Pks+ Escherichia Coli More Prevalent in Benign than Malignant Colorectal Tumors

Carmina Villariba Tolentino
University of Santo Tomas
Ana Maria Cariño
University of Santo Tomas
Kin Israel Notarte
University of Santo Tomas Faculty of Medicine and Surgery
Imee Macaranas
University of Santo Tomas Faculty of Medicine and Surgery
Allan Fellizar
University of Santo Tomas
Rock Christian Tomas
University of the Philippines Los Banos
Lara Mae Angeles
University of Santo Tomas Hospital
Lorenzo Abanilla
Divine World Hospital
Antonio Lim
Divine World Hospital
Ma. Kristina Carmela Aguilar
University of Santo Tomas
Pia Marie Albano (psalbano@ust.edu.ph)
University of Santo Tomas College of Science https://orcid.org/0000-0002-5024-8819

Research Article

Keywords: pks+ E. coli, colorectal cancer, clbB, clbN, clbA

Posted Date: July 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-437863/v1

Version of Record: A version of this preprint was published at Molecular Biology Reports on July 23rd, 2021. See the published version at https://doi.org/10.1007/s11033-021-06552-1.

Pks^{+} *Escherichia Coli* More Prevalent in Benign than Malignant Colorectal Tumors

Carmina Villariba-Tolentino[1,2,3], Ana Maria Cariño[1,2,4], Kin Israel Notarte[5], Imee Macaranas[5], Allan Fellizar[1,2,6], Rock Christian Tomas[7], Lara Mae Angeles[5,8], Lorenzo Abanilla[9], Antonio Lim[9], Ma. Kristina Carmela Aguilar[1,2], and Pia Marie Albano[1,2,10*]

[1]*The Graduate School, University of Santo Tomas, Manila 1015, Philippines*

[2]*Research Center for the Natural and Applied Sciences, University of Santo Tomas, Manila 1015, Philippines*

[3]*Manuel S. Enverga University Foundation, Lucena City 4301, Philippines*

[4]*Quirino State University, Diffun 3401, Philippines*

[5]*Faculty of Medicine and Surgery, University of Santo Tomas, Manila 1015, Philippines*

[6]*Mariano Marcos Memorial Hospital and Medical Center, Batac 2906, Philippines*

[7]*University of the Philippines Los Baños, Los Baños, Laguna 4031, Philippines*

[8]*University of Santo Tomas Hospital, Manila 1015, Philippines*

[9]*Divine Word Hospital, Tacloban City 6500, Philippines*

[10]*College of Science, University of Santo Tomas, Manila, Philippines*

Corresponding author:

*psalbano@ust.edu.ph; p.albano2013@gmail.com

ABSTRACT

Background: Some *E. coli* strains that synthesize the toxin colibactin within the 54-kb *pks* island are being implicated in colorectal cancer (CRC) development. Here, the prevalence of *pks*$^+$ *E. coli* in malignant and benign colorectal tumors obtained from selected Filipino patients was compared to determine the association of *pks*$^+$ *E. coli* with CRC in this population.

Methods and Results: A realtime qPCR protocol was developed to quantify *uidA, clbB, clbN,* and *clbA* genes in formalin fixed paraffin embedded colorectal tissues. The number of malignant tumors (44/62; *71%*) positive for the *uidA* gene was not significantly different (p=0.3428) from benign (38/62; *61%*) tumors. Significantly higher number of benign samples (p<0.05) were positive for all three colibactin genes *(clbB, clbN, and clbA)* compared with malignant samples. There was also higher prevalence of *pks*$^+$ *E. coli* among older females and in tissue samples taken from the rectum.

Conclusion: Hence, *pks*$^+$ *E. coli* may not be associated with CRC development among Filipinos.

Keywords: *pks*$^+$ *E. coli*, colorectal cancer, *clbB, clbN, clbA*

Introduction

In 2018, the World Health Organization reported that about 1 in every 6 deaths worldwide is due to cancer, with colorectal cancer (CRC) ranking second in mortality rate, presenting great economic burden particularly in low- and middle-income countries (LMIC) [1]. In the Philippines, CRC ranks third in incidence for both sexes and across all age groups in 2020 [2]. Notably, mortality and morbidity related to CRC can be prevented by early screening as it takes up to 20 years for a non-cancerous growth to become malignant. Non-modifiable factors such as age, sex, hereditary mutations, and modifiable factors which include smoking, alcohol intake, medications, obesity, and diet have been identified to increase the risk of developing CRC. Of particular importance is diet since the gut microbiota heavily depends on nutraceuticals for their survival [3]. The gut microbiota has been of great interest lately because of its apparent role in the pathogenesis of immune-mediated/autoimmune, metabolic, cardiovascular, neuropsychiatric, and uremic diseases, and also cancer [4]. *Bacteroides fragilis, Fusobacterium nucleatum*, *Streptococcus gallolyticus*, *Helicobacter pylori, Enterococcus faecalis, Clostridioides difficile, Clostridium septicum*, and *Escherichia coli* were among the species associated with CRC development [5–10].

E. coli strains are the first to colonize the gastrointestinal tract within hours of birth, which later harmoniously coexist with its human host. However, some strains synthesize toxins, including cyclomodulins, which are known genotoxins and modulators of cellular differentiation, apoptosis, and proliferation [11]. Among these cyclomodulins is colibactin, which is synthesized by a hybrid non-ribosomal peptide synthetase-polyketide synthase (NRPS-PKS) assembly line found within the 54-kb *pks* island containing 19 genes (*clbA* to *clbS*) [12]. Colibactin has been shown to induce DNA double-strand breaks and transient G2-M cell cycle arrest in host mammalian cells. The infected host cells can eventually survive but the incomplete DNA repair may lead to higher mutation rates that drive tumorigenesis [13]. In

addition, infected host cells can also secrete growth factors that can be stimulatory to non-infected neighboring cells leading to their abnormal proliferation and tumor development [14]. Hence, this *pks* island that codes for colibactin has been associated with CRC development.

A study on Italian samples revealed that *pks*$^+$ *E. coli* colonize precancerous lesions, polyp lesions, and the normal tissues adjacent to these lesions but not the healthy normal mucosa [15]. Increased amounts of *pks*$^+$ *E. coli* among ulcerative colitis and CRC cases have also been noted [16]. Moreover, a meta-analysis presented that *pks*$^+$ *E. coli* strains were more prevalent among CRC and IBD patients compared with clinically healthy controls [17]. Yet, the prevalence of *pks*$^+$ *E. coli* in colonic lavage samples from Japanese patients with CRC was not significantly different from clinically healthy controls [18].

In the present study, the prevalence of *pks*$^+$ *E. coli* in malignant and benign colorectal tumors obtained from selected Filipino patients was compared. A realtime qPCR protocol was also developed, including design of primers to quantify select colibactin genes (*clbB, clbN* and *clbA*) using formalin fixed paraffin embedded (FFPE) biopsies.

Materials and methods

Study samples and sample preparation

A total of 140 FFPE colorectal tissue samples (70 malignant paired with 70 benign) collected from January 2015 to August 2018 were retrieved from the repositories of the hospital study sites. Samples from patients with history of inflammatory bowel disease (IBD), polyposis syndromes, and Lynch syndrome were excluded from this study.

Tissue sectioning was carried out with a microtome (Leica RM2235, Germany) following standard protocols. The outer sections were stained with hematoxylin and eosin (H&E), and then sent to two (2) external evaluators (pathologists) blinded of the original diagnosis to confirm presence or absence of cancer cells. The inner slices were collected in

nuclease-free tubes and stored at room temperature until DNA extraction and molecular analyses. Only the tissue samples in diagnostic concordance among external evaluators and original diagnosis of the study sites were considered for further molecular analysis.

Anti-contamination protocols were strictly followed. The working area and the entire microtome were cleaned with 70% ethyl alcohol and acetone, respectively; and the gloves and microtome blades were changed after every specimen processed. Pure paraffin blocks were also sectioned every ten (10) tissue samples to assess any cross-contamination.

DNA extraction

The number and thickness of sections generated were based on tissue size: ten 5 µm-thick sections if ≤0.2 cm^2 in size; five 5µm-thick sections if >0.2 cm^2 but ≤3.0 cm^2; and three 5µm-thick sections if >3.0 cm^2. DNA was extracted using an in-house protocol as described [19]. Briefly, the tissues were treated with proteinase K dissolved in aqueous solution of tris-HCl, sodium EDTA, and Tween 20, and incubated for 24 hr at 56°C with constant mixing at 300 RPM. Proteinase K was then inactivated at 72°C for 10 min and the mixtures were centrifuged at maximum speed for 2 min at 4ºC to allow the paraffin layer to form. The clear aqueous phase below the paraffin film was aspirated, transferred to a new nuclease-free tube, and stored at -20°C until use.

dsDNA control and primer set design for detection of *pks+ Escherichia coli*

The novel primer sequences that aimed for production of <100 bp amplicons were defined and experimentally tested. Sequence variation in each target gene (*uidA, clbB, clbN* and *clbA*) was assessed utilizing BLASTn and primer blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast). The database generated 5 to 10 pairs of primers indicating the number of base pair, GC content, position of sequence, length, and self-

complementarity [20]. Due to degradation of DNA from formalin fixation and paraffin embedding, the primers were designed to amplify target sequence <100 bp in length, with similar melting temperatures (Tm), 21-23 bp long, and 40-60% G/C content. *In silico* primer tests were performed with the NetPrimer software by Premier Biosoft (http://www.premierbiosoft.com/netprimer/index.html). Gene sequences were sent to IDT Integrated DNA Technologies, Inc. (Singapore) for synthesis.

To further check the efficiency of the designed primers, gradient run with annealing temperatures from 59 to 61°C was performed in a conventional PCR (T100 Thermal Cycler, BioRad, USA) using random colorectal DNA samples described above. Ten µl of the reaction mixture consisted of 1 µl DNA sample, 1 µl each of 10 µmol/l forward and reverse primers, 5 µl GoTaq Green Master Mix (Promega, USA), and 2 µl nuclease-free water. PCR products were subjected to 2% agarose gel electrophoresis for 45 mins at 135V and visualized under Gel Doc EZ Gel Documentation System (BioRad, USA).

Optimization of PCR conditions

Optimization of realtime qPCR conditions were done through gradient runs using 2 µl of a ten-fold serially diluted synthetic dsDNA (1 ng) in 20 µl reaction mixtures and 2 µl of random colorectal DNA sample in identical reaction mixtures. Optimum annealing temperatures were chosen by observing which temperature provided the earliest amplification and stable plateau of the synthetic dsDNA and DNA sample. Accuracy and efficiency of each primer pair were established through reproducible standard curves with 90-110% efficiency readings and 0.98-0.99 correlation coefficient (r^2) values derived using CFX Maestro software version 1.0 (Bio-Rad, USA).

Detection of pks+ *E. coli* by real-time qPCR

The *uidA* and colibactin genes (*clbB, clbN,* and *clbA*) were quantified in FFPE tissues by realtime qPCR (CFX96 Touch™ Real-Time PCR Detection System, Bio-Rad, USA). Twenty (20) µl reaction mixtures consisting of 2 µl template DNA, 0.4 µl each of 100 µM/l forward and reverse primer solutions, 7.2 µl nuclease-free water, and 10 µl iTaq Universal SYBR Green Supermix (Bio-Rad, USA) were used for analysis.

Data analysis

To determine the copy number present per nanogram (ng) sample of each synthetic dsDNA *uidA*, *clbB, clbN,* and *clbA*, the formula

$$N_C^{syn} = \frac{m \cdot N_A}{l \cdot k \cdot m_{DNA}} \quad (1)$$

was used; where N_C^{syn} is the gene copy number as a function of a variable amount of the synthetic dsDNA m in nanograms (ng) and number of base pairs l; while N_A, m_{DNA}, and k are constants, in which N_A is the Avogadro's number, m_{DNA} is the mass of one mole of a base pair assumed as 650 g, and k is a dimensionality constant equal to 1×10^9 units, which converts the units of m to grams [21].

To determine the copy number of *uidA*, *clbB, clbN,* and *clbA* present in the colorectal tumor tissues, the formula

$$N_C^{tis}(x) = \alpha + \beta \ln(x) \quad (2)$$

was used; where N_C^{tis} is the copy number of the target gene based on the Cq value x, while α and β are constants. The values of the constants were derived using the formulas

$$\alpha = \frac{\sum N_C^{syn} \cdot \sum x^2 - \sum x \cdot \sum (x \cdot N_C^{syn})}{n \cdot \sum x^2 - (\sum x)^2} \quad (3)$$

$$\beta = \frac{n \cdot \sum (x \cdot N_C^{syn}) - \sum x \cdot \sum N_C^{syn}}{n \cdot \sum x^2 - (\sum x)^2} \quad (4)$$

where n refers to the fold dilution series of the synthetic dsDNA. The Cq values and their corresponding cut-off values were internally determined by the CFX Maestro software.

To determine the prevalence of each gene in malignant and benign samples, their respective odds ratios were computed, and Fischer's exact t-test was used to determine if these values were statistically significant at 5% confidence. To further characterize the prevalence of each gene, the quantity and percentage of malignant and benign cases for each considered patient characteristic (age, sex, tumor site, and tumor grade) were tabulated. Only the samples which tested positive to all colibactin genes were evaluated for prevalence.

Results

Characteristics of the study participants and their samples

Of the 70 pairs of malignant and benign FFPE colorectal tissue blocks retrieved from the biological repositories, 62 pairs had concordant histopathologic readings by all evaluators and were included for molecular analysis. Of the 62 malignant samples, 43 were matched with the adjacent cancer-free tissues retrieved from the same CRC patients (*n*=43). The remaining 19 malignant samples were matched with cancer-free tissues removed from other patients who were of the same age and sex as the CRC patients. Median age at diagnosis of the CRC patients (*n*=62) was 72 years old (range: 22-88 y/o). More samples came from female participants (*n*=37/62; 59%). More malignant tumor samples originated from the rectum (41/62; *66%*) and were diagnosed as adenocarcinoma (55/62; *89%*); while for the benign tumors, they were mostly derived from the colon (56/62; *90%*) as lines of resection (43/62; *69%*). Only 39 samples had available information on their tumor grade, which were mostly poorly differentiated (25/39; *64%*) (Table 1).

In-house realtime qPCR protocol for detection of *uidA, clbB, clbN and clbA*

Accession numbers of the target genes *uidA* (KT311783.1)*, clbB* (JX280405.1), *clbN* (JX280402.1), *and clbA* (JX280403.1) obtained from NCBI website were noted to specify the

locations of each gene in the whole genome database. Several primers to detect the aforementioned genes have been reported [13, 14, 29, 18, 22–28]; however, new sets were designed to amplify PCR products that were <100 bp in length since the DNA samples were extracted from FFPE tissues. Optimum PCR conditions were achieved through several gradient runs using different annealing temperatures (50-63°C) set in 40 cycles. Efficiency ratings of 90 to 100% and r^2 ranging from 0.98 to 0.99 were obtained after several assays and replications. The primer sequences and PCR conditions designed and optimized in this study are listed in Table 2.

Prevalence of *pks*$^+$ *E. coli* in malignant versus benign colorectal tissues

Of the 43 same patient paired samples, 35 and 29 malignant and paratumorous (benign) samples tested positive for *E. coli uidA* gene, respectively. Among the 19 case control pairs, 9 malignant and 9 benign tissues tested positive to *uidA*. Of the *uidA* positive same patient samples, 9/35 and 25/29 tumorous and paratumorous (benign) samples tested positive to all three colibactin genes, *clbB, clbN,* and *clbA*. Among the *uidA* positive case control samples, 3/9 malignant and 2/9 benign samples were positive to all three colibactin genes. Combining the same patient and case control tissue pairs, no statistically significant difference (p=0.3428) has been noted as to the number of malignant (44/62; *71%*) and benign (38/62; *61%*) colorectal tissue samples that tested positive to *uidA*. However, significantly higher number of benign colorectal tissues ($p<1.3325 \times 10^{-4}$) were positive for all three colibactin genes compared with malignant samples (Table 3).

Correlation between *pks*$^+$ *E. coli* and clinical characteristics

The prevalence of *pks*$^+$ *E. coli* in colorectal tissues as deduced by quantitative PCR of the telltale genes was further correlated with the patient's profile. However, results were only

reported in frequency or percentages and no statistical analysis was done due to the limited sample size. Results showed that the prevalence of *uidA* in malignant and benign samples was not different regardless of age, sex, and tumor site. However, the colibactin genes were more prevalent among female patients, especially those above 60 years of age, and from samples taken from the rectum (Table 4).

Discussion

This study tried to investigate the possible association of *pks*$^{+}$ *E. coli* with colorectal tumor development among selected Filipinos. The malignant and benign colorectal tumor samples were initially sent to external reviewers and only the FFPE samples where all pathologists agreed on their diagnosis were included for further molecular analysis.

The extracted DNA were then analyzed for the *uidA* gene which encodes the enzyme beta-*D*-glucuronidase. Molina et al. (2015) reported that this gene is highly specific for detecting *E. coli* strain K12 by PCR [30]. Furthermore, the same gene is also used for detecting urosepsis strains of *E. coli* encoding cyclomodulins [31]. In this study, only the *uidA*-positive samples were analyzed for *clbB*, *clbN* and *clbA*, the colibactin synthesis genes of *E. coli* [28]. These three genes were chosen to represent the 19 colibactin genes found in the *pks* island. Specifically, *clbA* is located at the beginning of the assembly line and is responsible for the activation of the non-ribosomal peptide synthetase genes, *clbB* and *clbN* [12]. Colibactin *B* and *N* are carrier proteins that tether the growing colibactin chains [32]. The biosynthesis of colibactin in *E. coli* is reported to induce DNA interstrand crosslinks *in cellulo* and contributes to bacterial virulence [33]. Due to its cytopathic effects in infected epithelial cells, colibactin has been implicated in CRC development by increasing epithelial cell proliferation and tumor invasion [34].

BLAST® sequence analysis of the *pks* marker genes showed 100% homology with *E. coli* strains EcPF5, EcPF14, SCU-488, SCU-306, HB37, SCU-101, UPEC129, P14, NS-NP030, and KS-P019. Among the 19 colibactin genes that BLAST has primer sequences, only *clbA, clbB, clbM, clbN, clbQ, clbP, clbR,* and *clbS* are provided, which may also amplify the colibactin genes of *Klebsiella pneumoniae* and *Citrobacter koseri*. However, it must be noted that *E. coli* (*53.6%*) has significantly higher relative percentage of *pks* genes compared with *K. pneumoniae* (*17.9%*) and *C. koseri* (*7.1%*). Furthermore, among these three bacteria, the proteins encoded by the *pks* genes have only been detected so far in *E. coli* strains [23].

A significantly higher number of benign colorectal tissues tested positive for all three colibactin genes compared with malignant samples. This is in contrast to studies done in the UK, France, and Malaysia which reported that *pks*$^+$ *E. coli* was significantly higher in CRC patients compared to non-cancer patients [34–36]. However, a study conducted among a Japanese population reported no significant difference in the prevalence of *pks*$^+$ genes between CRC cases and healthy controls [18]. It is postulated that microbiota composition varies according to geographic area; thus, this might explain the discrepancy in the distribution of the *pks* island in the Philippines, Malaysia, Japan, and Europe (UK and France) [37]. Moreover, the difference in the type of tissue samples that were subjected to molecular analysis might have affected the results. Both the UK and France studies made use of fresh biopsy tissues; Malaysia used *in vitro* assays; whereas, Japan and this study utilized colonic lavage and FFPE colorectal tissues, respectively [35]. Several studies have reported that there was little variation in the bacterial communities present along the colon of CRC patients, whether it be a malignant or benign region. Same clonal *E. coli* has been documented along the colon of patients regardless of the presence of tumor [36]. Additionally, a study on cyclomodulin-producing *E. coli* B2 strains isolated from patients with colon cancer and diverticulitis were able to form biofilm but showed poor invasive and adherent activities. This proved that *pks*$^+$ *E. coli* could

colonize intestinal mucosa of the colon but the malignant transformation remains a question [38].

Both age and sex showed to affect the gut microbiota component of an individual. This study observed that the older female patients, whether with malignant or benign colorectal tumors, had higher prevalence of *pks*$^+$ *E. coli* than males. It is worth noting that the incidence of CRC per 100,000 population in the Philippines is at 23.7% for males and 15.1% for females [39]. In another study, the pathogenic cyclomodulin-positive *E. coli* strains were more prevalent in the mucosa of male patients with late-stage CRC [40].

In addition, the old age group has been observed to carry increased Proteobacteria population, which explains why the older patients in this study registered higher amounts of *E. coli* [41]. It would be interesting to know whether a more sedentary lifestyle associated with age is a risk factor for colonization with *pks*$^+$ *E. coli*.

This study also noted that *pks*$^+$ *E. coli* was more abundant in samples taken from the rectal tumors, whether benign or malignant. In the beginning of the study, when the tumor sections were retrieved, there were more malignant samples from the rectum which could have influenced the frequency of tumors that tested positive for *uidA*. In contrast, a study among Malaysian CRC patients showed that *pks*$^+$ *E. coli* was more abundant in the distal than proximal parts of the colon [36].

Results also showed that *pks*$^+$ *E. coli* is more predominant in poorly differentiated malignant tissues. Conversely, a study conducted in a Malaysian population noted higher prevalence of *pks*$^+$ *E. coli* in early stage CRC than later stages [36]. In another study, no significant difference was seen between CRC patients with *pks*$^+$ *E. coli* and *pks*$^-$ *E. coli* in terms of bacterial colonization, inflammatory score, neoplastic stage, and tumor node metastases grade [14]. This higher ratio of *pks*$^+$ bacteria in tumor cells may result to dormancy of tumor cells since *pks*$^+$ *E. coli* can induce cellular senescence. Meanwhile, a low bacteria : tumor ratio

induces senescence only to a smaller area of the tumor, while at the same time releasing growth factors that negates senescence leading to further tumor proliferation [42].

To our knowledge, this is the first study on the prevalence of *pks*$^{+}$ *E. coli* in malignant and benign colorectal tumors obtained from selected Filipino patients. Compared to previous studies that analyzed fresh biopsies and colonic lavage, the present study was limited to the use of archived FFPE colorectal tissue samples. FFPE remains to be widely used in molecular assays because of technical ease in tissue processing as well as it offers economic advantage for longitudinal tissue specimen storage [43]. The use of FFPE is also advantageous in settings where ethical clearance is a crucial consideration because archived specimens from accredited repositories can be retrieved and analyzed without the need for recruiting new patients [44]. However, one major challenge in using FFPE in molecular assays is the low quality and quantity of nucleic acids extracted from FFPE tissue blocks. Thus, the need to optimize and develop new protocols in extracting and amplifying DNA products from FFPE specimens is necessary [43]. In this study, a realtime qPCR protocol was developed, including the design of primers that can amplify <100 bp long sequence of the *uidA,* and colibactin genes. In addition, this study was only able to include a limited number of samples even though it is easier to retrieve FFPE samples than do perspective sampling. The lack of financial capabilities in developing countries such as the Philippines poses a great limitation in the collection and storage of FFPE samples; thus these practices are not a routine practice in hospitals [19]. As in this study, samples were collected only from two hospitals. Thus, a follow-up study that includes a greater number of participants should be explored to confirm any or lack of association of *pks*$^{+}$ *E. coli* with CRC development among Filipinos. Nonetheless, results of this study can provide baseline data on the prevalence of *pks*$^{+}$ *E. coli* in malignant and benign colorectal tissues from Filipino patients.

Funding

This study has been funded by the Commission on Higher Education (CHED) of the Philippines through its Discovery-Applied Research and Extension for Trans/Inter-disciplinary Opportunities (DARE TO) Research Grants.

Declarations

Authors' contribution

Conceptualization - CVT, PMA; *Data curation* - CVT, AMC, AF, MKCA, PMA; *Formal analysis* - CVT, RCT, PMA; *Funding acquisition* - CVT, AMC, PMA; *Investigation* - CVT, AMC, LMA, LA, AL, MKCA; *Methodology* - CVT, RCT, PMA; *Project administration* - CVT, AMC, AF, AL, MKCA, PMA; *Resources* - CVT, AMC, AF, LA, AL; *Supervision* – PMA; *Validation*; CVT, AMC, LMA, LA, AL, MKCA; *Visualization* - CVT, KIN, IM, RCT, PMA; *Writing – original draft* - CVT, KIN, IM, RCT, PMA; *Writing – review & editing* - CVT, AMC, KIN, IM, AF, RCT, LMA, LA, AL, MKCA, PMA

Conflicts of interest

The authors declare no conflict of interest.

Ethics approval

All procedures were performed in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethical clearance was obtained from the Mariano Marcos Memorial Hospital and Medical Center (MMMH-MC; Protocol No. MMMH-RERC-17-013) in Batac, Philippines and University of Santo Tomas Hospital (USTH; Protocol No. IRB-2018-04-058-IS) in Manila, Philippines prior to specimen retrieval.

Consent to Participate

Written informed consent from the patients or their legal guardians had been waived by the Research Ethics Review Committees (RERC) of the hospital study sites since the study was restricted to the use of archived FFPE specimens and did not involve additional invasive procedures nor pose risk or harm to subjects. Clinical data of patients were retrieved from the medical records of the hospital study sites.

Consent to Publish

Publication of the results of this work has been approved by the RERC of the hospital study sites while maintaining the identity of the owners of the FFPE samples confidential.

References

[1] World HealthOrganization. Cancer, https://www.who.int/news-room/fact-sheets/detail/cancer (2018, accessed 2 February 2021).

[2] International Agency for Research on Cancer (IARC). Philippines. *Global Cancer Statistics*, https://gco.iarc.fr/today/data/factsheets/populations/608-philippines-fact-sheets.pdf (2020, accessed 2 February 2021).

[3] Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Gastroenterol Rev* 2019; 14: 89–103.

[4] Kho ZY, Lal SK. The Human Gut Microbiome – A Potential Controller of Wellness and Disease. *Front Microbiol* 2018; 9: 1835.

[5] Gagnière J, Raisch J, Veziant J, et al. Gut microbiota imbalance and colorectal cancer. *World J Gastroenterol 2016* 2016; 22: 501–518.

[6] Boleij A, Hechenbleikner EM, Goodwin AC, et al. The *Bacteroides fragilis* toxin gene is prevalent in the colon mucosa of colorectal cancer patients. *Clin Infect Dis* 2015; 60:

208–215.

[7] Viljoen KS, Dakshinamurthy A, Goldberg P, et al. Quantitative profiling of colorectal cancer-associated bacteria reveals associations between *Fusobacterium spp*., enterotoxigenic *Bacteroides fragilis* (ETBF) and clinicopathological features of colorectal cancer. *PLoS ONE* 2015; 10: 1–21.

[8] Guevarra Jr L, Afable AC, Belza PJ, et al. Immunogenicity of a *Fap2* peptide mimotope of *Fusobacterium nucleatum* and its potential use in the diagnosis of colorectal cancer. *Infect Agent Cancer*; 13. Epub ahead of print 2018. DOI: 10.1186/s13027-018-0184-7.

[9] Magat EM, Balanag GA, Cariño AM, et al. Clostridioides difficile antibody response of colorectal cancer patients versus clinically healthy individuals. *Biosci Microbiota, Food Heal* 2020; 39: 123–127.

[10] Butt J, Romero-Hernández B, Pérez-Gómez B, et al. Association of *Streptococcus gallolyticus* subspecies *gallolyticus* with colorectal cancer: Serological evidence. *Int J Cancer* 2016; 138: 1670–1679.

[11] Buc E, Dubois D, Sauvanet P, et al. High Prevalence of Mucosa-Associated *E. coli* Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS ONE*; 8. Epub ahead of print 2013. DOI: 10.1371/journal.pone.0056964.

[12] Faïs T, Delmas J, Barnich N, et al. Colibactin: More than a new bacterial toxin. *Toxins*; 10. Epub ahead of print 2018. DOI: 10.3390/toxins10040151.

[13] Cuevas-ramos G, Petit CR, Marcq I, et al. *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *PNAS* 2010; 107: 11537–11542.

[14] Cougnoux A, Dalmasso G, Martinez R, et al. Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut*

2014; 1–11.

[15] Sarshar M, Scribano D, Marazzato M, et al. Genetic diversity, phylogroup distribution and virulence gene profile of *pks* positive *Escherichia coli* colonizing human intestinal polyps. *Microbial Pathogenesis* 2017; 275: 1–10.

[16] Prorok-Hamon M, Friswell MK, Alswied A, et al. Colonic mucosa-associated diffusely adherent *afaC+ Escherichia coli* expressing lpfA and pks are increased in inflammatory bowel disease and colon cancer. *Gut* 2014; 63: 761–70.

[17] Song M, Chan AT, Sun J. Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer. *Gastroenterology* 2020; 158: 322–340.

[18] Shimpoh T, Hirata Y, Ihara S, et al. Prevalence of *pks*-positive *Escherichia coli* in Japanese patients with or without colorectal cancer. *Gut Pathog* 2017; 9: 1–8.

[19] Albano PM, Holzinger D, Salvador C, et al. Low prevalence of human papillomavirus in head and neck squamous cell carcinoma in the northwest region of the Philippines. *PLOS ONE* 2017; 12:e0172240.

[20] Ye J, Coulouris G, Zaretskaya I, et al. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*; 13.

[21] Staroscik A. Calculator for determining the number of copies of a template. *URI Genomics & Sequencing Center*, http://cels.uri.edu/gsc/cndna.html (2004, accessed 20 March 2020).

[22] Oliero M, Calvé A, Fragoso G. Inulin and Galacto-oligosaccharides Increase the Genotoxic Effect of Colibactin Produced by *pks*$^{+}$ *Escherichia coli* Strains. *BMC Cancer* 2020; 1–13.

[23] Morgan RN, Saleh SE, Farrag HA, et al. Prevalence and pathologic effects of colibactin and cytotoxic necrotizing factor-1 (*Cnf 1*) in *Escherichia coli*: Experimental and bioinformatics analyses. *Gut Pathogens* 2019; 11: 1–18.

411 [24] Lee E, Lee Y. Prevalence of *Escherichia coli* carrying pks islands in bacteremia
412 patients. *Ann Lab Med* 2018; 38: 271–273.
413 [25] Sarshar M, Scribano D, Marazzato M, et al. Genetic diversity, phylogroup distribution
414 and virulence gene profile of *pks* positive *Escherichia coli* colonizing human intestinal
415 polyps. *Microb Pathog* 2017; 112: 274–278.
416 [26] Secher T, Payros D, Brehin C, et al. Oral tolerance failure upon neonatal gut
417 colonization with *Escherichia coli* producing the genotoxin colibactin. *Infect Immun*
418 2015; 83: 2420–2429.
419 [27] Suresh A, Ranjan A, Jadhav S, et al. Molecular genetic and functional analysis of pks-
420 Harboring, extra-intestinal pathogenic *Escherichia* colifrom India. *Front Microbiol*
421 2018; 9: 1–8.
422 [28] Johnson JR, Johnston B, Kuskowski MA, et al. Molecular epidemiology and
423 phylogenetic distribution of the *Escherichia coli* pks genomic island. *J Clin Microbiol*
424 2008; 46: 3906–3911.
425 [29] Bossuet-Greif N, Vignard J, Taieb F, et al. The colibactin genotoxin generates DNA
426 interstrand cross-links in infected cells. *mBio* 2018; 9: 1–15.
427 [30] Walker DI, McQuillan J, Taiwo M, et al. A highly specific *Escherichia coli* qPCR and
428 its comparison with existing methods for environmental waters. *Water Research* 2017;
429 126: 101–110.
430 [31] Dubois D, Delmas J, Cady A, et al. Cyclomodulins in urosepsis strains of *Escherichia*
431 *coli*. *Journal of Clinical Microbiology* 2010; 48: 2122–2129.
432 [32] Miyanaga A, Kudo F, Eguchi T. Protein-protein interactions in polyketide synthase-
433 nonribosomal peptide synthetase hybrid assembly lines. *Natural Product Reports*
434 2018; 35: 1185–1209.
435 [33] Faïs T, Delmas J, Barnich N, et al. Colibactin: More than a new bacterial toxin. *Toxins*

436 2018; 10: 16–18.
437 [34] Buc E, Dubois D, Sauvanet P, et al. High Prevalence of Mucosa-Associated *E. coli*
438 Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS ONE* 2013; 8:
439 e56964.
440 [35] Arthur JC, Perez-Chanona E, Muhlbauer M, et al. Intestinal Inflammation Targets
441 Cancer-Inducing Activity of the Microbiota. *Science* 2012; 338: 120–123.
442 [36] Iyadorai T, Mariappan V, Vellasamy KM, et al. Prevalence and association of *pks*$^+$
443 *Escherichia coli* with colorectal cancer in patients at the University Malaya Medical
444 Centre, Malaysia. *PLoS ONE* 2020; 15: e0228217.
445 [37] Lozupone CA, Stombaugh JI, Gordon JI, et al. Diversity, stability and resilience of the
446 human gut microbiota. *Nature* 2012; 489: 220–230.
447 [38] Raisch J, Buc E, Bonnet M, et al. Colon cancer-associated B2 *Escherichia coli*
448 colonize gut mucosa and promote cell proliferation. *World J Gastroenterol* 2014; 20:
449 6560–6572.
450 [39] Globocan 220. *Philippine Fact Sheets (Cancer)*. 2020.
451 [40] Bonnet M, Buc E, Sauvanet P, et al. Colonization of the Human Gut by *E . coli* and
452 Colorectal Cancer Risk. *Clinical Cancer Research* 2014; 20: 859–867.
453 [41] Rinninella E, Raoul P, Cintoni M, et al. What is the healthy gut microbiota
454 composition? A changing ecosystem across age, environment, diet, and diseases.
455 *Microorganisms* 2019; 7: 14.
456 [42] Dalmasso G, Cougnoux A, Delmas J, et al. The bacterial genotoxin colibactin
457 promotes colon tumor growth by modifying the tumor microenvironment. *Gut*
458 *Microbes* 2015; 5: 675–680.
459 [43] Zhang P, Lehmann BD, Shyr Y, et al. The Utilization of Formalin Fixed-Paraffin-
460 Embedded Specimens in High Throughput Genomic Studies. *Int J Genomics*; Article

ID. Epub ahead of print 2017. DOI: 10.1155/2017/1926304.

[44] Mathieson W, Thomas G. Using FFPE Tissue in Genomic Analyses: Advantages, Disadvantages and the Role of Biospecimen Science. *Curr Pathobiol Rep* 2019; 7: 35–40.

Table 1. Tumor characteristics

Characteristics	**Malignant (*n* = 62)***	**Benign tumor (*n* = 62)***
Tumor site		
Colon	21 *(34%)*	56 *(90%)*
Rectum	41 *(66%)*	6 *(10%)*
Classification**	Adenocarcinoma (*n*=55)	Lines of resection (*n*=43)
	Mucinous carcinoma (*n*=3)	Tubulovilous adenoma (*n*=13)
	Carcinoma signet ring cell (*n*=3)	Chronic active colitis (*n*=4)
	Adenosquamous cell (*n*=1)	Benign colonic mucosa (*n*=2)
Tumor grade**	Well differentiated (*n*=5)	
	Moderately differentiated (*n*=9)	
	Poorly differentiated (*n*=25)	
	No data available (*n*=23)	

*The FFPE colorectal tissue samples collected from the hospital study sites were subjected to another round of histopathologic analysis by two (2) external evaluators (pathologists). Only the tissue samples in diagnostic concordance among external evaluators and original diagnosis of the study sites were considered for further molecular analysis.

**Information were retrieved from medical records of the hospital study sites.

472 **Table 2.** Primers designed to detect the *uidA* and colibactin genes

Accession number	Target gene	Primer sequence	qPCR conditions
KT311783.1	*uidA* (74 bp)	(F) 5'-GTGAATCCACACCTCTGGCA-3' (R) 5'-TGTCTGGCTTTTGGCTGTGA-3'	95°C 3 mins, 95°C 10 secs, 52.7 °C 30 secs 39x
JX280405.1	*clbB* (77 bp)	(F) 5'-CGTATTACCCGGCCCACATT-3' (R) 5'-GTGCTGTTTAGCGAAGGTGC-3'	95°C 3 mins, 95°C 10 secs, 62.5 °C 30 secs 39x
JX280402.1	*clbN* (76 bp)	(F) 5'-CGGCAACACTTTCAGCACAA-3' (R) 5'-TGCGTTGGTTTACGCAGTTG-3'	95 °C 3 mins, 95 °C 10 secs, 61.7 °C 30 secs, 39x
JX280403.1	*clbA* (81 bp)	(F) 5'-ACTCCACAGGAAGCTACTTAACA-3' (R) 5'-TGAGCGTCCACATTTTCCAA-3'	95°C 3 mins, 95°C at 10 secs, 60 °C 30 secs, 39x

473

Table 3. Prevalence of pks$^+$ *E. coli* in malignant (*n*=62) and benign (*n*=62) colorectal tumors

Gene	Malignant tissue		Benign tissue		OR (95% CI)	*p*-value*
	n	%	n	%		
uidA	44/62	*71*	38/62	*61*	0.65 (0.31-1.37)	0.3428
clbB, clbN, and *clbA*	12/44	*27*	27/38	*71*	6.5455 (2.4931-17.1844)	1.3325×10^{-4}

*Fisher's exact test; $p<0.05$ is considered significant. ***Note****:* Of the 62 tissue pairs, 43 were same patient samples and 19 were case controls. Of the 43 pairs, 35 and 29 malignant and paratumorous (benign) samples were positive for *uidA*, respectively. Among the 19 case control samples, 9 malignant and 9 benign samples tested positive for *uidA*. Of the *uidA* positive same patient samples, 9/35 and 25/29 tumorous and paratumorous (benign) samples tested positive to all three colibactin genes, *clbB, clbN,* and *clbA*. Meanwhile, among the case control samples, 3 malignant and 2 benign tissues were positive to the three colibactin genes.

Table 4. Association of patient characteristics with prevalence of pks+ *E. coli*

Characteristics		***uidA* (+)**				***clbB, clbN, and clbA* (+)**			
		Malignant		**Benign**		**Malignant**		**Benign**	
		n = 44	*(%)*	*n* = 38	*(%)*	*n* = 12	*(%)*	*n* = 27	*(%)*
Age	*< 60 yrs old*	14	*32*	12	*32*	2	*17*	8	*30*
	> 60 yrs old	30	*68*	26	*68*	10	*83*	19	*70*
Sex	*female*	28	*64*	25	*66*	8	*67*	18	*67*
	male	16	*36*	13	*34*	4	*33*	9	*33*
Tumor site	*rectum*	29	*66*	25	*66*	8	*67*	16	*59*
	colon	15	*34*	13	*34*	4	*33*	11	*41*
Tumor grade	*well*	5	*17*			1			
	moderate	7	*24*			2			
	poor	17	*59*			6			
	No information	15				3			