Identification of genes associated with renal cell carcinoma using gene expression profiling analysis

TING YAO1*, QINFU WANG2*, WENYONG ZHANG3, AIHONG BIAN4 and JINPING ZHANG5

1Physical Examination Center; Departments of 2Chronic Non-Communicable Diseases Control and Prevention, 3Health Education, 4Health Inspection and 5Communicable Diseases Control and Prevention, Laiwu Center for Disease Control and Prevention, Laiwu, Shandong 271100, P.R. China

Received March 29, 2015; Accepted April 22, 2016

DOI: 10.3892/ol.2016.4573

Abstract. Renal cell carcinoma (RCC) is the most common type of kidney cancer in adults and accounts for ~80% of all kidney cancer cases. However, the pathogenesis of RCC has not yet been fully elucidated. To interpret the pathogenesis of RCC at the molecular level, gene expression data and bio-informatics methods were used to identify RCC associated genes. Gene expression data was downloaded from Gene Expression Omnibus (GEO) database and identified differentially coexpressed genes (DCGs) and dysfunctional pathways in RCC patients compared with controls. In addition, a regulatory network was constructed using the known regulatory data between transcription factors (TFs) and target genes in the University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu) and the regulatory impact factor of each TF was calculated. A total of 258,0427 pairs of DCGs and target genes were identified. The regulatory network contained 1,525 pairs of regulatory associations between 126 TFs and 1,259 target genes and these genes were mainly enriched in cancer pathways, ErbB and MAPK. In the regulatory network, the 10 most strongly associated TFs were FOXC1, GATA3, ESR1, FOXL1, PATZ1, MYB, STAT5A, EGR2, EGR3 and PELP1. GATA3, ERG and MYB serve important roles in RCC while FOXC1, ESR1, FOXL1, PATZ1, STAT5A and PELP1 may be potential genes associated with RCC. In conclusion, the present study constructed a regulatory network and screened out several TFs that may be used as molecular biomarkers of RCC. However, future studies are needed to confirm the findings of the present study.

Introduction

Kidney cancer that forms in tissues of the kidney is not a single disease, instead it comprises a number of different types of cancer of which renal cell carcinoma (RCC) is the most common type in adults, responsible for ~80-90% of cases (1,2). The diagnosis of RCC is a challenging and difficult task, and RCC is one of the most therapy-resistant types of cancer. RCC is regarded as a highly lethal cancer and ~35% of patients succumb to the disease after 5 years (3).

Understanding the pathogenesis and biological mechanism of RCC may improve the current diagnosis, treatment and prognosis of RCC. However, the pathogenesis of RCC is extremely complex and remains to be fully elucidated. Smoking, obesity and mutations in specific genes increase the risk of developing RCC (4-6). The best characterized oncogenic gene in human RCC is the tumor suppressor gene von Hippel-Lindau (VHL). VHL along with elongin B, elongin C and cullin 2 form a E3 ubiquitin-ligase complex, and are considered to serve an important role in RCC (7,8). The VHL complex targets hypoxia-inducible factor (HIF) for ubiquitin-mediated degradation (9). Mutations in VHL can result in the over accumulation of HIF, and its target genes such as VEGF (vascular endothelial growth factor), PDGF (platelet derived growth factor) and EGRF (epidermal growth factor receptor), thus resulting in carcinogenesis (10-12).

Mutations of either tuberous sclerosis 1 (TSC1) or TSC2 are associated with aberrant activation of mammalian target of rapamycin (mTOR) pathway, which increases the risk of RCC (13,14). Owing to the above reasons, VEGF, PDGF, EGRF and mTOR may serve as potential target molecules for the treatment of RCC. In addition, dysregulation of membrane MHC class I chain-related gene A (MICA), cyclooxygenase-1 (COX1), TGF-β-activated Kinase-1 (TAK1), and cell division cycle 25B (CDC25B) serve critical roles in RCC progression (15-18). Previous studies have predominantly focused on single genes, and paid little attention to the dysregulation of transcription factors or differentially coexpressed genes; therefore the present study aimed to address this issue to provide novel insights into RCC.

DNA microarray has previously been used to identify gene expression patterns in RCC (19,20). In order to achieve a more comprehensive understanding of the molecular mecha-
nisms underlying RCC, gene expression profiling analysis was applied to identify DCGs and regulatory network analysis was used to identify potential associations.

**Materials and methods**

**Affymetrix microarray data.** The gene expression profile dataset GSE6344 (21), including 10 RCC samples and 10 patient-matched normal control samples, was obtained from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. The probe-level data was converted into the corresponding gene symbol using the annotation information derived from platform GPL96. For genes corresponding with multiple probe sets that had multiple expression values, the mean expression values of those probe sets were obtained. As a result, 12,754 genes were obtained from this dataset.

**Construction of regulatory network.** The data of TF binding sites of human genome hg18 and the coordinates of genes were downloaded from UCSC (http://genome.ucsc.edu). TF binding sites within the area from 1 kb upstream to 0.5 kb downstream of the transcription initiation site of target genes were selected to construct the regulatory network.

**Differentially coexpressed genes analysis.** For any pair of genes (X,Y), the absolute difference (Diff) of the Pearson's correlation coefficient in the normal state (P-normal) and in the cancer state (P-tumor) was calculated. The pair of genes was regarded as DCGs only when their Diff was >1. Diff value >1 indicated that the pair of genes had a negative correlation (P-normal ∈ [-1,0] ) and a positive correlation (P-tumor ∈ [0,1]) under the normal and tumor state respectively, or vice versa. The difference was calculated using the following equation 1: Diff=abs (r1ij-r2ij). In this equation, r1ij and r2ij indicate the Pearson correlation coefficients between gene i and gene j under the normal state and the state of cancer, respectively.

**Measurement of RIF.** Regulatory impact factors (RIF) (22), which is a robust and effective methodology to identify the regulatory impact factor of TF, was applied to identify the TF with the largest contribution to differential expression of genes in two biological conditions. RIF was calculated using the following equation 2:

\[
RIF_i = \frac{1}{n_{de}} \sum_{j=1}^{n_{de}} [(e_1 \times r_{1j})^2 - (e_2 \times r_{2j})^2]
\]

In this equation, n_{de} is the number of DEGs; e1 and e2j indicate the expression value of the DEG in conditions 1 and 2, respectively; r_{1j} and r_{2j} indicate the correlation coefficient for the i TF and the j DEG in conditions 1 and 2, respectively.

**Pathway enrichment analysis.** For functional analysis of the large gene lists in the regulatory network, the DCGs were inputted into Database for Annotation, Visualization and Integrated Discovery (DAVID) (23) for Kyoto Encyclopedia of Genes and Genomes (KEGG) (24) pathway enrichment analysis. By calculating the hypergeometric test P-value for probability of random association between a given list of genes and a pathway, DAVID identifies canonical pathways associated with this set of genes. FDR <0.05 was used as the cutoff criteria.

**Results**

**Identification of differentially coexpressed genes in RCC.** The gene expression profile dataset GSE6344 was downloaded from the GEO database and formula 1 was used to identify DCGs with Diff >1 between 10 RCC samples and 10 control samples. Finally, a total of 2,580,427 DCGs were screened out (Table I).

**Construction of regulatory network.** Based on the known regulatory data from UCSC, TFs and their corresponding target genes from DCGs were selected to construct a regulatory network. The network contained a total of 1,525  pairs of regulatory associations between 126 TFs and 1,259 target genes. Using Cytoscape (25), the regulatory associations were integrated and visualized in Fig. 1.

**KEGG pathway enrichment.** The DCGs with FDR <0.05 were inputted into DAVID for KEGG pathway enrichment analysis. The results are presented in Table II, from which it was identified that DCGs were predominantly enriched in cancer pathways, ErbB, mitogen-activated protein kinase (MAPK) and other important pathways.

### Table I. Part of the differentially co-expressed genes.

| Gene1     | Gene2 | Diff   |
|-----------|-------|--------|
| AAGAB     | A1CF  | 1.031645 |
| ABCD4     | A1CF  | 1.0116908 |
| ACCN2     | A1CF  | 1.071472 |
| ACTR5     | A1CF  | 1.039394 |
| ADAM22    | A1CF  | 1.014619 |
| AHCTF1    | A1CF  | 1.194273 |
| AIP       | A1CF  | 1.130951 |
| AK2       | A1CF  | 1.069488 |
| ALKBH1    | A1CF  | 1.034613 |
| AMD1      | A1CF  | 1.083589 |
| AMELX     | A1CF  | 1.278415 |
| AMH       | A1CF  | 1.040918 |
| ANKRD12   | A1CF  | 1.137963 |

Diff indicates the absolute difference of Pearson’s correlation coefficient. AAGAB, α- and γ-adaptin binding protein; ABCD4, adenosine triphosphate binding cassette subfamily D member 4; ACCN2, acid-sensing (proton-gated) ion channel 1; ACTR5, ARPS actin-related protein 5 homolog (yeast); ADAM22, ADAM metallopeptidase domain 22; AHCTF1, AT-hook containing transcription factor 1; AIP, aryl hydrocarbon receptor interacting protein; AK2, adenylyl kinase 2; ALKBH1, AlkB homolog 1, histone H2A dioxygenase; AMD1, adenosylmethionine decarboxylase 1; AMELX, amelogenin, X-Linked; AMH, anti-mullerian hormone; ANKRD12, ankyrin repeat domain 12; A1CF, APOBEC1 complementation factor.
Analysis of transcription factor impact. First, total 4,793 differentially expressed genes (DEGs) with FDR <0.05 were identified between normal and tumor samples by linear models for microarray data (limma) method (26). Subsequently, 469 overlapping DEGs were collected by comparing these 4,793 DEGs with the 1,259 target genes in the network. To further investigate which TFs were significant, the RIF of each TF targeting to the overlapping DEGs was targeted. The top 10 were forkhead box C1 (FOXC1), GATA-binding protein 3 (GATA3), estrogen receptor 1 (ESR1), FOXL1, POZ (BTB) and AT hook containing zinc finger 1 (PATZ1), v-myb avian myeloblastosis viral oncogene homolog (MYB), signal transducer and activator of transcription 5A (STAT5A), early growth response 2 (EGR2), EGR3 and proline, glutamate and leucine rich protein 1 (PELP1) (Table III). Of these TFs, GATA3, MYB, EGR2, and EGR3 have previously been identified to be associated with RCC and the regulatory associations of them with their targets are presented in Fig. 2. Occurrence of RCC is likely caused by the abnormal changes of these regulatory associations.
Molecular biomarkers are useful to improve diagnosis, clinical predictive capability and novel therapeutic efficacy. Because the emergence of microarray technology makes it possible to investigate the expression levels of thousands of genes simultaneously (27), it has been widely used in discovery of disease biomarkers (28-30). However, the majority of previous studies are based on a single gene differential expression analysis. In the current study, differentially coexpressed genes analysis was used to find the dysregulated gene pairs in RCC. The differentially coexpressed genes analysis provides novel analytical perspective to studies including identification of differentially coexpressed genes and marker genes of disease, construction of differential coexpression network and analysis of module (31,32).

Besides, a regulatory network was constructed and the RIF score of each TF targeting to the DEGs with FDR <0.05 was calculated. Finally, the top 10 ranked TFs were identified which were FOXC1, GATA3, ESR1, FOXL1, PATZ1, MYB, STAT5A, EGR2, EGR3 and PELP1. They may be potential molecular biomarkers of RCC.

Table III. The top 10 ranked TFs.

| TF        | RIF score | RIF rank |
|-----------|-----------|----------|
| FOXC1     | 7.804438725 | 1        |
| GATA3     | 6.908779522 | 2        |
| ESR1      | 6.32301186  | 3        |
| FOXL1     | 4.242514268 | 4        |
| PATZ1     | 3.800727043 | 5        |
| MYB       | 3.507929833 | 6        |
| STAT5A    | 3.467483166 | 7        |
| EGR2      | 3.361578969 | 8        |
| EGR3      | 3.337751915 | 9        |
| PELP1     | 2.818195935 | 10       |

TF represents the transcription factor in the regulatory network. RIF represents the regulatory impact factor of TF. Rank represents the impact rank of TF. TF, transcription factor; FOXC1, forkhead box C1; GATA3, GATA-binding protein 3; ESR1, estrogen receptor 1, FOXL1, forkhead box L1; PATZ1, POZ (BTB) and AT hook containing zinc finger 1; MYB, v-myb avian myeloblastosis viral oncogene homolog, STAT5A, signal transducer and activator of transcription 5A, EGR2/3, early growth response 2 or 3; PELP1, proline, glutamate and leucine rich protein 1.

Discussion

Molecular biomarkers are useful to improve diagnosis, clinical predictive capability and novel therapeutic efficacy. Because the emergence of microarray technology makes it possible to investigate the expression levels of thousands of genes simultaneously (27), it has been widely used in discovery of disease biomarkers (28-30). However, the majority of previous studies are based on a single gene differential expression analysis. In the current study, differentially coexpressed genes analysis was used to find the dysregulated gene pairs in RCC.
ated genes, and the PATZ1, STAT5A and PELP1 are hormone associated genes. The latest article reported that hormones may inhibit the occurrence of renal carcinoma (40), so that PATZ1, STAT5A and PELP1 may serve as potential renal cancer-associated genes.

Table II identifies genes of the regulatory network which were enriched in various cancer pathways, just as the above discussion that the top 10 TFs were involved in a number of types of cancers. In addition, ErbB and MAPK pathways were overrepresented. The MAPK signaling pathways serve vital roles in cell proliferation and differentiation. Recently, Huang et al (41) reported that suppression of one or more MAPK signaling pathways by inhibitor of MAPK kinases (MKKs) reduced RCC cell proliferation in vitro and inhibited RCC growth in vivo. The ErbB protein family is a family containing 4 structurally associated receptor tyrosine kinases, ErbB1 (also termed EGFR), ErbB2, ErbB3 and ErbB4. ErbB signaling pathway has been implicated in the development of a wide variety of types of tumor, including RCC (42,43). Further analysis of these pathways will contribute to an improved understanding of the roles of the differentially expressed genes and the underlying molecular mechanism of RCC.

In conclusion, the present study used microarray gene expression profiling and regulatory network analysis to explore the molecular mechanism of RCC. The top 10 ranked TFs were identified, which were FOXC1, GATA3, ESR1, FOXL1, PATZ1, MYB, EGR2, EGR3 and PELP1. GATA3, ERG and MYB are considered RCC associated genes while FOXC1, STAT5A and PELP1 may also be potential genes associated with RCC. The present study indicates that the above TFs may be used as biomarkers of RCC for accurate diagnosis, prognosis or as predictive markers for treatment efficiency. However, further experiments are needed to confirm these result.

References

1. Yi Z, Fu Y, Zhao S, Zhang X and Ma C: Differential expression of miRNA patterns in renal cell carcinoma and nontumorous tissues. J Cancer Res Clin Oncol 136: 855-862, 2010.
Pathway signature and cellular differentiation in clear cell renal cell carcinoma

Lichtenfels R: Lottspeich F, Marincola FM, Baumgärtner M, Atkins D and Yu XY, Zhang Z, Zhang GJ, Guo KF and Kong CZ: Asian Pac J Cancer Prev potential prognostic marker for clear cell renal cell carcinoma genease-1 correlates with poor prognosis in renal cell carcinoma Gnarra Linehan of renal cell carcinoma Lipworth carcinoma

Constitutive activation of hypoxia -inducible genes related to expression of hypoxia-inducible factor-Ialpha in clear cell renal carcinomas. Cancer Res 61: 5215-5222, 2001.

Koenig HC and Sinha A: Gene expression profile analysis by DNA microarrays: Promise and pitfalls. JAMA 286: 2280-2288, 2001.

Guttula SV, Allam A and Gumpeny RS: Analyzing microarray data of alzheimer's using cluster analysis to identify the biomarker genes. Int J Alzheimers Dis 2012: 694-4056, 2012.

Wu MY, Dai DQ, Shi Y, Yan H and Zhang XF: Biomarker identification and cancer classification based on microarray data using Laplace naive bayes model with mean shrinkage. IEEE/ACM Trans Comput Biol Bioinform 9: 1649-1662, 2012.

Arias CR, Yeh HY and Soo VW: Biomarker identification for prostate cancer and lymph node metastasis from microarray data using gene prioritization method. Scientific World Journal 2012: 842727, 2012.

Yang CY: Differential expression analysis of gene expression data: mRNA-tag filtering, comparison of t-type tests and their gene-wide co-expression based adjustments. Int J Bioinform Res Appl 6: 353-365, 2010.

Weng Y, Chen F and Duh FM: Preferential activation of hypoxia-inducible factor-1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.

Zsiger LA and Herould JL: The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.

Zsiger LA and Herould JL: The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.

Zsiger LA and Herould JL: The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.

Zsiger LA and Herould JL: The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.

Zsiger LA and Herould JL: The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.

Zsiger LA and Herould JL: The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.

Zsiger LA and Herould JL: The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.

Zsiger LA and Herould JL: The forkhead box transcription factor FOXC1 promotes breast cancer invasion by inducing matrix metalloprotease 7 (MMP7) expression. J Biol Chem 287: 24631-24640, 2012.

Chung TK, Lau TS, Cheung TH, Yim SF, Lo KW, Siu NS, Chan LC, Yu MY, Kwong J, Doran G, et al: Dysregulation of microRNA-204 mediates migration and invasion of endometrial cancer by regulating FOXC1. Tumour Biol 34: 941-946, 2013.