Genomics Data 14 (2017) 132–140

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC)

Lushun Yuan[a], Liang Chen[a], Kaiyu Qian[a,b], Guofeng Qian[c], Chin-Lee Wu[d], Xinghuan Wang[a,*], Yu Xiao[a,e,f,**]

[a] Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, China
[b] Department of Urology, The Fifth Hospital of Wuhan, Wuhan, China
[c] Department of Endocrinology, The First Affiliated Hospital of Zhejiang University, Hangzhou, China
[d] Department of Urology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[e] Laboratory of Precision Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China
[f] Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan, China

ARTICLE INFO

Keywords:
Clear cell renal cell carcinoma (ccRCC)
Co-expression network analysis
Hub genes
Progression
Prognosis

ABSTRACT

Human clear cell renal cell carcinoma (ccRCC) is one of the most common types of malignant adult kidney tumors. We constructed a weighted gene co-expression network to identify gene modules associated with clinical features of ccRCC (n = 97). Six hub genes (*CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A* and *UBE2C*) were identified in both co-expression and protein-protein interaction (PPI) networks, which were highly correlated with pathologic stage. The significance of expression of the hub genes in ccRCC was ranked top 4 among all cancers and correlated with poor prognosis. Functional analysis revealed that the hub genes were significantly enriched in cell cycle regulation and cell division. Gene set enrichment analysis suggested that the samples with highly expressed hub gene were correlated with cell cycle and p53 signaling pathway. Taken together, six hub genes were identified to be associated with progression and prognosis of ccRCC, and they might lead to poor prognosis by regulating p53 signaling pathway.

1. Introduction

Renal cell carcinoma (RCC) is the most common type of malignant adult kidney tumors, accounting for > 90% of all adult renal tumors. Up to one-third of patients with RCC already suffered with a distant metastasis at the time of diagnosis [1]. Clear cell RCC (ccRCC), taking up about 75%–85% of RCC, is the most common subtype [2]. At present, ccRCC is usually resistant to chemotherapy. Targeted therapies have been exploited for their target specificity and low toxicity so they can be the best choice of non-surgical treatments [3]. Therefore, many biomarkers for clear cell renal cell carcinoma have been discovered including *VHL*, *VEGF*, *CAIX* and *HIF1α/2α* mutations. Some of which could predict therapeutic effect and clinical prognosis [4]. We know that carcinogenesis is not the result of deregulation of several oncogenes or tumor suppressors; it is the outcome of complex mechanisms, including the high interconnection between genes with similar expression patterns [5]. Thus, it is urgently needed to identify novel molecular biomarkers that can predict disease stage and clinical outcome of ccRCC patients, which could help understand its pathogenesis and provide personalized treatment.

Rapid technological breakthroughs of genome-wide sequencing have shed new light on the research of clinical issues and related pathological mechanisms in various cancers [6]. Nowadays, most studies just concentrated on the screening of differentially expressed genes and not attached enough attention to the high degree of interconnection between genes, where genes with similar expression patterns may be functionally related. The algorithm, weighted gene co-expression network analysis (WGCNA), can construct free-scale gene co-expression networks to explore the relationships between different gene sets or between gene sets and clinical features [7]. WGCNA has been widely

Abbreviations: ccRCC, clear cell renal cell carcinoma; HPA, human protein atlas; GSEA, enrichment analysis and gene set enrichment; WGCNA, weighted gene co-expression network analysis; TCGA, the cancer genome atlas; DEGs, differentially expressed genes; TOM, topological overlap matrix; MEs, module eigengenes; GS, gene significance; MS, module significance; STRING, search tool for the retrieval of interacting genes; PPI, protein-protein interaction; DAVID, Database for Annotation, Visualization and Integrated Discovery; DEG, differentially expressed gene; SAM, significance analysis of microarrays

* Corresponding author.
** Correspondence to: Y. Xiao, Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan, China.
E-mail addresses: wangxinghuan@whu.edu.cn (X. Wang), yu.xiao@whu.edu.cn (Y. Xiao).

https://doi.org/10.1016/j.gdata.2017.10.006
Received 10 July 2017; Received in revised form 12 October 2017; Accepted 25 October 2017
Available online 04 November 2017

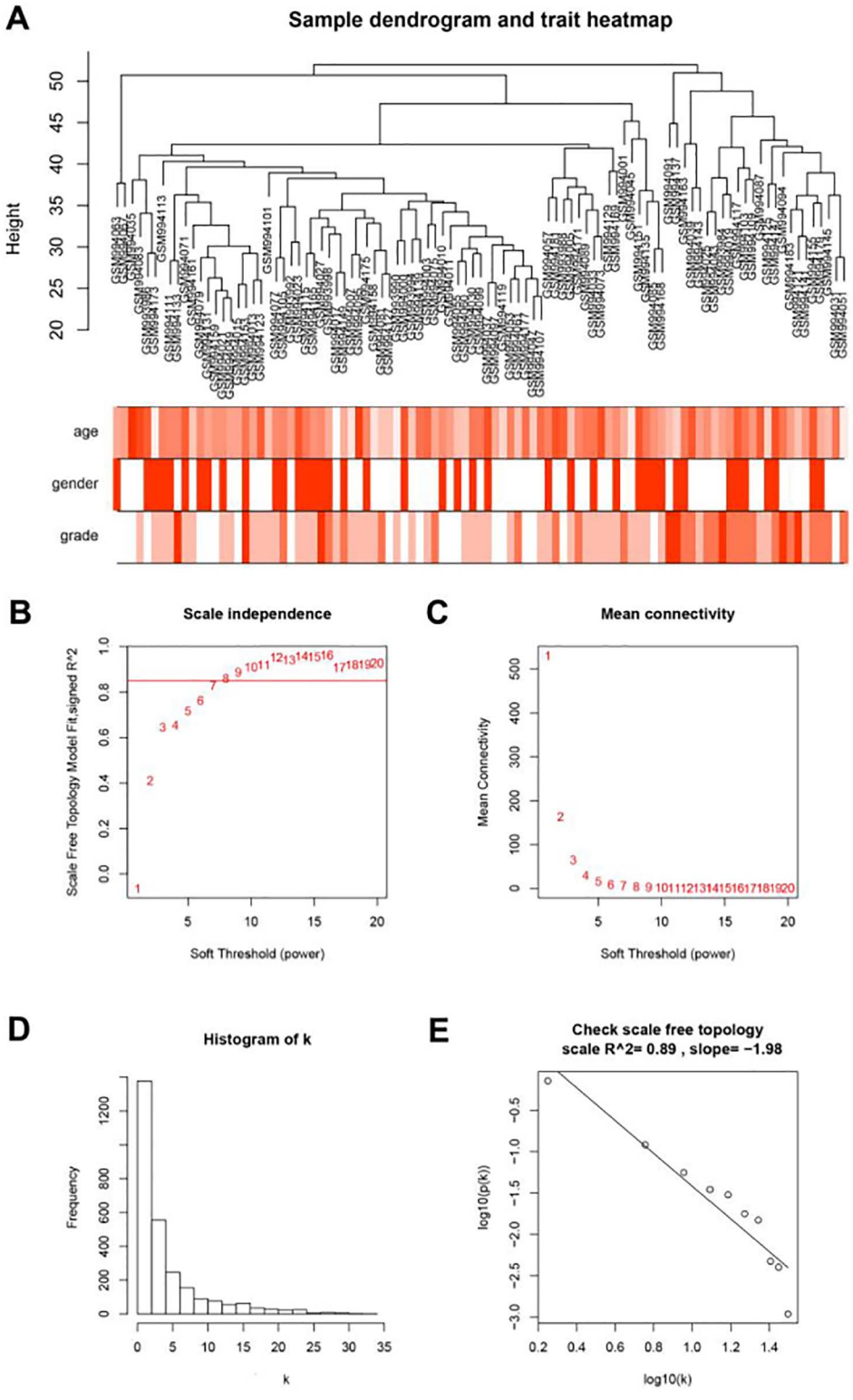

Fig. 1. Clustering dendrogram of 97 tumor samples and the clinical traits. (A) The clustering was based on the expression data of differentially expressed genes between tumor samples and non-tumor samples in ccRCC. The color intensity was proportional to older age and higher Furhman grade. (B-D) Determination of soft-thresholding power in the weighted gene co-expression network analysis (WGCNA). (B) Analysis of the scale-free fit index for various soft-thresholding powers (β). (C) Analysis of the mean connectivity for various soft-thresholding powers. (D) Histogram of connectivity distribution when β = 8. (E) Checking the scale free topology when β = 8. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

applied to finding the hub genes associated with clinical feature in different cancer types [8–10].

In this study, WGCNA and other analysis methods are adopted to jointly analyze clinical information and microarray data of ccRCC patient samples to identify key genes associated with clinical features (age, gender, and tumor grade). These key genes may have important clinical implications and serve as diagnostic and prognostic biomarkers or therapeutic targets.

2. Materials and methods

2.1. Data collection

Expression profiles of mRNA and related clinical data of clear cell renal carcinoma were downloaded from Gene Expression Omnibus (GEO) database (*http://www.ncbi.nlm.nih.gov/geo/*). Dataset GSE40435 performed on Illumina HumanHT-12 V4.0 expression beadchip was used as a training set to construct co-expression networks and identify