Published in final edited form as:

Arch Oral Biol. 2007 January ; 52(1): 30–35. doi:10.1016/j.archoralbio.2006.08.014.

Characterization of Salivary RNA by cDNA Library Analysis

Noh Jin Park^a, Xiaofeng Zhou^{a,b,c}, Tianwei Yu^a, Brigitta M.N. Brinkman^a, Bernard Zimmerman^a, Visu Palaniswamy^a, and David T. Wong^{a,b,c,d,e,*}

^aUCLA Dental Research Institute

^bUCLA School of Dentistry

^cUCLA Jonsson Comprehensive Cancer Center

^dUCLA Division of Head and Neck Surgery/Otolaryngology

^eUCLA Henry Samueli School of Engineering

Abstract

Oral fluid (saliva) meets the demands for a noninvasive and accessible diagnostic medium. Recent reports by our group and others described the presence and use of human RNA in saliva as a diagnostic or forensic tool, including the use for oral cancer detection. To gain insights into the integrity of salivary RNA, we examined in detail the integrity of salivary RNA by generating a cDNA library from pooled supernatant saliva of 10 healthy donors. From a library with a primary library titer of 1.3×10^6 cfu/mL of which 95 % of the clones had inserts, we successfully sequenced 117 random colonies containing recombinant clones. BLAST search results indicated that all of these clones contained sequences of human origin. Most of the salivary RNAs appeared to be endonucleolytically cleaved at random positions as indicated by comparisons to respective full length parental RNAs from the Genbank. Twelve of the insert sequences matched to the normal salivary core transcriptome sequences, which are highly abundant mRNAs present in healthy individuals. This study provides an in-depth molecular analysis of the saliva transcriptome and should be a useful resource for future basic and translational studies of RNA in human saliva. In addition this paper presents unequivocal evidence for the presence of RNA in saliva as determined by the use of diverse techniques such as reverse transcriptase quantitative polymerase chain reaction (RT-qPCR), *in vitro* translation, and the construction of a salivary cDNA library.

Keywords

saliva; RNA; cDNA library; RNA degradation; translational research

Introduction

Previous studies have demonstrated that the supernatant phase of human saliva contains on average, about 3,000 different mRNA species (1). Of these, 185 mRNAs were consistently detected in saliva of healthy subjects, referred as normal salivary core transcriptome (NSCT).

*Corresponding author at: UCLA School of Dentistry, Dental Research Institute, 73-017 CHS, 10833 Le Conte Avenue, Los Angeles, CA 90095, Phone: (310) 206-3048, Fax: (310) 825-7609, dtww@ucla.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Human salivary mRNA has been used for various applications. Forensic scientists use a small set of saliva-specific mRNAs for body fluid identification (2,3). Inflammatory mRNA markers were detected in whole saliva to monitor the status of periodontal disease in type II diabetic patients (4). Studies from our laboratory have demonstrated the feasibility of utilizing salivary mRNA for detection of oral cancer (5,6). Recently, we showed that salivary RNA is protected from degradation due to its association with macromolecules (7). Salivary RNA appears to enter the oral cavity from different sources including the three major salivary glands, gingival crevice fluid, and desquamated oral epithelial cells. RT-qPCR data indicated that there is not a preferential degradation pattern of salivary RNA from either 5' nor 3' ends, suggesting that endonucleases are largely responsible for salivary RNA degradation (7).

To further evaluate the integrity of salivary RNA, we have generated a cDNA library using RNA isolated from the supernatant phase of saliva. Our analysis is based on the randomly sequenced 117 colonies and subsequent bioinformatics analyses.

Materials and Methods

Participants

A total of 14 healthy volunteers were recruited for this study. For the generation of salivary cDNA library, the saliva was collected from 10 volunteers. For the RT-qPCR and in vitro translation, the saliva was collected from 4 volunteers. The age of the subjects ranged between 22 to 45 years old and with an average age of 31 years. These volunteers consisted of 2 Caucasians and 12 Asians; 4 females and 10 males. The oral health status of all participants was good at the moment of saliva collection. The volunteers have no history of malignancy, immunodeficiency, autoimmune disorders, hepatitis or HIV infection.

Saliva RNA extraction and nuclease treatment

Salivary RNA used for cDNA library construction was isolated from 80 mL of pooled supernatant saliva from 10 healthy donors, which yielded ~5 µg of total RNA. The RNA isolation method was described previously (5). Salivary RNA used in Figures 1 and 2 were isolated from the supernatant phase of saliva using RNeasy Micro Kit (QIAGEN Inc., Valencia, CA) following centrifugation (2,600 g, 15 min, 4 °C). Briefly, total RNA was isolated from 15 mL of supernatant saliva from 4 donors. RLT lysis buffer, a component of RNeasy Micro Kit, was added to the saliva sample in a 3.5:1 ratio without addition of β-mercaptoethanol. Next, 2.5x original saliva volume of 100 % ethanol was added to the sample. The remaining protocol was followed as suggested by the manufacturer. The samples were divided into three aliquots; one aliquot was left untreated, one aliquot underwent on-column RNase treatment and the last aliquot underwent on-column DNase I treatment. Per 5 mL of saliva, 6 columns were used to extract RNA. Therefore, about 830 µL of saliva was loaded per column and each column was eluted with 40 µL thus yielding total of 240 µL of extracted salivary RNA. RNase treatment consisted of adding 2 µL of RNase cocktail (Ambion Inc., Austin, TX, USA) mixed with 78 µL of DNase buffer to the column, and incubation of the column for 15 minutes at room temperature. DNase I treatment consisted of adding 80 µL of DNase I solution according to the manufacturer's instruction (QIAGEN Inc.). Human breast cancer MCF7 cell line RNA was isolated using RNeasy Mini kit according to the manufacturer's instruction (QIAGEN Inc.). Agilent 2100 Bioanalyzer was used to visualize RNA (Agilent Technologies, Palo Alto CA). The basic concept of Agilent 2100 Bioanalyzer is size separation by the micro-scale gel matrix coupled with the continuous UV spectrophotometer reading of the flow through that gives rise to a virtual gel of nucleic acids based on their sizes and quantities.

cDNA library construction

The cDNA library was generated in collaboration with DNA Technologies Incorporation (Gaithersburg, MD). Our strategy is aimed to detect all of the RNAs, including those that lacked the poly(A) tail due to degradation. Accordingly, a poly(A) tail was added before cDNA synthesis using the Poly(A) Tailing kit according to the manufacturer's instruction (Ambion, Austin, TX). Reverse transcription was performed with oligo-dT primers linked to a Xho I site at its 5' end. Methyl dCTP was used instead of dCTP to protect the internal Xho I sites during the subcloning. pBluescript SK (Stratagene, La Jolla, CA) was used as the cloning vector, and inserts less than 200 bp were removed during the size fractionation. cDNA clones were then transformed into *Escherichia coli*. The primary library titer was 1.3×10^6 cfu/mL and 1×10^8 cfu/mL after amplification. About 95 % of the library had inserts.

Sequencing

A total of 117 colonies were sequenced successfully. The requisite for successful sequencing is when the sequence alignment search using the BLAST of the insert sequences results in a matching parental RNA in the sequence database with the BLAST score over 100, and the highest scored parental RNA was selected. Of 117 colonies, 85 colonies were sequenced bi-directionally by Agencourt (Beverly, MA) using M13 forward and reverse primers. Thirty two colonies were sequenced in one direction by UCLA Sequencing and Genotyping Core Facility using M13 forward primers. Degradation status reported in Figure 3 contains results from clones that were sequenced bidirectionally.

RT-qPCR

Primers were designed using Primer 3 software (8), and synthesized from Sigma-Genosys (The Woodlands, TX). RT was performed with ABI cDNA synthesis kit (ABI, Foster City, CA, USA). Final concentrations of the RT reaction are 2 mM (mmol/L) MgCl₂, 1 mM 10x PCR buffer II, 1 mM dNTP mix, 1 U/μL RNase Inhibitor, 2.5 U/μL MuLV reverse transcriptase, and 2.5 μM (μmol/L) anchored oligo(dT)18 and 30 μM random hexamer as RT primers. For the RT- control, reaction mix was inactivated at 95 °C for 5 min prior to sample addition and mock RT was performed. Twenty μL of PCR reactions were performed with Power SYBR®Green PCR mix (ABI) with 300 nM of each primer and 2 μL of RT product. QPCR was performed with an iCycler™ thermal Cycler (Bio-Rad, Hercules, CA, USA). The primers used in the qPCR are as follows. interleukin 8 (IL-8): GAGGGTTGTGGAGAAGTTTTTG (F), CTGGCATCTTCACTGATTCTTG (R); interleukin 1 beta (IL-1β): GTGCTGAATGTGGACTCAATCC (F), ACCCTAAGGCAGGCAGTTG (R). The Beta-actin (β-actin) primers and method used for the RNA-specific PCR have been previously described (1,7).

In vitro translation

In vitro translation reactions were performed using the TNT Coupled Reticulocyte Lysate System (Promega, Madison, WI). Briefly, 7 μL of purified salivary RNA, corresponding to 145 μL of supernatant saliva, was mixed with 17 μL of rabbit reticulocyte lysate, 0.5 μL of ³⁵S labeled methionine (Amersham Biosciences, Piscataway, NJ) and 0.5 μL of RNase inhibitor in a total volume of 25 μL in accordance with the manufacturer's instructions. This mixture was incubated at 30 °C for 3 hrs. After the incubation, RNase A (0.2 mg/mL) was added and incubated at 30 °C for another 10 min. The reaction products were separated on an 8-16 % SDS-PAGE and visualized by Storm 840 (Amersham Bioscience). RNase cocktail (Ambion Inc., Austin, TX)-treated samples were prepared by substituting DNase I with RNase cocktail during the on-column DNase I digestion step.

Results

Salivary RNA

Supernatant phase of saliva from 10 healthy donors were pooled to generate enough total RNA for cDNA library construction. The quality of the pooled salivary RNA is shown as a virtual gel in Figure 1A which is a typical visualization of the extracted supernatant saliva nucleic acids using the Agilent 2100 Bioanalyzer (see Materials and Methods for an explanation of the virtual gel). For a comparison, we extracted nucleic acids from the human cell line MCF-7 in a similar fashion (Figure 1B). The salivary RNA is degraded (RNA integrity number, RIN < 2.5), with an average fragment length around 100 nt and an apparent maximum length around 1500 nt. Two separate peaks were observed in the electropherograms of untreated and RNase-treated samples (* and ** in Fig. 1A). The larger fragment peak is likely to be DNA contamination since DNase treatment removed this smear and RNase treatment did not. The RNase treatment of the samples was effective, as judged by the complete disappearance of the 28S and 18S bands in the cell line RNA preparation (Fig. 1B). The lower peak in Figure 1A is likely to consist of both DNA and RNA fragments, as there are residual bands after DNase and RNase digestions. Two additional control experiments were performed on the DNase I-treated pooled salivary RNA to ensure 1) there is no genomic DNA contamination and 2) there is genuine mRNA in the saliva. To rule out the presence of genomic DNA, RT-qPCR was performed for 3 NSCT RNAs; IL-1 β , IL-8, and β -actin. Figure 2A shows that only RT+ samples show the amplification of PCR products. RT- samples on the other hand, showed no amplification product. Next we demonstrated that there is genuine mRNA in the pooled saliva by performing *in vitro* translation with the pooled saliva RNA. The *in vitro* translation was carried out with rabbit reticulocyte in the presence of ³⁵S methionine. The result in Figure 2B shows the autoradiogram of *in vitro* translation products of pooled saliva samples that are untreated, DNase-treated or RNase-treated and ran on the polyacrylamide gel electrophoresis. Both untreated (lane 1) and DNase treated (lane 2) saliva samples showed protein bands after translation, indicating the presence of intact RNA in these samples. This is confirmed by the complete absence of translation products in the RNase treated sample (lane 3). These data clearly shows that there is no genomic contamination in our samples, and genuine mRNA is present in saliva.

Salivary cDNA library analysis

A cDNA library was constructed using the pooled total RNA. Pooling from 10 healthy subject was necessary to yield sufficient amount of total salivary RNA since each subject yielded ~0.5 μ g of RNA per donation (~5mL of supernatant saliva). In order to detect all of the RNAs present in saliva, poly(A) tail was added to the salivary RNA using bacterial poly(A) polymerase, and oligo dT primers was used for first strand cDNA synthesis. This approach was successfully utilized previously where the poly(A) tail was added to influenza virus RNA for oligo dT-based cDNA synthesis (9). The sub-cloned plasmids were transformed into *Escherichia coli*. The primary library has a titer of 1.3×10^6 cfu/mL and 1×10^8 cfu/mL after amplification with 95 % transformants containing a cDNA insert.

We randomly picked colonies from a plated library and 117 cDNA insert containing colonies were sequenced. Of these 117, 85 were sequenced bidirectionally and 32 were sequenced in one direction. Table 1 summarizes the sequencing results. For the corresponding parental RNA identification, the NCBI BLAST search was used (10) and only those that had the highest scores with minimum BLAST score of 100 were selected. The BLAST results indicated that all of the recombinant insert sequences are of human origin with 104 clones matched to nucleus-encoded mRNAs, 4 matched to mitochondria-encoded mRNAs, 2 matched to mitochondria-encoded 16S rRNA, and 7 matched to human sequences that have no known quotation. Also, 7 NSCT were found. Of these NSCT, translation elongation factor alpha 1 (EEF1A1) was found

four times, 2 NSCT were found twice, and 4 NSCT were found once (see Supplemental Data). Six genes that are not the member of NSCT were also found twice. None of the 117 clones were duplication products as comparisons of 5' and 3' ends of these clones showed that they are indeed independent clones of the same genes, not duplication products. Eighteen genes were reported to have pseudogenes as determined by Online Mendelian Inheritance in Man (OMIM) and GeneCards web searches suggesting a potential genomic DNA contamination of our salivary RNA samples. However, a direct comparison of the cDNA clone sequences to parental and pseudogene sequences showed that 17 of 18 salivary cDNA sequences aligned only with parental mRNA sequences and showed disparities with the pseudogene sequences. The only gene that was not verified was ribosomal protein S2 because of its very large number of pseudogenes. These data suggest that most of the RNAs in supernatant saliva are genuine human mRNAs and particularly high abundant ones are nucleus-encoded human mRNAs, and NSCT.

Salivary RNA degradation status

To understand the degradation status of salivary RNA, the exact locations of cDNA inserts on the respective full length parental mRNA were analyzed. The insert sequences of 62 clones that contained nucleus-encoded mRNA sequences and sequenced bidirectionally were used for this analysis (see Supplemental Data for accession numbers and gene names of these clones). Figure 3 shows the relative length and positions of cDNA inserts compared to their respective parental mRNAs. It appears that there are only 5 cDNA inserts that show over 90 % in length compared to their parental mRNAs (arrowed), suggesting that most of the salivary mRNAs are partially degraded. Figure 3 also shows that in general, there are more 3' fragments compared to the 5' fragments. This might be due to more degradation activities toward the 5' end. It should be noted, however, that the 5' regions are usually underrepresented during the cDNA library synthesis because once the reverse transcriptase binds to the oligo-dT during the RT reaction, not all the reverse transcriptases can transcribe to the 5' end of the RNA. These data suggest that most of the salivary RNAs are partially degraded and the degradation pattern appears to be random with potentially higher degradation activity toward the 5' end.

We also investigated the potential correlation between the length of the parental mRNA and the length of cDNA inserts. If endonucleases are primarily involved in salivary RNA degradation, we would expect to see longer salivary transcripts being cleaved more often than the shorter transcripts. Figure 4 shows the correlation between the lengths of salivary mRNA expressed as a percentage of their respective full length parental mRNA. It shows that salivary mRNAs tend to be close to full length if the respective parental mRNA is shorter, and are more severely degraded if the respective parental mRNA is longer. The average length of the parental mRNAs is ~3000 nt and the average length of the corresponding cDNA inserts is ~1300 nt. This suggests that on average, salivary mRNA represents 42% of its respective full length mRNA and those long transcripts have an increased chance of being degraded than shorter transcripts.

Discussion

For RNA in plasma, 5' fragments are more abundantly present compared to 3' fragments (11). Therefore, clinical usages of plasma RNA will likely focus on the 5' region of RNA for detection. For a similar reasoning, it is important to understand the degradation status of salivary RNA because of its application on oral cancer and potentially other disease detection (1,4,6). To measure the degradation status of salivary RNA, we created a cDNA library from pooled supernatant phase of saliva from 10 healthy subjects and analyzed 117 random colonies by sequencing. Our data indicate that most of the supernatant saliva RNAs are genuine mRNAs of human origin and many of these are present in partially degraded forms. We did not detect

any bacterial RNA in saliva probably due to removal of bacteria during the centrifugation of the saliva.

It should be noted that the amount of mRNA in saliva is so low that we have not been able to use mRNA specific RT-PCR technologies for salivary mRNA detection (12). However, we have shown by three independent methods that our salivary RNA samples are free of genomic DNA contamination. First, we showed that even though many of the genes identified in the salivary cDNA library have pseudogenes, sequence comparisons to the parental RNA showed no presence of DNA pseudogene. Second, we showed that reverse transcription step is required for PCR product amplification during the RT-qPCR reaction. Lastly, by means of *in vitro* translation, we showed that RNase treatment of the salivary RNA samples completely abolished the translation products. However, it is important to note that DNase I treatment of salivary RNA samples is required to completely remove the genomic DNA since we observed RT-qPCR signals from DNase I-untreated samples (data not shown). These data showed that our cDNA library is true representation of human salivary RNA.

Our data suggest that endonucleases are largely responsible for the partial degradation of salivary RNA. This is consistent with our previous RT-PCR result where we showed that RT-PCR that generates long PCR amplicons which span nearly the full length of their respective mRNAs are not always detected in saliva, whereas short amplicon products are always detected suggesting that most of the salivary mRNAs are partially degraded (7). Knowing that the oligo dT-based RT can lead to 3' bias, data in Figure 3 suggest that, in general, salivary RNA degradation appears to be random if not slightly more degradation takes place at the 5' end. RT-qPCR analysis of salivary β -actin and IL-8 mRNAs with primers that target 5', middle, and 3' ends showed that β -actin showed no preferential degradation but IL-8 mRNA showed lower amount of 3' region compared to the 5' region (7). Therefore, it is possible that different mRNAs may experience different types of degradation depend on their sequences and secondary structure formation. Nonetheless, we did not detect significant degradation toward one direction as seen in the plasma (11) which is a typical pattern for exonucleolytic RNA digestion. Also, we observed a correlation between the length of the full length parental mRNA and the degradation status of respective salivary RNA. We detected more degradation as the parental mRNA gets longer. This is consistent with endonuclease activity because the chance of endonucleolytic cleavage is directly proportionally to the length of the target RNA.

In conclusion, our results indicate that the supernatant saliva contains genuine human mRNAs. These mRNAs are partially degraded and there is no preferential degradation at either ends. Our study also presented a useful resource for future basic and translational studies of transcriptomes in human saliva.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Lauren Kimlin for careful reading of our manuscript. Research described in this article was supported by NIH grants RO1 DE15970, UO1 DE15018, UO1 DE17790 and UO1 16275 to D.T.W. Noh Jin Park was supported by an NIH NRSA training grant T32 DE07296.

Abbreviations

NSCT

normal salivary core transcriptome

RT-qPCR

reverse transcriptase quantitative polymerase chain reaction

IL-8

interleukin 8

IL-1 β

interleukin 1 beta

 β -actin

beta-actin

References

1. Li Y, Zhou X, St John MA, Wong DT. RNA profiling of cell-free saliva using microarray technology. *J Dent Res* 2004;83(3):199–203. [PubMed: 14981119]
2. Juusola J, Ballantyne J. Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. *Forensic Sci Int* 2003;135(2):85–96. [PubMed: 12927408]
3. Nussbaumer C, Gharehbaghi-Schnell E, Korschneck I. Messenger RNA profiling: a novel method for body fluid identification by Real-Time PCR. *Forensic Sci Int* 2006;157(23):181–6. [PubMed: 16289614]
4. Gomes MA, Rodrigues FH, Afonso-Cardoso SR, Buso AM, Silva AG, Favoreto S Jr, et al. Levels of immunoglobulin A1 and messenger RNA for interferon gamma and tumor necrosis factor alpha in total saliva from patients with diabetes mellitus type 2 with chronic periodontal disease. *J Periodontal Res* 2006;41(3):177–83. [PubMed: 16677285]
5. Li Y, St John MA, Zhou X, Kim Y, Sinha U, Jordan RC, et al. Salivary transcriptome diagnostics for oral cancer detection. *Clin Cancer Res* 2004;10(24):8442–50. [PubMed: 15623624]
6. St John MA, Li Y, Zhou X, Denny P, Ho CM, Montemagno C, et al. Interleukin 6 and interleukin 8 as potential biomarkers for oral cavity and oropharyngeal squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg* 2004;130(8):929–35. [PubMed: 15313862]
7. Park NJ, Li Y, Yu T, Brinkman BM, Wong DT. Characterization of RNA in saliva. *Clin Chem* 2006;52(6):988–94. [PubMed: 16601067]
8. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000;132:365–86. [PubMed: 10547847]
9. Romanova J, Katinger D, Ferko B, Vcelar B, Sereinig S, Kuznetsov O, et al. Live cold-adapted influenza A vaccine produced in Vero cell line. *Virus Res* 2004;103(12):187–93. [PubMed: 15163508]
10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403–10. [PubMed: 2231712]
11. Wong BC, Chiu RW, Tsui NB, Chan KC, Chan LW, Lau TK, et al. Circulating placental RNA in maternal plasma is associated with a preponderance of 5' mRNA fragments: implications for noninvasive prenatal diagnosis and monitoring. *Clin Chem* 2005;51(10):1786–95. [PubMed: 16081506]
12. Hurteau GJ, Spivack SD. mRNA-specific reverse transcription-polymerase chain reaction from human tissue extracts. *Anal Biochem* 2002;307(2):304–15. [PubMed: 12202248]

Figure 1.

Virtual gel derived from Agilent Bioanalyzer Analysis. Shown are electropherograms of RNA extracted from saliva (A) and MCF-7 breast cancer cells (B) before and after treatment with DNase or RNase as described in the materials and methods section. The salivary RNA was concentrated 10x before it was run on a Picochip. The two peaks that were observed in the Agilent profile of saliva are indicated with a * and ** on the right side of the saliva electropherogram. The cell line RNA was run on a Nanochip.

Figure 2A

Figure 2B

Figure 2.

A. RT-qPCR amplification plots for RT+ and RT- reactions of IL-1 β , IL-8 and β -actin mRNA.

B. *In vitro* translation of saliva RNA. Lane 1 – untreated salivary RNA. Lane 2- salivary RNA treated with DNase I. Lane 3- salivary RNA treated with RNase cocktail. The product of each *in vitro* translation reaction was run on an 8-16% SDS-PAGE protein gel.

Figure 3.

Degradation status of human mRNA in saliva. Results of bidirectional sequencing of colonies containing nucleus-encoded human mRNA sequences (total of 62 clones). Length of each cDNA insert is expressed to its respective full length parental mRNA. The positions of each bar represents the location of cDNA inserts on their respective full length parental mRNA. Arrows (left) indicate 3 cDNA clones that are >90% full length.

Figure 4.
Correlation between the length of salivary mRNA to the respective full length parental mRNA. The relative length of salivary mRNAs (cDNA inserts) are plotted against its respective full length parental mRNA.

Table 1- Summary of cDNA library sequencing

Colonies sequenced	117
Sequences of human origin	117
Nucleus-encoded mRNA	104
Mitochondria-encoded mRNA	4
Mitochondria-encoded 16S rRNA	2
Unknown regions	7
NSCT	7
Genes Identified more than once	9