scientific reports

Check for updates

OPEN

Multi-staged gene expression profiling reveals potential genes and the critical pathways in kidney cancer

Hamed Ishaq Khouja[1]✉, Ibraheem Mohammed Ashankyty[1], Leena Hussein Bajrai[2,3], P. K. Praveen Kumar[4], Mohammad Amjad Kamal[5,6,7], Ahmad Firoz[8] & Mohammad Mobashir[9]✉

Cancer is among the highly complex disease and renal cell carcinoma is the sixth-leading cause of cancer death. In order to understand complex diseases such as cancer, diabetes and kidney diseases, high-throughput data are generated at large scale and it has helped in the research and diagnostic advancement. However, to unravel the meaningful information from such large datasets for comprehensive and minute understanding of cell phenotypes and disease pathophysiology remains a trivial challenge and also the molecular events leading to disease onset and progression are not well understood. With this goal, we have collected gene expression datasets from publicly available dataset which are for two different stages (I and II) for renal cell carcinoma and furthermore, the TCGA and cBioPortal database have been utilized for clinical relevance understanding. In this work, we have applied computational approach to unravel the differentially expressed genes, their networks for the enriched pathways. Based on our results, we conclude that among the most dominantly altered pathways for renal cell carcinoma, are PI3K-Akt, Foxo, endocytosis, MAPK, Tight junction, cytokine-cytokine receptor interaction pathways and the major source of alteration for these pathways are MAP3K13, CHAF1A, FDX1, ARHGAP26, ITGBL1, C10orf118, MTO1, LAMP2, STAMBP, DLC1, NSMAF, YY1, TPGS2, SCARB2, PRSS23, SYNJ1, CNPPD1, PPP2R5E. In terms of clinical significance, there are large number of differentially expressed genes which appears to be playing critical roles in survival.

Renal cell carcinoma (RCC) is the most common type of kidney cancer in adults, responsible for approximately 90–95% of cases and it is one of the leading causes of cancer death. Its occurrence shows mainly male predominance over women with a ratio of 1.5:1. RCC, a kidney cancer originates in the lining of the proximal convoluted tubule which is the part of the very small tubes in the kidney and transport primary urine[1,2]. High-throughput data is created at a large scale in order to understand complex diseases like cancer, and it has aided in research and diagnostic advancement[3–6]. However, extracting useful knowledge from such vast datasets for a complete and detailed understanding of cell phenotypes and disease pathophysiology remains a difficult task, and the molecular events that contribute to disease initiation and progression are still poorly understood[7–9]. The advancement of the post-genomics period has resulted in a huge amount of "big data" in biological sciences, which has led to a multitude of interdisciplinary applications in recent decades[5,10]. There are a number of biological databases that

[1]Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. [2]Special Infectious Agents Unit-BSL3, King Fahad Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia. [3]Biochemistry Department, Sciences College, King Abdulaziz University, Jeddah, Saudi Arabia. [4]Department of Biotechnology, Sri Venkateswara College of Engineering, Sriperumbudur 602105, India. [5]West China School of Nursing/Institutes for Systems Genetics, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China. [6]King Fahd Medical Research Center, King Abdulaziz University, P. O. Box 80216, Jeddah 21589, Saudi Arabia. [7]Enzymoics, Novel Global Community Educational Foundation, 7 Peterlee Place, Hebersham, NSW 2770, Australia. [8]Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia. [9]SciLifeLab, Department of Oncology and Pathology, Karolinska Institutet, Box 1031, 171 21 Stockholm, Sweden. ✉email: hkhoja@kau.edu.sa; m.mobashir@cdslifesciences.com

house various types of datasets. TCGA, oncomine, nephroseq, and GEO (gene expression omnibus) are the most widely used databases in biological sciences[11]. These databases mainly GEO store vast amount of datasets related with cancer, diabetes, and other biological problems[8,12–16].

The identification of pathogenetically distinct tumour types poses a significant challenge in the treatment of complex diseases (especially cancer)[17–19]. The improvement in tumor classification always helps in the improvement during therapeutic approaches[20,21]. In target specific therapy, effectiveness can be maximised while toxicity is reduced by using enhanced classification. To access biological datasets from these databases previously, a variety of tools/approaches were used. For molecular classification of cancer Golub TR et al.,[22] have divided cancer classification into two challenges as class discovery and class prediction.

A number of oncogenes and tumour suppressor genes that are changed in RCC, resulting in pathway dysregulation, need to be identified and investigated further[23–25]. Copy number, gene sequencing, expression pattern, and methylation in primary RCC are all possible avenues for achieving this goal. With continued breakthroughs in omics technology, the application of molecular markers for early diagnosis and prognosis deserves further attention[1,2,26–30].

We have selected RCC dataset with samples from two stages (stages I and II) for the purpose of understanding how gene expression patterns vary and how altered gene expression patterns lead to possible changes in the respective inferred functions as tumour stage I to II changes and from affymetrix platforms (U133A to U133B). Different cancer stages help in describing where a cancer could be located, how far it has spread, and whether it is affecting other parts of the body[31–33]. Healthy tissue usually contains many different types of cells grouped together. If the cancer looks similar to healthy tissue and contains different cell groupings, it is called differentiated or low-grade tumor and when the cancerous tissue looks very different from the healthy tissue, it is termed as poorly differentiated or high-grade tumor. The cancer's grade may help the clinician to predict how quickly the cancer will spread. In general, the lower the tumor's grade, the better the prognosis. Different types of cancer have different methods to assign a cancer grade[7,34–37]. In general, it is very hard to detect most of the cancers at early stage so the main focus was on exploring the gene expression pattern alterations and its functional consequences and further to avoid biasedness, we have incorporated TCGA dataset also which have the samples from all the grades.

Here, we have selected a dataset from gene expression omnibus (GEO) where the samples are from human with two tumor stages (I and II). We have organized the samples in the order such as stage I normal versus tumor and stage II normal versus tumor for the affymetrix platforms U133A and U133B and analyzed the tumor samples with respect to their respective controls (normal sample of the same stage) for the gene expression alterations and evolved functions with the increase in tumor percentage. Based on our work, we conclude that irrespective of the tumor stage PI3K-Akt, Foxo, endocytosis, MAPK, Tight junction, cytokine-cytokine receptor interaction pathways and the major source of alteration for these pathways are MAP3K13, CHAF1A, FDX1, ARHGAP26, ITGBL1, C10orf118, MTO1, LAMP2, STAMBP, DLC1, NSMAF, YY1, TPGS2, SCARB2, PRSS23, SYNJ1, CNPPD1, PPP2R5E. In addition, we have also studied the clinical significance and observe that there are large number of differentially expressed genes which appears to be playing critical roles in for survival such as ARHGAP6, TGM4, CD248, SLC13A3, EPO, PARD6A, CLCA2, UBE2S, ERAL1, FGFR1, MRVI1, DYNC1I2, CDCA7.

Results

In the first step, we have selected the data of our interest (raw expression dataset) GSE6344[30,38], organized the samples in the order such as stage I normal versus tumor and stage II normal versus tumor for the affymetrix platforms U133A and U133B and processed it until normalization and log2 values for all the mapped genes as mentioned in the workflow Fig. 1a. This dataset contains 40 samples (5 normal and 5 tumor for two stages I and II from U133A and U133B platforms). For differential gene expression analysis, we have compared the tumor samples with normal samples of the respective stages and the respective platforms that it gives us four DEGs lists.

Gene expression profiling and the associated functions for varying tumor percentages. In this study, the initial focus of our goal was to understand the gene expression pattern between the different stages for normal versus tumor samples. For this purpose, the total number of the DEGs, up, and down regulated genes have been calculated (Fig. 1b) and the number of down-regulated genes are higher than the up-regulated genes and further, we observe that the number of down regulated genes are comparatively high in all the four DEGs list (Fig. 1c). For U133A dataset, we observe very high number of DEGs for same stage and shares 1147 genes between stage I and II with respect to U133B which is 606 genes and stage I and II specific genes are also high in both the platforms U133A and U133B. Similar to the DEGs distribution, the enriched pathways are also distributed in the similar trend as shown in Fig. 1d (p-values < 0.05) and even after applying strict cut-off of p-value as shown in Fig. 1e (p-values < 0.001). Most of the shared genes between different stages and platforms have been shown with their fold changes and these genes are known to be associated with the critical pathways which are very important for multiple type of cancers (Fig. 1f). In addition, we have also mapped the known association between all these genes (from Fig. 1f) for which one list of all these DEGs have been combined to single DEGs list and finally, these genes have been mapped by using the network database in the form of network as s in Fig. 1g. Figure 1g presents the network of DEGs and their connectivity with each other where there are four smaller clusters and these clusters are connected by a core cluster of SYNJ1, MAPT, YY1, NSMAF, and FNBP4 genes and among the highly connected genes SYNJ1, LAMP2, SCARB2, FDX1, HDLBP, CHAF1A, MAPT, and FNBP4.

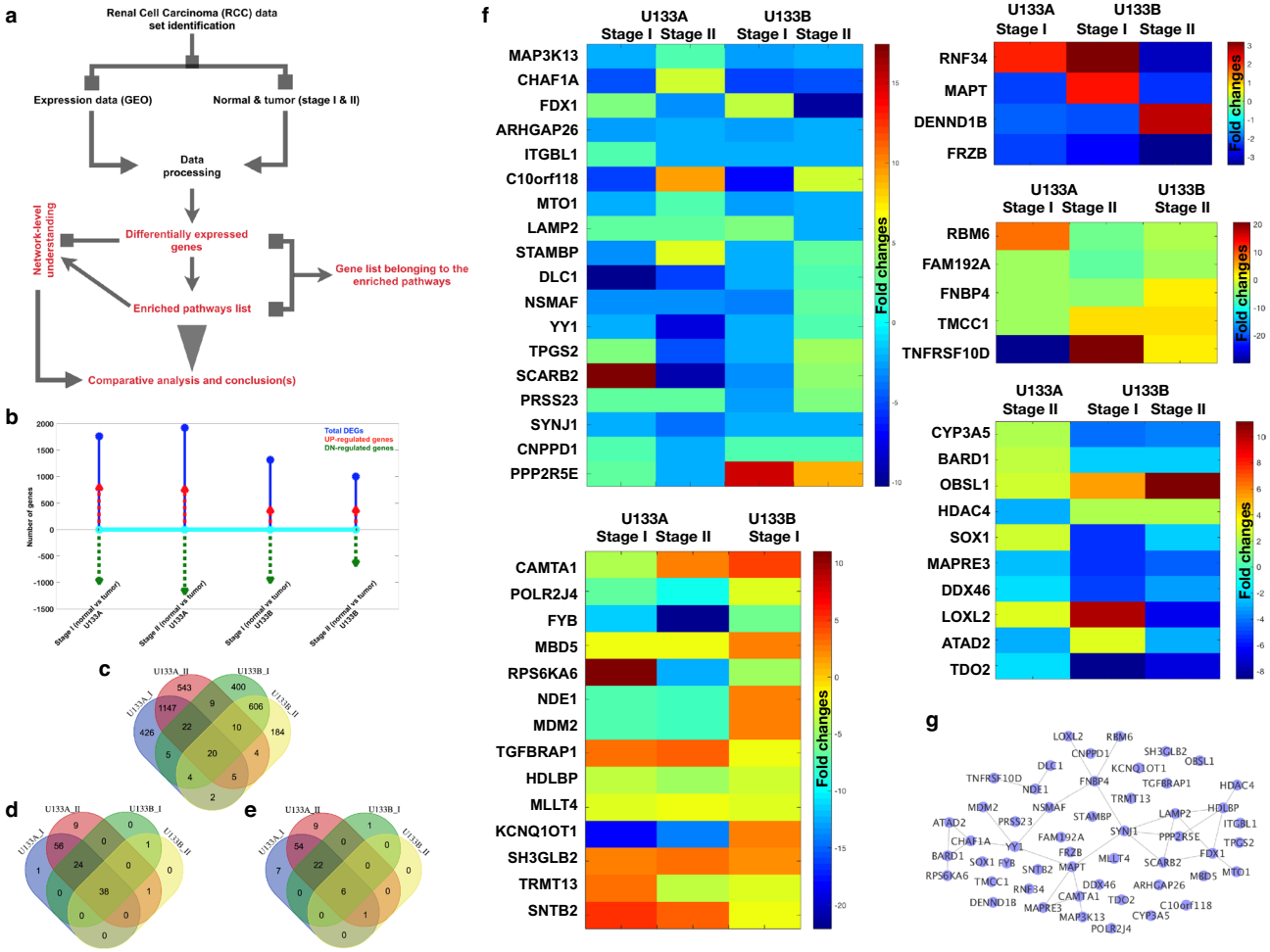

Figure 1. Gene expression profiling. (**a**) Workflow: from raw dataset to analysis. (**b**) Number of DEGs, up- and down-regulated genes. (**c**) Venn diagram to display the DEGs. (**d**) Venn diagram for enriched pathways for the DEGs (p-value ≤ 0.05). (**e**) Venn diagram for enriched pathways for the DEGs (p-value ≤ 0.001). (**f**) Common genes between different stages and the array chips (U133A and U133B) with their fold changes. (**g**) Mapped network for all the genes in (**f**). For venn diagram plotting, freely available webserver (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used and the heatmaps and bar plot were generated by using MATLAB 2017b by using imagesc and plot commands, respectively.

Top-ranked enriched pathways for the respective DEGs list. After analyzing the number of DEGs and the enriched pathways, we have analyzed the enriched pathways and the genes which are altered in different RCC tumor stages (Table 1). We observe that MAPK, cytokine, Akt, Wnt, hippo, Hif1, metabolic signaling pathways are among the top-ranked pathways which are frequently altered and their potential source of alterations are MAP3K13, CHAF1A, FDX1, ARHGAP26, ITGBL1, C10orf118, MTO1, LAMP2, STAMBP, DLC1, NSMAF, YY1, TPGS2, SCARB2, PRSS23, SYNJ1, CNPPD1, PPP2R5E. These genes and the pathways are known to play the potential roles directly or indirectly in case of cancer.

Network-level understanding of the DEGs. Based on the venn diagram of the enriched pathways, we have prepared the list of the pathways in five groups (commonly enriched) and matched the genes with these pathways lists from all the four DEGs list (normal versus tumor in stage I and II for the U133A and U133B datasets). In Fig. 2, the networks have been shown for stage I of U133A, Stage I and II of U133B datasets. The networks shown are for those DEGs which are matching to different pathways lists obtained during venn diagram drawing. The major pathways have been highlighted on the top of the figure and in the left side the tumor stage have been mentioned. Since most of the networks for stage II of U133A dataset were densely connected so for such networks we have presented top 30 genes in terms of connectivity within the network (Fig. 3). Here, we have also shown the connectivity of the genes for those networks where the connections are not clearly visible. For more details of the list of the genes and the pathways used for the network-level analysis were supplied in the Supplementary Table S1.

U133A: stage I and II U133B: stage I and II	Pl3K-Akt signaling, endocytosis, FoxO signaling, MAPK signaling, tight junction, cytokine–cytokine receptor interaction
U133A: stage I and II U133B: stage I	Neurotrophin signaling, insulin signaling, phospholipase-D signaling pathway, adipocytokine signaling, AMPK signaling, thyroid hormone synthesis, Rap1 signaling, oxytocin signaling, phagosome, apelin signaling, thyroid hormone signaling, gap junction, cGMP-PKG signaling, prolactin signaling, longevity regulating pathway, signaling pathways regulating pluripotency of stem cells, progesterone-mediated oocyte maturation, Protein processing in endoplasmic reticulum, purine metabolism, Platelet activation, axon guidance, ubiquuitin mediated proteolysis
U133A: stage I U133B: stage I and II	Ras signaling pathway
U133A: stage I U133B: stage I	Circadian entrainment, sphingolipid signaling, cell adhesion molecules (CAMs), cell cycle, TGF-beta signaling, cAMP signaling, osteoclast differentiation, TNF signaling, adrenergic signaling in card iomyocytes, peroxisome, T cell receptor signaling, hematopoietic cell lineage, natural killer cell mediated cytotoxicity, Fc gamma R-mediated phagocytosis, HIF-1 signaling, oocyte meiosis, NF-kappa B signaling, leukocyte transendothelial migration, Fc epsilon Rl signaling, valine leucine and isoleucine degradation, ECM-receptor interaction, phosphatidylinositol signaling system, estrogen signaling, ErbB signaling, pyrimidine metabolism, long-term potentiation, Regulation of actin cytoskeleton, retrograde endocannabinoid signaling, vascular smooth muscle contraction, inflammatory mediator regulation of TRP channels, RNA transport, apoptosis, Focal adhesion, notch signaling, renin secretion, Jak-STAT signaling, melanogenesis, calcium signaling, VEGF signaling, B cell receptor signaling, oxidative phosphorylation, Spliceosome, long-term depression, drug metabolism—cytochrome P450, glycerophospholipid metabolism, neuroactive ligand-receptor interaction, tryptophan metabolism, p53 signaling pathway, Antigen processing and presentation, Wnt signaling, Hippo signaling, toll-like receptor signaling, GnRH signaling pathway, adherens junction
U133A: Stage I	Olfactory transduction, PPAR signaling, inositol phosphate metabolism, RIG-I-like receptor signaling, lysine degradation, taste transduction, ovarian steroidogenesis
U133B: Stage I	Butanoate metabol ism, synaptic vesicle cycle, tyrosine metabolism, drug metabolism-other enzymes, mRNA surveillance pathway, steroid hormone biosynthesis, proteasome, metabolism of xenobiotics by cytochrome P450, retinol metabolism
U133A: Stage II	Ribosome

Table 1. Enriched pathways grouped either common or specific to the conditions. These pathways have been generated after plotting the venn diagram.

Clinical significance of the differentially expressed genes. Additionally, we have selected the top-ranked genes (based on the fold change 15 up and 15 down) and analyzed the patients survival (Kaplan–Meier plot) for the patient samples from TCGA database and the dataset was TCGA kidney renal clear cell carcinoma (source data from GDAC Firehose) which contains 538 samples[5,36,39]. We observe that most of the top-raked genes (from selected 30 DEGs) mainly up-regulated genes show very high significance on the patients survival (Fig. 4). In this figure, we have also shown the mutations in these top-ranked DEGs for clear renal cell carcinoma in the TCGA databse. There are few genes(ERBB4, SLC13A3, TGM4, and FGFR1) which are mutated at very high rate as shown in Fig. 4a,b. Further, we have also selected different dataset (GSE68417[40]) which contains the samples for adjacent normal, low grade, and high grade and compared the differentially expressed genes and the enriched pathways with each other (Fig. 5a). This shows that the DEGs of adjacent normal versus low grade tumor samples share majority of the DEGs of adjacent normal versus high grade tumor samples and both these list share few DEGs with low grade versus high grade DEGs list and as expected there was no shared enriched pathways at all because there appears only few genes which have gene expression with fold change $\geq +1.5$ (up regulated) or ≤ -1.5 (down regulated) in case of low grade versus high grade. Kaplan–Meier plots show the clinical significance and that is a large number of differentially expressed genes appear to be potentially significant in terms of survival and some of the selected genes are ARHGAP6, TGM4, CD248, SLC13A3, EPO, PARD6A, CLCA2, UBE2S, ERAL1, FGFR1, MRVI1, DYNC1I2, CDCA7 (additional data shown in supplementary Figs. S1–S6). Moreover, Fig. 5b has been presented with the list of genes and the respective p-values for survival analysis and here only those genes have been shown which are clinically significant and the overall pathways associated with these genes and further specific assocations were shown in Fig. 5c. Additionally, the expressions (RNA and protein) have been shown in supplementary data S7. We have checked the expression of these clinically relevant genes by using protein atlas where most of these genes are expressed in case of RCC and act as biomarkers and only TGM4 and GGN were not expressed.

Discussion

Renal cell carcinoma is one of the most common cancers, and it is one of the leading causes of cancer death[14,15,41]. In terms of therapy and diagnosis, therapeutic and clinical outcomes differ between the individuals with even close similarity in clinical and pathological characteristics (tumor type, grades, and stages) and despite tremendous efforts to identify molecular biomarkers (prognostic and predictive) and with improved precision compared to clinical and pathological predictors only few molecular tests have been introduced into oncological practice[29]. So it is important to understand and unravel different levels (such as gene expression pattern, epigenetics, protein expression) of diversities in cancer[42,43]. We gathered the previously published dataset for this purpose and conducted a detailed and precise study ranging from gene expression profiling to functional changes, including networks mapped from the human protein network database.

Our work leads to the conclusion that irrespective of the tumor stage PI3K-Akt, Foxo, endocytosis, MAPK, Tight junction, cytokine-cytokine receptor interaction pathways and the major source of alteration for these pathways are MAP3K13, CHAF1A, FDX1, ARHGAP26, ITGBL1, C10orf118, MTO1, LAMP2, STAMBP, DLC1, NSMAF, YY1, TPGS2, SCARB2, PRSS23, SYNJ1, CNPPD1, PPP2R5E. Networks of DEGs for the enriched

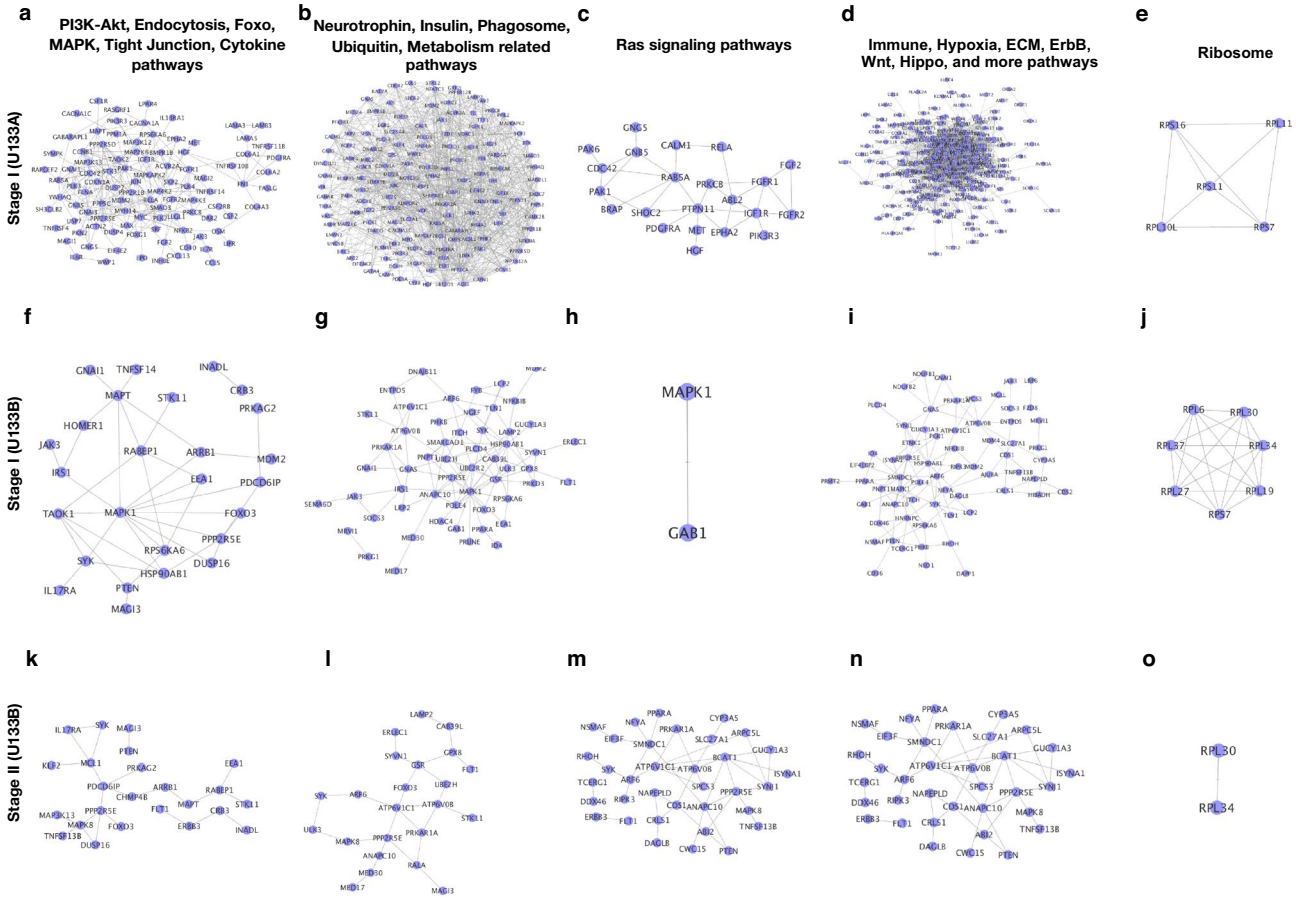

Figure 2. Networks for the genes matched with those pathways which are enriched (p-values ≤ 0.001) and common shown in venn diagram for all the four DEGs list (Stage I and II for U133A and U1333B). In this figure, we have selected those pathways which are commonly enriched pathways and mapped the genes belonging to these pathways from the DEGs list and finally mapped out the networks. (**a**–**e**) It represents the networks for Stage I of U133A platform data for the list of pathways. (**f**–**j**) It represents the networks for Stage I of U133B platform data for the list of pathways. (**k**–**o**) It represents the networks for Stage II of U133B platform data for the list of pathways. All these networks, were drawn by using cytoscape software.

pathways show that there are large number of genes from few specific pathways are altered such as Ras signaling pathways(Fig. 2c,h,m), immune sysytems, Wnt, hippo, (Fig. 2d,i,n) Akt pathways (Fig. 2a,f,k). Here, we observe that critical pathways altered in RCC are wnt, hippo, regulation of actin cytoskeleton, ECM, infection and inflammation, metabolic, and more cancer related pathways. From the mapped network, we observe that the highly connected genes infer the potential pathways or in other works the top ranked genes based on connectivity refer to those pathways which are directly or indirectly associated either with RCC or other types of cancer.

In terms of clinical significance, we looked at the rate of mutations for the top ranked genes (based on fold change) and patients' survival for changes in gene expression, with Kaplan–Meier plots indicating clinical significance. We conclude that a large number of differentially expressed genes tend to be potentially important in terms of survival, with ARHGAP6, TGM4, CD248, SLC13A3, EPO, PARD6A, CLCA2, UBE2S, ERAL1, FGFR1, MRVI1, DYNC1I2, CDCA7 among the genes chosen. Using the publicly available datasets, we have investigated the gene expression profiling for renal cell carcinoma. In the previous work, it has been focused on selected genes and pathways. Here, we have investigated the list of critical pathways and the genes which appear to be clinically highly significant in case of renal cell carcinoma. These clinically significant genes lead to potential alteration in PI3K-Akt, foxo, endocytosis, MAPK, tight junction, cytokine-cytokine receptor interaction pathways. Our work will help in diagnosing the renal cell carcinoma patients because here, we have presented the differentially expressed genes, their inferred pathways, and the clinical impact of the selective genes. Since, our finding is from overall perspective including clinical relevance so this study will help in future for diagnostic also.

This work also appears to be more unique in comparison to the previous study that we potentially explored grade I and II of RCC and further explored the clinical relevance. Healthy tissue usually contains many different types of cells grouped together and if the cancer looks similar to healthy tissue and contains different cell groupings, it is called differentiated or low-grade tumor and when the cancerous tissue looks very different from the healthy tissue, it is termed as poorly differentiated or high-grade tumor. The cancer's grade may help the clinician to predict how quickly the cancer will spread. In general, the lower the tumor's grade, the better the

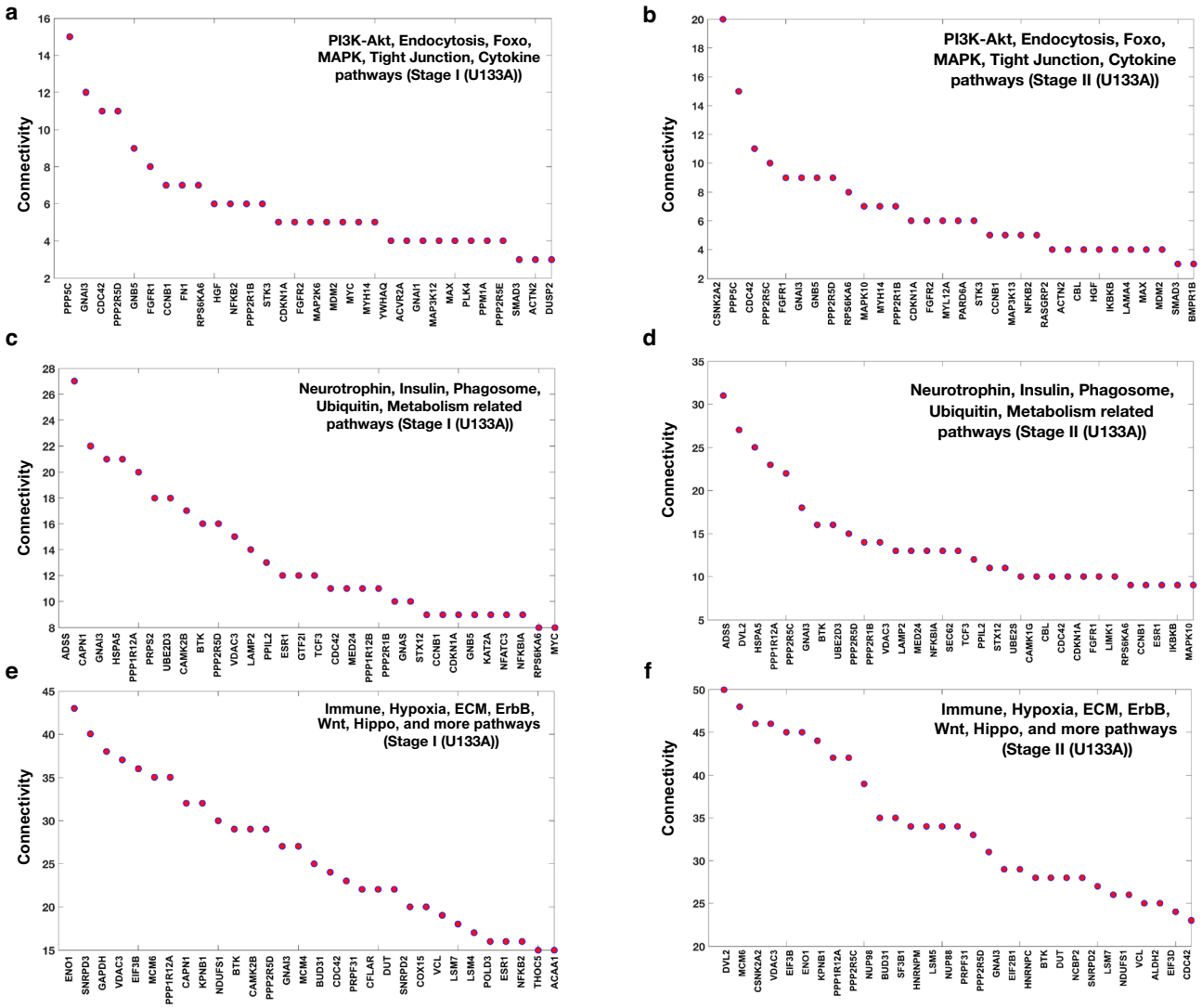

Figure 3. Connectivity in the selected networks (where the gene connectivity is not visible), for the top 30 genes matched with those pathways which are enriched for the DEGs list. (**a**–**f**) Connectivity of the genes in the network for Stage I and II of U133A. These sub-figures were plotted by using MATLAB 2017b by using plot command and afterward dot option was selected and the line was removed.

prognosis. Different types of cancer have different methods to assign a cancer grade[7,34–37] and the different tumor stages could help in describing the severeness, tumor propagation speed, and its impact on the other organs[31–33].. In general, it is very hard to detect most of the cancers at early stage so the main focus was on exploring the gene expression pattern alterations and its functional consequences and further to avoid biasedness, we have incorporated TCGA dataset also which have the samples from all the grades. Further, we have also investigated the expression of these clinically relevant genes by using protein atlas (https://www.proteinatlas.org/)[44–48]. We observe that most of these genes are expressed in case of RCC and act as biomarkers and only TGM4 and GGN were not expressed. This study will be an important step for the understanding of early stage tumor propagation and also will be helpful for clinical aspect.

Conclusions

Based on our findings, we conclude that PI3K-Akt, Foxo, endocytosis, MAPK, Tight junction, and cytokine-cytokine receptor interaction pathways are among the most commonly altered pathways in renal cell carcinoma, and that MAP3K13, CHAF1A, FDX1, ARHGAP26, ITGBL1, C10orf118, MTO1, LAMP2, STAMBP, DLC1, NSMAF, YY1, TPGS2, SCARB2, PRSS23, SYNJ1, CNPPD1, and PPP2R5E are the major sources of alteration for these pathways. Wnt, hippo, actin cytoskeleton control, ECM, infection and inflammation, metabolic, and other cancer-related pathways are among the most important pathways altered in RCC. ARHGAP6, TGM4, CD248, SLC13A3, EPO, PARD6A, CLCA2, UBE2S, ERAL1, FGFR1, MRVI1, DYNC1I2, CDCA7 are some of the genes that were chosen after survival study.

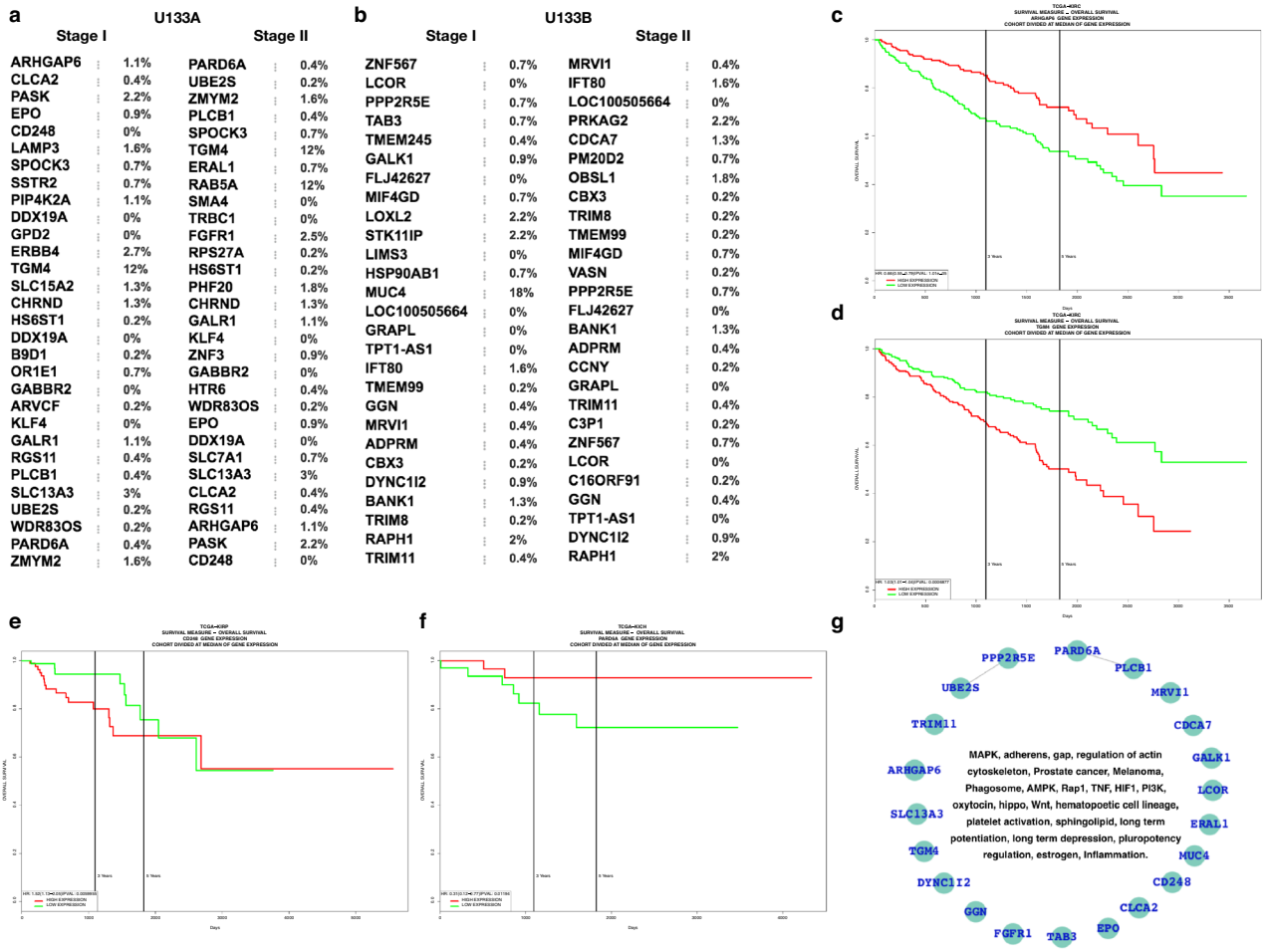

Figure 4. Clinical significance of the top ranked genes. (**a**,**b**) Top 30 (15 up and down) DEGs (based on fold change) with the rate of mutation in kidney renal clear cell carcinoma (TCGA) with their mutations. (**c**–**f**) Survival plots for the selected top ranked genes. (**g**) Network of the clinically significant genes and the associated pathways. Here, (**a**) and (**b**) were drawn by using cBioPortal, (**c**–**f**) by using PROGgeneV2 (http://www.progtools.net/gene/index.php), and (**g**) by cytoscape.

Methods

Here, GSE6344 dataset was used for the study which contains the samples of stage I and II of gene expression for tumor kidney cancer[30,38]. In the first step, we selected the raw expression dataset GSE6344 and processed it until normalisation and log2 values of all mapped genes were achieved, as shown in Fig. 1a of the workflow. These 40 samples in this dataset were 5 normal and 5 tumor for two stages I and II from U133A and U133B platforms. We have compared the tumor samples with standard samples of the respective stages and platforms for differential gene expression analysis, yielding four DEGs lists.

In short the basic steps involved for the entire study are raw file processing, intensity calculation and normalization. For normalization[49–51], GCRMA[52–56], RMA, and EB are the most commonly used approaches. Here, we have used EB for raw intensity normalization. After normalization, we proceed for our goal which is to understand the gene expression patterns[14,57] and its inferred functions[57,58].

To prepare the list of DEGs and analysis, we have our own in-built codes. The samples were placed into two groups such as COVID-19 positive and negative and then normal and the tumor samples. The selection criteria were placed by the fold change and p-values which have been calculated and for the selection of genes as differentially expressed the threshold of fold changes and p-values applied were ± 2 and 0.05, respectively and then KEGG database[59–61] have been used for pathway analysis and for which there is our own code designed[62]. In summary, for differential gene expression prediction and statistical analysis, MATLAB2017 functions (e.g., mattest) were applied and further for pathway analysis, we used KEGG[61] database[62–65].

For generating DEGs network, FunCoup2.0[66] hasbeen used for all the networks throughout the work and cytoscape[67] has been used for network visualization. For most of our coding and calculations MATLAB has been used[62–65]. Furthermore, FunCoup2.0[66] database and cytoscape and its applications[68] were used for network visualization to understand the network and the connectivity of the genes within thenetwork of DEGs[69,70]. The basic concept of FunCoup network database is that it predicts four different classes of functional coupling

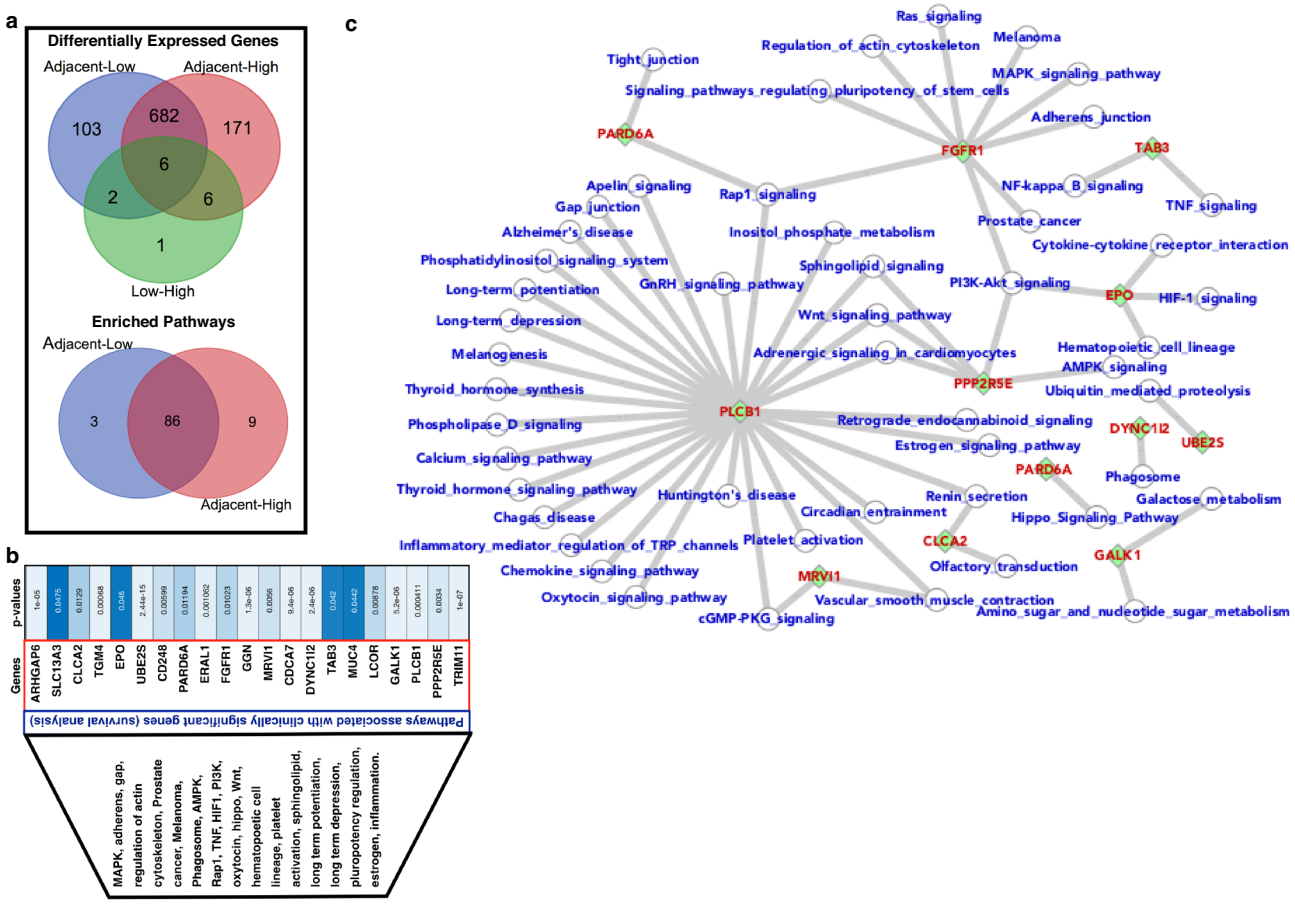

Figure 5. Cross-analysis and clinical relevance. (**a**) Expression profiling of the genes for early (low grade) and advanced stages (high grade) and the tumors with adjacent normal tissues. (**b**) Survival analysis. Genes appear to be clinically significant in terms of survival analysis and the critical pathways associated with them and (**c**) their associations and to draw such association, the KEGG database was used. The green color and diamond shape node represent the gene and the circular node represent the pathway[59–61,71].

or associations such as protein complexes, protein–protein physical interactions, metabolic, and signaling pathways[66]. MATLAB 2017b codes and the command line codes have been used for figure plotting and during analysis. For the network level-analysis such as the number of connectivity per gene and the genes belonging to different number of pathways, the codes have been written in MATLAB and finally it has been plotted also by the codes written in MATLAB[64,65]. For venn diagram plotting, freely available webserver (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used[72–74].

Data availability

We have utilized the publicly available datasets (main data source) which are freely available and have mentioned it in method section with proper references. The analyzed details have been supported by the supplementary data.

Received: 13 May 2021; Accepted: 11 October 2021
Published online: 04 May 2022

References

1. Cairns, P. Renal cell carcinoma. *Cancer Biomark* **9**, 461–473 (2010).
2. Hsieh, J. J. *et al.* Renal cell carcinoma. *Nat. Rev. Dis. Primers.* **3**, 1–19 (2017).
3. Swanton, C. Cancer evolution: The final frontier of precision medicine?. *Ann. Oncol.* **25**, 549–551 (2014).
4. Hiley, C., de Bruin, E. C., McGranahan, N. & Swanton, C. Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. *Genome Biol.* **15**, 453 (2014).
5. Werner, H. M. J., Mills, G. B. & Ram, P. T. Cancer systems biology: a peek into the future of patient care?. *Nat. Rev. Clin. Oncol* **11**, 167–176 (2014).
6. Wang, E. Understanding genomic alterations in cancer genomes using an integrative network approach. *Cancer Lett.* **340**, 261–269 (2013).
7. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
8. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671–679 (2005).
9. Gaulton, K. J. *et al.* Genetic fine mapping and genomic annotation. *Nat. Genet.* **47**, 1415–1425 (2015).

10. Hornberg, J. J., Bruggeman, F. J., Westerhoff, H. V. & Lankelma, J. Cancer: A systems biology disease. *Biosystems* **83**, 81–90 (2006).
11. Yuan, Y. *et al.* AssessingClinicalUtilityCancerGenomicProteomicDatTumorTypes2014NatBiotech. *Nat. Biotechnol.* 1–11 (2014). https://doi.org/10.1038/nbt.2940
12. Li, B. & Li, J. Z. A general framework for analyzing tumor subclonality using SNP array and DNA sequencing data. *Genome Biol.* **15**, 473 (2014). https://doi.org/10.1186/s13059-014-0473-4
13. Rybak, A. P., Bristow, R. G. & Kapoor, A. Prostate cancer stem cells: deciphering the origins and pathways involved in prostate tumorigenesis and aggression. *Oncotarget* **6**, 1900–1919 (2015).
14. Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 811–816 (2004).
15. Roth, R. B. *et al.* Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics* **7**, 67–80 (2006).
16. Ko, J.-H. *et al.* Expression profiling of ion channel genes predicts clinical outcome in breast cancer. *Mol. Cancer* **12**, 106 (2013).
17. Aparicio, S. & Mardis, E. Tumor heterogeneity: next-generation sequencing enhances the view from the pathologist's microscope. *Genome Biol.* **15**, 463 (2014). https://doi.org/10.1186/s13059-014-0463-6
18. Navin, N. E. Tumor evolution in response to chemotherapy: Phenotype versus genotype. *Cell Rep.* **6**, 417–419 (2014).
19. Strandmann, von, E. P., Reinartz, S., Wager, U. & Müller, R.,. Tumor-host cell interactionsin ovarian cancer: Pathwaysto therapy failure. *Trends Cancer* **3**, 137–148 (2017).
20. Yap, T. A., Swanton, C. & de Bono, J. S. Personalization of prostate cancer prevention andtherapy: Are clinically qualified biomarkers in thehorizon?. *EPMA J.* **3**, 3 (2012).
21. Jia, Z. *et al.* Diagnosis of prostate cancer using differentially expressed genes in stroma. *Can. Res.* **71**, 2476–2487 (2011).
22. Golub, T. R. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
23. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–799 (2004).
24. Murai, M. & Oya, M. Renal cell carcinoma: Etiology, incidence and epidemiology. *Curr. Opin. Urol.* **14**, 229–233 (2004).
25. Terris, M., Klaassen, Z. & Kabaria, R. Renal cell carcinoma: Links and risks. *IJNRD* **45**. https://doi.org/10.2147/IJNRD.S75916 (2016).
26. Tiwari, P., Kumar, L., Singh, G., Seth, A. & Thulkar, S. Renal cell cancer: Clinicopathological profile and survival outcomes. *Indian J. Med. Paediatr. Oncol.* **39**, 23 (2018).
27. Navai, N. & Wood, C. G. Environmental and modifiable risk factors in renal cell carcinoma. *Urol. oncol.* **30**, 220–224 (2012).
28. Maruschke, M. *et al.* Expression profiling of metastatic renal cell carcinoma using gene set enrichment analysis. *Int. J. Urol.* **21**, 46–51 (2013).
29. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
30. Tun, H. W. *et al.* pathway signature and cellular differentiation in clear cell renal cell carcinoma. *PLoS ONE* **5**, e10696 (2010).
31. Cheung, K. J. & Ewald, A. J. A collective route to metastasis: Seeding by tumor cell clusters. *Science* **352**, 167–169 (2016).
32. Jackson, T., Koh, G. Y. & Zheng, X. A continuous model of angiogenesis: Initiation, extension, and maturation of new blood vessels modulated by vascular endothelial growth factor, angiopoietins, platelet-derived growth factor-B, and pericytes. *DCDS-B* **18**, 1109–1154 (2013).
33. Reis, P. P. *et al.* A gene signature in histologically normal surgical margins is predictive of oral carcinoma recurrence. *BMC Cancer* **11**, 437–511 (2011).
34. Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* 1–22. https://doi.org/10.1038/nature20788 (2017).
35. Penault-Llorca, F. & Radosevic-Robin, N. Biomarkers of residual disease after neoadjuvant therapy for breast cancer. *Nat. Rev. Clin. Oncol.* 1–17. https://doi.org/10.1038/nrclinonc.2016.1 (2016).
36. Liu, J. *et al.* An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400-416.e11 (2018).
37. Suzuki, H. *et al.* Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* 1–14 (2015). https://doi.org/10.1038/ng.3273
38. Gumz, M. L. *et al.* Secreted frizzled-related protein 1 loss contributes to tumor phenotype of clear cell renal cell carcinoma. *Clin. Cancer Res.* **13**, 4740–4749 (2007).
39. Ellrott, K. *et al.* Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst.* **6**, 271-281.e7 (2018).
40. Thibodeau, B. J. *et al.* Characterization of clear cell renal cell carcinoma by gene expression pro. *Urol. Oncol.* 1–9 (2015). https://doi.org/10.1016/j.urolonc.2015.11.001
41. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24**, 227–235 (2000).
42. Swanton, C. Intratumor heterogeneity: Evolution through space and time. *Can. Res.* **72**, 4875–4882 (2012).
43. Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256–259 (2014).
44. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367** (2020).
45. Uhlén, M. *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4**, 1920–1932 (2005).
46. Uhlén, M. *et al.* A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **366,** (2019).
47. Uhlén, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357,** (2017).
48. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).
49. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet.* **32**, 496–501 (2002).
50. Simon, R. Microarray-based expression profiling and informatics. *Curr. Opin. Biotechnol.* **19**, 26–29 (2008).
51. Ideker, T., Thorsson, V., Siegel, A. F. & Hood, L. E. Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *J. Comput. Biol.* **7**, 805–817 (2000).
52. Reimers, M. Making informed choices about microarray data analysis. *PLoS Comput. Biol.* **6**, e1000786 (2010).
53. Chen, K.-H. *et al.* Gene selection for cancer identification: A decision tree model empowered by particle swarm optimization algorithm. *BMC Bioinf.* **15**, 1–10 (2014).
54. Bild, A. H. *et al.* An integration of complementary strategies for gene-expression analysis to reveal novel therapeutic opportunities for breast cancer. *Breast Cancer Res.* **11**, R55 (2009).
55. Salomonis, N. *et al.* GenMAPP 2: New features and resources for pathway analysis. *BMC Bioinformatics* **8**, 217 (2007).
56. Girke, T. Microarray analysis. 1–42 https://faculty.ucr.edu/~tgirke/HTML_Presentations/Manuals/Microarray/arrayBasics.pdf (2011).
57. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545 (2005).
58. Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. PANTHER version 10: Expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* **44**, D336–D342 (2016).
59. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2007).

60. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360 (2009).
61. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2011).
62. Eldakhakhny, B. M., Sadoun, Al, H., Choudhry, H. & Mobashir, M. In-Silico Study of immune system associated genes in case of type-2 diabetes with insulin action and resistance, and/or obesity. *Front. Endocrinol.* **12,** 1–10 (2021).
63. Warsi, M. K., Kamal, M. A., Baeshen, M. N., Izhari, M. A. & Mobashir, M. Comparative study of gene expression profiling unravels functions associated with pathogenesis of dengue infection. *Curr. Pharmaceut. Des.* **26**(41), 5293–5299 https://doi.org/10.2174/1381612826666201106093148 (2020).
64. Kamal, M. A. *et al.* Gene expression profiling and clinical relevance unravel the role hypoxia and immune signaling genes and pathways in breast cancer: Role of hypoxia and immune signaling genes in breast cancer. *jimsa* **1,** (2020).
65. Krishnamoorthy, P. K. P. *et al.* Informatics in Medicine Unlocked. *Inf. Med. Unlocked* **20**, 100422 (2020).
66. Alexeyenko, A. & Sonnhammer, E. L. L. Global networks of functional coupling in eukaryotes from comprehensive data integration. *Genome Res.* **19**, 1107–1116 (2009).
67. Okawa, S., Angarica, V. E., Lemischka, I., Moore, K. & del Sol, A. A differential network analysis approach for lineage specifier prediction in stem cell subpopulations. *npj Syst Biol Appl* 1–8 (2015). https://doi.org/10.1038/npjsba.2015.12
68. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
69. Mobashir, M., Schraven, B., & Beyer, T. Simulated evolution of signal transduction networks. *PloS one* **7**(12), e50905. https://doi.org/10.1371/journal.pone.0050905 (2012).
70. Mobashir, M., Madhusudhan, T., Isermann, B., Beyer, T. & Schraven, B. Negative interactions and feedback regulations are required for transient cellular response. *Sci. Rep.* **4**, 3718. https://doi.org/10.1038/srep03718 (2014).
71. Kanehisa, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, 277D – 280 (2004).
72. Helmi, N., Alammari, D. & Mobashir, M. Role of potential COVID-19 immune system associated genes and the potential pathways linkage with type-2 diabetes. *Comb. Chem. High Throughput Screen.* https://doi.org/10.2174/1386207324666210804124416 (2021).
73. Bajrai, L. H. *et al.* Understanding the role of potential pathways and its components including hypoxia and immune system in case of oral cancer. *Sci. Rep.* **11**(1), 19576. https://doi.org/10.1038/s41598-021-98031-7 (2021).
74. Bajrai, L. H. *et al.* Gene Expression Profiling of Early Acute Febrile Stage of Dengue Infection and Its Comparative Analysis With Streptococcus pneumoniae Infection. *Front. Cell. Infect. Microbiol.* **11**, 707905. https://doi.org/10.3389/fcimb.2021.707905 (2021).

Acknowledgements

HIK, IMA, LHB, PKPK, MAK, and MM designed the experiment, performed calculations, analyzed the results and written the manuscript. HIK, IMA, LHB, PKPK, MAK, and MM contributed in designing the experiment, analysis, and manuscript writing. HIK, MAK, and MM contributed in experiment designing, analysis, and manuscript writing. The work has been supported by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia funded this project, under grant no. (422-800).

Author contributions

H.I.K., I.M.A., L.H.B., P.K.P.K., M.A.K., A.F., and M.M. designed the experiment, performed calculations, analyzed the results and written the manuscript. H.I.K., I.M.A., L.H.B., P.K.P.K., M.A.K., A.F., and M.M. contributed in designing the experiment, analysis, and manuscript writing. H.I.K., M.A.K., and M.M. contributed in experiment designing, analysis, and manuscript writing.

Funding

The work has been supported by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia funded this project, under grant no. (422-800). The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-11143-6.

Correspondence and requests for materials should be addressed to H.I.K. or M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.