Research status and hotspots of tight junctions and colorectal cancer: A bibliometric and visualization analysis

A bibliometric of TJs and CRC
Abstract

BACKGROUND
Colorectal cancer (CRC) is the third most common cancer worldwide and the second leading cause of cancer-related death. Over the past two decades, numerous researchers have provided important evidence regarding the role of tight junction (TJ) proteins in the occurrence and progression of CRC. The causal relationship between the presence of specific TJ proteins and the development of CRC has also been confirmed. Despite the large number of publications in this field, a bibliometric study to review the current state of research and highlight the research trends and hotspots in this field has not yet been performed.

AIM
This study used bibliometrics to analyze research on TJs and CRC, summarize the field’s history and current status, and predict future research directions.

METHODS
We searched the SCI-E database for all literature on CRC and TJs from 2001-2023. We used bibliometrics to analyze the data of these papers, such as authors, countries, institutions, and references. Co-authorship, co-citation, and co-occurrence analyses were the main methods of analysis. CiteSpace and VOSviewer were used to visualize the results.

RESULTS
A total of 205 studies were ultimately identified. The number of publications on this topic has steadily increased since 2007. China and the United States have made the largest contributions to this field. Anticancer Research was the most prolific journal, publishing 8 articles, while the journal Oncogene had the highest average citation rate (68.33). Professor Dhawan P was the most prolific and cited author in this field. Co-occurrence analysis of keywords revealed that "tight junction protein expression,"
"colorectal cancer," "intestinal microbiota," and "inflammatory bowel disease" had the highest frequency of occurrence, revealing the research hotspots and trends in this field.

CONCLUSION
This bibliometric analysis evaluated the scope and trends of TJ proteins in CRC, providing valuable research perspectives and future directions for studying the connection between the two. It is recommended to focus on emerging hotspots such as the correlations among intestinal microbiota, inflammatory bowel disease, TJ protein expression, and CRC.

**Key Words:** Colorectal cancer; Tight junctions; Bibliometric analysis; Research trends; Hot spots

Li HM, Liu Y, Hao MD, Liang XQ, Yuan DJ, Huang WB, Li WJ, Ding L. Research status and hotspots of tight junctions and colorectal cancer: A bibliometric and visualization analysis. *World J Gastrointest Oncol* 2024; In press

**Core Tip:** This bibliometric analysis evaluated the scope and trends of TJ proteins in CRC, providing valuable research perspectives and future directions for studying the connection between the two. It is recommended to focus on emerging hotspots such as the correlations among intestinal microbiota, inflammatory bowel disease, TJ protein expression, and CRC.

**INTRODUCTION**
Colorectal cancer (CRC) ranks as the third most common cancer worldwide and the second leading cause of cancer-related mortality. In 2018, it was estimated that there were approximately 1.8 million new cases and 881,000 deaths globally[1]. With its continued rise in Western countries, it is projected that by 2030, the global incidence of CRC will increase to 2.2 million new cases and 1.1 million deaths[2]. In China, it is
estimated that there are more than 376,000 new cases and 191,000 deaths annually[3]. Despite regional variations and declining trends, the burden of CRC remains high due to population growth. Environmental factors such as obesity, Western dietary habits, smoking, and alcohol consumption are associated with CRC[4]. Additionally, genetic mutations and epigenetic changes are known to be involved in the occurrence and progression of CRC[5,6]. There is increasing evidence indicating a close association between tight junction (TJ) protein expression and colorectal cancer[7,8].

TJs are crucial intercellular connections essential for building epithelial barriers and maintaining epithelial polarity[9]. They regulate tissue homeostasis and integrity by controlling paracellular permeability and polarity. Beyond the static expression that promotes barrier function, TJs dynamically participate in regulating a range of cellular processes, including proliferation, migration, plasticity, and differentiation, all of which are central to cancer initiation and progression[10]. Numerous studies have confirmed that disrupting TJ integrity leads to cell invasion and proliferation and consequently, the development of CRC and even distant metastasis[11-13].

Current research has focused on various members of the claudin family, elucidating their roles in CRC development. Cldn-2 is upregulated in CRC patients and is associated with poor survival. Depletion of Cldn-2 significantly promotes the transcription of N-myc downstream-regulated gene 1 (NDRG1), leading to the termination of CRC growth and metastasis in vitro and in vivo[14]. Moreover, Cldn-2 has been shown to be associated with CRC stem-like renewal[15]. Abnormal expression and distribution of claudin-7, a key protein in epithelial cells, have been reported in various human malignancies, including lung, colon, ovarian, breast, gastric, esophageal, and prostate cancers, and are associated with cancer progression and metastasis[16]. Claudin-7 interacts with integrin β1 to inhibit CRC cell proliferation and migration[17]. These studies collectively demonstrate the intimate relationship between TJ proteins and the metastasis of CRC, suggesting their potential as novel targets for CRC therapy.

Bibliometric analysis is a statistical method based on public literature databases (such as Web of Science) used to analyze and visualize research trends[18]. Bibliometric
analysis has become one of the most widely used methods for assessing the credibility, quality, and impact of academic work[19,20]. In recent years, an increasing number of studies have focused on TJ proteins and CRC. However, systematic research on the relationship between TJ proteins and CRC through bibliometric analysis has yet to be conducted. Through bibliometric analysis, researchers can better understand the current status, trends, and hotspots in this field. This study aimed to reveal new trends in TJ protein and CRC research, such as articles, journals, and cooperation patterns among key authors and institutions, and explore the research hotspots and future directions in the field through bibliometric analysis.

MATERIALS AND METHODS

Search strategy
This study utilized the Web of Science as the data source. All potentially relevant publications were collected based on the topic (TS) using the following search strategy: #1: [TS = (colorectal* OR colon* OR rectum* OR rectal*)] #2: TS = (cancer* OR neoplasm* OR carcinoma* OR adenoma*) #3: TS = (claudins* OR CLDNs* OR claudin* OR CLDN*), Final dataset: #1 AND #2 AND #3]. To capture as many data sources as possible, wildcards () were used, which can substitute for any other character and allow keywords to have variable endings. For example, "cancer" would return both "cancer" and "cancers". English articles published from January 1, 2001, to December 31, 2023, were retrieved.

Study selection
The selection criteria and literature screening process for this study are illustrated in Figure 1. In brief, we inputted search terms for the initial search, and then two researchers reviewed the publications identified in the initial search and excluded those that did not meet the following inclusion criteria: (1) publications were limited to English language only; (2) document types included articles and reviews, excluding letters, comments, and conference abstracts; (3) publications were sourced from the
Science Citation Index Expanded (SCI-E) database; (4) the search timeframe was from January 1, 2001, to December 31, 2023, spanning 23 years; (5) included publications focused on patients with CRC (including patients with CRC, preoperative and postoperative patients with CRC), CRC animal models, and CRC cell models and evaluated the relevance of the study subjects to intestinal Tj proteins; and (6) to avoid bias due to daily updates of the database, searches and screening of publications were conducted and completed on the same day.

**Data extraction**

Two researchers independently reviewed all included publications, downloaded them, and exported them into different file formats for analysis. The following indicators were extracted: The number of publications, citation frequency, countries, institutions, journals, authors, keywords, and the journal's impact factor for the year 2023.

**Data analysis**

VOSviewer (version 1.6.20) was used to create, visualize, and explore collaboration networks of countries, journals, and authors, where each point represents a country/region, institution, or author, with the size of the point determined by the number of publications and the strength of connections between points determined by the number of collaborations.

CiteSpace (version 6.2.R4) is a visual tool for analyzing keyword trends over time.

**RESULTS**

**Global publication and collaboration trends**

From January 1, 2001, to December 31, 2023, a total of 677 records were identified. A total of 239 records were excluded because they were conference abstracts, letters, and ongoing papers. The remaining 438 records were further assessed through abstract or full-text reading. Ultimately, 205 studies that met the inclusion and exclusion criteria were included in the bibliometric analysis (Figure 1). These papers were published by
1337 authors from 355 institutions in 32 countries and appeared in 127 journals. These articles cited 7060 references from 1361 journals. The number of annual and cumulative publications has significantly increased over the past 23 years. The peak number of papers was published in 2023 (24 papers) (Figure 2A, B).

**Country/Region analysis**

A total of 32 countries/regions have published research on TJ proteins and CRC. Table 1 shows the five most productive countries. China had the highest number of publications, with 66 papers, followed by the United States (46 papers) and Japan (30 papers). The United States had the highest citation rate (3227 times), followed by Japan (1186 times) and China (994 times). Germany (74 times) and the United States (70.15 times) had the highest average citation rates.

The collaboration network among these countries is shown in Figure 3A. All countries contributed at least 3 publications in this field, with 10 countries represented in the network map. Figure 3B also illustrates the years of collaboration among countries, showing that China and the United States can be considered the central countries in the network and have had the closest mutual connections in recent years. Furthermore, the United States has connections with almost all countries in the network, with particularly close ties to Japan, South Korea, Canada, and Spain. Compared to the United States, China has fewer connections with other countries, with some connections to countries such as Japan and Germany.

**Authors and journals**

Additionally, we evaluated the most productive authors of the articles included in this study. Professor Dhawan P published the most papers, with a total of 17, followed by Singh AB (15 papers) and Ding Lei (9 papers). Similarly, Professor Dhawan P had the highest total number of citations (922), while Professor Sharma A had the highest average number of citations (83.6) (Table 2).
Next, researchers analyzed the journals in which the articles were published from 2001 to 2023. A total of 127 journals published articles in this field, with the 9 journals with the greatest publication frequency listed in Table 3. The top three journals were Anticancer Research (8, 3.90%), International Journal of Molecular Sciences (7, 3.41%), and Oncogene (6, 2.92%). Among the top 9 journals, Oncogene had the greatest number of citations, with 410 citations and an average citation rate of 68.33, making it an influential journal in this field.

Co-citation analysis

Co-citation analysis refers to the relationship between two (or more) papers that are simultaneously cited by one or more subsequent papers, forming a co-citation relationship[1]. This research method is used to measure the degree of relationship between references. Together, all articles cited a total of 7060 references, with 46 papers cited at least 15 times. Papers cited more than 15 times were imported into VOSviewer for co-citation analysis and visualization (Figure 4A). The focus was divided into two main clusters, represented in green and red. Table 4 Lists the top 6 research articles cited from 2001 to 2023 (data from the Web of Science database). The most cited article was "Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer," authored by Dhawan P, with 86 citations; this article focused on the expression and role of claudin-1 in colorectal carcinoma and metastasis. The next most cited articles were authored by Miwa N and Resnick MB, cited 48 and 35 times, respectively. Analysis revealed that these six articles focused mainly on the expression of claudin-1 in CRC and its role in the development of CRC. Additionally, the expression and function of claudin-1 and claudin-7 in breast cancer were analyzed.

Together, all papers cited references from 1361 journals. We visualized the journals cited more than 50 times using VOSviewer, obtaining a co-citation network of cited journals (Figure 4B). Table 5 Lists the 6 journals cited most frequently from 2001 to 2023 (data from the Web of Science database). The top three journals were Cancer Research (377 times), Oncogene (319 times), and Gastroenterology (285 times). The cited journals
are represented by clusters of three different colors. The green cluster primarily consisted of basic journals such as Cell Biology and Molecular Biology. These journals were cited to review current research findings and provide theoretical support for their studies. The blue and red clusters consisted of clinical-oriented journals in the field of gastrointestinal tumors.

**Keyword visualization**

The purpose of keyword co-occurrence analysis is to study the co-occurrence relationships between keywords in a set of publications, reflecting the popular topics. A total of 205 research articles that met the inclusion and exclusion criteria were exported from the Web of Science database in pure text format, and keyword analysis was performed using VOSviewer, with a threshold set to 10. Combining some frequently recurring keywords and synonyms, a total of 47 keywords were identified. Figure 5A and 5B show that "expression" and "colorectal cancer" were the most prominent keywords. All identified keywords can be divided into 3 clusters, represented in red, green, and blue. These clusters represent the most prominent themes in this research field to date.

Figure 5C displays the average year of the appearance of keywords. The use of keywords such as “migration”, “apoptosis”, and “inflammatory bowel disease” has increased in recent years. Recent research has focused on the mechanisms of migration and apoptosis related to TJ proteins and their roles in the development of inflammatory bowel disease.

Keyword clusters consist of one or more keywords that have specific relationships with each other. Cluster analysis of keywords was performed using CiteSpace software, and Figure 5D shows a Q value of 0.6723 ( > 0.3) and an S value of 0.8887 ( > 0.7), further indicating that the clustering results are robust and reasonable[21]. There were a total of 10 clustering patterns, 7 of which were selected for analysis. The cluster keywords included "microbiome" (Cluster 0), "progression" (Cluster 1), "cell adhesion molecules" (Cluster 2), "Clostridium perfringens enterotoxin" (Cluster 3), "prognosis"
(Cluster 4), "epithelial-mesenchymal transition" (Cluster 5), and "colorectal cancer" (Cluster 6). There are many connections between the nodes of these clusters, indicating a high degree of keyword co-occurrence in this field.

Keyword timeline visualization was used to display the classification and publication time of the keywords (Figure 6A). Most keywords appeared after 2005, with increasing focus on "activation," "invasion," "metastasis," and "gene expression" in recent years. Additionally, this figure visualizes the publication times of clusters and the relationships between different clusters. Through cluster analysis in this study, it was determined that the current hotspots in this field mainly focus on the intestinal microbiota, cell adhesion molecules, prognosis, and epithelial-mesenchymal transition.

If certain keywords are concentrated in a particular period, they are called burst keywords. Burst keywords can reflect different stages of development in a field. CiteSpace was used to detect burst keywords to identify research frontiers in this field, with blue bands indicating the start and end times of the appearance of the word and red bands indicating the start and end years of the burst. The five most common groundbreaking keywords extracted from CRC- and TJ-related papers via CiteSpace are shown in Figure 6B. The keyword with the highest burst intensity was "inflammatory bowel disease" (burst intensity of 4.86), which has been a research hotspot in recent years. This keyword first appeared in 2010 and showed explosive growth in 2021, indicating that an increasing number of studies in recent years have focused on exploring the role of TJ proteins in the process of inflammation-induced carcinogenesis.

**DISCUSSION**

Studies[22-25] have shown a close association between intestinal TJ proteins and CRC, leading to an increasing number of studies over the past two decades exploring the role of TJs in the occurrence and development of CRC. Bibliometrics was used to analyze authors, institutions, countries, and references in the SCIE literature database to understand research areas; trends were visualized using CiteSpace and VOSviewer. This research method provides a more comprehensive analysis of the literature and
produces more intuitive results than conventional systematic reviews. To date, there has been no bibliometric analysis specifically focused on the relationship between TJ proteins and CRC. Therefore, for the first time, this study used bibliometrics to explore the applications and developments in this field from 2001 to 2023 and speculated on future research trends.

In this bibliometric analysis, we found an increase in the number of annual publications over the past 23 years. This trend indicates growing interest among researchers worldwide in this field. Until 2006, the number of publications related to CRC and TJ proteins increased slowly, followed by a significant increase in related publications after 2007, particularly in recent years (Figure 2). This phenomenon suggests recent rapid development in this field.

We analyzed the most influential countries, authors, and journals in this field. The top three countries accounted for approximately 60% of all publications. This result reveals significant research disparities among countries worldwide, with leading countries having a decisive advantage over others. There were 66 articles contributed by Chinese scholars, accounting for approximately 30% of all publications and producing an average citation rate of 15.06. The United States ranked second in terms of the number of publications, with 46 papers, which were cited 3227 times with an average citation rate of 70.15. This finding indicates the central role of the United States in this field. In addition to China and the United States, the top five countries include Japan, South Korea, and Germany, indicating progress in these countries, likely as a result of significant government support for medical research, as most countries are developed nations. China is the only developing country with the highest publication output, which can be attributed to the rapid growth in financial investment by the Chinese government in medical research, especially in cancer-related studies; this investment even surpasses that of most other countries except the United States. Additionally, with the world's largest population, China provides ample patient resources for clinical research, which may contribute to the significant progress in medical research in China[26,27]. Notably, although China is the most productive in
research related to TJ proteins and CRC, its collaboration with other countries/regions in this field is limited; therefore, China needs to strengthen its collaboration with other countries/regions. Figure 3A shows the collaboration among different countries; the United States is positioned at the center of the collaboration network, collaborating with multiple countries, including extensive collaboration with China. We also found that collaboration has become increasingly more closer in recent years.

In terms of authorship analysis, Dhawan P from the United States had the highest number of publications, making him the most prolific scholar in this field. His research primarily focused on elucidating the mechanisms underlying the involvement of the TJ protein family member Cldn-1 in the development and progression of CRC. Among all the journals publishing related articles, Anticancer Research had the highest number of publications, followed by the International Journal of Molecular Sciences; the former is a journal focusing on oncology research, and the latter is a journal focused on molecular biology. Among the cited journals, Cancer Research ranked first in terms of citation frequency, with a maximum impact factor of 11.2, followed by Oncogene and Gastroenterology, indicating the high academic influence of these journals in this research field. Notably, in the analysis of the articles cited in this study, the article titled "Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer[28]" published in 2005 had the highest number of citations. This study demonstrated increased expression of claudin-1 in human primary CRC, metastases, and corresponding cell lines, with claudin-1 expression detected in the cell nucleus. Claudin-1 induces phenotypic changes in CRC cell lines, leading to alterations in the structure and function of epithelial-mesenchymal transition markers. Furthermore, additional data suggest that the regulation of E-cadherin expression and β-catenin/Tcf signaling may be potential mechanisms underlying claudin-1-dependent changes.

In terms of keyword frequency, "colorectal cancer" and "expression" were the most prominent keywords, indicating that the majority of studies focused on the correlation between TJ protein expression and CRC. Additionally, "inflammatory bowel disease" is currently a hot topic. Keyword cluster analysis, keyword temporal distribution, and
keyword burst analysis revealed that early-stage researchers are still in a relatively macroscopic and superficial stage of exploration regarding the association between CRC and TJ proteins. Researchers have attempted to determine the relationships between TJ protein expression and human colorectal adenomas and cancer. In the midterm, research directions gradually extended to the pathogenesis level, and to some extent, certain interactions between CRC and TJ proteins have been identified, with research primarily focusing on the TJ protein family members Cldn-1 and Cldn-7. In recent years, numerous studies have elucidated the relevance of TJ proteins to inflammatory bowel disease in the occurrence of CRC, further exploring the role of the intestinal microbiota in the association between TJ proteins and CRC occurrence.

This study has certain limitations. First, we included only articles published in English, excluding non-English publications from the bibliometric analysis, which may have interfered with the results. Second, the publications were retrieved only from the Web of Science database, which may have resulted in some publications not included in the database being overlooked. The Web of Science database is the most widely used database for bibliometric analysis and was designed for this type of analysis[29]. Additionally, the Web of Science database rigorously evaluates publications, ensuring high-quality literature[30,31]. Third, recent high-quality publications have only been published for a short time, which may result in lower citation counts. Therefore, citation frequency may not accurately reflect the quality of publications.

CONCLUSION
This bibliometric analysis evaluated the scope and trends of TJ proteins in CRC, providing valuable research perspectives and future directions for studying the connection between the two. The number of publications in the TJs and CRC field has grown steadily over the past two decades. It is recommended to focus on emerging hotspots such as the correlations among intestinal microbiota, inflammatory bowel disease, TJ protein expression, and CRC.
1. www.frontiersin.org
   Internet
   278 words — 7%

   Internet
   137 words — 4%

   Crossref
   37 words — 1%

4. Qinhui Tang, Limin Li, Yantong Li, Wenmeng Yin, Xiaolin Zhong. "Association between inflammatory bowel disease and primary sclerosing cholangitis: Insights from bibliometric analysis", Medicine, 2024
   Crossref
   35 words — 1%

   Crossref Posted Content
   23 words — 1%

   Internet
   23 words — 1%
Yanling Liu, Manman Tan, Chunjuan Fang, Xiaoyan Chen, Huan Liu, Ying Feng, Yujuan Zhang, Weiping Min. "A novel multifunctional gold nanorod-mediated and tumor-targeted gene silencing of GPC-3 synergizes photothermal therapy for liver cancer", Nanotechnology, 2021