Research Inventy: International Journal of Engineering And Science
Issn (e): 2278-4721, Issn (p):2319-6483, PP: 48-53
www.researchinventy.com

An in-silico transcriptomic study to highlight the prevalence of the miR23b in *BCL2L1* gene of oral cancer

Agnik Haldar[1], Ajay Kumar Singh[1*]
[1]*Department of Bioinformatics, Central University of South Bihar, Gaya, Bihar*
**Corresponding Author*

Abstract
Oral cancer is considered one of the most prevalent cancers affecting the globe. Often the high usage of tobacco, poor hygiene, and viral infections are usually the cause of oral cancer. Upon delving into the epigenetic aspects of the disease, we can also deduce that several epigenetic processes like chromatin remodeling and histone modification cause dysregulation of genes. Dysregulation of genes can be explained in a specific manner as to when the genes tend to behave abnormally and causes a chain of events that causes an anomaly in the biological system. These anomalies also include the effect of non-coding RNA (ncRNA) on the genes. ncRNAs are non-protein coding RNAs, i.e. they do not code for proteins. However, they have been reported to act as epigenetic modifiers which ultimately play a role in the dysregulation of genes. There have been numerous reports over the years which have identified the dysregulation of genes as one of the major causes at play in the mechanisms of cancer and the involvement of ncRNAs like microRNA (miRNA) and long non-coding RNA (lncRNA) cannot be ignored. Here in this article, we have aimed to identify and understand the prevalence of the miRNA, miR23b in oral cancer and how they affect the BCL2L1 gene in correlation to oral cancer.

Keywords: miRNA, BCL2L1, Transcriptomic analysis, NGS Data analysis, Epigenetics, non-coding RNA

I. Introduction

miR-23b has been known to be associated with a number of cancer hallmarks like cell apoptosis, cell invasion and cell proliferation (Hao & Yu, 2017). For years it has been used as a biomarker in various cancers. However, there seems to be little evidence of its association with oral cancer and the *BCL2L1* gene. The *BCL2* family of genes have a long proven history of being associated with apoptosis. Since its discovery, the *BCL2* family of genes have been thoroughly divided into 3 categories, anti-apoptotic, BH3-only (pro-apoptotic), and pore-forming or 'executioner' (pro-apoptotic) proteins (Levine et al., 2008). However, there still exists a gap in our knowledge regarding the function of *BCL*-2 family isoforms. In our research study after an extensive transcriptomic data analysis, we were able to find correlation between the *BCL2L1* gene, which is an isoform of the *BCL2* family and oral cancer. Upon further investigation we were able to correlate the prevalence of the miR23b on the *BCL2L1* gene. Thus leading us to speculate that the miRNA hsa-miR23b-3p targets the *BCL2L1* gene, which is a confirmed anti-apoptotic target for oral cancer.

With the advent of next-generation sequencing brought along with it the means to analyze gene expression and correlate their prevalence with several diseases, more specifically cancer. Before its invention, basic gene expression analysis techniques did exist, however for rudimentary purposes. Alizadeh et al in 2000 (Alizadeh et al., 2000) documented one of the first major studies to use gene expression analysis in the diagnosis of diffuse large B-cell lymphoma, thus making it one of the first studies to use transcriptomics to diagnose cancer.

Subsequent studies have thereafter been done profiling various gene expression studies involving protein-coding as well as non-protein-coding genes. Over the years, researchers considered protein-coding genes to be useful data and non-protein-coding genes were treated as junk data, due to their inability to code for proteins. However, it was later discovered that they contain invaluable information related to various regulations and pathways involved in a variety of diseases, one of the most notorious being cancer. In oral cancer specifically, numerous studies have been done concerning noncoding genes and their implications for the disease. The invention of NGS methodologies has certainly made it feasible to perform such analysis without significant problems. The first generation of NGS technologies involved Sanger sequencing, which was developed in 1977, based on the principle of the chain termination method or the dideoxynucleotide method. The second generation of NGS included the microarray and RNA sequencing techniques, both of which were considered significant advancements over its predecessor and slowly rendered Sanger sequencing obsolete barring a few specific exception cases. However, we are now entering the age of third-generation

sequencing, which is still in its infancy, but it is on its way to toppling its predecessors (Heather & Chain, 2016; Slatko et al., 2018).

Here we used RNA sequencing to find the expressed genes in oral cancer data-sets which we will be using to analyze the pathways of those expressed genes. Further we were able to correlate the *BCL2L1* gene with miR23b and for a miRNA-gene network to portray the apoptotic nature of the gene prevalent in oral cancer.

II. Materials and Methods

Data Collection

The requisite files, namely the FASTQ files and the GTF files, for the RNA seq data analysis were downloaded from NCBI's Gene Expression Omnibus (GEO). The sequencing data deposited in the NCBI GEO, consisting of oral cancer tissues of 30 patients, were downloaded and stored in a separate folder. The respective files were extracted from their zip files and the zip files were kept as backup.

Alignment

The reads were then aligned with the help of a splice-aware aligner HISAT2 (Kim et al., 2015) with the hg19 reference genome. We were able to achieve a >90% alignment rate for all the samples. Of the 53 million paired-end reads per sample, approximately, 50 million reads were aligned successfully to the hg19 reference genome. Low and inconsistent reads were subsequently removed.

Data Analysis

Finally, for differential gene expression data analysis, we used the DESeq2 tool (Love et al., 2014) and we were able to report that the gingivobuccal cancer datasets yielded 32869 differentially expressed genes. Wald Test was employed for statistical analysis which is a way to find out if the explanatory variable in a model is significant.Excluding the outliers, low counts, and keeping an adjusted p-value threshold of < 0.001, we sorted out the dysregulated genes based on our threshold value of < 0.0001 and obtained 1351 significant dysregulated genes. Dysregulation is explained as the involvement of the genes in the disruption of normal pathways. The dysregulated genes provide us with an incentive to further explore the causality and their involvement in the disease.

Constructing a miRNA-gene regulatory network

We aimed to highlight the prevalence of the miR23b, which is a known tumor suppressor (Fukumoto et al., 2016), and its incidence on the *BCL2L1* gene. We employed the usage of the miRWalk database (Sticht et al., 2018) to find correlation and binding site location of the miR23b and *BCL2L1* gene. We had also proven its incidence in our recently published paper where we were able to correlate the anti-apoptotic properties of the *BCL2L1* gene which came into play after binding with the miR23b miRNA. We also analysed the relation between the miRNA and the gene in the miRNet database (Fan et al., 2016) and we generated a miRNA-gene regulatory network.

III. Discussion

Transcriptomic analysis

The dysregulated genes which were obtained after comparing the normal dataset with the tumor dataset yielded data which can be used to further identify and explore the relational nature of the genes to the disease. The dysregulated genes serve as an indication of the disruption of normal genomic processes. This disruption can be traced back to the source of the disease at hand and thus can be further analyzed and be used as a biomarker or subsequent target for drug identification. For identification and analysis of dysregulated genes, comparison between the expression in normal and affected tissues will confirm the involvement of those genes in the disease or the lack thereof. Similarly the dysregulated genes will help us understand their relevance in tumor growth and maintenance.

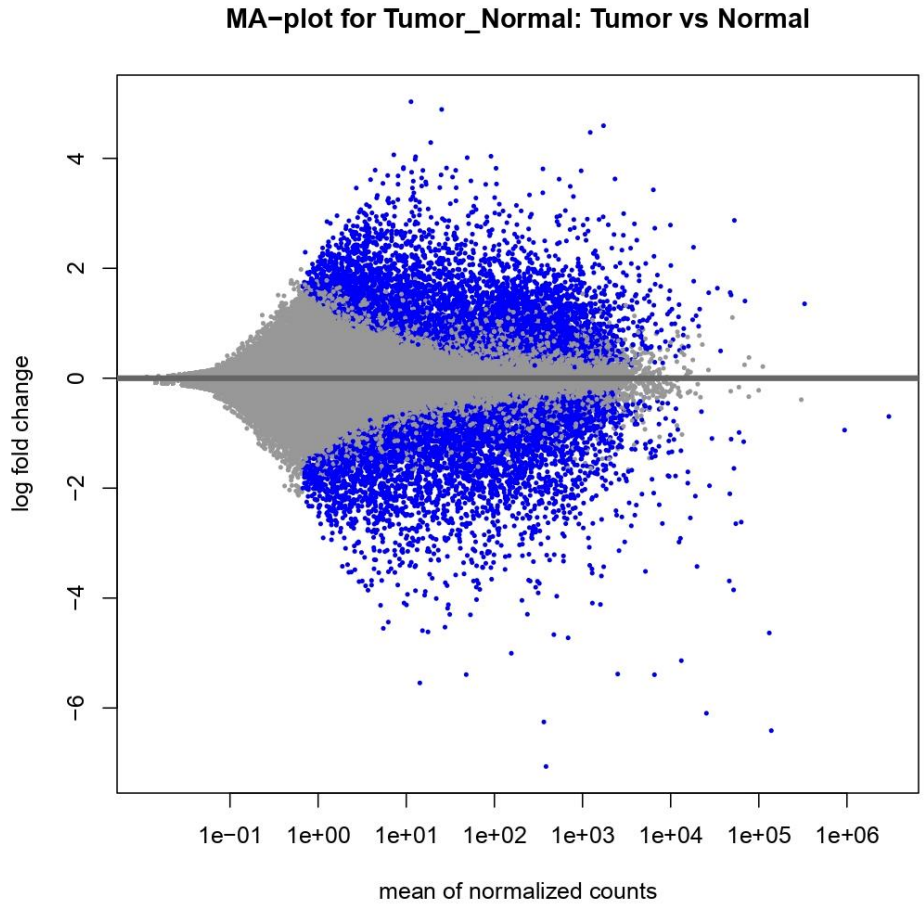

Fig 1: Graph representing DE genes (red points) against mean expression levels. The blue points indicate genes with p-adj (FDR) < 0.05

We were able to analyze 32869 differentially expressed genes from the data sets. Out of the 32869 dysregulated genes we excluded the outliers, low counts, and kept an adjusted p-value threshold of < 0.001.

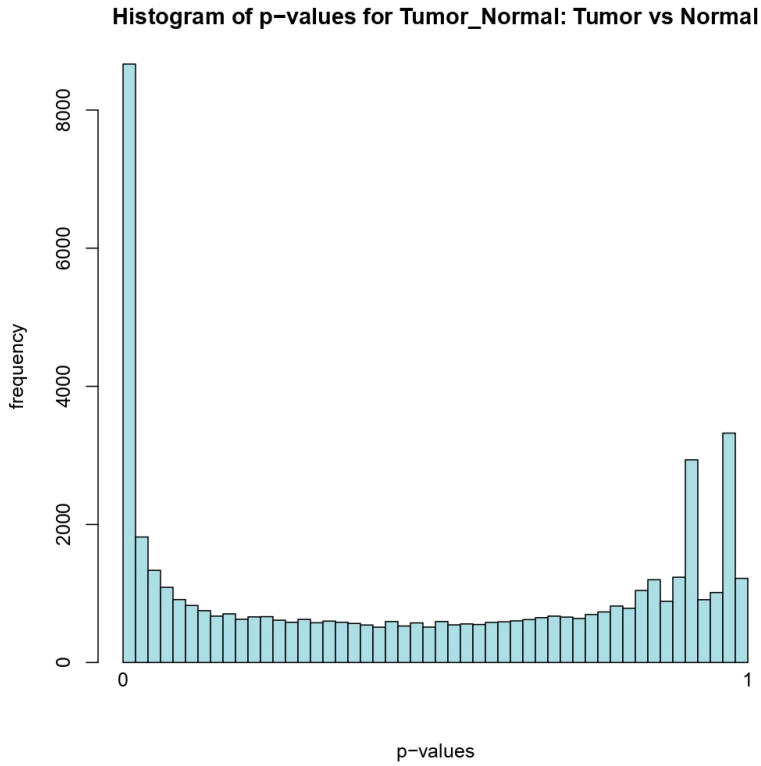

Fig 2: Histogram representing the p-values of the genes for Tumor vs Normal datasets

As a result we were able to sort out the dysregulated genes based on our threshold value of < 0.0001 and obtained 1351 significant dysregulated genes. Among those 1351 highly dysregulated genes, we were able to report the presence of the *BCL2L1* gene.

***BCL2L1* gene in Oral Cancer**

The *BCL*-2 family is known for its pro or anti apoptotic regulators which are seen to influence a number of cellular and epigenetic mechanisms. A recent paper by Alam and Mishra (Alam & Mishra, 2021) highlighted the incidence of *BCL*-2 expression in the progression, and recurrence of oral cancer. They were able to demonstrate the role of *BCL*-x and AP-1 (Fra-2), causing OSCC progression. We recently reported the prevalence of the *BCL2L1* gene and its anti apoptotic properties in our recent paper. Further the *BCL2L1* gene is primarily responsible for the inhibition of cell death which we confirmed in the PANTHER (Mi et al., 2019) database along with a literature study (Li et al., 1998; Makkoch et al., 2016). Similarly, the presence of miR23b plays a major role in the expression of genes in oral cancer. We have for quite some time now known the importance of non-coding RNAs. They were earlier considered as junk data. However over the recent years, their importance has slowly been realized. The miR23b were reported to be significantly reduced in cancer tissues. Few reports indicated the functional analyses of these clustered miRNAs in oral squamous cell carcinoma (OSCC) (Hao & Yu, 2017) . It was also revealed that the miR23b significantly inhibited cancer cell migration and invasion. In our latest paper we were able to highlight the correlation among the miRNA and the *BCL2L1* gene. Thus a correlation between the miR23b and *BCL2L1* gene certainly gains traction for further analysis.

The miRNA-gene disease correlation sheds some light on the regulatory nature of the disease and helps to better understand the mechanism behind the causality of the disease. We were able to analyze with a p-value of 0.107 the incidence of miR23b on the *BCL2L1* gene was significantly high. Therefore we were able to correlate a higher degree of association between the miR23b and *BCL2L1* gene in oral squamous cell carcinoma. The association of miR23b with the *BCL2L1* gene leads to the conclusion that the *BCL2L1* gene is primarily responsible for the inhibition of cell death and has been reported to be significantly up-regulated in oral cancer. The binding of the miR23b plays a significant part as it is speculated to be involved in cell death and helps in regulating the *BCL2L1* gene.

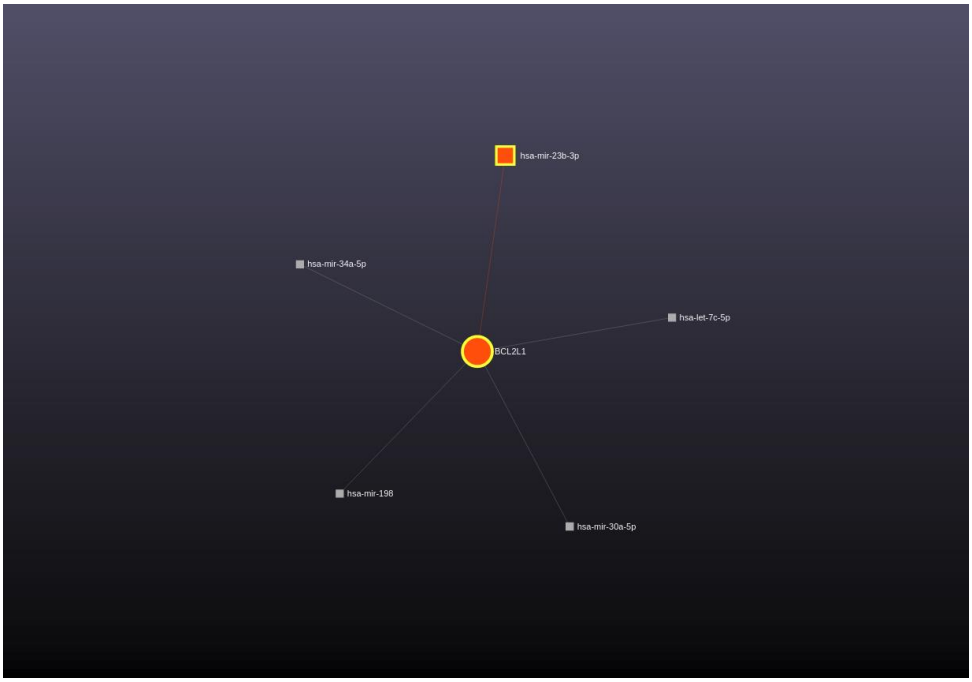

Fig 3: A miRNA-gene co-regulatory network showing the incidence of the miR23b and the *BCL2L1* gene.

IV. Conclusion

Different noncoding RNAs (ncRNAs) that negatively regulate gene expression, such as the microRNAs and the long ncRNAs (lncRNAs), have been associated with cell invasiveness and cell dissemination, tumor recurrence, and metastasis.Increasing evidence points towards the need to explore the possibilities of genome scale expression of ncRNAs in cancer. It would also be beneficial to gain knowledge about their potential biological functions as information is severely lacking in these sectors. In this article we were able to analyze a RNA-Seq dataset of normal and tumor tissue data. We extracted the information about the dysregulated genes and found a total of 1352 highly differentially expressed genes. Among them existed the *BCL2L1* gene of our attention. Furthermore, we generated an miRNA–gene regulatory network with which we were able to isolate the *BCL2L1* gene and fathom its correlation with miRNAs in the regulatory network. The *BCL2L1* gene is primarily responsible for the inhibition of cell death and has been reported to be significantly up-regulated in oral cancer. KEGG analysis also showed that the *BCL2L1* gene is also involved in the PI3K-Akt signaling pathway, JAK-STAT signaling pathway, and NF-kappa B signaling pathway. Moreover, hsa-miR-23b, an miRNA known to target the oral cancer cells by inhibiting cancer cell migration and invasion. It was also found to play a key role in the regulation of the *BCL2L1* gene. These results were validated using literature data, therefore the absence of randomized clinical trials and experimental assays pose limitations.

With the recurrence of oral cancers being one of the most important aspects of the disease. Identifying factors that affect the recurrence of these cancers to reduce postoperative recurrence is an emerging issue in clinics. Since ncRNAs have been linked with the cause of recurrence, a detailed analysis might lead us to the identification of prognostic biomarkers related to the recurrence gains significance of paramount proportions.

References

[1]. Alam, M., & Mishra, R. (2021). Bcl-xL expression and regulation in the progression, recurrence, and cisplatin resistance of oral cancer. Life Sciences, 280, 119705. https://doi.org/10.1016/j.lfs.2021.119705

[2]. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., … Staudt, L. M. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 403(6769), 503–511. https://doi.org/10.1038/35000501

[3]. Fan, Y., Siklenka, K., Arora, S. K., Ribeiro, P., Kimmins, S., & Xia, J. (2016). miRNet—Dissecting miRNA-target interactions and functional associations through network-based visual analysis. Nucleic Acids Research, 44(W1), W135-141. https://doi.org/10.1093/nar/gkw288

[4]. Fukumoto, I., Koshizuka, K., Hanazawa, T., Kikkawa, N., Matsushita, R., Kurozumi, A., Kato, M., Okato, A., Okamoto, Y., & Seki, N. (2016). The tumor-suppressive microRNA-23b/27b cluster regulates the MET oncogene in oral squamous cell carcinoma. International Journal of Oncology, 49(3),1119–1129.https://doi.org/10.3892/ijo.2016.3602

[5]. Hao, L., & Yu, H. (2017). MiR-23b inhibits cell migration and invasion through targeting PDE7A in colon cancer cells. International Journal of Clinical and Experimental Pathology, 10(9), 9436–9443.

[6]. Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics, 107(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003
[7]. Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. Nature Methods,12(4),357–360. https://doi.org/10.1038/nmeth.3317
[8]. Levine, B., Sinha, S., & Kroemer, G. (2008). Bcl-2 family members. Autophagy, 4(5), 600–606.
[9]. Li, H., Zhu, H., Xu, C. J., & Yuan, J. (1998). Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. Cell, 94(4), 491–501.https://doi.org/10.1016/s0092-8674(00)81590-1
[10]. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550. https://doi.org/10.1186/s13059-014-0550-8
[11]. Makkoch, J., Poomipak, W., Saengchoowong, S., Khongnomnan, K., Praianantathavorn, K., Jinato, T., Poovorawan, Y., & Payungporn, S. (2016). Human microRNAs profiling in response to influenza A viruses (subtypes pH1N1, H3N2, and H5N1). Experimental Biology and Medicine (Maywood, N.J.), 241(4), 409–420. https://doi.org/10.1177/1535370215611764
[12]. Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research, 47(Database issue),D419–D426. https://doi.org/10.1093/nar/gky1038
[13]. Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing Technologies. Current Protocols in Molecular Biology, 122(1), e59. https://doi.org/10.1002/cpmb.59
[14]. Sticht, C., Torre, C. D. L., Parveen, A., & Gretz, N. (2018). miRWalk: An online resource for prediction of microRNA binding sites. PLOS ONE, 13(10), e0206239. https://doi.org/10.1371/journal.pone.0206239