Proceedings of the 6th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/2026.37943

Human Tumor Microbiomes are Site-Specific Across Five Body Locations

Haoxuan Deng

The High School Affiliated to Renmin University of China, Beijing, China
Haoxuandeng0822@163.com

Abstract. Tumor-associated microbial communities may influence cancer development, yet their patterns across anatomical sites remain incompletely characterized. The analysis used 424 samples from The Cancer Microbiome Atlas (TCMA), including tumors from the colon, esophagus, head/neck, rectum, and stomach, along with healthy tissues. Comparisons covered microbial abundance, alpha and beta diversity, core taxa, and differential abundance. Random forest models were used to assess classification. Tumor and healthy samples had similar overall community profiles, with no significant difference in Shannon diversity and little beta-diversity variation explained by tumor status. Anatomical site accounted for more variation than any other variable examined. Colon and rectum showed similar communities, while esophagus, head/neck, and stomach formed another group. Diversity measures, core and differential taxa, and predictive models supported this division. These results suggest that anatomical location has a stronger association with microbial composition than tumor status and should be considered when developing microbial biomarkers.

Keywords: tumor microbiome, body-site specificity, cancer, random forest

1. Introduction

More than 19 million new cancer cases are reported worldwide each year [1], yet early detection remains difficult for many cancer types. This has encouraged interest in additional biomarkers that reflect tumor biology. The presence of microorganisms in several cancer types suggests that microbial signatures may provide information useful for this purpose [2].

Microbial effects on cancer can involve immune and metabolic pathways within the tumor microenvironment [3]. Studies of treatment response have linked Fusobacterium nucleatum to chemoresistance in colorectal cancer [4] and commensal microbiome composition to anti-PD-1 response in melanoma [5]. For biomarker development, the consistency of such signals across cohorts is a concern. Detection of F. nucleatum differs considerably across colorectal cancer datasets, reflecting both cohort and technical effects [6]. This variability makes validation central to the assessment of microbial biomarkers, as emphasized in large cross-cohort analyses [7].

Even when a microbial signature is associated with cancer, its interpretation may depend on anatomical location and disease stage. Cross-cohort studies and meta-analyses have identified colorectal cancer-associated signatures [7, 8], with additional differences between disease stages [9].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.37943

Pan-cancer work has also described microbial ecologies specific to cancer types [2]. These observations make it useful to compare several sites directly and assess how much variation is associated with tumor status relative to anatomical location.

Using TCMA data from five cancer sites, this study compared tumor (PT) and healthy (STN) communities and examined factors associated with beta diversity. Shared and site-specific microbial patterns were then assessed for their value in microbiome-based classification.

2. Materials and methods

The analysis proceeded in three stages: comparison of PT and STN communities, evaluation of variables associated with beta-diversity structure, and cross-site comparison of the five tumor groups. Random forest models were then used to assess the predictive value of the observed microbial patterns.

2.1. Data collection

Raw microbiome data were obtained from The Cancer Microbiome Atlas (TCMA; https://research.repository.duke.edu/concern/datasets/4x51hk100?locale=en). The repository contained 424 samples: 93 from colon, 41 from esophagus, 143 from head/neck, 40 from rectum, 52 from stomach, and 55 healthy samples (STN).

2.2. Statistical analysis

2.2.1. Abundance profiling

MicrobiomeAnalyst was used to summarize microbial abundance across tumor and healthy samples, with Hiplot used for visualization.

2.2.2. Alpha diversity

Alpha diversity was quantified with the Shannon index, which incorporates richness and evenness. Group differences were assessed with Welch's t-test or ANOVA in MicrobiomeAnalyst and visualized as boxplots in Hiplot.

2.2.3. Beta diversity

Beta diversity was evaluated using Bray-Curtis distances and principal coordinates analysis (PCoA). PERMANOVA was used to test group differences and compare the explanatory strength of candidate variables through p-values and R-squared values.

2.2.4. Core microbiome analysis

Core microbiome analysis in MicrobiomeAnalyst identified taxa that were widespread and relatively stable across the five PT groups and the STN group.

2.2.5. Linear discriminant analysis effect size (LEfSe)

Linear discriminant analysis effect size (LEfSe) was used to identify differentially abundant taxa with a log LDA threshold of 2. Hiplot was used to generate heatmaps and overlap visualizations.

DOI: 10.54254/2753-8818/2026.37943

2.2.6. Prediction model

Random forest classifiers were implemented in JupyterLab to distinguish PT from STN samples and to evaluate the proposed cancer-site groupings. Performance was assessed with ROC curves and area under the curve (AUC).

3. Results

The dataset comprised 424 samples: 369 tumor samples across five cancer sites and 55 healthy samples. The full distribution is shown in Table 1.

3.1. Differences between PT and STN microbial communities

Table 1. Sample distribution of healthy (STN) and tumor (PT) tissues across the five cancer sites

Variable	OverallN = 424	ColonN = 93	Esophagus N = 41	Head/NeckN = 143	RectumN = 40	StomachN = 52	STNN = 55
Sample							
PT	369 (87.0%)	93 (100%)	41 (100%)	143 (100%)	40 (100%)	52 (100%)	0 (0%)
STN	55 (13.0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	55 (100%)

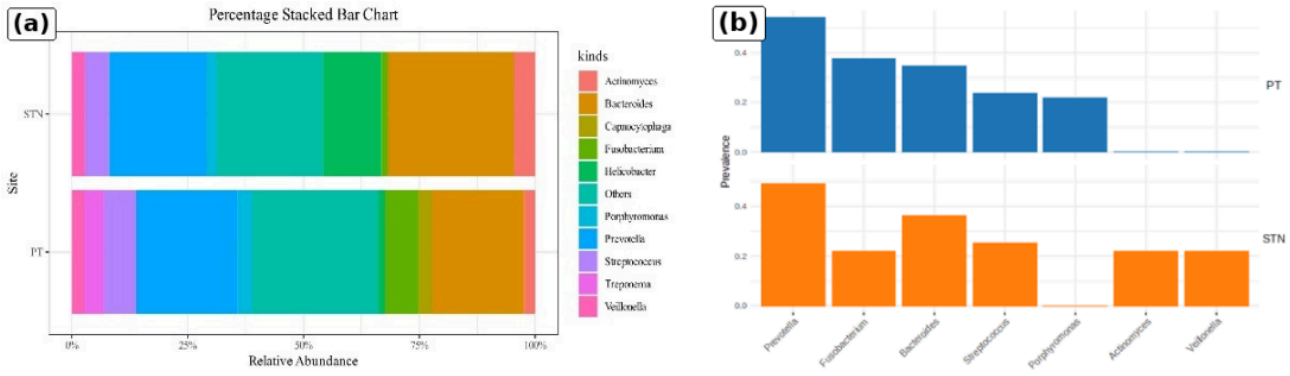


Figure 1. Comparison of microbiome composition between healthy and tumor samples: (a) relative abundance profiles; (b) core-taxon prevalence profiles

PT and STN samples showed broadly similar dominant microbial compositions (Figure 1). They also shared several core taxa, and Shannon diversity did not differ significantly between groups.

Beta diversity differed statistically but weakly (R-squared = 0.006, p = 0.018), indicating that tumor status explained only 0.6% of the observed community variation. Prevotella, Fusobacterium, and Bacteroides were among the dominant and core taxa in tumor samples. Overall, the PT-STN separation was limited.

DOI: 10.54254/2753-8818/2026.37943

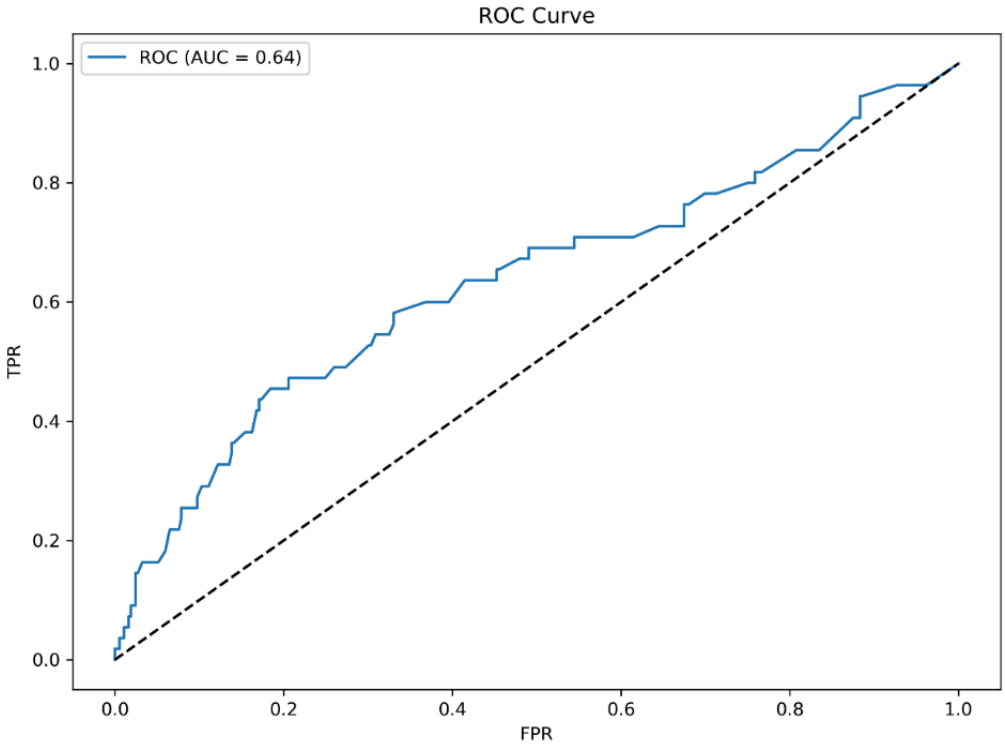


Figure 2. ROC curve generated using a random forest model to distinguish healthy from tumor samples

The random forest model for PT-STN classification produced an AUC of 0.64 (Figure 2). This moderate performance was consistent with the limited compositional separation observed in the diversity analyses.

3.2. Site of disease as the strongest explanatory factor

Table 2. R-squared values for beta diversity under different experimental factors

Experimental factor	R-squared	p-value
site_of_disease	0.24389	0.001
percent_normal_cells	0.069581	0.001
percent tumor cells	0.043875	0.001
percent tumor_nuclei	0.039178	0.001
oct_embedded	0.031842	0.001
percent stromal cells	0.031515	0.001
percent_necrosis	0.03063	0.001
necrosis_percent	0.026664	0.001
shortest dimension	0.022263	0.001
percent_monocyte_infltration	0.020555	0.001

DOI: 10.54254/2753-8818/2026.37943

Table 2. (continued)

percent_neutrophil_infltration	0.013939	0.065
percent lymphocyte infiltration	0.012917	0.095
weight	0.01242	0.11
intermediate dimension	0.011674	0.164
longest_dimension	0.0094574	0.126
tumor_nuclei_percent	0.0075258	0.373

PERMANOVA was used to compare the contribution of site of disease, tumor dimensions, weight, and other sample characteristics to beta-diversity variation (Table 2).

Site of disease showed the largest R-squared value (0.244, p = 0.001), whereas all other variables explained substantially less variation. Anatomical site was therefore the strongest measured determinant of tumor microbiome structure.

3.3. Shared and specific microbial characteristics across cancer groups

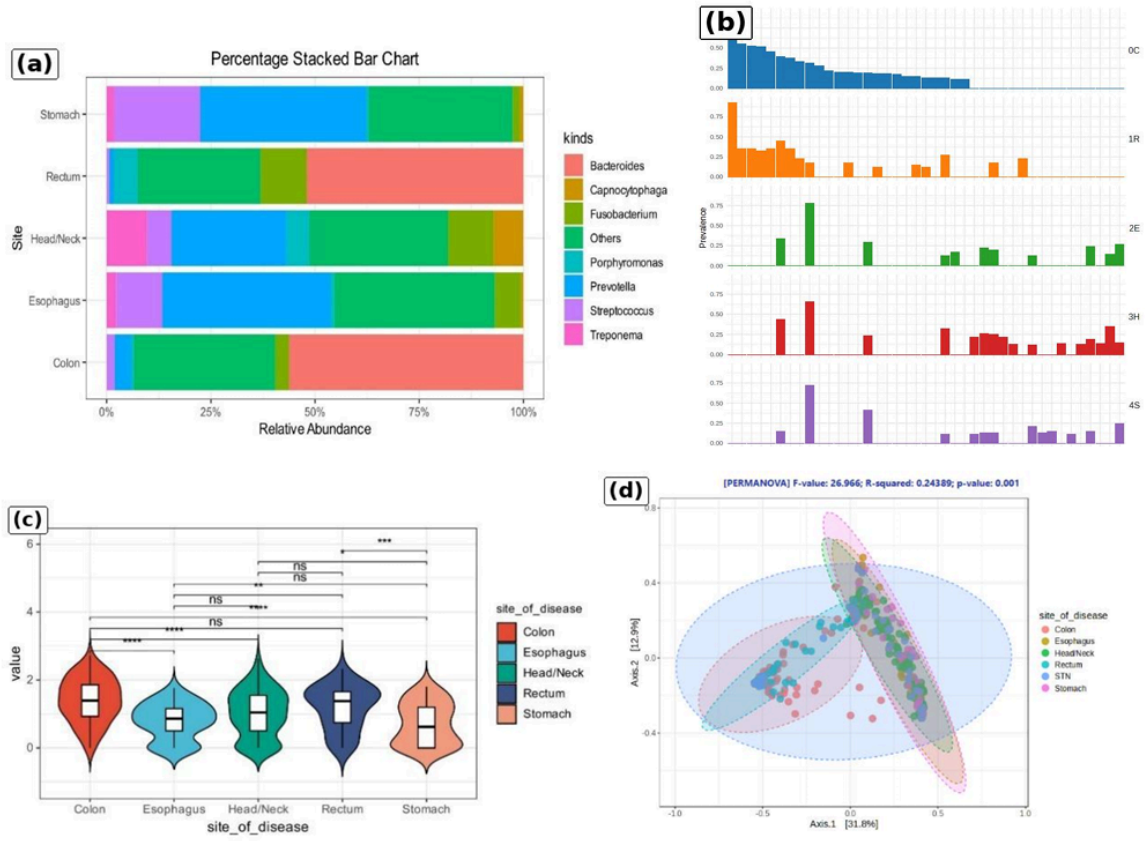


Figure 3. Comparison of microbiome composition among five cancer sites: (a) relative abundance profiles; (b) taxon prevalence profiles; (c) Shannon diversity; (d) PCoA beta diversity

Because anatomical site was the strongest explanatory factor, the five tumor groups were compared directly using beta diversity, alpha diversity, taxonomic composition, and differential abundance (Figure 3; Tables 3-4).

DOI: 10.54254/2753-8818/2026.37943

Table 3. Pairwise R-squared values across the five cancer sites

Pair	R-squared	P-value
Colon vs Stomach	0.33576	0.001
Colon vs Esophagus	0.32738	0.001
Esophagus vs Rectum	0.30674	0.001
Rectum vs Stomach	0.30024	0.001
Colon vs Head/Neck	0.26288	0.001
Head/Neck vs Rectum	0.16535	0.001
Head/Neck vs Stomach	0.036447	0.001
Esophagus vs Head/Neck	0.020543	0.002
Colon vs Rectum	0.01686	0.024
Esophagus vs Stomach	0.013525	0.266

Table 4. Mean Shannon diversity values across the five cancer sites

Cancer site	Mean Shannon index
Colon	1.3977
Rectum	1.1795
Head/Neck	1.0106
Esophagus	0.7894
Stomach	0.7155

Pairwise beta-diversity results revealed two broad similarity groups. Colon and rectum were closely related (R-squared = 0.0169), while esophagus, head/neck, and stomach showed relatively small pairwise differences, especially esophagus versus stomach and esophagus versus head/neck (Table 3).

Shannon diversity supported the same pattern: colon and rectum had higher mean diversity, whereas esophagus, head/neck, and stomach had lower values (Table 4). Dominant and core taxa were also more similar within the two proposed groups.

Differential-abundance clustering further separated the sites into colon-rectum and esophagus-head/neck-stomach groups (Figure 4). Substantial within-group overlap of differential taxa provided additional support for this classification.

DOI: 10.54254/2753-8818/2026.37943

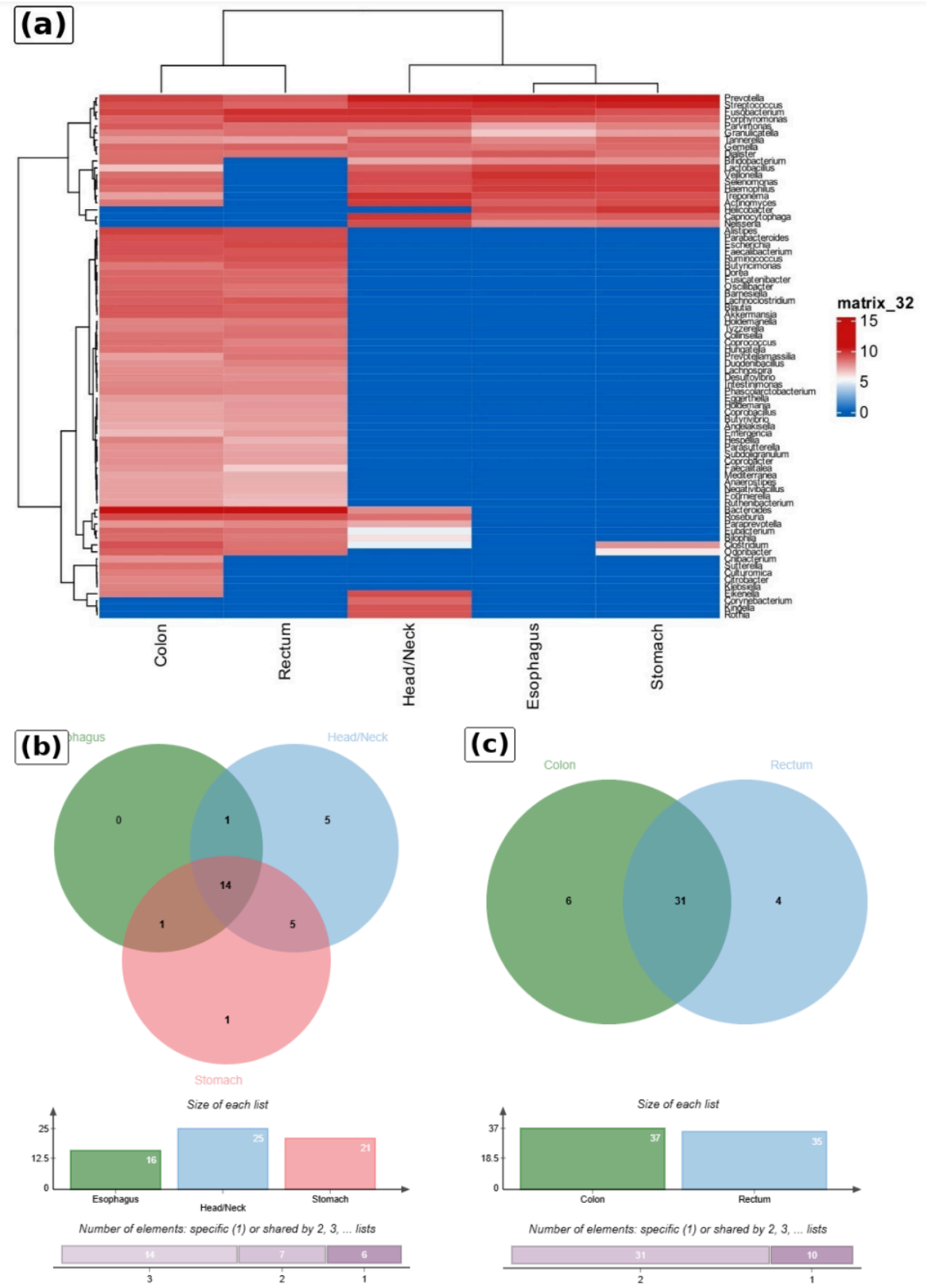


Figure 4. Differential microbial composition across five cancer sites: (a) differential-abundance heatmap; (b) overlap among esophagus, head/neck, and stomach; (c) overlap between colon and rectum

DOI: 10.54254/2753-8818/2026.37943

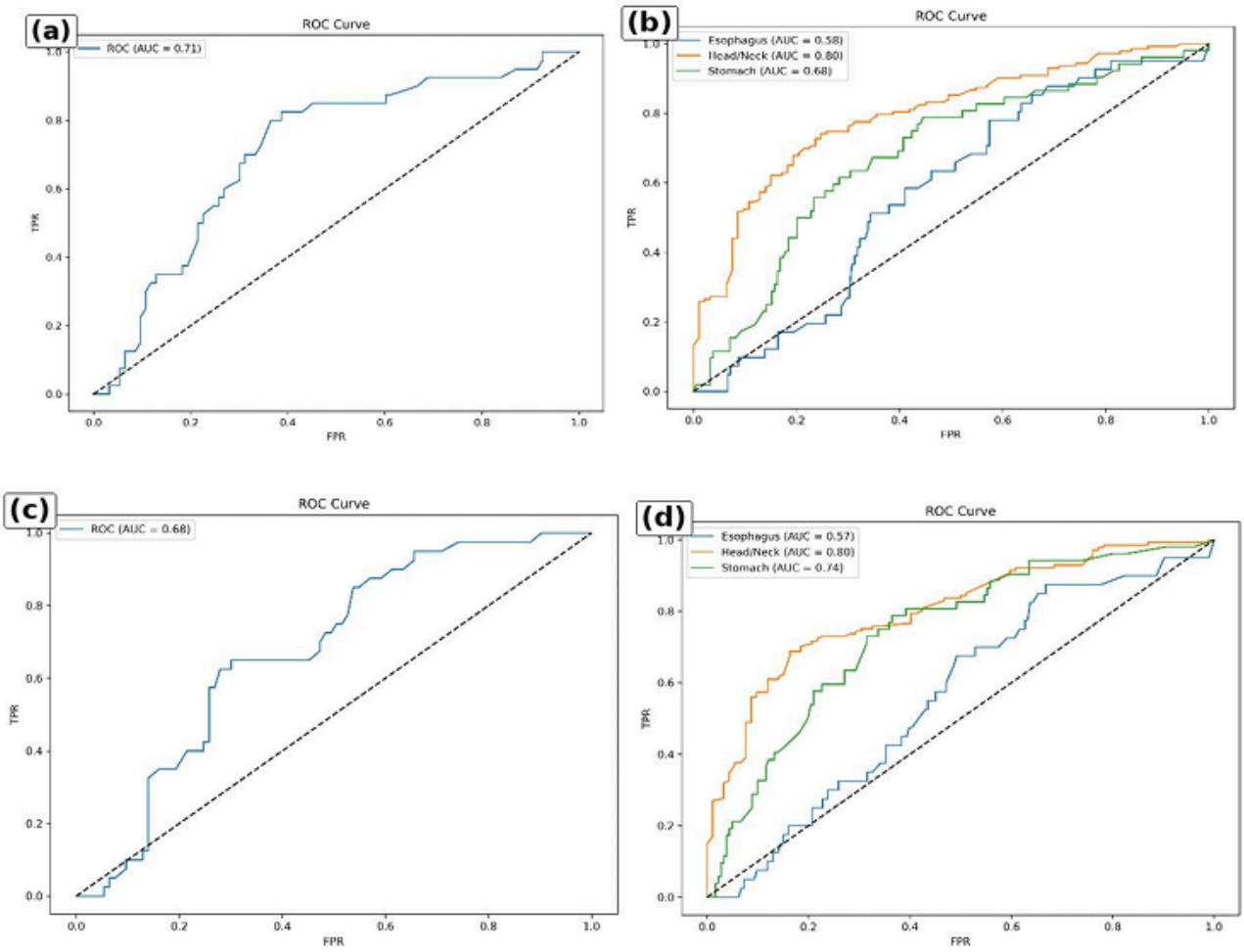


Figure 5. ROC curves generated using a random forest model to classify the two proposed cancer-site groups: (a) all microbial features for the colon-rectum group; (b) all microbial features for the esophagus-head/neck-stomach group; (c) differentially abundant features for the colon-rectum group; and (d) differentially abundant features for the esophagus-head/neck-stomach group

Random forest models were then used to test whether microbial features could recover the proposed groupings (Figure 5). Models using all taxa showed satisfactory discrimination, and feature filtering retained a smaller set of differential markers while improving interpretability. Taken together, the diversity, taxonomic, differential-abundance, and predictive analyses consistently supported the same two-group structure.

4. Discussion

Three findings emerged from the TCMA analysis. First, PT and STN communities differed only modestly: alpha diversity was not significant, and tumor status explained 0.6% of beta-diversity variation. Second, anatomical site accounted for the most variation among the factors tested. Third, the analyses consistently grouped colon with rectum and esophagus with head/neck and stomach.

Colon and rectum share conditions that could favor similar microbial communities, including anaerobic environments, near-neutral pH, and relatively slow transit [9-11]. Their grouping is

DOI: 10.54254/2753-8818/2026.37943

therefore biologically plausible, and the higher Shannon diversity at both sites is consistent with this shared ecology.

For esophagus, head/neck, and stomach, local ecology is another possible explanation for the observed similarity. Work on the small intestine shows how rapid uptake and conversion of simple carbohydrates can shape the microbiota [12]. Distinct bacterial communities have been described in the distal esophagus [13] and stomach [14], while oral microbiome studies relate local ecology to oral and gastrointestinal cancers [15]. These observations suggest that anatomical environments could contribute to the grouping found here.

Because the STN group was much smaller than the PT group and lacked healthy samples matched to each cancer site, sampling differences may confound the tumor-versus-healthy comparison. Interpretation at individual sites is limited for the same reason. The proposed groups were assessed in just one dataset and have not been clinically validated.

Independent cohorts with matched healthy tissues and clinical covariates are needed to test these patterns. Further analyses should establish whether smaller microbial marker sets retain predictive value across datasets.

5. Conclusion

Analysis of 424 TCMA samples showed that anatomical site explained substantially more microbiome variation than tumor status. PT and STN samples were only weakly separated, whereas the five tumor sites formed a reproducible two-group pattern: colon with rectum, and esophagus with head/neck and stomach. This structure was supported by alpha and beta diversity, core and differential taxa, and random forest modeling. The findings therefore support a site-specific view of tumor-associated microbiomes and provide a focused basis for future validation of microbial biomarkers across anatomical locations.

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249. https://doi.org/10.3322/caac.21660

[2] Narunsky-Haziza, L., Sepich-Poore, G. D., Livyatan, I., Asraf, O., Martino, C., Nejman, D., Gavert, N., Stajich, J. E., Amit, G., González, A., Wandro, S., Perry, G., Ariel, R., Meltser, A., Shaffer, J. P., Zhu, Q., Balint-Lahat, N., Barshack, I., Dadiani, M., Gal-Yam, E. N., Patel, S. P., Bashan, A., Swafford, A. D., Pilpel, Y., Knight, R., & Straussman, R. (2022). Pan-cancer analyses reveal cancer-type-specific fungal ecologies and bacteriome interactions. *Cell, 185*(20), 3789–3806.e17. https://doi.org/10.1016/j.cell.2022.09.005

[3] Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., Meltser, A., Douglas, G. M., Kamer, I., Gopalakrishnan, V., Dadosh, T., Levin-Zaidman, S., Avnet, S., Atlan, T., Cooper, Z. A., Arora, R., … Straussman, R. (2020). The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science (New York, N.Y.), 368*(6494), 973–980. https://doi.org/10.1126/science.aay9189

[4] Yu, T., Guo, F., Yu, Y., Sun, T., Ma, D., Han, J., Qian, Y., Kryczek, I., Sun, D., Nagarsheth, N., Chen, Y., Chen, H., Hong, J., Zou, W., & Fang, J. Y. (2017). Fusobacterium nucleatum promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell, 170*(3), 548–563.e16. https://doi.org/10.1016/j.cell.2017.07.008

[5] Matson, V., Fessler, J., Bao, R., Chongsuwat, T., Zha, Y., Alegre, M. L., Luke, J. J., & Gajewski, T. F. (2018). The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science (New York, N.Y.), 359*(6371), 104–108. https://doi.org/10.1126/science.aao3290

[6] Ito, M., Kanno, S., Nosho, K., Sukawa, Y., Mitsuhashi, K., Kurihara, H., Igarashi, H., Takahashi, T., Tachibana, M., Takahashi, H., Yoshii, S., Takenouchi, T., Hasegawa, T., Okita, K., Hirata, K., Maruyama, R., Suzuki, H., Imai, K., Yamamoto, H., & Shinomura, Y. (2015). Association of Fusobacterium nucleatum with clinical and molecular

DOI: 10.54254/2753-8818/2026.37943

features in colorectal serrated pathway. *International Journal of Cancer, 137*(6), 1258–1268. https://doi.org/10.1002/ijc.29488

[7] Thomas, A. M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., Pozzi, C., Gandini, S., Serrano, D., Tarallo, S., Francavilla, A., Gallo, G., Trompetto, M., Ferrero, G., Mizutani, S., Shiroma, H., Shiba, S., … Segata, N. (2019). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nature Medicine, 25*(4), 667–678. https://doi.org/10.1038/s41591-019-0405-7

[8] Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J. S., Voigt, A. Y., Palleja, A., Ponnudurai, R., Sunagawa, S., Coelho, L. P., Schrotz-King, P., Vogtmann, E., Habermann, N., Niméus, E., Thomas, A. M., Manghi, P., Gandini, S., Serrano, D., … Zeller, G. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine, 25*(4), 679–689. https://doi.org/10.1038/s41591-019-0406-6

[9] Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., Watanabe, H., Masuda, K., Nishimoto, Y., Kubo, M., Hosoda, F., Rokutan, H., Matsumoto, M., Takamaru, H., Yamada, M., Matsuda, T., Iwasaki, M., Yamaji, T., Yachida, T., Soga, T., … Yamada, T. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine, 25*(6), 968–976. https://doi.org/10.1038/s41591-019-0458-7

[10] Donaldson, G. P., Lee, S. M., & Mazmanian, S. K. (2016). Gut biogeography of the bacterial microbiota. *Nature Reviews. Microbiology, 14*(1), 20–32. https://doi.org/10.1038/nrmicro3552

[11] Evans, D. F., Pye, G., Bramley, R., Clark, A. G., Dyson, T. J., & Hardcastle, J. D. (1988). Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut, 29*(8), 1035–1041. https://doi.org/10.1136/gut.29.8.1035

[12] Zoetendal, E. G., Raes, J., van den Bogert, B., Arumugam, M., Booijink, C. C., Troost, F. J., Bork, P., Wels, M., de Vos, W. M., & Kleerebezem, M. (2012). The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *The ISME Journal, 6*(7), 1415–1426. https://doi.org/10.1038/ismej.2011.212

[13] Pei, Z., Bini, E. J., Yang, L., Zhou, M., Francois, F., & Blaser, M. J. (2004). Bacterial biota in the human distal esophagus. *Proceedings of the National Academy of Sciences of the United States of America, 101*(12), 4250–4255. https://doi.org/10.1073/pnas.0306398101

[14] Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F., Perez-Perez, G., Blaser, M. J., & Relman, D. A. (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proceedings of the National Academy of Sciences of the United States of America, 103*(3), 732–737. https://doi.org/10.1073/pnas.0506655103

[15] Ahn, J., Chen, C. Y., & Hayes, R. B. (2012). Oral microbiome and oral and gastrointestinal cancer risk. *Cancer Causes & Control : CCC, 23*(3), 399–404. https://doi.org/10.1007/s10552-011-9892-7