INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

A Transcriptomic Analysis to Identify Prevalent lncRNAs in Gingivobuccal Oral Cancer

Agnik Haldar[1], **Ajay Kumar Singh**[1]*

1 Department of Bioinformatics, Center for Biological Sciences (Bioinformatics), Central University of South Bihar, Panchanpur Road,Tekari, Gaya, Fathehpur, 824236, India

OPEN ACCESS

Received: 13-11-2022

Accepted: 25-01-2023

Published: 12-04-2023

Citation: Haldar A, Singh AK (2023) A Transcriptomic Analysis to Identify Prevalent lncRNAs in Gingivobuccal Oral Cancer. Indian Journal of Science and Technology 16(14): 1082-1089. https://doi.org/10.17485/IJST/v16i14.2163

* **Corresponding author**.

ajaysingh@cusb.ac.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: Gingivobuccal cancer is a subtype of oral cancer, prevalent in developing countries where the use of tobacco and Areca (betel) nut is rampant. This study aims to search through the publically available NCBI data to find comparisons between the normal and cancer-affected transcripts of RNA, which are molecules that copied the genetic information from DNA and that will be "read" as instructions to produce other molecules in the body. Using bioinformatic tools, we analyzed the variations in the transcripts and examined how they were correlated with other molecular processes in the body. We specifically investigated the role of "long noncoding RNAs (lncRNAs)" because it is speculated that they play an active role in the causality of cancer. Our analysis found a correlation of CTA-384D8.35 lncRNA with gingivobuccal cancer after analyzing the Hazard Ratio using the Cox Proportional-Hazard Model and is expected to pave the way for future studies. **Methods**: In this study, transcriptomic analysis was performed on a publicly available gingivobuccal cancer dataset obtained from NCBI. We developed a novel method for identification of annotated lncRNAs, which helped us select and correlate the lncRNA transcripts with the gingivobuccal cancer disease. Furthermore, the results were validated using the GEPIA2 tool for integrated analysis of co-lncRNA and GEPIA databases. **Findings**: From 24 RNA-Seq samples (12 normal samples and 12 tumor samples), a total of 16,065 potential lncRNA transcripts were recognized. From these 16,065 lncRNA transcripts, 1201 lncRNA transcripts were further filtered out based on a threshold adjusted p value of <0.01. We were able to for the first time identify these 1201 lncRNA transcripts, which were not annotated for gingivobuccal squamous cell carcinoma previously. Integrated analysis of co-lncRNA and GEPIA databases revealed that a total of 11 lncRNAs were upregulated, whereas 29 lncRNAs were downregulated. Furthermore, CTA-384D8.35 showed a significant correlation with disease-free survival time in HNSC. CTA-384D8.35 which is downregulated in gingivobuccal cancer has been correlated with tumor stage, metastasis, cancer status, clinical stage, and grade**. Novelty:** In this report we are for the first time correlating the CTA-384D8.35 lncRNA with the gingivobuccal cancer disease with the help of statistical and survival analysis. Furthermore,

1201 lncRNA transcripts, which have previously not been annotated for gingivobuccal squamous cell carcinoma, were identified in this study based on a threshold adjusted p value of <0.01. Finally we have also introduced a novel method of identifying annotated lncRNAs in a disease as well as improved upon the NGS data analysis pipeline by determining the strandness of the raw RNA-Seq data. Differential expression of lncRNAs could possibly play a significant role in carcinogenesis. Therefore the identification of lncRNAs gains priority, which are speculated to play a role in the mechanism and formation of various forms of cancer. For identification and analysis of lncRNAs, a comparison of their expression in normal and affected tissues is likely to confirm their involvement in the disease or the lack thereof. The introduction of a novel methodology along with the identification of CTA-384D8.35 lncRNA paves the way for further analysis pertaining to lncRNAs as prognostic or diagnostic biomarkers. Since CTA-384D8.35 lncRNA has previously been correlated with a number of cancer hallmarks, it gains significant traction to be considered as an important aspect in the biomarker identification pertaining to gingivobuccal cancer. This study has reported for the first time the lncRNAs that are prevalent in gingivobuccal squamous cell carcinoma.

Keywords: Gingivobuccal cancer; lncRNA; oral cancer; epigenetics; NGS data analysis

1 Introduction

Long noncoding RNAs (lncRNAs) play a significant role in cancer, as evidenced by numerous reports over the years[1]. They are essential transcriptional and post-transcriptional regulators that mediate gene expression at multiple levels. Cancers are caused by genetic alterations and often originate from aberrant gene expression. Oral cancer, which is associated with tissues in the mouth and throat, is a subtype of a larger group of cancers known as head and neck cancer. Although there are various forms of oral cancer, approximately 90% of the oral cancers are classified as squamous cell carcinomas because the mouth and throat are lined almost entirely with squamous cells[2,3]. Gingivobuccal complex cancer is a type of oral squamous cell cancer and has been reported to be the most common form of oral cancer in India among men and is quite predominant in countries where tobacco use is rampant. Only few studies have been performed on lncRNA expression profiling related to gingivobuccal cancer[4]. Thus, there is a need to properly identify and validate the expression of lncRNA in the gingivobuccal complex. Despite the presence of glaring evidence for lncRNA playing a critical role in the biological processes of human diseases, very few efforts have been made to identify their association with gingivobuccal cancer and assess their prognostic[2,5]. Furthermore the interest in the identification of lncRNA signatures in Gingivobuccal Squamous Cell Carcinoma (GBSCC) stems from the idea of identification of non-invasive biomarkers present in the saliva for early detection of oral cancer[6,7].

Recently the research on the implication of lncRNAs on cancer epidemiology has gained significant traction. This is due to the fact that lncRNAs, which were earlier considered as junk data has now been proven to have epigenetic implications in the mechanistic properties of cancer. There has been several research regarding coding genes and their implications in the tissues and cell lines[8,9]. However, studies pertaining to lncRNAs specifically impacting gingivobuccal cancer has been extremely rare and dated. A majority of the studies reporting the non-coding RNAs have used the dated hg19 reference genome for their alignment. Whereas the stable latest version of

the reference genome available to us is the hg38. As a result for this study we have used the hg38 reference genome as well as the latest iteration of the annotated lncRNAs from various databases like NONCODE and LNCipedia were used to construct the GTF file.

Cancer biomarker identification based on lncRNAs is a relatively new research area. The genesis of next generation sequencing (NGS) data analysis coupled with the renewed interest of noncoding genes has bought forth the idea of narrowing down and concentrating on specific noncoding signature genes for potential biomarker identification. Li et al[5] recently identified 6 novel lncRNA biomarkers in oral cancer using statistical methodologies. TSPEAR-AS2 is another lncRNA which was recently identified as a novel prognostic biomarker for OSCC. It was also reported to behave as a competing endogenous RNA (ceRNA) by sponging the miR-487a-3p[10]. Therefore, constant identification of lncRNAs via in silico methods are being performed in a bid to reduce the amount of targets which requires attention and further studies.

In this study, we have for the first time identified and correlated the CTA-384D8.35 lncRNA with gingivobuccal oral squamous cell carcinoma, which is a relatively novel lncRNA seen to be prevalent in only clear cell renal cell carcinoma[11]. We have also improved upon the existing data analysis pipeline with a more efficient approach for the detection of annotated non-coding RNAs from raw FASTA data files. The methodology adopted is seen to reduce the time as well as increase the accuracy of the data obtained when compared with other identification pipelines.

2 Methodology

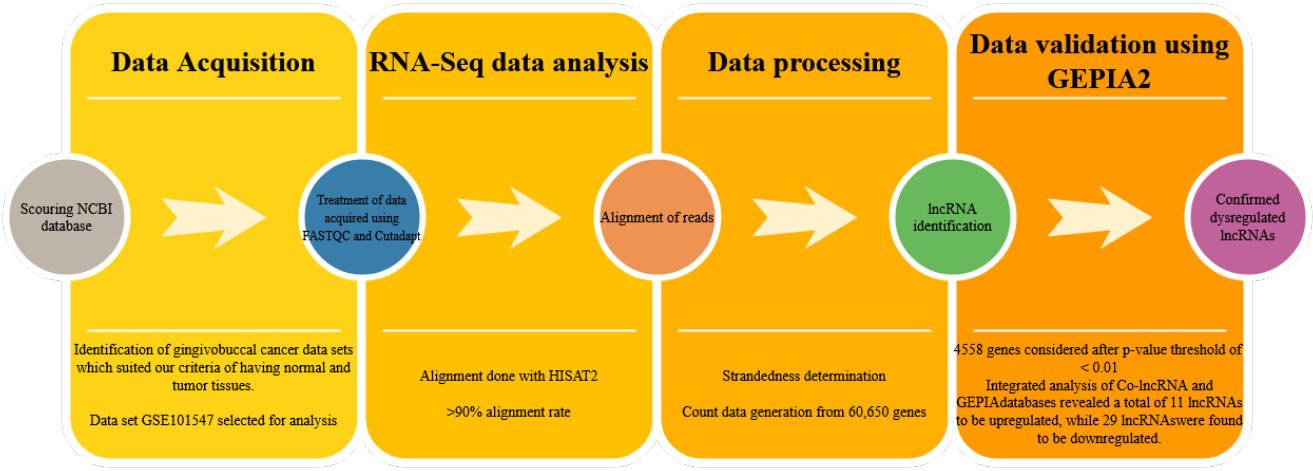

Fig 1. Graphical representation of the computational pipeline performed here

2.1 Data Collection

For this study, as mentioned in the workflow diagram (Figure 1), the NCBI database for gingivobuccal cancer datasets was scoured and those datasets that suited our criteria of having normal and tumor tissues were selected. The available dataset GSE101547 in the GEO database, which consisted of 12 normal tissues and 12 patients samples of gingivobuccal cancer tissues were used[12].

2.2 Data Processing

The FASTQ files were processed using standard transcriptomic bioinformatics analysis pipeline to obtain dysregulated genes from the data. The dysregulated genes provided us with the opportunity to further understand the relational nature of the genes with the disease. The raw sequence reads were initially checked for quality using FASTQC.

2.3 Alignment

The reads were then aligned with the help of a splice-aware aligner HISAT2[13] using the hg38 reference genome. It is quite pertinent to use the latest stable version of the reference genome because using an outdated reference genome runs the risk of

missing out potential transcripts. Over 90% alignment rate was achieved for all the samples. Of the 53 million paired-end reads per sample, approximately 50 million reads were aligned successfully with the hg38 reference genome. Low and inconsistent reads were subsequently removed.

2.4 Count data generation

For the generation of count data, the HTSeq-count tool[14] was used, which provided the count data for the overlapping exons of each gene. To sort the lncRNAs, an lncRNA GTF annotation file was used, which was prepared in-house with the help of the GENCODE GTF file. LncRNA annotations were selected from various lncRNA databases, which helped us sort all the lncRNAs present in the reads and obtain the count data.

2.5 Data Analysis

The data available in the dataset lacked information on the strandness of the reads; moreover, it was not considered in the study by Singh et al. Hence, the strandness of the raw reads was determined in this study because it is an important aspect when looking for count data. The Salmon quantification tool[15] was used to identify the dataset to be unstranded. Finally, for differential gene expression data analysis, the DESeq2 tool[16] was used. For differential gene expression data analysis we compared the normal with tumor tissues. A gene is considered as differentially expressed if the false discovery rate (FDR)<0.01 and log2 fold change is >= 2 or <= -2. We also performed multidimensional scaling (MDS) based sample clustering (where every individual gene's expression is taken as an independent vector, distance metric: Euclidian). For identification and analysis of dysregulated genes, comparison between the expression in normal and affected tissues (Figure 2 A) 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.

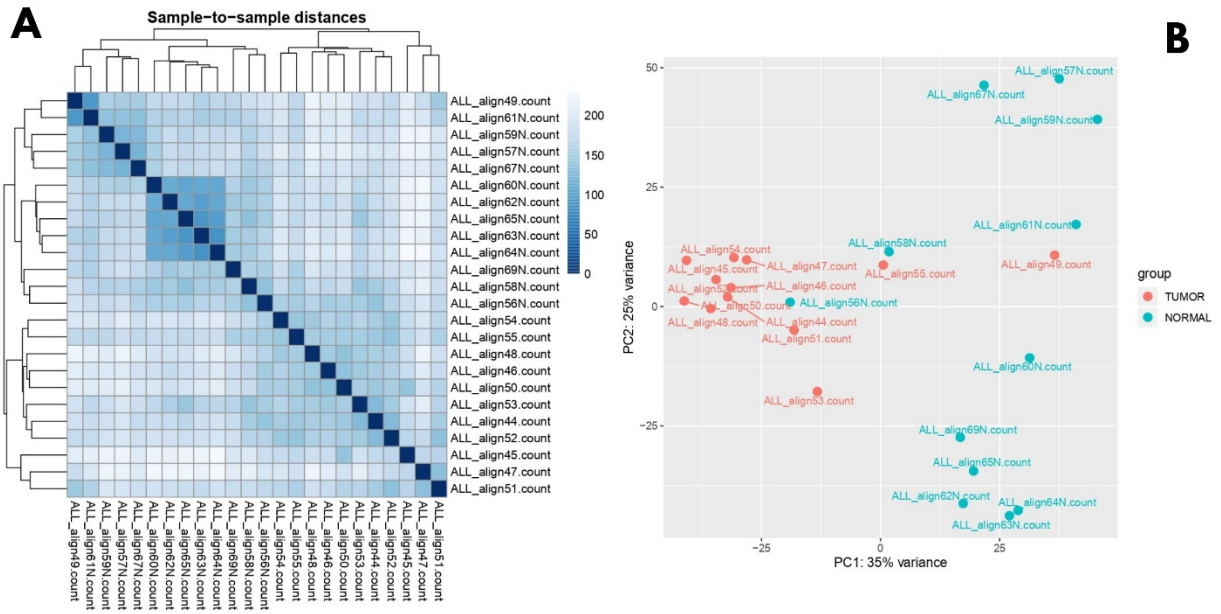

Fig 2. (A) Sample to sample distance plot validating the quality of the samples with respect to each other. (B) PCA Plot with a PCI variance of 35% and PC2 variance of 25% for the 24 samples.

The cluster analysis sheds light on the behavior of the normal and tumor samples. A major part of the tumor samples are seen to be rigid in nature and are based around the PC2.

Principle component analysis can successfully separate the normal and the tumor tissues. As evidenced in Figure 2 (B), the plot displays good separation between normal tissue and diseased tissue from the gingivobuccal cancer tissues.

The gingivobuccal cancer datasets yielded 16,065 differentially expressed lncRNA genes. Upon excluding the outliers and low counts and setting an adjusted p-value threshold of <0.01, the lncRNAs were sorted based on the threshold value and 1201 significant dysregulated lncRNAs were obtained. Dysregulation of lncRNAs indicated the involvement of those genes in the

disruption of normal pathways. The dysregulated genes provided us with an incentive to further explore the causality and their involvement in the disease. A comparison of the methods was also performed to validate the authenticity of the method, and it was found that sorting lncRNAs using a specific lncRNA GTF annotation file yielded better results. The standard lncRNA identification protocol, which includes identification of transcript length and using a coding potential calculator to identify the coding potential of the genes, was able to detect 988 lncRNA genes. Whereas our proposed method of using a single GTF file to identify lncRNA transcripts within the raw alignment data yielded 1201 genes.

2.6 GEPIA2 Analysis

GEPIA2[17] is an updated version of the GEPIA (Gene Expression Profiling Interactive Analysis) web server and is used mainly for the comparative analysis of gene expression between cancerous and normal tissues fetched from the TCGA and the GTEx databases. Here, the GEPIA2 database was used, and head and neck squamous cell carcinoma entries present in it were thoroughly analyzed. The CTA-384D8.35 lncRNAs were specifically isolated owing to a combination of the results of prevalence obtained from our NGS data analysis pertaining to Hazard Ratio analysis using Cox PH Model and literature review, which highlighted their association with cancer hallmarks.

2.7 Disease-free survival analysis

The disease-free survival (DFS) of CTA-384D8.35 was analyzed with the help of GEPIA2 which is an online tool and database for analysis, evaluation and download of cancer-related genomics data sets. We were able to generate a survival map of the CTA-384D8.35 lncRNA to understand and filter the prognostic values in gingivobuccal oral cancer. We set a parameter of "0.05" as the significance level, and the group cutoff was set as "Median". The expression level of the CTA-384D8.35 was sorted by the best cutoff and months of survival were selected as survival time units.

3 Results and Discussion

The dysregulated genes which were obtained after comparing the normal dataset with the tumor dataset yielded data that was used to further identify and explore the relational nature of the genes to the disease.

The dysregulated genes (Figure 3) serve as an indication of the disruption of the normal genomic process. 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 the identification of lncRNAs expressed in gingivobuccal cancer, we assembled the transcripts for each of the 24 RNA-Seq samples. We were able to report a total of 16065 potential transcripts. Setting a threshold p-value of < 0.01, we were able to sort and consider 4558 potential dysregulated genes as statistically significant. Further we set a series of stringent parameters on these 4558 genes for identification of known lncRNAs from our analysis and we were for the first time able to identify 1201 lncRNA transcripts based on a threshold p-value of < 0.01 which were not previously annotated for gingivobuccal squamous cell carcinoma. Additionally our novel methodology of identifying lncRNAs based on the GTF file significantly reduced the time and resources to identify the dysreglated lncRNAs. There have been studies[18] pertaining to lncRNA-mRNA network identification in head and neck cancer based on TCGA data, however lncRNA identification specifically for gingivobuccal cancer from patient data obtained from the Indian subcontinent is severely lacking.

The lncRNA transcripts that were obtained from our data analysis were then annotated and identified. For validation purposes, we employed the usage of the GEPIA2 database since it also gave our study the validation of survival analysis. The GEPIA2 database was thoroughly analyzed, for head and neck squamous carcinoma entries. From a total of 2077 HNSC entries, we compared the co-lncRNA data with our findings of lncRNAs and we can report that 40 lncRNA transcripts were shown to be highly prevalent in GBSC. The results along with information pertaining to their chromosomal location are provided in Supplementary Table 1. Integrated analysis of Co-lncRNA and GEPIA2 databases revealed a total of 11 lncRNAs to be upregulated, while 29 lncRNAs were found to be downregulated.

The 40 dysregulated lncRNAs were analyzed in the lncSEA database[19] for disease association and functional annotation. Out of the 40 novel transcripts, 6 lncRNAs returned positive disease associations as mentioned in Table 1.

Next we went for a disease-free survival analysis which showed that correlation was obtained in downregulated lncRNAs which reported a longer disease-free survival time in HNSC as evidenced from the graph in Figure 4. The GEPIA2 dataset was explored to find a correlation between the lncRNA expression and survival time. We were able to analyze that CTA-384D8.35 among all the 40 lncRNAs showed a significant correlation to disease-free survival time in HNSC.

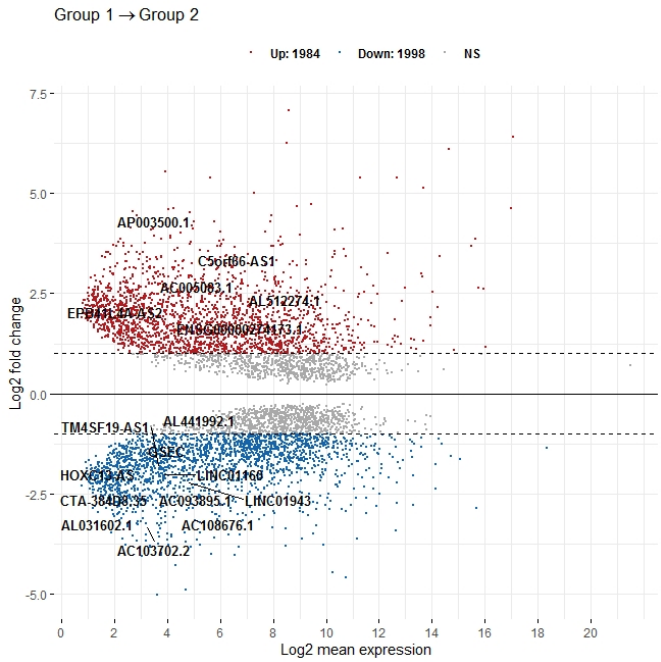

Fig 3. MA Plot representing DE genes (red dots denoting upregulated genes and blue dots denoting downregulated genes) against mean expression levels with a False Discovery Rate (FDR) < 0.01.

Table 1. LncSEA analysis results on the 40 dysregulated lncRNAs

Set	Class	Count	LncRNA	P-value	FDR	Bonferroni
Non_small_cell_lung_cancer	Disease	4	DUXAP8 SNHG1 AGAP2-AS1 PART1	3.27E-06	0.000494	0.000494
Glioma	Disease	3	LINC01116 SNHG1 AGAP2-AS1	2.44E-05	0.00184	0.00368
Glioblastoma	Disease	3	LINC01116 AGAP2-AS1 PART1	4.06E-05	0.0156	0.0156
Prostate_cancer	Disease	3	LINC01116 SNHG1 PART1	4.18E-05	0.0021	0.00631
Esophageal_cancer	Disease	2	DUXAP8 SNHG1,	0.000136	0.00513	0.0205

In this report, we have used a novel methodology of identifying annotated lncRNAs by creating a separate GTF file which was made in-house consisting of all the annotated lncRNAs from lncRNA data repositories like NONCODE and matching them with the alignment file generated from the experiment. The proposed model significantly reduces the time taken to identify the dysregulated lncRNAs in the disease. Plus the addition of the strandness determination in the pipeline also increases the accuracy of the data obtained. Previous studies pertaining to the data analysis of dysregulated genes often skip the step which can lead to biasness setting in the datasets.

The dysregulated lncRNAs were then implemented in the GEPIA2 database. Upon further inspection, we were able to correlate our findings with the GEPIA2 database yielding 40 dysregulated lncRNAs out of which 29 were downregulated and 11 were upregulated.

CTA-384D8.35 was found among the downregulated lncRNAs to be the most significantly dysregulated and also upon survival analysis it was correlated to show a shorter disease-free survival rate. This is an important perspective since the idea of a potential non-invasive biomarker for cancer diagnosis has been seen and discussed in previous studies [5,10,20] in oral cancer patients. In the present study, we are able to report a relatively novel lncRNA CTA-384D8.35 which was seen to be prevalent in GBSCC and also showed that the regulation of the lncRNA is related to the survival of the patients. Furthermore, we can also infer from this finding that downregulation of the CTA-384D8.35 gene is correlated to the occurrence of gingivobuccal carcinoma. Further analysis has also revealed the CTA-384D8.35 is prevalent in various cancers like clear cell Renal Cell

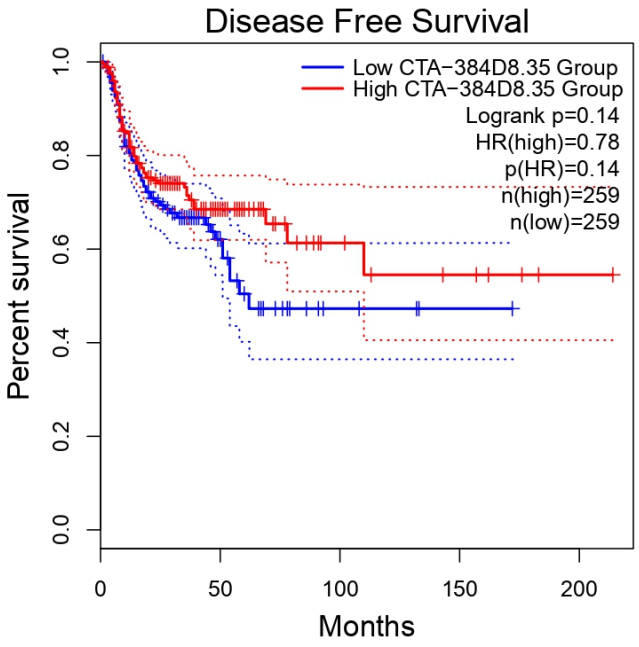

Fig 4. Disease free survival of CTA-384D8.35 lncRNA generated with the help of GEPIA2

Carcinoma (ccRCC) and it is related to tumor stage, metastasis, cancer status, clinical stage, and grade [11]. Our data analysis was also able to report about numerous oncogenes as well as tumor suppressor genes that are seen present in the 8p and 8q arm of the chromosome in oral cancers. Significant non-coding RNAs such as PCAT1 which were found to be highly altered in chromosome 8 led us to believe that a high number of genes were altered in the chromosome 8 region [21]. It has been earlier reported in numerous articles about gains and losses in chromosome 8 [22,23] among various other cancers like breast, colorectal, and prostate cancer.

4 Conclusion

Different noncoding RNAs that negatively regulate gene expressions, such as the microRNAs and the lncRNAs, have been associated with cell invasiveness and cell dissemination, tumor recurrence, and metastasis. Thus a comparison between the expression of lncRNAs in normal and affected tissues will confirm the involvement of lncRNAs in the disease or the lack thereof. Increasing evidence points towards the need explore the possibilities of genome-scale expression of lncRNAs in cancer. It would also be beneficial to gain knowledge about their potential biological functions as information is severely lacking in these sectors.

Furthermore, we have introduced a novel methodology pertaining to identification of annotated lncRNAs using the GTF file which consisted of all the annotated lncRNAs from data repositories like NONCODE and matching them with the alignment file generated from the experiment. The method adopted significantly reduced the time required to obtain the dysregulated lncRNAs seen to be implicated in the disease.

Additionally we have also improved upon the accuracy rate for the identification of dysregulated genes with the implementation of strandness determination step in the data analysis pipeline. The strandness of the raw reads was determined in this study because it is an important aspect due to the fear of biasness setting in the count data.

Finally the identification of the CTA-384D8.35 was found among the downregulated lncRNAs to be the most significantly dysregulated and also upon survival analysis it was correlated to show a shorter disease-free survival rate. Further analysis also revealed that the CTA-384D8.35 is prevalent in various cancers like clear cell Renal Cell Carcinoma (ccRCC) and it is related to tumor stage, metastasis, cancer status, clinical stage, and grade.

A detailed analysis regarding the lncRNA might lead us to the identification of prognostic biomarkers related to the recurrence, which gains significance of paramount proportions. Our next step of action is to functionally annotate the lncRNA as a majority of their functions are still unknown and understand its mechanistic properties. The results obtained in this report certainly points to the fact that the CTA-384D8.35 lncRNA is a target of interest which shows its potential as a prominent aspect

of research in the disease pertaining to non-invasive biomarker identification.

References

1) Liu H, Wang D, Kan S, Hao M, Chang L, Lu P. The role of lncRNAs and XIST in oral cancer. . Available from: https://www.frontiersin.org/articles/10.3389/fcell.2022.826650.
2) Xu Y, Jiang E, Shao Z, Shang Z. Long Noncoding RNAs in the Metastasis of Oral Squamous Cell Carcinoma. *Frontiers in Oncology*. 2021;10:3014. Available from: https://doi.org/10.3389/fonc.2020.616717.
3) Borse V, Konwar AN, Buragohain P. Oral cancer diagnosis and perspectives in India. *Sensors International*. 2020;1:100046. Available from: https://doi.org/10.1016/j.sintl.2020.100046.
4) Farah CS. Molecular landscape of head and neck cancer and implications for therapy. *Annals of Translational Medicine*. 2021;9(10):915. Available from: https://doi.org/10.21037/atm-20-6264.
5) Li Y, Cao X, Li H. Identification and Validation of Novel Long Non-coding RNA Biomarkers for Early Diagnosis of Oral Squamous Cell Carcinoma. *Frontiers in Bioengineering and Biotechnology*. 2020;8:256. Available from: https://doi.org/10.3389/fbioe.2020.00256.
6) Suresh PS. Emerging Roles of Salivary lncRNAome in Oral Cancer. *Indian Journal of Clinical Biochemistry*. 2021;36(2):246–247. Available from: https://doi.org/10.1007/s12291-019-00869-2.
7) Abdeldayem E, Rashed L, Ali S. Salivary expression of lncRNA DQ786243 and IL-17 in oral lichen planus: case–control study. *BMC Oral Health*. 2022;22(1):240. Available from: https://doi.org/10.1186/s12903-022-02277-0.
8) Wu HTT, Chen WTT, Li GWW, Shen JXX, Ye QQQ, Zhang MLL, et al. Analysis of the Differentially Expressed Genes Induced by Cisplatin Resistance in Oral Squamous Cell Carcinomas and Their Interaction. *Frontiers in Genetics*;10. Available from: https://www.readcube.com/articles/10.3389/fgene.2019.01328.
9) Ghosh A, Das C, Ghose S, Maitra A, Roy B, Majumder PP, et al. Integrative analysis of genomic and transcriptomic data of normal, tumour, and co-occurring leukoplakia tissue triads drawn from patients with gingivobuccal oral cancer identifies signatures of tumour initiation and progression. *The Journal of Pathology*. 2022;257(5):593–606. Available from: https://doi.org/10.1002/path.5900.
10) Xia YCC, Cao J, Yang J, Zhang Y, Li YSS. lncRNA TSPEAR-AS2, a Novel Prognostic Biomarker, Promotes Oral Squamous Cell Carcinoma Progression by Upregulating PPM1A via Sponging miR-487a-3p. *Disease Markers*. 2021;2021:1–11. Available from: https://doi.org/10.1155/2021/2217663.
11) Zeng JH, Lu W, Liang L, Chen G, Lan HH, Liang XY, et al. Prognosis of clear cell renal cell carcinoma (ccRCC) based on a six-lncRNA-based risk score: an investigation based on RNA-sequencing data. *Journal of Translational Medicine*. 2019;17(1):281. Available from: https://doi.org/10.1186/s12967-019-2032-y.
12) Singh R, De Sarkar N, Sarkar S, Roy R, Chattopadhyay E, Ray A, et al. Analysis of the whole transcriptome from gingivo-buccal squamous cell carcinoma reveals deregulated immune landscape and suggests targets for immunotherapy. *PLOS ONE*. 2017;12(9):e0183606. Available from: https://doi.org/10.1371/journal.pone.0183606.
13) Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*. 2019;37(8):907–915. Available from: https://doi.org/10.1038/s41587-019-0201-4.
14) Anders S, Pyl PT, Huber W. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–175. Available from: https://doi.org/10.1093/bioinformatics/btu638.
15) Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat Methods*. 2017;14(4):417–426. Available from: https://doi.org/10.1038/nmeth.4197.
16) Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550–550. Available from: https://doi.org/10.1186/s13059-014-0550-8.
17) Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Research*. 2019;47(W1):W556–W560. Available from: https://doi.org/10.1093/nar/gkz430.
18) Zhong Z, Hong M, Chen X, Xi Y, Xu Y, Kong D, et al. Transcriptome analysis reveals the link between lncRNA-mRNA co-expression network and tumor immune microenvironment and overall survival in head and neck squamous cell carcinoma. *BMC Medical Genomics*. 2020;13(1):57. Available from: https://doi.org/10.1186/s12920-020-0707-0.
19) Chen J, Zhang J, Gao Y, Li Y, Feng C, Song C, et al. LncSEA: a platform for long non-coding RNA related sets and enrichment analysis. *Nucleic Acids Research*. 2021;49(D1):D969–D980. Available from: https://doi.org/10.1093/nar/gkaa806.
20) Chaudhary R, Wang X, Cao B, Iglesia JDL, Masannat J, Song F. Long noncoding RNA, LINC00460, as a prognostic biomarker in head and neck squamous cell carcinoma (HNSCC). *Am J Transl Res*. 2020;12(2):684–96. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7061833/.
21) Bardhan A, Banerjee A, Basu K, Pal DK, Ghosh A. PRNCR1: a long non-coding RNA with a pivotal oncogenic role in cancer. *Human Genetics*. 2022;141(1):15–29. Available from: https://doi.org/10.1007/s00439-021-02396-8.
22) Dehner C, Moon CI, Zhou Z, Zhang X, Miller C, Xu H, et al. Chromosome 8 gain is associated with high-grade transformation in MPNST. *JCI Insight*;6(6):146351. Available from: https://doi.org/10.1172/jci.insight.146351.
23) Wilson C, Kanhere A. 8q24.21 Locus: A Paradigm to Link Non-Coding RNAs, Genome Polymorphisms and Cancer. *International Journal of Molecular Sciences*. 2021;22(3):1094. Available from: https://doi.org/10.3390/ijms22031094.