Genome Wide Gene Expression Profiling and Molecular Classification of Renal Cell Cancer Subtypes

Gökmen Zararsız, Gözde Ertürk Zararsız, and Ahmet Özt ürk Department of Biostatistics, Erciyes University Faculty of Medicine, Kayseri, Turkey Email: gokmenzararsiz@hotmail.com,{gozdeerturk, aozturk }@erciyes.edu.tr

Din çer G öks ül ük and Sel çık Korkmaz Department of Biostatistics, Hacettepe University Faculty of Medicine, Ankara, Turkey Email: {selcuk.korkmaz, dincer.goksuluk}@hacettepe.edu.tr

Vahap Eldem Department of Biology, Istanbul University Faculty of Science, Istanbul, Turkey Email: vahap.eldem@istanbul.edu.tr

Halit Karaca

Department of Medical Oncology, Erciyes University Faculty of Medicine, Kayseri, Turkey Email: karacah@erciyes.edu.tr

Tunahan Taşçi Vocational School of Health Services, Istanbul Bilgi University, Istanbul, Turkey Email: tunahan.tasci@bilgi.edu.tr

Abstract-Renal Cell Carcinoma (RCC) is one of the most common malignancy and the sixth cause of leading death from the cancer. In this study, our aim is to identify biomarker candidates expressed at different levels between subtypes of RCC for potential new therapy for RCC. A total of 1,020 raw RNA-Seq counts of renal cancer dataset including 606 KIRP, 323 KIRC and 91 KICH were obtained from the Cancer Genome Atlas (TCGA) Project. Five statistical learning methods including support vector (SVM), random forests (RF), penalized machines discriminant analysis (PDA), na we bayes (NB) and knearest neighbors (KNN) were applied on the purpose of classification. Gene ontology (GO) and Kyoto encyclopaedia of genes and genomes (KEGG) pathway analysis were also conducted for a deeper understanding of differentially expressed genes in three renal cancer subtypes. We identified 181 up-regulated and 69 down-regulated genes that showed a significant differential expression between cancer samples. Classification analysis resulted that integrating these genes in a statistical learning framework perform between 90.7-94.6% accuracy, 81.4-91.2% sensitivity and 92.9-97.0% specificity. Our findings showed a few DEGs including NPK, HRNPLT, ARF1 and TTR might contain a critical role in RCC. Furthermore, study demonstrated using the DEGs as the high rate hit distinguish and classification for RCC subtypes.

Index Terms—gene expression profiling, kidney cancer, next-generation sequencing, renal cell cancer.

I. INTRODUCTION

Renal Cell Carcinoma (RCC) is one of the most common malignant cancer types representing 2-3% of all malignancies with an expectation of more than 200,000 new cases annually. In 2014, approximately 64,000 new cases of renal cancers are expected in the USA, which makes the RCC the sixth largest cause of cancer deaths. These new diagnosed cases correspond approximately 5% of all malignancies within males and 3% within females [1-4]. Kidney renal clear cell (KIRC), kidney renal papillary cell (KIRP) and kidney chromophobe (KICH) carcinomas are the most common subtypes of the RCC, accounting for approximately 95% of total malignant kidney tumors in adults. KIRC is the most common (70-80%) and deadly subtype of this disease with 5-year survival rate of 44-69% and metastasis rate of 94%. For KIRP and KICH subtypes, incidence rates are about 14-17% and 4-8%, 5-year survival rates are about 82-92% and 78-87%, metastasis rates are about 4% and 2%, respectively [5, 6].

Moreover, distinguishing renal cell cancer subtypes are crucial due to following reasons: (i) histological subtypes are associated with the prognosis and survival differences, (ii) it is possible to detect individual chances for optimal treatment responses and (iii) it may help understanding the etiology of this disease, since potential biomarkers may have roles in different pathways for different disease

Manuscript received July 1, 2016; revised October 21, 2016.

subtypes [7]. With the recent advances in molecular biology, it is possible to understand the molecular mechanisms that underlie the tumorigenesis and progression of cancer diseases. In this way, reliable candidate biomarkers can be detected and novel potential targets can be identified for targeted cancer therapies. Several studies are reported regarding to gene-expression based differential diagnosis of the RCC [1, 8-10]. Lately, various genomics, transcriptomics and epigenomics related studies have been developed in genome-wide characterization of RCC [11-16]. However, there is still an urgent need to determine new genome wide biomarkers and diagnostic schemes for the RCC.

In this study, we aimed to detect the differentially expressed genes among the RCC subtypes to identify candidate biomarkers and novel potential targets for targeted RCC therapies. We also aimed to classify corresponding subtypes, using the significant genes, which may play a significant role in the diagnosis of the RCC.

II. MATERIALS AND METHODS

A. Experimental Data

We obtained a real RNA-Seq dataset composed of level 3, RNASeqV2 data from the Cancer Genome Atlas (TCGA) Project [17]. The data contains a total of 1,020 renal cancer samples containing 606 KIRP, 323 KIRC and 91 KICH, respectively. The data for each sample is the raw read counts that resulted from 20 billion reads and mapped to human reference genome. Each sample contains the counts of 20,531 known human RNA transcripts as generated by one of the Genome Sequencing Center of TCGA, University of North Carolina at Chapel Hill. We selected the top 5,000 genes with the highest variances to filter non-informative genes before starting analysis.

B. Data Normalization and Transformation

To adjust the sample specific differences, we applied trimmed mean of M values (TMM) normalization to the raw count data. Next, we applied a voom transformation to estimate the mean-variance relationship of counts and to convert the data in continuous form because of several advantages such as: (i) count based mean-variance estimation using discrete distributions does not totally match to its expected value, (ii) voom is more precise method in mean-variance modeling, (iii) voom based differential expression analysis are able to control type I error and has the highest statistical power [18, 19].

C. Differential Expression Analysis

Linear models for microarray and RNA-Seq data (limma) method was applied to the transformed data to identify the most significantly differentially expressed genes (DEGs) among the RCC subtypes. Log-cpm (log counts per million) values and the associated weights for each observation and each gene were taken as input data for the limma method. p values were adjusted for multiple testing problem using Benjamin-Hochberg procedure. Top 250 genes with lowest false discovery

rate values were stored in all pairwise comparisons. A Venn diagram is constructed to identify the commonly differentially expressed genes in all pairwise comparisons [20, 21].

D. Classification Analysis

To identify the predictive performances of multiple significant genes and distinguish different subtypes of the RCC, we performed classification analysis. Five statistical learning methods including support vector machines (SVM), random forests (RF), penalized discriminant analysis (PDA), na we bayes (NB) and knearest neighbors (KNN) were applied [22]. A ten-fold cross-validation technique is used for model validation. Diagnostic accuracies, sensitivity, specificity, positive and negative predictive values are recorded to assess the model performances.

E. Gene Ontology & Pathway Analysis

For a deeper understanding of differentially expressed genes in three renal cancer subtypes, Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis [23] were performed using GeneCodis 3.0 program [24, 25]. Additionally, a hypergeometric test was used to map all differentially expressed genes to terms in the GO database for identifying significantly enriched GO terms in our gene list (*p*-values were calculated by using this hypergeometric test and corrected by FDR method). Subsequently, all identified GO terms were categorized according to three main GO categories; biological process (BP), molecular function (MF) and cellular component (CC).

III. RESULTS

A. Differentially Expressed Genes Among Renal Cell Cancer Subtypes

A venn diagram is constructed for all three pairwise limma differential expression results (KIRC vs. KICH, KIRP vs. KICH, KIRC vs. KIRP) and shown in Fig. 1. We detected the top 250 genes that were overlapped in all pairwise analysis. A total of 181 genes in this gene list were found as significantly up-regulated (p<0.001), while the remaining 69 were significantly down-regulated (p<0.001). Top 20 up-regulated and down-regulated genes with their corresponding fold change values are given in Table I and Table II.

B. Functional Annotation Results

To gain a comprehensive insight into the biological role of DEGs, we applied GO analysis to significantly upregulated and down-regulated genes in three renal carcinoma tissues. A total of 250 DEGs were assigned to 345 unique GO terms, which are distributed under the three main categories; biological process (204 GO terms, 59.13%), molecular function (60 GO terms, 17.39%) and cellular components (81 GO terms, 23.47%). For upregulated genes, "gene expression" (GO:0010467) with thirty genes was found to be most abundant in biological process, followed by "cellular protein metabolic process" (GO:0044267), "RNA splicing (GO:0008380)" and

"mRNA metabolic process" (GO:0016071). Regarding molecular functions, "protein binding" (GO:0005515, 99 genes) and "nucleotide binding" (GO:0000166, 48 genes) were first and second most represented GO terms. Under cellular component category, "cytoplasm" (GO:0005737, 104 genes), "nucleus" (GO:0005634, 77 genes), "cytosol" (GO:0005829, 69 genes)."membrane" (GO:0016020, 45 genes) were the most highly represented categories. As for down-regulated genes, GO terms associated with "transport" (GO:0006810). "transmembrane transport" (GO:0055085), "multicellular organismal development" (GO:0007275) and "ion transport" (GO:0006811) were the most common annotated biological processes. In the molecular function category, GO terms; "protein binding" (GO:0005515), "calcium ion binding" (GO:0005509) and "transporter activity" (GO:0005215) were predominant in down-regulated genes. The genes within cellular component category were found to be mostly associated with GO terms; "extracellular region" (GO:0005576, 26 genes). (GO:0005737, "cytoplasm" 18 genes), "membrane" (GO:0016020, 16 genes) and "plasma membrane" (GO:0005886, 16 genes). To identify metabolic pathways involved in kidney renal carcinomas in which DEGs were involved, metabolic pathway-based analysis was applied using the KEGG database. Our analysis showed that a total of 61 unique pathways were significantly enriched (p < 0.05), including; "protein processing in endoplasmic reticulum" (Kegg:04141, 18 genes), "neurotrophinsignaling pathway" (Kegg:04722, "regulation of actin cytoskeleton" 11 genes). (Kegg:04810, 10 genes) and "tight junction" (Kegg:04530, 10 genes), (Fig. 2).



Figure 1. Venn diagram displaying the overlapping and non-overlapping gene numbers differentially expressed among RCC subtypes

Classification Performances of Differentially Expressed Genes in Renal Cell Cancer Subtypes DEGs were included in the classification analysis to identify the multiple effects of these genes in separating the RCC subtypes. The classification methods used in this study, including SVM, RF, PDA, NB and KNN, are performed on both pairwise and overall RCC subtypes and the classification performances are given in Table III.

When the three cancer subtypes are classified together, we obtained classification accuracies between 90.7-94.6%. Sensitivity measures vary between 81.4-91.2% and specificity measures vary between 92.9-97.0%. In addition to overall classification, each cancer subtype is discriminated from other cancer subtypes. Classification accuracies are obtained between 92.8-95.2% for KIRC versus KICH, 91.4-96.8% for KIRC versus KIRP and 97.1-99.3% for KIRP versus KICH. The highest discrimination is observed between KIRC and KICH subtypes. Among the selected classification methods, SVM performed the best for most of the situations (Table III).

IV. DISCUSSION

systematic comparison and evaluation Α of transcriptome dynamics between different tissues at genome-wide scale enables us to determine the candidate genes whose expressions are highly induced during carcinogenesis. From possible candidate genes, the identification of most DEGs by a robust statistical method can be a good starting point for researchers, interested in early cancer diagnosis. In this study, a procedure for determining most DEGs in three subtypes of renal cell carcinoma (RCC) was applied using the large volume of RNA-Seq data. Gene expression profiles of 1,020 renal cell cancer samples were downloaded from the TCGA database and analyzed.

As explained above (Table I), two heterogeneous nuclear ribonucleoproteins (HNRNPK, HNRNPL) involved in multiple processes of gene expression, transcription and mRNA splicing were found to be most significantly up-regulated expressed genes and associated with von Hippel-Lindau (VHL) disease [26, 27]. As an autosomal-dominant inherited tumor syndrome, von Hippel-Lindau (VHL) disease is caused by mutations of the von Hippel-Lindautumor suppressor gene (VHL) and the mutations are linked to renal cell carcinoma [28]. It was previously well reported that the mRNA expression of splicing factors such as; HNRNPK were increased in tumor samples and modulate the expression of many genes that control cellular proliferation and tumorigenesis. In addition, a research showed that one of RNA-binding protein (HNRNP A) is the potential direct targets of VHL [26]. Therefore, we concluded that the overexpression of above-mentioned ribonucleoproteins in cellular splicing machinery disrupt somehow suppressor activity of VHL causing renal tumor progression.

Of note, ADP-ribosylation factor 1 (ARF1) was identified among the top DEGs. ARF1 belongs to the Ras superfamily of small GTPases and plays a crucial role in mediating vesicular transports and enhancing cell proliferation [29]. There was a significant correlation between cell growth inhibition and Golgi disruption. It was demonstrated that AMF-26, a novel anticancer drug, inhibits the Golgi system by targeting ARF1 activation, inducing complete regression of human breast cancer [30]. An evidence from a prostate cancer study showing that the overexpression of ARF1 activate the expression of ERK1/2, whereas its silencing with shRNA leads to inhibition of ERK1/2. Therefore, in the progression of prostate cancer, ARF1 may play an important role by modulating the expression of MAPK ERK1/2 [31]. Another study showed that ARF1 was found to be significantly upregulated in gastric cancer tissue and its high level expression was evaluated as a biomarker for

gastric cancer detection [32]. This accumulating evidence points to an important role of ARF1 in tumorigenesis and

ARF1 thus hold a potential as a biomarker in determining renal carcinomas subtypes because of its high expression.

TABLE I. LOG-FOLD CHANGES OF TOP 20 UP-REGULATED GENES DIFFERENTIALLY EXPRESSED AMONG RENAL CELL CANCER O	GROUPS
---	--------

Gene	KIRC-KICH	KIRP-KICH	KIRC-KIRP	F	adj.p
HNRNPK	9.30	9.27	9.49	368,710	<0.001
HNRNPL	8.03	8.13	8.03	351,879	< 0.001
ARF1	9.43	9.48	9.27	293,340	<0.001
XRCC6	8.57	8.34	8.41	290,324	<0.001
C20orf43	7.43	7.23	7.22	273,707	<0.001
HNRNPC	8.60	8.98	8.99	271,592	<0.001
PSMF1	7.42	7.63	7.57	260,625	<0.001
UBE2V1	7.57	7.51	7.45	257,556	<0.001
PTBP1	8.23	8.33	8.08	251,607	<0.001
ACTB	12.38	12.57	12.29	246,595	<0.001
GNB1	8.59	8.85	9.11	245,281	<0.001
VCP	8.76	8.69	8.53	244,585	<0.001
CIA01	7.07	7.13	7.02	242,216	<0.001
SF3B2	8.27	8.10	7.87	238,306	<0.001
RAB5B	8.37	8.05	8.03	234,695	<0.001
KIAA0174	7.72	7.83	7.59	230,611	<0.001
YWHAE	9.33	9.73	9.51	230,581	<0.001
PPP2R1A	9.10	9.10	8.64	230,176	<0.001
MRFAP1	8.84	8.55	8.57	227,909	< 0.001
SARIA	7.24	7.02	7.20	225,560	< 0.001

KIRC: Kidney renal clear cell carcinoma, KIRP: Kidney renal papillary cell carcinoma, KICH: Kidney chromophobe

TABLE II. LOG-FOLD CHANGES OF TOP 20 DOWN-REGULATED GENES DIFFERENTIALLY EXPRESSED AMONG RENAL CELL CANCER GRC	OUPS.
--	-------

Gene	KIRC-KICH	KIRP-KICH	KIRC-KIRP	F	adj.p
CALML5	-5.88	-5.38	-5.67	5,647	< 0.001
ORM2	-5.54	-5.68	-5.49	3,914	< 0.001
CYP2A6	-4.98	-5.14	-4.95	3,970	< 0.001
NXF2	-5.03	-5.25	-5.19	2,969	< 0.001
CASP14	-5.53	-5.44	-5.51	2,245	< 0.001
S100A7	-5.74	-5.33	-5.44	2,064	< 0.001
KRT20	-5.16	-3.84	-5.16	1,391	< 0.001
ORM1	-3.87	-4.77	-3.67	1,132	< 0.001
C4orf7	-4.82	-4.77	-4.80	1,129	< 0.001
ADH4	-4.41	-4.35	-4.36	947	< 0.001
ITIH1	-3.48	-4.84	-3.58	790	< 0.001
LRRTM1	-4.70	-4.52	-4.53	694	< 0.001
PAEP	-3.95	-4.64	-3.99	677	< 0.001
ATP6V1G3	-3.97	-4.57	-3.86	658	< 0.001
CALCA	-4.13	-4.08	-4.06	459	< 0.001
VSTM2A	-5.34	-0.53	-5.33	446	< 0.001
HHATL	-4.34	-3.64	-4.18	439	< 0.001
CRABP1	-3.79	-4.05	-3.77	436	< 0.001
TTR	-3.34	-4.18	-3.34	402	< 0.001
KLK4	-4.07	-2.18	-4.16	345	< 0.001

KIRC: Kidney renal clear cell carcinoma, KIRP: Kidney renal papillary cell carcinoma, KICH: Kidney chromophobe

Among the down-regulated genes with significant fold-differences, a particular attention was given to transthyretin (TTR) gene, a transport protein that is required for retinol-binding protein and thyroxin and also works as a rapid turnover protein. Recently, a proteomic study indicated that the expression level of TTR in Pca tissue and its serum level correlate with Pca (advanced prostate cancer) malignancy. This report also showed the ACC: Accuracy rate, SEN: Sensitivity, SPE: Specificity, PPV: Positive predictive value, NPV: Negative predictive value, SVM: Support vector machine, RF: Random forests, PDA: Penalized discriminant analysis, NB: Naive bayes, KNN: k-nearest neighbors.

High expression level of TTR in Pca tissue may allow Pca cells to escape from immunological surveillance, whereas its down regulation can activate the immune system, causing cancer cell death. Therefore, researchers pointed out that the expression levels of TTR serve as biomarkers for the prognosis of Pca [33]. Another study, using LDI-TOF-MS peptide mass fingerprinting, found that TTR expression was found to be decreased in sera of lung cancer and benign lung disease compared with normal sera and suggested that TTR are potential biomarkers for lung cancer [34]. Additionally, the TTR was weakly expressed in some human cancers such as; bronchopulmonary carcinoid and lung squamous carcinoma [35]. In contrary, another proteomic study using 2-DE and MALDI-TOF/MS demonstrated that the protein level of transthyretin (TTR) was up-regulated more than 2-fold in pancreatic cancer juice compared with chronic pancreatitis and choledocholithiasis. The western blot experiments in the same study also verified

that TTR was found to be abundant in the pancreatic cancer juice [19]. As correlated with the previous study, we found the expression level of TTR gene was significantly down regulated when comparing the renal carcinoma subtypes to each other. As reported in elsewhere, the kidneys are important organ in the metabolism and catabolism of TTR and its concentration is increased in patients with reduced kidney function [34]. Thus, we deduced that the reduced expression of TTR could be a candidate biomarker signature that distinguishes renal carcinoma subtypes.

Moreover, classification analysis was carried out to identify the predictive performances of these DEGs in classifying the RCC subtypes. Based on the results of five statistical learning algorithms, we obtained very high predictive performances, mostly for SVM method. Using SVM, these three RCC subtypes can be differentiated with 94.6% classification accuracy, 91.2% sensitivity and 97.0% specificity. In addition to overall classification, each cancer subtype is discriminated successively. The highest discrimination was obtained between KIRP and KICH groups. SVM performed 95.2% accuracy, 97.1% sensitivity and 86.7% specificity in classifying KIRC versus KICH; 96.8% accuracy, 97.1% sensitivity and 96.6% specificity in classifying KIRC versus KICP; 99.3% accuracy, 100.0% sensitivity and 91.7% specificity in classifying KIRP versus KICH. These genes combined with SVM method may be useful in molecular diagnosis of the RCC.

TABLE III. CLASSIFICATION PERFORMANCES OF STATISTICAL LEARNING ALGORITHMS IN DISTINGUISHING RENAL CELL CANCER SUBTYPES.

Classifier	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)
KIRC vs. KICH					
SVM	95.2	97.1	86.7	97.1	86.7
RF	94.0	97.1	80.0	95.7	85.7
PDA	95.2	98.5	80.0	95.7	92.3
NB	94.0	97.1	80.0	95.7	85.7
kNN	92.8	94.1	86.7	97.0	76.5
KIRC vs. KIRP					
SVM	96.8	97.1	96.5	94.4	<i>98.3</i>
RF	95.2	94.2	95.7	92.9	96.6
PDA	94.6	89.9	97.4	95.4	94.2
NB	91.4	91.3	91.5	86.3	94.4
kNN	93.6	88.4	96.6	93.9	93.4
KIRP vs. KICH					
SVM	<i>99.3</i>	100.0	91.7	99.2	100.0
RF	<i>99.3</i>	100.0	91.7	99.2	100.0
PDA	98.6	99.2	91.7	99.2	91.7
NB	97.1	96.9	100.0	100.0	75.0
kNN	98.6	99.2	91.7	99.2	91.7
Total					
SVM	94.6	91.2	97.0	92.8	97.0
RF	91.2	82.3	94.2	89.7	95.7
PDA	93.1	88.6	95.7	90.9	96.3
NB	90.7	89.5	94.7	87.4	94.2
KNN	90.7	81.4	92.9	86.3	94.5



Figure 2. Number of hits for each GO term for differentially expressed genes.

In conclusion, we detected 250 genes as differentially expressed, where 181 were up regulated and 69 were down regulated, among RCC subtypes. Our analysis identified several DEGs may have crucial role on RCC including HRNPK, HRNPL, ARF1 and TTR. Further, we demonstrated that DEGs have very high classification accuracies and could be useful in distinguishing RCC subtypes. We believe that this study will provide groundwork to the most widespread renal tumor cancer disease for further experiments. It will also provide important insights in identification of candidate biomarkers and novel potential targets for targeted RCC therapies.

ACKNOWLEDGMENT

This work was supported by the Research Fund of Ercives University [TDK-2015-5468].

REFERENCES

- M. Takahashi, D. R. Rhodes, K. A. Furge, H. Kanayamat, S. Kagawa, B. B. Haab, et al, "Gene expression profiling of clear cell renal cell carcinoma: Gene identification and prognostic classification," *P Natl Acad Sci USA.*, vol. 98, no.1, pp. 9754-9759, 2001.
- [2] S. Bulut, B. K. Aktas, A. E. Erkmen, C. Ozden, C. S. Gokkaya, M. M. Baykam, et al, "Metabolic syndrome prevalence in renal cell cancer patients," *Asian Pac J Cancer P.*, vol. 15, no. 18, pp. 7925-7928, 2014.
- [3] N. Pirincci, I. Gecit, M. Gunes, M. Kaba, S. Tanik, M. B. Yuksel, et al, "Levels of serum trace elements in renal cell carcinoma cases," *Asian Pac J Cancer P.*, vol. 14, no. 1, pp, 499-502, 2013.
- [4] Cancer.org [www.cancer.org]. USA: American Cancer Society, Web-Report, 2014. Available from: www.cancer.org
- [5] R. Goyal, E. Gersbach, X. M. J. Yang, S. M. Rohan, "Differential diagnosis of renal tumors with clear cytoplasm," *Arch Pathol Lab Med*, vol. 137, no. 4, pp. 467-480, 2013.
- [6] J. R. Young, D. Margolis, S. Sauk, A. J. Pantuck, J. Sayre, S. S. Raman, "Clear cell renal cell carcinoma: Discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT," *Radiology*, vol. 267, no. 2, pp. 444-453, 2013.
- [7] S. G. C. Kroeze, A. M. Bijenhof, J. L. H. R. Bosch, J. J. M. Jans, "Diagnostic and prognostic tissuemarkers in clear cell and papillary renal cell carcinoma," *Cancer Biomark*, vol. 7, no. 6, pp. 261-268, 2010.
- [8] F. Kosari, A. S. Parker, D. M. Kube, C. M. Lohse, B. C. Leibovich, M. L. Blute, et al, "Clear cell renal cell carcinoma: Gene expression analyses identify a potential signature for tumor aggressiveness," *Clin Cancer Res*, vol. 11. no. 14, pp.5128-5139, 2005.
- [9] J. X. Chen, M. C. Cheng, Y. Li, C. B. Jiang, "Relationship between CYP1A1 genetic polymorphisms and renal cancer in China," *Asian Pac J Cancer P*, vol. 12, no. 9, pp. 2163-2166, 2011.
- [10] M. B. Wozniak, F. Le Calvez-Kelm, B. Abedi-Ardekani, G. Byrnes, G. Durand, C. Carreira C, et al, "Integrative Genome-wide gene expression profiling of clear cell renal cell carcinoma in Czech republic and in the United States," *Plos One*, vol. 8, 2013.
- [11] J. J. Kin, M. Labots, L. Marchionni, S. Begum, G. A. Meijer, H. M. W. Verheul, et al, "Genome-wide methylation profiling to identify potential epigenetic biomarkers associated with response to sunitinib in metastatic renal cell cancer (mRCC) patients (pts)," *J Clin Oncol*, vol. 31, no. 15, 2013.
- [12] A. Florcken, M. Grau, A. Wolf, A. Weilemann, J. Kopp, B. Dorken, et al, "Gene expression profiling of peripheral blood mononuclear cells during treatment with a gene-modified allogeneic tumor cell vaccine in advanced renal cell cancer: Tumor-induced immunosuppression and a possible role for NF-kappa B," *Int J Cancer*, vol. 136, no. 8, pp. 1814-1826, 2015.

- [13] L. F. Zerbini, M. K. Bhasin, J. F. de Vasconcellos, J. D. Paccez, X. S. Gu, A. L. Kung, et al, "Computational repositioning and preclinical validation of pentamidine for renal cell cancer," *Mol Cancer Ther*, vol. 13, no. 7, pp. 1929-1941, 2014.
- [14] J. Heinzelmann, A. Unrein, U. Wickmann, S. Baumgart, M. Stapf, A. Szendroi A, et al, "MicroRNAs with prognostic potential for metastasis in clear cell renal cell carcinoma: A comparison of primary tumors and distant metastases," *Ann Surg Oncol*, vol. 21, no. 3, pp. 1046-1054, 2014.
- [15] A. M. Czarnecka, W. Kukwa, A. Kornakiewicz, F. Lian, C. Szczylik, "Clinical and molecular prognostic and predictive biomarkers in clear cell renal cell cancer," *Future Oncol*, vol. 10, no. 15, pp. 2493-2508, 2014.
- [16] G. G. Malouf, J. P. Zhang, Y. Yuan, E. Comperat, M. Roupret, P. Cussenot, et al, "Characterization of long non-coding RNA transcriptome in clear-cell renal cell carcinoma by next-generation deep sequencing," *Mol Onco*, vol. 9, no. 1, pp. 32-43, . 2015.
 [17] T. J. Giordano, "The cancer genome atlas research network: A
- [17] T. J. Giordano, "The cancer genome atlas research network: A sight to behold. endocr pathol," vol. 9, no. 1, pp. 32-43, 2014.
- [18] M.D. Robinson, A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biol*, vol. 11, no. 3, 2010.
- [19] C. W. Law, Y. S. Chen, W. Shi, G. K. Smyth, "voom: precision weights unlock linear model analysis tools for RNA-seq read counts," *Genome Biol*, vol, 15, no. 2, 2014.
- [20] Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, I. Golani, "Controlling the false discovery rate in behavior genetics research," *Behavioural brain research*, vol. 125, no. 1, pp. 279-284, 2001.
- [21] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate - a practical and powerful approach to multiple testing," *J Roy Stat Soc B Met*, vol. 57, no. 1, pp. 289-300, 1995.
- [22] T. Hastie, R. Tibshirani, J. Friedman. 2dn. ed, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York, 2009.
- [23] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, "KEGG as a reference resource for gene and protein annotation," *Nucleic Acids Res*, vol. 44, no. D1, pp. D457-D462, 2016.
- [24] S. Y. Wei, Y. F. Wang, H. Xu, Y. Kuang, "Screening of potential biomarkers for chemoresistant ovarian carcinoma with miRNA expression profiling data by bioinformatics approach," *Oncol Lett*, vol. 10, no. 4, pp. 2427-2431, 2015.
- [25] D. Tabas-Madrid, R. Nogales-Cadenas, A. Pascual-Montano, "GeneCodis3: A non-redundant and modular enrichment analysis tool for functional genomics," *Nucleic Acids Res.*, vol. 40, no. W1 pp. W478-W483i 2012.
- [26] P. A. Pioli, W. F. C. Rigby, "The von Hhippel-Lindau protein interacts with heteronuclear ribonucleoprotein A2 and regulates its expression," *J Biol Chem*, vol. 276, no. 43, pp. 40346-40352 2001.
- [27] A. Piekielko-Witkowska, H. Wiszomirska, A. Wojcicka, P. Poplawski, J. Boguslawska, Z. Tanski, et al, "Disturbed expression of splicing factors in renal cancer affects alternative splicing of apoptosis regulators, oncogenes, and tumor suppressors," *Plos One, vol.* 276, no. 43, pp. 40346-40352, 2010.
- [28] W. G. Kaelin, "The von Hippel-Lindau gene, kidney cancer, and oxygen sensing," J Am Soc Nephrol, vol. 14, no. 11, pp. 2703-2711, 2003.
- [29] C. D'Souza-Schorey, P. Chavrier, "ARF proteins: roles in membrane traffic and beyond," *Nat Rev Mol Cell Bio*, vol. 7, no. 5, pp. 347-358, 2006.
- [30] Y. Ohashi, H. Iijima, N. Yamaotsu, K. Yamazaki, S. Sato, M. Okamura, et al, "AMF-26, a novel inhibitor of the golgi system, targeting ADP-ribosylation factor 1 (Arf1) with potential for cancer therapy," *J Biol Chem*, vol. 287, no. 6, pp. 3885-3897, 2012.
- [31] J. Davis, Y. Teng, J. H. Guo, G. Y. Wu, "Upregulation of ADPribosylation factor 1 contributes to the enhanced activation of the MAPK ERK1/2 in prostate cancer cells," *Faseb J*, vol. 28 no. 1, 2014.
- [32] M. M. Tsai, P. Y. Lin, W. L. Cheng, C. Y. Tsai, H. C. Chi, C. Y. Chen, et al, "Overexpression of ADP-ribosylation factor 1 in human gastric carcinoma and its clinicopathological significance," *Cancer Sci*, vol. 103, no. 6, ppi 1136-1144, 2012.
- [33] D. H. Wang, H. Liang, X. P. Mao, W. Liu, M. T. Li, S. P. Qiu, "Changes of Transthyretin and Clusterin after Androgen Ablation

Therapy and Correlation with Prostate Cancer Malignancy," *Transl Oncol*, vol. 5 no. 2, pp. 124-U78, 2012.

- [34] A. Henze, J. Raila, C. Kempf, P. Reinke, A. Sefrin, U. Querfeld, et al, "Vitamin A metabolism is changed in donors after livingkidney transplantation: an observational study," *Lipids Health Dis*, vol. 10, 2011.
- [35] D. W. Huang, B. T. Sherman, R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nat Protoc*, vol. 4, no. 1, pp. 44-57, 2009.



Gökmen Zararsız is an assistant professor in Biostatistics department of Erciyes University. His main working areas include next-generation sequencing data analysis, machine learning, genomics and transcriptomics. With the other co-authors of this study, S.K. and D.G., they started the 'BIOSOFT Project' to develop algorithms and free, user-friendly, easy-to-use, up-to-date and comprehensive tools in the area of biostatistics and bioinformatics. Some of

these tools can be reached at http://www.biosoft.hacettepe.edu.tr/ . G.Z. currently works on developing novel applications for RNA-Seq classification.



Halit Karaca is an associate professor in Medical School of Erciyes University. His main working area is medical oncology. Background works of H.K., is colorectal cancer, colon cancer, renal cell carcinoma and molecular oncology.



Selcuk Korkmaz is a research assistant in Biostatistics department of Hacettepe University. His working area and expertise is mostly related with proteomics, drug discovery, virtual screening and machine learning. S.K. currently works in University of California, San Diego Supercomputing Center with the project on developing new approaches for the prediction of biological assemblies and proteinprotein interfaces.



Gözde Ertürk Zararsız is a lecturer in Erciyes University, Dept. of Biostatistics. Her research interests focus on proteomics, survival analysis, competing risks, multivariate analysis, image processing and fMRI data analysis. GEZ is currently working as a visiting scholar in Dr Von Hauner University Children's Hospital.



Tunahan Taşçi is a lecturer in Vocational School of Health Services of Istanbul Bilgi University and he is Phd student in Istanbul University. His main working areas including male reproductive system, molecular biology, cell culture, phylogeny and bioinformatics.



Vahap Eldem is a research assistant in Biology department of Istanbul University. For more than 5 years, V.E. has been working for computational biology and bioinformatics. His other research interests include marine genomics, small RNAs and phylogenetics. V.E. currently works on de novo transcriptome assembly analysis of European anchovy (Engraulis encrasicolus).



Ahmet Ozturk is an associate professor in Biostatistics department of Erciyes University. A.O. has great expertise and more than a hundred papers in the field of biostatistics. His expertise area includes biostatistics, pediatrics, epidemiology, machine learning and clinical trials.



Dincer Goksuluk is a research assistant in Biostatistics department of Hacettepe University. His research mostly focuses on software engineering, machine learning, transcriptomics, proteomics and metabolomics. D.G. currently works for developing novel applications for various problems in transcriptomics field.