Computational and Structural Biotechnology Journal 23 (2024) 834–842

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Microbial functional pathways based on metatranscriptomic profiling enable effective saliva-based health assessments for precision wellness

Eric Patridge [a], Anmol Gorakshakar [a,1], Matthew M. Molusky [a,1], Oyetunji Ogundijo [a], Angel Janevski [a], Cristina Julian [a], Lan Hu [a], Momchilo Vuyisich [b], Guruduth Banavar [a,*]

[a] Viome Research Institute, Viome Life Sciences Inc., New York City, USA
[b] Viome Research Institute, Viome Life Sciences Inc., Seattle, USA

ARTICLE INFO

Keywords:
Saliva
Metatranscriptomics
Oral microbiome
Functional pathways
Oropharyngeal disease
Precision wellness

ABSTRACT

It is increasingly recognized that an important step towards improving overall health is to accurately measure biomarkers of health from the molecular activities prevalent in the oral cavity. We present a general methodology for computationally quantifying the activity of microbial functional pathways using metatranscriptomic data. We describe their implementation as a collection of eight oral pathway scores using a large salivary sample dataset (n = 9350), and we evaluate score associations with oropharyngeal disease phenotypes within an unseen independent cohort (n = 14,129). Through this validation, we show that the relevant oral pathway scores are significantly worse in individuals with periodontal disease, acid reflux, and nicotine addiction, compared with controls. Given these associations, we make the case to use these oral pathway scores to provide molecular health insights from simple, non-invasive saliva samples, and as molecular endpoints for actionable interventions to address the associated conditions.

1. Introduction

Billions of people are impacted by oral diseases such as dental caries, periodontal gum disease, oral pre-malignancies like leukoplakia, and oral cancers [1,2]. The number of people affected by them continues to grow for a variety of reasons, including the availability and affordability of food with high sugar content [3], a lack of reliable oral cancer screening tools [4], and limited success of effective oral health care services in communities [5], including regular access to fluoride [6,7].

The oral cavity is the primary gateway to the body, and it hosts a complex environment which plays vital functional roles [8,9]. Both the oral microbiome and the oral immune system defend against a vast array of pathogens [10–13], but when either of these components is impaired, the oral cavity can support pathogenic activity, leading to chronic oral inflammation [14–17]. Since digestion begins in the oral cavity [18,19], the impairment of either component can lead to digestion-related health issues, including putrefaction of foods and host proteins within the mouth [20]. Halitosis (bad breath), gum disease, and oropharyngeal cancers are just a few of the conditions or diseases research suggests may coincide with an impaired oral microbiome and/or oral immune system [21–26].

Poor oral health may point to underlying health issues, since bi-directional associations exist between oral health and systemic health [27]. For example, left unchecked, chronic oral inflammation may advance to gum disease, and is linked to several conditions beyond the oral cavity, including diabetes, cardiovascular diseases, and Alzheimer's disease [28–32]. Similarly, chronic halitosis is associated with *Helicobacter pylori* infections, liver disease, and gastroesophageal reflux disease [33–35]. Bi-directional associations like these are opportunities for health professionals to act on findings from early risk assessments and encourage proactive healthcare measures [36].

To improve oral health, and indirectly, some aspects of systemic health, the existing prophylactic and therapeutic healthcare efforts can be augmented with non-invasive wellness tools that comprise molecular tests, diet changes, and supplement use [37,38]. As many have already shown, oral diagnostics are non-invasive tools to identify and address existing health issues [4], but these are reactive measures and generally not useful to proactive healthcare efforts [39–41]. In fact, there are still relatively few options which individuals or health professionals can leverage in the pursuit of proactive efforts for general healthcare [42].

* Corresponding author.
E-mail address: guru@viome.com (G. Banavar).
[1] Equal contribution.

https://doi.org/10.1016/j.csbj.2024.01.018
Received 1 November 2023; Received in revised form 25 January 2024; Accepted 25 January 2024
Available online 29 January 2024

We believe that an important step towards effective general healthcare is to provide deep health insights into the molecular activities prevalent in the oral cavity. Our recent efforts include a saliva-based test that measures the entire oral metatranscriptome [43] as well as the application of this test for personal wellness. A major aspect of our personalized precision wellness application is to provide each individual a detailed assessment of their molecular activities as a set of oral pathway scores. Microbial functions are far more important for human health than taxonomy [44–47], so each oral pathway score leverages the high-resolution detection of microbial functional features (i.e., microbial gene expression) captured as KEGG Orthologs [48,49], or KOs, within the oral metatranscriptome. In the present context, each pathway score is defined as a set of well-understood microbial biochemical functions that have been tied to various health conditions in the literature or clinical domain. The scope of the score and decisions to include or exclude specific functions are curated by domain experts using associations with various health conditions. Through this approach, these pathway scores focus on oral host-microbiome interactions that maintain human health and can assist in developing methods to modulate microbial activities for improved health.

This paper aims to describe (1) oral pathway scores designed to assess common oral and systemic health issues, developed through a well-defined, data-driven process, (2) the computational development methodology of saliva-based oral pathway scores using a large development cohort (n = 9350), and (3) their validation against well-known oropharyngeal and related diseases within a large independent cohort (n = 14,129).

2. Materials and methods

All samples and metadata were obtained from customers who purchased Viome's Full Body Intelligence kits. These individuals were at least 18 years old at the time of sample collection, and they either completed a research Informed Consent Form (approved by an HHS-registered Institutional Review Board; IRB Number: IRB00011543; IORG Number: IORG0009710) or agreed to have their data analyzed in the terms and conditions during their purchase.

Full Body Intelligence kits are shipped directly to participants' homes. Each kit includes a saliva collection tube, along with a built-in insert that serves two functions: 1) a funnel to ease saliva collection, and 2) a press-sealed chamber with RNA Preservation Buffer (RPB) to preserve the RNA integrity at the time of collection. RPB has been clinically validated to preserve RNA for up to 28 days at room temperature [43].

After (unstimulated) saliva collection, removal of the built-in insert releases RPB. The user then caps the saliva collection tube, mixes the RPB with saliva, and ships the sample at room temperature. Upon arrival in the lab, the saliva/RPB mixture is aliquoted into cryogenic tubes and stored at − 80 °C in the biobank.

At the time of sample collection, participants also answer a comprehensive questionnaire describing their lifestyle, oral hygiene, dietary preferences, and health history. Relevant lifestyle questions include tobacco use and alcohol intake. Relevant oral hygiene questions include mouthwash use, brushing habits, and flossing habits. Relevant dietary preferences include caffeine intake. Health history is collected from both a multi-select question and open text response, resulting in hundreds of self-reported condition and disease phenotypes. Relevant self-reported health history includes gingivitis, acid reflux, nicotine addiction, dry mouth, kidney disease, kidney cyst, and Sjögren's syndrome. All study data are de-identified; data analysis team members have no access to personally identifiable information.

2.1. Cohort development

To facilitate health insights from saliva samples, we develop and validate wellness scores in the context of reference cohorts that are representative of the general adult population. Each cohort is composed of approximately 10,000 samples and excludes outliers based on age, birth sex, and body mass index.

For the score development cohort, additional criteria serve to diminish artifacts related to sequencing depth or data sparseness; cohort samples contain a total count of more than 1700 KOs, with a total read count of at least 70,000. The final cohort for score development is composed of 9350 saliva samples from adults (59.4% female; see Table 1). For independent validation of scores, the final cohort is composed of 14,129 saliva samples from adults (62.9% female; see Table 1) that were not part of the score development cohort.

2.2. Metatranscriptomic analysis

Saliva samples are collected and analyzed from individuals who fast and do not brush their teeth or use a mouthwash for at least 8 h. The complete set of transcripts (RNA molecules) from each saliva sample is quantified using previously reported metatranscriptomic techniques [43], yielding both the primary sequence and read count for each transcript. The bioinformatics method aligns each sequencing read to microbial genomes as follows: Viome maintains a custom reference catalog which includes 32,599 genomes from NCBI RefSeq release 205 'complete genome' category, 4644 representative human gut genomes of UHGG [50], ribosomal RNA (rRNA) gene sequences, and the human genome GRCh38 [51]. These genomes cover archaea, bacteria, fungi, protozoa, phages, viruses, and the human host. The microbial genomes

Table 1
Representative anthropometric, sociodemographic, oral hygiene, and lifestyle metrics of the development and validation cohorts. (Mean ± SD for continuous variables).

	Score Development Cohort (n = 9350 saliva samples)	Independent Validation Cohort (n = 14,129 saliva samples)
Female	59.4%	62.9%
Age (years)	46.2 ± 13.0	46.0 ± 13.1
BMI (kg/m^2)	26.1 ± 5.4	25.8 ± 5.2
Ethnicity	74.2% White	73.4% White
	From the remaining 25.8%:	From the remaining 26.6%:
	28.1% Multi-ethnic	26.7% Multi-ethnic
	28.9% Hispanic or Latino	29.1% Hispanic or Latino
	19.2% Asian	19.9% Asian
	11.1% Black/African American	10.6% Black/African American
	5.6% Arab or Middle Eastern	6.7% Arab or Middle Eastern
	7.0% Other	6.9% Other
Tobacco use	8.7% Current smoker	8.3% Current smoker
	13.3% Former smoker	13.4% Former smoker
	78.0% Never smoker	78.2% Never smoker
	0.0% No answer	0.1% No answer
Alcohol intake	25.5% Heavy drinker	22.9% Heavy drinker
	60.6% Light/Moderate drinker	62.5% Light/Moderate drinker
	9.2% Former drinker	9.8% Former drinker
	4.5% Abstainer	4.7% Abstainer
	0.2% No answer	0.1% No answer
Caffeine intake (drinks)	1.5% High intake	1.3% High intake
	19.2% Moderate intake	20.1% Moderate intake
	63.1% Low intake	62.2% Low intake
	16.1% Do not consume	16.3% Do not consume
	0.1% No answer	0.1% No answer
Mouthwash use	23.8% Yes	23.7% Yes
	75.9% No	76.1% No
	0.3% No answer	0.2% No answer
Brushing habit	64.9% Two or more times a day	67.1% Two or more times a day
	30.9% Once a day	29.3% Once a day
	4.0% Not every day	3.4% Not every day
	0.2% No answer	0.2% No answer
Flossing habit	9.0% Twice or more per day	9.7% Twice or more per day
	35.5% Once a day	36.6% Once a day
	36.7% Less than once a day	35.6% Less than once a day
	18.6% Do not floss	18.0% Do not floss
	0.2% No answer	0.1% No answer

have 98,527,909 total annotated genes. We adopt KEGG Orthology (KO) [48,49] to annotate the microbial gene functions using eggNOG-mapper [52]. The microbiome pipeline maps paired-end reads to this catalog using Centrifuge [53] for taxonomy classification (at all taxonomic ranks). Reads mapped to the host genome and rRNA sequences are tracked for monitoring, but excluded from further analysis. Reads mapped to microbial genomes are processed with an Expectation-Maximization (EM) algorithm [54] to estimate the expression level (or activity) in the sample. Respective taxonomic ranks (strains, species, genera, etc.) can be easily aggregated from the genomes. For this study, we use species activity in the downstream analyses. These genome mapped reads are extracted and mapped to only gene or open reading frame (ORF) regions for molecular function or KO annotation and quantification. The KOs we identify from saliva samples are used for downstream analyses including score development, validation, and eventually, scoring of new samples.

2.3. Score development framework

With the goal of delivering health insights based exclusively on microbial functions (i.e., gene expression) from the oral microbiome, we adopt an iterative and multistep process for developing wellness scores. All scores described in this paper comprise only microbial KOs. Within the finalized scores, each KO contributes a certain weight to the score, either in a positive or negative direction. To develop the scores, we examine domain concepts such as the activities associated with microbial colonization, consumption of salivary proteins, consumption of dietary mono and disaccharides, destruction of host tissues in the mouth, etc. We identify the microbial KOs associated with these physiological processes. The normalized expression value of each KO is aggregated with computationally determined weights applied to each, based on the first component of Principal Component Analysis (PCA). The final scores, when applied to a large cohort, exhibit a gaussian-like curve which stratify health insights across the population (shown in Supplemental Fig. S1). As shown in Fig. 1, the development steps for each score include: 1) domain exploration to identify KOs and associated phenotypes; 2) curation of self-reported metadata to enable case/control difference analyses; 3) signal definition to align each score for broad wellness assessment; 4) feature selection which involves iterative curation of score KOs; and 5) pathway activity quantification of the KOs selected for the score. This section explains each step in further detail using the score development cohort.

Step 1: Domain Exploration.

For each wellness score, we prioritize explainability of the overall score and its full set of KOs. We therefore begin by exploring the domain in order to understand the biological and clinical phenomena within the scope of each score, such as inflammation within periodontal pockets. While surveying knowledge sources, scientific manuscripts, and clinical literature, we collect an exhaustive set of KOs that are related to the score concepts and associated phenotypes, either based on their biological functions or based on reported correlations of their transcriptional regulation. Furthermore, iterating over this domain exploration step also allows us to prioritize a subset of KOs and phenotypes to be utilized in downstream development steps.

Step 2: Metadata Curation.

Part of the score development process requires metadata labels for case/control difference analyses. To create these labels, we curate and normalize the phenotype data from individual surveys, including anthropometrics, socioeconomic, demographics, oral hygiene, and lifestyle characteristics. For the current manuscript, these sources are self-reported and specifically include health conditions/diseases, which enable us to explore the association between these wellness scores and diseases. If an individual reported no diseases or conditions, then they were identified as "lacking all disease phenotypes." During the score development process, controls for each "case" definition are randomly sampled from the larger population, excluding anyone with a "case" label.

Step 3: Signal Definition.

As we begin defining wellness scores for health insights, we use both the development cohort and the exhaustive set of curated KOs from Step 1. The metadata labels from Step 2 guide us during case/control difference analyses. Importantly, we make every effort to design scores for general wellness rather than diagnostics, so we avoid anchoring a score design to a single, individual disease. The definitive "signal" we pursue is a KO set that consistently differentiates related phenotypes across the development cohort. However, published gene expression efforts report variable findings across populations as well as reported measures [55–57], so it is not known which KOs will consistently differentiate signals across a large number of saliva samples. Therefore, we use a heuristic approach to prioritize KOs based on the numerous labels we create in Step 2. There are several metrics we utilize to define successful signals for phenotypes and health insights, and these include: 1) score difference between cases and controls; 2) a normalized score distribution; and 3) independence between score and sample sequencing depth. The output of the signal definition step is a minimal set of KOs, which sets a foundation for us to build upon for the final KO selection.

Step 4: Feature Selection.

The final KO set for each score is defined through an iterative curation process, which uses all of the KOs and metadata labels identified through the above steps. We do not intentionally aim for each score to contain a specific number of KOs. Instead, iteration continues until scores reach an optimum, as determined through case/control difference analyses, stratifying individuals from the cohort into lower and higher scores, such that the phenotype labels remain consistent. If the overall "signal" is not consistent, or the score distribution is far from a normal curve, or the score is highly correlated with sequencing depth, then the iterative process continues.

Step 5: Pathway Activity Quantification.

Once the feature set (KOs) for a given pathway score is defined, the goal of this step is to combine the expression levels of the selected features to arrive at an aggregate quantification of the entire pathway. We derive a pathway score as a weighted function Score $= C1F1 + C2F2$

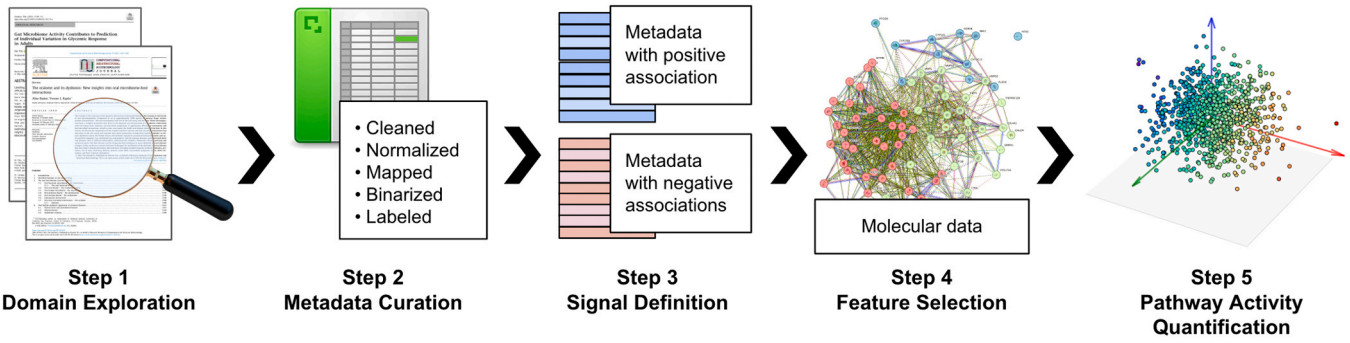

Fig. 1. Score Development Framework.

$+ \ldots + CnFn$, where *Fi* is the expression level of the feature and *Ci* is its weight. After experimenting with multiple weight computation methods, including a manual approach based on domain knowledge and decision tree methods, we settled on an algorithmic method that learns the weights based on the simultaneous expression (i.e., covariance) of the selected features, which intuitively corresponds to the activity of the entire pathway. Since the first principal component of PCA captures the largest variance in the original dataset, we believe that it provides a reasonable approximation of the simultaneous expression of the selected features. Furthermore, weighting features by the first principal component also enhances explainability of the score's complete set of features.

2.4. Score validation framework

The validation cohort of 14,129 saliva samples is used to validate the reproducibility of the signals found in score development in an unseen independent sample set. To validate score designs, we ask whether scores can assess common oral health issues within the validation cohort. Domain knowledge provides evidence that many oral diseases increase risk for others: i.e., periodontal disease with halitosis [58]; periodontal disease with Sjögren's [59]; acid reflux with both periodontal disease and cavities [60]. Therefore, we pursue a general wellness approach and examine the performance of all eight oral pathway scores with several diseases known to directly impact oropharyngeal health.

We specifically focus on three phenotypes – gingivitis, acid reflux, and nicotine addiction – for the validation process presented in this paper. To test scores with each disease phenotype, we perform case/control differential analyses. Case labels constructed for the development cohort are also applied to the validation cohort. Controls are defined as all individuals with no self-reported health conditions/diseases; they lacked all disease phenotype labels. Controls are matched to cases (1:10) on age, birth sex, and BMI. For age, we match individuals ≤ 10 year difference. For BMI, we match within BMI categories as follows: BMI < 18.5 (defined as 'underweight'), $18.5 \leq$ BMI < 24.9 (defined as 'healthy'), $24.9 \leq$ BMI < 29.9 (defined as 'overweight'), $29.9 \leq$ BMI (defined as 'obese'). To perform case/control difference analyses, we use the Mann-Whitney U test along with the Benjamini-Hochberg correction to control the False Discovery Rate.

3. Results

3.1. Cohort characteristics

For score design and validation, we use a score development cohort and an independent validation cohort, respectively. The development cohort consists of 9350 saliva samples from adults (59.4% Female) with an average age of 46.2 ± 13.0 years and BMI of 26.1 ± 5.4 kg/m^2. The validation cohort consists of 14,129 saliva samples from adults (62.9% Female) with an average age of 46.0 ± 13.1 years and BMI of 25.8 ± 5.2 kg/m^2.

Anthropometric, sociodemographic (including ethnicity), oral hygiene (including mouthwash use, brushing habits and flossing habits), and lifestyle characteristics (including tobacco use, alcohol intake, and caffeinated drink intake), all self-reported at the time of sample collection, are shown in Table 1 for the development and validation cohorts.

3.2. Metatranscriptomic analysis

Transcriptome metrics for score development (n = 9350) and validation cohorts (n = 14,129) demonstrate the range and quality of sample data generated across both cohorts. Total single reads, or the number of reads in paired-end fastq files after demultiplexing, for the development and validation cohorts combined is 8.09 ± 3.34 in millions. The number of unique KOs (KO richness) associated with each sample for the development cohort and validation cohort combined is 3276 ± 488 KOs. The distribution for the number of unique species (Species richness) associated with each sample for the development cohort and validation cohorts combined is 349 ± 53 species.

3.3. Score Development

All eight oral pathway scores presented in this paper are shown in Table 2, and each score is intended to assess a biological or clinical phenomena. During score development, authors selected a name for each score to represent the score's intended functionality while remaining accessible to the general population. Across these scores we include 234 distinct KOs, and we make every effort to minimize feature overlap between the scores; there is a maximum of two KOs shared between any two scores. Across all eight scores, 221 of the KOs appear only once, while the remaining 13 KOs appear 2–4 times.

Score development is a five-step process that begins with domain exploration to identify microbial KOs potentially relevant to the score concepts and associated phenotypes. Here, we detail the score development process using *OralUreaseActivityPathways* as an illustrative example. (Each of the other scores follow a similar process.).

As part of domain exploration, we identify key components and functions that we consider to be "in scope" for the

Table 2
High level overview of Oral Pathway Scores. Score names are shown alongside the high level descriptions of their scope, and the number of KOs in each score. In a pairwise comparison of each score, there is a maximum of two KOs shared between any two scores.

Score Name	Key functions	Microbial activities	KOs
OralAmmonia-Production-Pathways	Ammonia productionArginine biosynthesis Nitrogen metabolism	Homeostatic levels of ammonia within the oral cavity which protect against acids and impart an anti-cariogenic effect.	26
OralButyrate-Production-Pathways	Butanoate products Lysine degradation Short chain fatty acid metabolism	Production of beneficial butyrate as well as butyrate from pathogens.	30
OralCariogenic-Pathways	Acid production Chelation activity Sucrose metabolism	Cariogenic and chelation activities, as well as other activities which promote dental cavities.	28
OralFlagellar-Assembly-Pathways	Bacterial secretion system Chemotaxis Flagellar assembly Quorum sensing	Flagellar components, related motility, informative of pathogenic activity.	49
OralLPS-Biosynthesis-Pathways	Biofilm formation Biosynthesis of nucleotide sugars Lipopolysaccharide biosynthesis	Production of pro-inflammatory lipopolysaccharide (LPS).	35
OralPeriodontal-Mucin-Degradation-Pathways	Amino sugar metabolism Galactose metabolism Glycoprotein metabolism Nucleotide sugar metabolism	Degradation of salivary mucin and diminishment of oral health barrier defenses.	30
OralPeriodontalPro-Inflammatory-Pathways	Glyoxylate metabolism Lipoic acid metabolism Porphyrin metabolism Protease/peptidase activity	Inflammatory factors and degradation of periodontal gum health.	26
OralUrease-ActivityPathways	Acid-stress Nickel transport Urate metabolism Urea metabolism	Acid-stress and pH balancing activities as well as others which protect against an acidic environment and related pathogenic activity.	29

OralUreaseActivityPathways score (Table 3), which focuses on the oral microbiome's acid/base activities related to urea.

After extensive research covering the scope of the score, we identify a long list of relevant KOs (in the hundreds or thousands), and we supplement these with additional relevant KOs which we identify from the development cohort. At this point in the process, all KOs are thought to directly or indirectly contribute to a "key component" or "key function" shown in Table 3.

The second and third steps of the Score Development Framework involve metadata curation and signal definition. The metadata labels we consider to be central to the *OralUreaseActivityPathways* score include Dry mouth, Kidney disease, Kidney cyst, and Sjögren's syndrome, from a total of more than 500 labels available.

The fourth step is feature selection, and the output of this step is the final set of KOs and their weights (loadings) to the score. Using *OralUreaseActivityPathways* as an example, Table 4 shows the included KOs and their respective weights. As presented here, the KOs with positive loadings are indicative of beneficial contributions with respect to the oral microbiome's acid/base status and related to urea.

3.4. Score validation

We set out to determine whether there is a significant association between scores and disease phenotypes within an unseen independent cohort. Using the validation cohort, matched case/control difference analyses are conducted with oropharyngeal diseases, demonstrating that the final scores are able to differentiate phenotypes across a set of samples independent from those we use in score development.

To minimize complexity, we assess the oral scores with a small set of diverse phenotypes. The selected oropharyngeal and related disease phenotypes (Gingivitis, Acid Reflux, and Nicotine Addiction) have varied origins and are known to be associated with altered oral microbiomes [61–65].

Cases for three oropharyngeal and related disease phenotypes were identified based on self-reported information in response to the question: "Please list all of the illnesses you are currently suffering from or diagnosed with."

The selected phenotypes are used to validate oral pathway scores through case/control difference analyses (Fig. 2). Cases of gingivitis ($n = 150$; c=1462), cases of acid reflux ($n = 450$; c=4268), and cases of nicotine addiction ($n = 67$; c=665) are all matched (1:10). As seen in Fig. 2, the differences between "case" and "control" samples are variably significant across the eight scores ($p \leq 0.05$ to $p \leq 0.001$) according to the Mann-Whitney U test with Benjamini-Hochberg correction for multiple hypothesis testing (FDR < 0.05).

4. Discussion

Oral microbiomes are associated with both oral and systemic diseases. Therefore, our starting point in this study was that molecular data obtained from the oral microbiome may provide useful biomarkers of health and disease. Further, it is clear from multiple studies [66,67] that the functional aspects of the microbiome are likely more impactful on human physiology than composition (taxonomy), which only measures the functional potential. Our metatranscriptomic method quantifies gene expression, that is the activity of microbial functions, and makes them available as KEGG Orthologs or KOs, which we use as the basis of developing oral functional pathway scores.

In the design of our oral pathway scores, both the development and independent validation cohorts are considerably important to the process. Further, the large number of samples in each cohort is essential to identifying sufficient numbers of "case" metadata labels, which enable the case/control difference analyses presented herein. A comparison of the two cohorts (Table 1) indicates they are similar and comparable in terms of anthropometrics, sociodemographics, oral hygiene (including mouthwash use, brushing habits and flossing habits), and lifestyle metrics (including tobacco use, alcohol intake, and caffeinated drink intake), supporting our goal of validating the reproducibility of the signals found in score development within an unseen independent validation cohort.

While developing each oral pathway score, we prioritize health insights for oropharyngeal disease phenotypes, leveraging the connection between the oral microbiome and oral health. The key functions we identify for each score (Table 2) are relevant to disease phenotypes, and these functions take central roles in the score designs and guide the selection of specific KOs. In the case of *OralUreaseActivityPathways*, the expanded set of key functions (Table 3) point to several urea-related KOs

Table 4
Score design for *OralUreaseActivityPathways*. Loadings originate from the first principal component of PCA analysis using the development cohort.

Loadings	KEGG Ortholog (KO) and Description
0.23282	K00101 (lldD); L-lactate dehydrogenase (cytochrome)
0.21168	K03746 (hns); DNA-binding protein H-NS
0.20788	K01420 (fnr); CRP/FNR family transcriptional regulator
0.20710	K11605 (sitC); manganese/iron transport system permease
0.17742	K11606 (sitD); manganese/iron transport system permease
0.10082	K00605 (gcvT,AMT); aminomethyltransferase
-0.00871	K01895 (ACSS1_2,acs); acetyl-CoA synthetase
-0.01382	K15585 (nikB,cntB); nickel transport system permease
-0.01632	K15586 (nikC,cntC); nickel transport system permease
-0.03862	K01941 (uca); urea carboxylase
-0.04997	K03187 (ureE); urease accessory protein
-0.07264	K07770 (cssR); two-component OmpR, response regulator
-0.09710	K01755 (argH,ASL); argininosuccinate lyase
-0.10985	K01218 (gmuG); mannan endo-1,4-beta-mannosidase
-0.13493	K07650 (cssS); two-component OmpR, histidine kinase
-0.14145	K14977 (ylbA,UGHY); (S)-ureidoglycine aminohydrolase
-0.15198	K01466 (allB); allantoinase
-0.17054	K02083 (allC); allantoate deiminase
-0.17658	K01428 (ureC); urease subunit alpha
-0.18129	K01430 (ureA); urease subunit gamma
-0.20003	K09477 (citT); citrate:succinate antiporter
-0.21234	K01429 (ureB); urease subunit beta
-0.21740	K03190 (ureD,ureH); urease accessory protein
-0.22403	K03188 (ureF); urease accessory protein
-0.26017	K14048 (ureAB); urease subunit gamma/beta
-0.26711	K00366 (nirA); ferredoxin-nitrite reductase
-0.27764	K22373 (larA); lactate racemase
-0.29448	K00370 (narG,narZ,nxrA); nitrate reductase, alpha subunit
-0.31953	K03191 (ureI); acid-activated urea channel

Table 3
Scope of the *OralUreaseActivityPathways* Score.

Key components	Organic acids	Citrate, succinate, fumarate, 2-oxoglutarate
	Proteins	Urease, urea carboxylase, urea transporter, organic acid transporter, stress regulator, argininosuccinate lyase
	Purines	Uric acid, xanthine, hypoxanthine, adenine, guanine
	Enzyme cofactors	Nickel, nickel-pincer cofactor, manganese, iron
	Amino acids	Arginine, citrulline, ornithine, glutamine, glutamate
	Other	Urea, allantoate, ureidoglycine, N-acetylglutamate, Acetyl-CoA, ammonia, nitrate, nitrite
Key functions	Nitrogen waste	Urea metabolism, urate metabolism, ammonia production
	Signaling	Acid-stress, acid tolerance
	Homeostasis	Protection, carbon metabolism
	Other	pH balance, nutrient deprivation

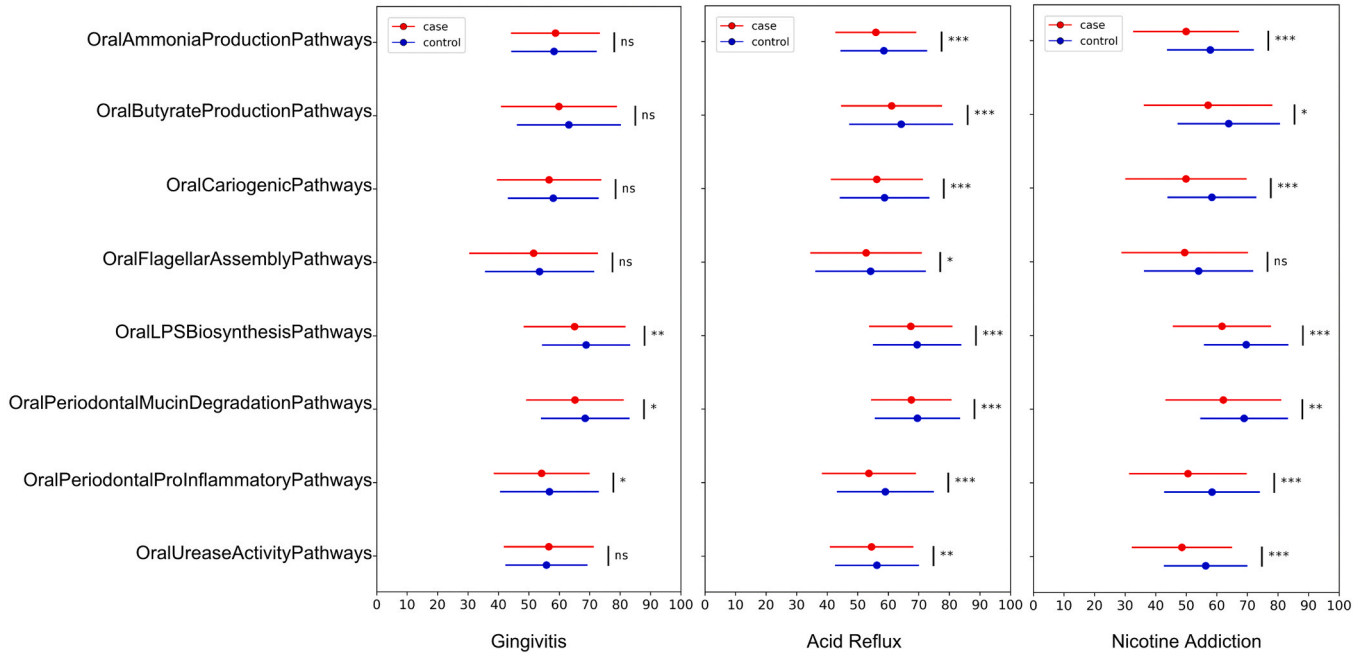

Fig. 2. Oropharyngeal and related diseases negatively impact oral scores. Oral scores from our independent validation cohort comparing people with self-reported oropharyngeal and related diseases (case in red) versus people with no self-reported comorbidities (control in blue). Cases include gingivitis (n = 150; c=1462), acid reflux (n = 450; c=4268), and nicotine addiction (n = 67; c=665). Cases and controls are matched (1:10) by age, sex, and BMI. The color line represents mean ± SD of each score in case and control respectively; *p ≤ 0.05, * *p ≤ 0.01, * **p ≤ 0.001, ns p > 0.05, from the Mann-Whitney U test with Benjamini-Hochberg correction for multiple hypothesis testing (FDR < 0.05).

which are up-regulated at low pH or during acid stress, and these guide the design for this score. While finalizing each KO set, we minimize correlations with the sequencing depth (r = 0.08 for *OralUreaseActivityPathways*) so that scores are robust (less sensitive to the amount of data generated). We also work towards a normal distribution of scores across the development cohort (See Supplemental Fig. S1) to maximize their utility as wellness scores. The case/control difference analyses with oropharyngeal and related disease phenotypes (Fig. 2) indicate these wellness scores are effective for oral health insights, and this signal is reproducible in a very large cohort of independent samples.

The choice of the eight oral pathway scores presented here was based on the biochemical and biological functions that underlie the disease phenotypes we focused on for this study – gingivitis, acid reflux, and nicotine addiction. While we focus on a few health conditions, the development and validation methodology presented here is general enough to quantify a large array of functional pathways. Indeed, we are currently in the process of developing and validating not only the full oral microbiome, but also the gut microbiome and the human blood transcriptome scores.

Our approach to validating oral pathway scores is to examine whether the scores differentiate cases and controls as expected; a case cohort with the disease phenotype of interest is compared with a healthy control cohort without that disease phenotype. The key question in this study, of course, is whether a signal found during score development is reproducible in independent collections of samples. Both the case and control validation cohorts come from a sample set that was independent and unseen at score development time. We recognize that there could be other ways to validate the mechanistic basis of these pathways, for example, via independent laboratory analysis of the resulting metabolite measurements for each pathway – this could be possible future work.

Among cases of gingivitis, three of the expected oral pathway scores demonstrated significant differences with respect to controls (Fig. 2), *OralLPSBiosynthesisPathways*, *OralPeriodontalMucinDegradationPathways*, and *OralPeriodontalProInflammatoryPathways*. These scores were designed to assess microbial activities which strongly align with symptoms of gingivitis (Table 2). In its earliest form, before progressing to full periodontal disease, gingivitis symptoms include irritation, redness, and swelling of the gums (inflammation), although these are reversible with proper care and maintenance. The *OralPeriodontalMucinDegradationPathways* score is focused on salivary barrier defenses, which preserve the health of the oral cavity, and mucin degradation is a known factor in periodontal disease [68–71]. The design of the *OralPeriodontalProInflammatoryPathways* score is heavily focused on both inflammation and periodontal destruction, which are defining symptoms of gingivitis and periodontal disease [72–74]. The *OralLPSBiosynthesisPathways* score overlaps with proinflammatory signals, but it is designed to typify inflammatory activities, specifically highlighting contributions of a detrimental Gram-negative biofilm [75–78]. As gingivitis develops into full periodontal disease, we anticipate other scores would gain significance, and we look forward to examining this in future work.

In looking at cases of acid reflux, all eight of the oral pathway scores demonstrated significant differences with respect to controls (Fig. 2). These results are consistent with an oral environment associated with acid reflux [64,65]. In addition, mounting evidence indicates that acid reflux (and the low pH of gastric acid) diminishes both dental health [79–81] and oral soft tissue health [82–84]. Reports also show disease associations increase with gastric acid contact [85,86]. Three of the oral pathway scores, *OralAmmoniaProductionPathways*, *OralCariogenicPathways*, and *OralUreaseActivityPathways*, are strongly aligned with the acidic environment fostered by the low pH of gastric acid. And, as already indicated, several scores are strongly aligned with gingival damage and inflammation. In future studies, we look forward to examining how interventions designed to address the low pH and/or contact time of gastric acid will impact the scores.

We also found significant differences for seven of the oral pathway scores between cases of nicotine addiction and controls (Fig. 2), and these results are consistent with an oral environment disrupted by tobacco products [62,63]. Among cases (individuals who self-reported a nicotine addiction) and controls, 92.5% and 8.4% were current tobacco users, respectively. Tobacco smoke contains over four thousand compounds, most of which are considered toxic in both gaseous and solid

form [87,88], and nicotine itself is known to impact many systems across the human body [89–91]. With regards to oral health, it is well-established that tobacco products have a detrimental impact on every oral condition as well as the success of many oral treatments [92]. Specifically, it has been shown that tobacco products can diminish salivary pH [93,94], reduce salivary flow [94,95], and exacerbate bone loss and inflammation [96–98]. Long term tobacco use is also known to increase risk for dental caries [99], periodontitis [100], oral cancer [101], and many other oral conditions [102,103]. Thus, the overall impact of tobacco-related oral conditions aligns with the design and performance of our scores.

Besides affecting oral health, there is increasing evidence that the oral microbiome is linked to systemic health. Our extensive catalog of over 500 systemic disease phenotype labels affords us the opportunity to explore connections to systemic health as an additional aspect of oral pathway scores. For example, with regards to the *OralUreaseActivityPathways* score, there are well-established connections between salivary urea levels and kidney function [104,105]. Therefore, during score development, we prioritize phenotypes like kidney disease and kidney cyst, which retain high explainability with the *OralUreaseActivityPathways* score. Given this, we can do a case/control difference analysis for *OralUreaseActivityPathways* in our independent validation cohort. Here, "case" samples are defined using an aggregate of multiple systemic disease phenotypes which are based on reported associations to urea or uric acid levels in saliva (or blood) [106,107], including: alcohol addiction [108,109], ankylosing spondylitis [110], diverticulosis [111], gout [107], hyperlipidemia [112,113], kidney cyst [114,115], kidney disease [116–118], pancreatitis [119,120]. "Control" samples are selected from those lacking all disease phenotype labels in the validation cohort. We did such an initial analysis, with case ($n = 388$) and control ($n = 3633$) matched (1:10) by age, sex, and BMI. Although this effort is outside the scope of the current paper, the result was a significant difference in score values ($p \leq 0.001$) using the Mann-Whitney U test with Benjamini-Hochberg correction for multiple hypothesis testing (FDR < 0.05). Similar analyses could be done for each score presented here and will be attempted as future work.

The study presented here has some limitations. Our validation employs case/control difference analyses which do not account for all possible confounders. Our analyses do not evaluate causality; the KO features that make up the scores are most likely a combination of causal and consequential. Prospective interventional trials will identify the causal features in follow-on studies. While our cohorts are very large and include many demographics, they may not perfectly represent every group in the USA or other countries. Finally, our metadata labels are based on self-reported data that may be misreported; however, given the size of our cohorts, such effects should be negligible.

One important objective of this work is to develop an iterative process for continuously improving the design of these scores (the selection of features and their weights). We will continue to improve the utility of the scores as we collect more data from more samples and individuals, and as the field of molecular science advances to include additional insights. The goal for each score is to provide the most meaningful health insights given the latest science and data.

Another important future objective of this work is to show the impact of lifestyle factors and specific interventions on functional pathways, over the course of a longitudinal re-testing regime. For example, there are several interventions for periodontal gum disease that could be initiated by an individual as part of a self-care routine, or by a dental professional who sees acute or chronic issues. An important question is how these oral pathway scores change longitudinally from a pre-intervention state to a post-intervention state. We look forward to evaluating the effect of a range of interventions on molecular functional pathway scores in the future.

5. Conclusions

With the increasing prevalence of many types of chronic diseases due to lifestyle factors, and the advent of broadly available molecular testing, it is important to establish a systematic methodology for assessing microbial functional pathways relevant to these diseases. Chronic diseases being complex and multifactorial, it is necessary for such a methodology to assess a group of related molecular markers together (rather than individual molecular markers). In this paper, we have presented such a methodology, in the context of the oral microbiome providing functional pathway insights associated with oropharyngeal disease phenotypes. We have described the development of eight oral functional pathway scores using a large cohort, and shown that the scores are able to distinguish oropharyngeal disease phenotype cases from controls in an unseen independent validation cohort. The same methodology can be applied to other contexts, such as the gut microbiome providing functional pathway insights associated with gastrointestinal and related disease phenotypes.

To our knowledge, these are the first reported wellness scores based on the oral metatranscriptome, and they are designed to provide molecular health insights from simple, non-invasive saliva samples. In this context, it is possible to aggregate many oral scores as presented here, into an overall score that represents oral health at the highest level; indeed, this is the approach we have taken in our implementation within our wellness product. Within this product, these scores also facilitate a personal timeline of health insights across multiple time points, for individuals who re-test their oral health. Furthermore, the health insights delivered by these oral pathway scores can drive hygiene, dietary, lifestyle, and pharmaceutical recommendations [121].

CRediT authorship contribution statement

Eric Patridge: Conceptualization, Data curation, Investigation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, and Writing – review & editing. **Anmol Gorakshakar:** Data curation, Investigation, Formal Analysis, Methodology, Software, Visualization, Writing – original draft, and Writing – review & editing. **Matthew M. Molusky:** Data curation, Investigation, Formal Analysis, Methodology, Software, Visualization, Writing – original draft, and Writing – review & editing. **Oyetunji Ogundijo:** Data curation, Formal analysis, Software. **Angel Janevski:** Software. **Cristina Julian:** Investigation, Validation, Writing – review & editing. **Lan Hu:** Investigation, Software, Writing - review & editing. **Momchilo Vuyisich:** Funding acquisition, Investigation, Resources, and Writing – review & editing. **Guruduth Banavar:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

No AI-assisted technology was used in the writing process.

Declaration of Competing Interest

The material represents original research, has not been previously published (is deposited as a preprint in **biorxiv.org**) and has not been submitted for publication elsewhere while under consideration in CSBJ. All authors of this manuscript were employees of Viome Life Sciences Inc at the time of their contributions, and held stock options in the company. MV and GB hold management positions within the company. The work was funded by Viome Life Sciences Inc.

Data Availability

The raw data used in this study cannot be shared publicly due to

privacy and legal reasons. However, if data is specifically requested, we may be able to share a summary and/or portions of the data. Researchers requiring more data for non-commercial purposes can request via: https://www.viomelifesciences.com/data-access. Viome may provide access to summary statistics through a Data Transfer Agreement that protects the privacy of participants' data.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.01.018.

References

[1] WHO. Report of the second meeting of the Strategic and Technical Advisory Group for Noncommunicable Diseases: virtual meeting [Internet]. 2022. Available from: ⟨https://apps.who.int/iris/bitstream/handle/10665/370424/9789240072176-eng.pdf⟩.
[2] Richards D. Oral diseases affect some 3.9 Billion people. Évid-Based Dent 2013;14(2). 35–35.
[3] Mobley C, Marshall TA, Milgrom P, Coldwell SE. The contribution of dietary factors to dental caries and disparities in caries. Acad Pedia 2009;9(6):410–4.
[4] Banavar G, Ogundijo O, Julian C, Toma R, Camacho F, Torres PJ, et al. Detecting salivary host and microbiome RNA signature for aiding diagnosis of oral and throat cancer. Oral Oncol 2023;145:106480.
[5] Kassebaum NJ, Smith AGC, Bernabé E, Fleming TD, Reynolds AE, Vos T, et al. Global, regional, and national prevalence, incidence, and disability-adjusted life years for oral conditions for 195 countries, 1990–2015: a systematic analysis for the global burden of diseases, injuries, and risk factors. J Dent Res 2017;96(4):380–7.
[6] Pitts NB, Zero DT, Marsh PD, Ekstrand K, Weintraub JA, Ramos-Gomez F, et al. Dental caries. Nat Rev Dis Prim 2017;3(1):17030.
[7] Cate JM ten. Contemporary perspective on the use of fluoride products in caries prevention. Br Dent J 2013;214(4):161–7.
[8] Welch JLM, Ramírez-Puebla ST, Borisy GG. Oral microbiome geography: micron-scale habitat and niche. Cell Host Microbe 2020;28(2):160–8.
[9] Arnold RR, Ribeiro AA. How fermented foods feed a healthy gut microbiota. Nutr. Continuum 2019:141–53.
[10] Sedghi L, DiMassa V, Harrington A, Lynch SV, Kapila YL. The oral microbiome: role of key organisms and complex networks in oral health and disease. Periodontol 2000 2021;87(1):107–31.
[11] Şenel S. An overview of physical, microbiological and immune barriers of oral mucosa. Int J Mol Sci 2021;22(15):7821.
[12] Mattos-Graner RO, Klein MI, Alves LA. The complement system as a key modulator of the oral microbiome in health and disease. Crit Rev Microbiol 2023:1–30.
[13] Ebersole JL, Peyyala R, Gonzalez OA. Biofilm-induced profiles of immune response gene expression by oral epithelial cells. Mol Oral Microbiol 2019;34:1.
[14] Pasman R, Krom BP, Zaat SAJ, Brul S. The role of the oral immune system in oropharyngeal candidiasis-facilitated invasion and dissemination of staphylococcus aureus. Front Oral Heal 2022;3:851786.
[15] Dong J, Li W, Wang Q, Chen J, Zu Y, Zhou X, et al. Relationships between oral microecosystem and respiratory diseases. Front Mol Biosci 2022;8:718222.
[16] Sudhakara P, Gupta A, Bhardwaj A. Oral dysbiotic communities and their implications in systemic diseases. Dent J 2018;6(2):10.
[17] Radaic A, Kapila YL. The oralome and its dysbiosis: new insights into oral microbiome-host interactions. Comput Struct Biotechnol J 2021;19:1335–60.
[18] Maddu N. Functions of Saliva [Internet]. Saliva and Salivary Diagnostics. IntechOpen; 2019. Available from: https://doi.org/10.5772/intechopen.84709.
[19] Deo PN, Deshmukh R. Oral microbiome: unveiling the fundamentals. J Oral Maxillofac Pathol Jomfp 2019;23(1):122–8 (Available from), ⟨https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6503789/⟩.
[20] Foo LH, Balan P, Pang LM, Laine ML, Seneviratne CJ. Role of the oral microbiome, metabolic pathways, and novel diagnostic tools in intra-oral halitosis: a comprehensive update. Crit Rev Microbiol 2021;47(3):359–75.
[21] Hampelska K, Jaworska MM, Babalska ZŁ, Karpiński TM. The role of oral microbiota in intra-oral halitosis. J Clin Med 2020;9(8):2484.
[22] Barrak I, Stájer A, Gajdács M. Urbán E. Small, but smelly: the importance of Solobacterium moorei in halitosis and other human infections. Heliyon 2020;6(10):e05371.
[23] Siddiqui R, Badran Z, Boghossian A, Alharbi AM, Alfahemi H, Khan NA. The increasing importance of the oral microbiome in periodontal health and disease. Futur Sci OA 2023;0(0):FSO856.
[24] Ray RR. Periodontitis: an oral disease with severe consequences. Appl Biochem Biotechnol 2023;195(1):17–32.
[25] Domnich M, Riedesel J, Pylaeva E, Kürten CHL, Buer J, Lang S, et al. Oral neutrophils: underestimated players in oral cancer. Front Immunol 2020;11:565683.
[26] Li Z, Liu Y, Zhang L. Role of the microbiome in oral cancer occurrence, progression and therapy. Micro Pathog 2022;169:105638.
[27] Iyer P. Oral cavity is the gateway to the body: role of oral health professionals: a narrative review. J Calif Dent Assoc 2023;51(1).
[28] Bansal M, Rastogi S, Vineeth NS. Influence of periodontal disease on systemic disease: inversion of a paradigm: a review. J Med Life 2012;6(2):126–30.
[29] Kane SF. The effects of oral health on systemic health. Gen Dent 2017;65(6):30–4.
[30] Machado V, Aguilera EM, Botelho J, Hussain SB, Leira Y, Proença L, et al. Association between periodontitis and high blood pressure: results from the study of periodontal health in almada-seixal (SoPHiAS). J Clin Med 2020;9(5):1585.
[31] Oral Spirito FD. and systemic health in the elderly. Appl Sci 2022;12(22):11718.
[32] Herrera D, Sanz M, Shapira L, Brotons C, Chapple I, Frese T, et al. Association between periodontal diseases and cardiovascular diseases, diabetes and respiratory diseases: Consensus report of the Joint Workshop by the European Federation of Periodontology (EFP) and the European arm of the World Organization of Family Doctors (WONCA Europe). J Clin Periodo 2023;50(6):819–41.
[33] Memon MA, Memon HA, Muhammad FE, Fahad S, Siddiqui A, Lee KY, et al. Aetiology and associations of halitosis: a systematic review. Oral Dis 2023;29(4):1432–8.
[34] Kudo Y, Kudo S, Miyachi H, Ichimasa K, Ogawa Y, Kouyama Y, et al. Changes in halitosis value before and after Helicobacter pylori eradication: a single-institutional prospective study. J Gastroenterol Hepatol 2022;37(5):928–32.
[35] Issitt T, Wiggins L, Veysey M, Sweeney ST, Brackenbury WJ, Redeker K. Volatile compounds in human breath: critical review and meta-analysis. J Breath Res 2022;16(2):024001.
[36] Gordon SC, Riedy CA, Stohler CS, Vujicic M. Trends in scope of practice for oral health care: future transformative effects. JDR Clin Transl Res 2022;7(1_suppl):31S–9S.
[37] Pittman TW, Decsi DB, Punyadeera C, Henry CS. Saliva-based microfluidic point-of-care diagnostic. Theranostics 2023;13(3):1091–108.
[38] Delrue C, Bruyne SD, Speeckaert MM. Unlocking the diagnostic potential of saliva: a comprehensive review of infrared spectroscopy and its applications in salivary analysis. J Pers Med 2023;13(6):907.
[39] Legg M. The role of informatics in the shift from reactive to proactive healthcare. EPMA J 2014;5(Suppl 1):A50.
[40] Polverini PJ. Personalized oral health care. Concept Des Clin Pract 2016:145–53.
[41] Polverini PJ. Personalized medicine and the future of dental practice. Pers Med 2018;15(6):449–51.
[42] Pritchard DE, Moeckel F, Villa MS, Housman LT, McCarty CA, McLeod HL. Strategies for integrating personalized medicine into healthcare practice. Pers Med 2017;14(2):141–52.
[43] Toma R, Cai Y, Ogundijo O, Hu L, Gline S, Demusaj D, et al. A clinically validated human saliva metatranscriptomic test for global systems biology studies. Biotechniques 2023;74(1):31–44.
[44] Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. Metatranscriptomics of the human oral microbiome during health and disease. Mbio 2014;5(2):e01012–4.
[45] Jovel J, Nimaga A, Jordan T, O'Keefe S, Patterson J, Thiesen A, et al. Metagenomics versus metatranscriptomics of the murine gut microbiome for assessing microbial metabolism during inflammation. Front Microbiol 2022;13:829378.
[46] Visconti A, Roy CIL, Rosa F, Rossi N, Martin TC, Mohney RP, et al. Interplay between the human gut microbiome and host metabolism. Nat Commun 2019;10(1):4505.
[47] Eisenstein M. The hunt for a healthy microbiome. Nature 2020;577(7792):S6–8.
[48] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000 1;28(1):27–30.
[49] Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci 2019;28(11):1947–51.
[50] Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol 2021;39(1):105–14.
[51] Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res 2017;27(5):849–64.
[52] Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Mering C von, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol 2017;34(8):2115–22.
[53] Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res 2016;26(12):1721–9.
[54] Henderson NC, Varadhan R. Damped anderson acceleration with restarts and monotonicity control for accelerating EM and EM-like algorithms (Available from) J Comput Graph Stat [Internet] 2019;28(4):834–46. https://doi.org/10.1080/10618600.2019.1594835.
[55] Keys KL, Mak ACY, White MJ, Eckalbar WL, Dahl AW, Mefford J, et al. On the cross-population generalizability of gene expression prediction models. PLoS Genet 2020;16(8):e1008927.
[56] Kelly DE, Hansen MEB, Tishkoff SA. Global variation in gene expression and the value of diverse sampling. Curr Opin. Syst Biol 2017;1:102–8.
[57] Daca-Roszak P, Zietkiewicz E. Transcriptome variation in human populations and its potential application in forensics. J Appl Genet 2019;60(3–4):319–28.
[58] Lee YH, Hong JY. Oral microbiome as a co-mediator of halitosis and periodontitis: a narrative review. Front Oral Heal 2023;4:1229145.
[59] Lin TC, Tseng CF, Wang YH, Yu HC, Chang YC. Patients with chronic periodontitis present increased risk for primary Sjögren syndrome: a nationwide population-based cohort study. PeerJ 2018;6:e5109.
[60] Lechien JR, Chiesa-Estomba CM, Henriquez CC, Mouawad F, Ristagno C, Barillari MR, et al. Laryngopharyngeal reflux, gastroesophageal reflux and dental disorders: a systematic review. PLoS One 2020;15(8):e0237581.

[61] Yost S, Duran-Pinedo AE, Teles R, Krishnan K, Frias-Lopez J. Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis. Genome Med 2015;7(1):27.
[62] Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, et al. Cigarette smoking and the oral microbiome in a large study of American adults. ISME J 2016;10(10):2435–46.
[63] Kumar PS, Matthews CR, Joshi V, Jager M de, Aspiras M. Tobacco smoking affects bacterial acquisition and colonization in oral biofilms. Infect Immun 2011;79(11):4730–8.
[64] Qian J, Yang M, Xu D, Zhang G, Cai Y, Yang B, et al. Alterations of the salivary microbiota in gastroesophageal reflux disease. J Oral Biosci 2023.
[65] Yang L, Lu X, Nossa CW, Francois F, Peek RM, Pei Z. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. Gastroenterology 2009;137(2):588–97.
[66] Heintz-Buschart A, Wilmes P. Human gut microbiome: function matters. Trends Microbiol 2018;26(7):563–74.
[67] Djemiel C, Maron PA, Terrat S, Dequiedt S, Cottin A, Ranjard L. Inferring microbiota functions from taxonomic genes: a review. GigaScience 2022;11(1):giab090.
[68] Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA. Mucins in the mucosal barrier to infection. Mucosal Immunol 2008;1(3):183–97.
[69] Kejriwal S, Bhandary R, Thomas B, Kumari S. Estimation of levels of salivary mucin, amylase and total protein in gingivitis and chronic periodontitis patients. J Clin Diagn Res 2014;8(10):ZC56–60.
[70] Giannobile WV, Beikler T, Kinney JS, Ramseier CA, Morelli T, Wong DT. Saliva as a diagnostic tool for periodontal disease: current state and future directions. Periodontol 2000 2009;50(1):52–64.
[71] Stafford GP, Sharma A. Periodontal pathogen sialometabolic activity in periodontitis (Available from) Cham: Springe Int Publ 2020:187–94. https://doi.org/10.1007/978-3-030-42990-4_12.
[72] Pérez-Carrasco V, Uroz-Torres D, Soriano M, Solana C, Ruiz-Linares M, Garcia-Salcedo JA, et al. Microbiome in paired root apices and periapical lesions and its association with clinical signs in persistent apical periodontitis using next-generation sequencing. Int Endod J 2023;56(5):622–36.
[73] Kato T, Takahashi N, Kuramitsu HK. Sequence analysis and characterization of the Porphyromonas gingivalis prtC gene, which expresses a novel collagenase activity. J Bacteriol 1992;174(12):3889–95.
[74] Loos BG, Dyke TEV. The role of inflammation and genetics in periodontal disease. Periodontol 2000 2020;83(1):26–39.
[75] Mcilwaine C, Strachan A, Harrington Z, Jerreat M, Belfield LA, Sandor V, et al. Comparative analysis of total salivary lipopolysaccharide chemical and biological properties with periodontal status. Arch Oral Biol 2020;110:104633.
[76] Jain S, Darveau RP. Contribution of Porphyromonas gingivalis lipopolysaccharide to periodontitis. Periodontol 2000 2010;54(1):53–70.
[77] Pussinen PJ, Kopra E, Pietiäinen M, Lehto M, Zaric S, Paju S, et al. Periodontitis and cardiometabolic disorders: the role of lipopolysaccharide and endotoxemia. Periodontol 2000 2022;89(1):19–40.
[78] Zhang D, Chen L, Li S, Gu Z, Yan J. Lipopolysaccharide (LPS) of Porphyromonas gingivalis induces IL-1β, TNF-α and IL-6 production by THP-1 cells in a way different from that of Escherichia coli LPS. Innate Immun 2008;14(2):99–107.
[79] Picos AM, Poenar S, Opris A, Chira A, Bud M, Berar A, et al. Prevalence of dental erosions in GERD: a pilot study. Clujul Méd (1957) 2013;86(4):344–6.
[80] Linnett V, Seow W, Connor F, Shepherd R. Oral health of children with gastro-esophageal reflux disease: a controlled study. Aust Dent J 2002;47(2):156–62.
[81] Ersin NK, Onçağ O, Tümgör G, Aydoğdu S, Hilmioğlu S. Oral and dental manifestations of gastroesophageal reflux disease in children: a preliminary study. Pedia Dent 2006;28(3):279–84.
[82] Song JY, Kim HH, Cho EJ, Kim TY. The relationship between gastroesophageal reflux disease and chronic periodontitis. Gut Liver 2014;8(1):35–40.
[83] Watanabe M, Nakatani E, Yoshikawa H, Kanno T, Nariai Y, Yoshino A, et al. Oral soft tissue disorders are associated with gastroesophageal reflux disease: retrospective study. BMC Gastroenterol 2017;17(1):92.
[84] Jajam M, Bozzolo P, Niklander S. Oral manifestations of gastrointestinal disorders. J Clin Exp Dent 2017;9(10):e1242–8.
[85] Valena V, Young W. Dental erosion patterns from intrinsic acid regurgitation and vomiting. Aust Dent J 2002;47(2):106–15.
[86] Johansson A, Lingström P, Imfeld T, Birkhed D. Influence of drinking method on tooth-surface pH in relation to dental erosion. Eur J Oral Sci 2004;112(6):484–9.
[87] Maiyo AK, Kibet JK, Kengara FO. A review of the characteristic properties of selected tobacco chemicals and their associated etiological risks. Rev Environ Heal 2023;38(3):479–91.
[88] G T, O P, M AW, P K, E GT, A S. The effect of nicotine on oral health. Balt J Health Phys Act 2018;2(10):7–13.
[89] Mishra A, Chaturvedi P, Datta S, Sinukumar S, Joshi P, Garg A. Harmful effects of nicotine. Indian J Méd Paediatr Oncol 2015;36(01):24–31.
[90] Audrain-McGovern J, Benowitz N. Cigarette smoking, nicotine, and body weight. Clin Pharm Ther 2011;90(1):164–8.
[91] Benowitz NL, Burbank AD. Cardiovascular toxicity of nicotine: implications for electronic cigarette use. Trends Cardiovasc Med 2016;26(6):515–23.
[92] Chaffee BW, Couch ET, Vora MV, Holliday RS. Oral and periodontal implications of tobacco and nicotine products. Periodontol 2000 2021;87(1):241–53.
[93] Grover N, Sharma J, Sengupta S, Singh S, Singh N, Kaur H. Long-term effect of tobacco on unstimulated salivary pH. J Oral Maxillofac Pathol: JOMFP 2016;20(1):16–9.
[94] Kanwar A, Sah K, Grover N, Chandra S, Singh RR. Long-term effect of tobacco on resting whole mouth salivary flow rate and pH: An institutional based comparative study. Eur J Gen Dent 2013;2(03):296–9.
[95] Rad M, Kakoie S, Brojeni FN, Pourdamghan N. Effect of long-term smoking on whole-mouth salivary flow rate and oral health. J Dent Res, Dent Clin, Dent Prospect 2010;4(4):110–4.
[96] Park G, Kim Y, Kang K, Bae S, Baek H, Auh Q, et al. Effects of sirtuin 1 activation on nicotine and lipopolysaccharide-induced cytotoxicity and inflammatory cytokine production in human gingival fibroblasts. J Periodontal Res 2013;48(4):483–92.
[97] Johnson GK, Guthmiller JM, Joly S, Organ CC, Dawson DV. Interleukin-1 and interleukin-8 in nicotine- and lipopolysaccharide-exposed gingival keratinocyte cultures. J Periodontal Res 2010;45(4):583–8.
[98] Rosa GM, Lucas GQ, Lucas ON. Cigarette smoking and alveolar bone in young adults: a study using digitized radiographs. J Periodo 2008;79(2):232–44.
[99] Jiang* X, Jiang* X, Wang# Y, Huang# R. Correlation between tobacco smoking and dental caries: a systematic review and meta-analysis. Tob Induc Dis 2019;17(April):34.
[100] Leite FRM, Nascimento GG, Scheutz F, López R. Effect of smoking on periodontitis: a systematic review and meta-regression. Am J Prev Med 2018;54(6):831–41.
[101] Chaturvedi P, Singh A, Chien CY, Warnakulasuriya S. Tobacco related oral cancer. BMJ 2019;365:l2142.
[102] Patil PB, Bathi R, Chaudhari S. Prevalence of oral mucosal lesions in dental patients with tobacco smoking, chewing, and mixed habits: a cross-sectional study in South India. J Fam Community Med 2013;20(2):130–5.
[103] Warnakulasuriya S, Dietrich T, Bornstein MM, Peidró EC, Preshaw PM, Walter C, et al. Oral health risks of tobacco use and effects of cessation. Int Dent J 2010;60(1):7–30.
[104] Pandya D, Nagrajappa AK, Ravi KS. Assessment and correlation of urea and creatinine levels in saliva and serum of patients with chronic kidney disease, diabetes and hypertension– a research study. J Clin Diagn Res 2016;10(10):ZC58–62.
[105] Kovalčíková AG, Pavlov K, Lipták R, Hladová M, Renczés E, Boor P, et al. Dynamics of salivary markers of kidney functions in acute and chronic kidney diseases. Sci Rep 2020;10(1):21260.
[106] Jaiswal A, Madaan S, Acharya N, Kumar S, Talwar D, Dewani D. Salivary uric acid: a noninvasive wonder for clinicians? Cureus 2021;13(11):e19649.
[107] Riis JL, Bryce CI, Matin MJ, Stebbins JL, Kornienko O, Huisstede L van, et al. The validity, stability, and utility of measuring uric acid in saliva. Biomark Med 2018;12(6):583–96.
[108] Nakamura K, Sakurai M, Miura K, Morikawa Y, Yoshita K, Ishizaki M, et al. Alcohol intake and the risk of hyperuricaemia: a 6-year prospective study in Japanese men. Nutr, Metab Cardiovasc Dis 2012;22(11):989–96.
[109] Yamamoto T, Moriwaki Y, Takahashi S. Effect of ethanol on metabolism of purine bases (hypoxanthine, xanthine, and uric acid). Clin Chim Acta 2005;356(1–2):35–57.
[110] Sargın B, Gürer G, Bozbaş GT, Öztürk H. Association between serum uric acid and inflammation markers in ankylosing spondylitis patients treated with tumor necrosis factor-α or nonsteroidal anti-inflammatory drugs (Available from) Eur Res J [Internet] 2019;5(1):104–8. https://doi.org/10.18621/eurj.359504.
[111] Hu JM, Chang YT, Shih CW, Hsu CH, Lin TC, Lai CY, et al. Association between serum uric acid levels and colonic diverticulosis in terms of sex. PLoS ONE 2022;17(8):e0269978.
[112] Yu W, Cheng JD. Uric acid and cardiovascular disease: an update from molecular mechanism to clinical perspective. Front Pharm 2020;11:582680.
[113] Peng TC, Wang CC, Kao TW, Chan JYH, Yang YH, Chang YW, et al. Relationship between hyperuricemia and lipid profiles in US adults. BioMed Res Int 2015;2015:127596.
[114] Zhao D, Chen M, Yang X, Xie J, Wang S. Association between serum uric acid levels and simple renal cyst risk in a nondiabetic population: a nested case-control study. Clin Chim Acta 2023;540:117237.
[115] Han Y, Zhang M, Lu J, Zhang L, Han J, Zhao F, et al. Hyperuricemia and overexcretion of uric acid increase the risk of simple renal cysts in type 2 diabetes. Sci Rep 2017;7(1):3802.
[116] Miyata Y, Obata Y, Mochizuki Y, Kitamura M, Mitsunari K, Matsuo T, et al. Periodontal disease in patients receiving dialysis. Int J Mol Sci 2019;20(15):3805.
[117] Šedý J, Horká E, Foltán R, Špačková J, Dušková J. Mechanism of increased mortality in hemodialysed patients with periodontitis. Méd Hypotheses 2010;74(2):374–6.
[118] Craig R. Interactions between chronic renal disease and periodontal disease. Oral Dis 2008;14(1):1–7.
[119] Su A, Yuan X, Zhu G, Jiang X, Shu R, Yang W, et al. Association between baseline uric acid and the risk of acute pancreatitis. Pancreas 2022;51(8):966–71.
[120] H Y, F Z, S T, C L, S J. Correlation of uric acid with disease severity, prognosis, and complications in patients with severe acute pancreatitis. Front Med Sci Res 2023;5(4).
[121] Connell J, Toma R, Ho C, Shen N, Moura P, Cai Y, et al. Evidence-based precision nutrition improves clinical outcomes by analyzing human and microbial molecular data with artificial intelligence. Biorxiv 2021. 2021.04.24.441290.