Research Paper

Identification of cancer prognosis-associated functional modules using differential co-expression networks

Wenshuai Yu[1], Shengjie Zhao[1,2], Yongcui Wang[3], Brian Nlong Zhao[4], Weiling Zhao[5] and Xiaobo Zhou[6,7,8]

[1]Key Laboratory of Embedded System and Service Computing, College of Electronics and Information Engineering, The Ministry of Education, Tongji University, Shanghai, China

[2]College of Software Engineering, Tongji University, Shanghai, China

[3]Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, China

[4]Shanghai High School International Division, Shanghai, China

[5]Department of Radiology and Comprehensive Cancer Center, Wake Forest University School of Medicine, Winston Salem, NC, USA

[6]College of Electronics and Information Engineering, Tongji University, Shanghai, China

[7]Center for Big Data Sciences and Network Security, Tongji University, Shanghai, China

[8]Center for Bioinformatics and System Biology, Wake Forest University School of Medicine, Winston Salem, NC, USA

Correspondence to: *Shengjie Zhao,* ***email:*** *shengjiezhao@tongji.edu.cn*
Xiaobo Zhou, ***email:*** *zhouxb2015@163.com*

Keywords: *co-expression network; prognosis; HO-GSVD; gene module; cancer*

Abbreviations: DAVID: the Database for Annotation, Visualization, and Integrated Discovery; TCGA: The Cancer Genome Atlas; KEGG: Kyoto Encyclopedia of Genes and Genomes

Received: *July 12, 2017* ***Accepted:*** *November 15, 2017* ***Published:*** *December 04, 2017*

ABSTRACT

The rapid accumulation of cancer-related data owing to high-throughput technologies has provided unprecedented choices to understand the progression of cancer and discover functional networks in multiple cancers. Establishment of co-expression networks will help us to discover the systemic properties of carcinogenesis features and regulatory mechanisms of multiple cancers. Here, we proposed a computational workflow to identify differentially co-expressed gene modules across 8 cancer types by using combined gene differential expression analysis methods and a higher-order generalized singular value decomposition. Four co-expression modules were identified; and oncogenes and tumor suppressors were significantly enriched in these modules. Functional enrichment analysis demonstrated the significantly enriched pathways in these modules, including ECM-receptor interaction, focal adhesion and PI3K-Akt signaling pathway. The top-ranked miRNAs (mir-199, mir-29, mir-200) and transcription factors (*FOXO4*, *E2A*, *NFAT*, and *MAZ*) were identified, which play an important role in deregulating cellular energetics; and regulating angiogenesis and cancer immune system. The clinical significance of the co-expressed gene clusters was assessed by evaluating their predictability of cancer patients' survival. The predictive power of different clusters and subclusters was demonstrated. Our results will be valuable in cancer-related gene function annotation and for the evaluation of cancer patients' prognosis.

INTRODUCTION

The rapid accumulation of cancer-related data owing to high-throughput technologies has provided unprecedented choices to understand the progression of cancer and discover functional networks in multiple cancers. The vast majority of cancer-related studies have focused on a single cancer, but always ignored the common traits across different cancer types. Different cancers usually share common hallmarks, such as evading growth suppressors, resisting cell death and inducing angiogenesis. Moreover, the methods based on biological networks including gene co-expression networks, metabolic networks, protein-protein interaction networks and genetic regulatory networks can infer regulatory mechanisms related to biological processes. The network-based methods to search biological processes related to cancer hallmarks will help us in identifying the characterizations of tumor biology. Few studies focus on genome-scale networks across different cancer types. Thus, the network analysis may help us to unveil common traits involved in multiple cancers.

The majority of the network-based methods are performed to identify distinct patterns within one cancer [1, 2]. Compared with only analyzing one cancer, several methods have been used to identify common patterns shared by two or more cancers. Zhang et al. used a network mining algorithm to build tightly connected gene co-expression networks from the microarray datasets spanning 33 cancer types. Their results indicate that the commonly recognized characteristics of cancers are supported by highly coordinated transcriptomic activities [3]. Yang et al. did the weighted correlation network analysis (WGCNA) to highlight common properties of prognostic genes in four cancer types [4]. Li et al. described a network method to analyze the driver mutation by integrating both cancer genomes and transcriptomes and identified a significant correlation of genotype to phenotype in six solid tumors [5].

Several computationally efficient methods have been developed for construction of the networks, such as Generalized Singular Value Decomposition (GSVD) [6] and higher-order GSVD (HO-GSVD) [7]. GSVD is used for identifying common structures across two conditions [6]. The analysis based on HO-GSVD can extract common gene modules in two or more conditions [7–9]. Ponnapalli et al. first proposed HO-GSVD to analyze common structures shared by multiple datasets [7]. Xiao et al. developed a new HO-GSVD method for analyzing common and tissue-specific modules from seven rat tissues and four human brain regions [8]. Wang et al. applied a simple mathematical framework of HO-GSVD for analysis of multiple tissues [9]. These studies indicate that HO-GSVD is a valuable method in common gene pattern discovery among different tissues. Motivated by this approach, we applied HO-GSVD to pan-cancer analysis in this study. To our best of knowledge, this approach has not been used for pan-cancer research.

Differential gene expression analysis has been widely used for identifying differentially expressed genes between conditions. The commonly used methods include edgeR [10], limma [11], SAMseq [12] and DESeq [13]. Most studies only use one method to analyze gene expression patterns. Soneson et al. compared the commonly used methods for differential expression analysis and found that no single method is optimal under all circumstances and the method of choice in a particular situation depends on the experimental conditions [14].

In this study, we applied the above four methods for differential expression analysis to get common genes in eight types of cancers through a three-step procedure. First, the differentially expressed gene set shared by four caner types was selected using one method. Second, the commonly expressed gene set selected by four methods was set as a candidate gene set. We then converted candidate gene IDs to the corresponding DAVID gene IDs and removed the non-mapped genes. The gene expression matrices of multiple cancers were decomposed by the HO-GSVD method for identifying the common modules across different cancers. We chose the vectors with top eigenvalues in the right basis matrix as candidates of co-expression genes. The co-expression genes were selected based on the assumption that a small proportion of genes in candidate vectors is highly similar. We used the DAVID tools to validate the functional significance of the modules. The gene modules involved in the significant pathways were retained for further analysis. Multiple types of enrichment analysis were performed, including gene ontology terms, KEGG pathways, cell type enrichment, disease association and miRNA and transcription factor enrichment analysis [15–17]. The functional interaction networks were constructed for the identified modules. The survival analysis was then applied for the prognostic properties of modules across cancers. By following the procedure shows in the Figure 1, we found that the co-expression gene modules are enriched with oncogenes and tumor suppressors, which play an important regulatory role in multiple cancers. The genes in the main- and sub-modules are closely associated with the prognosis of multiple cancers.

RESULTS

Identifying co-expression gene modules

We analyzed differentially expressed genes using the raw count data and constructed co-expression networks using the FPKM count data. 5229 differentially expressed genes were detected in breast invasive carcinoma (BRCA), kidney renal clear cell carcinoma (KIRC), prostate adenocarcinoma (PRAD), thyroid carcinoma (THCA), lung adenocarcinoma (LUAD), bladder urothelial carcinoma (BLCA), colon adenocarcinoma (COAD) and stomach adenocarcinoma (STAD) using limma, edgeR, DESeq, and SAMseq methods (Figure 2A). 4973 genes were used for construction of co-expression networks and pathway enrichment analysis. Five clusters with significant pathways were identified (Supplementary

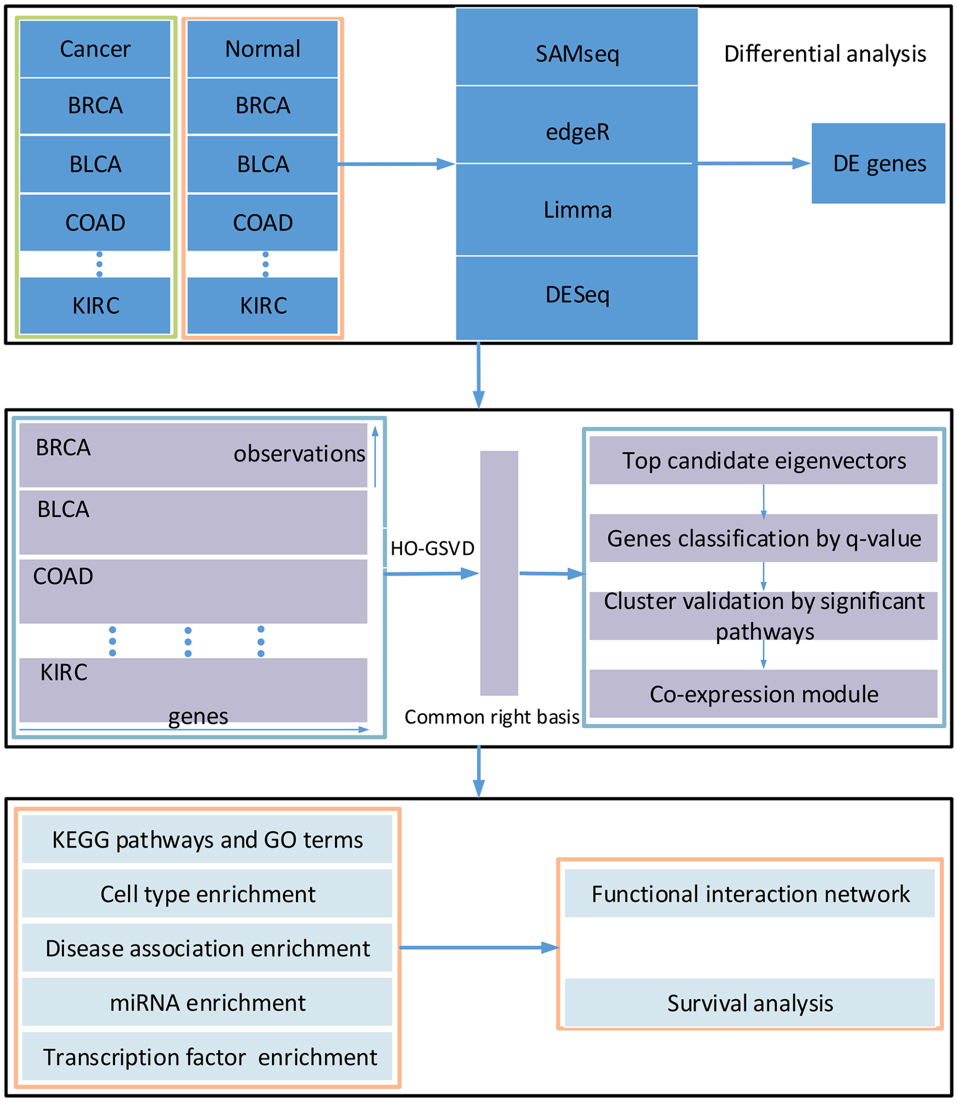

Figure 1: Overview of the workflow. There are three main steps including gene differential expression analysis, identification of co-expression modules and significant enrichment analysis.

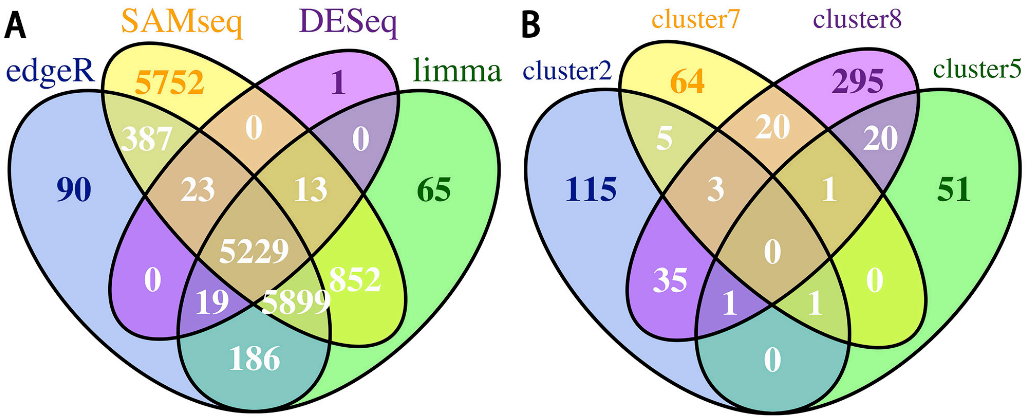

Figure 2: Identification of differentially expressed genes and co-expression modules. **(A)** Venn diagram showing the overlap between differentially expressed genes selected by the four methods. **(B)** Venn diagram showing the overlap between the genes in the four co-expression modules.

Table 1). The significant pathways enriched in the smallest cluster 10 were almost identical to the pathways in the cluster 5. Then five clusters except cluster 10 were kept for further analysis. The highly specific genes were included in each cluster. But there were a few genes shared by these clusters (Figure 2B). The results showed that these co-expression modules with specific genes may share similar biological information. To comprehensively investigate the characterization of the identified co-expression modules, we applied multiple types of enrichment analysis.

We applied functional enrichment analysis to identify the KEGG pathways (Figure 3A–3D) and Gene Ontology terms (Supplementary Table 2) in the four clusters. For cluster 5, there are only six significant KEGG pathways as shown in the Figure 3B. ECM-receptor interaction, focal adhesion and PI3K-Akt signaling pathways were presented in the cluster 8 and cluster 5, consistent with the previous results obtained by pan-cancer analysis [18]. Moreover, these pathways are related to the deregulation of cellular energetics [19]. Abnormal ECM affects cancer progression by directly promoting cellular transformation and metastasis [20]. Focal

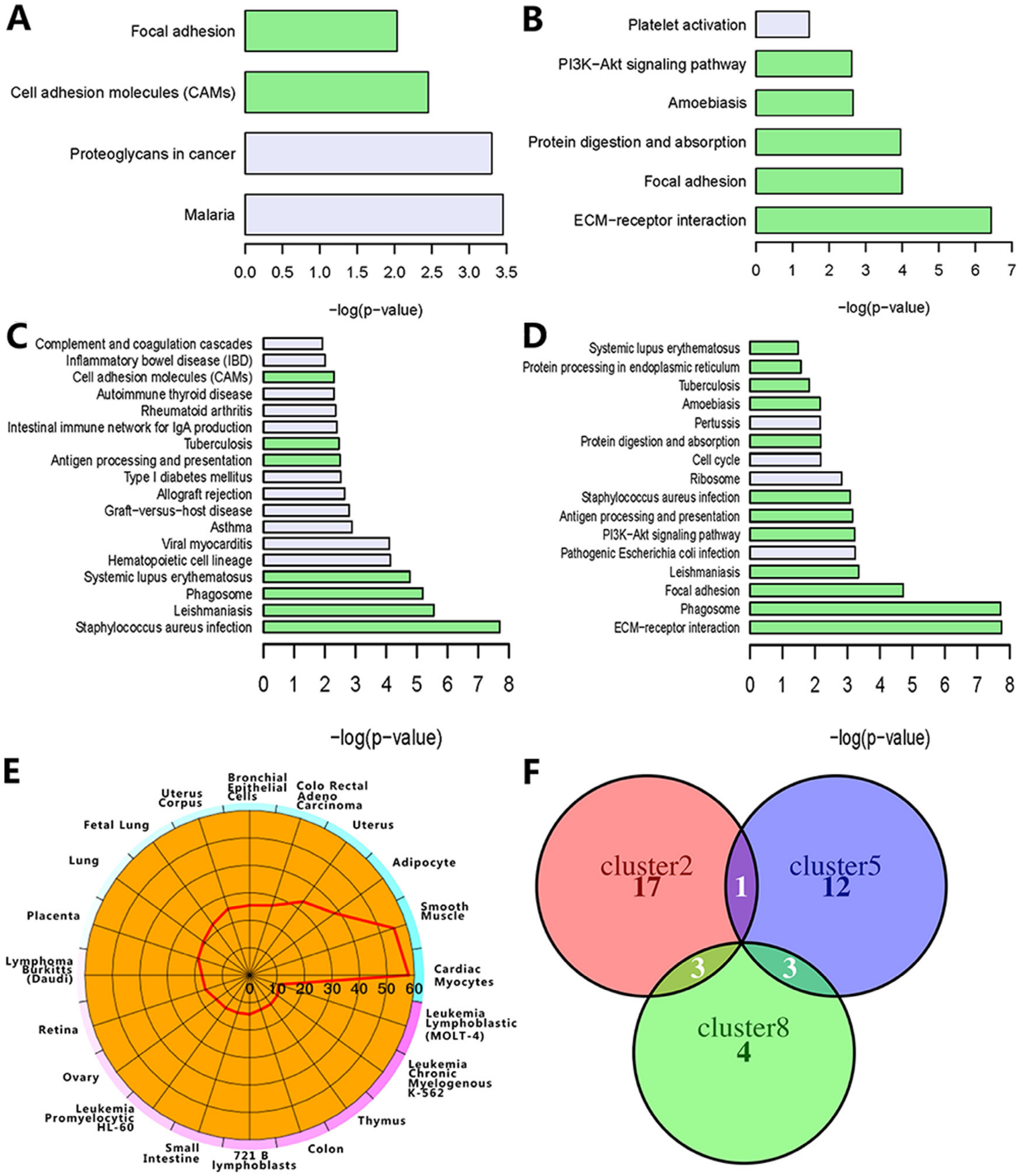

Figure 3: Enrichment of co-expression modules. The four bar charts display the pathway enrichment results of cluster 2 **(A)**, cluster 5 **(B)**, cluster 7 **(C)** and cluster 8 **(D)**. The pathway shared by at least two cluster are colored with light green. The cell-type enrichment analysis of cluster 8 **(E)** and other clusters (Supplementary Figure 2) is shown as –log10 (Benjamini-Hochberg corrected p-value). The overlap of the members in the top 5-ranked miRNA families is shown as the Venn diagram **(F)**.

adhesion kinase, a protein tyrosine kinase, regulates cellular adhesion, motility, proliferation and survival in cancer cells, thereby promoting cancer progression and metastasis [21]. PI3K-Akt signaling pathway has been reported as one of the most important pathways in cancer metabolism and growth [22]. Figure 3C and Figure 3D show the enriched pathways in the cluster 7 and 8, respectively. Only focal adhesion was shared by the cluster 2, 5 and 8 (Figure 3A, 3B, 3D). These results indicated that the enriched pathways in the clusters are associated with tumorigenesis. To further evaluate whether the functional features in these clusters are also important in other cancers, we did analysis for another four cancers, including uterine corpus endometrial carcinoma (UCEC), head and neck squamous cell carcinoma (HNSC), rectum adenocarcinoma (READ) and liver hepatocellular carcinoma (LIHC) (Supplementary Figure 1). 8619 genes were identified. We converted the candidate gene IDs to the corresponding DAVID gene IDs. If two IDs were corresponding to the same gene, we removed the redundant one. Therefore, 490 redundant genes were removed and 8129 genes used for further analysis. We got two clusters with significant pathways (Supplementary Table 3) using HO-GSVD approach. We found that cluster 6 in the four cancers was highly enriched in focal adhesion $(p_6 = 5.7\text{e}-11)$, ECM-receptor interaction $(p_6 = 2.6\text{e}-10)$ and PI3K-Akt signaling $(p_6 = 2.7\text{e}-5)$ pathways. For the eight cancers, these clusters partially shared some KEGG pathways. We observed some similar patterns in the following results of the GO biological process (BP) enrichment analysis. The top-ranked five enriched BPs in the cluster 5 included collagen fibril organization $(p_5 = 3.5\text{e}-6)$, extracellular matrix organization $(p_5 = 4.5\text{e}-6)$, collagen catabolic process $(p_5 = 4.9\text{e}-5)$, skeletal system development $(p_5 = 3.2\text{e}-3)$ and cell adhesion $(p_5 = 1.3\text{e}-2)$. The top three enriched BPs in the cluster 8 were the same as those in the cluster 5. It was reported that the collagen fibrils reorganization within an extracellular matrix facilitated tumorigenesis and invasion [23]. We found that the BPs in the cluster 7 are mainly involved in T cell and immune response [24]. The highly enriched BPs in the cluster 2 are related to angiogenesis. In addition to cluster 7, these commonly enriched BPs are associated with extracellular matrix organization and cell adhesion. Similarity to the KEGG pathways, the corresponding enriched BPs in these clusters also play an important role in cancers. As expected, all the above results indicated that the identified clusters with biological functions related to cellular energetics, angiogenesis and anti-cancer immunity are important to the cancer development.

To further explore the cell specificity and disease association of the differentially expressed genes, we analyzed cell type enrichment (Cten) and disease association for the genes in four clusters (Supplementary Tables 4, 5). Enriched cell types in these clusters were closely associated with all eight cancer types, including BRCA, KIRC, PRAD, THCA, LUAD, BLCA, COAD, STAD. The top three cell types for the cluster 8 included cardiac myocytes (score as –log10 (Benjamini-Hochberg corrected p-value), score = 57), smooth muscle (score = 55) and adipocyte (score = 38) (Figure 3E). Adipocytes can support tumorigenesis by secreting adipokines and producing energy [25]. The immune cells in the cluster 7 and cluster 8 included CD14+ Monocytes, CD33+ Myeloid and BDCA4+ Dentritic Cells. CD14+ Monocytes (score = 4.88 in the cluster 7 and score = 7.81 in cluster 8) has been reported to affect dendritic cell differentiation and T-cell function in cancer patients [26]. These results indicated that the enriched cell types in the clusters were associated with tumor progression. Moreover, disease association analysis suggested that the genes in the four clusters were mainly enriched in cancer-related diseases with high significance (Supplementary Table 5, $p \le 8.39\text{e}-8$). All the above results further suggested that the four clusters which play important roles in biological processes related to tumor progression and immunity are associated with multiple cancers.

To further demonstrate the potential regulatory mechanisms, we also analyzed miRNAs and transcription factors using the WebGestalt tool (Supplementary Tables 6, 7). Three of the top ranked miRNA families were identical in the cluster 2, 5 and 8, including mir-199, mir-29 and mir-200 (Figure 3F). Mir-199a was only presented in the cluster 2 and cluster 5 $(p_2 = 0.004,\ p_5 = 0.0035)$. Mir-29a, mir-29b and mir-29c $(p_5 = 0.0009,\ p_8 = 5.88\text{e}-5)$ in the cluster 5 and 8 are the mature members of the mir-29 family. Three of the main mir-200 family members $(p_2 = 0.0004,\ p_8 = 5.88\text{e}-5)$ in the cluster 2 and 8 have been reported to associate with tumor progression [27, 28]. The let-7 family of miRNAs $(p_5 = 0.0009)$ in the cluster 5 is involved in tumorigenesis [29]. We also found that some clusters were high significantly enriched with transcription factors. *FOXO4* gene was presented in the cluster 2, 5 and 7 $(p_2 = 2.82\text{e}-9,\ p_5 = 4.29\text{e}-5,\ p_7 = 0.0027)$ [30]. The clusters 2, 5 and 8 shared three transcription factors, including *E2A* $(p_2 = 3.66\text{e}-8,\ p_5 = 1.06\text{e}-9,\ p_8 = 7.38\text{e}-15)$, *NFAT* $(p_2 = 8.79\text{e}-11,\ p_5 = 2.38\text{e}-5,\ p_8 = 3.51\text{e}-17)$, and *MAZ* $(p_2 = 2.47\text{e}-8,\ p_5 = 2.40\text{e}-6,\ p_8 = 2.01\text{e}-15)$. *E2A*, *FOXO4*, *NFAT* and *MAZ* play important roles in tumor cell growth and metastasis in multiple cancers, such as colorectal cancer, breast cancer, prostate cancer and lung cancer [31–33]. Taken together, these results revealed that common regulators shared by the four clusters may cooperate with each other to regulate the biological processes of multiple cancers.

Functional network analysis of individual co-expression modules

To further analyze the characterization of single identified modules, we constructed the functional interaction networks. These networks were built based on human PPIs, fly PPIs, worm PPIs, yeast PPIs, domain interaction, Lee's Gene Expression, Prieto's Gene Expression, GO BP sharing and PPIs from GeneWays

[34]. We found that the corresponding proportions of the genes linked to the functional interaction (FI) networks in the four modules were 80.63%, 83.78%, 76.60% and 87.20%, respectively (Supplementary Table 1). This result indicated that the four clusters were highly conserved at the functional protein level. According to the results from the above subsections, all clusters were closely associated with the eight cancers. Among of them, most of the enriched pathways in the cluster 5 were consistent with the previous pan-cancer outcomes [18], therefore, we did further functional analysis of the cluster 5-based FI network as shown in the Figure 4.

The FI network was clustered into small modules and eight modules were annotated as M1~M8 (Supplementary Table 8). We performed the enrichment analysis for GO BP and found that some high significantly BPs (FDR < 0.001) were consistent with the above results. Extracellular matrix disassembly, collagen catabolic process, extracellular matrix organization, collagen fibril organization, skeletal system development and cell adhesion were enriched in the M2. The BPs in the M7 are associated with the negative regulation of extracellular matrix disassembly and endothelial cell migration. Endothelial cell migration has been reported to contribute the entry of cancer cells into the circulatory system [35]. There were two BPs in the M5, including O-glycan processing and protein O-linked glycosylation. It has been reported that alterations in glycosylation impacted cell cycle and may support neoplastic progression [36]. Interestingly, M1 had only one enriched pathway, that is, the notch signaling pathway. Notch signaling pathway plays an important in regulating stem cell self-renewal and the pathogenesis of breast cancer [37]. The most enriched pathway in the M2 was ECM-receptor interaction. Phenylalanine metabolism was the top enriched pathway in the M3. Nicotinic acetylcholine receptor signaling pathway in M4 and amino acid metabolism pathway in M6 are associated with cancer growth [38–40]. These results indicated that distinct BPs enriched in different sub-modules contribute to the development of cancers.

We then defined an individual gene in the modules with at least ten neighbors as one hub gene, and obtained 20 hub genes, including 15 linker genes and 5 module genes. Over half of the hub genes were enriched in the M2, including *FN1*, *COL1A1*, *COL1A2*, *COL3A1* and *COL6A3*. *FN1* is a FDA-approved drug target gene against cancer [41, 42]. *COL1A1*, *COL1A2*, *COL3A1* and *COL6A3* are members of the collagen family. These five

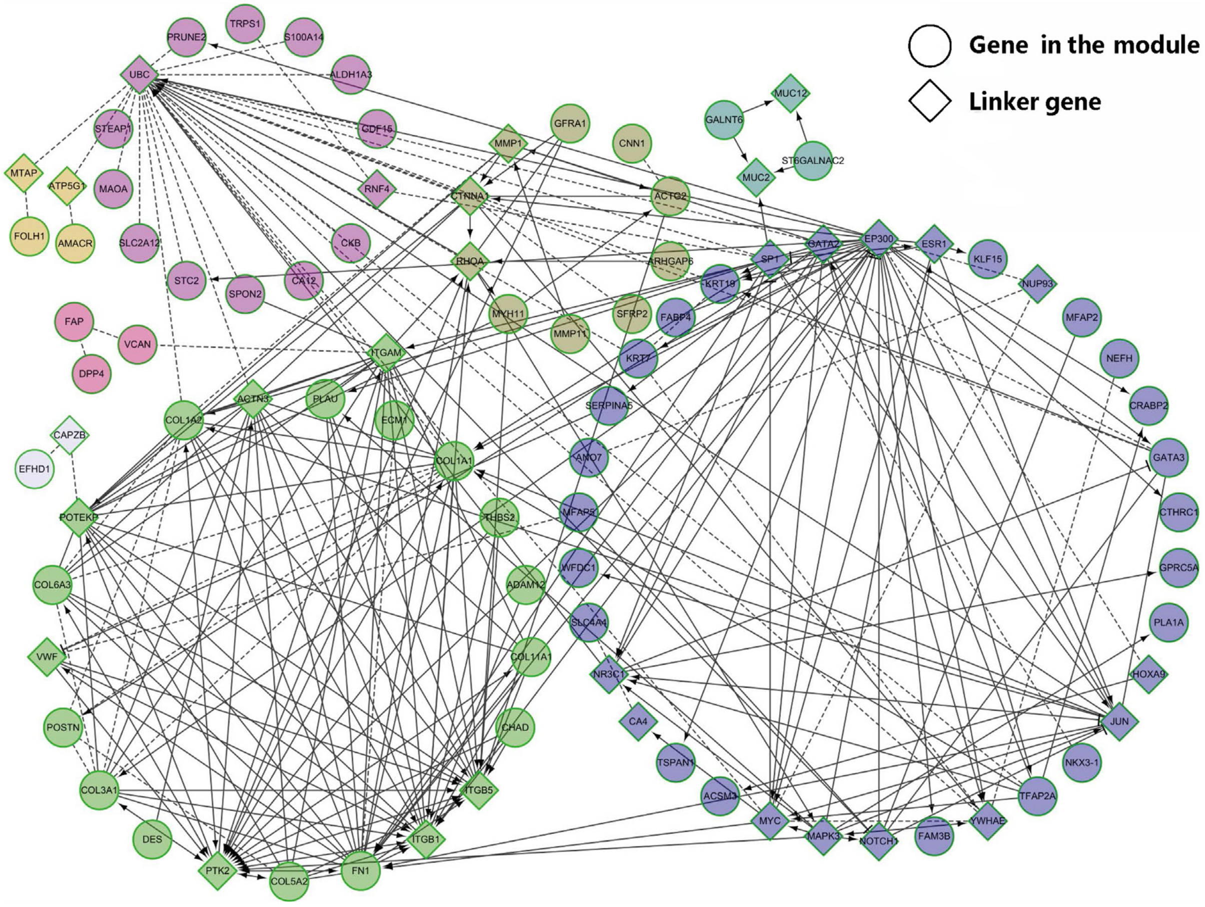

Figure 4: Network visualization of the cluster 5. The functional interaction network consists of eight sub-modules marked with different colors. The genes and link genes in the modules are represented as circles and diamonds, respectively.

genes were enriched in pathways such as ECM-receptor interaction, protein digestion and absorption, integrin signaling pathway and PI3K-Akt signaling pathway. *DES*, *THBS2*, *PLAU* and *POSTN* in the M2 have been associated with multiple cancers [43–50]. *DES* encodes a muscle-specific class III intermediate filament and is related with colorectal cancer, breast cancer, prostate cancer, kidney cancer and lung cancer [43–46]. As the member of *THBS* family, *THBS2* plays an important role in cancer progression [47]. *PLAU* is involved in cancer cell migration [48]. The protein encoded by *POSTN* has been reported to function in cancer stem cell [49, 50].

Survival analysis of gene co-expression modules

To further test the clinical significance of the co-expressed gene clusters, we did survival analysis for the eight types of cancers. The patients were clustered into

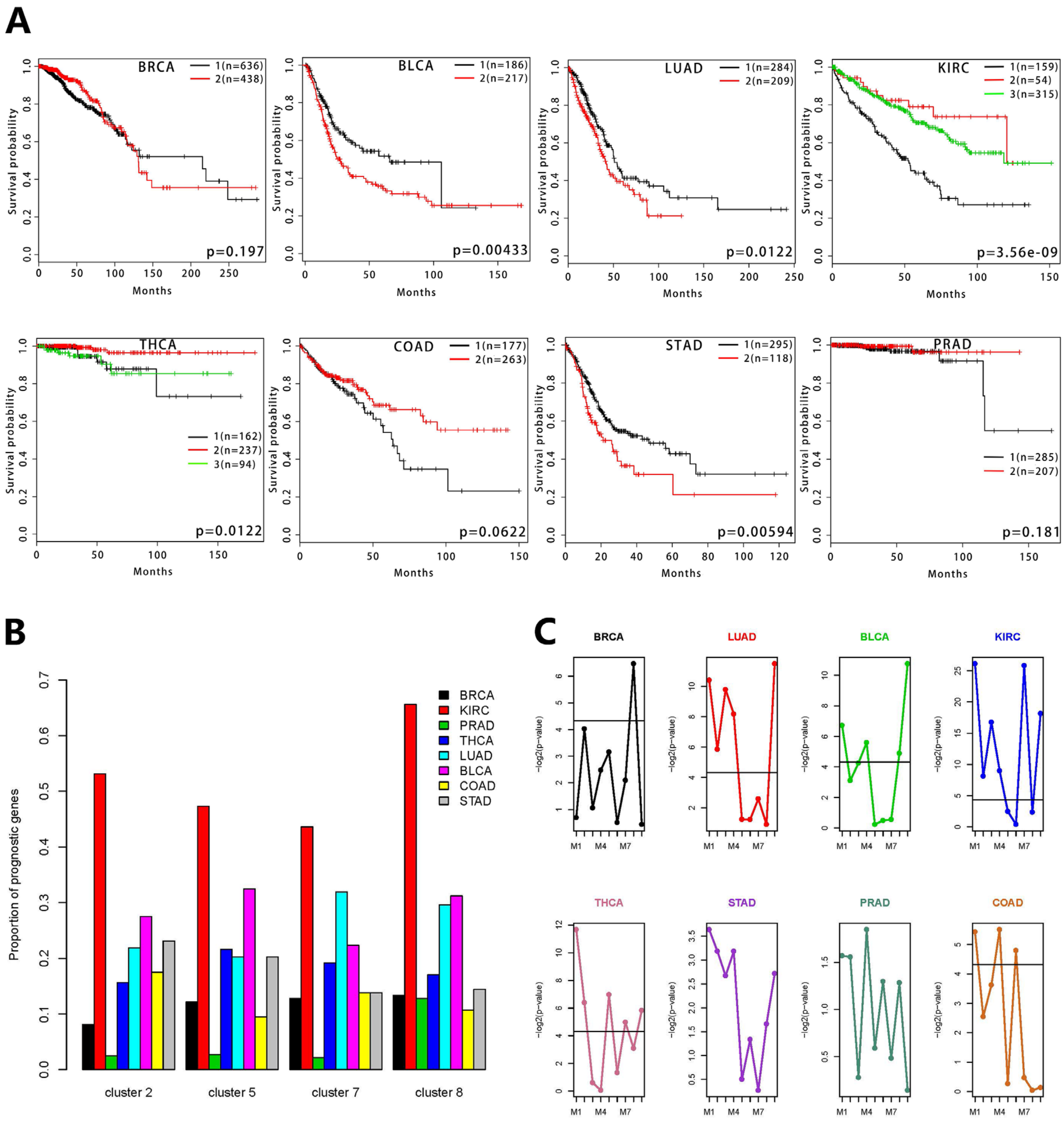

Figure 5: Survival analysis of gene co-expression modules for 8 types of cancers. (A) Survival analysis of the cluster 5 using Kaplan-Meier curves. The calculation of the log-rank p-values is based on the split of patient groups using non-negative matrix factorization. The number of patients in each group is also labeled in each panel. **(B)** The distributions of prognostic genes for the eight cancer types in each cluster. The y-axis represents the gene proportion of each cancer in the corresponding cluster. The eight cancers are marked with different colors. **(C)** The differential significance distributions of nine sub-modules for each cancer. The y-axis represents the log-rank p-value and the x-axis is annotated with the sub-modules grouped by the network clustering. The solid line represents $-\log_2(0.05)$.

different groups using non-negative matrix factorization (NMF). The cluster 2 and cluster 5 were able to predict patient survival of various cancers, including KIRC, THCA, LUAD, BLCA, COAD and STAD (Figure 5A, Supplementary Figure 3). The cluster 7 and cluster 8 could predict survival of LUAD, KIRC and BLCA cancer patients (Supplementary Figures 4, 5). The cox models were used to identify prognostic genes in various cancers. The distinct proportions of prognostic genes in the eight cancers were observed in the four clusters (Figure 5B). Among all of the clusters, KIRC contains the largest proportion of prognostic genes. We also carried out the survival analysis on the eight sub-modules in the cluster 5, as well as the module 9 containing the genes that weren't connected to the networks in the cluster 5 (Supplementary Tables 8, 9). The patients were grouped using the k-means method when the number of genes in the sub-module was one. In the survival analysis of individual sub-modules, all sub-modules showed statistically significant differences in survival probabilities (Figure 5C). M1, M2, M4 and M9 were significantly associated with the patients' survival in at least three cancers. But only one sub-module showed a significant difference in the survival analysis of BRCA and none of the sub-modules was significant in PRAD. Analysis of STAD showed that individual sub-modules couldn't predict these patients' survival but the corresponding cluster 5 had the predictive power. The results of the sub-modules in the survival analysis were mostly consistent with the cluster 5. All these results revealed that the sub-modules associated with known biological processes cooperate with each other to contribute to the prognosis in multiple cancers.

In order to further validate the prognosis of cluster 5 in multiple cancers, we performed survival analysis on HNSC, READ, LIHC and UCEC. The results showed that the cluster 5 was able to predict patient survival in HNSC, LIHC and UCEC (Figure 6A), but not in READ. The corresponding sub-modules obtained from the functional network analysis had statistically significant differences in survival analysis of the four cancers (Figure 6B). M2 had predictive power for patient survival in three cancers. We also applied survival analysis on other eight cancers, including glioblastoma multiforme (GBM), brain lower grade glioma (LGG), ovarian serous cystadenocarcinoma (OV), skin cutaneous melanoma (SKCM), adrenocortical carcinoma (ACC), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), kidney renal papillary cell carcinoma (KIRP) and lung squamous cell carcinoma (LUSC). As shown in the Supplementary Figure 6, the cluster 5 can predict patients' survival of LGG, SKCM, ACC and KIRP cancers. These results confirmed the prognostic ability of cluster 5 in 7 cancer types. We also applied survival analysis on the above twelve cancer types for the cluster 2, 7 and 8. All of the three clusters can predict patients' survival of LIHC, UCEC, LGG, SKCM, ACC and KIRP cancers. Additionally, cluster 2 can predict the survival of HNSC and CESC cancer patients.

DISCUSSION

In this study, we applied four methods to analyze the data and obtained 4973 differentially expressed genes shared by the eight cancers. We constructed the co-expression network using HO-GSVD and identified

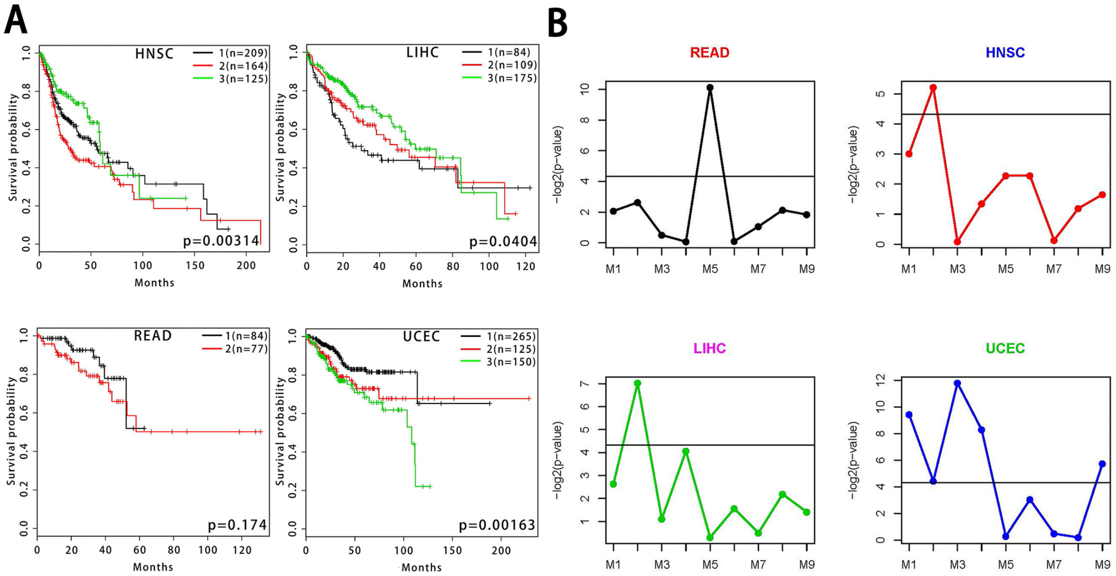

Figure 6: Survival analysis of gene co-expression modules for HNSC, LIHC, READ and UCEC. **(A)** Survival analysis of the cluster 5 using Kaplan-Meier curves for HNSC, LIHC, READ and UCEC. **(B)** The differential significance distributions of nine sub-modules for each cancer.

modules and associated pathways. Some of these pathways have been reported in the pan-cancer analysis [18], such as focal adhesion, PI3K-Akt signaling pathway, ECM-receptor interaction and Systemic lupus erythematosus. We constructed the functional interaction networks for further analyzing individual co-expression modules. These sub-modules were enriched with different pathways and cooperated with each other across cancers. We found that these modules are associated with patients' survival in multiple cancers. In addition, the individual sub-modules in various cancers had different prognostic capability.

Gene differential expression analysis is a commonly used method for identifying differentially expressed genes without considering the relationship between genes. There is no consensus on the best method for differential expression analysis [14, 51]. Limma, edgeR, DESeq, and SAMseq have been commonly used for gene differential expression analysis. Each of them has advantages and disadvantages. The limma method, based on linear models and the voom transformation, is developed for analyzing RNA-seq data [52]. The negative binomial model and empirical Bayes methods in edgeR are used to detect differential gene expression, and then the gene-wise dispersions is estimated by conditional maximum likelihood [10]. DESeq models the count data using the negative binomial distribution and estimates the mean-variance relationship of each gene. In contrast to edgeR, it allows a widely data-driven parameter in the statistical test [13]. SAMseq, a non-parametric method, uses the Wilcoxon rank statistic and resampling procedure to identify differential expressed genes [12]. edgeR becomes liberal for small sample sizes with default settings to a certain extent and keeps a better balance between speed and accuracy than DESeq [14, 51]. Limma and DESeq are described as the safest choices in some cases in terms of the consistency of differentially expressed genes when analyzing the complete data. SAMseq can identify a large number of genes that are usually not detected with the other methods [53]. The combination of these methods can compensate for the shortcomings of a single method. The results of differential expression analysis obtained using these four methods are robust. We chose the co-expression networks to analysis cancer-related genes on the system level. Then we applied the simple framework of HO-GSVD to identify co-expression modules, which has been proved to be a simple, parameter-free and reproducible method. The results in the functional enrichment analysis showed the all the identified modules may play regulatory roles in multiple cancers. We found that the enriched BPs and pathways in the three clusters are associated with the cancer hallmarks including deregulating cellular energetics and inducing angiogenesis [54]. The enriched pathways in the cluster 5 and 8 are related to deregulating cellular energetics, including focal adhesion, PI3K-Akt signaling pathway and ECM-receptor interaction. Focal Adhesion Kinase through ECM promotes the activation of PI3K-AKT signaling [55]. Then P13K-AKT signaling pathway increases glycolysis in metabolic processes [19]. The enriched BPs in the cluster 2 are related to inducing angiogenesis, such as platelet degranulation, positive regulation of angiogenesis, leukocyte migration and angiogenesis. Angiogenesis plays an important role during macroscopic and microscopic neoplastic progression [54]. The microenvironment-related BPs in the cluster 7 are closely associated with anti-cancer immunity. Not surprisingly, we found that the genes in the modules were implicated in diseases, such as neoplastic processes, immune system diseases, cancer or viral infections and neovascularization. Some miRNA and transcription factors were also enriched in the modules. The mir-200 family has been identified as a biomarker in cancer [56]. Mir-199a has been associated with various cancers, including kidney, breast, bladder, bronchial squamous and stomach cancers [57–61]. Studies indicate that mir-29 family members may cooperatively or separately contribute to the development of breast and colon cancer [62]. *FOXO4* is reported to suppress tumor proliferation and metastasis in stomach carcinoma, and its clinical significance is observed in multiple cancers [63–65]. The *NFAT*-related roles have been studied in the tumor microenvironment [33]. *MAZ* promotes the tumor progression in glioblastoma, breast cancer, prostate cancer and hepatocellular cancer [31, 32]. The regulatory and clinical significance of *E2A* are studied in colorectal cancer [66]. Our study indicates that the genes in the identified modules play a cooperate role in multiple cancers through miRNA and transcription factors.

The sub-modules in the cluster 5 were analyzed for further understanding the regulatory mechanism in multiple cancers. We found that some genes in sub-modules play vital roles in multiple cancers, such as tumor suppressors and oncogenes. There was only one pathway in the largest sub-module M1 but the sub-module was highly enriched in cancer-related genes. Hub genes enriched in M2 have been identified to play a prognostic role in multiple cancers. Genes in M3 are involved in metastasis and prognosis. These sub-modules may help in discover new genes related with multiple cancers.

The identified modules showed different predictive power in prognosis of cancers. Some modules were able to predict survival in six cancers, such as cluster 2 and cluster 5. These modules were not significantly related to survival in BRCA and PRAD. Sample size and the number of deaths are two important factors in survival analysis. There exists the unbalanced number of genes involved in multiple cancers. These factors may result in the poor performance of prognostic capability. Some sub-modules in the cluster 5 showed the statistical significance of survival analysis in at least three cancers, such as M1, M2, M4 and M9. Other sub-modules also revealed the prognostic capability in different cancers. The prognosis of cluster 5 was validated in other cancers. Importantly, the cluster 5 could be prognostic in 12 cancer types in total.

In summary, we identified the functional modules and co-expression networks for the systematic analysis of the carcinogenic properties and regulatory mechanisms of multiple cancer. Further analysis indicates that these co-expression modules have a strong ability in predicting the survival of cancer patients. The results will be helpful in identifying new targets associated with cancer treatment. Our results will be valuable in cancer-related gene function annotation, and for the evaluation of cancer patients' prognosis.

MATERIALS AND METHODS

Differential gene expression analysis

We downloaded TCGA gene expression data from the Gene Expression Omnibus under accession number GSE62944. Eight types of cancers were used to construct the co-expression network, including BRCA, KIRC, PRAD, THCA, LUAD, BLCA, COAD and STAD. After the construction of the co-expression network, we also analyzed UCEC, HNSC, READ and LIHC for further assessing the functional significance of the modules (Supplementary Table 9).

Four methods were used for differential gene expression analysis, including limma, edgeR, DESeq, and SAMseq. The sets of differentially expressed genes from each cancer were pooled together to increase the statistical power.

The genes with multiple test corrected p-value < 0.05 were considered to be significantly differentially expressed. For SAMseq, the number of permutations used to estimate false discovery rates was set to 200 and the number of resamples used to construct test statistic was set to 100. For limma, we used the TMM method of the edgeR package and the voom transformation.

We proposed a three-step procedure for selecting the common genes shared by the eight cancer types (BRCA, KIRC, PRAD, THCA, LUAD, BLCA, COAD, STAD). First, genes which were significantly differentially expressed in at least four types of cancers analyzed by one method were clustered into one gene set. Second, genes shared by all four gene sets obtained by four methods were selected as a candidate gene set. At last, the genes that were not mapped to the corresponding DAVID genes were removed. We detected genes shared by UCEC, HNSC, READ, and LIHC cancers by following the rules similar to the above procedure, but with modification. Three methods (edgeR, SAMseq, and Limma) were used and significant genes shared by at least two cancer types with one method were clustered in one gene set.

Co-expression network construction

The gene expression data expressed as fragments per kilobase of exon per million reads mapped (FPKM) and normalized on log2 scale were represented by matrices. We then used HO-GSVD to extract common modules through matrix decomposition. The basic idea behind this approach is using spectral decomposition to identify common structures (subnetworks) in multiple datasets.

For the tth cancer $\left(t=1,2,..,T\right)$, the input data $D_t \in R^{n_t \times p}$ is the real matrix represented by the rows denoted by the samples n_t and the columns denoted by the genes p. The mathematical form of the HO-GSVD framework of T real matrices is given by

$$D_1 = U_1 \Sigma_1 V^T, D_2 = U_2 \Sigma_2 V^T, \ldots, D_T = U_T \Sigma_T V^T \ (1)$$

where $U_t \in R^{n_t \times p}$ is composed of normalized left basis vectors, $\Sigma_t \in R^{p \times p}$ is a non-negative diagonal matrix consisted of the higher-order generalized singular values and $V \in R^{p \times p}$ is composed of normalized right basis vectors. As the previous method [8], the right basis vectors V were defined as the solution of the eigensystem of the matrix S:

$$S = \frac{1}{T(T-1)} \sum_{t=1}^{T-1} \sum_{r=t+1}^{T} (E_t E_r^{-1} + E_r E_t^{-1})(2)$$

where the covariance matrix $E_t = D_t^T D_t$ can be treated as the co-expression matrix. Importantly, the common HO-GSVD subspace is spanned by the right basis vectors V. Then we used the right basis vectors to select common structures shared by all cancer types. The advantage of this approach for identification of co-expressed structures across cancer types is able to reproduce accurately common structures without relying on any predefined parameters.

Based on the eigenvalues of the eigen-decomposition of S, we chose the top ten eigenvectors to identify co-expression gene modules. A small part of the genes is supposed to have significantly similarity with each other and these genes can be regarded as the co-expression genes. Similarly to the strategy used in [67], the selected eigenvectors were decomposed into two components and modeled using Gaussian Mixture Model (GMM). In addition, the small weight component of bimodal distribution can identify the small proportional genes with high similarity. We calculated the tail area-based false discover rate (q-value) using the R package fdrtool to identify co-expression genes. Then we clustered the genes into the co-expression gene module through the cut-off of q-value (0.001).

Enrichment analysis

Functional enrichment analysis of the co-expressions gene modules was performed using DAVID [15]. Benjamini-Hochberg method was used for the multiple test correction of p-values. KEGG pathways and Gene Ontology terms with Benjamini-Hochberg corrected p-value less than 0.05 were considered as significantly enriched. To validate the functional significance of modules, only gene modules involved in significant pathways were retained for further analysis. The cell type enrichment was analyzed using Cten [17]. Disease association analysis was conducted using the gene set analysis toolkit WebGestalt. In addition, the enrichment of miRNAs and transcription factors was performed using WebGestalt [16].

Functional network analysis

We used module genes to construct the functional interaction network within the Cytoscape FI plugin [34]. Linker genes were used to maximize the number of connected genes in the module. Then we clustered the FI network into small modules. The functions of small modules were analyzed for GO terms and pathway enrichment.

Survival analysis

We downloaded overall survival data from the Firehose. The patients with multiple repeated samples were not included in the survival analysis. The R package 'survival' was used for the Cox proportional hazards (PH) model and Kaplan-Meier survival analysis. To detect prognostic genes, we calculated P-values based on the raw Wald test for the Cox PH model. Based on the expression of module genes, we classified tumor samples into clusters through NMF [68]. We then estimated the significance of differences between different clusters with patients' survival using log-rank test of the Kaplan-Meier method. We also performed survival analysis on small modules generated from the FI network.

Author contributions

WY collected and analyzed the data, and wrote the manuscript; XZ and WY conceived the idea. BZ participated in this work and analyzed the data. SZ participated in this work, supervised the whole work and revised the manuscript. YW, WZ and XZ revised the manuscript extensively. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no potential conflicts of interest.

FUNDING

This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2014CB340404, in part by the National Science Foundation of China under Grant 61471267, Grant 61373105 and Grant 61672422, in part by the Fundamental Research Funds for the Central Universities, and in part by National Institutes of Health [1U01CA166886, 1U01AR069395 and 1R01GM123037].

REFERENCES

1. Borkowska E, Constantinou M, Jędrzejczyk A, Traczyk M, Banaszkiewicz M, Pietrusiński M, Marks P, Rożniecki M, Kruk A, Kałużewski B. Artificial neural network in predicting bladder cancer recurrence. Hered Cancer Clin Pract. 2012; 10:A3. https://doi.org/10.1186/1897-4287-10-S1-A3.
2. Mishra M, Kumar A. Computational analysis of genetic network involved in pancreatic cancer in human. BMC bioinformatics. 2011; 12:A11. https://doi.org/10.1186/1471-2105-12-S11-A11.
3. Zhang J, Lu K, Xiang Y, Islam M, Kotian S, Kais Z, Lee C, Arora M, Liu HW, Parvin JD, Huang K. Weighted frequent gene co-expression network mining to identify genes involved in genome stability. PLOS Comput Biol. 2012; 8:e1002656. https://doi.org/10.1371/journal.pcbi.1002656.
4. 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. https://doi.org/10.1038/ncomms4231.
5. Li W, Wang R, Bai L, Yan Z, Sun Z. Cancer core modules identification through genomic and transcriptomic changes correlation detection at network level. BMC Syst Biol. 2012; 6:64. https://doi.org/10.1186/1752-0509-6-64.
6. Alter O, Brown PO, Botstein D. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. Proc Natl Acad Sci USA. 2003; 100:3351–56. https://doi.org/10.1073/pnas.0530258100.
7. Ponnapalli SP, Saunders MA, Van Loan CF, Alter O. A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms. PloS one. 2011; 6:e28072. https://doi.org/10.1371/journal.pone.0028072.
8. Xiao X, Moreno-Moral A, Rotival M, Bottolo L, Petretto E. Multi-tissue analysis of co-expression networks by higher-order generalized singular value decomposition identifies functionally coherent transcriptional modules. PLoS Genet. 2014; 10:e1004006. https://doi.org/10.1371/journal.pgen.1004006.
9. Wang Y, Zhao W, Zhou X. Matrix factorization reveals aging-specific co-expression gene modules in the fat and muscle tissues in nonhuman primates. Sci Rep. 2016; 6:34335. https://doi.org/10.1038/srep34335.
10. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139-40. https://doi.org/10.1093/bioinformatics/btp616.

11. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004; 3:e3. https://doi.org/10.2202/1544-6115.1027.

12. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat Methods Med Res. 2013; 22:519–36. https://doi.org/10.1177/0962280211428386.

13. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106.

14. Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics. 2013; 14:91. https://doi.org/10.1186/1471-2105-14-91.

15. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44-57. https://doi.org/10.1038/nprot.2008.211.

16. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. 2005; 33:W741-48. https://doi.org/10.1093/nar/gki475.

17. Shoemaker JE, Lopes TJ, Ghosh S, Matsuoka Y, Kawaoka Y, Kitano H. CTen: a web-based platform for identifying enriched cell types from heterogeneous microarray data. BMC genomics. 2012; 13:460. https://doi.org/10.1186/1471-2164-13-460.

18. Neapolitan R, Horvath CM, Jiang X. Pan-cancer analysis of TCGA data reveals notable signaling pathways. BMC cancer. 2015; 15:516. https://doi.org/10.1186/s12885-015-1484-6.

19. Pickup MW, Mouw JK, Weaver VM. The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. 2014; 15:1243-53. https://doi.org/10.15252/embr.201439246.

20. Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol. 2012; 196:395-406. https://doi.org/10.1083/jcb.201102147.

21. McLean GW, Carragher NO, Avizienyte E, Evans J, Brunton VG, Frame MC. The role of focal-adhesion kinase in cancer—a new therapeutic opportunity. Nat Rev Cancer. 2005; 5:505-15. https://doi.org/10.1038/nrc1647.

22. Fruman DA, Rommel C. PI3K and cancer: lessons, challenges and opportunities. Nat Rev Drug Discov. 2014; 13:140-56. https://doi.org/10.1038/nrd4204.

23. Provenzano PP, Eliceiri KW, Campbell JM, Inman DR, White JG, Keely PJ. Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med. 2006; 4:38. https://doi.org/10.1186/1741-7015-4-38.

24. Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, Lichtor T, Decker WK, Whelan RL, Kumara HMCS, Signori E, Honoki K, Georgakilas A, et al. Immune evasion in cancer: Mechanistic basis and therapeutic strategies. Seminars in cancer biology. 2015; 35:S185-98.

25. Nieman KM, Romero IL, Van Houten B, Lengyel E. Adipose tissue and adipocytes support tumorigenesis and metastasis. Biochim Biophys Acta. 2013; 1831:1533-41. https://doi.org/10.1016/j.bbalip.2013.02.010.

26. Mytar B, Baj-Krzyworzeka M, Majka M, Stankiewicz D, Zembala M. Human monocytes both enhance and inhibit the growth of human pancreatic cancer in SCID mice. Anticancer research. 2008; 28:187-92.

27. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 2008; 10:593-601. https://doi.org/10.1038/ncb1722.

28. Li X, Roslan S, Johnstone CN, Wright JA, Bracken CP, Anderson M, Bert AG, Selth LA, Anderson RL, Goodall GJ, Gregory PA, Khew-Goodall Y. MiR-200 can repress breast cancer metastasis through ZEB1-independent but moesin-dependent pathways. Oncogene. 2014; 33:4077-88. https://doi.org/10.1038/onc.2013.370.

29. Boyerinas B, Park SM, Hau A, Murmann AE, Peter ME. The role of let-7 in cell differentiation and cancer. Endocr Relat Cancer. 2010; 17:F19–36. https://doi.org/10.1677/ERC-09-0184.

30. Fu Z, Tindall DJ. FOXOs, cancer and regulation of apoptosis. Oncogene. 2008; 27:2312-9. https://doi.org/10.1038/onc.2008.24.

31. Luo W, Zhu X, Liu W, Ren Y, Bei C, Qin L, Miao X, Tang F, Tang G, Tan S. MYC associated zinc finger protein promotes the invasion and metastasis of hepatocellular carcinoma by inducing epithelial mesenchymal transition. Oncotarget. 2016; 7:86420–32. https://doi.org/10.18632/oncotarget.13416.

32. Maity G, Sarkar S, Dhar K, Dhar G, Haque I, Banerjee SK, Banerjee S. Abstract 2330: Transcription factor MAZ promotes cell growth and aggressive behavior of human pancreatic cancer cells. Cancer Res. 2014; 74. https://doi.org/10.1158/1538-7445.AM2014-2330.

33. Pan MG, Xiong Y, Chen F. NFAT gene family in inflammation and cancer. Curr Mol Med. 2013; 13:543–54. https://doi.org/10.2174/1566524011313040007.

34. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. Genome Biol. 2010; 11:R53. https://doi.org/10.1186/gb-2010-11-5-r53.

35. van Zijl F, Krupitza G, Mikulits W. Initial steps of metastasis: cell invasion and endothelial transmigration. Mutat Res. 2011; 728:23-34. https://doi.org/10.1016/j.mrrev.2011.05.002.

36. Stowell SR, Ju T, Cummings RD. Protein glycosylation in cancer. Annu Rev Pathol. 2015; 10:473-510. https://doi.org/10.1146/annurev-pathol-012414-040438.

37. Rizzo P, Osipo C, Foreman K, Golde T, Osborne B, Miele L. Rational targeting of Notch signaling in cancer. Oncogene. 2008; 27:5124-31. https://doi.org/10.1038/onc.2008.226.

38. Dang N, Meng X, Song H. Nicotinic acetylcholine receptors and cancer. Biomed Rep. 2016; 4:515-8. https://doi.org/10.3892/br.2016.625.

39. Improgo MR, Soll LG, Tapper AR, Gardner PD. Nicotinic acetylcholine receptors mediate lung cancer growth. Front Physiol. 2013; 4:251. https://doi.org/10.3389/fphys.2013.00251.

40. Ananieva E. Targeting amino acid metabolism in cancer growth and anti-tumor immune response. World J Biol Chem. 2015; 6:281–89. https://doi.org/10.4331/wjbc.v6.i4.281.

41. Ewald JA, Downs TM, Cetnar JP, Ricke WA. Expression microarray meta-analysis identifies genes associated with Ras/MAPK and related pathways in progression of muscle-invasive bladder transition cell carcinoma. PloS one. 2013; 8:e55414. https://doi.org/10.1371/journal.pone.0055414.

42. Eissa S, Zohny SF, Zekri AR, El-Zayat TM, Maher AM. Diagnostic value of fibronectin and mutant p53 in the urine of patients with bladder cancer: impact on clinicopathological features and disease recurrence. Med Oncol. 2010; 27:1286–94. https://doi.org/10.1007/s12032-009-9375-9.

43. Arentz G, Chataway T, Price TJ, Izwan Z, Hardi G, Cummins AG, Hardingham JE. Desmin expression in colorectal cancer stroma correlates with advanced stage disease and marks angiogenic microvessels. Clin Proteomics. 2011; 8:16. https://doi.org/10.1186/1559-0275-8-16.

44. Ayala G, Tuxhorn JA, Wheeler TM, Frolov A, Scardino PT, Ohori M, Wheeler M, Spitler J, Rowley DR. Reactive stroma as a predictor of biochemical-free recurrence in prostate cancer. Clin Cancer Res. 2003; 9:4792-801.

45. Abdou AG, Kandil M, Elshakhs S, El-Dien MS, Abdallah R. Renal cell carcinoma with rhabdoid and sarcomatoid features presented as a metastatic thigh mass with an unusual immunohistochemical profile. Rare Tumors. 2014; 6:5037. https://doi.org/10.4081/rt.2014.5037.

46. Popper H, Wirnsberger G, Hoefler H, Denk H. Immunohistochemical and histochemical markers of primary lung cancer, lung metastases, and pleural mesotheliomas. Cancer Detect Prev. 1987; 10:167-74.

47. Wang X, Zhang L, Li H, Sun W, Zhang H, Lai M. THBS2 is a Potential Prognostic Biomarker in Colorectal Cancer. Sci Rep. 2016; 6:33366. https://doi.org/10.1038/srep33366.

48. Sudol M. From Rous sarcoma virus to plasminogen activator, src oncogene and cancer management. Oncogene. 2011; 30:3003-10. https://doi.org/10.1038/onc.2011.38.

49. Li Z, Zhang X, Yang Y, Yang S, Dong Z, Du L, Wang L, Wang C. Periostin expression and its prognostic value for colorectal cancer. Int J Mol Sci. 2015; 16:12108-18. https://doi.org/10.3390/ijms160612108.

50. Lee MJ, Heo SC, Shin SH, Kwon YW, Do EK, Suh DS, Yoon MS, Kim JH. Oncostatin M promotes mesenchymal stem cell-stimulated tumor growth through a paracrine mechanism involving periostin and TGFBI. Int J Biochem Cell Biol. 2013; 45:1869–77. https://doi.org/10.1016/j.biocel.2013.05.027.

51. Guo Y, Li CI, Ye F, Shyr Y. Evaluation of read count based RNAseq analysis methods. BMC Genomics. 2013; 14:S2. https://doi.org/10.1186/1471-2164-14-S8-S2.

52. Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014; 15:R29. https://doi.org/10.1186/gb-2014-15-2-r29.

53. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform. 2015; 16:59–70. https://doi.org/10.1093/bib/bbt086.

54. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011; 144:646-74. http://dx.doi.org/10.1016/j.cell.2011.02.013.

55. Tai YL, Chen LC, Shen TL. Emerging roles of focal adhesion kinase in cancer. BioMed Res Int. 2015; 2015:690690. https://doi.org/10.1155/2015/690690.

56. Park SM, Gaur AB, Lengyel E, Peter ME. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. Genes Dev. 2008; 22:894–907. https://doi.org/10.1101/gad.1640608.

57. Huang J, Dong B, Zhang J, Kong W, Chen Y, Xue W, Liu D, Huang Y. miR-199a-3p inhibits hepatocyte growth factor/c-Met signaling in renal cancer carcinoma. Tumour Biol. 2014; 35:5833–43. https://doi.org/10.1007/s13277-014-1774-7.

58. Li W, Wang H, Zhang J, Zhai L, Chen W, Zhao C. miR-199a-5p regulates β1 integrin through Ets-1 to suppress invasion in breast cancer. Cancer Sci. 2016; 107:916-23. https://doi.org/10.1111/cas.12952.

59. Shatseva T, Lee DY, Deng Z, Yang BB. MicroRNA miR-199a-3p regulates cell proliferation and survival by targeting caveolin-2. J Cell Sci. 2011; 124:2826-36. https://doi.org/10.1242/jcs.077529.

60. Gu S, Chan WY. Flexible and versatile as a chameleon-sophisticated functions of microRNA-199a. Int J Mol Sci. 2012; 13:8449–66. https://doi.org/10.3390/ijms13078449.

61. Wang Z, Ma X, Cai Q, Wang X, Yu B, Cai Q, Liu B, Zhu Z, Li C. MiR-199a-3p promotes gastric cancer progression by targeting ZHX1. FEBS Lett. 2014; 588:4504–12. https://doi.org/10.1016/j.febslet.2014.09.047.

62. Kriegel AJ, Liu Y, Fang Y, Ding X, Liang M. The miR-29 family: genomics, cell biology, and relevance to renal and cardiovascular injury. Physiol Genomics. 2012; 44:237-44. https://doi.org/10.1152/physiolgenomics.00141.2011.

63. Su L, Liu X, Chai N, Lv L, Wang R, Li X, Nie Y, Shi Y, Fan D. The transcription factor FOXO4 is down-regulated and inhibits tumor proliferation and metastasis

in gastric cancer. BMC Cancer. 2014; 14:378. https://doi.org/10.1186/1471-2407-14-378.

64. Wang Y, Zhou Y, Graves DT. FOXO transcription factors: their clinical significance and regulation. Biomed Res Int. 2014; 2014:925350. https://doi.org/10.1155/2014/925350.

65. Li J, Jiang Z, Han F, Liu S, Yuan X, Tong J. FOXO4 and FOXD3 are predictive of prognosis in gastric carcinoma patients. Oncotarget. 2016; 7:25585–92. https://doi.org/10.18632/oncotarget.8339.

66. Huang A, Zhao H, Quan Y, Jin R, Feng B, Zheng M. E2A predicts prognosis of colorectal cancer patients and regulates cancer cell growth by targeting miR-320a. PloS one. 2014; 9:e85201. https://doi.org/10.1371/journal.pone.0085201.

67. Xiang T, Gong S. Spectral clustering with eigenvector selection. Pattern Recognit. 2008; 41:1012-29. https://doi.org/10.1016/j.patcog.2007.07.023.

68. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci USA. 2004; 101:4164–69. https://doi.org/10.1073/pnas.0308531101.