

Article

# Compositional and Functional Differences between Microbiota and Cervical Carcinogenesis as Identified by Shotgun Metagenomic Sequencing

Minji Kwon <sup>1,†</sup>, Sang-Soo Seo <sup>2,†</sup>, Mi Kyung Kim <sup>1,\*</sup>, Dong Ock Lee <sup>2</sup> and Myoung Cheol Lim <sup>2</sup>

<sup>1</sup> Division of Cancer Epidemiology and Prevention, National Cancer Center, 323, Ilsan-ro, Ilsandong-gu, Goyang-si 10408, Korea; 74433@ncc.re.kr

<sup>2</sup> Center for Uterine Cancer, National Cancer Center, 323, Ilsan-ro, Ilsandong-gu, Goyang-si 10408, Korea; ssseomd@ncc.re.kr (S.-S.S.); dolee@ncc.re.kr (D.O.L.); mclim@ncc.re.kr (M.C.L.)

\* Correspondence: alrud@ncc.re.kr; Tel.: +82-31-920-2202; Fax: +82-31-920-2006

† These authors contributed equally to this work.

Received: 18 December 2018; Accepted: 28 February 2019; Published: 5 March 2019



**Abstract:** Recent studies have reported the potential role of microbiomes in cervical disease. However, little is known about the microbiome composition and function in cervical carcinogenesis. We aimed to identify the compositional and functional alterations of cervical microbiomes in cases of cervical carcinogenesis of Korean women using shotgun metagenomic sequencing. In this study, using shotgun sequencing, we sequenced the cervical metagenomes of cervical intraepithelial neoplasia 2/3 ( $n = 17$ ), cervical cancer ( $n = 12$ ), and normal controls ( $n = 18$ ) to identify the microbial abundances and enriched metabolic functions in cervical metagenomes. At the genus level, the microbiota of cervical cancer were differentially enriched with genera *Alkaliphilus*, *Pseudothromotoga*, and *Wolbachia*. Cervical intraepithelial neoplasia (CIN) 2/3 were enriched with *Lactobacillus*, *Staphylococcus*, and *Candidatus Endolissoclinum*. The normal group was enriched with *Pseudoalteromonas* and *Psychrobacter*. Further characterization of the functionalities of the metagenomes may suggest that six Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologies (KOs) that are involved in 10 pathways are associated with an increased risk of CIN2/3 and cervical cancer. Specifically, cervical metagenomes were enriched in the course of peptidoglycan synthesis and depleted by dioxin degradation and 4-oxalocrotonate tautomerase. The Cluster of Orthologous Groups (COG) category ‘Defense mechanisms’ was depleted in cervical cancer patients. Our findings based on shotgun metagenomic sequencing suggest that cervical microbiome community compositions and their metagenomics profiles differed between cervical lesions and normal subjects. Future studies should have larger sample sizes and/or aggregate their results to have sufficient power to detect reproducible and significant associations.

**Keywords:** cervical cancer; cervical intraepithelial neoplasia; shotgun metagenomic sequencing; microbiome

## 1. Introduction

Cervical cancer arises from multi-stage epithelium infection, persistent HPV infection, the progression of persistently HPV-infected epithelium to cervical precursor cells, and invasion through the epithelial basement membrane in the uterine cervix [1]; however, such mechanisms are not sufficient to explain the development of cervical cancer [2]. Various variables such as smoking, oral contraceptive use, sexually transmitted diseases, parity, and dietary factors affect cervical cancer progression with human papillomavirus (HPV) [3]. Recent studies have reported that microbiomes can affect the HPV-related process of cervical carcinogenesis [4]. There is increasing evidence that uterine and vaginal microbiomes play important roles in the carcinogenesis process of the uterine cervix [5–7].

The dominance of vaginal microbiota by *Lactobacillus* species is associated with the maintenance of a healthy reproductive state and relatively low vaginal pH (<4.5) [8]. Bacterial vaginosis (BV) is also associated with cervical intraepithelial neoplasia [9]. BV is characterized by the massive breeding of anaerobic bacteria such as *Gardnerella vaginalis*, *Mobiluncus* species, *Prevotella* species, *Mycoplasma hominis*, and *Atopobium vaginae*, as well as the loss of native *Lactobacillus*-dominant microflora [10].

In a previous study, we pyrosequenced the 16S rRNA gene and showed that cervical microbial patterns rich in *Atopobium vagiane*, *Lactobacillus iners*, and *Gardnerella vaginalis* and not rich in *Lactobacillus crispatus* were at high risk for cervical intraepithelial neoplasia (CIN) [11–13]. However, it is unclear whether cervical cancer and precancerous lesions are associated with alterations in the functional composition of cervical metagenome. Shotgun metagenomic DNA sequencing has provided valuable insight into the phylogeny, biodiversity, metabolic abilities and functional diversity of a variety of organisms [14]. This technology has the potential to provide researchers and clinicians with a better understanding of the pathogenesis of cervical disease and identify unknown pathogens to aid the formulation of both therapeutic and prevention strategies for microbial disease agents [15]. To the best of our knowledge, no study on cervical carcinogenesis has examined the relationship between diagnosis and the compositional or functional alterations of shotgun metagenomes.

In this study, using shotgun sequencing, we sequenced the cervical metagenomes of cervical intraepithelial neoplasia 2/3, cervical cancer, and healthy controls to identify microbial abundances and enriched metabolic functions.

## 2. Results

### 2.1. Characteristics of Subjects

The general characteristics of the 47 subjects are provided in Table 1. The study included 18 normal (group A), 17 CIN 2/3 (group B), and 12 cervical cancer (group C) patients. Significant differences were observed for age ( $p = 0.0004$ ): the mean ages of the groups were 45, 41, and 55 years, respectively. No inter-group differences were observed for other variables such as body mass index, oral contraceptive use, smoking, or drinking status.

**Table 1.** Characteristics of Normal, Cervical Intraepithelial Neoplasia (CIN) 2/3, and Cervical Cancer Subjects.

Characteristics	Normal (Group A, n = 18)	CIN 2/3 (Group B, n = 17)	Cervical Cancer (Group C, n = 12)	$p^a$
Age (year) <sup>b</sup>	45.6 (7.72)	41.1 (6.98)	55.9 (10)	0.0004
BMI (kg/m <sup>2</sup> ) <sup>b</sup>	21.3 (2.66)	21.3 (2.56)	22.9 (4.03)	0.4852
Energy intake (kcal/day) <sup>b</sup>	1892 (367)	1839 (435)	1953 (452)	0.8232
HPV <sup>c</sup>	11 (61.1)	12 (80)	8 (72.7)	0.5435
Post-menopausal <sup>c</sup>	7 (41.2)	3 (18.7)	7 (63.6)	0.0604
Oral contraceptive use (Yes&ever) <sup>c</sup>	3 (16.7)	5 (29.4)	4 (33.3)	0.5343
Hormone treatment (Yes&ever) <sup>c</sup>	4 (23.6)	0 (0)	1 (9.1)	0.1552
Current smoker <sup>c</sup>	2 (11.8)	2 (12.5)	1 (9.1)	1.0
Passive smoker <sup>c</sup>	10 (58.8)	13 (81.3)	7 (63.6)	0.3587
Alcohol drinker <sup>c</sup>	12 (70.6)	14 (87.5)	6 (54.5)	0.0615

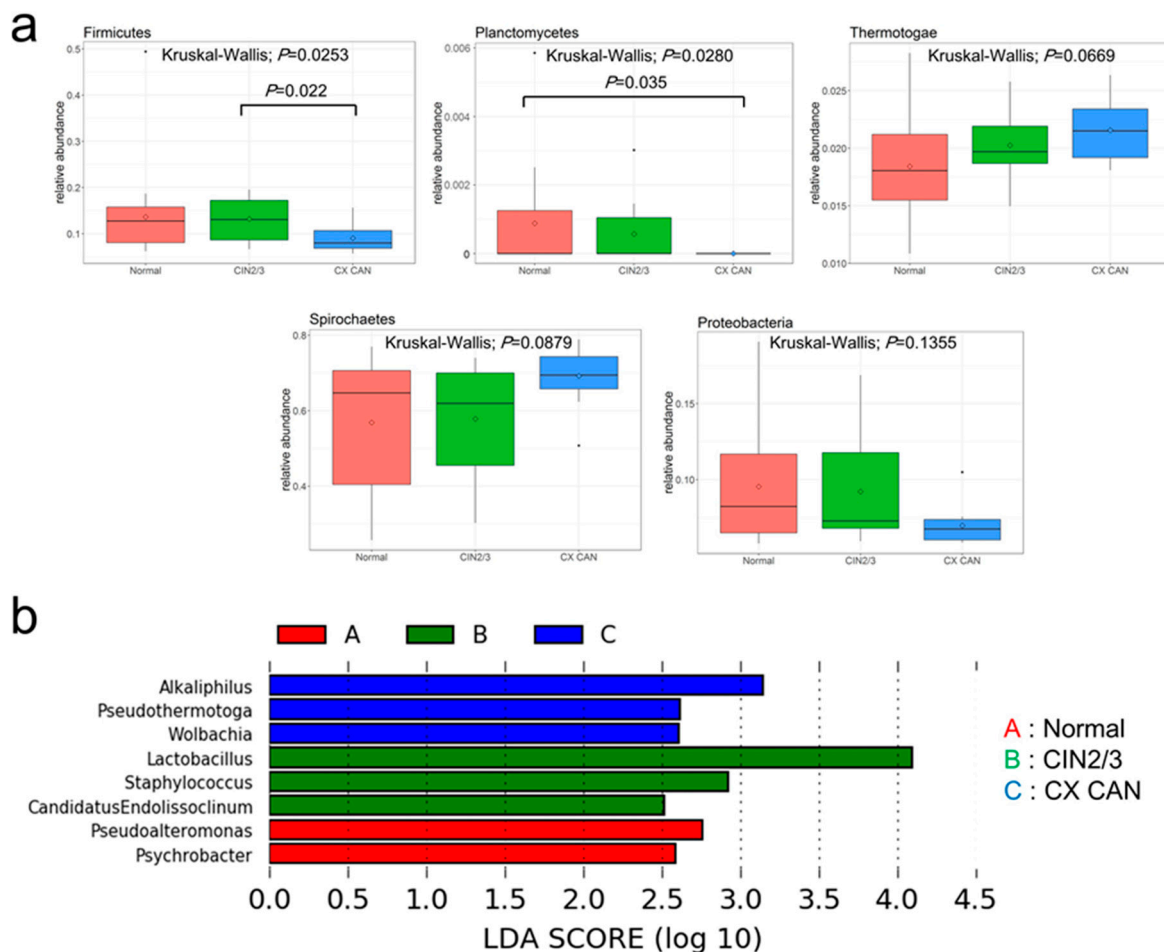
<sup>a</sup>  $p$  value calculated by *chi-squared* test for categorical variables and by *Kruskal–Wallis* test for continuous variables. Fisher's exact test was performed when the categorical variable was more than 25% of the cells with an expected frequency of five or less. <sup>b</sup> Mean (standard deviation) <sup>c</sup> n (%) BMI: body mass index; HPV: human papillomavirus.

### 2.2. Taxonomic Characterization of Cervical Microbiome

The majority ( $73 \pm 12\%$ ) of the reads were bacterial and dominated by the phyla *Spirochaetes* and *Firmicutes*, representing 60% and 12% of the microbiota, respectively, followed by *Proteobacteria* (8.8%), *Actinobacteria* (4.4%), and *Bacteroidetes* (4.3%) (Supplementary Table S1). The most abundant genera in our cohort were *Leptospira*, *Gardnerella*, *Ehrlichia*, *Lactobacillus*, *Clostridium*, and *Streptococcus* (Supplementary Figure S1).

### 2.3. Microbial Composition among Normal, Cervical Intraepithelial Neoplasia 2/3, and Cervical Cancer

The phyla *Firmicutes* and *Planctomycetes* were decreased in cervical cancer patients ( $p < 0.05$ ) relative to the normal or CIN 2/3 group. *Proteobacteria* also was decreased (normal, CIN 2/3 versus cancer, Wilcoxon rank sum test  $p = 0.04$ ), whereas *Spirochaetes* (normal, CIN 2/3 versus cancer, Wilcoxon rank sum test  $p = 0.02$ ) was enriched (Figure 1a). The taxonomic lists that were differentially abundant among the groups were identified by Linear discriminant analysis (LDA) effect size (LEfSe) (minimum LDA score: 2.5) (Figure 1b). At the genus level, the microbiota of cervical cancer were differentially enriched with the genera *Alkaliphilus* ( $p = 0.03$ ), *Pseudothermotoga* ( $p = 0.02$ ), and *Wolbachia* ( $p = 0.01$ ). Meanwhile, CIN 2/3 were enriched with *Lactobacillus* ( $p = 0.02$ ), *Staphylococcus* ( $p = 0.03$ ), and *Candidatus Endolissoclinum* ( $p = 0.01$ ). The normal group was enriched with *Pseudoalteromonas* ( $p = 0.04$ ) and *Psychrobacter* ( $p = 0.03$ ) (Supplementary Figure S2). A principal component analysis revealed that the microbial phyla (and species) abundance did not distinguish the disease groups (healthy controls, CINs, and cervical cancer, Bray  $p = 0.087$  and Jaccard  $p = 0.094$ ) (Supplementary Figure S3). Alpha diversity according to the Shannon and Simpson indices did not differ at the phylum or genera level (Supplementary Figure S4).



**Figure 1.** Microbial compositions among normal group, cervical intraepithelial neoplasia 2 or 3, and cervical cancer. (a) The differences are in the relative abundances of assembly-based taxonomy for the five major phyla levels. *Firmicute* and *Planctomycetes* showed statistically significant differences ( $p = 0.0253$  (B–C:  $p = 0.022$ );  $p = 0.0280$  (A–C:  $p = 0.035$ )). (b) Genus showed statistically significant difference as a result of Linear discriminant analysis (LDA) effect size (LEfSe) analysis. (Logarithmic LDA score  $>2.5$ ; alpha value  $<0.05$ ). The post hoc analysis used the Bonferroni method.

#### 2.4. Distribution and Differences in Relative Abundances of COG Categories

To obtain insight into the functional properties of cervical microbes, a significant detection gene was assigned to the Cluster of Orthologous Groups (COG) database [16] for microbial function analysis. The major COG categories detected in our subjects belonged to the functional categories “Cellular processes and signaling” (mainly ([D] cell-cycle control, cell division, chromosome partitioning, [O] post-translational modification, protein turnover, and chaperones, [Z] Cytoskeleton) and “Metabolism” (mainly ([G] Carbohydrate transport and metabolism) (Figure 2a, Supplementary Table S2). The COG categories that were differentially abundant among the groups were identified by LefSe. Of the 25 COG categories, 22 showed an LefSe LDA score of 2.5 or higher (Figure 2b). The COG categories of [V] Defense mechanisms, [T] Signal transduction mechanisms, and [K] Transcription were depleted in cervical cancer patients. Meanwhile, the dominant categories “Cellular processes and signaling” ([D] cell-cycle control, cell division, chromosome partitioning, [O] post-translational modification, protein turnover, chaperones, and [Z] Cytoskeleton) and “Metabolism” ([G] Carbohydrate transport and metabolism) were enriched in cervical cancer patients. The 17 COG categories with lesser abundances were significantly enriched in the normal controls.

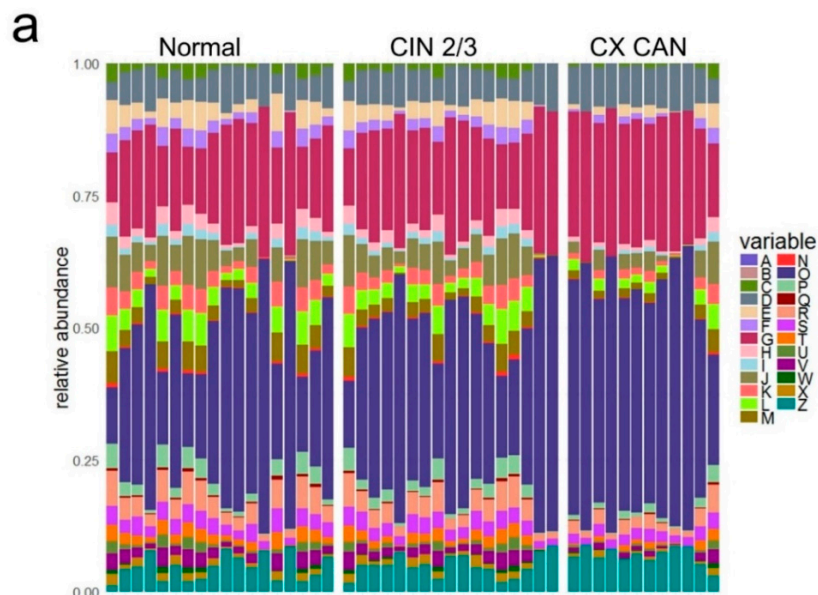
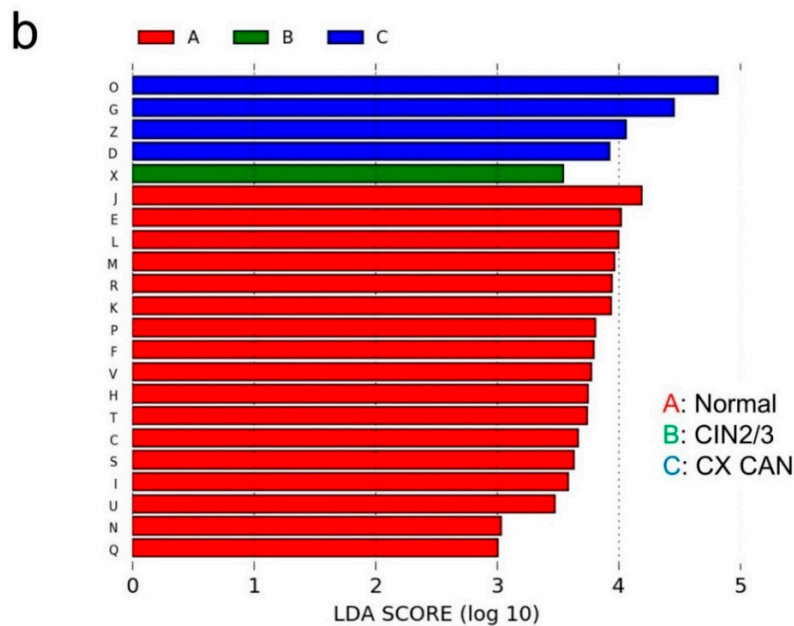


Figure 2. Cont.



**Figure 2.** Distribution of relative abundances of Cluster of Orthologous Groups (COG) categories with significant differences in LefSe analysis. (a) Distribution in relative abundances of COG categories. (b) Differences in relative abundances of COG categories by LefSe analysis. (Logarithmic LDA score >2.5; alpha value <0.05). [A] RNA processing and modification; [B] Chromatin structure and dynamics; [C] Energy production and conversion; [D] Cell-cycle control, cell division, chromosome partitioning; [E] Amino acid transport and metabolism; [F] Nucleotide transport and metabolism; [G] Carbohydrate transport and metabolism; [H] Coenzyme transport and metabolism; [I] Lipid transport and metabolism; [J] Translation, ribosomal structure, and biogenesis; [K] Transcription; [L] Replication, recombination, and repair metabolism; [M] Cell-wall/membrane/envelope biogenesis; [N] Cell motility; [O] Post-translational modification, protein turnover, and chaperones; [P] Inorganic ion transport and metabolism; [Q] Secondary metabolites biosynthesis, transport, and catabolism; [R] General function prediction only; [S] Function unknown; [T] Signal transduction mechanisms; [U] Intracellular trafficking, secretion, and vesicular transport; [V] Defense mechanisms; [W] Extracellular structures; [X] Mobilome: prophages, transposons; [Z] Cytoskeleton.

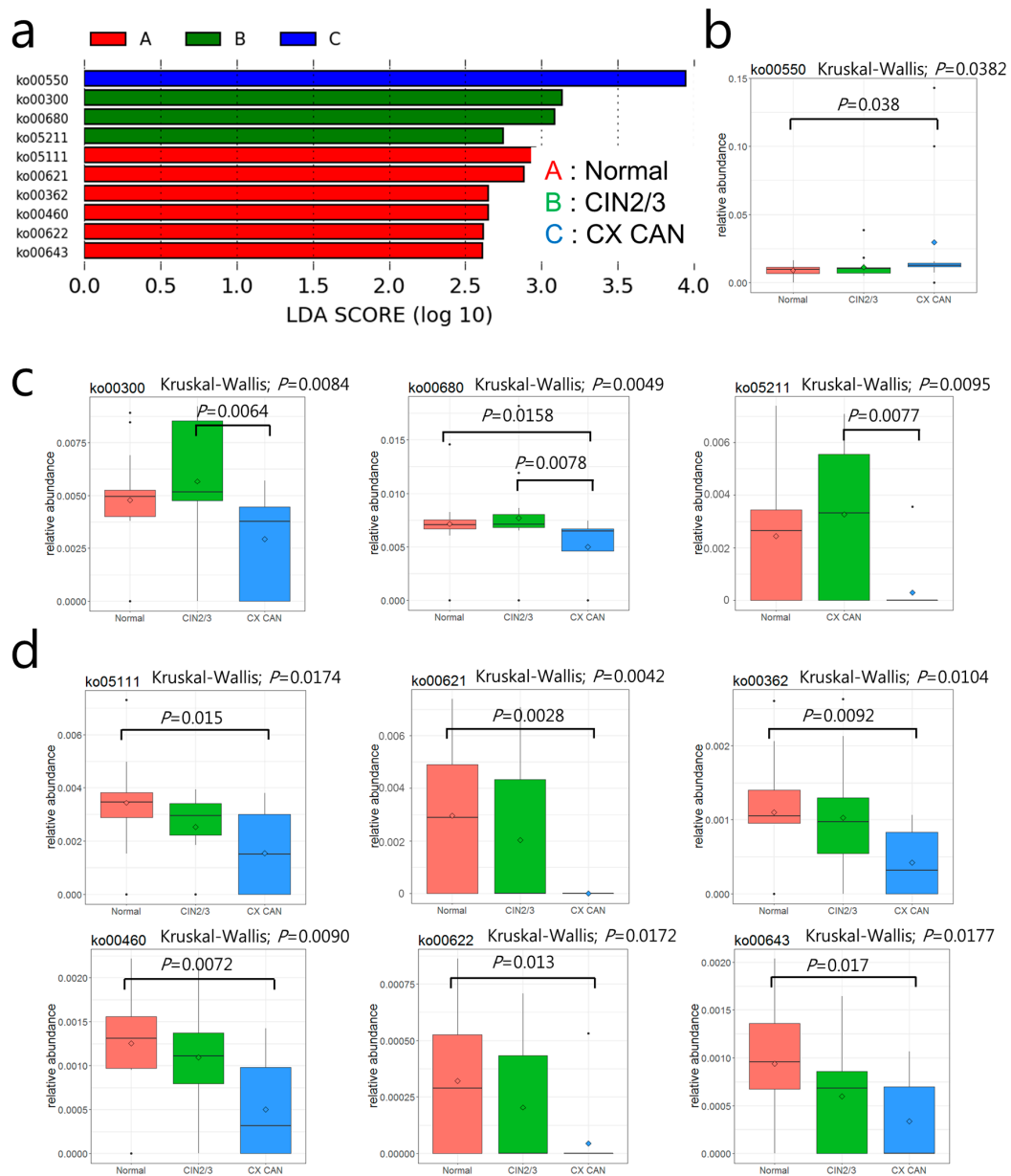
### 2.5. Metabolic Functions of Cervical Microbiota

We conducted a LefSe analysis to distinguish among the disease groups and 224 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Ten KEGG pathways were presented significantly differently among the groups (Figure 3a): six pathways were enriched in the cervical metagenomes of the controls, while three pathways and one pathway were enriched in the CIN2/3 and cancer subjects, respectively (*Kruskal–Wallis* test). Of the 2860 KEGG orthologies (KOs), 118 were found to be statistically significant in the LefSe analysis (Supplementary Figure S5). Of the 118 KOs, six belonging to 10 KEGG pathways were observed (Table 2).

The KEGG pathway of peptidoglycan biosynthesis (ko00550) was enriched in cervical cancer subjects ( $p = 0.03$ ) (Figure 3b), while ko00300 (Lysine biosynthesis,  $p = 0.01$ ), ko00680 (Methane metabolism,  $p = 0.005$ ), and ko05211 (Renal cell carcinoma,  $p = 0.01$ ) were enriched in CIN2/3 (adj.  $p < 0.05$ , Figure 3c). We also found that the pathways of Biofilm formation in *Vibrio cholerae* (ko05111,  $p = 0.01$ ), dioxin degradation (ko00621,  $p = 0.004$ ), benzoate degradation (ko00362,  $p = 0.01$ ), cyanoamino acid metabolism (ko00460,  $p = 0.09$ ), xylene degradation (ko00622,  $p = 0.01$ ), and styrene degradation (ko00643,  $p = 0.01$ ) were abundant in the normal controls.

Several metabolites were involved in these KEGG pathways (Table 2): K00625 (phosphate acetyltransferase, Figure 4a) and K01007 (pyruvate, water dikinase), which are involved in ko00680 (methane metabolism) were decreased in cervical cancer, as were K00215 (4-hydroxy-

tetrahydrodipicolinate reductase, Figure 4b) and K01778 (diaminopimelate epimerase), which are involved in ko00300 (lysine biosynthesis), and K01679 (fumarate hydratase, class II, Figure 4c), which is involved in Ko05211 (renal cell carcinoma). Consequently, we also found that a lower abundance of K01821 (4-oxalocrotonate tautomerase, Figure 4d) in cervical cancer was involved in the pathways of K01821 (4-oxalocrotonate tautomerase, Figure 4d) in ko00621 (dioxin degradation), Ko00362 (Benzoate degradation), and Ko00622 (cyanoamino acid metabolism). A principal component analysis showed that the abundances of the KEGG pathway by group (Supplementary Figure S6) did not distinguish the disease groups (controls, CIN 2/3, and cervical cancer).

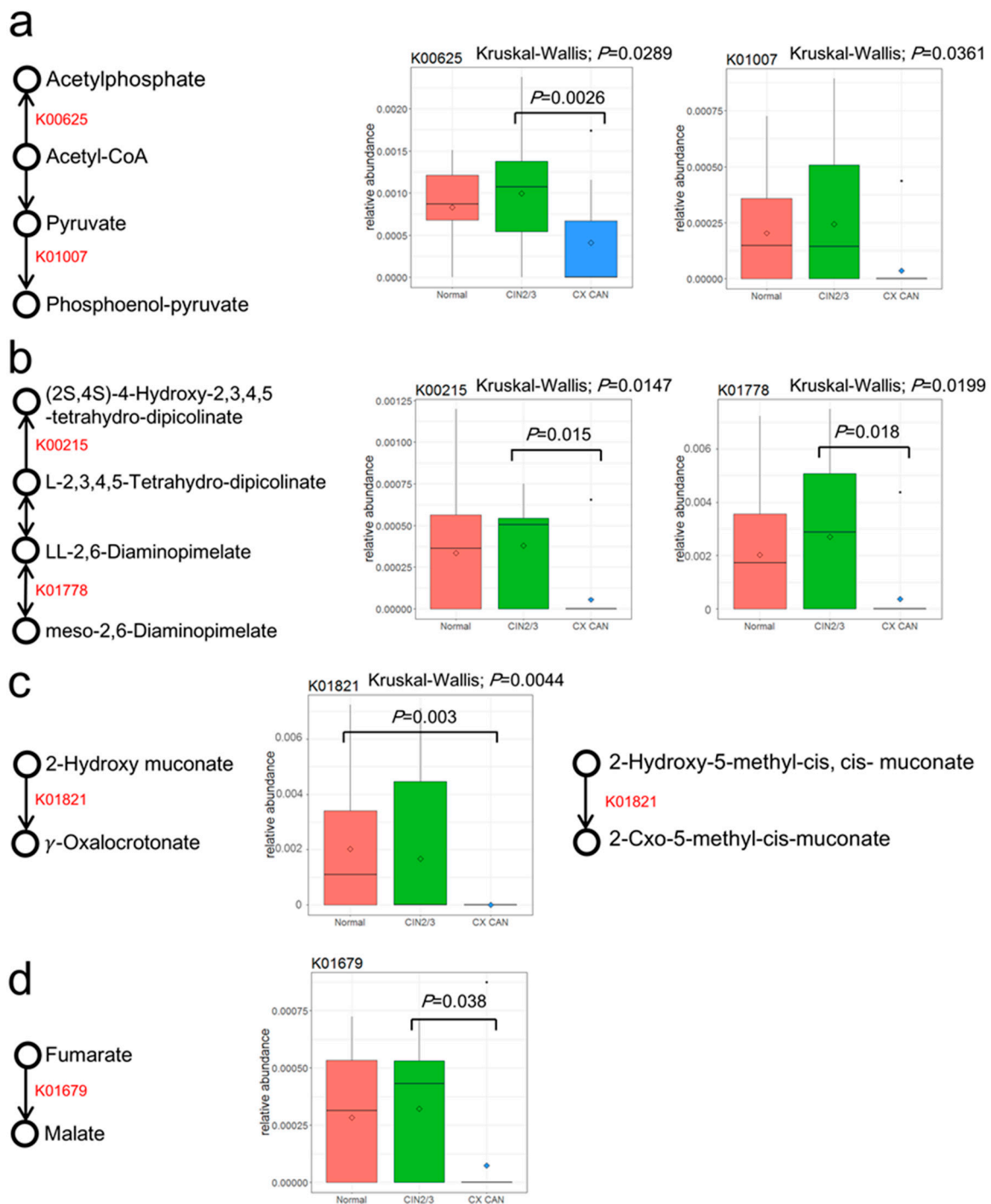


**Figure 3.** Differences in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway profiles among groups. (a) LefSe results of KEGG pathway. (b) Relative abundances of ko00550 were significantly enriched in cervical cancer. (c) KEGG pathways were significantly enriched in CIN 2 or 3. (d) KEGG pathways were enriched in normal subjects (logarithmic LDA score >2.5; alpha value <0.05). The post hoc analysis used the Bonferroni method.

**Table 2.** Mean relative abundance of 10 KEGG pathways and six KEGG orthologies (KOs).

KEGG Pathway	Pathway Name	Normal	CIN2/3	Cervical Cancer	<i>p</i> <sup>a</sup>	KEGG Orthology	KO Name	Normal	CIN2/3	Cervical Cancer	<i>p</i> <sup>a</sup>
ko00550	Peptidoglycan biosynthesis	0.8881	1.1465	2.9481	0.0382						
ko00300	Lysine biosynthesis	0.4785	0.5669	0.2940	0.0084	K00215	4-Hydroxy-tetrahydrodipicolinate reductase	0.0334	0.0380	0.0055	0.0146
						K01778	Diaminopimelate epimerase	0.0202	0.0271	0.0036	0.0199
ko00680	Methane metabolism	0.7161	0.7716	0.4996	0.0049	K00625	Phosphate acetyltransferase	0.0831	0.0997	0.0407	0.0289
						K01007	Pyruvate, water dikinase	0.0203	0.0243	0.0036	0.0361
ko05211	Renal cell carcinoma	0.0244	0.0327	0.0030	0.0095	K01679	Fumarate hydratase, class II	0.0284	0.0322	0.0073	0.0318
ko05111	Biofilm formation in <i>Vibrio cholerae</i>	0.3439	0.2532	0.1552	0.0174						
ko00621	Dioxin degradation	0.0296	0.0204	0	0.0042	K01821	4-Oxalocrotonate tautomerase	0.0202	0.0167	0	0.0044
ko00362	Benzoate degradation	0.1100	0.1027	0.0421	0.0104	K01821	4-Oxalocrotonate tautomerase	0.0202	0.0167	0	0.0044
ko00460	Cyanoamino acid metabolism	0.1251	0.1093	0.0502	0.0090						
ko00622	Xylene degradation	0.0320	0.0204	0.0044	0.0172	K01821	4-Oxalocrotonate tautomerase	0.0202	0.0167	0	0.0044
ko00643	Styrene degradation	0.0939	0.0596	0.0339	0.0177						

<sup>a</sup> *p* value calculated by *Kruskal–Wallis* test.



**Figure 4.** Differences among KEGG orthologies (KOs). (a) KOs are associated with ko00680. K00625 [phosphate acetyltransferase] and K01007 [pyruvate, water dikinase] enriched in CIN 2 or 3 ( $p = 0.02893$  (B–C:  $p = 0.0026$ );  $p = 0.0361$ ). (b) KOs are associated with ko00300. K00215 [4-hydroxy-tetrahydrodipicolinate reductase] and K01778 [diaminopimelate epimerase] enriched in CIN 2 or 3 ( $p = 0.0147$  (B–C:  $p = 0.015$ );  $p = 0.0199$  (B–C:  $p = 0.018$ )). (c) K01679 is associated with ko05211. K01679 [fumarate hydratase, class II] enriched in CIN 2 or 3 ( $p = 0.0318$  (B–C:  $p = 0.038$ )). (d) K01821 [4-oxalocrotonate tautomerase] is associated with ko00621, ko00362, and ko00622. K01821 enriched in normal subjects ( $p = 0.0044$  (A–C:  $p = 0.003$ )). The post hoc analysis used the Bonferroni method.



### 3. Discussion

This study performed shotgun metagenomic sequencing of cervical-swab samples from Korean women. Thereby, we observed significant microbiome profile differences and functionalities among a normal group, CIN 2/3, and cervical cancer. An LEfSe analysis showed statistically significant differences among eight genera and 10 pathways. Further, six KOs within those pathways were found. The COG category 'Defense mechanisms' was depleted in cervical cancer patients. Furthermore, this category has been reported to be deficient in ileal Crohn's disease as compared with healthy controls [17].

Bacteria, as a key player, were increased in number in the initiation and progression of a number of malignant tumors; *Fusobacteria* has been implicated for its role as a possible pro-carcinogenic bacterial phylum. Our findings showed that *Alkaliphilus*, *Pseudothermotoga*, and *Wolbachia* were statistically significantly enriched in cervical cancer. *Alkaliphilus* has been reported to be more numerous in a hepatic encephalopathy negative group [18]. *Wolbachia* has been reported for its immunological role in filarial infection, especially in humans [19]. In the present study, *Lactobacillus*, *Staphylococcus*, and *Candidatus Endolissoclinum* were found in CIN 2 or 3. *Lactobacillus* is a genus containing *L. crispatus* and *L. iners*. *L. crispatus* is predominant in healthy females, and *L. iners* is predominant in unhealthy individuals [11,20]. *Staphylococcus* spp. has been associated with colon carcinoma, catheter-related bloodstream infections, and neutropenic patients [21–23]. In the current study, the normal group were found to have been enriched with the genera *Pseudoalteromonas* and *Psychrobacter*. *Pseudoalteromonas* has been reported to be lower in *Helicobacter pylori*-infected patients [24], and *Psychrobacter* was reduced in colorectal carcinoma patients [25]. However, the associations of *Pseudothermotoga* and *Candidatus Endolissoclinum* with human diseases is not yet clear.

Among our results, the alpha diversity Shannon and Simpson indices were found to be lower in women with cervical cancer, but these findings were without statistical significance. This result is consistent with previous studies showing no difference between healthy normal cases and precancerous lesions of the cervix [5,20]. Other results have shown that *Lactobacillus* with low bacterial diversity and low pH is dominant in the healthy vagina, whereas in the unhealthy vagina, it shows high bacterial diversity with low counts of lactobacilli and high counts of anaerobic bacteria [26]. This difference seems to be due to the sequencing method employed [15,27].

In the KEGG pathway, peptidoglycan biosynthesis is enriched in cases of cervical cancer. Peptidoglycan biosynthesis is also enriched in patients with symptomatic atherosclerosis [28], which indicates that the increased production of peptidoglycan from gut microbiota improves neutrophil functionality and the innate immune system, thus leading to symptomatic atherosclerosis. Inflammation has been identified as an important contributor to the pathogenesis of atherosclerosis. Our results suggest that the cervical metagenome might contribute to the development of cervical cancer and precancerous lesions by acting as a regulator of host inflammatory pathways. A bacterial cell-wall polymer, peptidoglycan, in its maintenance of cell integrity against osmotic pressure, is essential for cell survival [29].

Among the other results of the present study, lysine biosynthesis, methane metabolism, and renal cell carcinoma were enriched in CIN 2 or 3. Lysine biosynthesis has been found to be diminished in ulcerative colitis and Crohn's disease [30]. Lysine is an amino acid that is essential for humans, and is synthesized in bacteria via the diaminopimelate pathway, which acts in the first and second enzymatic reactions in the biosynthesis of isoleucine, methionine, and threonine. In these amino acids' biosynthesis, the initial enzymatic step is catalyzed by aspartokinase [31]. Methane metabolism is one of the major metabolic processes of colorectal carcinoma [32]. Methane, as one of the end products of fermentation in the gastrointestinal system, along with fermented gases including hydrogen, readily appears in breath [33]. Methane metabolism by bacteria in the large intestine has been reported as early as 1977. This report showed that the excretion of methane in breath occurred twice as frequently in patients with colonic cancer as in normal individuals [34]. This suggests a difference in the anaerobic intestinal flora between patients and normal subjects, and implies that colorectal cancer can be caused

by carcinogens formed by anaerobic bacteria's dehydrogenation of nuclear bile acid in the colon [35]. It is possible that these and other functional roles are related to cervical cancer, but further functional studies are needed before any firm conclusions can be drawn. Renal cell carcinoma is one of the significantly down-regulated pathways related to Parkinson's disease [36]. Meanwhile, some other pathways related to human diseases also have been identified as down-regulated in Parkinson's disease. The most important finding was that several cancer-related pathways were uncovered, including colorectal cancer, renal cell carcinoma, and endometrial cancer. This is consistent with previous reports of low cancer rates among patients with Parkinson's disease. It has been suggested that the high levels of total body potassium in Parkinson's patients might be a protective factor against cancer [37].

Biofilm formation in *Vibrio cholerae*, dioxin degradation, styrene degradation, benzoate degradation, xylene degradation, and cyanoamino acid metabolism were enriched in the present study's normal subjects. Biofilm formation in *Vibrio cholerae*, correspondingly, has been identified in associated pathways in Type 2 diabetes (T2D) patients [38]. Particularly, in an independent comparative study of the microbiomes of European T2D patients [38], cysteine and methionine metabolism was one of the most significant T2D-related pathways identified. A significant association with T2D only when applying a modified pipeline also has been found in biofilm formation in *Vibrio cholera*, both of which had high levels of blood sugar. These results suggest that identifying functional variation that is often masked by standard metagenomic bioinformatic processing can improve our ability to identify important associations between microbiome functionality and disease. A loss of the vitamin D receptor in rats has been reported to reduce the dioxin degradation pathway [39]. Benzoate degradation and styrene degradation were reported in *Helicobacter pylori*, which is associated with gastric cancer [40]. Xylene degradation [41] as well as polycyclic aromatic hydrocarbon and xylene degradation, have been found to be abundant in current smokers. These chemicals are components of cigarette smoke [42], and thus alterations in the ability to break down these substances can have toxic consequences for the host. It is surprising that some of the xenobiotic degradation pathways are depleted by smokers, due to the need for the bacterial up-regulation of these pathways to detoxify tobacco smoke. Cyanoamino acid metabolism, for instance, appears to be up-regulated in stage II rectal carcinoma tissues [43].

Our study has several strengths. First, the use of whole-genome shotgun metagenomic sequencing for microbiome analysis allowed us to not only comprehensively study overall bacterial community composition and specific oral taxon abundances, but also perform a functional analysis of the COG and KEGG pathways. Analysis using shotgun metagenomic sequencing can find and identify unknown microbiota that can play a role in the process of cervical carcinogenesis. It has been reported that there are differences in 16S rRNA sequencing and whole-genome shotgun metagenomic sequencing [15,27]. 16S rRNA sequencing can be biased due to unequal amplification of the 16S rRNA gene in species, although shotgun metagenomic sequencing might not be deep enough to detect the 16S rRNA gene of rare species. Second, in this study, we examined the compositional and functional alterations of the cervical microbiomes in cervical carcinogenesis for normal group, CINs as precancerous lesions, and cervical cancer. The present report is, as far as we know, the first to have been based on the utilization of shotgun metagenomic sequencing to study the compositional and functional alterations of cervical microbiomes in cervical carcinogenesis. However, the study on which the report is based was limited by an insufficient sample size for determination of the association of cervical metagenomes with cervical carcinogenesis. Further studies that boast larger sample sizes or pooled cross-study data and that are reproducible and have sufficient power to detect significant associations are needed. HPV infection status, HPV genotype, and sexual activity were not considered as potential variables in this study, and their exclusion could have biased the results [44]. Finally, the case-control studies were limited to one point in time in order to prevent the direction of a series of events. Moreover, case-control studies are inherently limited in their causality-inference utility.

## 4. Materials and Methods

### 4.1. Study Design and Subjects

From March 2006 to the present, the Korean HPV cohort study including women aged 18–65 years has been ongoing. Details on the cohort design criteria are available in a previous paper [45]. The subjects enrolled in the study were patients who had been given a diagnosis of CIN or cervical cancer that was histologically proven. Subjects receiving any therapy or surgery or using immunosuppressive agents were excluded at enrollment. Detailed self-administered health and lifestyle questionnaires, including questions on behavior related to alcohol consumption, were completed at enrollment. From the patients with CIN 2/3 or cervical cancer, a cervical swab was obtained at the first visit prior to any treatment such as Large Loop Excision of the Transformation Zone (LLETZ), surgical or radiation therapy, or chemotherapy. Cervical swabs were collected for Papanicolaou smear tests and tumor HPV DNA tests using a Cervical Sampler Brush (Cervical Sampler, Digene Co., MD, USA). A total of 47 women were randomly selected from 18 normal subjects, 17 CIN 2 or 3 subjects, and 12 cervical cancer patients. All of the study participants consented in writing according to the requirements of the Institutional Review Board. This study was approved by the Institutional Ethics Committee of the National Cancer Center of Korea (IRB No. NCC2016-0147).

### 4.2. DNA Extraction, Sequencing, and Quality Check

Metagenomic DNA samples were extracted using the FastDNA Spin extraction kit (MP Biomedicals, Santa Ana, CA, USA). All of the samples were sequenced by Illumina HiSeq2500 at Chunlab, Inc. in Seoul, Korea. Paired-end reads were generated with 250 bp. The average insert size of the libraries was 350 bp.

After sequencing, FastQC [46] was used to check the read quality. Since the raw data includes host (human) genome and adapter sequences, raw sequence reads were filtered and trimmed with KneadData [47] in order to reduce biases in the analysis. The Trimmomatic [48] option in KneadData was used with 'ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36'. The analysis was carried out using the non-host, non-adapter sequence.

### 4.3. Taxonomical Analysis

K-mer-based taxonomical assignment was performed with CLAssifier based on Reduced K-mers (CLARK) [49]. For each sample, all of the reads were classified using the National Center for Biotechnology Information (NCBI) RefSeq database (bacteria and viruses) from phylum to species. The classification algorithm was based on discriminative k-mers, and the classification was performed with the parameter '-m 0 (full mode)'. After classification, the abundance was estimated using the count, and the proportion of each target was identified using CLARK with the given parameter: '-highconfidence'.

### 4.4. De Novo Assembly and Gene Prediction

De novo assembly was performed for each sample with IDBA-UD [50]. Based on the de Bruijn graph approach, the optimal result was selected from among the multiple k-mers (73, 83, 93, and 101). The assembled contigs were classified with Fragment Classification Package (FCP) [51] using the NCBI RefSeq and taxonomy databases. First, classification with Naïve Bayes (NB) was performed. Next, classification with Epsilon-NB was performed, after which the numbers of fragments and base pairs assigned to the different taxonomic categories were calculated.

Finding genes and assigning their functions are two of the main purposes of shotgun metagenomic sequencing. Gene prediction was performed with Prodigal [52] in the metagenomics mode. After gene prediction, the obtained genes were searched against the COG [16] databases using Basic Local

Alignment Search Tool (BLAST) [53] with an  $E$ -value threshold of  $1 \times 10^{-5}$ , in order to carry out functional annotation.

Finally, to identify which pathways the predicted genes are involved in, the KEGG [54–56] database was used with BLAST under the same conditions. Bacteria and virus sequences from the KEGG database were used for the analysis. Also, the number of genes matched for each KO were found using an in-house script and compared by sample.

#### 4.5. Statistical Analysis

Differences in the demographic and clinical characteristic of the participants were subjected to the *Kruskal–Wallis* test for continuous variables. Categorical variables were analyzed by the *chi-squared* test, and Fisher's exact test was performed when more than 25% of the cells with an expected frequency of less than five were used. The alpha diversities of the Shannon and Simpson indices were calculated for each group. Beta diversity was calculated with principal coordinates analysis (PCoA) according to the Bray–Curtis and Jaccard distances. A permutational multivariate analysis of variance (PERMANOVA) was implemented to determine significance in distance. Diversity and the PERMANOVA results were analyzed using the R package's "vegan" [57]. To determine the taxonomic and genetic features that were differentially abundant, LEfSe was performed with  $\alpha = 0.5$  and an LDA score = 2.5 as the standard [58]. The post hoc analysis used the Bonferroni method with the R package's "PMCMR" [59]. The visualization used the ggplot2 package [60]. The analysis was performed with SAS 9.4 (SAS Institute, Cary, NC, USA) and the R platform (version 3.4.3) (The R Foundation for Statistical Computing, Vienna, Austria).

## 5. Conclusions

In summary, in this case-control study of cervical metagenomics using shotgun metagenomic sequencing, we observed significantly different microbial abundance and enriched metabolic functions in cervical metagenomes between normal, CIN 2/3, and cervical cancer. At the genus level, the microbiota of cervical cancer was differentially enriched with the genera *Alkaliphilus*, *Pseudothermotoga*, and *Wolbachia*; CIN 2/3 were enriched with *Lactobacillus*, *Staphylococcus*, and *Candidatus Endolissoclinum*, and the normal group was enriched with *Pseudoalteromonas* and *Psychrobacter*. Further characterization of the functional capacities of the metagenomes revealed that six KOs involved in 10 pathways are associated with an increased risk of CIN2/3 and cervical cancer. Specifically, cervical metagenomes were enriched in peptidoglycan synthesis and depleted in dioxin degradation and 4-oxalocrotonate tautomerase. The COG category of 'Defense mechanisms' was depleted in cervical cancer patients. This is the first study using shotgun metagenomic sequencing for the cervical carcinogenesis stage. Even though our study cannot provide evidence of direct causation, these findings demonstrate differences in cervical microbiomes and their functional bacteria-involved pathways among normal, CIN 2 or 3, and cervical cancer patients.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-6694/11/3/309/s1>, Figure S1: Box plots of relative abundances of major genuses, Figure S2: Differences among genus level, Figure S3: PCoA in phylum, Figure S4: Alpha indices in genera, Figure S5: Comparison of KOs using LEfSe to identify differentially abundant KEGG pathways, Figure S6: PCoA in KEGG pathway, Table S1: Mean cervical microbial abundances in 31 phyla, Table S2: Mean relative abundances of COG categories.

**Author Contributions:** All of the authors contributed significantly to this work. M.K. and S.-S.S. designed the research study; S.-S.S., D.O.L., and M.C.L. enrolled the subjects; M.K. analyzed the results; M.K., S.-S.S., and M.K.K. drafted the article. All authors read and approved the final manuscript.

**Funding:** This work was funded by the Korea National Cancer Center (Grant No. 1910180).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

CIN	cervical intraepithelial neoplasia
LEfSe	Linear discriminant analysis (LDA) effect size
COG	Cluster of Orthologous Groups
KEGG	Kyoto Encyclopedia of Genes and Genomes
PCoA	principal coordinates analysis
PERMANOVA	Permutational multivariate analysis of variance
Kos	KEGG orthologies
LLETZ	Large Loop Excision of the Transformation Zone

## References

- Schiffman, M.; Castle, P.E.; Jeronimo, J.; Rodriguez, A.C.; Wacholder, S. Human papillomavirus and cervical cancer. *Lancet* **2007**, *370*, 890–907. [[CrossRef](#)]
- Munoz, N.; Bosch, F.X.; de Sanjose, S.; Herrero, R.; Castellsague, X.; Shah, K.V.; Snijders, P.J.; Meijer, C.J. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.* **2003**, *348*, 518–527. [[CrossRef](#)] [[PubMed](#)]
- Gadducci, A.; Barsotti, C.; Cosio, S.; Domenici, L.; Riccardo Genazzani, A. Smoking habit, immune suppression, oral contraceptive use, and hormone replacement therapy use and cervical carcinogenesis: A review of the literature. *Gynecol. Endocrinol.* **2011**, *27*, 597–604. [[CrossRef](#)] [[PubMed](#)]
- Kyrgiou, M.; Mitra, A.; Moscicki, A.B. Does the vaginal microbiota play a role in the development of cervical cancer? *Transl. Res.* **2017**, *179*, 168–182. [[CrossRef](#)] [[PubMed](#)]
- Mitra, A.; MacIntyre, D.A.; Lee, Y.S.; Smith, A.; Marchesi, J.R.; Lehne, B.; Bhatia, R.; Lyons, D.; Paraskevaidis, E.; Li, J.V.; et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. *Sci. Rep.* **2015**, *5*, 16865. [[CrossRef](#)] [[PubMed](#)]
- Zhang, C.; Liu, Y.; Gao, W.; Pan, Y.; Gao, Y.; Shen, J.; Xiong, H. The direct and indirect association of cervical microbiota with the risk of cervical intraepithelial neoplasia. *Cancer Med.* **2018**, *7*, 2172–2179. [[CrossRef](#)] [[PubMed](#)]
- Aviles-Jimenez, F.; Yu, G.; Torres-Poveda, K.; Madrid-Marina, V.; Torres, J. On the Search to Elucidate the Role of Microbiota in the Genesis of Cancer: The Cases of Gastrointestinal and Cervical Cancer. *Arch. Med. Res.* **2017**, *48*, 754–765. [[CrossRef](#)] [[PubMed](#)]
- Antonio, M.A.; Hawes, S.E.; Hillier, S.L. The identification of vaginal Lactobacillus species and the demographic and microbiologic characteristics of women colonized by these species. *J. Infect. Dis.* **1999**, *180*, 1950–1956. [[CrossRef](#)] [[PubMed](#)]
- Gillet, E.; Meys, J.F.; Verstraelen, H.; Verhelst, R.; de Sutter, P.; Temmerman, M.; Vanden Broeck, D. Association between bacterial vaginosis and cervical intraepithelial neoplasia: Systematic review and meta-analysis. *PLoS ONE* **2012**, *7*, e45201. [[CrossRef](#)] [[PubMed](#)]
- Livengood, C.H. Bacterial Vaginosis: An Overview for 2009. *Rev Obstet. Gynecol.* **2009**, *2*, 28–37. [[PubMed](#)]
- Oh, H.Y.; Kim, B.S.; Seo, S.S.; Kong, J.S.; Lee, J.K.; Park, S.Y.; Hong, K.M.; Kim, H.K.; Kim, M.K. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. *Clin. Microbiol. Infect.* **2015**, *21*, 674.e1–674.e9. [[CrossRef](#)] [[PubMed](#)]
- Seo, S.S.; Oh, H.Y.; Lee, J.K.; Kong, J.S.; Lee, D.O.; Kim, M.K. Combined effect of diet and cervical microbiome on the risk of cervical intraepithelial neoplasia. *Clin. Nutr.* **2016**, *35*, 1434–1441. [[CrossRef](#)] [[PubMed](#)]
- Oh, H.Y.; Seo, S.S.; Kong, J.S.; Lee, J.K.; Kim, M.K. Association between Obesity and Cervical Microflora Dominated by Lactobacillus iners in Korean Women. *J. Clin. Microbiol.* **2015**, *53*, 3304–3309. [[CrossRef](#)] [[PubMed](#)]
- Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **2012**, *486*, 207–214. [[CrossRef](#)] [[PubMed](#)]
- Shah, N.; Tang, H.; Doak, T.G.; Ye, Y. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac. Symp. Biocomput.* **2011**, *16*, 165–176.
- Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; et al. The COG database: An updated version includes eukaryotes. *BMC Bioinform.* **2003**, *4*, 41. [[CrossRef](#)] [[PubMed](#)]

17. Erickson, A.R.; Cantarel, B.L.; Lamendella, R.; Darzi, Y.; Mongodin, E.F.; Pan, C.; Shah, M.; Halfvarson, J.; Tysk, C.; Henrissat, B.; et al. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS ONE* **2012**, *7*, e49138. [[CrossRef](#)] [[PubMed](#)]
18. Yildiz, S.; Dogan, I.; Dogruman-Al, F.; Nalbantoglu, U.; Ustek, D.; Sarzhanov, F.; Yildirim, S. Association of Enteric Protist *Blastocystis* spp. and Gut Microbiota with Hepatic Encephalopathy. *J. Gastrointestin. Liver Dis.* **2016**, *25*, 489–497. [[CrossRef](#)] [[PubMed](#)]
19. Simón, F.; Prieto, G.; Morchón, R.; Bazzocchi, C.; Bandi, C.; Genchi, C. Immunoglobulin G Antibodies against the Endosymbionts of Filarial Nematodes (Wolbachia) in Patients with Pulmonary Dirofilariasis. *Clin. Diagn. Lab. Immunol.* **2003**, *10*, 180–181. [[CrossRef](#)] [[PubMed](#)]
20. Piyathilake, C.J.; Ollberding, N.J.; Kumar, R.; Macaluso, M.; Alvarez, R.D.; Morrow, C.D. Cervical Microbiota Associated with Higher Grade Cervical Intraepithelial Neoplasia in Women Infected with High-Risk Human Papillomaviruses. *Cancer Prev. Res.* **2016**, *9*, 357–366. [[CrossRef](#)] [[PubMed](#)]
21. Noguchi, N.; Fukuzawa, M.; Wajima, T.; Yokose, K.; Suzuki, M.; Nakaminami, H.; Kawai, T.; Moriyasu, F.; Sasatsu, M. Specific clones of *Staphylococcus lugdunensis* may be associated with colon carcinoma. *J. Infect. Public Health* **2017**, *11*, 39–42. [[CrossRef](#)] [[PubMed](#)]
22. Vassallo, M.; Genillier, P.L.; Dunais, B.; Kaphan, R.; Saudes, L.; Duval, Y.; Rolland, F.; Jullien, V.; Weiss, N.; Blanchouin, E.; et al. Short-course daptomycin lock and systemic therapy for catheter-related bloodstream infections: A retrospective cohort study in cancer patients with surgically implanted devices. *J. Chemother* **2017**, *29*, 232–237. [[CrossRef](#)] [[PubMed](#)]
23. Gonzalez-Barca, E.; Carratala, J.; Mykietiuk, A.; Fernandez-Sevilla, A.; Gudiol, F. Predisposing factors and outcome of *Staphylococcus aureus* bacteremia in neutropenic patients with cancer. *Eur. J. Clin. Microbiol. Infect. Dis.* **2001**, *20*, 117–119. [[CrossRef](#)] [[PubMed](#)]
24. Peng, X.; Zhou, L.; Gong, Y.; Song, Z.; He, L.; Lin, S.; Zhang, J. Non-*pylori* *Helicobacters* (NHPHs) Induce Shifts in Gastric Microbiota in *Helicobacter pylori*—Infected Patients. *Front. Microbiol.* **2017**, *8*, 1038. [[CrossRef](#)] [[PubMed](#)]
25. Gao, Z.; Guo, B.; Gao, R.; Zhu, Q.; Qin, H. Microbiota dysbiosis is associated with colorectal cancer. *Front. Microbiol.* **2015**, *6*, 20. [[CrossRef](#)] [[PubMed](#)]
26. Brotman, R.M. Vaginal microbiome and sexually transmitted infections: An epidemiologic perspective. *J. Clin. Investig.* **2011**, *121*, 4610–4617. [[CrossRef](#)] [[PubMed](#)]
27. Vogtmann, E.; Hua, X.; Zeller, G.; Sunagawa, S.; Voigt, A.Y.; Hercog, R.; Goedert, J.J.; Shi, J.; Bork, P.; Sinha, R. Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. *PLoS ONE* **2016**, *11*, e0155362. [[CrossRef](#)] [[PubMed](#)]
28. Karlsson, F.H.; Fak, F.; Nookaew, I.; Tremaroli, V.; Fagerberg, B.; Petranovic, D.; Backhed, F.; Nielsen, J. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat. Commun.* **2012**, *3*, 1245. [[CrossRef](#)] [[PubMed](#)]
29. Rogers, H.J.; Perkins, H.R.; Ward, J.B. Biosynthesis of Peptidoglycan. In *Microbial Cell Walls and Membranes*; Rogers, H.J., Ed.; Chapman and Hall: London, UK, 1980; pp. 239–297.
30. Morgan, X.C.; Tickle, T.L.; Sokol, H.; Gevers, D.; Devaney, K.L.; Ward, D.V.; Reyes, J.A.; Shah, S.A.; LeLeiko, N.; Snapper, S.B.; et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* **2012**, *13*, R79. [[CrossRef](#)] [[PubMed](#)]
31. Umbarger, H.E. Amino acid biosynthesis and its regulation. *Annu. Rev. Biochem.* **1978**, *47*, 533–606. [[CrossRef](#)] [[PubMed](#)]
32. Zamani, Z.; Arjmand, M.; Vahabi, F.; Eshaq Hosseini, S.M.; Fazeli, S.M.; Iravani, A.; Bayat, P.; Oghalayee, A.; Mehrabanfar, M.; Haj Hosseini, R.; et al. A metabolic study on colon cancer using (1)h nuclear magnetic resonance spectroscopy. *Biochem. Res. Int.* **2014**, *2014*, 348712. [[CrossRef](#)] [[PubMed](#)]
33. Holma, R.; Korpela, R.; Sairanen, U.; Blom, M.; Rautio, M.; Poussa, T.; Saxelin, M.; Osterlund, P. Colonic methane production modifies gastrointestinal toxicity associated with adjuvant 5-fluorouracil chemotherapy for colorectal cancer. *J. Clin. Gastroenterol.* **2013**, *47*, 45–51. [[CrossRef](#)] [[PubMed](#)]
34. Haines, A.; Dilawari, J.; Metz, G.; Blendis, L.; Wiggins, H. Breath-methane in patients with cancer of the large bowel. *Lancet* **1977**, *2*, 481–483. [[CrossRef](#)]
35. Piqué, J.M.; Pallarés, M.; Cusó, E.; Vilar-Bonet, J.; Gassull, M.A. Methane production and colon cancer. *Gastroenterology* **1984**, *87*, 601–605. [[PubMed](#)]

36. Zhang, B.; Xia, C.; Lin, Q.; Huang, J. Identification of key pathways and transcription factors related to Parkinson disease in genome wide. *Mol. Biol. Rep.* **2012**, *39*, 10881–10887. [[CrossRef](#)] [[PubMed](#)]
37. Jansson, B.; Jankovic, J. Low cancer rates among patients with Parkinson's disease. *Ann. Neurol.* **1985**, *17*, 505–509. [[CrossRef](#)] [[PubMed](#)]
38. Manor, O.; Borenstein, E. Revised computational metagenomic processing uncovers hidden and biologically meaningful functional variation in the human microbiome. *Microbiome* **2017**, *5*, 19. [[CrossRef](#)] [[PubMed](#)]
39. Jin, D.; Wu, S.; Zhang, Y.G.; Lu, R.; Xia, Y.; Dong, H.; Sun, J. Lack of Vitamin D Receptor Causes Dysbiosis and Changes the Functions of the Murine Intestinal Microbiome. *Clin. Ther.* **2015**, *37*, 996–1009. [[CrossRef](#)] [[PubMed](#)]
40. Mohandhas, B.; Vennila, J.J.; Singh, S. Drug target identification in Helicobacter pylori associated with Gastric Cancer: An insilico approach. *J. Pharm. Res.* **2012**, *5*, 4303–4306.
41. Wu, J.; Peters, B.A.; Dominianni, C.; Zhang, Y.; Pei, Z.; Yang, L.; Ma, Y.; Purdue, M.P.; Jacobs, E.J.; Gapstur, S.M.; et al. Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J.* **2016**, *10*, 2435–2446. [[CrossRef](#)] [[PubMed](#)]
42. Rodgman, A.; Perfetti, T.A. *The Chemical Components of Tobacco and Tobacco Smoke*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2016.
43. Zhao, H.; Shi, Z.-Z.; Jiang, R.; Zhao, D.-B.; Zhou, H.-T.; Liang, J.-W.; Bi, X.-Y.; Zhao, J.-J.; Li, Z.-Y.; Zhou, J.-G. Metastasis associated genomic aberrations in stage II rectal cancer. *Genes Genom.* **2016**, *38*, 1085–1094. [[CrossRef](#)]
44. Oh, H.Y.; Seo, S.S.; Kim, M.K.; Lee, D.O.; Chung, Y.K.; Lim, M.C.; Kim, J.Y.; Lee, C.W.; Park, S.Y. Synergistic effect of viral load and alcohol consumption on the risk of persistent high-risk human papillomavirus infection. *PLoS ONE* **2014**, *9*, e104374. [[CrossRef](#)] [[PubMed](#)]
45. Hwang, J.H.; Lee, J.K.; Kim, T.J.; Kim, M.K. The association between fruit and vegetable consumption and HPV viral load in high-risk HPV-positive women with cervical intraepithelial neoplasia. *Cancer Causes Control* **2010**, *21*, 51–59. [[CrossRef](#)] [[PubMed](#)]
46. Andrews, S. FastQC: A Quality-Control Tool for High-Throughput Sequence Data. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 18 December 2018).
47. Huttenhower Lab. KneadData. Available online: <http://huttenhower.sph.harvard.edu/kneaddata> (accessed on 18 December 2018).
48. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
49. Ounit, R.; Wanamaker, S.; Close, T.J.; Lonardi, S. CLARK: Fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genom.* **2015**, *16*, 236. [[CrossRef](#)] [[PubMed](#)]
50. Peng, Y.; Leung, H.C.; Yiu, S.M.; Chin, F.Y. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012**, *28*, 1420–1428. [[CrossRef](#)] [[PubMed](#)]
51. Parks, D.H.; MacDonald, N.J.; Beiko, R.G. Classifying short genomic fragments from novel lineages using composition and homology. *BMC Bioinform.* **2011**, *12*, 328. [[CrossRef](#)] [[PubMed](#)]
52. Hyatt, D.; LoCascio, P.F.; Hauser, L.J.; Uberbacher, E.C. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **2012**, *28*, 2223–2230. [[CrossRef](#)] [[PubMed](#)]
53. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
54. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)] [[PubMed](#)]
55. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [[CrossRef](#)] [[PubMed](#)]
56. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [[CrossRef](#)] [[PubMed](#)]
57. Oksanen, J.; Blanchet, F.; Kindt, R.; Legendre, P.; Minchin, P.; O'Hara, R.; Simpson, G.; Solymos, P.; Stevens, M.; Wagner, H. vegan: Community Ecology Package, Version 2.4-1. Available online: <https://cran.r-project.org/web/packages/vegan/index.html> (accessed on 18 December 2018).
58. Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W.S.; Huttenhower, C. Metagenomic biomarker discovery and explanation. *Genome Biol.* **2011**, *12*, R60. [[CrossRef](#)] [[PubMed](#)]

59. Pohlert, T. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). R Package. 2014. Available online: <http://CRAN.R-project.org/package=PMCMR> (accessed on 18 December 2018).
60. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).