Transcriptome analysis of oral samples

Dr. B. Nandlal[1*] and Dr. M V S S T SubbaRao[2*]

1. Dept. Of Pedodontics and Preventive Dentistry, JSS Dental College and Hospital.
2. CEMR Laboratory, Dept. Of Biochemistry, JSS Medical College, JSS University, Mysore.

* Authors for Correspondence: dr_nandlal@yahoo.com; mvsstsubbarao@jssuni.edu.in; madhunapantulas@yahoo.com

ABSTRACT:

Analysis of transcriptome of cells present in saliva and buccal cheek provide key information about the expression of genes in healthy individuals. Knowledge about the expression profile of genes helps to determine the key players involved in keeping the oral health in good condition. However, to date, not much is known about the ideal sample, salivary cells or buccal cheek cells, to analyze for measuring the expression profile of genes. Therefore, in this study, we have compared the expression profile (at mRNA level) of cells collected from saliva of normal healthy individuals with the cells harvested from buccal cheek of the same individual.

INTRODUCTION:

Oral health diseases are serious, lifelong multifactorial conditions characterized by inflammatory changes due to the microbiological , local and dietary factors. There is no doubt that genetic components play a key role in the pathogenesis of with a high heritability index. However, the genes that have been associated with health and diseased only relate to only a small portion of the cases .Due to the wide range of genetic heterogeneity reported, it is quite possible that different causes and various underlying molecular mechanisms may lead to a common set of changes in the oral tissues resulting in a similar behavioral profile. Nonetheless, analysis of the transcriptome, as an intermediate phenotype, still could provide valuable insights for investigation of such complex cases. Careful comparison of the gene expression profiles in oral conditions patients with normal subjects could provide important clues for the development of and also the molecular mechanisms and crossroads involved. A review of the literature shows that no single gene or particular set of genes has been used consistently as reference in gene expression studies in these oral conditions. Surprisingly, as we will further outline in the discussion, almost none of these studies have embarked on a systematic validation of the reference gene(s) prior to the start of the experiment. Total RNA was isolated from the cells present in saliva and buccal cheek of 15 young adults and quality assessed using Tape Station. The quality RNA was processed as shown in Image and differentially expressed genes identified. Genes were considered significantly differentially expressed, if their q values < 0.05 from ANOVA tests and log2Fold Change value is above and below 1.

MATERIALS AND METHODS:

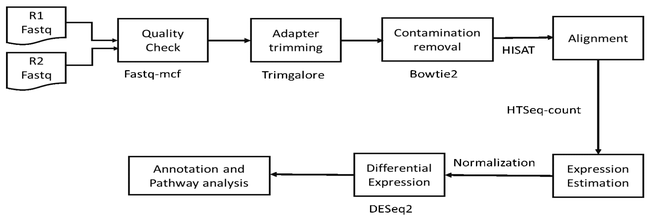


Figure 1: Workflow for the bioinformatics analysis

Sequenced 15 samples on Illumina HiSeq 4000 machine. This report provides the transcriptome analysis of these 15 oral samples. The workflow involves (Figure 1) Quality check, adapter trimming and alignemnt to the reference genome (hg19) followed differential gene expression analysis.

Trimming: The Sequence reads are trimmed where necessary to only retain high quality sequence for further analysis. In addition, the low-quality sequence reads are excluded from the analysis. The adapter trimming was performed using Trim Galore.

Contamination Removal: For the RNA-Seq analysis, unwanted sequences such as mitochondrial genome sequences, ribosomal RNAs, transfer RNAs, adapter sequences are removed using Bowtie2 (version - 2.2.4).

Read Alignment: The paired-end reads are aligned to the reference human genome Feb. 2009 release downloaded from UCSC database (GRCh37/hg19). The GTF file was downloaded from ENSEMBL. Alignment was performed using Hisat (0.1.7) and the aligned reads were counted using HTSeq-0.6.1. Raw Read count data were normalized using DESeq2 for all the samples. In all samples, we observed adequate data and an acceptable level of read mapping (>90%).

Identification of differentially expressed genes: We normalized each data set internally using DESeq2, and then by using gene expression levels in the DMSO treated cells. The results of these comparisons were used for understanding the combinatorial effect of the molecule and the stimulant. All zero expression genes were excluded from this analysis. We also excluded genes whose normalized expression values across all treatment conditions was zero. Genes were considered significantly differentially expressed, if their q values < 0.05 from ANOVA tests and log2FoldChange values is above and below 1.

Functional Annotation: The differentially expressed genes are subjected to Gene Set Enrichment Analysis. The statistically significant one are annotated in terms of gene ontology terms (Molecular Function, Biological Process and Cellular component) and Reactome pathways.

QUALITY AND YEILD OF RNA:

SAMPLE ID	SAMPLE NAME	Nano Drop (ng/µL)	A 260/280 Ratio	A 260/230 Ratio	QUBIT (ng/µL)	Volume (µl)	RIN NO
63093	301_S	12.7	1.79	0.03	11.2	35	3.1
63096	302_S	10.5	2.08	0.02	5.7	35	5.9
63099	303_S	4.7	1.63	0.02	Too Low	35	1.4
63102	304_S	24.4	1.89	0.04	22.6	35	3.3
63105	305_S	10.2	1.82	0.02	11.2	35	4.6
63108	306_S	33.5	1.56	0.02	Too Low	35	-
63111	307_S	22.6	1.88	0.03	20.4	35	3.4
63114	308_S	4.7	2.25	0.01	Too Low	35	1.6
63117	309_S	55.4	1.53	0.04	14.8	35	2.7
63120	310_S	96.1	1.61	0.07	36.2	35	5.6
63123	311_S	19	1.59	0.02	Too Low	35	2.6
63126	312_S	77.9	1.56	0.06	25.8	35	6.3
63129	313_S	61.7	1.47	0.06	15.0	35	5.8
63132	314_S	31.6	1.84	0.05	30.0	35	5.5
63135	315_S	47.7	1.63	0.07	31.2	35	3.8

Table 1 : The yield of RNA isolated from Saliva.

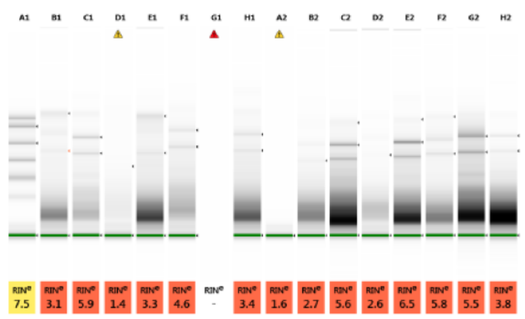


Figure 2: Gel image showing quality of RNA isolated from saliva.

SAMPLE ID	SAMPLE NAME	Nano Drop (ng/µL)	A 260/280 Ratio	A 260/230 Ratio	QUBIT (ng/µL)	Volume (µl)	RIN NO
63095	301_BS	78.4	1.43	0.31	40.6	70	1.9
63098	302_BS	10.1	1.19	-0.17	22.8	70	4.4
63101	303_BS	20.0	1.39	0.55	31.8	70	3.8
63104	304_BS	58.2	1.58	0.52	61.6	70	2.2
63107	305_BS	67.6	1.48	0.32	47.2	70	4.2
63110	306_BS	3.2	0.66	-0.07	TOO LOW	70	4.9
63113	307_BS	343.0	1.81	2.39	55.4	70	3.8
63116	308_BS	25.5	1.48	0.75	35.2	70	2.1
63119	309_BS	41.1	1.43	-0.96	36	70	3.5
63122	310_BS	71.1	1.61	4.42	76.2	70	3.8
63125	311_BS	22.9	1.32	-0.42	30	70	3.4
63128	312_BS	12.9	1.13	-0.29	31.4	70	3.6
63131	313_BS	45.4	1.41	1.09	44	70	2.5
63134	314_BS	41.9	1.48	-3.09	50.6	70	4.9
63137	315_BS	37.6	1.55	0.59	22	70	2.5

Table 2 : The yield of RNA isolated from Buccal Swab.

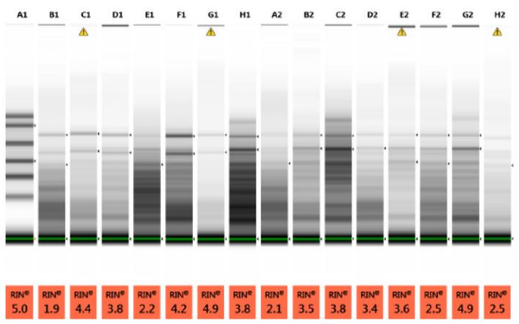


Figure 3: Gel image showing quality of RNA isolated from Buccal Swab.

Alignment Statistics:

Sample IDs	Raw Data. Size(Gb)	Data Size After.AT	DataSize After.CR		Alignment %(Bacterial Genome)
308_BS	7.177963	6.897212	6.204913	65.24	**13.708**
303_BS	7.184943	6.907914	6.265079	63.16	**34.96%**
305_S	6.076094	5.772480	5.271871	56.24	**27.3754**
302_S	8.008164	7.735948	7.102439	51.36	**39.5926**
314_S	7.309745	6.844309	6.284525	51.00	**43.0016**
313_S	7.482948	7.163580	6.730743	41.39	**67.1761**
311_S	9.618220	8.769612	8.637582	8.68	**63.6291**
304_S	7.153951	6.405836	6.160605	3.48	**39.0273**
312_S	9.674642	9.111988	8.978378	3.21	**72.853**
310_S	5.976139	5.425055	5.371978	1.63	**55.7915**

Table 3: Indicates the data size after each step of the preprocessing like Adapter trimming (AT) and Contamination removal (CR). The alignment percentages to the reference genome and bacterial genomes is also indicated.

RESULTS AND DISCUSSION:

It is seen that the alignment percentage for sample varies from 0.5-72.3%. The samples were checked for bacterial contamination by aligning to the bacterial genome database. Many of the samples are seen to have around 20-60% alignment percentage. Only 6 samples in which the alignment percentage is above 20% have been taken forward for further analysis.

Differential Gene Expression Analysis:

The normalized read counts for the differentially expressed genes are given as heat maps for each comparisons. The following table indicates the up and down regulated genes in each of the comparisons.

Comparisons	Up	down	Total
comparison	15	125	140

Table 4: Differentially expressed gene counts.

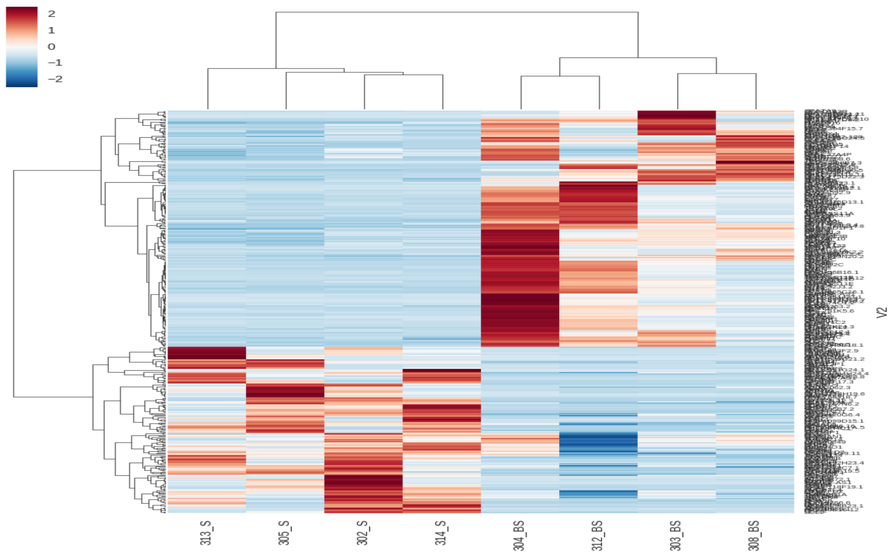


Figure 4: The differentially expressed genes showing Up and downregulated genes for comparison

Among a total of 140 RNAs differentially expressed between saliva and buccal cheek cells, 15 genes found to be upregulated and 125 genes appeared to be down regulated.

CONCLUSION AND SIGNIFICANCE:

In conclusion, results of this study demonstrate that the ideal sample for RNA isolation and subsequent transcriptome analysis is buccal cheek but not the saliva.