Japanese Journal of Gastroenterology and Hepatology

Research Article ISSN: 2435-1210 | Volume 8

Advancement in Colorectal Cancer (CRC) Treatment in The Past Two Decades: A High-Level Text Mining Analysis of Publications

Eyal Klang[1,2,3], Ben Boursi [4,5], Benjamin S Glicksberg [6,7], Yiftach Barash [1,3], Shelly Soffer [8] and Adi Lahat[9*]

[1]Department of Diagnostic Imaging, Sheba Medical Center, Tel Hashomer, Israel, and Sackler Medical School, Tel Aviv University, Tel Aviv, Israel

[2]Sheba Talpiot Medical Leadership Program, Tel Hashomer, Israel, and Sackler Medical School, Tel Aviv University, Tel Aviv, Israel

[3]DeepVision Lab, Sheba Medical Center, Tel Hashomer, Israel

[4]Department of Oncology, Sheba Medical Center, Tel Hashomer, Israel, and Sackler Medical School, Tel Aviv University, Tel Aviv, Israel

[5]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA, USA.

[6] Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY USA

[7]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY USA

[8]Internal Medicine B, Assuta Medical Center, Ashdod, Israel, and Ben-Gurion University of the Negev, Be'er Sheva, Israel

[9]Department of Gastroenterology, Sheba Medical Center, Tel Hashomer, Israel, and Sackler Medical School, Tel Aviv University, Tel Aviv, Israel

***Corresponding author:**

Adi lahat,
Chaim sheba Medical Center, Ramat Gan, Israel. 52621, Tel:+972526667501,
E-mail: zokadi@gmail.com

Received: 21 Mar 2022
Accepted: 08 Apr 2022
Published: 14 Apr 2022
J Short Name: JJGH

Citation:

Adi lahat, Advancement in Colorectal Cancer (CRC) Treatment in The Past Two Decades: A High-Level Text Mining Analysis of Publications.
J Gstro Hepato.. V8(12): 1-10

Keywords:

Colorectal Cancer; PubMed; Treatment; Surgery, Colorectal; Chemotherapy

List of Abbreviations:

Colorectal Cancer (CRC); Complementary Medicine (CAM); Microsatellite Instability (MSI); Mismatch Repair (MMR); Epidermal Growth Factor Receptor (EFGR); Human Epidermal Growth Factor Receptor 2 (HER-2)

1. Abstract

1.1. Background

During the last two decades, there is a constant search for new therapeutic options for colorectal cancer (CRC). We evaluated PubMed literature on CRC different treatment options during the past two decades. Special attention was attributed to specific treatment options according to biological markers.

1.2. Methods

We have queried PubMed for all available CRC-related entries published during 2000-2020. For each entry we retrieved the title, abstract, and keywords. A gastrointestinal specialist and a CRC oncology specialist decided in consensus on a list of terms to classify entries. The terms belonged to seven treatment groups: chemotherapy, biologic, surgery, immunotherapy, radiation, dietary or complementary medicine (CAM) and microbiota modulation. Sub-analyses were performed for disease stages and genetic mutations. Annual trends of publications during 2000-2020 were plotted for different treatment types and sub-analyses.

1.3. Results

Overall, 162,196 CRC related entries were published between 2000 and 2020. Surgical treatment showed the highest number of publications (18.5%), followed by chemotherapy (14.3%), radiation (5.8%), diet/CAM (5.0%), biologic (3.8%), microbiome research (1.4%), and immunotherapy (0.1%). Sub-analysis by mutation types showed that for KRAS, NRAS, BRAF, EGFR and HER2 main research topics were first biologic and second chemotherapy treatments. For MSI, MMR, main research topics were surgery, chemotherapy, and immunotherapy treatments.

1.4. Conclusions

We observed publication trends in CRC treatment over the past two decades. These decades, and mainly the last one, certainly deserve the reference as the "personalized medicine era". During the last ten years, we see a clear and steep elevation in specific patient-customized treatment publications.

2. Introduction

Colorectal cancer (CRC) is one of the leading causes of cancer-related deaths worldwide. This disease accounted for almost 10% of all cancer cases and 8.5% of total cancer deaths. The incidence of CRC is expected to rise by up to 60% by 2030 and reach 2.2 million new cases and 1.1 million deaths per year [1,2]. In the United States, CRC is the second leading cause of cancer related death in men and women and accounts for yearly health care costs of approximately 14 billion USD [3]. During the last two decades, there is a continual improvement in patients' prognosis [2]. At least part of this improvement results from new treatment strategies and therapeutic options, many of which target specific signaling pathways and biological markers [4]. Available data in literature diverse extensively, and includes clinical studies exploring various treatment options targeted explicitly at disease biological markers. Despite the important advancement in therapeutic options, almost 50% of patients suffering from CRC experience tumor recurrence [5]. These high recurrence rates affect survival and pose an important therapeutic goal of maintaining long-term disease remission after achieving complete remission [5]. Thus, treatment strategies aim to achieve clinical remission and prevent recurrence in patients with localized disease [6]. Naturally, in patients with disseminated disease, initial goals differ according to disease severity [6]. Text mining is a computational method that enables a broad-scale data extraction [7]. This method excerpt information from texts using computational statistical modes [7]. Text mining can be employed to characterize trends and explore dynamics in research fields [8-11]. We applied text-mining to evaluate published literature on CRC different treatment options during the past two decades. Special attention was attributed to specific treatment options according to biological markers.

3. Methods

3.1. Dataset

The U.S. National Center for Biotechnology Information (NCBI) provides public application programming interfaces (APIs) that allow programmatic access to the PubMed database. We have used the publicly available PyMed Python package to query the PubMed API. The following data were extracted for each entry: PubMed unique article ID (PMID), title, publishing journal, abstract text, keywords (if any), and authors' affiliations. Data were collected up until May 12, 2021.

3.2. Inclusion Criteria

The entire MEDLINE/PubMed database was used as the source for this article. We retrieved all available colorectal cancer related entries. The search was conducted in entries' titles, abstracts, and keywords using the terms "CRC" OR combinations using the terms "colorectal", "colon", "rectal" and the terms "cancer", "malignancy", "carcinoma", "adenocarcinoma". We have limited the entries to publications between January 1, 2000, to December 31, 2020.

3.3. Data Processing

The data processing and result visualization were written in Python (Ver. 3.6.5, 64 bits). A gastrointestinal specialist (AL) and an oncology specialist (BB) decided on consensus list of terms to classify entries (Table 1). The terms belonged to seven treatment groups: Chemotherapy, Biologic, Surgery, Immunotherapy, Radiation, Dietary or Complementary medicine (CAM) and microbiota modulation. Each entry was categorized by querying the title, abstract, and keywords for terms belonging to the treatment groups. Entries could belong to more than one group. Sub-analyses were performed for different conditions. This was done by querying the entries for disease stages (Table 2) and genetic mutations (Table 3).

Table 1: List of terms used to classify entries into seven treatment groups: Chemotherapy, Biologic, Surgery, Immunotherapy, Radiation, Microbiome altering, and Dietary or Complementary medicine (CAM).

Chemotherapy	5FU, 5-FU, Fluorouracil, Capecitabine, FOLFOX, OxMdG, mFOLFOX, Oxaliplatin, Irinotecan, Leucovorin, IROX, Roswell Park, AVEX, Trifluridine, Tipiracil, CAPOX, xeloda, adjuvant, neoadjuvant
Biologic	Bevacizumab, Cetuximab, Panitumumab, Encorafenib, Regorafenib, Larotrectinib, Entrectinib, Ziv-aflibercept, Ramucirumab, Trastuzumab, Pertuzumab, Lapatinib, Deruxtecan
Surgery	Surgery, Surgical treatment, Surgical therapy, Colectomy, Sigmoidectomy, Abdominoperineal resection, Low anterior resection, Total neoadjuvant therapy, Total mesorectal excision, Cytoreductive Surgery, Hyperthermic intraperitoneal chemotherapy, Oligometastatic, Hepatectomy, Metastasectomy, APR, LAR, TNT, TME, CRS, HIPEC
Immunotherapy	Pembrolizumab, Nivolumab, Ipilimumab
Radiation	Radiotherapy, Chemoradiation, Chemo-radiation, Short-course, Long-course, Radio-frequency ablation, Stereotactic body radiation, Chemoembolization, SBRT
Microbiome	Microbiome", Microbiota, Fusobacterium, F.nucleatum, Fusobacterium nucleatum, Genotoxic e.coli, bacteroides fragilis, Probiotic, Prebiotic
Diet	Diet, Nutrition
CAM	Cannabis, Marijuana, Rick simpson, Tetrahydrocannabinol, Cannabinol, Complementary medicine, Alternative medicine, Curcumin, Turmeric, Herbal, Herbs

Table 2: List of terms used to classify entries into four disease stages.

stage 1, stage i
stage 2, stage ii
stage 3, stage iii
stage 4, stage iv

Table 3: List of terms used to classify entries into seven biologic markers.

MSI, microsatellite instability
MMR, mismatch repair
EGFR, epidermal growth factor receptor
HER2, HER-2, HER 2, human epidermal growth factor receptor 2
KRAS, K-Ras, k-ras
NRAS, neuroblastoma ras viral oncogene homolog
BRAF, B-RAF, b-raf

3.4. Statistical Analysis

All analyses were conducted with Python (Python software foundation, Version 3.6.5). Statistical significance was established at a 2-sided $P < .05$. Descriptive statistics were reported using counts with percentages for categorical variables. Annual trends of publications for 2000-2020 were plotted for different treatment types and sub-analyses. The slopes of publication trends were calculated by fitting linear regression lines to the annual number of publications in the years 2000-2020 (with X being calendar year and Y being annual publications count). P-values and standard errors (SE) were calculated for the linear regression lines.

4. Results

Out of 31,850,051 PubMed entries available, 201,952 (0.6%) were CRC related (Figure 1). 162,196 / 201,952 (80.3%) of the entries were published between 2000 and 2020.

4.1. Publication Trends

Trends of CRC publications by treatment types are presented in Figure 2. Surgical treatment showed the highest number of total publications 30,033 / 162,196 (18.5%), followed by chemotherapy 23,202 / 162,196 (14.3%), radiation 9348 / 162,196 (5.8%), diet/CAM 8,074 / 162,196 (5.0%), biologic 6,149 / 162,196 (3.8%), microbiome research 2267 / 162,196 (1.4%), and immunotherapy 190 / 162,196 (0.1%). The slope of the trend of surgical treatment was also the steepest (112.3±5.8 publications/year, $p<0.001$), followed by chemotherapy (66.82.9 publications/year, $p<0.001$), radiation (34.7±1.7 publications/year, $p<0.001$), biologic (30.7±2.3 publications/year, $p<0.001$), diet/CAM (23.7±1.2 publications/year, $p<0.001$), microbiome (18.9±2.7 publications/year, $p<0.001$), and immunotherapy (4.4±1.1 publications/year, $p=0.002$). Sub-analysis by disease stages (Figures 3 A-D) showed that for stage I, treatment types included mainly surgical treatment 492 / 1148 (42.9%), followed by chemotherapy 21.7% and radiation 10.4%.

For both stage II and stage III, treatment types mainly included chemotherapy (stage II 1104/2916, 55.7%, stage III 1569/2819, 53.3%), followed by surgery (stage II 37.4%, stage III 37.9%) and radiation (stage II 16.2%, stage III 11.1%). For stage IV, main treatment types included surgery 814 / 1896 (42.9%), chemotherapy 24.9%, radiation 8.3% and biologic treatment 8.0%. Sub-analysis by mutation types (Figures 4 A-G) showed that for KRAS, NRAS, BRAF, EGFR and HER2 main research topics were first biologic and second chemotherapy treatments. For MSI, MMR, main research topics were surgery, chemotherapy, and immunotherapy treatments. Table 4 presents the total number of publications related to immunotherapy and biologic treatment, stratified by mutation type.

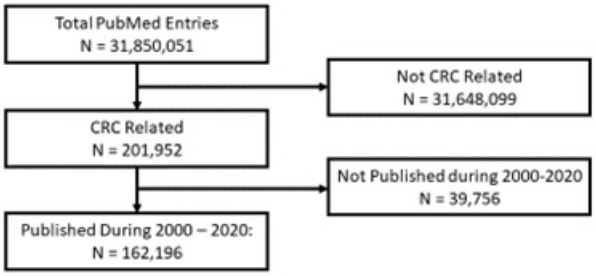


Figure 1: Study inclusion chart.

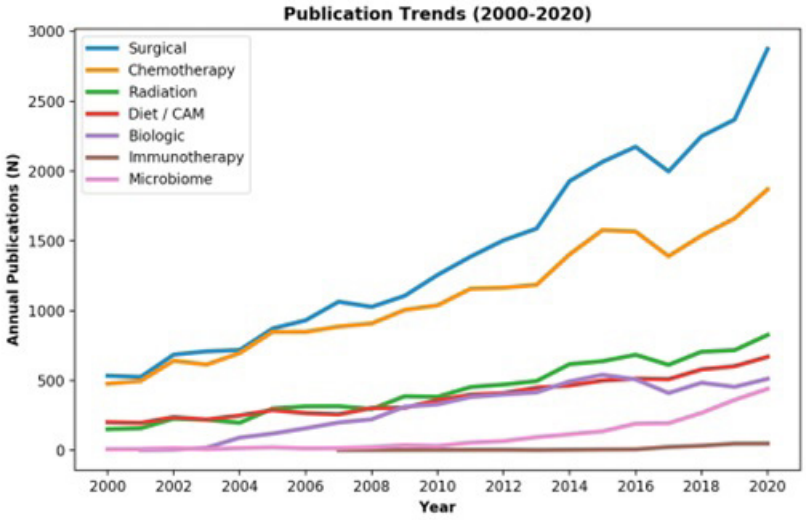
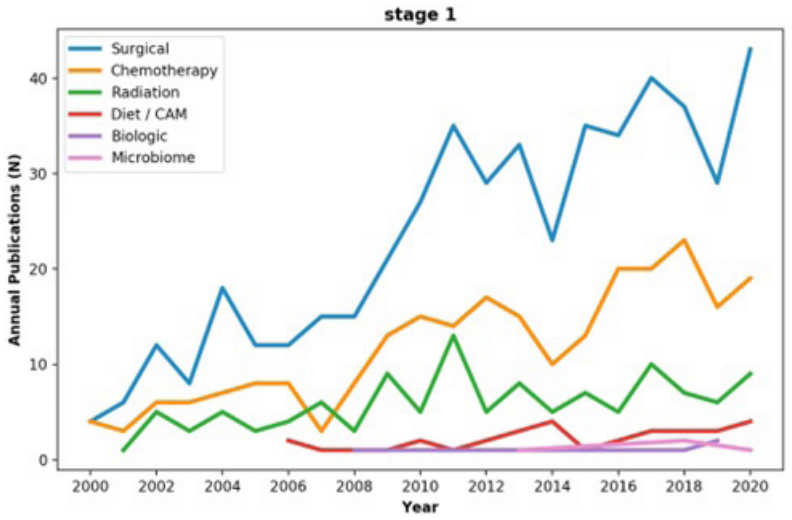
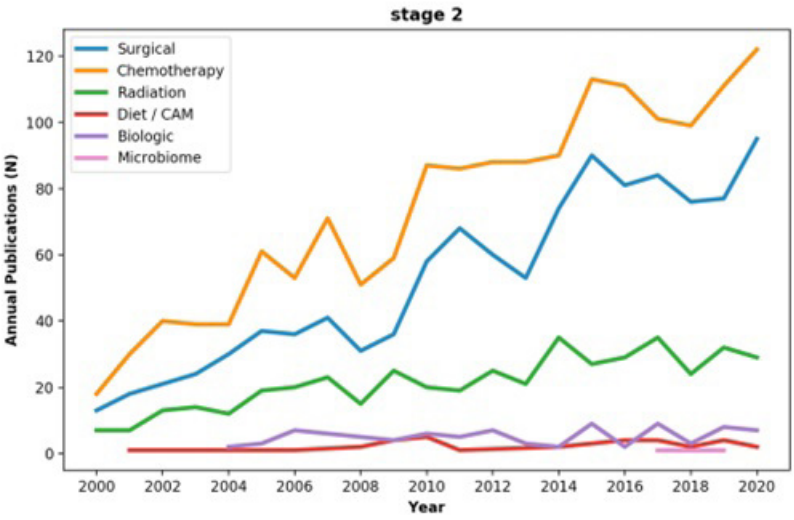
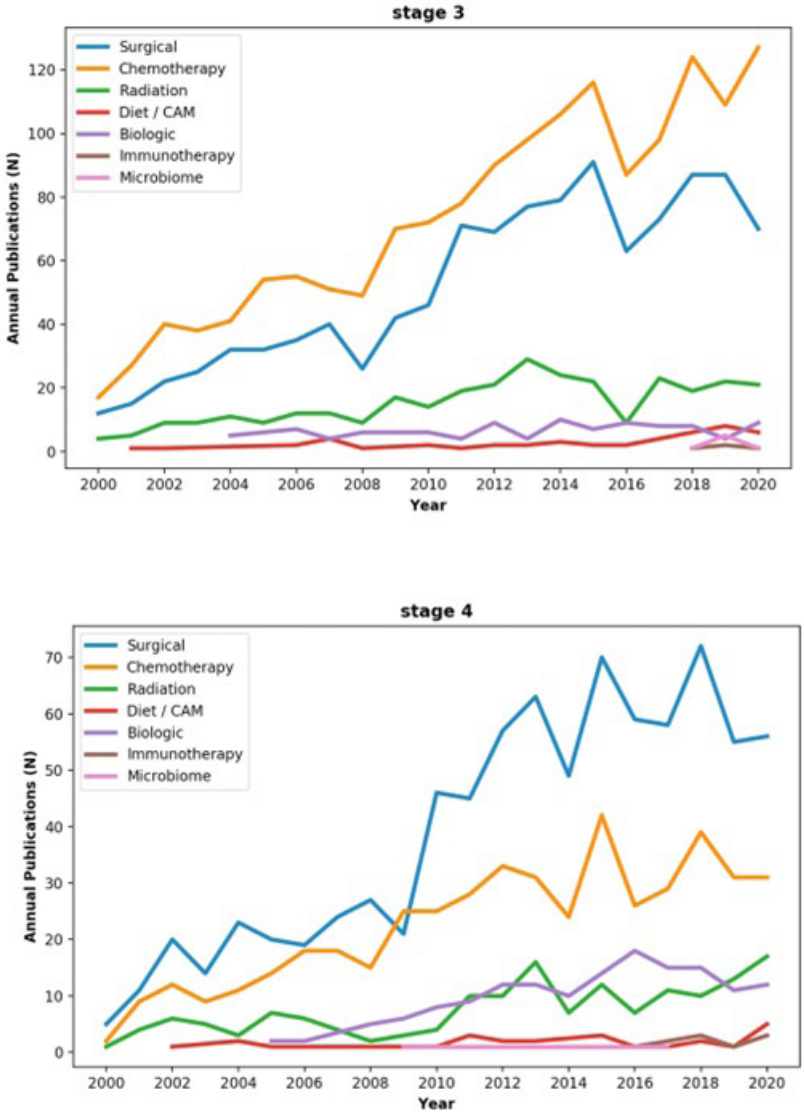


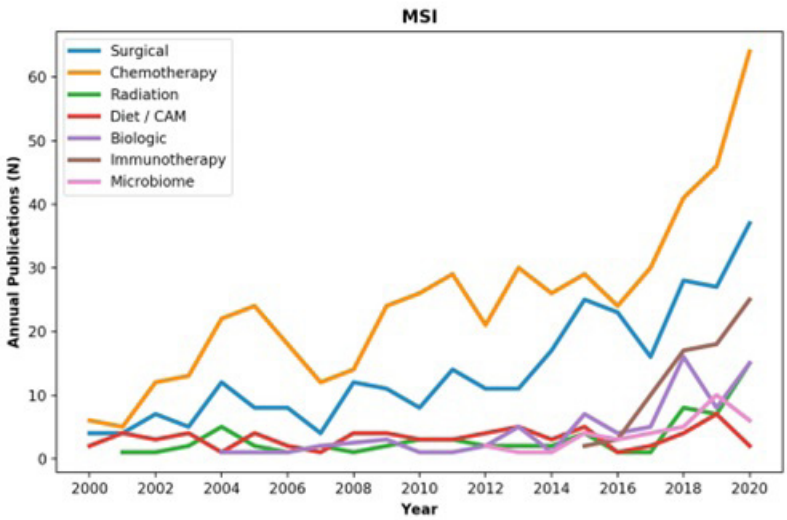
Figure 2: Trends of colorectal cancer (CRC) publications during 2000-2020, grouped by treatment type: Chemotherapy, Biologic, Surgery, Immunotherapy, Radiation, Microbiome altering, and Dietary or Complementary medicine (CAM).

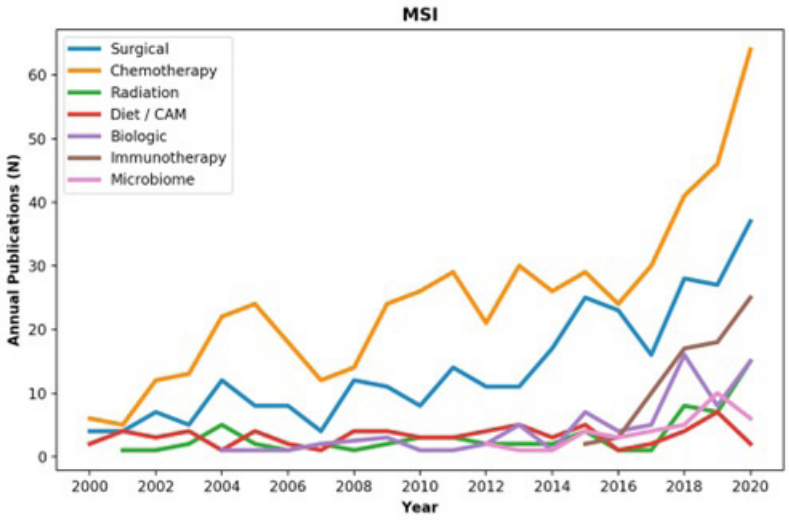


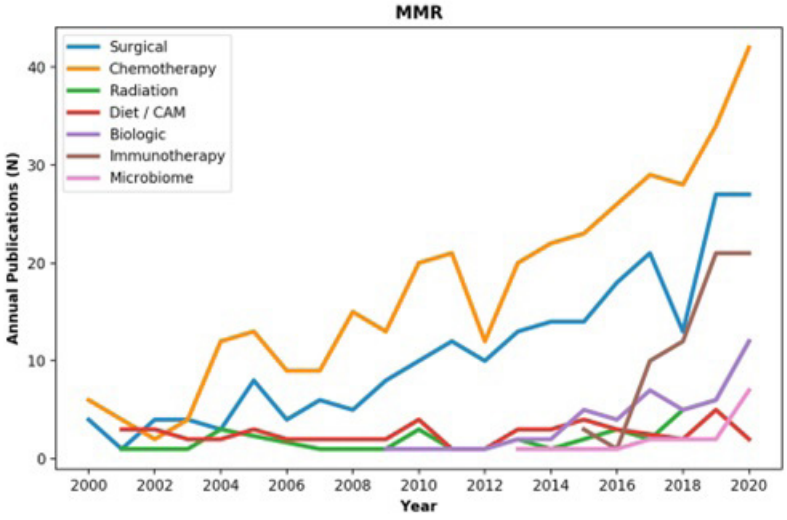


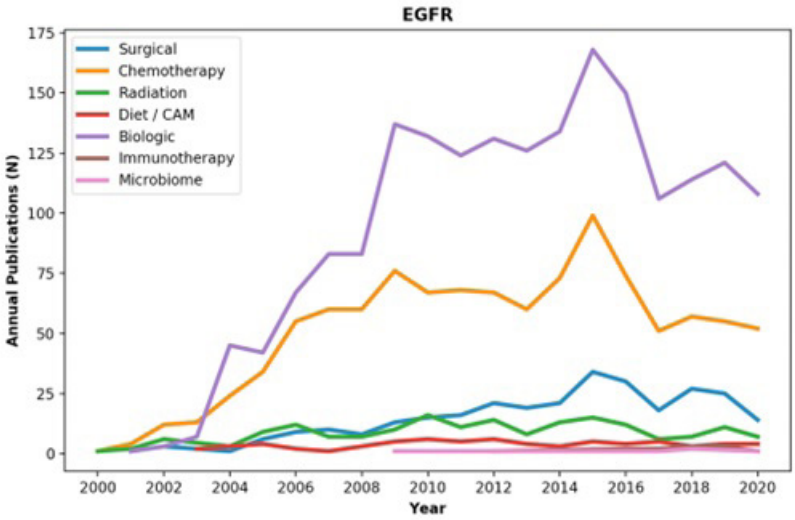


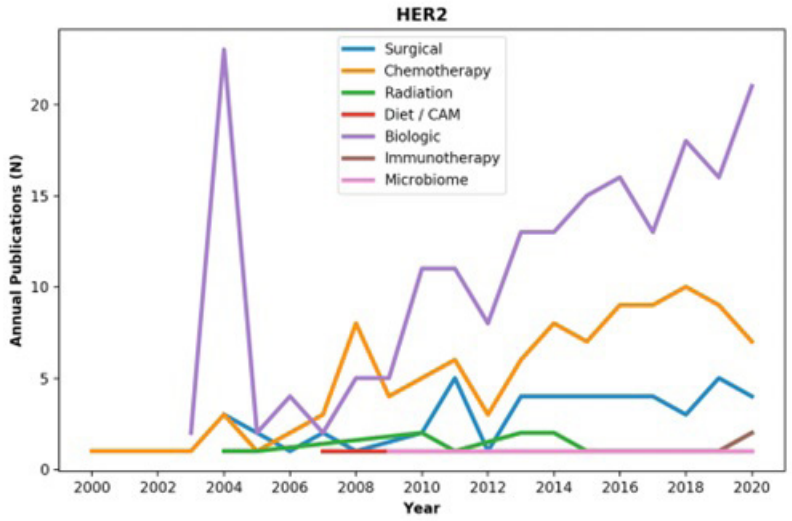
Figures 3 A-D: Trends of colorectal cancer (CRC) publications during 2000-2020, grouped by treatment types, stratified by disease stage: I – IV.

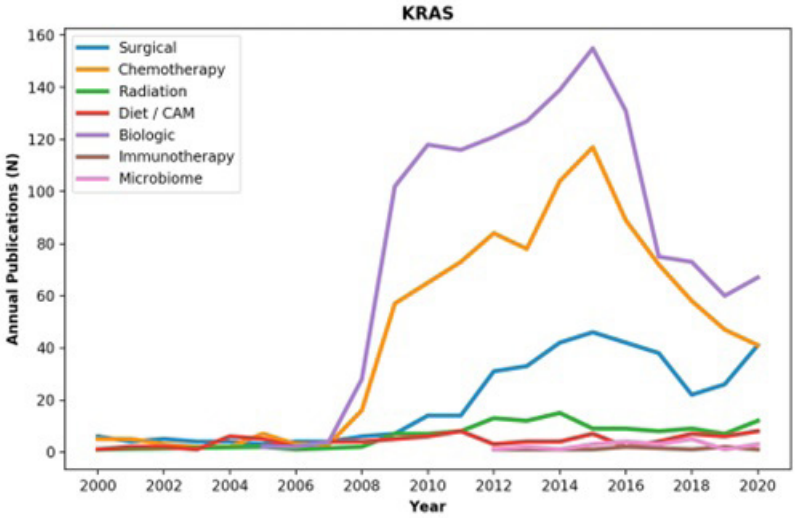


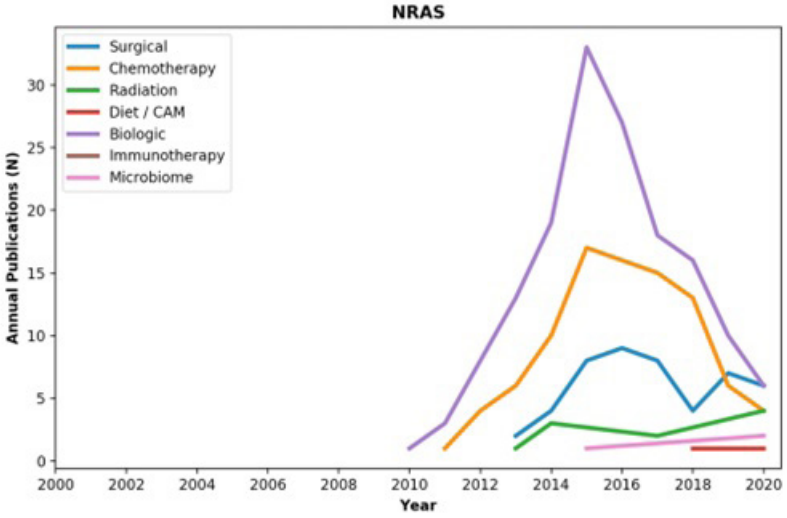

MSI
Surgical
Chemotherapy
Radiation
Diet / CAM
Biologic
Immunotherapy
Microbiome
Annual Publications (N)
Year

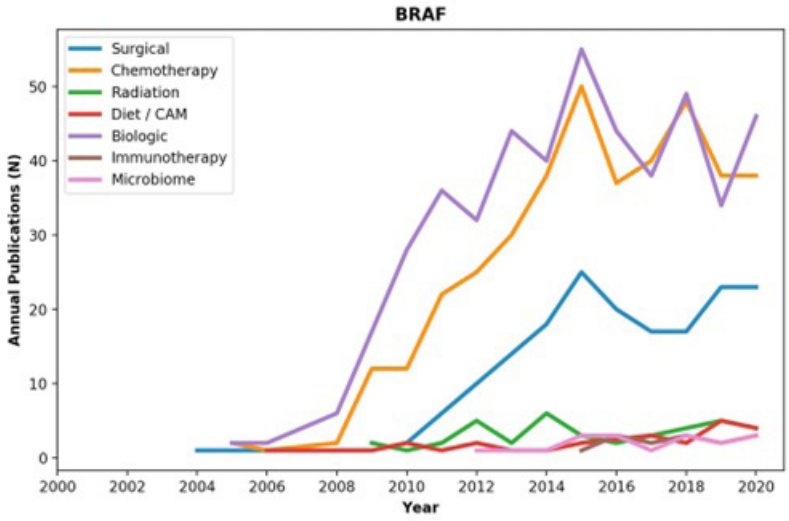

MMR
Surgical
Chemotherapy
Radiation
Diet / CAM
Biologic
Immunotherapy
Microbiome
Annual Publications (N)
Year


EGFR
Surgical
Chemotherapy
Radiation
Diet / CAM
Biologic
Immunotherapy
Microbiome
Annual Publications (N)
Year


HER2
Surgical
Chemotherapy
Radiation
Diet / CAM
Biologic
Immunotherapy
Microbiome
Annual Publications (N)
Year


KRAS
Surgical
Chemotherapy
Radiation
Diet / CAM
Biologic
Immunotherapy
Microbiome
Annual Publications (N)
Year


NRAS
Surgical
Chemotherapy
Radiation
Diet / CAM
Biologic
Immunotherapy
Microbiome
Annual Publications (N)
Year



Figures 4 A-G: Trends of colorectal cancer (CRC) publications during 2000-2020, grouped by treatment types, stratified by biologic markers: A - microsatellite instability (MSI); B – mismatch repair (MMR); C – epidermal growth factor receptor (EFGR); D – human epidermal growth factor receptor 2 (HER-2); E - KRAS; F - NRAS; G – BRAF.

Table 4: Overall number of publications related to biologic and immunotherapy treatment in different mutations.

Biologic marker	Immunotherapy publications (N)	Biologic treatment publications (N)
MSI	100	90
MMR	85	56
EGFR	13	2051
HER-2	7	210
KRAS	10	1385
NRAS	1	165
BRAF	15	529

5. Discussion

While the 5 – years survival rates are approaching 90% for patients with an early CRC stage, they decrease to less than 10% in patients with disseminated disease and distant metastases [12]. This dismal prognosis emphasizes the need for better patient stratification and identifying biomarkers to improve and navigate therapeutic decisions.

Thus, in recent years, as understanding the mechanism behind carcinogenesis and the mutual influence of gene mutations and epigenetic modifications on cancer development and progression widened, individual personalized treatment based on the presence or absence of specific genetic biomarkers became common [13-18].

Simultaneously, major progress was achieved in developing particular biologic therapies aimed at specific gene mutations [13-18]. Thus, while for early disease stage surgery remained the first and major therapeutic option, advanced stages are now treated with specific mutations-based treatment [6]. Furthermore, with the growing emphasis on personalized medicine, complementary and lifestyle modifications in addition to conventional therapy gathered popularity in CRC treatment [19,20]. In our current study, we applied a text mining approach to observe and analyze CRC treatment publications in the past two decades, aiming to achieve some comprehensions regarding treatment trends and development over the years. Hence, 80% of the literature published on CRC treatment was issued in the last two decades, increasing each year. Surgical treatment was the most studied treatment, comprising 18.5% of all treatment publications. Surgical therapy also showed the steepest slope (112.3 publications/year, $p<0.001$), which indicates the most rapid growth in publications rate. There is a clear increase in biologic therapy publications in stage IV, with the highest increase since 2010. This, too, represents current guidelines for treating disseminated disease [13-15] and the progress achieved in the last decade. An interesting trend that clearly emerges from our data is the steady rise along the years in the study of complementary medicine and diet modifications for CRC treatment. This reflects the holistic approach towards the treatment of malignant diseases. As patients' well-being during therapy and the understanding of the importance of full commitment and involvement of the patient in therapeutic decisions, CAM and nutrition became more acceptable as supplementary treatment and now are an integral part of the management in many oncologic institutions. According to current literature, up to 51% of cancer patients use CAM during their treatment [21, 22]. Younger age, high education, high income,

and female gender are independent predictors of CAM treatment [22]. Most patients use CAM to improve their general health and response to treatment and to treat therapy's adverse effects [22]. One of the most studied therapeutic option in this aspect is medical marijuana. Cannabis therapy is now among the first palliative treatments offered to patients in many cases [23-27]. Further sub-analysis was performed according to mutation type. As shown in Figures 4 A-G, biologic therapy was the most studied subject for mutations EGFR, HER-2, KRAS, NRAS, and BRAF. In most mutation types, there is an impressive and significant increase in publications during the last decade. This reflects the clear direction of individualized medicine, which became more prevalent and a common practice during the last ten years, along with the approval of various specific treatments. Thus, in 2008- the significance of KRAS to anti EGFR treatment was established and approved [28], and in 2020 the treatment of BRAF inhibitors combined with anti EGFR became available [29]. For the biological markers MSI- H and MMR, the most studied treatment options during the last two decades were first chemotherapy followed by surgical treatment. However, since 2016, there is a steep increase in publications assessing immunotherapy- along with the approval of immunotherapy to MMR in 2017 [30]. This, again, emphasizes the progression in the understanding of tumor biology and focusing on personalized medicine. Our study has limitations. This analysis only provides a high-level look at the field. The sheer number of publications prohibits a manual analysis of the records. A list of terms was determined based on current data in the literature and consensus between a senior oncologist specialist in CRC and a senior gastroenterologist. However, different terms might have achieved different results. The data was extracted from MEDLINE/Pubmed. Other options as google scholar were not included and might have reached different results. In conclusion, in our current study, we observed publication trends in CRC treatment over the past two decades. According to our findings, these decades, and mainly the last one, certainly deserve the reference as the "personalized medicine era". During the last ten years, we see a clear and steep elevation in specific patient-customized treatment publications. As more data regarding cancer pathogenesis accumulated, the understanding of specific mechanisms for carcinogenesis gradually intensified, and the ability to treat specific mutations and reach better therapeutic results increased accordingly. Furthermore, the steady increase in CAM and nutritional therapy also falls under personalized medicine as we perceive its' wider meaning. This therapeutic method values patients' preferences and regard at the patient in holistic eye; Thus, enabling tailor-made specific combined therapeutic approach to each patient. We predict this trend will further increase in the upcoming years.

References

1. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017; 66(4): 683-691.
2. Torre LA, Bray F, Siegel RL, Ferlay J. Lortet-Tieulent J and Jemal A. Global cancer statistics. *CA Cancer J Clin*. 2012; 65: 87–108.
3. Simon K. Colorectal cancer development and advantages in screening. Clin Interv aging. 2016; 11: 967-976.
4. Gupta R, Bhatt LK, Johnston TP, Prabhavalkar KS. Colon cancer stem cells: Potential target for the treatment of colorectal cancer.Cancer Biol Ther. 2019; 20(8): 1068-1082.
5. El-Shami K, Oeffinger KC, Erb NL, Willis A, Bretsch JK, Pratt-Chapman ML, et al. American Cancer Society Colorectal Cancer Survivorship Care Guidelines CA Cancer J Clin. 2015; 65(6): 428-5.
6. Benson AB. Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology, J Natl Compr Canc Netw. 2021 Mar 2;19(3):329-359.
7. Thuraisingham B. Data mining: technologies,techniques, tools, and trends. Boca Raton: CRC. Press. 2014.
8. Barash Y, Klang E, Tau N. Evolution of inflammatory bowel disease research from a bird's-eye perspective: a text-mining analysis ofpublication trends and topics. Inflamm Bowel Dis. Epub ahead of print. 2020.
9. Wang S-H, Ding Y, Zhao W. Text mining for identifying topics in the literatures about adolescent substance use and depression. BMCPublic Health. 2016; 16: 279.
10. Zhang Y, Tao J, Wang J. Trends in diatomresearch since 1991 based on topic modeling. Microorganisms. 2019; 7: 213.
11. Song M and Kim SY. Detecting the knowledgestructure of bioinformatics by mining full-text collections. Scientometrics. 2013; 96: 183-201.
12. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020.CA Cancer J Clin. 2020; 70(3): 145-164.
13. Modest DP, Pant S, Sartore-Bianchi A. Treatment sequencing in metastatic colorectal cancer. Eur J Cancer. 2019; 109: 70-83.
14. Morse MA, Hochster H, Benson A. Perspectives on Treatment of Metastatic Colorectal Cancer with Immune Checkpoint Inhibitor Therapy. Oncologist. 2020; 25(1): 33-45.
15. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. Nat Rev Gastroenterol Hepatol. 2019; 16(6): 361-375.
16. Lech G, Słotwiński R, Słodkowski M, Krasnodębski IW. Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances.. World J Gastroenterol. 2016; 22(5): 1745-55.
17. Sveen A, Kopetz S, Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in complexity. Nat Rev Clin Oncol. 2020; 17(1): 11-32.
18. Coppedè F, Lopomo A, Spisni R, Migliore L. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. World J Gastroenterol. 2014; 20(4): 943-956.
19. Lin YC, Huang WT, Ou SC, Hung HH, Cheng WZ, Lin SS, et al. Neural network analysis of Chinese herbal medicine prescriptions for patients with colorectal cancer. Complement Ther Med. 2019; 42: 279-285.

20. Wu R, Wang L, Yin R, Hudlikar R, Li S, Kuo HD, et al. Epigenetics/epigenomics and prevention by curcumin of early stages of inflammatory-driven colon cancer. Mol Carcinog. 2020; 59(2): 227-236.

21. West HJ. Complementary and Alternative Medicine in Cancer Care. JAMA Oncol. 2018; 4(1): 139.

22. Keene MR, Heslop IM, Sabesan SS, Glass BD. Complementary and alternative medicine use in cancer: A systematic review. Complement Ther Clin Pract. 2019; 35: 33-4.

23. Turgeman I, Bar-Sela G. Cannabis for cancer - illusion or the tip of an iceberg: a review of the evidence for the use of Cannabis and synthetic cannabinoids in oncology. Expert Opin Investig Drugs. 2019; 28(3): 285-296.

24. Chung M, Kim HK, Abdi S. Update on cannabis and cannabinoids for cancer pain. Curr Opin Anaesthesiol. 2020; 33(6): 825-831

25. Hauser W, Welsch P, Klose P, Radbruch L, Fitzcharles MA. Efficacy, tolerability and safety of cannabis-based medicines for cancer pain: A systematic review with meta-analysis of randomised controlled trials. Schmerz. 2019; 33(5): 424-436.

26. Lal S, Shekher A, Puneet, Narula AS, Abrahamse H, Gupta SC. Cannabis and its constituents for cancer: History, biogenesis, chemistry and pharmacological activities. Pharmacol Res. 2021; 163: 105302.

27. Sawtelle L, Holle LM. Use of Cannabis and Cannabinoids in Patients With Cancer. Ann Pharmacother. 2021; 55(7): 870-890.

28. G Milano, MC Etienne-Grimaldi, L Dahan, M Francoual, JP Spano, D Benchimol, et al. Formento1Epidermal growth factor receptor (EGFR) status and K-Ras mutations in colorectal cancer. Ann Oncol. 2008; 19(12): 2033-2038.

29. FDA approves encorafenib in combination with cetuximab for metastatic colorectal cancer with a BRAF V600E mutation. US Food and Drug Administration. 2020.

30. Leigh Marcus, Steven J, Lemery, Patricia Keegan, Richard Pazdur. FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. Clin Cancer Res. 2019; 25(13): 3753-3758.