



OPEN ACCESS

EDITED BY

Xinpei Deng,
Sun Yat-sen University Cancer Center
(SYSUCC), China

REVIEWED BY

Tamer A. Addissouky,
University of Menoufia, Egypt
Xiyue Wang,
Stanford University, United States

*CORRESPONDENCE

Chaofan Chen,
✉ ccf901441@163.com

RECEIVED 27 January 2026

REVISED 11 April 2026

ACCEPTED 13 April 2026

PUBLISHED 30 April 2026

CITATION

Su H, Wang T and Chen C (2026) AI-enabled multi-omics integration in colorectal cancer: from molecular stratification to clinical translation. *Front. Cell Dev. Biol.* 14:1797221. doi: 10.3389/fcell.2026.1797221

COPYRIGHT

© 2026 Su, Wang and Chen. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

AI-enabled multi-omics integration in colorectal cancer: from molecular stratification to clinical translation

Honghua Su¹, Tiangui Wang¹ and Chaofan Chen^{2*}

¹Department of General Surgery, Kunming Municipal Hospital of Traditional Chinese Medicine, The Third Affiliated Hospital of Yunnan University of Chinese Medicine, Kunming, China, ²Department of Anorectal, Kunming Municipal Hospital of Traditional Chinese Medicine, The Third Affiliated Hospital of Yunnan University of Chinese Medicine, Kunming, China

Colorectal cancer (CRC) remains a heterogeneous disease for which improved molecular stratification is needed across the clinical pathway. Multi-omics technologies have expanded insight into CRC biology, and artificial intelligence (AI) has created new possibilities for integrating molecular, pathological, imaging, and clinical data. This review examines how these approaches are being applied across screening, diagnosis, treatment, and prognosis, with particular emphasis on their clinical relevance and translational limitations. We argue that, despite encouraging advances in biomarker discovery and risk prediction, most current studies remain retrospective and are constrained by heterogeneity of data sources, limited standardisation, weak interpretability, and insufficient external or prospective validation. AI-enabled multi-omics integration has substantial potential in CRC, but meaningful clinical impact will require rigorous validation and implementation frameworks suited to routine care.

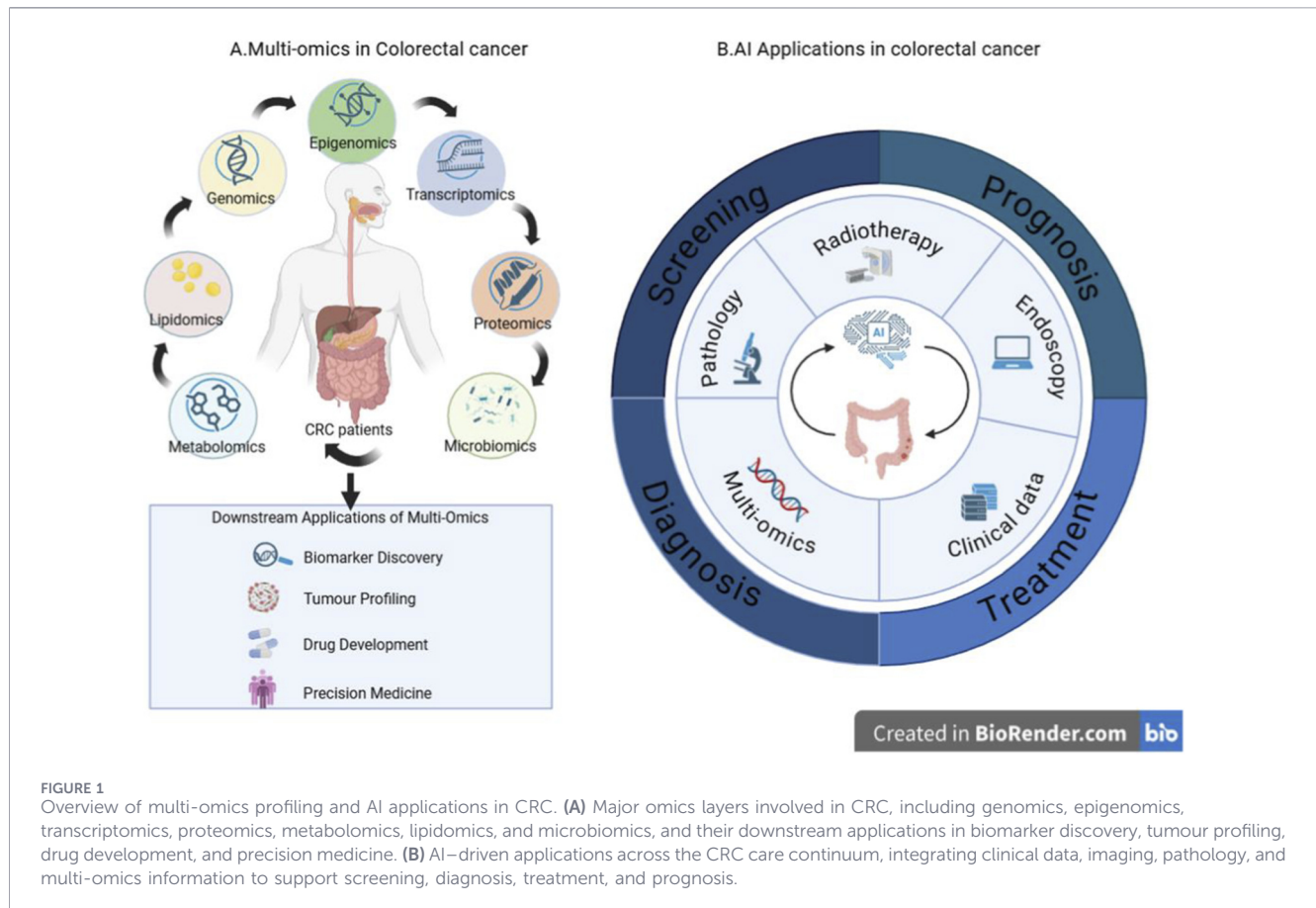
KEYWORDS

artificial intelligence, biomarker discovery, clinical decision support, colorectal cancer, machine learning, multi-omics, precision medicine

1 Introduction

Colorectal cancer (CRC) is a major global health challenge, representing nearly 10% of the 19.3 million cancer cases worldwide in 2020 (Sung et al., 2021). It is the third most common malignancy and the second leading cause of cancer-related mortality, with 3.2 million new cases and 1.6 million deaths projected by 2040 (Yoshino et al., 2021). CRC is a highly heterogeneous disease driven by complex genetic, epigenetic, and microenvironmental alterations, which substantially influence therapeutic response and clinical outcome (Punt et al., 2017; Tepeli et al., 2021). Although surgery, chemotherapy, and radiotherapy remain the cornerstones of CRC treatment and have improved survival in many patients (Biller and Schrag, 2021), their long-term benefits are limited for a considerable proportion of individuals, largely owing to interpatient molecular heterogeneity and treatment resistance (Punt et al., 2017; Andrei et al., 2022; Zhao et al., 2024). Together, these features make CRC a challenging disease to manage and highlight the need for more precise biologically informed approaches to classification and treatment (Punt et al., 2017; Buikhuisen et al., 2020).

Advances in high-throughput multi-omics technologies including genomics, transcriptomics, epigenomics, proteomics, metabolomics, microbiomics, and lipidomics have enabled systematic interrogation of CRC across multiple molecular



layers in a spatiotemporal context (Tang et al., 2019). While genomic profiling has facilitated the identification of CRC subtypes (Zhao et al., 2022), DNA-level alterations do not necessarily translate into functional changes in gene expression or protein activity (Xu J. Y. et al., 2020). Transcriptomic analyses have led to the definition of consensus molecular subtypes (CMS) (Guinney et al., 2015), yet the relationship between mRNA, protein abundance, and treatment response remains incompletely understood (Xu J. Y. et al., 2020).

Integrated multi-omics analyses provide a holistic view of cancer biology and enable the identification of molecular networks that drive tumour initiation and progression (Karczewski and Snyder, 2018; Hasin et al., 2017). In particular, proteome-centric integration strategies, supported by mass spectrometry-based proteomics, offer critical insights into functional dysregulation beyond genomic alterations and hold promise for biomarker discovery and personalized therapy development (Xu J. Y. et al., 2020; Li S. et al., 2023; Liu et al., 2024; Zhang et al., 2025).

Artificial intelligence (AI), especially deep learning (DL), has emerged as a powerful tool for extracting clinically actionable information from high-dimensional, multimodal datasets (Esteva et al., 2019; Kann et al., 2021). By integrating imaging, genomic, and molecular data, AI-based approaches have shown utility across several CRC tasks, including real-time polyp detection, radiogenomic prediction of molecular features, pathology-based prognosis modelling, and multimodal risk stratification; however, reported performance and generalisability vary substantially across

data modality, cohort composition, and validation setting (Kann et al., 2021).

Against this background, the present review is organised around the major omics layers relevant to CRC and their integration with AI across screening, diagnosis, treatment, and prognosis. The overall conceptual framework is outlined in Figures 1A,B.

This article was conducted as a structured narrative review. A targeted literature search was performed in PubMed to identify studies related to colorectal cancer, multi-omics technologies, and artificial intelligence. We prioritized recent and representative studies according to clinical relevance, methodological relevance, and their contribution to understanding AI applications and cross-omics integration in CRC. This review was not designed as a formal systematic review; therefore, no PRISMA-based study selection workflow was applied.

2 Multi-omics in CRC

2.1 Genomics of CRC

Genomic profiling has become a cornerstone in CRC management, enabling precise diagnosis and treatment. Key mutations, such as those in KRAS, BRAF, and PIK3CA, are crucial for understanding tumour biology and tailoring therapies. For instance, KRAS mutations occur in approximately 40% of CRC cases, leading to continuous activation of the RAS-MAPK pathway

and resistance to anti-EGFR therapies like cetuximab and panitumumab (Ahmad et al., 2021; Vitiello et al., 2019). Similarly, BRAF V600E mutations, found in 8%–10% of patients, contribute to more aggressive disease and resistance to standard therapies. Targeted inhibitors of BRAF and MEK, such as vemurafenib, offer more personalized treatment options (Dankner et al., 2018; Sanchez et al., 2018). Furthermore, PIK3CA mutations, which activate the PI3K-AKT-mTOR pathway, are associated with resistance to chemotherapy and anti-EGFR therapies in 15%–20% of CRC patients (Hu et al., 2021; Rascio et al., 2021). These insights allow clinicians to optimize treatment choices and avoid ineffective therapies.

Epigenetic modifications, including DNA methylation and histone modifications, also play a crucial role in CRC development and progression. Altered methylation patterns, such as hypermethylation of tumour suppressor genes and hypomethylation of oncogenes, contribute to carcinogenesis and microsatellite instability (MSI) (Zamani et al., 2018; Berdasco and Esteller, 2019). Non-invasive biomarkers, including methylation in cfDNA and CTCs, hold potential for early CRC detection (Galanopoulos et al., 2017; Krishnamurthy et al., 2017). The role of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) in CRC is increasingly recognised, with emerging studies suggesting their value in diagnostics and prognosis (Belkhir and El-Rifai, 2014; Ichimura et al., 2015; Hur et al., 2014). These epigenetic markers, along with genomic alterations, provide important insights into CRC's molecular landscape.

In CRC, genomic abnormalities such as single nucleotide variations (SNVs), copy number variations (CNVs), and microsatellite instability are prevalent. KRAS mutations are a major driver of CRC proliferation and metastasis (Zhang et al., 2009). Moreover, somatic CNVs detected in circulating tumour DNA have emerged as early biomarkers for CRC detection (Molparia et al., 2018). MSI, a key biomarker in CRC, also predicts the efficacy of PD-1 inhibitors in metastatic CRC patients, offering new therapeutic avenues (Lin et al., 2020). Together, these genomic and epigenetic insights are pivotal in refining CRC diagnosis, treatment, and prognosis, and in advancing precision medicine for this disease.

2.2 Transcriptomics of CRC

Transcriptomic analysis has proven essential in uncovering biomarkers within the CRC microenvironment, particularly by evaluating the composition of immune and stromal cell populations. Tumours of different molecular subtypes exhibit distinct tumour microenvironment (TME) characteristics, which correlate with tumour mutational burden, affecting prognosis and the efficacy of immunotherapies (Petitprez et al., 2018). Recent advancements, particularly single-cell RNA sequencing (scRNA-seq), have revealed insights into the involvement of neutrophils in iron metabolism and migration in CRC (Gao et al., 2022; Zhang Y. et al., 2020). This technology holds promise in providing more specific diagnostic and prognostic biomarkers, along with feasible therapeutic targets (Zhang et al., 2014). RNA-seq has also facilitated the identification of differentially expressed genes in CRC, enabling the development of risk score models for prognosis (Zhang X. et al., 2023). Pathways such as Wnt/ β -catenin, PI3K/Akt, and TGF- β are

crucial in CRC progression and serve as potential biomarkers (Sebio et al., 2014; Duan et al., 2018). Further, genes like c-MYC, KRAS, BRAF, and PIK3CA have been identified as key predictive markers for CRC prognosis (Marmol et al., 2017).

Despite its potential, transcriptomics in CRC research faces challenges. Current technologies have not yet achieved true single-cell resolution, and capturing the structural and functional diversity of transcripts, such as alternative splicing and modifications, remains a limitation. Moreover, the vast scale and complexity of transcriptomic data pose significant hurdles in data processing, integration, annotation, and interpretation. The development of more efficient algorithms and tools is critical to overcoming these challenges. While transcriptomics is pivotal for identifying biomarkers and assessing prognosis, its clinical applications face issues such as standardization, reproducibility, and sensitivity.

Genomic instability and epigenetic changes in CRC often lead to aberrant alterations in the transcriptome, reflecting the cell's activity at any given time (Wang et al., 2009; Habermann et al., 2008). High-throughput RNA-Seq has enabled a more comprehensive mapping of the CRC transcriptome, revealing novel transcripts, alternative splicing events, and gene fusion products (Wang et al., 2009; Morozova et al., 2009). Studies using RNA-Seq to compare CRC tumour tissues with adjacent normal tissues have identified key activated pathways, including extracellular matrix remodeling and metabolic pathways, which are associated with tumour metastasis (Wu et al., 2012). In another study, the most frequently altered genes in CRC, including APC, TP53, and KRAS, were identified through high-throughput sequencing (Han et al., 2013). RNA-Seq is increasingly becoming the preferred method for profiling the CRC transcriptome, offering deeper insights into its genetic alterations (Wang et al., 2009).

2.3 Proteomics of CRC

Proteomics has emerged as a critical approach for identifying functional biomarkers in CRC, particularly proteins that can be detected in accessible biofluids such as blood, urine, and saliva (Tjalsma, 2010; Alvarez-Chaver et al., 2014). Compared with genomic or transcriptomic alterations, the proteome represents the functional output of the genome and is directly linked to tumour behaviour. Numerous studies have demonstrated that CRC is characterized by distinct proteomic signatures, including dysregulated expression and post-translational modifications (PTMs), which play essential roles in early diagnosis, prognostic stratification, and therapeutic intervention (Zhu et al., 2022). Recent advances combining mass spectrometry with machine learning (ML) have further enhanced biomarker discovery; for example, a convolutional neural network model identified a five-protein panel (SAA2, APCS, APOA4, F2, and AMBP) for early and advanced CRC detection (Ahn et al., 2019).

Proteomic profiling of tumour tissues and stromal compartments has revealed proteins involved in CRC initiation, progression, and invasion. Differential expression of ACTBL2, DPEP1, cyclophilin A, annexin A2, and aldolase A has been consistently reported across tissue-based proteomic studies using LC-MS platforms (Ghazanfar et al., 2017; Hao et al., 2017; Yamamoto et al., 2016). Proteomics has also enabled the

identification of stromal and microenvironment-related biomarkers, including LTBP2, OLFML3, CDH11, CALU, and FSTL1, which are implicated in tumour migration and invasion (Torres et al., 2013). Moreover, proteome-level analyses of patient-derived organoids revealed that activation of MYC and enrichment of the TRiC chaperonin complex predicted sensitivity to palbociclib, highlighting the potential of proteomics for guiding treatment selection (Papaccio et al., 2023). Large-scale proteomic classification studies further identified biologically distinct CRC subtypes associated with immune regulation and therapeutic response, including resistance to radiotherapy and anti-EGFR therapy mediated by HSF1 and HDAC6, respectively (Li Y. et al., 2023).

Circulating proteomics has shown particular promise for non-invasive CRC screening, risk stratification, and disease monitoring. Multiple blood-based protein panels have demonstrated favorable diagnostic performance, with reported sensitivities exceeding 70% and specificities approaching 90% (Ivancic et al., 2020; Bhardwaj et al., 2019). Proteins such as EGFR, LRG1, ITIH4, HPX, SOD3, osteopontin, and amphiregulin have been repeatedly validated in plasma and serum cohorts (Ivancic et al., 2020; Bhardwaj et al., 2019; Chantaraamporn et al., 2020). Chemokines including CXCL8 have also shown superior diagnostic performance compared with carcinoembryonic antigen (CEA) (Paczek et al., 2020). Importantly, integration of blood proteomics with polygenic risk scores and clinical models has improved CRC risk stratification and informed individualized screening strategies (Sun et al., 2024).

In metastatic CRC, proteomic analyses of circulating tumour cells revealed heterogeneous epithelial–mesenchymal transition (EMT) states, identifying ERBB2, COL6A1, and CAVIN1 as promising metastasis-associated biomarkers and potential therapeutic targets (Huang et al., 2022). Despite these advances, effective integration of proteomics with other omics layers remains a major challenge, underscoring the need for standardized, scalable, and clinically translatable analytical frameworks.

2.4 Microbiomics of CRC

Microbiome profiling has identified CRC-associated microbial signatures with potential value in screening, diagnosis, and treatment-response prediction. Advances in 16S rRNA sequencing, shotgun metagenomics, and quantitative PCR have enabled increasingly detailed characterisation of gut microbial alterations in CRC (Liu et al., 2023; Holmes et al., 2012; Clos-Garcia et al., 2020; Zeller et al., 2014; Zackular et al., 2014; Yang et al., 2020).

Accumulating evidence indicates that the gut microbiome plays a critical role in CRC screening, diagnosis, and therapeutic response prediction. Specific microbial taxa, such as *Fusobacterium nucleatum*, are enriched in patients with adenomas and CRC and can be detected in fecal samples, highlighting their potential as non-invasive screening biomarkers (Guo et al., 2018; McQuade et al., 2019). Beyond local intestinal effects, recent studies have demonstrated that gut-derived microbial DNA can be detected in the circulation and may discriminate between different cancer types, supporting the concept of microbiome-based liquid biopsy approaches (Li et al., 2022). These findings underscore the

systemic relevance of gut dysbiosis in CRC pathogenesis and clinical management.

Integration of microbiomics with metabolomics has further strengthened biomarker discovery in CRC. Microbial metabolites, including short-chain fatty acids, bile acids, and polyamines, have been consistently linked to colorectal carcinogenesis (Weir et al., 2013; Brown et al., 2016; Lin et al., 2019). For example, altered levels of acetate, butyrate, and secondary bile acids have been observed in CRC tissues and fecal samples, demonstrating diagnostic potential. Large-scale serum metabolomic analyses have identified reproducible metabolite signatures associated with gut microbiota that distinguish healthy individuals, adenoma patients, and CRC patients with high diagnostic accuracy (Chen et al., 2022). Collectively, these data suggest that microbiome–metabolome interactions are central to CRC development and represent promising targets for early detection, prevention, and personalized intervention strategies (Thomas et al., 2017; Nakatsu et al., 2015; Kwong et al., 2018).

2.5 Metabolomics of CRC

In CRC, metabolomics provides a functional readout of tumour-associated biochemical reprogramming and has become increasingly relevant for biomarker discovery using minimally invasive biospecimens such as blood, urine, stool, and saliva (Troisi et al., 2020; Zhang et al., 2015; Dunn et al., 2011; Ivanisevic and Want, 2019). Both targeted and untargeted metabolomic approaches have been applied to characterise metabolic alterations associated with CRC development and progression (Vignoli et al., 2019; Commisso et al., 2013).

Accumulating metabolomic studies across diverse CRC biospecimens have consistently demonstrated profound metabolic reprogramming during tumour initiation and progression (Wang et al., 2010; Nannini et al., 2020). Using analytical platforms such as nuclear magnetic resonance (NMR), gas chromatography–mass spectrometry (GC–MS), and liquid chromatography–mass spectrometry (LC–MS), investigators have identified dysregulation of pathways including glycolysis, the tricarboxylic acid cycle, amino acid metabolism, lipid metabolism, bile acid metabolism, and polyamine synthesis (Udo et al., 2020; Ning et al., 2021). Several studies have shown that metabolite profiles can distinguish healthy individuals, patients with benign polyps, and those with CRC, as well as differentiate early-stage from advanced-stage disease (Gu et al., 2019; Kim et al., 2019). Notably, polyamines (e.g., putrescine and cadaverine), short-chain fatty acids, bile acids, and sphingolipids have repeatedly been reported as candidate biomarkers for CRC diagnosis and prognosis (Clos-Garcia et al., 2020; Lin et al., 2019; Yang et al., 2019; Song et al., 2018).

At a functional level, the metabolome directly reflects biochemical reaction outputs and is highly responsive to both intrinsic oncogenic signaling and extrinsic environmental stimuli (Martinez-Reyes and Chandel, 2021; Wishart, 2019). Tumour cells undergo extensive metabolic rewiring to sustain uncontrolled proliferation, while immune cell differentiation and function are likewise governed by metabolic programs (Schmidt et al., 2021; Johnson et al., 2016; Boothby and Rickert, 2017). These metabolic vulnerabilities generate characteristic metabolic phenotypes that can be exploited for early cancer detection, disease monitoring, and

therapeutic targeting (Schmidt et al., 2021). Recent clinical and preclinical studies have further highlighted the value of metabolomics in drug discovery and pharmacodynamic evaluation, including the identification of metabolite signatures associated with treatment response (Zhang et al., 2018; Yin et al., 2022). Nevertheless, no single omics layer sufficiently captures the complexity of CRC biology, underscoring the necessity of integrating metabolomics with other omics technologies to achieve robust biomarker validation and mechanistic insight.

2.6 Lipidomics of CRC

Lipidomic profiling has revealed clinically relevant metabolic perturbations in CRC, with mass spectrometry-based approaches enabling detailed characterisation of lipid signatures associated with tumour progression and prognosis (Burla et al., 2018; Wang R. et al., 2020; Kyle, 2020). These analyses provide insight into CRC-related lipid remodelling and may support biomarker discovery and risk stratification (Wang R. et al., 2020; Kyle, 2020).

In CRC, growing evidence indicates that dysregulated lipid metabolism plays a critical role in tumour initiation and progression. Recent studies have demonstrated altered serum and tissue lipid profiles in CRC patients, characterized by increased very-long-chain fatty acids (VLCFAs) and reduced long-chain fatty acids (LCFAs), potentially driven by elongation of very-long-chain fatty acid (ELOVL) enzymes (Pakiet et al., 2019; Kondo et al., 2011). Lipidomic analyses have further shown that saturated triacylglycerols account for a substantial proportion of lipid perturbations during CRC progression, likely reflecting lipid droplet accumulation (Liu et al., 2019). Additional alterations, including decreased levels of glycolipids, glycerophosphocholine, and acylcarnitines, have also been reported in CRC serum samples (Tevini et al., 2022). Importantly, tissue-based lipidomic signatures, particularly triglyceride profiles, have emerged as independent prognostic markers capable of discriminating CRC subgroups according to mismatch repair status, oncogenic mutations (KRAS/BRAF), and tumour grade (Ecker et al., 2021).

Beyond biomarker discovery, lipidomics provides unique insights into lipid-protein, lipid-gene, and lipid-lipid interaction networks that underlie cancer biology. As dysregulated lipid metabolism is increasingly recognised as a hallmark of cancer (Snaebjornsson et al., 2020), clinical lipidomics has emerged as a translational subfield aiming to improve disease prediction, diagnosis, and therapeutic stratification (Dehairs et al., 2015; Moreno et al., 2022). Compared with traditional diagnostic approaches, lipidomics offers a systems-level view of metabolic dysregulation and enables precise identification of therapeutic targets and diagnostic biomarkers (Lv et al., 2018). Nonetheless, substantial challenges remain, including methodological heterogeneity, lack of standardization, and difficulties in cross-study comparison, which currently limit clinical translation. Addressing these challenges will be essential for integrating lipidomics into precision oncology and advancing its application in CRC management.

2.7 Epigenomics of CRC

Epigenetic dysregulation plays a central role in the initiation, progression, and therapeutic resistance of CRC, and epigenomic

analyses have increasingly been applied to identify clinically relevant biomarkers and targets. Aberrant DNA methylation patterns have been associated with CRC screening, staging, and disease stratification. Tissue-based methylation markers such as SFMBT2, ITGA4, THBD, and ZNF304 have shown potential for early CRC detection, whereas methylation of KCNJ12, VAV3-AS1, and EVC has been linked to disease staging. In addition, NEUROD1 and FAM72C have been proposed as dual biomarkers to distinguish non-advanced precancerous lesions from inflammatory bowel disease (Zhang X. et al., 2020). Fecal DNA methylation markers, including SEPT9 and SDC2, have demonstrated value as non-invasive tools for early CRC detection (Zhao et al., 2020). These findings highlight the translational relevance of epigenetic alterations in CRC diagnosis and risk stratification.

Beyond DNA methylation, multiple layers of epigenetic regulation contribute to CRC pathogenesis. Histone modifications are increasingly recognised as key determinants of malignant phenotypes. For example, ubiquitination-mediated degradation of HDAC3 reduces histone acetylation and promotes cancer stem cell-associated gene expression, revealing a novel mechanism of epigenetic regulation in CRC (Zhan et al., 2020). Moreover, VprBP kinase-mediated phosphorylation of histone H2A at threonine 120 drives oncogenic transcriptional programs and enhances colon cancer cell proliferation, identifying VprBP as a potential therapeutic target (Ghate et al., 2021). Advances in epigenomic profiling technologies—including bisulfite sequencing for DNA methylation (Singer, 2019), chromatin immunoprecipitation sequencing (ChIP-seq) for histone modifications (Neganova et al., 2022), and multiple high-throughput methods for RNA modifications such as N6-methyladenosine (m6A) mapping (Dominissini et al., 2012; Linder et al., 2015; Molinie et al., 2016; Meyer, 2019; Zhang Z. et al., 2019; Wang Y. et al., 2020)—have enabled comprehensive characterization of epigenetic landscapes in CRC.

Despite these advances, substantial challenges remain. CRC exhibits pronounced intertumour and intratumour heterogeneity, with distinct genomic and epigenetic profiles across patients and even within individual tumours, complicating the identification of universally applicable biomarkers or therapeutic targets. Furthermore, most epigenomic studies have focused on coding regions, whereas the roles of non-coding elements—including regulatory regions and long non-coding RNAs—remain incompletely understood in CRC pathogenesis. Importantly, unlike genetic mutations, epigenetic alterations are potentially reversible, making them attractive targets for therapeutic intervention (Jung et al., 2020). The growing interest in epigenetic therapies, such as DNA methyltransferase inhibitors, reflects this therapeutic promise and underscores the need for deeper mechanistic insight into how epigenetic and genetic factors interact to drive CRC development and metastasis (Muller and Gyorffy, 2022; Rezapour et al., 2019; Thomas and Marcato, 2018). To provide a more clinically relevant overview of the major omics layers in CRC, Table 1 summarises representative biomarkers, sample types, clinical applications, and the translational evidence already described in the current literature.

TABLE 1 Clinical applications of multi-omics technologies in CRC.

Omics layer	Representative sample types	Key biomarkers/signatures	Main clinical applications	Representative clinical/translational evidence already described in the manuscript	Major limitations
Genomics	Tumour tissue, plasma cfDNA, ctDNA	KRAS, BRAF, PIK3CA, APC mutations; MSI status	Treatment selection, prognosis stratification, immunotherapy eligibility	Genomic profiling is described as a cornerstone of CRC management; MSI is described as a key biomarker that predicts PD-1 inhibitor efficacy in metastatic CRC	Limited functional insight; mutations do not always translate into phenotypic changes
Epigenomics	Tumour tissue, stool DNA, cfDNA	DNA methylation panels (e.g., SEPT9, SDC2, SFMBT2, KCNJ12)	Early screening, disease staging, non-invasive diagnosis	Tissue-based methylation markers are described as having potential for early detection and staging; stool SEPT9 and SDC2 are described as having demonstrated value as non-invasive early detection tools	Tumour heterogeneity; lack of standardized validation pipelines
Transcriptomics	Bulk tumour tissue, scRNA-seq	CMS subtypes; immune-related gene signatures; ncRNAs	Molecular subtyping, tumour microenvironment profiling, prognosis	scRNA-seq is described as revealing tumour microenvironment features and clinically relevant biomarkers; RNA-seq is described as enabling risk score models for prognosis	Batch effects; incomplete mRNA–protein concordance
Proteomics	Tumour tissue, plasma, serum	Protein panels (EGFR, LRG1, SAA2, ITIH4); PTMs	Early detection, risk stratification, therapy response assessment	Multiple blood-based protein panels are described as showing sensitivities exceeding 70% and specificities approaching 90%; several proteins are described as repeatedly validated in plasma and serum cohorts	Technical variability; limited generalisability across platforms
Metabolomics	Serum, plasma, urine, stool	Polyamines, bile acids, SCFAs, amino acid signatures	Early detection, disease monitoring, metabolic phenotyping	Metabolite profiles are described as distinguishing healthy individuals, benign polyps, and CRC, and as differentiating early-stage from advanced-stage disease	High biological variability; complex data interpretation
Lipidomics	Serum, tumour tissue	VLCFA/LCFA imbalance; glycerophospholipids; sphingolipids	Prognosis assessment, metabolic vulnerability identification	Tissue-based lipidomic signatures, particularly triglyceride profiles, are described as independent prognostic markers that discriminate CRC subgroups by mismatch repair status, KRAS/BRAF mutations, and tumour grade	Methodological heterogeneity; limited cohort validation
Microbiomics	Stool, tumour tissue	<i>Fusobacterium nucleatum</i> , microbial metabolite profiles	Screening, prognosis, treatment response prediction	Specific microbial taxa are described as enriched in adenoma and CRC and detectable in faecal samples; microbiome–metabolome signatures are described as having diagnostic potential	Causality unclear; inter-individual variability

3 Artificial intelligence (AI) in CRC

3.1 AI methodological framework for CRC multimodal analysis

AI methodologies used in CRC research can be understood across three complementary dimensions: learning paradigm, model

architecture, and interpretability. At the broadest level, AI refers to computational approaches that support data-driven prediction and decision-making (Koteluk et al., 2021). Machine learning, a major subfield of AI, enables algorithms to identify patterns from large datasets and apply learned rules to new data for classification or prediction (Lapiente-Santana et al., 2021; Langley, 2011). Deep learning represents a further subset of ML based on multilayer

neural networks, which can learn hierarchical feature representations directly from high-dimensional inputs (Hamet and Tremblay, 2017; Bengio, 2009).

From the perspective of learning paradigm, supervised learning remains the dominant strategy in clinically oriented CRC research because it relies on labelled data and is well suited to diagnostic, prognostic, and treatment-prediction tasks (Zhu et al., 2020; Alizadehsani et al., 2021; Bilal et al., 2021). Common supervised algorithms include logistic regression, support vector machine, naïve Bayes, gradient boosting, decision trees, random forests, and neural networks. Unsupervised learning, by contrast, identifies latent structure in unlabelled data and is mainly used for clustering, dimensionality reduction, and exploratory analyses (Hamet and Tremblay, 2017; Guo et al., 2023). Semi-supervised learning extends this framework by combining labelled and unlabelled data, thereby reducing annotation burden while improving model performance. Closely related weakly supervised approaches are particularly relevant to CRC digital pathology, where slide-level labels are often more feasible than exhaustive patch-level annotation. Reinforcement learning represents a distinct paradigm in which learning is guided by rewards or penalties linked to expected outcomes (Zhu et al., 2020; Alizadehsani et al., 2021).

Model selection in CRC is also shaped by data modality. Classical ML algorithms are most commonly applied to structured clinical variables, complete blood count indices, serum or stool biomarkers, and selected omics features after feature engineering. Their main advantages include lower data requirements, greater transparency, and easier integration with conventional clinical prediction frameworks. By contrast, DL models are most widely used for image-rich tasks, particularly in endoscopy, radiology, digital pathology, and surgical video analysis. Among these, convolutional neural networks (CNNs) have become central to image classification, detection, and segmentation, whereas recurrent neural networks (RNNs) and related sequential architectures have been explored for temporally ordered data. Compared with handcrafted-feature pipelines, DL approaches often achieve stronger predictive performance, but they typically require larger annotated datasets and are more vulnerable to limited interpretability, domain shift, and centre-specific artefacts (Hamet and Tremblay, 2017).

Recent methodological development has increasingly focused on reducing dependence on manual labels. Self-supervised learning enables models to learn transferable and discriminative feature representations from unlabelled data using internal learning signals rather than expert annotations (Niehues et al., 2023; Claudio Quiros et al., 2024). This shift has helped to establish the conceptual basis for foundation models, which are large-scale pretrained models that can be adapted to multiple downstream tasks with limited fine-tuning (Chen et al., 2024; Wang et al., 2024). Such approaches are particularly promising for CRC pathology and multimodal prediction because they may improve transferability across tasks and cohorts while lowering annotation requirements.

Recent advances in general-purpose foundation models and large multimodal models have further extended this paradigm. Medical variants of the Gemini family, including Med-Gemini, are based on multimodal architectures capable of processing and reasoning across text, images, and code, thereby supporting more complex tasks such as clinical reasoning, literature synthesis, report

generation, and early diagnostic assistance (Alhur, 2024; Sonoda et al., 2024; Yang et al., 2024; Wagner et al., 2020; Fedus et al., 2022). In computational pathology, models such as CHIEF, RetCCL, and CTransPath illustrate how large-scale pretraining can improve transferability across diverse downstream tasks. CHIEF, developed through unsupervised tile-level pretraining and weakly supervised whole-slide pretraining on large-scale histopathology data, has shown strong generalisability across centres, scanners, and specimen types, and has demonstrated robust performance in tumour origin classification, genomic profile prediction, MSI prediction in CRC, and survival stratification (van der Laak et al., 2021; Song et al., 2023; Fu et al., 2020; Kather et al., 2020; Echle et al., 2021; Wagner et al., 2023; Azizi et al., 2023). RetCCL further highlights the value of clustering-guided contrastive learning for whole-slide representation learning and interpretable image retrieval without task-specific fine-tuning (Wang et al., 2023). Likewise, CTransPath combines a convolutional neural network backbone with a multi-scale Swin Transformer and uses semantically relevant contrastive learning to derive transferable histopathology representations across multiple downstream tasks, including weakly supervised whole-slide classification and colorectal adenocarcinoma gland segmentation (Wang et al., 2022). In radiology, Merlin represents a 3D vision-language foundation model for CT imaging that was pretrained using paired CT scans, radiology reports, and structured electronic health record data, with both internal and external validation across a broad range of clinically relevant tasks. Collectively, these models suggest that large-scale pretraining, self-supervised or weakly supervised learning, and multimodal alignment may improve feature reuse, generalisation, and label efficiency across pathology and radiology tasks relevant to CRC.

PROGPATH, CRCFound, and DINOPath represent recent advances in foundation AI models for cancer prognosis and personalized treatment. PROGPATH integrates histopathological image features with routinely collected clinical variables to achieve pancancer survival prediction using a weakly supervised deep learning framework. Trained on 7,999 whole-slide images from 6,670 patients across 15 cancer types and validated on 7,374 images from 4,441 patients in 17 external cohorts, it demonstrated robust performance across cancer types, stages, treatment modalities, and biomarker-defined subgroups, while interpretability analyses identified key morphological patterns driving risk stratification (Yuan et al., 2025). CRCFound, a self-supervised CT image foundation model for colorectal cancer pretrained on 5,137 unlabelled scans, achieved strong adaptability and generalisation across six diagnostic and two prognostic tasks, addressing the challenge of limited annotated data (Yang et al., 2025). DINOPath, trained on over 130 million patches from 104,876 whole-slide images of gastrointestinal cancers, accurately predicted disease-free and disease-specific survival, stratified patients into clinically relevant risk groups, and predicted benefit from adjuvant chemotherapy, demonstrating the potential for foundation models to complement clinicopathologic factors and inform personalized therapy (Wang X. et al., 2025). Collectively, these models highlight the promise of large-scale, self-supervised, and multimodal AI for robust, generalizable, and clinically actionable cancer prognosis and treatment planning.

Interpretability represents a separate but equally important dimension. In principle, AI models can be viewed as explainable, partially explainable, or black-box systems according to the extent to which their internal decision process can be understood by human experts (Keyl et al., 2025; Binder et al., 2021; Shephard et al., 2024). Explainable models aim to provide insight into the features or patterns driving prediction, whereas black-box models—often highly complex architectures—may achieve strong performance at the cost of limited transparency (Coudray et al., 2018). Accordingly, the methodological value of an AI system in CRC should not be judged by predictive accuracy alone, but also by its suitability for the input modality, annotation burden, interpretability requirements, and potential for robust clinical translation.

3.2 Applications in CRC screening

Screening is essential for reducing CRC incidence and mortality by enabling early detection and timely intervention (Siegel et al., 2021). Although colonoscopy remains the most effective screening tool (Issa and Nouredine, 2017), its performance is limited by operator dependence and a substantial miss rate, particularly for small or flat lesions (van Rijn et al., 2006). AI-assisted endoscopy has emerged as a promising solution, substantially improving the detection, classification, localisation, and segmentation of colorectal polyps while reducing inter-observer variability. Deep learning-based systems have demonstrated high sensitivity and specificity for real-time identification of diminutive and non-polypoid lesions, shortening reading time and reducing missed diagnoses (Yamada et al., 2019; Park et al., 2024; Wang J. et al., 2025).

Beyond endoscopy, AI-driven risk stratification models using large-scale population data have expanded CRC screening strategies. Machine learning approaches integrating demographic variables (Hilsden et al., 2018), complete blood count indices (Hornbrook et al., 2017; Kinar et al., 2016), serum or stool biomarkers (Ivancic et al., 2020; Pan et al., 2021), and multi-omics data have shown robust predictive performance (Nazari et al., 2021; Wan et al., 2019). In primary care settings, CBC-based models achieved AUCs above 0.80 and significantly increased case detection when combined with FOBT, supporting their role as scalable triage tools for colonoscopy referral (Park et al., 2024; Hornbrook et al., 2017; Kinar et al., 2016). In parallel, blood-based protein and N-glycan biomarkers analysed with machine learning have demonstrated high sensitivity for detecting advanced adenomas and CRC (Ivancic et al., 2020; Pan et al., 2021), while plasma cfDNA sequencing combined with ML enabled accurate detection of early-stage CRC (AUC up to 0.92) (Wan et al., 2019).

Mismatch repair deficiency (dMMR) and MSI are critical biomarkers in CRC, present in approximately 15% of cases and closely linked to immunotherapy response (De'Angelis et al., 2018; Vilar and Gruber, 2010; Saillard et al., 2023). Recent studies have shown that deep learning models can directly predict MSI status from histopathological images, achieving strong discriminatory performance without the need for immunohistochemistry or genetic testing (Jiang et al., 2022). More recent work has extended this field from proof-of-concept prediction towards benchmarking of pathology foundation models for MSI

assessment in CRC, highlighting the importance of external generalisability and model comparison across architectures (Bilal et al., 2026). AI-based tools such as MSIIntuit further support clinical MSI screening by reliably excluding microsatellite-stable tumours and accurately identifying MSI cases (Saillard et al., 2023). Nevertheless, challenges remain, including limited generalisability, lack of standardised thresholds across laboratories, and insufficient long-term outcome data. Future CRC screening strategies should therefore focus on integrating AI technologies into evidence-based pathways while carefully balancing early detection benefits against the risks of overdiagnosis (Mori et al., 2021; Ozawa et al., 2020).

3.3 Applications in CRC diagnosis and staging

The diagnosis of CRC encompasses both qualitative confirmation and disease staging, aiming to establish pathological evidence of malignancy and to assess tumour extent and severity (Song et al., 2019). Conventional qualitative diagnosis relies on biopsy obtained during endoscopy or surgery followed by histopathological evaluation, whereas staging primarily depends on radiological imaging, including CT and MRI. The integration of AI into diagnostic workflows seeks to enhance efficiency, reduce clinician workload, improve image interpretability, and ultimately decrease rates of misdiagnosis and missed diagnosis.

Histopathology remains the gold standard for CRC diagnosis and underpins postoperative staging (Acs et al., 2020), yet diagnostic accuracy is affected by inter- and intra-observer variability. AI applications in digital pathology have therefore focused on gland segmentation and tumour classification. Deep learning-based models have demonstrated robust performance in segmenting colorectal glands and distinguishing benign from malignant tissues, achieving high accuracy and reproducibility across heterogeneous datasets (Ding et al., 2020; Yuan et al., 2024). Machine learning classifiers, including support vector machines and hybrid feature-based approaches, have further improved CRC detection and grading accuracy, in some cases exceeding 95% (Yuan et al., 2024; Rathore et al., 2015). In addition, AI-assisted analysis of immunohistochemistry images has shown promise in predicting lymph node metastasis, suggesting potential complementary roles to conventional histopathology (Takamatsu et al., 2019). Likewise, multi-institutional clinical validation studies have begun to extend digital pathology AI towards more specific staging-related tasks, including automated detection of lymph node metastasis in CRC specimens (Giammanco et al., 2024). Emerging studies combining pathological images with immune and genetic data further support AI-enabled individualized diagnostic strategies relevant to targeted therapy and immunotherapy (Ge et al., 2019).

Radiological imaging constitutes the cornerstone of CRC staging, and AI-driven radiomics has become a major focus of diagnostic research (Yun and Guangwei, 2020). CT-based computer-aided detection systems have demonstrated high sensitivity for identifying flat or subtle lesions, supporting their role as non-invasive screening and diagnostic adjuncts (Taylor et al., 2008). Convolutional neural network-based models applied to CT have also shown performance comparable to expert radiologists in

detecting hepatic metastases, while offering improved diagnostic confidence (Khalili et al., 2020). Notably, radiomic approaches have enabled the non-invasive prediction of molecular features such as KRAS mutation status directly from CT images, providing a potential alternative to invasive tissue sampling (González-Castro et al., 2020). More recent original studies continue to support the feasibility of CT-based radiomics for non-invasive KRAS prediction in CRC, reinforcing the translational potential of radiogenomic modelling (Wang W. et al., 2025). In rectal cancer, deep learning applied to MRI has achieved high accuracy in tumour and lymph node segmentation and metastasis prediction, with markedly reduced interpretation time compared with radiologists (Lu et al., 2018; Trebeschi et al., 2017).

Despite these advances, most AI applications in CRC diagnosis remain focused on pathological and radiological images (Acs et al., 2020; Bad et al., 2021). Challenges include limited interpretability of deep learning “black-box” models (Zhao and Hastie, 2019), variability in image acquisition quality, lack of standardized public datasets, and insufficient integration of clinical data. Moreover, AI-driven diagnostic tools linked to targeted therapies and immunotherapy remain underexplored, and robust prospective validation is still lacking. Nevertheless, AI-assisted diagnostic systems have shown promising gains in selected CRC tasks, particularly in digital pathology and radiology, but reported performance remains dependent on image quality, task definition, and validation setting, and robust prospective evidence is still limited (Mori et al., 2018; Wang P. et al., 2020; Liang et al., 2016).

3.4 Applications in CRC treatment

Treatment of CRC relies on chemotherapy, surgery, and multidisciplinary comprehensive management strategies (Bondevén et al., 2015). The integration of AI into CRC treatment has primarily focused on enhancing treatment safety, optimizing surgical management, and supporting clinical decision-making.

One important clinical application of AI in systemic therapy is the management of treatment-related toxicity. Irinotecan (CPT-11), a commonly used chemotherapeutic agent for CRC, is associated with substantial adverse effects that often limit its clinical use. Machine learning models incorporating clinical variables and serum biomarkers have been developed to predict irinotecan-induced toxicities, including diarrhea, leukopenia, and neutropenia, thereby assisting clinicians in individualized risk assessment and treatment planning (Oyaga-Iriarte et al., 2019). In contrast, AI-based toxicity management strategies for other standard CRC agents remain underexplored.

AI has demonstrated particularly strong potential in CRC surgery. Deep learning-based systems have been developed for automatic recognition of surgical phases and actions during laparoscopic and robotic colorectal procedures, achieving high accuracy and enabling real-time analysis of surgical workflows (Kitaguchi et al., 2020; Kitaguchi et al., 2022; Kolbinger et al., 2024). Image-based AI models have further been applied to identify anatomical structures and tissue planes, supporting intraoperative guidance and surgical standardization (Igaki et al., 2022). In addition, AI-assisted analysis of intraoperative perfusion

imaging and perioperative clinical data has shown promise in predicting surgical complications, such as anastomotic leakage, thereby contributing to risk stratification and perioperative decision-making (Park et al., 2020; Mazaki et al., 2021).

Beyond intraoperative applications, AI technologies have also been applied to surgical training and postoperative management. Deep learning models trained on large-scale surgical video datasets can objectively assess surgical skills and procedural quality, offering valuable tools for colorectal surgical education (Kitaguchi et al., 2021). Moreover, machine learning models integrating perioperative clinical data have been developed to predict postoperative outcomes, including length of hospital stay, readmission, and short-term mortality, supporting postoperative care planning and healthcare resource optimization (Masum et al., 2022).

Overall, although AI applications in CRC treatment remain at an early stage and are limited by small datasets and lack of prospective validation (Huang et al., 2020), current evidence suggests that AI may improve selected aspects of treatment-related risk estimation, surgical workflow analysis, and perioperative management; however, the clinical magnitude of benefit has not yet been established through prospective evaluation. Future efforts should prioritize multimodal data integration, standardized model development, and real-world clinical validation to facilitate the translation of AI technologies into routine CRC treatment pathways.

3.5 Applications in CRC prognosis

Prognosis in CRC mainly involves predicting recurrence and long-term survival (Huang et al., 2020). Traditional prognostic models based on clinicopathological variables and conventional statistics are limited in their ability to capture the biological heterogeneity and complex, nonlinear interactions that drive disease progression. AI offers a data-driven framework capable of integrating multidimensional information and extracting high-order features, thereby enabling more accurate and individualized prognostic assessment.

AI-based models have been widely applied to recurrence prediction following curative resection, which is critical for postoperative surveillance and adjuvant treatment selection (Weiser et al., 2008). Early nomogram-based approaches improved risk stratification beyond standard staging systems but remained constrained by intrastage heterogeneity. More recent machine learning and deep learning methods have demonstrated superior performance in stage-specific recurrence prediction, particularly in stage II and IV CRC, by leveraging clinical variables and imaging-derived features (Takenaka et al., 2020; Xu Y. et al., 2020; Li et al., 2019). These approaches allow for finer risk discrimination and may help identify patients most likely to benefit from intensified follow-up or additional therapy.

Survival prediction represents another major focus of AI-driven CRC prognosis. Deep learning models applied to digitized haematoxylin-eosin-stained pathology slides have shown that prognostic information can be directly extracted from tumour histomorphology, achieving expert-level performance in outcome prediction (Kather et al., 2019). In addition, AI-based quantification of spatial immune features in whole-slide images has yielded clinically relevant prognostic biomarkers, with CD3CT emerging as a stage-independent predictor of overall survival in CRC (Cai

et al., 2024). Recent multimodal studies suggest that combining histopathology with additional clinical and imaging information may further improve prognostic stratification beyond single-modality pathology models (Shi et al., 2025). For example, PRISM-CRC integrated histopathology, radiology, endoscopy, and clinical data within a deep learning framework, achieving a concordance index of 0.82 for 5-year disease-free survival prediction and an AUC of 0.91 for microsatellite instability classification, while also showing that performance may decline under domain shift (Shi et al., 2025). In parallel, AI models using structured clinical data have modestly but consistently outperformed clinician-based prognostication in estimating long-term survival (Sailer et al., 2015). More biologically informed multimodal approaches have further linked radiological phenotypes to RNA sequencing profiles and histopathological patterns, revealing immune and stromal features associated with differential benefit from adjuvant chemotherapy in stage II CRC (Xie et al., 2025). Tumour microenvironment-related information has also been incorporated into multimodal frameworks to improve prediction of both survival and chemotherapy benefit in large CRC cohorts (Ji et al., 2025). Taken together, these studies suggest that multimodal prognostic frameworks may offer stronger predictive performance and finer risk stratification than single-modality models, although differences in robustness, interpretability, and implementation complexity remain important. Despite these advances, most studies remain retrospective, and challenges related to model interpretability, generalisability, and prospective validation continue to limit clinical translation (Kinar et al., 2016; Tai et al., 2018).

3.6 The combination of AI and multi-omics in CRC

In oncology, the rapid development of AI and big data analytics has enabled large-scale analysis of multi-omics datasets, thereby accelerating precision cancer research (Zafari et al., 2023). In CRC, AI-driven platforms have demonstrated the capacity to link molecular features with clinical phenotypes. For example, the MOMA platform can infer genetic alterations and predict prognosis directly from histopathological patterns (Tsai et al., 2023). Machine-learning models integrating genomics, transcriptomics, proteomics, and metabolomics data have generally shown better predictive discrimination than single-modality approaches in retrospective CRC studies, although the magnitude of improvement varies across cohorts, endpoints, and integration framework (Zhang S. L. et al., 2023; Tong et al., 2020; Yang et al., 2022). In addition, single-cell transcriptomic analyses have identified clinically relevant biomarkers (Zhang G. L. et al., 2019), while serum metabolomics and spatial proteomics have enabled robust patient stratification and biomarker discovery in CRC cohorts (Troisi et al., 2022; Levy et al., 2023).

Despite these advances, substantial challenges remain in applying AI to multi-omics data integration. The high dimensionality, heterogeneity, and scale of multi-omics datasets increase the risk of noise, overfitting, and reduced model interpretability if irrelevant features are not adequately controlled. Feature selection and dimensionality reduction techniques are therefore essential to enhance model performance.

Moreover, differences in data types, distributions, and sample sizes across omics layers complicate integration and often require data transformation or mapping before downstream analysis. Additional challenges include class imbalance and missing data, which can be addressed through sampling strategies, cost-sensitive learning, or imputation methods such as random forests and k-nearest neighbors (Picard et al., 2021).

To address these issues, three principal strategies for multi-omics integration have been proposed: early, intermediate, and late integration. Early integration concatenates features from multiple omics layers but is prone to the curse of dimensionality, whereas late integration analyzes each omics layer independently before combining results. Intermediate integration, by contrast, projects heterogeneous data into shared or complementary latent representations, enabling more effective capture of cross-omics interactions and biological complexity (Cai et al., 2022). Within this framework, AI and machine-learning approaches have been widely applied to overcome data heterogeneity, dimensionality, missing values, scalability, and class imbalance (Mirza et al., 2019). Importantly, AI-based integration of multi-omics data with non-omics information such as clinical variables, laboratory tests, imaging, pathology, and electronic health records offers a powerful means to bridge genotype and phenotype.

However, the translational value of these models depends not only on predictive performance, but also on data quality, computational infrastructure, and robustness beyond the development setting. These requirements are particularly difficult to meet in CRC multi-omics research, where matched datasets are often limited in size, incomplete across modalities, and vulnerable to batch effects or cohort-specific bias (Abadir et al., 2020; Auger et al., 2020). Variability in clinical presentation, treatment pathways, and endpoint definition may further reduce model stability when algorithms are applied outside the conditions under which they were trained (Auger et al., 2020; Chen et al., 2019). Interpretability also remains a major challenge. Classical machine-learning models based on structured clinical or omics variables may offer greater transparency, whereas deep learning and multimodal fusion models can capture more complex nonlinear interactions at the cost of reduced interpretability. In such settings, the “black-box” problem becomes particularly important, because apparently strong performance may obscure unrecognised confounding, technically driven artefacts, or biologically implausible associations (Auger et al., 2020; Narla et al., 2018). Collectively, these considerations suggest that AI-driven multi-omics integration is highly promising for precision oncology in CRC, but its clinical value will depend on reproducibility, biological plausibility, and robust external validation. To compare AI applications across the CRC clinical continuum more critically, Table 2 summarises the major data modalities, analytical methods, representative performance reported in the manuscript, and the main validation or translational limitations currently highlighted in the field.

4 Outlook and challenges

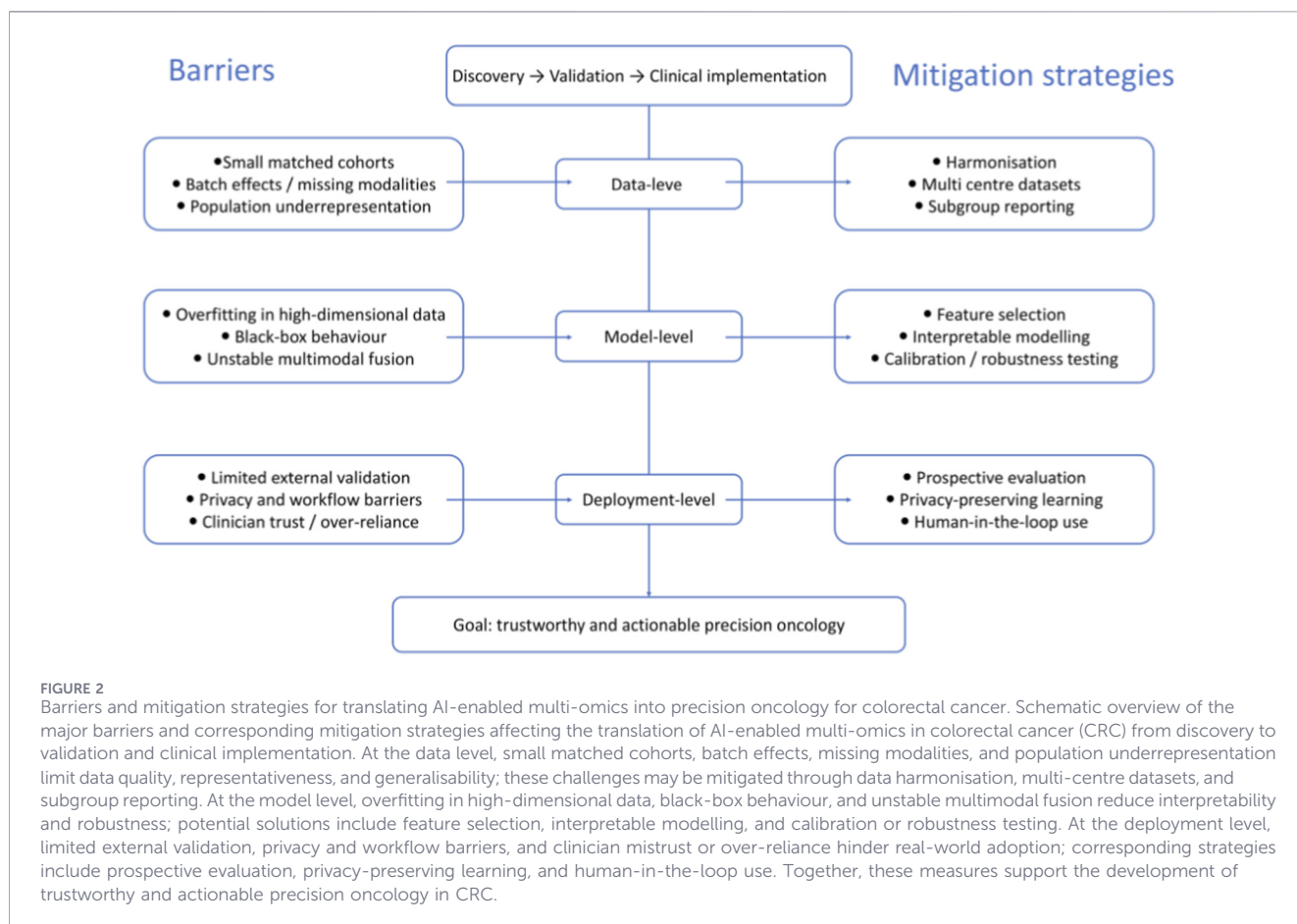
The integration of multi-omics technologies has substantially accelerated biomarker discovery in CRC by enabling systematic interrogation of molecular alterations across multiple biological

TABLE 2 AI applications across the CRC clinical continuum.

Clinical stage	Data modality	AI methodology	Representative application	Representative performance reported in the manuscript	Validation/key limitation described in the manuscript
Screening	Endoscopy images	CNN, deep learning	Real-time detection and classification of diminutive and non-polypoid colorectal polyps	High sensitivity and specificity for real-time identification; reduced missed diagnoses	Limited generalisability and insufficient long-term outcome data are described as ongoing challenges for AI screening applications
Screening	Blood/stool biomarkers	Machine learning (LR, RF, SVM)	Risk stratification using CBC indices, serum proteins, and stool markers	CBC-based models achieved AUCs above 0.80; blood-based protein and N-glycan biomarkers showed high sensitivity for advanced adenomas and CRC	Prospective validation not reported in the cited studies beyond the need for evidence-based integration into screening pathways
Screening	Plasma cfDNA	ML-based sequencing models	Early-stage CRC detection using cfDNA fragmentation and mutation patterns	AUC up to 0.92 for early-stage CRC detection	Prospective validation not reported in the cited studies
Diagnosis	Pathology whole-slide images	CNN, DL	Gland segmentation, tumour classification, MSI/dMMR status prediction	High accuracy and reproducibility across heterogeneous datasets; some ML classifiers exceeded 95%; MSI prediction showed strong discriminatory performance	Limited interpretability, lack of standardized public datasets, and insufficient robust prospective validation
Staging	CT/MRI imaging	Radiomics, DL	Lymph node metastasis prediction; tumour stage assessment	High accuracy for tumour and lymph node segmentation/metastasis prediction; reduced interpretation time compared with radiologists	Variability in image acquisition quality; robust prospective validation still lacking
Molecular profiling	Radiology images	ML-based radiogenomics	Non-invasive prediction of molecular features such as KRAS mutation status	Promising non-invasive molecular prediction described in current text	Prospective validation not reported in the cited studies
Treatment (surgery)	Surgical video streams	CNN, DL	Surgical phase recognition, anatomical structure identification	High accuracy for automatic recognition of surgical phases and actions; real-time workflow analysis	AI treatment applications are described as early stage, limited by small datasets and lack of prospective validation
Prognosis	Pathology + clinical data	ML, DL survival models	Recurrence risk stratification and survival prediction	Deep learning on H&E slides achieved expert-level performance; structured clinical-data models modestly but consistently outperformed clinician-based prognostication	Most studies remain retrospective; interpretability, generalisability, and prospective validation remain limiting factors
Integration	Multi-omics + clinical data	ML, DNN	Precision stratification and personalised clinical decision support	Multi-omics models improved accuracy for CRC outcomes and survival compared with single-modality approaches	High dimensionality, heterogeneity, missing data, and class imbalance complicate integration

layers. However, elucidating the mechanisms underlying biomarker generation and establishing the diagnostic and therapeutic relevance of candidate biomarkers remain challenging and are both critical for drug development. Integrative strategies that combine genomics, proteomics, and metabolomics have shown promise in improving tumour characterisation and supporting anticancer drug development, particularly with the continued advancement of high-throughput sequencing technologies (Eylem et al., 2020; Nam et al., 2021; Correa-Aguila et al., 2022).

Despite this potential, several barriers continue to limit the clinical translation of multi-omics discoveries. Omics approaches remain costly and technically demanding, while data quality is often compromised by non-standardised sampling, heterogeneous experimental protocols, and limited cohort sizes. In particular, metabolomics faces substantial analytical challenges because of the complexity and dynamic variability of metabolite profiles. Moreover, the absence of standardised evaluation frameworks and large-scale validation cohorts hampers reliable assessment of



biomarker sensitivity and specificity. Tumour heterogeneity further complicates biomarker validation, highlighting the need for higher-resolution approaches, including single-cell and spatially resolved analyses.

The convergence of AI and multi-omics has reshaped cancer research, yet reliable translation still requires rigorous data standardisation, harmonisation, and quality control. Differences in data types, scales, and preprocessing pipelines across omics platforms may introduce bias or information loss during integration. Advanced computational strategies are therefore needed to handle dimensionality, missingness, and cross-platform heterogeneity, including feature selection, data augmentation, and biologically informed modelling (Amann et al., 2020; Grinsztajn et al., 2022; Dash et al., 2022). Improved data accessibility and sharing are also essential, as limited access to high-quality multi-omics datasets continues to constrain progress, despite established resources such as TCGA and GTEx.

Interpretability, dataset bias, and fairness represent additional and closely linked challenges. Many high-performing deep learning systems lack transparency, limiting clinical trust and impeding routine adoption (Amann et al., 2020). Although approaches such as tree-based models, SHAP-based explanations, and domain knowledge-guided modelling may improve interpretability (Grinsztajn et al., 2022; Dash et al., 2022), their practical value depends on whether such explanations support reliable human judgement rather than superficial reassurance. Recent work has increasingly emphasised both explainability and causability, arguing

that AI systems should not only generate accurate outputs but also provide explanations that enhance human understanding and support actionable clinical decisions (Holzinger et al., 2019). This issue is particularly relevant in complex diagnostic contexts, such as histopathology, where clinicians require interpretable and biologically coherent evidence to justify clinical action.

Bias in multi-omics datasets may further impair model transportability and ethical acceptability. Differences in sequencing platforms, proteomic workflows, image acquisition quality, pathology staining, preprocessing pipelines, and cohort composition may cause models to learn centre-specific or platform-specific signals rather than disease biology. Incomplete multimodal datasets may also introduce selection bias when analyses are restricted to patients with complete data, while imputation strategies may propagate hidden uncertainty if not carefully validated (Picard et al., 2021). Underrepresentation of non-European populations in current datasets raises additional concern that AI systems may perform unevenly across demographic groups, thereby exacerbating existing inequities in precision oncology (Carter et al., 2020; Murdoch, 2021; Peterson et al., 2019; Martin et al., 2019). These limitations are not merely technical; they directly affect clinical reliability, fairness, and the safe deployment of AI across heterogeneous real-world populations.

Rigorous validation remains essential before widespread implementation. Most current AI studies in CRC, including many models for prognosis and treatment prediction, remain retrospective and show substantial variability in reported

sensitivity, specificity, and overall accuracy (Kinar et al., 2016; Huang et al., 2020; Tai et al., 2018; Wang et al., 2019). Broader deployment will therefore require validation beyond the originating centres and, where feasible, prospective or randomised evaluation to determine whether AI-assisted decision-making confers genuine clinical benefit (Abadir et al., 2020; Plana et al., 2022). In parallel, privacy protection must remain a central priority, particularly as academic, clinical, and commercial data ecosystems become increasingly interconnected (Tom et al., 2020). Privacy-preserving frameworks, including federated learning, swarm learning, and other distributed strategies, may support responsible data sharing and model development (Warnat-Herresthal et al., 2021; Cao et al., 2022), but such approaches must still be accompanied by transparent reporting, subgroup performance assessment, and careful regulatory oversight. Finally, trust should be regarded as an outcome of evidence rather than an assumption. Although explainable AI is often proposed as a solution to the black-box problem, poorly calibrated explanations may paradoxically increase blind reliance on algorithmic outputs and reduce critical clinical appraisal (Jacobs et al., 2021; Cabitza et al., 2023). Thus, future progress in AI-driven multi-omics for CRC will depend not only on more powerful algorithms, but also on prospective validation, transparent reporting, biologically plausible interpretation, equitable dataset design, and clinically feasible implementation pathways (Reis-Filho and Kather, 2023). The major challenges and potential solutions are summarised in Figure 2.

5 Conclusion

CRC management increasingly depends on integrative molecular stratification, and multi-omics approaches have expanded the range of clinically relevant biomarkers and biological insights. However, the clinical translation of these data remains limited by heterogeneity, high dimensionality, and incomplete interpretability.

AI provides a powerful analytical framework to integrate multi-omics data with imaging and clinical information, facilitating early screening, accurate diagnosis, individualized treatment, and refined prognostic assessment. Emerging evidence suggests that AI-driven multi-omics integration may outperform single-modality approaches in selected CRC applications, particularly in risk stratification and prognosis modelling, although current evidence remains heterogeneous and is still dominated by retrospective studies.

Despite these advances, significant challenges remain, including data standardisation, model generalisability, interpretability, and

ethical considerations related to privacy and fairness. Addressing these issues will require multidisciplinary collaboration and robust clinical validation. Ultimately, the convergence of AI and multi-omics is poised to transform CRC management by bridging molecular complexity with actionable clinical decision-making.

Author contributions

HS: Writing – original draft, Writing – review and editing. TW: Conceptualization, Writing – review and editing. CC: Writing – review and editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abadir, A. P., Ali, M. F., Karnes, W., and Samarasekera, J. B. (2020). Artificial intelligence in gastrointestinal endoscopy. *Clin. Endosc.* 53 (2), 132–141. doi:10.5946/ce.2020.038
- Acs, B., Rantalainen, M., and Hartman, J. (2020). Artificial intelligence as the next step towards precision pathology. *J. Intern. Med.* 288 (1), 62–81. doi:10.1111/ijom.13030
- Ahmad, R., Singh, J. K., Wunnavu, A., Al-Obeid, O., Abdulla, M., and Srivastava, S. K. (2021). Emerging trends in colorectal cancer: dysregulated signaling pathways (Review). *Int. J. Mol. Med.* 47 (3), 14. doi:10.3892/ijmm.2021.4847
- Ahn, S. B., Sharma, S., Mohamedali, A., Mahboob, S., Redmond, W. J., Pascovici, D., et al. (2019). Potential early clinical stage colorectal cancer diagnosis using a proteomics blood test panel. *Clin. Proteomics* 16, 34. doi:10.1186/s12014-019-9255-z
- Alhur, A. (2024). Redefining healthcare with Artificial Intelligence (AI): the contributions of ChatGPT, Gemini, and co-pilot. *Cureus* 16 (4), e57795. doi:10.7759/cureus.57795
- Alizadehsani, R., Khosravi, A., Roshanzamir, M., Abdar, M., Sarrafzadegan, N., Shafie, D., et al. (2021). Coronary artery disease detection using artificial intelligence techniques: a survey of trends, geographical differences and diagnostic features 1991–2020. *Comput. Biol. Med.* 128, 104095. doi:10.1016/j.compbiomed.2020.104095
- Alvarez-Chavez, P., Otero-Esteviz, O., Paez de la Cadena, M., Rodriguez-Berrocal, F. J., and Martinez-Zorano, V. S. (2014). Proteomics for discovery of candidate colorectal cancer biomarkers. *World J. Gastroenterol.* 20 (14), 3804–3824. doi:10.3748/wjg.v20.i14.3804

- Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., and Precise4Q consortium (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med. Inf. Decis. Mak.* 20 (1), 310. doi:10.1186/s12911-020-01332-6
- Andrei, P., Battuello, P., Grasso, G., Rovera, E., Tesio, N., and Bardelli, A. (2022). Integrated approaches for precision oncology in colorectal cancer: the more you know, the better. *Semin. Cancer Biol.* 84, 199–213. doi:10.1016/j.semcancer.2021.04.007
- Auger, S. D., Jacobs, B. M., Dobson, R., Marshall, C. R., and Noyce, A. J. (2020). Big data, machine learning and artificial intelligence: a neurologist's guide. *Pract. Neurol.* 21 (1), 4–11. doi:10.1136/practneurol-2020-002688
- Azizi, S., Culp, L., Freyberg, J., Mustafa, B., Baur, S., Kornblith, S., et al. (2023). Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. *Nat. Biomed. Eng.* 7 (6), 756–779. doi:10.1038/s41551-023-01049-7
- Badic, B., Tixier, F., Cheze Le Rest, C., Hatt, M., and Visvikis, D. (2021). Radiogenomics in colorectal cancer. *Cancers (Basel)* 13 (5), 973. doi:10.3390/cancers13050973
- Belkhir, A., and El-Rifai, W. (2014). 5-Methylcytosine hydroxylation-mediated LINE-1 hypomethylation: a novel mechanism of proto-oncogenes activation in colorectal cancer? *Gut* 63 (4), 538–539. doi:10.1136/gutjnl-2013-305176
- Bengio, Y. J. F. (2009). Learning deep architectures for AI. 2 (1), 1–127.
- Berdasco, M., and Esteller, M. (2019). Clinical epigenetics: seizing opportunities for translation. *Nat. Rev. Genet.* 20 (2), 109–127. doi:10.1038/s41576-018-0074-2
- Bhardwaj, M., Gies, A., Weigl, K., Tikik, K., Benner, A., Schrotz-King, P., et al. (2019). Evaluation and validation of plasma proteins using two different protein detection methods for early detection of colorectal cancer. *Cancers (Basel)* 11 (10), 1426. doi:10.3390/cancers11101426
- Bilal, M., Raza, S. E. A., Azam, A., Graham, S., Ilyas, M., Cree, I. A., et al. (2021). Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. *Lancet Digit. Health* 3 (12), e763–e772. doi:10.1016/S2589-7500(21)00180-1
- Bilal, M., Gulzar, M. A., Jaffar, N., Alabuljabbar, A., Altherwy, Y., Alsuhaibani, A., et al. (2026). Benchmarking pathology foundation models for predicting microsatellite instability in colorectal cancer histopathology. *Comput. Med. Imaging Graph* 127, 102680. doi:10.1016/j.compmedimag.2025.102680
- Billir, L. H., and Schrag, D. (2021). Diagnosis and treatment of metastatic colorectal cancer: a review. *JAMA* 325 (7), 669–685. doi:10.1001/jama.2021.0106
- Binder, A., Bockmayr, M., Hägele, M., Wienert, S., Heim, D., Hellweg, K., et al. (2021). Morphological and molecular breast cancer profiling through explainable machine learning. *Nat. Mach. Intell.* 3 (4), 355–366. doi:10.1038/s42256-021-00303-4
- Bondeven, P., Laurberg, S., Hagemann-Madsen, R. H., and Ginnerup Pedersen, B. (2015). Suboptimal surgery and omission of neoadjuvant therapy for upper rectal cancer is associated with a high risk of local recurrence. *Colorectal Dis.* 17 (3), 216–224. doi:10.1111/codi.12869
- Boothby, M., and Rickert, R. C. (2017). Metabolic regulation of the immune humoral response. *Immunity* 46 (5), 743–755. doi:10.1016/j.immuni.2017.04.009
- Brown, D. G., Rao, S., Weir, T. L., O'Malia, J., Bazan, M., Brown, R. J., et al. (2016). Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. *Cancer Metab.* 4, 11. doi:10.1186/s40170-016-0151-y
- Buikhuisen, J. Y., Torang, A., and Medema, J. P. (2020). Exploring and modelling colon cancer inter-tumour heterogeneity: opportunities and challenges. *Oncogenesis* 9 (7), 66. doi:10.1038/s41389-020-00250-6
- Burla, B., Arita, M., Arita, M., Bendt, A. K., Cazenave-Gassiot, A., Dennis, E. A., et al. (2018). MS-based lipidomics of human blood plasma: a community-initiated position paper to develop accepted guidelines. *J. Lipid Res.* 59 (10), 2001–2017. doi:10.1194/jlr.S087163
- Cabitza, F., Campagner, A., Ronzio, L., Cameli, M., Mandoli, G. E., Pastore, M. C., et al. (2023). Rams, hounds and white boxes: investigating human-AI collaboration protocols in medical diagnosis. *Artif. Intell. Med.* 138, 102506. doi:10.1016/j.artmed.2023.102506
- Cai, Z., Poulos, R. C., Liu, J., and Zhong, Q. (2022). Machine learning for multi-omics data integration in cancer. *iScience* 25 (2), 103798. doi:10.1016/j.isci.2022.103798
- Cai, M., Zhao, K., Wu, L., Huang, Y., Zhao, M., Hu, Q., et al. (2024). Artificial intelligence-based analysis of tumor-infiltrating lymphocyte spatial distribution for colorectal cancer prognosis. *Chin. Med. J. Engl.* 137 (4), 421–430. doi:10.1097/CM9.0000000000002964
- Cao, H., Zhang, Y., Baumbach, J., Burton, P. R., Dwyer, D., Koutsouleris, N., et al. (2022). dsMTL: a computational framework for privacy-preserving, distributed multi-task machine learning. *Bioinformatics* 38 (21), 4919–4926. doi:10.1093/bioinformatics/btac616
- Carter, S. M., Rogers, W., Win, K. T., Frazer, H., Richards, B., and Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *Breast* 49, 25–32. doi:10.1016/j.breast.2019.10.001
- Chantaraamporn, J., Champattanachai, V., Khongmanee, A., Verathamjamras, C., Prasongsook, N., Mingkwan, K., et al. (2020). Glycoproteomic analysis reveals aberrant expression of complement C9 and fibronectin in the plasma of patients with colorectal cancer. *Proteomes* 8 (3), 26. doi:10.3390/proteomes8030026
- Chen, D., Liu, S., Kingsbury, P., Sohn, S., Storlie, C. B., Habermann, E. B., et al. (2019). Deep learning and alternative learning strategies for retrospective real-world clinical data. *NPJ Digit. Med.* 2, 43. doi:10.1038/s41746-019-0122-0
- Chen, F., Dai, X., Zhou, C. C., Li, K. X., Zhang, Y. J., Lou, X. Y., et al. (2022). Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma. *Gut* 71 (7), 1315–1325. doi:10.1136/gutjnl-2020-323476
- Chen, R. J., Ding, T., Lu, M. Y., Williamson, D. F. K., Jaume, G., Song, A. H., et al. (2024). Towards a general-purpose foundation model for computational pathology. *Nat. Med.* 30 (3), 850–862. doi:10.1038/s41591-024-02857-3
- Claudio Quiros, A., Coudray, N., Yeaton, A., Yang, X., Liu, B., Le, H., et al. (2024). Mapping the landscape of histomorphological cancer phenotypes using self-supervised learning on unannotated pathology slides. *Nat. Commun.* 15 (1), 4596. doi:10.1038/s41467-024-48666-7
- Clos-Garcia, M., Garcia, K., Alonso, C., Iruarizaga-Lejarreta, M., D'Amato, M., Crespo, A., et al. (2020). Integrative analysis of fecal metagenomics and metabolomics in colorectal cancer. *Cancers (Basel)* 12 (5), 1142. doi:10.3390/cancers12051142
- Commisso, M., Strazzer, P., Toffali, K., Stocchero, M., and Guzzo, F. (2013). Untargeted metabolomics: an emerging approach to determine the composition of herbal products. *Comput. Struct. Biotechnol. J.* 4, e201301007. doi:10.5936/csbj.201301007
- Correa-Aguila, R., Alonso-Pupo, N., and Hernandez-Rodriguez, E. W. (2022). Multi-omics data integration approaches for precision oncology. *Mol. Omics* 18 (6), 469–479. doi:10.1039/d1mo00411e
- Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyo, D., et al. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* 24 (10), 1559–1567. doi:10.1038/s41591-018-0177-5
- Dankner, M., Rose, A. A. N., Rajkumar, S., Siegel, P. M., and Watson, I. R. (2018). Classifying BRAF alterations in cancer: new rational therapeutic strategies for actionable mutations. *Oncogene* 37 (24), 3183–3199. doi:10.1038/s41388-018-0171-x
- Dash, T., Chitlangia, S., Ahuja, A., and Srinivasan, A. (2022). A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Sci. Rep.* 12 (1), 1040. doi:10.1038/s41598-021-04590-0
- De'Angelis, G. L., Bottarelli, L., Azzoni, C., De'Angelis, N., Leandro, G., Di Mario, F., et al. (2018). Microsatellite instability in colorectal cancer. *Acta Biomed.* 89 (9–S), 97–101. doi:10.23750/abm.v89i9-S.7960
- Dehairs, J., Derua, R., Rueda-Rincon, N., and Swinnen, J. V. (2015). Lipidomics in drug development. *Drug Discov. Today Technol.* 13, 33–38. doi:10.1016/j.ddtec.2015.03.002
- Ding, H., Pan, Z., Cen, Q., Li, Y., and Chen, S. J. N. (2020). Multi-scale fully convolutional network for gland segmentation using three-class classification. *Neurocomputing* 380, 150–161. doi:10.1016/j.neucom.2019.10.097
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485 (7397), 201–206. doi:10.1038/nature11112
- Duan, S., Huang, W., Liu, X., Liu, X., Chen, N., Xu, Q., et al. (2018). IMPDH2 promotes colorectal cancer progression through activation of the PI3K/AKT/mTOR and PI3K/AKT/FOXO1 signaling pathways. *J. Exp. Clin. Cancer Res.* 37 (1), 304. doi:10.1186/s13046-018-0980-3
- Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R., and Griffin, J. L. (2011). Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* 40 (1), 387–426. doi:10.1039/b906712b
- Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., and Kather, J. N. (2021). Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br. J. Cancer* 124 (4), 686–696. doi:10.1038/s41416-020-01122-x
- Ecker, J., Benedetti, E., Kindt, A. S. D., Horing, M., Perl, M., Machmuller, A. C., et al. (2021). The colorectal cancer lipidome: identification of a robust tumor-specific lipid species signature. *Gastroenterology* 161 (3), 910–923.e19. doi:10.1053/j.gastro.2021.05.009
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., et al. (2019). A guide to deep learning in healthcare. *Nat. Med.* 25 (1), 24–29. doi:10.1038/s41591-018-0316-z
- Eylem, C. C., Yilmaz, M., Derkus, B., Nemutlu, E., Camci, C. B., Yilmaz, E., et al. (2020). Untargeted multi-omic analysis of colorectal cancer-specific exosomes reveals joint pathways of colorectal cancer in both clinical samples and cell culture. *Cancer Lett.* 469, 186–194. doi:10.1016/j.canlet.2019.10.038
- Fedus, W., Zoph, B., and Shazeer, N. (2022). Switch transformers: scaling to trillion parameter models with simple and efficient sparsity. *J. Machine Learn. Res.* 23 (120), 1–39.
- Fu, Y., Jung, A. W., Torne, R. V., Gonzalez, S., Vohringer, H., Shmatko, A., et al. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* 1 (8), 800–810. doi:10.1038/s43018-020-0085-8
- Galanopoulos, M., Tsoukalas, N., Papanikolaou, I. S., Tolia, M., Gazouli, M., and Mantzaris, G. J. (2017). Abnormal DNA methylation as a cell-free circulating DNA

- biomarker for colorectal cancer detection: a review of literature. *World J. Gastrointest. Oncol.* 9 (4), 142–152. doi:10.4251/wjgo.v9.i4.142
- Gao, J., Wu, Z., Zhao, M., Zhang, R., Li, M., Sun, D., et al. (2022). Allosteric inhibition reveals SHP2-mediated tumor immunosuppression in colon cancer by single-cell transcriptomics. *Acta Pharm. Sin. B* 12 (1), 149–166. doi:10.1016/j.apsb.2021.08.006
- Ge, P., Wang, W., Li, L., Zhang, G., Gao, Z., Tang, Z., et al. (2019). Profiles of immune cell infiltration and immune-related genes in the tumor microenvironment of colorectal cancer. *Biomed. Pharmacother.* 118, 109228. doi:10.1016/j.biopha.2019.109228
- Ghate, N. B., Kim, S., Spiller, E., Kim, S., Shin, Y., Rhie, S. K., et al. (2021). VprBP directs epigenetic gene silencing through histone H2A phosphorylation in colon cancer. *Mol. Oncol.* 15 (10), 2801–2817. doi:10.1002/1878-0261.13068
- Ghazanfar, S., Fatima, I., Aslam, M., Musharraf, S. G., Sherman, N. E., Moskaluk, C., et al. (2017). Identification of actin beta-like 2 (ACTBL2) as novel, upregulated protein in colorectal cancer. *J. Proteomics* 152, 33–40. doi:10.1016/j.jpro.2016.10.011
- Giammanco, A., Bychkov, A., Schallenberg, S., Tsvetkov, T., Fukuoka, J., Prylukhin, A., et al. (2024). Fast-track development and multi-institutional clinical validation of an artificial intelligence algorithm for detection of lymph node metastasis in colorectal cancer. *Mod. Pathol.* 37 (6), 100496. doi:10.1016/j.modpat.2024.100496
- González-Castro, V., Cernadas, E., Huelga, E., Fernández-Delgado, M., Porto, J., Antunez, J. R., et al. (2020). CT radiomics in colorectal cancer: detection of KRAS mutation using texture analysis and machine learning. *Appl. Sci.* 10 (18), 6214. doi:10.3390/app10186214
- Grinsztajn, L., Oyallon, E., and Varoquaux, G. (2022). “Why do tree-based models still outperform deep learning on typical tabular data?,” in *advances in neural information processing systems 35: 36th conference on neural information processing systems (NeurIPS 2022)*. Editors S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh. (Red Hook, NY: Curran Associates, Inc), 507–520.
- Gu, J., Xiao, Y., Shu, D., Liang, X., Hu, X., Xie, Y., et al. (2019). Metabolomics analysis in serum from patients with colorectal polyp and colorectal cancer by (1)H-NMR spectrometry. *Dis. Markers* 2019, 3491852. doi:10.1155/2019/3491852
- Guinney, J., Dienstmann, R., Wang, X., de Reynies, A., Schlicker, A., Soneson, C., et al. (2015). The consensus molecular subtypes of colorectal cancer. *Nat. Med.* 21 (11), 1350–1356. doi:10.1038/nm.3967
- Guo, S., Li, L., Xu, B., Li, M., Zeng, Q., Xiao, H., et al. (2018). A simple and novel fecal biomarker for colorectal cancer: ratio of *Fusobacterium nucleatum* to probiotics populations, based on their antagonistic effect. *Clin. Chem.* 64 (9), 1327–1337. doi:10.1373/clinchem.2018.289728
- Guo, J., Cao, W., Nie, B., and Qin, Q. (2023). Unsupervised learning composite network to reduce training cost of deep learning model for colorectal cancer diagnosis. *IEEE J. Transl. Eng. Health Med.* 11, 54–59. doi:10.1109/JTEHM.2022.3224021
- Habermann, J. K., Bader, F. G., Franke, C., Zimmermann, K., Gemoll, T., Fritzsche, B., et al. (2008). From the genome to the proteome-biomarkers in colorectal cancer. *Langenbecks Arch. Surg.* 393 (1), 93–104. doi:10.1007/s00423-007-0230-1
- Hamet, P., and Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism* 69S, S36–S40. doi:10.1016/j.metabol.2017.01.011
- Han, S. W., Kim, H. P., Shin, J. Y., Jeong, E. G., Lee, W. C., Lee, K. H., et al. (2013). Targeted sequencing of cancer-related genes in colorectal cancer using next-generation sequencing. *PLoS One* 8 (5), e64271. doi:10.1371/journal.pone.0064271
- Hao, J. J., Zhi, X., Wang, Y., Zhang, Z., Hao, Z., Ye, R., et al. (2017). Comprehensive proteomic characterization of the human colorectal carcinoma reveals signature proteins and perturbed pathways. *Sci. Rep.* 7, 42436. doi:10.1038/srep42436
- Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biol.* 18 (1), 83. doi:10.1186/s13059-017-1215-1
- Hilsden, R. J., Heitman, S. J., Mizrahi, B., Narod, S. A., and Goshen, R. (2018). Prediction of findings at screening colonoscopy using a machine learning algorithm based on complete blood counts (ColonFlag). *PLoS One* 13 (11), e0207848. doi:10.1371/journal.pone.0207848
- Holmes, E., Li, J. V., Marchesi, J. R., and Nicholson, J. K. (2012). Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab.* 16 (5), 559–564. doi:10.1016/j.cmet.2012.10.007
- Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Muller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 9 (4), e1312. doi:10.1002/widm.1312
- Hornbrook, M. C., Goshen, R., Choman, E., O’Keeffe-Rosetti, M., Kinar, Y., Liles, E. G., et al. (2017). Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. *Dig. Dis. Sci.* 62 (10), 2719–2727. doi:10.1007/s10620-017-4722-8
- Hu, H., Zhu, J., Zhong, Y., Geng, R., Ji, Y., Guan, Q., et al. (2021). PIK3CA mutation confers resistance to chemotherapy in triple-negative breast cancer by inhibiting apoptosis and activating the PI3K/AKT/mTOR signaling pathway. *Ann. Transl. Med.* 9 (5), 410. doi:10.21037/atm-21-698
- Huang, S., Yang, J., Fong, S., and Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Cancer Lett.* 471, 61–71. doi:10.1016/j.canlet.2019.12.007
- Huang, M. S., Fu, L. H., Yan, H. C., Cheng, L. Y., Ru, H. M., Mo, S., et al. (2022). Proteomics and liquid biopsy characterization of human EMT-related metastasis in colorectal cancer. *Front. Oncol.* 12, 790096. doi:10.3389/fonc.2022.790096
- Hur, K., Cejas, P., Feliu, J., Moreno-Rubio, J., Burgos, E., Boland, C. R., et al. (2014). Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. *Gut* 63 (4), 635–646. doi:10.1136/gutjnl-2012-304219
- Ichimura, N., Shinjo, K., An, B., Shimizu, Y., Yamao, K., Ohka, F., et al. (2015). Aberrant TET1 methylation closely associated with CpG island methylator phenotype in colorectal cancer. *Cancer Prev. Res. (Phila)* 8 (8), 702–711. doi:10.1158/1940-6207.CAPR-14-0306
- Igaki, T., Kitaguchi, D., Kojima, S., Hasegawa, H., Takeshita, N., Mori, K., et al. (2022). Artificial intelligence-based total mesorectal excision plane navigation in laparoscopic colorectal surgery. *Dis. Colon Rectum* 65 (5), e329–e333. doi:10.1097/DCR.0000000000002393
- Issa, I. A., and Nouredine, M. (2017). Colorectal cancer screening: an updated review of the available options. *World J. Gastroenterol.* 23 (28), 5086–5096. doi:10.3748/wjg.v23.i28.5086
- Ivancic, M. M., Megna, B. W., Sverchkov, Y., Craven, M., Reichelderfer, M., Pickhardt, P. J., et al. (2020). Noninvasive detection of colorectal carcinomas using serum protein biomarkers. *J. Surg. Res.* 246, 160–169. doi:10.1016/j.jss.2019.08.004
- Ivanisevic, J., and Want, E. J. (2019). From samples to insights into metabolism: uncovering biologically relevant information in LC-HRMS metabolomics data. *Metabolites* 9 (12), 308. doi:10.3390/metabo9120308
- Jacobs, M., Pradier, M. F., McCoy, T. H., Jr., Perlis, R. H., Doshi-Velez, F., and Gajos, K. Z. (2021). How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection. *Transl. Psychiatry* 11 (1), 108. doi:10.1038/s41398-021-01224-x
- Jiang, W., Yang, K., Xiao, C., Ji, H., Yan, B., Zhao, S., et al. (2025). Multimodal tumor microenvironment signature of colorectal cancer for prediction prognosis and chemotherapy benefit. *NPJ Precis. Oncol.* 9 (1), 270. doi:10.1038/s41698-025-01069-3
- Jiang, W., Mei, W. J., Xu, S. Y., Ling, Y. H., Li, W. R., Kuang, J. B., et al. (2022). Clinical actionability of triaging DNA mismatch repair deficient colorectal cancer from biopsy samples using deep learning. *EBioMedicine* 81, 104120. doi:10.1016/j.ebiom.2022.104120
- Johnson, M. O., Siska, P. J., Contreras, D. C., and Rathmell, J. C. (2016). Nutrients and the microenvironment to feed a T cell army. *Semin. Immunol.* 28 (5), 505–513. doi:10.1016/j.smim.2016.09.003
- Jung, G., Hernandez-Illan, E., Moreira, L., Balaguer, F., and Goel, A. (2020). Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat. Rev. Gastroenterol. Hepatol.* 17 (2), 111–130. doi:10.1038/s41575-019-0230-y
- Kann, B. H., Hosny, A., and Aerts, H. (2021). Artificial intelligence for clinical oncology. *Cancer Cell* 39 (7), 916–927. doi:10.1016/j.ccell.2021.04.002
- Karczewski, K. J., and Snyder, M. P. (2018). Integrative omics for health and disease. *Nat. Rev. Genet.* 19 (5), 299–310. doi:10.1038/nrg.2018.4
- Kather, J. N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C. A., et al. (2019). Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med.* 16 (1), e1002730. doi:10.1371/journal.pmed.1002730
- Kather, J. N., Heij, L. R., Grabsch, H. I., Loeffler, C., Ehle, A., Muti, H. S., et al. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* 1 (8), 789–799. doi:10.1038/s43018-020-0087-6
- Keyl, J., Keyl, P., Montavon, G., Hosch, R., Brehmer, A., Mochmann, L., et al. (2025). Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. *Nat. Cancer* 6 (2), 307–322. doi:10.1038/s43018-024-00891-1
- Khalili, K., Lawlor, R. L., Pourafkari, M., Lu, H., Tyrrell, P., Kim, T. K., et al. (2020). Convolutional neural networks versus radiologists in characterization of small hypoattenuating hepatic nodules on CT: a critical diagnostic challenge in staging of colorectal carcinoma. *Sci. Rep.* 10 (1), 15248. doi:10.1038/s41598-020-71364-5
- Kim, E. R., Kwon, H. N., Nam, H., Kim, J. J., Park, S., and Kim, Y. H. (2019). Urine-NMR metabolomics for screening of advanced colorectal adenoma and early stage colorectal cancer. *Sci. Rep.* 9 (1), 4786. doi:10.1038/s41598-019-41216-y
- Kinar, Y., Kalkstein, N., Akiva, P., Levin, B., Half, E. E., Goldshtein, I., et al. (2016). Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. *J. Am. Med. Assoc.* 315 (5), 879–890. doi:10.1093/jama/ocv195
- Kitaguchi, D., Takeshita, N., Matsuzaki, H., Takano, H., Owada, Y., Enomoto, T., et al. (2020). Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg. Endosc.* 34 (11), 4924–4931. doi:10.1007/s00464-019-07281-0
- Kitaguchi, D., Takeshita, N., Matsuzaki, H., Igaki, T., Hasegawa, H., and Ito, M. (2021). Development and validation of a 3-dimensional convolutional neural network for automatic surgical skill assessment based on spatiotemporal video analysis. *JAMA Netw. Open* 4 (8), e2120786. doi:10.1001/jamanetworkopen.2021.20786

- Kitaguchi, D., Takeshita, N., Matsuzaki, H., Hasegawa, H., Igaki, T., Oda, T., et al. (2022). Deep learning-based automatic surgical step recognition in intraoperative videos for transanal total mesorectal excision. *Surg. Endosc.* 36 (2), 1143–1151. doi:10.1007/s00464-021-08381-6
- Kolbinger, F. R., Bodenstedt, S., Carstens, M., Leger, S., Krell, S., Rinner, F. M., et al. (2024). Artificial intelligence for context-aware surgical guidance in complex robot-assisted oncological procedures: an exploratory feasibility study. *Eur. J. Surg. Oncol.* 50 (12), 106996. doi:10.1016/j.ejso.2023.106996
- Kondo, Y., Nishiumi, S., Shinohara, M., Hatano, N., Ikeda, A., Yoshie, T., et al. (2011). Serum fatty acid profiling of colorectal cancer by gas chromatography/mass spectrometry. *Biomark. Med.* 5 (4), 451–460. doi:10.2217/bmm.11.41
- Koteluk, O., Wartecki, A., Mazurek, S., Kolodziejczak, I., and Mackiewicz, A. (2021). How do machines learn? Artificial intelligence as a new era in medicine. *J. Pers. Med.* 11 (1), 32. doi:10.3390/jpm11010032
- Krishnamurthy, N., Spencer, E., Torkamani, A., and Nicholson, L. (2017). Liquid biopsies for cancer: coming to a patient near you. *J. Clin. Med.* 6 (1), 3. doi:10.3390/jcm6010003
- Kwong, T. N. Y., Wang, X., Nakatsu, G., Chow, T. C., Tipoe, T., Dai, R. Z. W., et al. (2018). Association between bacteremia from specific microbes and subsequent diagnosis of colorectal cancer. *Gastroenterology* 155 (2), 383–390.e8. doi:10.1053/j.gastro.2018.04.028
- Kyle, J. E. (2020). Extracting biological insight from untargeted lipidomics data. *Methods Mol. Biol.* 2104, 121–137. doi:10.1007/978-1-0716-0239-3_7
- Langley, P. (2011). The changing science of machine learning. *Mach. Learn.* 82 (3), 275–279. doi:10.1007/s10994-011-5242-y
- Lapiente-Santana, O., van Genderen, M., Hilbers, P. A. J., Finotello, F., and Eduati, F. (2021). Interpretable systems biomarkers predict response to immune-checkpoint inhibitors. *Patterns (N Y)* 2 (8), 100293. doi:10.1016/j.patter.2021.100293
- Levy, J. J., Zavras, J. P., Veziroglu, E. M., Nasir-Moin, M., Kolling, F. W., Christensen, B. C., et al. (2023). Identification of spatial proteomic signatures of Colon tumor metastasis: a digital spatial profiling approach. *Am. J. Pathol.* 193 (6), 778–795. doi:10.1016/j.ajpath.2023.02.020
- Li, H., Boimel, P., Janopaul-Naylor, J., Zhong, H., Xiao, Y., Ben-Josef, E., et al. (2019). Deep convolutional neural networks for imaging data based survival analysis of rectal cancer. *Proc. IEEE Int. Symp. Biomed. Imaging.* 2019, 846–849. doi:10.1109/ISBI.2019.8759301
- Li, J., Zhang, A. H., Wu, F. F., and Wang, X. J. (2022). Alterations in the gut microbiota and their metabolites in colorectal cancer: recent progress and future prospects. *Front. Oncol.* 12, 841552. doi:10.3389/fonc.2022.841552
- Li, S., Yuan, L., Xu, Z. Y., Xu, J. L., Chen, G. P., Guan, X., et al. (2023a). Integrative proteomic characterization of adenocarcinoma of esophagogastric junction. *Nat. Commun.* 14 (1), 778. doi:10.1038/s41467-023-36462-8
- Li, Y., Wang, B., Ma, F., Jiang, D., Wang, Y., Li, K., et al. (2023b). Proteomic characterization of the colorectal cancer response to chemoradiation and targeted therapies reveals potential therapeutic strategies. *Cell Rep. Med.* 4 (12), 101311. doi:10.1016/j.xcrm.2023.101311
- Liang, C., Huang, Y., He, L., Chen, X., Ma, Z., Dong, D., et al. (2016). The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. *Oncotarget* 7 (21), 31401–31412. doi:10.18632/oncotarget.8919
- Lin, Y., Ma, C., Bezabeh, T., Wang, Z., Liang, J., Huang, Y., et al. (2019). (1) H NMR-based metabolomics reveal overlapping discriminatory metabolites and metabolic pathway disturbances between colorectal tumor tissues and fecal samples. *Int. J. Cancer* 145 (6), 1679–1689. doi:10.1002/ijc.32190
- Lin, A., Zhang, J., and Luo, P. (2020). Crosstalk between the MSI status and tumor microenvironment in colorectal cancer. *Front. Immunol.* 11, 2039. doi:10.3389/fimmu.2020.02039
- Linder, B., Grozhik, A. V., Olerin-George, A. O., Meydan, C., Mason, C. E., and Jaffrey, S. R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12 (8), 767–772. doi:10.1038/nmeth.3453
- Liu, T., Peng, F., Yu, J., Tan, Z., Rao, T., Chen, Y., et al. (2019). LC-MS-based lipid profile in colorectal cancer patients: TAGs are the main disturbed lipid markers of colorectal cancer progression. *Anal. Bioanal. Chem.* 411 (20), 5079–5088. doi:10.1007/s00216-019-01872-5
- Liu, Y., Lau, H. C., Cheng, W. Y., and Yu, J. (2023). Gut microbiome in colorectal cancer: clinical diagnosis and treatment. *Genomics Proteomics Bioinforma.* 21 (1), 84–96. doi:10.1016/j.gpb.2022.07.002
- Liu, Q., Zhang, J., Guo, C., Wang, M., Wang, C., Yan, Y., et al. (2024). Proteogenomic characterization of small cell lung cancer identifies biological insights and subtype-specific therapeutic strategies. *Cell* 187 (1), 184–203.e28. doi:10.1016/j.cell.2023.12.004
- Lu, Y., Yu, Q., Gao, Y., Zhou, Y., Liu, G., Dong, Q., et al. (2018). Identification of metastatic lymph nodes in MR imaging with faster region-based convolutional neural networks. *Cancer Res.* 78 (17), 5135–5143. doi:10.1158/0008-5472.CAN-18-0494
- Lv, J., Zhang, L., Yan, F., and Wang, X. (2018). Clinical lipidomics: a new way to diagnose human diseases. *Clin. Transl. Med.* 7 (1), 12. doi:10.1186/s40169-018-0190-9
- Marmol, I., Sanchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E., and Rodriguez Yoldi, M. J. (2017). Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *Int. J. Mol. Sci.* 18 (1), 197. doi:10.3390/ijms18010197
- Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., and Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51 (4), 584–591. doi:10.1038/s41588-019-0379-x
- Martinez-Reyes, I., and Chandel, N. S. (2021). Cancer metabolism: looking forward. *Nat. Rev. Cancer* 21 (10), 669–680. doi:10.1038/s41568-021-00378-6
- Masum, S., Hoppgood, A., Stefan, S., Flashman, K., and Khan, J. (2022). Data analytics and artificial intelligence in predicting length of stay, readmission, and mortality: a population-based study of surgical management of colorectal cancer. *Discov. Oncol.* 13 (1), 11. doi:10.1007/s12672-022-00472-7
- Mazaki, J., Katsumata, K., Ohno, Y., Udo, R., Tago, T., Kasahara, K., et al. (2021). A novel predictive model for anastomotic leakage in colorectal cancer using auto-artificial intelligence. *Anticancer Res.* 41 (11), 5821–5825. doi:10.21873/anticancer.15400
- McQuade, J. L., Daniel, C. R., Helmink, B. A., and Wargo, J. A. (2019). Modulating the microbiome to improve therapeutic response in cancer. *Lancet Oncol.* 20 (2), e77–e91. doi:10.1016/S1470-2045(18)30952-5
- Meyer, K. D. (2019). DART-seq: an antibody-free method for global m(6)A detection. *Nat. Methods* 16 (12), 1275–1280. doi:10.1038/s41592-019-0570-0
- Mirza, B., Wang, W., Wang, J., Choi, H., Chung, N. C., and Ping, P. (2019). Machine learning and integrative analysis of biomedical big data. *Genes (Basel)* 10 (2), 87. doi:10.3390/genes10020087
- Molinie, B., Wang, J., Lim, K. S., Hillebrand, R., Lu, Z. X., Van Wittenberghe, N., et al. (2016). m(6)A-LAIC-seq reveals the census and complexity of the m(6)A epitranscriptome. *Nat. Methods* 13 (8), 692–698. doi:10.1038/nmeth.3898
- Molparia, B., Oliveira, G., Wagner, J. L., Spencer, E. G., and Torkamani, A. (2018). A feasibility study of colorectal cancer diagnosis via circulating tumor DNA derived CNV detection. *PLoS One* 13 (5), e0196826. doi:10.1371/journal.pone.0196826
- Moreno, L. O., Sanchez, P. N., and Abalo, R. (2022). Lipidomics as tools for finding biomarkers of intestinal pathology: from irritable bowel syndrome to colorectal cancer. *Curr. Drug Targets* 23 (6), 636–655. doi:10.2174/1389450122666210707122151
- Mori, Y., Kudo, S. E., Misawa, M., Saito, Y., Ikematsu, H., Hotta, K., et al. (2018). Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. *Ann. Intern. Med.* 169 (6), 357–366. doi:10.7326/M18-0249
- Mori, Y., Bretthauer, M., and Kalager, M. (2021). Hopes and hypes for artificial intelligence in colorectal cancer screening. *Gastroenterology* 161 (3), 774–777. doi:10.1053/j.gastro.2021.04.078
- Morozova, O., Hirst, M., and Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annu. Rev. Genomics Hum. Genet.* 10, 135–151. doi:10.1146/annurev-genom-082908-145957
- Muller, D., and Gyorffy, B. (2022). DNA methylation-based diagnostic, prognostic, and predictive biomarkers in colorectal cancer. *Biochim. Biophys. Acta Rev. Cancer* 1877 (3), 188722. doi:10.1016/j.bbcan.2022.188722
- Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med. Ethics* 22 (1), 122. doi:10.1186/s12910-021-00687-3
- Nakatsu, G., Li, X., Zhou, H., Sheng, J., Wong, S. H., Wu, W. K., et al. (2015). Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat. Commun.* 6, 8727. doi:10.1038/ncomms9727
- Nam, A. S., Chaligne, R., and Landau, D. A. (2021). Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat. Rev. Genet.* 22 (1), 3–18. doi:10.1038/s41576-020-0265-5
- Nannini, G., Meoni, G., Amedei, A., and Tenori, L. (2020). Metabolomics profile in gastrointestinal cancers: update and future perspectives. *World J. Gastroenterol.* 26 (20), 2514–2532. doi:10.3748/wjg.v26.i20.2514
- Narla, A., Kuprel, B., Sarin, K., Novoa, R., and Ko, J. (2018). Automated classification of skin lesions: from pixels to practice. *J. Invest. Dermatol.* 138 (10), 2108–2110. doi:10.1016/j.jid.2018.06.175
- Nazari, E., Aghemiri, M., Avan, A., Mehrabian, A., and Tabesh, H. (2021). Machine learning approaches for classification of colorectal cancer with and without feature selection method on microarray data. *Gene Rep.* 25, 101419. doi:10.1016/j.genrep.2021.101419
- Neganova, M. E., Klochkov, S. G., Aleksandrova, Y. R., and Aliev, G. (2022). Histone modifications in epigenetic regulation of cancer: perspectives and achieved progress. *Semin. Cancer Biol.* 83, 452–471. doi:10.1016/j.semcancer.2020.07.015
- Niehues, J. M., Quirke, P., West, N. P., Grabsch, H. I., van Treeck, M., Schirris, Y., et al. (2023). Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: a retrospective multi-centric study. *Cell Rep. Med.* 4 (4), 100980. doi:10.1016/j.xcrm.2023.100980
- Ning, W., Qiao, N., Zhang, X., Pei, D., and Wang, W. (2021). Metabolic profiling analysis for clinical urine of colorectal cancer. *Asia Pac. J. Clin. Oncol.* 17 (4), 403–413. doi:10.1111/ajco.13591

- Oyaga-Iriarte, E., Insausti, A., Sayar, O., and Aldaz, A. (2019). Prediction of irinotecan toxicity in metastatic colorectal cancer patients based on machine learning models with pharmacokinetic parameters. *J. Pharmacol. Sci.* 140 (1), 20–25. doi:10.1016/j.jphs.2019.03.004
- Ozawa, T., Ishihara, S., Fujishiro, M., Kumagai, Y., Shichijo, S., and Tada, T. (2020). Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. *Ther. Adv. Gastroenterol.* 13, 1756284820910659. doi:10.1177/1756284820910659
- Paczek, S., Lukaszewicz-Zajac, M., Gryko, M., Mroczko, P., Kulczynska-Przybyk, A., and Mroczko, B. (2020). CXCL-8 in preoperative colorectal cancer patients: significance for diagnosis and cancer progression. *Int. J. Mol. Sci.* 21 (6), 2040. doi:10.3390/ijms21062040
- Pakiet, A., Kobiela, J., Stepnowski, P., Sledzinski, T., and Mika, A. (2019). Changes in lipids composition and metabolism in colorectal cancer: a review. *Lipids Health Dis.* 18 (1), 29. doi:10.1186/s12944-019-0977-8
- Pan, Y., Zhang, L., Zhang, R., Han, J., Qin, W., Gu, Y., et al. (2021). Screening and diagnosis of colorectal cancer and advanced adenoma by Bionic Glycome method and machine learning. *Am. J. Cancer Res.* 11 (6), 3002–3020.
- Papaccio, F., Garcia-Mico, B., Gimeno-Valiente, F., Cabeza-Segura, M., Gambardella, V., Gutierrez-Bravo, M. F., et al. (2023). Proteotranscriptomic analysis of advanced colorectal cancer patient derived organoids for drug sensitivity prediction. *J. Exp. Clin. Cancer Res.* 42 (1), 8. doi:10.1186/s13046-022-02591-z
- Park, S. H., Park, H. M., Baek, K. R., Ahn, H. M., Lee, I. Y., and Son, G. M. (2020). Artificial intelligence based real-time microcirculation analysis system for laparoscopic colorectal surgery. *World J. Gastroenterol.* 26 (44), 6945–6962. doi:10.3748/wjg.v26.i44.6945
- Park, D. K., Kim, E. J., Im, J. P., Lim, H., Lim, Y. J., Byeon, J. S., et al. (2024). A prospective multicenter randomized controlled trial on artificial intelligence assisted colonoscopy for enhanced polyp detection. *Sci. Rep.* 14 (1), 25453. doi:10.1038/s41598-024-77079-1
- Peterson, R. E., Kuchenbaecker, K., Walters, R. K., Chen, C. Y., Popejoy, A. B., Periyasamy, S., et al. (2019). Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell.* 179 (3), 589–603. doi:10.1016/j.cell.2019.08.051
- Petitprez, F., Vano, Y. A., Becht, E., Giraldo, N. A., de Reynies, A., Sautes-Fridman, C., et al. (2018). Transcriptomic analysis of the tumor microenvironment to guide prognosis and immunotherapies. *Cancer Immunol. Immunother.* 67 (6), 981–988. doi:10.1007/s00262-017-2058-z
- Picard, M., Scott-Boyer, M. P., Bodein, A., Perin, O., and Droit, A. (2021). Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* 19, 3735–3746. doi:10.1016/j.csbj.2021.06.030
- Plana, D., Shung, D. L., Grimshaw, A. A., Saraf, A., Sung, J. J. Y., and Kann, B. H. (2022). Randomized clinical trials of machine learning interventions in health care: a systematic review. *JAMA Netw. Open* 5 (9), e2233946. doi:10.1001/jamanetworkopen.2022.33946
- Punt, C. J., Koopman, M., and Vermeulen, L. (2017). From tumour heterogeneity to advances in precision treatment of colorectal cancer. *Nat. Rev. Clin. Oncol.* 14 (4), 235–246. doi:10.1038/nrclinonc.2016.171
- Rascio, F., Spadaccino, F., Rocchetti, M. T., Castellano, G., Stallone, G., Netti, G. S., et al. (2021). The pathogenic role of PI3K/AKT pathway in cancer onset and drug resistance: an updated review. *Cancers (Basel)* 13 (16), 3949. doi:10.3390/cancers13163949
- Rathore, S., Hussain, M., Aksam Iftikhar, M., and Jalil, A. (2015). Novel structural descriptors for automated colon cancer detection and grading. *Comput. Methods Programs Biomed.* 121 (2), 92–108. doi:10.1016/j.cmpb.2015.05.008
- Reis-Filho, J. S., and Kather, J. N. (2023). Overcoming the challenges to implementation of artificial intelligence in pathology. *J. Natl. Cancer Inst.* 115 (6), 608–612. doi:10.1093/jnci/djad048
- Rezapour, S., Hosseinzadeh, E., Marofi, F., and Hassanzadeh, A. (2019). Epigenetic-based therapy for colorectal cancer: Prospect and involved mechanisms. *J. Cell Physiol.* 234 (11), 19366–19383. doi:10.1002/jcp.28658
- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *CA Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654
- Sailer, F., Pobiruchin, M., Bochum, S., Martens, U. M., and Schramm, W. (2015). Prediction of 5-Year survival with data mining algorithms. *Stud. Health Technol. Inf.* 213, 75–78.
- Saillard, C., Dubois, R., Tchita, O., Loiseau, N., Garcia, T., Adriansen, A., et al. (2023). Validation of MSIntuit as an AI-based pre-screening tool for MSI detection from colorectal cancer histology slides. *Nat. Commun.* 14 (1), 6695. doi:10.1038/s41467-023-42453-6
- Sanchez, J. N., Wang, T., and Cohen, M. S. (2018). BRAF and MEK inhibitors: use and resistance in BRAF-mutated cancers. *Drugs* 78 (5), 549–566. doi:10.1007/s40265-018-0884-8
- Schmidt, D. R., Patel, R., Kirsch, D. G., Lewis, C. A., Vander Heiden, M. G., and Locasale, J. W. (2021). Metabolomics in cancer research and emerging applications in clinical oncology. *CA Cancer J. Clin.* 71 (4), 333–358. doi:10.3322/caac.21670
- Sebio, A., Kahn, M., and Lenz, H. J. (2014). The potential of targeting Wnt/beta-catenin in colon cancer. *Expert Opin. Ther. Targets* 18 (6), 611–615. doi:10.1517/14728222.2014.906580
- Shephard, A. J., Bashir, R. M. S., Mahmood, H., Jahanifar, M., Minhas, F., Raza, S. E. A., et al. (2024). A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia. *NPJ Precis. Oncol.* 8 (1), 137. doi:10.1038/s41698-024-00624-8
- Shi, R., Sun, J., Zhou, Z., Su, Q., and Shu, Y. (2025). Deep multimodal fusion of patho-radiomic and clinical data for enhanced survival prediction for colorectal cancer patients. *NPJ Digit. Med.* 9 (1), 86. doi:10.1038/s41746-025-02210-z
- Singer, B. D. (2019). A practical guide to the measurement and analysis of DNA methylation. *Am. J. Respir. Cell Mol. Biol.* 61 (4), 417–428. doi:10.1165/rcmb.2019-0150TR
- Snaebjornsson, M. T., Janaki-Raman, S., and Schulze, A. (2020). Greasing the wheels of the cancer machine: the role of lipid metabolism in cancer. *Cell Metab.* 31 (1), 62–76. doi:10.1016/j.cmet.2019.11.010
- Song, E. M., Byeon, J. S., Lee, S. M., Yoo, H. J., Kim, S. J., Lee, S. H., et al. (2018). Fecal fatty acid profiling as a potential new screening biomarker in patients with colorectal cancer. *Dig. Dis. Sci.* 63 (5), 1229–1236. doi:10.1007/s10620-018-4982-y
- Song, C. L., Vardaki, M. Z., Goldin, R. D., and Kazarian, S. G. (2019). Fourier transform infrared spectroscopic imaging of colon tissues: evaluating the significance of amide I and C-H stretching bands in diagnostic applications with machine learning. *Anal. Bioanal. Chem.* 411 (26), 6969–6981. doi:10.1007/s00216-019-02069-6
- Song, A. H., Jaume, G., Williamson, D. F., Lu, M. Y., Vaidya, A., Miller, T. R., et al. (2023). Artificial intelligence for digital and computational pathology. *Nat. Rev. Bioeng.* 1 (12), 930–949. doi:10.1038/s44222-023-00096-8
- Sonoda, Y., Kurokawa, R., Nakamura, Y., Kanzawa, J., Kurokawa, M., Ohizumi, Y., et al. (2024). Diagnostic performances of GPT-4o, Claude 3 Opus, and Gemini 1.5 Pro in “Diagnosis Please” cases. *Jpn. J. Radiol.* 42 (11), 1231–1235. doi:10.1007/s11604-024-01619-y
- Sun, J., Liu, Y., Zhao, J., Lu, B., Zhou, S., Lu, W., et al. (2024). Plasma proteomic and polygenic profiling improve risk stratification and personalized screening for colorectal cancer. *Nat. Commun.* 15 (1), 8873. doi:10.1038/s41467-024-52894-2
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660
- Tai, Y. H., Chang, W. K., Wu, H. L., Chan, M. Y., Chen, H. H., and Chang, K. Y. (2018). The effect of epidural analgesia on cancer progression in patients with stage IV colorectal cancer after primary tumor resection: a retrospective cohort study. *PLoS One* 13 (7), e0200893. doi:10.1371/journal.pone.0200893
- Takamatsu, M., Yamamoto, N., Kawachi, H., Chino, A., Saito, S., Ueno, M., et al. (2019). Prediction of early colorectal cancer metastasis by machine learning using digital slide images. *Comput. Methods Programs Biomed.* 178, 155–161. doi:10.1016/j.cmpb.2019.06.022
- Takenaka, Y., Miyoshi, N., Fujino, S., Takahashi, Y., Nishimura, J., Yasui, M., et al. (2020). Development of a novel prediction model for recurrent stage II colon cancer. *Surg. Today* 50 (4), 389–395. doi:10.1007/s00595-019-01897-4
- Tang, X., Huang, Y., Lei, J., Luo, H., and Zhu, X. (2019). The single-cell sequencing: new developments and medical applications. *Cell Biosci.* 9, 53. doi:10.1186/s13578-019-0314-y
- Taylor, S. A., Iinuma, G., Saito, Y., Zhang, J., and Halligan, S. (2008). CT colonography: computer-aided detection of morphologically flat T1 colonic carcinoma. *Eur. Radiol.* 18 (8), 1666–1673. doi:10.1007/s00330-008-0936-7
- Tepeli, Y. I., Unal, A. B., Akdemir, F. M., and Tastan, O. (2021). PAMOGK: a pathway graph kernel-based multiomics approach for patient clustering. *Bioinformatics* 36 (21), 5237–5246. doi:10.1093/bioinformatics/btaa655
- Tevini, J., Eder, S. K., Huber-Schonauer, U., Niederseer, D., Streibinger, G., Gostner, J. M., et al. (2022). Changing metabolic patterns along the colorectal adenoma-carcinoma sequence. *J. Clin. Med.* 11 (3), 721. doi:10.3390/jcm11030721
- Thomas, M. L., and Marcato, P. (2018). Epigenetic modifications as biomarkers of tumor development, therapy response, and recurrence across the cancer care continuum. *Cancers (Basel)* 10 (4), 101. doi:10.3390/cancers10040101
- Thomas, S., Izard, J., Walsh, E., Batich, K., Chongsathidkiet, P., Clarke, G., et al. (2017). The host microbiome regulates and maintains human health: a primer and perspective for non-microbiologists. *Cancer Res.* 77 (8), 1783–1812. doi:10.1158/0008-5472.CAN-16-2929
- Tjalsma, H. (2010). Identification of biomarkers for colorectal cancer through proteomics-based approaches. *Expert Rev. Proteomics* 7 (6), 879–895. doi:10.1586/epi.10.81
- Tom, E., Keane, P. A., Blazes, M., Pasquale, L. R., Chiang, M. F., Lee, A. Y., et al. (2020). Protecting data privacy in the age of AI-enabled ophthalmology. *Transl. Vis. Sci. Technol.* 9 (2), 36. doi:10.1167/tvst.9.2.36

- Tong, D., Tian, Y., Zhou, T., Ye, Q., Li, J., Ding, K., et al. (2020). Improving prediction performance of colon cancer prognosis based on the integration of clinical and multi-omics data. *BMC Med. Inf. Decis. Mak.* 20 (1), 22. doi:10.1186/s12911-020-1043-1
- Torres, S., Bartolome, R. A., Mendes, M., Barderas, R., Fernandez-Acenero, M. J., Pelaez-Garcia, A., et al. (2013). Proteome profiling of cancer-associated fibroblasts identifies novel proinflammatory signatures and prognostic markers for colorectal cancer. *Clin. Cancer Res.* 19 (21), 6006–6019. doi:10.1158/1078-0432.CCR-13-1130
- Trebeschi, S., van Griethuysen, J. J. M., Lambregts, D. M. J., Lahaye, M. J., Parmar, C., Bakers, F. C. H., et al. (2017). Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Sci. Rep.* 7 (1), 5301. doi:10.1038/s41598-017-05728-9
- Troisi, J., Cavallo, P., Colucci, A., Pierri, L., Scala, G., Symes, S., et al. (2020). Metabolomics in genetic testing. *Adv. Clin. Chem.* 94, 85–153. doi:10.1016/bs.acc.2019.07.009
- Troisi, J., Tafuro, M., Lombardi, M., Scala, G., Richards, S. M., Symes, S. J. K., et al. (2022). A metabolomics-based screening proposal for colorectal cancer. *Metabolites* 12 (2), 110. doi:10.3390/metabolites12020110
- Tsai, P. C., Lee, T. H., Kuo, K. C., Su, F. Y., Lee, T. M., Marostica, E., et al. (2023). Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients. *Nat. Commun.* 14 (1), 2102. doi:10.1038/s41467-023-37179-4
- Udo, R., Katsumata, K., Kuwabara, H., Enomoto, M., Ishizaki, T., Sunamura, M., et al. (2020). Urinary charged metabolite profiling of colorectal cancer using capillary electrophoresis-mass spectrometry. *Sci. Rep.* 10 (1), 21057. doi:10.1038/s41598-020-78038-2
- van der Laak, J., Litjens, G., and Ciompi, F. (2021). Deep learning in histopathology: the path to the clinic. *Nat. Med.* 27 (5), 775–784. doi:10.1038/s41591-021-01343-4
- van Rijn, J. C., Reitsma, J. B., Stoker, J., Bossuyt, P. M., van Deventer, S. J., and Dekker, E. (2006). Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am. J. Gastroenterol.* 101 (2), 343–350. doi:10.1111/j.1572-0241.2006.00390.x
- Vignoli, A., Ghini, V., Meoni, G., Licari, C., Takis, P. G., Tenori, L., et al. (2019). High-throughput metabolomics by 1D NMR. *Angew. Chem. Int. Ed. Engl.* 58 (4), 968–994. doi:10.1002/anie.201804736
- Vilar, E., and Gruber, S. B. (2010). Microsatellite instability in colorectal cancer-the stable evidence. *Nat. Rev. Clin. Oncol.* 7 (3), 153–162. doi:10.1038/nrclinonc.2009.237
- Vitiello, P. P., Cardone, C., Martini, G., Ciardiello, D., Belli, V., Matrone, N., et al. (2019). Receptor tyrosine kinase-dependent PI3K activation is an escape mechanism to vertical suppression of the EGFR/RAS/MAPK pathway in KRAS-mutated human colorectal cancer cell lines. *J. Exp. Clin. Cancer Res.* 38 (1), 41. doi:10.1186/s13046-019-1035-0
- Wagner, P., Strodtthoff, N., Boussetjot, R.-D., Kreiseler, D., Lunze, F. I., Samek, W., et al. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Sci. Data* 7 (1), 154. doi:10.1038/s41597-020-0495-6
- Wagner, S. J., Reisenbuchler, D., West, N. P., Niehues, J. M., Zhu, J., Foersch, S., et al. (2023). Transformer-based biomarker prediction from colorectal cancer histology: a large-scale multicentric study. *Cancer Cell* 41 (9), 1650–1661.e4. doi:10.1016/j.ccell.2023.08.002
- Wan, N., Weinberg, D., Liu, T. Y., Niehaus, K., Ariazi, E. A., Delubac, D., et al. (2019). Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. *BMC Cancer* 19 (1), 832. doi:10.1186/s12885-019-6003-8
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10 (1), 57–63. doi:10.1038/nrg2484
- Wang, H., Tso, V. K., Slupsky, C. M., and Fedorak, R. N. (2010). Metabolomics and detection of colorectal cancer in humans: a systematic review. *Future Oncol.* 6 (9), 1395–1406. doi:10.2217/fon.10.107
- Wang, Y. H., Nguyen, P. A., Islam, M. M., Li, Y. C., and Yang, H. C. (2019). Development of deep learning algorithm for detection of colorectal cancer in EHR data. *Stud. Health Technol. Inf.* 264, 438–441. doi:10.3233/SHTI190259
- Wang, R., Li, B., Lam, S. M., and Shui, G. (2020a). Integration of lipidomics and metabolomics for in-depth understanding of cellular mechanism and disease progression. *J. Genet. Genomics* 47 (2), 69–83. doi:10.1016/j.jgg.2019.11.009
- Wang, Y., Xiao, Y., Dong, S., Yu, Q., and Jia, G. (2020b). Antibody-free enzyme-assisted chemical approach for detection of N(6)-methyladenosine. *Nat. Chem. Biol.* 16 (8), 896–903. doi:10.1038/s41589-020-0525-x
- Wang, P., Liu, X., Berzin, T. M., Glissen Brown, J. R., Liu, P., Zhou, C., et al. (2020c). Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CAdE-DB trial): a double-blind randomised study. *Lancet Gastroenterol. Hepatol.* 5 (4), 343–351. doi:10.1016/S2468-1253(19)30411-X
- Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., et al. (2022). Transformer-based unsupervised contrastive learning for histopathological image classification. *Med. Image Anal.* 81, 102559. doi:10.1016/j.media.2022.102559
- Wang, X., Du, Y., Yang, S., Zhang, J., Wang, M., Zhang, J., et al. (2023). RetCCL: clustering-guided contrastive learning for whole-slide image retrieval. *Med. Image Anal.* 83, 102645. doi:10.1016/j.media.2022.102645
- Wang, X., Zhao, J., Marostica, E., Yuan, W., Jin, J., Zhang, J., et al. (2024). A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature* 634 (8035), 970–978. doi:10.1038/s41586-024-07894-z
- Wang, X., Jiang, Y., Yang, S., Wang, F., Zhang, X., Wang, W., et al. (2025). Foundation model for predicting prognosis and adjuvant therapy benefit from digital pathology in GI cancers. *J. Clin. Oncol.* 43 (32), 3468–3481. doi:10.1200/JCO-24-01501
- Wang, J., Liu, M., Liao, H., Fan, J., Zhu, H., Ma, T., et al. (2025). Development of deep learning-based narrow-band imaging endocytoscopic classification for predicting colorectal lesions from a retrospective study. *Nat. Commun.* 16 (1), 8351. doi:10.1038/s41467-025-63812-5
- Wang, W., Zhang, Q., Fan, S., Wang, Y., Le, X., Ai, M., et al. (2025). Prediction of KRAS gene mutations in colorectal cancer using a CT-based radiomic model. *Front. Med. (Lausanne)* 12, 1592497. doi:10.3389/fmed.2025.1592497
- Warnat-Herresthal, S., Schultze, H., Shastry, K. L., Manamohan, S., Mukherjee, S., Garg, V., et al. (2021). Swarm learning for decentralized and confidential clinical machine learning. *Nature* 594 (7862), 265–270. doi:10.1038/s41586-021-03583-3
- Weir, T. L., Manter, D. K., Sheflin, A. M., Barnett, B. A., Heuberger, A. L., and Ryan, E. P. (2013). Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS One* 8 (8), e70803. doi:10.1371/journal.pone.0070803
- Weiser, M. R., Landmann, R. G., Kattan, M. W., Gonen, M., Shia, J., Chou, J., et al. (2008). Individualized prediction of colon cancer recurrence using a nomogram. *J. Clin. Oncol.* 26 (3), 380–385. doi:10.1200/JCO.2007.14.1291
- Wishart, D. S. (2019). Metabolomics for investigating physiological and pathophysiological processes. *Physiol. Rev.* 99 (4), 1819–1875. doi:10.1152/physrev.00035.2018
- Wu, Y., Wang, X., Wu, F., Huang, R., Xue, F., Liang, G., et al. (2012). Transcriptome profiling of the cancer, adjacent non-tumor and distant normal tissues from a colorectal cancer patient by deep sequencing. *PLoS One* 7 (8), e41001. doi:10.1371/journal.pone.0041001
- Xie, C., Ning, Z., Guo, T., Yao, L., Chen, X., Huang, W., et al. (2025). Multimodal data integration for biologically-relevant artificial intelligence to guide adjuvant chemotherapy in stage II colorectal cancer. *EBioMedicine* 117, 105789. doi:10.1016/j.ebiom.2025.105789
- Xu, J. Y., Zhang, C., Wang, X., Zhai, L., Ma, Y., Mao, Y., et al. (2020). Integrative proteomic characterization of human lung adenocarcinoma. *Cell* 182 (1), 245–261.e17. doi:10.1016/j.cell.2020.05.043
- Xu, Y., Ju, L., Tong, J., Zhou, C. M., and Yang, J. J. (2020). Machine learning algorithms for predicting the recurrence of stage IV colorectal cancer after tumor resection. *Sci. Rep.* 10 (1), 2519. doi:10.1038/s41598-020-59115-y
- Yamada, M., Saito, Y., Imaoka, H., Saiko, M., Yamada, S., Kondo, H., et al. (2019). Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci. Rep.* 9 (1), 14465. doi:10.1038/s41598-019-50567-5
- Yamamoto, T., Kudo, M., Peng, W. X., Takata, H., Takakura, H., Teduka, K., et al. (2016). Identification of aldolase A as a potential diagnostic biomarker for colorectal cancer based on proteomic analysis using formalin-fixed paraffin-embedded tissue. *Tumour Biol.* 37 (10), 13595–13606. doi:10.1007/s13277-016-5275-8
- Yang, Y., Misra, B. B., Liang, L., Bi, D., Weng, W., Wu, W., et al. (2019). Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. *Theranostics* 9 (14), 4101–4114. doi:10.7150/thno.35186
- Yang, J., Li, D., Yang, Z., Dai, W., Feng, X., Liu, Y., et al. (2020). Establishing high-accuracy biomarkers for colorectal cancer by comparing fecal microbiomes in patients with healthy families. *Gut Microbes* 11 (4), 918–929. doi:10.1080/19490976.2020.1712986
- Yang, M., Yang, H., Ji, L., Hu, X., Tian, G., Wang, B., et al. (2022). A multi-omics machine learning framework in predicting the survival of colorectal cancer patients. *Comput. Biol. Med.* 146, 105516. doi:10.1016/j.compbiomed.2022.105516
- Yang, L., Xu, S., Sellergren, A., Kohlberger, T., Zhou, Y., Ktena, I., et al. (2024). Advancing multimodal medical capabilities of Gemini. *arXiv*. doi:10.48550/arXiv.2405.03162
- Yang, J., Cai, D., Liu, J., Zhuang, Z., Zhao, Y., Wang, F. A., et al. (2025). CRCFound: a colorectal cancer CT image foundation model based on self-supervised learning. *Adv. Sci. (Weinh)* 12 (41), e07339. doi:10.1002/adv.202407339
- Yin, F. T., Zhou, X. H., Kang, S. Y., Li, X. H., Li, J., Ullah, I., et al. (2022). Prediction of the mechanism of Dachengqi Decoction treating colorectal cancer based on the analysis method of “into serum components -action target-key pathway”. *J. Ethnopharmacol.* 293, 115286. doi:10.1016/j.jep.2022.115286
- Yoshino, T., Argiles, G., Oki, E., Martinelli, E., Taniguchi, H., Arnold, D., et al. (2021). Pan-Asian adapted ESMO Clinical Practice Guidelines for the diagnosis treatment and follow-up of patients with localised colon cancer. *Ann. Oncol.* 32 (12), 1496–1510. doi:10.1016/j.annonc.2021.08.1752
- Yuan, L., Zhou, H., Xiao, X., Zhang, X., Chen, F., Liu, L., et al. (2024). Development and external validation of a transfer learning-based system for the pathological diagnosis of colorectal cancer: a large emulated prospective study. *Front. Oncol.* 14, 1365364. doi:10.3389/fonc.2024.1365364
- Yuan, W., Chen, Y., Zhu, B., Yang, S., Zhang, J., Mao, N., et al. (2025). Pancancer outcome prediction via a unified weakly supervised deep learning model. *Signal Transduct. Target Ther.* 10 (1), 285. doi:10.1038/s41392-025-02374-w
- Yun, L., and Guangwei, L. (2020). Application, problems and countermeasures of artificial intelligence in colorectal. *Cancer* 40, 271–274.

- Zackular, J. P., Rogers, M. A., Ruffin, M. T., and Schloss, P. D. (2014). The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev. Res. (Phila)* 7 (11), 1112–1121. doi:10.1158/1940-6207.CAPR-14-0129
- Zafari, N., Bathaei, P., Velayati, M., Khojasteh-Leylakoochi, F., Khazaei, M., Fiuji, H., et al. (2023). Integrated analysis of multi-omics data for the discovery of biomarkers and therapeutic targets for colorectal cancer. *Comput. Biol. Med.* 155, 106639. doi:10.1016/j.compbiomed.2023.106639
- Zamani, M., Hosseini, S. V., and Mokarram, P. (2018). Epigenetic biomarkers in colorectal cancer: premises and prospects. *Biomarkers* 23 (2), 105–114. doi:10.1080/1354750X.2016.1252961
- Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. I., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* 10 (11), 766. doi:10.15252/msb.20145645
- Zhan, W., Liao, X., Liu, J., Tian, T., Yu, L., and Li, R. (2020). USP38 regulates the stemness and chemoresistance of human colorectal cancer via regulation of HDAC3. *Oncogenesis* 9 (5), 48. doi:10.1038/s41389-020-0234-z
- Zhang, F., Gu, W., Hurles, M. E., and Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* 10, 451–481. doi:10.1146/annurev.genom.9.081307.164217
- Zhang, S., Wei, J. S., and Khan, J. (2014). “Chapter 4 - the significance of transcriptome sequencing in personalized cancer medicine,” in *Cancer genomics*. Editors G. Dellaire, J. N. Berman, and R. J. Arceci (Boston: Academic Press), 49–64.
- Zhang, A., Sun, H., Yan, G., Wang, P., and Wang, X. (2015). Metabolomics for biomarker discovery: moving to the clinic. *Biomed. Res. Int.* 2015, 354671. doi:10.1155/2015/354671
- Zhang, A., Sun, H., and Wang, X. (2018). Mass spectrometry-driven drug discovery for development of herbal medicine. *Mass Spectrom. Rev.* 37 (3), 307–320. doi:10.1002/mas.21529
- Zhang, Z., Chen, L. Q., Zhao, Y. L., Yang, C. G., Roundtree, I. A., Zhang, Z., et al. (2019). Single-base mapping of m(6)A by an antibody-independent method. *Sci. Adv.* 5 (7), eaax0250. doi:10.1126/sciadv.aax0250
- Zhang, G. L., Pan, L. L., Huang, T., and Wang, J. H. (2019). The transcriptome difference between colorectal tumor and normal tissues revealed by single-cell sequencing. *J. Cancer* 10 (23), 5883–5890. doi:10.7150/jca.32267
- Zhang, Y., Song, J., Zhao, Z., Yang, M., Chen, M., Liu, C., et al. (2020). Single-cell transcriptome analysis reveals tumor immune microenvironment heterogeneity and granulocytes enrichment in colorectal cancer liver metastases. *Cancer Lett.* 470, 84–94. doi:10.1016/j.canlet.2019.10.016
- Zhang, X., Wan, S., Yu, Y., Ruan, W., Wang, H., Xu, L., et al. (2020). Identifying potential DNA methylation markers in early-stage colorectal cancer. *Genomics* 112 (5), 3365–3373. doi:10.1016/j.ygeno.2020.06.007
- Zhang, X., Yang, L., Deng, Y., Huang, Z., Huang, H., Wu, Y., et al. (2023). Single-cell RNA-seq and bulk RNA-seq reveal reliable diagnostic and prognostic biomarkers for CRC. *J. Cancer Res. Clin. Oncol.* 149 (12), 9805–9821. doi:10.1007/s00432-023-04882-0
- Zhang, S. L., Cheng, L. S., Zhang, Z. Y., Sun, H. T., and Li, J. J. (2023). Untangling determinants of gut microbiota and tumor immunologic status through a multi-omics approach in colorectal cancer. *Pharmacol. Res.* 188, 106633. doi:10.1016/j.phrs.2022.106633
- Zhang, Q., Xu, X., Jiang, D., Wang, Y., Wang, H., Zhu, J., et al. (2025). Integrated proteogenomic characterization of ampullary adenocarcinoma. *Cell Discov.* 11 (1), 2. doi:10.1038/s41421-024-00742-4
- Zhao, Q., and Hastie, T. (2019). Causal interpretations of black-box models. *J. Bus. Econ. Stat.* 2019. doi:10.1080/07350015.2019.1624293
- Zhao, G., Liu, X., Liu, Y., Li, H., Ma, Y., Li, S., et al. (2020). Aberrant DNA methylation of SEPT9 and SDC2 in stool specimens as an integrated biomarker for colorectal cancer early detection. *Front. Genet.* 11, 643. doi:10.3389/fgene.2020.00643
- Zhao, Q., Wang, F., Chen, Y. X., Chen, S., Yao, Y. C., Zeng, Z. L., et al. (2022). Comprehensive profiling of 1015 patients' exomes reveals genomic-clinical associations in colorectal cancer. *Nat. Commun.* 13 (1), 2342. doi:10.1038/s41467-022-30062-8
- Zhao, N., Lai, C., Wang, Y., Dai, S., and Gu, H. (2024). Understanding the role of DNA methylation in colorectal cancer: mechanisms, detection, and clinical significance. *Biochim. Biophys. Acta Rev. Cancer* 1879 (3), 189096. doi:10.1016/j.bbcan.2024.189096
- Zhu, W., Xie, L., Han, J., and Guo, X. (2020). The application of deep learning in cancer prognosis prediction. *Cancers (Basel)* 12 (3), 603. doi:10.3390/cancers12030603
- Zhu, G., Jin, L., Sun, W., Wang, S., and Liu, N. (2022). Proteomics of post-translational modifications in colorectal cancer: discovery of new biomarkers. *Biochim. Biophys. Acta Rev. Cancer* 1877 (4), 188735. doi:10.1016/j.bbcan.2022.188735