Identification of key modules and prognostic markers in adrenocortical carcinoma by weighted gene co-expression network analysis

YONG ZOU and LUANLIAN JING

Department of Oncology, The People's Hospital of Hanchuan, Hanchuan, Hubei 431600, P.R. China

Received January 24, 2019; Accepted June 12, 2019

DOI: 10.3892/ol.2019.10725

Abstract. Adrenocortical carcinoma (ACC) is a rare and aggressive cancer with a high relapse rate and limited treatment options. Therefore, the identification of potential prognostic markers in patients with ACC may improve early detection, survival rates and may additionally provide novel insights into the early detection of recurrence. In the present study, clinical traits and RNA-seq data of 79 patients with ACC were obtained from The Cancer Genome Atlas (TCGA). Weighted gene co-expression network analysis was carried out and 17 distinct co-expression modules were built to examine the association between the modules and the clinical traits. Of the 17 modules, two co-expression modules, which contained 214 and 168 genes, were significantly correlated with two clinical traits, tumor stage and vital status. Functional enrichment analysis was performed on the selected modules. The results showed that one of the modules was primarily enriched in cell division and the other module was enriched in metabolic pathways, suggesting their involvement in tumor progression. Furthermore, cyclin dependent kinase 1 (CDK1) and ubiquitin C (UBC) were identified as hub genes in both modules. Survival analysis revealed that the high expression of the hub genes significantly correlated with the poor survival rate of patients, suggesting that CDK1 and UBC have vital roles in the progression of ACC. In the present study, a co-expression gene module of ACC was provided and the prognostic genes that may serve as new diagnostic markers in the future were defined.

Introduction

The adrenal glands reside above the kidneys and produce multiple hormones essential for development (1). Each gland consists of an inner medulla and an outer cortex that produce catecholamines and steroid hormones, respectively (2). Adrenocortical carcinoma (ACC) is a rare tumor of the adrenal cortex with an estimated incidence of 1 patient per million/year (3,4). There is a higher prevalence of ACC in females and an increased incidence in the first and fourth to fifth decades of life (5). There are no notable clinical phenotypic characteristics in patients with ACC during the early stage and the majority of patients are diagnosed with advanced stage ACC in the first instance (6). Patients with ACC do not respond favorably to chemotherapy and radiotherapy (7) and the patients frequently have to undergo surgical resection where the 5-year survival rate is >35%. Mitotane (o, p'-dichlorodiphenyldichloroethane) has been used since the 60s for treating patients with ACC, despite its toxicity and narrow therapeutic index (8,9). An improved understanding of the genes associated with ACC may improve treatment options by identifying potential therapeutic targets. Weighted gene co-expression network analysis (WGCNA) is a frequently used method to explore the association between genes and phenotypes (10). Gene expression data are transformed into co-expression modules and provide insights into the signaling networks that may underlie certain phenotypes. WGCNA is widely used to improve understanding of various biological processes such as cancer and its progression (11,12). Yang et al (13) identified candidate biomarkers and molecular mechanisms involved in glioblastoma multiforme using WGCNA (13). WGCNA compares differentially expressed genes and identifies key interactions among different co-expression modules (12).

Next generation sequencing is used to detect genomic alterations which could be used to guide targeted therapies for treating patients with ACC and several targets have been discovered (14,15). However, the molecular diagnostic parameters are still not entirely known and there are only small number of studies that have cataloged relevant expression modules in patients with ACC, which has limited understanding of the disease and its mechanisms (16-18). Using WGCNA, it is possible to construct gene networks and analyze the connectivity between genes and clinical traits (19). The master regulators identified in the gene regulation network will typically exhibit important functions.

In this study, it was hypothesized that distinct ACC co-expression modules are associated with different clinical outcomes and the highly connected genes in the modules can represent the biological feature of this module and have the potential as prognosis markers.
and clinical data was downloaded from the TCGA database of 79 patients with ACC at different stages. The candidate mRNAs related to tumor progression were identified by co-expression analysis. Furthermore, 2,472 differentially expressed genes from ACC and normal tissues were downloaded from the Gene expression profiling interactive analysis (GEPIA) website (20), and ACC and ACC co-expression modules were constructed (20). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on the selected modules, and the hub gene in each module was identified, which assisted in determining the primary function(s) of the genes in each module. These findings may play a significant role in recognizing the malignant potential of these genes as well as the prognosis of patients with ACC.

Materials and methods

Patients and TCGA data retrieval. The clinical and gene expression data of 79 patients with ACC were obtained from TCGA (https://tcga-data.nci.nih.gov/tcga/) (21), where the expression profile was obtained based on the IlluminaHiSeq RNA-seq platform (Illumina, Inc.). The Bioconductor (22) packages TCGAbiolinks (23) based on the R software (24) both for the normal adrenal gland tissue and the ACC tissue were downloaded from Recount2 database (29), including 159 normal adrenal gland samples and 79 ACC samples respectively. The ACC samples were labeled as ‘positive’ and the normal adrenal gland sample were labeled as ‘negative’. Subsequently, the ROCR (30) package based on the R (version 3.5.2) software system was used to plot the ROC curve and calculated the AUC for the hub genes.

WGCNA. Co-expression networks were constructed with the identified DEGs. Pearson's correlation coefficients were calculated for all the genes in the dataset and the correlation matrix of the entire gene dataset was obtained. The power β was used to remove weakly correlated genes, while retaining the strongly correlated ones. The process produced an adjacent matrices weighted network that was converted into a topological overlapping matrices (TOM) network as previously described (10,25). After constructing the TOM network, hierarchical clustering was used to generate a cluster dendrogram with branches corresponding to the gene co-expression modules. The WGCNA (10) algorithm was used to identify the co-expression modules. WGCNA was implemented using R (10). The TOM representing the overlap in shared neighbors and the soft thresholding power were calculated according to a previous study (25). Colored heatmaps were used to analyze the strength of the interactions.

Construct module-trait associations of ACC. Module-trait associations were estimated using the association between the module signature and the phenotype (clinical traits) as previously described (12), thus allowing easy identification of the expression set (module) that strongly correlated with the phenotype. For each expression profile, the gene significance (GS) was calculated as the absolute value of the association between the expression profile and each clinical trait, whereas the module membership (MM) was defined as the correlation between the expression profile and each module signature. Functional analysis of network module genes. To identify the underlying biological mechanisms responsible for the progression of ACC, the DEGs derived from the brown and yellow modules were used for GO and KEGG pathway enrichment analyses, because the two modules are both correlated with the patient clinical traits. The DEGs annotated in the GO database were used to classify the GO functions in the ClusterProfiler package (26), and the DEGs for KEGG enrichment analysis were mapped to the KEGG database. After the enrichment analysis, the significant KEGG pathway and GO terms were selected according to the cut-off criterion of adjusted P<0.05.

Protein-protein-interaction (PPI) network construction. The PPI data was retrieved from a previous study, which contained protein interaction associations from 15 databases (27). The overall PPI network was based on 16,081 nodes and 231,633 interactions in these databases. The genes involved in the brown and yellow modules were mapped to the overall PPI network to obtain the module specific interaction network. Furthermore, the degree of each node was calculated using the IGRAPH package (28). The widely linked genes (hub genes) in the PPI network were more closely related to most proteins.

Survival analysis. Survival analysis for all genes in the brown and yellow modules was performed using the R survival package (https://CRAN.R-project.org/package=survival; version 2.41-3). A log-rank test was performed to determine whether the expression of a gene correlated with the overall survival. For the overall survival rate, a log-rank test was used to test for significance in the univariate analysis between each subgroup. Unless otherwise specified, P<0.05 was considered to indicate a statistically significant difference.

Results

Pre-processing of datasets and construction of ACC co-expression modules. The clinical information, including the age, sex, race, ethnicity and vital status, and the normalized RNA-seq data of 79 patients were obtained from TCGA. The DEGs in normal and ACC tissues were retrieved from GEPIA. A total of 2,472 DEGs were identified with a Q-value <0.01 and a log2FC>1. The expression values of the 2,472 genes in the 79 ACC samples were used to construct the co-expression module by WGCNA. The flashClust package (31) was used to perform the cluster analysis.
(Fig. 1). A total of four abnormal samples were removed and 75 samples were used for further analysis.

The power value, which primarily affects the independence and the average connectivity degree of the co-expression module of the most critical parameters in the network, is considered an important parameter. Therefore, the power value of the data was assessed (Fig. 2A). A power value of 3, indicated the independent degree was up to 0.8 and the average connectivity degree was higher (Fig. 2B). The degree distribution of the node and the fitting relationship between the node degree and its corresponding proportion were calculated (Fig. 2C and D). The results revealed that the constructed network was a scale-free network conforming to biological characteristics. Therefore, a power value of 3 was used to construct the co-expression modules and the results generated 17 distinct gene co-expression modules in the ACC samples. These co-expression modules were constructed and depicted in different colors (Fig. 3A). They were arranged from large to small by the number of genes they included and the interactions of the 17 co-expression modules are shown in Fig. 3B.

Gene co-expression modules correspond to clinic traits. The data on the clinical traits were obtained from TCGA. The modules with common expression patterns that were associated with certain traits were identified based on the association between the module eigengene and the clinical traits (Fig. 3C; Tables SI and SII). The dendrogram and heat map of the eigengene were both used to identify groups of correlated eigengenes and clinical characteristics. The brown module clustered with two important clinical indexes, the tumor stage and vital status (Fig. 4). Moreover, the Pearson’s correlation coefficient (PCC) between the yellow module and the tumor stage is -0.34 (P=0.003), and the PCC between the yellow module and the vital status is -0.47 (P=2x10^-5). Furthermore, the yellow module negatively correlated with these two indexes, indicating that genes in this module may be associated with the prognosis of patients (Fig. 3B). Therefore, both modules were defined as core modules for further study.

GO and KEGG enrichment analysis of genes in brown and yellow modules. The association between the genes and the clinical characteristics were calculated along with the MM and GS, where MM was the correlation of the gene expression profile with the eigengenes and GS was the absolute value of the association between the gene and the external traits. MMs and GSs with high thresholds were chosen to avoid false positive prognostic genes, and the top 50% of genes in the brown and yellow modules were selected as candidate genes. GO and KEGG analyses were performed to explore the biological functions of the candidate genes in the brown and yellow modules. All significant terms in the annotated systems were represented as colored bars to compare the relative significance of the enriched terms, where the length and color saturation of each term were proportional to the gene count/ratio and the P-value obtained from the enrichment analyses (Figs. 5 and 6). According to the results, GO enrichment analysis indicated that the genes in the brown module were primarily involved in cell division and protein kinase activity (Fig. 6A), which was consistent with the KEGG pathway enrichment results (Fig. 5A). For the yellow module, the genes were enriched in various metabolic pathways in GO enrichment analysis such as valine, creatinine and isoleucine metabolic processes (Fig. 6B) and various tumor-associated pathways including acute myeloid leukemia and renal cell carcinoma in KEGG enrichment analysis (Fig. 5B).

PPI network-based prognostic gene identification. The brown and yellow modules contained 214 and 168 genes, respectively. To identify the prognostic genes based on the PPI network, the genes were mapped to the pre-built PPI network, and the nodes
that failed to map to the PPI network were ignored. Cytoscape (V3.6.1) (32) software was used to construct the module and to calculate the intramodular connectivity. The intramodular connectivity was calculated for each gene by the connection strength with other module genes and genes with high intramodular connectivity were considered intramodular hub genes. The hub genes in the two modules are represented with larger dots in Fig. 7. The brown module subnetwork contained 105 nodes and 216 edges (Fig. 7A), whereas the yellow module subnetwork contained 92 nodes and 126 edges (Fig. 7B).

To further validate the results, three genes with the highest degree, in the brown (CCNB1, CDC2 and CDK1) and the yellow (PRKCA, RAD23A and UBC) modules were selected to investigate their correlation with the overall survival of the patients (Fig. 8). The nodes with the highest degrees in the two modules were CDK1 and UBC, and their elevated expression levels were significantly associated with poor overall survival (P<0.0001; Fig. 8A and D). Other genes (CCNB1, CDC2, PRKCA and RAD23A) were also significantly associated with the overall survival and they may serve as prognostic genes in ACC (Fig. 8B, C, E and F). The relationships between all nodes and the overall survival were calculated (data not shown). In addition, as shown in the Fig. 9, ROC curve validated that the high degree gene in the brown module (Fig. 9A) and the yellow module (Fig. 9B) exhibited good diagnostic efficiency for normal and tumor tissues.

Discussion

Compared with other methods such as differential expression analysis, WGCNA places a focus on the association between co-expression modules and clinical traits, and therefore, the results that are considered more reliable yield relevant biological significance (12). In the present study, a total of 17 distinct gene co-expression modules were identified from 79 patients with ACC by WGCNA to determine the association between the transcriptome of patients with ACC and the clinical traits. After examining the associations between the modules and the clinical traits, two modules were correlated with clinical traits. Several hub genes in the network were identified which
confirmed that CDK1 and UBC serve important roles in the progression of ACC.

Early diagnosis and specific markers are important for managing and limiting tumor development. Electrochemical immunosensors have been developed to detect dehydroepiandrosterone sulfate in blood plasma samples for early diagnosis of patients with pediatric ACC (33). Mohan et al (34) demonstrated that evaluation of G0S2 hypermethylation identified a subgroup of patients with ACC with a rapidly progressive disease course which is feasible for clinical treatment options (34). Novel therapeutic regimens are frequently based on specific gene or protein targets present in the disease. Identification of programmed cell death protein 1 function has improved the development of cancer immunotherapy (35,36). The identification of potential biomarkers in the present study may serve as novel targets for drugs or in diagnostic methods clinically.

The power value was the most critical parameter affecting the independence and the average connectivity degree of the co-expression modules in WGCNA (10,37). A higher average connectivity degree appeared when the power value =3. The brown module clustered with tumor stage and vital status, whereas the yellow module negatively correlated with these two indices. The Tumor-Node-Metastasis classification system was used in ACC staging and stage III ACC was characterized by infiltration in the surrounding tissues (38,39). As discussed above tumor stage and vital status were important clinical parameters, and they may also reflect tumor prognosis (40-42). Functional enrichment analysis for candidate genes in the brown module showed that these genes were primarily enriched in pathways associated with cell division. Uncontrolled self-renewal capacity and aberrant regulation of genetic material may promote tumor cell progression and recurrence (43). Tripartite motif-containing protein 3 (TRIM3) has been used as a tumor suppressor due to its ability to regulate asymmetric cell division in glioblastoma and expression of TRIM3 additionally attenuates the stem-like quality of primary glioblastoma cultures (44). A combination of mitochondrial division inhibitor 1 and platinum agents produced a synergistic pro-apoptotic effect in drug-resistant tumor cells (45). These results indicate that cell division is associated with tumor progression and genes in the brown module may serve an important role in the regulation of cell division in patients with ACC. Furthermore, genes in the yellow module were primarily enriched in metabolic processes and lysosome membrane-associated pathways, which may influence the T cell-mediated immune response, tumor invasion, and malignancy (46-48).

Cytoscape was used to construct a co-expression network for the brown and yellow modules, and the genes with high intramodular connectivity were considered as hub genes. The genes with the highest degree of connectivity in the two modules were CDK1 and UBC. CDK1 is a master regulator of the cell cycle and overexpression of CDK1 increases the spheroid-forming ability of tumor cells and the tumor-initiating capacity, whereas the inhibition of CDK1 reduced these characteristics (49-51). CDK1 was previously identified as a biomarker in patients with ACC using PCR, western blotting
Figure 6. Gene ontology enrichment analysis of the genes in the (A) brown module and (B) yellow module.

Figure 7. Protein-protein interaction network of the genes involved in the (A) brown module and (B) yellow module. The size of the node represents the degree of connectivity of the node and the edges represents, and interaction between two genes.
and immunofluorescence by Xiao et al (52). UBC is a highly conserved protein which is involved in the selective proteolysis of abnormal proteins (53). Hao et al (54) identified UBC as a differential node protein which may serve as a key regulator in the response of non-small cell lung cancer A549 cells to phycocyanin (54). Furthermore, knockdown of UBC and UBB with mixed small hairpin RNAs suppressed the growth and radio sensitivity of H1299 lung cancer cells (55). UBC was also associated with the regulation of Toxoplasma gondii rhoptry protein 18, which serves a key regulatory role in cell immunity and apoptosis (56). Survival analysis revealed that increased expression of CDK1 and UBC resulted in poorer overall survival of patients with ACC, which is consistent with previous studies (49,50).

However, there were some limitations in the present study. The finally identified hub genes in the network requires experimental verification. At present, there are no studies investigating the function of UBC in ACC, to the best of our knowledge. Therefore, experiments based on clinical samples are required to verify the results of the present study.
In conclusion, the brown and yellow modules were identified as the most critical modules in the progression of ACC. The hub genes CDK1 and UBC were the most significantly expressed genes in the two modules, and they may serve as potential diagnostic and prognostic biomarkers of patients with ACC in the future. The selected candidate genes may serve as targets for the development of novel therapeutics.

Acknowledgements
Not applicable.

Funding
No funding was received.

Availability of data and materials
The datasets analyzed in the present study are available in the Cancer Genome Atlas repository, (https://cancergenome.nih.gov/).

Authors’ contributions
Cancer Genoma Atlas repository, (https://cancergenome.nih.gov/). The datasets analyzed in the present study are available in the The

Ethics approval and consent to participate
Not applicable.

Patient consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References


