Fakultät für Medizin, Technische Universität München

Potential role of the microbiome in Barrett's Esophagus and Esophageal Adenocarcinoma

Nikole Fjoralba Radani

Vollständiger Abdruck der von der Fakultät für Medizin der Technischen Universität München zur Erlangung des akademischen Grades einer Doktorin der Medizin genehmigten Dissertation.

Vorsitz: Prof. Dr. Wolfgang Weber

Prüfer*innen Dissertation:

1. Prof. Dr. Michael Quante
2. Prof. Dr. Susanne Kossatz

Die Dissertation wurde am 21.04.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 12.10.2021 angenommen.

Technische Universität München

Klinik und Poliklinik für Innere Medizin II

Klinikum rechts der Isar

Potential role of the microbiome in Barrett's Esophagus and Esophageal Adenocarcinoma

Nikole Fjoralba Radani

Dissertation
zur Erlangung des akademischen Grades Dr. med.

2021

Abstract

Esophageal Adenocarcinoma (EAC) incidence has increased faster than any other cancer in Western nations, which is likely due to multiple extrinsic cofactors. Shifts in microbial community composition may display external influences, so that the microbiome may present a biomarker or contributing factor for malignant progression. We aimed to investigate whether different disease stages in EAC carcinogenesis are associated with a distinct local tissue, salivary or fecal microbiome, and whether smoking and reflux as the strongest known risk factors for EAC are linked to definite microbiota patterns also found in advanced stages.

We conducted a case-control study in a cohort of 48 subjects with Barrett's Esophagus (BE), dysplasia (DP), esophageal adenocarcinoma (EAC), and controls. Epidemiological and clinical-pathological data of patients were gathered via questionnaire and the clinic database. Biospecimens including local esophageal and cardia mucosa biopsies, saliva and fecal samples were collected and their microbiomes were profiled using 16S rRNA gene sequencing. The epidemiological, clinical-pathological and microbial information gained was processed, merged and analyzed via a bioinformatical pipeline.

The microbiota composition in biopsies (not in saliva or feces) was associated with diagnosis (beta diversity, p=0.001). The two most abundant phyla present in biopsies were significantly altered in EAC in comparison to BE: Firmicutes was enriched, while Actinobacteria was depleted. In the disease groups, Bradyrhizobium and Streptococcus were reduced, Fusobacterium and Ralstonia were increased, and Granulicatella was reduced in EAC in comparison to BE. In saliva, several Streptococcus and Granulicatella species were differentially abundant in the different diagnosis groups. In feces, in DP and EAC in comparison to BE, Betaproteobacteriales was enriched, Eubacterium coprostanoligenes was reduced; and Faecalibacterium was reduced in EAC in comparison to BE. The microbial profiles of BE patients were associated to smoking and reflux symptoms in all three sample types. In BE smokers, Actinobacteria was reduced in biopsies, one Streptococcus species was reduced in saliva and Faecalibacterium was increased in feces. All of these taxa were linked in the same or inverse direction to EAC.

The diagnosis and exposure to smoking and reflux is associated with alterations of microbiome in all three sample types. Major shifts were identified in local biopsies and in subjects with a history of smoking. In conclusion, local biopsy-associated microbiome presents a promising field for further research.

Zusammenfassung

Die Inzidenz des Adenokarzinoms des Ösophagus (EAC) in den westlichen Ländern hat schneller zugenommen als jede andere Krebsart, was wahrscheinlich auf multiple extrinsische Cofaktoren zurückzuführen ist. Verschiebungen in der Zusammensetzung der mikrobiellen Gemeinschaft können äußere Einflüsse anzeigen, sodass das Mikrobiom einen Biomarker oder einen beitragenden Faktor für die maligne Progression darstellen kann. Unser Ziel war es zu untersuchen, ob verschiedene Krankheitsstadien in der EAC-Karzinogenese mit einem bestimmten lokalen Gewebe-, Speichel- oder Stuhlmikrobiom assoziiert sind, und ob Rauchen und Reflux als die stärksten bekannten Risikofaktoren für EAC mit bestimmten Mikrobiota-Mustern verbunden sind, die auch in fortgeschrittenen Stadien zu finden sind.

Wir führten eine Fall-Kontroll-Studie in einer Kohorte von 48 Probanden mit Barrett-Ösophagus (BE), Dysplasie (DP), Adenokarzinom des Ösophagus (EAC) und Kontrollen durch. Epidemiologische und klinisch-pathologische Daten der Patienten wurden mittels Fragebogen und der Klinikdatenbank erhoben. Bioproben einschließlich lokaler Biopsien der Speiseröhren- und Kardia-Schleimhaut, Speichel- und Stuhlproben wurden gesammelt, und ihre Mikrobiome wurden mittels 16S rRNA-Gensequenzierung profiliert. Die gewonnenen epidemiologischen, klinisch-pathologischen und Mikrobiom Daten wurden über eine bioinformatische Pipeline verarbeitet, zusammengeführt und analysiert.

Die Zusammensetzung des Mikrobioms in den Biopsien (nicht in Speichel oder Stuhl) war mit der Diagnose assoziiert (beta-Diversität, p=0.001). Die am häufigsten vorkommende Phyla in Biopsien waren beim EAC in Vergleich zu BE signifikant verändert: Firmicutes war erhöht, während Actinobacteria erniedrigt war. In den Krankheitsgruppen waren Bradyrhizobium und Streptococcus reduziert; Fusobacterium und Ralstonia erhöht; in EAC Granulicatella im Vergleich zu BE reduziert. In Speichel waren einzelne Streptococcus und Granulicatella Arten in den verschiedenen Diagnosegruppen unterschiedlich häufig vorhanden. In Stuhl war in DP und EAC im Vergleich zu BE, Betaproteobacteriales angereichert, Eubacterium coprostanoligenes reduziert; und in EAC im Vergleich zu BE Faecalibacterium reduziert. Die mikrobiellen Profile von BE Patienten waren mit Rauchen und Reflux Symptomen in allen drei Probentypen assoziiert. Bei BE-Rauchern war Actinobacteria in Biopsien reduziert, Streptococcus im Speichel reduziert und Faecalibacterium im Stuhl angereichert. All diese Taxa waren ebenfalls mit EAC assoziiert.

Die Diagnose und die Exposition gegenüber Rauchen und Reflux ist mit Veränderungen des Mikrobioms in allen drei Probentypen verbunden. Größere Veränderungen wurden in lokalen Biopsien und bei Personen mit einer Rauchergeschichte festgestellt. Zusammenfassend lässt sich sagen, dass das lokale Biopsie-assoziierte Mikrobiom ein vielversprechendes Feld für die weitere Forschung darstellt.

Table of Contents

List of Abbreviations

16S rRNA	16S ribosomal RNA
AEG	Adenocarcinomas of the esophagogastric junction
AJCC	American Joint Committee on Cancer
ANOVA	Analysis of variance
BE	Barrett's Esophagus
c	Class
CLE	Columnar-lined esophagus
CRC	Colorectal cancer
CT	Computer tomography
Ctrl.	Control
DCA	Deoxycholic acid
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
DP	Dysplasia
EAC	esophageal adenocarcinoma
e.g.	for example
EGD	Esophagogastroduodenoscopy
EMR	Endoscopic Mucosa Resection
ESD	Submucosal Dissection
EUS	Endoscopic ultrasound
f	Family
FDG-PET scan	Fluorodeoxyglucose positron emission tomography
FFPE	Formalin fixed paraffin-embedded
FXR	Farnesoid-X-Rezeptor
G	Grade
g	Genus
GC	Goblet cell
GERD	Gastroesophageal reflux disease
H&E	Hematoxylin and Eosin
H. p.	Helicobacter pylori
H. pylori	Helicobacter pylori
IMNGS	Integrated microbial NGS platform
IMSE	Institute of Medical Informatics, Statistics and Epidemiology
IWGCO	International Working Group for the Classification of Esophagitis

KORA	Kooperative Gesundheitsforschung in der Region Augsburg
LPS	Lipopolysaccharides
M	Metastasis
MDS	Multidimensional Scaling plot
N	Node
NF-κB	nuclear factor 'kappa-light-chain-enhancer'
NMDS	Non-metric Multidimensional Scaling plot
NR	No reflux symptoms
NS	Never-smoker
NSAIDS	Non-steroidal anti-inflammatory drug
o	Order
OTU	Operational taxonomic unit
p	Phylum
PAS	Periodic Acid-Schiff
PFPE	PAXgene fixed and paraffin-embedded
R	Reflux symptoms
RDP	Ribosomal Database Project
RFA	Radiofrequency Ablation
RNA	Ribonucleic acid
S	Smoker
SCC	Squamous cell carcinoma
SCJ	Squamocolumnar junction
T	Tumor
T-BMCA	Tauro-β-muricholic acid
WGS	Whole genome sequencing

List of Figures

List of Tables

1 Introduction

1.1 Barrett's Esophagus

1.1.1 Definition, epidemiology and pathogenesis

Barrett's Esophagus (BE) is a facultative premalignant condition of the distal esophagus, defined by the replacement/transformation of squamous epithelium with/into columnar epithelium with similarity to intestinal mucosa, therefore it is also called intestinal metaplasia of the esophagus [101]. Metaplasia is a term used in the field of pathology, denoting the reversible replacement of one specialized cell type by another specialized cell type, usually occurring as a response to chronic inflammation of tissue [92]. Gastroesophageal reflux disease (GERD) is the most well-known risk factor for developing BE, as it causes chronic inflammation at the esophagogastric junction. The development of BE is only detected in approximately 5–8% of patients suffering from GERD [147]. A metaplastic area may progress into a dysplastic area, which is classified as low-grade (LGD) or high-grade Dysplasia (HGD) depending on the dimension of cytological and architectural changes [88]. Dysplasia, in contrast to metaplasia, is defined by the proliferation of atypical cells that have not yet fulfilled the criteria for cancer but are likely to become malignant cells. The last stage in the BE progression cascade is esophageal adenocarcinoma. In a study undertaken in 2000, which analyzed estimated risk for BE progression based on several studies performed between 1966 and 1998, the progression rate of BE to esophageal adenocarcinoma (EAC) was stated to be 0.5% per patient-year [119]. Nevertheless, recent studies report lower progression rates of BE to cancer (0.10 and 0.13% per year) [101]. The data situation referring to BE progression to EAC remains ambiguous, however, as it depends on several factors such as BE definition and diagnosis, size of the analyzed cohort, and possible publication bias [119]. To identify progression of BE in the clinical routine, patients attend a surveillance program involving repetitive control endoscopies at fixed time intervals. Given the low progression rates of Barrett's Esophagus, the invasiveness of endoscopy, and the subsequent high costs for the health system, it is essential to identify biomarkers for risk stratification among BE patients.

The incidence of esophageal adenocarcinoma in the United States has risen by approximately 460% since 1970, for reasons that remain unclear [13]. The timing of this incidence increase is some years after the introduction and clinical use of the first antibiotics

(the first clinical use of Penicillin being in 1941), which can be seen as an indication for a possible role played by antibacterial therapy and thus of the microbiome in pathogenesis [75].

There are some well investigated epidemiologic factors associated with BE development and its progression to esophageal adenocarcinoma, including age, sex, ethnic background, obesity (Waste-to-Hip-Ratio), smoking, medication, Helicobacter pylori, and reflux symptoms [105]. Large metanalyses confirm that men are affected by BE twice as much as women, and that the highest disease prevalence is found in Caucasians [24]. In recent ten years, more focus has been given to obesity as a risk factor for BE [65]. A study involving 1,102 BE cases and 1,400 controls revealed a significant association between abdominal obesity and BE in both men and women [65]. An association with patients' BMI was not found [65]. Abdominal obesity is more prevalent among men than women [35], which correlates with the higher prevalence of BE among men. Another explanation for the higher prevalence of BE among men could be that fertile women might benefit from the anti-inflammatory properties of estrogen [4], decreasing the risk of developing BE. Regarding obesity as a risk factor for BE, experts assert that this association is provided through a simple mechanical mechanism by which a high abdominal pressure induces the relaxation of the lower esophagus sphincter and consequently causes GERD and BE [25, 27, 44]. However, studies have found a slight association between GERD and abdominal obesity; therefore, the mechanism by which obesity influences BE development remains unclear [65] and requires further explanation.

A possible explanation for this association could be the link between abdominal obesity and a high-fat diet, incidence of which has increased in recent decades in western countries. Studies using IL1β-mouse, a mouse model that undergoes stages of esophagitis, metaplasia, dysplasia, and esophageal adenocarcinoma due to the overexpression of the proinflammatory cytokine IL-1β in the esophagus [101], demonstrated the presence of an alternated microbiome in mice fed with high-fat chow compared to mice fed with conventional chow [80]. In addition, mice fed with chow containing a higher percentage of fat showed an advanced phenotype [80]. Accordingly, a high-fat diet, and thus microbiome alternation, can be seen as influencing factors for disease progression.

Although much is known about epidemiological factors linked to BE and EAC, their pathogenesis remains unclear and controversial, since there is divided opinion on this. Two particular theories of the origin of BE have been researched most [103]. The first involves the tissue wounding model, which proposes that microinjuries to the mucosal epithelium

occur in squamocolumnar junction (SCJ) due to chronic exposure to a severe environment. The injured epithelium is replaced by a reparative epithelium consisting of adjacent squamous and columnar cell progenitors. In the course of the disease, columnar cells dominate due to their selection advantage. The columnar cells of the reparative epithelium most likely originate from cardia stem cells, which expand driven by luminal factors such as acid-biliary reflux, microbiome, and the inflammatory tissue microenvironment to the esophagogastric junction; subsequently, the columnar progenitor cells evolve into glands and differentiate into Barrett's glands in order to adapt to the local conditions (Figure 1) [103]. This hypothesis is also supported using lineage tracing methods in BE mouse models [94, 102, 103]. Gastric cardia stem cells (labeled by LGR5) expanded proximally to the esophagus driven by inflammation and application of bile acids to drinking water in the IL-1β mice [102].

The second model explains BE's origin through a transformation of the differentiation lineage of squamous cells of the esophagus to a columnar cell lineage. It was shown, that after daily exposure of an in vitro cell line model to acid and bile, transcription factors that are lineage-determining for columnar epithelium like SOX9 and CDX2 were upregulated and TAp63, determiner for squamous epithelium, was downregulated at the mRNA or/and protein level [77].

Finally, in a very small percentage of patients, BE progresses to esophageal adenocarcinoma. It is likely that the progression is also driven by the inflammation inducing a number of genetic mutations in the Barrett's niche. These mutations are not yet fully studied, but the most frequent mutations in EAC tissue seem to be chromosomal instability and DNA hypermethylation [83].

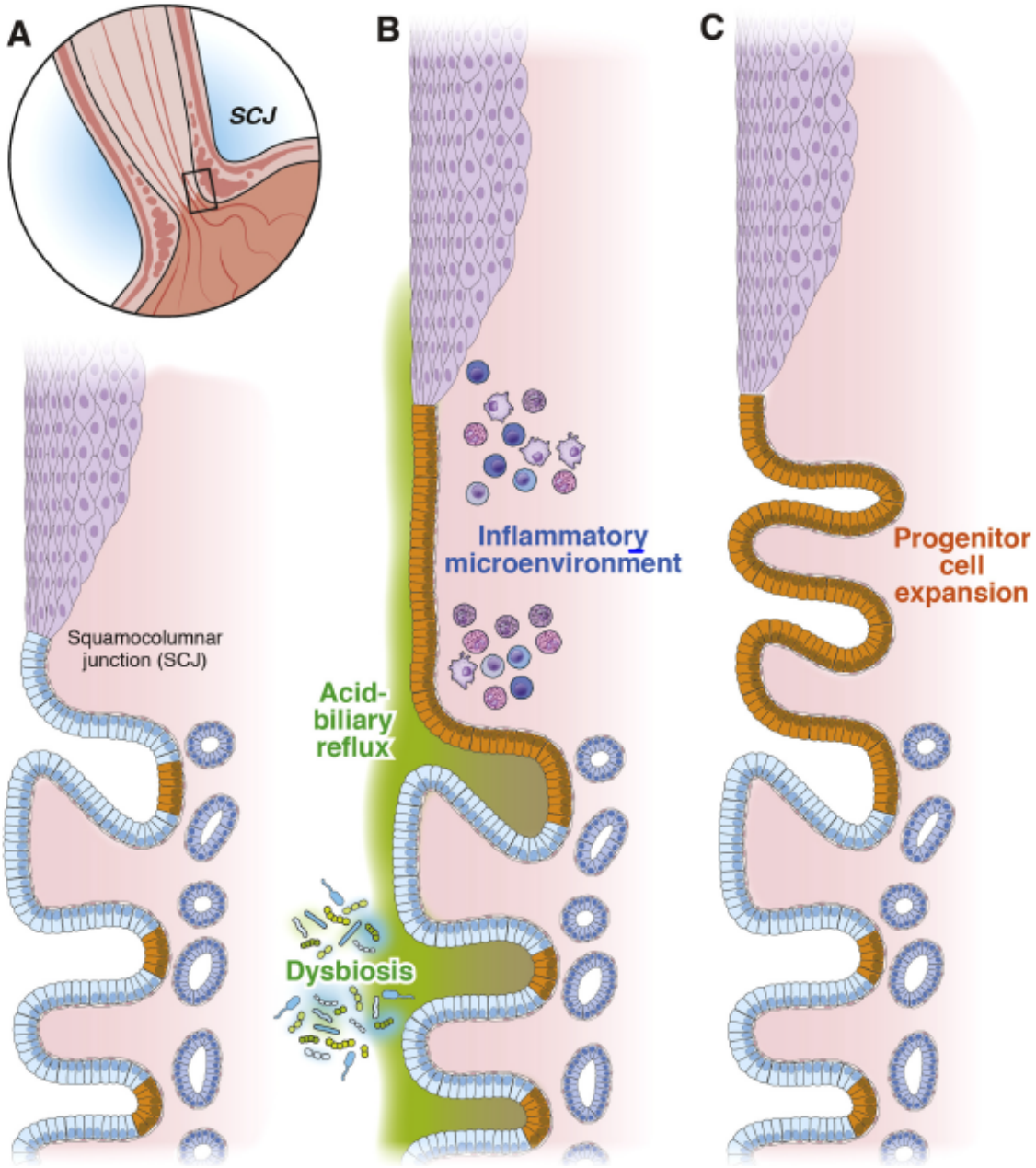


Figure 1. Luminal factors and inflammatory immune response mediate stem cell expansion from gastric cardia [103].
(A) The squamocolumnar junction (SCJ) is the boundary between the squamous epithelium of the esophagus and the columnar epithelium of the stomach and is consistent with the anatomical boundary between the esophagus and the stomach. (B) Stem cells in cardia glands (orange) expand toward the SCJ, influenced by acid-biliary reflux and dysbiosis from the luminal side and present the origin of the columnar epithelium in the junction. (C) The next stage is the evolution of glands from these progenitor cells, which are more resistant to the severe conditions at the SCJ than the squamous epithelium.

1.1.2 Diagnostics and surveillance

An Esophagogastroduodenoscopy (EGD) remains the standard procedure for diagnosing BE. The normal esophageal tissue appears endoscopically pale due to its squamous epithelium, while BE areas appear salmon pink and velvety [8]. To confirm the diagnosis histologically, biopsies from suspicious areas are taken during endoscopy.

Previous diagnostic classification classified BE into long (>3 cm) or short (<3 cm) segment Barrett based on the 3-cm length threshold [21]. This simple classification of various endoscopic findings was assessed in the clinical practice as too crude to describe the endoscopic appearance of BE [21].

The Prague classification system was created by the International Working Group for the Classification of Esophagitis (IWGCO) in 2004 to standardize the length of BE in endoscopy; it is based on measures of the circumferential (C) and maximal (M) extents of BE [120]. This classification is currently the most common classification system used in the clinical routine to describe the extent of BE.

In the current German S2k guidelines for gastroesophageal reflux disease, regular endoscopic controls are recommended at defined intervals due to the chance of early detection of possible progression to dysplasia or cancer. According to this surveillance program, high-resolution video endoscopy controls, possibly with chromoendoscopy (with acetic acid 1.5-3% for EAC), are recommended at predetermined intervals. Biopsies from suspicious areas and four quadrant biopsies every 1–2 cm in columnar-lined epithelium are taken each time. The frequency of control EGD is dependent on the presence of dysplasia and on its grade. Conforming to German surveillance guidelines, in case of BE diagnosis in the index examination, a control upper endoscopy is recommended after one year. Upon the first control assessment, if non-dysplastic BE is still present, control EGDs are required every 3–4 years. If low-grade dysplasia is diagnosed, control EGDs are recommended every 6 months within the first year, and if low-grade dysplasia has not progressed in the first year, annual control EGDs are recommended. In the event that the BE region is endoscopically conspicuous, a radiofrequency ablation (RFA) or an endoscopic mucosa resection (EMR) can be conducted. Once HGD is diagnosed, a similar therapeutic approach to that used for early carcinomas is recommended because of the risk of synchronous adenocarcinoma or the high risk of rapid progression [43].

1.2 Esophageal Adenocarcinoma

1.2.1 Definition and epidemiology

Adenocarcinoma of the esophagus represents the second most prevalent esophageal cancer worldwide after squamous cell carcinoma (SCC) and occurs almost 75% of the time in the distal esophagus. Studies in the last two decades have shown that esophageal adenocarcinoma (EAC) has overtaken SCC in prevalence in industrialized nations [117, 130]. Compared to other tumors, esophageal carcinoma is a tumor with aggressive features, representing the eighth most prevalent cancer and the sixth most frequent cause of cancer-associated deaths in the world [48]. Importantly, in western countries EAC is one of the tumor identities with the fastest-growing incidence rates, and it has shown a 500% increase since 1970 [26].

Adenocarcinomas of the esophagogastric junction (AEG), defined as tumors whose center is localized within 5 cm proximal or distal of anatomical cardia, were first described by Siewert and Hölscher as particular tumor identities [121]. The 7th edition of the TNM classification categorized adenocarcinomas of the esophagogastric junction as esophageal carcinomas if the tumor center is localized 5 cm proximal or distal of the esophagogastric junction and the tumor infiltrates the esophagus simultaneously [106]. Tumors infiltrating the first 5 cm of the stomach without infiltration of the esophagus, or those with a greater extension than 5 cm from the esophagogastric junction, are ranked as gastric cancers [106].

1.2.2 Diagnostics and staging

Unfortunately, symptoms of EAC are non-specific and appear mostly for the first time in advanced tumor stages. If it occurs after the age of 40, dysphagia is a warning symptom of esophageal cancer [47]. In such cases, an upper control endoscopy is recommended to exclude the existence of tumors, because at this age carcinoma represents the most frequent cause of esophageal stenosis [47]. The other non-specific symptoms are pain and the sensation of pressure retrosternal and in the spine, reflux, unintended weight loss, anorexia, gastrointestinal bleeding, anemia, and pale skin.

The primary diagnostic tool for identification of EAC is often an esophagogastroduodenoscopy, by which most carcinomas are found incidentally [76]. To confirm the diagnosis histologically, more than eight biopsies from tumor suspicious areas should be

performed [45], as the probability of detecting carcinoma from biopsies increases with the number of biopsies performed [32]. Endoscopic mucosa resection offers an alternative tool to carcinoma detection [45].

A pretherapeutic accurate staging is important for prognosis and decision-making for an individual therapeutic strategy, as a clear correlation has been shown between TNM stage and prognosis [46]. The clinical stage can be determined based on the TNM classification system developed by the American Joint Committee on Cancer (AJCC) (Table 1) [82].

Stage	T	N	M	Therapeutic options
0	Tis	N0	M0	Local ablative therapy
I	T1	N0	M0	Surgery
IIA	T2	N0	M0	Surgery
	T3	N0	M0	
IIB	T1	N1	M0	Neoadjuvant therapy with or without surgery
	T2	N1	M0	
III	T3	N1	M0	Neoadjuvant therapy with or without surgery
	T4	Any N	M0	
IVA	Any T	Any N	M1a	Chemotherapy or radiation therapy with or without surgery
IVB	Any T	Any N	M1b	Palliative treatment

Table 1. Tumor stage and therapeutic options of esophageal adenocarcinoma [82].

Table 2 presents the TNM system, which is used to describe esophageal tumor expansion. T describes the depth of tumor invasion across the organ wall layers, N the affection of lymph nodes, and M the existence of metastasis. For the assessment of tumor stage, different imaging options are available, including endoscopic ultrasound (EUS), computed tomography (CT) and 18F-fluorodeoxyglucose positron emission tomography (FDG-PET scan) [62]. EUS represents an imaging modality to evaluate the tumor infil-

tration within the organ layers and lymph nodes. Several studies have revealed the superiority of EUS in comparison to CT in terms of assessment of primary tumor extension within the organ wall layers, nearby invasion (T stage), and lymph node metastasis (N) [127]. In addition, EUS was evaluated as being the most accurate modality for the assessment of tumor stages I to III, as an accurate differentiation between T1, T2, and T3 infiltration is not possible using CT [109]. Nevertheless, a computed tomography is generally performed in order to identify local dissemination to adjacent structures [82] and distant metastasis. Lymphatic metastasis is observed in early tumor stages, while hematogenic metastasis in states of advanced disease. 20–30% of patients have distant metastasis at the time of primary diagnosis [106]. The preferred metastasis localizations include liver, lung and bone [62].

TNM	Description
Tis	Carcinoma in situ
T1	Tumors invade lamina propria or submucosa
T2	Tumors invade muscularis propria
T3	Tumors invade adventitia
T4	Tumors invade adjacent structures
N0	No regional lymph node metastases
N1	Metastases of 1-2 lymph nodes
N2	Metastasis of 3-6 lymph nodes
N3	Metastasis in 7 or more lymph nodes
M0	No distant metastasis
M1	No distant metastasis

Table 2. Clinical TNM classification system of esophageal adenocarcinoma [110].

Another accepted classification of EAC separates adenocarcinoma of the esophagogastric junction (AEG) into type I, II, and III based on the anatomic position of the tumor center: type I AEG are carcinomas of the distal esophagus infiltrating the esophagogastric junction from above; type II carcinomas present the so-called true cardia carcinomas; type III represent subcardial gastric carcinomas infiltrating the esophagogastric junction and esophagus from below [122]. This classification is given in order to choose the right surgical approach.

However, there is disagreement regarding the surgical approach [78]. In the diagnostics of carcinomas of the esophagogastric junction, the exact statement of tumor and adjacent structures via imaging processes is important due to the classification into three different types [78]. The endoscopy, representing one of the imaging diagnostic tools of AEG carcinomas, should be performed anterograde and retrograde in order to assess rf the largest tumor mass localization [78]. However, it is difficult to distinguish these three types of AEG in the endoscopy; the final allocation can occur intraoperative and through evaluation of the resected parts [78]. Additionally, the lymph node metastasis localization distinguishes between these three types: while in AEG I metastasis are frequently found in superior mediastinum with infiltration of tracheal bifurcation and above, in the case of AEG II and III metastasis occurs in the inferior mediastinum and parts of the coeliac trunk, which shows a strong similarity between types II and III [78].

1.2.3 Therapy

Esophageal adenocarcinoma represents a challenging tumor identity, not only because of difficulties in early diagnosis, but also due to complex therapeutic decisions. Although different treatment modalities are available—including endoscopic, surgical, chemotherapeutic, and radiation-based modalities, mostly combined in a multimodality treatment strategy—the prognosis remains poor. The decision for the appropriate individual therapy is usually taken by an interdisciplinary team (tumor board), who consider several factors such as tumor stage and patient's health condition.

Endoscopic therapy

Endoscopic therapy represents an approach with curative intent, and it is used as therapy for early carcinomas (T1 carcinomas). Studies revealed that endoscopic therapy provides a cancer-free survival rate with lower morbidity similar to that of surgical therapy [98]. The diagnosis frequency of an early esophageal carcinoma is relatively low, given with one in eight [64]. In the endoscopic approach, resection techniques such as EMR and submucosal dissection (ESD) can be distinguished from ablative techniques such as radiofrequency ablation (RFA) [5].

There are different indications for the use of each technique. EMR has an advantage in providing large tissue specimens, which can be used for histological diagnostic and for a better staging [79]. EMR is usually used for nodular BE, BE with high grade dysplasia, and T1a lesions (mucosal lesions) [5]. ESD can be used for the same indications as EMR, but is used more commonly for large areas of dysplasia or T1b lesions (submucosal lesions) [5]. Ablative techniques such as RFA have superficial ablation properties, and thus a limited use for treatment of early carcinomas as monotherapy [84].

Surgical therapy

For esophageal carcinomas, there are diverse surgical approaches, the most frequently performed ones being the transthoracic and the transhiatal esophagectomy [81]. The objectives of the surgical approach are the R0 resection with the complete excision of tumorous tissue and resection of adjacent lymph nodes. The surgical approach for type I AEG consists of a transthoracic en bloc esophagectomies including two-field lymphadenectomy (posterior mediastinum, upper abdomen), as better long-term survival rates were associated with this approach compared to the transhiatal approach [90]. Nevertheless, in terms of the oncologic outcome, for other esophageal cancer localizations with the exception of AEG I, no significant differences were found between the transthoracic and transhiatal approaches [90].

Referring to 2018 German guidelines for esophageal carcinoma, the extent of the lymphonodectomy depends on the location of the primary tumor; whereby three fields are distinguished; the 2-field lymphadenectomy represents the standard approach [68]. The extent of lymph node dissection for esophageal adenocarcinoma is still matter of debate [81]. The preoperative evidence of distant metastasis presents a contraindication for a surgical approach; if limited metastasis is identified as intraoperative, a resection of metastasis and primary tumor can be considered [68]. The local and distant recurrence rate is unfortunately high. In the case of recurrence, a multimodal therapeutic approach and a reevaluation of chemotherapy and radiation is necessary.

Radio-/Chemotherapy

In terms of a multimodal therapeutic approach, chemotherapy and radiation can be combined adjuvant or neoadjuvant with surgical therapy [82]. Multiple studies have suggested that a surgical therapeutic approach, in conjunction with neoadjuvant chemotherapy, provides a better treatment of EAC than surgery alone [1]. Referring to German S3 guidelines for esophageal cancer published in 2015, in the event of tumor respectability in cT2 stage, preoperative chemotherapy should be conducted, and in cT3 or respectable cT4 stage preoperative chemotherapy or radio-chemotherapy is recommended. The recommended chemotherapeutic substance combinations for the adequate chemotherapy from this guideline are (i) 5-Fluorouracil (5-FU)/ Cisplatin; (ii) Carboplatin/Paclitaxel; and (iii) FOLFOX including Oxaliplatin, Folic acid and 5-FU.

Considering EAC's rapidly increasing incidence rate, its detection in advanced stages, and the limited therapeutic success and poor prognosis, there is a pressing clinical need to design dependable risk-prediction models and to reveal causal factors of disease progression.

1.3 BarrettNET study

The BarrettNET study, which has previously been described in great detail [141], is a multi-center prospective cohort study in Germany, that aims among others to estimate the risk of malignant progression in BE patients and identify potential biomarkers of this malignant progression. To achieve this objective, BE patients, who visited the clinic for follow-up endoscopies, were included in this study.

The BarettNET study consists of collecting patient-related information (through a questionnaire) and various biosamples, including blood samples, tissue biopsies, saliva and

fecal samples. Per patient 2-6 tissue biopsies were taken: 1-4x from BE/dysplasia/EAC-region, 1x from the inconspicuous esophageal squamous mucosa (proximal of the lesion), 1x from inconspicuous cardia mucosa (distal of the lesion). The collection of saliva and fecal samples from patients of the BarrettNET study was initiated with the present microbiome project. The frequency of sample collection follows the individual patient surveillance plan as recommended by German guidelines, rather than a fixed scheme predetermined by the study [141]. After approval of the study by the local ethics committee (#5428/12), participants are consecutively asked to participate [141]. The recruitment of patients takes place at the Department of Gastroenterology, Interdisciplinary Endoscopy, Klinik und Poliklinik für Innere Medizin II, Klinikum rechts der Isar, Munich, Germany and 20 external study centers in Germany [141]. Participants are informed by a trained physician of the study purpose, the sampling, the risks and benefits of participation, their rights, and the content and extent of the epidemiological questionnaire. All participants grant written, informed consent. Those eligible to participate are patients aged 18-80 with an endoscopic or histologic confirmed BE (minimum extension of C0M1 according to the *Prague C&M criteria guidelines*), LGD, HGD, or EAC prior to inclusion, who are participating in a surveillance program for BE and capable of giving consent [141]. Exclusion criteria are contraindication for specimen sampling (thrombocytes <50000 Tc/µl, Quick <60 percent, pTT >50 seconds), any other tumor disease, and general health [141]. With the implementation of fecal and saliva sample collection, further exclusion criteria are considered, such as antibiotic use in the last six weeks, chronic inflammatory bowel disease, diarrhea, vegetarian or vegan eating habits, and diseases of the liver and gallbladder. A diagnosis of HGD/EAC or withdrawal of informed consent are the endpoints.

Around 560 patients are currently taking part of BarrettNET [141]. The follow up period is defined as more than 10 years or, in case of progression, until EAC is diagnosed [141].

1.4 Microbiota and esophageal adenocarcinoma

The abrupt increase in esophageal adenocarcinoma incidence rates in western countries suggests that, apart from host factors, environmental factors may influence pathogenesis [30, 143]. Lifestyle factors such as smoking and alcohol consumption, as well as high-fat diet, increased hygiene, and antibiotic use might possibly cause—directly or indirectly, through acid-biliary reflux induction—changes in the esophageal microbiome and generate an inflammatory microenvironment [103]. Microbiota composition and diversity may probably embody of a range of external influences on the esophagogastric junction.

Cigarette smoking has been repeatedly associated with changes in microbiota. Laryngeal tissue biopsies of smokers reveal a lower microbial diversity and alterations in the abundance of several bacteria, such as a greater abundance of Streptococcus, compared to those of non-smokers [57]. Smoking-related microbiota perturbations were also found in oral wash and salivary samples of large patient cohorts [9, 142].

It is well known that microbes inhabiting our bodies play a causal role in many human diseases. Considering the localization of 99% of the human microbiome within the gastrointestinal tract [118] and the several functions of microbial communities in the host organism, it is plausible to assume a role for microbiota in pathogenesis of digestive tract diseases. In recent decades, many studies in this field have revealed links between microbiota, inflammatory bowel disease, and cancer [36, 118]. Bacteria are able to exert local effects through various mechanisms, inter alia by toxin release, genotoxic effects, and local inflammation, as well as systemic effects through metabolizing ability and systemic inflammation [3].

From 1994 onwards, H. pylori has been considered by the International Agency for Research on Cancer as a class I carcinogen based on epidemiology, history, animal models, and human studies [53]. Compared to gastric cancer, it seems that in CRC dysbiosis plays a potential role in cancer initiation, since germ-free mice develop a lower number of tumors in their colon [108]. Nevertheless, in recent years, single bacteria with the ability to cause damage at DNA-level were also identified in CRC. In this context, there is a lack of clarity regarding the question of whether a singular bacterial strain, the dysbiosis, or both at once promote carcinogenesis [104].

Even though the role of the microbiome in the development of gastric cancer and CRC has been widely investigated, its role regarding BE and EAC remains poorly explored. Studies on IL1β mice have shown some indications of microbiota being involved in disease pathogenesis. In IL1β mice, BE development was associated with Infiltration of CD11b+Gr1+ myeloid cells [53]. These cells are not only known as suppressors of T cells in cancer tissue (myeloid-derived suppressor cells), but also as early immune system responders to bacteria [15, 132]. High-fat diets enhanced IL1β mice phenotypes and altered their intestinal microbiome [80]. Furthermore, germ-free IL-1β-mice showed lower histopathological scoring compared to conventional mice. Other animal experiments have revealed that lipopolysaccharides (LPS), which are an integral part of the cell membrane of gram-negative germs, can cause a relaxation of the lower esophagus sphincter, with reflux as a consequence [17, 34]. LPS binding also seems to increase NF-kβ expression levels, which correlate with tissue inflammation and progression to esophageal

cancer [86]. In summary, based on these data, it could be assumed that in the case of EAC the effects of bacteria consist of modulation of inflammation.

These findings derived from animal models also needed to be shown in human cohorts, as animal-based microbiome experimental approaches do not reflect the complex composition of the human microbiome [118]. Several studies have demonstrated that sampling techniques and microbiome identification methods have a large impact on the range of species identified. In order to begin the present microbiome study on humans, it was essential to be confidential with the sampling and microbiome profiling. The first studies on the profiling of the esophageal microbiome took place in the 1980s and used culture-based methods. In these first attempts, in which samples of luminal esophageal washes were analyzed, no species, or only a few, could be identified, leading to the postulation that the esophagus is almost free of bacteria [39, 92]. In tissue biopsies, it was possible to identify more bacteria. Differences in the number of species identified in samples gained by esophageal luminal washes and those gained by tissue biopsies can be explained by the fact that only transient bacteria were detected by luminal washes, while the species associated with the mucosal layer were not detected [95]. Studies using microscopy found that the bacteria are related to the mucosa [91]. Since then, using the 16S rRNA gene sequencing method, 95 species-level operational taxonomic units (OTU) have been revealed from biopsies of the normal esophagus [95]. Researchers assume that not all species inhabiting the esophagus have yet been identified, the estimated number of species being around 139 [95]. The higher number of taxa identified by 16S rRNA gene sequencing in comparison to culture-based methods indicates that not every species is culturable, or that some require special cultivation settings [95]. These facts underline the importance of choosing the appropriate material and technique to sample and identify OTU in the esophageal microbiome.

The most common method that is currently being used to identify microbiome OTU is 16S ribosomal RNA (16S rRNA) gene sequencing. One of the reasons why the 16S rRNA gene is used to identify the bacterial communities is that it can be found in almost all bacteria [56], thus providing a genetic marker for bacteria. Other reasons are (i) the presence of high variable regions within the gene, making it possible to identify different bacteria [22], in the best scenario at species level; (ii) the presence of conserved gene regions within the 16S rRNA gene, which can be used to produce suitable primers for amplification [116]; (iii) the presence of several databases in which the sequencing datasets can be compared, making it possible to assign taxonomy to sequences [6]; and (iiii) the presence of well-established primers that are specific for bacteria [123, 133]. Operational taxonomic units represent groups of similar sequence variants of 16S rRNA

gene [94]. Comparing the sequences in the available databases, a taxonomic assignment can be provided to each OTU. Studies on human esophageal biopsies, involving small cohorts, have so far shown that the microbial composition of subjects with reflux esophagitis or BE differs from that of the normal esophagus, with BE biopsies showing an overabundance of Gram-negative bacteria [96, 143].

Based on this knowledge, in this study we aimed at evaluating the role of the microbiome as a biomarker for esophageal adenocarcinoma and researching the potential link between EAC risk factors and microbiota. Epidemiologic factors, when separately considered, are not reliable for risk stratification among patients with or without predisposition for malignant progression. For example, even though reflux is considered as the main cause of BE, studies involving a large number of patients describe more cases of BE, HGD, or EAC among patients without reflux symptoms than among patients with reflux symptoms [33], showing that reflux symptoms are not an adequate predictor for disease development or progression. For the generation of prediction models, it is important to consider the prediction priority of the evaluation of multiple factors compared to that of single factors [105]. In this sense, it is important to acquire knowledge of the epidemiological background, host genetics, and microbial community to design a comprehensive risk prediction model [105] for such a multifactorial disease as BE. Considering the high-throughput sequencing methods, the microbiome might currently outline a feasible marker. Snider et al. showed that different sampling locations such as squamous epithelium, BE, or cardia have similar biota diversity and composition [126]. This emphasizes the potential role of microbiota in risk stratification among Barrett's patients, where the diagnostics are currently limited to local biopsy sampling and its histological evaluation, with a permanent remaining risk of missing a progression, and the likelihood of unnecessary effort for patients without progression.

Furthermore, we also aimed at identifying microbiome composition in materials gained by non-invasive sampling such as saliva and feces. Several studies support the notion that the esophageal microbiome is similar to the oropharynx microbiome, suggesting saliva as a possible and readily accessible biomarker for pathologies of distal esophagus [61, 95].

The intestinal microbiome might have a systemic impact by changing the inflammatory status in the human organism. Consumption of a high-fat diet has been linked to changes in the fecal microbiome and simultaneously to BE progression in the IL1β mice [80]. The bile acid pool size in blood has been shown to depend on intestinal bacteria, which metabolize primary bile acids such as tauro-β-muricholic acid (T-BMCA) to secondary bile

acids. Secondary bile acids are agonists of the Farnesoid-X-Rezeptor (FXR); this receptor can be found among other organs also in the intestine. FXR activation through secondary bile acids in the intestine induces the release of FRF-15, which inhibits the key enzyme for bile acid synthesis in the liver (cholesterol 7α-hydroxylase (CYP7a1)) and in this way regulates bile acid blood levels [111]. Changes in microbiome composition or abundance may influence bile acid blood levels. In the IL1β-mice, the damaging potential of unconjugated bile acids submitted to drinking water has been shown, as it enhances BE progression to EAC [102]. Bile acids may exert their damaging influences on esophageal tissue locally through bile acid reflux or possibly systemically through high unconjugated bile acid blood levels. To summarize, microbiota linked to a high-fat diet or bile acid metabolism might serve as potential markers for systemic inflammation and disease progression.

It would be useful if future studies could evaluate the functional role of microbiota in esophageal pathologies. If the findings were to show that the microbiome has a causal role in pathogenesis, it could potentially be used for chemoprevention or treatment. The microbiome has simple modifiable properties and might thus present a readily accessible therapeutic target. Antibiotics, probiotics, prebiotics, or microbiota transplants can be applied to modify the microbiome [7]. Antibiotics with a limited germ spectrum, non-absorbable antibiotics, or the combination of antibiotics with chemotherapy [19] should be researched with caution. For example, a 2013 study showed that the effect of oxaliplatin and immunotherapy was reduced in antibiotic-treated- and germ-free mice [54].

2 Materials and Methods

2.1 Study design, epidemiological data, and sample acquisition

2.1.1 Study design

The present study was a case-control study involving 48 participants, with cases diagnosed with BE (n=15), BE-associated dysplasia (n=4), esophageal adenocarcinoma (n=15), and controls with no BE-related pathologies (n=14). The study was performed in Klinikum rechts der Isar from March 2017 to February 2018 and was part of the BarrettNET study described in detail in Section 1.3. This study was approved by the local ethics committee (#5428/12). All subjects underwent upper endoscopy for clinical indications, including the control subjects. At the time of recruiting patients in the study, the diagnosis of Barrett's Esophagus, dysplastic Barrett's Esophagus and Esophageal adenocarcinoma had already been confirmed by a pathologist from previous patient surveillance visits. The final classification of patients into a diagnostic group was determined based on the most advanced current or previous histological diagnosis, after the biopsies of the most current patients' visit were evaluated by a pathologist.

The objectives of this project were i) the establishment of a methodology for studying the human microbiome, from biospecimen sampling to the final bioinformatical microbiome data analysis; ii) the investigation of the association between different stages of disease and local tissue-, fecal-, and salivary-microbiomes to identify the microbiome linked to disease progression; iii) the investigation of the association between smoking history/reflux symptoms and microbiota to assess the potential relationship between the microbiome and the most proximate noxious substances linked to esophageal adenocarcinoma development.

Materials collected in this study included a questionnaire, patients' epidemiological data, blood samples, tissue biopsies, and saliva and fecal samples. This was the first attempt to characterize the biopsy-associated salivary and fecal microbiomes in a small group of patients within the BarrettNET study. The collection of fecal and saliva samples is now an integral part of BarrettNET and continued even after finalization of this project, with the objective of collecting information about the microbiome composition from a larger patient cohort.

The first step to recruitment of possible study patients consisted of researching the respective internal center database, initially with the help of trained study staff. Strict inclusion and exclusion criteria were considered to obtain a relatively homogeneous patient group. The inclusion and exclusion criteria are described in Section 1.3 of this work.

In this project, patients were recruited at the Department of Gastroenterology, Interdisciplinary Endoscopy, Klinik und Poliklinik für Innere Medizin II. Patients were consecutively approached to participate in the study according to the regulations of the ethics committee. They were informed about the study's purpose and sampling, its risks, the benefits of participation, and participants' rights, as well as the content and extent of the epidemiological questionnaire. All participants gave written, informed consent. To protect the identity of participants, each patient was assigned a barcode, which was used to identify all samples and epidemiological information provided that patient. Study inclusion and epidemiological survey data were subsequently registered in an online study database. The blood samples and tissue biopsies were provided by the physician performing the endoscopy. Additionally, saliva- and fecal-sample collection kits containing written instructions, shipment boxes, and all the necessary components for sample collection in home conditions were provided to patients. To summarize, this study involved acquiring epidemiological, clinical-pathological, and microbiota information, and integrating the acquired information for each patient. It represents a concept of individualized medicine that aims to develop a more individual and comprehensive risk stratification for BE patients, and presents simultaneously a tool for evaluating the potential role of microbiome composition and diversity in BE and its malignant progression. Figure 2 presents the working methodology of the current microbiome study starting from sample collection to final bioinformatic microbiome and metadata integration and analysis.

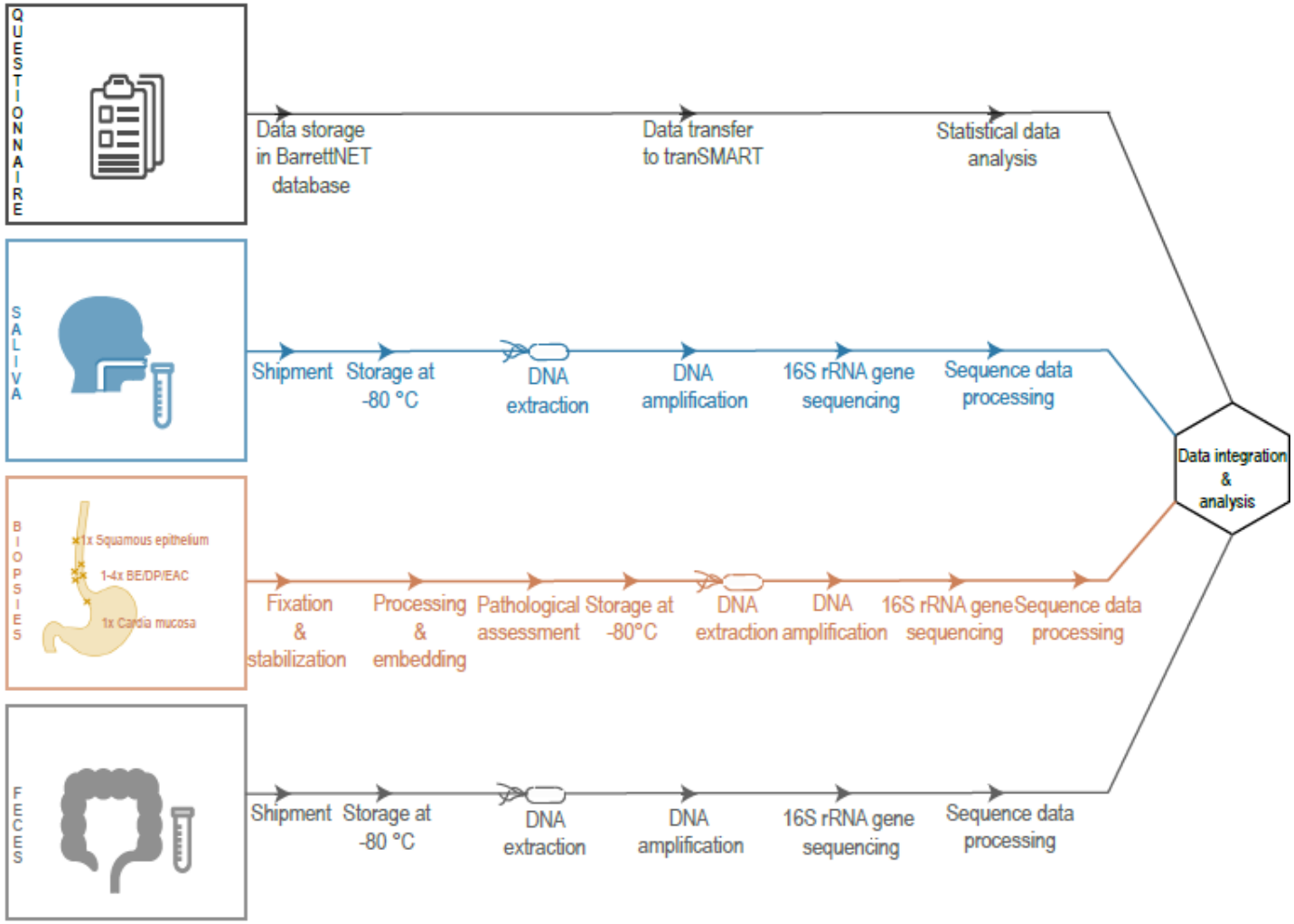


Figure 2. Study workflow, from sample collection to the final bioinformatical microbiome data analysis.

2.1.2 Epidemiological and clinical-pathological data

The questionnaire completed by the patients in this study corresponds to the survey designed for the BarrettNET study, which is a very comprehensive survey—no additional information was required in the context of this microbiome project. The questionnaire of the BarrettNET study contains 10 main topics, including demographics, general medical history, BE history and family history, epigastric pain in terms of days/week and intensity, chronic diseases, medication, physical activity, smoking, alcohol consumption, and diet [141]. Each topic contains precise questions and response options, thereby allowing a specification and an exact evaluation of patient statements. The design of this survey was modified according to items of the KORA (Cooperative Health Research in the Region of Augsburg, Southern Germany) study [49, 140] for reasons of comparison with a population-based German cohort. For non-German speakers, an English version of the questionnaire was provided. Patients were supported in the event of queries or required clarification. The survey answers were recorded and stored via a mobile study database (BarrettNET DIS database).

The clinical-pathological data was provided by searching the clinic database and summarizing the latest information found from endoscopic imaging, radiological imaging, pathology reports, and physicians' letters. Among other things, this was an important step for determining the pretherapeutic cancer stage in patients diagnosed with EAC.

For the statistical analysis of the data stemming from the epidemiologic questionnaire, and for hypothesis generation, we used tranSMART, a Data Warehouse, where data from different source systems are stored, incorporated, and statistically evaluated. In order to make constructive use of tranSMART, we took part in a training course held by a member of the IMSE team (Institute for Medical Informatics, Statistics and Epidemiology) at the Klinikum rechts der Isar. After the epidemiological information was obtained from all patients, the data was exported from the BarrettNET DIS database, and after necessary modifications it was uploaded to tranSMART. In tranSMART, to determine statistical differences between the analyzed groups, the Fisher exact test was used for categorical variables, whereas analysis of variance (ANOVA) was used for numerical variables.

2.1.3 Biospecimen acquisition and preanalytical workflow

Tissue

In addition to the routinely performed biopsies, two to six forceps biopsies were taken during upper endoscopy from each patient in this cohort: one to four biopsies from suspected Barrett or neoplastic tissue, if present; one from the inconspicuous region proximal to the lesion (squamous epithelium of distal esophagus); and one from the inconspicuous region distal to it (cardia region of the stomach). For the preanalytical preparation of tissue, the PAXgene tissue system, a formalin-free tissue fixative and stabilizer, was used. It consists of a PAXgene tissue container including two chambers, one containing a fixative and the other a stabilizing solution. Biopsies were placed immediately after provision into a cassette with six compartments for up to six biopsies and after that into the PAXgene tissue fixative chamber of the dual-chamber tissue container. After this procedure, the PAXgene container was scheduled with a patient barcode, and the corresponding batch card was sent to the tissue biobank in the Department of Pathology at Klinikum rechts der Isar. After two hours of fixation, the cassette was transferred into the PAXgene tissue stabilizer chamber. The length of stabilization was two hours at RT. Downstream tissue processing and paraffin-embedding were conducted based on established SOPs. For long-term archiving, the tissue biopsies were stored at -80°C. Part of the biopsy block cuts had been stained by our cooperation team for necessary pathological assessment by one gastrointestinal pathologist.

In addition to the biopsy cuts, which were necessary for purposes of histological assessment by a pathologist, additional tissue biopsy cuts were generated to perform stains for morphology review and a goblet cell ratio evaluation. From two to six tissue blocs provided from each patient, cuts of 2.5µm thickness were processed using a microtome and placed into object slides (Super Frost Plus, Microm, Walldorf). Slides were air-dried for 12 hours at room temperature and subsequently baked at 60°C in an oven for one hour (Functionline T12, Hereaus, Hanau). This step removes the paraffin. Hematoxylin-eosin (H&E) and Alcian blue and periodic acid-Schiff (PAS) stains were then applied.

Saliva samples

Saliva samples required for salivary microbiota analysis were self-sampled by patients via the Stratec SalivaGene® Collector with DNA stabilizer. For the self-collecting and shipment procedure, a saliva collection kit with a SalivaGene® collection tube containing 150mg of dry stabilization buffer, disposable gloves, an illustrated description of the col-

lection procedure, as well as a shipment box, was provided. The collection occurred either prior to endoscopy or at home. Collection prior to endoscopy was favored, to avoid bias, as fasting was a requirement of this procedure, for which the subjects had to abstain from eating, drinking, smoking, or gum chewing during the 30 minutes before saliva provision. However, some patients preferred to process sampling at home. All subjects received written and verbal instructions about the sampling procedure of saliva samples. Subjects were required to provide 2mL of saliva and to shake the tube gently in order to dissolve the dry DNA stabilization buffer (150mg of stabilization buffer) (STRATEC Molecular GmbH, 2018). This process takes a few minutes and requires a high patient compliance, as patients need to rub their cheeks from the outside and press them against their teeth to encourage saliva secretion. The shipment was addressed to one of the biobank laboratories in MRI. Upon receipt, samples were stored at -80°C.

It is noteworthy that the collection of saliva samples was initiated after the collection of fecal samples. For this reason, the first patients included in this study could not obtain a saliva collection kit on their first visit in person to the clinic. These subjects were called at a later date and invited to provide their saliva sample for study purposes. Upon patients' approval, a saliva collection kit and a shipment box for specimen return was sent.

Fecal samples

Fecal samples required for fecal microbiome analysis were self-sampled by patients using the Stratec® stool collection tube with DNA stabilizer mostly in at-home conditions, or at the clinic in cases where patients were hospitalized. For the self-collecting and shipment procedure, a stool collection kit was provided, including a stool collection tube with integrated spoon and 8ml of stool DNA stabilizer reagent, a second light impermeable tube, a special stool specimen collection unit, disposable gloves, and an illustrated description of the collection procedure, along with a shipment box. In addition to the written and illustrated Information, patients were given verbal instructions regarding the collection procedure, the optimal stool sample size, and shipping possibilities. The sampling procedure consisted of taking a small amount of stool (small bean size) with the integrated spoon, placing it in the stool collection tube, shaking the mixture, and placing it in the light impermeable tube (STRATEC Molecular GmbH, 2018). For the hospitalized patients, the stool samples were collected during their stay in the clinic by the patients themselves and picked up by the study team in the corresponding ward. The samples were shipped to the same biobank laboratory in MRI as for the saliva samples and were stored at -80°C.

2.2 Tissue staining

2.2.1 Hematoxylin and eosin (H&E)

Two to six slides per patient were stained with H&E to control the morphology and to provide the diagnosis. For this purpose, a deparaffinization using xylol (2 x 20 minutes) and a rehydration step using the decreasing alcohol solutions (2 x 100%, 2 x 96%, 2 x 70%, 2 minutes each) were applied. In this procedure, slides were washed for 3 minutes using distillated water (dH2O) and transferred to Mayer modified hematoxylin solution for 3 minutes to stain acid tissue structures. After a short washing step using tap water, the samples were transferred to an ethanolic eosin solution 0.33% for 3.5 minutes to stain the basic tissue structures. For dehydration, 96% ethanol and isopropanol was used. A deparaffination step with xylol (2 x 1.5 minutes) followed. Finally, slides were covered using Pertex®.

2.2.2 Alcian blue and Periodic Acid-Schiff (PAS)

Alcian blue and PAS staining was performed on all available human tissue biopsies to verify the diagnosis histologically and to evaluate tissue differentiation calculating the goblet cell ratio. Alcian blue visualizes sulfated and carboxylated acid mucopolysaccharides [139], thus highlighting the goblet cells of Barrett's metaplasia. Alcian blue stained parts appear blue to bluish-green. PAS stain is a reaction used to highlight tissue neutral mucopolysaccharides. Periodic acid oxidizes carbohydrates. After this reaction, aldehyde groups are formed, which create a magenta tone in reaction with Schiff's agent [71]. Initially, mounted slides were deparaffinized in xylol (2 x 20 minutes) and rehydrated using the decreasing alcohol solutions (2 x 100%, 2 x 96%, 2 x 70%, 2 minutes each). A dH2O washing step followed, and the slides were then placed for 20 seconds in Alcian blue to stain acidic polysaccharides, washed under running tap water, rinsed briefly in dH2O, and dipped in periodic acid (0.5%) for 5 minutes if fresh, or 15 minutes if very old solution. After a washing step using running tap water and a short rinse using dH2O, the slides were transferred to Schiff's reagent for 15 minutes. They were washed under running tap water and given a short rinse in dH2O, followed by a counterstain with Gill III modified Hematoxylin, a further washing step, dehydration, and deparaffination. Pertex® was used as mounting medium.

A previous case-control study of our research group revealed a decreased goblet cell ratio from the BE region of patients diagnosed with high grade dysplasia and early carcinoma compared to the BE region of patients with non-dysplastic BE, thus suggesting goblet cell ratio as a risk prediction factor for malignant progression [115]. Goblet cells are considered as highly differentiated cells unlikely to transform to cancer cells, and their rate correlates with the degree of tissue differentiation [115]. To ascertain whether results of the present patient cohort were consistent with the findings of our previous study with approximately 190 patients, goblet cell ratio was calculated. In this project, each Alcian blue and PAS-stained biopsy containing BE region from patients with a reliable diagnosis of BE, dysplasia, or cancer was evaluated. The total number of Barrett crypts regardless of the presence of goblet cells and the number of Barrett crypts containing goblet cells were counted. Subsequently, the quotient of the number of Barrett crypts with goblet cells to the total number of Barrett crypts was formed. A Barrett crypt was defined as any lumen lined with a columnar epithelium and depending on whether a crypt contains goblet cells or not, it is evaluated as positive or negative for goblet cells; glands opened to the biopsy surface were not considered, to avoid counting pseudo-goblet cells [115]. Notably, no Barrett's region could be identified in biopsies of patients with the reliable diagnosis of EAC, and thus no goblet cell ratio was calculated for this group of patients; only biopsies from the BE and dysplasia patient groups were evaluated. Eight biopsy specimens from patients with the diagnosis of BE and low-grade or HGD were analyzed using Aperio ImageScope, a slide viewing software with analyzing tools. Using Aperio ImageScope, an overview of the overall biopsy and a detailed view was provided. Since three to six biopsies were taken from each patient with suspicious morphology of the distal esophagus, for some patients there was more than one biopsy available containing BE region. Therefore, for those patients, the mean value from the determined GC ratio from each biopsy was calculated. To evaluate whether the determined GC ratio differed significantly between the two diagnosed groups, a two-paired t-test was performed.

2.3 DNA extraction for 16S rRNA sequencing

DNA was extracted from all available samples, including tissue, feces, and saliva. For each sample type a different DNA extraction protocol was used. Additionally, water samples were processed for DNA and used as negative controls. In order to measure the concentration and purity of the extracted DNA, we used NanoDrop® Spectrophotometer ND-1000 (ThermoFisher Scientific, USA).

2.3.1 Tissue biopsies

For the extraction of DNA from PAXgene-fixed, paraffin-embedded biopsies, we used the PAXgene® tissue DNA Kit (Preanalytix), following the manufacturer's instructions.

Sections of 10µm thickness were generated from each PFPE biopsy, and the first two or three sections were discarded because of block exposure to air. Two to three sections of all available biopsies of a patient were pooled together. The excess paraffin surrounding the tissue was discarded during cryotomy by removing it with needles.

The samples were then placed in a 1.5 ml microcentrifuge tube; EtOH and 650 µl Applichem, which is a xylene substituent that does not precipitate, were added to the sample and incubated for 3 minutes. 650µl ethanol of 100% purity was added and a centrifugation for 5 minutes at maximum speed was performed. After this, the supernatant was removed by pipetting. The tube was then opened and incubated for 5 minutes at 37°C using a shaker-incubator for dispersal of remaining alcohol. Following this, 150ul Buffer TR1 was added, and the pellet was resuspended by vertexing, after which 130ul of RNase-free water and 20ul proteinase K were added. The tube was then incubated for 5 minutes at 45°C, using a shaker-incubator at 1400rpm. After incubation, a short centrifugation was performed to remove drops from inside the tube. 200ul Buffer TR1 were added, and after a centrifugation step the sample was pipetted into a PAXgene DNA spin column and centrifuged. The sample was stored at room temperature (15–25°C) for up to 24 hours for DNA purification. After 24 hours, 350 µl Buffer TD3 was added to the PAXgene tissue DNA spin column and centrifuged for 1 minute at 6000 x *g*. The eluate was placed in a new processing tube and the flow-through was discarded. After this, 25µl proteinase K was mixed with 50µl Buffer TD3 and centrifuged. This mixture was pipetted onto the PAXgene DNA spin column and the eluate was incubated for 30 minutes at ambient temperature (20–30°C). After this, 350 µl Buffer TD3 and 500 µl Buffer TD4 were added to the PAXgene DNA spin column and centrifuged; the old processing tube containing the flow-through was discarded and a new one was used. To dry the membrane, a 3-minute centrifugation step at maximum speed followed, and the flow-through was discarded. Following this, 20–200 µl Buffer TD5 was added to the PAXgene DNA spin column membrane and centrifuged for 1 minute. The precipitation was performed via sodium acetate precipitation. For the resuspension, dH2O was used. During library preparation the samples were stored at 4°C.

2.3.2 Saliva

The DNA extraction from the saliva samples was performed using the DNA mini kit (Qiagen), following the manufacturer's instructions for DNA purification from saliva stabilized in RNAprotect® Saliva Reagent. The first 38 saliva samples obtained from the patients of this cohort underwent the DNA extraction procedure. The DNA was extracted from the total amount of saliva available from each patient (normally 2mL saliva).

The following steps prescribed from the manufacturer were followed: Initially, saliva samples were balanced to RT before the DNA extraction procedure, as they had been stored at -80ºC. The mixture of saliva and RNAprotect® Saliva Reagent was centrifuged for 10 minutes at 10,000 x g, the supernatant was eliminated, and the pellet was loosened. After this, 350 µL Buffer RLT, which contains among other substances guanidine isothiocycanate to facilitate the binding of DNA to the silica membrane, was added and the pellet was completely dissolved by vortexing. The suspension was subsequently transferred to a QIAamp Mini spin column placed in a 2 ml collection tube and centrifuged for 30 seconds at 8000 x g. A new 2ml collection tube was taken, and the spin column was stored there briefly at room temperature. Then, 350 µl Buffer AW1 was added to the spin column and centrifuged for 15 seconds in order to wash the spin column. The exact composition of the buffers (Buffer RLT, AW1, AW2) is confidential and not precisely described by the manufacturer. 20 µl Proteinase K mixed with 60 µl distilled water was added to the spin column membrane and incubated at 56ºC for 10 minutes. After this, the spin column membrane was washed using 350 µl of Buffer AW1. After throwing away the flow-through, an additional washing step using 500 µl Buffer AW2 was performed. A 3-minute centrifugation at maximum speed followed, in order to remove the ethanol, which had been used to suspend the washing buffers (AW1 and AW2), since ethanol may interfere with downstream reactions. 100 µl of Buffer AE (comprising 10 mM Tris-Cl and 0.5 mM EDTA; pH 9.0) was added two times to the spin column and centrifuged for one minute each time to elute DNA. The applied centrifugal force for all centrifugation steps was 8000 x g except for the last washing step, in which the highest adjustable force was used. During library preparation, the samples were stored at 4°C.

2.3.3 Feces

The DNA extraction from patient fecal samples was conducted using silica beads and subsequently based on a modified version of the Godon et al. protocol [41]. The first 42 fecal samples obtained underwent DNA extracting procedures. The mixture of feces and

DNA stabilizer (Sarstedt, Stool Collection Tubes® with Stool DNA Stabilizer, Cat. 1038111300) had been frozen at -80°C since sample arrival in the human biobank laboratory for long-term storage.

In the first step, aliquots of 700µL were dispensed into cryovial tubes on dry ice, and 250 µL of 4 M guanidine thiocyanate - 0.1 M Tris (pH 7.5) and 500 µL of 5% N-lauroyl sarcosine - 0.1 M phosphate buffer (pH 8.0) were then added to the aliquots and vortexed, followed by incubation for 1 hour at 70°C. For mechanical bacterial membrane disruption, sterile silica beads (0.1 mm, Biospec products) were used in a FastPrep-24 bead beater (MP biomedicals) with cooling adaptor. After this, 15 mg of Polyvinylpolypyrroli-done (PVPP, Sigma Aldrich), which served as a polyphenol adsorbent, was added. This suspension was then centrifuged at 15.000g at 4°C for 3 minutes. The supernatant was collected and once again centrifuged at 15.000xg at 4°C for 3 minutes. 2µL RNAse (10 mg/mL) were added to the supernatant (500 µl) and the solution was incubated at 37°C for 30 minutes with continuous shaking at 700 rpm in order to discard bacterial RNA.

The final steps involved purification of the genomic DNA with the NucleoSpin gDNA Clean-up kit (Macherey Nagel) following the manufacturers protocol: To 500µL elute, 300µL Tris EDTA buffer (10 mM Tris/HCl pH 7.5, 0.1 mM EDTA) and 200 µL NT buffer (guanidinium thiocyanate) were added and the solution was vortexed. For each sample, one NucleoSpin® gDNA Cleanup XS Column was placed into a collection tube (2 mL). 500 µL of the mixture was placed in the column and centrifuged for 30 s at 11,000 x g, and the supernatant was removed. The remaining mixture was added to the column again and centrifuged for 30 s at 11,000 x g. The silica membrane was then washed using 100 µL Buffer B5 to remove contamination and centrifuged for 2 minutes at 11,000 x g. The NucleoSpin® gDNA Clean-up XS Column was placed into a 1.5 mL microcentrifuge tube, and then 6–15 µL BE buffer (5 mMTris/HCl, pH 8.5) was twice added directly to the center of the membrane and centrifuged for 1 min at 11,000 x g each time. Finally, the residual ethanol was removed by incubating the elute with an open lid at 90 °C for 8 minutes. During library preparation, the samples were stored at 4°C.

2.4 16S rRNA gene sequencing, sequencing data processing, and microbiome profiling

To characterize the microbiota, the variable regions V3–V4 of the 16S rRNA gene were amplified. The 16S rRNA gene contains around 1550 bp and includes nine variable regions across the conserved 16 S gene sequence, which enables the identification of

bacteria [58]. Here, we amplified only the variable regions V3–V4, which are about 450 bp long, to identify the microbial communities. The amplification of the V3–V4 region of 16S rRNA gene was performed through 25 cycles for fecal and saliva samples, and 15 cycles for tissue samples from the processed DNA (12ng from each sample of tissue, saliva, and feces processed) using the bacteria-specific primers 341F and 785R [63] and a two-step protocol. As Caporaso et al. 2011 described, the 5' and 3' primers used for the V4 region of bacterial 16S gene, contain generally three parts: an adapter, which provides binding to the DNA; a linker and pad sequence, which improves linking of the primer to the DNA; and the sequence complementary to the target gene region (V3/V4) [18]. In addition, the 3' primer contains barcode sequences to provide assignment of the sequencing reads to each sample [18]. The purification of the replicated material was processed using the AMPure XP system (Beckman-Coulter, MA, USA). 16S rRNA gene sequencing was conducted in a run for the first 124 samples collected (43 tissue biopsy samples, 38 saliva samples, and 42 fecal samples) in paired-end modus utilizing the MiSeq system (Illumina, CA, USA).

The next step was to process the raw data provided from the sequencing. For this purpose we utilized the integrated microbial NGS platform (IMNGS), created by ZIEL Institute for Food and Health, Core Facility NGS/Microbiome, Technische Universität München, Freising, Germany [67]. The raw amplicon dataset was first demultiplexed to the samples they belonged to, based on the mapping file incorporating the barcodes of each sample [85], permitting a maximum of two errors in the barcodes. After that, each sample was processed using the UPARSE approach [29]. Cutoff values were set to provide an accurate processing of the sequencing data. The determination of cutoff values was based on the experience of our collaboration laboratory in Freising, Germany, with the procession of sequencing data originating from human specimens, and on the information provided from negative controls (in this case, water samples). The first step of the UPARSE pipeline consisted of trimming the first base at both sequence ends to avoid possible bias [67]. The reads were then paired. After this step, sequences containing less than 400 nucleotides and more than 600 nucleotides were excluded. Pared sequences, in which >3 errors were found, were eliminated and the sequences left were trimmed by five nucleotides at both sequence ends. This was done to prevent GC bias and other composition-based bias. It is already known that GC-rich or GC-poor DNA material is associated with inaccurate sequencing results, mostly arising from bias during PCR as a result of inadequate amplification [14], [55], [100]. The pared sequences were then de-replicated. To monitor for chimeras on the de-replicated sequences, UCHIME

was utilized. The next step comprised the fusion of all sequences of all samples, arranged by abundance. Clustering of sequences into OTUs was done when sequences shared a 97% identity. Finally, all sequences were assigned to corresponding OTUs, and this way an OTU table was created. The OTU table contains in the first column the names of the OTU (named OTU_1 to OTU_999, a total of 682 OTUs); the first row shows the barcodes of each sample, and the next rows show the abundance of each OTU in each specific sample. The OTUs with a relative abundance higher than 0.5% in all sequences in at least one sample are included, and the other OTUs are excluded from the final table. The taxonomy of each OTU is presented in the last column. The assignment of taxonomy in IMNGS was done at an 80% confidence level by comparing the present sequences to those existing in the RDP classifier. The output files of the IMNGS platform consisted of one OTUs table, one file containing the sequences and one phylogenetic tree (constructed using fasttree). We used the SILVA database for ribosomal RNA genes (project 50) and the EzTaxon database at an 80% confidence level for taxonomic classification [67].

After modification of the OTU table, Rhea R-package was performed for downstream microbiome analysis. Rhea package includes a series of R scripts (six scripts) and is programed from the Core Facility Microbiome/NGS of Technical University of Munich, Freising, Germany. The analysis steps using the Rhea pipeline included: (1) a normalization of the OTUs table to compensate for bias deriving from unequal sequencing depths between samples; (2) downstream analysis to determine alpha diversity; (3) downstream analysis to determine beta diversity; (4) taxonomic-binding at kingdom, phyla, class, order, family, and genus level; (5) serial-group-comparisons to provide comparisons between groups determined by the user; and (6) correlations between metavariables and taxonomic variables [66]. Rhea scripts all have the same structure, which facilitates the application. They consist of the initialization stage, in which the user must change parameters accordingly; the comments, which assist in an understanding of every step of the analysis; and the main part. The input files must be processed by the user prior to analysis according to the individual project needs. In our analysis, the input files, including the OTUs table, the tree file, and the mapping files were processed prior to each analysis. The parameters of the initialization part of the scripts were changed according to the analyses we intended to conduct. In the main part of the script, the color settings of plots, as well as the order in which the groups were intended to be visualized, were changed manually.

The microbiome analyses consisted of (i) sample-type guided microbiota analysis, comparing microbiome diversity and composition in saliva, tissue, and feces; (ii) diagnosis-

guided microbiota analysis comparing microbiome diversity, composition and taxa's'/OTU's relative abundance in BE, DP, EAC and controls – this analysis was conducted separately for all three sample-types; and (iii) risk factors guided analyses. The risk factor-guided analyses compromised two separately analyses: (1) Analysis 1 included the comparison of subjects sub-stratified by (1i) smoking history (smokers vs. never-smokers); (1ii) presence of reflux symptoms (subjects with reflux symptoms vs. without reflux symptoms) – independent of the diagnosis; (2) Analysis 2 included the comparison of subjects sub-stratified by (2i) diagnosis and smoking history (BE smoker vs. BE never-smoker, DP smoker vs. DP never-smoker, EAC smoker vs. EAC never-smoker, Ctrl. smoker vs. Ctrl. never-smoker); (2ii) diagnosis and presence of reflux symptoms (BE reflux symptoms vs. BE no reflux symptoms, DP reflux symptoms vs. DP no reflux symptoms, EAC reflux symptoms vs. EAC no reflux symptoms, Ctrl. reflux symptoms vs. Ctrl. no reflux symptoms) – these analyses were performed separately for all three sample-types. Last, we identified bacteria, that were altered in both – patients with exposure to a risk factor and in – patients with progressed phenotype.

The first Rhea script performed was the normalization script, which calculates relative abundances of each OTU in order to offset the different sequencing depths between samples. We used the rarefication curves provided in this step to identify samples that were not sufficiently sequenced. Here, one fecal sample out of a total of 124 samples (including saliva, tissue, and feces) was excluded from downstream analysis. The normalized OTUs table containing the relative abundances of each OTU, which were computed here, was used in the following analysis steps.

In script 2, alpha diversity within each sample was determined. Alpha diversity is a parameter that shows the diversity of OTUs within a sample. To determine alpha diversity, only OTUs with an abundance greater than 0.5 counts were considered. There are different dimensions that can be used to express the alpha diversity within a sample, including Species Richness, Shannon and Simpson diversity indices, and effective number of species. Species Richness is calculated by counting the number of OTUs found in a single sample without considering the abundance of an individual OTU. In contrast, Shannon and Simpson diversity indices also consider the structure of the bacterial community by taking into account the abundance of every single OTU. After analyzing the limitations of indices, Jost proposed using Shannon and Simpson to effectively determine alpha diversity rather than indices [59]. The effective number of species represents the number of equally abundant species needed to reach the associated index value and is less distorted by the number of species found only rarely. All dimensions described above were calculated for each sample. We chose to visualize Richness and Simpson

effective number of species (i.e., not all indices of alpha diversity) in the above-mentioned groups to prevent misconception of data. A one-way ANOVA was applied to determine significance. If the result of the ANOVA was significant, the groups were analyzed pairwise using the Welch Two Sample t-test.

In script 3, beta diversity between samples was determined. Beta diversity describes the diversity of OTUs through different samples [66]. To determine beta diversity, distances between OTUs were calculated by using generalized UniFrac distances, which provide a precise calculation of differences between bacterial community members (OTUs) based on their genomic distances and considering the abundance of each OTU. We compared the microbial profiles after sub-stratifying samples as previously described. A PERMANOVA test was used to define the statistical significance of microbial profile distances between groups. The final step was the visualization of the provided distances by applying Multidimensional Scaling (MDS) and the non-metric version of MDS (NMDS).

Script 4 provides a calculation of the relative abundance of taxa that share the same taxonomic classification at a given level—for example, kingdom, phylum, class, order, family, genus, and species in a single sample—providing an overview of microbial composition at a higher taxonomic level than OTUs level. The output files of this script consist of six tables with respectively relative abundance of taxa in kingdom, phylum, class, order, family, and genus level for each sample; one taxonomic summary table including all levels for each sample; and a graphical illustration.

Script 5 compares the relative abundances of taxa and OTUs between predefined groups. The statistical tests used to define significance among groups were the Kruskal-Wallis Rank Sum test (a non-parametric analysis of variance) and the Fisher test, followed by a pairwise comparison of groups using the Mann-Whitney Test (a non-parametric test). The taxa relative abundances are not usually normally distributed, hence non-parametric tests were used for statistical analyses. Part of this script was also the correction for multiple testing utilizing the Benjamini-Hochberg method [10]. A corrected p-value was given beside the p-value calculated from the pairwise comparison.

Script 6 offers the option to run correlations between taxonomic and numeric metavariables. We did not perform this analysis in this work. The reasons for this included our intension to avoid such correlations analyses and keep our results descriptive. Additionally, the main research question in this work involves establishing methodology and characterizing the microbiome among groups sharing a fixed characteristic; and, in this study, these characteristics comprise (i) sample type; (ii) diagnosis; and (iii) exposure to risk

factors such as smoking and reflux complaints, all of which are categorical and not numerical variables.

3 Results

3.1 Epidemiological and clinical-pathological characteristics of the patient cohort

The demographics, lifestyle, medication and diet characteristics of this cohort are summarized in Table 3 to provide an overview of distribution in the different diagnosis groups. The mean age of this cohort was 62.02 years, ranging from 33 to 80 years. The gender distribution was 82.74% male and 17.26% female. The controls were age and gender matched. The diagnosis groups did not differ significantly in regard to age and gender. The only significant differences between the diagnosis groups were with regard to smoking and PPI usage: the number of smokers (versus non-smokers) was notably higher in the BE group and in the esophageal adenocarcinoma group than in the control group (BE vs. control p= 0.012, EAC vs. control p=0.008); and the PPI intake was significantly higher in the BE patients and in the DP patients than in the controls (BE vs. control p= 0.035, DP vs. control (p=0.015). No remarkable differences were revealed with regard to BMI and alcohol consumption. Intake of analgesics such as ASS, Ibuprofen, and Paracetamol, and heart medication, did not differ significantly among groups; however, there was a tendency toward a higher usage in EAC patients.

Characteristics	Control n=14	Barrett's esophagus n=15	Low-/high-grade dysplasia n=4	Esophageal adenocarcinoma n=15	p-value
Age, years					**0.294**
· Mean (SD)	56.4 (16.1)	61.7 (12.3)	65.0 (10.4)	65.3 (9.9)	
· Range	33-80	37-79	51-76	46-80	
Gender, n (%)					**0.405**
· Male	9 (64.3%)	11 (73.3%)	4 (100%)	13 (86.7%)	
· Female	5 (35.7%)	4 (26.7%)	0 (0.0%)	2 (13.3%)	
BMI, kg/m²					**0.299**
· Mean (SD)	25.5 (3.57)	27.6 (3.35)	26.3 (2.27)	27.7 (3.53)	
· Range	19,95 - 31,01	23.11 - 32.85	24.44 - 29.24	22.28 - 34.57	
Smoking, n (%)					**0.022**
· Smoker	3 (21.4%)	10 (66.7%)	3 (75.0%)	11 (73.3%)	
· Never-smoker	10 (71.4%)	4 (26.7%)	1 (25.0%)	4 (26.7%)	
Missing information	1 (7.1%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	
Alcohol consumption, n (%)					**0.065**
· Daily/almost daily	0 (0.0%)	2 (13.3%)	2 (50.0%)	4 (26.7%)	
· Rarer than daily/almost daily	13 (92.9%)	12 (80.0%)	2 (50.0%)	11 (73.3%)	
Missing information	1 (7.1%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	

Reflux symptoms, n (%)					**0.854**
· Yes	6 (42.9%)	7 (46.7%)	1 (25.0%)	6 (40.0%)	
· No	5 (35.7%)	7 (46.7%)	3 (75.0%)	9 (60.0%)	
Missing information	3 (21.4%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	
Medication, n (%)					
· ASS	1 (7.1%)	2 (13.3%)	1 (25.0%)	6 (40.0%)	**0.162**
· Paracetamol/Ibuprofen	2 (14.3%)	1 (6.7%)	0 (0.0%)	1 (6.7%)	**0.774**
· PPI	4 (28.6%)	10 (66.7%)	4 (100.0%)	9 (60.0%)	**0.04**
· Diuretics	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	**1.0**
· Corticoids	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	**1.0**
· Heart medication	3 (21.4%)	2 (13.3%)	0 (0.0%)	5 (33.3%)	**0.564**
· Hormones	0 (0.0%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	**0.659**
Missing Information	1 (7.1%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	
Diet					
Meat consumption, n (%)					**0.089**
· almost daily or several times a week	7 (50.0%)	6 (40.0%)	4 (100.0%)	12 (80.0%)	
· about once a week or rarely	5 (35.7%)	8 (53.3%)	0 (0.0%)	3 (20.0%)	
Sausage consumption, n (%)					**0.527**
· almost daily or several times a week	7 (50.0%)	10 (66.7%)	4 (100.0%)	11 (73.3%)	
· about once a week or rarely	5 (35.7%)	4 (26.7%)	0 (0.0%)	4 (26.7%)	
Missing information	2 (14.3%)	1 (6.7%)	0 (0.0%)	0 (0.0%)	

Table 3. Epidemiological characteristics of the cohort patients.

The DP and EAC groups, and the control group, included subjects with inhomogeneous clinical-pathological features. In the dysplasia group, 2/4 subjects were diagnosed with low-grade dysplasia and 2/4 with HGD. In the EAC group, 5/15 subjects showed T1 stage, 1/15 T2 stage, 9/15 T3 stage, 9/15 lymph node metastasis, and 2/15 distant metastasis at the time of their inclusion in the study. Regarding the histologic grading, 4/15 EAC subjects showed a well differentiated carcinoma (G1), 8/15 a moderately differentiated carcinoma (G2), and 3/15 a poorly differentiated carcinoma (G3) (Table 4).

The control patients attended the clinic for reasons of chronic upper abdominal pain (10/14) or due to reflux symptoms (4/14). For the latter group, the decision to perform an upper endoscopy was made by a clinic physician. Importantly, study participation did not interfere with the decision of whether or not to undergo an endoscopy. In fact, 9/14 controls were diagnosed with chronic gastritis, with 2 additionally diagnosed with reflux esophagitis, while 5/14 controls had no pathological finding in the upper endoscopy (Table 5).

A generally good patient compliance was observed regarding the range of information and samples provided. Patients showed a great willingness to cooperate in the collection

of saliva and feces, and 45 out of 48 patients included in the study provided saliva and fecal samples. Tissue biopsies were taken from 47 out of 48 patients.

Characteristics	n	(%)
Gender		
· Male	14	(93.3%)
· Female	1	(6.7%)
Depth of invasion		
· T1	5	(33.3%)
· T2	1	(6.7%)
· T3	9	(60.0%)
Lymph node metastasis		
· Yes	9	(60.0%)
· No	6	(40.0%)
Distant metastasis		
· Yes	2	(13.3%)
· No	13	(86.7%)
Histologic grade		
· Well differentiated (G1)	4	(26.7%)
· Moderately differentiated (G2)	8	(53.3%)
· Poorly differentiated (G3)	3	(20.0%)
AJCC TNM stage		
· I	4	(26.7%)
· II	4	(26.7%)
· III	5	(33.3%)
· IV	2	(13.3%)

Table 4. Demographic and clinical-pathological characteristics of cohort patients diagnosed with esophageal adenocarcinoma.

Characteristics	n	(%)
Gender		
· Male	9	(64.3%)
· Female	5	(35.7%)
Main symptom		
· Upper abdominal pain	10	(71.4%)
· Reflux	4	(28.6%)
Endoscopic diagnosis		
· Corpus-/antrum gastritis	9	(64.3%)
· Reflux esophagitis	2	(14.3%)
· No pathological finding	5	(35.7%)
Pathology report		
· Low grade chronic gastritis	5	(35.7%)
· High grade chronic gastritis	4	(28.6%)
· H. pylori +	4	(28.6%)
· H. pylori -	5	(35.7%)
No pathology report	5	(35.7%)

Table 5. Demographic and clinical-pathological characteristics of control patients.

3.2 Histopathological tissue analysis

Histopathological diagnosis differed partially from the endoscopic diagnosis

A total of 160 biopsies (68 biopsies from patients with BE, 17 biopsies from patients with dysplasia, 53 from patients with EAC, and 26 from controls) were stained with H&E to evaluate the morphology and to determine histologic diagnosis. In Figure 3, respectively, one image per localization is presented from H&E stained PAXgene-fixed biopsies of the present cohort: from the esophageal squamous epithelium, cardia epithelium, BE, DP, EAC from tubular type, and EAC from mucinous type.

It is noteworthy that only 15 out of 34 biopsies (44.1%) from BE subjects, 8 out of 9 biopsies (88.9%) from DP subjects, and 12 out of 23 biopsies (52.2%) from EAC subjects—which were taken from suspicious areas for BE, DP, or EAC—actually contained metaplastic, dysplastic, or cancer tissue (Table 6).

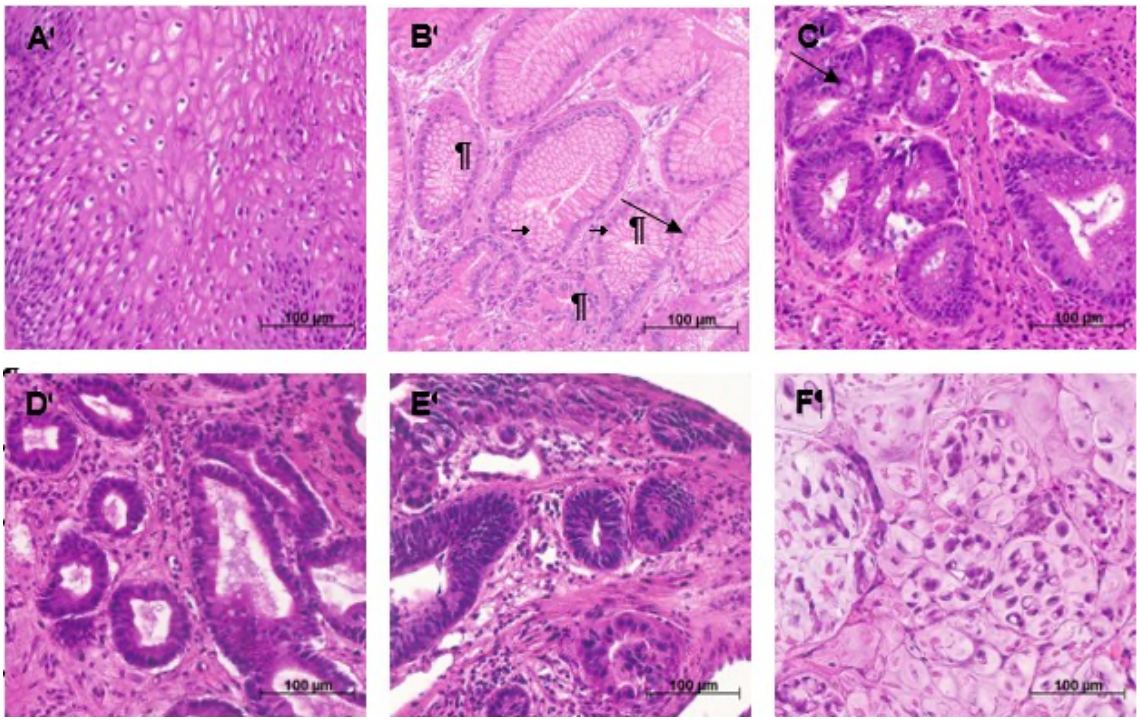


Figure 3. Representative Hematoxylin and Eosin (H&E) staining of PAXgene-fixed paraffin-embedded biopsies of the present cohort.
(A) stratified squamous epithelium of the esophagus; (B) columnar-lined epithelium of the cardia mucosa (arrow indicates the gastric pits); (C) columnar-lined epithelium with single goblet cells from a biopsy taken at a suspicious region at the esophagogastric junction, histologically well compatible with intestinal metaplasia of esophagus (Barrett's Esophagus); (D) columnar-lined epithelium containing cell atypia from a biopsy taken at a suspicious area at the esophagogastric junction, histologically compatible with low-grade dysplasia; (E) tubular adenocarcinoma of the intestinal type (Laurén-classification) of esophagus; (F) mucinous adenocarcinoma of esophagus.

Phenotype	Obtained biopsies	Obtained biopsies from cardiac mucosa	Obtained biopsies from normal esophageal mucosa	Obtained biopsies from suspicious areas at the esophagogastric junction	Biopsies with a histological diagnosis of metaplasia	Biopsies with a histological diagnosis of dysplasia	Biopsies with a histological diagnosis of carcinoma
BE	64	15	15	34	15	0	0
DP	17	4	4	9	7	1	0
EAC	53	15	15	23	0	3	12
Control	26	13	13	0	0	0	0
Total	160	47	47	66	22	4	12

Table 6. Number of biopsies and their histological assessment.

The Goblet cell ratio did not differ significantly between BE and DP patients

All available biopsies were stained with Alcian blue and PAS for the confirmation of the diagnosis and the calculation of the goblet cell ratio. Figure 4 presents images from Alcian blue and PAS-stained biopsies of patients diagnosed with BE, low-grade dysplasia, and esophageal adenocarcinoma. The goblet cell ratio provided from the metaplastic region of BE subjects was 0.61, of low-/high-grade dysplasia subjects 0.47 (Figure 5). Even through no statistical significance could be determined between BE and DP, a trend toward lower goblet cell ratio among patients with dysplastic BE was observed.

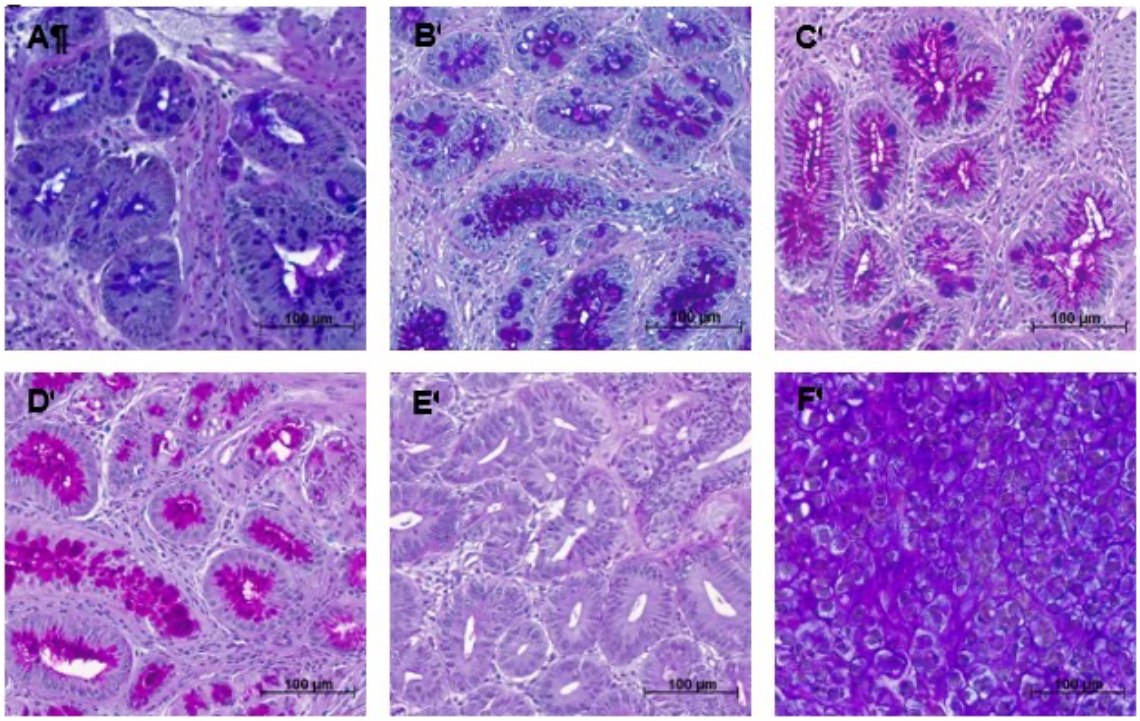


Figure 4. Representative PAS and Alcian blue Staining of PAXgene-fixed paraffin-embedded biopsies of the present cohort.
(A), (B) BE crypts containing goblet cells; (C) BE crypts containing partially goblet cells; (D) BE crypts containing no goblet cells; (E) esophageal adenocarcinoma glands, which have lost the ability to produce mucus; (F) esophageal mucinous adenocarcinoma containing no goblet cells.

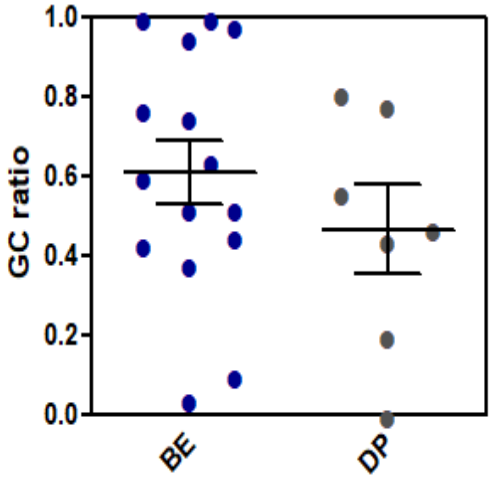


Figure 5. Goblet cell ratio in biopsies of BE and DP subjects.

3.3 Microbiome profiling

3.3.1 Sample type-guided analyses

Microbiota diversity and composition was associated with the sample type

In total, 124 samples from 137 samples collected were examined, including 44 tissue biopsies sets, 42 fecal samples, and 38 saliva samples (Table 7).

First, we investigated the diversity in all three sample types (Alpha diversity). This was assessed by different alpha diversity parameters, including Richness and Simpson effective. On average, we identified 90 OTUs in saliva samples, 156 OTUs in tissue biopsies, and 147 OTUs in fecal samples (Richness). Statistical analyses revealed that species richness was significantly lower in saliva samples compared to tissue biopsies and feces samples ($p<0.0001$), while no significant difference was determined between tissue and feces samples (Figure 6.A). While comparing Simpson effective between sample types—which, apart from the species number, also considers the abundance of each species within a sample—significant differences between all three sample types were revealed (saliva vs. tissue $p<0.0001$, saliva vs. feces $p<0.0001$, and tissue vs. feces $p<0.0491$) (Figure 6.B). As a result, fecal samples of this cohort showed the highest species diversity, followed by esophageal tissue biopsies and saliva.

Second, we compared the microbial composition between the three sample types (Beta diversity). The graphical illustration of beta diversity between saliva, tissue, and fecal

samples was provided by a non-metric multidimensional scaling (NMDS) plot of microbial profiles (Figure 6.C). The beta diversity statistical analysis revealed that fecal, salivary, and esophageal tissue microbial profiles differ significantly from each other ($p<0.0001$). Nevertheless, a slight overlap of tissue biopsy and salivary microbial profiles was observed, whereas fecal microbial communities were very different to both salivary and tissue-associated microbial communities.

Sample type had a greater impact on microbiome composition as patients' diagnosis

Figure 7 shows a NMDS plot of microbial profiles of samples, which were sub-stratified by the sample type and the diagnosis (BE, DP, EAC, controls). In this analysis, as expected, a clear separation was observed between biospecimen types, as saliva, tissue, and fecal samples clustered apart from each other, independent of the diagnosis of the subjects from which the sample had come.

On a closer inspection, exclusively in tissue, the diagnosis groups clustered together. This suggests that dysbiosis is present in local biopsies, and that the microbiome could sub-stratify and reflect diagnosis related to malignant progression. In addition to this brief overview of the relationship between microbiome composition and patients' diagnoses, deeper insights into microbiome alterations in disease onset and progression are provided in the next sections.

In total, 11 phyla were identified in tissue biopsies. The most present phyla were Firmicutes (21%), Actinobacteria (20%), Proteobacteria (19%), Bacteroidetes (17%), and Fusobacteria (15%) (Figure 8).

In saliva, 7 phyla were identified, all of which were also identified in tissue biopsies. The most prevalent ones were: Firmicutes (35%), Actinobacteria (20%), Bacteroidetes (16%), Proteobacteria (15%), and Fusobacteria (11%) (Figure 8).

In fecal samples, 7 phyla were revealed. The most prevalent ones were Firmicutes (37%), Bacteroidetes (19%), Proteobacteria (18%), and Actinobacteria (18%) (Figure 8).

Number of samples collected					
	Control	**BE**	**DP**	**EAC**	**Total**
Saliva	12	15	4	14	45
Tissue	13	15	4	15	47
Feces	11	15	4	15	45
Number of samples sequenced					
	Control	**BE**	**DP**	**EAC**	**Total**
Saliva	10	14	3	11	38
Tissue	11	14	4	15	44
Feces	9	15	4	14	42

Table 7. Total number of samples collected and sequenced.

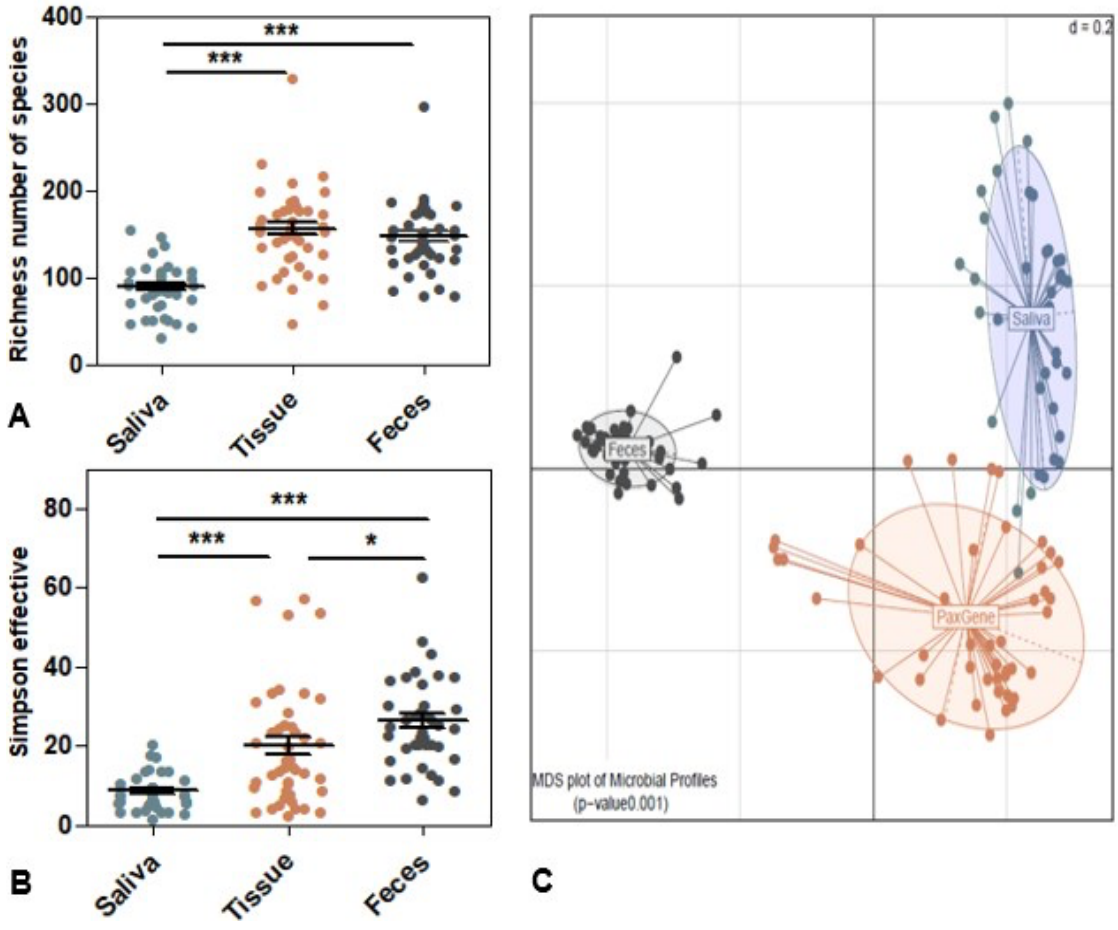


Figure 6. Alpha and beta diversity of microbiota in saliva, biopsies, and feces.

(A) Alpha diversity assessed through species richness, showing a significant difference in richness of number of species between saliva and tissue biopsies (p<0.0001), as well as between saliva and feces (p<0.0001). (B) Alpha diversity assessed through Simpson effective number of species, showing significant differences in number of species between saliva and tissue (p<0.0001), saliva and feces (p<0.0001), and tissue and feces (p=0.0491). (C) Beta diversity illustrated by the multidimensional scaling plot (MDS plot) of microbial profiles obtained in saliva, tissue, and feces, showing significant difference in the overall microbiome composition between the three sample types (p=0.001).

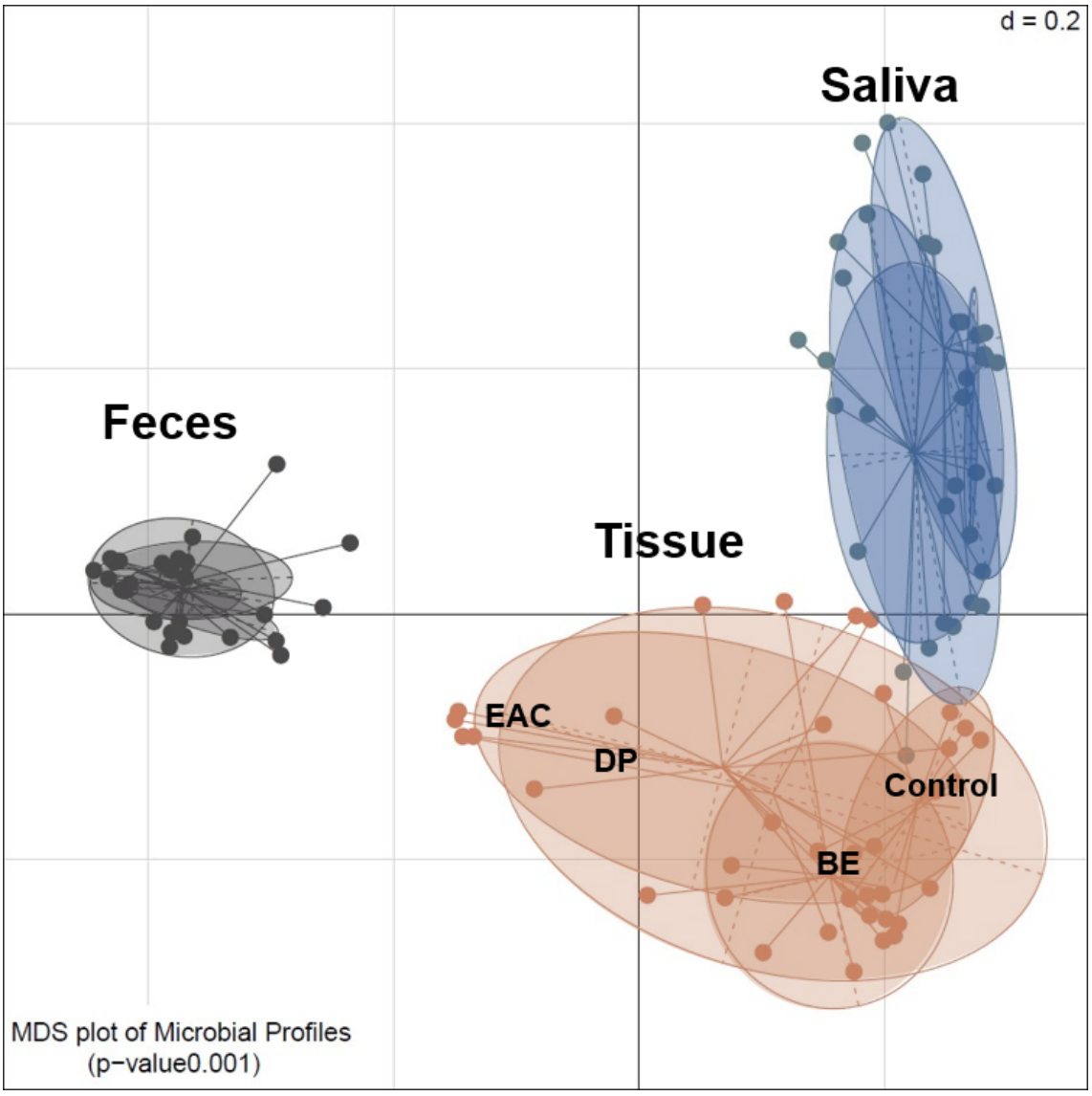


Figure 7. Beta diversity of microbiota in saliva, biopsies, and fecal samples after sub-stratifying the samples by diagnosis.
Beta diversity illustrated by a multidimensional scaling plot (MDS), showing a significant difference between microbial profiles in different sample types (p=0.001), as well as a greater impact of the sample type in the microbiome composition as the clinical diagnosis.

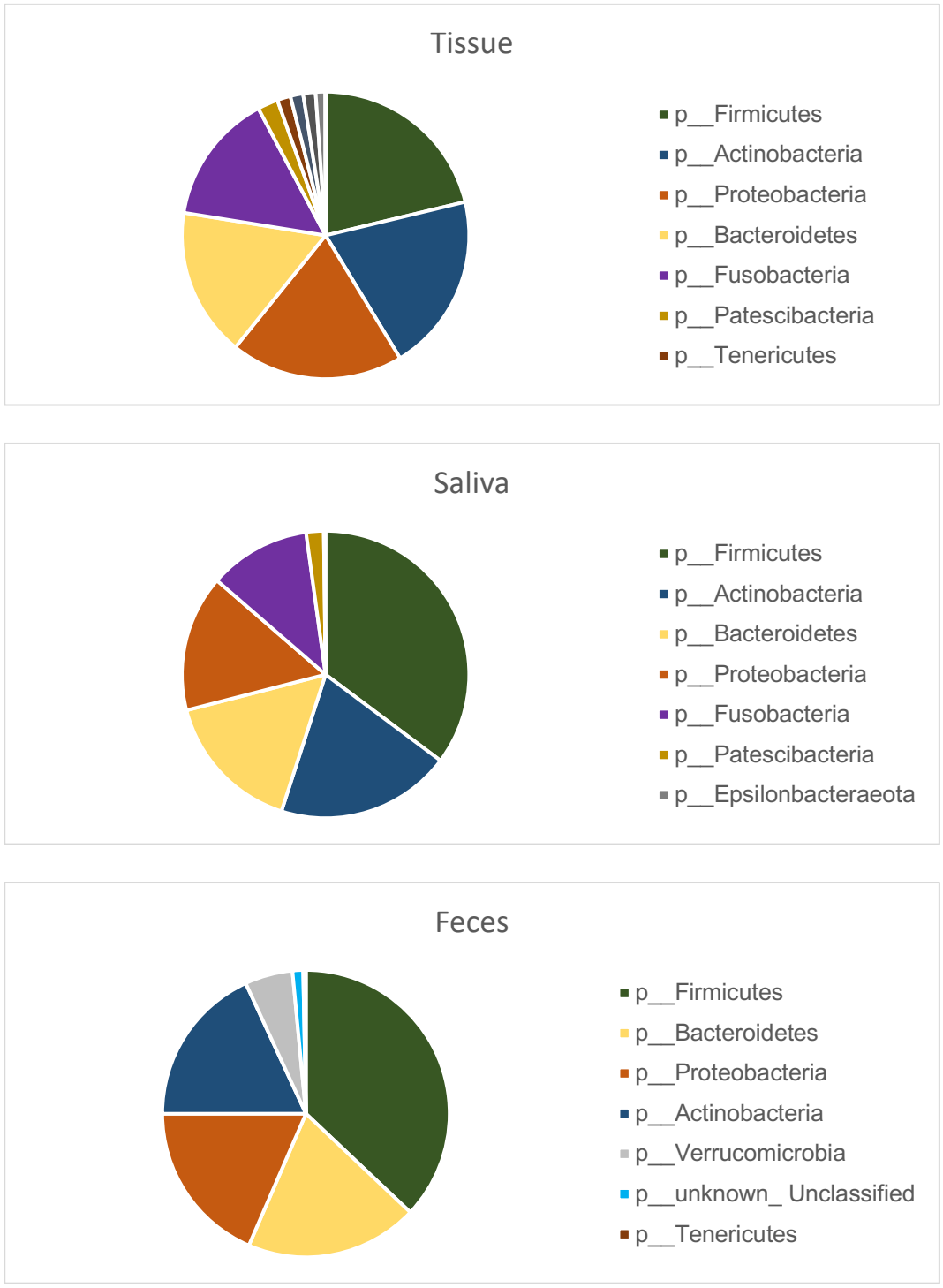


Figure 8. The most prevalent taxa in tissue, saliva, and fecal samples.

3.3.2 Diagnosis-guided analyses

Tissue biopsies

Microbiota diversity and composition in tissue biopsies was different between the diagnosis groups

Forty-three of 47 collected tissue biopsy sets (each set containing two to six biopsies taken from esophageal mucosa, cardia mucosa, and endoscopically suspicious areas for BE, dysplasia, or EAC) from this patient cohort were sequenced (Table 8). The quality control demonstrated that all tissue biopsy samples were sufficiently sequenced, and thus all sequenced biopsies were included in the downstream microbiome analysis. For the quality control of the sequencing, rarefication curves of all samples were provided (Figure 9). They demonstrate whether the sequencing depth is sufficient or not and, in this case, they showed that a plateau of number of species was achieved in each sample after all available reads were assigned (Figure 9).

The alpha diversity was assessed by different parameters, including Richness and Simpson effective, which are visualized in Figures 10 and 11. The Richness number of species did not differ between disease (including BE, DP, and EAC) and control, while there was a significant higher Simpson effective in disease ($p= 0.0053$, unpaired t-test with Welch's correction), (Figure 10).

Next, all four diagnosed groups were compared with each other. The richness number of species was significantly higher in DP than in Control or EAC, while the Simpson effective number of species did not show any significant difference between the phenotypes (Figure 11A, B).

Beta diversity statistical analysis revealed that samples from subjects with the same diagnosis cluster together, and that the microbial community composition differs significantly between the diagnosis groups. Moreover, it is observed that the microbial community composition shows high interindividual variation in biopsies of subjects with EAC.

The results of the pairwise beta diversity analysis between the diagnosis groups showed significant difference between the control group and all other disease phenotypes (Control vs. BE, Control vs. DP, Control vs. EAC) and between BE vs. EAC; no significant separation of microbial profiles was present between DP vs. BE or DP vs. EAC (Figure 12).

Phenotype	Number of samples collected	Number of samples sequenced
Control (n=14)	13	11
BE (n=15)	15	14
Dysplasia (n=4)	4	4
EAC (n=15)	15	15
Total	47	44

Table 8. Total number of tissue biopsy sets collected and sequenced.

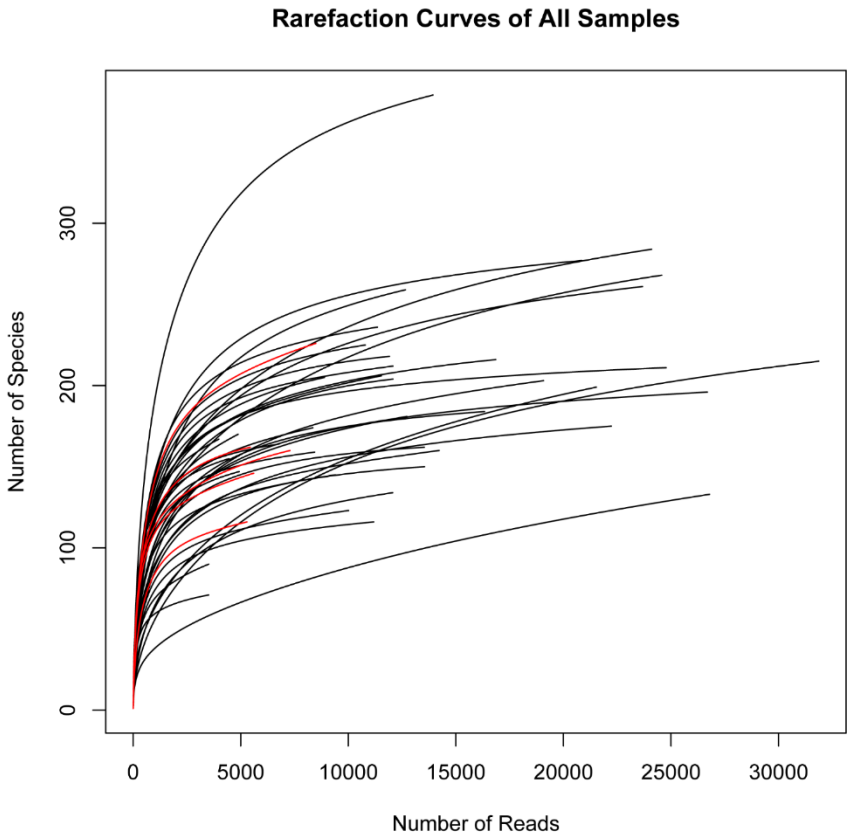


Figure 9. Rarefaction curves of biopsy samples sequenced.

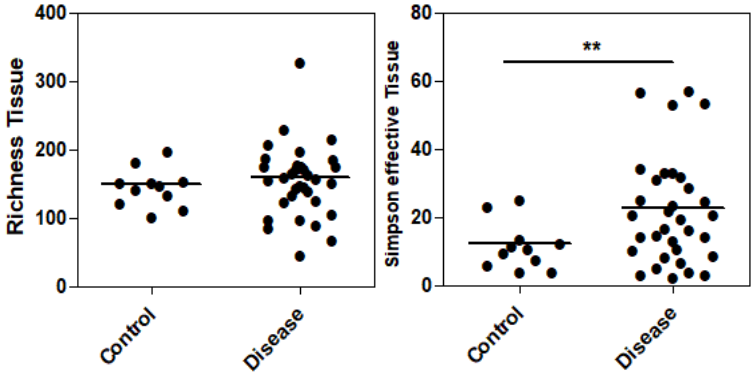


Figure 10. Alpha diversity in tissue biopsies of controls and patients with disease phenotype including BE, DP, and EAC.
(A) Alpha diversity assessed through species richness, showing no difference between health and disease. (B) Alpha diversity assessed through Simpson effective, showing a significant higher number of species in biopsies of patients with disease phenotype as in healthy controls.

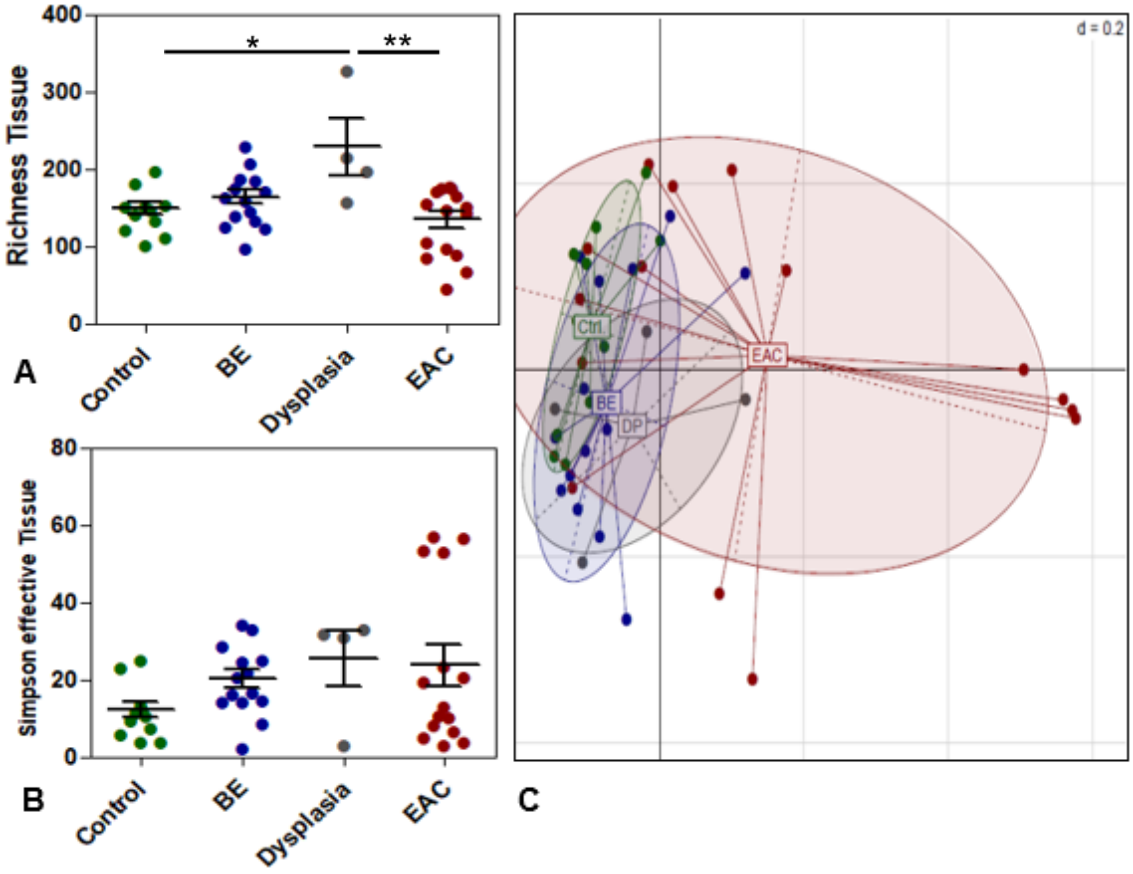


Figure 11. Alpha and beta diversity in tissue biopsies of patients diagnosed with BE, DP, EAC, and controls.
(A) Alpha diversity assessed through species richness, showing no significant difference between the diagnosis groups. (B) Alpha diversity assessed by Simpson effective number of species, showing no significant difference between the diagnosis groups. (C) Beta diversity illustrated by a multidimensional scaling plot (MDS plot) of microbial profiles obtained from tissue

biopsies of patients diagnosed with BE, DP and EAC, and from controls, showing a significant difference in the overall microbiome composition between the diagnosis groups (p=0.001).

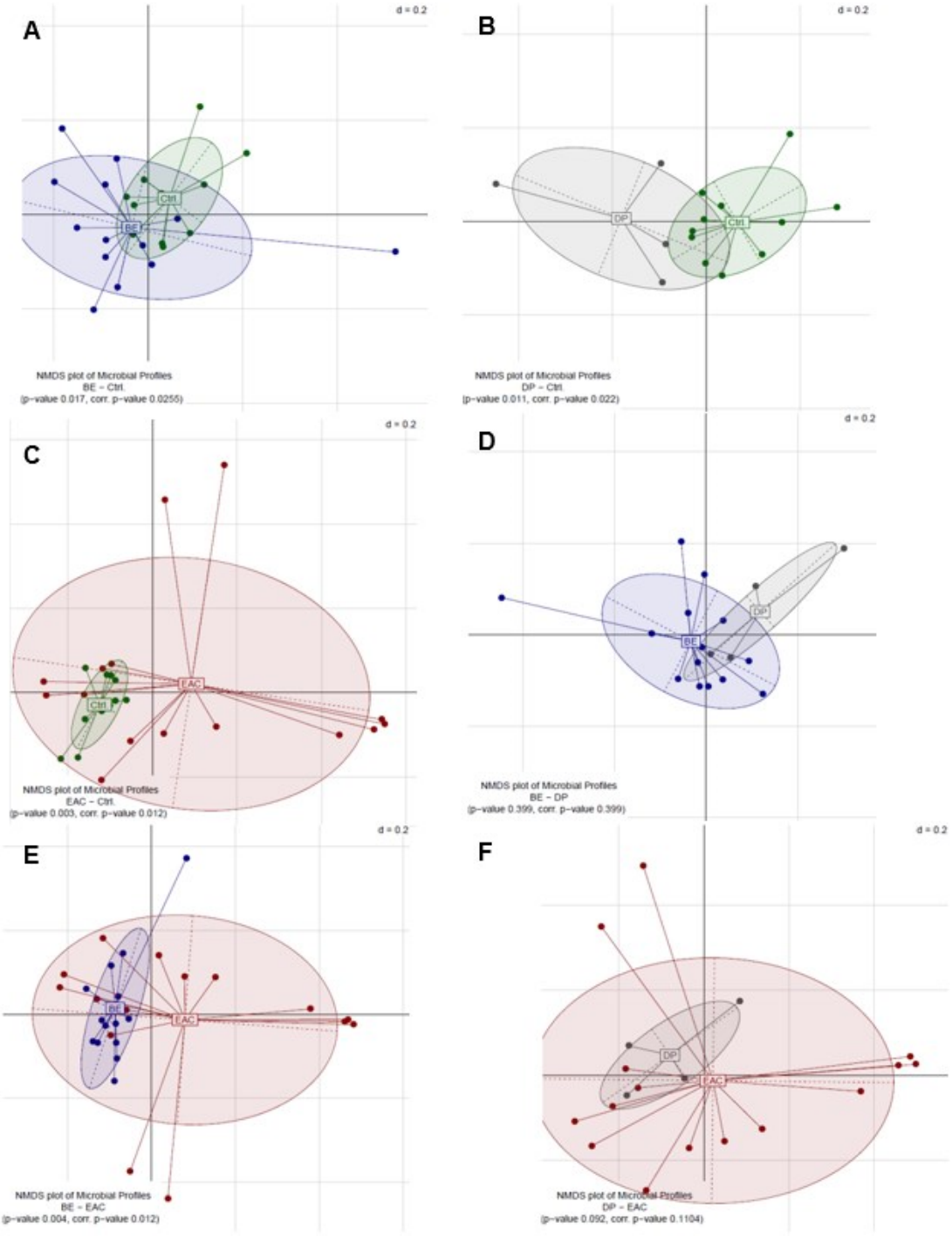


Figure 12. Pairwise comparison of microbial profiles in biopsies of different diagnosis groups. Beta diversity, illustrated by a non-metric multidimensional scaling plot, comparing the microbial profiles of patients with different diagnosis pairwise, showing significant differences between Ctrl. vs. BE (p=0.017), Ctrl. vs DP (p=0.011), Ctrl. vs. EAC (p=0.003), and BE vs. EAC (p=0.012).

In EAC, Firmicutes was more abundant than in BE, while Actinobacteria was less abundant than in BE

In addition to significant differences between the diagnosis groups in terms of overall microbial diversity and composition, single taxa were differently abundant among groups at phylum, genus, and OTUs level.

Four phyla—Actinobacteria, Firmicutes, Tenericutes, and Patescibacteria—showed significant differences between diagnosis groups (Figure 13). The relative abundance of Actinobacteria was significantly lower in EAC compared to all other groups (Ctrl., BE, DP), while the abundance of Firmicutes was higher in EAC than in BE. Tenericutes was only present in EAC, while Patescibacteria was less present in EAC than in BE and dysplasia (Figure 13).

At genus level, Bradyrhizobium and Streptococcus were significantly underabundant in disease phenotypes compared to controls (Figure 13). While Bradyrhizobium was underabundant in all disease phenotypes (BE, DP, EAC), Streptococcus was only underabundant in BE and EAC compared to controls. Furthermore, Actinomyces and Granulicatella were underabundant in EAC compared to controls. Granulicatella abundance was also significantly lower in EAC compared to BE (Figure 13). At genus level, Fusobacterium and Ralstonia were significantly enriched in disease phenotypes compared to controls. While Fusobacterium was enriched in all disease phenotypes (BE, DP, EAC), Ralstonia was only enriched in BE and EAC (Figure 13). Additionally, an overabundance of Sphingomonas and Porphyromonas in BE compared to controls was observed (Figure 13).

Several taxa were exclusively present in specific diagnosis groups

Figure 14 shows the taxa that were present exclusively in certain diagnosis groups. Some taxa were almost only present in samples of EAC patients and included Anaeroplasma (6/15 EAC patients); ASF 356 (5/15 EAC patients); Rikenellaceae RC9 gut group (5/15 EAC patients); Clostridiales vadin BB60 group (6/15 EAC patients); Rominiclostridium (5/15 EAC patients); and Odoribacter (6/15 EAC patients and 1/11 controls). The genera of Rothia and Haemophilus were less prevalent in EAC than in other phenotypes.

The genera of Bifidobacterium, Campylobacter, Unknown Lachnospiracae, and one uncultured unknown genus were, with few exceptions, only found in disease phenotypes.

The genera of one unknown Saccharimonadaceae and Atoponium were, with few exceptions, only found in BE.

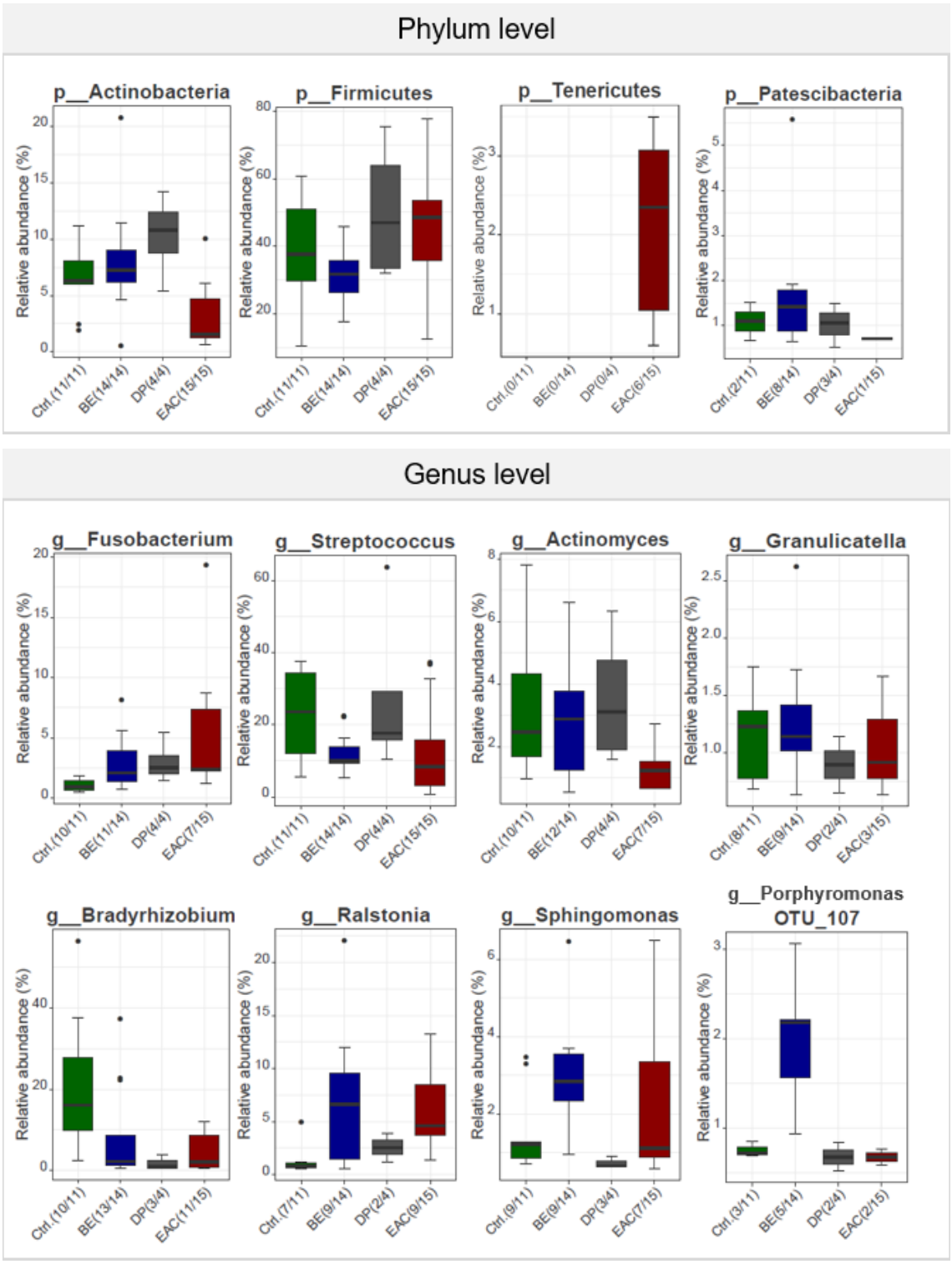


Figure 13. Taxa present in varying abundance in tissue biopsies of different diagnosis groups.

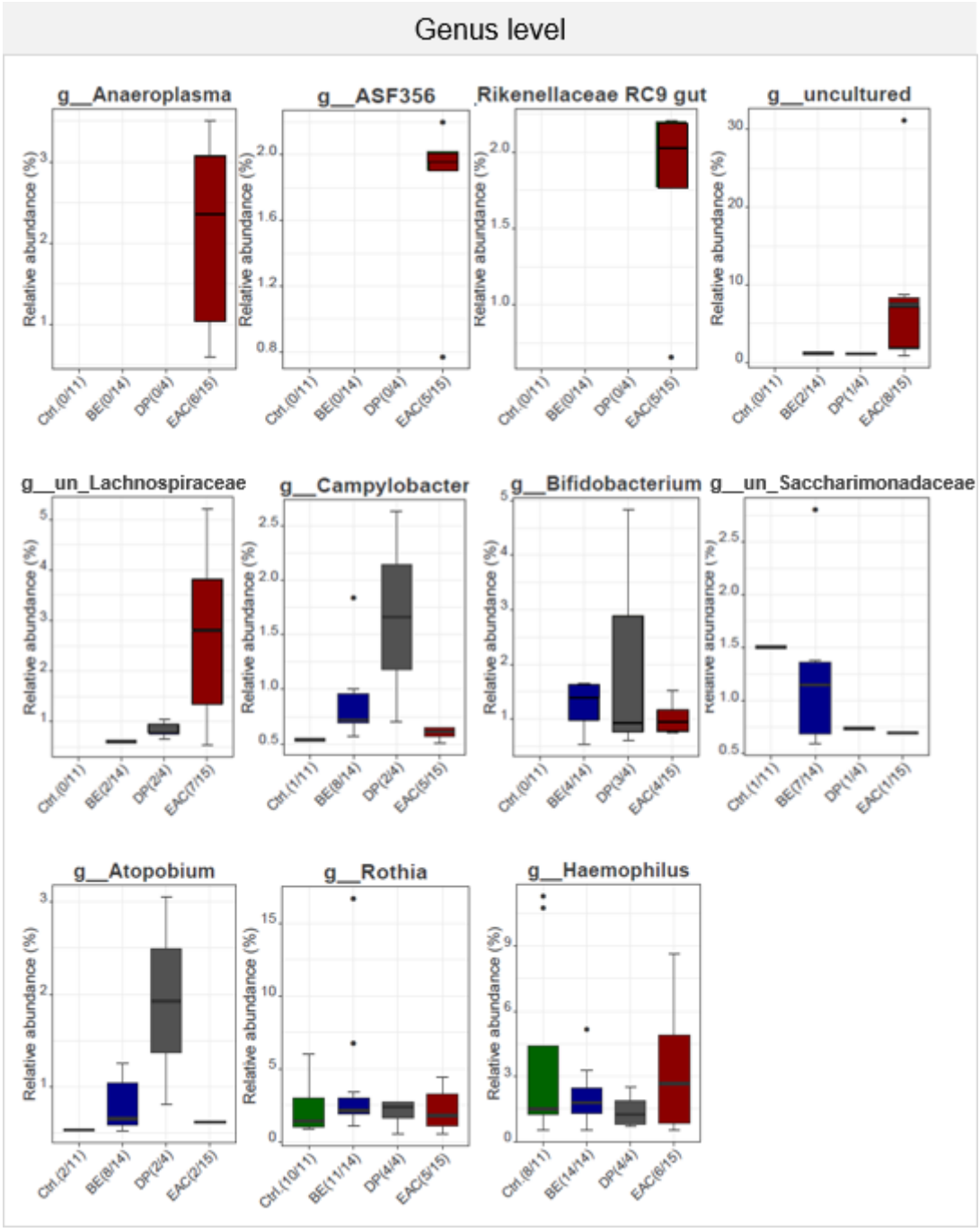


Figure 14. Taxa that occur exclusively in biopsies of certain groups.

Saliva

Salivary microbial community diversity and composition did not differ between the diagnosis groups

Thirty-eight of 45 saliva samples collected from the patients of this cohort were sequenced and analyzed (Table 9).

The quality control demonstrated that all saliva samples were sufficiently sequenced, and thus all sequenced samples were included in the downstream microbiome analysis. For the quality control of the sequencing, rarefication curves of all samples were provided (Figure 15) and they show a plateau of number of species in each sample after all available reads were assigned (Figure 15).

The alpha diversity was assessed by different parameters, including Richness and Simpson effective, which are visualized in Figure 16A, B. There was no significant difference in either Richness or Simpson effective number of species between the four diagnosis groups. A high interindividual variation regarding the number of species could be determined in saliva samples, varying from 36 to over 150 OTUs identified per sample.

Statistical analyses of the beta diversity in saliva did not find any significant difference between the diagnosis groups (Figure 16.C). A high overlap of the microbial profiles of patients belonging to different diagnosis groups shows that saliva microbiome diversity cannot separate between the phenotypes.

Streptococcus and Granulicatella were differently abundant in different diagnosis groups

Although there was no significant difference in overall diversity and composition at the phylum and genus levels, salivary microbial community composition showed alterations between phenotypes in the abundance of Streptococcus and Granulicatella at OTU level (Figure 17). The relative abundance of some species belonging to the Streptococcus genus differed significantly among groups, with a downward tendency in EAC. Additionally, a continuous fall of the relative abundances of species belonging to the Granulicatella genus from controls to esophageal adenocarcinoma was determined.

Phenotype	Number of samples collected	Number of samples sequenced
Control (n=14)	12	10
BE (n=15)	15	14
Dysplasia (n=4)	4	3
EAC (n=15)	14	11
Total	45	38

Table 9. Total number of saliva samples collected and sequenced.

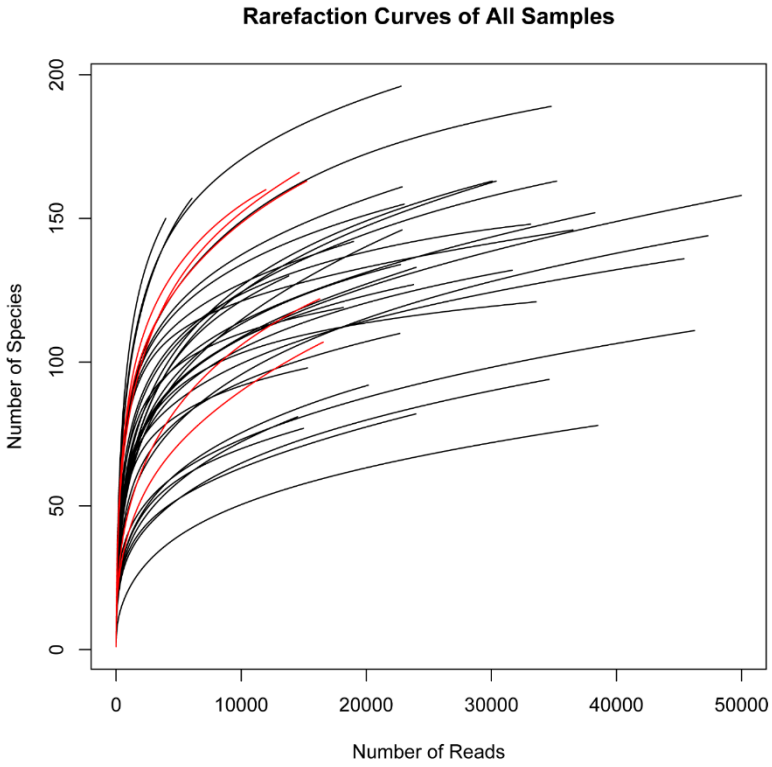


Figure 15. Rarefaction curves of saliva samples sequenced.

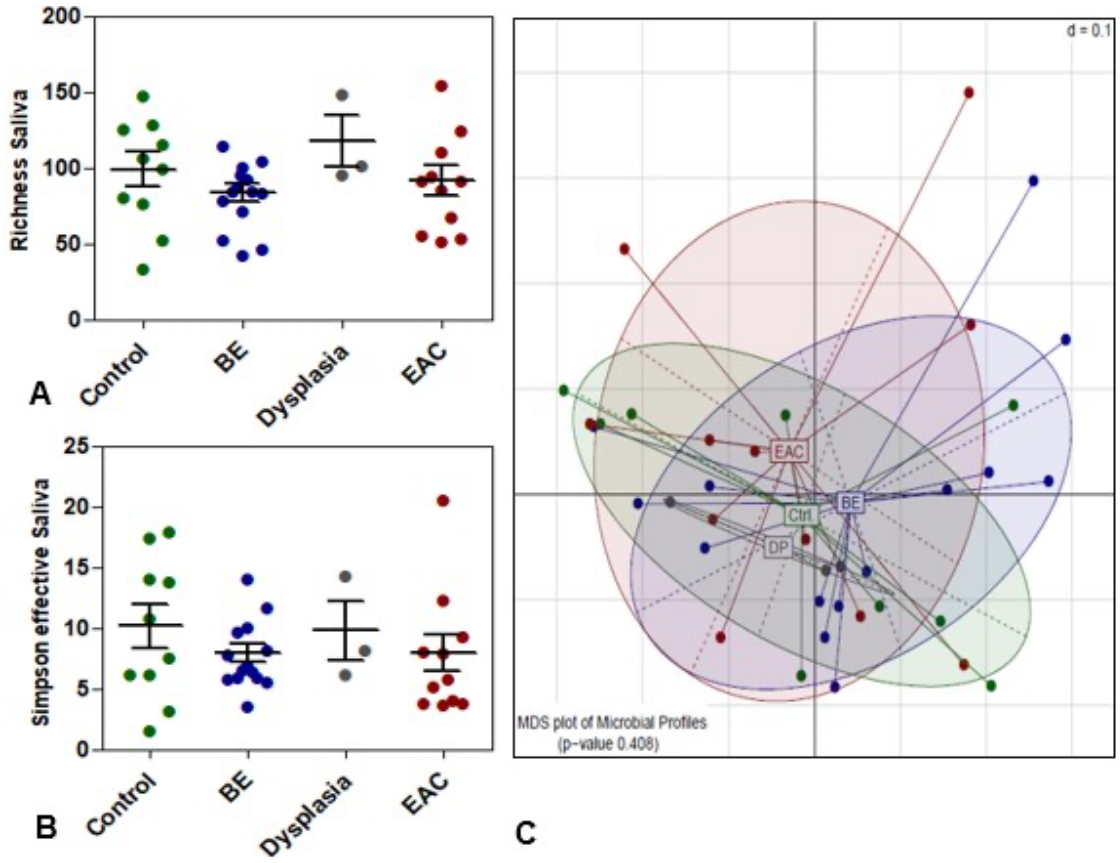
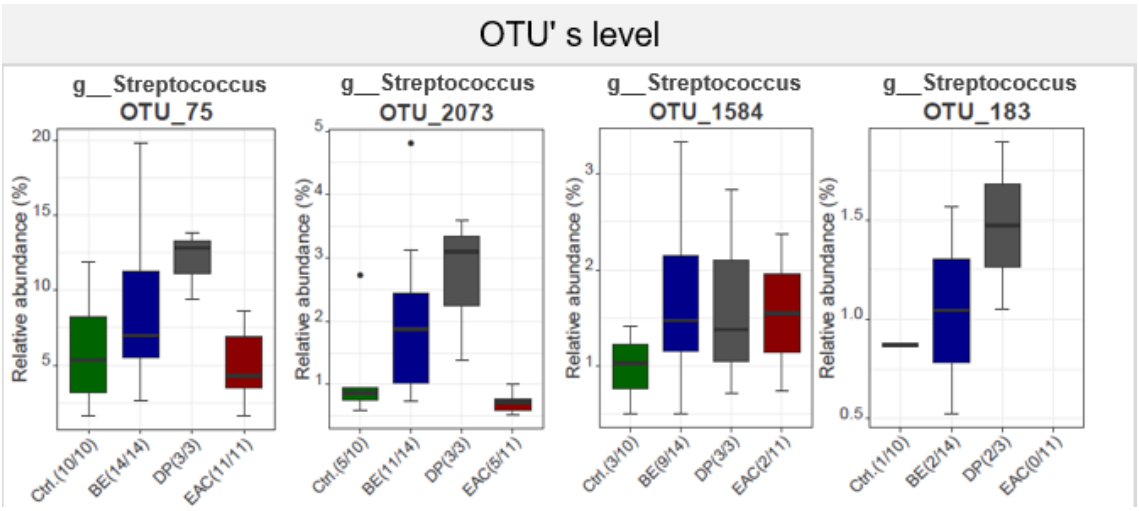


Figure 16. Alpha and beta diversity of microbiota in saliva samples of patients diagnosed with BE, DP, and EAC, and controls.
(A) Alpha diversity assessed through species richness, showing no significant difference between the diagnosis groups. (B) Alpha diversity assessed by Simpson effective number of species, showing no significant difference between the diagnosis groups. (C) Beta diversity illustrated by a multidimensional scaling plot (MDS plot) of microbial profiles obtained in saliva samples of patients diagnosed with BE, DP, and EAC, and controls, showing no significant difference in the overall microbiome composition between the diagnosis groups (p=0.408).



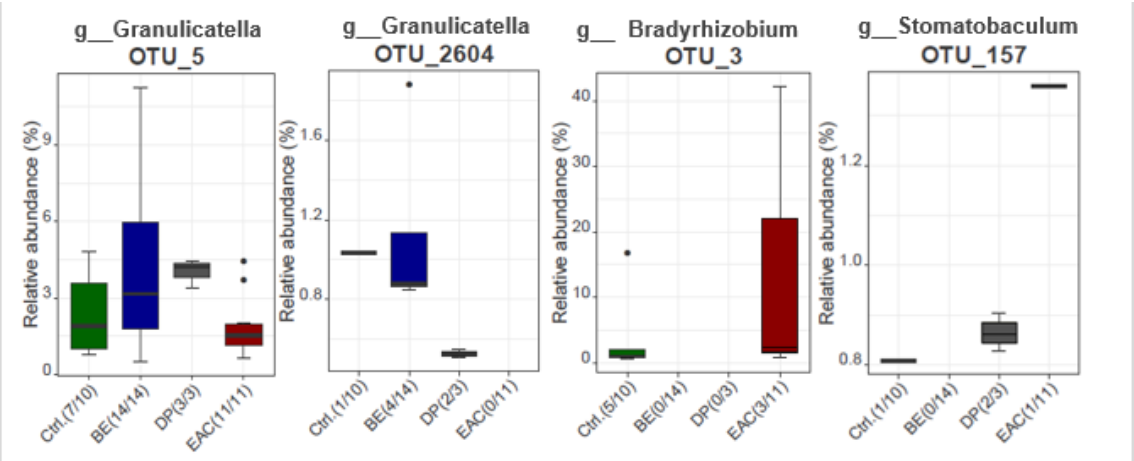


Figure 17. OTUs present in varying abundance in saliva samples of different diagnosis groups.

Feces

Fecal microbiota diversity and composition did not differ between the diagnosis groups

Forty-two of 45 fecal samples collected were sequenced (Table 10).

The quality control demonstrated that one out of 42 fecal samples was not sufficiently sequenced (Figure 18). This insufficiently sequenced sample belonged to a patient of the EAC diagnosis group and was excluded from downstream analysis.

The alpha diversity was assessed by different parameters, including Richness and Simpson effective. There was no significant difference in either Richness or Simpson effective number of species between the four diagnosis groups (Figure 19A, B). In contrast to other sample types (tissue, saliva), the fecal samples presented only a low interindividual variation with regard to the number of OTUs identified per sample.

Statistical analysis of beta diversity in fecal samples showed no significant separation between the diagnosis groups, with a high overlap of the samples from different diagnosis groups (Figure 19C).

Betaproteobacteriales was enriched in DP and EAC in comparison to BE, while Eubacterium coprostanoligenes was depleted in DP and EAC in comparison to BE

In the different diagnosis groups, several taxa were presented in different abundances (Figure 20). An overabundance of the Coriobacteriia class was revealed in EAC in comparison to controls, as well as an overabundance of the Burkholderiaceae family and the Betaproteobacteriales order in DP and EAC in comparison to BE.

At genus level, Faecalibacterium was significantly underabundant in BE and EAC compared to controls, but overabundant in DP compared to BE and EAC. Phascolarctobacterium was underabundant in EAC compared to controls, and Eubacterium coprostanoligenes was underabundant in DP and EAC in comparison to BE.

Several taxa were exclusively present in specific diagnosis groups

Several taxa were differently represented between groups (Figure 21). Acidaminococcus and Sutterella were more commonly represented in disease phenotypes than in controls. The genera of Parasaturella and Rominococcaceae UCG-005, as well as two OTUs that contained species of Romincoccaceae UCG-002 and Bacteroides genera, were more abundant in EAC compared to BE, while the genus of Lachnospira was less abundant in EAC compared to controls. Holdemanella was found in seven out of 15 BE subjects and in one out of 11 controls identified, while in dysplasia and EAC this genus was not found.

Phenotype	Number of samples collected	Number of samples sequenced
Control (n=14)	11	9
BE (n=15)	15	15
Dysplasia (n=4)	4	4
EAC (n=15)	15	14
Total	45	42

Table 10. Total number of fecal samples collected and sequenced.

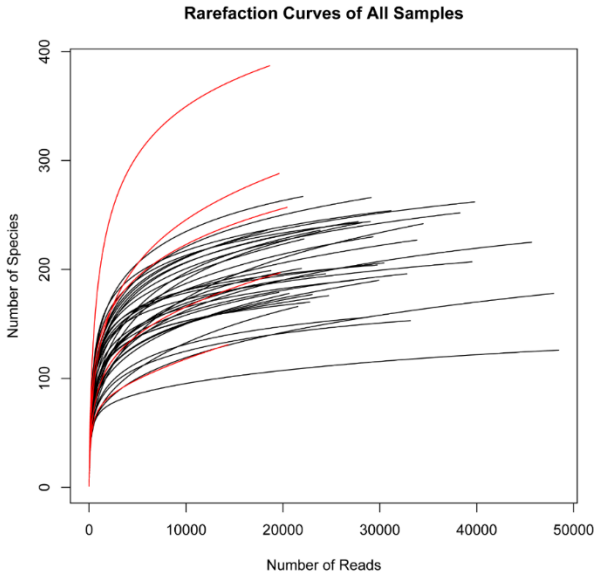


Figure 18. Rarefaction curves of fecal samples sequenced.

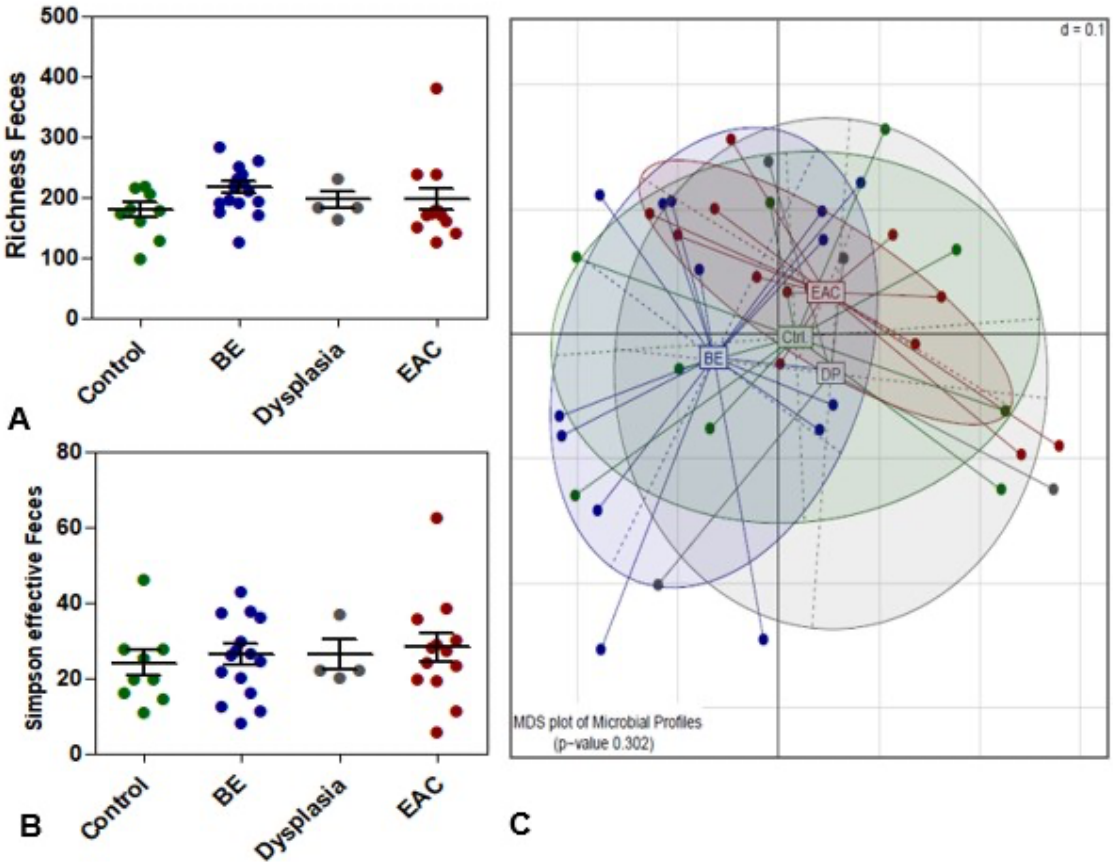


Figure 19. Alpha and beta diversity of microbiota in fecal samples of patients diagnosed with BE, DP, and EAC, and controls.
(A) Alpha diversity assessed through species richness, showing no significant difference between the diagnosis groups. (B) Alpha diversity assessed by Simpson effective number of species, showing no significant difference between the diagnosis groups. (C) Beta diversity illus-

trated by a multidimensional scaling plot (MDS plot) of microbial profiles obtained in fecal samples of patients diagnosed with BE, DP, and EAC, and controls, showing no significant difference in the overall microbiome composition between the diagnosis groups (p=0.302).

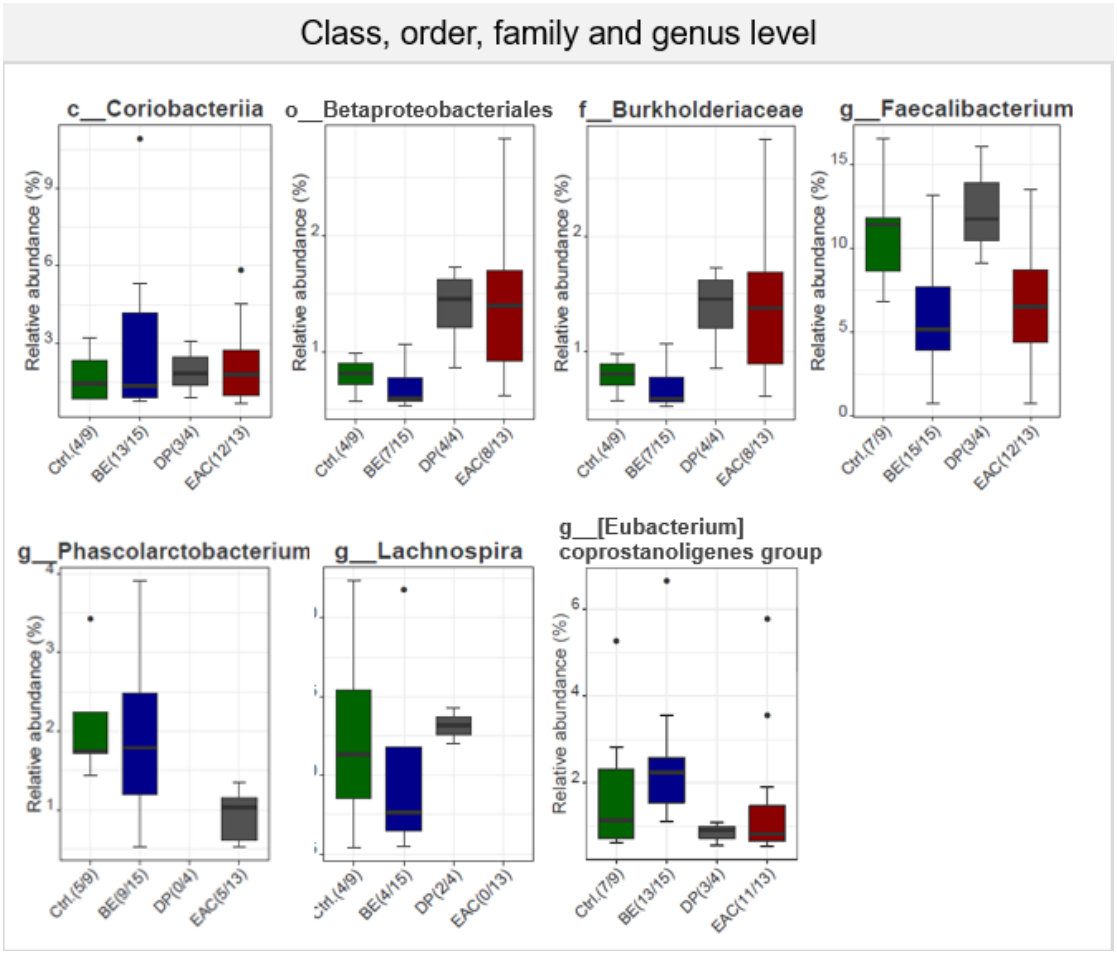
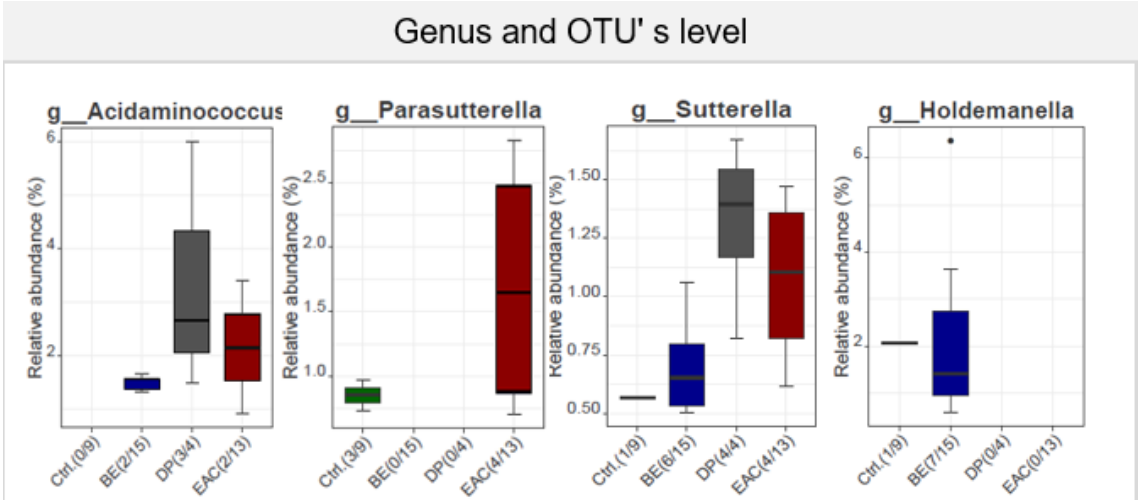


Figure 20. Taxa present in varying abundance in fecal samples of different diagnosis groups.



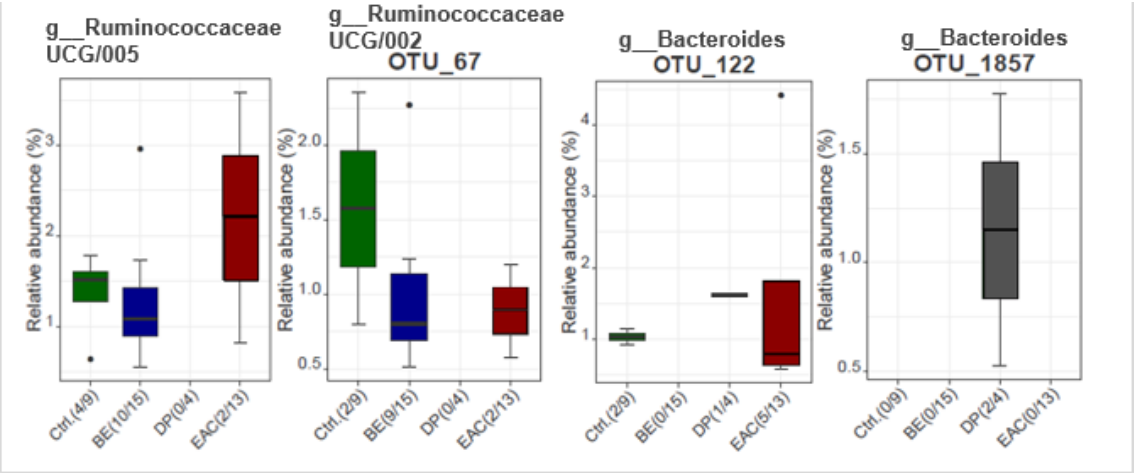


Figure 21. Taxa that occur exclusively in fecal samples of certain groups.

3.3.3 Diagnosis- and risk factor-guided analyses

Smoking and gastroesophageal reflux were related to microbiota alternations

One of the aims of this study was to investigate whether the most-known risk factors related to esophageal adenocarcinoma are associated with specific microbiota alternations. In order to achieve this, the patients' epidemiological data were merged with the microbiome sequencing data.

First, we analyzed the microbiome of smoker vs. never smoker and patients with reflux symptoms vs. without reflux symptoms independent of the diagnosis running alpha and beta diversity analysis. A significant difference was observed between the microbial composition of patients with reflux symptoms vs. those without reflux symptoms in saliva material (beta diversity, p=0.022), while the stratification based on smoking did not show any significant result (Figure 22).

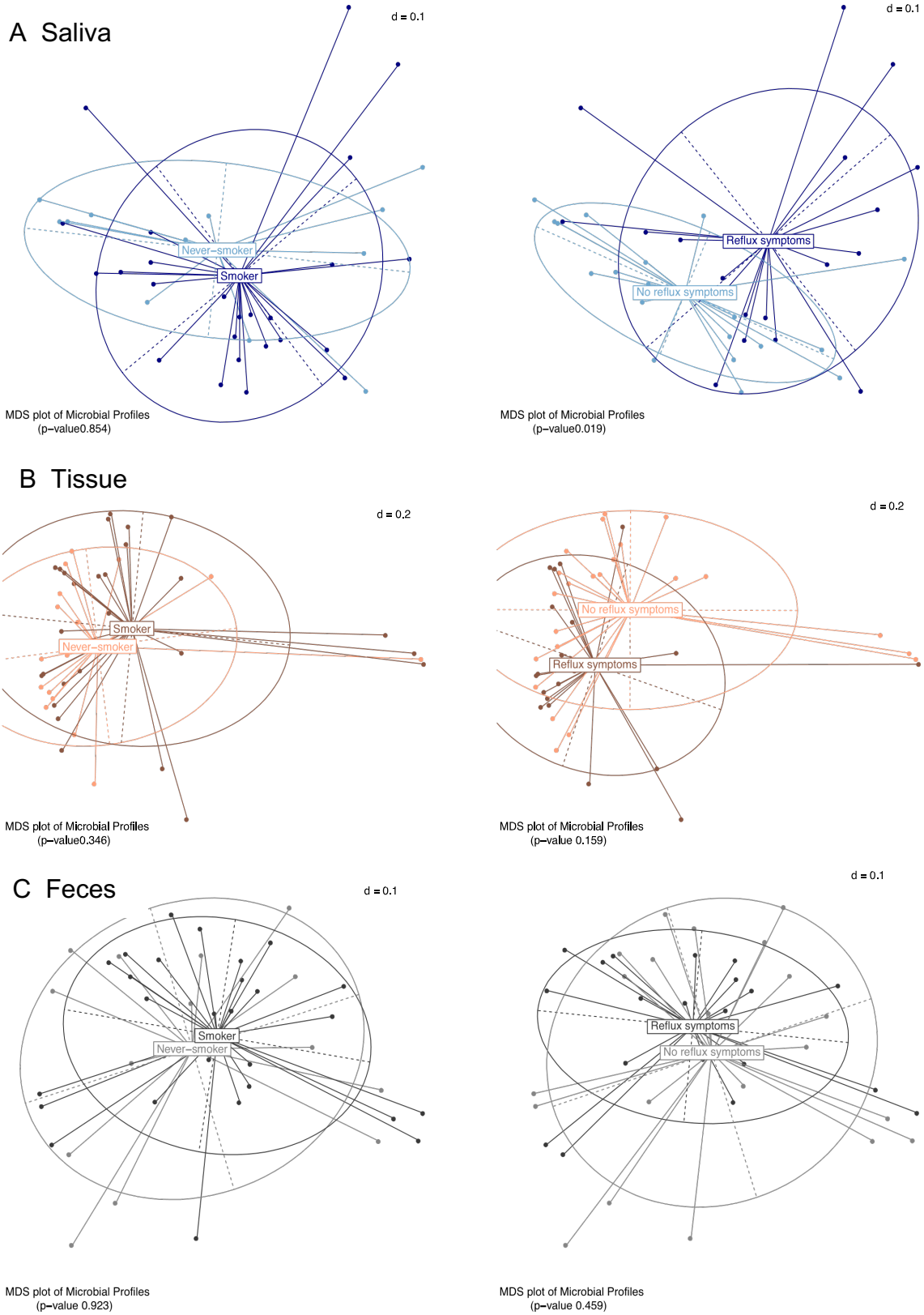


Figure 22. The presence of reflux symptoms correlates with microbiome diversity in saliva. (A) Beta diversity in saliva samples of patients divided by smoking history or alcohol consumption or GERD. (B) Beta diversity in tissue samples of patients divided by smoking history or alcohol consumption of GERD. (C) Beta diversity in fecal samples of patients divided by smoking history or alcohol consumption of GERD.

When stratifying subjects based on their smoking history/reflux symptoms and clinical diagnosis simultaneously, we observed significant differential microbial composition not only between different diagnosis groups, but also within diagnosis group at the BE stage of disease: BE subjects with reflux symptoms had different microbial profiles compared to BE subjects without reflux symptoms in saliva; BE smoker had different microbial profiles compared to BE never-smoker in tissue biopsies; and BE subjects with reflux symptoms vs. BE subjects without reflux symptoms, as well as BE smokers vs. BE never smokers showed different microbial profiles in fecal samples, proposing an impact of these risk factors in the microbiome composition at the BE disease stage (Figure 23, 24).

We run group comparison analysis at taxon/OTU level between these 8 subgroups and found here several differentially abundant bacteria; however, none of them was consistently associated with smoking or reflux. We revealed (i) a trend towards a decreased abundance of Actinobacteria in smokers (BE smoker vs. BE never-smoker p= 0.0240) and in patients with progressed phenotype in tissue biopsies; (ii) a trend toward enrichment of one Streptococcus species in BE subjects with reflux symptoms (BE reflux symptoms vs. BE no reflux symptoms p=0.0200) but a loss in patients with progressed phenotype in saliva; (iii) a trend toward higher abundance of Faecalibacterium in smokers (BE smokers vs. BE never smokers p= 0.0264) but a decreased abundance in EAC in feces (Figure 25).

Next, we analyzed the relative abundances of taxa/OTUs exclusively in BE patients; the sub stratification was conducted as described above in BE smoker vs. BE never smoker; BE reflux symptoms vs. BE without reflux symptoms; all significant taxa of this analysis are presented in Supplementary Table 1. These taxa/OTUs were compared to the significant ones from the previously conducted diagnosis-guided analysis; a summary of altered bacteria in both Barrett's subjects with exposure to a risk factors and subjects with progressed phenotype are presented in Supplementary table 2. Here, we could confirm the results described in the upper paragraph: (i) in tissue biopsies the phylum of Actinobacteria was decreased among BE smokers, which was also the case in esophageal adenocarcinoma samples compared to BE and to other phenotypes; (ii) in saliva, one Streptococcus species was decreased in smokers, and in esophageal adenocarcinoma compared to other phenotypes; another Streptococcus species was increased in subjects with reflux symptoms but decreased in esophageal adenocarcinoma; (iii) in feces, Faecalibacterium was increased in BE smokers but decreased in EAC (Figure 26).

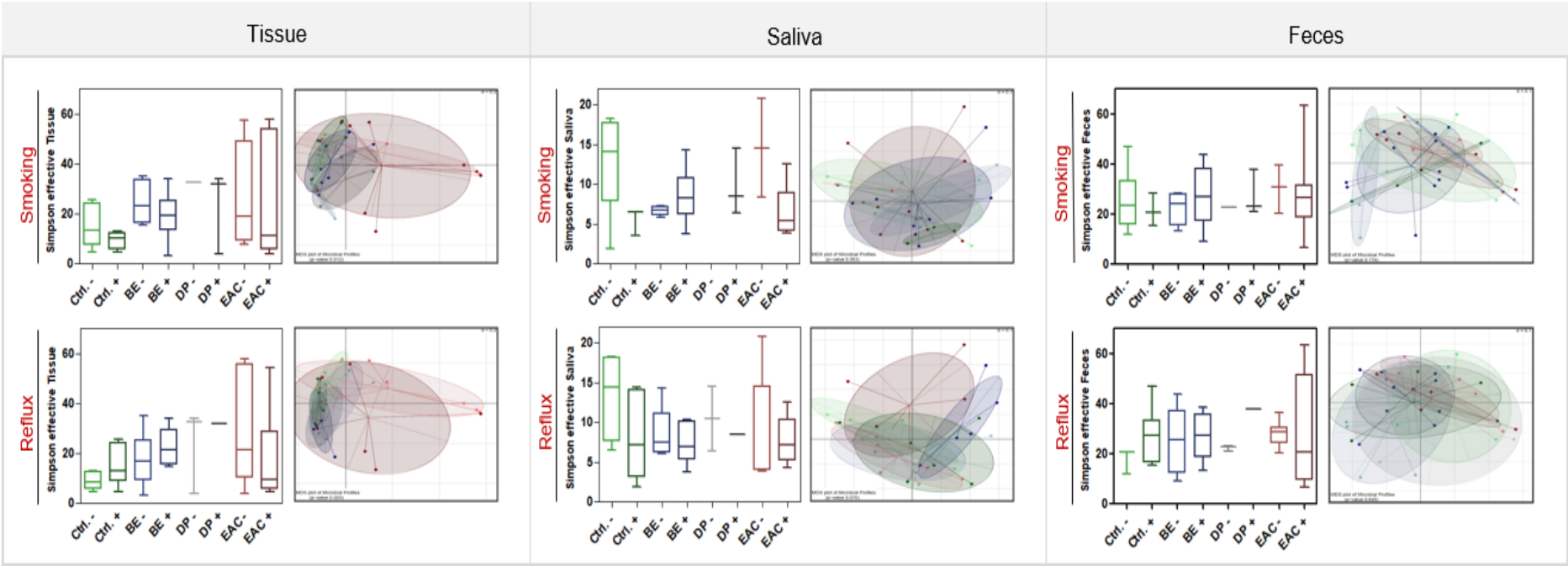


Figure 23. Alpha- and beta diversity of microbiota in samples sub-stratified by diagnosis and exposure to smoking or reflux symptoms. Alpha diversity assessed by Simpson effective showing a tendency toward higher interindividual variations between subjects with progressed phenotype. Beta diversity visualized by MDS plots of generalized Unifrac distances showing significant separation between subgroups with different diagnosis and between BE smokers and BE never-smokers in tissue; significant separation between BE with reflux symptoms vs. BE without reflux symptoms in saliva; as well as between BE smokers vs. BE never smokers, BE with reflux symptoms vs. BE without reflux symptoms in feces.

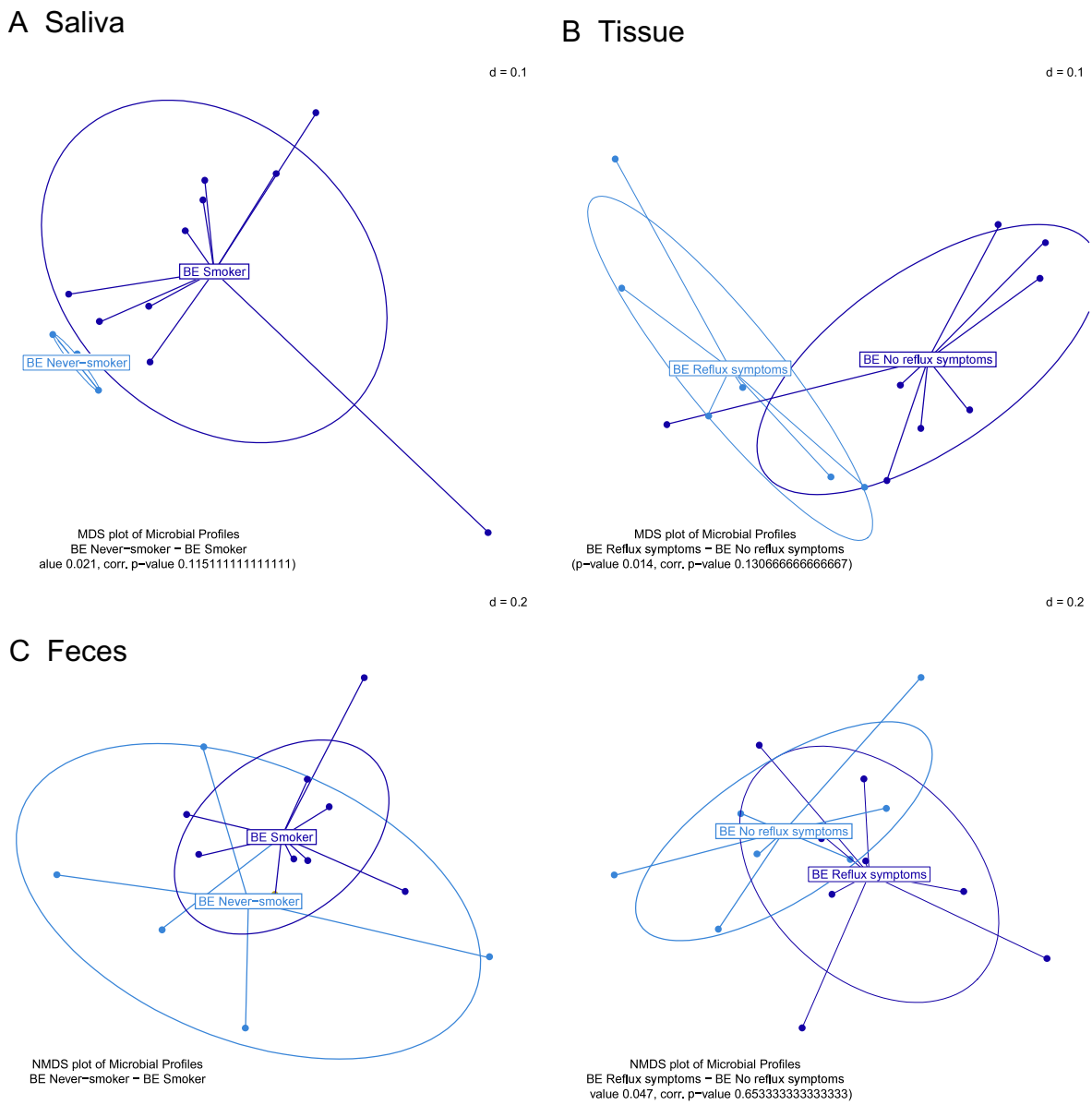


Figure 24. Risk factors for EAC such as smoking and GERD correlate with microbiome diversities in Barrett's Esophagus.
(A) Beta diversity in saliva samples of BE subjects with reflux symptoms vs. without reflux symptoms. (B) Beta diversity in tissue biopsies of BE smokers vs. BE never-smokers. (C) Beta diversity in fecal samples of BE smokers vs. BE never-smokers and BE subjects with reflux symptoms vs. without reflux symptoms.

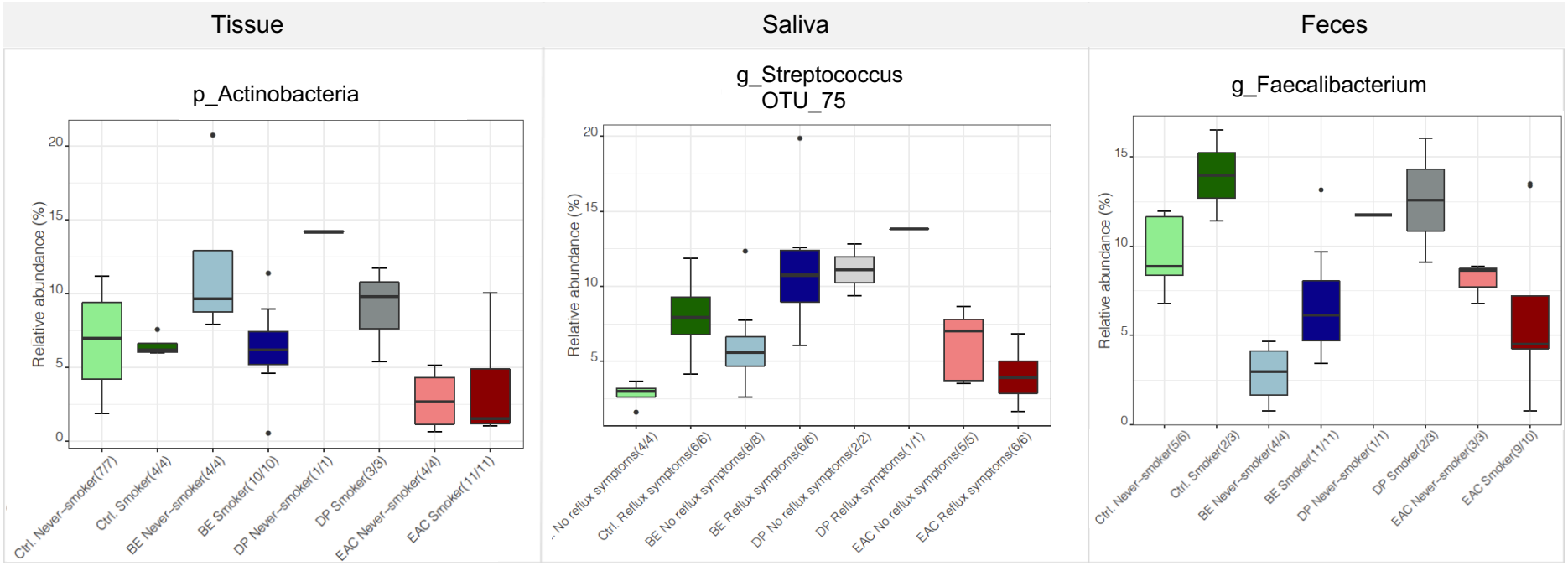


Figure 25. Single Taxa and OTUs are presented in same or reverse abundance tendency in subjects with exposure to a risk factor and in subjects with advanced phenotype.

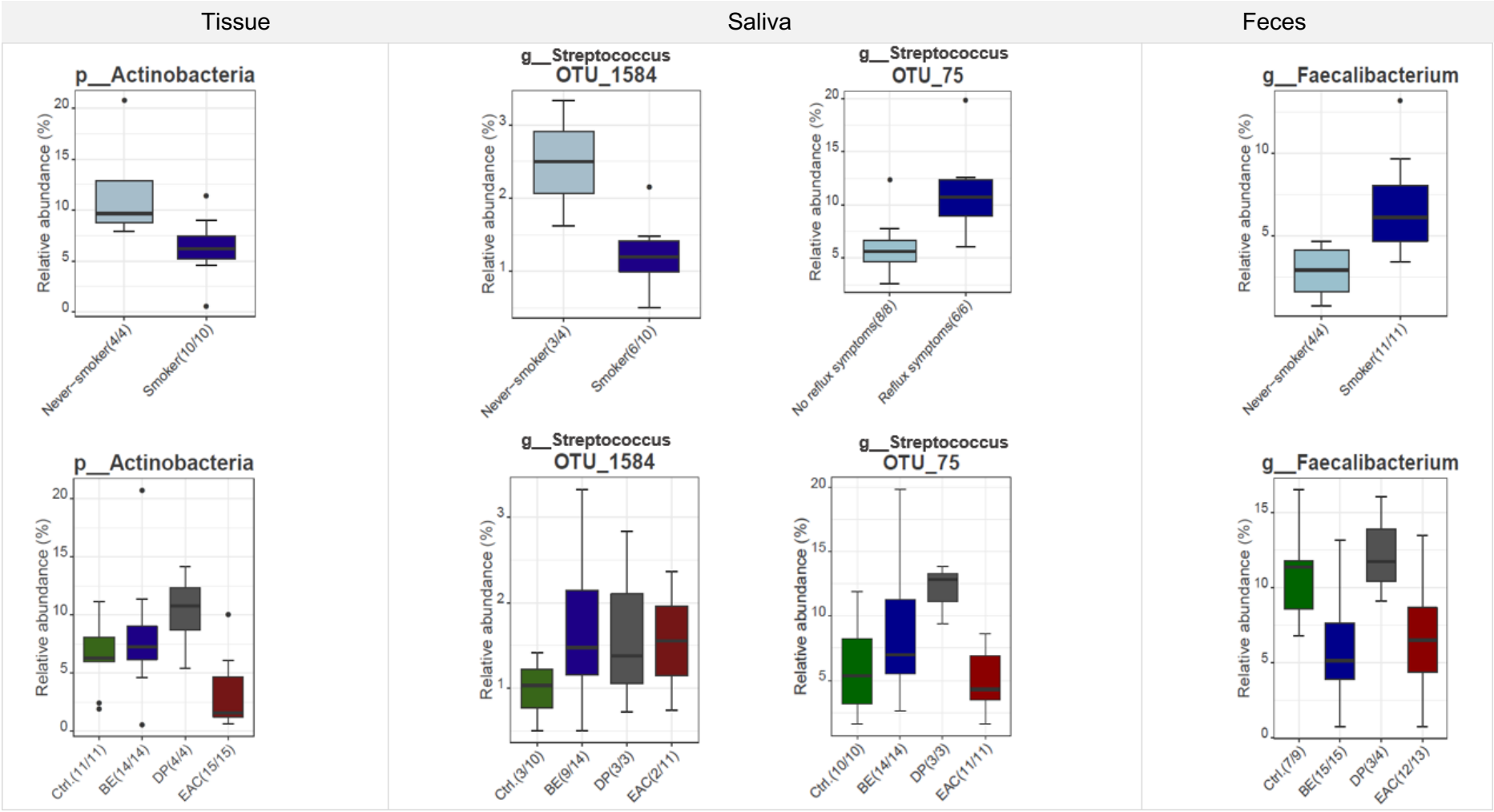


Figure 26. Single Taxa and OTUs presented in same or reverse abundance tendency in BE subjects with exposure to a risk factor and in subjects with progressed phenotype.

4 Discussion

4.1 A study with standardized methodical instruments made investigating associations between microbiota, disease and risk factors feasible

In this project we established the methodology of profiling the human esophageal and cardia mucosa microbiome from biopsy material, salivary and fecal microbiome in a cohort of 48 patients. We conducted a case-control study on cases diagnosed with diseases of the esophageal adenocarcinoma cascade, including BE, DP and EAC, and controls with no history of BE associated pathologies using 16S rRNA amplicon sequencing. The microbiome characterization was achieved by stratifying the samples according to sample type, the patient's clinical diagnosis, and the patient's exposure to well-known risk factors for the development of esophageal adenocarcinoma. We found clear associations between the microbiota and all metavariables analyzed, which are discussed in the following sections. The pipeline established within the scope of this work, is currently a firm part of the BarrettNET Study and continues collecting samples for further investigation of the microbiome in a larger cohort.

Active participation and cooperation were notable in this cohort, reflected in the broad spectrum of information and samples provided. Reasons for participation based on patients' comments included interest in the study purpose, altruism, and the recruiter's approach. Only a small percentage declined to participate, mostly due to patients' insecurities in terms of higher risk associated with additional biopsy sampling for study purposes or due to patients' subjective assessment of their own poor general condition.

Aside from the esophageal tissue-associated microbiome, the present study represents one of the first efforts to discover the salivary and fecal microbiome of patients with a pathology of the distal esophagus. The saliva and fecal microbiome have potential etiological implementations in disease initiation and progression, and knowledge about the microbial population in these materials at different disease stages could generate novel hypothesis and elucidate further aspects of EAC pathogenesis important for prevention and risk stratification. The identification of shifts in salivary or fecal microbiota among the disease cascade would open new avenues for non-invasive risk assessment among BE patients at risk of progression, and thus present one of the few opportunities to screen for EAC progression without the effort of endoscopy.

Beside this, patients with low-/high-grade dysplasia and esophageal adenocarcinoma at different stages of tumorigenesis and different tumor grades were included in the study. Whereas the microbial community composition of normal esophagus, GERD and BE patients has been widely investigated, until recently there was a gap with regard to microbiome composition among patients with enhanced phenotypes, which this study aimed to address.

Another strength of this study was the exclusion of patients with antibiotic use in the last 6 weeks, diarrhea, and vegetarian diets, all of which might influence the microbial diversity and composition [52] [11, 124, 144].

The study also captures comprehensive patient epidemiologic data, providing an accurate overview of distribution of possible risk factors for disease development among patients with different clinical diagnosis. The patient-related epidemiological information enabled us to examine the possible effects of known risk factors for EAC development in microbial population diversity and composition: firstly, between Barrett's subjects with different exposure to reflux and smoking, and finally between all cohort subjects divided by diagnosis and exposure to these two risk factors.

The standardized sampling and preanalytical phase, the diagnosis confirmation by certified gastrointestinal pathologists, and the DNA extraction procedure performed by following well-established protocols by a trained scientific staff member contributed to the quality assurance of this study and thus presents another strength. We used the PAXgene tissue system, a formalin free preparation system, for biopsy tissue fixation and stabilization. The advantages of using this fixation and stabilization system is the preservation of both morphology, antigenicity and other biomolecule [99]. It enables histopathological and biomolecular analysis to be conducted from a single tissue-biopsy sample, which allowed us to perform both histological tissue evaluation and DNA extraction for microbiome characterization from each single biopsy sample. PAXgene-fixed, paraffin-embedded colon cancer tissue showed comparable histopathological morphology results to formalin-fixed, paraffin-embedded mirrored tissue in a blinded randomized trial [42]. In particular, the nuclear structures could be better recognized from PFPE tissues [42]. Both quality and yield of PAXgene fixed tissue RNA were comparable with those of the snap frozen tissue RNA – a marked improvement over the formalin fixed tissue [135]. Snap-freezing in liquid nitrogen, which is the most popular method used for molecular analytical purposes to date, may damage tissue morphology and antigenicity. It requires a high degree of methodical, logistical effort and is cost-intensive, making its

integration into the clinical routine difficult. DNA extraction was conducted using a different extensive nucleic acid extraction protocol for each sample type, all of which showed sufficient DNA quality with the exception of one fecal sample, which was excluded for downstream analysis due to low sequencing depth.

Other strengths of this study are the standardized sequence data processing using the Integrated Microbial Next Generation Sequencing platform, followed by comparison of the three largest and most known taxonomy databases, such as the latest version of RDP classifier, SILVA and EzTaxon databases to improve the taxonomic assignment of OTUs. In addition, the utilization of the Rhea pipeline offered a systematic testing of all OTUs and taxa, providing a range of parameters for accessing the microbiome and applying suitable statistical tests for microbiome data [66].

This study has also some limitations. Among them, is the small cohort size; in particular the low number of patients with LG-/HG-IEN, which can be explained by the rare diagnosis of this intermediate disease stage before neoplastic progression to EAC. These limitations might reduce the power of the statistical tests and restrict the generalizability of the results. To prevent this, larger and ethnically different cohorts are necessary.

A further study limitation is the notably large number of controls (80%) diagnosed endoscopically and histologically with chronic gastritis. Gastritis may impact on microbiome diversity and composition of the esophagus and thus present a confound. Gastritis induced by Helicobacter pylori (H.pylori) infection leads, after chronification, to atrophy of the gastric mucosa, thus reducing gastric acid production (Hypochlorhydrie). This might be relevant in cases with reflux symptoms since H.pylori gastritis may alter the composition and pH of the refluxing gastric fluid, altering therefore the microbial community inhabiting the esophagogastric junction; this was also suggested by Pei et al. [95]. H. pylori is known to influence the gastric microbiome [85]. Some authors attribute a crucial role to H pylori in the composition of esophageal microbial community. Additionally, the fecal microbiome of H. pylori positive subjects has been found to be altered relative to H. pylori negative subjects and showed a higher microbiota diversity relative to H. pylori negative subjects [37]. Other studies, however, did not find any alteration in diversity from fecal samples in connection with H. pylori colonization. In our study, the most subjects diagnosed with gastritis, who also complained about reflux symptoms, did not manifest H. pylori colonization. Two control subjects were diagnosed simultaneously with chronic gastritis and reflux esophagitis. However, in neither subjects' gastric biopsies was Helicobacter pylori found. To summarize the pathology findings, the number of control subjects with H. pylori turned out to be equal to the number of these without H. pylori

in gastric biopsies. Additionally, 5/9 controls had slight corpus antrum gastritis and 4/9 controls had severe corpus antrum gastritis, and none of the controls manifested atrophic gastritis. Since the distribution of H. p infection in the control cohort is balanced, the majority of controls with reflux had no H. p associated gastritis and none of controls showed atrophic gastritis, it suggests that no great shifts are to be expected in esophageal microbiota diversity and composition. The fact that approximately the half of the control subjects of this study had an H. pylori infection is consistent with the finding that 50% of the population worldwide are also known to have a H. pylori colonization [131].

Control subjects were patients of the gastroenterology department of Klinikum rechts der Isar, undergoing a scheduled upper endoscopy examination for clinical indications, such as upper abdominal pain or heartburn symptoms, and had no history of BE associated lesions or other tumor diseases. Difficulties in including healthy controls in the study were encountered due to the invasiveness of the examination and strict indication for an upper endoscopy in Germany. The esophagogastroduodenoscopy (EGD) is not part of routine check-up examinations, unlike a precautionary colonoscopy, which is in Germany intended for all patients over the age of 55. Hence, only patients with prolonged or severe symptoms receive an upper endoscopy.

Further, using the 16S rRNA amplicon sequencing method to access the microbiome is a common and cost-effective method with very high throughput, which has some disadvantages compared to whole genome sequencing (WGS); among other issues, the lower taxonomic resolution makes it impossible to distinguish between taxa with a very similar sequence [107]. In this study, however, we aimed primarily to reveal whether dysbiosis occurs at the tumor niche, as well as in saliva and fecal samples. We therefore utilized 16S amplicon sequencing. We aimed to provide a comprehensive description of the alterations of the microbial population, without intending to define the one species responsible for carcinogenesis. This will require larger cohorts and longitudinally designed studies. Facts to date, even for the H. pylori case, suggest that it is the dysbiosis which contributes to cancer development rather than a single species.

We chose a case-control study design, an observational study design which compares cases to controls with regard to the exposure to one or more risk factors and describes the associations observed between potential risk factors and disease. Here, we observed (i) the association of the microbial community to disease, (ii) the association of epidemiologic factors to disease, as well as (iii) the association of risk factors to the microbial community at different stages of disease. A case-control study might present problems

of validity, typically caused by recall bias, interviewer bias and precision of the information required [128]. Patients might be influenced by the disease, their general condition, the interviewer, unclarities in the survey or other factors to not provide accurate information [128]. Here, we used the same standardized multiple-choice survey to interview both cases and controls, and the patients were guaranteed privacy and the possibility to contact the study staff in case of unclarities to avoid interviewer bias or missing/ambiguous data. Another difficulty affecting a case-control study is the selection of the control group [128]. The controls in our study were age, gender and race matched to cases to make groups comparable. Utilizing a case-control study design does not enable differentiation of genesis from consequence [85]. To get closer to the goal of determining causality between microbiota and disease/risk factors at different disease stages in a human study setting, a prospectively conducted longitudinal cohort study is needed. This would discover the alterations to the esophageal, salivary and fecal microbiome, occurring at different timepoints in the cascade of EAC development. However, the longitudinal cohort study design requires prolonged follow up periods and might thus lead to attrition bias (loss to follow-up) [128].

4.2 This cohort was a representative cohort of patients with BE associated pathologies in the population based on epidemiologic information

The epidemiologic characteristics of patients diagnosed with diseases of the esophageal adenocarcinoma cascade suggests that this cohort was a small representative sample of this patient group in the population, while the controls were age and gender matched. The percentage of smokers was higher in patients diagnosed with Barrett's Esophagus and esophageal adenocarcinoma than in the control group, which conforms with known epidemiologic information about BE related lesions. Barrett's and dysplasia subjects also differed to controls with regard to higher proton-pump inhibitor usage. However, it is still not clear whether PPIs are cancer-protective and the data in this respect is very controversial [23]. In a recent metanalysis of a total of 9 studies, no significant risk reduction of esophageal adenocarcinoma in BE patients was shown to be associated with PPI therapy (unadjusted OR 0.43, 95% CI 0.17-1.08) [23] [50]. One interesting observation in our study was, although not significant, that the number of subjects with reflux complaints was relatively lower in the dysplasia and esophageal adenocarcinoma group than in the control and BE groups, given that 3/4 DP and 9/15 EAC subjects did not report reflux symptoms. This is also consistent with previous data, where more cases of dysplasia

and esophageal adenocarcinoma have been found among patients undergoing the upper endoscopy for reasons other than reflux symptoms vs them underdoing upper endoscopy due to reflux symptoms [33]. These findings suggest that a number of patients with malignant progression unfortunately do not show reflux symptoms and are diagnosed in advanced stages due to development of other symptoms such as dysphagia. This insight is very important in shifting our focus to other epidemiological factors for BE associated diseases in order to perform screening endoscopies or other screening strategies in an additional target group other than just reflux patients. The clinical pathological data of patients diagnosed with EAC show that at timepoint of primary diagnosis, the majority of them already had stage three carcinoma. Studies to date also report that the esophageal adenocarcinomas are frequently diagnosed in advanced stages and the five-year survival after diagnosis is approximately 17% [114]. This fact makes clear the urgent necessity of developing novel cancer screening and early detection strategies.

4.3 PAXgene fixed biopsies provided the histological diagnosis and sufficient microbial DNA quality and quantity

We used the PAXgene stabilization and fixation system, a formalin free system for pre-analytical preparation of our biopsies sampled in a clinical setting, as we aimed to check the morphology, assess the goblet cell ratio and identify the tissue-associated microbiome. To this end, H&E as well as PAS and Alcian blue stainings were performed from each biopsy sample in addition to the stainings required for histological evaluation by the pathologist. All microscopically examined biopsies contained evaluable tissue with well-preserved morphology represented by high-contrast colors.

The goblet cell ratio was not found to differ significantly between non-dysplastic and dysplastic Barrett's samples, but a trend toward lower goblet cell ratio among dysplastic BE biopsies was apparent. As mentioned in the introduction to this dissertation, a previous study by our research group revealed a significant lower goblet cell ratio in endoscopic mucosa resection (EMR) samples of T1 esophageal adenocarcinoma than in non-dysplastic Barrett tissue samples by using the same calculation approach as in the present study [115]. The explanation for this might be, in addition to the higher number of samples analyzed in the previous study, that a resection specimen from EMR was analyzed, whose sample size is much higher than that of biopsies. The results of our study show that the determination of goblet cell ratio in biopsies might not operate alone as a risk predictor for disease progression. It is usually only patients with conspicuous BE such

as nodular BE, dysplastic Barrett's or esophageal adenocarcinoma T1a who are scheduled for an EMR for therapeutic and rarely for diagnostic purposes, as this carries a higher intervention risk and requires higher technological and endoscopist experience in comparison to biopsies. Therefore, EMR specimens do not play a crucial role in the clinical routine for risk prediction in early BE stages, emphasizing the need for easily readable prediction markers, which are safe and independent of the endoscopist's experience. The need for these kinds of markers is also underscored by the fact that the vast majority of patients diagnosed with Barrett's Esophagus do not progress to esophageal adenocarcinoma (approximately 0.2-0.3%progress annually) [51], and the detection rates of cancer in early stages is unfortunately low. The characterization of tumor niche alterations, which are manifested in the entire lower esophagus, independent of lesion localization, such as microbiota alterations are, will possible be essential components in future risk prediction models.

The quality of the DNA extracted using the PAXgene® tissue DNA Kit (Preana-lytix) from PAXgene-fixed-paraffin-embedded biopsies was satisfactory as showed by DNA ratios in 260/280 nm with mean ± SD value of 1.9 ± 0.07 and in 230/260nm with mean ± SD value of 1.51 ± 0.25 identified by NanoDrop from biopsy samples of BE, DP and EAC – these results have been already presented in the BarrettNET paper [141]. All biopsy samples were of good quality and we did not exclude any of the samples from downstream microbiota analysis. In summary, the PAXgene system, as demonstrated in previous studies, preserves both the morphology and provides sufficient DNA quality and quantity, and is feasible for future clinical research, which aims simultaneously at histopathological and biomolecular analysis.

4.4 The microbiome of mucosal biopsies, saliva and feces differed significantly from each other, suggesting a unique microbial composition of the mouth, esophagus and intestine

We found a significantly different number of species in samples originating from different body compartments, ranging from fecal samples with the highest Simpson effective number of species to esophageal mucosa tissue biopsies and finally saliva samples. Beside differences in the number of species, they also differed in the composition of the microbial community. We observed 3 clusters, one for each sample type: The samples originating from the same sampling compartment clustered together and were significantly different regarding the microbial profile composition from the other clusters. However, a

very slight non- significant overlap in microbial community structure was observed between the salivary and esophageal tissue biopsy samples.

The overlap found between saliva and esophageal biopsy microbiota is consistent with data from previous studies, which also revealed a closeness of the normal human esophageal microbiota population to the oral microbiota, which also encouraged us to examine the salivary microbiome for BE associated esophageal pathologies [95].

The microbial profile of saliva and fecal samples were found to be substantially different in an older study (2008), which compared the PCR-denaturing gradient gel electrophoresis analysis (DGGE) profiles of these two materials in healthy subjects as well as in a recent study, which assessed the transcriptionally active bacteria along the gastrointestinal tract including amongst other mucosal tissue samples as well as saliva and fecal samples in healthy subjects [74, 134].

The same 2008 study, found a significantly higher number of amplicons in feces than in saliva, which is compatible with our results [74]. In contrast to this, the study which analyzed the transcriptionally active microbiota found a higher number of species in saliva compared to feces [134]. The decreasing number of species found in the lower GI tract in comparison to the upper GI tract is attributed by Vasapolli et al. to the physiologic selection processes through the passage into the gastrointestinal tract [134]. In keeping with this logic, the higher number of species in fecal relative to saliva samples in our study, and the 2008 study, can be explained through the identification of bacterial DNA from feces, originating from the whole gastrointestinal tract and not being functionally active as in saliva due to loss of function. Additionally, it is important to note that the cohort in this study consisted of three quarters of unhealthy patients, i.e., those with a premalignant lesion and patients with esophageal adenocarcinoma, and thus makes our results not directly comparable to those for healthy cohorts. Even though we did not find any differences between controls and disease patients with regard to the species number in saliva or feces, this nevertheless renders our results incomparable in precise terms to those of the aforementioned studies.

This study is, to the best of our knowledge, one of the first studies to compare the microbiota originating from three different strategic body sites of patients with BE associated pathologies. The sampled niches, including the oral and colonic compartment represented by saliva and feces, are at the entrance and the end of the gastrointestinal tract and are therefore readily accessible microbiomes, which may serve to non-invasively lighten esophageal pathologies relative to sampling local esophageal mucosa biopsies.

These findings suggest that each of these three body niches harbors a unique microbial community. This was till recently unclear, since the esophagus was thought to be either sterile or have only transient bacteria corresponding to the microbiota originating from oral cavity and stomach [40, 93, 95]. The results of this study and previous studies finding that the bacterial biota inhabiting the esophageal and oral cavity differ from each other, illustrates that the microbiome of the esophagus is unique, and suggests that the esophageal tissue luminal and mucosal side environment has an influence on its biota [96].

The conclusion of our study, that these three niches are unique, and the knowledge of similarities/overlaps is important for decisions about where to sample the human microbiome efficiently and reasonably and which targets need further investigation, depending on the disease, as many authors recommend a selective sampling based on up-to-date knowledge of the human microbiome [134].

4.5 The microbial population in tissue was highly associated with disease phenotype

One of the aims of this study was to provide a comprehensive description of microbiota found in esophageal/cardia biopsies, as well as in saliva and fecal samples at different disease stages of esophageal adenocarcinoma cascade. Further, we sought to assess the usefulness of microbiota sampling from saliva and feces as a non-invasive tool compared to biopsy. For this purpose, we performed a phenotype-guided analysis on microbial data from all three sample types and were able to determine – for each sample type – microbes that were differently abundant at different disease stages, sub-stratifying and reflecting disease phenotype. Based on our results, the majority of significant differences could be determined in tissue biopsies. Here, notably, we found great shifts in the overall diversity and microbial community composition throughout the stages of the EAC cascade.

We found an increased microbial diversity regarding the richness and number of species (alpha diversity) in biopsies of dysplasia subjects compared to all other diagnostic groups. However, the Simpson effective, another index to estimate alpha diversity, which considers beside the absolute number of species as well as the community structure through abundance information, showed no significant difference between diagnostic groups. In early esophageal microbiome studies, which included subjects at early stages of the Barrett's Esophagus progression cascade, biopsies of gastrointestinal reflux disease and Barrett's Esophagus subjects have been reported to have a significantly higher

microbial diversity than those of healthy controls [73, 96, 143]. One of the few previous studies analyzing the esophageal microbiota from biopsy material at advanced disease stages found a decreased diversity in esophageal adenocarcinoma compared to controls, whereas dysplasia biopsies were not sampled [31]. From cytosponge material, a decreased alpha diversity has been revealed in dysplasia [31]. However, Snider et al., who more recently analyzed the microbiota from esophageal brushings of a cohort consisting of Barrett's Esophagus, LGD, HGD and esophageal adenocarcinoma subjects did not find any significant differences in alpha diversity between the phenotypes [126]. The results of the aforementioned studies are inconsistent, and the dysplasia subgroup of our study consisted of only 4 patients in total – this small number of samples might reduce the power of the statistical analysis. The aforementioned studies used different methods for sampling, preanalytical processing, and analyzing the esophageal microbiome, making the findings not directly comparable. To date, there is still no standard method for investigating the esophageal microbiome, which makes additionally data comparison difficult. In summary, however, it is important to conclude, with reference to the data available so far, that the alpha diversity of a sample is not a reliable parameter for separation between the disease phenotypes and risk estimation.

Biopsies samples from BE, DP, EAC patients and controls showed different microbial profiles; in the pairwise analysis controls differed significantly from BE, DP and EAC, while BE differed significantly from EAC. Similar studies have also found distinct microbial profiles at different disease stages [31, 126]. The distance between samples was determined using the generalized UniFrac, which considers the genetic distance between OTUs found in each sample to the OTUs in all other samples, as well es the relative abundance of OTUs, and defines what is known as beta-diversity across samples [20, 66].

We did not find a difference between microbial profiles of BE and DP subjects, but only between BE and EAC, in our study, possibly due to the small number of DP samples analyzed. In one previous study altered microbial profiles have been already identified in dysplasia stage [126]. If this could be validated, it would enable clinicians to fish out patients whose microbial profile clusters are distinct from Barrett's Esophagus and closer to dysplasia microbial profile. This would serve as an early cancer detection measure, most importantly in the cases where dysplastic progression is not detectable in biopsies.

The samples of esophageal adenocarcinoma subjects showed the highest inter-individual variations. This is potentially attributable to the different stages of the esophageal adenocarcinoma in this cohort, extending from stage 1 to stage 4. It would be of interest

to validate the promising role of the beta diversity for risk stratification and early detection in Barrett's Esophagus progression cascade in larger cohorts.

As we pooled biopsies from inconspicuous squamous esophageal mucosa, cardia and injured lesions for each patient, this method is not dependent on the accuracy of biopsy sampling from the conspicuous area, in contrast to histology. The success rates in sampling the pathologic area are unfortunately low and dependent on endoscopist experience, which additionally emphasizes the importance of these findings. Several studies have proved that the alternations in microbiota concern the whole lower esophagus and are not limited to the histologically modified area [28, 31], This clarifies the role tumor microenvironment might have, part of which is also the microbiome, in accurate risk stratification and early detection strategies. Importantly, we sampled also biopsies originating from gastric cardia due to the data of our research group which suggest that Barrett's Esophagus and esophageal adenocarcinoma may arise from gastric cardia progenitor cells [102].

In tissue biopsies, several single taxa were found to be differently abundant between the disease phenotypes and controls. These included the genera of Fusobacterium and Ralstonia, which were enriched in esophageal adenocarcinoma relative to controls, while Bradyrrhizobium, Streptococcus, Actinomyces and Granulicatella were represented in higher abundance in controls as compared to disease phenotypes. Both Fusobaterium and Ralstonia belong to gram-negative genera and are in conformity with previous findings, which have repeatedly reported an overabundance of gram-negative genera in BE [75].

It should be noted that some taxa were significantly differently present or abundant in BE and EAC and may have importance as biomarkers for risk stratification among BE patients. The genera of Actinobacteria and Granulicatella were less abundant in EAC as compared to BE, while the phylum of Firmicutes was overabundant in EAC compared to BE. These taxa with the potential to distinguish between BE patients with or without increased risk of malignant progression need to be validated in a prospective longitudinal follow-up study.

An overabundance of gram-negative strains has been also isolated much earlier in subjects with colorectal cancer. In particular, Fusobacterium has received increased attention among other taxa in the last few years due to its association with colorectal cancer (CRC). However, several authors focusing on the esophageal microbiome in EAC and its premalignant conditions have also reported an enrichment of Fusobacterium in BE

[69]. For CRC, the mechanism by which Fusobacterium may initiate and progress carcinogenesis are widely investigated [12]. Fusobacterium promotes cell proliferation through binding of its FadA protein to E-cadherin of mucosal cells and subsequently activation of ß-Catenin and WNT pathway, and through activation of TLR4 and NF-κB [88]. Once the tumor is present, Fusobacterium nucleatum can bind the Gal-GalNAc protein that is expressed in the surface of tumor cells and thus enrich its abundance in the tumor niche and promote tumor progression [88].

Further, our analysis showed that Streptococcus presents the most abundant taxon in both saliva and esophageal tissue microbial community and its relative abundance is decreased in the tissue of BE and EAC patients. This suggests that dysbiosis with large shifts in several single taxa occurs in the esophagus during BE and EAC pathogenesis.

In addition, several taxa were exclusively present in biopsies of patients with disease phenotype or solely in EAC, and included, inter alia, Campylobacter, unknown Lachnospiracae and Atobium in disease phenotypes and Anaeroplasma, ASF 356, Rikenellacae RC9, Clostridiales vadin BB60 and Ruminoclostridium in EAC. In fact, several previous studies profiling the local esophageal microbiome reported the presence of Campylobacter exclusively in biopsies of subjects with the disease phenotype, whereas none of the controls showed a Campylobacter colonization in their esophagus [73]. As such, questions with regard to Campylobacter's role in EAC pathogenesis and hypothesis of its significance in EAC to the one of H. pylori in gastric cancer have been raised [60]. Campylobacter species possess special virulence properties which may contribute to carcinogenesis, such as toxin production, cellular invasion and intracellular survival. Colonization with Campylobacter concisus was associated with higher levels of IL-18 and IL-8 expression [87]. Furthermore, macrophages infected with Campylobacter concisus produce bacterial DNA recognizing macromolecular patterns and, as a result, trigger innate immune response. The infection with this species has also been shown to affect several genes and proteins related to EAC pathogenesis. For example, the known NF-κB and STAT3 pathways are upregulated during colonization with C. concisus. The TGF-b1 pathway, loss of which is associated with BE development, is downregulated during C. concisus infection.

These examples and others support the hypothesis that this single taxon might have the potential to contribute to EAC pathogenesis. What is intriguing with regard to our results showing higher abundance and prevalence of Campylobacter and Fusobacterium in disease phenotypes, is the potential of Campylobacter to aggregate with Fusobacterum and Leptotricha species, as reported by some authors studying the microbiome in colorectal

cancer [138]. Also intriguing is that some Campylobacter species may also colonize the human oral cavity, and that Fusobacterium nucleatum primarly inhabits the oral cavity and has been associated with periodontal disease [7]. This raises the question why we did not find alterations of the relative abundance in saliva samples. The saliva microbiota is only a representation of the oral microbiome, but it might be that these microbiotas prefer to reside and stay stacked in the mucosa, which makes them undetectable in saliva.

In saliva, the overall microbial community diversity and composition did not show separation between phenotypes, but at the OTU's level, alternations were observed in the abundance of Streptococcus and Granulicatella. Snider et al. also failed to reveal any significant separation of the overall diversity in their case control study with 49 patients between BE and controls, while they identified an overabundance of Firmicutes and underabundance of Proteobacteria at the phylum level in BE patients relative to controls [125]. Additionally, Snider et al. suggest the possibility of diagnosing BE non-invasively by analyzing the salivary microbiome with a sensitivity of 97% and specificity of 88% using a three-taxon model including Streptococcus, Lautropia and unspecified genus of the Bacteroidales order [125]. In the present study, we could not determine any significant enrichment of Firmicutes at the phylum level among BE patients compared to controls, nor alterations of Lautropia or Bacteroidales. One of the first prospective study of oral microbiome also showed no differences between phenotypes in terms of the overall diversity but revealed a relation between Tannerella forsythia and higher EAC risk as well as an inverse relation between Streptococcus pneumoniae and EAC [97]. Tanerella forsythia has been associated in previous studies with periodontitis. However, in the aforementioned study the association of EAC with periodontitis was not consistent between the two parallel cohorts. A lower abundance of Streptococcus pneumoniae was also determined in dysplasia and EAC from biopsies by Zaidi et al. [145]. The inconsistent results reported from different studies profiling the salivary microbiome suggest that the salivary microbiome is exposed to many influences, such as smoking, eating, mouth hygiene and oral pathologies, and much higher sample sizes are needed to lower the effect of all these factors. In summary, it will be a challenging task to determine well-validated salivary taxa in order to distinguish between disease progression phenotypes and to replace invasive endoscopic examination.

To the best of our knowledge, this is the first human study to characterize the fecal microbiome in the BE progressions cascade. Previously our group has found some exiting links of the high-fat diet in the L2-IL1B Barrett's Esophagus mouse model with alternations of the gut microbiota and disease progression [80]. In this human study, samples

from different disease phenotypes did not differ in terms of overall and diversity or composition. However, we revealed an enrichment of Burkholderiaceae and a loss of Faecalibacterium and Phascolarctobacterium in EAC fecal samples in comparison to controls, as well as several taxa only prevalent in samples of diseased subjects, proving that the gut microbiota in BE pathogenesis is altered. It remains for future studies to investigate the functional roles of these bacteria, as beside important local effects in the intestine, they might also exert systemic effects, such as bile acids - metabolism and - blood levels. Bile acids have been associated with colon and esophageal cancer [38, 102]. Deoxycholic acid (DCA) results from the metabolism of primary bile acids by bacterial enzymes. In fact, Zheng et al. found increased bile acids in blood levels 12h after feeding mice with a high fat diet, and 24h later, an alteration of gut microbial community composition [148]. Therefore, since we were able to show that the gut microbiota in the BE progressions cascade is altered in this human cohort, our groups' future projects will focus on the metabolomics analysis of these samples to identify taxa involved in the metabolism of bile acids. If these links can be proved, there is every prospect of future development of prevention or risk stratification algorithms, which also take into consideration the fecal microbiota composition for malignant progression in BE.

In the end, the mechanisms by which bacteria contribute to cancer development in BE are poorly understood. There are two most commonly discussed principles: The first implements bacterial crosstalk with the immune system, leading to the release of inflammatory cytokines followed by activation of cancer-promoting signaling pathways – this has been supported by studies focusing on cytokine releasing (e.g. IL-8), myeloid-derived suppressor cells, bacterial LPS, Toll-like receptors and NF-κB pathway [2, 15, 72, 80, 86, 118, 132, 145]; the second implements the ability of bacteria to produce metabolites that can promote cancer – an example is the elevation of bile acid blood levels by bile acid-metabolizing bacteria, which have also been linked to BE carcinogenesis, or bacteria associated with obesity or higher glucose/lactate serum levels [16, 70, 146, 148]. Supporting studies exist for both of the aforementioned principles, although metabolomic studies have been limited to small cohort numbers or have been so far descriptive and had a statistical design rather than a mechanistic one. Based on current data, it is highly suggestive that bacteria promote carcinogenesis at least partly through their crosstalk with the host immune system.

4.6 Most-known risk factors for esophageal adenocarcinoma were associated to microbiota alternations

In the course of this study, questions have been raised as to whether the patient's exposure to the strongest risk factors for esophageal adenocarcinoma affected their microbiota composition. Since we obtained comprehensive epidemiological data from our patients utilizing the BarrettNET survey, we had the facility to integrate metadata into the microbial data and conduct further advanced analyses. Cigarette/tobacco smoking (former or current smoking) and the presence of gastroesophageal reflux symptoms during the previous week (burning sensation behind the sternum or acidulous taste in the mouth or regurgitation of stomach contents) were taken into consideration, as the most investigated and established risk factors for the onset of esophageal cancer.

In our epidemiological data statistical analysis, the proportion of habitual smokers (present or former smokers) was found to be higher among Barrett's Esophagus and esophageal adenocarcinoma subjects than in controls even for this small cohort of patients. In contrast, the proportion of current smokers in the different diagnostic groups did not vary significantly, revealing that the majority of diseased subjects had quit smoking. Unfortunately, only a slightly reduced risk for EAC development was revealed in a large meta-analysis in former smokers, who have quit smoking for 20 or more years, compared to current smokers, showing that former smokers' risk is not reduced to that of those who have never smoked [23, 137]. In contrast to this, quitting smoking in subjects with esophageal squamous cell carcinoma was strongly associated with lower risk of cancer [137]. For this reason, we conducted our analysis by sub-stratifying our cohort into smokers and never-smokers to identify microbiota alternations that are smoking-related regardless of whether the subjects are current or former smokers.

In our epidemiological data statistical analysis, the number of patients reporting reflux symptoms did not differ significantly between the diagnosis groups. This can probably be explained by the high portion of controls reporting reflux symptoms and by the significant higher intake of PPI in the disease groups (BE and DP), which might have led to symptom relief.

We found a significant separation between the microbial profiles of patients with reflux symptoms and patients without reflux symptoms in saliva, but no significant difference with regard to the number of species. These results conform the results of one 2019 conducted study, which also did not reveal any significant difference in the number of

species between GERD and non-GERD patients in saliva, but a significant difference in microbial community composition [149]. The altered microbial community composition in patients with reflux symptoms suggests a direct influence of the gastroesophageal refluxate reaching the mouth on the oral microbiota.

In risk factor- and diagnosis-directed microbial analysis only the microbial profiles of BE patients (not of other phenotypes) showed an association with smoking or reflux in all three sample types: in tissue an association to smoking was found, in saliva an association to reflux symptoms and in feces an association to smoking and reflux symptoms. Remarkably, in tissue the microbial profiles of non-smokers were closely clustered, while those of smokers showed high interindividual variations - EAC patients also showed high interindividual variations. This did not apply to saliva and fecal samples. To the best of our knowledge, there are no studies that analyzed the esophageal biopsy-associated microbiome with focus on smoking history, so accurate comparisons with the results of other studies are not possible. However, there are studies that have analyzed this matter in saliva and in esophageal balloon cytology samples. In saliva, higher inter-individual variability was found among smokers compared to non-smokers [113]. In the other study, the upper gastrointestinal microbiome of 278 Chinese men was analyzed from esophageal balloon cytology samples – here current smokers showed a higher number of unique species and different microbial profiles compared to non-smokers, while former smokers showed no significant difference from non-smokers [136].

In the abundance analyses we found a decrease of Actinobacteria in biopsies of BE smokers and EAC patients. Also, in saliva samples we found a decrease of one Streptococcus species in BE smokers and EAC patients. A previous study comparing the microbial profiles of smokers and non-smokers in saliva showed that Streptococcus was the most common genus in both groups and that it was present in a higher percentage in non-smokers than in smokers [112] – this result is not fully consistent with our result as we did not find differences between both groups for Streptococcus at genus level but only for a specific Streptococcus species. As regards reflux symptoms, we identified one Streptococcus species in saliva which was enriched in patients which reported reflux symptoms but depleted in EAC patients. These results show that certain bacteria associated with smoking in saliva and biopsies are associated with an advanced phenotype in the same direction; while bacteria associated with reflux are associated with an advanced phenotype in the opposite direction, suggesting that smoking may be a causal factor in microbiota shifts and malignant progression.

Cigarette smoking is a well-established risk factor for EAC and BE, but the mechanism by which smoking is involved in pathogenesis are still unknown. Currently there are indications that smoking leads to DNA damage in the BE mucosa [89]. In addition, it is also known that smoking shuttles the immune system down [129]. In recent years, given the higher feasibility of accessing the microbiome and the broad range of alterations in microbiota revealed as associated with smoking in saliva, the focus has been directed to the effects of smoking on microbiota composition and contribution to disease development.

The microbiota alterations caused by reflux in saliva were inversed correlated with disease progression, which suggests a protective role of reflux induced bacteria in pathogenesis, strengthened also by the fact that the most cancer patients in this cohort did not report reflux symptoms. The problem with investigating reflux as risk factor is that it encompasses a number of subjective complaints; and PPI usage is a potential confound in that it leads to relief of reflux symptoms and to microbiota alterations [126] for their part.

Our results indicate that some risk factors may be linked to microbiota shifts and clinical diagnosis, and at this point it can only be hypnotized that this connection is casual. Still, these results should be interpreted with caution as we considered here the microbiomes of a small representative cohort in the cascade of BE carcinogenesis.

4.7 Clinical and scientific implications and suggestions for future research

In this study, we hypothesized that microbiota may have a significance in both the diagnostics and pathogenesis of esophageal adenocarcinoma. We were able to show that the microbiota composition and abundance in tissue biopsies differs between controls, BE, DP and EAC subjects, independently of the biopsy sampling location. This was thanks to the fact that we pooled biopsies from the cardia, esophagus and injured regions of each patient together, thus demonstrating the potential role of microbiota in diagnostics. The association of bacterial biota with EAC pathogenesis can be explained by the potential of several environmental factors, in addition to host genetic and epigenetic, to alter the microbiota and the cancerogenic or protecting properties of microbiota as revealed in previous studies. In fact, we were able to identify taxa altered in subjects sharing the same diagnosis but having exposure to smoking or reflux complaints, amongst others, for taxa which have previously been linked to cancer.

The most frequently asked question is whether microbiota perturbations contribute to cancer pathogenesis or whether it is the cancerogenic tissue perturbating the microbiota that inhabits that niche [75]. The overgrowth or deletion of specific taxa related to the subject's diagnosis may also be influenced by microscopic structure or release of toxic/stimulating substances in the associated Barrett's lesion niche, thus encouraging bacterial selectivity. In this sense, the results of this study should be interpreted with caution. Single taxa might present potential biomarkers of cancer progression and some of them might even have partly a causative role for development or lack of EAC.

The results of this study may open up new doors for promising future findings. For this initially, prospective longitudinal large cohort human studies are needed to identify the alternations of microbiome at different progression and rest stages of disease and validate possible biomarkers.

The performance of 16S rRNA gene sequencing is relatively expensive for clinical routine and requires intensive laboratory work. For this reason, a PCR test of keystone taxa might offer an alternative to identifying these faster and more cost-effectively [125].

In addition, the pathways driven by microbiota in the host organism are promising areas, which should be further investigated using animal experimental approaches and human translational studies to better uncover the mechanism by which microbiota contribute to pathogenesis.

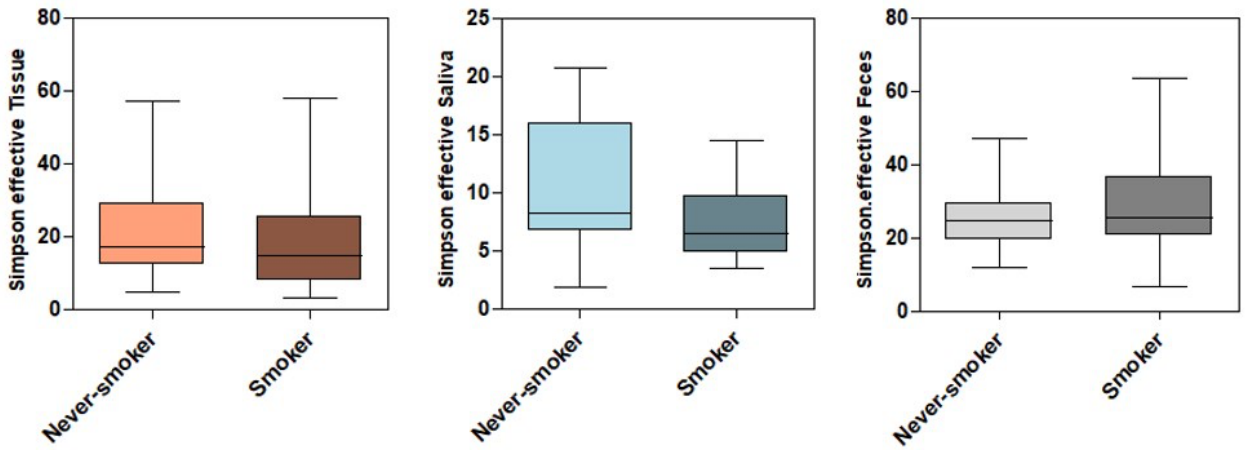
Attachment

Tissue		Saliva		Feces	
BE S vs. NS	**BE R vs. NR**	**BE S vs. NS**	**BE R vs. NR**	**BE S vs. NS**	**BE R vs. NR**
p__Actinobacteria	f__Beijerinckiaceae	OTU_1584 g_Streptococcus	p__Proteobacteria	p__Proteobacteria	p__Actinobacteria
p__Bacteroidetes	f__Family XI	OTU_4168_g_Streptococcus	c__Clostridia	c__Gammaproteobacteria	c__Actinobacteria
p__Proteobacteria	g__Methylobacterium	OTU_65 g_Actinomyces	c__Gammaproteobacteria	f__Bacteroidaceae	c__Erysipelotrichia
c__Actinobacteria			c__Negativicutes	g__Bacteroides	o__Bifidobacteriales
c__Bacteroidia			o__Actinomycetales	g__Dorea	o__Erysipelotrichales
c__Gammaproteobacteria			o__Bacillales	g__Faecalibacterium	f__Bifidobacteriaceae
o__Actinomycetales			o__Clostridiales	g__uncultured	f__Erysipelotrichaceae
f__Actinomycetaceae			o__Pasteurellales	OTU_12 g_Subdoligranulum	f__Prevotellaceae
f__Burkholderiaceae			o__Selenomonadales	OTU_16 g_Blautia	f__Veillonellaceae
g__Actinomyces			f__Actinomycetaceae	OTU_44 g_Dorea	g__Bifidobacterium
OTU_2 g_Streptococcus			f__Pasteurellaceae		g__CAG−56
OTU_29 g_Actinomyces			f__Veillonellaceae		g__Erysipelotrichaceae UCG−003
OTU_65 g_Actinomyces			g__Actinomyces		OTU_136 g_CAG-56
			g__Gemella		OTU_52 g_Erysipelotrichaceae UCG-003
			g__Haemophilus		
			g__Veillonella		
			OTU_1583 g_Streptococcus		
			OTU_1584 g_Streptococcus		
			OTU_2073 g_Streptococcus		
			OTU_23 g_Gemella		
			OTU_261 g_Actinomyces		
			OTU_31 g_Haemophilus		
			OTU_33 g_Veillonella		
			OTU_4168 g_Streptococcus		
			OTU_75 g_Streptococcus		

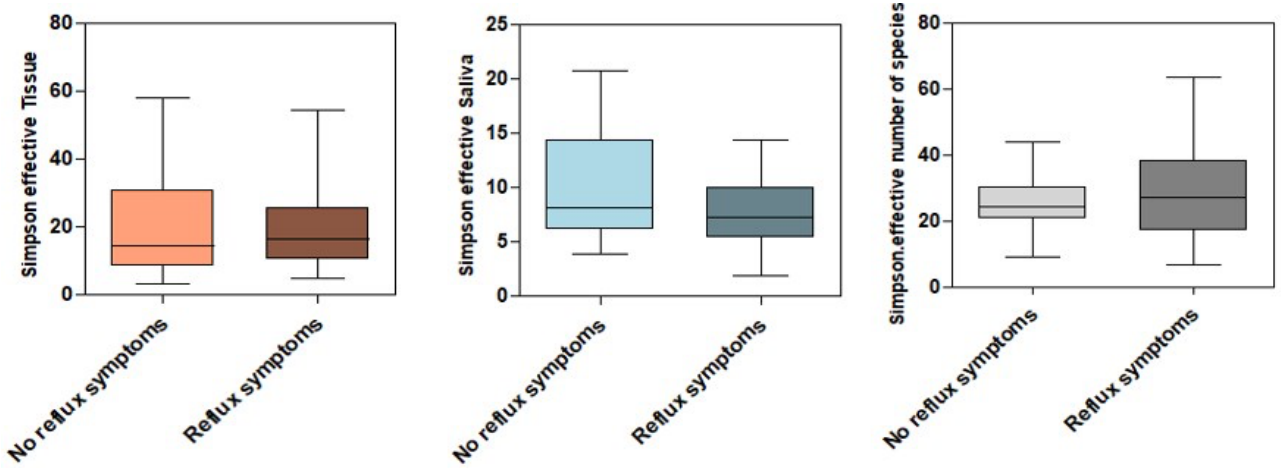
Supplementary table 1. Summary of taxa presented in a significantly distinct abundance between Barrett's Esophagus subjects with/without exposure to a risk factor.

Tissue		Saliva		Feces	
BE S vs. NS	**BE R vs. NR**	**BE S vs. NS**	**BE R vs. NR**	**BE S vs. NS**	**BE R vs. NR**
p__Actinobacteria		OTU_1584 g_Streptococcus	OTU_1584 g_Streptococcus	g__Faecalibacterium	
c__Actinobacteria			OTU_2073 g_Streptococcus		
o__Actinomycetales			OTU_75 g_Streptococcus		
f__Actinomycetaceae					
g__Actinomyces					
OTU_2 g_Streptococcus					
OTU_29 g_Actinomyces					

Supplementary table 2. Summary of taxa presented in significantly distinct abundance in both Barrett's subjects with/without exposure to a risk factor and subjects with progressed phenotypes.



Supplementary figure 1. Alpha diversity assessed by Simpson number of species in smokers and never-smokers in tissue, saliva and feces.



Supplementary figure 2. Alpha diversity assessed by Simpson number of species in smokers and never-smokers in tissue, saliva and feces.

References

1. Almhanna, K., R. Shridhar, and K.L. Meredith, *Neoadjuvant or adjuvant therapy for resectable esophageal cancer: is there a standard of care?* Cancer Control, 2013. **20**(2): p. 89-96.
2. Arora, M., et al., *TLR4/MyD88-induced CD11b+Gr-1 int F4/80+ non-migratory myeloid cells suppress Th2 effector function in the lung.* Mucosal Immunol, 2010. **3**(6): p. 578-93.
3. Arthur, J.C., et al., *Intestinal inflammation targets cancer-inducing activity of the microbiota.* Science, 2012. **338**(6103): p. 120-3.
4. Asanuma, K., K. Iijima, and T. Shimosegawa, *Gender difference in gastro-esophageal reflux diseases.* World J Gastroenterol, 2016. **22**(5): p. 1800-10.
5. ASGE Standards of Practice CommitteeJohn A.EvansMDDayna S.EarlyMDVinayChandraskharaMDKrishnavel V.ChathadiMDRobert D.FanelliMD(SAGES Representative)Deborah A.FisherMDKimberly Q.FoleyRN, B., CGRN(SGNA Representative)Joo HaHwangMDTerry L.JueMDShabana F.PashaMDRaviSharafMDAmandeep K.ShergillMDJason A.DominitzMD, MPH(Prior Chair)Brooks D.CashMD(Committee Chair)The role of endoscopy in the assessment and treatment of esophageal cancer, *The role of endoscopy in the assessment and treatment of esophageal cancer.* Gastrointestinal Endoscopy, 2013. **77**: p. Pages 328-334.
6. Ashelford, K.E., et al., *At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies.* Appl Environ Microbiol, 2005. **71**(12): p. 7724-36.
7. Baba, Y., et al., *Review of the gut microbiome and esophageal cancer: Pathogenesis and potential clinical implications.* Ann Gastroenterol Surg, 2017. **1**(2): p. 99-104.
8. Barrett, N.R., *Chronic peptic ulcer of the oesophagus and 'oesophagitis'.* Br J Surg, 1950. **38**(150): p. 175-82.
9. Belstrøm, D., et al., *Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status.* J Oral Microbiol, 2014. **6**.
10. Benjamini, Y. and Y. Hochberg, *Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing.* J. Royal Statist. Soc., Series B, 1995. **57**: p. 289-300.
11. Blaser, M.J., *Antibiotic use and its consequences for the normal microbiome.* Science, 2016. **352**(6285): p. 544-5.
12. Brennan, C.A. and W.S. Garrett, *Fusobacterium nucleatum - symbiont, opportunist and oncobacterium.* Nat Rev Microbiol, 2019. **17**(3): p. 156-166.
13. Brown, L.M., S.S. Devesa, and W.H. Chow, *Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age.* J Natl Cancer Inst, 2008. **100**(16): p. 1184-7.
14. Browne, P.D., et al., *GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms.* Gigascience, 2020. **9**(2).

15. Brudecki, L., et al., *Myeloid-derived suppressor cells evolve during sepsis and can enhance or attenuate the systemic inflammatory response.* Infect Immun, 2012. **80**(6): p. 2026-34.
16. Buas, M.F., et al., *Candidate serum metabolite biomarkers for differentiating gastroesophageal reflux disease, Barrett's esophagus, and high-grade dysplasia/esophageal adenocarcinoma.* Metabolomics, 2017. **13**(3).
17. Calatayud, S., et al., *Downregulation of nNOS and synthesis of PGs associated with endotoxin-induced delay in gastric emptying.* Am J Physiol Gastrointest Liver Physiol, 2002. **283**(6): p. G1360-7.
18. Caporaso, J.G., et al., *Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.* Proceedings of the National Academy of Sciences, 2011. **108**(Supplement 1): p. 4516.
19. Carmody, R.N. and P.J. Turnbaugh, *Host-microbial interactions in the metabolism of therapeutic and diet-derived xenobiotics.* J Clin Invest, 2014. **124**(10): p. 4173-81.
20. Chen, J., et al., *Associating microbiome composition with environmental covariates using generalized UniFrac distances.* Bioinformatics, 2012. **28**(16): p. 2106-13.
21. Choe, J.W., et al., *Application of the Prague C and M criteria for endoscopic description of columnar-lined esophagus in South Korea.* World J Gastrointest Endosc, 2016. **8**(8): p. 357-61.
22. Claesson, M.J., et al., *Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions.* Nucleic Acids Res, 2010. **38**(22): p. e200.
23. Coleman, H.G., S.H. Xie, and J. Lagergren, *The Epidemiology of Esophageal Adenocarcinoma.* Gastroenterology, 2018. **154**(2): p. 390-405.
24. Cook, M.B., C.P. Wild, and D. Forman, *A systematic review and meta-analysis of the sex ratio for Barrett's esophagus, erosive reflux disease, and nonerosive reflux disease.* Am J Epidemiol, 2005. **162**(11): p. 1050-61.
25. Corley, D.A. and A. Kubo, *Body mass index and gastroesophageal reflux disease: a systematic review and meta-analysis.* Am J Gastroenterol, 2006. **101**(11): p. 2619-28.
26. Corley, D.A., et al., *Race, ethnicity, sex and temporal differences in Barrett's oesophagus diagnosis: a large community-based study, 1994-2006.* Gut, 2009. **58**(2): p. 182-8.
27. de Vries, D.R., et al., *Gastroesophageal pressure gradients in gastroesophageal reflux disease: relations with hiatal hernia, body mass index, and esophageal acid exposure.* Am J Gastroenterol, 2008. **103**(6): p. 1349-54.
28. Deshpande, N.P., et al., *Signatures within the esophageal microbiome are associated with host genetics, age, and disease.* Microbiome, 2018. **6**(1): p. 227.
29. Edgar, R.C., *UPARSE: highly accurate OTU sequences from microbial amplicon reads.* Nat Methods, 2013. **10**(10): p. 996-8.
30. Edgren, G., et al., *A global assessment of the oesophageal adenocarcinoma epidemic.* Gut, 2013. **62**(10): p. 1406-14.

31. Elliott, D.R.F., et al., *A non-endoscopic device to sample the oesophageal microbiota: a case-control study.* The lancet. Gastroenterology & hepatology, 2017. **2**(1): p. 32-42.
32. Evans, J.A., et al., *The role of endoscopy in the assessment and treatment of esophageal cancer.* Gastrointest Endosc, 2013. **77**(3): p. 328-34.
33. Fan, X. and N. Snyder, *Prevalence of Barrett's esophagus in patients with or without GERD symptoms: role of race, age, and gender.* Dig Dis Sci, 2009. **54**(3): p. 572-7.
34. Fan, Y.P., et al., *Inducible and neuronal nitric oxide synthase involvement in lipopolysaccharide-induced sphincteric dysfunction.* Am J Physiol Gastrointest Liver Physiol, 2001. **280**(1): p. G32-42.
35. Ford, E.S., et al., *Trends in obesity and abdominal obesity among adults in the United States from 1999-2008.* Int J Obes (Lond), 2011. **35**(5): p. 736-43.
36. Franzosa, E.A., et al., *Gut microbiome structure and metabolic activity in inflammatory bowel disease.* Nat Microbiol, 2019. **4**(2): p. 293-305.
37. Frost, F., et al., *Helicobacter pylori infection associates with fecal microbiota composition and diversity.* Sci Rep, 2019. **9**(1): p. 20100.
38. Fu, T., et al., *FXR Regulates Intestinal Cancer Stem Cell Proliferation.* Cell, 2019. **176**(5): p. 1098-1112.e18.
39. Gagliardi, D., et al., *Microbial flora of the normal esophagus.* Dis Esophagus, 1998. **11**(4): p. 248-50.
40. Gagliardi, D., et al., *Microbial flora of the normal esophagus.* Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus, 1998. **11**(4): p. 248-250.
41. Godon, J.J., et al., *Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis.* Appl Environ Microbiol, 1997. **63**(7): p. 2802-13.
42. Gündisch, S., et al., *Evaluation of colon cancer histomorphology: a comparison between formalin and PAXgene tissue fixation by an international ring trial.* Virchows Arch, 2014. **465**(5): p. 509-19.
43. H. Koop1, K.H.F., J. Labenz3, P. Lynen Jansen4, H. Messmann5, S. Miehlke6, W. Schepp7, T. Wenzl8 und die Mitarbeiter der Leitliniengruppe, *Gastroösophageale Refluxkrankkheit*, in *Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V*. 2014.
44. Hampel, H., N.S. Abraham, and H.B. El-Serag, *Meta-analysis: obesity and the risk for gastroesophageal reflux disease and its complications.* Ann Intern Med, 2005. **143**(3): p. 199-211.
45. Harrison, R., et al., *Detection of intestinal metaplasia in Barrett's esophagus: an observational comparator study suggests the need for a minimum of eight biopsies.* Am J Gastroenterol, 2007. **102**(6): p. 1154-61.
46. Heger, U., et al., *Is preoperative chemotherapy followed by surgery the appropriate treatment for signet ring cell containing adenocarcinomas of the esophagogastric junction and stomach?* Ann Surg Oncol, 2014. **21**(5): p. 1739-48.
47. Herold, G., *Innere Medizin*. 2016.

48. Herszenyi, L. and Z. Tulassay, *Epidemiology of gastrointestinal and liver tumors.* Eur Rev Med Pharmacol Sci, 2010. **14**(4): p. 249-58.
49. Holle, R., et al., *KORA--a research platform for population based health research.* Gesundheitswesen, 2005. **67 Suppl 1**: p. S19-25.
50. Hu, Q., et al., *Proton Pump Inhibitors Do Not Reduce the Risk of Esophageal Adenocarcinoma in Patients with Barrett's Esophagus: A Systematic Review and Meta-Analysis.* PLoS One, 2017. **12**(1): p. e0169691.
51. Hvid-Jensen, F., et al., *Incidence of adenocarcinoma among patients with Barrett's esophagus.* N Engl J Med, 2011. **365**(15): p. 1375-83.
52. Ianiro, G., H. Tilg, and A. Gasbarrini, *Antibiotics as deep modulators of gut microbiota: between good and evil.* Gut, 2016. **65**(11): p. 1906-1915.
53. IARC, W.G., *Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994.* IARC Monogr Eval Carcinog Risks Hum, 1994. **61**: p. 1-241.
54. Iida, N., et al., *Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment.* Science, 2013. **342**(6161): p. 967-70.
55. Jakobsen, T.H., et al., *Complete genome sequence of the cystic fibrosis pathogen Achromobacter xylosoxidans NH44784-1996 complies with important pathogenic phenotypes.* PLoS One, 2013. **8**(7): p. e68484.
56. Janda, J.M. and S.L. Abbott, *16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls.* J Clin Microbiol, 2007. **45**(9): p. 2761-4.
57. Jette, M.E., et al., *The human laryngeal microbiome: effects of cigarette smoke and reflux.* Sci Rep, 2016. **6**: p. 35882.
58. Johnson, J.S., et al., *Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis.* Nature Communications, 2019. **10**(1): p. 5029.
59. Jost, L., *Jost, L. Entropy and diversity. Oikos.* Oikos, 2006. **113**: p. 363-375.
60. Kaakoush, N.O., et al., *Is Campylobacter to esophageal adenocarcinoma as Helicobacter is to gastric adenocarcinoma?* Trends Microbiol, 2015. **23**(8): p. 455-62.
61. Kazor, C.E., et al., *Diversity of bacterial populations on the tongue dorsa of patients with halitosis and healthy patients.* J Clin Microbiol, 2003. **41**(2): p. 558-63.
62. Kim, T.J., et al., *Multimodality assessment of esophageal cancer: preoperative staging and monitoring of response to therapy.* Radiographics, 2009. **29**(2): p. 403-21.
63. Klindworth, A., et al., *Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.* Nucleic Acids Res, 2013. **41**(1): p. e1.
64. Koch-Institut, G.d.e.K.e.V.G.u.Z.f.K.-r.Z.i.R., *Krebs in Deutschland, ed 9. Berlin*. 2013, Gesellschaft der epidemiologischen Krebsregister e.V. (GEKID) und Zentrum für Krebs-registerdaten (ZfKD) im Robert Koch-Institut.
65. Kubo, A., et al., *Sex-specific associations between body mass index, waist circumference and the risk of Barrett's oesophagus: a pooled*

analysis from the international BEACON consortium. Gut, 2013. **62**(12): p. 1684-91.
66. Lagkouvardos, I., et al., *Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons.* PeerJ, 2017. **5**: p. e2836.
67. Lagkouvardos, I., et al., *IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies.* Sci Rep, 2016. **6**: p. 33721.
68. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, D.K., AWMF), *S3-Leitlinie Diagnostik und Therapie der Plattenepithelkarzinome und Adenokarzinome des Ösophagus, Langversion 2.0 - Dezember 2018, AWMF-Registernummer: 021/023OL*. 2018.
69. Liu, N., et al., *Characterization of bacterial biota in the distal esophagus of Japanese patients with reflux esophagitis and Barrett's esophagus.* BMC Infect Dis, 2013. **13**: p. 130.
70. Lopetuso, L.R., et al., *Esophageal microbiome signature in patients with Barrett's esophagus and esophageal adenocarcinoma.* PLoS One, 2020. **15**(5): p. e0231789.
71. Ltd, T.S.A.P., *Periodic Acid–Schiffs Stain*
72. Lv, J., et al., *Alteration of the esophageal microbiota in Barrett's esophagus and esophageal adenocarcinoma.* World J Gastroenterol, 2019. **25**(18): p. 2149-2161.
73. Macfarlane, S., et al., *Microbial colonization of the upper gastrointestinal tract in patients with Barrett's esophagus.* Clin Infect Dis, 2007. **45**(1): p. 29-38.
74. Maukonen, J., et al., *Intra-individual diversity and similarity of salivary and faecal microbiota.* Journal of medical microbiology, 2008. **57**(Pt 12): p. 1560-1568.
75. May, M. and J.A. Abrams, *Emerging Insights into the Esophageal Microbiome.* Curr Treat Options Gastroenterol, 2018. **16**(1): p. 72-85.
76. Meves, V., A. Behrens, and J. Pohl, *Diagnostics and Early Diagnosis of Esophageal Cancer.* Viszeralmedizin, 2015. **31**(5): p. 315-8.
77. Minacapelli, C.D., et al., *Barrett's metaplasia develops from cellular reprograming of esophageal squamous epithelium due to gastroesophageal reflux.* Am J Physiol Gastrointest Liver Physiol, 2017. **312**(6): p. G615-g622.
78. Monig, S.P. and A.H. Holscher, *Clinical classification systems of adenocarcinoma of the esophagogastric junction.* Recent Results Cancer Res, 2010. **182**: p. 19-28.
79. Moss, A., et al., *Endoscopic resection for Barrett's high-grade dysplasia and early esophageal adenocarcinoma: an essential staging procedure with long-term therapeutic benefit.* Am J Gastroenterol, 2010. **105**(6): p. 1276-83.
80. Munch, N.S., et al., *High-fat Diet Accelerates Carcinogenesis in a Mouse Model of Barrett's Esophagus via IL8 and Alterations to the Gut Microbiome.* Gastroenterology, 2019.
81. Nafteux, P., et al., *Principles of esophageal cancer surgery, including surgical approaches and optimal node dissection (2- vs. 3-field).* Ann Cardiothorac Surg, 2017. **6**(2): p. 152-158.

82. Napier, K.J., M. Scheerer, and S. Misra, *Esophageal cancer: A Review of epidemiology, pathogenesis, staging workup and treatment modalities.* World J Gastrointest Oncol, 2014. **6**(5): p. 112-20.
83. Network, T.C.G.A.R., *Integrated genomic characterization of oesophageal carcinoma.* Nature, 2017. **541**(7636): p. 169-175.
84. Nomura, T., et al., *Argon plasma coagulation for the treatment of superficial esophageal carcinoma.* J Nippon Med Sch, 2007. **74**(2): p. 163-7.
85. Noto, J.M. and R.M. Peek, Jr., *The gastric microbiome, its interaction with Helicobacter pylori, and its potential role in the progression to stomach cancer.* PLoS Pathog, 2017. **13**(10): p. e1006573.
86. O'Riordan, J.M., et al., *Proinflammatory cytokine and nuclear factor kappa-B expression along the inflammation-metaplasia-dysplasia-adenocarcinoma sequence in the esophagus.* Am J Gastroenterol, 2005. **100**(6): p. 1257-64.
87. O'Sullivan, K.E., et al., *The role of inflammation in cancer of the esophagus.* Expert Rev Gastroenterol Hepatol, 2014. **8**(7): p. 749-60.
88. Odze, R.D., *Diagnosis and grading of dysplasia in Barrett's oesophagus.* J Clin Pathol, 2006. **59**(10): p. 1029-38.
89. Olliver, J.R., et al., *Risk factors, DNA damage, and disease progression in Barrett's esophagus.* Cancer Epidemiol Biomarkers Prev, 2005. **14**(3): p. 620-5.
90. Omloo, J.M., et al., *Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the mid/distal esophagus: five-year survival of a randomized clinical trial.* Ann Surg, 2007. **246**(6): p. 992-1000; discussion 1000-1.
91. Osias, G.L., et al., *Esophageal bacteria and Barrett's esophagus: a preliminary report.* Dig Dis Sci, 2004. **49**(2): p. 228-36.
92. Pajecki, D., et al., *Megaesophagus microbiota: a qualitative and quantitative analysis.* J Gastrointest Surg, 2002. **6**(5): p. 723-9.
93. Pajecki, D., et al., *Megaesophagus microbiota: a qualitative and quantitative analysis.* Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract, 2002. **6**(5): p. 723-729.
94. Paull, A., et al., *The histologic spectrum of Barrett's esophagus.* N Engl J Med, 1976. **295**(9): p. 476-80.
95. Pei, Z., et al., *Bacterial biota in the human distal esophagus.* Proc Natl Acad Sci U S A, 2004. **101**(12): p. 4250-5.
96. Pei, Z., et al., *Bacterial biota in reflux esophagitis and Barrett's esophagus.* World J Gastroenterol, 2005. **11**(46): p. 7277-83.
97. Peters, B.A., et al., *Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers.* Cancer Res, 2017. **77**(23): p. 6777-6787.
98. Prasad, G.A., et al., *Endoscopic and surgical treatment of mucosal (T1a) esophageal adenocarcinoma in Barrett's esophagus.* Gastroenterology, 2009. **137**(3): p. 815-23.
99. PreAnalytiX, *PAXgene® Tissue FIX Container (50 ml) Handbook* 2019. p. 28.
100. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC Genomics, 2012. **13**: p. 341.

101. Quante, M., et al., *Barrett esophagus: what a mouse model can teach us about human disease.* Cell Cycle, 2012. **11**(23): p. 4328-38.
102. Quante, M., et al., *Bile acid and inflammation activate gastric cardia stem cells in a mouse model of Barrett-like metaplasia.* Cancer Cell, 2012. **21**(1): p. 36-51.
103. Quante, M., T.A. Graham, and M. Jansen, *Insights Into the Pathophysiology of Esophageal Adenocarcinoma.* Gastroenterology, 2018. **154**(2): p. 406-420.
104. Quante, M., et al., *The gastrointestinal tumor microenvironment.* Gastroenterology, 2013. **145**(1): p. 63-78.
105. Quante, M.B., T., *Pathogenese und Prävention von ösophagealen Adenokarzinomen.* Der Gastroenterologe, 2017. **12**: p. 376–387.
106. Quint, L.E., et al., *Incidence and distribution of distant metastases from newly diagnosed esophageal carcinoma.* Cancer, 1995. **76**(7): p. 1120-5.
107. Ranjan, R., et al., *Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing.* Biochem Biophys Res Commun, 2016. **469**(4): p. 967-77.
108. Reddy, B.S., et al., *Colon carcinogenesis with azoxymethane and dimethylhydrazine in germ-free rats.* Cancer Res, 1975. **35**(2): p. 287-90.
109. Reed CE , E.M., *New techniques for staging esophageal cancer. .* The Surgical Clinics of North America 2002: p. 82(4):697-710.
110. Rice, T.W., D.T. Patil, and E.H. Blackstone, *8th edition AJCC/UICC staging of cancers of the esophagus and esophagogastric junction: application to clinical practice.* Ann Cardiothorac Surg, 2017. **6**(2): p. 119-130.
111. Ridlon, J.M., et al., *Bile acids and the gut microbiome.* Curr Opin Gastroenterol, 2014. **30**(3): p. 332-8.
112. Rodriguez-Rabassa, M., et al., *Cigarette Smoking Modulation of Saliva Microbial Composition and Cytokine Levels.* Int J Environ Res Public Health, 2018. **15**(11).
113. Rodríguez-Rabassa, M., et al., *Cigarette Smoking Modulation of Saliva Microbial Composition and Cytokine Levels.* Int J Environ Res Public Health, 2018. **15**(11).
114. Rustgi, A.K. and H.B. El-Serag, *Esophageal carcinoma.* N Engl J Med, 2014. **371**(26): p. 2499-509.
115. Schellnegger, R., et al., *Goblet Cell Ratio in Combination with Differentiation and Stem Cell Markers in Barrett Esophagus Allow Distinction of Patients with and without Esophageal Adenocarcinoma.* Cancer Prev Res (Phila), 2017. **10**(1): p. 55-66.
116. Schloss, P.D., D. Gevers, and S.L. Westcott, *Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies.* PLoS One, 2011. **6**(12): p. e27310.
117. Schmassmann, A., M.G. Oldendorf, and J.O. Gebbers, *Changing incidence of gastric and oesophageal cancer subtypes in central Switzerland between 1982 and 2007.* Eur J Epidemiol, 2009. **24**(10): p. 603-9.
118. Schwabe, R.F. and C. Jobin, *The microbiome and cancer.* Nat Rev Cancer, 2013. **13**(11): p. 800-12.
119. Shaheen, N.J., et al., *Is there publication bias in the reporting of cancer risk in Barrett's esophagus?* Gastroenterology, 2000. **119**(2): p. 333-8.

120. Sharma, P., et al., *The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria.* Gastroenterology, 2006. **131**(5): p. 1392-9.
121. Siewert JR, H.A., Becker K, Gössner W, *Versuch einer therapeutisch relevanten Klassifikation.* . Chirurg, 1987: p. 58: 25–32.
122. Siewert JR, H.A., Becker K, Gössner W, *Versuch einer therapeutisch relevanten Klassifikation.* . Chirurg 1987: p. 58: 25–32.
123. Sim, K., et al., *Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing.* PLoS One, 2012. **7**(3): p. e32543.
124. Singh, R.K., et al., *Influence of diet on the gut microbiome and implications for human health.* J Transl Med, 2017. **15**(1): p. 73.
125. Snider, E.J., et al., *Barrett's esophagus is associated with a distinct oral microbiome.* Clin Transl Gastroenterol, 2018. **9**(3): p. 135.
126. Snider, E.J., et al., *Alterations to the Esophageal Microbiome Associated with Progression from Barrett's Esophagus to Esophageal Adenocarcinoma.* Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 2019. **28**(10): p. 1687-1693.
127. Sobin LH, G.M., Wittekind C, *TNM Classification of Malignant Tumors, ed 7. Oxford.* Wiley-Blackwell, 2010.
128. Song, J.W. and K.C. Chung, *Observational studies: cohort and case-control studies.* Plast Reconstr Surg, 2010. **126**(6): p. 2234-42.
129. Sopori, M., *Effects of cigarette smoke on the immune system.* Nat Rev Immunol, 2002. **2**(5): p. 372-7.
130. Stavrou, E.P., et al., *Adenocarcinoma of the oesophagus: incidence and survival rates in New South Wales, 1972-2005.* Med J Aust, 2009. **191**(6): p. 310-4.
131. Suerbaum, S. and P. Michetti, *Helicobacter pylori infection.* N Engl J Med, 2002. **347**(15): p. 1175-86.
132. Tam, J.W., et al., *CD11b+ Ly6Chi Ly6G- immature myeloid cells recruited in response to Salmonella enterica serovar Typhimurium infection exhibit protective and immunosuppressive properties.* Infect Immun, 2014. **82**(6): p. 2606-14.
133. Tyler, A.D., M.I. Smith, and M.S. Silverberg, *Analyzing the human microbiome: a "how to" guide for physicians.* Am J Gastroenterol, 2014. **109**(7): p. 983-93.
134. Vasapolli, R., et al., *Analysis of Transcriptionally Active Bacteria Throughout the Gastrointestinal Tract of Healthy Individuals.* Gastroenterology, 2019. **157**(4): p. 1081-1092.e3.
135. Viertler, C., et al., *A new technology for stabilization of biomolecules in tissues for combined histological and molecular analyses.* J Mol Diagn, 2012. **14**(5): p. 458-66.
136. Vogtmann, E., et al., *Association between tobacco use and the upper gastrointestinal microbiome among Chinese men.* Cancer Causes Control, 2015. **26**(4): p. 581-8.
137. Wang, Q.L., et al., *Smoking Cessation and Risk of Esophageal Cancer by Histological Type: Systematic Review and Meta-analysis.* J Natl Cancer Inst, 2017. **109**(12).

138. Warren, R.L., et al., *Co-occurrence of anaerobic bacteria in colorectal carcinomas.* Microbiome, 2013. **1**(1): p. 16.
139. Wendi Edgett, B., HTL *Alcian Blue - H&E - Metanil Yellow Stain for Diagnosing Barrett's Esophagus.* HistoLogic. **vol 37 no. 2, 2004**.
140. Wichmann, H.E., C. Gieger, and T. Illig, *KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes.* Gesundheitswesen, 2005. **67 Suppl 1**: p. S26-30.
141. Wiethaler, M., et al., *BarrettNET—a prospective registry for risk estimation of patients with Barrett's esophagus to progress to adenocarcinoma.* Diseases of the Esophagus, 2019. **32**(8): p. 0, 1-10.
142. Wu, J., et al., *Cigarette smoking and the oral microbiome in a large study of American adults.* Isme j, 2016. **10**(10): p. 2435-46.
143. Yang, L., et al., *Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome.* Gastroenterology, 2009. **137**(2): p. 588-97.
144. Youmans, B.P., et al., *Characterization of the human gut microbiome during travelers' diarrhea.* Gut Microbes, 2015. **6**(2): p. 110-9.
145. Zaidi, A.H., et al., *Associations of microbiota and toll-like receptor signaling pathway in esophageal adenocarcinoma.* BMC Cancer, 2016. **16**: p. 52.
146. Zhang, J., et al., *Metabolomics study of esophageal adenocarcinoma.* J Thorac Cardiovasc Surg, 2011. **141**(2): p. 469-75, 475.e1-4.
147. Zhang, Y., *Epidemiology of esophageal cancer.* World J Gastroenterol, 2013. **19**(34): p. 5598-606.
148. Zheng, X., et al., *Bile acid is a significant host factor shaping the gut microbiome of diet-induced obese mice.* BMC Biol, 2017. **15**(1): p. 120.
149. Ziganshina, E.E., et al., *Comparison of the Microbiota and Inorganic Anion Content in the Saliva of Patients with Gastroesophageal Reflux Disease and Gastroesophageal Reflux Disease-Free Individuals.* BioMed Research International, 2020. **2020**: p. 2681791.

Danksagung

Ich bedanke mich herzlich bei Prof. Dr. med. Quante für die Möglichkeit in seiner Arbeitsgruppe zu promovieren und für die Vergabe dieses interessanten Forschungsthemas. Zudem möchte ich mich besonders für seine dauerhafte Unterstützung bedanken.

Ein herzlicher Dank geht an meinen Betreuer Prof. Dr. rer. nat. Janssen für seine Unterstützung und wertvolle Hilfestellung.

Zudem bedanke ich mich bei allen Beteiligten der AG Quante, AG Janssen, AG Haller, der Abteilung für interdisziplinäre Endoskopie und Abteilung für Pathologie vom Klinikum rechts der Isar für die gute Zusammenarbeit.

Ein großer Dank geht an meine Eltern, Loreta und Agron Radani, die meine größte Motivation und Kraft waren, diese Arbeit erfolgreich zu absolvieren. Ein großes Dankeschön geht auch an Yannick Mayer, der mich stets ermutigt hat.

Dieses Projekt wurde von SFB 1371 unterstützt.