Saliva as a sample type for the evaluation of miRNA biomarkers using small RNA sequencing

ORAcollect™•RNA

Christina Dillane, Mike Tayeb
DNA Genotek Inc., Ottawa, Ontario, Canada

Introduction

Sequencing of salivary microRNA (miRNA) is rapidly becoming a means to explore disease biomarkers. Saliva sampling offers a key advantage over other sample types because it is non-invasive and easy to self-collect; although, salivary RNA is prone to degradation if not stabilized immediately. Saliva collected using ORAcollect™•RNA (ORE-100) self-collection kits overcomes this obstacle and is a scalable way to collect and stabilize high-quality miRNA.

In this study, we demonstrate that stabilized ORAcollect™•RNA saliva samples provide RNA of sufficient quality for miRNA sequencing when libraries are prepared using the QIAseq® miRNA Library Kit (QIAGEN). Data analysis using the GeneGlobe-integrated RNA-seq Analysis Portal (QIAGEN)[1] enables fast and easy detection of miRNA biomarkers, including miRNAs with reported links to cancer,[2,3] head injury,[4] mental health[5] and endometriosis.[6]

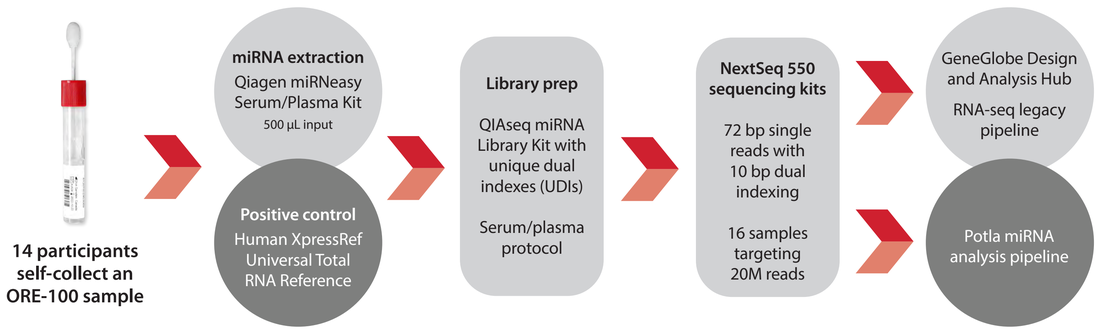

Figure 1. *Workflow of collection, extraction, library prep, sequencing and analysis.*

Methods

Saliva was collected from 14 participants (aged 27-50) using the DNA Genotek ORAcollect™•RNA device and RNA was extracted using the miRNeasy Serum/Plasma kit (QIAGEN). Two ORAcollect™•RNA aliquots of 250 µL were prepared and loaded onto the column sequentially. Sequencing libraries were generated using the QIAseq miRNA Library Kit with UDIs (QIAGEN) without normalization of the input RNA. No miRNA library gel extraction was performed. Ten nanograms (10 ng) of XpressRef Universal Human RNA Reference (QIAGEN) acted as a positive control, and a negative no template control (NTC) was included to assess contamination during library prep. Libraries were quantified by PicoGreen™ (Invitrogen) and library size was assessed by TapeStation (Agilent). Libraries were sequenced on the NextSeq 550 (Illumina) targeting a mean of 20 million reads per sample. FASTQ files were uploaded to GeneGlobe Design & Analysis Hub (QIAGEN) and annotated using miRBase and piRNAdb databases. The RNA-seq Legacy Analysis Pipeline (QIAGEN)[7] was used to evaluate contaminating RNA reads. Additionally, a miRNA-seq analysis pipeline based off the protocol by Potla et al.[8] was used to validate GeneGlobe results. Read counts were normalized by counts per million.

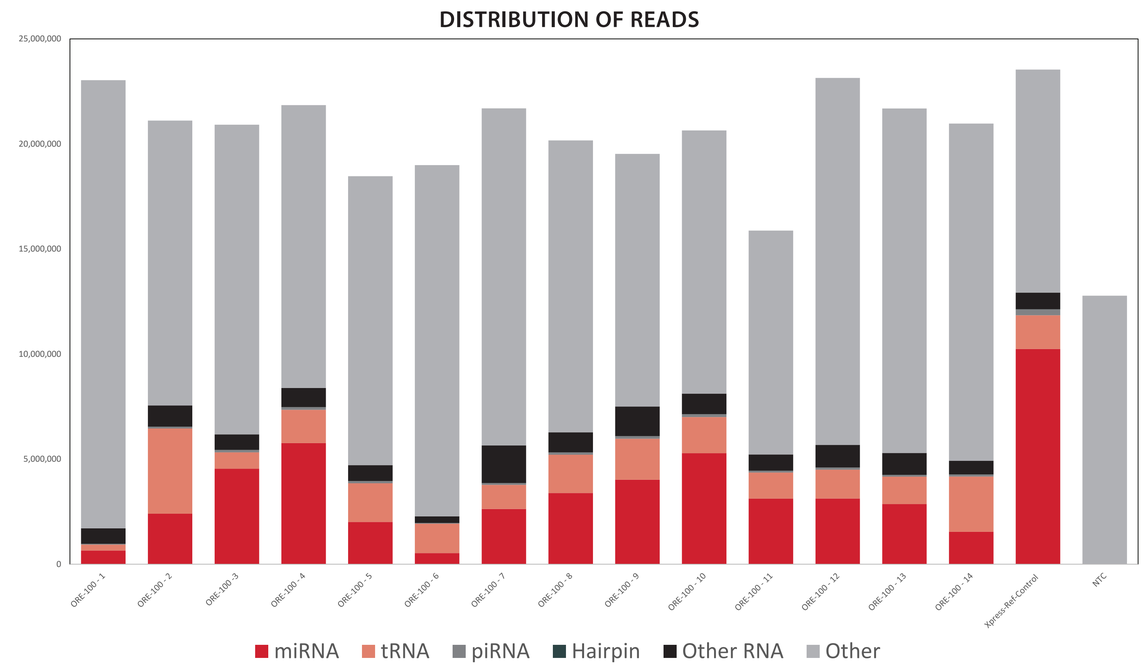

Figure 2. *Reads that were annotated with miRBase or piRNAdb ranged from 10% to 50% of the total reads per saliva sample. No miRNAs were found in no-template controls. The reads that are categorized as "Other" include reads that did not contain a UMI or adaptor, were too short, mapped to the human genome or were un-mappable.*

Results

QIAseq miRNA library prep

Sequencing libraries were successfully prepared for all samples. In all cases, library concentrations exceeded the 4 nM threshold for sequencing. Library sizes were evaluated using the TapeStation and all libraries were between 158-176 bp in size, including the XpressRef control.

Data analysis

GeneGlobe Design and Analysis Hub was used to investigate sequencing data quality, identify the number of pre-miRNA annotation records, and evaluate the expression of miRNAs in the samples. The mean number of reads per sample was 20.7 million and Q scores of Unique Molecular Index (UMI) reads are above 30 demonstrating a high sequencing performance in all libraries. On average, 47% of the reads per sample did not contain a UMI, which was observed across both the saliva and control samples. These consist of reads that are too short, are missing adapters or are UMI defective.

Of the reads containing UMIs, 10%-51% of the reads in each saliva sample were annotated with miRBase or piRNAdb, compared to 69% annotated UMI reads for the XpressRef control. The range of annotated reads in saliva samples is expected given the variable miRNA input that is inherent of saliva samples.

The RNA-seq Legacy Analysis Pipeline was used to further characterize the non-annotated UMI reads in each sample. The bulk of non-annotated UMI reads (3.7M) were found to not be mappable; although, a further average of 1.7M mapped to the human genome, 0.84M were human mRNA and rRNA, and 1.7M mapped to RNA. (Figure 2).

While the NTC contained several sequencing reads, the vast majority did not contain an adapter which is consistent with adapter dimers.

Notably, the majority of ORAcollect™•RNA samples include in excess of 900 miRbase pre-miRNA annotation records, while the Xpress Ref Control, comprised of 20 different human adult and fetal major organs, contains only a slight increase of 1,112. (Table 1).

	miRBase (Homo sapiens)	piRNA (piRNAdb_hsa.v1_7_6)
Records in source	1,917 (pre-miRNAs)	27,700
ORE-100 - 1	835 (43.56%)	1,484 (5.36%)
ORE-100 - 2	800 (41.73%)	1,142 (4.12%)
ORE-100 - 3	967 (50.44%)	1,445 (5.22%)
ORE-100 - 4	963 (50.23%)	1,200 (4.33%)
ORE-100 - 5	794 (41.42%)	1,298 (4.69%)
ORE-100 - 6	676 (35.26%)	1,227 (4.43%)
ORE-100 - 7	891 (46.48%)	1,103 (3.98%)
ORE-100 - 8	939 (48.98%)	1,284 (4.64%)
ORE-100 - 9	941 (49.09%)	1,346 (4.86%)
ORE-100 - 10	942 (49.14%)	1,335 (4.82%)
ORE-100 - 11	928 (48.41%)	1,283 (4.63%)
ORE-100 - 12	952 (49.66%)	1,563 (5.64%)
ORE-100 - 13	967 (50.44%)	1,285 (4.64%)
ORE-100 - 14	934 (48.72%)	1,711 (6.18%)
Xpress-Ref-Control	1,112 (58.01%)	3,345 (12.08%)

Table 1. *Number of pre-miRNA annotated records found in miRBase and piRNA base per sample.*

Comparisons

Recently, a retrospective cross-sectional miRNA sequencing study utilizing 1,225 saliva samples, primarily collected from ORAcollect™•RNA and Oragene™•RNA devices, was published and gives greater insight into the expected expression of salivary miRNAs from multiple cohorts.[9] We used expression profiles from the Sullivan et al. study and compared them to our own data to evaluate concordance between the two. After accounting for sequencing depth, we found a 3.5-fold increase in annotated miRNA reads in our samples, likely due to the addition of UMIs. This demonstrates that the QIAseq miRNA Library Kit performs well with ORAcollect™•RNA saliva samples.

The 4 most highly expressed miRNAs in each study were distinctive, though a 50% concordance was observed between the top 10 most expressed miRNAs in each study. Differences are likely due to the Sullivan study size, age range (0-35 vs. 27-50), participant demographics, or differing collection methods. The increased performance from UMIs in our study could also account for the observed differences.

The top 10 miRNAs were concordant between the Potla and GeneGlobe analysis methods; although, GeneGlobe consistently annotated a higher percentage of the top expressed miRNAs. This was expected as the Potla analysis contained an increased percentage of annotated records.

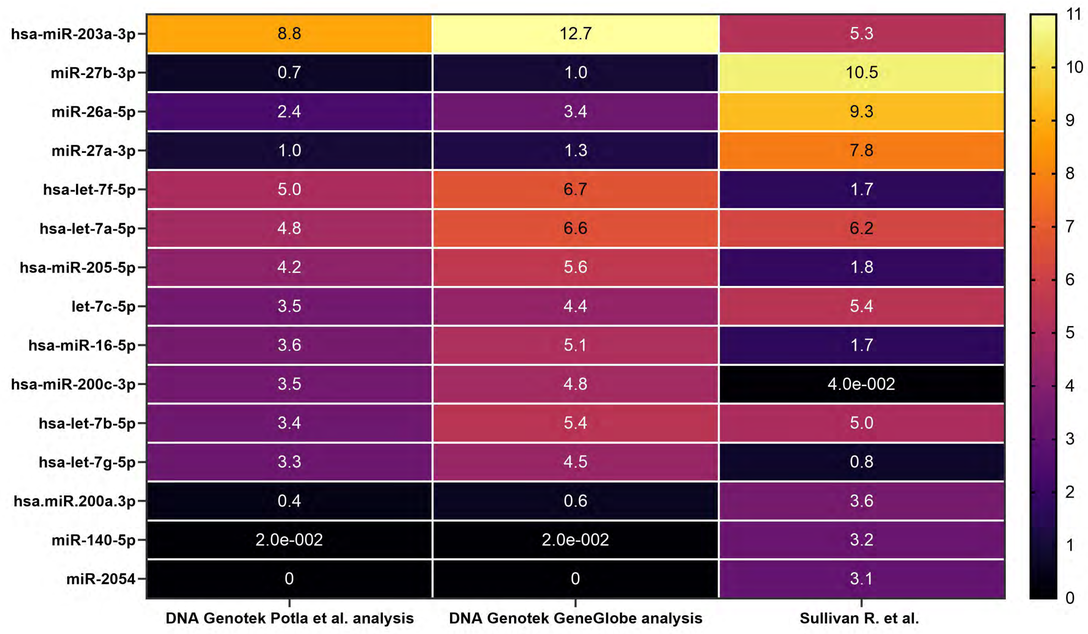

Figure 3. *The top 10 miRNAs were selected from each study and compared. The Potla analysis method was used in this study to validate GeneGlobe results. Although GeneGlobe consistently annotated a higher percentage (36%) pf the top miRNAs, results from the 2 analysis pipelines were comparable.*

Conclusions

- ORAcollect™•RNA devices stabilize miRNAs from saliva and provide sufficient yields for miRNA sequencing when using the QIAseq miRNA Library Kit with UDIs without the need for gel extraction of miRNA libraries. Avoiding a gel extraction during library prep significantly speeds up wet lab time.
- Using GeneGlobe Design and Analysis Hub, bioinformatic analysis can be easily completed by the researcher.
- The most prevalent annotated miRNA records identified in this study (Figure 3) have previously been identified as significant biomarkers for disease states such as endometriosis, depression, concussions or traumatic brain injury, and lung, head and neck cancers. Collection of tissue samples for these diseases is often invasive; therefore, the use of a non-invasive sample type such as saliva, with corresponding biomarkers, is crucial.
- ORAcollect™•RNA saliva provides a high number of annotated reads and covers up to 50% of the total number of pre-miRNA annotation records in miRbase.
- When compared to current literature, the most highly expressed miRNAs in each study are distinctive, though a 50% concordance between the top 10 miRNAs from each study is observed.
- The top 10 miRNAs were concordant between the Potla and GeneGlobe analysis methods signifying that GeneGlobe provides accurate results, though the Potla pipeline may detect additional rare miRNAs.
- These results demonstrate that ORAcollect™•RNA stabilized saliva is a suitable sample type for miRNA biomarker detection using small RNA sequencing.

Acknowledgments

We would like to acknowledge Jonathan Shaffer and Samuel Rulli from QIAGEN for their advice and insights throughout this study.

References

1 QIAGEN. GeneGlobe Data Analysis Center (RRID:SCR_021211). https://geneglobe.qiagen.com/us/analyze

2 Fadhil RS, Wei MQ, Nikolarakos D, et al. (2020). Salivary microRNA miR-let-7a-5p and miR-3928 could be used as potential diagnostic bio-markers for head and neck squamous cell carcinoma. *PLoS One*. 24;15(3):e0221779. doi: 10.1371/journal.pone.0221779.

3 Naakka E, Barros-Filho MC, Adnan-Awad S, et al. (2022). miR-22 and miR-205 drive tumor aggressiveness of mucoepidermoid carcinomas of salivary glands. *Front Oncol*, vol. 11. ISSN 2234-943X. doi: 10.3389/fonc.2021.786150

4 Di Pietro V, O'Halloran P, Watson CN, et al. (2021). Unique diagnostic signatures of concussion in the saliva of male athletes: the study of concussion in rugby union through microRNAs (SCRUM). *Br J Sports Med*, vol. 55. doi: 10.1136/bjsports2020-103274

5 Ferrúa CP, Giorgi R, Camerini da Rosa L, et al. (2019). MicroRNAs expressed in depression and their associated pathways: A systematic review and a bioinformatics analysis. *J Chem Neuroanat,* vol. 100. doi: 10.1016/j.jchemneu.2019.101650

6 Perricos A, Proestling K, Husslein H, et al. (2022). Hsa-mir-135a shows potential as a putative diagnostic biomarker in saliva and plasma for endometriosis. *Biomolecules*. 12(8):1144. doi: 10.3390/biom12081144

7 QIAGEN RNA-seq Analysis Portal 4.0 (QIAGEN, Aarhus, Denmark), https://rnaportal.qiagen.com

8 Potla P, Ali SA, Kapoor M, et al. (2021). A bioinformatics approach to microRNA-sequencing analysis. *Osteoarthr and Cartil Open*, vol. 3, issue 1. ISSN 2665-9131. doi: 10.1016/j.ocarto.2020.100131

9 Sullivan R, Montgomery A, Scipioni A, et al. (2022). Confounding factors impacting microRNA expression in human saliva: methodological and biological considerations. *Genes*, 13, 1874. doi: 10.3390/genes13101874

Scan here for a digital copy of this poster

DNAgenotek™ *Superior samples • Proven performance*

www.dnagenotek.com • support@dnagenotek.com