ORALMICROBIOME SIGNATURE FOR HEMATOLOGICAL CANCER

by

Holden Christopher Langdon

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Biology

Charlotte

2020

Approved by:

_____

Dr. Christine Richardson

_____

Dr. Farah Mougeot

_____

Dr. Jean-Luc Mougeot

_____

Dr. Eugenia Lo

ABSTRACT

HOLDEN CHRISTOPHER LANGDON. Oral Microbiome Signature for Hematological Cancer (Under the direction of Dr. FARAH MOUGEOT and Dr. JEAN-LUC MOUGEOT)

Background

The endogenous microbiome of healthy individuals in the oral cavity is diverse, representing over 700 different bacterial species. Some of these species may become opportunistic if certain elements in the microenvironment and host-response in the oral cavity are altered. Imbalance in microbiome composition, defined by higher or lower levels of relative abundance and microbial gene expression changes, has been linked to different forms of hematological cancer.

Hypothesis

We hypothesize we will identify unique oral microbiome profiles in hematological cancer patients when compared to healthy controls. Additionally, we expect to determine significant differences in the oral microbiome *beta*-diversity of lymphoma patients when compared to acute myelogenous leukemia (AML) patients.

Objective

1) To identify unique oral microbiome profiles of hematological cancer patients when compared to healthy control subjects.
2) To compare the oral microbiome profiles of lymphoma patients and AML patients prior to cancer treatment.

Methods

Saliva samples and swabs of buccal mucosa, supragingival plaque and tongue were collected from hematological cancer patients (N=51), prior to cancer treatment, and healthy control subjects (N=38). Next generation sequencing (16S-rRNA gene V3-V4

region) was used to determine the relative abundance of bacterial taxa present at the genus and species levels. Differences in oral microbiome *beta*-diversity were tested using PERMANOVA by comparing hematological cancer patients *vs.* healthy controls patients (N=38) [Monte-Carlo corrected p values (α=0.05)]. Linear discriminant analysis (LDA) effect size (LEfSe; Log LDA threshold >0.005) analysis was performed to identify differentiating bacterial genus and species probes for the above mentioned pairwise comparisons.

Results

There were significant differences in the oral microbiome *beta*-diversity of hematological cancer patients compared to healthy controls (p=0.0001). LEfSe analysis showed significant LDA scores for 51 probes differentiating hematological cancer patients from healthy controls. Furthermore, there was a significant difference in the beta diversity of the oral microbiome of both lymphoma and AML patients when compared to each other as well as to healthy control subjects.

Conclusion

Hematological cancer patients had distinct oral microbial profiles compared to healthy controls. Additionally, oral microbiome profiles of lymphoma and AML patients were significantly different. Further investigation into the mechanistic interaction of the oral microbiome with the microenvironment in oral cavity and the host immune system is warranted.

ACKNOWLEDGEMENTS

I would like to express my appreciation to my advisors and directors of research Dr. Farah Mougeot and Dr. Jean-Luc Mougeot for their continuous support of my Master study and research. It was their mentoring and constant availability that allowed this project to be possible. I also want to thank all additional members of the Oral Medicine Laboratory at Carolinas Medical Center, Micaela Beckman, Craig Stevens and Darla Morton for their continued guidance and assistance with this project. Additionally, I would like to thank the oral medicine clinical staff at Carolinas Medical Center as well as Forsyth Institute. Furthermore, I would like to thank my family for their continued support and encouragement throughout past and present programs.

Finally, I would like to express my thankfulness to my committee members Dr. Christine Richardson and Dr. Eugenia Lo for their continued support and mentoring throughout this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: BACKGROUND AND SIGNIFICANCE

1.1 Background

Cancer

This year in the United States alone an estimated 1.8 million people will be diagnosed with cancer with 600,000 people projected to succumb from cancer-related complications (16). Because of this, cancer has been a leading topic among researchers for the past 25 years since the advent of modern molecular and cellular biology. Before we are able to discuss the complexity of cancer formation, it is important to understand the different forms of cancer and their cellular basis. In its most basic form, cancer is the continual unregulated proliferation of any cell resulting in the invasion of healthy tissue (19). This results in more than one hundred different forms of cancer which all fall into four main groups; carcinomas, sarcomas, and hematological cancers such as leukemias and lymphomas (19). Carcinomas originate from epithelial tissue and are the most abundant form of cancer consisting of approximately 90 percent of all diagnosis. Sarcomas, being the rarest form of cancer, originate from connective tissue such as muscle and bone. Lastly, lymphomas and leukemias account for roughly eight percent of all diagnosis and originate from blood forming cells and cells of the immune system.

A fundamental characteristic of all cancer is the development of a tumor from a single progenitor cell with an abnormal proliferation rate. The progenitor cell that gives rise to the tumor is not necessarily classified as a cancer cell, as it may have not acquired all the proper characteristics. Through a series of alterations, the cellular profile begins to favor increased capacity for proliferation and invasion ultimately resulting in cancer formation. This is normally a multistage process occurring over many years as indicated by most diagnoses occurring in older individuals.

Over the past decade there have been many advancements in cancer diagnostics as well as treatment protocols for patients. Even with these advancements, there are still many shortcomings in understanding the true underlying causes and factors affecting the progression of many cancers. This is in part due to the lack of knowledge on the interaction of the human microbiome and carcinogenesis (8). With current advancements being made in genomic analysis, new avenues are able to be explored that before may have not been possible such as the determination of microbial profiles of cancer patients.

16S rRNA Gene

The ability to identify microbiota at the species and genus level has become increasingly more efficient and cost effective over the past number of years. This is primarily due to the advancement in Next Generation Sequencing (NGS) technology. At the forefront of this research has been the idea that ribosomal RNA (rRNA) genes are highly conserved and evolutionarily stable, except for their hypervariable region making them ideal for microbial genus and species identification. Due to the attention 16S rRNA has received and advancements in genome sequencing technology, the current rRNA gene databases have been significantly improved making phylogenetic studies more efficient. A specific focus of ribosomal genome sequencing has been the 16S rRNA gene making it an ideal candidate for universal primer identification. The 16S rRNA gene contains nine hypervariable regions totaling 1500bp in length, with each region having the ability to determine certain taxonomical relationships. The most prominently used region is the V4 region, which is less than 300 bp long and can determine majority of identified microbes down to the genus and species level.

Human Microbiome

It was not until 2008 that the Human Microbiome Project was launched with a goal of obtaining a better understanding of the microbial flora responsible for health and

disease (15). Recent advancements in NGS have allowed us to analyze microbiota, including species that have not yet been isolated from live cultures (17). In conjunction with the Human Microbiome Project, there have been several breakthroughs in cancer research concerning a possible link between cancer and microbiome. It is important to understand how the human microbiome contributes to diseases such as cancers, human papillomavirus (HPV) infections and heart disease, amongst many others, but also how the human microbiome contributes to health.

Oral Microbiome

The oral cavity of healthy individuals has a diverse microbial profile, representing over 700 different bacterial species (1). Many of these species may become opportunistic if certain elements of the microenvironment and host-response in the oral cavity are altered. There are four main phyla of oral microbiota that play a role in health and disease: Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria (7). More specifically, the most prevalent genera in the oral cavity of healthy subjects are *Prevotella*, *Neisseria*, *Streptococcus*, *Haemophilus*, and *Fusobacterium* (18). However, an imbalance of each of these genera, *i.e.*, higher or lower levels of relative abundance and associated gene expression are linked to many different forms of cancers from hematological to solid state tumors (10,12,13).

Oral Microbiome and Cancer

a) Solid State Cancers

Due to the lack of diagnostic tools and low survival rate, pancreatic cancer has been widely studied within the microbial science community. Studies have noted significant increases in 31 different bacterial species in saliva samples collected from pancreatic cancer patients (5). Among those with increased levels are *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans*, both found to be known causes of tooth decay and heart disease (4). It is believed that the increased chance of pancreatic

cancer and liver cirrhosis is possibly due to an elevation in the blood serum antibodies against *P. gingivalis* (5). Another prevalent genus of bacterium was found to be *Fusobacterium*, present in tongue microbiome. This genus was found not to be a good biomarker for diagnosis, but was associated with a worse prognosis (11).

While *Fusobacterium* is a poor biomarker for pancreatic cancer, it has been identified as an indicator of colorectal carcinogenesis. *F*. *nucleatum* is present not only in the stool samples but also in tumor biopsies of colorectal cancer patients (2). This bacterium has become such a consistent biomarker for colorectal cancer that it has begun to be used in screening patients and giving prognosis in Japan (9).

Gastric cancer has also been linked to changing levels in the four main phyla of microbiota in oral cavity. There is a significant increase, roughly 18 percent, in *Proteobacteria* in gastric cancer patients when compared to healthy controls (6). In addition to the increased *Proteobacteria,* there was a decrease in *Fusobacterium*, *Neisseria*, and *Haemophilus* by roughly six percent with the exception of *Porphyromonas* which decreased by three percent (6).

Esophageal cancer has also shown a link between the oral microbiome and carcinogenesis. When comparing patients with esophageal squamous cell carcinomas to healthy controls, there was an increase in *Prevotella, Streptococcus, and Porphyromonas* (3). Through further analysis, researchers were able to predict the pathway through which *Streptococcus* contributes to carcinogenesis of esophageal tissue. As with many carcinogens, the carcinogenesis pathway activation occurs through the initiation of inflammation and development of dysplasia (3).

While solid state tumors have been studied for linkages between the oral microbiome and carcinogenesis over the recent years, there are still areas that lack significant research. For example, there is little evidence for lung and breast cancers showing associations with possible microbiome biomarkers for these diseases.

b) Hematological Cancers

Very little research has focused on the association of the oral microbiome with hematological cancers. The data that has been collected on blood cancers such as multiple myeloma, acute lymphoblastic leukemia, and many others suggest that the human microbiota could be playing a role in tumorigenesis or the prevention of that thereof. While there has been evidence showing that leukemia patients have a less diverse microbiome, there has not been significant evidence showing specific bacterial associations between the two (14).

1.2 Significance

While cancer development has been a leading topic of research for many years, there is still a large deficit of information regarding mechanisms of pathogenesis and host response in conjunction with the role of oral microbiome in carcinogenesis. The understanding of the complex interaction and cohabitation of the oral microbiome with the host is largely based on an understudied theory on tumor formation and progression. While there has been some research done on such correlation, the sample sizes used in previous studies were too small to draw a reliable conclusion (20). With our sample size, hematological cancer (n=148 Samples, N= 51 patients) and healthy control (n=141 samples, N= 38 patients), we will be able to conclude with higher confidence that our results pertain to a larger cancer population.

Our main objective is to show an association between oral microbiome profiles and hematological cancer types. Due to the complex and diverse genetic make-up of cancer, the investigation of cancer formation pathways in the context of the oral microbiome is challenging. However, once an association has been established, certain mechanisms for tumorigenesis could be investigated in the context of host response involving the innate and adaptative immune system. Ultimately, biomarkers may be identified for diagnosis and for the prediction or monitoring of rate of progression and

response to treatment. Thus, this research provided a better understanding of the oral microbiome in both healthy controls and hematological cancer populations.

Hypothesis: We expect to identify unique oral microbiome profiles in hematological cancer patients compared to healthy controls. Additionally, we expect to see significant differences in the oral microbiome *beta*-diversity of lymphoma patients when compared to AML patients.

1.3 Objectives

Objective 1) To compare oral microbiome profiles between hematological cancer patients and healthy controls prior to cancer treatment.

Objective 2) To compare the oral microbiome profiles of AML and lymphoma patients prior to cancer treatment.

CHAPTER 2: MATERIALS AND METHODS

2.1 Patient Recruitment

Patients (N=51; 19 females, 32 males; age range= 18-76) diagnosed with hematological cancers and were to receive conditioning treatment to prepare for hematopoietic stem cell transplant were recruited by Carolinas Medical Center, Charlotte, North Carolina. Additionally, healthy control subjects (N=38; 27 females, 11 males; age range= 24-84) were recruited to obtain an oral microbial profile baseline. Hematological cancer diagnosis included Acute Lymphoblastic Leukemia (ALL), Acute Myelogenous Leukemia (AML), Lymphoma, Myelodysplastic Syndrome (MDS), Myelofibrosis and Myeloma. Due to data collection being used initially for a longitudinal study, samples were collected before conditioning therapy at baseline, one to three weeks post-transplantation, and one-year post-transplantation. For the purpose of our research we analyzed only those at baseline (pre-cancer treatment).

2.2 Sample Collection and Processing

Oral samples (saliva sample, buccal mucosa, tongue and superficial supragingival plaque swabs) were collected from each patient at baseline (pre-cancer treatment). Initial saliva collection was performed while chewing unsweetened and unflavored gum (The Wrigley Company, Mars, Inc., Chicago, IL) for a period of two minutes into a 50mL conical BD Falcon polypropylene centrifuge tube (Corning, NY), and kept on ice for no longer than 30 minutes before processing or being stored at -80°C.

Buccal mucosal samples were collected by swabbing both sides of the buccal mucosa for 10 seconds each. Tongue samples were obtained by swabbing a 1 cm^2 region on both sides of the middorsal region of the tongue for 5 seconds. Finally, the superficial supragingival plaque (SSP) samples were obtained using OmniSwabs (GE

Life Sciences-Buckinghamshire, UK) across the lateral surfaces of all maxillary and mandibular teeth at the junction of the tooth and gingiva.

Bacterial genomic DNA was extracted from oral samples using QIAamp DNA Mini Kit procedure (QIAGEN, Valencia, CA) per manufacturer's instructions. Identification of bacterial genus and species utilized Human Oral Microbe Identification using Next Generation Sequencing (HOMINGS), which employs a ProbeSeq program for species detection through recognition of the 16S rRNA gene (V3-V4 region). During sample preparation, 50ng of genomic DNA was used for PCR in which the 16s rRNA (V3-V4) region was amplified, followed by purification and processing methods described by Caporaso et al. using Miseq (Illumina, Inc., San Diego, CA.)(18). ProbeSeq sequence identification used rRNA-based in silico probes in a BLAST program to determine the species frequency. The sequence-reads matched one of the 737 probes (620 ProbeSeq species probes and 117 genus probes). Results were provided as Excel spreadsheets displaying total probe hits (number of matches per oligomer) ranging from 0 – 300,000.

2.3 Analysis Preparation

Results for each patient set (hematological patients and healthy controls) were provided in Excel files (2 hematological cancer files, 1 healthy control file). For the purpose of our study, we were only interested in baseline timepoint. Once all data was extracted from each patient set, sorted and combined into two separate excel files, i.e. one for hematological cancer patients and one for healthy control subjects, we converted all data to relative abundance by dividing each probe by the total number of hits per patient. Each patient set was also associated with demographic information that had to be added to each Excel file once the conversion to relative abundance was complete. Additionally, all Excel files needed to be individually formatted according to Primer v7 program (PRIMER-E Ltd., Ivybridge, UK).

2.4 Statistical Analysis

A chi-squared test was run on each group to ensure there were no significant differences in frequencies of saliva sample, buccal mucosa, tongue and superficial supragingival plaque swabs between the three groups.

a) PERMANOVA

Primer v7 program (PRIMER-E Ltd., Ivybridge, UK) was used to run a multivariate permutational analysis of variance (PERMANOVA) using reduced-model with unrestricted permutation of raw data, 9,999 permutations and type III partial sum of squares. Four different groupings were analyzed, hematological cancer vs. healthy control (**G1**), AML *vs.* healthy control (**G2**), Lymphoma vs. healthy control (**G3**) and AML *vs.* Lymphoma (**G4**) in order to compare the beta-diversity in a cross-sectional analysis of relative abundance data. RA data were square root transformed and converted to Bray-Curtis similarity matrices prior to PERMANOVA analysis. Each PERMANOVA grouping was run from paired data of the four oral site samples, namely saliva (S), buccal mucosa (B), tongue (T) and superficial supragingival plaque (P). For all groupings, PERMANOVA Monte-Carlo-corrected p-values ($p<0.05$) were obtained using both fixed and random variables. The factors “condition” (Hematological cancer and Healthy control) and “sample site” (S, B, T, P) were set as fixed variables, while “diagnosis” (ALL, AML, Lymphoma, MDS, Myelofibrosis, Myeloma) and “antibiotic treatment were set as random variables for G1. The factors “condition” and “sample site” were set as fixed variables for G2, G3, and G4. For the analysis involving random variables *(i.e.*, antibiotic treatment, cancer stage), these were nested in the factor “condition”. For depiction of PERMANOVA design refer to **(Figure 1)**.

b) *Principal coordinate analysis (PCoA) and non-metric multidimensional scaling (nMDS)*

For the analyses that were statistically significant to visualize the comparisons, non-metric multi-dimensional scaling nMDS and PCoA plots were formed from RA data using Primer v7 (PRIMER-E Ltd., Ivybridge, UK).

c) Linear discriminate analysis (LDA) effective size (LEfSe)

Analysis was performed on each grouping using online tool Galaxy v1.0, in order to determine differentiating features at the genus and species level, as demonstrated by Segata et al. (21). Taxonomy levels were manually added to all groupings within the text files that were formatted according to Galaxy V1.0 required formatting. Data formatting and input consisted of diagnosis (hematological cancer and healthy control) as class and patient ID as subject. Analysis strategy 'one-against-all' was used for multiclass analysis, the factorial Kruskal-Wallis test as well as pairwise Wilcoxon signed-rank test for all groupings was set at a Monte-Carlo (α=0.05). Results were displayed as histograms and cladograms representing taxa with an LDA>4.0 threshold for groupings (G1, G2, G3) and LDA>3.0 for grouping (G4).

CHAPTER 3: RESULTS

3.1 Abundance Data and Species Detection

ProbSeq results representing matched sequence reads of the 737 total probes (620 species and 117 genus probes) for hematological cancer patients and healthy control subjects are represented in **(Supplemental Table 1)**. Relative abundance data was used to analyze most prevalent species and genus for hematological cancer patients, as well as healthy control subjects. The most abundant genus present in both hematological cancer and healthy control subjects was *Streptococcus*. Furthermore, *Rothia dentocariosa* accounted for the largest difference in relative abundance between hematological cancer and healthy controls **(Figure 2)**.

3.2 Cross-sectional Analysis

Hematological cancer patients had a distinct oral microbial profile when compared to healthy controls (p=0.0001). Initial analysis of G2 comparison noted significant interaction between condition and sample site, possibly indicating an imbalance in the two groupings. Secondary analysis of G2 was performed after sample data was balanced, *i.e*. sample number and male to female distribution with new sample data as follows; AML (n= 56 samples M= 32, F= 24) and healthy control (n=56 samples, M= 32, F= 24). The secondary analysis indicated a significant difference in oral microbiome beta-diversity when comparing AML to healthy control subjects (p=0.0001). Additionally, G3 and G4 analyses indicated a significant difference in the oral microbial profile of Lymphoma patients when compared to healthy control subjects (p=0.0001), as well as AML patients (N=20 patients; n=56) (p=0.0148). As expected, there were significant differences in the oral microbial beta-diversity when comparing sample sites (p=0.0001), as well as comparison of individuals that had been treated with antibiotics to

those who had not for all patient sets (hematological cancer and healthy control) (p=0.0245).

Significant comparisons are expressed in PCoA plots and nMDS plots with the least amount of variation observed among G4 (AML vs lymphoma), with a total of three principal coordinates of 36.3 percent. The highest variation was observed among G3 (lymphoma vs helathy control), with a total of three principal coordinates of 40.8 percent. **(Figure 3 and 4)**

Detailed PERMANOVA results

G1: Hematological cancer *vs*. healthy control subjects

- Analysis 1- Condition- fixed, Sample Site- fixed, Diagnosis random (Sa, Co)
    - Co 0.0001 (main variable)
    - Sa 0.0001, Di 0.0107 (secondary variables)
- Analysis 2- Condition- fixed, Sample Site- fixed, Antibiotics random (Sa, Co)
    - Co 0.0001 (main variable)
    - Sa 0.0001, An 0.0245 (secondary variables)

G2: AML vs. healthy control subjects

- Analysis 1- Condition- fixed, Sample Site- fixed
    - Co 0.0001 (main variable)
    - Sa 0.0001 (secondary variable)

G3: Lymphoma vs. healthy control subjects

- Analysis 1- Condition- fixed, Sample Site- fixed
    - Co 0.0001 (main variable)
    - Sa 0.0001 (secondary variable)

G4: AML vs. Lymphoma

- Analysis 1- Condition- fixed, Sample Site- fixed

- Co 0.0148 (main variable)
- Sa 0.0001 (secondary variable)

Condition= G1- Hematological cancer, healthy control subjects; G2- AML, healthy control subjects; G3- Lymphoma, healthy control subjects; G4- AML, Lymphoma patients

Diagnosis= ALL, AML, Lymphoma, MDS, Myelofibrosis and Myeloma

Antibiotic= Yes, No

Sample site= Buccal mucosa, saliva, superficial supragingival plaque, tongue

3.3 LEfSe Analysis

LEfSe analysis identified 52 probes differentiating hematological cancer patients from healthy controls. The most differential features being among those diagnosed with MDS, noting *Streptacoccaceae* (log LDA ≈ 5.3) as the most differential feature **(Figure 5)**. Additionally, there was an indication that there may be a relationship between the microbial communities of AML, ALL and MDS **(Figure 6)**. Further analysis indicated 21 features differentiating AML patients from healthy controls and 24 features differentiating Lymphoma patients from healthy controls. Bacilli (log LDA ≈ 4.8) were the leading differential feature of AML and Lymphoma patients when compared to healthy control subjects. Additionally, there were 19 features differentiating Lymphoma patients from patients who were diagnosed with AML **(Figure 5)**.

CHAPTER 4: DISCUSSION AND CONCLUSION

4.1 Discussion

This is one of the largest studies that has compared the effects of hematological cancers on oral microbiome profiles to that of healthy individuals. In agreeance with Hu Jei et al.(6) the most prevalent genera in the oral cavity of our healthy control subjects were *Prevotella*, *Neisseria*, *Streptococcus*, *Haemophilus*, and *Fusobacterium.* Additionally, we analyzed five different forms of hematological cancers independently as well as a group (hematological cancer). All analysis of taxa abundance was done using V3-V4 region 16s rRNA gene next generation sequencing in conjunction with ProbeSeq species and genera identification program.

There were significant differences in the oral microbiome profile of hematological cancer patients when compared to healthy controls. LEfSe analysis indicated there were 52 features differentiating hematological cancer patients from healthy control subjects. The most prominent of the differential features was *Streptococcaceae* belonging to MDS patient group. Additionally, analysis of cladogram structure indicated a possible relationship between the communities of AML, ALL and MDS.

Furthermore, AML and Lymphoma had a distinct oral microbiome beat-diversity when compared to healthy controls. We identified 21 differential features of AML and 24 differential features of lymphoma, the most prevalent being those among the healthy control subjects. Bacilli were indicated as the most prevalent differentiating feature (log LDA ≈ 4.8) for both AML and lymphoma patients when compared to healthy control subjects. Additionally, there was indication that there may be a relationship between the oral microbiome profile of AML patients and healthy control subjects.

Relative abundance data comparison showed a significant increase of *Rothia dentocariosa* and *Actinomyces* genus in hematological cancer patients when compared

to healthy control subjects. *Actinomyces* is an opportunistic pathogen most well-known for causing oral abscesses and on occasion, abscesses in the lungs and gastric tract. We were also able to note a significant decrease in *Haemophilus parainfluenzae* and *Streptococcus* genus. Many different species of *Streptococcus* are present in the oral cavity and as a genus account for the largest RA of healthy individuals.

4.2 Conclusion

In conclusion, we noted significant differences in oral microbiome *beta*-diversity when comparing hematological cancer patients to healthy control subjects. We were also able to note differences among hematological cancer patient's oral microbiome profiles when comparing patients diagnosed with AML to patients diagnosed with lymphoma. More research is necessary to better understand how the oral microbiome interacts with its microenvironment and the host immune system of cancer patients in order to elucidate possible mechanistic pathways.

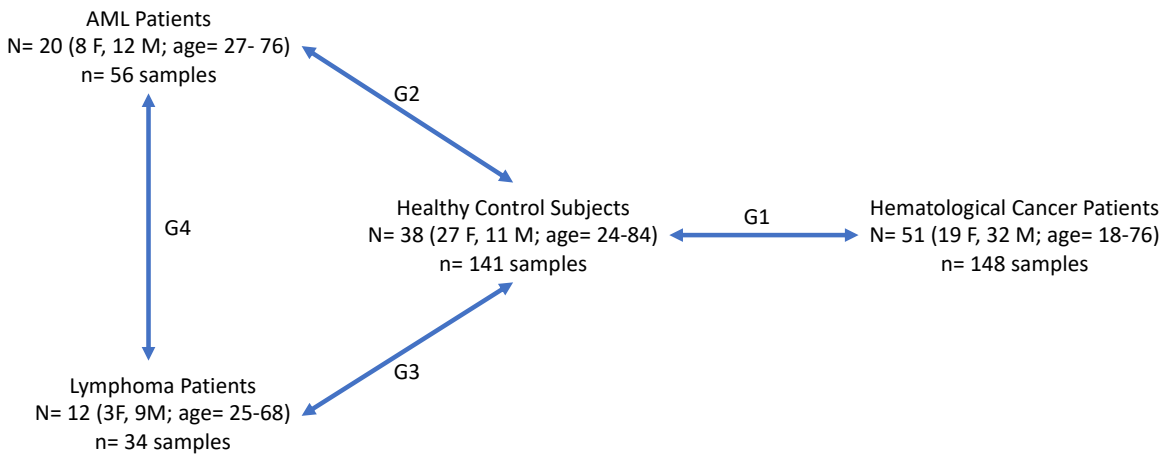

Figure 1. PERMANOVA design to assess differences in oral microbiome *beta*-diversity between groups

Oral microbiome *beta*-diversity differences were determined using four different groupings. Group 1 (G1) analysis compared hematological cancer patients to healthy control subjects. Group 2 (G2), Group 3 (G3) and Group 4 (G4) used subgroup diagnosis of hematological cancer group as follows. G2 analysis compared acute myelogenous leukemia (AML) patients to healthy control subjects. G3 analysis compared lymphoma patients to healthy control subjects. G4 analysis compared AML to lymphoma patients.

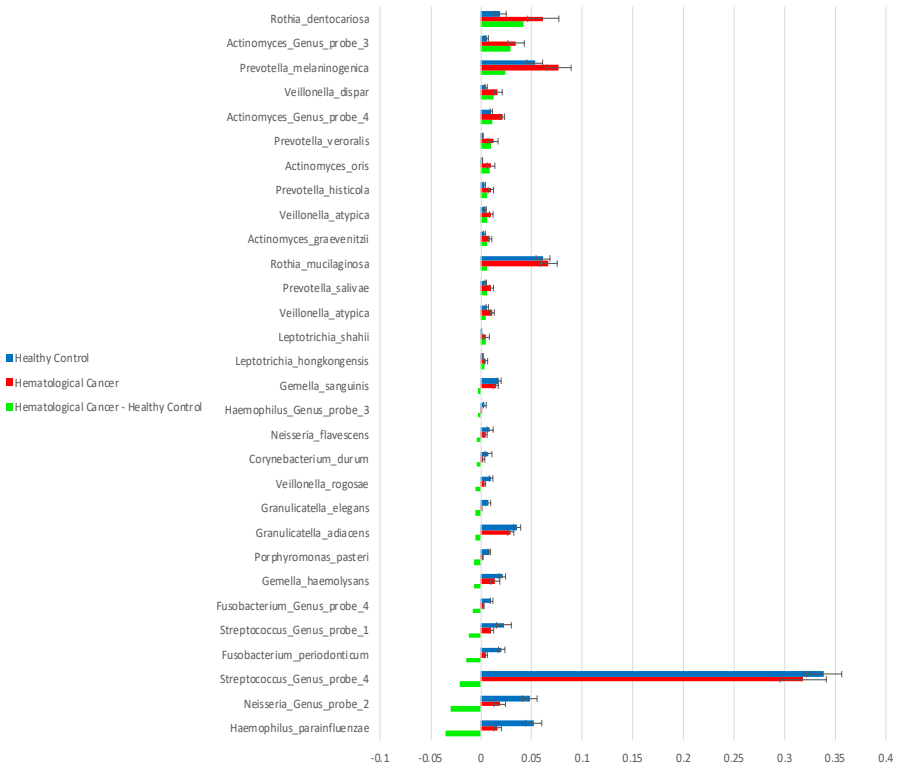

Figure 2. Clustering of cancer patients and healthy controls oral samples based on oral microbiome relative abundances

Oral bacterial genus and species relative abundance (RA) changes in oral samples (saliva sample, buccal mucosa, tongue and superficial supragingival plaque swabs) for hematological cancer patients and healthy control subjects are represented. The difference between hematological cancer patients and healthy control subjects is represented by green bars.

Standard error of the mean (Standard deviation/Square root of N) values are represented for each RA with the probe identifier shown for each genus and species.

The largest relative abundance difference when comparing hematological cancer patients to healthy control subjects is *Rothia dentocariosa*.

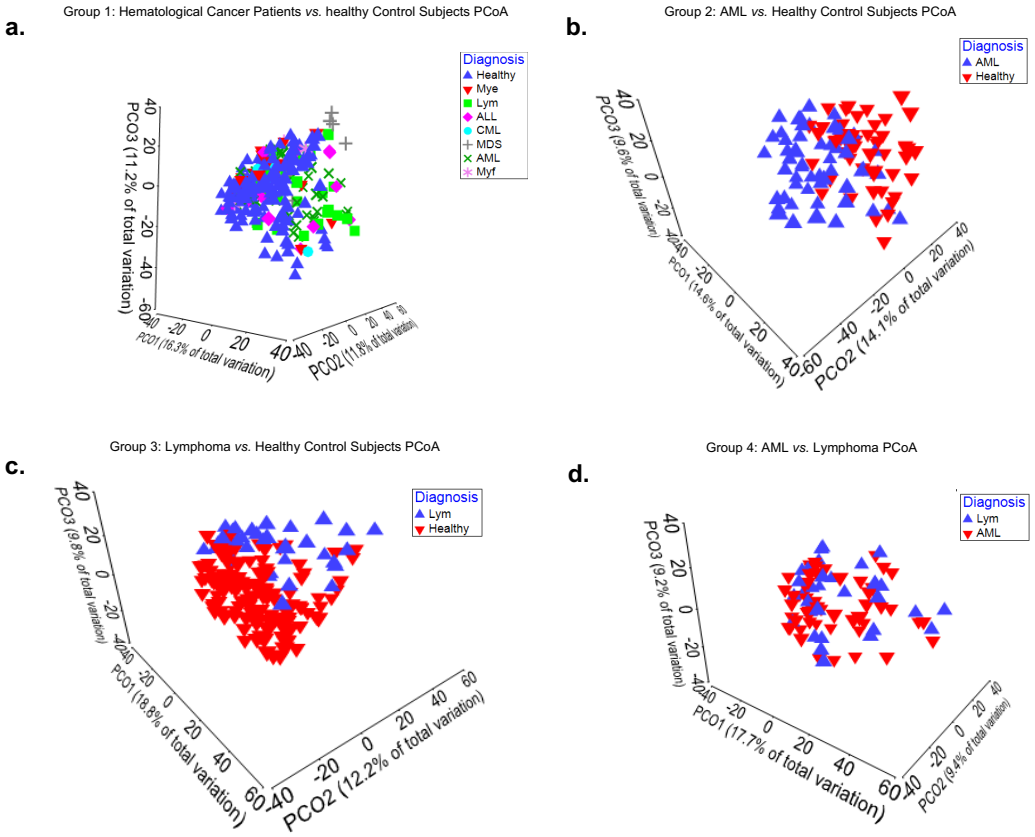

Figure 3. PCoA cross-sectional analysis of hematological cancer patients and healthy control subjects

PERMANOVA testing for the PcoA comparisons shown in figure resulted in Monte-Carlo $p<0.05$. PERMANOVA was based on all 737 probes comprised of 620 species and 117 genus probes.

Patients having next generation sequencing data for saliva sample, buccal mucosa, tongue and superficial supragingival plaque swabs. **a.** hematological cancer patients *vs.* healthy control subjects with a total of three principal coordinates of 39.3 percent variation; **b.** AML vs healthy control subjects with a total of three principal coordinates of 39.4 percent variation; **c.** Lymphoma patients *vs.* healthy control subjects with a total of all principal coordinates being 40.8 percent variation representing the largest variation; **d.** AML *vs.* lymphoma patients with a total of all principal coordinates of 36.3 percent variation representing the smallest variation.

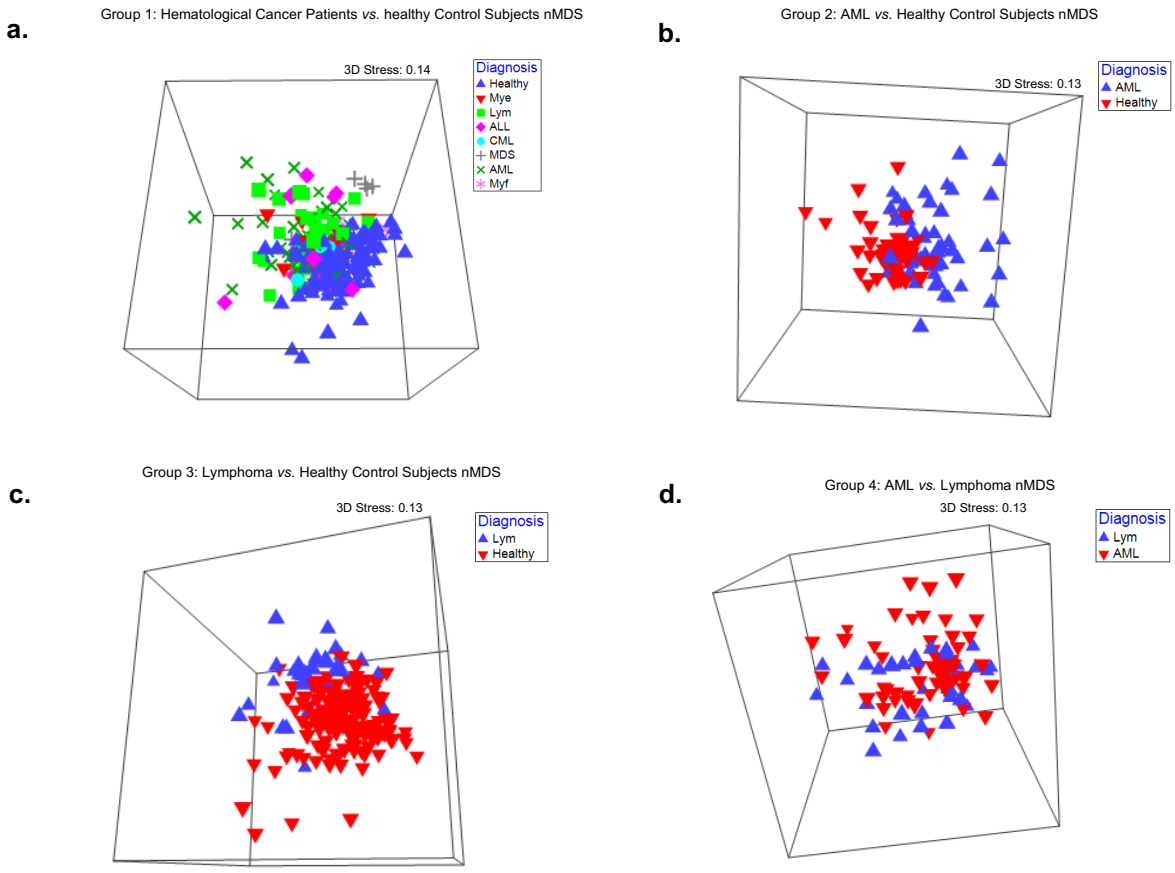

Figure 4. nMDS for hematological cancer patients and healthy control subjects

Nonmetric multidimensional (nMDS) scaling of PERMANOVA performed on Bray-Curtis dissimilarity matrices determined from square root transformed relative abundance data using PRIMER$_{v7}$ (PRIMER-E Ltd, Ivybridge, UK). PERMANOVA testing resulted in Monte-Carlo $p<0.05$ for the comparisons shown and was based on all 737 probes comprised of 620 species and 117 genus probes. All groupings displayed differences in 3D clustering of samples when comparing hematological cancer samples to healthy control subjects as well as AML to lymphoma patients.

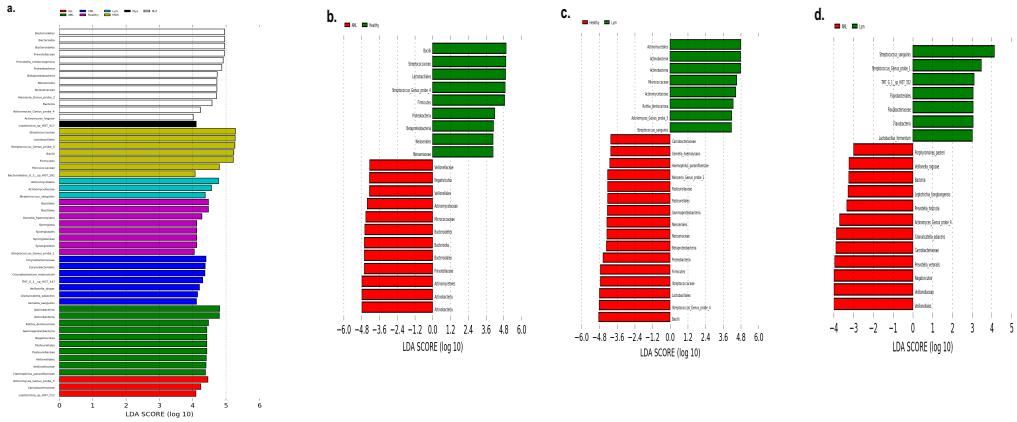

Figure 5. LEfSe results for hematological cancer patients and healthy control subjects displayed as horizontal histograms

Linear discriminant analysis Effect Size (LEfSe) was performed to determine distinct oral microbiome features in samples (saliva sample, buccal mucosa, tongue and superficial supragingival plaque swabs) from hematological cancer patients and healthy control subjects.

Data formatting and input consisted of diagnosis (hematological cancer and healthy control) as class and patient ID as subject.

a. LEfSe results showing a histogram of hematological cancer patient samples and helathy control subjects with most differential features belong to MDS. A total of 52 differential features were identified 13 Myf, 1 Mye, 7 MDS, 3 Lym, 7 CML, 10 AML, 3 ALL and 8 healthy controls.

b. Horizontal histogram comparing differential features of AML patients (red) and healthy control subjects (green) representing top 21 differential features

c. Horizontal histogram displaying top 24 differential features of lymphoma patients(green) and healthy control subjects (red).

d. Horizontal histogram displaying top 19 differential features of AML (red) and lymphoma (green) patients.

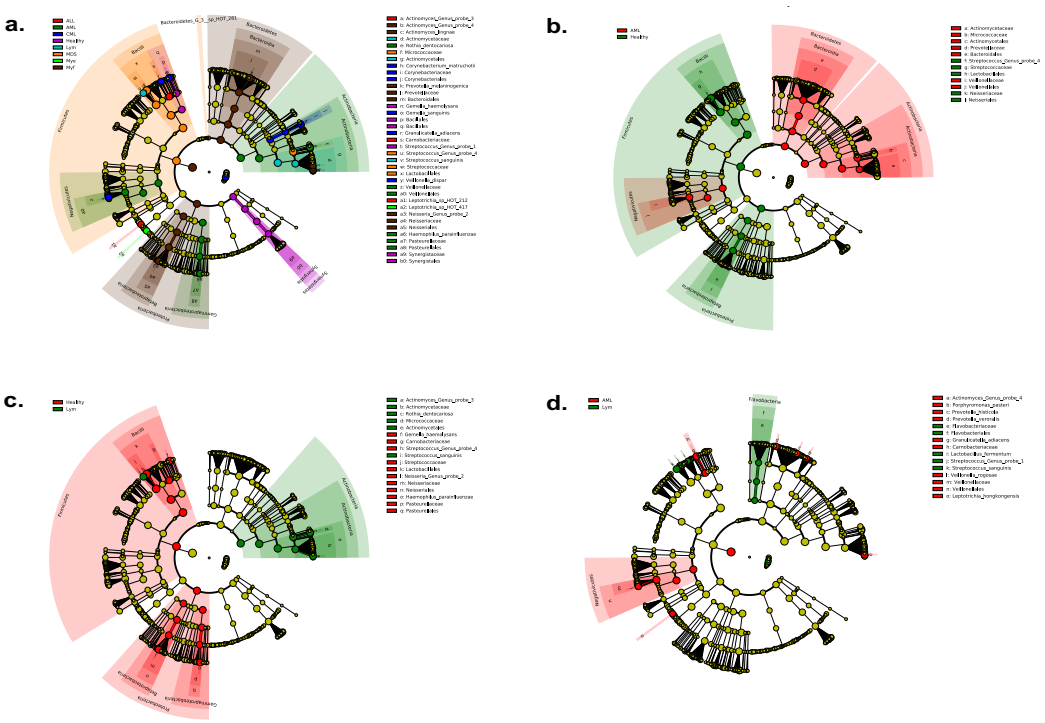

Figure 6. Cladograms representing discriminative features of hematological cancers and healthy control subjects

a. LEfSe Cladogram with 52 differential features identified by LEfSe. Largest group is the Firmicutes that show the differential features of MDS. Within that group falls Negativicutes and Bacilli making subgroups that differentiate with AML and MDS. There may be some relationship between the communities of AML, ALL, and MDS. Actinobacteria are the largest differential feature of Lym with a subgroup of differential features of CML. There may be overlap in microbial communities within these two diagnosis. Myf has differential features coinciding with Bacteroidetes and Proteobacteria. **b.** LEfSe Cladogram with 21 differential features identified by LEfSe. Largest group is the Firmicutes that show the differential features of healthy controls. Within that group fall the Negativicutes and Bacilli making subgroups that differentiate healthy controls from AML. Actiniobacteria and Bacteroidetes are the largest groups that show differential features of AML. **c.** LEfSe Cladogram with 24 differential features identified by LEfSe. Largest group is the Firmicutes that show the differential features of healthy controls. Actiniobacteria are the largest groups that show differential features of Lym. **d.** LEfSe Cladogram with 19 differential features identified by LEfSe. Largest group is the Negativicutes that show the differential features of AML. Flavobacteria is the largest group that show differential features of Lym.

Table 1. Demographics table for hematological cancer patients and healthy control subjects

Criteria	Hemotological Cancer	Healthy Control
Male	32	11
Femal	19	27
Total Patients	51	38
Ethnisity/Sex	**Hemotological Cancer**	**Healthy Control**
Male		
American Indian	0	1
Asian	0	2
Black	12	0
Caucasion	20	7
Hispanic	0	1
Other	0	0
Total Males	32	11
Female		
American Indian	0	0
Asian	0	2
Black	5	2
Caucasion	13	22
Hispanic	0	1
Other	1	0
Total Females	19	27
Total Patients	**51**	**38**

REFERENCES CITED

1. Ahn, Jiyoung et al. “Oral microbiome and oral and gastrointestinal cancer risk” *Cancer causes & control : CCC* vol. 23,3 (2012): 399-404.

2. Brennan CA, Garrett WS ”Fusobacterium nucleatum - symbiont, opportunist and oncobacterium” Nat Rev Microbiol. 2019 Mar;17(3):156-166. doi: 10.1038/s41579-018-0129-6. PMID:30546113

3. Chen, Xingdong et al. “Oral Microbiota and Risk for Esophageal Squamous Cell Carcinoma in a High-Risk Area of China” *PloS one* vol. 10,12 e0143603. 7 Dec. 2015, doi:10.1371/journal.pone.0143603 PMID: 26641451

4. Farrell, J. J., Zhang, L., Zhou, H., Chia, D., Elashoff, D., Akin, D., Paster, B. J., Joshipura, K. Wong, D. T. (2011). Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. *Gut*, *61*(4), 582-8. PMID: 21994333

5. Hiba Mohammed, Elena Maria Varoni, Andrea Cochis, Massimo Cordaro, Patrizia Gallenzi, Romeo Patini, Edoardo Staderini, Carlo Lajolo, Lia Rimondini, and Vincenzo Rocchetti, Oral Dysbiosis in Pancreatic Cancer and Liver Cirrhosis: A Review of the Literature Biomedicines. 2018 Dec; 6(4): 115. Published online 2018 Dec 11. doi: 10.3390/biomedicines6040115 PMCID: PMC6316311 PMID: 30544974

6. Hu Jei, Shuwen Han, Yan Chen, and Zhaoning Ji , “Variations of Tongue Coating Microbiota in Patients with Gastric Cancer” doi: 10.1155/2015/173729 PMCID: PMC4589578 PMID: 26457297

7. Jia, G., Zhi, A., Lai, P. F. H., Wang, G., Xia, Y., Xiong, Z., . . . Ai, L. (2018). “The oral microbiota - a mechanistic role for systemic diseases.” British Dental Journal, 224(6), 447-455. doi:http://dx.doi.org.prox.lib.ncsu.edu/10.1038/sj.bdj.2018.217 PMID: 29569607

8. Karpiński, Tomasz M. Role of Oral Microbiota in Cancer Development *Microorganisms* vol. 7,1 20. 13 Jan. 2019, doi:10.3390/microorganisms7010020 PMID: 30642137

9. Klimesova K, Jiraskova Zakostelska Z, Tlaskalova-Hogenova H. Oral Bacterial and Fungal Microbiome Impacts Colorectal Carcinogenesis. Front Microbiol. 2018 Apr 20;9:774. doi: 10.3389/fmicb.2018.00774. eCollection 2018. Review. PMID: 29731748

10. Michaud, Dominique S et al. “Plasma antibodies to oral bacteria and risk of pancreatic cancer in a large European prospective cohort study.” Gut vol. 62,12 (2012): 1764-70. doi:10.1136/gutjnl-2012-303006 PMID:22990306

11. Mitsuhashi K, Nosho K, Sukawa Y, Matsunaga Y, Ito M, Kurihara H “Association of Fusobacterium species in pancreatic cancer tissues with molecular features and prognosis” Oncotarget. 2015 Mar 30;6(9):7209-20. PMID:25797243

12. Olsen I, Yilmaz Ö “Possible role of Porphyromonas gingivalis in orodigestive cancers” J Oral Microbiol. 2019 Jan 9;11(1):1563410. doi: 10.1080/20002297.2018.1563410. eCollection 2019 PMID: 30671195

13. Peters BA, Wu J, Pei Z, Yang L, Purdue MP, Freedman ND, Jacobs EJ, Gapstur SM, Hayes RB, Ahn J “Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers” Cancer Res. 2017 Dec 1;77(23):6777-6787. doi: 10.1158/0008-5472.CAN-17-1296. PMID 29196415

14. Wang, Yan et al. “Oral microbiota distinguishes acute lymphoblastic leukemia pediatric hosts from healthy populations” PloS one vol. 9,7 e102116. 15 Jul. 2014, doi:10.1371/journal.pone.0102116 PMID: 25025462

15. “About the Human Microbiome.” *NIH Human Microbiome Project - About the Human Microbiome*, www.hmpdacc.org/overview/.

16. Zhao H, Chu M, Huang Z, Variations in oral microbiota associated with oral cancerSci Rep. 2017 Sep 18;7(1):11773. doi: 10.1038/s41598-017-11779-9. PMID:28924229

17. “Cancer Facts & Figures 2019.” *American Cancer Society*, www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019.html.

18. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS. 2011;108(Suppl 1):4516–22

19. Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. The Development and Causes of Cancer. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9963/

20. Lim, Yenkai et al. “The Performance of an Oral Microbiome Biomarker Panel in Predicting Oral Cavity and Oropharyngeal Cancers.” Frontiers in cellular and infection microbiology vol. 8 267. 3 Aug. 2018, doi:10.3389/fcimb.2018.00267 PMID: 30123780

21. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011.Metagenomic biomarker discovery and explanation. Genome Biol. 12(6):R60.

22. Jean-Luc C Mougeot, Craig B Stevens, Darla S Morton, Michael T Brennan, Farah B Mougeot, Oral Microbiome and Cancer Therapy-Induced Oral Mucositis, *JNCI Monographs*, Volume 2019, Issue 53, August 2019, lgz002, https://doi.org/10.1093/jncimonographs/lgz002