Pathology International

WILEY | Pathology International

ORIGINAL ARTICLE OPEN ACCESS

A Homology-Based Application to Diagnose Colorectal Adenoma and Early-Stage Cancer

Yuhki Yokoyama[1] | Kazuki Kanayama[2] | Miwa Matsuda[3] | Nobuaki Hatori[4] | Taiki Asano[5] | Eiichi Morii[5] | Hirofumi Yamamoto[1] | Kazuaki Nakane[1]

[1]Department of Molecular Pathology, Division of Health Sciences, Graduate School of Medicine, The University of Osaka, Osaka, Japan | [2]Department of Biophysics, Graduate School of Health Sciences, Kobe University, Hyogo, Japan | [3]Department of Medical Science and Technology, Hiroshima International University, Hiroshima, Japan | [4]The Research Foundation for Microbial Diseases of Osaka University, Osaka, Japan | [5]Department of Diagnostic Pathology, Osaka University Hospital, Osaka, Japan

Correspondence: Yuhki Yokoyama (yyokoyama@sahs.med.osaka-u.ac.jp)

Received: 26 August 2025 | **Revised:** 29 April 2026 | **Accepted:** 15 June 2026

Funding: Japan Society for the Promotion of Science, Grant/Award Number: JP 26310209; Ministry of Education, Culture, Sports, Science and Technology

Keywords: adenoma | colorectum | early-stage cancer | homology

ABSTRACT

Homology is a mathematical concept involving two Betti numbers—b_0, representing the number of independent objects, and b_1, representing the number of holes formed by surrounding objects—which can be used to quantify the contact degree. We previously demonstrated the use of a homology-based method to diagnose advanced colorectal cancer. It is clinically important to discriminate adenoma, especially high-grade adenoma, and early-stage cancer from non-cancerous tissue; thus, it was of interest to evaluate whether our homology-based method could also diagnose adenoma and early-stage cancer. In the present study, we developed a homology-based application to calculate b_1 values for a region of interest, and used it to analyze normal tissue, adenoma, and early-stage cancer from the colorectum. Area correction was required for proper comparison of b_1 values. We found that area-corrected b_1 values were significantly higher in high-grade adenoma and early-stage cancer, compared to normal tissue. Although there remains a need for further development of the algorithm to discriminate low-grade adenoma and inflammatory regions, our present findings indicate that our homology-based method using Betti numbers can be useful for distinguishing high-grade adenoma and early-stage cancer from normal tissue, as well as advanced colorectal cancer.

1 | Introduction

Colorectal cancer is responsible for 903 859 deaths annually, and was the second highest cause of cancer-related mortality in 2022 [1]. A definitive cancer diagnosis must be made by a specifically certified pathologist, but there is a global shortage of such pathologists. In Japan, there are fewer than 3000 pathologists; therefore, many hospitals do not employ a full-time pathologist. The scarcity of pathologists results in an increased workload per pathologist, and can delay therapeutic interventions while waiting for a pathological diagnosis.

Although it would be desirable, it is not easy to increase the number of pathologists. When training to become a full-fledged pathologist, it takes quite a long time to accumulate experience and hone skills, even after obtaining professional licenses. To supplement the work of pathologists, the use of artificial intelligence (AI) for pathological diagnosis is attracting attention [2–6]. Currently, AI research is experiencing a third boom with the development of deep learning, also termed machine learning [7]. By analyzing large numbers of X-ray images and corresponding pathological information, a machine program can identify certain common patterns, and become able to

Abbreviations: AI, Artificial intelligence; ROI, Region of interest.

report a correct pathological diagnosis based on a new X-ray image [8]. Since it is relatively easy to find common patterns on X-ray images, AI using machine learning is being actively developed in this research field. However, with regards to histological diagnosis, it is more difficult to identify common patterns due to the wide variety of cancer tissue morphology.

Rather than pursuing conventional machine learning-based morphological evaluation, we have developed a system that automatically identifies cancerous regions by quantifying the loss of contact inhibition (disordered proliferation), which is an essential characteristic of cancer. In normal cells, the property called contact inhibition causes cells to stop proliferating when they make contact with each other [9]. Cancer cells lose this property, and rapidly accumulate [10]. We hypothesized that it might be possible to quantify the alteration of contact degree due to loss of contact inhibition, using the mathematical concept of homology theory.

Specifically, we found that two indicators called Betti numbers—b_0, representing the number of independent objects, and b_1, representing the number of "holes (empty space)" formed by surrounding objects—can be used together to determine the contact degree (Figure S1). Pathology images can be appropriately binarized (black and white conversion), and thereby regarded as mathematical figures, making Betti number calculation relatively straightforward. By counting the Betti number, we have successfully accomplished the diagnosis of advanced colorectal cancer and lung cancer, based on discrimination from normal epithelial cells [11–13].

It is widely known that most colorectal cancers develop through an adenoma-to-carcinoma sequence [14]. In our previous study, we demonstrated that our homology method was well-applicable to the diagnosis of advanced colorectal cancer, for which the entire cancerous field shares a roughly uniform histology [11]. In the present study, we further investigated and aimed to clarify whether homology diagnosis could also be useful for the discrimination of pre-cancerous lesions, i.e., adenomas and early-stage cancer.

2 | Materials and Methods

2.1 | Tissue Samples

Tissue samples of early-stage colorectal cancer (38 cases) were collected from patients who underwent surgery at the Osaka University Hospital from 1990 to 2004. Table 1 summarizes the clinicopathological characteristics of the samples. Signed informed consent was obtained from all patients, and this study was approved by the Ethics Committee of Osaka University Hospital (No. 15144). This study was performed in accordance with the Declaration of Helsinki.

The tissue samples were stained with hematoxylin and eosin (H&E), and scanned using a Nanozoomer2.0-HT C9600-13 (Hamamatsu Photonics, Shizuoka, Japan). The Nanozoomer2.0-HT function enables an additional 2× magnification, which is equivalent to using a 40× objective lens. Scanning was performed at 40× magnification. Due to the nature of WSI data, it is possible to extract images at 10× or 20× magnifications.

TABLE 1 | Clinicopathological characteristics of samples.

Characteristics	Total number 38
Age in years, mean (range)	62 (36–81)
Sex	
Male	29
Female	9
Lesion location	
Cecum	1
Ascending	3
Transverse	3
Descending	1
Sigmoid	13
Rectum	17
Tumor size (cm), mean (range)	2.5 (0.6–7)
Clinical stage	
I	36
IIIa	1
Histological diagnosis	
Well	20
Moderate	16
Muc	1

Note: Well, well differentiated adenocarcinoma; Modearate, moderately differentiated adenocarcinoma; Muc, mucinous carcinoma.
*Some data could not be obtained because old samples were used.

A whole-slide image (WSI) was created using two patterns of scanning: the entire slide, and the region of interest (ROI) only. These slides had already been pathologically diagnosed as early-stage cancer, but areas of normal tissue and areas diagnosable as adenoma were also present surrounding the cancerous tissue. In this study, from each whole slide, the pathologist selected ROIs containing areas of normal tissue, adenoma, or cancer (Figure S2). The images of the ROI were converted to bmp format using NDP.view2 (Hamamatsu Photonics) with a 10× objective, and then exported.

2.2 | Homology Computation in Histological Images

Homology is a mathematical method used to quantify the connectivity structure of geometrically defined objects. However, direct computation of homology from histological images is computationally challenging. Therefore, in this study, images were converted into a mathematically tractable form by transforming them into binary images using grayscale values.

Grayscale values were calculated from the red, green, and blue (RGB) components of each pixel using the following equation:

$$\text{Grayscale} = 0.2989 \times \text{R} + 0.5870 \times \text{G} + 0.1140 \times \text{B}$$

The resulting values range from 0 (black) to 255 (white). Pixels with grayscale values exceeding the threshold were assigned as white, whereas those below the threshold were assigned as black, thereby producing a binary image. The binarization

threshold was determined by analyzing the RGB distribution of the entire ROI image (Figure S3). The detailed algorithm used to calculate the threshold is proprietary and is therefore not disclosed in this study.

All computations were performed using bitmap images. Images in other formats were converted to bitmap format prior to processing.

A bitmap image represents an image as a grid of small squares (pixels). Pixel adjacency was defined using the 8-neighbor connectivity scheme, in which each pixel is considered adjacent to the eight surrounding pixels (horizontal, vertical, and diagonal directions). In this framework, contact was defined when two pixels shared an edge, whereas pixels sharing only a vertex were not considered to be in contact.

Based on this definition, the connectivity structure and enclosed regions in the binary image were evaluated, and the Betti numbers b_0 and b_1 were calculated.

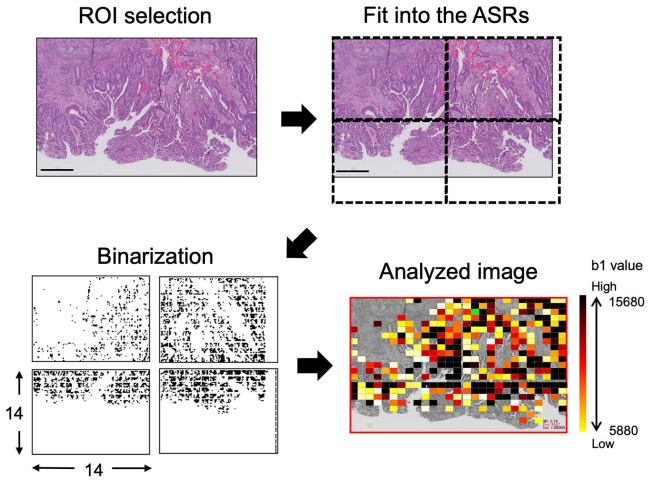

FIGURE 1 | Illustration of the analysis step by the AutoPatho application. Scale bar: 400 μm.

2.3 | Technical Remark for System

To ensure the reproducibility and portability of the AutoPatho system, the following factors should be considered:

Spatial Scaling: Differences between imaging devices could be addressed by normalizing the spatial resolution (i.e., pixel size per unit length).

Staining Standardization: Variations in staining conditions across institutions are managed by analyzing the RGB distribution (Figure S3), with dynamic adjustment of the binarization threshold to maintain consistent feature extraction.

Parameter Optimization: Since Betti numbers are sensitive to image resolution and magnification, these parameters must be optimized based on the histological characteristics of the target organ (e.g., colon, stomach, or prostate).

Device Compatibility: While the area standardization rectangle (ASR) size in this study was optimized for our whole-slide scanners, the method remains compatible with other imaging systems by re-defining the scale based on the sub-segment size.

2.4 | Statistical Analysis

Data were compared using Kruskal–Wallis tests, and the Steel–Dwass method (multiple comparison tests for non-parametric data). A p value of < 0.05 was considered statistically significant. All statistical analyses were performed using Statcel 4.

3 | Results

3.1 | Homology Analysis Using AutoPatho

In this study, we developed a homology analysis application named AutoPatho (Figure 1).

First, regions of interest (ROI) captured at an equivalent of 10× objective magnification were divided into frames of identical size, referred to as Area Standardization Rectangles (ASRs). The size of each ASR is 883.2×496.8 μm at 10× magnification, with a pixel resolution of 0.46 μm/pixel. ASR regions located outside the ROI were treated as blank, and their b_1 and b_0 values were set to 0. Since a 10% occupancy in a subsegment corresponds to a minute area of approximately 224 μm^2 (equivalent to only a few cells), this threshold was implemented to prevent the overestimation of values caused by factors such as staining precipitates or dust. While this approach may introduce a conservative bias in extremely sparse regions, the structural continuity of tumor tissue ensures that diagnostic features are captured in adjacent subsegments, thereby enhancing the overall robustness of the diagnostic threshold.

Next, each ASR was subdivided into 196 subsegments (14×14). The b_1 and b_0 values were calculated for each subsegment. However, if the amount of tissue within a subsegment was less than 10%, both values were set to 0.

In the AutoPatho system, the raw Betti numbers calculated for each subsegment were multiplied by a coefficient corresponding to the total number of subsegments ($14 \times 14 = 196$) to convert them into values representing the entire ASR scale. These scaled values were used as the final classification indicators. All numerical results presented in this paper are values after applying this scaling factor and therefore differ from the raw Betti numbers.

This scaling assumes a hypothetical situation in which a microstructural region is expanded to the size of the entire image. At the same time, it also functions as an engineering approach to improve the robustness of classification. When raw Betti numbers are small, classification thresholds become susceptible to noise and minor fluctuations, potentially reducing analytical stability. Therefore, the scaling factor was applied to expand the numerical range of Betti numbers and stabilize diagnostic threshold determination.

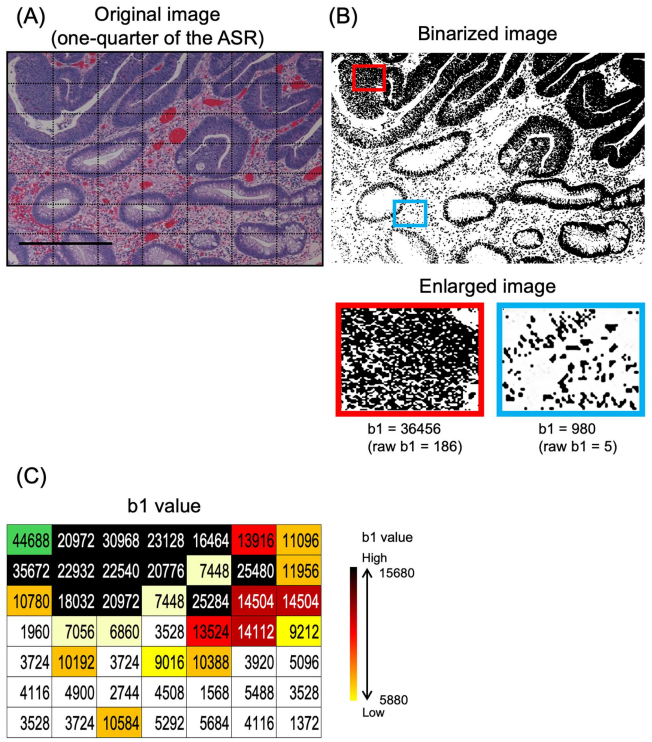

FIGURE 2 | Example of b_1 calculation. (A) One-quarter of the area standardization rectangle (ASR) image, with 49 sub-segments indicated. Scale bar: 200 μm. (B) Top: The corresponding image after binarization using the standardized method. Only the tumor region (red square) and normal region (blue square) were cropped for calculation of the b_1 values (raw $b_1 \times 14 \times 14$). Bottom-Left: An enlarged image of the binarized tumor region, showing an increased b_1 value (36 456, compared to a raw b_1 of 186), resulting from cancer-associated features, such as nuclear enlargement and blurred boundaries. Bottom-Right: An enlarged image of the binarized normal region, which exhibits a minimal number of holes due to the small and appropriately distributed nuclei ($b_1 = 980$; raw $b_1 = 5$). (C) The grid shows the b_1 values of each subsegment from (A). Colors correspond to the displayed color bar. Green indicates very high b_1 values (> 39 200).

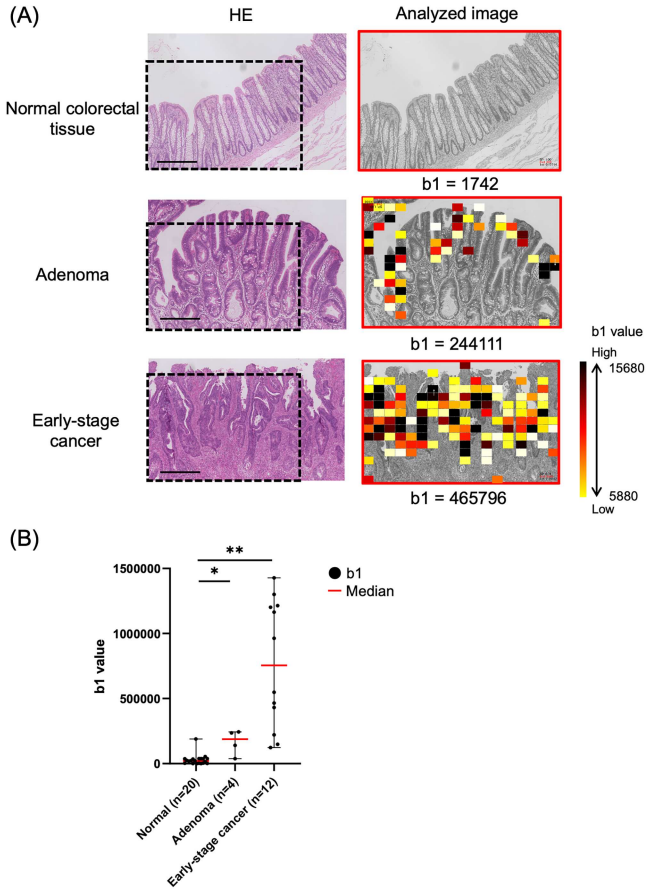

FIGURE 3 | (A) Representative images of normal colorectal tissue, adenoma, and early-stage cancer, with H&E staining and analyzed by AutoPatho. The b_1 values are presented below each image. Dashed lines indicate the sizes of the area standardization rectangles (ASRs). The sub-segment colors correspond to the color bar. Scale bar: 400 μm. (B) Comparison of b_1 values from first set analysis of normal colorectal tissue, adenoma, and early-stage cancer. *$p < 0.05$, **$p < 0.01$.

For visualization of the analysis results, each subsegment was color-coded according to its b_1 values. Subsegments with b_1 values below 5880 ($14 \times 14 \times 30$) were left uncolored, whereas higher values were assigned colors according to a color bar. Particularly high values exceeding 39 200 ($14 \times 14 \times 200$) were displayed in green. Representative calculation results and numerical examples are shown in Figure 2. In Figure 2, high-magnification (20×) images are used to facilitate visual comparison between pathological diagnosis and the numerical results. In this system, the b_1 values serve as a quantitative indicator of the "contact degree" within the tissue structure, building upon the methodological framework previously established [11].

3.2 | First Set Analysis

For the first set analysis, we used 19 cases, and selected ROIs from 20 normal colorectal tissues, 4 adenomas, and 12 early-stage cancers. Table S1 presents details regarding how ROIs were selected from each case, and Figure 3A and Figure S4 show representative analyzed images.

In normal colorectal tissues, most segments were not colored (b_1 values < 5880), with the exception of a few segments where epithelial cells accumulated. Adenoma and early-stage cancer tissues exhibited increased b_1 values, and the colored regions matched well with the tumor nests. Additionally, most of the interstitial fields were not colored, although b_1 values were increased by inflammatory cell infiltration in lymph follicles and lamina propria. For each sample type, the median b_1 value and range were as follows: normal tissue, median 21 583, range 0–189 480; adenoma, median 188 106, range 38 066–244 111; and early-stage cancer, median 755 653, range 122 902–1 427 607. We observed a step-wise increase of *b1* value from normal to adenoma to early-stage cancer, with significant differences between normal and adenoma ($p < 0.05$), and between normal and early-stage cancer ($p < 0.01$) (Figure 3B). However, we determined that this comparison was inaccurate, due to differences in the sizes of each ROI (area and aspect ratio) and the proportions of that area containing tissues. Thus, it was deemed necessary to perform proper correction of the b_1 values according to the tissue occupancy rate (area correction).

Area correction was performed as follows. For all ASRs, the tissue occupancy rate (%area) was calculated using imageJ software (https://imagej.net/ij/). Then the number of sub-segments for each ROI was multiplied by the tissue occupancy rate, and the product defined as tissue area. Finally, the b_1 value was divided by the tissue area, yielding the area-corrected b_1. Figure 4 presents examples of images showing large and small tissue areas. The overall Betti numbers of these ROIs exhibited completely different b_1 values: 1 201 820 versus 221 058. However, the area-corrected b_1 values were comparable: 2448 versus 2456, respectively.

When we analyzed the first set of samples with area correction, the area-corrected b_1 values were as follows: normal tissue, median 56, range 0–298; adenoma, median 850, range 79–1370; and early-stage cancer, median 1879, range 876–2623. We observed a step-wise increase in the values from normal to adenoma to early-stage cancer, and statistically significant differences between normal and adenoma ($p < 0.05$), and between normal and early-stage cancer ($p < 0.01$) (Figure 5A).

3.3 | Validation Analysis

Based on the results of the first set analysis, we decided to evaluate all ROI by calculating the area-corrected b_1 values. For the validation analysis, we further divided adenomas into low-grade and high-grade adenomas, and then selected the following ROIs from 19 cases: 24 normal tissue, 13 low-grade adenoma, 8 high-grade adenoma, and 29 early-stage cancer. Table S2 presents details regarding how these ROIs were selected from each case. The median b_1 values and ranges for each region were as follows: normal tissue, median 732, range 4–1981; low-grade adenoma, median 1252, range 41–3713; high-grade adenoma, median 2523, range 538–3757; and early-stage cancer, median 2590, range 968–5739. The validation analysis results indicated significant differences between normal and high-grade adenoma ($p < 0.01$), between normal and early-stage cancer ($p < 0.01$), and between low-grade adenoma and early-stage cancer ($^*p < 0.05$) (Figure 5B). To evaluate the discriminatory ability of this method for a pathologist's diagnosis, we calculated the sensitivity, specificity, and accuracy for high-grade adenoma versus normal tissue and for early-stage cancer

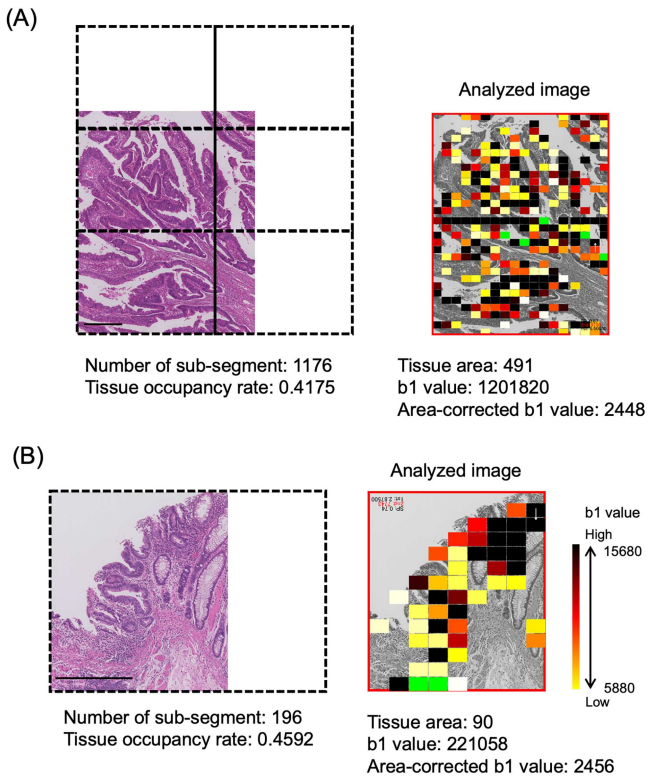

FIGURE 4 | Examples of area correction, with representative images of large tissue (A) and small tissue (B). Dashed lines indicates the sizes of the area standardization rectangles (ASRs). The sub-segment colors correspond to the color bar. Green indicates very high b_1 values (> 39 200). Scale bar: 400 μm.

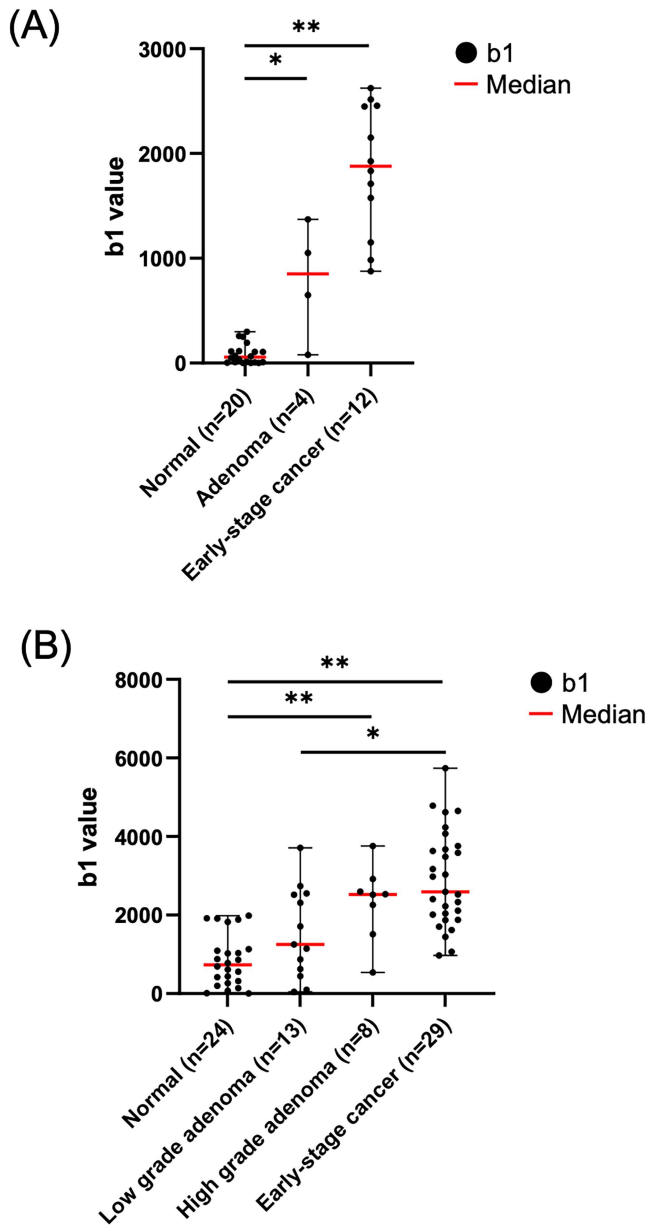

FIGURE 5 | (A) Comparison of area-corrected b_1 values from first set analysis of normal colorectal tissue, adenoma, and early-stage cancer. $^*p < 0.05$, $^{**}p < 0.01$. (B) Comparison of area-corrected b_1 values from validation analysis of normal colorectal tissue, adenoma, and early-stage cancer. $^*p < 0.05$, $^{**}p < 0.01$.

TABLE 2 | The discriminatory ability of homology method for pathologist's diagnosis.

	Cut off values ≧1400		
	Sensitivity (%)	**Specificity (%)**	**Accuracy (%)**
High grade adenoma	87.5	79.2	81.3
vs			
Normal tissue			
Early-stage cancer	93.1	79.2	86.8
vs			
Normal tissue			

versus normal tissue. The best results were achieved when the threshold value for b_1 was set as 1400 (Table 2).

4 | Discussion

Advances in molecular biology have revealed that multiple genetic alterations are involved in colorectal cancer development and progression [15–17]. This carcinogenic pathway is termed multistep carcinogenesis, or the adenoma–carcinoma sequence [14]. In this process, mutation or deletion of the *APC* gene results in a low-grade adenoma, and subsequent *KRAS* gene mutation induces development of a high-grade adenoma. The additional deletion of the *TP53* gene leads to carcinogenesis [14].

Endoscopic treatment is usually not performed for adenomas of < 5 mm, but is conducted when the tumor size reaches ≥ 6 mm. In contrast to adenomas, early-stage cancer requires the fastest possible diagnosis and therapeutic intervention. Thus, it is clinically important to differentiate between non-cancerous tissue, adenomas (especially high-grade adenomas), early-stage cancer, and advanced colorectal cancer. In the present study, we evaluated whether our homology-based diagnostic system could distinguish between normal colorectal tissue, adenoma, and early-stage cancer.

During the first set analysis, we realized that area correction was necessary, due to the bias caused by the differing sizes and tissue occupancy rates of each ROI. After area correction, our data still showed a gradual increase of b_1 values from normal tissue to adenoma to early-stage cancer. Moreover, our validation analysis revealed that the b_1 values of high-grade adenomas and early-stage cancers were significantly higher than for normal tissues. Compared to normal tissues, high-grade adenomas and cancer tissues are considered to have a higher degree of contact. In tumors, disordered cell proliferation and loss of contact inhibition cause the increased cell density and irregular overlapping of cells. An enlarged nucleus and high nuclear-cytoplasmic ratio are also the hallmark of tumor cells. Furthermore, dysregulated proliferation causes marked structural distortion of tissue components such as glandular structures, resulting in the loss of normal tissue function. In normal tissues, cells and tissue structures are arranged with a certain spatial order, and a relatively uniform state of contact is maintained throughout the tissue. Tumorigenesis disrupts this spatial order. These factors increase the opportunities for nuclear contacts and consequently become a major factor that increases b_1 values. When we classified the early-stage cancer cases in this study into well-differentiated adenocarcinoma (tub1) and moderately differentiated adenocarcinoma (tub2), and compared the b_1 values including adenomas, we observed a trend of gradually increasing b_1 values from adenomas to tub1 and tub2 (Figure S5). Although these results are still preliminary and require further detailed examination, this result suggests that the complexity of the glandular structure may contribute to the increase in b_1 values.

In the present method, even when direct physical contact between nuclei is absent, several factors can contribute to an increase in b_1 values. These include the apparent overlap of nuclei and indistinct nuclear boundaries caused by the two-dimensional projection of three-dimensional structures. Although we believe that nuclear contact is a significant factor in the increase in b_1 values, eosinophilic staining associated with reduced mucin in the cytoplasm and tumor necrotic background can also contribute to an increase in b_1 values (as a result of grayscale binarization, which captures eosinophilic staining in addition to hematoxylin-stained nuclei).

In this study, individual tissue components such as nucleus, cytoplasm, and background are not strictly separated and measured individually. Instead, each segment is binarized as a whole and processed as a single mathematical object. This operation corresponds to quantitatively calculating the average topological connectivity density, which integrates local variations within the segment. This approach enhances the robustness of the analysis and enables the complex phenomenon of tumor to be evaluated as statistically stable numerical values.

The results of present study suggest that this average topological connectivity density tends to show higher values in high-grade adenomas and early-stage cancers, while remaining low in normal tissue. This finding is consistent with established pathological knowledge and indicates that this index may serve as a highly effective indicator for distinguishing high-grade adenoma and early-stage cancer from normal tissue.

On the other hand, the b_1 value of low-grade adenoma was not significantly different from either normal tissue or high-grade adenoma. Overall, our results indicated that the homology method could discriminate high-grade adenoma and early-stage cancer from normal tissue, while it was difficult to diagnose a low-grade adenoma with this technique.

As shown in Tables S1 and S2, multiple ROIs were selected from the same cases in this study. Notably, ROIs from the same case are likely to show similar histology; therefore, further validation with more varied cases is required to strengthen the robustness of our present results. The size of ROIs was left arbitrary to ensure the regions selected by pathologist on each slide—classified as normal tissue, adenoma, or carcinoma—contained histologically as pure a component as possible and the impact of the analysis scale on the results was not examined. Although this study established an identical size ASR and performed area

correction, the dependence on ROI selection and the analysis scale remains an important topic for future investigation. Due to the applied segmentation strategy, some sub-segments may contain a mixture of both malignant and benign tissue, occasionally resulting in low b_1 values despite the presence of a lesion. However, the majority of tumor areas were accurately delineated, suggesting that these localized exceptions do not significantly compromise the overall diagnostic utility. We anticipate that in the future, advanced image analysis methods—for example, magnifying areas with similar chromatic properties or optimizing the segmentation approach—will yield results with greater precision. Furthermore, because the gastrointestinal mucosa has a strong structural orientation (vertical crypt axes *vs.* horizontal layering), the influence of directional bias should also be considered. Although the subsegments are sufficiently small to minimize the influence of directional bias, this point also requires further investigation.

Current efforts to use AI for pathological diagnosis are mainly focused on the creation of a shape-based algorithm for tissue image analysis using deep learning. Algorithms that focus on an object's shape and size can recognize slight deformations of shape, and identify them as a different object. However, since the tissue shapes of adenoma and cancer can vary even within the same case, a large amount of training data is required to create shape-based algorithms. On the other hand, the homology method does not require a large amount of training data, because it quantitatively evaluates the number of holes, which changes depending on the degree of contact.

Even if significant differences in b_1 values are observed, when the ranges of b_1 values overlap, it is difficult to apply this method in a clinical setting. Therefore, it is necessary to reduce the overlapping zone.

The challenge of our present homology method is that the methodology is exclusively based on the degree of contact. Thus, it can potentially generate false positives in cases of, for example, epithelial cell accumulation, inflammatory cell infiltration in lymph follicles and lamina propria, and artifacts. This suggests that the results may be impacted by the conditions of sectioning and the patient's inflammatory status. On the other hand, intracellular mucin may cause false negatives. Figure S6 presents examples of these limitations. Our method detected regions exhibiting accumulations of epithelial cells and inflammatory cells, and artifacts, while it failed to detect a poorly differentiated adenocarcinoma, which was characterized by abundant mucin accumulation. Thus, there remains an urgent need to develop new homology-based algorithms that can distinguish inflammation and artifacts from tumor tissues, and detect poorly differentiated tumor cells. Since normal tissue including the stroma tend to have low b_1 values, automatically excluding low-b_1 regions from the entire sample should significantly narrow down the areas likely to contain lesions. Furthermore, by combining this approach with AI technology specifically designed to detect poorly differentiated adenocarcinomas, we can build a system that ensures no cases are missed. In addition, we believe that accuracy can be further improved by effectively integrating the present methodology with AI technology, e.g., for size and shape recognition.

In conclusion, the present study showed that a homology method using Betti numbers could discriminate high-grade adenoma and early-stage cancer from normal tissue. Further research is needed to improve the discrimination of low-grade adenoma and to differentiate between inflammation and tumor tissue.

Author Contributions

Yuhki Yokoyama: conceptualization, writing – original draft, and writing – review and editing. **Kazuki Kanayama:** formal analysis, writing – original draft, and writing – review and editing. **Miwa Matsuda:** data curation and formal analysis. **Nobuaki Hatori:** data curation and formal analysis. **Taiki Asano:** data curation and formal analysis. **Eiichi Morii:** conceptualization and resources. **Hirofumi Yamamoto:** conceptualization, supervision, and writing – review and editing. **Kazuaki Nakane:** conceptualization, methodology, funding acquisition, and writing – review and editing.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP 26310209 and Ministry of Education, Culture, Sports, Science and Technology (MEXT) 'Next-Generation Cancer Professional Training Plan'.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. F. Bray, M. Laversanne, H. Sung, et al., "Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians* 74 (2024): 229–263.

2. S. Shafi and A. V. Parwani, "Artificial Intelligence in Diagnostic Pathology," *Diagnostic Pathology* 18 (2023): 109.

3. G. Campanella, M. G. Hanna, L. Geneslaw, et al., "Clinical-Grade Computational Pathology Using Weakly Supervised Deep Learning on Whole Slide Images," *Nature Medicine* 25 (2019): 1301–1309.

4. J. van der Laak, G. Litjens, and F. Ciompi, "Deep Learning in Histopathology: The Path to the Clinic," *Nature Medicine* 27 (2021): 775–784.

5. V. Baxi, R. Edwards, M. Montalto, and S. Saha, "Digital Pathology and Artificial Intelligence in Translational Medicine and Clinical Practice," *Modern Pathology* 35 (2022): 23–32.

6. P. Schüffler, K. Steiger, and W. Weichert, "How to Use AI in Pathology," *Genes, Chromosomes and Cancer* 62 (2023): 564–567.

7. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature* 521 (2015): 436–444.

8. G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis* 42 (2017): 60–88.

9. R. Mayor and C. Carmona-Fontaine, "Keeping in Touch With Contact Inhibition of Locomotion," *Trends in Cell Biology* 20 (2010): 319–328.

10. A. Seluanov, C. Hine, J. Azpurua, et al., "Hypersensitivity to Contact Inhibition Provides a Clue to Cancer Resistance of Naked

Mole-Rat," *Proceedings of the National Academy of Sciences* 106 (2009): 19352–19357.

11. K. Nakane, A. Takiyama, S. Mori, and N. Matsuura, "Homology-Based Method for Detecting Regions of Interest in Colonic Digital Images," *Diagnostic Pathology* 10 (2015): 36.

12. T. Qaiser, Y. W. Tsang, D. Taniyama, et al., "Fast and Accurate Tumor Segmentation of Histology Images Using Persistent Homology and Deep Convolutional Features," *Medical Image Analysis* 55 (2019): 1–14.

13. M. Nishio, K. Nakane, T. Kubo, et al., "Automated Prediction of Emphysema Visual Score Using Homology-Based Quantification of Low-Attenuation Lung Region," *PLoS One* 12 (2017): e0178217.

14. B. Vogelstein, E. R. Fearon, S. R. Hamilton, et al., "Genetic Alterations During Colorectal-Tumor Development," *New England Journal of Medicine* 319 (1988): 525–532.

15. E. R. Fearon and B. Vogelstein, "A Genetic Model for Colorectal Tumorigenesis," *Cell* 61 (1990): 759–767.

16. M. Miyaki, M. Seki, M. Okamoto, et al., "Genetic Changes and Histopathological Types in Colorectal Tumors From Patients With Familial Adenomatous Polyposis," *Cancer Research* 50 (1990): 7166–7173.

17. K. Tariq, K. Ghias, K. Tariq, and K. Ghias, "Colorectal Cancer Carcinogenesis: A Review of Mechanisms," *Cancer Biology & Medicine* 13 (2016): 120–135.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Figure S1: Schematic image of Betti numbers: b_0 denotes the number of independent figures, while b_1 denotes the number of areas enclosed by a figure. In the left figure, $b_0 = 3$, $b_1 = 0$; in the right figure, $b_0 = 1$, $b_1 = 1$.

Supporting Figure S2: Representative examples of ROI selection. From each whole slide, the pathologist selected ROIs containing areas of normal tissue, adenoma, or cancer. Scale bar: 2 mm.

Supporting Figure S3: Example of the red-green-blue (RGB) distribution of a given tissue image. This information enables standardization of the image's color properties. When following the present methodology, these standardized data were used to determine the binarization parameters.

Supporting Figure S4: Representative images of normal colorectal tissue (A), adenoma (B), and early-stage cancer (C), with H&E staining and analyzed by AutoPatho. The b_1 values are indicated for each image. Dashed lines indicate the sizes of area standardization rectangles (ASRs). The sub-segment colors correspond to the color bar. Green indicates very high b_1 values (> 39 200). Scale bar: 400 μm.

Supporting Figure S5: (A) Comparison of area-corrected b_1 values from first set analysis of adenoma, well differentiated carcoinoma and moderately differentiated carcoinoma. $^*p < 0.05$. (B) Comparison of area-corrected b_1 values from validation analysis of low grade adenoma, high grade adenoma, well differentiated carcoinoma and moderately differentiated carcoinoma. $^*p < 0.05$.

Supporting Figure S6: The limitations of our homology method. Regions containing accumulations of epithelial cells and inflammatory cells, and artifacts, showed high b_1 values (false positives). On the other hand, poorly differentiated adenocarcinomas in the central region were not detected (false negatives). The sub-segment colors correspond to the color bar. Green indicates very high b_1 values (> 39 200). Scale bar: 200 μm.

Supporting Table S1: pin70143-sup-0007-Yokoyama_Y_et_al_TableS1.xlsx.

Supporting Table S2: pin70143-sup-0008-Yokoyama_Y_et_al_TableS2.xlsx.