A clinically validated human saliva metatranscriptomic test for global systems biology studies

Ryan Toma[1*], Ying Cai[1], Oyetunji Ogundijo[1], Stephanie Gline[1], Diana Demusaj[1], Nathan Duval[1], Guruduth Banavar[1], Momchilo Vuyisich[1]

[1]Viome, Inc., Bothell and Bellevue, WA/New York, NY, USA, viome.com

*Corresponding author: Ryan Toma, ryan.toma@viome.com

Abstract

We report here the development of a high throughput, automated, inexpensive, and clinically validated saliva metatranscriptome test that requires less than 100 microliters of saliva. The RNA in the samples is preserved at the time of collection. Samples can then be stored and transported at ambient temperatures for up to 28 days without significantly affecting the test results. Another important feature of the sample preservative is that it inactivates all classes of pathogens, enabling safe transportation globally. Given the unique set of convenience, low cost, safety, and technical performance, this saliva metatranscriptomic test can be integrated into longitudinal, global scale, systems biology studies that will lead to an accelerated development of precision medicine diagnostic and therapeutic tools.

Method Summary

This report describes an important improvement to saliva transcriptome analysis. While current methods are complicated and expensive, the method reported here includes at-home sample collection, global shipping at ambient temperatures and pathogen inactivation at the point of collection; it uses fully automated, clinically validated and licensed laboratory and bioinformatic analyses.

Introduction

Chronic diseases are the leading cause of morbidity and mortality globally. Yet, the majority of them have unknown etiologies, with weak genetic contributions. The human microbiome and alterations to human and microbial gene expression patterns have been identified as the underlying driver of many chronic diseases. To identify the etiology of chronic diseases and develop more effective preventative measures, a comprehensive gene expression analysis of the human body and associated microbiomes is needed, in addition to accurate strain-level taxonomy of all microorganisms. Saliva is easy to obtain and contains a diverse microbiome involved in many fundamental aspects of human physiology [1–3]. Saliva is also a suitable clinical specimen for the identification of human pathogens, such as SARS-CoV-2 [4]. Human transcripts in saliva can also produce an informative snapshot of human gene expression and their possible roles in human health and disease [5,6].

Human saliva contains a vast number of both commensal and potentially pathogenic microorganisms performing a wide variety of metabolic functions, with beneficial and

potentially harmful effects to the host [7]. Human saliva also contains a substantial amount of human transcripts that can reveal specific interactions between the microbiome and human tissues [8]. Therefore, saliva presents an opportunity to investigate the oral microbiome and oral human transcriptome and their roles in disease pathogenesis [7,9]. In the era of precision medicine, methods to non-invasively assess an individual's human and microbial gene expression levels can reveal predictive markers of disease, inform the best choice of therapy, and enable the development of novel treatment options. For example, using systems biology approaches to interrogate the transcriptome is now a standard strategy to stratify cancer patients and select the best available therapies [10–12]. Methods to analyze the transcriptome are becoming invaluable in the understanding of chronic diseases with unknown etiologies, but large scale adoption of metatranscriptomic analysis has not yet been possible due to the lack of low-cost and scalable methods.

The saliva microbiome has been associated with numerous chronic diseases including Alzheimer's Disease [13], Parkinson's Disease [14], Autism Spectrum Disorder [5], cancers [15], cardiovascular disease [16], diabetes [17], obesity [18], and autoimmune disorders [19,20]. As an example of the direct effect of salivary microorganisms on disease, *Fusobacterium nucleatum* originating in the oral cavity has been shown to promote colorectal cancer development [21,22]. The oral microbiome has also been shown to have a strong influence on the gut microbiome and the immune system, which are involved in a variety of different chronic diseases [23]. In addition to the clear role of the saliva microbiome in human health and disease, there is substantial information to be gained from human gene expression patterns in saliva. Alterations in salivary human gene expression and epigenetic markers have been observed in a range of disorders such as Autism Spectrum Disorder [6], Parkinson's Disease [24], and Traumatic Brain Injury [25]. Saliva is clearly an informative sample type that can offer highly predictive insights into human health and disease and could provide information about the molecular mechanisms of disease onset and progression.

Despite the abundance of literature showing clear connections between the oral microbiome, human gene expression, and chronic diseases, most of the evidence is based on 16S ribosomal RNA gene sequencing or metagenomic (shotgun DNA) sequencing methods. These approaches limit the biomarker potential and actionability of interventions that can be made. 16S rRNA methods can provide taxonomic resolution to the genus level at best [26,27] and do not measure the biochemical functions of the microorganisms [28] or distinguish between living and dead microorganisms. In addition, traditional 16S rRNA gene sequencing excludes an analysis of human genes and excludes many bacteria and archaea, and all eukaryotic microorganisms and viruses [29], resulting in a limited view of the saliva microbiome ecosystem and minimizing the ability to discover novel microbial or human derived biomarkers.

Metagenomic sequencing can provide strain-level resolution of all DNA-based microorganisms [28], but it does not detect any RNA viruses (human, plant, and bacteriophages). Importantly, metagenomic analysis does not measure gene expression, it can only identify the *potential* biochemical functions of a microbial community. This is a disadvantage for studying disease states, which are affected more by gene expression and the biochemical outputs of the microbiome than the taxonomic identity of the microorganisms. This has already been demonstrated in several gut and oral microbiome studies [30,31]

The lack of comprehensive taxonomic classifications and the inability to investigate gene expression has limited the utility of the conventional methodologies (16S and metagenomics) in

investigating the role of the saliva microbiome and human transcriptome in health and disease [32]. This is particularly apparent in the context of microbial functions, which have been shown to be more informative in disease states than taxonomy alone [30]. This showcases the utility of metatranscriptomic analysis of saliva, as it presents a more comprehensive and relevant snapshot of the microbial and human transcripts compared to traditional methodologies.

Saliva metatranscriptomic tests have already been used by our laboratory to create comprehensive and accurate diagnostic indicators of oral cancer [33]. Importantly, these predictive models rely on microbial taxonomy, microbial gene expression, and human gene expression signatures. Only through the use of a comprehensive sample analysis method, such as metatranscriptomics, can such a wide variety of data be analyzed.

In addition to their utility in biomarker discovery, functional outputs are also critical in developing a comprehensive understanding of disease etiology and subsequent treatment options [34]. Gene expression analyses can also provide information about specific microbial pathogenicity that is lost when simply looking at the microbiome composition. *Porphyromonas gingivalis* for example can be present in small amounts, but have highly expressed functions with a strong negative impact on the microbial community and ultimately the host [30,35].

Metatranscriptomic analyses address the limitations of both 16S rRNA gene sequencing and metagenomics. Metatranscriptomics provides insights into the biochemical activities of the microbiome by quantifying microbial gene expression levels, allowing for the assessment of biochemical pathway activities, while providing strain-level taxonomic resolution for all metabolically active microorganisms and viruses [36,37]. Metatranscriptomics also provides a comprehensive analysis of human genes in saliva. To date, metatranscriptomic methods have been limited due to the cost and complexity of both laboratory and bioinformatic methods [38]. Effective RNA preservation has been challenging, traditionally requiring a cold chain that is expensive and complicated. Another challenge of working with RNA is that about 97% of total RNA is ribosomal (rRNA) and 1% transfer RNA (tRNA), whereas messenger RNA (mRNA), which is the most informative RNA species, makes up the remaining 2%. By removing the non-informative prokaryotic and eukaryotic rRNAs and tRNA, more valuable transcriptome data can be generated with less sequencing depth [36,39], resulting in reduced per-sample sequencing costs and higher data resolution.

Here we present a comprehensive (sample collection-to-result) method for the quantitative metatranscriptomic analysis of saliva samples that can easily be applied to clinical studies and trials globally. The method is automated, high-throughput, inexpensive, clinically validated, and includes a fully automated bioinformatics suite for strain-level taxonomic classification of all microorganisms and human genes and their quantitative gene expression levels. An RNA preservation buffer mixed with the saliva sample at the point of collection inactivates all pathogens (bacteria, fungi, and viruses), preserves RNA integrity, enabling safe transportation globally at ambient temperatures for up to 28 days.

Methods

Ethics Statement

All procedures involving human subjects were performed in accordance with the ethical standards and approved by a federally-accredited Institutional Review Board (IRB) committee. Informed consent was obtained from all participants, who were residents of the USA at the time.

Sample collection

Figure 1. At-home saliva collection device. The insert contains the sample preservative that inactivates all types of pathogens, and preserves both RNA and DNA for 28 days at ambient temperature. Once saliva is deposited inside the tube, the funnel and insert are removed, which releases the preservative and mixes it with the specimen.

Saliva can be collected in any format that enables mixing of the Viome RNA preservative buffer (RPB) with the sample as soon as it is collected. Viome has developed a patented and convenient sample collection device that allows anyone, even children, to easily collect sufficient volumes of saliva and preserve it for metatranscriptomic analysis (Figure 1).

Participants were instructed to fast for 8 hours prior to collection, to collect the samples within one hour of waking up, and to gently rinse their mouth with water for ten seconds before collecting the sample.

Method Validation

The performance of the saliva method was assessed by determining the method precision, sample stability, and longitudinal changes in saliva. Precision was assessed in four technical replicates of saliva from ten donors. Sample stability was assessed by storing samples for 7 and 28 days at room temperature (RT) compared to the technical replicates analyzed on the day of collection (day 0). A subset of replicate samples were shipped and returned to Viome to assess the impact of shipping on sample stability. All sample stability analyses were performed with four technical replicates between three donors. Longitudinal changes and helinger distances in saliva were assessed by collecting saliva weekly for five weeks from eight donors.

To develop a comprehensive saliva cohort, saliva samples were collected from 1,102 unique individuals. For the saliva cohort, the average age was 48 years (range: 12-88 years old) with 436 males, 665 females, and 1 participant identifying as another gender.

Laboratory analysis

Saliva samples were lysed by bead beating in the RNA preservative buffer (RPB), as previously described [40,41]. RNA was extracted using silica beads and a series of washes, followed by elution in molecular biology grade water. DNA was degraded using RNase-free DNase. Prokaryotic and human rRNAs were removed using a subtractive hybridization method previously described [40]. Briefly, biotinylated DNA probes with sequences complementary to human and microbial rRNAs were added to total RNA, the mixture was heated and cooled, and the probe-rRNA complexes were removed using magnetic streptavidin beads. The remaining RNAs were converted to directional sequencing libraries with unique dual-barcoded adapters and ultrapure reagents. Libraries were pooled and quality controlled with dsDNA Qubit (Thermo Fisher Scientific) and Fragment Analyzer (Advanced Analytical) methods. Library pools were sequenced on Illumina NovaSeq instruments using 300 cycle kits.

Bioinformatic analysis

Viome's bioinformatics methods include quality control, strain-level taxonomic classification, microbial gene expression, and human gene expression characterizations. The quality control includes per-sample and per-batch metrics, such as the level of barcode hopping, batch contamination, positive and negative process controls, DNase efficacy, and number of reads obtained per sample. Following the quality control, the paired-end reads are aligned to a catalog containing ribosomal RNA (rRNA), human transcriptome, and 53,660 genomes spanning archaea, bacteria, fungi, protozoa, and viruses. Reads that map to rRNA are filtered out. Strain-level relative activities were computed from mapped reads via the expectation-maximization (EM) algorithm [42]. Relative activities at other levels of the taxonomic tree were then computed by aggregating according to taxonomic rank. Relative activities for the biological functions were computed by mapping paired-end reads to a catalog of 52,324,420 genes, quantifying gene-level relative activity with the EM algorithm and then aggregating gene-level activity by KEGG Ortholog (KO) annotation [43]. The identified and quantified active microbial species and KOs for each sample were then used for downstream analysis.

Data analysis

Statistical parameters, including transformations and significance, are reported in the figures and figure legends. To compare pairs of samples, we report Spearman correlations (which are invariant to absolute expression levels of the genes and only consider the similarity of ranked expression) and Hellinger distance (an appropriate distance measure for compositional data). Multiplicative replacement method [44] is employed to deal with missing values, the data is transformed with centered log-ratio (CLR) [45] and Spearman correlation coefficients were computed. CLR transformation is commonly done to reduce false discoveries due to the compositional nature of sequencing data. Statistical analyses were performed in Python.

Results

Saliva Cohort Metrics

For the cohort of 1,102 saliva samples, the average number of sequencing reads per sample was 24,353,310. On average, there were 487 species, 2,539 microbial KOs, and 5,174

human transcripts detected per sample. See supplementary tables 1-5 for a rank ordered list of all detected genera, species, strains, microbial KOs, and human transcripts in the saliva cohort.

Method precision

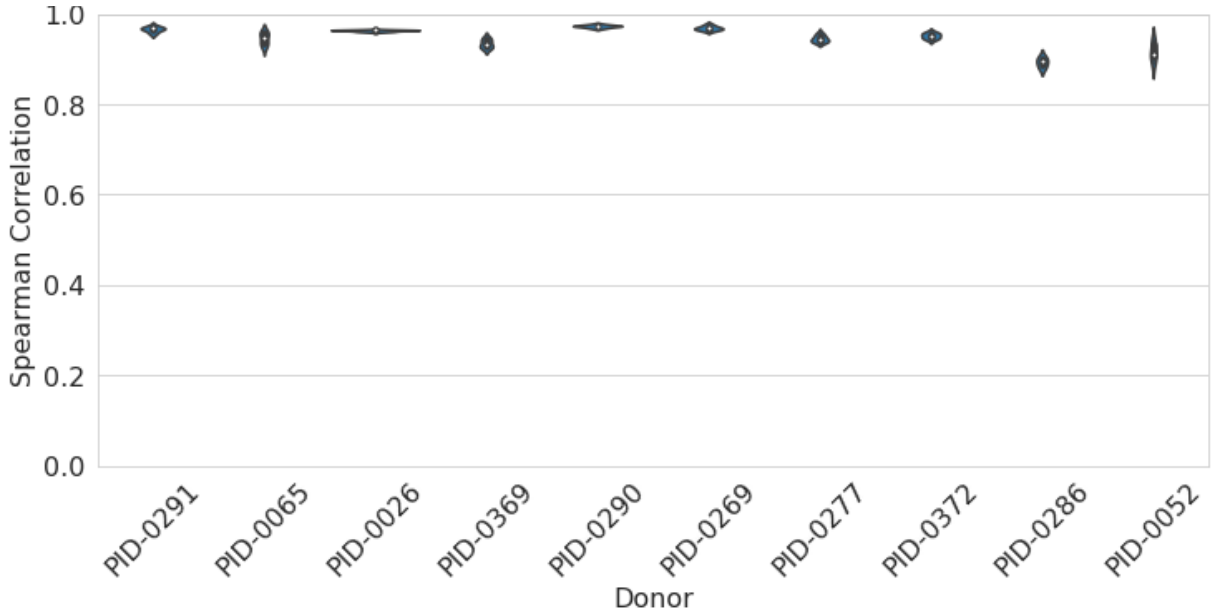
For longitudinal clinical studies in large populations that measure gene expression changes as a function of health and disease states, a method with high precision is extremely important, as it allows for the measurement of small changes in the levels of gene expression that are associated or predictive of chronic disease. To determine the precision of the metatranscriptomic method, Spearman correlations were calculated for microbial taxonomy (species) and microbial functions (KOs) for four technical replicates from ten participants (Figure 2a and 2b). For one participant (PID-0286), one replicate was determined to be an outlier with the log transformed effective sequencing depth (the number of sequencing reads aligned to microbial species) being greater than three standard deviations away from the mean and was removed from the analysis as it failed quality control. The method precision for microbial taxonomy and microbial functions is high, showing that the method is able to reliably reproduce the results (Figure 2a and 2b).

Human transcripts are present in low amounts in saliva and therefore require higher sequencing depth to analyze. Samples from four donors were re-sequenced with higher sequencing depth to determine the method precision for human genes. The effective sequencing depth (ESD, the number of reads aligned to human transcripts) needs to be about 1 million reads to yield high quality human gene expression data (table 1, Figure 2c). This demonstrates that the technology does allow for high precision human gene expression analyses from saliva, but requires higher sequencing depth.

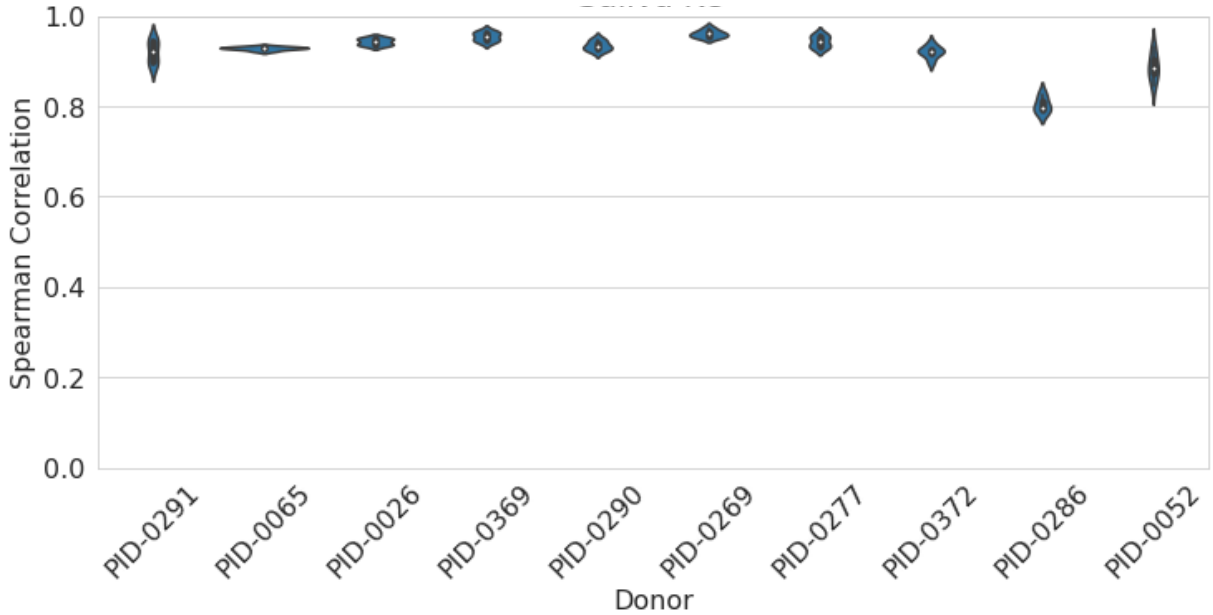
Table 1: Total reads and ESD of re-sequenced saliva samples for human gene expression precision analyses

Donor ID	Average Total Reads	Total Reads Standard Deviation	Average ESD	ESD Standard Deviation
PID-0026	73653434	38925007	1155273.803	592131.0757
PID-0290	110747517	27390256	1075291.804	381364.1083
PID-0291	84602795	45127258	1674916.386	861131.1174
PID-0372	87167604	17886807	1765820.818	586940.0628

A.


1.0
0.8
0.6
0.4
0.2
0.0
Spearman Correlation
PID-0291
PID-0065
PID-0026
PID-0369
PID-0290
PID-0269
PID-0277
PID-0372
PID-0286
PID-0052
Donor

B.


1.0
0.8
0.6
0.4
0.2
0.0
Spearman Correlation
PID-0291
PID-0065
PID-0026
PID-0369
PID-0290
PID-0269
PID-0277
PID-0372
PID-0286
PID-0052
Donor

C.

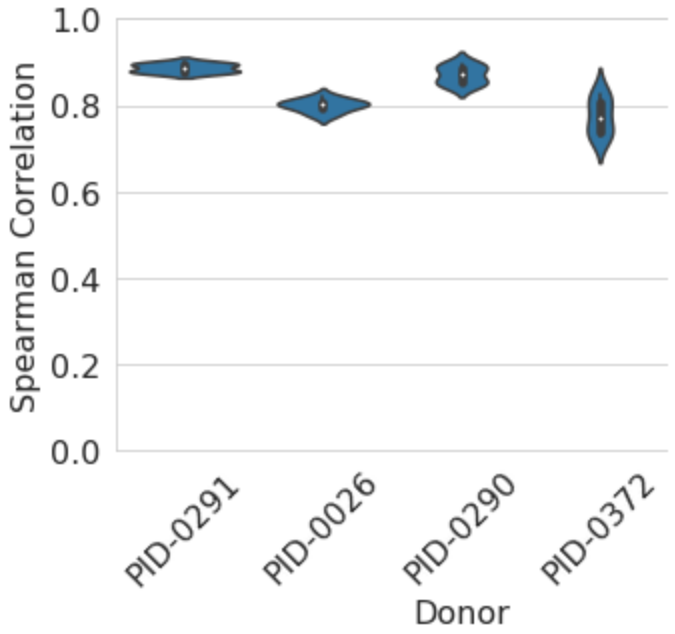


Figure 2. The method precision is shown comparing technical replicates from ten donors for microbial taxonomy (A), microbial functions, (B), and from four donors for human transcripts (C)

In addition to the Spearman correlations, two random technical replicates were compared from each participant for microbial taxonomy, microbial KOs, and human gene expression (Figure 3). These data show that Viome's metatranscriptomic saliva test is able to produce data with high precision.

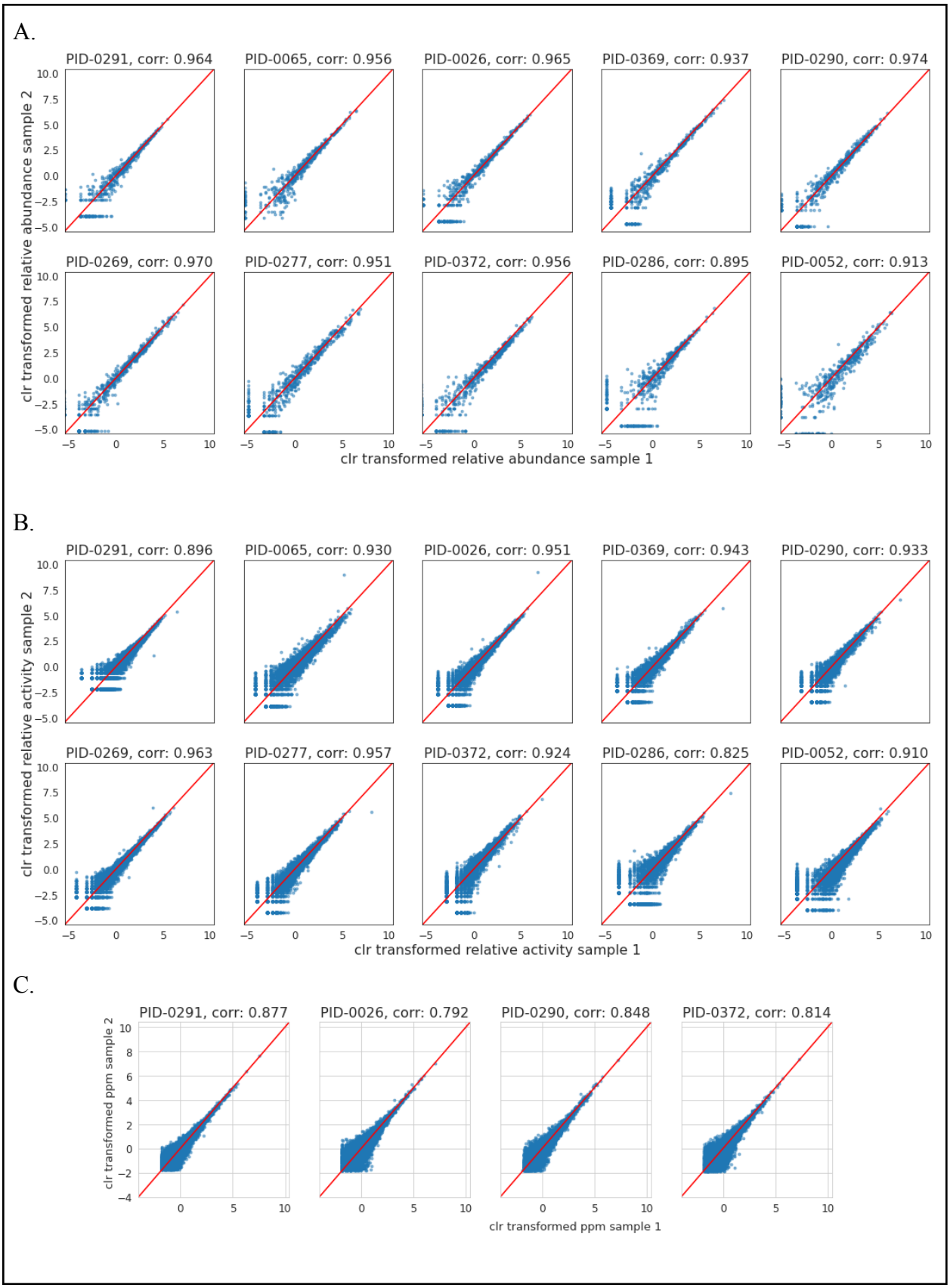

A.
PID-0291, corr: 0.964
PID-0065, corr: 0.956
PID-0026, corr: 0.965
PID-0369, corr: 0.937
PID-0290, corr: 0.974
PID-0269, corr: 0.970
PID-0277, corr: 0.951
PID-0372, corr: 0.956
PID-0286, corr: 0.895
PID-0052, corr: 0.913
clr transformed relative abundance sample 2
clr transformed relative abundance sample 1
B.
PID-0291, corr: 0.896
PID-0065, corr: 0.930
PID-0026, corr: 0.951
PID-0369, corr: 0.943
PID-0290, corr: 0.933
PID-0269, corr: 0.963
PID-0277, corr: 0.957
PID-0372, corr: 0.924
PID-0286, corr: 0.825
PID-0052, corr: 0.910
clr transformed relative activity sample 2
clr transformed relative activity sample 1
C.
PID-0291, corr: 0.877
PID-0026, corr: 0.792
PID-0290, corr: 0.848
PID-0372, corr: 0.814
clr transformed ppm sample 2
clr transformed ppm sample 1

Figure 3. Relative abundance of taxons and transcripts are shown between two technical replicates from ten donors for microbial taxonomy (A), microbial functions (B), and from four donors for human genes (C)

Sample stability

A major concern when performing metatranscriptomic analysis is the rapid and facile degradation of RNA, due to self-cleavage and action of ribonucleases [46]. To preserve RNA, our method uses an RNA preservative buffer (RPB) with multiple functions. RPB's detergents quickly dissolve lipid bilayers and inactivate all microorganisms, the denaturants halt all enzymatic activity by disrupting the tertiary structures of proteins, while other ingredients prevent RNA self-cleavage by keeping the 2' oxygen protonated. To validate the ability of RPB to preserve the saliva transcriptome at ambient temperatures, saliva samples were collected from three participants and stored at ambient temperature (72ºF, 22ºC) for 0, 7, and 28 days, with and without shipping. The shipping consisted of collecting the samples at the laboratory, sending them to an address via US Postal Service, then shipping them back to the laboratory using the same shipping method. Four technical replicates per condition were analyzed using our test, and microbial taxonomy and microbial functions were compared among all samples. The correlation between microbial species in saliva samples in all conditions is very high, with Spearman correlation coefficients above 0.928 for all conditions tested (Figure 4a). The correlation between microbial functions in saliva samples in all conditions is also very high, with Spearman correlation coefficients above 0.901 for all conditions tested (Figure 4b). This shows that Viome's saliva metatranscriptomic analysis method can adequately preserve RNA for up to 28 days at ambient temperature, inclusive of sample shipping.

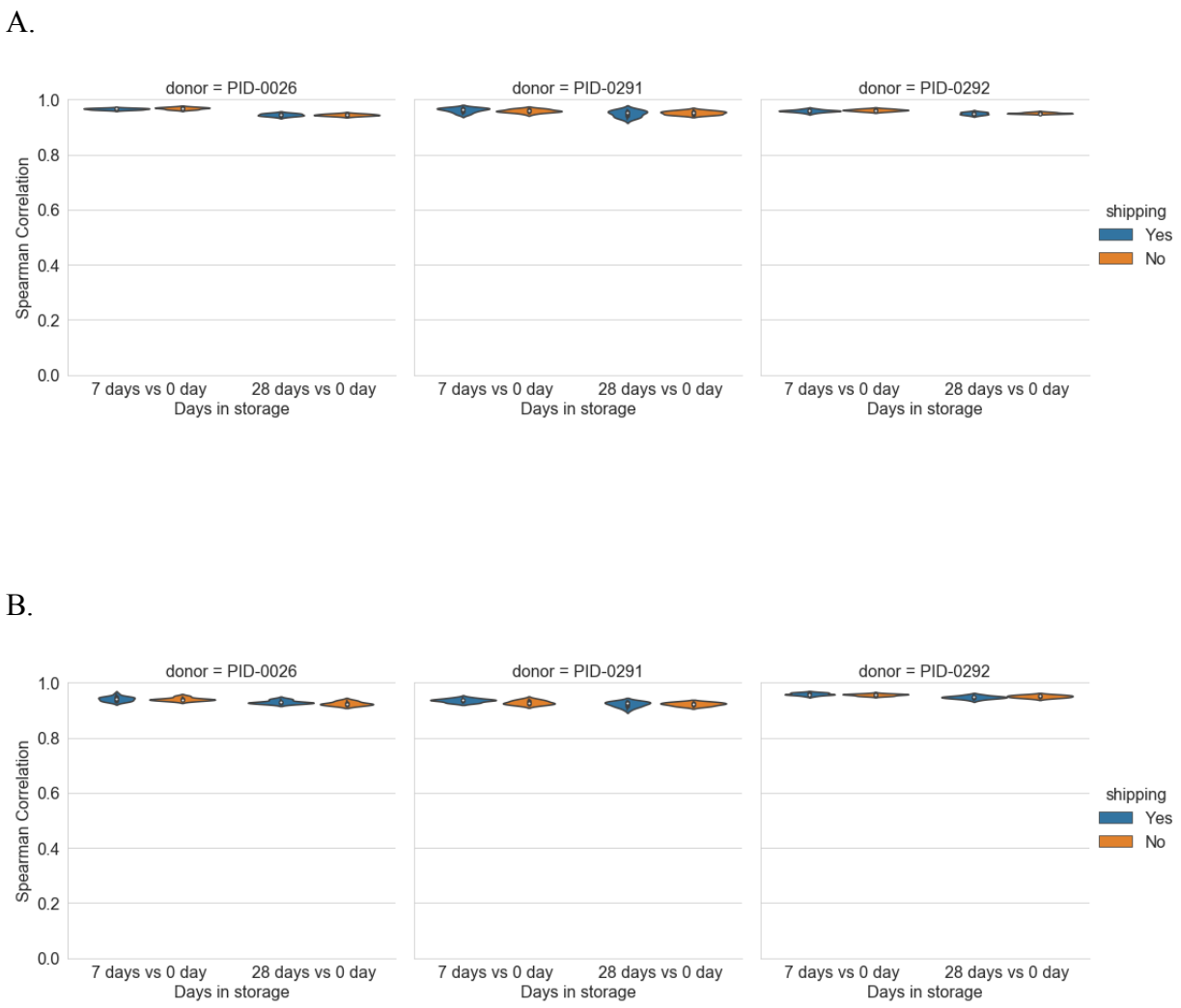


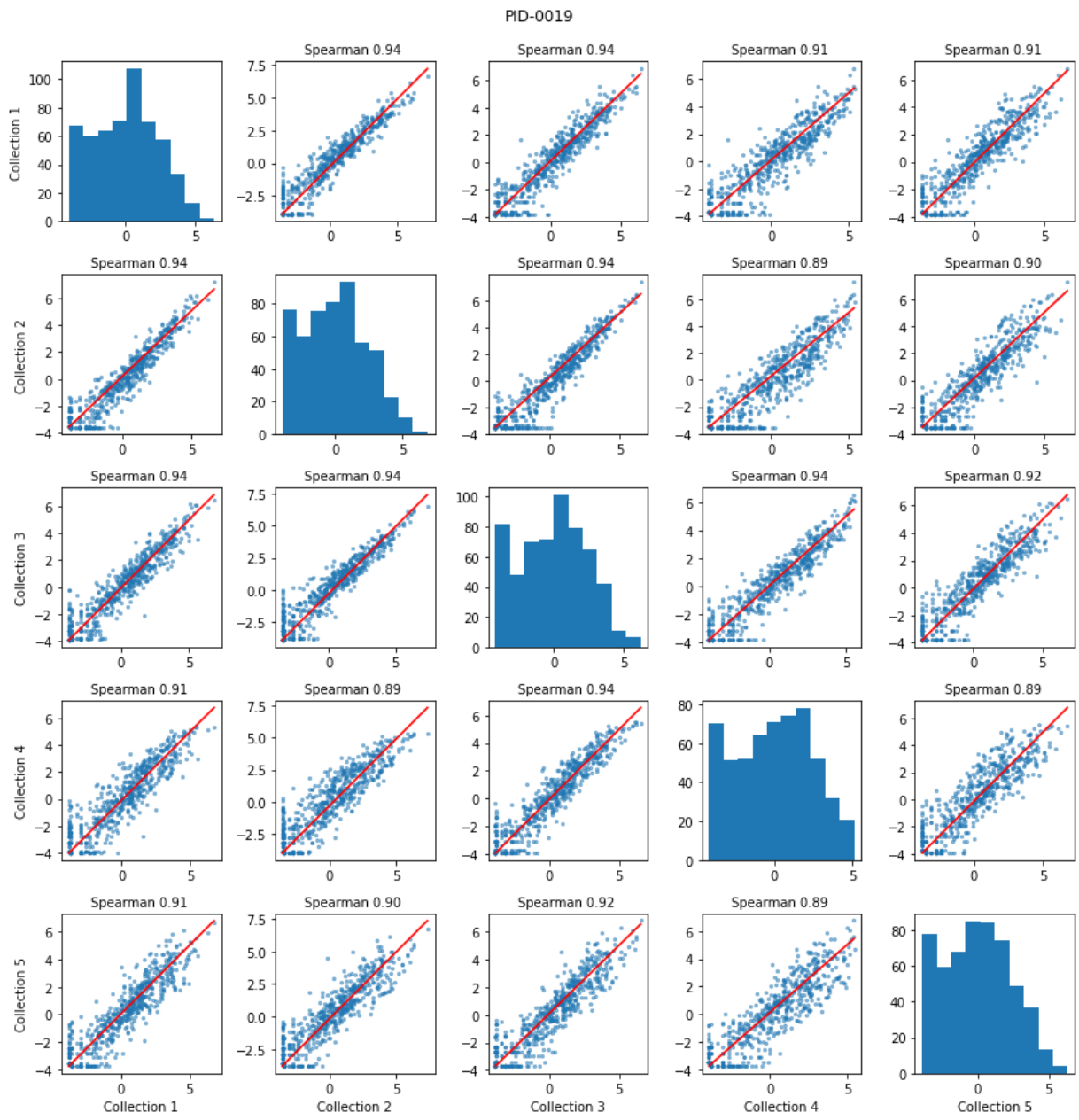
Figure 4. Spearman correlation coefficients of taxonomy (4a), and microbial functions (4b) for sample stability of saliva samples stored at ambient temperature for 0, 7, and 28 days, with and without shipping conditions.

Longitudinal changes in the saliva transcriptome

It is important to understand the longitudinal stability of the salivary transcriptome so that large scale studies can be reliably carried out. Towards this goal, we recruited eight subjects that collected saliva samples weekly for five weeks. On average, 537.18 microbial species and 2693.54 microbial KOs were detected per sample. For all of the taxa and KOs that were detected, the correlation remains very high across five weeks for each participant (Figure 5 and Supplemental Figure 1). The correlations observed between the 1 week and 5 week samples for taxonomy ranged from 0.87 to 0.94 (Figure 5A and Supplemental Figure 1 A-H) and for KOs ranged from 0.88 to 0.94 (Figure 5B and Supplemental Figure 1 I-P). These data show that the

saliva microbiome is longitudinally stable both in terms of composition (taxonomy) and activity (KOs).

A.



B.

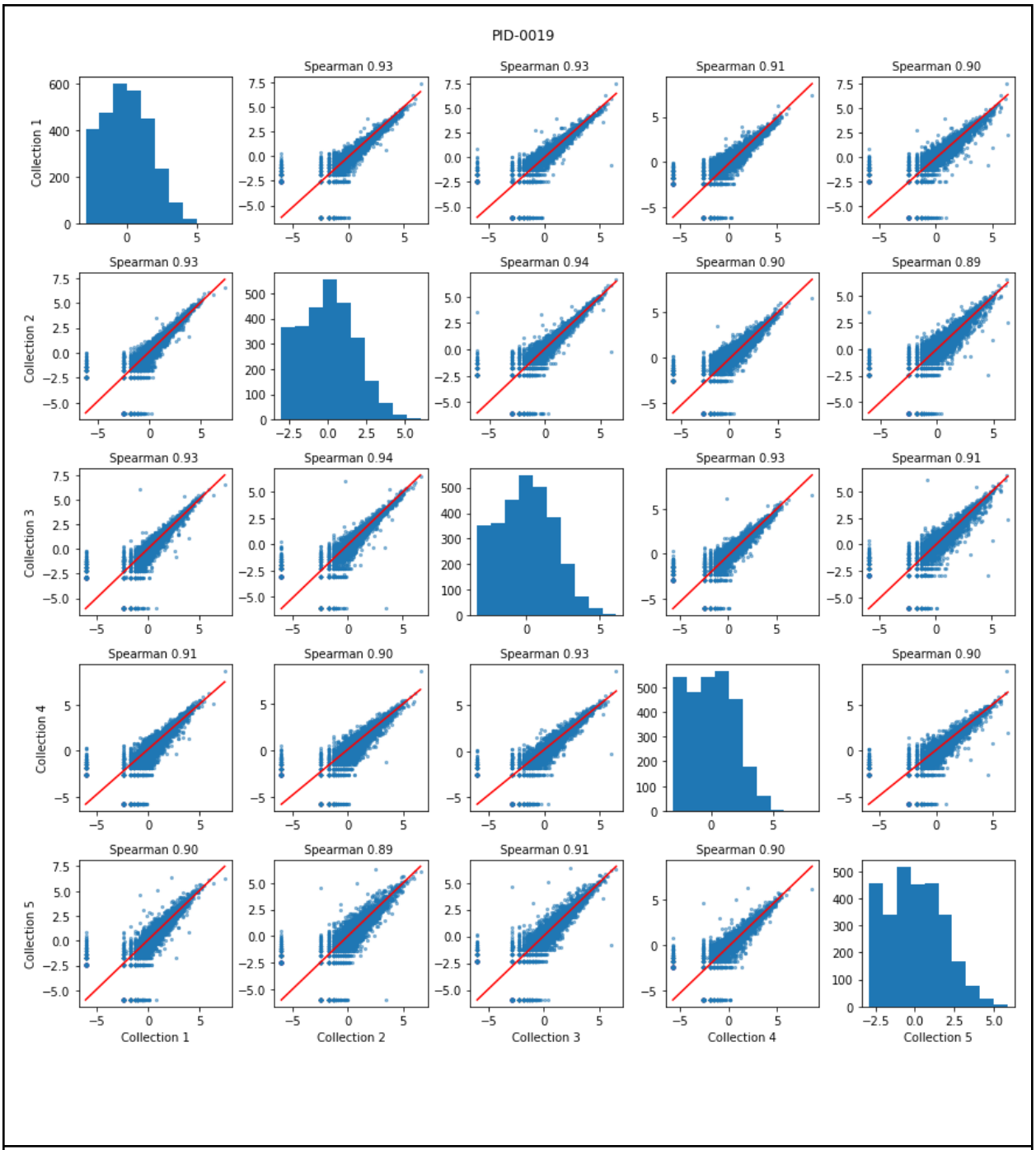
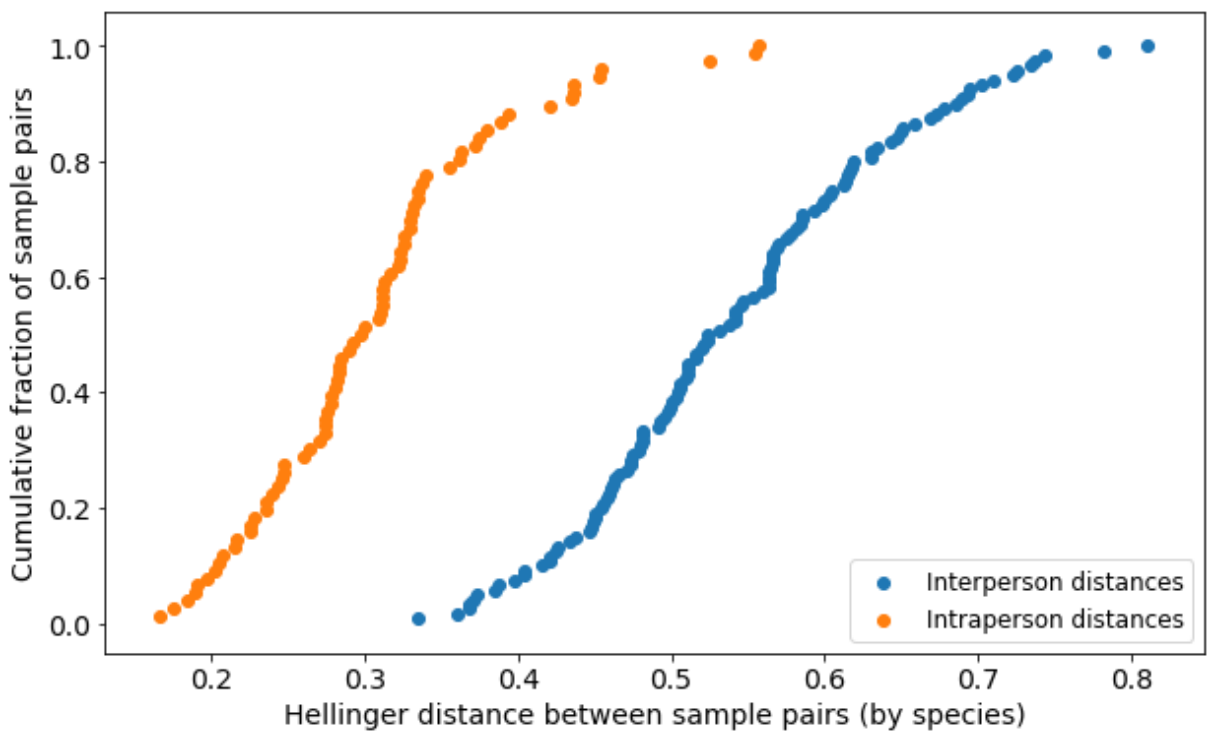


Figure 5. Longitudinal stability of saliva transcriptome in eight study participants over a 5 week period (one participant is shown here, for all eight participants see Supplementary Figure 1). (A) Scatter plots comparing the relative abundance values of each taxon at each collection time. (B) The same analysis was repeated for KOs comparing sum transcripts per million (tpm) of each KO.

Method Precision

One of the more important parameters of the saliva test is to be sufficiently precise to distinguish very small changes among thousands of measured features. Such high precision would enable the identification of transcriptome changes over time, and changes related to health and disease. To assess the test precision in this context, we compared Hellinger distances among Biological samples from the same person collected over time (intraperson distances in Figure 6), and Biological samples obtained from different people (interperson distances in Figure 6). Empirical cumulative distribution function (ECDF) plots of Hellinger distance show that when taking microbial species (Figure 6A), microbial functions (Figure 6B), and human genes (Figure 6C) into account, samples taken from one individual over five different weeks (intraperson distances) tend to be more similar than pairs of samples coming from two different participants (interperson distances).

A.



B.

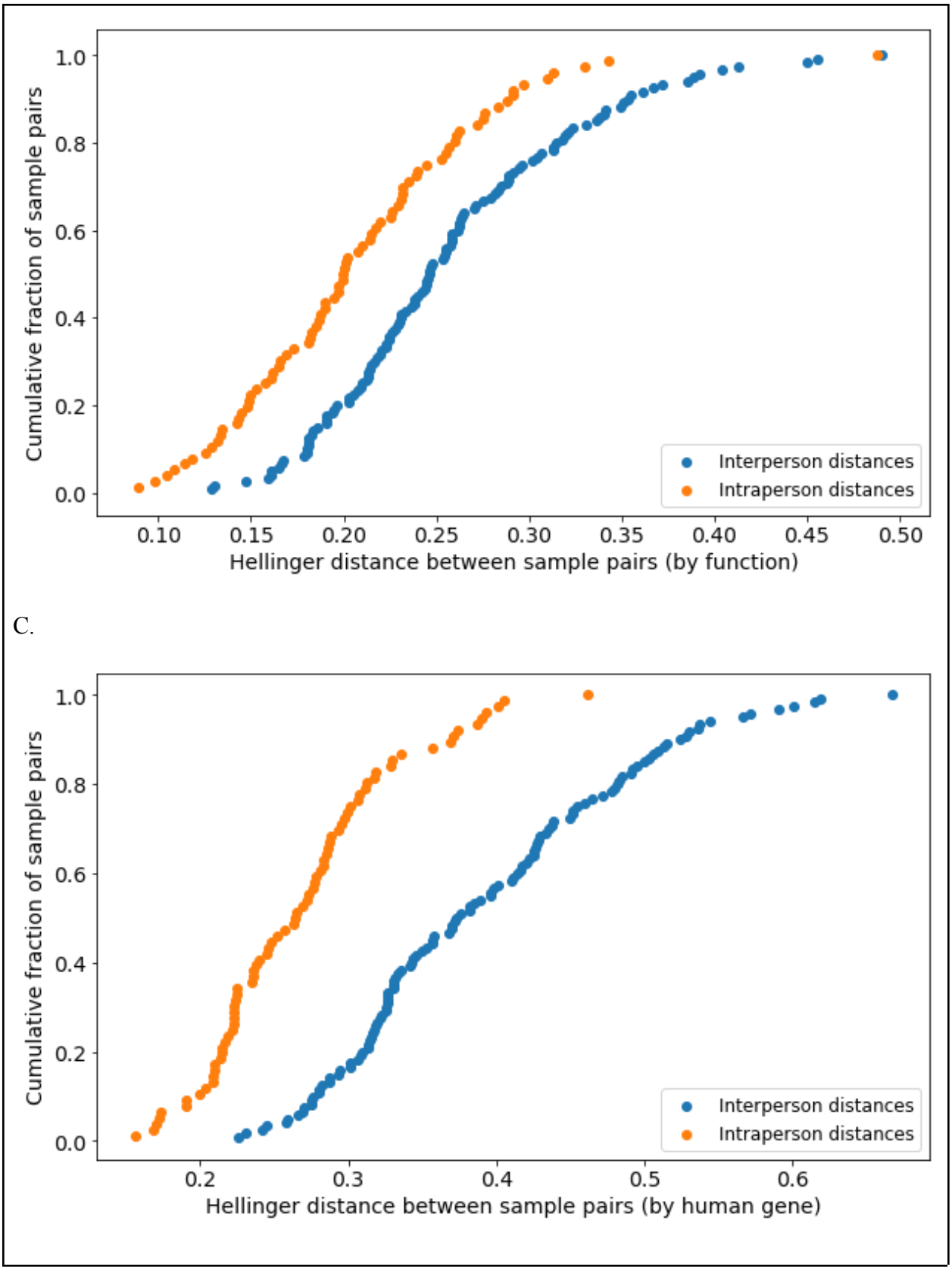

Cumulative fraction of sample pairs
1.0
0.8
0.6
0.4
0.2
0.0
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
Hellinger distance between sample pairs (by function)
Interperson distances
Intraperson distances
C.
Cumulative fraction of sample pairs
1.0
0.8
0.6
0.4
0.2
0.0
0.2
0.3
0.4
0.5
0.6
Hellinger distance between sample pairs (by human gene)
Interperson distances
Intraperson distances

Figure 6. Empirical cumulative distribution function (ECDF) plots of Hellinger distance shows that when taking microbial species (A), microbial functions (B), and human genes (C) into account, samples taken from one individual over five different weeks (intraperson distances) tend to be more similar than pairs of samples coming from two different participants (interperson distances).

Discussion

Chronic, non-communicable diseases are the leading cause of death globally [47]. There is strong evidence that human health, aging, and chronic diseases are heavily influenced by nutrition and microbiome functions. To understand the root causes of chronic diseases, we need to study the human body as an ecosystem, which requires a systems biology approach. Gene expression levels, for both the human and microbiome, play a very important role in chronic disease onset and progression, and need to be studied in the context of an integrated systems biology platform. While scalable and clinically validated gut microbiome and blood transcriptome gene expression tests have been developed [41,48], a human saliva transcriptome test with the same features and performance is also needed.

Here we describe a novel saliva transcriptome test that is clinically validated and globally scalable. This is the first saliva test capable of producing high quality microbial taxonomic classifications and both microbial and human gene expression profiles. The saliva sample (87.5 microliters) can be collected by anyone, anywhere, and can be stored and shipped at ambient temperature to a reference laboratory. The method is automated, inexpensive, and high throughput. The included preservative (RPB) inactivates pathogens, which is very important for international shipping, as it prevents disease spread. RPB also preserves RNA for up to 28 days at ambient temperature, which obviates the need for cold storage and specialized shipping, which are expensive and not easily available. The method precision is high and sufficient to study populations.

Future Directions

This test, along with the previously published stool and blood metatranscriptome tests, can be integrated into a comprehensive systems biology platform that can be applied to population-scale longitudinal studies. Such large studies will identify mechanisms of chronic disease onset and progression, improve or bring new diagnostic and companion diagnostic tools, and enable precision nutrition and microbiome engineering to prevent and cure chronic diseases.

Author Contributions

M Vuyisich and R Toma conceived and designed the methods. R Toma, D Demusaj, and N Duval performed data collection. Y Cai, O Ogundijo, S Gline, and G Banavar performed data analysis. All authors contributed to data interpretation. R Toma and N Duval contributed to the writing of the manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Financial and competing interests disclosure

All authors are current or former employees of Viome, Inc. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References:

1. Nikitakis NG, Papaioannou W, Sakkas LI, Kousvelari E. The autoimmunity–oral microbiome connection. *Oral Dis.* 23(7), 828–839 (2017).
2. Solbiati J, Frias-Lopez J. Metatranscriptome of the Oral Microbiome in Health and Disease. *J. Dent. Res.* 97(5), 492–500 (2018).
3. Vanhatalo A, Blackwell JR, L'Heureux JE, *et al.* Nitrate-responsive oral microbiome modulates nitric oxide homeostasis and blood pressure in humans. *Free Radic. Biol. Med.* 124, 21–30 (2018).
4. Wyllie AL, Fournier J, Casanovas-Massana A, *et al.* Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. *N. Engl. J. Med.* 383(13), 1283–1286 (2020).
5. Hicks SD, Uhlig R, Afshari P, *et al.* Oral microbiome activity in children with autism spectrum disorder. *Autism Res.* 11(9), 1286–1299 (2018).
6. Hicks SD, Carpenter RL, Wagner KE, *et al.* Saliva MicroRNA Differentiates Children With Autism From Peers With Typical and Atypical Development. *J. Am. Acad. Child Adolesc. Psychiatry*. 59(2), 296–308 (2020).
7. Willis JR, Gabaldón T. The Human Oral Microbiome in Health and Disease: From Sequences to Ecosystems. *Microorganisms*. 8(2), 308 (2020).
8. Kinser HE, Pincus Z. MicroRNAs as modulators of longevity and the aging process. *Hum. Genet.* 139(3), 291–308 (2020).
9. Inchingolo F, Martelli FS, Gargiulo Isacco C, *et al.* Chronic Periodontitis and Immunity, Towards the Implementation of a Personalized Medicine: A Translational Research on Gene Single Nucleotide Polymorphisms (SNPs) Linked to Chronic Oral Dysbiosis in 96 Caucasian Patients. *Biomedicines*. 8(5), 115 (2020).
10. Cabanero M, Tsao MS. Circulating tumour DNA in EGFR-mutant non-small-cell lung cancer. *Curr. Oncol. Tor. Ont*. 25(Suppl 1), S38–S44 (2018).
11. Chauhan AK, Bhardwaj M, Chaturvedi PK. Molecular Diagnostics in Liver Cancer [Internet]. In: *Molecular Diagnostics in Cancer Patients*. Shukla KK, Sharma P, Misra S (Eds.), Springer, Singapore, 293–303 (2019) [cited 2021 Jan 13]. Available from: https://doi.org/10.1007/978-981-13-5877-7_17.
12. Rosell R, Carcereny E, Gervais R, *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* 13(3), 239–246 (2012).

13. Bathini P, Foucras S, Dupanloup I, *et al.* Classifying dementia progression using microbial profiling of saliva. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* 12(1), e12000 (2020).
14. Mihaila D, Donegan J, Barns S, *et al.* The oral microbiome of early stage Parkinson's disease and its relationship with functional measures of motor and non-motor function. *PLOS ONE*. 14(6), e0218252 (2019).
15. Acharya A, Chan Y, Kheur S, Jin L, Watt R, Mattheos N. Salivary microbiome in non-oral disease: A summary of evidence and commentary. *Arch. Oral Biol.* 83 (2017).
16. Pietiäinen M, Liljestrand JM, Kopra E, Pussinen PJ. Mediators between oral dysbiosis and cardiovascular diseases. *Eur. J. Oral Sci.* 126 Suppl 1, 26–36 (2018).
17. Long J, Cai Q, Steinwandel M, *et al.* Association of oral microbiome with type 2 diabetes risk. *J. Periodontal Res.* 52(3), 636–643 (2017).
18. Wu Y, Chi X, Zhang Q, Chen F, Deng X. Characterization of the salivary microbiome in people with obesity. *PeerJ*. 6, e4458 (2018).
19. Sharma D, Sandhya P, Vellarikkal SK, *et al.* Saliva microbiome in primary Sjögren's syndrome reveals distinct set of disease-associated microbes. *Oral Dis.* 26(2), 295–301 (2020).
20. Tong Y, Zheng L, Qing P, *et al.* Oral Microbiota Perturbations Are Linked to High Risk for Rheumatoid Arthritis. *Front. Cell. Infect. Microbiol.* [Internet]. 9 (2020). Available from: https://www.frontiersin.org/articles/10.3389/fcimb.2019.00475/full?report=reader.
21. Brennan CA, Garrett WS. Fusobacterium nucleatum — symbiont, opportunist and oncobacterium. *Nat. Rev. Microbiol.* 17(3), 156–166 (2019).
22. Rubinstein MR, Baik JE, Lagana SM, *et al.* Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. *EMBO Rep.* [Internet]. 20(4) (2019). Available from: https://www.embopress.org/doi/abs/10.15252/embr.201847638.
23. Kitamoto S, Nagao-Kitamoto H, Jiao Y, *et al.* The Intermucosal Connection between the Mouth and Gut in Commensal Pathobiont-Driven Colitis. *Cell*. 182(2), 447-462.e14 (2020).
24. Cressatti M, Juwara L, Galindez JM, *et al.* Salivary microR-153 and microR-223 Levels as Potential Diagnostic Biomarkers of Idiopathic Parkinson's Disease. *Mov. Disord.* 35(3), 468–477 (2020).
25. LaRocca D, Barns S, Hicks SD, *et al.* Comparison of serum and saliva miRNAs for identification and characterization of mTBI in adult mixed martial arts fighters. *PloS One*. 14(1), e0207785 (2019).
26. Knight R, Vrbanac A, Taylor BC, *et al.* Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16(7), 410–422 (2018).
27. Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLOS ONE*. 9(4), e93827 (2014).
28. Langille MGI, Zaneveld J, Caporaso JG, *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31(9), 814–821 (2013).
29. Raymann K, Moeller AH, Goodman AL, Ochman H. Unexplored Archaeal Diversity in the Great Ape Gut Microbiome. *mSphere* [Internet]. 2(1) (2017). Available from: https://msphere.asm.org/content/2/1/e00026-17.
30. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. Metatranscriptomics of the Human Oral Microbiome during Health and Disease. *mBio* [Internet]. 5(2) (2014). Available from: https://mbio.asm.org/content/5/2/e01012-14.
31. Koh A, Mannerås-Holm L, Yunn N-O, *et al.* Microbial Imidazole Propionate Affects Responses to Metformin through p38γ-Dependent Inhibitory AMPK Phosphorylation. *Cell*

Metab. 32(4), 643-653.e4 (2020).
32. Hidayat MFH, Milne T, Cullinan MP, Seymour GJ. Feasibility of the salivary transcriptome as a novel biomarker in determining disease susceptibility. *J. Periodontal Res.* 53(3), 369–377 (2018).
33. Microbiome test spots oral cancers. *Nat. Biotechnol.* 39(6), 650–650 (2021).
34. Kadosh E, Snir-Alkalay I, Venkatachalam A, *et al.* The gut microbiome switches mutant p53 from tumour-suppressive to oncogenic. *Nature*. 586(7827), 133–138 (2020).
35. Olsen I, Singhrao SK. Porphyromonas gingivalis infection may contribute to systemic and intracerebral amyloid-beta: implications for Alzheimer's disease onset. *Expert Rev. Anti Infect. Ther.* 18(11), 1063–1066 (2020).
36. Bashiardes S, Zilberman-Schapira G, Elinav E. Use of Metatranscriptomics in Microbiome Research. *Bioinforma. Biol. Insights*. 10, 19–25 (2016).
37. Gosalbes MJ, Durbán A, Pignatelli M, *et al.* Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota. *PLOS ONE*. 6(3), e17447 (2011).
38. Knight R, Jansson J, Field D, *et al.* Unlocking the potential of metagenomics through replicated experimental design. *Nat. Biotechnol.* 30(6), 513–520 (2012).
39. He S, Wurtzel O, Singh K, *et al.* Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods*. 7(10), 807–812 (2010).
40. Hatch A, Horne J, Toma R, *et al.* A Robust Metatranscriptomic Technology for Population-Scale Studies of Diet, Gut Microbiome, and Human Health. *Int. J. Genomics*. 2019, 1718741 (2019).
41. Toma R, Duval N, Pelle B, *et al.* A clinically validated human capillary blood transcriptome test for global systems biology studies. *BioTechniques*. 69(4), 289–301 (2020).
42. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B*. 39(1), 1–38 (1977).
43. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28(1), 27–30 (2000).
44. Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V. Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. *Math. Geol.* 35(3), 253–278 (2003).
45. Aitchison J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B Methodol.* 44(2), 139–177 (1982).
46. Souza MF de, Kuasne H, Barros-Filho M de C, *et al.* Circulating mRNAs and miRNAs as candidate markers for the diagnosis and prognosis of prostate cancer. *PLOS ONE*. 12(9), e0184094 (2017).
47. About Chronic Diseases | CDC [Internet]. (2021). Available from: https://www.cdc.gov/chronicdisease/about/index.htm.
48. Hatch A, Horne J, Toma R, *et al.* A Robust Metatranscriptomic Technology for Population-Scale Studies of Diet, Gut Microbiome, and Human Health. *Int. J. Genomics*. 2019, 1–9 (2019).