Translational Oncology 14 (2021) 101050

Contents lists available at ScienceDirect

Translational Oncology

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

Original Research

Distinct tumor bacterial microbiome in lung adenocarcinomas manifested as radiological subsolid nodules

Yi Ma[1], Mantang Qiu[1,*], Shaodong Wang, Shushi Meng, Fan Yang[*], Guanchao Jiang[*]

Department of Thoracic Surgery, Peking University People's Hospital, Beijing 100044, China

ARTICLE INFO

Keywords:
Lung adenocarcinoma
Microbiome
Subsolid nodules
Biomarker

ABSTRACT

Objectives: Increasing evidence indicates that microbiota dysbiosis in the human body may play vital roles in carcinogenesis. However, the relationship between microbiome and lung cancer remains unclear. In this study, we aimed to characterize the microbiome in early stage of lung adenocarcinoma (LUAD), which presented as subsolid nodules (SSN) or solid nodules (SN).

Materials and Methods: We performed 16S rRNA sequencing of 35 pairs (10 SSN and 25 SN) of LUAD tumor tissues and paired adjacent normal tissues. Machine learning was used to identify microbial signatures and construct predictive models.

Results: SSN has higher microbiome richness and diversity compared with SN (richness $p = 0.017$, Shannon index $p = 0.17$), and the microbiome composition of SSN is distinct from that of SN (Bray-Curtis $p = 0.013$, unweighted unifrac $p = 0.001$). Phylum *Chloroflexi* ($p = 0.009$), *Gemmatimonadetes* ($p = 0.018$) and genus including *Cloacibacterium* ($p = 0.003$), *Subdoligranulum* ($p = 0.002$), and *Mycobacterium* ($p = 0.034$) were significantly increased in SSN. Tumor and normal tissues had similar richness and diversity, as well as overall microbiome composition. Probiotics with anti-cancer potential, like *Lactobacillus*, showed elevated levels in normal tissues ($p = 0.018$). A random forest model with 20 genera-based biomarkers achieved high accuracy for LUAD prediction (area under curve, AUC = 0.879). Meanwhile, a five genera-based signature can accurately discriminate SSN between SN (AUC = 0.950). Cross-validation of these two models also showed high predictive performance (LUAD AUC = 0.813, SSN AUC = 0.933).

Conclusions: This study demonstrates, for the first time, the tumor bacterial microbiome composition of LUAD manifested as SSN is distinct from that presented as SN, which adds new knowledge to SSN in the perspective of microbiome. Furthermore, microbiome signatures showed good performance to predict LUAD or SSN.

1. Introduction

Lung cancer accounts for the highest cancer-related mortality worldwide [1]. Despite encouraging results achieved with targeted therapy and immunotherapy in recent years, the 5-year overall survival remains poor, mainly as a result of late diagnosis and a high frequency of recurrence and metastasis [2]. Non-small cell lung cancer (NSCLC) accounts for over 80% of lung cancer while lung adenocarcinoma (LUAD) is the major subtype [3]. With the wide application of high-resolution CT in lung cancer screening, ground-glass opacity nodules (GGNs), which also called subsolid nodules (SSNs), have emerged as a new type of lung cancer [4]. Over 80% of early-stage lung cancers manifest as radiological SSN in clinic [4,5]. SSNs have been considered as a special clinical subtype and are less aggressive than pure solid LUAD. Unlike solid lung cancer, many GGNs did not harbor the driver mutations that commonly occurred in LUAD [6]. Meanwhile, clinical and epidemiologic studies have suggested a strong association among chronic infection, inflammation, and lung cancer [7,8]. Therefore, we hypothesis that microbiome may play vital roles in the initiation and progression of LUAD.

Human commensal microbiota plays essential roles in varieties of biological processes, which have an extensive impact on human health. Dysbiosis of the human microbiota was commonly found in metabolic diseases, such as obesity and diabetes [9,10]. Recently, more and more studies revealed that microbiota may contribute to carcinogenesis and affect response to cancer therapies through various biological mechanism, including inflammation, metabolism, and immune responses [11–13]. However, most studies focus on digestive system-related cancer [12,13], rare studies have been conducted to clarify the interplay between lung cancer and microbiota. A special bond between gut micro-

* Corresponding authors.
E-mail addresses: qiumantang@163.com (M. Qiu), yangfan@pkuph.edu.cn (F. Yang), jiangguanchao@263.net (G. Jiang).
[1] These authors contributed to the work equally and should be regarded as co-first authors.

https://doi.org/10.1016/j.tranon.2021.101050
Received 2 October 2020; Received in revised form 29 January 2021; Accepted 19 February 2021

Table 1
Baseline clinical characteristics of objects in this study.

Clinical characteristics	SSN (n =10)	SN (n =25)	P-value
Age (years; mean ± SD)	57.80 ± 9.27	57.48 ± 11.46	0.938
Gender (M/F)	3/7	11/14	0.264
BMI (kg/m2, mean ± SD)	25.13 ± 3.21	23.24 ±2.41	0.065
Smoking history (Y/N)	2/8	4/21	1.000
Location (upper/middle or lower lobe)	3/7	13/12	0.285
TNM stage (I/II/III)	10/0/0	19/2/4	0.454
Lymph node metastasis (N0/N1)	10/0	19/6	0.152
Pleural invasion (Y/N)	10/0	17/8	0.073
NLR (mean ± SD)	2.16 ± 1.27	1.92 ± 0.89	0.540
PLR (mean ± SD)	114.93 ± 31.90	123.38 ± 42.47	0.575
LMR (mean ± SD)	4.15 ± 1.54	4.25 ±1.44	0.866
AMC (10^9/L, mean ± SD)	0.42 ± 0.09	0.44 ± 0.15	0.654
CHOL (mmol/L, mean ± SD)	5.00 ± 1.18	4.79 ± 1.06	0.604
TG (mmol/L, mean ± SD)	1.34 ±0.58	2.08 ±2.83	0.419
HDL (mmol/L, mean ± SD)	1.19 ± 0.37	1.10 ± 0.19	0.345
LDL (mmol/L, mean ± SD)	3.24 ± 0.84	3.04 ± 0.65	0.462

Abbreviations: SSN, subsolid nodules; SN, solid nodules; M/F, male/female; BMI, body mass index; Y/N: Yes/No; NLR: neutrophil-lymphocyte ratio; PLR, platelet-lymphocyte ratio; LMR, lymphocyte-monocyte ratio; AMC, absolute monocyte count; CHOL, total cholesterol; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SD, standard deviation.

biome and lung was constructed as “Gut-Lung Axis” [14]. Dysbiosis of gut microbiota was found to correlate with lung cancer. As lungs are no longer considered sterile, emerging evidence suggests that microbiota in lungs may play a more important role in the lung carcinogenic process [15,16]. Therefore, a better understanding of the relationship between the lung microbiome and lung cancer will promote the development of novel diagnosis and treatment strategies.

Here, we compared the microbiota profiles between LUAD tumor and normal lung tissues, as well as solid LUAD and subsolid LUAD, through the 16 s ribosomal RNA (rRNA) gene sequencing analysis. We also identified the specific microbiome signatures potentially useful for the diagnosis of lung cancer.

2. Materials and methods

2.1. Patients and tissue samples

35 lung tumor tissues and paired normal lung tissues for 16 s rRNA sequencing were obtained from patients who underwent surgical resection at Peking University People’s Hospital between April 2017 and 2018. Tumor and normal lung specimens were frozen in a liquid nitrogen tank immediately after resection in a sterile environment, and then transferred to −80 ˚C until processing for DNA extraction. Pathological diagnosis was confirmed by pathologists while staging was performed according to the TNM staging system of the IASLC, version 8 [17]. All patients were diagnosed as lung adenocarcinoma, and they were further divided into 2 groups based on thin-section CT findings including subsolid nodules (SSN group, n =10) and solid nodules (SN group, n = 25). Age, gender and body mass index (BMI) were matched between these two groups, and comprehensive clinical characteristics for the enrolled patients are summarized in Table 1. This study was approved by the ethics committee of Peking University People’s Hospital (2018PHB012-01). Written informed consent was obtained from all patients.

2.2. DNA extraction and 16s gene sequencing

Microbial DNA was extracted from the samples using the E.Z.N.A.® soil DNA Kit(Omega Bio-tek, Norcross, GA, U.S.) according to manufacturer’s protocols. The final DNA concentration and purification were determined by NanoDrop 2000 UV–vis spectrophotometer (Thermo Scientific, Wilmington, USA), and DNA quality was checked by 1% agarose gel electrophoresis. The V3–V4 hypervariable regions of the bacteria 16S rRNA gene were amplified with primers 338 F (5’- ACTCCTACGGGAGGCAGCAG-3’) and 806R (5’-GGACTACHVGGGTWTCTAAT-3’) by thermocycler PCR system (GeneAmp 9700, 110 ABI, USA). The amplified products were extracted from a 2% agarose gel and further purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) and quantified using QuantiFluor ™ -ST (Promega, USA) according to the manufacturer’s protocol. Purified amplicons were pooled in equimolar and paired-end sequenced on an Illumina HiSeq 2500 platform (Illumina, San Diego, USA). To address potential environmental contaminations, we introduced 3 negative controls that were processed with the samples. qPCR was used to evaluated the bacterial load in negative controls with 16 s primers (Forward: CCTACGGGNGGCWGCAG; Reverse: GACTACHVGGGTATCTAATCC).

2.3. Microbiome sequences analysis

Raw fastq files were demultiplexed, quality-filtered by Trimmomatic software (version 0.33) [18] and merged by FLASH software (version 1.2.7) [19]. Chimeric sequences were detected and removed using UCHIME software (version 4.2) [20]. Operational taxonomic units (OTUs) were clustered with 97% similarity cutoff using UCLUST in QIIME software (version 1.8.0) [21]. The OTUs were filtered with a threshold of 0.005% of all sequence numbers. The taxonomy of each 16 s rRNA gene sequence was analyzed by Ribosomal Database Project Classifier (version 2.2, http://sourceforge.net/projects/rdpclassifier/) algorithm [22] against the Silva 16 s rRNA database (Release128, http://www.arb-silva.de) [23] using a confidence threshold of 80%.

Alpha diversity is implemented to display the diversity of bacteria among the different samples. Richness (number of OTUs) and the Shannon diversity index in our samples were calculated with QIIME and displayed by R software (version 3.6.1). Beta diversity analysis was used to evaluate global differences between microbial community composition. Bray-Curtis and unweighted unifrac distance for principal co-ordinates analysis (PCoA) were carried out to visualize the Beta diversity between different groups. Permutational multivariate analysis of variance (PERMANOVA) test was performed to analyze microbiome similarity between different groups. We also performed non-parametric factorial Wilcoxon sum-rank test to detect species-related significant abundance differences in two groups. The linear discriminant analysis (LDA) ef-

fect size (LEfSe) method (http://huttenhower.sph.harvard.edu/lefse/) [24] was used to detect microbial biomarkers that differentiate different groups. Significant correlations ($P<0.05$, and |correlation r| >0.3) between differential genera and clinical characteristics were calculated by Spearman's correlation analyses. The random forest (RF) algorithm was applied to construct predictive models based on microbiota signatures.

Finally, the microbiota functional compositions were predicted using PICRUSt [25]. The OTU abundance table was normalized and then the metagenome functional predictions were performed and mapped on Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologues pathways [26]. STAMP software (version 2.0.8) was used to differential analysis and visualize the results [27].

2.4. Statistical analyses

Statistical analyses were implemented using QIIME, *R* and STAMP software. Fisher's exact test was used to evaluate categorical variables. Unpaired sample Student's *t*-test or Wilcoxon rank-sum test was used to compare the differences of continuous variables between two groups. Statistical tests were all two-sided, and differences with $P < 0.05$ were considered statistically significant.

3. Results

3.1. Participant characteristics

The demographic and clinical characteristics of SSN patients and SNs are shown in Table 1. All patients are northern Chinese and clinical characteristics, including age ($p = 0.938$), gender ($p = 0.264$), BMI ($p = 0.065$), smoking history ($p =1$), tumor location ($p = 0.285$), and TNM stage ($p = 0.454$) were paired between these two groups. Although lymph node metastasis and pleural invasion were only found in SN group, the differences were not significant. Meanwhile, inflammation and lipid-related parameters also showed no significant difference.

Microbial profiling of lung adenocarcinoma and paired normal tissue

A total of 56,16,903 effective taxon tags were analyzed after sequence de-noising, trimming and chimera filtration. The reads were clustered into 1436 OTUs at 97% similarity cutoff, with 1418 OTUs in lung cancer tissues and 1431 in normal tissues (Fig. 1A). The sequencing depths were evaluated by plotting the rarefaction curve for each sample (Fig. S1A). Most of the samples reached plateau, indicating that the sequencing depth was adequate. Cq values of 16s were 39.6, 39.4 and not detectable in three negative controls, which indicated the influence of contamination could be neglected. Furthermore, we compared the alpha diversity between LUAD tumor and normal groups. LUAD tumor tissue samples tend to have a lower microbial richness (number of OTUs, $p =0.95$, Fig. 1B) and diversity (Shannon index, $p =0.13$, Fig. 1C, Fig. S1B) than paired normal tissues, although the differences were non-significant. We then assessed the overall microbiome composition between the two groups. PCoA and PERMANOVA analysis did not find significant differences between tumor and normal tissues when considering Bray Curtis (R^2=0.019, $p = 0.097$, Fig. 1D–E) and unweighted unifrac distance (R^2=0.013, $p = 0.616$, Fig. S2). However, normal tissue seemed to have a left shift in Fig. 1D.

We next compared the relative abundance of microbiota composition of LUAD tumor and normal lung tissues. At phylum level, the proportions of *Actinobacteria* and *Proteobacteria* increased, but the proportions of *Firmicutes* ($p = 0.036$) and *Bacteroidetes* decreased in LUAD tumor tissues compared to that in normal tissues (Fig. 1F). At genus level, *Rhodococcus* and *Ochrobactrum* were main compositions in both tumor and normal tissues (Fig. 1G). Furthermore, we found 54 statistically significant different genera between these two groups (Fig. 2A). The relative abundance of probiotics, such as *Lactobacillus* ($p = 0.018$) showed a notable decline in tumor tissues. Moreover, we performed LEfSe analysis to identify biomarkers that may discriminate between LUAD tumor and normal tissues, which revealed 40 discriminative features (LDA > 2, $p < 0.05$, Fig. 2B–C) at the class ($n = 4$), order ($n = 6$), family ($n = 10$) and genus ($n = 20$). Interestingly, we found that tumor group was associated with *Parvibaculales, Parvibaculaceae*, and *Parvibaculum* at order, family and genus level, respectively. *Renibacterium* and *Ancylobacter* were prominent genus level biomarkers. For normal tissues, *Lactobacillus* was the strongest marker regardless of species levels.

KEGG pathways enrichment analyses were performed to explore functional differences of the microbiome in tumor and normal tissues. Although the functional composition of these two groups was highly similar, cGMP-PKG signaling pathway was significantly less abundant in tumor than normal tissues ($p = 0.03$, Fig. 2D).

Associations between lung microbial species and clinical characteristics

To further explore the potential roles microbiota played in LUAD, we investigated the relationships between the relative abundance of genus features ($n = 20$) and several clinical indices using Spearman's correlation analyses (Fig. 2E). The abundance of LUAD-enriched genus, *[Eubacterium]_nodatum_group*, was positively correlated with neutrophil-lymphocyte ratio (NLR, $r = 0.361$, $p = 0.033$) and platelet-lymphocyte ratio (PLR, $r = 0.458$, $r = 0.006$). The abundance of *Finegoldia* was negatively correlated with pleural invasion ($r = -0.353$, $p = 0.038$). And we also found a negative correlation between *Renibacterium* and lymph node metastasis ($r = -0.343$, $p = 0.043$). In addition, the lipid-related parameters total cholesterol (CHOL), triglyceride (TG), and low-density lipoprotein (LDL) were positively or negatively correlated with different genera.

3.4. Detection of LUAD based on the lung microbiota

Based on features identified by LEfSe analysis, we constructed a random forest model to differentiate tumor and normal tissues. All 71 samples were randomly split into the training set (70%) and test set (30%) for model calibration and validation. All the 20 genera were selected based on the feature evaluation step. In the training set, this model had an AUC of 0.879 (95% CI, 0.786–0.972, Fig. 2F). We used the test set to further evaluate the model and reached acceptable predictive power (AUC = 0.806, 95% CI: 0.624–0.987, Fig. 2F), albeit lower than that in the training set. Moreover, 10-fold cross-validation was performed five times for model assessment, resulting in an average AUC of 0.813. These results suggest that our model has stable performance and may provide a possibility for clinical transformation.

3.5. Distinct tumor bacterial microbiome between subsolid and solid nodules

Microbiome profiles are different between SSN and SN. SSN group owns 102 distinctive OTUs while SN group has 232 unique OTUs (Fig. 3A, Fig. S3). At phylum level, both SSN and SN were mainly composed of *Actinobacteria, Proteobacteria, Firmicutes, and Bacteroidetes* (Fig. 3B). SSN had significantly higher abundances of *Chloroflexi* ($p = 0.009$) and *Gemmatimonadetes* ($p = 0.018$). At genus level, *Rhodococcus* and *Ochrobactrum* are main compositions in these two groups (Fig. 3C). Furthermore, 53 diverse abundance genera were found using Wilcoxon rank-sum test (Fig. 3D). Remarkably, SSN group has higher abundances of *Cloacibacterium* ($p = 0.003$), *Subdoligranulum* ($p = 0.002$), and *Achromobacter* ($p = 0.001$).

For alpha-diversity, SSN group has greater bacterial richness ($p = 0.017$, Fig. 4A) and diversity ($p =0.17$, Fig. 4B, Fig. S1C). We then assessed the microbiome composition of all tumor tissues with beta-diversity values. Both PERMANOVA and PCoA analyses confirmed SSN and SN have statistically different microbiome composition when considering unweighted unifrac distance (R^2=0.085 $p = 0.001$, Fig. 4C–

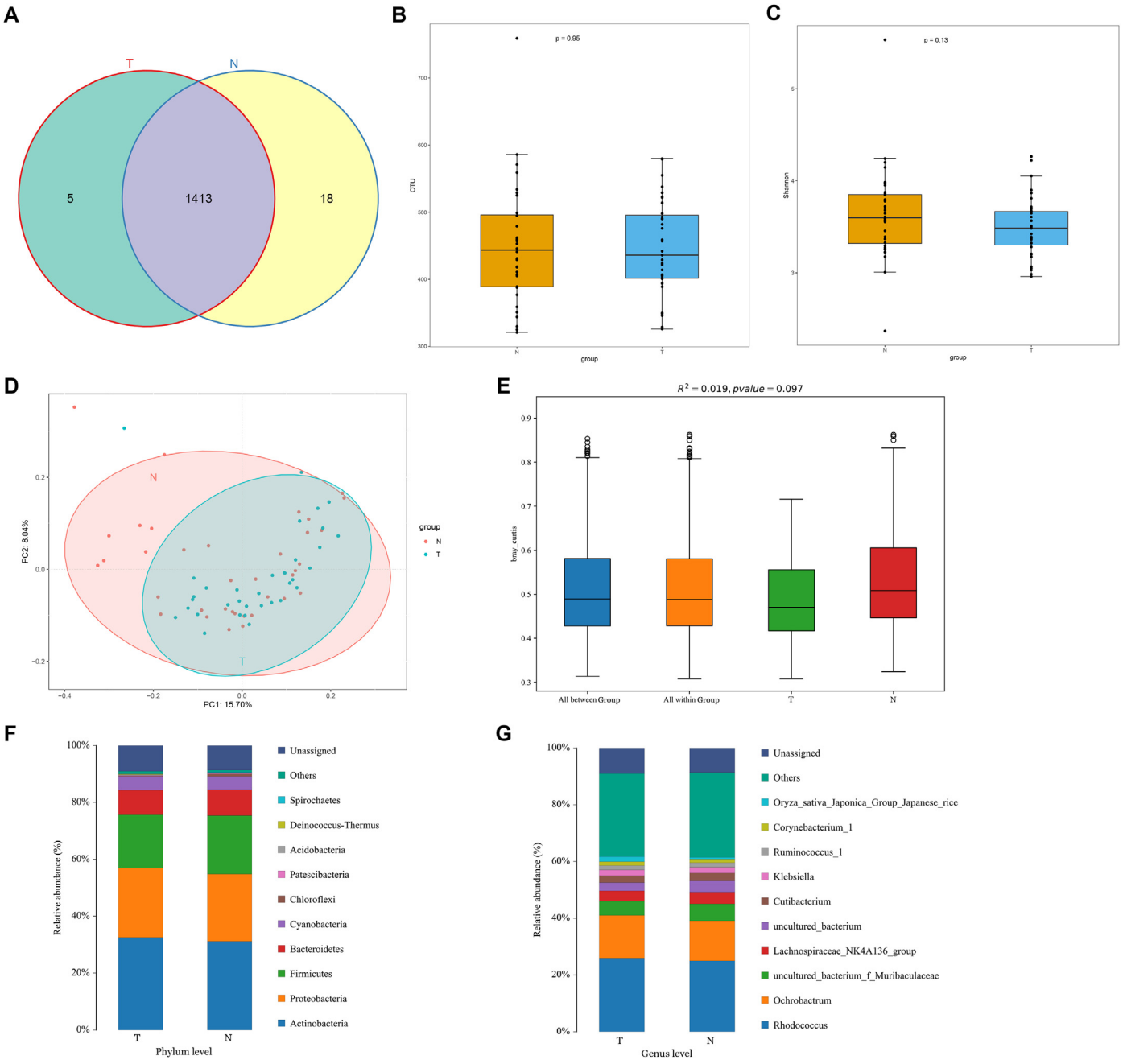

Fig. 1. Microbial profiles of lung adenocarcinoma and paired normal tissue. (A) Different operational taxonomic units (OTUs) between tumor and normal tissues. (B) Richness was evaluated by the number of OTUs. (C) Shannon index of tumor and normal tissues. (D) Principal coordinate analysis (PCoA) using Bray Curtis method. (E) PERMANOVA analysis using Bray Curtis method. (F) Species composition at phylum level. (G) Species composition at genus level.

D) and Bray Curtis (R^2=0.05, p = 0.013, Fig. 4E–F). The microbiome composition and species abundance similarity of all tumor tissues was shown in Fig. S4. Moreover, LEfSe analysis revealed 54 features that may discriminate SSN and SN (LDA > 2.5, p < 0.05, Fig. 5A–B). Consistent with composition differences, *Chloroflexi* and *Gemmatimonadetes* were detected as potential markers for SSN at phylum level. At genus level, increased bacteria such as *Cloacibacterium, Subdoligranulum*, and *Mycobacterium* and decreased bacteria like *Lachnoclostridium* were strong discriminative features for SSN.

Furthermore, we constructed a predictive model to discriminate SSN from SN. According to LEfSe result, 26 genus features were selected as candidate signatures. After feature elimination, the best performing model was selected as the final model, with five genera signatures (Fig. 5C). SSN could be accurately discriminated from SN using random forest algorithm with a sensitivity of 1, specificity of 0.8, and AUC of 0.950 (95% CI, 0.852–1.000, Fig. 5D). To further verify the accuracy of this model, we performed 10-fold cross-validation for five times, resulting in a mean AUC of 0.933.

16s functional predictive analysis also showed differences between SSN and SN groups. KEGG pathway comparisons found oral microbiome related pathways, salivary secretion (p<0.001) and gastric acid secretion (p<0.001) were significantly enriched in SSN but not in SN (Fig. 5E). In addition, endocrine and other factor-regulated calcium reabsorption and two other pathways were also different between these groups.

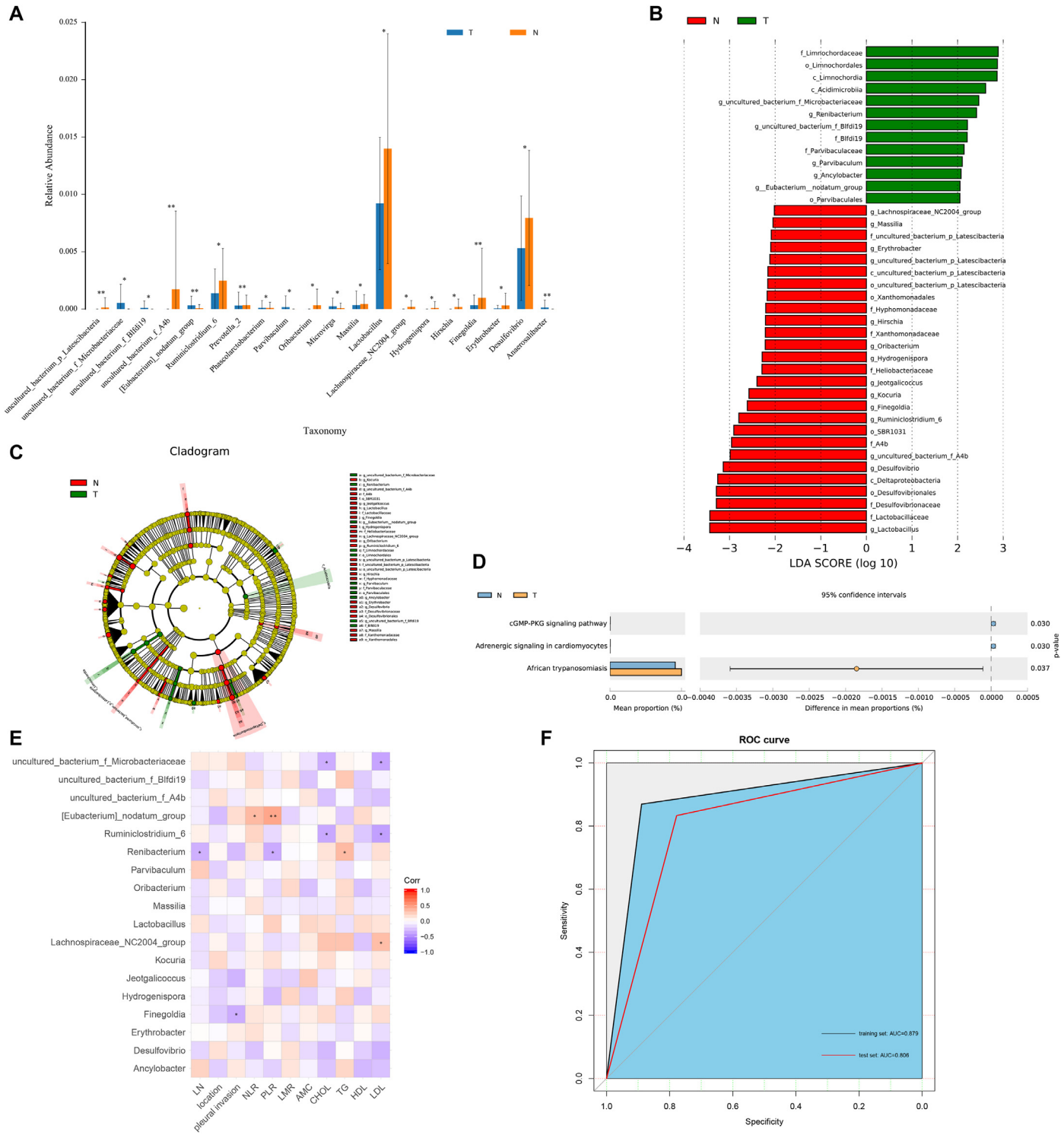

Fig. 2. Identification of predictive biomarkers for lung adenocarcinoma and functional prediction. (A) Top 20 differential genus between tumor and normal tissues. (B) LEfSe analysis identified 40 markers at different levels between tumor and normal tissues. The first letter indicates taxonomy level, p: phylum, c: class, o: order, f: family, g: genus. (C) Cladogram of bacterial lineages with significantly different representations in tumor and normal tissues. (D) Differential pathways at KEGG pathway level 3. (E) The relationships among clinical indicators and genera features. (F) ROC curve for the training set and test set. $^{*}p < 0.05$, $^{**}p < 0.01$. ROC, receiver operating characteristic; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Discussion

This study demonstrates, for the first time, the distinct tumor bacterial microbiome between subsolid and solid nodules. Increasing evidence suggests that reduced microbiome diversity is associated with poor survival of cancer patients [28,29]. Our findings that SSN tissues have higher microbiome diversity than SN is consistent with the fact that SSN is less invasive compared with SN and patients with SSN have longer recurrence-free survival [4]. Therefore, this study provides new clues to the indolent nature of SSN in the perspective of microbiome.

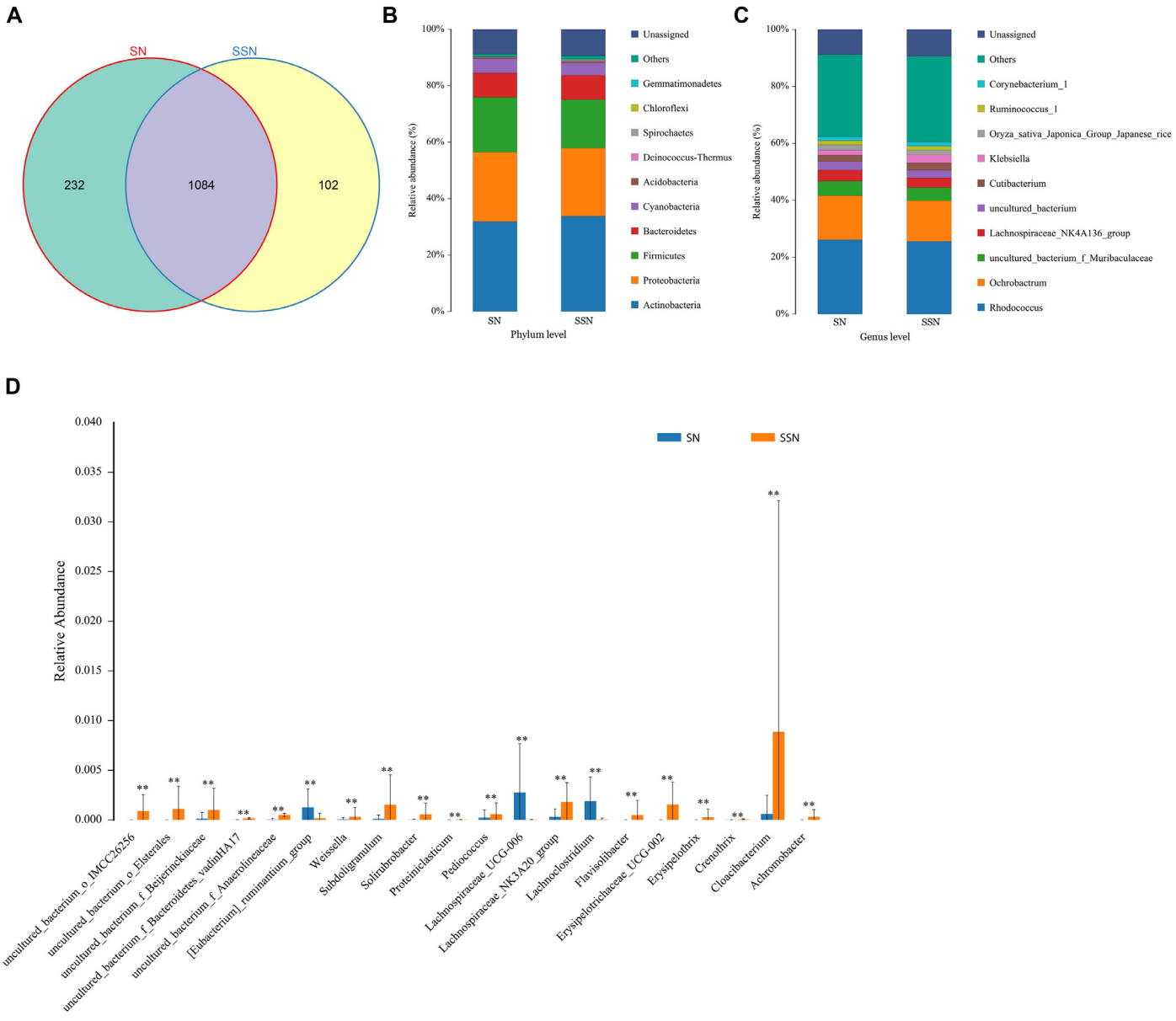

Fig. 3. Microbiome profiles of LUAD manifested as SSN and SN. (A) Different OTUs between SSN and SN. (B) Species composition at phylum level. (C) Species composition at genus level. (D) Top 20 differential genus between SSN and SN. $^{**}p < 0.01$. LUAD, lung adenocarcinoma; SSN, subsolid nodules; SN, solid nodules; OTUs: operational taxonomic units.

In addition, consistent with previous findings [16,30,31], we also found similar microbiome composition between LUAD tumor and normal tissues. Furthermore, microbiome signatures showed good performance to predict LUAD or SSN, which may provide a possibility and reliability for clinical transformation.

Most previous studies on the lung microbiota were based on bronchoalveolar lavage, bronchoscopic brushing, or sputum samples, and few directly characterized the microbiome in lung tumor tissues [16,32]. Increasing evidence revealed that alpha diversity is higher in non-malignant lung tissues than in tumor tissues [16,31]. Beta diversity, or overall microbiome composition is not significantly different between normal and tumor tissues. A study by Yu and colleagues demonstrated that *Proteobacteria, Firmicutes, Bacteroidetes*, and *Actinobacteria* were core microbiota in lung tissues at phylum level [33]. In this study, we also found a declining alpha diversity in LUAD tumor tissues, and core microbiota was similar to Yu's finding. Although overall microbiome composition and functional pathways were similar between tumor and normal tissues, we still found some differences. *Lactobacillus* was reported to show anti-cancer effects in various cancers [34,35]. *Lactobacillus* could exert direct cytotoxic effects on cancer cells and has various immunoregulatory functions [34]. Our results showed an obviously decreased abundance of *Lactobacillus* in LUAD tumor tissues. cGMP-PKG signaling pathway was known to regulate cell apoptosis and proliferation [36]. Several studies demonstrated that cGMP-PKG signaling pathway could suppress Wnt/β-catenin signaling to inhibit tumor cell growth [36–38]. In this study, we found that cGMP-PKG signaling pathway was significantly less abundant in LUAD tumor than normal tissues.

Lung adenocarcinomas manifested as radiological subsolid nodules were defined as a special clinical subtype [4]. SSN are considered to be more indolent and less aggressive than LUAD manifested as SN. The prevalence of lymphatic metastasis was lower in patients with SSN than in those with SN. In this study, we also found the lymph node metastasis or pleural invasion were more frequency in SN. However, the microbiome properties of SSN remain unclear. Only one study tried to char-

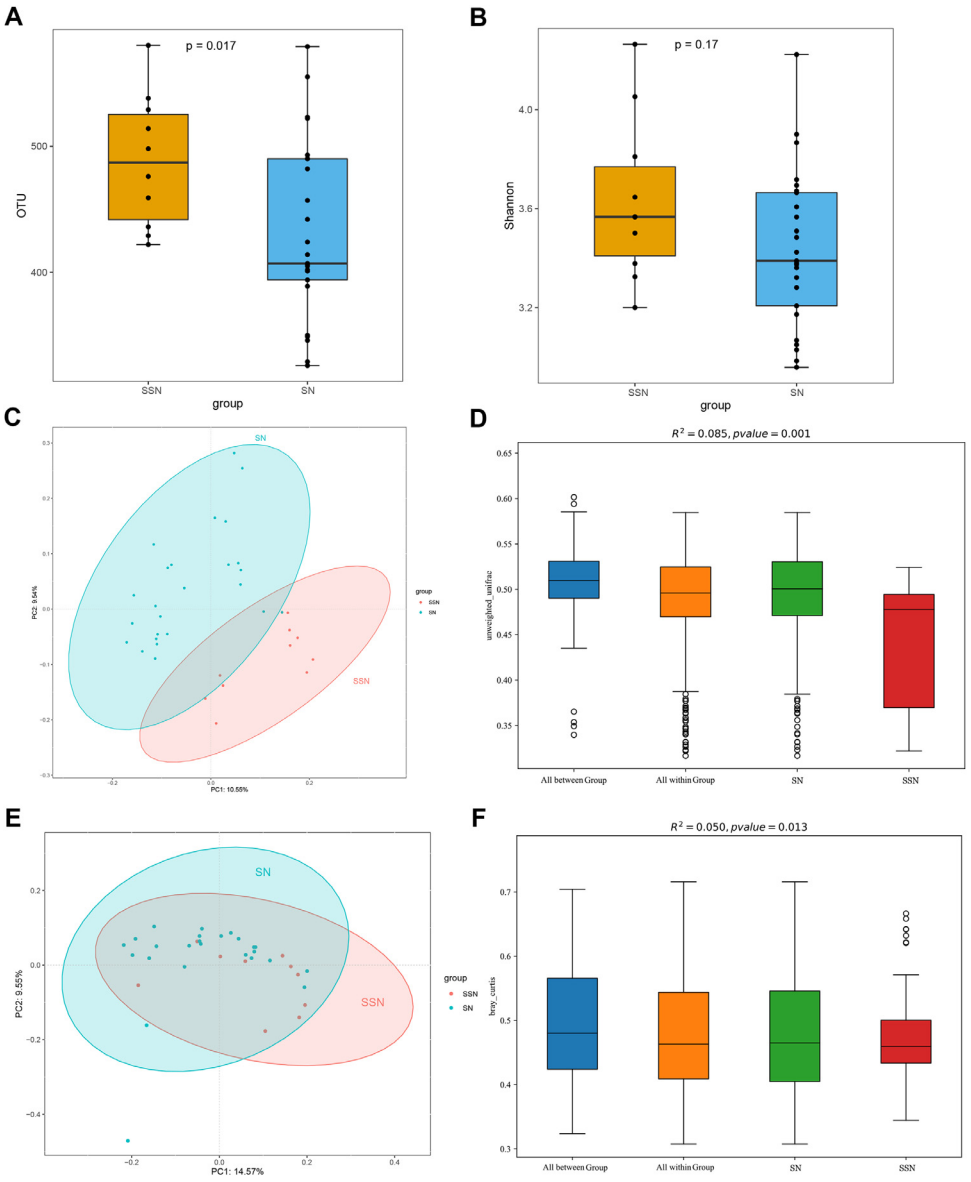

Fig. 4. Microbiome diversity between LUAD manifested as SSN and SN. (A) Richness was evaluated by the number of operational taxonomic units. (B) Shannon index of SSN and SN. (C) PCoA using unweighted unifrac method. (D) PERMANOVA analysis using unweighted unifrac method. (E) PCoA using Bray Curtis method. (F) PERMANOVA analysis using Bray Curtis method. LUAD, lung adenocarcinoma; SSN, subsolid nodules; SN, solid nodules; PCoA, Principal coordinate analysis.

acterize the microbiota in SSN [39]. They found that *Mycobacterium* was the core microbiota found in SSN, and two functional pathways, Secondary Metabolism and the Serine Threonine protein kinase were significantly increased in SSN over normal tissue samples. Nevertheless, they did not compare the microbiome between SSN and SN. In the present study, we, for the first time, characterized the different microbiome between SSN and SN. Higher diversity of SSN than SN suggests SSN is less malignant, which also supports the observation that SSN is more indolent than SN. PCoA and PERMANOVA showed significantly different overall microbiome composition between SSN and SN, which also provided evidence that SSN is a special subtype. Moreover, we found elevated *Chloroflexi* and *Gemmatimonadetes* in SSN group. Interestingly, increased genera, including *Cloacibacterium, Subdoligranulum*, and *Mycobacterium* were detected as potential biomarkers for SSN. *Mycobacterium tuberculosis* and *Mycobacterium avium* complex are considered to relate to lung cancer [40,41]. Another study found that 6.5% of patients with pulmonary nontuberculous mycobacterial disease, also had lung cancer [42]. These epidemiologic links are mechanically interpreted by chronic inflammation-associated carcinogenesis [43]. The persistent infection of mycobacterial organisms in the lung can induce a pro-inflammatory response to local tissues and then trigger the release of factors that may support the outgrowth of premalignant cells. Besides, *Cloacibacterium, Subdoligranulum* were also reported to relate to chronic inflammation [44,45]. And *Achromobacter* infection may mimic lung cancer [46]. These results indicated that microbiota in the local microenvironment may also contribute to the initiation and progression of SSN. We also found function differences between SSN and SN. Salivary secretion and gastric acid secretion were significantly enriched in SSN but not in SN, which means that SSN may contain more oral-derived microbiota. Furthermore, endocrine and other factor-regulated calcium reabsorption was different between SSN and SN. Calcium ion is an important second messenger within cells, which can participate in the development of tumors in various ways [47]. Thus, the abnormal absorption of calcium ions in cells may associate with tumorigenesis.

Recently, Poore et al. proposed a novel cancer diagnostic approach with preferable accuracy through microbiome analyses of blood and tissues [48]. They found unique microbial signatures in tissue and blood within and between most major types of cancer. Likewise, in current study, we identified microbial signatures to differentiate LUAD tumor and normal tissues, as well as SSN and SN. LEfSe analysis was used to identify potential markers and machine learning was applied to identify microbial signatures. Although the microbiome diversity between LUAD

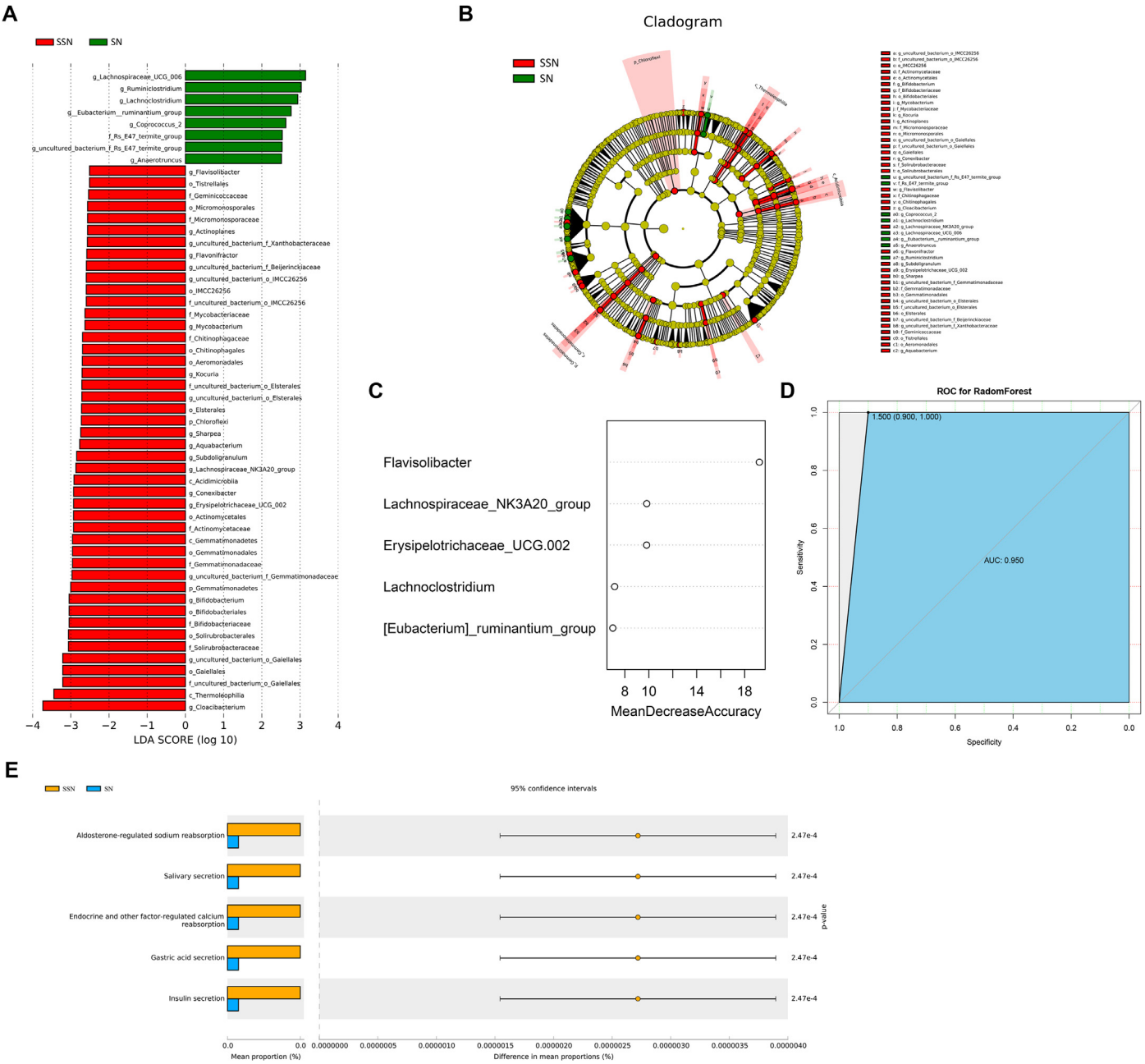

Fig. 5. Identification of biomarkers and functional prediction. (A) LEfSe analysis identified 54 markers at different levels between SSN and SN. The first letter indicates taxonomy level, p: phylum, c: class, o: order, f: family, g: genus. (B) Cladogram of bacterial lineages with significantly different representations in SSN and SN groups. (C) 5 genus signatures to construct the predictive model. (D) ROC curve for the random forest model. (E) Differential pathways at KEGG pathway level 3. SSN, subsolid nodules; SN, solid nodules; ROC, receiver operating characteristic; KEGG, Kyoto Encyclopedia of Genes and Genomes.

tumor and normal tissues were similar, some species had different abundances between these two groups. A 20 genera-based signature showed a good diagnostic accuracy to discriminate LUAD from normal tissues in both training and test sets. Meanwhile, a five genera-based signature can accurately discriminate SSN and SN. In addition, cross-validation of these two models results in reasonably high predictive powers as well. Our results indicated that microbiota may become a novel tool for cancer diagnostic.

We also recognize several limitations of our study. Although 16s rRNA gene sequencing is widely used for microbiota characterization, it has great limitations in species level analysis and functional prediction. Further study using metagenomic sequencing may provide a deeper understanding of the characteristics and function of microbiota in lung cancer. In addition, the number of patients and samples included are insufficient. Bacterial derived metabolites play important roles in tumor development and may strengthen the accuracy to diagnosis lung adenocarcinoma. However, we didnot make metabolome analysis in this study. Future studies with large sample sizes and more rigorous designs are expected to further comprehend microbiota in lung cancer.

5. Conclusions

To our knowledge, this is the first study that investigated tumor microbiome diversity between subsolid and solid lung adenocarcinoma. Early stage LUAD manifested as radiological SSN has elevated microbiome richness and diversity compared with SN, and the microbiome

composition of SSN is distinct from that of SN. Our finding is consistent with the fact that SSN is less invasive compared with SN. Therefore, this study provides new clues to the indolent nature of SSN in the perspective of microbiome. In addition, we found similar diversity and microbiome composition between LUAD tumor and normal tissues. We also uncover the microbial signatures associated with LUAD or SSN and develop accurate genera-based predictors for clinical diagnosis.

Funding

This study was supported by the National Natural Science Foundation of China (81871879, 81772469, and 61877001).

Authors' contributions

MQ, FY and GJ conceived and designed the study; YM, SW and SM conducted the experiments, collected and analyzed the data; YM, FY and MQ interpreted the data. YM, SW and MQ wrote and revised the manuscript.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2021.101050.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 68 (6) (2018) 394–424.
[2] J. Polanski, B. Jankowska-Polanska, J. Rosinczuk, M. Chabowski, A. Szymanska-Chabowska, Quality of life of patients with lung cancer, Onco Targets Ther. 9 (2016) 1023–1028.
[3] J.R. Molina, P. Yang, S.D. Cassivi, S.E. Schild, A.A. Adjei, Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship, Mayo Clin. Proc. 83 (5) (2008) 584–594.
[4] T. Ye, L. Deng, S. Wang, J. Xiang, Y. Zhang, H. Hu, Y. Sun, Y. Li, L. Shen, L. Xie, W. Gu, Y. Zhao, F. Fu, W. Peng, H. Chen, Y. Shen, Lung adenocarcinomas manifesting as radiological part-solid nodules define a special clinical subtype, J. Thorac. Oncol. 14 (10) (2019) S546–S547.
[5] L. Fan, Y. Wang, Y. Zhou, Q. Li, W. Yang, S. Wang, F. Shan, X. Zhang, J. Shi, W. Chen, S.Y. Liu, Lung cancer screening with low-dose CT: baseline screening results in Shanghai, Acad. Radiol. 26 (10) (2019) 1283–1291.
[6] Y.J. Ren, S.J. Huang, C.Y. Dai, D. Xie, L. Zheng, H.K. Xie, H. Zheng, Y.L. She, F.Y. Zhou, Y. Wang, P.P. Li, K. Fei, G.N. Jiang, Y. Zhang, B. Su, E.A. Sweet-Cordero, N.L. Tran, Y.N. Yang, J.N. Patei, C. Rolfo, G. Rocco, A.F. Cardona, A. Tuzi, M.B. Suter, P. Yang, W.N. Xu, C. Chen, Germline predisposition and copy number alteration in pre-stage lung adenocarcinomas presenting as ground-glass nodules, Front Oncol. 9 (2019) 288.
[7] M. Gomes, A.L. Teixeira, A. Coelho, A. Araujo, R. Medeiros, The role of inflammation in lung cancer, Inflamm. Cancer 816 (2014) 1–23.
[8] E.A. Engels, Inflammation in the development of lung cancer: epidemiological evidence, Expert Rev. Anticancer 8 (4) (2008) 605–615.
[9] J. Vangipurapu, L.F. Silva, T. Kuulasmaa, U. Smith, M. Laakso, Microbiota-related metabolites and the risk of type 2 diabetes, Diabetes Care 43 (6) (2020) 1319–1325.
[10] R.X. Liu, J. Hong, X.Q. Xu, Q. Feng, D.Y. Zhang, Y.Y. Gu, J. Shi, S.Q. Zhao, W. Liu, X.K. Wang, H.H. Xia, Z.P. Liu, B. Cui, P.W. Liang, L.Q. Xi, J.B. Jin, X.Y. Ying, X.L. Wang, X.J. Zhao, W.Y. Li, H.J. Jia, Z. Lan, F.Y. Li, R. Wang, Y.K. Sun, M.L. Yang, Y.X. Shen, Z.Y. Jie, J.H. Li, X.M. Chen, H.Z. Zhong, H.L. Xie, Y.F. Zhang, W.Q. Gu, X.X. Deng, B.Y. Shen, X. Xu, H.M. Yang, G.W. Xu, Y.F. Bi, S.H. Lai, J. Wang, L. Qi, L. Madsen, J.Q. Wang, G. Ning, K. Kristiansen, W.Q. Wang, Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention, Nat. Med. 23 (7) (2017) 859.
[11] B. Routy, E. Le Chatelier, L. Derosa, C.P.M. Duong, M.T. Alou, R. Daillere, A. Fluckiger, M. Messaoudene, C. Rauber, M.P. Roberti, M. Fidelle, C. Flament, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragon, N. Jacquelot, B. Qu, G. Ferrere, C. Clemenson, L. Mezquita, J.R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.C. Soria, E. Deutsch, Y. Loriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M.D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors, Science 359 (6371) (2018) 91.
[12] K. Mima, S. Nakagawa, H. Sawayama, T. Ishimoto, K. Imai, M. Iwatsuki, D. Hashimoto, Y. Baba, Y. Yamashita, N. Yoshida, A. Chikamoto, H. Baba, The microbiome and hepatobiliary-pancreatic cancers, Cancer Lett. 402 (2017) 9–15.
[13] J.L. Nugent, A.N. McCoy, C.J. Addamo, W. Jia, R.S. Sandler, T.O. Keku, Altered tissue metabolites correlate with microbial dysbiosis in colorectal adenomas, J. Proteome Res. 13 (4) (2014) 1921–1929.
[14] R. Bingula, M. Filaire, N. Radosevic-Robin, M. Bey, J.Y. Berthon, A. Bernalier–Donadille, M.P. Vasson, E. Filaire, Desired turbulence? gut-lung axis, immunity, and Lung cancer, J. Oncol. 2017 (2017).
[15] A.G. Ramirez-Labrada, D. Isla, A. Artal, M. Arias, A. Rezusta, J. Pardo, E.M. Galvez, The influence of lung microbiota on lung carcinogenesis, immunity, and immunotherapy, Trends Cancer 6 (2) (2020) 86–97.
[16] Q.X. Mao, F. Jiang, R. Yin, J. Wang, W.J. Xia, G.C. Dong, W.D. Ma, Y. Yang, L. Xu, J.Z. Hu, Interplay between the lung microbiome and lung cancer, Cancer Lett. 415 (2018) 40–48.
[17] F.C. Detterbeck, V. Bolejack, D.A. Arenberg, J. Crowley, J.S. Donington, W.A. Franklin, N. Girard, E.M. Marom, P.J. Mazzone, A.G. Nicholson, V.W. Rusch, L.T. Tanoue, W.D. Travis, H. Asamura, R. Rami-Porta, I.S.P.F. C, A. Boards, M.P.S. Workgrp, The IASLC lung cancer staging project: background data and proposals for the classification of lung cancer with separate tumor nodules in the forthcoming eighth edition of the TNM classification for lung cancer, J. Thorac. Oncol. 11 (5) (2016) 681–692.
[18] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114–2120.
[19] T. Magoc, S.L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies, Bioinformatics 27 (21) (2011) 2957–2963.
[20] R.C. Edgar, B.J. Haas, J.C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection, Bioinformatics 27 (16) (2011) 2194–2200.
[21] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Tumbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 7 (5) (2010) 335–336.
[22] Q. Wang, G.M. Garrity, J.M. Tiedje, J.R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, Appl. Environ. Microb. 73 (16) (2007) 5261–5267.
[23] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glockner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic Acids Res. 41 (D1) (2013) D590–D596.
[24] N. Segata, J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W.S. Garrett, C. Huttenhower, Metagenomic biomarker discovery and explanation, Genome Biol. 12 (6) (2011).
[25] G.M. Douglas, R.G. Beiko, M.G.I. Langille, Predicting the functional potential of the microbiome from marker genes using PICRUSt, Methods Mol. Biol. 1849 (2018) 169–177.
[26] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi, KEGG for linking genomes to life and the environment, Nucleic Acids Res. 36 (2008) D480–D484.
[27] D.H. Parks, G.W. Tyson, P. Hugenholtz, R.G. Beiko, STAMP: statistical analysis of taxonomic and functional profiles, Bioinformatics 30 (21) (2014) 3123–3124.
[28] Y.P. Jin, H. Dong, L.L. Xia, Y. Yang, Y.Q. Zhu, Y. Shen, H.J. Zheng, C.C. Yao, Y. Wang, S. Lu, The diversity of gut microbiome is associated with favorable responses to anti-programmed death 1 immunotherapy in Chinese patients with NSCLC, J. Thorac. Oncol. 14 (8) (2019) 1378–1389.
[29] E. Riquelme, Y. Zhang, L.L. Zhang, M. Montiel, M. Zoltan, W.L. Dong, P. Quesada, I. Sahin, V. Chandra, A. San Lucas, P. Scheet, H.W. Xu, S.M. Hanash, L. Feng, J.K. Burks, K.A. Do, C.B. Peterson, D. Nejman, C.W.D. Tzeng, M.P. Kim, C.L. Sears, N. Ajami, J. Petrosino, L.D. Wood, A. Maitra, R. Straussman, M. Katz, J.R. White, R. Jenq, J. Wargo, F. McAllister, Tumor microbiome diversity and composition influence pancreatic cancer outcomes, Cell 178 (4) (2019) 795.
[30] R. Bingula, E. Filaire, I. Molnar, E. Delmas, J.Y. Berthon, M.P. Vasson, A. Bernalier–Donadille, M. Filaire, Characterization of microbiota in saliva, bronchoalveolar lavage fluid, non-malignant, peritumoural and tumor tissue in non-small cell lung cancer patients: a cross-sectional clinical trial, Respir. Res. 21 (1) (2020) 129.
[31] B.A. Peters, R.B. Hayes, C. Goparaju, C. Reid, H.I. Pass, J. Ahn, The microbiome in lung cancer tissue and recurrence-free survival, Cancer Epidemiol. Biomark. 28 (4) (2019) 731–740.
[32] S.K. Patnaik, E.G. Cortes, E.D. Kannisto, A. Punnanitinont, S.S. Dhillon, S. Liu, S. Yendamuri, Lower airway bacterial microbiome may influence recurrence after resection of early-stage non-small cell lung cancer, J. Thorac. Cardiovasc. Surg. 161 (2) (2021) 419–429 e16.
[33] G. Yu, M.H. Gail, D. Consonni, M. Carugno, M. Humphrys, A.C. Pesatori, N.E. Caporaso, J.J. Goedert, J. Ravel, M.T. Landi, Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features, Genome Biol. 17 (1) (2016) 163.
[34] A. Abedin-Do, Z. Taherian-Esfahani, S. Ghafouri-Fard, S. Ghafouri-Fard, E. Motevaseli, Immunomodulatory effects of Lactobacillus strains: emphasis on their effects on cancer cells, Immunotherapy 7 (12) (2015) 1307–1329.
[35] G.L. Banna, F. Torino, F. Marletta, M. Santagati, R. Salemi, E. Cannarozzo, L. Falzone, F. Ferrau, M. Libra, Lactobacillus rhamnosus GG: an overview to explore the rationale of its use in cancer, Front Pharmacol. 8 (2017) 603.
[36] K. Lee, A.P. G, The interaction between the Wnt/beta-catenin signaling cascade and PKG activation in cancer, J. Biomed. Res. 31 (3) (2017) 189–196.
[37] D.D. Browning, I.K. Kwon, R. Wang, cGMP-dependent protein kinases as potential

targets for colon cancer prevention and treatment, Future Med. Chem. 2 (1) (2010) 65–80.
[38] N. Li, Y. Xi, H.N. Tinsley, E. Gurpinar, B.D. Gary, B. Zhu, Y. Li, X. Chen, A.B. Keeton, A.H. Abadi, M.P. Moyer, W.E. Grizzle, W.C. Chang, M.L. Clapper, G.A. Piazza, Sulindac selectively inhibits colon tumor cell growth by activating the cGMP/PKG pathway to suppress Wnt/β-catenin signaling, Mol. Cancer Ther. 12 (9) (2013) 1848–1859.
[39] Y.J. Ren, H. Su, Y.L. She, C.Y. Dai, D. Xie, S. Narrandes, S.J. Huang, C. Chen, W.N. Xu, Whole genome sequencing revealed microbiome in lung adenocarcinomas presented as ground-glass nodules, Transl. Lung Cancer R 8 (3) (2019) 235.
[40] V. Pilaniya, K. Gera, S. Kunal, A. Shah, Pulmonary tuberculosis masquerading as metastatic lung disease, Eur. Respir. Rev. 25 (139) (2016) 97–98.
[41] L. Lande, D.D. Peterson, R. Gogoi, G. Daum, K. Stampler, R. Kwait, C. Yankowski, K. Hauler, J. Danley, K. Sawicki, J. Sawicki, Association between pulmonary mycobacterium avium complex infection and lung cancer, J. Thorac. Oncol. 7 (9) (2012) 1345–1351.
[42] T.K. Marras, C.L. Daley, Epidemiology of human pulmonary infection with nontuberculous mycobacteria, Clin. Chest Med. 23 (3) (2002) 553–567.
[43] H.Y. Liang, X.L. Li, X.S. Yu, P. Guan, Z.H. Yin, Q.C. He, B.S. Zhou, Facts and fiction of the relationship between preexisting tuberculosis and lung cancer risk: a systematic review, Int. J. Cancer 125 (12) (2009) 2936–2944.
[44] A. Hirano, J. Umeno, Y. Okamoto, H. Shibata, Y. Ogura, T. Moriyama, T. Torisu, S. Fujioka, Y. Fuyuno, Y. Kawarabayasi, T. Matsumoto, T. Kitazono, M. Esaki, Comparison of the microbial community structure between inflamed and non-inflamed sites in patients with ulcerative colitis, J. Gastroenterol. Hepatol. (2018).
[45] J. Sundin, I. Rangel, D. Repsilber, R.J. Brummer, Cytokine response after stimulation with key commensal bacteria differ in post-infectious irritable bowel syndrome (PI-IBS) patients compared to healthy controls, PLoS One 10 (9) (2015) e0134836.
[46] S.L. Claassen, J.M. Reese, V. Mysliwiec, S.D. Mahlen, Achromobacter xylosoxidans infection presenting as a pulmonary nodule mimicking cancer, J. Clin. Microbiol. 49 (7) (2011) 2751–2754.
[47] G.R. Monteith, N. Prevarskaya, S.J. Roberts-Thomson, The calcium-cancer signalling nexus, Nat. Rev. Cancer 17 (6) (2017) 367–380.
[48] G.D. Poore, E. Kopylova, Q.Y. Zhu, C. Carpenter, S. Fraraccio, S. Wandro, T. Kosciolek, S. Janssen, J. Metcalf, S.J. Song, J. Kanbar, S. Miller-Montgomery, R. Heaton, R. Mckay, S.P. Patel, A.D. Swafford, R. Knight, Microbiome analyses of blood and tissues suggest cancer diagnostic approach, Nature 579 (7800) (2020) 567.