Early Prognosis of Human Renal Cancer with Kaplan-Meier Plotter Data Analysis Model

Saif Almansouri1* and Susan Zwyea2*

1 Department of Computer Engineering, College of Engineering, Al-Nahrain University, Baghdad, Iraq
2 Department of Radiological Techniques, College of Health and Medical Technology/Baghdad, Middle Technical University, Baghdad, Iraq

*The authors have contributed equally to this work

Correspondence: Dr.susanzwyea@mtu.edu.iq

Abstract. Clinical data analysis is one of the powerful learning methods in cancer research. Several analysis methods have been used for detection purposes in computational pathology. However, little information is known about the model features. Here, we described Kaplan-Meier plotter analysis model as a powerful tool with new features. The model combines follow up threshold, disease stage, and race to ensure better validation for genes as prognostic biomarkers in early disease stages. The proposed model is evaluated for the relevance role of Rab1A, an oncogene, in renal cancer early prognosis on the benchmark datasets from The Human Protein Atlas. We found Rab1A overexpression in human renal cancer has potential role in early prognosis of the disease and it is associated with poor prognosis (p<0.05). Our model results were also confirmed in an independent dataset in The Human Protein Atlas. Together, our studies emphasize the role of Rab1A in human malignancies and identify Rab1A as a new prognostic predictor for human renal cancer.

Keywords: Kaplan-Meier plotter, Prognosis, Biomarker, Renal cancer, Rab1A

Introduction
Renal cancer is the tenth most common leading cause of cancer death in both men and women in the United States [1, 2]. The incidence of renal cancer in the United States is 73,000 new cases and 14000 people died from this cancer yearly [2]. The survival rate of the disease after successful treatment in the early stages is high (94%). However, if renal carcinoma is not detected in its early stages, then the treatment cannot ensure a successful and full recovery. Therefore the early detection and prognosis of a disease is very important. As a result, many researchers from different fields such as biomedical and the bioinformatics fields improve many methods that can be utilized in medical research as a model to facilitate an early and accurate prognosis and detection of a disease type [3]. Among these most important techniques are Machine Learning (ML), clinical images analysis, and miRNA profiling. Moreover, these techniques can predict a cancer type from complex datasets [4-10].

Machine learning is one of the areas in artificial intelligence. It used different techniques and algorithms with two phases of learning process. The first phase uses a dataset to estimate unknown dependencies in a system. The second phase is the prediction of new outputs from estimated dependencies of the given system. Machine learning has a variety of successful applications in different fields such as biomedical research. In the last few years, machine learning shows good performance for cancer
susceptibility, prognosis and prediction tasks [11-16]. Due to several complexities, the machine learning detection method cannot detect cancer in early stages [17]. So researchers tend to identify diseases from clinical images. Clinical images analysis, a magnified image of the cancer lesions, are used to enhance the visual effects of the regions of interest for obtaining very detailed, deeper levels of lesion regions. It has been reported that magnified images show a higher accuracy compared to the naked eye [10,18]. In addition, using naked eye tends to be very time consuming and error-prone process. This accuracy may drop drastically if clinicians have not been adequately trained [10,18]. Yang et al. [19] recently showed that in the clinical images cameras are being used varied in orientations, illuminations, lighting conditions, and other artifacts make this problem difficult to analyze with automated approaches. As a result, this issue causes low recognition accuracy in identifying diseases from various clinical images. Therefore, some scientists used another type of cancer detection technique name miRNA profiling. miRNA profiling is a class for cancer detection and identification that use gene expression signatures to predict a cancer type outcome. Yet this type of detection method shows high performance in cancer detection, it has low sensitivity regarding its use in screening at early disease stages. Moreover, it shows difficulties in discriminating benign from malignant tumors [6-8].

In the last few years, these previously mentioned techniques show superior performance in cancer detection [20-22]. However, an appropriate level of validation is needed in order for these techniques to be considered in clinical practice. In this paper, we propose and apply an improved tool for renal carcinoma prognosis. To accomplish our goal, the proposed model is evaluated with a set of experiments on dataset deposited in The Cancer Genome Atlas Research Network (TCGA) database as shown in Table 1. The contributions of this work can be summarized as follows:

- The prognosis model is proposed and applied to a kidney cancer on TCGA (RNA-seq) dataset.
- The impact of different integrated features in renal cancer early prognosis is investigated.
- The fusion (or integrated) features were then entered as input to the Kaplan–Meier analysis model for identification a new prognostic marker task.
- In order to evaluate the performance of the model, the impact of Rab1A, a member of RAS oncogene family, from TCGA datasets to The Human Protein Atlas dataset and clinical images is evaluated for the prognosis task.

Table 1. Different features inputs for Kaplan–Meier analysis model result in different prognosis outcomes

| Method                              | Cancer type       | Important features as inputs       | Outputs                                                                 |
|-------------------------------------|-------------------|-----------------------------------|-------------------------------------------------------------------------|
| Classical (or basic) Kaplan Meier analysis model | Renal cancer       | Rab1A gene probe expression, overall survival, 60 months threshold follow up | Prognosis of renal cancer                                               |
| New enhanced Kaplan Meier analysis model | Renal cancer       | Rab1A gene probe expression, overall survival, 120 months threshold follow up, race, cancer stage 1 | Validation of Rab1A gene as prognostic marker and prognosis of renal cancer in early stages (stage 1) in white people. |

2. Preliminaries
2.1 Kaplan–Meier analysis new model for early prognosis of renal cancer

In 1958, Edward L. Kaplan and Paul Meier collaborated to evaluate the prediction of cancer survival [23]. The integration and extension of the set of features in Kaplan–Meier analysis model plays a crucial role in renal cancer prognosis. The basic or classical Kaplan–Meier analysis model has only 15 features. Among these most important features are: fellow up threshold, cutoff value, censor at threshold, probe expression, median survival, trichotomization, survival, tumor grade (grade 1 to grade 4), mutation burden (high or low), and gender (male and female). In the fellow up threshold feature, several periods of time after patient’s treatment (from 2 months to 240 months) are available as an input data. For the cutoff value, it shows all possible cutoff value used to split patients between the upper and lower quartile. For censor at threshold, this feature allows patients surviving over the selected threshold to be censored instead of excluded. In the probe expression feature, single or multiple gene(s) symbols can be entered for validation of gene of interest as prognostic marker. For median survival, it helps computing median surviving when both cohorts reach median survival. In trichotomization, the feature allows to compare patients in upper quartile versus lower quartile [24]. However, the classic or the basic model cannot detect a cancer type in early stages. For most cancer types if it is not detected in early stages, the treatment cannot ensure a successful and full recovery. In order to enhance the model detection tasks of a cancer type at the early stages, we integrated new restriction features to the analysis along with the basic features so the analysis can be even more restricted to the subtypes such as cancer histological stage (range from stage 1 to stage 4) and race (white, Asian, and African American). The integration of new features to the model is very crucial in producing better and more accurate outputs. The proposed model and the advantages of the extension of the set of features are evaluated with a set of patient data that deposited in the TCGA dataset of 806 patients considered out of 7642.

2.2 Evaluation Kaplan–Meier analysis model results with the new integrated features to The Human Protein Atlas databases

Identification of a prognostic marker for a cancer type is one of the challenging tasks therefore the new model results were evaluated on the benchmark genomic data from The Human Protein Atlas databases (https://www.proteinatlas.org) before it considered for clinical trials or clinics. The website (proteina.png) was used to download data including RNA seq data. Then the data was used to analyze Rab1A mRNA expression in 877 renal cancer samples and 32 of normal renal tissue samples. The samples were obtained and prepared at the Pathology Clinic, Uppsala University Hospital (Uppsala, Sweden). The overall survival time was computed starting from the day of therapy to the end of the follow-up or the death date because of the recurrence and metastasis [25]. Patients with low Rab1A mRNA expression were all alive while patients with high Rab1A mRNA expression were 651 alive and 226 dead. The log-rank test was used to evaluate the significance of the correlation between Rab1A mRNA expression and patient survival in The Human Protein Atlas databases. Results are considered significant when p value is less than 0.05. The results were then compared to our model results for evaluation purposes.

2.3 Rab1A is overexpressed in human renal cancer

Genomic findings from previous studies suggest that Rab1A gene acts as an oncogene and is amplified in many cancer cases [26-33]. Little is known about its role of aberrant expression of Rab1A in renal cancer. Therefore, this study investigated Rab1A mRNA level in 877 kidney cancer samples and 32 normal kidney tissues that available in The Human Protein Atlas database (25).

In support the task of high Rab1A expression in signaling pathways and oncogenesis [26,27], we investigated the level of P-S6K1, the effector of mTORC1 which is the master regulator of proliferation in cancer cells, in cancer renal tissue and normal renal tissue samples. To investigate this phenomenon, we
checked the P-S6K1 gene expression levels of renal cancer patients by using RNA-seq data deposited in The Human Protein Atlas database (25). The samples for Rab1A and P-S6K1 mRNA were obtained and prepared at the Pathology Clinic, Uppsala University Hospital (Uppsala, Sweden). The standard RNA-seq protocol was used for measuring mRNA quantity. Then RNA was extracted from tissue samples. Samples of high-quality RNA were used in the sequencing.

3. Results and Discussion

3.1 Kaplan–Meier analysis model has potential role in early prognosis of renal cancer

The accurate prognosis of the possible outcomes of a disease is one of the most interesting and difficult tasks for physicians. Therefore, many models have been developed to enhanced detection of diseases. One of the most important models is Kaplan–Meier analysis model. According to the basic input features of Kaplan–Meier analysis model (such as cancer type, fellow up threshold, survival, probe gene and so on), the analysis can predict a cancer type but not in the early stage. Here in this paper, we have evaluated Rab1A as a probe to predict renal cancer using fellow up threshold 60 months and overall survival. Strikingly, the analysis results show renal cancer patients with high expression of Rab1A have overall poor survival compared to patients with low Rab1A expression (p=0.048). A pictorial representation of the model results is shown in Figure 1A. Interestingly, Rab1A has been reported previously to be highly overexpressed in human tongue squamous cell carcinomas, human prostate cancer, colorectal cancer (CRC), hepatocellular carcinoma (HCC), pancreatic cancer and gliomas [26-33]. However, basic input features of Kaplan–Meier analysis model cannot detect kidney cancer in early stages. Therefore, in this paper we propose a new model. Different new restriction features (fellow up threshold 120 months, histological stage and white race) were added to the analysis model to enhance the prediction of the disease at the early stages. The experiment results of the model with new added features help enhancing the sensitivity of the detection model at the early stages (p<0.05) (Figure 1B). The results show that Rab1A is associated with poor prognosis in white people with stage 1 kidney cancer. This model is able to produce a better prognosis for kidney cancer in the early stages compared to the basic or classical input features of Kaplan–Meier analysis model. Furthermore, we also tried to test whether the impact of gender type on the new prediction model. Our results show that the gender has no significant impact on the early prediction of kidney cancer (p>0.05) (Figure 1C). The new set of features (fellow up threshold 120 months, histological stage and white race) for Kaplan–Meier analysis model which was applied was more informative and effective. Our results demonstrate that the new model may have a potential role in early prognosis of other diseases.

Figure 1: Overexpression of Rab1A in renal cancer is associated with poor survival A.

Kaplan-Meier survival analysis model using the default (classical) model features comparing the overall survival time of renal cancer cases separated into two groups, low or high mRNA level of Rab1A.
The p value was computed by log rank test. **B. Meier survival analysis model with the two new added features (white race and stage1 of the disease) comparing the overall survival time of renal cancer cases separated into two groups, low or high mRNA level of Rab1A. C. Meier survival analysis model with three new added features (white race and stage1 of the disease, and male gender) comparing the overall survival time of renal cancer cases separated into two groups, low or high mRNA level of Rab1A.**

### 3.2 Genomic data from The Human Protein Atlas databases further confirm the potential role of Kaplan–Meier analysis model in detection of renal cancer

One of the most challenging and the scaring task for the physicians is the treatment option for renal cancer patients. Therefore, we validated our model results on the benchmark datasets from The Human Protein Atlas (25). Analysis of the primary data shows patients were separated into two groups by the median value for Rab1A expression level. The analysis reveals that overall survival of renal cancer patients with high Rab1A expression is significantly worse than those with low Rab1A expression in renal patients (p=0.048) (Figure 2). Thus, Rab1A level can provide predictive value for the outcome of renal cancer patients. These observations resemble those seen in our new model and previous studies [26-27]. Our analysis results further support the finding that Rab1A overexpression is significantly associated with an elevated risk of renal cancer related death. These findings demonstrate that Rab1A is an independent prognostic signature of poor survival in renal cancer patients.

![Figure 2: Overexpression of Rab1A in renal cancer is associated with poor survival. Survival analysis from The Human Protein Atlas database](https://www.proteinatlas.org/ENSG00000138069-RAB1A/pathology/tissue/renal+cancer) comparing the overall survival time of renal cancer cases separated into two groups, low or high mRNA level of Rab1A. The p value was computed by log rank test.

### 3.3 Kaplan–Meier analysis model support the role of Rab1A in oncogenesis and signaling pathways

Recent studies suggest that Rab1A gene acts as an oncogene and play a critical role in activating signaling pathways. Here in this study, we validate overexpression role Rab1A in renal cancer tissue samples. Analysis of the primary data that deposited in The Human Protein Atlas (25) shows Rab1A mRNA level were higher in renal cancer tissues compared to normal renal tissues (Figure 4A and 4B). Since the increase in Rab1A occurs at transcriptional level, this suggests that increased in Rab1A mRNA level is due to the change in Rab1A gene expression. Generally, our results indicate that high Rab1A expression level in renal cancer resemble those seen in other studies such as HCC and CRC [26,27]. Therefore, we think that the results of this study are useful as supporting evidence that may explain many of the findings related to the development of renal cancer. Moreover, we explore the role of Rab1A in mTORC1...
signaling pathway in particular its molecular effective, P-S6K1 protein. Protein primary data analysis reveals that P-S6K1 protein level is generally higher in renal cancer tissues compared to normal renal cancer tissues (Figure 3A and 3B). Altogether these outcomes demonstrate that enhancing mTORC1 signal may be crucial for Rab1A to advance the pathogenesis of renal cancer and that Rab1A could be a new guide for targeted cancer therapy as seen in other malignancies [34].

**Figure 3: Rab1A is overexpressed in human renal cancer tissues.**

RNA-seq data of Rab1A has been mapped using the number Fragments Per Kilobase of exon per Million reads (FPKMs). Shown is mRNA level in 32 renal normal tissues in The Human Protein Atlas database. B. RNA-seq data of Rab1A is proposed as average number Reads Per Kilobase gene model and Million mapped reads (FPKMs). Shown is mRNA level in renal cancer of 877 samples. Data is
downloaded from The Human Protein Atlas database (https://www.proteinatlas.org/ENSG00000138069-RAB1A/pathology/tissue/renal+cancer).

Figure 4: p-S6K is overexpressed in human renal cancer tissues.

A. Shown is RNA-seq data of S6K for the non-cancerous renal tissues in The Human Protein Atlas database (https://www.proteinatlas.org/ENSG00000108443-RPS6KB1/tissue/kidney). B. Shown are representative RNA-seq data of p-S6K1 mRNA expression of renal cancer tissues in the The Human Protein Atlas database (https://www.proteinatlas.org/ENSG00000108443-RPS6KB1/pathology/tissue/renal+cancer).

4. Conclusion

In this paper, we have proposed an improved Kaplan–Meier analysis model. The model was evaluated for renal cancer early prognosis task. The Kaplan–Meier analysis model with new integration features was applied and evaluated on the TCGA dataset. The results demonstrate better performance with the new
integrated features to the basic features of the Kaplan–Meier analysis model. In addition, the model results were validated on the benchmark to datasets from The Human Protein Atlas (25) for renal cancer prognosis task. Strikingly, The Human Protein Atlas results were resembled and support those results seen with our new model. Further investigations include a more experimental evaluation for the role of Rab1A gene in oncogenesis and signaling pathways. The results demonstrate that Rab1A may have a potential role in renal cancer progression in an mTORC1 dependent manner, and further emphasizes the role of Rab1A in human malignancies. Together, these results indicate that Rab1A overexpression could promote a new prognostic predictor for human renal cancer. Thus, studying the regulation of Rab1A in renal cancer is urgently needed.

Acknowledgments

We thank the staff of the Human Protein Atlas program for valuable contributions.

References

[1] Chow WH, Dong LM, Devesa SS: Epidemiology and risk factors for kidney cancer. *Nat Rev Urol.*;7(5):245–57 (2010).
[2] Siegel RL, Miller KD, Jemal A: Cancer Statistics, 2017. *CA Cancer J Clin.*;67(1):7–30 (2017).
[3] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell;144: 646–74 (2011).
[4] Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, Dimitrios I. Fotiadis. Computational and Structural Biotechnology Journal 13 8–17 (2015).
[5] Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Informat*;2:59(2006).
[6] Fortunato O, Boeri M, Verri C, Conte D, Mensah M, Suatoni P, et al. Assessment of circulating microRNAs in plasma of lung cancer patients. *Molecules*;19:3038–54 (2014).
[7] Heneghan HM, Miller N, Kerin MJ. MiRNAs as biomarkers and therapeutic targets in cancer. *Curr Opin Pharmacol*;10:543–50 (2010).
[8] Madhavan D, Cuk K, Burwinkel B, Yang R. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. *Front Genet*;4(2013).
[9] Zen K, Zhang CY. Circulating-microRNAs: a novel class of biomarkers to diagnose and monitor human cancers. *Med Res Rev*;32:326–48 (2012).
[10] Silveira, Margarida, Jacinto C. Nascimento, Jorge S. Marques, André RS Marêal, Teresa Mendonêa, Syogo Yamauchi, Junji Maeda, and Jorge Rozeira. "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images." IEEE Journal of Selected Topics in Signal Processing 3, no. 1: 35-45(2009).
[11] Cicchetti D. Neural networks and diagnosis in the clinical laboratory: state of the art. *Clin Chem*;38:9–10(1992).
[12] Cochran AJ. Prediction of outcome for patients with cutaneous melanoma. *Pigment Cell Res*;10:162–7 (1997).
[13] Exarchos KP, Goletsis Y, Fotiadis D. Multiparametric decision support system for the prediction of oral cancer recurrence. *IEEE Trans Inf Technol Biomed*;16: 1127–34 (2012).
[14] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med*;23:89–109 (2001).
[15] Park K, Ali A, Kim D, An Y, Kim M, Shin H. Robust predictive model for evaluating breast cancer survivability. *Engl Appl Artif Intell*;26:2194–205 (2013).
[16] Sun Y, Goodison S, Li J, Liu L, Farmerie W. Improved breast cancer prognosis through the combination of clinical and genetic markers. *Bioinformatics*;23:30–7(2007).
[17] Li Shen, Laurie R. Margolies, Joseph H. Rothstein, Eugene Fluder, Russell McBride and Weiva Sieh. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Scientific Reports 9:12495(2019).

[18] Kittler, Harold, H. Pehamberger, K. Wolff, and M. Binder. "Diagnostic accuracy of dermoscopy." The lancet oncology 3, no. 3: 159-165(2002).

[19] Yang, Jufeng, Xiaoxiao Sun, Jie Liang, and Paul L. Rosin. "Clinical skin lesion diagnosis using representations inspired by dermatologist criteria." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1258-1266. (2018).

[20] Bottaci L, Drew PJ, Hartley JE, Hadfield MB, Farouk R, Lee PWR, et al. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. Lancet;350:469–72 (1997).

[21] Maclin PS, Dempsey J, Brooks J, Rand J. Using neural networks to diagnose cancer. J Med Syst;15:11–9 (1991).

[22] Simes RJ. Treatment selection for cancer patients: application of statistical decision theory to the treatment of advanced ovarian cancer. J Chronic Dis ;38:171–86 (1985).

[23] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 53:457–81(1958).

[24] Á, Lánczyk A, Menyhárt O, Györfy B: Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets, Scientific Reports, 8:9277 (2018).

[25] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Alm T, Edqvist PH, Berling H, Tegl J, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. Tissue-based map of the human proteome. Science 347(6220):1260419 (2015).

[26] Thomas JD, Zhang YJ, Wei YH, Cho JH, Morris LE, Wang HY, Zheng XF. Rab1A is an mTORC1 activator and a colorectal oncogene. Cancer Cell. ; 26:754–69(2014).

[27] Xu BH, Li XX, Yang Y, Zhang MY, Rao HL, Wang HY, Zheng XF. Aberrant amino acid signaling promotes growth and metastasis of hepatocellular carcinomas through Rab1A-dependent activation of mTORC1 by Rab1A. Oncotarget. ; 28;6(25):20813-28(2015).

[28] Bao ZS, Li MY, Wang JY, Zhang CB, Wang HJ, Yan W, Liu YW, Zhang W, Chen L, and Jiang T. Prognostic value of a nine-gene signature in glioma patients based on mRNA expression profiling. CNS Neurosci Ther. ;20:112–18(2014).

[29] Abdlmageed ZY, Yang Y, Thomas R, Ranjan M, Mondal D, Moroz K, Fang Z, Rezk BM, Moparty K, Sikka SC, Sartor O,and Abdel-Mageed AB. Neoplastic reprogramming of patient-derived adipose stem cells by prostate cancer cell-associated exosomes. Stem Cells. ;32:983–97(2014).

[30] Sun T, Wang X, He HH, Sweeney CJ, Liu SX, Brown M, Balk S, Lee GS, and Kantoff PW. MiR-221 promotes the development of androgen independence in prostate cancer cells via downregulation of HECTD2 and RAB1A. Oncogene. ;33:2790–800(2014).

[31] Shimada K, Uzawa K, Kato M, Endo Y, Shiiba M, Bukawa H, Yokoe H, Seki N, and Tanzawa H. Aberrant expression of RAB1A in human tongue cancer. Br J Cancer. ;92:1915–21(2005).

[32] Zwyea S. The role of overexpression of Rab1A in human pancreatic cancer and its association with poor prognosis. Indian journal of public health research and development; 10(11): 1854-1859 (2019).
[33] Nikoshkov A, Broliden K, Attarha S, Sviatoha V, Hellström AC, Mints M, and Andersson S. Expression pattern of the PRDX2, RAB1A, RAB1B, RAB5A and RAB25 genes in normal and cancer cervical tissues. Int J Oncol.; 46:107–12 (2015).

[34] Meng LH and Zheng XS. Acta Pharmacol Sin; 36:1163-1169 (2015).