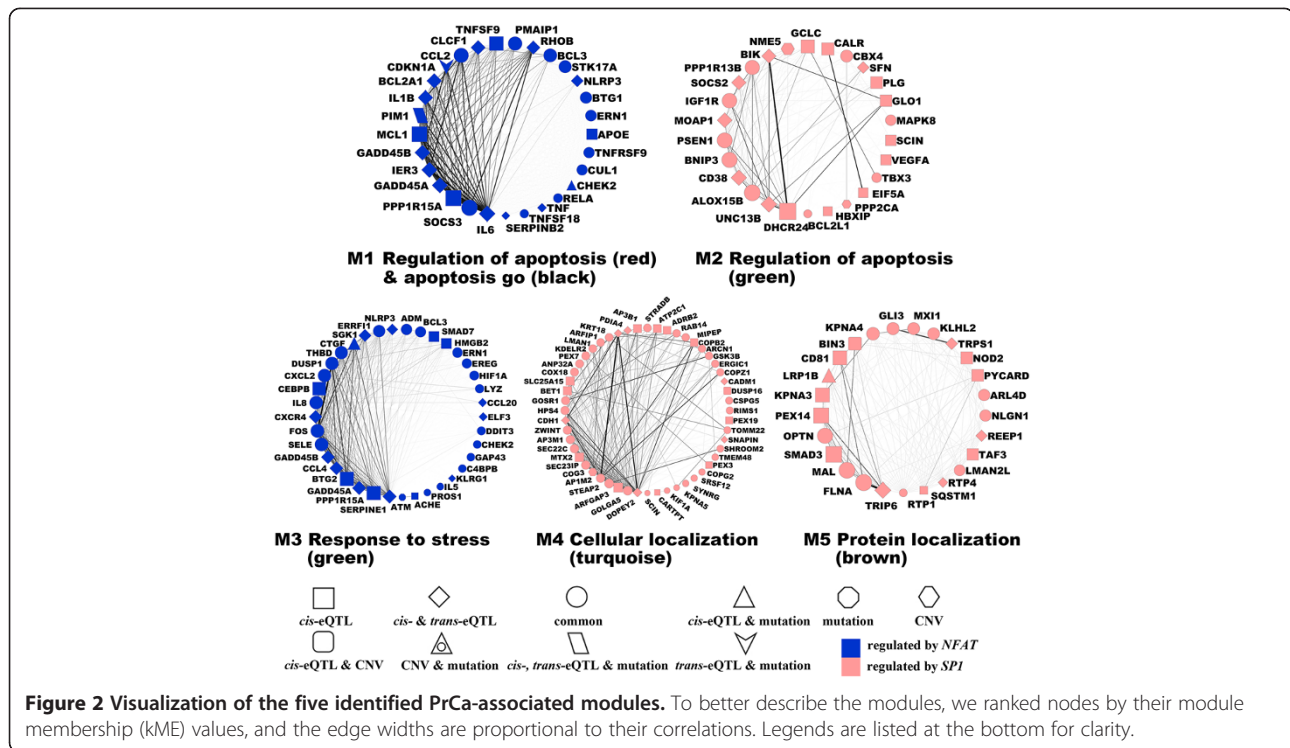
Moreover, we identified TF and microRNA regulators enriched in the modules, and the results are shown in Table 3. Some TFs could be seen as associated with several modules. For example, modules M1 and M3 were significantly enriched with the nuclear factor of activated T-cell transcription factor gene, *NFAT* ( $p_{\text{M1}} = 3.59 \times 10^{-7}$ ,  $p_{\text{M3}} = 2.20 \times 10^{-5}$ ). All the other three modules—M2, M4, and M5—were significantly enriched with the gene specificity protein 1, *SPI* ( $p_{\text{M2}} = 3.00 \times 10^{-4}$ ,  $p_{\text{M4}} = 4.00 \times 10^{-4}$ ,  $p_{\text{M5}} = 1.20 \times 10^{-3}$ ) (Figure 2). Both *NFAT* and *SPI* were known to be associated with PrCa [79,80].

To further search for evidence of the coordinated regulation of each identified module, we explored the

**Table 2 Overview of enrichment analyses of the five identified PrCa modules**

Module ID	Description	$p_{\text{Cor}}^a$	Size	$Z_{\text{summary}}$	<i>cis</i> -eQTL		<i>trans</i> -eQTL		SCNA		Somatic mutation		Prognostic genes	
					# genes	$p_{\text{cis-eQTL}}^a$	# genes	$p_{\text{trans-eQTL}}^a$	# genes	$p_{\text{SCNA}}^a$	# genes	$p_{\text{mutation}}^a$	# genes	$p_{\text{Prog}}^a$
M1	Regulation of apoptosis (red) Apoptosis go (black)	$6.23 \times 10^{-8}$	29	10.0	17	$3.66 \times 10^{-3}$	13	$1.01 \times 10^{-3}$	NA	NA	3	$1.99 \times 10^{-4}$	NA	NA
M2	Regulation of apoptosis (green)	$1.51 \times 10^{-2}$	26	6.1	15	$8.93 \times 10^{-3}$	7	0.11	2	0.040	NA	NA	2	0.081
M3	Response to stress (green)	$4.18 \times 10^{-11}$	37	8.3	19	0.017	14	$3.16 \times 10^{-3}$	NA	NA	2	$1.37 \times 10^{-3}$	NA	NA
M4	Cellular localization (turquoise)	$3.27 \times 10^{-9}$	70	5.7	35	$7.30 \times 10^{-3}$	9	0.85	6	0.024	NA	NA	10	$1.62 \times 10^{-3}$
M5	Protein localization (brown)	$8.04 \times 10^{-10}$	24	5.2	13	0.034	5	0.29	2	0.039	2	$6.84 \times 10^{-4}$	NA	NA

<sup>a</sup> $p$  values were adjusted by FDR method.



module eigengene in tumor and control samples for all five modules. As shown in Figure 3, the eigengenes of modules M1, M2, M3, and M4 were consistently over-expressed in PrCa tumor samples compared to control samples, while the eigengene of module M5 was under-expressed. These observations implied that the module genes might be co-regulated by the same or similar regulators, e.g., TFs (or their regulators) or miRNAs that regulate the expression of the module genes. We therefore examined the correlation between the candidate TFs (Table 3) with PrCa. Specifically, the TF NFAT family genes, such as *NFATC4* ( $p_{\text{cor1}} = 8.94 \times 10^{-7}$ ,  $p_{\text{cor2}} = 7.66 \times 10^{-4}$ ) and *NFATC1* ( $p_{\text{cor1}} = 2.25 \times 10^{-3}$ ,  $p_{\text{cor2}} = 3.61 \times 10^{-2}$ ), were associated with PrCa. The activation of partner gene *TRPV6* was reported to be critical to NFAT [79,81] in prostate cancer cells. Our specific examination showed that *TRPV6* was significantly over-expressed in PrCa tumor samples ( $p_{\text{cor1}} = 6.49 \times 10^{-6}$ ) in the training dataset, but not in the testing dataset ( $p_{\text{cor2}} = 0.12$ ). In addition, other enriched TFs showed promising evidence (Table 3) for association with PrCa, such as *SP1* ( $p_{\text{cor1}} = 4.60 \times 10^{-7}$  and  $p_{\text{cor2}} = 7.50 \times 10^{-5}$ ), *REL* ( $p_{\text{cor1}} = 8.08 \times 10^{-10}$ ,  $p_{\text{cor2}} = 0.010$ ), and *JUN* ( $p_{\text{cor1}} = 5.91 \times 10^{-15}$ ,  $p_{\text{cor2}} = 7.70 \times 10^{-5}$ ) in module M1; *RFX1* ( $p_{\text{cor1}} = 3.67 \times 10^{-5}$ ,  $p_{\text{cor2}} = 3.10 \times 10^{-6}$ ), *VSX1* ( $p_{\text{cor1}} = 1.55 \times 10^{-4}$ ,  $p_{\text{cor2}} = 0.036$ ), and *RORA* ( $p_{\text{cor1}} = 1.29 \times 10^{-3}$ ,  $p_{\text{cor2}} = 3.15 \times 10^{-11}$ ) in module M2; *NF1* ( $p_{\text{cor1}} = 1.14 \times 10^{-3}$ ,  $p_{\text{cor2}} = 6.01 \times 10^{-5}$ ) in module M4; and *CDC5L* ( $p_{\text{cor1}} = 0.012$ ,  $p_{\text{cor2}} = 8.07 \times 10^{-6}$ ) in module M5.

In Table 3, we listed the enriched miRNAs; several of them have been reported to be associated with PrCa, such as hsa-miR-19a [82] with modules M1 and M3 ( $p_{\text{M1}} = 3.50 \times 10^{-3}$ ,  $p_{\text{M3}} = 6.10 \times 10^{-3}$ ), hsa-miR-15a [83] with modules M4 and M5 ( $p_{\text{M4}} = 4.70 \times 10^{-3}$ ,  $p_{\text{M5}} = 3.00 \times 10^{-4}$ ), and hsa-miR-200b [84] with module M2 ( $p_{\text{M2}} = 9.00 \times 10^{-3}$ ). For the other microRNAs, experimental validation is needed to investigate their roles in PrCa.

In order to validate the regulatory TFs/miRNAs detected above, we retrieved RNASeqV2 and miRNASeq data for prostate adenocarcinoma from the TCGA portal to build the expression matrix. The R package edgeR [85] was applied to obtain the differentially expressed genes and miRNAs. The FDR method was applied to adjust  $p$  values for multiple testing. We found that the identified key TF regulators, including NFAT family genes ( $p_{\text{NFATC4}} = 7.85 \times 10^{-12}$ ,  $p_{\text{NFAT5}} = 6.76 \times 10^{-6}$ ,  $p_{\text{NFATC2}} = 6.79 \times 10^{-6}$ ,  $p_{\text{NFATC3}} = 6.95 \times 10^{-3}$ ), NFAT regulator *TRPV6* ( $p = 1.40 \times 10^{-6}$ ), and *SP1* ( $p = 9.49 \times 10^{-3}$ ), were highly differentially expressed, as well as the other enriched TFs, such as *REL* ( $p = 1.35 \times 10^{-5}$ ), *RORA* ( $p = 1.54 \times 10^{-13}$ ), and *NF1* ( $p = 6.24 \times 10^{-4}$ ). Similar patterns were also observed in the identified miRNAs, such as hsa-miR-19a ( $p = 1.72 \times 10^{-12}$ ), hsa-miR-15a ( $p = 8.10 \times 10^{-10}$ ), and hsa-miR-200b ( $p = 6.39 \times 10^{-3}$ ).

## Discussion

High-throughput genetic and genomic studies have demonstrated that GO terms are important prior knowledge

**Table 3 Transcription factors (TFs) and microRNAs (miRNAs) identified by the analyses of PrCa-associated modules (M1-M5)**

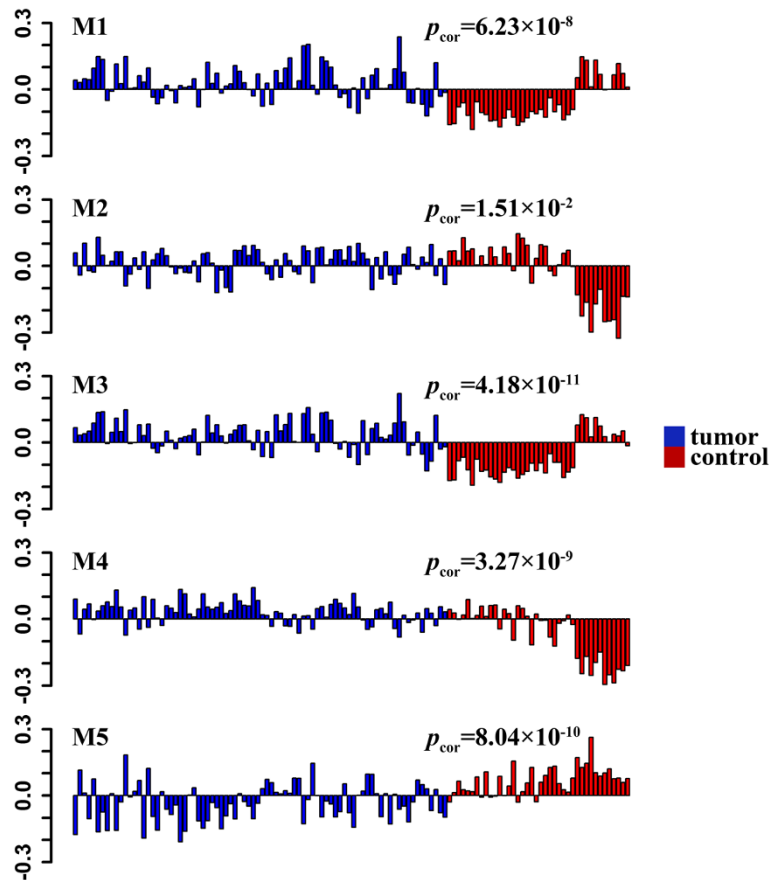
	No.	Module genes	TF symbol	$p^a$	Module genes	miRNA symbol	$p$
Module M1	1	<i>CEBPB CXCR4 CTGF PPP1R15A SELE CCL4 EREG SERPINE1 FOS THBD</i>	TATA	$1.23 \times 10^{-5}$	<i>BTG2 GAP43 ADM SMAD7</i>	hsa-miR-25	$2.70 \times 10^{-3}$
	2	<i>GADD45A CTGF HIF1A GADD45B CCL4 GAP43 SGK1 ERRFI1 FOS IL5 ADM</i>	NFAT [79]	$2.20 \times 10^{-5}$	<i>CEBPB GAP43</i>	hsa-miR-191	$4.00 \times 10^{-3}$
	3	<i>CEBPB NLRP3 GAP43 EREG HMGB2 ERRFI1 IL5 DUSP1</i>	UNKNOWN	$5.19 \times 10^{-5}$	<i>HMGB2 FOS ADM SMAD7</i>	hsa-miR-181a	$6.10 \times 10^{-3}$
	4	<i>ATM HIF1A FOS CCL20 EREG</i>	TCF1P	$7.40 \times 10^{-5}$	<i>SGK1 BCL3 CTGF EREG</i>	<i>hsa-miR-19a</i> [82]	$6.10 \times 10^{-3}$
	5	<i>BCL3 CTGF CCL20 GADD45B ERN1</i>	CREL	$7.40 \times 10^{-5}$	<i>ERRFI1 GADD45A GAP43</i>	hsa-miR-148a	$1.19 \times 10^{-2}$
Module M2	1	<i>DOPEY1 CARTPT RIMS1 CDH1 TLK1 STEAP2 ANP32A TOMM22 CADM1 ADRB2 COG3 XPO7 COG2 KDELR2 PDIA4</i>	UNKNOWN	$6.27 \times 10^{-5}$	<i>BET1 AP1G1 ARCN1 SEC62</i>	hsa-miR-409-3p	$3.50 \times 10^{-2}$
	2	<i>TMEM48 SEC62 CADM1 SYNRG AP3M1 DOPEY1 COPZ1 RIMS1 ARFIP1 ANP32A ARFGAP3 GSK3B COG3 TIMM17B LMAN1 AP1G1 SNAPIN</i>	SP1 [80]	$4.00 \times 10^{-4}$	<i>RAB14 AP1G1 ATP2C1</i>	hsa-miR-302b	$3.80 \times 10^{-3}$
	3	<i>SEC23IP ARFIP1 ATP2C1 TLK1 TOMM22 AP3B1 COG3 UHMK1 SYNRG MTX2</i>	ELK1	$1.50 \times 10^{-3}$	<i>RAB14 ARCN1 AP3M1 RERE</i>	hsa-miR-211	$4.70 \times 10^{-3}$
	4	<i>COPZ1 CARTPT ATP2C1 STEAP2 ANP32A CADM1 UHMK1 AP1G1 SYNRG ARCN1 MTX2 RERE</i>	NFAT	$2.60 \times 10^{-3}$	<i>STRADB CADM1 ADRB2 SYNRG ARCN1 TLK1</i>	<i>hsa-miR-15a</i> [83]	$4.70 \times 10^{-3}$
	5	<i>DOPEY1 CARTPT ERGIC1 AP1M2 CDH1 MIPEP SEC22A ANP32A CADM1 ADRB2 SHROOM2 AP3M1 RPAIN</i>	E12	$5.90 \times 10^{-3}$	<i>BET1 AP1G1 ARCN1 SEC62</i>	hsa-miR-1	$1.14 \times 10^{-2}$
Module M3	1	<i>SMAD3 KPNA4 KPNA3 LMAN2L REEP1 FLNA PEX14 OPTN TRPS1</i>	SP1	$1.20 \times 10^{-3}$	<i>SMAD3 MXI1 LRP1B KPNA4 TRPS1</i>	hsa-miR-524	$1.00 \times 10^{-4}$
	2	<i>RTP4 KPNA3 OPTN</i>	STAT1	$4.80 \times 10^{-3}$	<i>SMAD3 NLGN1 LRP1B KPNA4 LMAN2L</i>	<i>hsa-miR-15a</i>	$3.00 \times 10^{-4}$
	3	<i>MXI1 KPNA3 TRPS1</i>	CEBPA	$4.80 \times 10^{-3}$	<i>NLGN1 KPNA4 REEP1 TRPS1</i>	hsa-miR-493	$3.00 \times 10^{-4}$
	4	<i>SMAD3 BIN3 LMAN2L</i>	PITX2	$4.80 \times 10^{-3}$	<i>MXI1 KPNA3 REEP1</i>	hsa-miR-24	$2.40 \times 10^{-3}$
	5	<i>GLI3 MXI1 TRPS1</i>	CDC5L	$4.80 \times 10^{-3}$	<i>SMAD3 KPNA3</i>	hsa-miR-302b	$3.70 \times 10^{-3}$
Module M4	1	<i>CALR EIF5A UNC13B IGF1R TBX3 SOCS2 GLO1 BCL2L1 VEGFA PSEN1</i>	SP1	$3.00 \times 10^{-4}$	<i>BCL2L1 VEGFA DHCR24</i>	hsa-miR-377	$3.00 \times 10^{-3}$
	2	<i>EIF5A BCL2L1 PPP1R13B SOCS2</i>	UNKNOWN	$6.00 \times 10^{-4}$	<i>PPP2CA SOCS2</i>	hsa-miR-139	$9.00 \times 10^{-3}$
	3	<i>BCL2L1 VEGFA NME5 BIK</i>	RFX1	$6.00 \times 10^{-4}$	<i>PPP2CA VEGFA CBX4</i>	<i>hsa-miR-200b</i> [84]	$9.00 \times 10^{-3}$
	4	<i>PPP2CA CALR EIF5A PPP1R13B SFN</i>	VSX1	$1.90 \times 10^{-3}$	<i>VEGFA TBX3</i>	hsa-miR-140	$9.00 \times 10^{-3}$
	5	<i>PPP2CA EIF5A SOCS2</i>	RORA	$1.90 \times 10^{-3}$	<i>PPP2CA VEGFA PPP1R13B</i>	hsa-miR-29a	$9.60 \times 10^{-3}$
Module M5	1	<i>TNF BTG1 GADD45A CCL2 IL6 GADD45B PIM1 CDKN1A RELA RHOB IER3</i>	NFAT	$3.59 \times 10^{-7}$	<i>GADD45A RHOB SOCS3</i>	hsa-miR-527	$3.50 \times 10^{-3}$
	2	<i>BCL3 TNFRSF9 TNFSF18 ERN1 GADD45B IER3</i>	RELA	$3.59 \times 10^{-7}$	<i>BTG1 BCL3 RHOB SOCS3</i>	<i>hsa-miR-19a</i>	$3.50 \times 10^{-3}$
	3	<i>BCL3 TNFRSF9 TNFSF18 ERN1 GADD45B IER3</i>	REL	$3.59 \times 10^{-7}$	<i>NLRP3 RHOB</i>	hsa-miR-223	$8.40 \times 10^{-3}$
	4	<i>TNF BCL3 TNFRSF9 TNFSF18 ERN1 GADD45B</i>	NFKB	$3.59 \times 10^{-7}$	<i>BTG1 PIM1</i>	hsa-miR-183	$2.14 \times 10^{-2}$
	5	<i>TNFRSF9 GADD45A IL6 PPP1R15A GADD45B CDKN1A PIM1 RELA SERPINB2</i>	JUN	$4.86 \times 10^{-7}$	<i>CUL1 SOCS3</i>	hsa-miR-203	$2.89 \times 10^{-2}$

TFs and miRNAs in italics were those shared between modules and associated with prostate cancer.  
<sup>a</sup> $p$  values were adjusted by FDR method.

in facilitating and interpreting of discoveries in complex disease studies. In this study, we identified gene co-expression modules within GO\_BP terms for PrCa. We

found 118 GO\_BP terms that were preserved between training and testing datasets, some of them have been widely studied and reported, such as “programmed cell





**Figure 3** Expression of module eigengenes in five modules (M1-M5) across samples. Blue indicates tumor samples, while red indicates control samples.

death” [86], “cell-cell adhesion” [87], and “regulation of apoptosis” [87]. We applied WGCNA to the PrCa expression data sets and identified five co-expression modules which were preserved in the training and testing datasets and enriched with known PrCa genes. In our further evaluation of these modules, we identified several PrCa associated TFs and miRNAs as putative key regulators in PrCa genesis and progression.

To evaluate the performance of our approach, we applied another popular co-expression network reconstruction algorithm, K-means [88], to the GO\_BP-based expression matrices. For each GO\_BP term, the number of modules obtained from WGCNA was assigned to K-means. Among the 548 constructed modules, only 12 modules showed significant association with PrCa status ( $FDR < 0.05$ ) and were preserved in the testing dataset ( $Z_{summary} > 5$ ). Further enrichment tests showed that these 12 modules were poorly enriched in the collected PrCa genes (Additional file 3: Table S3). Since the module preservation calculation is computationally time-consuming, we did not perform other algorithms for comparison. Although more comparison with other methods

may be needed, the WGCNA approach seems to be effective on detecting the risk modules for PrCa.

The results revealed that the co-expression modules that belong to known cancer-related GO terms could play regulatory roles in PrCa, such as the two apoptosis-related candidate modules, M1 and M2. The results also indicated that those modules associated with general terms, e.g., “response to stress,” “cellular localization,” and “protein localization,” may contribute to PrCa risk in a synergistic way. As a core signaling pathway in cancers [89-92], apoptosis-blocking has proven to be very important in cell development [58,93] during the stages of progression from normal epithelial cells, to androgen-dependent tumor cells, and further to malignant androgen-independent ones. On the other hand, cells can be activated in various ways in response to stress during cell development, mainly to maintain the balance between cell death and cell proliferation [94]. Therefore, cells that experience too much stress, e.g., an over-expressed module M3, may bring down the rate of cancer cell death and thus result in the formation of cancer [95-97]. As indicated by previous studies [98,99], we then considered that the alternation of

expression for genes involved in cellular and protein localizations play critical roles during cell-division and cancer cell proliferation, such as through modules M4 and M5; studying the localizations of these genes and their encoded proteins can help us elucidate the molecular basis of cancer genesis and progression [100,101].

As we found, different modules may share the same genetic regulator, such as TF and miRNA. An intriguing example is the TF, NFAT, enriched by both modules M1 and M3. NFAT is reported to promote the epithelial cell proliferation of human primary PrCa [79,81] with store-independent  $Ca^{2+}$  entry via the TRPV6 channel. Significant expression changes of NFAT family genes and their regulator *TRPV6* were observed in PrCa in the datasets. Another TF, SP1, was found to be PrCa-associated and enriched in three modules: M2, M4, and M5. SP1 has been considered an important target for PrCa therapy, since it regulates important genes, like the androgen receptor gene (*AR*), *TGF- $\beta$* , *c-Met* and prostate specific antigen (*PSA*), and others. These genes are involved in cell cycle, proliferation, cell differentiation, and apoptosis [80]. Other enriched TFs, STAT1 and NFKB [102], and moderately enriched miRNAs, hsa-miR-15a [83] and hsa-miR-19a [82] (Table 3), are also reported to be associated with PrCa. Similar expression patterns were also observed in another independent TCGA dataset. Taken together, our findings suggested that these 5 modules and their TF and microRNA regulators are likely critical for the genesis and progression of PrCa. These modules and regulators may be molecular targets for the future development of drugs and new therapies.

In this study, we chose a relatively stringent significance level to detect candidate modules. The identified modules need to be enriched with both eQTL genes and SCNA or mutated genes, with a corrected  $p < 0.05$ . This might exclude moderately associated modules (e.g.,  $p < 0.2$ ), such as “response to chemical stimulus (in turquoise in Figure 1A4) ( $p_{trans-eQTL} = 0.086$ ,  $p_{SCNA} = 0.062$ ),” “regulation of apoptosis (turquoise) ( $p_{cis-eQTL} = 0.075$ ,  $p_{SCNA} = 0.039$ ),” and “apoptosis go (turquoise) ( $p_{cis-eQTL} = 0.071$ ,  $p_{SCNA} = 0.024$ )” (Additional file 1: Table S1).

In summary, our findings indicate that genes with same GO functions can cluster into several co-expressed modules to contribute to PrCa, as seen in modules M1 and M2. Modules across GO terms may act in the networks that are regulated by same genetic factors, such as modules M1 and M3. These findings indicate the importance of studying PrCa development at a systems level rather than at a single-gene level, offering insights into the underlying mechanisms of PrCa.

## Conclusions

Using GO\_BP terms to start, we conducted gene co-expression analysis of expression profiles of PrCa. Our

results revealed five modules that were differentially expressed between tumors and controls, preserved between independent expression datasets, and enriched with putative cancer genes. The enrichment analyses further identified TF and miRNA as key regulators in PrCa. Our study provides important insights for the future investigation of molecular functionality related to PrCa etiology and the development of PrCa diagnosis tools and targeted therapy strategies.

## Additional files

**Additional file 1: Table S1.** Enrichment analyses of 55 preserved prostate cancer associated modules and preservation summary of corresponding GO\_BP terms.

**Additional file 2: Table S2.** Summarization of modules M1 ~ M5.

**Additional file 3: Table S3.** Enrichment analyses of 12 preserved prostate cancer-associated modules by K-means algorithm.

## Competing interest

The authors declare that they have no competing interests.

## Authors' contributions

BS, ZZ, and PJ conceived and designed the experiments. JJ and PJ performed the experiments. JJ and PJ analyzed the data. JJ, PJ, ZZ, and BS wrote the paper. All authors read and approved the final manuscript.

## Acknowledgements

This work was partially supported by National Institutes of Health grants (R01LM011177, P50CA095103, P50CA098131, and P30CA068485) and Ingram Professorship Funds (to ZZ). We gratefully acknowledge financial support from National Natural Science Foundation of China grants (31470821, 91230117, and 31170795) and the program of China Scholarship Council (No. 201206920018).

## Author details

<sup>1</sup>Center for Systems Biology, Soochow University, Jiangsu, China. <sup>2</sup>Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN 37203, USA. <sup>3</sup>Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN 37232, USA. <sup>4</sup>Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN 37212, USA. <sup>5</sup>Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA.

Received: 8 July 2014 Accepted: 17 November 2014

Published: 24 November 2014

## References

1. Baade PD, Youlten DR, Krnjacki LJ: International epidemiology of prostate cancer: geographical distribution and secular trends. *Mol Nutr Food Res* 2009, **53**(2):171–184.
2. Siegel R, Ward E, Brawley O, Jemal A: Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin* 2011, **61**(4):212–236.
3. Gann PH: Risk factors for prostate cancer. *Rev Urol* 2002, **4**(Suppl 5):S3–S10.
4. Nelson WG, De Marzo AM, Isaacs WB: Prostate cancer. *N Engl J Med* 2003, **349**(4):366–381.
5. Chen J, Zhang D, Yan W, Yang D, Shen B: Translational bioinformatics for diagnostic and prognostic prediction of prostate cancer in the next-generation sequencing Era. *Biomed Res Int* 2013, **2013**:6589252.
6. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K: Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000, **343**(2):78–85.

7. Page WF, Braun MM, Partin AW, Caporaso N, Walsh P: **Heredity and prostate cancer: a study of World War II veteran twins.** *Prostate* 1997, **33**(4):240–245.
8. Jia Z, Wang Y, Sawyers A, Yao H, Rahmatpanah F, Xia XQ, Xu Q, Pio R, Turan T, Koziol JA, Goodison S, Carpenter P, Wang-Rodriguez J, Simoneau A, Meyskens F, Sutton M, Lernhardt W, Beach T, Monforte J, McClelland M, Mercola D: **Diagnosis of prostate cancer using differentially expressed genes in stroma.** *Cancer Res* 2011, **71**(7):2476–2487.
9. Li Y, Vongsangnak W, Chen L, Shen B: **Integrative analysis reveals disease-associated genes and biomarkers for prostate cancer progression.** *BMC Med Genet* 2014, **7**(Suppl 1):S3.
10. Wallace TA, Prueitt RL, Yi M, Howe TM, Gillespie JW, Yfantis HG, Stephens RM, Caporaso NE, Loffredo CA, Ambros S: **Tumor immunobiological differences in prostate cancer between African-American and European-American men.** *Cancer Res* 2008, **68**(3):927–936.
11. Wang Y, Chen J, Li Q, Wang H, Liu G, Jing Q, Shen B: **Identifying novel prostate cancer associated pathways based on integrative microarray data analysis.** *Comput Biol Chem* 2011, **35**(3):151–158.
12. Chen J, Wang Y, Shen B, Zhang D: **Molecular signature of cancer at gene level or pathway level? case studies of colorectal cancer and prostate cancer microarray data.** *Comput Math Methods Med* 2013, **2013**:909525.
13. Grisanzio C, Werner L, Takeda D, Awoyemi BC, Pomerantz MM, Yamada H, Sooriakumaran P, Robinson BD, Leung R, Schinzel AC, Mills I, Ross-Adams H, Neal DE, Kido M, Yamamoto T, Petrozziello G, Stack EC, Lis R, Kantoff PW, Loda M, Sartor O, Egawa S, Tewari AK, Hahn WC, Freedman ML: **Genetic and functional analyses implicate the NUDT11, HNF1B, and SLC22A3 genes in prostate cancer pathogenesis.** *Proc Natl Acad Sci U S A* 2012, **109**(28):11252–11257.
14. Zhang W, Zang J, Jing X, Sun Z, Yan W, Yang D, Guo F, Shen B: **Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer.** *J Transl Med* 2014, **12**:66.
15. Jiang J, Jia P, Shen B, Zhao Z: **Top associated SNPs in prostate cancer are significantly enriched in cis-expression quantitative trait loci and at transcription factor binding sites.** *Oncotarget* 2014, **5**(15):6168–6177.
16. Liu W, Sun J, Li G, Zhu Y, Zhang S, Kim ST, Wiklund F, Wiley K, Isaacs SD, Stattin P, Xu J, Duggan D, Carpten JD, Isaacs WB, Gronberg H, Zheng SL, Chang BL: **Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer.** *Cancer Res* 2009, **69**(6):2176–2179.
17. Li LC, Zhao H, Shiina H, Kane CJ, Dahiya R: **PGDB: a curated and integrated database of genes related to the prostate.** *Nucleic Acids Res* 2003, **31**(1):291–293.
18. Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H: **Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types.** *Nat Commun* 2014, **5**:3231.
19. Tang Y, Yan W, Chen J, Luo C, Kaipia A, Shen B: **Identification of novel microRNA regulatory pathways associated with heterogeneous prostate cancer.** *BMC Syst Biol* 2013, **7**(Suppl 3):S6.
20. Feldman BJ, Feldman D: **The development of androgen-independent prostate cancer.** *Nat Rev Cancer* 2001, **1**(1):34–45.
21. Chen C, Cheng L, Grennan K, Pibiri F, Zhang C, Badner JA, Gershon ES, Liu C: **Two gene co-expression modules differentiate psychotics and controls.** *Mol Psychiatry* 2012, **18**(12):1308–1314.
22. Stamatoyannopoulos JA: **The genomics of gene expression.** *Genomics* 2004, **84**(3):449–457.
23. Barabasi AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease.** *Nat Rev Genet* 2011, **12**(1):56–68.
24. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM: **Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer.** *Cancer Res* 2002, **62**(15):4427–4433.
25. Stuart JM, Segal E, Koller D, Kim SK: **A gene-coexpression network for global discovery of conserved genetic modules.** *Science* 2003, **302**(5643):249–255.
26. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Stat Appl Genet Mol Biol* 2005, **4**:Article17.
27. Ben-David E, Shifman S: **Networks of neuronal genes affected by common and rare variants in autism spectrum disorders.** *PLoS Genet* 2012, **8**(3):e1002556.
28. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9**:559.
29. Langfelder P, Luo R, Oldham MC, Horvath S: **Is my network module preserved and reproducible?** *PLoS Comput Biol* 2011, **7**(1):e1001057.
30. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH: **Functional organization of the transcriptome in human brain.** *Nat Neurosci* 2008, **11**(11):1271–1282.
31. Miller JA, Horvath S, Geschwind DH: **Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways.** *Proc Natl Acad Sci U S A* 2010, **107**(28):12698–12703.
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene ontology consortium.** *Nat Genet* 2000, **25**(1):25–29.
33. Wang Y, Xia XQ, Jia Z, Sawyers A, Yao H, Wang-Rodriguez J, Mercola D, McClelland M: **In silico estimates of tissue components in surgical samples based on expression profiling data.** *Cancer Res* 2010, **70**(16):6448–6455.
34. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH: **Transcriptomic analysis of autistic brain reveals convergent molecular pathology.** *Nature* 2011, **474**(7351):380–384.
35. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data.** *Nucleic Acids Res* 2003, **31**(4):e15.
36. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, Haussler D: **Knowledge-based analysis of microarray gene expression data by using support vector machines.** *Proc Natl Acad Sci U S A* 2000, **97**(1):262–267.
37. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**(43):15545–15550.
38. Langfelder P, Zhang B, Horvath S: **Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R.** *Bioinformatics* 2008, **24**(5):719–720.
39. Langfelder P, Horvath S: **Eigengene networks for studying the relationships between co-expression modules.** *BMC Syst Biol* 2007, **1**:54.
40. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A* 2003, **100**(16):9440–9445.
41. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG: **Common genetic variants account for differences in gene expression among ethnic groups.** *Nat Genet* 2007, **39**(2):226–231.
42. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurler ME, Dermitzakis ET: **Relative impact of nucleotide and copy number variation on gene expression phenotypes.** *Science* 2007, **315**(5813):848–853.
43. Choy E, Yelensky R, Bonakdar S, Plenge RM, Saxena R, De Jager PL, Shaw SY, Wolfish CS, Slavik JM, Cotsapas C, Rivas M, Dermitzakis ET, Cahir-McFarland E, Kieff E, Hafler D, Daly MJ, Altshuler D: **Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines.** *PLoS Genet* 2008, **4**(11):e1000287.
44. Price AL, Patterson N, Hancks DC, Myers S, Reich D, Cheung VG, Spielman RS: **Effects of cis and trans genetic ancestry on gene expression in African Americans.** *PLoS Genet* 2008, **4**(12):e1000294.
45. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE: **Common regulatory variation impacts gene expression in a cell type-dependent manner.** *Science* 2009, **325**(5945):1246–1250.
46. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET: **Transcriptome genetics using second generation sequencing in a Caucasian population.** *Nature* 2010, **464**(7289):773–777.
47. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veerrieras JB, Stephens M, Gilad Y, Pritchard JK: **Understanding mechanisms underlying human gene expression variation with RNA sequencing.** *Nature* 2010, **464**(7289):768–772.
48. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.** *Cancer Discov* 2012, **2**(5):401–404.
49. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G: **GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.** *Genome Biol* 2011, **12**(4):R41.

50. Zhang B, Kirov S, Snoddy J: **WebGestalt: an integrated system for exploring gene sets in various biological contexts.** *Nucleic Acids Res* 2005, **33**(Web Server issue):W741–W748.
51. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK: **Intrinsic disorder in cell-signaling and cancer-associated proteins.** *J Mol Biol* 2002, **323**(3):573–584.
52. Agarwal R: **Cell signaling and regulators of cell cycle as molecular targets for prostate cancer prevention by dietary agents.** *Biochem Pharmacol* 2000, **60**(8):1051–1059.
53. McArdle PA, Mir K, Almushat AS, Wallace AM, Underwood MA, McMillan DC: **Systemic inflammatory response, prostate-specific antigen and survival in patients with metastatic prostate cancer.** *Urol Int* 2006, **77**(2):127–129.
54. Benhar M, Engelberg D, Levitzki A: **ROS, stress-activated kinases and stress signaling in cancer.** *EMBO Rep* 2002, **3**(5):420–425.
55. Bode AM, Dong Z: **Post-translational modification of p53 in tumorigenesis.** *Nat Rev Cancer* 2004, **4**(10):793–805.
56. de Visser KE, Eichten A, Coussens LM: **Paradoxical roles of the immune system during cancer development.** *Nat Rev Cancer* 2006, **6**(1):24–37.
57. Haldar S, Chintapalli J, Croce CM: **Taxol induces bcl-2 phosphorylation and death of prostate cancer cells.** *Cancer Res* 1996, **56**(6):1253–1255.
58. Gurumurthy S, Vasudevan KM, Rangnekar VM: **Regulation of apoptosis in prostate cancer.** *Cancer Metastasis Rev* 2001, **20**(3–4):225–243.
59. Sonnenschein C, Olea N, Pasanen ME, Soto AM: **Negative controls of cell proliferation: human prostate cancer cells and androgens.** *Cancer Res* 1989, **49**(13):3474–3481.
60. Chung TD, Yu JJ, Spiotto MT, Bartkowski M, Simons JW: **Characterization of the role of IL-6 in the progression of prostate cancer.** *Prostate* 1999, **38**(3):199–207.
61. Pierconti F, Martini M, Pinto F, Cenci T, Capodimonti S, Calarco A, Bassi PF, Larocca LM: **Epigenetic silencing of SOCS3 identifies a subset of prostate cancer with an aggressive behavior.** *Prostate* 2011, **71**(3):318–325.
62. Ramachandran K, Gopisetty G, Gordian E, Navarro L, Hader C, Reis IM, Schulz WA, Singal R: **Methylation-mediated repression of GADD45alpha in prostate cancer and its role as a potential therapeutic target.** *Cancer Res* 2009, **69**(4):1527–1535.
63. Chen WW, Chan DC, Donald C, Lilly MB, Kraft AS: **Pim family kinases enhance tumor growth of prostate cancer cells.** *Mol Cancer Res* 2005, **3**(8):443–451.
64. Michaud DS, Daugherty SE, Berndt SJ, Platz EA, Yeager M, Crawford ED, Hsing A, Huang WY, Hayes RB: **Genetic polymorphisms of interleukin-1B (IL-1B), IL-6, IL-8, and IL-10 and risk of prostate cancer.** *Cancer Res* 2006, **66**(8):4525–4530.
65. Kibel AS, Suarez BK, Belani J, Oh J, Webster R, Brophy-Ebbers M, Guo C, Catalona WJ, Picus J, Goodfellow PJ: **CDKN1A and CDKN1B polymorphisms and risk of advanced prostate carcinoma.** *Cancer Res* 2003, **63**(9):2033–2036.
66. Loberg RD, Day LL, Harwood J, Ying C, St John LN, Giles R, Neeley CK, Pienta KJ: **CCL2 is a potent regulator of prostate cancer cell migration and proliferation.** *Neoplasia* 2006, **8**(7):578–586.
67. Matsushima-Miyagi T, Hatano K, Nomura M, Li-Wen L, Nishikawa T, Saga K, Shimbo T, Kaneda Y: **TRAIL and Noxa are selectively upregulated in prostate cancer cells downstream of the RIG-I/MAVS signaling pathway by nonreplicating Sendai virus particles.** *Clin Cancer Res* 2012, **18**(22):6271–6283.
68. Prendergast GC: **Actin' up: RhoB in cancer and apoptosis.** *Nat Rev Cancer* 2001, **1**(2):162–168.
69. Battista MC, Guimond MO, Roberge C, Doueik AA, Fazli L, Gleave M, Sabbagh R, Gallo-Payet N: **Inhibition of DHCR24/seladin-1 impairs cellular homeostasis in prostate cancer.** *Prostate* 2010, **70**(9):921–933.
70. Chen X, Gong J, Zeng H, Chen N, Huang R, Huang Y, Nie L, Xu M, Xia J, Zhao F, Meng W, Zhou Q: **MicroRNA145 targets BNIP3 and suppresses prostate cancer progression.** *Cancer Res* 2010, **70**(7):2728–2738.
71. Rochester MA, Riedemann J, Hellawell GO, Brewster SF, Macaulay VM: **Silencing of the IGF1R gene enhances sensitivity to DNA-damaging agents in both PTEN wild-type and mutant human prostate cancer.** *Cancer Gene Ther* 2005, **12**(1):90–100.
72. Jalava SE, Urbanucci A, Latonen L, Waltering KK, Sahu B, Janne OA, Seppala J, Lahdesmaki H, Tammela TL, Visakorpi T: **Androgen-regulated miR-32 targets BTG2 and is overexpressed in castration-resistant prostate cancer.** *Oncogene* 2012, **31**(41):4460–4471.
73. Day ML, Zhao X, Wu S, Swanson PE, Humphrey PA: **Phorbol ester-induced apoptosis is accompanied by NGFI-A and c-fos activation in androgen-sensitive prostate cancer cells.** *Cell Growth Differ* 1994, **5**(7):735–741.
74. Taichman RS, Cooper C, Keller ET, Pienta KJ, Taichman NS, McCauley LK: **Use of the stromal cell-derived factor-1/CXCR4 pathway in prostate cancer metastasis to bone.** *Cancer Res* 2002, **62**(6):1832–1837.
75. Obinata D, Takayama K, Urano T, Murata T, Ikeda K, Horie-Inoue K, Ouchi Y, Takahashi S, Inoue S: **ARFGAP3, an androgen target gene, promotes prostate cancer cell proliferation and migration.** *Int J Cancer* 2012, **130**(10):2240–2248.
76. Qiu LX, Li RT, Zhang JB, Zhong WZ, Bai JL, Liu BR, Zheng MH, Qian XP: **The E-cadherin (CDH1)–160 C/A polymorphism and prostate cancer risk: a meta-analysis.** *Eur J Hum Genet* 2009, **17**(2):244–249.
77. Lu S, Lee J, Revelo M, Wang X, Dong Z: **Smad3 is overexpressed in advanced human prostate cancer and necessary for progressive growth of prostate cancer cells in nude mice.** *Clin Cancer Res* 2007, **13**(19):5692–5702.
78. Prochownik EV, Eagle Grove L, Deubler D, Zhu XL, Stephenson RA, Rohr LR, Yin X, Brothman AR: **Commonly occurring loss and mutation of the MXI1 gene in prostate cancer.** *Genes Chromosomes Cancer* 1998, **22**(4):295–304.
79. Lehen'kyi V, Flourakis M, Skryma R, Prevarskaya N: **TRPV6 channel controls prostate cancer cell proliferation via Ca(2+)/NFAT-dependent pathways.** *Oncogene* 2007, **26**(52):7380–7385.
80. Sankpal UT, Goodison S, Abdelrahim M, Basha R: **Targeting Sp1 transcription factors in prostate cancer therapy.** *Med Chem* 2011, **7**(5):518–525.
81. Thebault S, Flourakis M, Vanoverbergh K, Vandermoere F, Roudbaraki M, Lehen'kyi V, Slomianny C, Beck B, Mariot P, Bonnal JL, Mauroy B, Shuba Y, Capiod T, Skryma R, Prevarskaya N: **Differential role of transient receptor potential channels in Ca2+ entry and proliferation of prostate cancer epithelial cells.** *Cancer Res* 2006, **66**(4):2038–2047.
82. Moltzahn F, Olshen AB, Baehner L, Peek A, Fong L, Stoppler H, Simko J, Hilton JF, Carroll P, Blesch R: **Microfluidic-based multiplex qRT-PCR identifies diagnostic and prognostic microRNA signatures in the sera of prostate cancer patients.** *Cancer Res* 2011, **71**(2):550–560.
83. Bonci D, Coppola V, Musumeci M, Addario A, Giuffrida R, Memeo L, D'Urso L, Pagliuca A, Biffoni M, Labbaye C, Bartucci M, Muto G, Peschle C, De Maria R: **The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities.** *Nat Med* 2008, **14**(11):1271–1277.
84. Kong D, Li Y, Wang Z, Banerjee S, Ahmad A, Kim HR, Sarkar FH: **miR-200 regulates PDGF-D-mediated epithelial-mesenchymal transition, adhesion, and invasion of prostate cancer cells.** *Stem Cells* 2009, **27**(8):1712–1721.
85. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26**(1):139–140.
86. Shyamsundar R, Kim YH, Higgins JP, Montgomery K, Jordan M, Sethuraman A, van de Rijn M, Botstein D, Brown PO, Pollack JR: **A DNA microarray survey of gene expression in normal human tissues.** *Genome Biol* 2005, **6**(3):R22.
87. LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V, Gerald WL: **Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease.** *Cancer Res* 2002, **62**(15):4499–4506.
88. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22**(3):281–285.
89. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffe EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, et al: **Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.** *Science* 2008, **321**(5897):1801–1806.
90. Lowe SW, Lin AW: **Apoptosis in cancer.** *Carcinogenesis* 2000, **21**(3):485–495.
91. Roth JA, Nguyen D, Lawrence DD, Kemp BL, Carrasco CH, Ferson DZ, Hong WK, Komaki R, Lee JJ, Nesbitt JC, Pisters KM, Putnam JB, Schea R, Shin DM, Walsh GL, Dolormente MM, Han CI, Martin FD, Yen N, Xu K, Stephens LC, McDonnell TJ, Mukhopadhyay T, Cai D: **Retrovirus-mediated wild-type p53 gene transfer to tumors of patients with lung cancer.** *Nat Med* 1996, **2**(9):985–991.
92. Bedi A, Pasricha PJ, Akhtar AJ, Barber JP, Bedi GC, Giardiello FM, Zehnbauser BA, Hamilton SR, Jones RJ: **Inhibition of apoptosis during development of colorectal cancer.** *Cancer Res* 1995, **55**(9):1811–1816.
93. Howell SB: **Resistance to apoptosis in prostate cancer cells.** *Mol Urol* 2000, **4**(3):225–229.
94. Lockshin RA, Zakeri Z: **Cell death in health and disease.** *J Cell Mol Med* 2007, **11**(6):1214–1224.

95. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**(5):646–674.
96. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, **100**(1):57–70.
97. Fulda S, Gorman AM, Hori O, Samali A: **Cellular stress responses: cell survival and cell death.** *Int J Cell Biol* 2010, **2010**:214074.
98. Abate-Shen C, Shen MM: **Molecular genetics of prostate cancer.** *Genes Dev* 2000, **14**(19):2410–2434.
99. Grauer LS, Lawler KD, Marignac JL, Kumar A, Goel AS, Wolfert RL: **Identification, purification, and subcellular localization of prostate-specific membrane antigen PSM' protein in the LNCaP prostatic carcinoma cell line.** *Cancer Res* 1998, **58**(21):4787–4789.
100. Thiebaut F, Tsuruo T, Hamada H, Gottesman MM, Pastan I, Willingham MC: **Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues.** *Proc Natl Acad Sci U S A* 1987, **84**(21):7735–7738.
101. Hu R, Dunn TA, Wei S, Isharwal S, Veltri RW, Humphreys E, Han M, Partin AW, Vessella RL, Isaacs WB, Bova GS, Luo J: **Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer.** *Cancer Res* 2009, **69**(1):16–22.
102. Shen H, Lentsch AB: **Progressive dysregulation of transcription factors NF-kappa B and STAT1 in prostate cancer cells causes proangiogenic production of CXC chemokines.** *Am J Physiol Cell Physiol* 2004, **286**(4):C840–C847.

doi:10.1186/1471-2164-15-1015

**Cite this article as:** Jiang et al.: Key regulators in prostate cancer identified by co-expression module analysis. *BMC Genomics* 2014 15:1015.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

