Transcriptomics: Open Access

Integration of Gene Coexpression Network, GO Enrichment Analysis for Identification Gene Expression Signature of Invasive Bladder Carcinoma

Hanaa Hibishy Gaballah*
Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Tanta University, Egypt.
Visiting researcher at Bioinformatic Centre, Chemical Research Lab, Kyoto University, Japan.

Abstract

Bladder carcinoma is the most common malignancy of the urinary tract. Identification of genetic biomarkers for tumor invasiveness will help in earlier diagnosis and proper treatment. The present study aimed to integrate co-expression network and GO enrichment analysis for identification of prognostic markers and key genes that contribute to bladder cancer initiation and progression using a DNA microarray dataset (GSE 37317), invasive and noninvasive bladder cancer genes were compared by applying weighted gene co-expression network, gene ontology and pathway analysis. This study identified candidate genes (PURA, SRPK2, TRAK1, BRD2, and UPF3) that might have significant role in progression and invasiveness of bladder carcinoma. These markers might aid in early diagnosis of muscle invasiveness of bladder cancer. In conclusion; these finding may provide better understanding of the molecular mechanism of bladder cancer progression and invasiveness.

Introduction

Alteration of the balance of cell proliferation and cell death is the hallmark of cancer, which cause reprogramming of cellular metabolism to support neoplastic proliferation and to avoid immune destruction Ghafouri-Fard et al., [1]. Bladder cancer is the commonest malignancy of the urinary tract, with the incidence being four times higher in men than in women Ploeg et al., [2]. Approximately 25% of the urothelial tumors are presented by muscle invasion (stage ≥T2), the rest are confined to the bladder mucosa (Ta) or lamina propria (T1) Cheung et al., [3]. Transurethral resection (TUR) of the bladder tumor is the standard method used to determine the local stage of the neoplasm which also provides treatment for patients with non-muscle-invasive disease and staging information for those with muscle-invasive disease Biagioli et al., [4]. Prior efforts have identified urinary protein biomarkers of bladder cancer Wang et al., [5,6]. However, markers that can predict tumor stage have not been identified. Such markers would provide pre-transurethral resection information that could be used for patient counseling and determination of the extent of the TUR required in cases where muscle-invasive disease is predicted Schiffer et al., [7]. The vast amount of available high-throughput gene expression data has provided excellent opportunities for studying gene functions on a global scale. Comparing the gene-expression profiles of cancerous and healthy tissues can help in the identification and characterization of susceptible genes associated with cancer Chin et al., [8]. Weighted coexpression network analysis is a system biology method for describing the correlation patterns among genes across microarray samples. It can be used for finding clusters (modules) of highly correlated genes, for summarizing such clusters, for relating modules to one another and to external sample traits, and for calculating module membership measures Horvath et al., [9,10]. Network-based gene screening methods can be used to identify candidate biomarkers or therapeutic targets and to identify key genes that contribute to the disease phenotypes Kristensen et al., [11]. By focusing on correlated gene modules rather than on individual genes, the network approach may provide a means to bridge the gap from individual genes to complex traits Stranger et al., [12]. Thus, we aimed to apply WGCNA to identify more prognosis markers and key genes that contribute to bladder cancer initiation and progression and to dissect its related biological pathways and networks by integration of coexpression network analysis and GO enrichment analysis.

Methods

Data description, processing and differential expression analysis

The data set with the series accession number (GSE 37317)

accessible at National Center for Biotechnology Information (NCBI), Gene Expression Omnibus data repository (http://www.ncbi.nlm.nih.gov/geo/) was used in this study. The 19 tissue samples for this microarray study consisted of 8 non-muscle invasive and 11 muscle invasive bladder cancers that were profiled for gene expression using the Affymetrix HG-U133A platform. The gene expression profile of one normal bladder tissue sample (GSM44682) was also used for the analysis of differential gene expression between normal and cancerous tissues. Raw gene expression data files were read into R statistical package using the Affy package and normalized using the Gene chip robust multichip average (GC-RMA) normalization method as implemented in Bioconductor Gautier et al., [13]. Then, the processed expression levels of each gene between normal samples and tumor samples were compared by linear modeling to identify a set of genes that are differentially expressed between normal and tumor samples. A false discovery rate (FDR) threshold of 5% was set to correct for multiple hypothesis testing. These analyses are performed in R statistical software with the Limma package Smyth [14].

Network construction and Module detection procedure

The expression profiles of these differentially expressed genes were inputted to construct weighted gene co-expression modules using the WGCNA R package Langfelder et al., [15]. Positive and negative correlations in gene expression may indicate different biological interactions (synergistic or antagonistic); hence, using the absolute value of the correlation may misinterpret biologically relevant information Song et al., [16]. Therefore, in order to identify direct gene-gene coexpression relationships among the set of differentially expressed genes, a signed weighted gene coexpression network was constructed by calculating the spearman's correlations between gene expression levels for all pairs of genes, then, a signed similarity (Sij) parameter was derived: $Sij=(1+cor\ (xi,xj))/2$, where xi and xj consist

*Corresponding author: Hanaa Hibishy Gaballah, Visiting Researcher at Bioinformatic Centre, Chemical Research Lab, Kyoto University, Japan, Tel: +201110077010; E-mail: hanaahibishy@hotmail.com

of the expression of genes i and j across microarray samples Mason et al., [17]. The signed similarity (Sij) was then raised to power β to represent the connection strength (aij); where aij=Sij^β. This step aims to emphasize strong correlations and reduce the emphasis of weak correlations on an exponential scale. A power of β=2 was chosen so that the resulting networks exhibited approximate scale-free topology Zhang et al., [18]. The topological overlap was used to organize the genes into clusters or modules. To calculate the topological overlap for a pair of genes, we compare them in terms of their connection strengths with all other genes in the network. Hence, the topological overlap-based metrics consider the correlation changes to all other genes to determine the similarity between two genes, this proximity measure can be used as input of clustering method followed by module extraction Yip et al., [19]. Hereby, my approach of identification of co-expressed gene modules was based on a modification of WGCNA method by applying the spectral clustering method, instead of the suggested hierarchical clustering, considering topological overlap matrix (TOM) as a similarity measure. Despite being applied in earlier studies Zhu et al., [5] and proved some utility, interpretation of hierarchical clustering is complex; besides expression patterns of individual gene sequences become less relevant as the clustering process progresses Li et al., [20]. Spectral clustering methods correspond to a family of unsupervised learning algorithms. The clustering information can be obtained from analyzing the eigenvalues and eigenvectors of a matrix derived from pairwise similarities of the data Li et al., [20]. These eigenvectors become a new representation of the data, where the clusters form a localized structure Huang et al., [21]. Following construction of a gene co-expression network, spectral clustering was run using specc function of the kernlab package implemented in R statistical software Karatzoglou et al., [22]. Number of centers was set to 18 and number of iterations to 200. Transcripts were clustered into 18 distinct gene clusters or modules whose number of genes ranged from 38 (the sixth cluster) to 71(the seventeenth cluster). Each module represents a group of genes with similar expression profiles across the samples and the expression profile pattern is distinct from those of the other modules. Singular value decomposition was performed using the module Eigengenes function in WGCNA package implemented in R software to summarize the expression levels of all genes in each module Xiao et al., [23]. The module eigengene roughly corresponds to the average of the signed normalized gene-expression values for a given sample. The module eigengene-based connectivity (k) for each gene was defined by correlating the expression profile of a gene with the module eigengene of its resident module. This measures how connected a given gene is to biologically interesting modules. The larger k_i, the more similar is gene i to the summary profile of the module Min et al., [24]. Furthermore, we assessed the trait-based gene significance by correlating the gene expression profile to the tumor stage, which was represented as a binary value (zero for the non-invasive and one for the invasive tumor). Gene significance specifies the biological significance of the gene, the higher this value, the more significant a gene is Puniya et al., [25]. Consequently, scatterplots, spearman correlation coefficients and the corresponding p-values as computed by the cor.test function in R statistical software were used to relate the gene significance to the intramodular connectivity. The physiological relevance of each module was assessed by measuring the module significance which was defined as the absolute value of the correlation between the tumor stage and the module eigengene. Functional annotation of the modules was performed on the basis of their gene composition using DAVID database**.** This software, which is available for download at (http://www.d.abcc.ncifcrf.gov/home.jsp), calculates the p value for the extent of enrichment of a given biological pathway/set by performing Fisher's exact test. 'KEGG_PATHWAY' and 'PANTHER_PATHWAY 'were selected for pathway enrichment analysis of coexpressed genes compared with the background list of all genes on the array. For characterization of modules, 'GO_BP_FAT' and 'GO_MF_FAT' were selected.

Results

The module significance was determined by correlating the trait with the module eigengene (Table 1). The three highest module significance (MS) scores observed were for the ninth module (MS=0.696, $p=9.2 \times 10^{-4}$), the thirteenth module (MS=0.648, $p=2.6 \times 10^{-3}$) and the twelfth module (MS=0.45, p=0.05). We also used the non-parametric Jonckheere-Terpstra trend test to evaluate the correlation between the module eigengenes and tumor stage. For a given physiological trait, a measure of gene significance was defined as the absolute value of the correlation between trait and gene expression values. For example, the tumor stage can be used to define a gene significance of the ith gene expression GS (i)=|cor (x (i), stage)| where x (i) is the gene expression profile of the ith gene. Furthermore, eigengene based gene connectivity (k) of the three modules, which showed the highest module significance values, was plotted against gene significance for all genes in each module. These plots, shown in (Figures 1-3), revealed that there is a direct relationship between the connectivity of a gene and the extent to which it is related to the tumor stage. The three modules showed a significant positive correlation presented by (r=0.566, $p=2. \times 10^{-6}$) for the ninth, (r=0.607, $p=1.54 \times 10^{-5}$) for the thirteenth and (r=0.373, $p=1.9 \times 10^{-3}$) for the twelfth module. So, it can be concluded that the connectivity within a physiologically relevant module is directly related to the gene significance. Using DAVID database, we tested each module for both enriched biological process and molecular function GO terms. The genes of the ninth module were markedly enriched for organelle organization ($p=1.2 \times 10^{-5}$), negative regulation of programmed cell death ($p=2.4 \times 10^{-4}$) and DNA replication ($p=3.3 \times 10^{-3}$). Moreover, the genes of the twelfth module are highly enriched for cell cycle regulation terms such as regulation of mitotic cell cycle ($p=3.4 \times 10^{-4}$), mitotic cell cycle checkpoints ($p=6.6 \times 10^{-4}$), cell cycle phase ($p=1.2 \times 10^{-3}$).

| Module number	Cluster size (number of genes in the modules)	Pearson correlation		JT test		
		r = correlation coefficient	p-value	JT test	p-value	Average gene significance
1	51	0.23	0.34	24	0.09	0.22
2	48	0.22	0.37	31	0.28	0.20
3	54	0.01	0.97	44	1	0.22
4	66	0.08	0.73	36	0.50	0.21
5	55	0.28	0.24	20	0.04	0.16
6	38	0.15	0.54	27	0.16	0.19
7	64	0.17	0.48	57	0.28	0.27
8	57	0.19	0.45	28	0.18	0.20
9	59	0.7	0.0009*	81	0.002*	0.23
10	45	0.11	0.64	26	0.13	0.29
11	42	0.41	0.07	20	0.09	0.20
12	66	0.45	0.04*	64	0.04*	0.25
13	43	0.65	0.002*	77	0.006*	0.23
14	60	0.24	0.31	65	0.08	0.31
15	68	0.14	0.55	36	0.50	0.26
16	71	0.33	0.16	19	0.06	0.22
17	71	0.07	0.79	39	0.67	0.22
18	42	0.02	0.92	45	0.93	0.20

*: P-value was considered significant at <0.05.

Table 1: Module significance determined by correlating the trait with the module eigen gene using Pearson correlation and JT test.
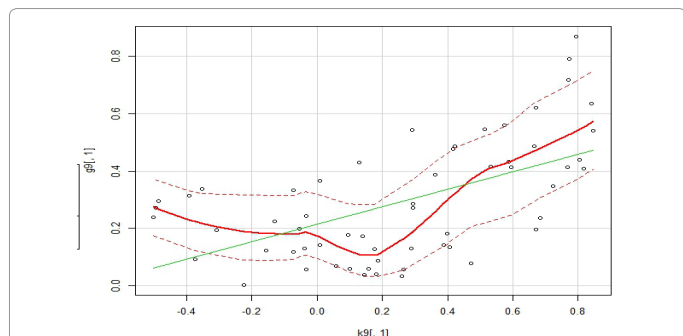
Figure 1: A scatter plot showing correlation between the gene connectivity (K value) of the ninth module genes on the X-axis and gene significance of the same module on the X-axis; showing a significantly high positive correlation ($r=0.566$, $p=2. \times 10^{-6}$).
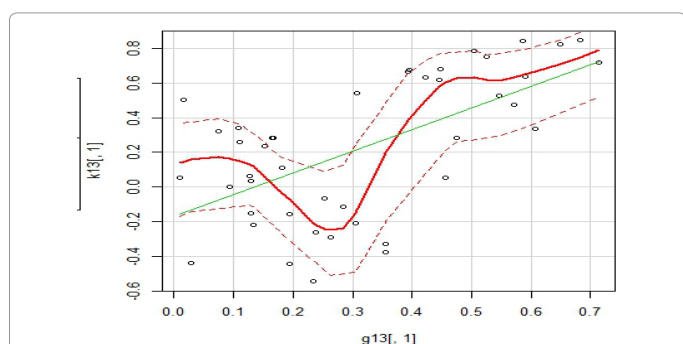
Figure 2: A scatter plot showing correlation between the gene connectivity (K value) of the thirteenth module genes on the X-axis and gene significance of the same module on the X-axis; showing a significantly high positive correlation ($r=0.607$, $p=1.54 \times 10^{-5}$).
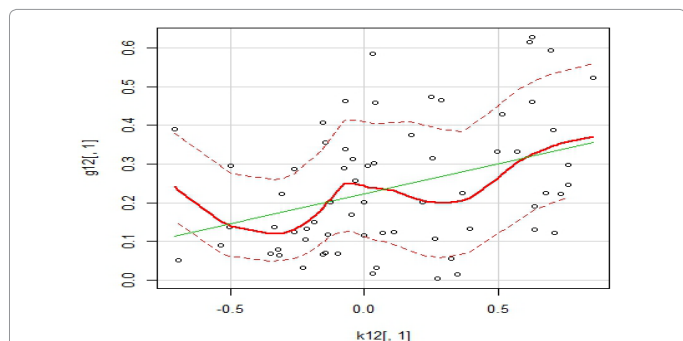
Figure 3: A scatter plot showing correlation between the gene connectivity (K value) of the twelfth module genes on the X-axis and gene significance of the same module on the X-axis; showing a significantly high positive correlation ($r=0.373$, $p=1.9 \times 10^{-3}$).

They were also enriched for DNA recombination ($p=8.0 \times 10^{-4}$) and nuclear mRNA splicing ($p=3.2 \times 10^{-3}$). Meanwhile, the genes of the thirteenth module were significantly enriched for cellular localization ($p=2.4 \times 10^{-2}$) and defense response to virus ($p=3.9 \times 10^{-2}$). Concomitantly, we conducted a pathway enrichment analysis for each module using DAVID database in order to obtain further insight into the functional significance of the gene modules. 'KEGG_PATHWAY' and 'PANTHER_PATHWAY' were selected for pathway enrichment analysis of co-expressed genes compared with the background list of all genes on the array. The ninth module was found to be enriched for genes within Pancreatic cancer pathway ($p=4.1 \times 10^{-2}$), B cell receptor signaling pathway ($p=4.4 \times 10^{-2}$) and Ras Pathway ($p=3.7 \times 10^{-2}$). Results for the functional analysis, those were significant at $p<0.05$

level, are shown in (Tables 2-5). Screening for genes that may have a possible role in the progression and invasiveness of bladder cancer was performed by ranking the gene expressions based on their correlations with the clinical tumor stage. Furthermore, as the three modules eigengenes showed significant correlations with the tumor stage (with p value ≤0.05), genes were ranked by their membership to each module or by their eigengene based connectivity (K) values. Furthermore, genes potentially involved in bladder cancer progression were screened by two methods the first is ranking genes within each significant module by their eigengene based connectivity (K) values. The second is assessing the trait-based gene significance measure correlating the gene expression profile to the tumor stage. Specifically, we used the 80th percentile of each screening variable, which resulted in six genes inside the ninth module and five genes in each of the twelfth and the thirteenth modules (Tables 6-8).

Discussion

Bladder carcinoma is the most common malignancy of the urinary tract. It is associated with high recurrence and mortality rates. The

9th cluster	Biological Process	p-value
1	Organelle organization	1.2E-5
2	mitochondrion organization	8.2E-5
3	regulation of cell death	2.4E-4
4	cellular component organization	4.9E-4
5	regulation of apoptosis	9.9E-4
6	negative regulation of cell death	1.0E-3
7	regulation of programmed cell death	1.1E-3
8	DNA replication	3.3E-3
9	negative regulation of apoptosis	5.7E-3
10	negative regulation of programmed cell death	6.0E-3
11	cellular metabolic process	9.7E-3
12	apoptosis	1.3E-2
13	cellular process	1.3E-2
14	programmed cell death	1.4E-2
15	DNA-dependent DNA replication	1.5E-2
16	negative regulation of cellular process	1.6E-2
17	regulation of cellular component organization	1.6E-2
18	chromosome organization	2.0E-2
19	protein complex assembly	2.4E-2
20	protein complex biogenesis	2.4E-2
21	DNA metabolic process	2.4E-2
22	positive regulation of biological process	2.7E-2
23	response to radiation	2.7E-2
24	cell death	2.8E-2
25	negative regulation of biological process	2.8E-2
26	cellular macromolecular complex subunit organization	2.9E-2
28	anti-apoptosis	2.9E-2
29	response to abiotic stimulus	3.1E-2
30	positive regulation of cellular process	3.2E-2
31	cellular nitrogen compound metabolic process	3.5E-2
32	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.6E-2
33	actin cytoskeleton organization	3.7E-2
Enriched molecular function		
1	protein binding	1.1E-3
2	ribonucleotide binding	8.8E-3
3	purine ribonucleotide binding	8.8E-3
4	purine nucleotide binding	1.2E-2
5	transferase activity, transferring phosphorus-containing groups	2.8E-2

Table 2: Enriched GO terms in the ninth module.

Pathway number	Pathway	Count	Genes	P- value
KEGG pathway hsa05212	Pancreatic cancer	3	BCL2-like 1 , v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog , v-ral simian leukemia viral oncogene homolog A (ras related)	4.1E-2
KEGG pathway hsa04662	B cell receptor signaling pathway	3	glycogen synthase kinase 3 beta , protein tyrosine phosphatase, non-receptor type 6 , v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	4.4E-2
PANTHER pathway P04393	Ras Pathway	3	glycogen synthase kinase 3 beta , v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog , v-ral simian leukemia viral oncogene homolog A	3.7E-2

Table 3: Enriched pathways in the ninth module.

Twelfth Cluster	Biological Process	p-value
1	regulation of mitotic cell cycle	3.4E-4
2	cell cycle checkpoint	6.6E-4
3	DNA recombination	8.0E-4
5	regulation of cell cycle	2.0E-3
6	nuclear mRNA splicing, via spliceosome	3.2E-3
7	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	3.2E-3
8	RNA splicing, via transesterification reactions	3.2E-3
9	cell cycle	3.3E-3
10	response to DNA damage stimulus	3.6E-3
11	chromosome segregation	4.1E-3
12	RNA splicing	5.4E-3
15	mRNA processing	8.9E-3
17	regulation of cell cycle process	1.1E-2
18	mRNA metabolic process	1.6E-2
20	regulation of nuclear division	2.1E-2
21	regulation of mitosis	2.1E-2
22	cellular response to stress	2.5E-2
23	DNA repair	2.6E-2
24	ribonucleoprotein complex assembly	3.1E-2
25	cellular macromolecular complex assembly	3.8E-2
Molecular Function		
1	protein N-terminus binding	3.3E-2
2	purine ribonucleotide binding	4.3E-2
3	ribonucleotide binding	4.3E-2

Table 4: Enriched GO terms in the twelfth module.

Thirteenth cluster	Biological Process	p-value
1	establishment of localization in cell	2.4E-2
2	cellular localization	3.4E-2
3	defense response to virus	3.9E-2
4	regulation of cell communication	4.9E-2
INTERPRO	WD40/YVTN repeat-like	2.0E-2

Table 5: Enriched GO terms in the thirteenth module.

Gene Symbol	Gene	Chr	ENSEMBL ID	Entrez ID	GS	Connectivity
VPS13D	vacuolar protein sorting 13 homolog D	1p36.22	ENSG00000048707	55187	0.540	0.847
ZZEF1	zinc finger, ZZ-type with EF-hand domain 1	17p13.2	ENSG00000074755	23140	0.620	0.671

purA	purine-rich element binding protein A	5q31	ENSG00000185129	5813	0.869	0.793
UBE2I	ubiquitin-conjugating enzyme E2I	16p13.3	ENSG00000103275	7329	0.789	0.772
DCAF8	WD repeat domain 42A	1q22-q23	ENSG00000132716	50717	0.634	0.841
MEIS3P1	Meis homeobox 3 pseudogene 1	17p12	ENSG00000179277	4213	0.716	0.771

Table 6: Gene screening strategy revealed candidate cancer driver genes in the ninth module.

Gene Symbol	Gene	Chr	ENSEMBL ID	Entrez ID	GS	Connectivity
LRRC37A2	leucine rich repeat containing 37, member A2	17q21.32	ENSG00000184525	474170	0.845	0.683584
CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	7q21.1	ENSG00000106258	1577	0.717	0.713264
WDFY3	WD repeat and FYVE domain containing 3	4q21.23	ENSG00000163625	23001	0.822	0.648671
upf3a	UPF3 regulator of nonsense transcripts homolog A	13q34	ENSG00000169062	65110	0.842	0.585826
ATP2A2	ATPase, Ca++ transporting, cardiac muscle, slow twitch2	12q23-q24.1	ENSG00000174437	488	0.753	0.525031

Table 7: Gene screening strategy revealed candidate cancer driver genes in the thirteenth module.

Gene symbol	Gene	Chr	ENSEMBL ID	Entrez ID	GS	Connectivity
SRPK2	SFRS protein kinase 2	7q22-q31.1	ENSG00000135250	6733	0.5940	0.6969
POLR2A	polymerase (RNA) II (DNA directed) polypeptide A, 220kDa	17p13.1	ENSG00000181222	5430	0.6275	0.6258
BCAP31	B-cell receptor-associated protein 31	Xq28	ENSG00000185825	10134	0.6151	0.6169
TRAK1	trafficking protein, kinesin binding 1	3p25.3-p24.1	ENSG00000182606	22906	0.4602	0.6281
BRD2	bromodomain containing 2	6p21.3	ENSG00000204256	6046	0.5224	0.8517

Table 8: Gene screening strategy revealed candidate cancer driver genes in the twelfth module.

basis for bladder cancer development and progression is complex and involves genetic abnormalities, therefore; development of accurate surveillance tests to evaluate disease aggressiveness is a major clinical need Frantzi et al., [26]. Network analysis has been emerged as an attractive approach to decipher the occurrence and progression of complex diseases as it provides means to bridge the gap from individual genes to systems biology Zeng et al., [27]. Weighted gene coexpression network measures correlations among all the gene pairs and extracts modules based on the density of gene connections Liu et al., [28].

Using this approach, demonstrated that the co-expression modules for Hepatitis B virus/Hepatitis C virus -induced hepatocellular carcinoma revealed distinct progression patterns from hepatitis infection to hepatocellular carcinoma He et al., [29]. Therefore, the current study aimed to apply WGCNA for exploring the significant coexpressed modules, genes and the functionally enriched pathways involved in progression of bladder carcinoma from the noninvasive to the invasive phenotype. Several gene co-expression modules in bladder cancer were explored in this study. These modules are clusters of highly correlated genes; this high correlation could be a result of transcriptional co-activation, or the co-regulation of mRNA stability, resulting in a complex genetic network of closely related genes coordinately acting to achieve a specific biological function Zhang et al., [16]. Among these co-expressed modules, 3 physiologically relevant or significant modules were detected, as their module eigengenes are found to be highly correlated with the tumor stage. Furthermore a gene screening strategy was applied for detecting the genetic drivers responsible for the invasiveness of bladder carcinoma, composed of two criteria: the first is the high association with the clinical tumor stage, i.e., high values of gene significance and the second is the value of k, i.e., membership in a tumor stage-related module. In this context, the current study demonstrated that most of the highly connected genes of a module have higher gene significance relationships to the malignant phenotype. Many of the identified genes by this screening method have been previously suggested as associated with tumorigenesis and cancer prognosis. For example, PURA gene encodes a single-stranded DNA-binding protein which binds preferentially to the purine-rich element at origins of replication and in gene flanking regions. Thus, it is implicated in the control of both DNA replication and transcription. Deletion of this gene has been associated with myelodysplastic syndrome and acute myelogenous leukemia Lezon-Geyda et al., [30]. Also, it was found to be upregulated in HCV-related hepatocarcinogenesis De Giorgi et al., [31]. Well in line, SRPK2 gene product belongs to a family of cell cycle regulated protein kinases which phosphorylate Serine/Arginine (SR) domain-containing proteins in nuclear speckles and mediate the pre-mRNA splicing. Of note, reported that SRPK2 gene was over-expresssed in lung adenocarcinoma and was associated with extensive stages which may be attributed to disruption of the splicing machinery Gout et al., [32]. Likewise, TRAK1 gene, encodes trafficking kinesin-binding protein 1, has been recently identified as an emerging novel cancer biomarker. An et al., [33] found that its elevated expression is correlated with poor prognosis in colorectal cancer patients. Noteworthy, BRD2 gene, encodes a transcriptional regulator belonging to the BET (the bromodomain and extra-terminal domain-containing proteins), is characterized by tandem bromodomains that interact with acetylated histones and influence gene expression, cell-cycle regulation, and development Belkina et al., [34]. In addition, BET proteins are associated with chromatin throughout mitosis, and thus facilitate rapid transcriptional reactivation of critical genes after mitosis. Recent studies showed that BET inhibitors inhibit growth in a range of hematopoietic malignant cell lines as well as prostate cancer Wyce et al., [35]. Concomitantly, UPF3 gene products are trans-acting factors of the nonsense-mediated mRNA decay pathway which is a surveillance mechanism that degrades transcripts containing nonsense mutations, preventing the translation of possibly harmful truncated proteins Chan et al., [36]. It has been reported that this pathway may have a role in hereditary gastric cancer by down regulation of the tumour suppressor E-cadherin transcripts in gastric cells Karam et al., [37]. Moreover, GO enrichement analysis of the significant coexpressed modules revealed that these modules are enriched with genes belonging to distinct cellular compartments. This finding is in agreement with previous findings by Wang et al., [38] who reported that proteins within the same compartment tend to be more highly connected. Pathway enrichment analysis of the co-expressed modules revealed activation of Pancreatic cancer pathway, B cell receptor signaling pathway and Ras Pathway. Conceivably, pancreatic cancer pathway, which involves apoptosis regulating genes, is implicated in many types of cancer and Ghosh et al., [39] suggested its major contribution in bladder cancer. Well in line, Arum et al., [40] have reported aberrant activation of B cell receptor signaling pathway in rat bladder carcinoma. Moreover, activated B cell receptor signaling pathway has been also linked to other types of human cancer as chronic lymphocytic leukemia Oppezzo et al., [41]. Concurring with the findings of this study, Ras signaling has been found to be activated in human bladder cancer Boulalas et al., [42], where its mutation cooperates with β-catenin activation to drive bladder tumourigenesis Ahmad et al., [43]. Altered Ras signaling was also cited as a contributing factor to various tumoral and nontumoral pathologies Fernández-Medarde et al., [44]. In conclusion, a gene coexpression network was constructed from DNA microarray gene expression data for human bladder cancer and reported significant genetic drivers and altered signaling pathways that might have a role in cancer initiation and progression. These finding shed light into the molecular mechanisms of bladder cancer, which is of great benefit for defining the driving molecules and pathways appropriate for novel-targeted therapies. However, these results need further experimental validation and can be considered as a starting point for research studies exploring the magnitude of the involvement of these candidate genes in progression of this type of cancer.

Acknowledgment

References

1. Ghafouri-Fard S, Nekoohesh L, Motevaseli E (2014) Bladder cancer biomarkers: review and update. Asian Pac J Cancer Prev 15: 2395-2403.

2. Ploeg M, Aben KK, Kiemeney LA (2009) The present and future burden of urinary bladder cancer in the world. World J Urol 27: 289-293.

3. Cheung G, Sahai A, Billia M, Dasgupta P, Khan MS (2013) Recent advances in the diagnosis and treatment of bladder cancer. BMC Med 11: 13.

4. Biagioli MC, Fernandez DC, Spiess PE, Wilder RB (2013) Primary bladder preservation treatment for urothelial bladder cancer. Cancer Control 20: 188-199.

5. Zhu XL, Ai ZH, Wang J, Xu YL, Teng YC (2012) Weighted gene co-expression network analysis in identification of endometrial cancer prognosis markers. Asian Pac J Cancer Prev 13: 4607-4611.

6. Wang HF, Wang JS (2012) Research progress in potential urinary markers for the early detection, diagnosis and follow-up of human bladder cancer. Asian Pac J Cancer Prev 13: 1723-1726.

7. Schiffer E, Vlahou A, Petrolekas A, Stravodimos K, Tauber R, et al. (2009) Prediction of muscle-invasive bladder cancer using urinary proteomics. Clin Cancer Res 15: 4935-4943.

8. Chin L, Hahn WC, Getz G, Meyerson M (2011) Making sense of cancer genomic data. Genes Dev 25: 534-555.

9. Horvath S, Dong J (2008) Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol 4: e1000117.

10. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559.

11. Kristensen VN, Lingjærde OC, Russnes HG, Vollan HK, Frigessi A, et al. (2014) Principles and methods of integrative genomic analyses in cancer. Nat Rev Cancer 14: 299-313.

12. Stranger BE, Stahl EA, Raj T (2011) Progress and promise of genome-wide association studies for human complex trait genetics. Genetics 187: 367-383.

13. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20: 307-315.

14. Smyth, GK (2005) In: Bioinformatics and Computational Biology Solutions using R and Bioconductor. R Gentleman, Springer, New York.

15. Song L, Langfelder P, Horvath S (2012) Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC Bioinformatics 13: 328.

16. Zhang W, Zang Z, Song Y, Yang H, Yin Q (2014) Co expression network analysis of differentially expressed genes associated with metastasis in prolactin pituitary tumors. Mol Med Rep 10: 113-118.

17. Mason MJ, Fan G, Plath K, Zhou Q, Horvath S (2009) Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. BMC Genomics 10: 327.

18. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4: Article17.

19. Yip AM, Horvath S (2007) Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics 8: 22.

20. Li L, Shiga M, Ching WK, Mamitsuka H (2010) Annotating gene functions with integrative spectral clustering on microarray expressions and sequences. Genome Inform 22: 95-120.

21. Huang GT, Cunningham KI, Benos PV, Chennubhotla CS (2011) A Genetic Spectral Clustering Algorithm. Journal of Computational Information Systems 7: 3245-3252.

22. Karatzoglou A, Smola A, Hornik K, Zeileis A (2004) kernlab-An S4 Package for Kernel Methods in R. Journal of Statistical Software 11: 1-20.

23. Xiao X, Moreno-Moral A, Rotival M, Bottolo L, Petretto E (2014) Multi-tissue analysis of co-expression networks by higher-order generalized singular value decomposition identifies functionally coherent transcriptional modules. PLoS Genet 10: e1004006.

24. Min JL, Nicholson G, Halgrimsdottir I (2012) Coexpression network analysis in abdominal and gluteal adipose tissue reveals regulatory genetic loci for metabolic syndrome and related phenotypes. PLoS Genet 8: e1002505.

25. Puniya BL, Kulshreshtha D, Verma SP, Kumar S, Ramachandran S (2013) Integrated gene co-expression network analysis in the growth phase of Mycobacterium tuberculosis reveals new potential drug targets. Mol Biosyst 9: 2798-2815.

26. Frantzi M, Makridakis M, Vlahou A (2012) Biomarkers for bladder cancer aggressiveness. Curr Opin Urol 22: 390-396.

27. Zeng T, Sun SY, Wang Y, Zhu H, Chen L (2013) Network biomarkers reveal dysfunctional gene regulations during disease progression. FEBS J 280: 5682-5695.

28. Liu R, Wang X, Aihara K, Chen L (2014) Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med Res Rev 34: 455-478.

29. He D, Liu ZP, Honda M, Kaneko S, Chen L (2012) Co-expression network analysis in chronic hepatitis Band C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. J Mol Cell Biol 4: 140-152.

30. Lezon-Geyda K, Najfeld V, Johnson EM (2001) Deletions of PURA, at 5q, and PURB, at 7p1, in myelodysplastic syndrome and progression to acute myelogenous leukemia. Leukemia 15: 954-962.

31. De Giorgi V, Monaco A, Worchech A, Tornesello M, Izzo F, et al. (2009) Gene profiling, biomarkers and pathways characterizing HCV-related hepatocellular carcinoma. J Transl Med 7: 85.

32. Gout S, Brambilla E, Boudria A, Drissi R, Lantuejoul S, et al. (2012) Abnormal expression of the pre-mRNA splicing regulators SRSF, SRSF, SRPK1 and SRPK2 in non small cell lung carcinoma. PLoS One 7: e46539.

33. An Y, Zhou Y, Ren G, Tian Q, Lu Y, et al. (2011) Elevated expression of MGb2-Ag/TRAK1 is correlated with poor prognosis in patients with colorectal cancer. Int J Colorectal Dis 26: 1397-1404.

34. Belkina AC, Blanton WP, Nikolajczyk BS, Denis GV (2014) The double bromodomain protein Brd2 promotes B cell expansion and mitogenesis. J Leukoc Biol 95: 451-460.

35. Wyce A, Degenhardt Y, Bai Y, Le B, Korenchuk S, et al. (2013) Inhibition of BET bromodomain proteins as a therapeutic approach in prostate cancer. Oncotarget 4: 2419-2429.

36. Chan WK, Bhalla AD, Le Hir H, Nguyen LS, Huang L, et al. (2009) A UPF3-mediated regulatory switch that maintains RNA surveillance. Nat Struct Mol Biol 16: 747-753.

37. Karam R, Carvalho J, Bruno I, Graziadio C, Senz J, et al. (2008) The NMD mRNA surveillance pathway downregulates aberrant E-cadherin transcripts in gastric cancer cells and in CDH1 mutation carriers. Oncogene 27: 4255-4260.

38. Wang J, Peng X, Peng W, Wu FX (2014) Dynamic protein interaction network construction and applications. Proteomics 14: 338-352.

39. Ghosh M, Brancato SJ, Agarwal PK, Apolo AB (2014) Targeted therapies in urothelial carcinoma. Curr Opin Oncol 26: 305-320.

40. Arum CJ, Anderssen E, Tømmerås K, Lundgren S, Chen D, et al. (2010) Gene expression profiling and pathway analysis of superficial bladder cancer in rats. Urology 75: 742-749.

41. Oppezzo P, Dighiero G (2013) Role of the B-cell receptor and the microenvironment in chronic lymphocytic leukemia. Blood Cancer J 3: e149.

42. Boulalas I, Zaravinos A, Karyotis I, Delakas D, Spandidos DA (2009) Activation of RAS family genes in urothelial carcinoma. J Urol 181: 2312-2319.

43. Ahmad I, Patel R, Liu Y, Singh LB, Taketo MM, et al. (2011) Ras mutation cooperates with β-catenin activation to drive bladder tumourigenesis. Cell Death Dis 2: e124.

44. Fernández-Medarde A, Santos E (2011) Ras in cancer and developmental diseases. Genes Cancer 2: 344-358.