Machine Learning Approach to Finding a MicroRNA Biomarker of Renal Cell Carcinoma with Therapeutic Abilities

In the United States, it is estimated that 200,000 people are living with renal cell carcinoma every year (Cairns, 2012). The mortality rate of renal cell carcinoma patients drastically increases from 19% to 26% to 47% and finally to 92% from stages 1, 2, 3, and 4, respectively, indicating the need for early detection and new treatments (Cairns, 2012).  Renal cell carcinoma is a malignant cancer of the kidneys with numerous subtypes: clear cell, collecting duct, chromophobe, mucinous tubular, multicular, papillary, and renal medullary. The cancer is thought to be created from a diverse number of specialized cells located along the nephron, the basic structure of the kidney, and the type of cell from which the cancer mutation stems from classifies it into the aforementioned types. Only 2-4% of renal cell carcinoma in patients have hereditary causes, but mutation in the tumor suppressor genes - von Hippel-Lindau, fumarate hydratase, folliculin, and succinate dehydrogenase genes-  which control abnormal cell growth, has been found to increase the likelihood of a person to develop renal cell carcinoma (American Cancer Society, 2017). The treatment options for patients with renal cell carcinoma are usually nephrectomy surgeries in which doctors try to cut out the kidney tissue with cancer. The real danger comes from having to remove two kidneys if the cancer is too far progressed; once the kidneys are removed, a person would need dialysis or a kidney transplant to survive (National Cancer Institute, 2019). The use of diagnostics is very important in order to treat renal cell carcinoma quickly. Recently, the enzyme carbonic anhydrase-IX (CAIX) has been found as a new biomarker for renal cell carcinoma by undergoing “increased expression in hypoxic environments created by tumor cells” (Kistler, 2017). Through microarray experimentation, miR-144-3p was found to be significantly upregulated in plasma of clear cell renal cell carcinoma patients, especially in those of advantaged stages of the cancer; thus, it was determined to be the first known microRNA biomarker of renal cell carcinoma of the clear cell type (Lou et al., 2017). A microRNA biomarker for various subtypes of renal cell carcinoma has yet to be researched, which is why a new method in diagnostics is so vital to the field.

As stated before, a common biomarker for cancers and other diseases is microRNA: non-coding, tiny RNAS that target messenger RNAS that regulate gene expression. They are transcribed from DNA sequences and do not code for protein like other RNAs; instead, they engage with target RNAs to suppress expression, sometimes create over expression, and act as signaling molecules to mediate cell communications (O’Brien et al., 2018). The dysfunction of microRNAs is commonly found to be present in either the tissue, blood, urine, or cerebrospinal fluid of disease ridden patients (Moldovan et al., 2014). Specifically, oncogenes - highly mutated genes found in tumor cells that can cause cancer - can potentially be made by overexpression of specific microRNAs, and those which have tumor suppressor abilities may be downregulated in cancer (Paranjape, Slack, & Weidhaas, 2009). For instance, microarray analysis was performed on breast cancer tissues to determine microRNA expressions and a downregulation of microRNA- 141 was found to be associated with the cell migration and invasion strength of the breast cancer (Li et al., 2017). Additionally, the down regulation of microRNA of the micro -143 and micro-145 groups were found to commonly be downregulated in cervical cancer and colon cancer cells (Akao, Nakawaga, & Naoe, 2006). MicroRNAs’ strong association with many different cancers has been found to make them viable biomarkers for a variety of diseases, including cancer. There are many known microRNA biomarkers of cancers such as miR-126-3p, miR-182-5p, miR-183-5p, and miR-210-3p for non-small cell lung cancer, which are specifically intuitive for diagnosing patients in early stages (Wang & Peng, 2018). These microRNA biomarkers can also be used to distinguish stages of cancer, as two miRNA precursors, pri-miR-944 and pri-miR-3662, were found to be able to distinguish non-small cell lung cancer through stages 1 through 3 (Wang & Peng, 2018).

The ability of machine learning to verify or find microRNA biomarkers in datasets has been investigated in recent years. Machine learning is the process in which computers learn to discern patterns from data with either supervised, unsupervised, or reinforcement algorithms. According to which type of data is being sorted or tested, a variety of different algorithms can be used for machine learning purposes based on whether the researcher is looking for structures, using labeled data to determine classifications for unlabeled data, maximizing a value based on a predescribed set of rewards (Golden, 2001).  The use of such algorithms for data of either structures, such as magnetic resonance imaging of brains; genome sequences; and functions of gene expression, such as microRNAs, is called bioinformatics (Luscombe et al., 2001). Through looking at the biology of large datasets of plasma, blood, urine, and resonance images with machine learning, people have determined predictive tools for a variety of diseases across massive samples. With datasets of non-images, machine learning algorithms such as random forest - a series of decision based tree designs - and PicTAR - a series of identifications of microRNA targets - are the most effective (Yue et al., 2009). In the field of cancer research specifically, machine learning has been used to identify biomarkers - substances, genetic structures, or processes that are indicators of any biological process (Strimbell & Tavel, 2011)

The repetitive regulatory network patterns that exist in gene regulation microRNAs are involved in can be coded for and, thus, make machine learning applicable and very useful in determining microRNA biomarkers. Specifically, there are many advantages to using machine learning to find microRNA biomarkers. The algorithms of machine learning are able to discern patterns between subtypes of a cancer. Additionally, circulating micrornas -  abnormal levels of unique miRNAs which can be observed at an early stage, during progression, and after metastasis of cancers - are hard to determine experimentally but are useful in distinguishing stages of cancer and tracking its progression. The use of machine learning can more easily identify such microRNAs and account for more heterogeneity among samples while providing statistically sound evidence of microrna biomarkers (Wang & Peng, 2018).

Even though the field of microRNA biomarkers is relatively new, there has been previous research to show successful attempts at selecting microRNA biomarkers from cancer datasets using machine learning.  For instance, a study used statistical analysis of a dataset of cancer microRNAs and classified them with random forest and support vector machine algorithms (Rehman et al., 2019). It concluded that machine learning could validate twelve microRNA biomarkers of breast cancer and determine hsa-mir-10b, has-let-7c, and hsa-mir-145 have the greatest predictive abilities (Rehman et al., 2019). Additionally, 20,000 microRNAs were used to train a machine learning predictive tool for microRNA target genes in multiple cancers to a high accuracy (Pla et al., 2018).  This research illustrates the ability to comb through existing data of microRNAs with machine learning to determine which microRNA targets can be used as diagnostic biomarkers. Similar approaches, therefore, can be used to look at the microRNAs themselves which are present in specific cancers such as renal cell carcinoma. Lastly, using Gene expression data from the Genotype-Tissue Expression project and the French Sarcoma Group, a machine learning random forest algorithm was used to determine the microRNa targeting of the HMMR gene as a novel biomarker of poor prognosis for liver, pancreatic, and lung cancer (Van IJzendoorn et al., 2019). There is substantial evidence to support the effectiveness of machine learning algorithms, specifically random forest, in determining highly predictive microRNAs for cancers. The random forest is a series of decision trees that output the predictive ability or classification of each tree, which is extremely useful in sorting and classifying genetic information. Random forest can also measure the relative importance of each feature in the prediction very easily. It does this by placing the best attribute of all the datasets at the root, or beginning, of the decision trees. Then the training set of data is split into subsets based on a common value for a chosen attribute, weighting the trees based on the perceived value of the attribute to the researcher. By continuously and randomly doing this, a huge collection of decision trees should be created, from which – as a whole, huge collection of trees- the predictive abilities of a regression or classification can be determined. The algorithm’s assumptions are that the whole training set is considered the root, features are non-continuous, and values are given recursively. The PicTAR is a type of decision tree algorithm for the classification of microRNA targets, similar to random forest, and can also be used to classify the data of microRNA by choosing the microRNA target as the best attribute of the root of the decision tree. Overall, both of these algorithms are extremely useful in determining biomarkers from classification of datasets, created through coding with the python or R coding language (Pla et al., 2018).

There is evidence to suggest that what was found to be a diagnostic microRNA of cancer can also be determined to have significant impacts on the effectiveness of patients’ treatments. The presence of higher miR-155-5p in pancreatic ductal adenocarcinomas was shown to have a direct relationship between chemo resistance and resistance for gemcitabine treatment in a patient (Garajova et al., 2014). There is also evidence that the up or down regulation - the reduction or gaining of specific microRNAs - of microRNA biomarkers is not only directly responsible for the resistance of cancer cells to treatment but the cancer growth itself. For instance, miR-128 was associated with a subtype of leukemia and found to create DNA damage when overly expressed in cells (Seca et al., 2014). This is an example of a direct manipulation of the levels of a microRNA biomarker of a disease in live cells, which is evidence for up or down regulation of the biomarker microRNA being a contributing factor to the cancer growth. Li and Zhang in 2014 did numerous trials in which they attempted at overexpressing mir-145 in osteacarma cells which resulted in higher proliferation, so there is a feasible model of how to overexpress or under expressed microRNA in a cancer with its precursor - it’s parent microRNA.  Thus, it was hypothesized that modifying the level of differentially expressed microRNAs believed to be a biomarker for renal cell carcinoma could also act as therapy to slow down the cancer growth.

The new investigation can answer questions remaining in the previous experimentation of microRNA biomarkers for renal cell carcinoma such as what is an upregulated or downregulated microRNA biomarker, if there is one, across the clear cell, papillary, and chromophore subtypes of renal cell carcinoma. Clear cell, papillary, and chromophore are the subtypes being investigated because they are the most common. Additionally, the experiment will determine if the overexpression or under expression of a microRNA found to be an upregulated or downregulated biomarker for renal cell carcinoma has a negative effect on the growth of the cancerous cells. The new experiment will use machine learning algorithms such as PicTAR and random forest in python code to find statistically supported biomarkers of renal cell carcinoma from datasets of the dbDEMC 2.0 and National Center for Biotechnology Information (NCBI) databases: 338 pairs of healthy and cancerous tissue microRNA sequences from dbDEMC 2.0 and 302 pairs of healthy and cancerous tissue microRNA sequences from the NCBI. The databases are sufficient for a machine learning project as they hold a lot of information of the expression of the microRNAs and the microRNAs’ predicted target for the algorithms to sort through. From the data collected, a possible biomarker will be tested for existence in the lab with a CRL- 1932 cell line of renal cell carcinoma. Upon detection, a precursor will be used to overexpress or under express the microRNA biomarker, and the growth of the cancer cells will be recorded. If a statistically significant difference is found in the growth of the cancer cell, a conclusion can be created to state whether there is evidence to suggest microRNA biomarkers of renal cell carcinoma also have therapeutic abilities. Renal cell carcinoma has had no previous machine learning applications and this experiment’s potential evidence of possible microRNA biomarkers and a new method of therapeutics could have implications for the entire field of cancer research.

References

Akao, Y., Nakagawa, Y., & Naoe, T. (2006). MicroRNAs 143 and 145 as possible common

microRNAs in human cancers. *Spandidos Publications*, 16, 845-850. doi: 10.3892

American Cancer Society. (2019). Retrieved from https://www.cancer.org/ cancer/ kidney-cancer

/detection-diagnosis-staging/survival-rates.html

Cairns, P. (2012). Renal cell carcinoma*. Cancer Biomar*k, 9, 461- 673. Doi:10.3233

Garajova, I., Le Large, T., Frampton, A. Rolfo, C., Voortman, J., & Giovannetti, E. (2014).

Molecular Mechanisms Underlying the Role of MicroRNAs in the Chemoresistance of Pancreatic Cancer . *Biomedical Research International,* 11, 136-145. doi:10.1155

Golden, R. (2001). Supervised, unsupervised, and reinforcement algorithms. *Principles,*

*Advances and Industrial Application*s, 4, 125 -146. doi: 10.1533

Kistler, A. (2017). Biomarkers in Renal Cell Carcinoma: Evidence and Challenges. Retrieved

from https://www.cancertherapyadvisor.com/home/cancer-topics/renal-cell-carcinoma

/biomarkers -in-renal-cell-carcinoma-evidence-and-challenges/

Li, E., & Zhang, J. (2014). miR-145 inhibits osteosarcoma cells proliferation and invasion by

targeting ROCK1. *Tumor Biology*, 35, 7645- 7650. doi: 10.1007

Li, P., Xu, T., Zhou, X., Liao, L., & Pang, G. (2017). Downregulation of miRNA-141 in breast

cancer cells is associated with cell migration and invasion: involvement of ANP32E targeting. *Cancer Medicin*e, 6, 662-672. doi: 10.1002

Lou, N., Ruan, A.M, & Qui, B. (2017). miR-144-3p as a novel plasma diagnostic biomarker for

clear cell renal cell carcinoma. *Urol Oncol*, 109, 36-37. doi: 10.1016

Luscombe, N., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed

definition and overview of the field. *Methods Informations Medicine*, 40, 346-358. Doi: 10.2348

Moldovan, L., Batte K., Trgovcich, J., Wisler, J., Marsh, C., & Piper, M. (2014). J Cell

Molecular Medicine, 18, 371-390. doi: 10.1111

National Cancer Institute. (2019). Renal Cell Carcinoma Treatment. Retrieved from

https://www.cancer.gov /types/kidney/patient/kidney-treatment-pdq

O’Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis,

Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*, 9, 402-410. doi: 10.3389

Paranjape, T., Slack, F., & Weidhaas, J. (2009). MicroRNAS: tools for cancer diagnostics.

*British Society of Gastrology*, 58, 1546-1554. Doi: 10.1136

Pla, A., Zhong, X., & Rayner, S. (2018). mirRAW: A deep learning-based approach to predict

microRNA targets by analyzing whole microRNA transcripts. *Public Library of Science: Computational Biology*, 14, 7-10. doi: 10.1371

Rehman, O., Zhuang, H., Ali, A., Ibrahim, A., & Li, Z. (2019). Validation of miRNAs as Breast

Cancer Biomarkers with a Machine Learning Approach. *Cancers*, 11, 431 - 441. doi: 10.3390

Seca, H., Lima, R., Almeida, G., Sobrinho-Simões, M., Bergantim, R., Guimaraes, J.,

Vasconcelos, M. (2014). Effect of miR-128 in DNA damage of HL-60 acute myeloid leukemia cells. *Curr Pharm Biotechnology*, 15, 492 -502. doi: 10.2484

Strimbu, K. & Tavel, J. (2011). What are Biomarkers? *Current Opin HIV AIDs*, 5, 463 - 466.

doi:10.1097

Van IJzendoorn, D., Szuhai, K., Briare-de Bruijn, I., Kostine, M., Kuijjer, M., & Bovee, J.  
 (2019). Machine learning analysis of gene expression data reveals novel diagnostic and

prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *Public*

*Library of Science: Computational Biology*, doi: 10.1371

Wang,  Z., Michalski, S. G., , Durka, W. (2018). miR-128 effect on the DNA damage of

leukemia. *Journal of Public Health*, 10, 213-221. doi:10.1128

Wang, H. & Peng, R. (2018). Circulating microRNAs as potential cancer biomarkers: the

advantage and disadvantage*. Clinical Epigenetics*, 10, 59 -61. doi: 10.1186

Yue, D., Liu, H., & Huang, Y. (2009). Survey of Computational Algorithms for MicroRNA

Target Prediction. *Current Genomics*, 10, 478 - 492. doi: 10.2174