Identifying miRNA targeting Prostate Cancer Gene in White Latino Populations

Hafidzhah Muhammad Aldino and Agustriawa David*
Bioinformatics Department, Indonesia International Institute for Life Science (i3L), East Jakarta, INDONESIA
*david.agustriawan@i3l.ac.id

Abstract
Cancer is one of the top 10 leading causes of death in the world, according to the worldwide cancer data. Cases throughout 2018 reached 1.8 million, making the disease a dangerous threat in the 21st century. Cancer is a disease that causes uncontrolled growth of cell tissues and organs such as the brain, breast, liver and prostate glands. Cancer can grow in almost all cell tissues in the human body including the prostate gland of a man. Prostate cancer is a carcinoma that occurs in the prostate gland of male individuals.

Researchers have found a way to inhibit cancer growth, especially prostate cancer which is by identifying a specific miRNA that might act as a tumor suppressor. miRNA is a short noncoding RNA that regulates the expression of a particular gene and is well conserved between different organisms. In this study, we try to identify several miRNAs and their specific target genes with the help of the pipeline that has been prepared. The statistical analysis result shows that hsa-miR-1286, hsa-miR-1294 andhsa-miR-3136 have the potential to be tumor suppressor genes.

Keywords: Cancer gene, Epigenetics, micro-RNA, Prostate cancer.

Introduction

Cancer is one of the top 10 leading causes of death in the world according to the worldwide cancer data. Cancer cases throughout 2018 reached 18 million, making the disease a dangerous threat in the 21st Century. In medical terms, cancer is a disease that causes uncontrolled growth and division of cells inside the body. Cancer can grow in almost all cell tissues and organs such as the brain, breast, liver, lungs and prostate glands.

In this study, we mostly address prostate cancer, which is one of the top five most common deadly cancer types. Statistics from WHO's worldwide cancer data for 2018 show that prostate cancer is the fourth most common deadly cancer. Prostate cancer is a carcinoma that occurs in the prostate gland of a male individual and if its treatment is not suitable or correct, it may lead to death.

A lot of research has been done to find treatments for cancer and scientists have discovered that miRNA might help in cancer treatment, especially prostate cancer. miRNA is a short noncoding RNA that regulates the expressions of a gene and is well conserved between different organisms[5]. It was identified in 1993 and since then, numerous studies have been conducted to show miRNAs critical role in regulating gene expression and also cellular processes[7]. miRNAs might act as tumor suppressors or oncogenes for any cancer types including prostate cancer[9]. It is important to understand what kind of interaction between the gene and miRNA can cause the miRNA to act as a tumor suppressor or oncogene?

When miRNAs are up-regulated, they will inhibit the growth of the cancer genes, for example, the prostate cancer genes of prostate adenocarcinoma (PRAD); these miRNAs that cause the cancer genes to be down-regulated are referred to as tumor-suppressing genes, while the miRNAs that when down-regulated help in the up-regulation of the cancer genes production are referred to oncogenes[2].

Several researchers have conducted numerous studies on miRNA gene interaction[1] compared miRNAs expressed in the prostate cancer of African-American and European-American men[4,8] examined the miRNA expressed in African-American and Caucasian men. Therefore, this study aims to observe whether miRNAs regulate the target genes that have been identified previously in a specific population (White, non-Hispanic, or Latino).

Material and Methods

Programming languages R and Python 3.6 (Pycharm) are the main tools used in this research as they are suitable for conducting statistical analyses. Here 'R' was used alongside with the TCGA-Assembler to help in retrieving The Cancer Genome Atlas (TCGA) data using TCGA barcode. The TCGA-Assembler is an open-source package for retrieving, assembling and processing public TCGA data and it contains the R-package[11].

The TCGA-Assembler allows for automatic download and integration of the data from the GDC Data Port al which makes the data retrieval much more efficient compared to manually downloading the TCGA data. Moreover, Python was used alongside its Pandas module for data science. The flowchart in figure 1 shows the complete steps that were used in this study.

Data Retrieval: In this study, the dataset is first retrieved. The metadata of a prostate cancer patient was used; it was acquired from the National Institute of Health Genomic Data Commons (NIH GDC) Portal), specifically from the repository data. For each category of files and cases, some indicators were chosen to filter the metadata that will be obtained later. The indicators that were chosen for the

repository data files include the transcriptome profiling (data category), miRNA expression quantification (data type), miRNA-Seq (experimental strategy) and the indicators that were chosen for cases in the repository data include the prostate gland (primary site), TCGA (program), TCGA PRAD (project), male (gender), White and non-Hispanic or Latino (race).

Data Processing: To start processing the data, we prepared the downloaded Python 3.6 (Pycharm) and also the Pandas module that has been installed in Python. They were used to obtain the TCGA barcode from the metadata file. Pycharm is an integrated development environment that is used in computer programming, specifically in the Python programming language and it also provides a lot of features, such as code analysis and debugging. The TCGA barcode was then used to download the genes and miRNAs data using the TCGA-Assembler package. Pandas data frame in Python will help to separate the normal and cancer data on both genes and miRNAs datasets.

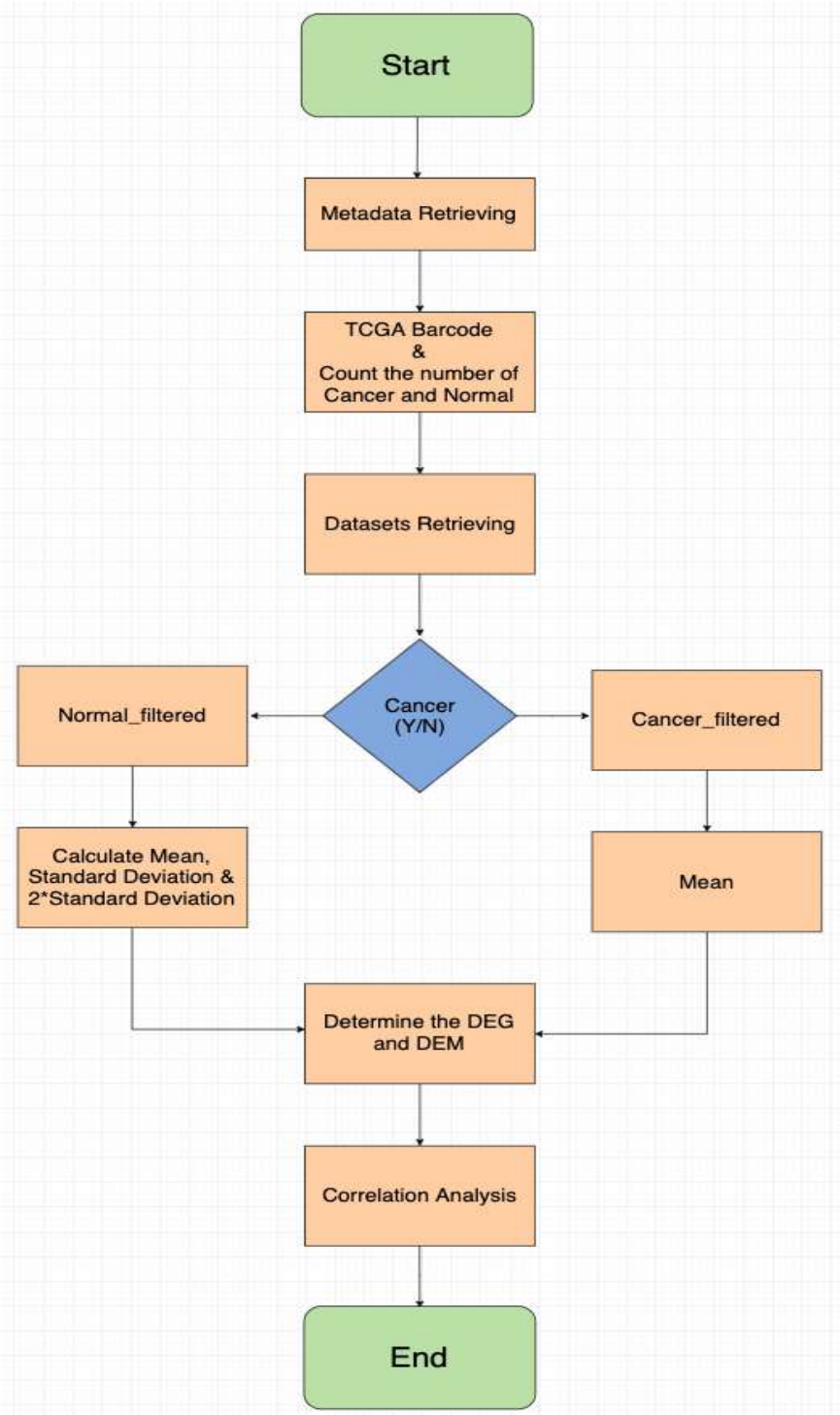

Figure 1: Shows how the process from gathering the data until the correlation analysis which involving several tools such as R, Python, Matlab and Microsoft Excel

Determination of The Differentially Expressed Genes and MiRNA: Here, we used Microsoft Excel and the Pandas module in Python to determine the up-regulated and down-regulated gene (cancer) and miRNA. The Pandas module was used to calculate the mean, standard deviation and2*standard deviation. The mean and standard deviation of normal genes were calculated and compared with the mean of cancer genes. Using Microsoft Excel, all Differentially Expressed Genes (DEG) and differentially expressed miRNAs (DEM) were determined using the formulas below:

1. Up-regulated formula:
Mean (Cancer) > [normal mean + (2 × normal standard deviation)]

2. Down-regulated formula:
Mean (Cancer) < [normal mean − (2× normal standard deviation)]

Once the results were acquired using these two formulas, we could finally filter the cancer and miRNA datasets into up-regulated or down-regulated datasets, which were separated into two different files.

Correlation Analysis: MATLAB was used for the correlation analysis. Using MATLAB, we aim to conduct the negative correlation test to determine the miRNAs targeting specific genes. Differentially expressed genes and miRNAs have been already identified in the previous step. Down-regulated genes and up-regulated miRNAs were selected based on the assumption that up-regulated miRNAs act as tumor suppressors for target genes and therefore, the genes were down-regulated. A correlation analysis was conducted for both the down-regulated and up-regulated miRNAs to further confirm the result.

The first method of correlation analysis used file cases that were selected and metadata files were downloaded from NIH GDC Portal (depending on the actual data); in this case, we used Spearman's rank correlation. The correlation analysis result was sorted to find the miRNA and its specific target genes.

Results and Discussion

All processes from metadata retrieving to the correlation analysis both for DEG and DEM were conducted. We acquired the up-regulated and down-regulated genes, both for the cancer gene and miRNA. From the datasets, we acquired 12,521 up-regulated cancers and 7,981 down-regulated cancers for the cancer gene.

Meanwhile, for the miRNA, we acquired 836 up-regulated miRNAs and 1,036 down-regulated miRNA; the details are presented in table 1. A correlation test was conducted to find the most significant miRNAs targeting cancer genes and vice versa. It was conducted in MATLAB version-R2018A. The correlation test was conducted between up-regulated miRNAs down-regulated genes and vice versa. To determine which miRNA-gene interactions are statistically significant, we used the parameters Rho and P-Value. If the Rho < -0.5 and the P-Value < 0.05, then it is statistically significant and if the Rho > -0.5 and the P-Value > 0.05, it is not statistically significant.

The correlation analysis result can be used to help determine which miRNA has the potential to be tumor suppressor genes and also which miRNA has the potential to be oncogenes. The result from the correlation analysis is presented in tables 2a. and 2b. and it shows the top three miRNAs that can be tumor suppressor genes or oncogenes.

Table 2a presents the top three potential candidates for tumor suppressor genes, which are hsa-miR-1286, hsa-miR-1294 and hsa-miR-3136. Meanwhile, in table 2b, the top three potential candidates for oncogenes are hsa-miR-1271, hsa-miR-126 and hsa-miR-106a. The results shown here were validated using miRTarBase, miRdb and Target Scan. These databases were used to experimentally validate miRNA-target interactions, predictions and also functional annotation[3,6,10].

Table 1
Shows the Compositions of DEG and DEM

Indicator	Genes	miRNAs
Upregulated	12,521	836
Downregulated	7,981	1,036
Total	20,502	1,872

Table 2a
Top 3 miRNA - Gene Interaction (miRNA Upregulated and Gene Downregulated)

Gene_ID	miRNA_ID	P-Value	Rho Value
MAT1A	hsa-miR-1286	0.0139575	-0.2051866
ADAMTS6	hsa-miR-1294	0.01635994	-0.2004806
ZEB2	hsa-miR-3136	0.00969451	-0.2156389

Table 2b
Top 3 miRNA - Gene Interaction (miRNA Downregulated and Gene Upregulated)

Gene_ID	miRNA_ID	P-Value	Rho Value
KRAS	hsa-miR-1271	0.007737655	-0.207388827
ZNF148	hsa-miR-126	0.012802899	-0.212195542
RNF2	hsa-miR-106a	0.003237179	-0.239282196

Conclusion

In-silico study of miRNAs and cancer genes helps in identifying several potential miRNAs that might act as a tumor suppressor in the prostate cancer gene of PRAD including hsa-mir-1286, hsa-mir-1294 and hsa-mir-3136. The data used in this study were limited to White non-Hispanic population only due to the insufficient amount of data from different races.

Moreover, some unknown pairs were discovered and have not been experimentally identified in the miRTarBase, Target Scan and miRdb. Therefore, the validation of the unknown pairs is required in further studies and the validation of the pairs might be possible using the RNA Hybrid and other available validation tools.

Acknowledgement

The authors would like to give a lot of thanks to the School of Bioinformatics and Center of Student Development Department of Indonesia International Institute for Life Sciences (i3L) for funding this research.

References

1. Amankwah E.K., MicroRNAs differentially expressed in prostate cancer of African-American and European-American men, *Cancer Translational Medicine*, **4(1)**, 28–34 **(2018)**

2. Cannell I.G., Kong Y.W. and Bushell M., How do microRNAs regulate gene expression?, *Biochemical Society Transactions*, **36(6)**, 1224–1231 **(2008)**

3. Chou C., Shrestha S., Yang C., Chang N., Lin Y., Liao K., Huang W.C., Sun T.H., Tu S.J., Lee W.H. and Chiew M.Y., MiRTarBase update 2018: A resource for experimentally validated microRNA-target interactions, *Nucleic Acids Research*, **46(D1)**, D296–D302 **(2017)**

4. Karakas C., Wang C., Deng F., Huang H., Wang D. and Lee P., Molecular mechanisms involving prostate cancer racial disparity, *American Journal of Clinical and Experimental Urology*, **5(3)**, 34–48 **(2017)**

5. Kehl T., Backes C., Kern F., Fehlmann T., Ludwig N., Meese E. and Keller A., About miRNAs, miRNA seeds, target genes and target pathways, *Oncotarget*, **8(63)**, 107167 **(2017)**

6. Riffo-Campos Á.L., Riquelme I. and Brebi-Mieville P., Tools for Sequence-Based miRNA Target Prediction: What to Choose?, *International Journal of Molecular Sciences*, **17(12)**, 1987 **(2016)**

7. Tan W., Liu B., Qu S., Liang G., Luo W. and Gong C., MicroRNAs and cancer: Key paradigms in molecular therapy, *Oncology Letters*, **15(3)**, 2735–2742 **(2017)**

8. Theodore S.C., Davis M., Zhao F., Wang H., Chen D., Rhim J., Dean-Colomb W., Turner T., Ji W., Zeng G. and Grizzle W., MicroRNA profiling of novel African American and Caucasian Prostate Cancer cell lines reveals a reciprocal regulatory relationship of miR-152and DNA methyltransferase 1, *Oncotarget*, **5(11)**, 3512 **(2014)**

9. Wahid F., Shehzad A., Khan T. and Kim Y.Y., MicroRNAs: Synthesis, mechanism, function andrecent clinical trials, *Biochimica et Biophysica Acta (BBA) -Molecular Cell Research*, **1803(11)**, 1231–1243 **(2010)**

10. Wong N. and Wang X., miRDB: an online resource for microRNA target prediction and functional annotations, *Nucleic Acids Research*, **43(D1)**, D146–D152 **(2015)**

11. Zhu Y., Qiu P. and Ji Y., TCGA-assembler: open-source software for retrieving and processing TCGA data, *Nature Methods*, **11(6)**, 599–600 **(2014)**.

(Received 10th August 2020, accepted 15th October 2020)