GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Abnormal gene expression profile reveals the common key signatures associated with clear cell renal cell carcinoma: a meta-analysis

H.J. Zhang, Z.Q. Sun, W.Q. Qian and L. Sheng

Department of Urology, Huadong Hospital, Fudan University, Shanghai, China

Corresponding author: L. Sheng
E-mail: shengludoctor@126.com

Genet. Mol. Res. 14 (1): 2216-2224 (2015)
Received January 21, 2014
Accepted October 21, 2014
Published March 27, 2015
DOI http://dx.doi.org/10.4238/2015.March.27.7

ABSTRACT. The aims of this study were to identify the common gene signatures of clear cell renal cell carcinoma (CCRCC), and to expand the respective protein-protein interaction networks associated with CCRCC regulation. For the latter, we utilized multiple gene expression data sets from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), with which we could analyze the aberrant gene expression patterns at the transcriptome level that distinguish cancer from normal samples. We obtained the GSE781 and GSE6344 clear cell renal cell carcinoma gene expression datasets from GEO, which contained a total of 37 cancer and 37 normal samples. Subsequent R language analysis allowed identification of the differentially expressed genes. The genes that exhibited significant up or downregulation in cancers were entered into the Database for Annotation, Visualization, and Integrated Discovery to perform analysis of gene functional annotations, resulting in the generation of two protein-protein interaction networks that included the most significantly up or downregulated genes in CCRCC. These allowed us to identify the key factor genes, which could potentially be utilized

to separate cancer versus normal samples. The differentially regulated genes are also highly likely to be functionally important regulatory factors in renal cell carcinoma: cell functions showing enrichment of these genes include amine biosynthetic and vitamin metabolic processes, ion binding, extracellular transport function, and regulation of biosynthesis. Together, the results from our study offer further reason to pursue diagnosis and therapy of CCRCC at the molecular level.

Key words: Gene expression; Microarray; Protein-protein interaction; Network; Renal cell carcinoma

INTRODUCTION

Renal cell carcinoma (RCC) is a kidney cancer that originates in the lining of proximal convoluted tubules, the very small tubes in the kidney that transport glomerular filtrate (GF) from the glomerulus to the descending limb of the nephron. RCC is the most common type of kidney cancer in adults, responsible for approximately 80% of cases (Mulders et al., 2008). RCC is also one of the most lethal urological cancers, with 40% of patients eventually dying because of metastatic progression of the disease (Ramana, 2012). RCC is classified based on morphological and genetic differences. This classification distinguishes metanephric adenoma oncocytoma and papillary adenoma as benign tumors from clear cell RCC (CCRCC), papillary/chromophilic RCC, chromophobic RCC (CHRCC), and collecting duct RCC. This classification is important because of its prognostic implications.

CCRCC represents the most common subtype (83%) of RCC; the cancer's most striking phenotypic feature is its clear cell morphology (Rezende et al., 1999; Schlenker et al., 2010). Gene expression data have been widely applied to analyze CCRCC. In this study, we applied current gene expression and computational resources to shed light on the key factors of CCRCC (Yao et al., 2005).

MATERIAL AND METHODS

Microarray data

In order to perform a robust analysis of CCRCC, we chose two microarray gene expression datasets as our source material: GSE781 from Lenburg et al. (2003) and GSE6344 from Gumz et al. (2007). Gene Expression Omnibus (GEO) is a National Center for Biotechnology Information (NCBI) database for gene expression (Barrett et al., 2013). GSE781 is composed of 17 normal and 17 cancer samples; GSE6344 contains 20 samples each from cancer patients and normal controls. These two datasets were both performed on two microarray platforms: GPL96 and GPL97.

Filters applied to differentially expressed genes (DEGs)

First, based on information provided in the original dataset reports, we separated the datasets into two groups: cancer and normal. The Student *t*-test was applied in the environment

of R language (Dessau and Pipper, 2008; http://www.R-project.org/), which is an open source statistical program. Finally, genes whose P values were less than 0.01 and with fold changes between groups of more than 1.5 were regarded as DEGs. Because we analyzed two datasets, we confirmed an intersection between the two experiments, which made the allocation as a DEG more reliable.

Heatmap plots of DEGs with clustering

Clustering is a process used to divide data into different groups. The objects in one group are similar, while objects in separate groups are distinctly different. To validate that the DEGs we identified would cluster the samples into two groups, we plotted heatmaps in R language to visualize our cluster analysis according to DEG expression.

Gene ontology (GO) analysis

Whole-genome sequencing has made it clear that a large fraction of the genes specifying core biological functions are shared by all eukaryotes (Ashburner et al., 2000). GO is a widely used tool for the bioinformatic research. The use of GO allows a better understanding of the function of the DEGs, with which we could explore the biological processes, molecular function, and cellular compartment in which the DEGs are involved.

Using STRING to build protein-protein interaction (PPI) networks

The final products of genes are proteins, through which we could build a PPI network. This would allow us to identify the important "key" regulatory proteins in the network, to determine how they and their genes control the disease. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; http://string-db.org/) is a database that integrates existing proteins and their information, or predicted proteins, from different resources (Jensen et al., 2009; Franceschini et al., 2013). We built two PPI networks that were made up of up- or downregulated genes in RCC.

Analysis of key factors with GeneCards

GeneCards (http://www.genecards.org/) is a new type of topic-specific overview resource that provides efficient access to distributed information for the purpose of developing models (Rebhan et al., 1998). Through the networks we created, we were able to identify the vital gene functions in RCC. We used the information provided by GeneCards to further examine every crucial gene in depth, and to develop potential models for the mechanisms underlying RCC.

RESULTS

DEGs

After pre-processing of the microarray data, we filtered out the DEGs that occurred in all datasets, as shown in the Venn diagram (Figure 1). There were 24 remaining upregulated

and 36 downregulated DEGs that were significantly different between groups. With these 60 genes, we were able to confidently proceed with further analyses.

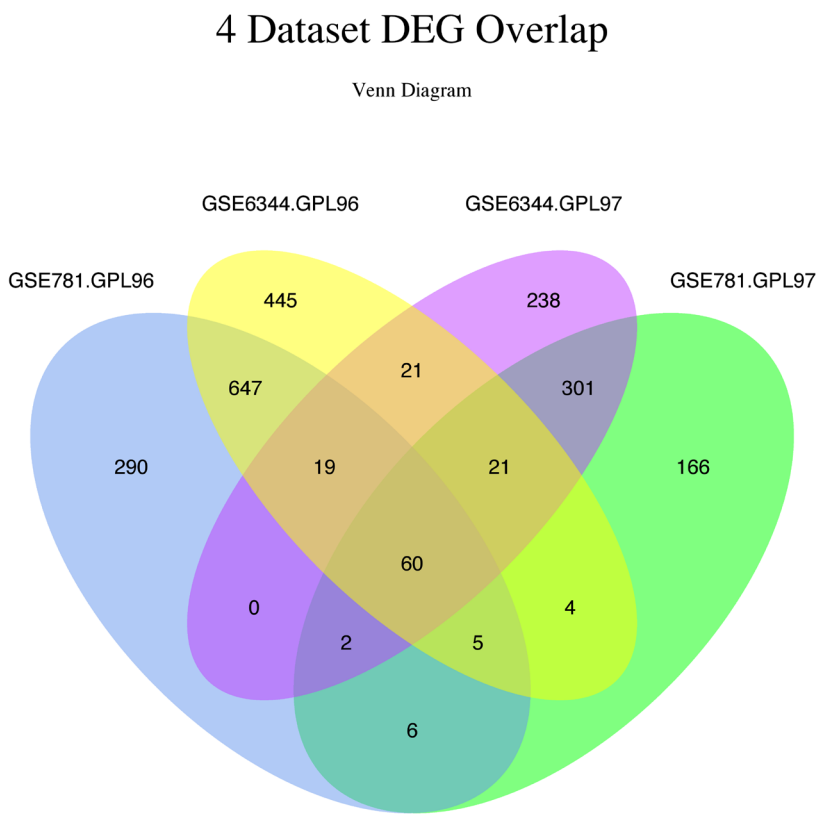

Figure 1. Venn diagram of the four datasets used in this study. GSE781 and GSE6344 datasets were used from clear cell renal cell carcinoma, which each has two platforms: GPL96 and GPL97. The overlap of the four datasets consists of 60 genes.

DEG clustering heatmap

DEGs chosen using the Venn diagram were used to generate a heatmap through cluster analysis via the R language (Dessau and Pipper, 2008) (Figure 2). All samples separated clearly into distinct cancer (left) and normal (right) clusters, which suggested that the DEGs that emerged from our filtering process are the genes crucial for distinguishing between normal tissue and RCC.

GO tables

We used the up- and downregulated genes to perform gene ontology analysis and obtained functional clusters (Tables 1 and 2). The tables indicated that the functions of the up-regulated genes were mainly focused on amino biosynthetic processes, water-soluble vitamin

metabolic processes, and transportation of certain ions through the membrane. At the same time, the downregulated genes' functions were concentrated on extracellular activities, regulation of cellular biosynthetic processes, and ion binding.

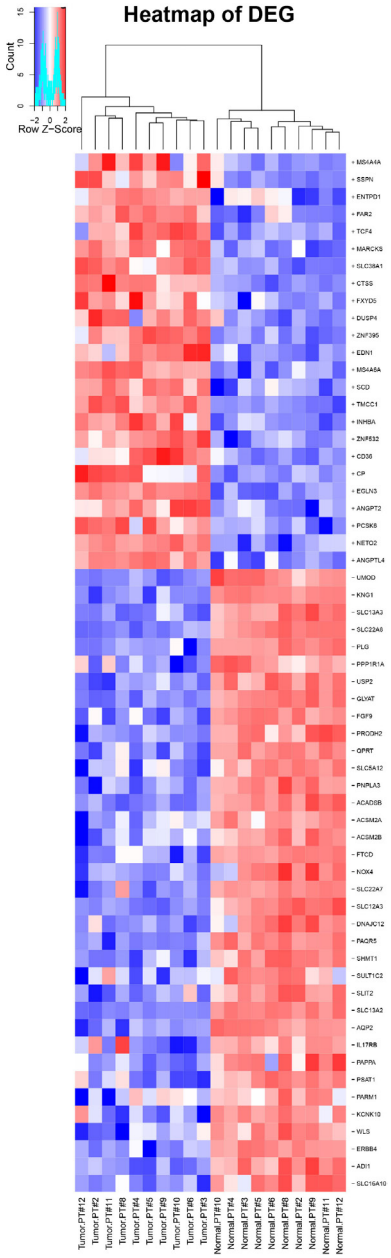

Figure 2. Heatmap of DEGs with clustering. The red/blue squares represent high/low expression, respectively. Up- (+) or downregulation (-) for each gene is indicated.

Table 1. GO analysis of upregulated genes.

	P value
Cluster 1 GO Term	
GO:0008652~cellular amino acid biosynthetic process	2.54E-04
GO:0009309~amine biosynthetic process	9.91E-04
GO:0016053~organic acid biosynthetic process	0.006
GO:0046394~carboxylic acid biosynthetic process	0.006
GO:0044271~nitrogen compound biosynthetic process	0.007
Cluster 2 GO Term	
GO:0006767~water-soluble vitamin metabolic process	0.004
GO:0044271~nitrogen compound biosynthetic process	0.008
GO:0006766~vitamin metabolic process	0.014
Cluster 3 GO Term	
GO:0008509~anion transmembrane transporter activity	4.43E-04
GO:0055085~transmembrane transport	0.002
GO:0031420~alkali metal ion binding	0.002
GO:0015296~anion:cation symporter activity	0.003
GO:0031402~sodium ion binding	0.003
GO:0006814~sodium ion transport	0.004
GO:0015293~symporter activity	0.004
GO:0006820~anion transport	0.005
GO:0015711~organic anion transport	0.005
GO:0015672~monovalent inorganic cation transport	0.007
GO:0006811~ion transport	0.010
GO:0005624~membrane fraction	0.014
GO:0044459~plasma membrane part	0.015
GO:0005626~insoluble fraction	0.016
GO:0015294~solute:cation symporter activity	0.023
GO:0030001~metal ion transport	0.026
GO:0006812~cation transport	0.044
GO:0000267~cell fraction	0.049
GO:0005886~plasma membrane	0.062
GO:0031224~intrinsic to membrane	0.161
GO:0005887~integral to plasma membrane	0.346
GO:0031226~intrinsic to plasma membrane	0.362

Table 2. GO analysis of downregulated genes.

	P value
Cluster 1 GO Term	
GO:0044421~extracellular region part	0.004
GO:0005615~extracellular space	0.004
GO:0001525~angiogenesis	0.016
GO:0048514~blood vessel morphogenesis	0.031
GO:0005576~extracellular region	0.036
GO:0001568~blood vessel development	0.041
GO:0001944~vasculature development	0.043
GO:0042592~homeostatic process	0.264
GO:0042981~regulation of apoptosis	0.290
GO:0043067~regulation of programmed cell death	0.295
GO:0010941~regulation of cell death	0.296
Cluster 2 GO Term	
GO:0010558~negative regulation of macromolecule biosynthetic process	0.163
GO:0031327~negative regulation of cellular biosynthetic process	0.170
GO:0009890~negative regulation of biosynthetic process	0.176
GO:0051173~positive regulation of nitrogen compound metabolic process	0.210
GO:0031328~positive regulation of cellular biosynthetic process	0.231
GO:0009891~positive regulation of biosynthetic process	0.236
GO:0010605~negative regulation of macromolecule metabolic process	0.255
Cluster 3 GO Term	
GO:0046872~metal ion binding	0.280
GO:0043169~cation binding	0.290
GO:0043167~ion binding	0.304
GO:0046914~transition metal ion binding	0.511

STRING PPI networks

We imported the select up or downregulated DEGs into STRING separately. The STRING network images revealed a general functional relationship between the genes' protein products (Figure 3). From the two regulation networks, we could identify certain key genes (proteins) that are connected with other genes, as a hub. We could easily determine that *EDN1*, *HIF1A*, *ANGPT2*, and *CD36* are genes that have close relationships with other genes in the PPI network of upregulated DEGs (Figure 3A). In the downregulated PPI network, we observed that *KNG1*, *PLG*, *EGF*, *SHMT1*, *PSAT1*, *SERPINE1*, and *SERPINE2* have more than four connecting lines with other genes. We deduced that these genes are key genes that play important roles in CCRCC.

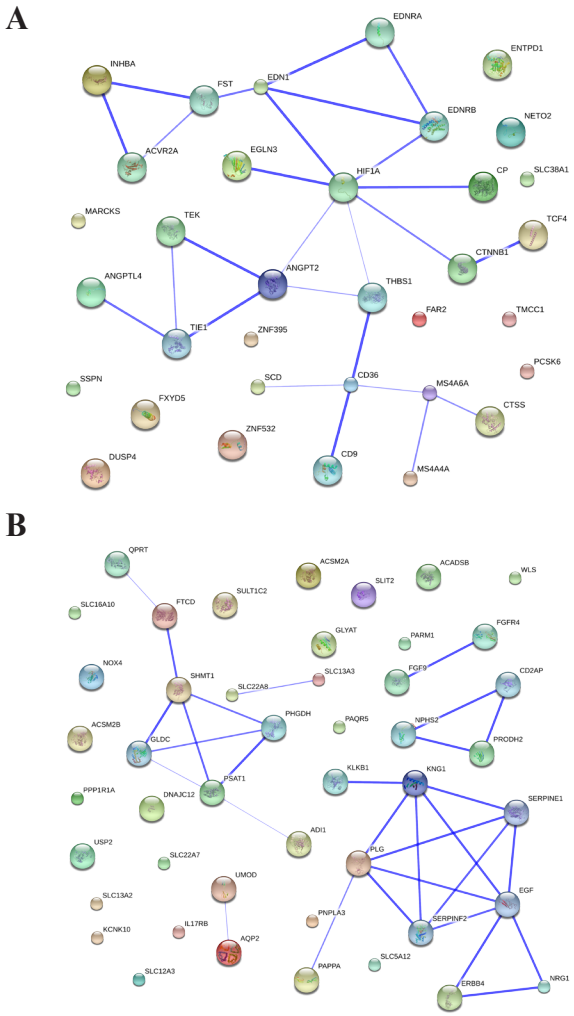

Figure 3. Relationship between each gene's protein in the PPI network. **A.** Upregulated gene PPI network. **B.** Downregulated gene PPI network.

DISCUSSION

CCRCC is an important cancer warranting many additional years of study. Microarray technology has been applied to cancer research and contributes substantially toward revealing the mechanisms of carcinoma. Combining gene expression profiling and modern bioinformatic analysis, we have developed a deep understanding of CCRCC.

We took two datasets and four kinds of experiments into consideration, which gave us a list of 60 DEGs in RCC. We then looked into the functions of these genes. The DEGs clustered cleanly into normal and cancer groups, setting a good standard for microarray data clustering analysis. The DEGs could therefore be regarded as biological marks potentially useful for distinguishing RCC samples.

Following the identification 60 DEGs that showed the greatest expression differential, these bio-markers were analyzed with the GO tool. The result indicated that these dysregulated genes were focused on the functions of amine biosynthetic processes, vitamin metabolic processes, ion binding, extracellular activities, and the regulation of biosynthetic processes. The elevated expression of the upregulated genes suggests that the associated functions are highly activated. Previous research has shown that serum amino acid profiles are a potentially useful biomarker in patients with RCC (Mustafa, et al., 2011), supporting the finding in our study that the amino acid biosynthesis pathway was markedly altered in RCC, and represents a key function for RCC analysis. Similarly, existing evidence has shown that intake of vitamins in a rich diet could reduce the risk of RCC (Hu et al., 2009). Therefore, in the staged progression of RCC, we might assume that the elevated gene expression implies that the organ was attempting to enhance the function of vitamin metabolic processes in order to control the tumor. We also understand that the kidney has a close relationship with ion transport. CCRCC, which represents 80% of diagnosed RCC, has dysfunctional ion transport. In turn, the downregulated genes might imply a low level of performance. The downregulation of extracellular activities and the attenuation of apoptosis both accord with the features of carcinoma.

In this study, we not only examined the clustering functions of the DEGs, but also created two PPI regulation networks of both the up- and downregulated genes. The hub genes that connected to other abnormal genes were clearly identified. In the network of upregulated DEGs, *HIF1A*, *EDN1*, *ANGPT2*, and *CD36* were predicted to be the most important hubs. *HIF1A* encodes the alpha subunit of transcription factor hypoxia-inducible factor-1 (HIF-1), which is a heterodimer composed of an alpha and a beta subunit. HIF-1 plays an essential role in tumor angiogenesis and the pathophysiology of ischemic disease (Ratcliffe, 2002); in RCC, it likely benefits tumor growth. It is also a core gene that affects other genes. In the downregulated PPI network, we found that the genes clustered into disparate groups, unlike the pattern in the PPI network of upregulated genes. *KNG1*, *PLG*, *EGF*, *SHMT1*, *PSAT1*, *SERPINE1,* and *SERPINE2* are the genes that appear to have the most influence on other genes. Of these, dysregulation of EGR has been associated with the growth and progression of certain cancers (Herbst, 2004). Its potential importance can be validated in our network through the observation that it has connections with six other genes.

CONCLUSION

Through integration of multiple microarray datasets and bioinformatic methods, we identified a list of 60 significant DEGs between cancer and normal samples from CCRCC pa-

tients. We were able to use these significant DEGs to distinctly separate samples into normal and cancer groups. GO analysis provided a clear explanation of the DEGs' potential functions, namely amine biosynthetic and vitamin metabolic processes, ion binding, etc. Finally, PPI networks revealed the key factors of CCRCC. We concluded that HIF1A, EDN1, ANGPT2, CD36, KNG1, PLG, EGF, SHMT1, PSAT1, SERPINE1, and SERPINE2 play major roles in CCRCC. More experiments need to be performed to validate the results from this deep analysis. The result may provide an important foundation for diagnosis and therapy for RCC.

REFERENCES

Ashburner M, Ball CA, Blake JA, Botstein D, et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet*. 25: 25-29.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, et al. (2013). NCBI GEO: archive for functional genomics data sets - update. *Nucleic Acids Res*. 41: D991-D995.

Dessau RB and Pipper CB (2008). "R" - project for statistical computing. *Ugeskrift for Laeger* 170: 328-330.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 41: D808-D815.

Gumz ML, Zou H, Kreinest PA, Childs AC, et al. (2007). Secreted frizzled-related protein 1 loss contributes to tumor phenotype of clear cell renal cell carcinoma. *Clin. Cancer Res.* 13: 4740-4749.

Herbst RS (2004). Review of epidermal growth factor receptor biology. *Int. J. Radiat. Oncol. Biol. Phys.* 59: 21-26.

Hu J, La Vecchia C, Negri E, DesMeules M, et al. (2009). Dietary vitamin C, E, and carotenoid intake and risk of renal cell carcinoma. *Cancer Causes Control* 20: 1451-1458.

Jensen LJ, Kuhn M, Stark M, Chaffron S, et al. (2009). STRING 8 - a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*. 37: D412-D416.

Lenburg ME, Liou LS, Gerry NP, Frampton GM, et al. (2003). Previously unidentified changes in renal cell carcinoma gene expression identified by parametric analysis of microarray data. *BMC Cancer* 3: 31.

Mulders PF, Brouwers AH, Hulsbergen-van der Kaa CA, van Lin EN, et al. (2008). Guideline 'Renal cell carcinoma'. *Ned. Tijdschr. Geneesk*. 152: 376-380.

Mustafa A, Gupta S, Hudes GR, Egleston BL, et al. (2011). Serum amino acid levels as a biomarker for renal cell carcinoma. *J. Urol.* 186: 1206-1212.

Ramana J (2012). RCDB: Renal Cancer Gene Database. *BMC Res. Notes* 5: 246.

Ratcliffe PJ (2002). From erythropoietin to oxygen: hypoxia-inducible factor hydroxylases and the hypoxia signal pathway. *Blood Purif.* 20: 445-450.

Rebhan M, Chalifa-Caspi V, Prilusky J and Lancet D (1998). GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 14: 656-664.

Rezende RB, Drachenberg CB, Kumar D, Blanchaert R, et al. (1999). Differential diagnosis between monomorphic clear cell adenocarcinoma of salivary glands and renal (clear) cell carcinoma. *Am. J. Surg. Pathol.* 23: 1532-1538.

Schlenker B, Seitz M, Bader MJ, Ganzer R, et al. (2010). Comparison of guideline recommendations with daily practice in patients with renal cell carcinoma. *Eur. J. Med. Res.* 15: 253-257.

Yao M, Tabuchi H, Nagashima Y, Baba M, et al. (2005). Gene expression analysis of renal carcinoma: adipose differentiation-related protein as a potential diagnostic and prognostic biomarker for clear-cell renal carcinoma. *J. Pathol.* 205: 377-387.