Differential microbiota network according to colorectal cancer lymph node metastasis stages

Yeuni Yu, Donghyun Han, Hyomin Kim, Yun Hak Kim, and Dongjun Lee

1Biomedical Research Institute, Pusan National University School of Medicine, Yangsan, Korea
2Department of Medicine, Pusan National University School of Medicine, Yangsan, Korea
3Interdisciplinary Program of Genomic Science, Pusan National University, Yangsan, Korea
4Periodontal Disease Signaling Network Research Center, School of Dentistry, Pusan National University, Yangsan, Korea
5Department of Anatomy, Pusan National University School of Medicine, Yangsan, Korea
6Department of Biomedical Informatics, Pusan National University School of Medicine, Yangsan, Korea
7Department of Convergence Medicine, Pusan National University School of Medicine, Yangsan, Korea
8Transplantation Research Center, Research Institute for Convergence of Biomedical Science and Technology, Pusan National University Yangsan Hospital, Yangsan, Korea

Purpose: Colorectal cancer (CRC) is a common malignancy worldwide and the second leading cause of cancer-related deaths. In addition, lymph node metastasis in CRC is considered an important prognostic factor for predicting disease recurrence and patient survival. Recent studies demonstrated that the microbiome makes substantial contributions to tumor progression, however, there is still unknown about the microbiome associated with lymph node metastasis of CRC. Here, we first reported the microbial and tumor-infiltrating immune cell differences in CRC according to the lymph node metastasis status.

Materials and Methods: Using Next Generation Sequencing data acquired from 368 individuals diagnosed with CRC (N0, 266; N1, 102), we applied the LEfSe to elucidate microbial differences. Subsequent utilization of the Kaplan-Meier survival analysis enabled the identification of particular genera exerting significant influence on patient survival outcomes.

Results: We found 18 genera in the N1 group and 3 genera in the N0 group according to CRC lymph node metastasis stages. In addition, we found that the genera Crenobacter (P=0.046), Maricaulis (P=0.093), and Arsenicicoccus (P=0.035) in the N0 group and Cecembia (P=0.08) and Asanoa (P=0.088) in the N1 group were significantly associated with patient survival according to CRC lymph node metastasis stages. Further, Cecembia is highly correlated to tumor-infiltrating immune cells in lymph node metastasized CRC.

Conclusion: Our study highlights that tumor-infiltrating immune cells and intratumoral microbe diversity are associated with CRC. Also, this potential microbiome-based oncology diagnostic tool warrants further exploration.

Key words: Colorectal neoplasms, The cancer genome atlas, Microbiota, Cancer microenvironment.
Introduction

Colorectal cancer (CRC) has become the third most common malignant tumor in the world [1]. The American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging system is the standard for determining the prognosis of patients with CRC and is highly correlated with 5-year overall survival. According to the TNM staging system, it has not spread to nearby lymph nodes (N0) or 1 to 3 nearby lymph nodes (N1) [2].

The microbiome is a complex ecosystem of microorganisms that live in and on the human body, particularly in the gut [3]. The gut microbiome plays a crucial role in human health, including digestion, vitamin synthesis, and immune system modulation [4]. The gut microbiome can modulate the immune response to cancer cells [5]. The microbiome can promote an inflammatory response that promotes the growth of cancer cells or stimulate an immune response that helps eliminate cancer cells [6]. Several studies demonstrated that gut microbiome dysbiosis is associated with tumorigenesis and/or tumor growth across CRCs [5,7]. Individuals with CRC have a less diverse microbiome, meaning there are fewer types of bacteria present. There is a higher abundance of specific bacterial species in individuals with CRC, such as Fusobacterium nucleatum [8,9]. Changes in the microbiome can contribute to the development and progression of CRC through toxin production, inflammation, and alteration of cellular signaling pathways [10,11]. Fusobacterium nucleatum in colorectal tumors promotes resistance to chemotherapy through modulating autophagy [12].

Internal organs were known to aseptic, but recent investigations have unveiled the existence of microorganisms, encompassing bacteria, viruses, and fungi, residing within tumors [13,14]. It is noteworthy that the phenomenon of microbiome invasion in the tumor microenvironment remains an actively researched area, with numerous aspects concerning its impact on tumor biology and response to treatment yet to be comprehensively elucidated. Nevertheless, exploration of the microbiome’s role within the tumor microenvironment holds considerable potential for advancing our understanding of cancer biology and facilitating the development of innovative therapeutic strategies.

Recent studies suggest that the composition of the microbiome may influence the effectiveness of cancer treatments [7]. The gut microbiome plays a role in the efficacy of immunotherapy, a type of cancer treatment that harnesses the immune system to target cancer cells [15-17]. Changes in the gut microbiome may contribute to the development and progression of CRC and may play a role in the efficacy of cancer treatment [12]. Thus, in this study, we understand the underlying molecular differences among the microbiome and tumor-infiltrating immune cells of CRC according to the lymph node metastasis stages.

Materials and Methods

1. Microbiome Datasets

We downloaded Kraken–TCGA (The Cancer Genome Analysis)–Raw Data (n=18,116) from microbiome count data [18]. We selected data whose sample_type is a primary tumor, investigation is TCGA-COAD (colon adenocarcinoma) and experimental_strategy is RNA sequencing. For this study, we also downloaded Metadata–TCGA (n=18,116) and Clinical data (n=7,579) to obtain patient data [19]. We screened samples whose pathologic N label is N0, N1, N1a, N1b, and N1c and merged three files using sample ID, case uuid, and case-submitter-ID (Fig. 1). Then, samples obtained from one person with the same case-submitter-ID were integrated using the mean. In duplicated samples, we selected preferentially clinical data including the TCGA barcode from "file name". In this study, processing and producing graphs were used with the R program (version 4.2.0).

2. Linear Discriminant Analysis (LDA) Effect Size (LEfSe) Analysis

To compare N0 and N1 groups and identify significantly different bacteria between them, we processed the merged dataset and inputted it into LEfSe via the Huttenhower Lab Galaxy Server. LEfSe computes effect sizes, enabling the quantification of the magnitude with which a specific microbe is linked to a particular group. This analytical approach extends beyond the mere identification of differences, providing insight into the significance and strength of these observed distinctions [20]. We used the LEfSe and adjusted the logarithmic LDA score cutoff to 2.0. Virus samples were excluded before input.

3. Boxplot and Survival Analysis

According to pathologic groups, we drew a boxplot graph of the genus screened by LEfSe [21]. A genus with excessive outlier value that might affect the overall comparative study and a genus whose median is 0 was excluded from the boxplot. But a genus with 0 medians was included in the survival analysis. Boxplot was drawn with data in the range of 0.01 to 0.97 quantile of count data.

To identify the genus that significantly affects the survival of
patients, we drew the Kaplan-Meier survival curve and represented the significance level (P-value). We used the data label “days_to_last follow up”. If one data status is not available, it is supplemented by adding a “days_to_death” value. In N1 group-rich genus, a genus whose P-value is under 0.10 in N1 group was selected as the meaningful genus. Likewise, a genus whose P-value is under 0.10 in the N0 group was selected for the meaningful genus in the N0 group-rich genus. Due to an insufficient number of identifiable strains at the 0.05 significance level, the threshold was adjusted to 0.1 to facilitate the selection of strains for subsequent sub-analysis. After all, we finally obtained the main significant bacteria.

4. Correlation Analysis
Correlation analysis between tumor infiltrate immune cells with microbiome was performed through TCGA query. We retrieved RNA expression data using the R package “TCGA bio links” [22]. Barcode was used preferentially if there was a detailed case submitter code in the file name, and if not, case-submitter-ID was used. In the case of duplicated samples because of different plate numbers, data was used with average values. Gene ensembles were converted to HGNC gene symbols using the org.Hs.dg.db package (version 3.15.0.). When one ensemble is converted to various symbols, we selected one that is first searched on the ENSEMBL website. If some ensembles are converted to one common symbol, data was processed by mean. After data processing, data were analyzed to see the expression of immune cells via the TIMER 2.0 server [23]. Based on the data that shows the association between immune cells and microbiome, we performed a co-occurrence network analysis using sparCC correlation. Hmisc and igraph packages were used to process statistics and results were visualized using Cytoscape software (P<0.05, R>0.15).

Results

1. Identifying the Unique Signature Microbiome of COAD with LEfSe
We obtained a final 368 samples that met our study criteria. Of these, 266 samples were classified as N0 stages, and 102 samples were classified as N1 stages (Table 1). Overall, there was no significant difference in patient status between the two groups, as determined by population numbers. However, the N0 group had significantly younger patients than the N1 group (N0:...
68.58647 years, and N1: 64.74510 years, \( P<0.05 \) (Table 1). Regarding the pathologic stage, stage 2A was the most common in the N0 group (137 samples) while stage 3B was the most common in the N1 group (51 samples). In race labels, samples from Germany were the most common (N0, 108; N1, 29 samples). There was no significant difference in days to the last follow-up that represents survival days (t-test, \( P=0.9756 \)) (Table 1).

To assess microbial diversity between the two groups, alpha diversity, and beta diversity analyses were conducted (Fig. 2). Alpha diversity was evaluated using the observed method, which measures microbial richness, as well as the Shannon and Simpson indices, which account for evenness. None of the three methods revealed any significant difference in alpha diversity between the two groups (Fig. 2A). Principal coordinate analysis of the Bray-Curtis dissimilarity index further demonstrated that there was no statistically significant variation in the microbial composition between the two groups (Fig. 2B).

2. Selecting a Genus Affecting Patient Survival

As a result of the boxplot that depicts genera screened by LEfSe (Fig. 3A), 18 genera in the N1 group, and 3 genera in the N0 group were found, excluding genera with a median 0 and those with a large influence of outlier (Fig. 3B and C). Although excluded in the boxplot, genera with 0 medians are analyzed in survival analysis.

To identify the genera that significantly affect survival in CRC, we performed survival analysis using the Kaplan–Meier method on the output data from the LEfSe analysis (Fig. 4). Based on a significance level of 0.1, we found that the genera Crenobacter (\( P=0.046 \)), Maricaulis (\( P=0.093 \)), and Arsenicicoccus (\( P=0.035 \)) in the N0 group and Cecembia (\( P=0.08 \)) and Asanoa (\( P=0.088 \)) in the N1 group were significantly associated with patient survival according to CRC lymph node metastasis stages (Fig. 4). In the N1 group, high levels of Asanoa hindered patient survival, while low levels of the other genera hindered survival in both groups. Each genus showed a specific survival outcome in each group, without affecting survival in the opposite group.

### Table 1. Demographic of sample

<table>
<thead>
<tr>
<th></th>
<th>N0 (n=266)</th>
<th>N1 (n=102)</th>
<th>( P )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>68.58647 (31-90)</td>
<td>64.74510 (37-90)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>145</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>121</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>29</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>119</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>10</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Not available</td>
<td>108</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Pathologic N stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1a</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1b</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1c</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days to last follow-up (day)</td>
<td>856.9019 (0-4,270)</td>
<td>853.9608 (0-4,502)</td>
<td>0.9756</td>
</tr>
</tbody>
</table>

Values are presented as mean (range) or number only.
N0, not spread to nearby lymph nodes; N1, 1 to 3 nearby lymph nodes.
-, not available.

![Fig. 2. Diversity of bacteria from the colorectal caner samples. (A) Alpha diversity for 564 cases. Methods are ‘Observed’, ‘Shannon’, and ‘Simpson’. (B) Beta diversity for 564 cases. There is no significant difference between N0 and N1 groups. PCoA, principal coordinate analysis.](image-url)
Fig. 3. LDA effect size (LEfSe) result and Boxplot. (A) LEfSe analysis between groups. Thirty-six genera were screened in the N1 group and Twenty-three genera were screened in the N0 group. (B) Box plot for N1 rich genus. (18 genera) Genera with zero median value or excessive outliers were excluded. The figure shows microbiomes with higher levels of distribution in the N1 group. (C) Box plot for N0 rich genus. (3 genera) Genera with zero median value or excessive outliers were excluded. The figure shows microbiomes with higher levels of distribution in the N0 group. LDA, linear discriminant analysis.

Fig. 4. Survival analysis using the Kaplan-Meier method. (A, B) Cecembia (N1 rich) for the N1 group and Asanoa (N1 rich) for the N1 group. P<0.1. (C-E) Crenobacter (N0 rich) for the N0 group, Maricaulis (N0 rich) for the N0 group, and Arsenicoccus (N0 rich) for the N0 group.
3. Tumor-Infiltrating Immune Cell Analysis

To identify mechanisms of action in vivo, immune estimation was performed using TIMER 2.0. We obtained estimation results consisting of the timer, Cibersort, Cibersort-abs, and Quantiseq. Utilizing TIMER results, we conducted differential estimation of immune cell populations and performed co-occurrence network analysis. Significant differences were observed between the N1 and N0 groups with regard to macrophage and CD4 T cells (P<0.05). Specifically, macrophage levels were higher in the N1 group, while CD4 T cell levels were higher in the N0 group (Fig. 5A). In the N0 group, Arsenicicoccus is associated with Maricaulis and these two genera showed no correlation with immune cells. Cecembia is N1 rich genus and is related to Myeloid dendritic cell activated, T cell CD4 memory resting, Mast cell activated, NK cell resting, M1 and M2 macrophage, and T cell regulatory Tregs (Fig. 5B and C).

Discussion

In this study, we first demonstrated microbiomes are associated with the survival of lymph node metastasis in CRC patients. Microbiomes communicate with gut mucosa and regulate gut immune function [15-17]. During the cancer progression, changes in pH and metabolism in the gut seem to occur and good microbiomes also disappear [24].

We observed no significant variation in the alpha and beta diversity of microorganisms between the N0 and N1 groups. However, we identified specific strains that exhibited statistically significant differences between the two groups. Notably, a greater number of bacteria with significantly higher abundance were found in the N1 group compared to the N0 group. Among the strains exhibiting differences between N0 and N1, the majority of strains showing higher abundance in N1 were identified as gram-negative bacteria. Actinobacillus, Enterococcus, and Chryseobacterium are known to induce inflammation in mammals and are associated with diseases [25-27]. Specifically, Biliverdin produced by Enterococcus stimulates tumor development in CRC and influences cell proliferation and angiogenesis [26]. Therefore, these bacterial strains may potentially influence lymph node metastasis in CRC.

In our survival analysis, interestingly, Asanoa worsened the prognosis of the N1 group. In addition, we found genera that have a protective effect on the survival of COAD patients. Genera with a higher population in the N0 group may be beneficial bacteria including Crenobacter, Maricaulis, and Arsenicicoccus. In addition, preserving Crenobacter, Maricaulis, Arsenicicoccus, and Cecembia would make a good improvement in survival, especially, Cecembia, N1 rich genus that has a defensive effect, showed a correlation with almost every immune cell, so it can be used as the main target for CRC patients with lymph node metastasis. Thus, the new treatment is possible by targeting new markers expressed in harmful microbiomes that correlate with immune cells. Also, it can be used as a basis for treatment or adjuvant therapy for prognosis improvement, such as using it in the development of probiotics that feed the microbiomes.

Microbiome and immune cell co-occurrence network analysis unveiled a heightened number of significant correlations within the N1 group. Notably, increased intercellular interaction among immune cells and a stronger correlation with Cecembia were observed within the N1 subgroup. In consideration of the correlation between Cecembia and Treg cells, as well as the elevated abundance of Cecembia and improved survival among N1 patients, these findings collectively indicate that Cecembia has the potential to influence the shift in the immunosuppressive
function of Tregs at the primary site of N1 CRC. Recent studies support the notion that the gut microbiome and its metabolites possess the capacity to regulate the anti-tumor immune response through modulation of T cell activation, exerting either promotional or inhibitory effects [28,29].

Also, tumor-infiltrating lymphocyte affects the prognosis of patients and accuracy is proportional to stage level [30] and we could apply a new score system that reflects the diagnosis and predicts future prognosis associated with the microbiome in CRC. It would help to understand the survival and therapy for CRC patients.

Acknowledgements

None.

Funding

This work was supported by grants from the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (NRF-2021R1A2C4001466, 2022R1A5A2027161, and 2021R1A6A3A01086785).

Authors’ Contributions

DL and YHK initiated the study and guided the work. DH and YY collected and normalized the data. DH, YY, and HK analyzed the experimental data and interpreted the data. All authors wrote the manuscript with input from all co-authors.

References


