

Research Article

Temporal Identification of Dysregulated Genes and Pathways in Clear Cell Renal Cell Carcinoma Based on Systematic Tracking of Disrupted Modules

Shao-Mei Wang,¹ Ze-Qiang Sun,² Hong-Yun Li,² Jin Wang,² and Qing-Yong Liu²

¹Center for Kidney Disease, Jinan Central Hospital Affiliated to Shandong University, Jinan, Shandong 250013, China

²Department of Urinary Surgery, Qianfoshan Hospital Affiliated to Shandong University, 16766 Jingshi Road, Jinan, Shandong 250014, China

Correspondence should be addressed to Qing-Yong Liu; qingyongliu_a@yeah.net

Received 18 June 2015; Revised 31 July 2015; Accepted 11 August 2015

Academic Editor: Konstantin Blyuss

Copyright © 2015 Shao-Mei Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. The objective of this work is to identify dysregulated genes and pathways of ccRCC temporally according to systematic tracking of the dysregulated modules of reweighted Protein-Protein Interaction (PPI) networks. **Methods.** Firstly, normal and ccRCC PPI network were inferred and reweighted based on Pearson correlation coefficient (PCC). Then, we identified altered modules using maximum weight bipartite matching and ranked them in nonincreasing order. Finally, gene compositions of altered modules were analyzed, and pathways enrichment analyses of genes in altered modules were carried out based on Expression Analysis Systematic Explored (EASE) test. **Results.** We obtained 136, 576, 693, and 531 disrupted modules of ccRCC stages I, II, III, and IV, respectively. Gene composition analyses of altered modules revealed that there were 56 common genes (such as *MAPK1*, *CCNA2*, and *GSTM3*) existing in the four stages. Besides pathway enrichment analysis identified 5 common pathways (glutathione metabolism, cell cycle, alanine, aspartate, and glutamate metabolism, arginine and proline metabolism, and metabolism of xenobiotics by cytochrome P450) across stages I, II, III, and IV. **Conclusions.** We successfully identified dysregulated genes and pathways of ccRCC in different stages, and these might be potential biological markers and processes for treatment and etiology mechanism in ccRCC.

1. Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common type of kidney cancer and accounts for approximately 60% to 70% of all renal tumors [1]. Patients with ccRCC comprise a heterogeneous group of patients with variable pathologic stage and grade, used to stratify patients and infer prognosis [2]. However, providing patients with reliable information about anticipated treatment response is challenging due to the molecular heterogeneity of ccRCC [3]. Delineating the pathogenesis of ccRCC by investigating the gene and epigenetic changes and their effects on key molecules and their respective biologic pathways is of crucial importance for the improvement of current diagnostics, prognostics, and drug development [4]. For example, studies suggest that ccRCC is closely associated with tumor suppressor von-Hippel Lindau

(*VHL*) gene mutations that lead to stabilization of hypoxia inducible factors (HIF-1 α and HIF-2 α , also known as HIF1A and EPAS1) in both sporadic and familial forms [5, 6].

With the advances of high-throughput experimental technologies, large amounts of Protein-Protein Interaction (PPI) data are uncovered, which make it possible to study proteins on a systematic level [7, 8]. In addition, a PPI network can be modeled as an undirected graph, where vertices represent proteins and edges represent interactions between proteins, to prioritize disease associated genes or pathways and to understand the modus operandi of disease mechanisms [9, 10]. But it has been noticed that PPI data are often associated with high false positive and false negative rates due to the limitations of the associated experimental techniques and the dynamic nature of protein interaction maps, which may have a negative impact on the performance

of complex discovery algorithms [11]. Many computational approaches have been proposed to assess the reliability of protein interactions data. An iterative scoring method proposed by Liu et al. [12] was selected to evaluate the reliability and predict new interactions, and it has been shown to perform better than other methods. However, studying multiple diseases simultaneously makes it challenging to discern clearly the intricate underlying mechanisms.

In addition, it is important to effectively integrate omics data into such an analysis; for example, Chu and Chen [13] combined PPI and gene expression data to construct a cancer perturbed PPI network in cervical carcinoma to study gain- and loss-of-function genes as potential drug targets. Magger et al. [14] combined PPI and gene expression data to construct tissue-specific PPI networks for 60 tissues and used them to prioritize disease genes. Beyond straightforward scoring genes in the gene regulatory network, it is crucial to study the behavior of modules across specific conditions in a controlled manner to understand the modus operandi of disease mechanisms and to implicate novel genes [15], since some of the important genes may not be identifiable through their own behavior, but their changes are quantifiable when considered in conjunction with other genes (e.g., as modules). What is required, therefore, is systematic tracking gene, pathways, and module behavior across specific conditions in a controlled manner.

Therefore, in this paper, we performed a temporal (stages I, II, III, and IV of ccRCC) analysis between normal controls and ccRCC patients to identify disrupted genes and pathways by systematically tracking the altered modules of reweighted PPI network. To achieve this, we firstly inferred normal and ccRCC cases of different stages PPI networks based on Pearson correlation coefficient (PCC); next, clique-merging algorithm was performed to explore modules in PPI network, and we compared these modules to identify altered modules; then gene composition of these modules was analyzed; finally, pathways enrichment analysis of genes in altered modules was carried out based on Expression Analysis Systematic Explored (EASE) test.

2. Materials and Methods

2.1. Inferring Normal and ccRCC PPI Network

2.1.1. PPI Network Construction. We utilized a dataset of human PPI network, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), which comprised 16730 genes and 1048576 interactions [16]. For STRING, self-loops and proteins without expression value were removed. The remaining largest connected component with score of more than 0.8 was kept as the selected PPI network, which consisted of 8590 genes and 53975 interactions.

2.1.2. Gene Expression Dataset and Dataset Preprocessing. A microarray expression profile, E-GEOD-53757, from Array Express database was selected for ccRCC related analysis. E-GEOD-53757 which existed on Affymetrix GeneChip Human Genome U133 Plus 2.0 Platform was divided into 4 groups according to tumor stage (stages I, II, III, and IV). There were

24, 19, 14, and 15 ccRCC patients at stages I, II, III, and IV, respectively; the number of normal controls in each stage was equaled to its patients' number.

The expression profile was preprocessed by standard methods, consisting of "rma" [17], "quantiles" [18], "mas" [19], and "medianpolish" [17]. To be specific, "rma" method was carried out for background correction to eliminate influences of nonspecific hybridization [17]. The quantile normalization algorithm was a specific case of the transformation $x'_i = F^{-1}(G(x_i))$, where we estimated G by the empirical distribution of each array and F using the empirical distribution of the averaged sample quantiles [18]. Perfect match (PM)/mismatch (MM) correction was conducted by "mas" method [19]. Summarization of the probe data was conducted by "medianpolish" [17]. A multichip linear model was fit to data from each probe set. In particular, for a probe set k with $i = 1, \dots, I_k$ probes and data from $j = 1, \dots, J$ arrays, we fitted the following model, $\log_2(\text{PM}_{ij}^k) = \alpha_i^k + \beta_j^k + \epsilon_{ij}^k$, where α_i was a probe effect and β_j was the \log_2 expression value.

Next, the data were screened by feature filter method of gene filter package, and the number of genes with multiple probes was 20102. At last, we obtained the gene expression value for each gene, including 20102 genes from 144 samples (72 normal controls and 72 ccRCC patients).

2.1.3. Reweighting Gene Interactions by PCC. Gene interactions in network based on ccRCC patients of different stages (stages I, II, III, and IV) and their normal controls were reweighted by PCC, which evaluated the probability of two coexpressed gene pairs. PCC is a measure of the correlation between two variables, giving a value between -1 and $+1$ inclusively [20]. The PCC of a pair of genes (x and y), which encoded the corresponding paired proteins (u and v) interacting in the PPI network, was defined as

$$\text{PCC}(x, y) = \frac{1}{s-1} \sum_{i=1}^s \left(\frac{g(x, i) - \bar{g}(x)}{\sigma(x)} \right) \cdot \left(\frac{g(y, i) - \bar{g}(y)}{\sigma(y)} \right), \quad (1)$$

where s was the number of samples of the gene expression data; $g(x, i)$ or $g(y, i)$ was the expression level of gene x or y in the sample i under a specific condition; $\bar{g}(x)$ or $\bar{g}(y)$ represented the mean expression level of gene x or y ; and $\sigma(x)$ or $\sigma(y)$ represented the standard deviation of expression level of gene x (or y).

The PCC of a pair of proteins (u and v) was defined as the same as the PCC of their corresponding paired genes (x and y), which was $\text{PCC}(u, v) = \text{PCC}(x, y)$. If $\text{PCC}(u, v)$ has a positive value, there is a positive linear correlation between u and v . In addition, we defined PCC of each gene-gene interaction as weight value of the interaction.

2.2. Identifying Modules from the PPI Network. In this paper, module-identification algorithm is based on clique-merging [21, 22] and is similar to the method proposed by Liu et al. [12]. It consisted of three steps; in the first step, it found

all the maximal cliques from the weighted PPI network. Maximal cliques were evaluated by a fast depth-first search with pruning-based algorithm proposed by Tomita et al. [23]. It utilized a depth-first search strategy to enumerate all maximal cliques and effectively pruned nonmaximal cliques during the enumeration process.

In the second step, we assigned a score to each clique; the score of a clique C was defined as its weighted density $d_W(C)$:

$$d_W(C) = \frac{\sum_{u \in C, v \in C} w(u, v)}{|C| \cdot (|C| - 1)}, \quad (2)$$

where $w(u, v)$ was the weight of the interaction between u and v . We ranked these cliques in nonincreasing order of their weighted densities $\{C_1, C_2, \dots, C_k\}$.

Finally, we went through this ordered list repeatedly merging highly overlapping cliques to build modules. For every clique C_i , we repeatedly looked for a clique C_j ($j > i$) such that the overlap $|C_i \cap C_j|/|C_j| \geq t$, where $t = 0.5$ was a predefined threshold for overlapping [15]. If such C_j was found, we calculated the weighted interconnectivity I_w between C_i and C_j as follows:

$$I_w(C_i, C_j) = \sqrt{\frac{\sum_{u \in (C_i - C_j)} \sum_{v \in C_j} w(u, v)}{|C_i - C_j| \cdot |C_j|} \cdot \frac{\sum_{u \in (C_j - C_i)} \sum_{v \in C_i} w(u, v)}{|C_j - C_i| \cdot |C_i|}}. \quad (3)$$

If $I_w(C_i, C_j) \geq t$, then C_j was merged into C_i forming a module; else C_j was discarded.

We captured the effect of differences in interaction weights between normal and ccRCC cases through the weighted density-based ranking of cliques. Weighted density assigned higher rank to larger and stronger cliques. Therefore, we expected cliques with lost proteins or weakened interactions to go down the rankings resulting in altered module generation, thereby capturing changes in modules between normal and ccRCC cases.

2.3. Comparing Modules between Normal and ccRCC Conditions. The approach to compare modules between normal and ccRCC conditions is similar to the method proposed by Srihari and Ragan [15]. In detail, H_N and H_T represented the PPI network of normal controls and ccRCC patients, identifying the sets of modules $S = \{S_1, S_2, \dots, S_m\}$ and $T = \{T_1, T_2, \dots, T_n\}$, respectively. For each $S_i \in S$, module correlation density $d_c(S_i)$ was defined as

$$d_c(S_i) = \frac{\sum_{x, y \in S_i} \text{PCC}((x, y), M)}{|S_i| \cdot (|S_i| - 1)}. \quad (4)$$

Correlation densities of ccRCC modules ($d_c(T_i)$) were calculated similarly.

Disrupted or altered module pairs were evaluated by modeling the set $Y(S, T)$ as maximum weight bipartite matching [24]. Firstly, we build a similarity graph $M = (V_M, E_M)$, where $V_M = \{S \cup T\}$ and $E_M = \cup\{(S_i, T_j) : J(S_i, T_j) \geq t_j, \Delta_C(S_i, T_j) \geq \delta\}$, whereby $J(S_i, T_j) = |S_i \cap T_j|/|S_i \cup T_j|$ was

the Jaccard similarity and $\Delta_C(S_i, T_j) = |d_c(S_i) - d_c(T_j)|$ was the differential correlation density between S_i and T_j , and t_j and δ were thresholds with $2/3$ and 0.05 [15]. $J(S_i, T_j)$ weighted every edge (S_i, T_j) . We next identified the disrupted module pairs $Y(S, T)$ by detecting the maximum weight matching in M , and we ranked them in nonincreasing order of their differential density Δ_C . At last, we inferred genes involved in ccRCC as $\Gamma = \{g: g \in S_i \cup T_j, (S_i, T_j) \in Y(S, T)\}$ and ranked in nonincreasing order of $\Delta_C(S_i, T_j)$. To identify altered modules, we matched normal and ccRCC modules by setting high t_j , which ensured that the module pairs either had the same gene composition or had lost or gained only a few genes.

2.4. Functional Enrichment Analysis. To further investigate the biological functional pathways of genes in altered modules from normal controls and ccRCC, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed by Database for Annotation, Visualization, and Integrated Discovery (DAVID) [25]. KEGG pathways with P value < 0.001 were selected based on EASE test implemented in DAVID. EASE analysis of the regulated genes indicated molecular functions and biological processes unique to each category [26]. The EASE score was used to detect the significant categories. In both of the functional and pathway enrichment analyses, the threshold of the minimum number of genes for the corresponding term > 2 was considered significant for a category

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}}, \quad (5)$$

where n was the number of background genes; a' was the gene number of one gene set in the gene lists; $a' + b$ was the number of genes in the gene list including at least one gene set; $a' + c$ was the gene number of one gene list in the background genes; a' was replaced with $a = a' - 1$.

3. Results

3.1. Analyzing Disruptions in ccRCC PPI Network. We obtained 20102 genes of normal and ccRCC cases after preprocessing and then investigated intersections between these genes' interactions and STRING PPI network and identified PPI networks of normal and ccRCC cases. The normal H_N and ccRCC H_T PPI networks of different stages (stages I, II, III, and IV) displayed equal numbers of nodes (8050) and interactions (49151). Although their interaction scores (weights) were different from each other, as shown in Figure 1, there was no statistical significance between normal and ccRCC cases in different stages in whole level based on Kolmogorov-Smirnov test ($P > 0.05$). However, the score distribution between the ccRCC networks and normal networks was different, especially for stages III and IV in the score distribution 0~0.3 (Figures 1(c) and 1(d)). Examining these interactions more carefully, distributions among different stages were also different, and changes of ccRCC networks and normal networks were more and more obvious from stage I to stage IV.

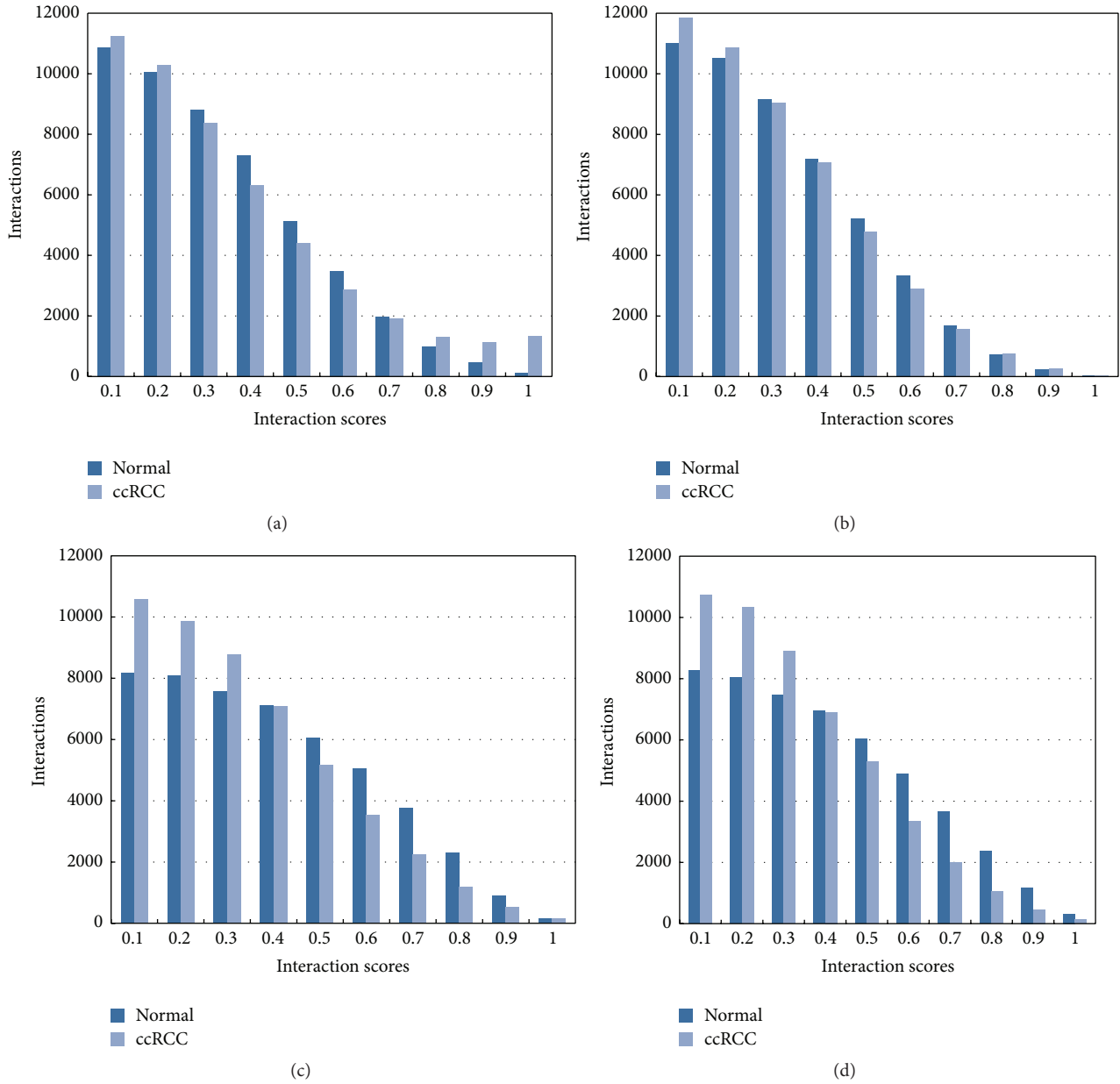


FIGURE 1: Score-wise distributions of interactions: normal versus ccRCC cases. (a) represents stage I of ccRCC, (b) represents stage II, (c) represents stage III, and (d) represents stage IV.

3.2. *Analyzing Disruptions in ccRCC Modules.* Clique-merging algorithm was selected to identify disrupted or altered modules from normal and ccRCC PPI network in this paper. In detail, we performed a comparative analysis between normal N and ccRCC T modules to understand disruptions at the module level. Maximal cliques of normal and ccRCC PPI network were obtained based on fast depth-first algorithm, and maximal cliques with the threshold of nodes > 5 were selected for module analysis. Overall, we noticed that the total number of modules (1895), as well as average module sizes (20.235), was almost the same across the two conditions and four stages. Table 1 showed overall changed rules of weighted interaction density between normal modules and ccRCC modules; we could find that

maximal and average weighted density of normal case was smaller than that of ccRCC for each stage; in detail, the average weighted density of stages III (0.075) and IV (0.089) was a little higher than that of stages I (0.068) and II (0.046), while, in the overall level, the difference of module density scores had no statistical significance between normal and ccRCC cases in different stages with $P > 0.05$. Further, the relationship between modules weighted density distribution and numbers of modules was illustrated in Figure 2. The module numbers were different when the interaction density ranged from 0.05 to 0.25, especially for stages II, III, and IV of ccRCC. These differences might be the reasons of weighted density changes of ccRCC from different stages (Table 1).

TABLE I: Correlations of normal and ccRCC modules of different stages.

Module set	Stage I		Stage II		Stage III		Stage IV	
	Normal	ccRCC	Normal	ccRCC	Normal	ccRCC	Normal	ccRCC
PCC correlation								
Maximal	0.315	0.324	0.254	0.278	0.324	0.339	0.294	0.326
Average	0.068	0.083	0.046	0.074	0.075	0.087	0.089	0.084
Minimum	-0.076	-0.072	-0.073	-0.073	-0.092	-0.074	-0.078	-0.057

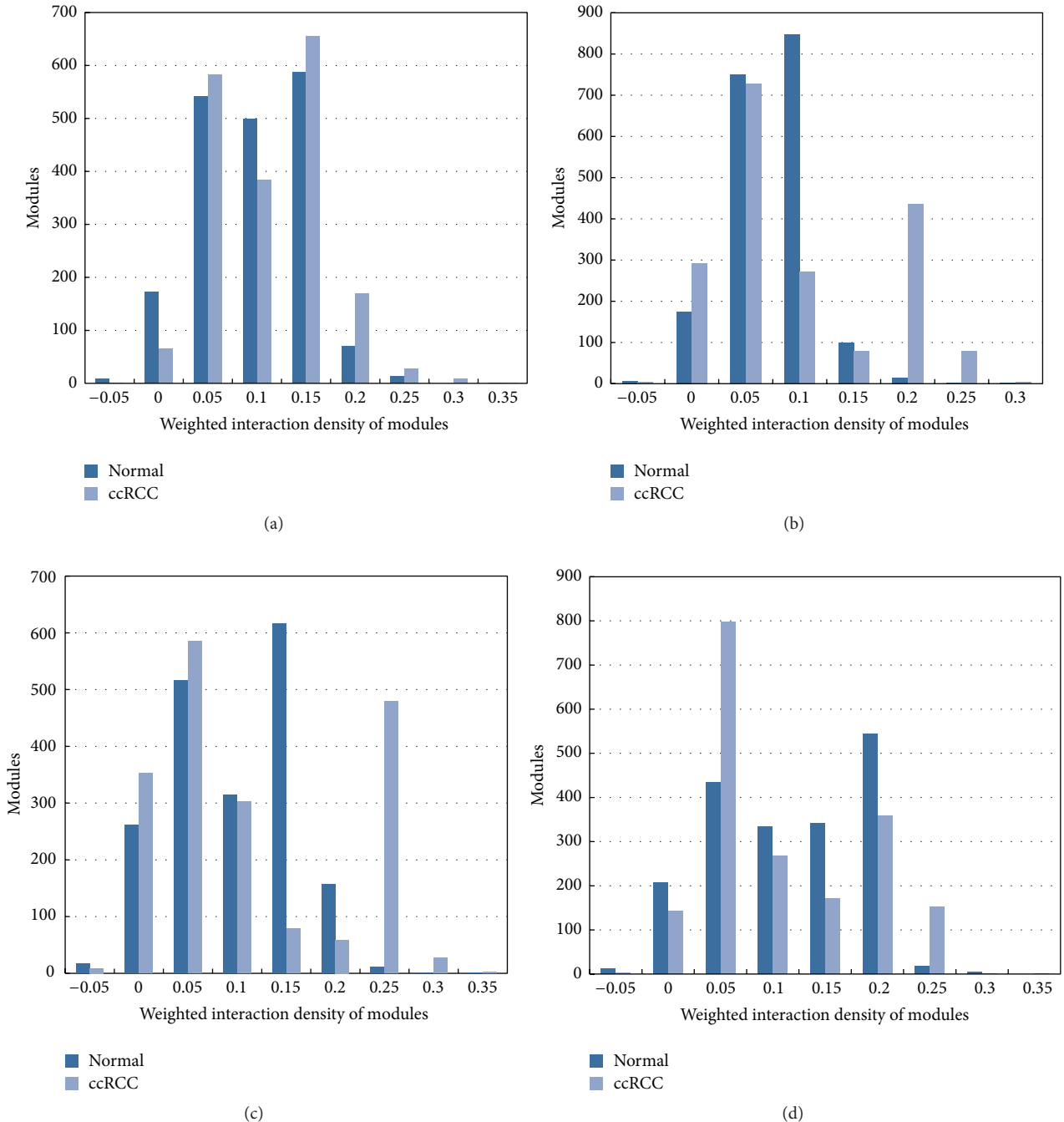


FIGURE 2: Weighted interaction density distribution of modules in normal and ccRCC cases. (a) represents stage I of ccRCC, (b) represents stage II, (c) represents stage III, and (d) represents stage IV.

TABLE 2: Overall conditions of changed module correlation density of ccRCC stages.

ccRCC stages	Changed module correlation density		
	Maximum	Average	Minimum
I	0.255	0.015	-0.195
II	0.254	0.028	-0.192
III	0.253	0.012	-0.246
IV	0.155	-0.006	-0.240

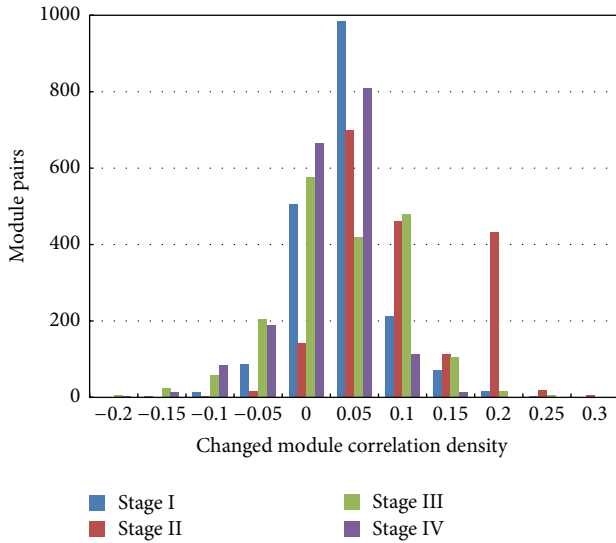


FIGURE 3: Module correlation density distributions of stage I, stage II, stage III, and stage IV.

Next, we obtained disrupted module pairs (ccRCC module and its relative normal module) based on modeling the set $Y(S, T)$ as maximum weight bipartite matching and then calculated their PCC difference values (also called changed module correlation density value). With the thresholds $t_j = 2/3$ and $\delta = 0.05$, the overall conditions of changed module correlation density of stages I, II, III, and IV in ccRCC had no significant difference ($P > 0.05$, Table 2). An overall decrease in maximum correlations of ccRCC modules with deepened stage was observed; besides minimum correlation density of stage III was the smallest among the four stages. In addition, changed module correlation density distributions were shown in Figure 3, and the number of modules was different in the same density interval of four stages, especially in the distribution interval of $-0.05 \sim 0.20$. For stage IV, module distributions firstly increased and then decreased with density increase; the maximum was reached at section of $0 \sim 0.05$.

3.3. In-Depth Analyses of Disrupted Modules. When restricting random inspection correction of modules under condition of $P < 0.01$, we obtained 136, 576, 693, and 531 disrupted modules of stages I, II, III, and IV, respectively. Meanwhile, a total of 1026 genes were obtained of these disrupted modules, in detail, 317 genes of stage I, 450 genes of stage II, 658 genes

TABLE 3: Common genes of disrupted modules based on four ccRCC stages.

Number	Genes
1	MAPK1
2	CDC6
3	CDKN1A
4	SF3B6
5	CPSF3
6	SRSF6
7	SRSF1
8	U2AF1
9	SRSF4
10	CCNB1
11	ESPL1
12	NCAPH
13	KIF11
14	BUB1B
15	CDC20
16	CCNA2
17	CCNB2
18	MAD2L1
19	CENPF
20	ALDH4A1
21	NCBP1
22	MGST1
23	GSTZ1
24	GSTM2
25	GSTM5
26	GSTM3
27	GSTA3
28	SNRPD3
29	CDKN1B
30	NDUFAB1
31	RNPS1
32	ALB
33	LPA
34	GOT1
35	GLUD1
36	FTCD
37	GLUD2
38	ALDH3A2
39	ALDH9A1
40	ALDH1B1
41	ALDH7A1
42	NAGS
43	GATM
44	ASNS
45	ACY3
46	ASPA
47	GOT2
48	ASS1
49	GAD2
50	ALDH2
51	MGST2
52	MGST3
53	GSTO1
54	GSTM4
55	RPA3
56	UPF3B

of stage III, and 690 genes of stage IV. Therefore, 56 common genes existing in four stages were explored (Table 3), such as *MAPK1*, *CCNA2*, and *GSTM3*.



FIGURE 4: Distribution of pathways in stages I, II, III, and IV. Pathways were identified by KEGG with $P < 0.001$. The light green square represented the notion that one pathway did not exist in the stage, while the dark one stood for the notion that the pathway existed in the stage.

As we all know, differentially expressed (DE) gene was usually selected to screen significant genes between normal controls and disease patients; thus we identified 2781 DE genes between normal controls and ccRCC patients of four stages based on Linear Models for Microarray Data package. Taking the intersection of common genes and DE genes into consideration, we obtained 19 genes (*ALB*, *ASS1*, *GSTM3*, *MAD2L1*, *ALDH1B1*, *ALDH4A1*, *MAPK1*, *GSTZ1*, *GATM*, *FTCD*, *CCNA2*, *CENPF*, *GSTM4*, *ASNS*, *CCNBI*, *NAGS*, *ACY3*, *GSTA3*, and *ESPL1*), which might play important roles in the ccRCC development.

Pathway enrichment analysis based on genes in disrupted modules of different stages was performed, and the results within threshold P value < 0.001 were shown in Figure 4; there were 5 common pathways (glutathione metabolism, cell cycle, alanine, aspartate, and glutamate metabolism, arginine and proline metabolism, and metabolism of xenobiotics by cytochrome P450) across stages I, II, III, and IV.

4. Discussions

The objective of this paper is to identify dysregulated genes and pathways in ccRCC from stage I to stage IV according to systematically tracking the dysregulated modules of reweighted PPI networks. We obtained reweighted normal and ccRCC PPI network based on PCC and then identified modules in the PPI network. By comparing normal and ccRCC modules in each stage, we obtained 136, 576, 693, and 531 disrupted modules of stages I, II, III, and IV, respectively. Furthermore, a total of 56 common genes (such as *MAPK1* and *CCNA2*) and 5 common pathways (e.g., cell cycle, glutathione metabolism, and arginine and proline metabolism) of the four stages were explored based on gene composition and pathway enrichment analyses. The common genes and pathways from stage I to stage IV were significant for ccRCC development; if we control these signatures and the biological progress in the early stage of the tumor, there might be positive effects on the therapy.

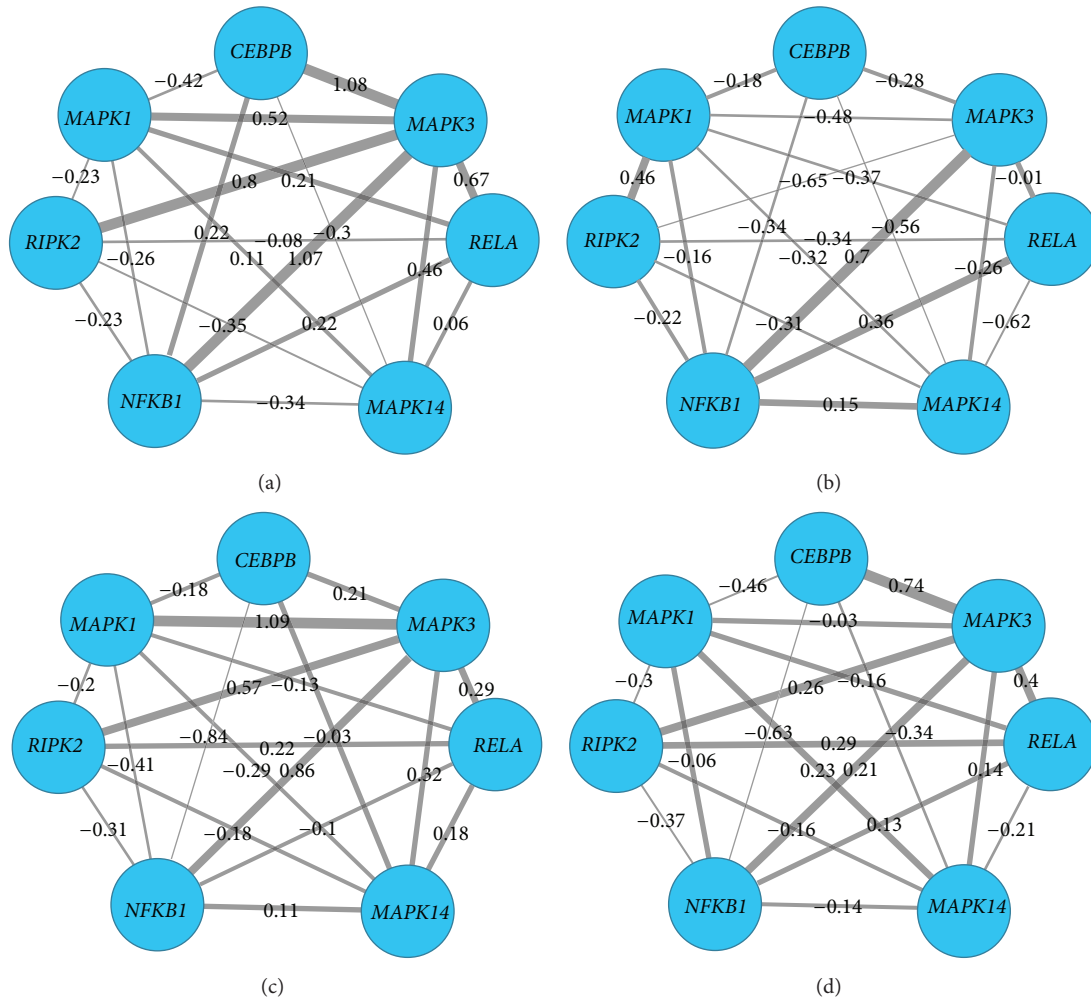


FIGURE 5: Swapping behavior in altered module (*MAPK1*, *CEBPB*, *MAPK3*, *RELA*, *MAPK14*, *NFKB1*, and *RIPK2*). Nodes stood for genes, and edges stood for the interactions of genes. The thickness of the edges represented the interaction scores or expression levels between two genes in the module, more thickness with higher value of expression scores. (a) represents stage I of ccRCC, (b) represents stage II, (c) represents stage III, and (d) represents stage IV.

MAPK1 (mitogen-activated protein kinase 1), which encoded a member of the *MAPK* family, acted as an integration point for multiple biochemical signals and was involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation, and development [27]. Roberts and Der had reported that aberrant regulation of *MAPK* contributed to cancer and other human diseases, such as ccRCC; in particular, the *MAPK* had been the subject of intense research scrutiny leading to the development of pharmacologic inhibitors for the treatment of cancer [28]. Moreover, *MAPK* participant biological processes were key signaling pathways involved in the regulation of normal cell proliferation and differentiation. For example, an increase in the activation of *MAPK* signal transduction pathway was observed as the cancer progresses [29]. *MAPK*/extracellular signal-related kinase pathway was activated in tumors and represented a potential target for therapy [30]. Therefore, ccRCC as a common tumor was related to *MAPK* closely.

Furthermore, we studied gene swapping behaviors in single altered module of four stages, taking *MAPK* family genes related altered modules as an example. As shown in Figure 5, we could discover that, for a module (*MAPK1*, *CEBPB*, *MAPK3*, *RELA*, *MAPK14*, *NFKB1*, and *RIPK2*) in stages I, II, III, and IV, its gene compositions (nodes) were the same, but the interaction scores (edges) were different. The interaction value between *MAPK1* and *MAPK3* was 0.52, -0.48, 1.09, and -0.03 in stages I, II, III, and IV, respectively, and there was a weak correlation of the two genes in stage II. It might explain differences of modules and existence of dysregulated modules. Swapping behavior in the altered module (*CCNA2*, *MND1*, *CDC45*, *RFC4*, *CCNB1*, and *CDK4*) was shown in Figure 6.

CCNA2, cyclin A2, was expressed in dividing somatic cells and regulated cell cycle progression by interacting with cyclin-dependent kinase (*CDK*) kinases [31]. Consistent with its role as a key cell cycle regulator, expression of *CCNA2*

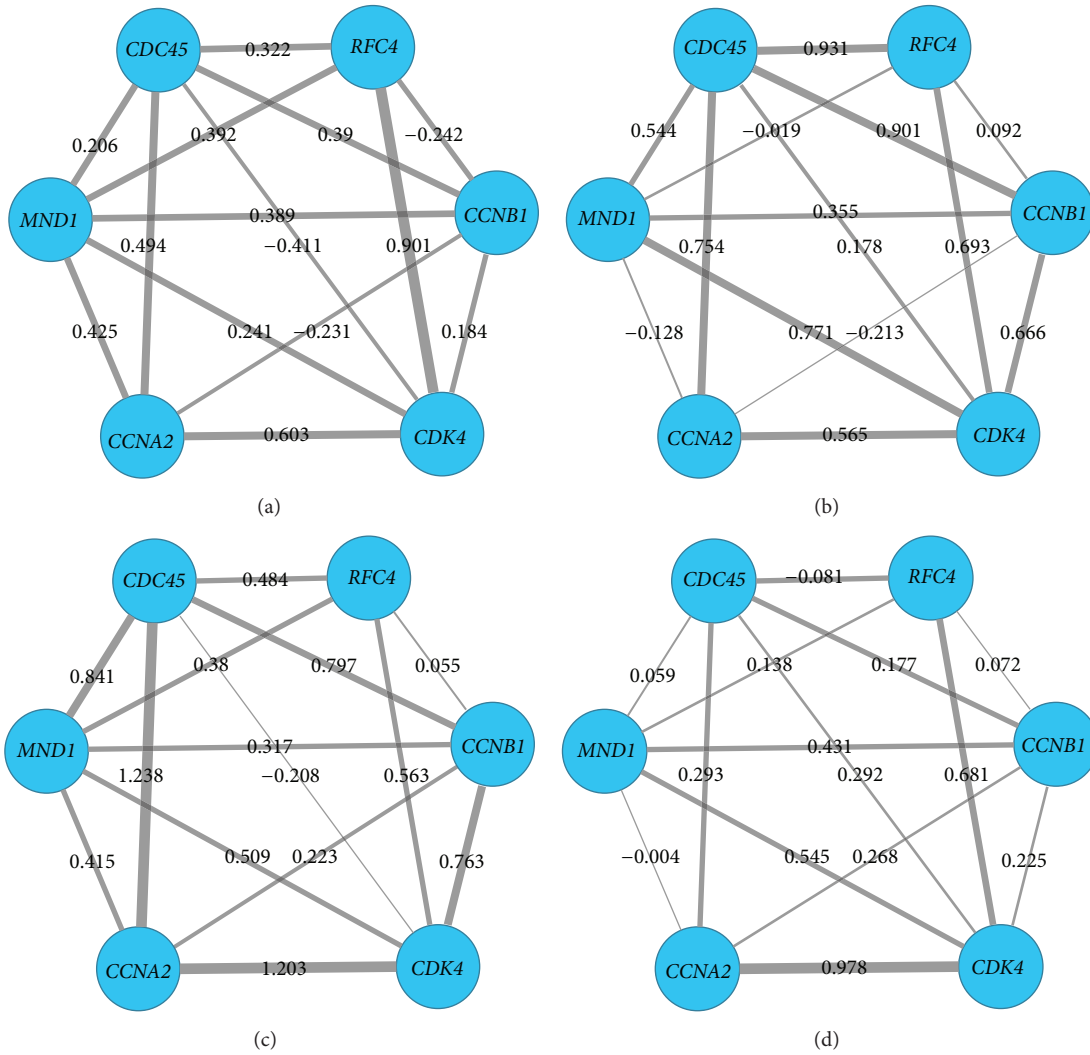


FIGURE 6: Swapping behavior in altered module (*CCNA2*, *CDK4*, *CDC45*, *RFC4*, *CCNB1*, and *MND1*). Nodes stood for genes, and edges stood for the interactions of genes. The thickness of the edges represented the interaction scores or expression levels between two genes in the module, more thickness with higher value of expression scores. (a) represents stage I of ccRCC, (b) represents stage II, (c) represents stage III, and (d) represents stage IV.

was found to be elevated in a variety of tumors such as breast, cervical, liver, and kidney tumors [32]. It was not clear whether increased expression of *CCNA2* was a cause or result of tumorigenesis; *CCNA2-CDK* contributed to tumorigenesis by the phosphorylation of oncoproteins or tumor suppressors [33]. In our paper, we had also proved that the correlation value between *CCNA2* and *CDK4* of the four stages was 0.603, 0.565, 1.203, and 0.978 in sequence (Figure 6). We might infer that cell cycle played a medium role in correlations of *CCNA2* and cancers; thus cell cycle was discussed next.

Cell cycle is the series of events that take place in a cell leading to its division and duplication, and dysregulation of the cell cycle components may lead to tumor formation [34]. It had been reported that alterations in activated proteins (cyclins and cyclin-dependent kinases, etc.), which led to failure of cell cycle arrest, may thus serve as markers of a more malignant phenotype and cell cycle-related genes aided

in discrimination of atypical adenomatous hyperplasia from early adenocarcinoma [35]. Chen et al. demonstrated that cell cycle progression effects on *NF-κB* activity represented a molecular basis underlying the aggressive tumor behavior [36]. Besides, cell cycle checkpoint inactivation allowed DNA replication in aneuploid cells and may favor oncogenic genomic [37], and a cell cycle regulator is potentially involved in genesis of many tumor types, such as ccRCC [38]. We could conclude that cell cycle played a key role in the ccRCC progress.

Our results also showed that ccRCC had close relationship with metabolism pathways, such as glutathione metabolism and arginine and proline metabolism. Glutathione metabolism which played both protective and pathogenic roles in cancers was crucial in the removal and detoxification of carcinogens [39]. And the present review highlighted the role of glutathione and related cytoprotective

effects in the susceptibility to carcinogenesis and in the sensitivity of tumors to the cytotoxic effects of anticancer agents [40]. Recently, Hao et al. discovered that three significant pathways related to ccRCC, namely, arginine and proline metabolism, aldosterone-regulated sodium reabsorption, and oxidative phosphorylation, were observed [41]. Arginine/proline metabolism is a significant pathway in ccRCC that had been discovered by Perroud et al. previously [42], and the results were in accordance with our analysis.

5. Conclusions

In conclusion, we successfully identified dysregulated genes (such as *MAPK1* and *CCNA2*) and pathways (such as cell cycle, glutathione metabolism, and arginine and proline metabolism) of ccRCC in different stages, and these genes and pathways might be potential biological markers and processes for treatment and etiology mechanism in ccRCC.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank the editors and reviewers for the insightful comments and suggestions on this work. They also thank Jinan Evidence Based Medicine Science-Technology Center for the help of statistical and computational advice.

References

- [1] H. An, Y. Zhu, L. Xu, L. Chen, Z. Lin, and J. Xu, "Notch1 predicts recurrence and survival of patients with clear-cell renal cell carcinoma after surgical resection," *Urology*, vol. 85, no. 2, pp. 483.e9–483.e14, 2015.
- [2] F. Vincenzo, G. Novara, A. Galfano et al., "The 'stage, size, grade and necrosis' score is more accurate than the University of California Los Angeles Integrated Staging System for predicting cancer-specific survival in patients with clear cell renal cell carcinoma," *BJU International*, vol. 103, no. 2, pp. 165–170, 2009.
- [3] V. R. Dondeti, B. Wubbenhorst, P. Lal et al., "Integrative genomic analyses of sporadic clear cell renal cell carcinoma define disease subtypes and potential new therapeutic targets," *Cancer Research*, vol. 72, no. 1, pp. 112–121, 2012.
- [4] A. H. Girgis, V. V. Iakovlev, B. Beheshti et al., "Multilevel whole-genome analysis reveals candidate biomarkers in clear cell renal cell carcinoma," *Cancer Research*, vol. 72, no. 20, pp. 5273–5284, 2012.
- [5] L. W. Marston, T. A. Rouault, J. Mitchell, and M. K. Cherukuri, "Nitroxide therapy for the treatment of von Hippel–Lindau disease (VHL) and renal clear cell carcinoma (RCC)," Google Patents, 2014.
- [6] T. Tanaka, T. Torigoe, Y. Hirohashi et al., "Hypoxia-inducible factor (HIF)-independent expression mechanism and novel function of HIF prolyl hydroxylase-3 in renal cell carcinoma," *Journal of Cancer Research and Clinical Oncology*, vol. 140, no. 3, pp. 503–513, 2014.
- [7] F. Jordán, T.-P. Nguyen, and W.-C. Liu, "Studying protein-protein interaction networks: a systems view on diseases," *Briefings in Functional Genomics*, vol. 11, no. 6, pp. 497–504, 2012.
- [8] S. Srihari and H. W. Leong, "Temporal dynamics of protein complexes in PPI networks: a case study using yeast cell cycle dynamics," *BMC Bioinformatics*, vol. 13, article S16, 2012.
- [9] J. Zhao, S. Zhang, L.-Y. Wu, and X.-S. Zhang, "Efficient methods for identifying mutated driver pathways in cancer," *Bioinformatics*, vol. 28, no. 22, pp. 2940–2947, 2012.
- [10] S. Srihari, C. H. Yong, A. Patil, and L. Wong, "Methods for protein complex prediction and their contributions towards understanding the organisation, function and dynamics of complexes," *FEBS Letters*, 2015.
- [11] C. Wu, J. Zhu, and X. Zhang, "Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes," *BMC Bioinformatics*, vol. 13, article 182, 2012.
- [12] G. Liu, J. Li, and L. Wong, "Assessing and predicting protein interactions using both local and global network topological metrics," *Genome Informatics*, vol. 21, pp. 138–149, 2008.
- [13] L.-H. Chu and B.-S. Chen, "Construction of a cancer-perturbed protein-protein interaction network for discovery of apoptosis drug targets," *BMC Systems Biology*, vol. 2, article 56, 2008.
- [14] O. Magger, Y. Y. Waldman, E. Ruppim, and R. Sharan, "Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks," *PLoS Computational Biology*, vol. 8, no. 9, Article ID e1002690, 2012.
- [15] S. Srihari and M. A. Ragan, "Systematic tracking of dysregulated modules identifies novel genes in cancer," *Bioinformatics*, vol. 29, no. 12, pp. 1553–1561, 2013.
- [16] D. Szklarczyk, A. Franceschini, M. Kuhn et al., "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Research*, vol. 39, no. 1, pp. D561–D568, 2011.
- [17] R. A. Irizarry, B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed, "Summaries of Affymetrix GeneChip probe level data," *Nucleic Acids Research*, vol. 31, no. 4, article e15, 2003.
- [18] B. M. Bolstad, R. A. Irizarry, M. Åstrand, and T. P. Speed, "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias," *Bioinformatics*, vol. 19, no. 2, pp. 185–193, 2003.
- [19] B. Ben, *affy: Built-in Processing Methods*, 2013.
- [20] N. Gerhard, "Pearson correlation coefficient," in *Dictionary of Pharmaceutical Medicine*, p. 132, Springer, 2009.
- [21] G. Liu, L. Wong, and H. N. Chua, "Complex discovery from weighted PPI networks," *Bioinformatics*, vol. 25, no. 15, pp. 1891–1897, 2009.
- [22] S. Srihari and H. W. Leong, "A survey of computational methods for protein complex prediction from protein interaction networks," *Journal of Bioinformatics and Computational Biology*, vol. 11, no. 2, Article ID 1230002, 2013.
- [23] E. Tomita, A. Tanaka, and H. Takahashi, "The worst-case time complexity for generating all maximal cliques and computational experiments," *Theoretical Computer Science*, vol. 363, no. 1, pp. 28–42, 2006.
- [24] H. N. Gabow, "An efficient implementation of Edmonds' algorithm for maximum matching on graphs," *Journal of the ACM*, vol. 23, no. 2, pp. 221–234, 1976.
- [25] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2008.

- [26] G. Ford, Z. Xu, A. Gates, J. Jiang, and B. D. Ford, "Expression Analysis Systematic Explorer (EASE) analysis reveals differential gene expression in permanent and transient focal stroke rat models," *Brain Research*, vol. 1071, no. 1, pp. 226–236, 2006.
- [27] E. Siljamäki, L. Raiko, M. Toriseva et al., "P38 δ mitogen-activated protein kinase regulates the expression of tight junction protein ZO-1 in differentiating human epidermal keratinocytes," *Archives of Dermatological Research*, vol. 306, no. 2, pp. 131–141, 2014.
- [28] P. J. Roberts and C. J. Der, "Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer," *Oncogene*, vol. 26, no. 22, pp. 3291–3310, 2007.
- [29] I. M. Subbiah, G. Varadhachary, A. M. Tsimberidou et al., "Abstract 4700: One size does not fit all: Fingerprinting advanced carcinoma of unknown primary through comprehensive profiling identifies aberrant activation of the PI3K and MAPK signaling cascades in concert with impaired cell cycle arrest," *Cancer Research*, vol. 74, no. 19, article 4700, 2014.
- [30] B. H. O'Neil, L. W. Goff, J. S. W. Kauh et al., "Phase II study of the mitogen-activated protein kinase 1/2 inhibitor selumetinib in patients with advanced hepatocellular carcinoma," *Journal of Clinical Oncology*, vol. 29, no. 17, pp. 2350–2356, 2011.
- [31] A. Cribier, B. Descours, A. L. C. Valadão, N. Laguette, and M. Benkirane, "Phosphorylation of SAMHD1 by cyclin A2/CDK1 regulates its restriction activity toward HIV-1," *Cell Reports*, vol. 3, no. 4, pp. 1036–1043, 2013.
- [32] C. H. Yam, T. K. Fung, and R. Y. C. Poon, "Cyclin A in cell cycle control and cancer," *Cellular and Molecular Life Sciences*, vol. 59, no. 8, pp. 1317–1326, 2002.
- [33] S. H. Woo, S.-K. Seo, S. An et al., "Implications of caspase-dependent proteolytic cleavage of cyclin A1 in DNA damage-induced cell death," *Biochemical and Biophysical Research Communications*, vol. 453, no. 3, pp. 438–442, 2014.
- [34] V. Hirt Bartholomäus, *Mathematical Modelling of Cell Cycle and Telomere Dynamics*, University of Nottingham, 2013.
- [35] Y. F. Ho, S. A. Karsani, W. K. Yong, and S. N. Abd Malek, "Induction of apoptosis and cell cycle blockade by helichrysetin in A549 human lung adenocarcinoma cells," *Evidence-Based Complementary and Alternative Medicine*, vol. 2013, Article ID 857257, 10 pages, 2013.
- [36] G. Chen, M. S. Bhojani, A. C. Heaford et al., "Phosphorylated FADD induces NF- κ B, perturbs cell cycle, and is associated with poor outcome in lung adenocarcinomas," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 35, pp. 12507–12512, 2005.
- [37] J. D. Gordan, P. Lal, V. R. Dondeti et al., "HIF- α effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma," *Cancer Cell*, vol. 14, no. 6, pp. 435–446, 2008.
- [38] Y. Yamada, H. Hidaka, N. Seki et al., "Tumor-suppressive microRNA-135a inhibits cancer cell proliferation by targeting the c-MYC oncogene in renal cell carcinoma," *Cancer Science*, vol. 104, no. 3, pp. 304–312, 2013.
- [39] G. K. Balendiran, R. Dabur, and D. Fraser, "The role of glutathione in cancer," *Cell Biochemistry and Function*, vol. 22, no. 6, pp. 343–352, 2004.
- [40] N. Traverso, R. Ricciarelli, M. Nitti et al., "Role of glutathione in cancer progression and chemoresistance," *Oxidative Medicine and Cellular Longevity*, vol. 2013, Article ID 972913, 10 pages, 2013.
- [41] J.-F. Hao, K.-M. Ren, J.-X. Bai et al., "Identification of potential biomarkers for clear cell renal cell carcinoma based on microRNA-mRNA pathway relationships," *Journal of Cancer Research and Therapeutics*, vol. 10, pp. C167–C172, 2014.
- [42] B. Perroud, J. Lee, N. Valkova et al., "Pathway analysis of kidney cancer using proteomics and metabolic profiling," *Molecular Cancer*, vol. 5, article 64, 2006.