Genes Associated with Disease-Free Survival Prognosis of Renal Cancers

A Computational Screening for Potential Biomarkers and Targets for Gene Therapy

Gideon Tay Yee Chuen, Timothy Lim Tyen Siang, Shannon Lee Xin Ying, Lee Wai Yeow and Caroline G. Lee

Abstract Renal Cell Carcinoma (RCC) incidence has consistently been on the rise in recent years. There are 4 main types of RCC, namely Bladder Urothelial Carcinoma (BLCA), Kidney Chromophobe (KICH), Kidney Renal Clear Cell Carcinoma (KIRC), and Kidney Renal Papillary Cell Carcinoma (KIRP). The aim of this investigation is to identify genes in the tumors across the various renal cancers that can best distinguish patients with good versus those with poor disease-free survival (DFS), and determine pertinent cancer-related pathways that genes associated with DFS reside in. We hypothesized that genes significantly associated with DFS are associated with pathways that can be targeted for gene therapy and be identified as potential biomarkers for RCC. Genes in the tumors of RCC patients significantly associated with DFS were identified from The Cancer Genome Atlas (TCGA) database using Kaplan-Meier analyses. Genes with high expression that are associated with poor survival of patients might serve as potential biomarkers and/or targets for gene therapy across renal associated cancers with the exception of BLCA.

Keywords Renal Cell Carcinoma (RCC) · Bladder Urothelial Carcinoma (BLCA) · Kidney Chromophobe (KICH) · Kidney Renal Clear Cell Carcinoma (KIRC) · Kidney Renal Papillary Cell Carcinoma (KIRP) · Gene therapy · Biomarkers · Disease-free survival · Prognosis

G. T. Y. Chuen · T. L. T. Siang · S. L. X. Ying
Anglo-Chinese School (Independent), Singapore, Singapore

L. W. Yeow · C. G. Lee (✉)
Department of Biochemistry, Graduate School for Integrative Sciences and Engineering, National University Singapore, Singapore, Singapore
e-mail: bchleec@nus.edu.sg

C. G. Lee
Division of Medical Sciences, National Cancer Centre Singapore, Singapore, Singapore

Cancer and Stem Cell Biology Programme, Duke-NUS Graduate Medical School, Singapore, Singapore

H. Guo et al. (eds.), *IRC-SET 2018*,
https://doi.org/10.1007/978-981-32-9828-6_19

1 Introduction

Renal Cell Carcinoma (RCC), a form of kidney cancer, originates from cells in the renal cortex [1]. 59% of Renal Cell Carcinoma (RCC) cases occur in developed countries, with statistics showing rates of diagnosis of RCC being three times lower in developing regions compared to developed regions. This could be due to obesity being a significant cause of renal cancer, as reflected in various studies [2]. With worldwide obesity rates nearly tripling since 1975, and almost 39% of adults aged 18 years and above being overweight, RCC incidence is naturally expected to increase with time [3]. In the last decade, there has been an increased risk of renal cancer in many countries, in particular, Korea, China, Hong Kong, Singapore and Japan [4]. Due to the growing significance and risk of RCC, it was selected as the focus of this study. Here, four subsets of RCC documented in The Cancer Genome Atlas (TCGA), a collaboration between the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), were studied. They are Bladder Urothelial Carcinoma (BLCA), Kidney Chromophobe (KICH), Kidney Renal Clear Cell Carcinoma (KIRC), and Kidney Renal Papillary Cell Carcinoma (KIRP).

In this investigation, we identify genes whose expression are significantly higher in the tumors of RCC patients with good disease-free survival (DFS) compared to those with poorer DFS, are involved in pertinent cancer-related pathways, and have functions that may affect cancer relapse. In this study, DFS has been defined as the time between the point of diagnosis and the first sign of relapse of RCC [5]. DFS was studied instead of Overall Survival given the high chance of cancer recurrence following nephrectomy [6].

It is hypothesized that genes significantly associated with DFS are associated with pathways that can be targeted for gene therapy or can serve as potential biomarkers for RCC.

2 Methodology

2.1 Overview

The flow of our methodology has been presented in Fig. 1. Based on data from TCGA database, genes with the most significant difference in relapse time ($p < 0.004$) between high and low expression of the genes were taken for each of the 4 subsets of RCC. RCC patients in the data set were arranged in order of their expression of a given gene, with the bottom quartile taken as the high expression population and the top quartile taken as the low expression population.

Kaplan-Meier plots were plotted for each of these genes, and a rigorous selection procedure was employed to allow for the screening of potential biomarkers and targets for gene therapy, which will be further explained below.

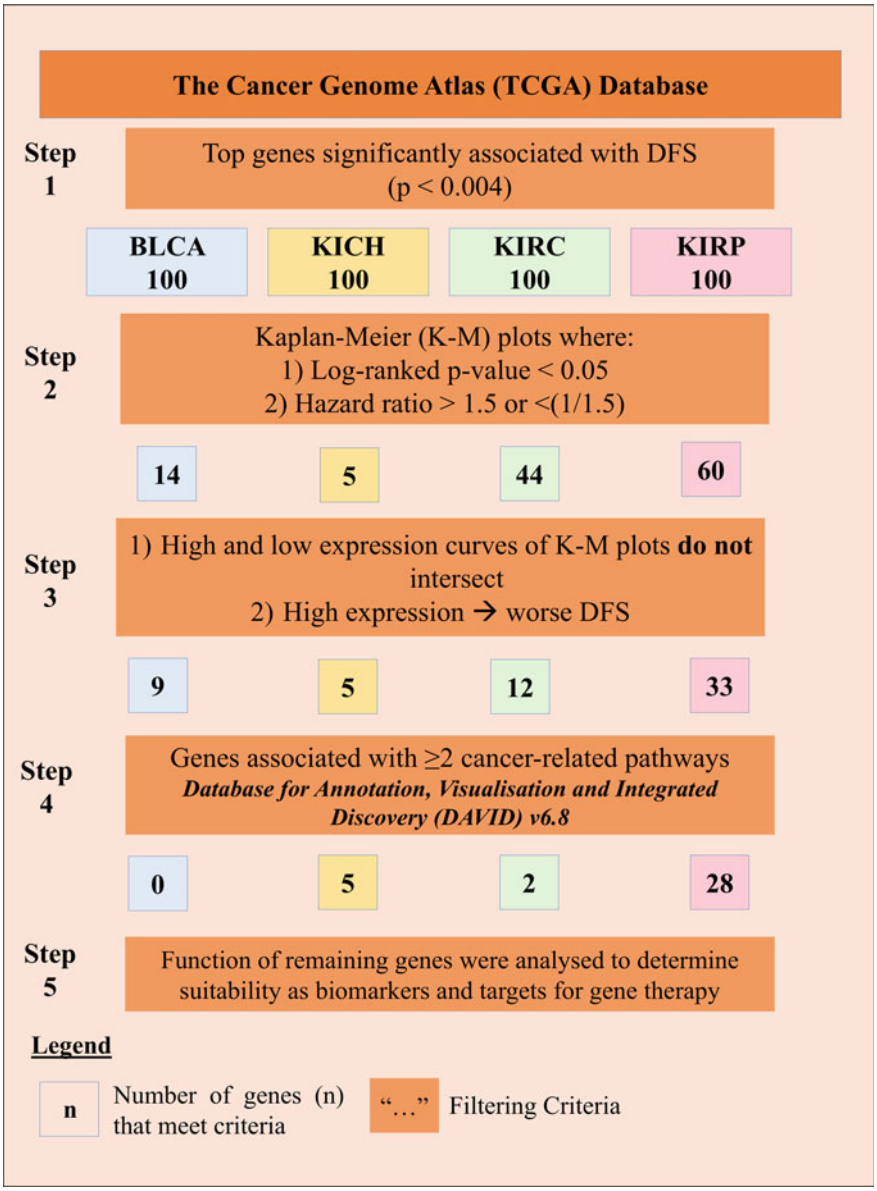

Fig. 1 Overview of methodology reflecting the number of genes selected for further downstream analysis at each step of our project. The orange boxes include the criteria used to select the genes while the colour-coded boxes reflect number of genes for each RCC subtype that meet criteria following the step taken

2.2 Selecting Differentially Expressed Genes from the Cancer Genome Atlas

Genes with most significant differential expression between tumor cells of patients with good and poor DFS were obtained from the TCGA database [7]. The most significantly differentially expressed genes ($p < 0.004$) were taken for each of the 4 subsets of RCC.

2.3 Kaplan-Meier Plots

From the raw data obtained from TCGA, Kaplan-Meier (K-M) survival curves were plotted for each gene obtained (Fig. 2). These curves represent processed data from a K-M statistical analysis, and are graphs of Percent Survival/probability against Time/months, and are hence used to estimate population survival over time. RCC patients in the data set were arranged in order of their expression of a given gene, with the highest 25% taken as the high expression population and the lowest 25% taken as the low expression population. Each population was monitored over time for relapse, allowing survival probability to be calculated according to (1):

$$\text{survival probability} = \frac{\text{number of patients yet to relapse}}{\text{total number of patients}}. \tag{1}$$

As can be seen in Fig. 2, each plot has two curves, and each curve represents the proportion of patients that have yet to experience a relapse with high gene expression (red line) or low gene expression (blue line) for the respective genes [8]. The curves are not smooth, but rather are a series of downward steps occurring each time a patient experiences a relapse [9].

From the K-M plots, we were able to determine whether higher expression or lower expression of the genes in the tumors were associated with better DFS of patients.

2.4 Log-Rank Test and Hazard Ratio

Each K-M plot also includes the log-rank p-value obtained from a log-rank test of the data. The log-rank test calculates the chi-square value for each event time (i.e. the time taken for RCC relapse to occur) for each curve and sums the results, which was then added to derive the final chi-square value to compare the two curves of each plot.

If the final p-value was less than 0.05, the survival times of the two plots were taken as significantly different from each other, indicating that level of gene expression of

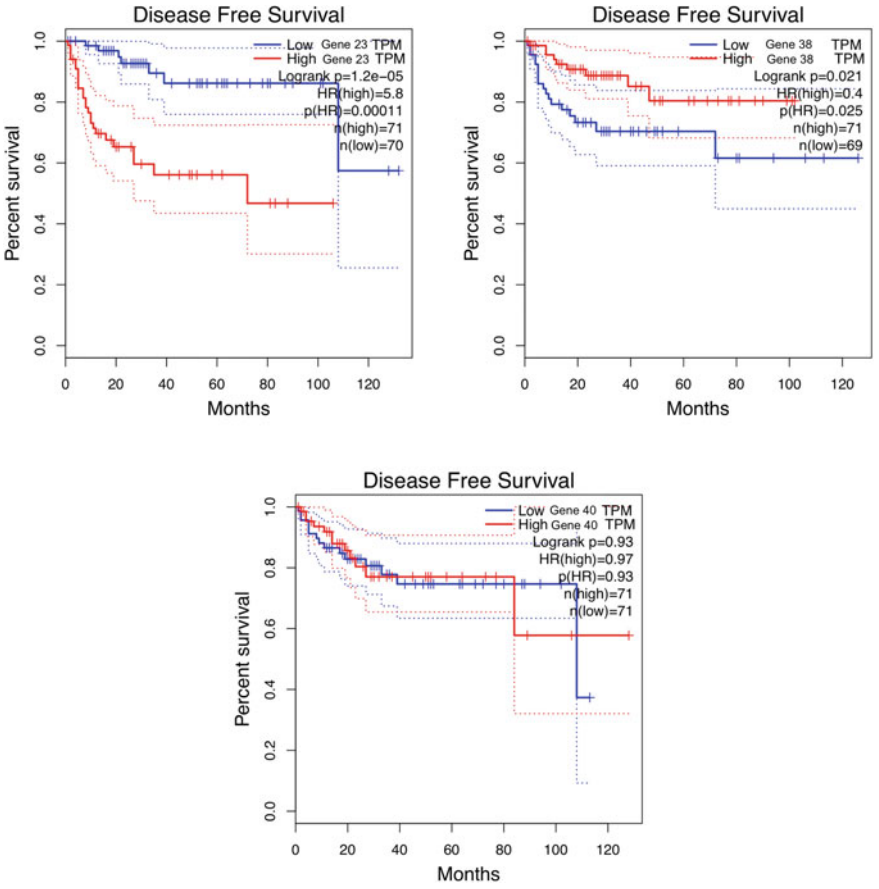

Fig. 2 Kaplan-Meier survival plots showing the difference in DFS between KIRP patients with high gene expression (red line) and low gene expression (blue line) of Genes 23, 38 and 40

the specific gene was more likely to be associated with the DFS of the patients. As such, genes with log-ranked p-values above 0.05 were removed from the list of genes for each cancer.

Each plot also has a hazard ratio, derived from the Cox proportional hazards model, which is a regression method for survival data [10]. It indicates the relative likelihood of the hazardous outcome occurring (i.e. relapse of RCC) in the group with patients experiencing high levels of gene expression compared to patients with low levels of gene expression. A hazard ratio of greater than or less than 1 indicates that DFS was better in one of the groups [11].

However, unlike the log rank test, there is no standard value for hazard ratios upon which the difference between the two groups would be considered significant. As such, to filter out genes not significantly associated with DFS in patients, the gene list was narrowed down to those with hazard ratios greater than 1.5 or less than 1/1.5.

Hence, only genes whose increased expression led to a 1.5 times increased likelihood of experiencing no DFS (hazard ratio greater than 1.5) or DFS (hazard ratio less than 1/1.5) were considered.

2.5 Criterion for Kaplan-Meier Plots

From the remaining genes (log-ranked p-value < 0.05, hazard ratio > 1.5 or < 1/1.5), those with high expression and low expression curves crossing in their respective Kaplan-Meier plots were not considered for future stages. This is because in such a case, the differential expression of the gene is unlikely to predict the DFS status. Furthermore, the curves do not indicate that there is a consistent effect of gene expression on DFS of the respective type of RCC, and thus the genes were rejected. An example of such a plot would be that of Gene 40 in Fig. 2.

Additionally, only genes whose high expression were found to lead to poorer DFS were further analyzed. Such genes are preferred as it is easier to downregulate gene expression rather than attempting to upregulate it in gene therapy, and products of genes with high expression would be more easily detected, making them also better potential biomarkers. This is reflected in the K-M plot as the high expression curve being below the low expression curve. An example of such a plot would be that of Gene 23 in Fig. 2.

Genes whose low expression were more likely to lead to poorer DFS were thus eliminated. This is reflected in the K-M plot as the low expression curve being below the high expression curve. An example of such a plot would be that of Gene 38 in Fig. 2.

Hence, using the examples given in Fig. 2, Gene 40 is rejected due to intersecting curves and gene 38 is rejected as high gene expression (red curve) led to better DFS compared to low gene expression (blue curve), while Gene 23 will be further analysed as it met both criteria set for Kaplan-Meier plots (Fig. 2).

2.6 Database for Annotation, Visualisation and Integrated Discovery (DAVID) v6.8

DAVID v6.8 is a web-based functional annotation tool that includes an integrated annotation knowledgebase, and provides various bioinformatics tools to analyze pathways enriched by specific gene sets [12, 13]. Using DAVID v6.8, genes whose expression can significantly predict the DFS status of patients with hazard ratios greater than 1.5 or less than $-1/1.5$ were analyzed, and then grouped according to their functional pathways. For each cancer, genes that were found in two or more pathways were identified as genes with great potential to be useful as targets for gene therapy or biomarkers for better prognosis.

3 Results and Discussion

3.1 Overview

In the discussion below, gene names have been substituted to alternative names due to intellectual property matters.

After being grouped into their functional pathways using the DAVID program, the open source software platform Cytoscape, in conjunction with the Agilent Literature Search Software, was used to illustrate the interaction network between the various genes in each group for easier visualization [14–17]. However, not all of the genes in this study were recognized by Agilent Literature Search Software due to the limitations of the databases linked, though they were still taken into account in later stages of the methodology, just not reflected in the respective pathway diagrams.

Aside from BLCA, significant genes in the other three RCC types were found to be associated with DFS. The resulting genes obtained are of interest as they can serve as potential biomarkers or as targets for gene therapy.

Biomarkers are used to better guide therapy and enable more accurate prognosis of renal cancers. Biomarkers are defined as "any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease" by the World Health Organization (WHO) [18]. Biomarkers are helpful in early prognosis, and treatment of various diseases, and is important for cancer. For example, the BRCA1 germline mutation has been used in estimating the risk of developing breast and ovarian cancer while the prostate specific antigen biomarker has been used in the screening for prostate cancer [19, 20]. However, there has yet to be a satisfactory biomarker (i.e. associated with survival of patients and easily measured) introduced for the prognosis of the different subsets of renal cancers, thereby adding to the significance of our work.

At the same time, the resulting genes obtained can also be targets of gene therapy. For instance, genes whose high expression are associated with poor DFS rates and are involved in pathways associated with the hallmarks of cancer can be targeted by therapy to downregulate their expression to prevent relapse in RCC patients. Given the rising prevalence of renal cancers, the possible useful applications of our results are thus laudable.

3.2 Kidney Chromophobe

Five genes, Genes 1, 2, 3, 4 and 5, fulfilled all given criteria (Fig. 3).

Given the high hazard ratios of these genes ranging from 4.9 to 10, their high expression is shown to negatively impact DFS of KICH patients. Furthermore, these genes all code for secretory proteins found in blood plasma, making them good candidates as KICH plasma biomarkers.

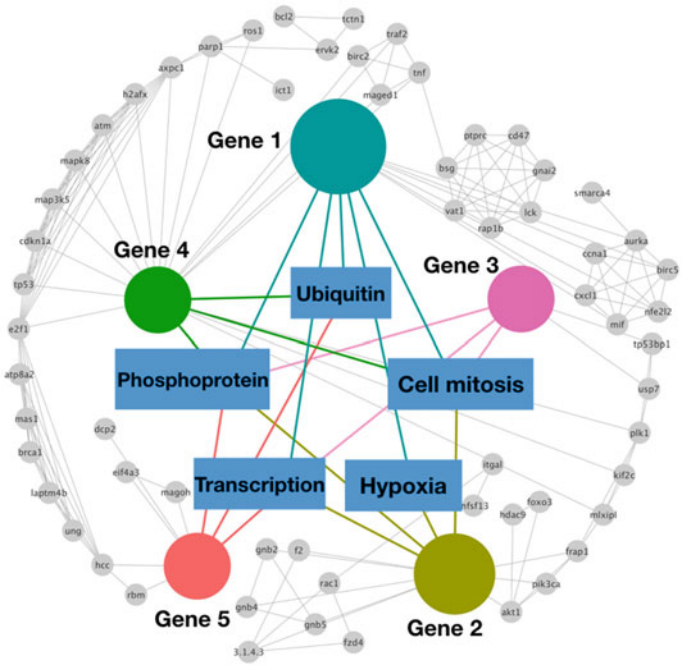

Fig. 3 Network of genes and pathways where higher expression of these genes is associated with poorer DFS of KICH patients. Blue rectangles represent pathways, while the circles represent genes. Non-grey circles represent the genes significantly associated with DFS while grey circles represent genes associated with the said genes. The larger the circle, the greater the number of pathways the gene is found in

Gene 1 and Gene 5, with high hazard ratios of 10 and 9.9, along with low log ranked p-value of 0.0077 and 0.011 respectively. These two genes are involved in cancer-related pathways: they affect ubiquitin function, and are involved in transcription and thus could regulate the expression of other cancer-related genes. The latter is the main function of Gene 5, which codes for a protein that is involved in mRNA splicing. The two genes also code for phosphoproteins, and excessive phosphorylation of proteins has been known to be associated with the emergence of cancers [21].

Furthermore, Gene 1 is also involved in cell mitosis and thus affects tumor growth, as well as cellular response to hypoxia, a pathway that is important in circumventing tumor hypoxia common in solid tumors. Specifically, Gene 1 functions to encode for cyclins involved in transition through cell cycle checkpoints, and upregulation of Gene 1 would thus result in increased proliferative signaling in the cell, which is a hallmark of cancer [22].

Given the high association of these two genes with DFS, and their apparent involvement in cancer-related pathways, they can potentially be used as targets for gene therapy. As the high expression of these two genes are strongly associated with

poorer DFS, it may be possible to employ gene therapy or other approaches to reduce the expression of these genes to better control the disease.

3.3 *Kidney Renal Clear Cell Carcinoma*

44 genes had significant p-values and hazard ratios. Out of these 44 genes, 12 had Kaplan-Meier plots without any intersecting survival curves, and when upregulated, were associated with poorer DFS. Out of the 12 genes, eight were recognized by Agilent Literature Search, and are reflected above (Fig. 4). As can be seen, only two genes were found to be associated with two or more pathways: Gene 6 and Gene 7.

Gene 6 is involved in the transcription and sprouting angiogenesis pathways. Sprouting angiogenesis is the mechanism of blood vessel growth, which includes growth toward tumor cells. This process provides oxygen and nutrients to enable further tumor growth, and the blood vessel proximity to the tumor facilitates metastasis [23]. Metastasis, a hallmark of cancer, will greatly increase chances of cancer relapse due to difficulty in eliminating all secondary tumor and circulating tumor

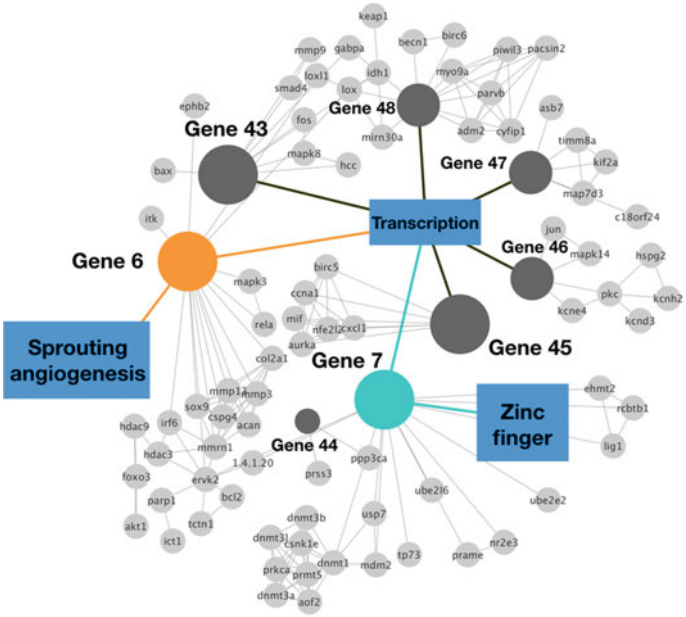

Fig. 4 Network of genes and pathways where higher expression of these genes is associated with poorer DFS of KIRC patients. Blue rectangles represent pathways, while the circles represent genes. Non-pale-grey circles represent the genes significantly associated with DFS while pale-grey circles represent genes associated with the said genes. Colourful circles represent genes that are associated with two or more pathways. The larger the circle, the greater the number of pathways the gene is found in

cells during treatment [24]. Promoting metastasis could be one of the reasons why high expression of Gene 6 is associated with poorer DFS.

Meanwhile, Gene 7 is involved in transcription and affects zinc finger function. Since zinc fingers are the largest transcription family in humans and are involved in RCC progression, this suggests Gene 7 affects cancer progression through affecting gene expression in cancer cells [25].

The hazard ratios of Gene 6 and Gene 7 are 2.7 and 2.2, while their log ranked p-value is 0.00027 and 0.0021 respectively.

Hence, they are clearly associated with DFS, and their function further suggests their association with relapse of KIRC. Since products of expression of Gene 6 and Gene 7 are not secreted, they can only function as biomarkers if there are circulating tumor cells which are extractable from the patient's blood. Additionally, these two genes could potentially be targeted for gene therapy, as their high expression is associated with the relapse of KIRC.

3.4 Kidney Renal Papillary Cell Carcinoma

60 genes had significant p-values and hazard ratios. Out of these 60 genes, 33 had Kaplan-Meier plots without any intersecting survival curves, and when upregulated, were associated with poorer DFS.

23 of the 33 remaining significant genes were recognized by Agilent Literature Search (Fig. 5). However, there were 28 genes that were associated with two or more pathways. Of these 28, there were 3 genes, Gene 8, Gene 9 and Gene 10, with hazard ratios more than or equal to 10, and log rank p-values of less than 10^{-7}. This signifies an extremely strong association with DFS.

All three of these genes are involved in transcription and cell mitosis pathways. Gene 8 encodes for a protein belonging to a dual specificity protein phosphatase family which regulates the cell cycle. Gene 9 codes for a component of the essential kinetochore-associated NDC80 complex, which is required for chromosome segregation, spindle checkpoint activity and kinetochore stability. Gene 10 codes for histone H3-like nucleosomal protein that is specifically found in centromeric nucleosomes.

Upregulation of these three genes will thus increase the proliferative rate of cells, which is a hallmark of cancer. The functions and pathways these 3 genes are involved in are intrinsically involved in the cell cycle, which in turn could affect cancer progression when upregulated and thus rates of cancer relapse. Hence, Genes 8, 9 and 10 present themselves as potential targets for gene therapy. Additionally, genes 9 and 10 codes for secretory proteins found in blood plasma, making them good candidates for biomarkers.

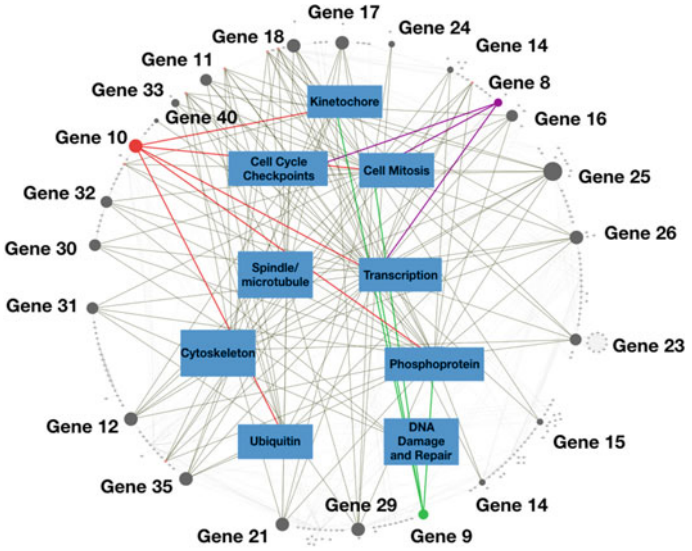

Fig. 5 Network of genes and pathways where higher expression of these genes is associated with poorer DFS of KIRP patients. Blue rectangles represent pathways, while the circles represent genes. Non-pale-grey circles represent the genes significantly associated with DFS while pale-grey circles represent genes associated with the said genes. Colourful circles represent genes that are associated with two or more pathways. The larger the circle, the greater the number of pathways the gene is found in

3.5 Bladder Urothelial Carcinoma

As can be seen in Fig. 6, only Gene 36 and 37 were found to be associated with one pathway, related to the centrosome.

As such, when genes that were found in less than two pathways were removed from the gene list, there were no genes left associated with the DFS of BLCA patients. Therefore, in this investigation, none of the significant genes of BLCA patients were associated with DFS of the patients, and thus no potential biomarkers or targets for gene therapy can be inferred for BLCA.

4 Conclusion

Out of the four cancers studied, three cancers (KICH, KIRC and KIRP) had genes that met all the set criteria of having suitable K-M plots, and were associated with two or more cancer-related pathways. The expression levels of these genes were found to be significantly associated with DFS, with high expression leading to poor DFS, and the genes are associated with two or more cancer-related pathways.

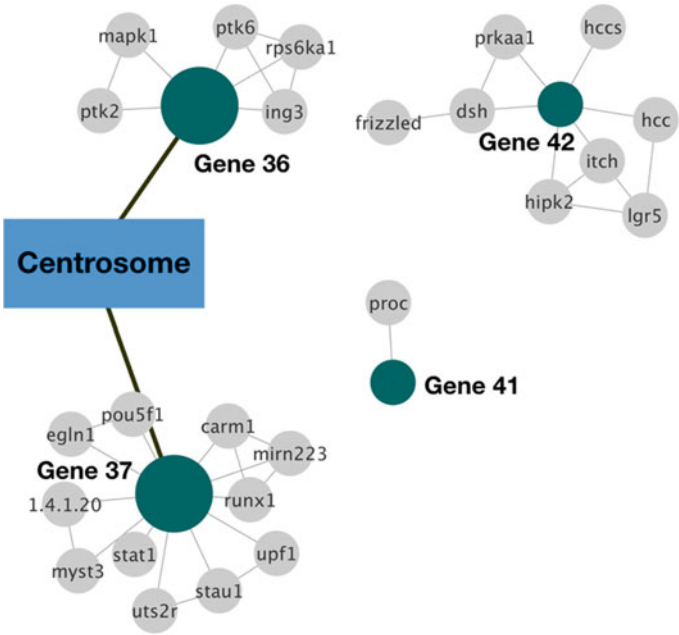

Fig. 6 Network of genes and pathways where higher expression of these genes is associated with poorer DFS of BLCA patients. Blue rectangles represent pathways, while the circles represent genes. Non-pale-grey circles represent the genes significantly associated with DFS while pale-grey circles represent genes associated with the said genes. The larger the circle, the greater the number of pathways the gene is found in

These genes obtained could be considered as potential biomarkers and/or targets for gene therapy. This would ensure better prognosis and more effective treatments for patients of the respective cancers.

In summary, genes identified as potential biomarkers and targets for gene therapy are Genes 1–5 for KICH, Genes 6 and 7 for KIRC, and Genes 8–10 for KIRP. Potential biomarkers of particular interest would be the protein products of Genes 1–5 for KICH and Genes 9 and 10 for KIRP, as these genes' protein products are secreted, making them more easily detectable given a blood sample. The results have been summarized in Table 1, with all the noteworthy genes having been included.

As can be seen in Table 1, most of the genes that have been identified are involved in the pathways of transcription, cell cycle and metastasis. It is also interesting to note that the genes significant to DFS of KIRC patients differ vastly from those significant to DFS of the other two cancers, in terms of the pathways that the genes are associated with. This could be due to different developmental pathways of the cancer.

Table 1 Summary table

Cluster	Pathway	Gene Number										Total in Pathway	Total in Cluster
		KICH					KIRC		KIRP				
		1	2	3	4	5	6	7	8	9	10		
Transcription	Transcription											9	21
	Phosphoprotein											7	
	Ubiquitin											4	
	Zinc Finger											1	
Cell cycle	Mitosis											7	10
	Kinetochore											2	
	Cell Cycle Checkpoints											1	
Metastasis	Hypoxia											2	3
	Angiogenesis											1	

Acknowledgements The authors thank their mentors, A/P Caroline Lee Guat Lay and Mr. Lee Wai Yeow for their constant guidance and support.

References

1. Motzer, R. J., Bander, N. H., & Nanus, D. M. (1996). Renal-cell carcinoma. *New England Journal of Medicine, 335*(12), 865–875. https://doi.org/10.1056/nejm199609193351207.
2. World Cancer Research Fund International. *Kidney cancer statistics*. Retrieved from http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/kidney-cancer-statistics.
3. World Health Organisation. (2017). *Obesity and overweight*. Retrieved from http://www.who.int/mediacentre/factsheets/fs311/en/.
4. Chow, W., Dong, L. M., & Devesa, S. S. (2010). Epidemiology and risk factors for kidney cancer. *Nature Reviews Urology, 7*(5), 245–257. https://doi.org/10.1038/nrurol.2010.46.
5. Denkert, C., et al. (2003). Elevated expression of cyclooxygenase-2 is a negative prognostic factor for disease free survival and overall survival in patients with breast carcinoma. *Cancer, 97*(12), 2978–2987. https://doi.org/10.1002/cncr.11437.
6. Chin, A. I., et al. (2006). Surveillance strategies for renal cell carcinoma patients following nephrectomy. *Reviews in Urology, 8*(1), 1–7.
7. Weinstein, J. N., et al. (2013). The Cancer Genome Atlas pan-cancer analysis project. *Nature Genetics, 45*(10), 1113–1120.
8. Chakraborty, R., Rao, C. R., & Sen, P. K. (2012). *Bioinformatics in human health and heredity* (Vol. 28). Newnes Books.
9. Barraclough, H., Simms, L., & Govindan, R. (2011). Biostatistics primer: What a clinician ought to know: Hazard ratios. *Journal of Thoracic Oncology, 6*(6), 978–982. https://doi.org/10.1097/jto.0b013e31821b10ab.
10. Rich, J. T., et al. (2010). A practical guide to understanding Kaplan-Meier curves. *Otolaryngology—Head and Neck Surgery, 143*(3), 331–336. https://doi.org/10.1016/j.otohns.2010.05.007.
11. Spruance, S. L., Reid, J. E., Grace, M., & Samore, M. (2004). Hazard ratio in clinical trials. *Antimicrobial Agents and Chemotherapy, 48*(8), 2787–2792. https://doi.org/10.1128/AAC.48.8.2787-2792.2004.
12. Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols, 4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211.

13. Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2008). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research, 37*(1), 1–13. https://doi.org/10.1093/nar/gkn923.
14. Shannon, P., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303.
15. Vailaya, A., et al. (2004). An architecture for biological information extraction and representation. *Bioinformatics, 21*(4), 430–438. https://doi.org/10.1093/bioinformatics/bti187.
16. King, J. Y., et al. (2005). Pathway analysis of coronary atherosclerosis. *Physiological Genomics, 23*(1), 103–118. https://doi.org/10.1152/physiolgenomics.00101.2005.
17. Ashley, E. A., et al. (2006). Network analysis of human in-stent restenosis. *Circulation, 114*(24), 2644–2654. https://doi.org/10.1161/circulationaha.106.637025.
18. World-Health-Organisation. (2001). *Biomarkers in risk assessment: Validity and validation*. Retrieved from http://www.inchem.org/documents/ehc/ehc/ehc222.htm.
19. Easton, D. F., et al. (1995). Breast and ovarian cancer incidence in BRCA1-mutation carriers. *American Journal of Human Genetics, 56*(1), 265–271.
20. Lin, K., Lipsitz, R., Miller, T., & Janakiraman, S. (2008). Benefits and harms of prostate-specific antigen screening for prostate cancer: An evidence update for the U.S. Preventive Services Task Force. *Annals of Internal Medicine, 149*(3), 192–199.
21. Singh, V., et al. (2017). Phosphorylation: Implications in cancer. *The Protein Journal, 36*(1), 1–6. https://doi.org/10.1007/s10930-017-9696-z.
22. Hanahan, D., & Weinberg, R. (2011). Hallmarks of cancer: The next generation. *Cell, 144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013.
23. Hillen, F., & Griffioen, A. W. (2007). Tumour vascularization: Sprouting angiogenesis and beyond. *Cancer and Metastasis Reviews, 26*(3–4), 489–502. https://doi.org/10.1007/s10555-007-9094-7.
24. Flemming, A. (2015). Cancer stem cells: Targeting the root of cancer relapse. *Nature Reviews Drug Discovery, 14*(165). https://doi.org/10.1038/nrd4560.
25. Jen, J., & Wang, Y.-C. (2016). Zinc finger proteins in cancer progression. *Journal of Biomedical Science, 23*(53). https://doi.org/10.1186/s12929-016-0269-9.