*Using Machine Learning to Identify Biomarkers with the Aid of MicroRNAs in Renal Cell Carcinoma*

Hanika Pandya

May 17th, 2019

With a mortality rate of 92% in its final stage, Renal Cell Carcinoma is a cancer that begins its formation in cells within the kidney (Hu et al., 2016). An estimated 73,820 people fall victim to this cancer each year in the United States alone, alongside an approximate 14,700 yearly deaths (American Cancer Society, 2019). Regardless of the high death rates, the only accepted biomarker for Renal Cell Carcinoma is miR-144-3p as this supposedly heightened the ability for these cancerous cells to rapidly divide and spread (Benej et al., 2015)., prompting the necessity to not only detect the cancer, but treat it as well (Hu et al., 2016). Renal Cell Carcinoma comes in the form of numerous subtypes, the most common being clear cell, papillary, and chromophobe; as these make up roughly 80% of all Renal Cell Carcinoma cases (American Cancer Society, 2019). Renal Cell is believed to be caused when malignant cancer cells line the small tubes within the kidney, this causes disruption in the body’s ability to clean blood and produce urine (National Cancer Institute, 2019). The majority of people diagnosed with Renal Cell Carcinoma are not due to hereditary reasons as this only constitutes for 2.3-4.5% of patients (American Cancer Society, 2019), this causes genes (fumarate hydratase, succinate dehydrogenase, and von Hippel-Lindau) who’s function it is to control the growth of cells are suppressed, leading to an abnormal and irregular cell growth rate (National Cancer Institute, 2019). Due to its rampant pace, the need to be able to identify a biomarker to aid with the diagnosis of Renal Cell Carcinoma is essential and urgent. While there is one biomarker known, there is none across varying subtypes of Renal Cell Carcinoma, and the usage of microRNA is key to identifying the new biomarker.

MicroRNAs are miniscule non-coding RNA molecules that play a key role in the regulation of gene expressions (MacFarlane & Murphy, 2015). This in turns mean that they are able to either down regulate or upregulate the ability for a cell to divide (Gulyaeva & Kushlinskiy, 2016). Downregulation is the classified as the decrease in microRNAs and vice versa in upregulation. These irregularities within microRNAs are not limited to Renal Cell Carcinoma, but across multiple different cancer types (Reddy, 2015). When genes are upregulated or overexpressed, they are classified as oncogenes, which can only decrease when a suppressing downregulated gene is intertwined. (Reddy, 2015). Over 699 microRNA expressions were analyzed in bladder urothelial carcinoma to understand the regulation of the genes. (Han et al., 2017). It was seen that has-miR-200b and has-miR-200c played a dominating role in the creation of abnormal cells (Han et al., 2017). The use of microRNA is still fairly advanced, as it began to be considered during the 2010s. Previously methods of biomarker identification were often insufficient, and now with the aid or microRNA new biomarkers are being found. For example, in Adenocarcinoma, there were no known biomarkers previous to the use of microRNA (Calvayrac et al., 2017) Once used, miR-210-3p along with others have been identified and used as early detection and diagnostic methods (Inamura & Ishikawa, 2016). This same foundation can be used to discover biomarkers for Renal Cell Carcinoma.

While the only known Renal Cell Carcinoma biomarker was found in the lab, this raises many questions. There is room for many errors as one lab result does not take irregularities and other variables into account. A more reliable method to identify biomarkers using microRNA expressions is the use of machine learning alongside appropriate database. Machine learning the way in which intelligence learns through a series of algorithms designed to complete a task often with a large number of variables and data (Brookings Education, 2018). In specifics, microRNA learning alongside machine learning is characterized as bioinformatics. The choice of algorithm is dependent on whether the data needs to be tested, classified, sorted or so on (Brookings Education, 2018).­ Decision tree algorithms such as Random Forest and Identifying PicTar have previously been used in relation to microRNA, and tend to provide better results (Xu et al., 2016). Machine learning is beneficial compared to lab experimentation as it is not only able to take numerous amounts of data into account, it adjusts to irregularities, therefore making the results more applicable to real life (Cornell University, 2015). The ability to adjust machine learning to take variables into account is essential as variations can be accounted for. Algorithms are an essential advantage as they are able to easily sort through data, narrow down information, and categorize across different subtypes of cancer. Additionally, they are able to track cancers that are challenging to do so in the experimentational setting. This provides more insight into the progression of the cancer as there is the ability to observe these unique miRNAs (Wang et al., 2018). In terms of databases, the more information the machine has, it is able to more accurately learn and give results.

Apart from success in using microRNAs as biomarkers, machine learning has been able to identify biomarkers for other cancer types. Within the algorithm miRAW, over 34,000 data pieces taught a machine to predict target microRNA genes across multiple cancers (Pla et al., 2018). A study used Random Forest to classify microRNAs in breast cancer to find biomarkers has-let7c, has-mir-145, and has-mir-10b (Rehman et al., 2019). These demonstrate the ability for an investigation that combines both the success of machine learning and microRNA to yield results.

Apart for microRNAs ability to prove useful in terms of diagnostics, theories have suggested that microRNAs can act as a therapeutic aspect in terms of treatment for the disease. It was seen that in MiR-33a, a precursor for metabolic diseases, the down regulation of this target gene decreased the spread of the cells, and restricted current cancer cells (Christopher et al., 2016). Through altering the regulation of target gene sequence there is a possibility that the same microRNA could not only act as a biomarker, but a therapeutic gene as well.

Previous studies showcase how microRNAs are able to locate biomarkers for disease previously unknown. However, the aspects of focusing on subtypes within Renal Cell Carcinoma along with therapeutic opportunities differentiate this investigation. By understanding the regulation of the gene expression in microRNAs of clear cell, papillary, and chromophore subtypes, the goal is to find a biomarker never discovered before. The use of machine learning contrasts the usual use of experimental trial in the lab setting, as machine learning has a more reliable set up. Furthermore, Renal Cell Carcinoma has yet to be tested using machine learning. The new investigation requires using algorithms such as Random Forest Generator and PicTar to identify the biomarkers across Renal Cell Carcinoma subtypes. The data for these algorithms would be collected by dbDEMC 2.0 along with the National Center for Biotechnology Information, a total of over 700 data. As the machine begins to learn, more data will be add to allow for a better analysis. Once a possible biomarker is identified, its credibility will be tested in a lab setting to ensure that it is true in both machine learning and experimental realms. Using the Renal Cell Carcinoma cell line CRL-1932, the biomarker will be assessed to see if it truly has the ability to act as a diagnostic for the series of Renal Cell Carcinoma subtypes. The expression of this biomarker will be altered (over and under) to observe if this causes any change in the rate of cancer cell growth in terms of therapy. Not only can the results of this investigation, if correct, alter the way Renal Cell Carcinoma patients are diagnosed; it opens the possibility for microRNA therapeutics across numerous cancer types, making the impact on a much larger scale.

References:

American Cancer Society (2019). *Key Statistics About Kidney Cancer*. Retrieved from https://www.cancer.org/cancer/kidney-cancer/about/key-statistics.html

Benej, M., Pastorekova, S., Pastorek, J. (2015), Carbonichydrase IX: Regulation and Role in Cancer. *Subcell Biochem, 75*, 199-219. doi: 10.1007/978-94-007-7359-2\_11

Brookings Education (2018). *What is Machine Learning*. Retrieved from https://www.brookings.edu/research/what-is-machine-learning/

Calvayrac, O., Pradines, A., Pons, E., Mazieres, J. & Guibert, N. (2017), Molecular Biomarkers for Lung Adenocarcinoma. *Eur Respir J, 5,* 49-53. Doi: 1 0.1183/13993003.01734-2016

Christopher, F.A, Kaur, R.P, …, Bansal, P. (2016). MicroRNA Therapeutics: Discovering Novel Targets and Developing Specific Therapy. *Perspect Clin Res, 7*, 68-74. doi: 10.4103/2229-3485.179431

Cornell University (2015). *A Review of Machine Learning Based Anomaly Detection Techniques*. Retrieved from https://arxiv.org/abs/1307.7286

Gulyaeva, L. F. & Kushlinskiy, N. E. (2016), Regulatory Mechanisms of microRNA Expression. *J Transi Med, 14*, 143-147. doi: 10.1186/s12967-016-0893-x

Han, Y., Chen, J.,…, Cai, Z. (2017), MicroRNA Expression Signatures of Bladder Cancer Revealed by Deep Sequencing. *Plos One, 10*, 186-204. doi: [10.1371/journal.pone.0018286](https://doi.org/10.1371/journal.pone.0018286)

Hu, S.L, Chang, A.,…, Weiss, R.H. (2016). The Nephrologist’s Tumor: Basic Biology and Management of Renal Cell Carcinoma. *J Am Soc Nephorol, 27*, 2227-2237. doi: 10.1681/ASN.2015121335

Inamura, K. & Ishikawa, Y. (2016), MicroRNA in Lung Cancer: Novel Biomarkers and Potential Tools for Treatment. *J Clin Med, 5*, 36-39. doi: 10.3390/jcm5030036

MacFarlane, L. & Murphy, P. R. (2015), MicroRNA: Biogenesis, Function and Role in Cancer. *Curr Genomics, 11*, 537-561. doi: 10.2174/138920210793175895

National Cancer Institute (2019).  *Renal Cell Cancer Treatment*. Retrieved from https://www.cancer.gov/types/kidney/patient/kidney-treatment-pdq

Pla, A., Zhong, X. & Rayner, S. (2018). miRAW: A Deep Learning-Based Approach to Predict microRNA Targets by Analyzing Whole microRNA Transcripts. *PLos Comput Biol, 14,* 637-677. doi: 10.1371/journal.pcbi.1006185

Reddy, K. B. (2015), MicroRNA (miRNA) in Cancer. *Cancer Cell International, 15*, 382-3994. doi: 10.1186/s12935-015-0185-1

Rehman, O., Zhuang, H., Muhamed., A, Ibrahim, A.. & Li, Z. (2019), Validation of miRNAs as Breast Cancer Biomarkers with a Machine Learning Approach. *Basel Cancers, 11*, 431. doi: 10.3390/cancers11030431

Wang, Z., Michalski S.G.,…, Durka, W. (2018). miR-128 effect on DNA damage of leukemia. *Journal of Public Health, 10*, 213-221. Doi:10.1128%2FJVLI.02386-12

Xu, W., Lucas, A.S., Wang, Z. & Liu, Y. (2016). Identifying microRNA Targets in Different Gene Regions. *BMC Bioinfromatics, 7*, 275- 312. doi: 10.1186/1471-2105-15-S7-S4