

Multimodal foundation models in colorectal cancer: from prediction to trustworthy clinical insight

Danial Gharai Amirabadi¹, Adib Miraki Feriz², Hossein Safarpour^{2,*}

¹Student Research Committee, Birjand University of Medical Sciences, PO Box 9717853577, Birjand, Ghaffari Street, South Khorasan Province, Iran

²Cellular and Molecular Research Center, Birjand University of Medical Sciences, PO Box 9717853577, Birjand, Ghaffari Street, South Khorasan Province, Iran

*Corresponding author. Cellular and Molecular Research Center, Birjand University of Medical Sciences, Birjand, Iran. E-mail: Safarpour701@yahoo.com

Abstract

Colorectal cancer (CRC) is characterized by profound, multi-layered heterogeneity that limits the precision of conventional single-modality clinical tools. The emergence of multimodal foundation models (MFMs) represents a conceptual paradigm shift, moving beyond static biomarkers to capture the dynamic and evolving nature of CRC. MFMs integrate histopathology, radiology, multi-omics data (including the critical regulatory layer of epigenomics), and clinical variables into shared high-dimensional representational spaces. This integration enables improved prognostication, refined molecular subtyping, and *in silico* simulation of therapeutic perturbations within the tumor's functional landscape, thereby supporting rational and model-driven drug development. In this review, we synthesize the rapidly expanding body of CRC-specific MFM research and critically examine the unresolved challenges that currently limit clinical translation. We place particular emphasis on the transition from correlation to causal inference, the establishment of cross-population generalizability, and the resolution of key issues related to trustworthiness and clinical interpretability. Finally, we propose an actionable roadmap outlining regulatory, data governance, and translational requirements, including the lab-in-the-loop paradigm, necessary to position MFMs as a robust and equitable framework in clinical oncology.

Keywords colorectal cancer, multimodal foundation models, functional space navigation, lab-in-the-loop, trustworthy AI, precision oncology

The clinical imperative: managing the heterogeneity of colorectal cancer

Colorectal cancer (CRC) is the third most commonly diagnosed malignancy and the second leading cause of cancer-related mortality worldwide [1]. Despite substantial advances in molecular characterization, therapeutic decision-making remains hindered by pronounced heterogeneity at multiple levels. Inter-patient heterogeneity results in markedly different disease trajectories and treatment responses among patients with identical TNM stages, while intra-patient heterogeneity encompasses both spatial variation (e.g. tumor core versus invasive front) and temporal evolution (e.g. pre- versus post-treatment biological states) [2].

Current clinical decision-making relies on a limited set of genetic biomarkers, including RAS (KRAS, NRAS), BRAF mutations, as well as mismatch repair deficiency (dMMR) or high microsatellite instability (MSI-H) [3, 4]. Although these biomarkers are clinically actionable, they apply to only a subset of patients and fail to capture the full biological complexity underlying CRC heterogeneity [5].

Comprehensive molecular profiling, particularly through RNA sequencing, has enabled broader biomarker discovery and the

development of consensus molecular subtypes (CMS) [6, 7]. However, these approaches remain constrained by their static snapshots as they capture tumor state at a single time point and provide limited insight into dynamic processes such as therapeutic adaptation, metastatic dissemination, and microenvironmental reprogramming [8].

This limitation have driven the development of computational 'virtual tumor' models, designed as high-fidelity digital representations of patient-specific tumors capable of simulating disease evolution and treatment response. Pioneering initiatives, such as the Center for the Development of a Virtual Tumor (CViT) platform at Massachusetts General Hospital, integrate multimodal data to generate predictive simulations of tumor behavior under therapeutic perturbations [9, 10]. While these frameworks underscore the clinical need for dynamic, patient-specific modeling, they remain constrained by fragmented data integration and lack of generalizability.

The advent of multimodal foundation models (MFMs) represents a paradigm shift beyond traditional static biomarker frameworks. Trained in a self-supervised manner on large and diverse biological and clinical datasets, MFMs capture fundamental principles governing molecular and cellular systems [11, 12]. Generative MFMs, in

Received: September 15, 2025. **Revised:** February 2, 2026. **Accepted:** March 5, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

particular, extend beyond correlation-based prediction by enabling simulation of tumor adaptation, resistance trajectories, and therapeutic perturbations within high-dimensional functional spaces [13].

Crucially, MFMs unify disparate data modalities, including histopathology, radiology, genomics, transcriptomics, proteomics, metabolomics, spatial omics, epigenomics, and electronic health records (EHRs), into shared representational frameworks [14, 15]. This integration enables CRC to be modeled not as collection of isolated data layers, but as an interconnected, evolving system. Early evidence supports this potential. Hu *et al.* (2025) demonstrated that a pan-cancer MFM integrating histopathology and genomics data improved prognostic accuracy across 18 tumor types, highlighting the presence of shared biological principles across malignancies [16].

By capturing both inter-patient variability (e.g. prognostic stratification) and intra-patient dynamics (e.g. spatial niche interactions and clonal evolution), MFMs pave the way for a future in which treatment decisions are informed by projected disease trajectories rather than retrospective assessments.

This review synthesizes the accelerating landscape of CRC-specific MFMs, critically evaluates unresolved challenges in causal inference, generalizability, and equity, and proposes a trustworthy translational roadmap, positioning CRC as a proving ground for MFMs in oncology.

From a practical clinical perspective, the most translation-ready MFMs rely on data that are already routinely collected within standard-of-care workflows, including hematoxylin and eosin (H&E)-stained whole-slide images (WSIs), CT, and MRI radiology scans, basic EHR variables (age, disease stage, tumor site, treatment history), and targeted gene-panel sequencing or immunohistochemical surrogates such as mismatch repair protein staining. Fully multimodal models that also ingest bulk or single-cell transcriptomics, epigenomics, or proteomics currently remain research-only because these assays are not universally performed; however, ongoing work aims to impute or predict these high-dimensional layers from routine H&E and clinical data alone, potentially eliminating the need for additional patient sampling.

To ensure relevance and coherence, models discussed in this review were selected according to predefined criteria. We prioritized architectures that (i) integrate two or more biomedical data modalities or provide a feasible pathway toward multimodal extension, (ii) demonstrate transferable utility across cancer types or large-scale biomedical tasks, (iii) report clinically meaningful endpoints (e.g. prognosis, MSI, treatment response), with external validation when available, and (iv) function as reusable or foundational backbones rather than single-purpose models. Architectural diversity was deliberately maintained, while models lacking transparency or clear translational relevance were deprioritized.

The technological revolution: the rise of foundation models in colorectal cancer

Foundation models (FMs) initially transformed natural language processing through large-scale self-supervised pretraining, enabling robust generalization across diverse downstream tasks [11]. This paradigm has since permeated biomedicine, driven by the exponential expansion of genomic, transcriptomic, imaging, and clinical datasets. In CRC, a disease defined by the complex interplay of somatic mutations, tumor microenvironment (TME) dynamics, and host-microbiota interactions, FMs provide a unifying framework not

only for predictive modeling but also for hypothesis generation and mechanistic simulation (Fig. 1).

FMs are broadly classified as discriminative or generative. Discriminative models, typically encoder-based (e.g. BioBERT [17], DNABERT [18], and ProteinBERT [19]), excel at extracting high-dimensional embeddings for downstream classification tasks such as MSI status prediction or biomarker identification. In contrast, generative architectures which are generally decoder-based (e.g. ProtGPT2 [20], ProGen [21], and ESM [22, 23]) synthesize biological sequences, impute missing data, or simulate biological perturbations, enabling applications in protein design and drug response forecasting. In CRC, this distinction translates into complementary strengths: discriminative FMs facilitate patient stratification, whereas generative FMs provide an unprecedented capacity to model tumor adaptation under therapeutic pressure, a capability largely absent from conventional biomarker-driven frameworks [13].

The architectural pillars of modern foundation models

Transformers and self-attention

The transformer architecture, introduced in 2017, replaced recurrence and convolutional operations with self-attention, enabling each token, whether a DNA k-mer, image patch, or gene expression value, to compute weighted relationships with all other tokens in the sequence [24]. This mechanism is particularly powerful in CRC, where long-range dependencies regulate both genomic control elements (e.g. distal enhancers) and tissue-level architecture (e.g. tumor budding patterns associated with metastasis risk).

Vision Transformers (ViTs) have established new performance benchmarks in computational pathology [25]. Wagner *et al.* (2023) analyzed thousands of WSI patches to predict MSI status, RAS/BRAF mutations, and CMS, achieving performance superior to state-of-the-art convolutional neural networks (CNNs) [26]. However, the necessity of billion-parameter ViTs remains a subject of debate. Distilled model variants achieve near-equivalent MSI prediction performance on The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) dataset while requiring <10% of the computational cost [27], raising critical questions regarding whether large-scale pathology FMs are over-engineered for clinically focused endpoints. Hierarchical architectures, such as Swin Transformers, partially mitigate quadratic computational scaling inherent to self-attention [28]. Nevertheless, domain shift arising from variability in staining protocols, scanners, and institutional practices remains a persistent barrier to real-world clinical deployment.

Graph neural networks

Unlike transformers, which primarily operate on sequential or grid-structured data, graph neural networks (GNNs) are explicitly designed to model relational topologies. Through iterative message-passing mechanism, node embeddings are updated based on information exchanged with neighboring nodes, making GNNs well suited for representing gene regulatory networks, signaling pathways, and multicellular communities within the TME [29]. In CRC, models such as Nicheformer leverage GNN-based representations to identify functional neighborhoods of interacting immune and stromal cells, revealing immunosuppressive niches that correlate with resistance to immunotherapy [30]. However, the predominantly static nature of most GNN implementations represents a fundamental limitation. These models typically assume fixed graph topologies and therefore

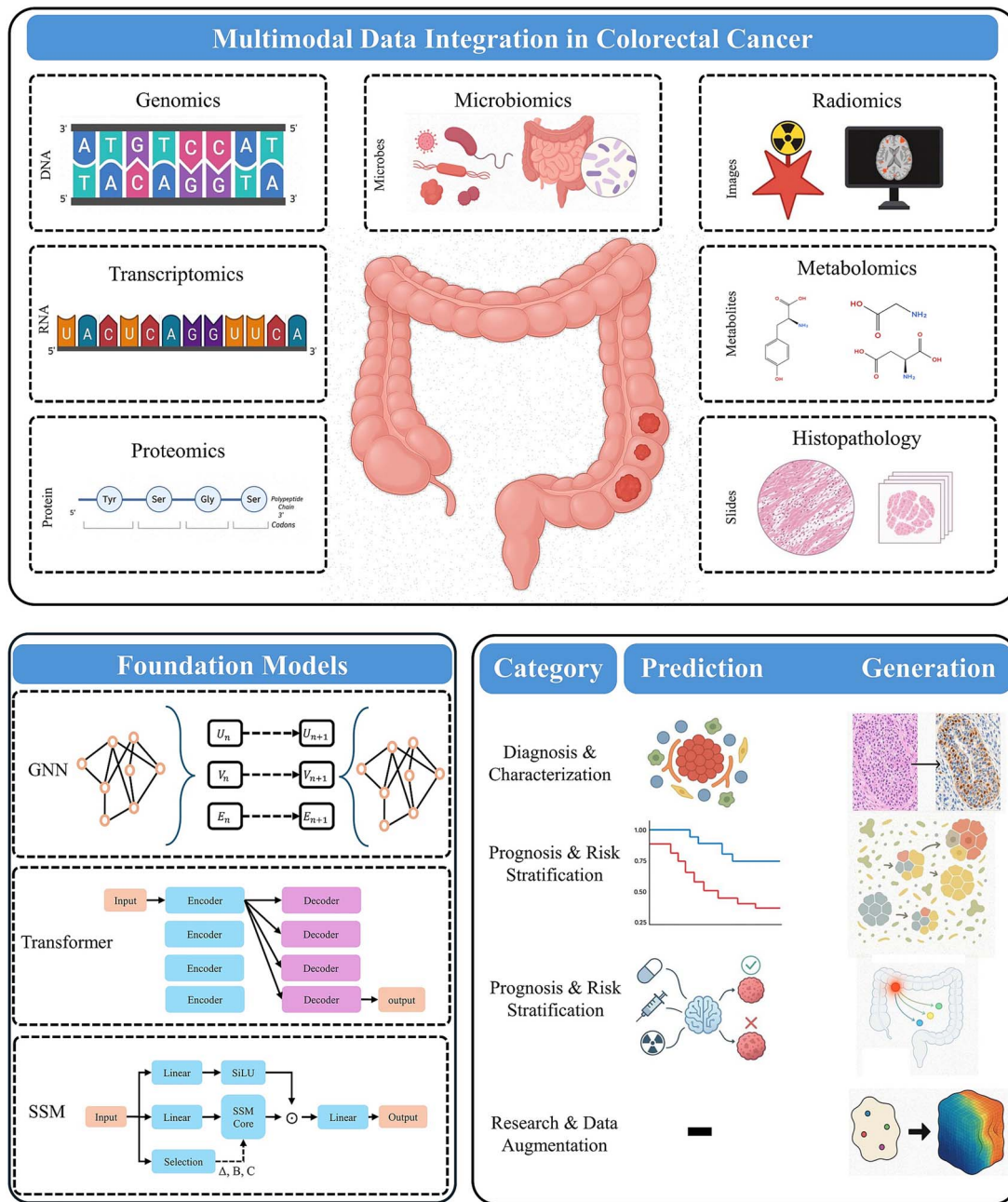


Figure 1 Conceptual overview of multimodal foundation models in colorectal cancer (CRC). Diagram illustrating how multimodal foundation models integrate genomic, transcriptomic, imaging, and clinical data to analyze colorectal cancer biology and generate predictive and mechanistic insights.

fail to capture clonal evolution, therapy-induced network rewiring, or temporal remodeling of the TME as processes that are central to CRC progression and therapeutic failure. Although residual connections and graph-rewiring strategies can partially alleviate over-smoothing [31, 32], temporal GNN architectures remain largely unexplored in oncology, representing a critical gap in the modeling dynamic disease trajectories.

State space models

State space models (SSMs), exemplified by architectures such as Mamba [33], offer a computationally efficient alternative attention-based models by representing sequences as latent dynamical systems

with selective recurrence. This linear-time complexity makes them particularly well suited for ultra-long inputs, including whole-chromosome variant profiles or longitudinal clinical records.

In CRC, approaches such as dnaGrinder apply SSM-based architectures to interpret non-coding variants across megabase-scale genomic regions, enabling the identification of regulatory elements that are often missed by shorter-context models [34]. Nevertheless, model expressivity persists, as SSMs tend to underperform transformer-based architectures when modeling highly irregular dependencies, such as focal genomic amplifications or burst-like gene expression dynamics [35, 36]. Although hybrid architectures that combine SSMs with attention mechanisms are beginning to emerge, their application

in CRC remains largely untested, leaving open the question of whether computational efficiency gains justify potential reductions in biological fidelity.

Hybrid and emerging architectures

The future of FMs increasingly lies in hybrid architectures that combine complementary strengths. CNN-transformer hybrids, such as Enformer, integrate local motif detection with long-range regulatory modeling to predict the functional effects of non-coding mutations [37].

Vision-language models align histopathology images with unstructured clinical reports, enabling multimodal reasoning across imaging data and EHR [38].

Spatially informed transformer architectures such as Nicheformer, extend attention mechanisms to tissue-level organization, enabling the identification of multicellular niches that shape prognosis and therapeutic response [30].

Graph-transformer hybrid models further embed prior biological knowledge, such as protein–protein interaction networks and host-microbiome relationships, into representational latent spaces, yielding more mechanistically grounded model outputs [39].

Autoencoder-based architectures compress high-dimensional single-cell datasets, enabling the identification of rare, drug-resistant subclonal populations that are often obscured in bulk analyses [40, 41].

Despite these advances, no MFM has yet demonstrated robust zero-shot transfer across staining protocols, ethnic cohorts, or longitudinal timepoints without task-specific fine-tuning [42, 43]. The path forward likely lies not in ever-larger models, but rather in modular and interpretable fusion frameworks coupled with continual learning strategies capable of adapting to real-world clinical drift.

Deconstructing the tumor: uni-modal foundation models in colorectal cancer

The language of histology and radiology

Digital pathology, centered on the analysis of WSIs of H&E-stained tissue, remains a primary source of morphological information reflecting underlying molecular processes [44, 45]. Early deep learning (DL) applications relied predominantly on CNNs, which achieved notable success in tasks such as tumor grading and the prediction of molecular features including MSI status [46–48]. However, CNNs are inherently limited by their local receptive fields, restricting their ability to model long-range dependencies across large WSIs [49].

The introduction of transformer-based architectures, particularly ViTs, has addressed this limitation and established a new standard in computational pathology. By employing self-attention mechanisms, transformers can evaluate the relative importance of all image patches simultaneously, enabling a global understanding of tissue architecture that is essential for analyzing complex WSIs in CRC [50–53]. This capability is clinically relevant, as features extending beyond the primary tumor also influence prognosis. For example, imaging-derived features from metastatic sites such as texture patterns in liver metastases, can independently impact patient outcomes regardless of primary tumor stage.

In parallel, convolutional architectures continue to advance pathology FMs through innovative design strategies. A notable example is ROSIE, a ConvNext-based model capable of performing *in*

silico multiplex immunofluorescence (mIF) by predicting biomarker expression directly from H&E-stained slides [54]. In a landmark study, Wagner *et al.* (2023) developed a transformer-based pipeline to predict key genomic and transcriptomic alterations in CRC from H&E slides across a multicenter cohort, demonstrating the potential of these approaches to function ‘digital molecular tests’ [26]. Beyond histopathology, attention-based models have also shown broad applicability in radiology, achieving near-expert-level accuracy in the challenging task of CRC tumor segmentation on computed tomography (CT) scans [55].

Despite these successes, substantial challenges remain. Domain shift arising from variability in slide preparation protocols, and scanning equipment across institutions can significantly degrade model performance, necessitating mitigation strategies such as stain normalization and training on large, diverse, multi-institutional datasets. Additionally, the sheer size of WSIs also presents a major computational challenge, requiring efficient patching strategies and hierarchical modeling approaches [56, 57].

An overview of FMs developed for histopathology and radiology applications in CRC is provided in [Supplementary Table S1](#).

The language of the genomic and transcriptomic

The genome serves as the fundamental blueprint of cancer, cataloging the landscape of somatic mutations and structural variants that underpin CRC classification [58, 59]. The ‘sequence-as-language’ paradigm, implemented through DNA language models like DNABERT, Enformer, and dnaGrinder [18, 34, 37], treats DNA sequences as linguistic constructs to learn the underlying grammar of gene regulation. This approach enables prediction of the functional impact of non-coding variants, extending beyond traditional variant calling.

While genomics provides the largely static blueprint, transcriptomics offers a dynamic snapshot of tumor activity, forming the basis of the landmark CMS framework [60]. More recently, AI-driven approaches have moved beyond established classifications. For example, the Gene Swin Transformer converts bulk transcriptomic data into synthetic image elements (SIEs) that can be processed by a vision transformer, enabling the identification of novel prognostic biomarkers such as *PEX10* [61].

Resolving tumor heterogeneity ultimately requires moving beyond the ‘blended’ perspective of bulk sequencing toward single-cell-level analysis. Single-cell foundation models (scFMs), such as Geneformer and scGPT [62, 63], are transformer-based architectures pretrained on millions of single cell profiles. This has led to the emergence of a broader ecosystem of models including scPRINT and scCello, which interrogate the TME at unprecedented resolution, enabling rapid cell-type annotation, inference of gene regulatory networks, and robust cross-species transfer learning [64–68].

Transcriptome FMs enable high-throughput, ontology-aware dissection of cell subpopulations, thereby refining our understanding of tumor heterogeneity and mechanisms of drug resistance. A key strength of these models lies in their ability to generate ‘digital’ surrogates for costly molecular assays, exemplified by the prediction of dMMR status directly from routine H&E slides [69].

The CRC TME represents a dynamic ecosystem in which spatial organization is critical to disease behavior. A new generation of FMs provides the tools to map this ecosystem through multi-stage analytical pipelines, including large-scale pre-trained models such as scFoundation [70] that function as universal cell annotators; spatial

FMs (e.g. scGPT-spatial) that map cells back to their tissue context [71]; and transformer-based models like Nicheformer [30] that identify multicellular ‘niches’ as functional neighborhoods of interacting cells.

Representative genomic, transcriptomic, and spatial FMs relevant to CRC are detailed in [Supplementary Tables S2 and S3](#).

The language of function, regulation, and the ecosystem

Epigenomics: the regulatory layer

Epigenomic alterations, particularly DNA methylation and histone modifications, represent a crucial regulatory layer that governing CRC progression and therapeutic response; yet, they remain underrepresented in current MFMs [72]. DNA methylation is a well-studied epigenetic mark in which the addition of a methyl group to cytosine residues regulates gene expression, often resulting in the silencing of tumor suppressor genes [72]. This process is fundamental to the CpG island methylator phenotype (CIMP), a distinct molecular subgroup of CRC characterized by widespread promoter hypermethylation [73]. Emerging FMs are being developed to process this complex, high-dimensional data by treating DNA methylation profiles as sequences. Models such as MethylGPT are transformer-based architectures pretrained on large collections of human DNA methylation profiles to learn the complex, non-linear dependencies between distant CpG sites [74]. Similarly, the Cytosine-phosphate-Guanine Pretrained Transformer leverages a specialized transformer to impute missing methylation values and is applicable for various cancer prediction and classification tasks [75]. Integrating these epigenomic FMs with other omics and imaging modalities is critical, as it provides a mechanistic link between regulatory alterations and the observable tumor phenotype, such as immune evasion within the TME. The relative scarcity of large, clinically annotated epigenomic datasets therefore represents one of the most significant bottlenecks in building truly pan-omic multimodal FMs for CRC. Bridging this gap through epigenomic imputation from H&E images, spatial transcriptomics, or cell-free DNA methylome profiling constitutes an active and clinically critical research frontier.

Proteomics: decoding the molecular machinery

Mass spectrometry (MS)-based proteomics provides a direct readout of protein abundance, post-translational modifications (PTMs), and signaling activities that collectively define cancer cell function. Initiatives such as the ‘Clinical Proteomic Tumor Analysis Consortium (CPTAC)’ have generated deep proteogenomic maps of CRC, offering unparalleled opportunities to integrate proteomic signatures with genomic alterations [76, 77]. However, a central challenge lies in extracting functional meaning from observed changes in protein levels; as differential abundance alone often fails to capture downstream mechanistic consequences. This gap is increasingly addressed by ‘protein FMs’. Large-scale transformer architectures, such as ‘ESM’, trained on billions of protein sequences across the tree of life, learn intrinsic biophysical principles governing protein folding, enzymatic function, and interaction potential [22].

A recent innovation in this space is KRONOS, a FM for spatial proteomics [78]. KRONOS is trained on 47 million image patches across diverse tissues and leverages self-supervised learning to enable segmentation-free cell phenotyping, tissue region classification, and patient stratification. Its ability to generalize across platforms and sample types underscores the value of spatial proteomic FMs in CRC,

where tumor-immune architecture and microenvironmental heterogeneity critically shape clinical outcomes. By coupling network-based reasoning with spatially resolved proteomic information, models such as KRONOS illustrate the next frontier of integrative CRC research. FMs for proteomics, protein structure, and molecular design are summarized in [Supplementary Table S4](#).

Metabolomics: mapping the biochemical circuitry

Cancer cells, including those in CRC, undergo profound metabolic reprogramming to sustain anabolic growth, evade immune surveillance, and maintain redox balance, thereby producing distinctive metabolic signatures [79]. Interpreting these signatures requires computational methods capable of handling the graph-like topology inherent to metabolic pathways [80]. GNNs are particularly well suited to this task, as metabolites can be represented as nodes and enzymatic reactions as edges, allowing information to propagate through metabolic networks and enabling mechanistically interpretable links between genotype and metabolic phenotype [81].

In parallel, artificial intelligence is reshaping how metabolomic data are acquired, annotated, and interpreted. Early transformer-based models such as DreaMS, trained on millions of unlabeled MS/MS spectra, demonstrated that pretraining can infer fragmentation grammars and substantially improve metabolite identification [82]. Building on this foundation, large-scale FMs for metabolomics have recently emerged. PRISM, trained on over one billion spectra, applies masked spectral modeling to learn transferable chemical representations, resulting in marked improvements in chemical property prediction and spectral retrieval [83]. Similarly, LSM1-MS2 employs transformer-based self-supervision on tandem MS data, enabling robust metabolite annotation and *de novo* molecular generation, even in low-data regimes [84]. Beyond MS, recent proposals for NMR-based FMs extend self-supervised learning to nuclear magnetic resonance data, highlighting the potential of multimodal frameworks in metabolomics.

Concurrently, a growing ecosystem of task-specific transformer models is advancing metabolite annotation. The RT-Transformer integrates 1D-transformer layers with graph attention networks to enhance retention time prediction [85]. MassFormer, a graph transformer, accurately predicts MS2 spectra by learning atom-bond relationships and experimental conditions [86]. Models such as MS2Prop predict chemical properties directly from spectra [87], while IDSL_MINT [88] and MIST [89] generate molecular fingerprints for efficient library matching.

In lipidomics, LipiDetective achieves improved annotation of novel lipid species from MS2 spectra [90], and MS2Mol enables *de novo* structural prediction directly from spectral data without reliance on reference libraries [91]. CRC-relevant metabolomics and cheminformatics FMs are summarized in [Supplementary Table S5](#).

Microbiomics: resolving the tumor-microbe axis

The gut microbiome is increasingly recognized as a pivotal determinant of CRC initiation, progression, and response to therapy [92]. However, advancing from correlative abundance profiles to causal mechanistic insights requires computational models that capture the community-level interactions within complex microbial consortia [93]. Graph transformer models, such as MICAHA (MICRObial Cancer-association Analysis using a Heterogeneous graph transformer), integrate abundance data, phylogenetic relationships, and metabolic capacities into a heterogeneous graph, enabling the identification

of cancer-associated microbial communities rather than isolated taxa [92]. To facilitate biological translation, interpretability frameworks are also emerging. For example, IMPACT projects microbial abundance data into image-like representations and applies saliency mapping techniques from computer vision to highlight not only key taxa predictive of CRC risk but also their associated functional attributes [94]. This integrated interpretability is critical for generating biologically testable hypotheses and identifying potential therapeutic targets [95].

Beyond these task-specific architectures, microbiome research has also benefited from the adoption of large pretrained models, which demonstrated the value of transfer learning on high-dimensional microbial data [96]. Such approaches established the feasibility of capturing community-level representations at scale, thereby providing a bridge toward the development of dedicated FMs such as the Microbial General Model (MGM) [97]. The MGM treats each microbiome sample as a ‘document’ (a sequence of tokenized taxonomic or functional units), with individual taxa/OTUs/ASVs represented as ‘tokens’, enabling self-supervised learning and attention mechanisms to model co-occurrence structure and ecological interactions. Pretrained on >260 000 microbiome samples, MGM encodes rich contextual information and captures complex interspecies interactions, enabling downstream tasks such as disease classification, biomarker discovery, and microbiome-host interaction analysis relevant to CRC. Building on this framework, MGM 2.0 extends the paradigm toward ‘generative microbiome modeling’ [98]. By reframing microbial communities as language-like sequences, MGM 2.0 introduces generative capabilities including colonization prediction, disease-conditioned microbiome synthesis, and fecal microbiota transplantation donor selection. Although not strictly a FM, MGM 2.0 illustrates how transformer-inspired architectures can progress from representation learning to simulation, highlighting future opportunities for CRC research into therapeutic microbiome engineering. FMs for microbiome and metagenome analyses are shown in [Supplementary Table S6](#).

Collectively, these functional omics layers and their associated AI models bridge the gap between genetic alterations and phenotypic manifestation. By capturing proteins, metabolites, and microbial networks as the executors and modulators of cancer biology, they provide a mechanistic framework that is essential for precision oncology.

Table 1 summarizes recent FMs and multi-omics models relevant to CRC, detailing their primary modality, key tasks, training scale, external validation, limitations, and clinical maturity.

Synthesizing the data: multimodal foundation models in colorectal cancer

The growing body of work in CRC FMs reveals a clear trajectory, progressing from unimodal predictors that excel within isolated data silos to emerging multimodal architectures that attempt to capture the disease in its full biological and clinical complexity. CRC is not defined by any single layer of biological or clinical information. Epigenomic alterations and DNA mutations propagate RNA expression, shape protein signaling, remodel the TME, and ultimately leave recognizable imprints in histological slides and radiological scans [8]. The movement toward MFMs is, in essence, an attempt to mirror this biological cascade within machine learning frameworks.

Early efforts demonstrated the potential of combining molecular and morphological features. Models such as TransSurv fused WSI with transcriptomic and copy number data to improve survival prediction, demonstrating that morphology and genomics together can yield more refined risk stratification than either modality alone [99]. Building on this, Gao and colleagues introduced a bridged fusion strategy capable of imputing missing modalities, representing a critical advance given the prevalence of incomplete clinical datasets [100]. Similarly, CATfusion scaled multimodal integration to encompass histology alongside a broad spectrum of omic data including mRNA, miRNA, methylation, mutations, and copy number, successfully grouping CRC patients into clinically meaningful subtypes [16]. Graph-based frameworks such as MOSEGCN offered an additional perspective by embedding molecular interactions and patient similarity networks, thereby producing predictions that are both robust and biologically interpretable [29]. Collectively, these examples represent early steps toward models that do not merely summarize data but instead attempt to capture the complex web of relationships underlying CRC heterogeneity.

In parallel, pathology and radiology FMs have matured to a point where they can serve as foundational pillars for CRC multimodal integration. Systems such as UNI [8], PLUTO [101], and TITAN [102] have been trained on millions of histopathology images, learning generic features that transfer across diagnostic tasks, while CT-CLIP [103] has pioneered cross-modal alignment between radiological images and textual data. More recent pathology FMs such as MUSK [104], Virchow [105], and PRISM have further expanded capabilities, supporting biomarker prediction and pan-cancer detection at scale. Although not CRC-specific, these imaging models provide the morphological backbone upon which multimodal integration can be constructed.

At the molecular level, omics-specific FMs are likewise converging toward integrative frameworks. Geneformer, scGPT, RNA-FM [106], and tGPT [107] capture cellular states at single-cell or bulk resolution, while molecular language models such as MolFM [108], GenePT [109], and DNABERT-2 [110] demonstrate that DNA, RNA, and protein sequences can be represented in shared vector spaces. Importantly, integrative multimodal approaches are already emerging in adjacent biomedical domains. Models such as BioMedGPT [111], and BiomedCLIP [112] show that image, text, and molecular data can co-exist within one representational framework, while graph-based methods illustrate how molecular knowledge networks can inform multimodal learning. Though developed at pan-cancer or pan-biology scales, these approaches establish architectural templates for the development of CRC-specific MFMs. Recently developed multimodal and clinical-integration models are summarized in [Supplementary Table S7](#).

The evolution from unimodal models to MFMs illustrates how CRC can be represented more faithfully through the alignment of molecular, morphological, and clinical data streams. Yet, the ultimate significance of these models lies not in technical elegance but in their potential to reshape patient care. The next step is to consider how these integrative architectures move from research into practice and how they can inform diagnosis, prognosis, treatment selection, and real-time monitoring in the clinic. Section 5 explores this clinical frontier, where the promise of CRC MFMs is tested against the practical realities of translation, validation, and patient impact.

Table 1 Representative FMs for CRC: CRC-related FMs with their corresponding tasks, applications, and limitations.

Model	Primary modality	Key tasks/CRC applications	Training scale/benchmarks	External validation	Key limitations	Clinical maturity
UNI (2024)	Histopathology (WSIs)	MSI prediction, outcome modeling, robust WSI embeddings	~100 K WSIs / ~100 M patches	Yes (multi-institution)	Domain shift persists; pathology-only	Preclinical (high maturity)
Virchow (2024)	Histopathology (gigapixel)	Pan-cancer detection incl. CRC, MSI prediction	1.5 M WSIs	Yes	Limited multimodality; interpretability challenges	Near-clinical
TITAN (2024)	Vision–language (WSI + reports)	Report-aware WSI retrieval, CRC biomarker prediction	335,645 WSIs with paired/synthetic text	Partial	Text variability reduces alignment	Preclinical
BiomedCLIP (2023)	Radiology + biomedical images + text	Radiology report grounding; CRC figure retrieval	15 M figure–caption pairs	Partial	General biomedical, not CRC-specific	Research
DNABERT-2 (2023)	DNA regulatory genomics	Promoters, enhancers, splice sites relevant to CRC	Multi-species genome corpora	Partial	DNA-only; limited tumor-specific fine-tuning	Research
Enformer (2021)	DNA → gene expression/epigenome	Noncoding variant effect for CRC risk loci	Long-range genomics (~200 kb receptive field)	Strong	Expensive inference; limited tumor context	Preclinical
scGPT / scGPT-spatial (2023–25)	Single-cell (+ spatial)	CRC TME cell states, spatial niches	33 M+ cells	Emerging	Very compute-heavy	Preclinical
Geneformer (2023–25)	Single-cell	Perturbation inference; CRC immune contexture	30 M single cells	Robust	Limited CRC-specific pretraining	Preclinical
CellFM (2025)	Single-cell (RetNet-based)	High-capacity cell-state embeddings for CRC	100 M cells	No	Very large model; limited public checkpoints	Early preclinical
ESM-2 (2022)	Protein sequences	Variant effect prediction, CRC protein function	Up to 15B params	Internal	No proteomics integration	High preclinical
KRONOS (2025)	Spatial proteomics	CRC cell phenotyping, TME architecture	Millions of IMC patches	Partial	Spatial proteomics not routine clinically	Preclinical
PRISM (2025)	Metabolomics (MS/MS)	CRC biomarker identification	>1B spectra	Growing	CRC metabolomics datasets remain small	Research
MGM / MGM 2.0 (2025)	Microbiome (metagenome)	CRC-associated dysbiosis, FMT response simulation	260 K+ microbiome datasets	Strong	Weak tumor-microbiome pairing in CRC	Preclinical
TransSurv (2023)	Multimodal (WSI + RNA + CNA + clinical)	CRC survival prediction	TCGA-CRC	Partial	No prospective evaluation	Translational preclinical
Subtype-Former (2022)	Multi-omics	CRC subtype discovery	TCGA (multi-omics)	No	Sensitive to missing modalities	Research
xTrimoGene (2023)	Pharmacogenomics	CRC drug-response prediction	Multi-omics drug screen datasets	No	In-vitro bias in drug screens	Research
CancerFoundation (2024)	Single-cell + drug response	Predicts chemotherapy response in CRC	Single-cell perturbation atlases	Partial	Lacks wet-lab validation loop	Early translational

The clinical frontier: translating multimodal foundation models into patient care

If MFMs embody the technical convergence of genomics, pathology, radiology, and clinical data, their ultimate value is measured by their potential to improve patient care. In the context of CRC, this potential is particularly compelling. Clinical decision-making remains fragmented: molecular tests identify subgroups such as MSI-H or RAS

mutations, histology informs staging and grading, radiology guides treatment monitoring, and clinical records provide essential contextual information. Each modality provides valuable insight, yet none alone captures the full heterogeneity of CRC. FMs have the potential to unify these perspectives into a single predictive and adaptive framework, thereby reshaping the processes of diagnosis, patient stratification, and treatment selection.

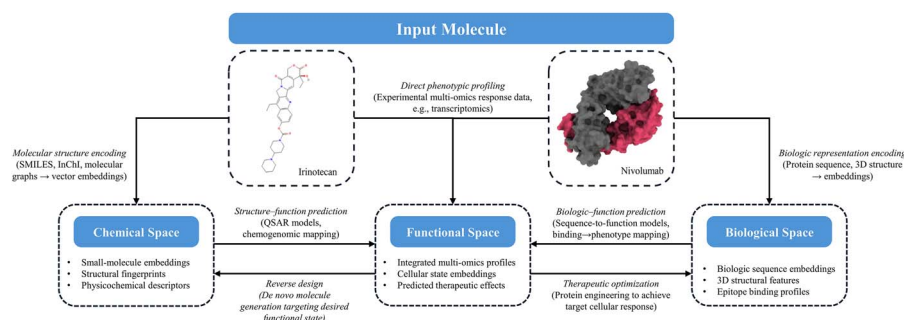


Figure 2 Conceptual framework for therapeutic design using multimodal foundation models (MFMs). Conceptual diagram showing a biological state space where cancer and healthy cell states are mapped, with therapeutic interventions represented as vectors guiding cells from diseased toward healthy states.

AI-driven biomarker discovery: from hypothesis generation to validation

FMs act as hypothesis-generating engines, capturing complex molecular and phenotypic patterns that may be overlooked by traditional analyses. By interrogating internal model representations, for instance, using attention maps to identify which genes, image regions, or pathways drive predictions, researchers can uncover novel biomarker candidates. A notable example is the Gene Swin Transformer, which applied this approach to nominate *PEX10* as a previously unrecognized prognostic gene in CRC, subsequently confirmed through experimental validation [61]. Such frameworks go beyond correlative analyses, enabling AI-guided discovery that is mechanistically grounded and clinically relevant.

Redefining prognostication and molecular subtyping

By integrating histology, multi-omics, and clinical data, MFMs can establish more robust and clinically relevant patient stratification systems than conventional methods. Models such as TransSurv consistently outperform conventional Cox regression models based solely on clinical variables, demonstrating a superior capacity to capture the multi-layered determinants of patient outcome. Likewise, transformer-based clustering frameworks such as Subtype-Former can identify novel molecular subtypes from multi-omics data, potentially outperforming the established CMS classification in predicting prognosis and therapeutic response [113]. Collectively, these advances are reshaping how oncologists conceptualize disease heterogeneity, with direct implications for personalized treatment selection.

Predicting treatment response

Forecasting therapeutic response is a central goal of clinical FMs. In immunotherapy, reliance on single biomarkers, such as MSI status, has proven inadequate to capture the full complexity of tumor-immune interactions. FMs enable the development of composite predictive signatures. Pathology-based systems such as Virchow can infer MSI-H status directly from H&E slides [105], whereas multimodal models like LORIS integrate genomic and clinical features to generate more accurate immune checkpoint inhibitor (ICI) response [114]. Vision-language models such as MUSK further extend this paradigm by integrating pathology images with information extracted from clinical reports, thereby improving predictive accuracy and interpretability [104].

Chemotherapy and targeted therapy pose equally formidable challenges due to the inherent heterogeneity of CRC. scFMs are uniquely positioned to address this complexity by capturing drug responses at the level of individual tumor cells. For example, CancerFoundation can analyze cell-level heterogeneity in drug sensitivity [115]. Applied to large-scale perturbation datasets such as Tahoe-100M, which profiles multiple CRC cell lines treated with hundreds of drugs, these models reveal transcriptomic signatures of sensitivity and resistance. These insights pave the way for highly personalized chemotherapy regimens [116]. Collectively, these advances represent a shift from reliance on static biomarker toward dynamic, model-driven prediction of treatment efficacy across modalities.

Moving in functional space: omics-assisted drug discovery and pharmacological modeling

Predicting prognoses and treatment responses marks a pivotal step forward, but the ultimate promise of MFMs is not just to interpret the state of a tumor, but to change it. This ambition moves the field from prediction toward therapeutic design, reframing pharmacology as a challenge of navigating a high-dimensional ‘functional space’ [117]. Within this paradigm, MFMs develop computational representation of cellular biology where diseased and healthy states are represented as distinct positions on a complex map. Therapeutic interventions, in this framework, become vector-like transformations capable of steering a cell’s trajectory away from a cancerous state toward a homeostatic one. This shift from descriptive correlation to mechanistic simulation exemplifies the true power of these models (Fig. 2).

This process begins with finding the right therapeutic target. Rather than relying on simple correlations, FMs integrate multi-omic data to uncover the causal drivers of disease. Early approaches such as CellOracle demonstrated how gene regulatory networks reconstructed from single-cell data can be perturbed *in silico* to reveal transcription factors critical for fate decisions [118]. Current FMs extend this concept by integrating proteogenomic and signaling data at scale, enabling the identification of central nodes within CRC networks that mediate resistance, such as compensatory pathways downstream of oncogenic KRAS or aberrant WNT/ β -catenin activation [119]. Once a causal target is identified, protein FMs, such as AlphaFold and ESM provide the next crucial step: rapid prediction of a protein’s 3D structure to assess its druggability [120, 121]. For CRC, this is especially important for historically ‘undruggable’ proteins such as β -catenin, where AI-predicted cryptic pockets may reveal novel therapeutic entry points. A model like EMOGI, when applied to CRC data from the CPTAC consortium,

could pinpoint kinases like PAK4 as central to compensatory signaling pathways that emerge under targeted therapy, revealing them as high-confidence combination targets [122]. This capacity to infer network-level causality exemplifies the next generation of therapeutic target discovery.

With a viable target identified, the framework transitions to virtual pharmacology, enabling simulation of drug effects prior to laboratory validation. By training on massive perturbation atlases such as Tahoe-100M, scFMs learn to predict the post-treatment transcriptomic state of a cell. Models such as CancerFoundation extend this approach by simulating cell-level heterogeneity in CRC, predicting not only which subpopulations are likely to respond but also those prone to evolve into resistant clones. For instance, in microsatellite-stable (MSS) CRC organoid models, scFMs trained on perturbation data can anticipate the survival of resistant epithelial subclones under MAPK inhibition, demonstrating the ability of virtual pharmacology to flag relapse trajectories before they manifest clinically. In KRAS-mutant MSS tumors, these simulations can predict persistence of resistant states driven by MAPK pathway feedback, thereby guiding rational design of first-line combination regimens that preempt relapse.

Finally, these computational frameworks accelerate the identification of novel drug synergies. Platforms such as xTrimoGene and MTLsynergy analyze large pharmacogenomic screens to predict potential drug interactions [123, 124]. By embedding small-molecule characteristics alongside multi-omic tumor data, these platforms can prioritize combinations most likely to achieve durable therapeutic responses. In CRC, this is particularly critical for treatment-refractory MSS populations, where synergistic strategies such as combining MAPK inhibitors with microenvironment-modulating agents are urgently needed. Pharmacogenomics and drug synergy prediction models relevant to CRC are summarized in [Supplementary Table S8](#).

By integrating causal target discovery, structural assessment, and virtual pharmacology into a single, cohesive framework, MFMs provide a strategy for designing therapies that are mechanistically grounded, patient-specific, and prospectively testable [125]. This integrative approach has the potential to reduce the high attrition rate in drug development and deliver rational, adaptive treatment strategies directly informed by the molecular logic of CRC.

Closing the loop: experimentally-guided artificial intelligence in colorectal cancer therapeutics

While FMs possess the capability to delineate a high-dimensional functional space and replicate pharmacological interventions, their predictive outcomes cannot solely rely on computational methodologies. The ‘lab-in-the-loop’ paradigm is emerging as an essential conduit between *in silico* design and practical application in the real world. In this approach, hypotheses generated by AI such as novel targets, druggable pockets, or synergistic combinations are continuously tested in experimental systems, with the resulting data fed back to refine and retrain the models. This iterative cycle transforms drug discovery from a linear pipeline into a dynamic feedback system (Fig. 3).

Significant initiatives within this domain exemplify its practicality. Genentech has innovated lab-in-the-loop discovery paradigms wherein robotic automation and artificial intelligence collaboratively engage in the design, testing, and iterative refinement of pharmaceutical candidates within a closed-loop system [126]. In a similar vein, NVIDIA has demonstrated AI-enhanced drug discovery

platforms that are seamlessly integrated with laboratory automation, highlighting the capacity for foundational models to adapt in near real-time to experimental results [127]. By integrating multi-faceted models with organoid cultures, patient-derived xenografts (PDXs), and high-throughput perturbation assays, CRC research stands to gain distinct advantages from such adaptive cycles. Tumors exhibiting heterogeneous resistance mechanisms, such as KRAS-mutant MSS CRC, serve as optimal test cases where iterative AI-guided design could swiftly lead to the development of rational combination therapies.

For CRC, embedding a lab-in-the-loop framework means that each computational prediction is not an endpoint, but rather the beginning of a refinement process. By aligning virtual pharmacology with automated experimentation, these systems not only accelerate discovery but also enhance trustworthiness by ensuring that AI-driven hypotheses are immediately confronted with biological reality. In doing so, lab-in-the-loop discovery may represent the most tangible pathway through which MFMs can directly shape therapeutic innovation in oncology.

The path to trustworthy artificial intelligence: challenges and a call for critical evaluation

Despite their immense promise, the path to clinical translation is paved with significant challenges spanning data, modeling, and implementation. Addressing these hurdles is essential for realizing the potential of FMs in routine cancer care. The unresolved domains discussed here are not merely technical limitations but represent fundamental bottlenecks that currently prevent MFMs from closing the loop between computational prediction and actionable clinical intervention.

Data-related challenges, generalizability, and scale contention

The performance of any FM is fundamentally limited by the quality and scale of its training data, especially when aiming for generalized clinical utility.

- **Data scarcity and harmonization:** while TCGA is a valuable resource, it lacks sufficient data on rarer but clinically important CRC subtypes. Data generated across different institutions frequently suffer from technical variation (e.g. inconsistent H&E staining protocols) leading to impaired model generalization. This is particularly critical for subtle features like assessing tumor budding at the invasive front of CRC, a key prognostic marker [128].
- **Generalizability and calibration:** the ‘Foundation’ concept implies broad utility, but robust performance across lower-resource global settings remains largely untested. Despite encouraging performance metrics reported in most studies, there remains a notable paucity of negative, null, or non-replicable findings in the literature. This publication bias limits fair benchmarking, inflates perceived generalizability, and obscures scenarios in which model performance degrades (e.g. domain shift, rare subgroups, low-resource cohorts). Transparent reporting of failures, misclassifications, and poor transferability is therefore essential for a realistic assessment of clinical readiness.

For example, while some DL models predict MSI from WSIs with strong in-domain performance, multi-center external validation has

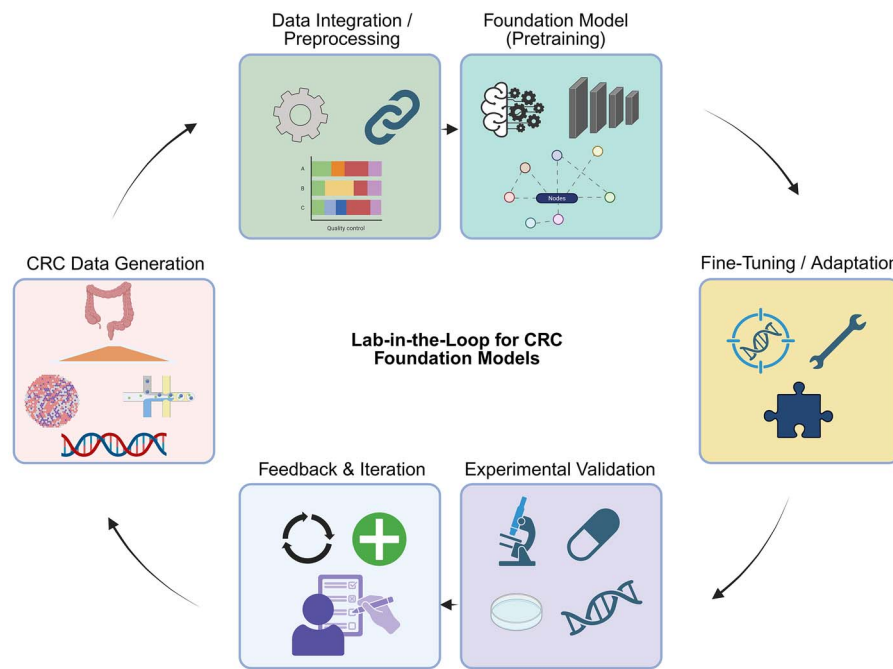


Figure 3 The lab-in-the-loop paradigm for AI-driven biomedical discovery. Diagram showing a feedback loop between AI models and laboratory experiments, where AI-generated hypotheses are tested experimentally and the results are used to improve the models.

revealed substantial gaps. In a self-supervised, attention-based multi-instance learning study, clinical-grade performance was achieved for MSI and BRAF, while KRAS and NRAS remained ‘clinically insufficient’ in external cohorts [129]. Another hybrid model combining pathology and clinical features achieved an AUC of 0.928 on its internal cohort (TCGA) but dropped to 0.811 in an external cohort [130]. Moreover, systematic reviews have consistently flagged the lack of external validation as a major limitation: many AI-based MSI/KRAS/BRAF prediction studies do not test on independent datasets [131]. Even beyond mutation prediction, a recent DL staging model showed AUC decline from 0.889 internally to 0.700 on external TCGA samples, underscoring how domain-shift and cohort heterogeneity challenge real-world deployment [132]. As FMs move toward clinical integration, rigorous validation across diverse populations, institutions, sequencing or imaging platforms, and disease stages will be critical to accurately determine robustness and true translational potential.

- **Contention on scale:** it remains unresolved whether billion-parameter, generic MFMs are required for CRC, or whether smaller, specialized multimodal expert models (MEMs) could achieve superior performance on specific clinical endpoints (e.g. MSI prediction) at a fraction of the cost.
- **Privacy and governance:** the use of sensitive patient data requires robust governance. Techniques like federated learning offer a promising solution to train a shared CRC prognostication model across hospitals in different countries without sharing raw patient data [133].

Model-related challenges: trust, transparency, and algorithmic bias

The complexity of FMs introduces difficulties related to interpretation, algorithmic integrity, and capturing the dynamic nature of disease.

- **The challenge of causal and dynamic inference:** current MFMs primarily function as sophisticated predictive models, often excelling at correlation while remaining limited in causal inference (e.g. determining ‘why’ a tumor feature drives resistance). Most MFMs are trained on static patient snapshots, limiting their ability to model the dynamic, evolving nature of CRC such as metastasis initiation, immune evasion, and clonal evolution under therapy.
- **Interpretable rationale versus explainable feature:** The ‘black box’ nature is a major barrier. Current eXplainable AI (XAI) methods often highlight ‘where’ the model looked (patch-level features) but rarely provide an actionable biological rationale that aligns with physician intuition and regulatory expectations (e.g. linking prediction to known pathway biology) [134, 135]. This partial interpretability raises the risk of automation bias and clinician over-reliance.
- **Robustness and bias:** rigorous external validation on independent, multi-institutional, and prospective cohorts is the gold standard [136]. If training data is not representative, models can exacerbate health disparities. For example, an AI model for CRC detection trained primarily on light-skinned patient colonoscopy images may exhibit reduced sensitivity for detecting subtle flat polyps in patients with darker skin tones [137].
- **Epigenomic integration:** the current MFM landscape is image- and sequence-dominant. The vital role of epigenomics (e.g. DNA methylation) in CRC plasticity and CIMP classification is under-represented. Integrating sparse, high-dimensional regulatory data with dense histology or radiology signals remains a major technical challenge.

Clinical and regulatory translation

Bridging the gap from a research model to a clinical tool involves overcoming formidable implementation and approval hurdles.

- Regulatory pathways: AI tools are medical devices requiring regulatory approval. The path for ‘continually learning’ models that adapt over time, which would be ideal for dynamic CRC risk models, is still being defined and presents a major regulatory challenge.
- Workflow integration: a model must fit seamlessly into the clinic. A successful CRC pathology AI tool must integrate with the laboratory information system (LIS), present results within the pathologist’s existing digital viewer, and allow them to seamlessly accept, reject, or modify the AI’s findings in their final report.
- Cost-effectiveness and reimbursement: for adoption, these tools must demonstrate value. An AI that predicts MSI status from a routine H&E slide could be highly cost-effective by reducing the number of patients who need expensive follow-up IHC and PCR testing, but this will require formal health economic evaluations to secure reimbursement from payers.

Future directions and the research agenda

The field is moving rapidly, with several key trends shaping the next generation of models that aim to resolve the current challenges.

- Architectural innovation: the rise of highly efficient architectures like state space models may enable the processing of entire genomes as a single input, potentially capturing
- complex structural variants and long-range epigenetic effects that are missed by current models in CRC [33, 138].
- The ‘Foundationless’ debate and critical evaluation: this debate highlights the need to rethink pre-training strategies and benchmarks for genomics, moving toward more biologically meaningful objectives, as some smaller, well-designed models can outperform massive pre-trained models for certain tasks [139].
- Continual learning for evolving data: future models will need to incorporate continual learning strategies, such as instruction-based knowledge distillation, to evolve over time without constant retraining from scratch [140], overcoming the ‘catastrophic forgetting’ of previous knowledge when updated with new clinical data.
- Causal inference and digital twins: moving beyond correlation to causation is a critical next step. Causal DL, using frameworks like do-calculus integrated with genomic data, aims to differentiate genes that are merely associated with CRC prognosis from those that mechanistically drive it [141]. This paves the way for the ultimate vision: a ‘digital twin’ of a patient’s tumor, used to simulate its response to different chemotherapy regimens [142].

Conclusion

MFMs represent more than an incremental advance in CRC; they signal a paradigm shift. By capturing the multilayered interplay of epigenomic and genomic alterations, transcriptomic programs, TME dynamics, and imaging features, MFMs hold the potential to unify previously fragmented approaches to diagnosis and therapy. CRC, with its well-defined yet heterogeneous subtypes such as MSI-H and MSS tumors, provides a uniquely challenging proving ground where the generalizability of oncology FMs can be rigorously tested.

The promise of these models extends beyond prognostic stratification. By embedding disease states within high-dimensional functional spaces, MFMs enable a new computational pharmacology where therapeutic interventions are modeled as transformations of cellular

trajectories. This reorientation from static biomarkers to dynamic, mechanistic simulations offers the possibility of rational, adaptive treatment design. The lab-in-the-loop paradigm further accelerates translation by iteratively coupling computational hypotheses with experimental validation, ensuring that AI-driven insights remain grounded in biological reality.

For such advances to impact patient care, trustworthiness must be treated as a cornerstone rather than an afterthought. Fairness, transparency, robustness, and accountability are essential for building confidence among clinicians and patients alike. Only models that can withstand rigorous external validation, explain their reasoning in clinically interpretable terms, and avoid exacerbating health inequities will be acceptable for integration into oncology practice.

Realizing this vision requires deep interdisciplinary collaboration. Computer scientists, systems biologists, experimental researchers, and clinicians must converge on shared frameworks that support both technical innovation and clinical feasibility. By embedding trustworthy AI within a lab-in-the-loop translational ecosystem, CRC research can set the precedent for how MFMs are deployed across oncology. The challenge now is not only to refine the models but to build the collaborative infrastructure that will carry them from computational elegance to clinical impact.

Key Points

- Colorectal cancer (CRC) is marked by profound molecular and clinical heterogeneity that limits the utility of single-modality biomarkers; multimodal foundation models (MFMs) provide a unified framework to capture this complexity.
- By integrating histopathology, radiology, multi-omics, and clinical records, MFMs enable accurate prognostication, molecular subtyping, and treatment response prediction that surpass conventional approaches.
- Beyond prediction, MFMs can simulate disease dynamics and therapeutic perturbations within high-dimensional ‘functional spaces’, supporting rational drug discovery and adaptive treatment design.
- The lab-in-the-loop paradigm, coupling artificial intelligence predictions with experimental validation, is central to building trustworthy, clinically actionable MFMs in CRC.
- Trustworthy translation requires interpretability, equity, and rigorous multi-institutional validation to ensure safe integration of MFMs into oncology practice.

Acknowledgements

We would like to thank the Clinical Research Development Unit, Imam Reza Hospital, Mashhad University of Medical Sciences, for their assistance in this manuscript.

Author contributions

Hossein Safarpour (Conceived the idea of the review and planned the outline, Revised and edited the manuscript, Contributed to discussions and approved the final version of the manuscript), Danial Gharai Amirabadi (Performed the literature search and wrote the initial draft under the supervision of H.S Contributed to discussions and approved the final version of the manuscript), and Adib Miraki Feriz (Performed

the literature search and wrote the initial draft under the supervision of H.S. Contributed to discussions and approved the final version of the manuscript)

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflicts of interest

The authors declare no competing interests.

Funding

None declared.

Data availability

No new data were generated or analyzed in this study. All information presented is derived from previously published studies cited within the manuscript.

Organization description

Cellular and molecular research center

The Cellular and Molecular Research Center at Birjand University of Medical Sciences conducts interdisciplinary research in molecular biology, cancer research, and translational medicine, with a focus on integrating experimental and computational approaches.

References

1. Sung H, Ferlay J, Siegel RL *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;**71**:209–49. <https://doi.org/10.3322/caac.21660>
2. Brunac AC, Fourquet J, Perot G *et al*. CINSARC signature outperforms gold-standard TNM staging and consensus molecular subtypes for clinical outcome in stage II–III colorectal carcinoma. *Mod Pathol* 2022;**35**:2002–10. <https://doi.org/10.1038/s41379-022-01166-9>
3. Venderbosch S, Nagtegaal ID, Maughan TS *et al*. Mismatch repair status and BRAF mutation status in metastatic colorectal cancer patients: a pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies. *Clin Cancer Res* 2014;**20**:5322–30. <https://doi.org/10.1158/1078-0432.CCR-14-0332>
4. Le DT, Uram JN, Wang H *et al*. PD-1 blockade in Tumors with mismatch-repair deficiency. *N Engl J Med* 2015;**372**:2509–20. <https://doi.org/10.1056/NEJMoa1500596>
5. Boussios S, Ozturk MA, Moschetta M *et al*. The developing story of predictive biomarkers in colorectal cancer. *J Pers Med* 2019;**9**:12. <https://doi.org/10.3390/jpm9010012>
6. Marisa L, de Reynies A, Duval A *et al*. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013;**10**:e1001453. <https://doi.org/10.1371/journal.pmed.1001453>
7. Phipps AI, Limburg PJ, Baron JA *et al*. Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology* 2015;**148**:77–87.e2. <https://doi.org/10.1053/j.gastro.2014.09.038>
8. Chen RJ, Ding T, Lu MY *et al*. Towards a general-purpose foundation model for computational pathology. *Nat Med* 2024;**30**:850–62. <https://doi.org/10.1038/s41591-024-02857-3>
9. Deisboeck TS, Wang Z, Macklin P *et al*. Multiscale cancer modeling. *Annu Rev Biomed Eng* 2011;**13**, 2011:127–55. <https://doi.org/10.1146/annurev-bioeng-071910-124729>
10. Deisboeck TS, Zhang L, Martin S. Advancing cancer systems biology: introducing the center for the development of a virtual tumor. *CViT Cancer Inform* 2007;**5**:1–8. <https://doi.org/10.1177/117693510700500001>
11. Bommasani R, Hudson DA, Adeli E *et al*. On the opportunities and risks of foundation models. arXiv preprint arXiv:210807258. 2021.
12. Nerella S, Bandyopadhyay S, Zhang J *et al*. Transformers and large language models in healthcare: a review. *Artif Intell Med* 2024;**154**:102900. <https://doi.org/10.1016/j.artmed.2024.102900>
13. Khan W, Leem S, See KB *et al*. A comprehensive survey of foundation models in medicine. *IEEE Rev Biomed Eng* 2025;**19**:283–304. <https://doi.org/10.1109/RBME.2025.3531360>
14. Xu H, Usuyama N, Bagga J *et al*. A whole-slide foundation model for digital pathology from real-world data. *Nature*. 2024;**630**:181–8. <https://doi.org/10.1038/s41586-024-07441-w>
15. Vanguri RS, Luo J, Aukerman AT *et al*. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer* 2022;**3**:1151–64. <https://doi.org/10.1038/s43018-022-00416-8>
16. Hu Y, Li X, Yi Y *et al*. Deep learning-driven survival prediction in pan-cancer studies by integrating multimodal histology-genomic data. *Brief Bioinform* 2025;**26**:bbaf121. <https://doi.org/10.1093/bib/bbaf121>
17. Lee J, Yoon W, Kim S *et al*. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020;**36**:1234–40. <https://doi.org/10.1093/bioinformatics/btz682>
18. Ji Y, Zhou Z, Liu H *et al*. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics* 2021;**37**:2112–20. <https://doi.org/10.1093/bioinformatics/btab083>
19. Brandes N, Ofer D, Peleg Y *et al*. ProteinBERT: a universal deep-learning model of protein sequence and function. *Bioinformatics* 2022;**38**:2102–10. <https://doi.org/10.1093/bioinformatics/btac020>
20. Ferruz N, Schmidt S, Hocker B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat Commun* 2022;**13**:4348. <https://doi.org/10.1038/s41467-022-32007-7>
21. Madani A, B MC, Naik N *et al*. Progen: Language modeling for protein generation. arXiv preprint arXiv:200403497. 2020.
22. Lin Z, Akin H, Rao R *et al*. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *BioRxiv* 2022;**2022**:500902.
23. Lin Z, Akin H, Rao R *et al*. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;**379**:1123–30. <https://doi.org/10.1126/science.ade2574>
24. Vaswani A, Shazeer N, Parmar N *et al*. Attention is all you need. *Adv Neural Inf Proces Syst* 2017;**30**:5998–6008.
25. Xu H, Xu Q, Cong F *et al*. Vision transformers for computational histopathology. *IEEE Rev Biomed Eng* 2024;**17**:63–79. <https://doi.org/10.1109/RBME.2023.3297604>
26. Wagner SJ, Reisenbuchler D, West NP *et al*. Transformer-based biomarker prediction from colorectal cancer histology: a

- large-scale multicentric study. *Cancer Cell* 2023;**41**:1650–1661.e4 e4. <https://doi.org/10.1016/j.ccell.2023.08.002>
27. Guo B, Li X, Yang M *et al.* Predicting microsatellite instability and key biomarkers in colorectal cancer from H&E-stained images: achieving state-of-the-art predictive performance with fewer data using Swin transformer. *J Pathol Clin Res* 2023;**9**:223–35. <https://doi.org/10.1002/cjp2.312>
 28. Liu Z, Lin Y, Cao Y *et al.* Swin transformer: hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 9992–10002.
 29. Wang J, Liao N, Du X *et al.* A semi-supervised approach for the integration of multi-omics data based on transformer multi-head self-attention mechanism and graph convolutional networks. *BMC Genomics* 2024;**25**:86. <https://doi.org/10.1186/s12864-024-09985-7>
 30. Schaar AC, Tejada-Lapuerta A, Palla G *et al.* Nicheformer: a foundation model for single-cell and spatial omics. *Nature Methods* 2025;**22**:2525–38. <https://doi.org/10.1101/2024.04.15.589472>
 31. Qureshi S. Limits of depth: over-smoothing and over-squashing in gnns. *Big Data Mining and Analytics* 2023;**7**:205–16.
 32. Wang Y, Hou W, Sheng N *et al.* Graph pooling in graph neural networks: methods and their applications in omics studies. *Artificial Intelligence Review* 2024;**57**:294. <https://doi.org/10.1007/s10462-024-10918-9>
 33. Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752. 2023.
 34. Zhao Q, Zhang C, Zhang W. dnaGrinder: a lightweight and high-capacity genomic foundation model. arXiv preprint arXiv:240915697. 2024.
 35. Wang P, Cai R, Wang Y *et al.* Understanding and mitigating bottlenecks of state space models through the lens of recency and over-smoothing. arXiv preprint arXiv:250100658. 2024.
 36. Popov M, Kallala A, Ramesh A *et al.* Leveraging state space models in long range genomics. arXiv preprint arXiv:250406304. 2025.
 37. Avsec Z, Agarwal V, Visentin D *et al.* Effective gene expression prediction from sequence by integrating long-range interactions. *Nat Methods* 2021;**18**:1196–203. <https://doi.org/10.1038/s41592-021-01252-x>
 38. AlSaad R, Abd-Alrazaq A, Boughorbel S *et al.* Multimodal large language models in health care: applications, challenges, and future outlook. *J Med Internet Res* 2024;**26**:e59505. <https://doi.org/10.2196/59505>
 39. Li B, Nabavi S. A multimodal graph neural network framework for cancer molecular subtype classification. *BMC Bioinformatics* 2024;**25**:27. <https://doi.org/10.1186/s12859-023-05622-4>
 40. Rahmanian M, Mansoori EG. MoVAE: multi-omics variational auto-encoder for cancer subtype detection. *IEEE Access* 2024;**12**:133617–31. <https://doi.org/10.1109/ACCESS.2024.3462543>
 41. Wang T, Mai D, Shu H *et al.* Enhancing cell subpopulation discovery in cancer by integrating single-cell transcriptome and expressed variants. *Fundam Res* 2025;**6**:88–98. <https://doi.org/10.1016/j.fmre.2025.01.001>
 42. de Jong ED, Marcus E, Teuwen J. Current pathology foundation models are unrobust to medical center differences. arXiv preprint arXiv:250118055. 2025.
 43. Tizhoosh HR. Why foundation models in pathology are failing. arXiv preprint arXiv:251023807. 2025.
 44. Waqas A, Naveed J, Shahnawaz W *et al.* Digital pathology and multimodal learning on oncology data. *BJR|artificial. Intelligence* 2024;**1**:ubae014. <https://doi.org/10.1093/bjrai/ubae014>
 45. Hua S, Yan F, Shen T *et al.* PathoDuet: foundation models for pathological slide analysis of H&E and IHC stains. *Med Image Anal* 2024;**97**:103289. <https://doi.org/10.1016/j.media.2024.103289>
 46. Wang CW, Muzakky H, Firdi NP *et al.* Deep learning to assess microsatellite instability directly from histopathological whole slide images in endometrial cancer. *NPJ Digit Med* 2024;**7**:143. <https://doi.org/10.1038/s41746-024-01131-7>
 47. Faa G, Coghe F, Pretta A *et al.* Artificial intelligence models for the detection of microsatellite instability from whole-slide imaging of colorectal cancer. *Diagnostics (Basel)* 2024;**14**:1605. <https://doi.org/10.3390/diagnostics14151605>
 48. Li H, Qin J, Li Z *et al.* Systematic review and meta-analysis of deep learning for MSI-H in colorectal cancer whole slide images. *NPJ Digit Med* 2025;**8**:456. <https://doi.org/10.1038/s41746-025-01848-z>
 49. Ben Hamida A, Devanne M, Weber J *et al.* Deep learning for colon cancer histopathological images analysis. *Comput Biol Med* 2021;**136**:104730. <https://doi.org/10.1016/j.combiomed.2021.104730>
 50. Raju ASN, Venkatesh K, Rajababu M *et al.* Colorectal cancer unmasked: a synergistic AI framework for hyper-granular image dissection, precision segmentation, and automated diagnosis. *BMC Med Imaging* 2025;**25**:283. <https://doi.org/10.1186/s12880-025-01826-7>
 51. Selvaraj J, Umapathy S, Rajesh NA. Artificial intelligence based real time colorectal cancer screening study: polyp segmentation and classification using multi-house database. *Biomed Signal Process Control* 2025;**99**:106928. <https://doi.org/10.1016/j.bspc.2024.106928>
 52. Kapuria S, Ikoma N, Chinchali S *et al.* Synthetic data-augmented explainable vision transformer for colorectal cancer diagnosis via surface tactile imaging. *Eng Appl Artif Intel* 2025;**151**:110633. <https://doi.org/10.1016/j.engappai.2025.110633>
 53. Raju ASN, Venkatesh K, Gatla RK *et al.* Expedited colorectal cancer detection through a dexterous hybrid CADx system with enhanced image processing and augmented polyp visualization. *IEEE Access* 2025;**13**:17524–53. <https://doi.org/10.1109/ACCESS.2025.3532807>
 54. Wu E, Bieniosek M, Wu Z *et al.* ROSIE: AI generation of multiplex immunofluorescence staining from histopathology images. *Nat Commun* 2025;**16**:7633. <https://doi.org/10.1038/s41467-025-62346-0>
 55. Hille G, Tummala P, Spitz L *et al.* Transformers for colorectal cancer segmentation in CT imaging. *Int J Comput Assist Radiol Surg* 2024;**19**:2079–87. <https://doi.org/10.1007/s11548-024-03217-9>
 56. Gulmez B. Deep learning based colorectal cancer detection in medical images: a comprehensive analysis of datasets, methods, and future directions. *Clin Imaging* 2025;**125**:110542. <https://doi.org/10.1016/j.clinimag.2025.110542>
 57. Lu MY, Williamson DFK, Chen TY *et al.* Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021;**5**:555–70. <https://doi.org/10.1038/s41551-020-00682-w>
 58. Cornish AJ, Gruber AJ, Kinnersley B *et al.* The genomic landscape of 2,023 colorectal cancers. *Nature* 2024;**633**:127–36. <https://doi.org/10.1038/s41586-024-07747-9>
 59. Berner AM, Murugaesu N. The evolving role of genomics in colorectal cancer. *Clin Oncol (R Coll Radiol)* 2025;**37**:103661. <https://doi.org/10.1016/j.clon.2024.10.033>
 60. Guinney J, Dienstmann R, Wang X *et al.* The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015;**21**:1350–6. <https://doi.org/10.1038/nm.3967>

61. Wang Y, Yue X, Lou S *et al.* Gene Swin transformer: new deep learning method for colorectal cancer prognosis using transcriptomic data. *Brief Bioinform* 2025;**26**:bbaf275. <https://doi.org/10.1093/bib/bbaf275>
62. Theodoris CV, Xiao L, Chopra A *et al.* Transfer learning enables predictions in network biology. *Nature* 2023;**618**:616–24. <https://doi.org/10.1038/s41586-023-06139-9>
63. Cui H, Wang C, Maan H *et al.* scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods* 2024;**21**:1470–80. <https://doi.org/10.1038/s41592-024-02201-0>
64. Yang F, Wang W, Wang F *et al.* scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat Mach Intell* 2022;**4**:852–66. <https://doi.org/10.1038/s42256-022-00534-z>
65. Levine D, Rizvi SA, Levy S *et al.* Cell2Sentence: teaching large language models the language of biology. *bioRxiv* 2024:2023.09.11.557287. <https://doi.org/10.1101/2023.09.11.557287>
66. Kalfon J, Samaran J, Peyre G *et al.* scPRINT: pre-training on 50 million cells allows robust gene network predictions. *Nat Commun* 2025;**16**:3607. <https://doi.org/10.1038/s41467-025-58699-1>
67. Pearce JD, Simmonds SE, Mahmoudabadi G *et al.* A cross-species generative cell atlas across 1.5 billion years of evolution: the TranscriptFormer single-cell model. *bioRxiv* 2025:2025.04.25.650731. <https://doi.org/10.1101/2025.04.25.650731>
68. Yuan X, Zhan Z, Zhang Z *et al.* Cell ontology guided transcriptome foundation model. *Adv Neural Inf Proces Syst* 2024;**37**:6323–66.
69. Aldera AP, Cifci D, Veldhuizen GP *et al.* Deep learning predicts microsatellite instability status in colorectal carcinoma in an ethnically heterogeneous population in South Africa. *J Clin Pathol* 2025;**79**:50–6. <https://doi.org/10.1136/jcp-2025-210053>
70. Hao M, Gong J, Zeng X *et al.* Large-scale foundation model on single-cell transcriptomics. *Nat Methods* 2024;**21**:1481–91. <https://doi.org/10.1038/s41592-024-02305-7>
71. Wang C, Cui H, Zhang A *et al.* scGPT-spatial: continual pretraining of single-cell foundation model for spatial transcriptomics. *bioRxiv* 2025:2025.02.05.636714. <https://doi.org/10.1101/2025.02.05.636714>
72. Sherif ZA, Ogunwobi OO, Ransom HW. Mechanisms and technologies in cancer epigenetics. *Front Oncol* 2024;**14**:1513654. <https://doi.org/10.3389/fonc.2024.1513654>
73. Mojarad EN, Kuppen PJ, Aghdaei HA *et al.* The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol-ogy Bed Bench* 2013;**6**:120.
74. Ying K, Song J, Cui H *et al.* MethylGPT: a foundation model for the DNA methylome. *bioRxiv* 2024. <https://doi.org/10.1101/2024.10.30.621013>
75. de Lima Camillo LP, Sehgal R, Armstrong J *et al.* CpGPT: a foundation model for DNA methylation. *bioRxiv* 2024. 2024.10. 24.619766
76. Tanaka A, Ogawa M, Zhou Y *et al.* Proteogenomic characterization of primary colorectal cancer and metastatic progression identifies proteome-based subtypes and signatures. *Cell Rep* 2024;**43**:113810. <https://doi.org/10.1016/j.celrep.2024.113810>
77. Li L, Jiang D, Liu H *et al.* Comprehensive proteogenomic profiling reveals the molecular characteristics of colorectal cancer at distinct stages of progression. *Cancer Res* 2024;**84**:2888–910. <https://doi.org/10.1158/0008-5472.CAN-23-1878>
78. Shaban M, Chang Y, Qiu H *et al.* A foundation model for spatial proteomics. arXiv preprint arXiv:250603373. 2025.
79. Qiu Y, Cai G, Zhou B *et al.* A distinct metabolic signature of human colorectal cancer with prognostic potential. *Clin Cancer Res* 2014;**20**:2136–46. <https://doi.org/10.1158/1078-0432.CCR-13-1939>
80. Liu Y, Zhang F, Ge Y *et al.* Application of LLMs/transformer-based models for metabolite annotation in metabolomics. *Health Metab* 2025;**7**. <https://doi.org/10.53941/hm.2025.100014>
81. Chen Y, Wang B, Zhao Y *et al.* Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. *Nat Commun* 2024;**15**:1657. <https://doi.org/10.1038/s41467-024-46043-y>
82. Bushuiev R, Bushuiev A, Samusevich R *et al.* Self-supervised learning of molecular representations from millions of tandem mass spectra using DreaMS. *Nat Biotechnol* 2025;**44**:630–40. <https://doi.org/10.1038/s41587-025-02663-3>
83. Healey D, Domingo-Fernández D, Taylor J *et al.* PRISM: a foundation model for life's chemistry. 2025. [Available from: <https://enveda.com/prism-a-foundation-model-for-lifes-chemistry/>]
84. Asher G, Campbell JM, Geremia J *et al.* LSM1-MS2: a self-supervised foundation model for tandem mass spectrometry applications, encompassing extensive chemical property predictions and spectral matching. *ChemRxiv* 2024. <https://doi.org/10.26434/chemrxiv-2024-k06gb>
85. Xue J, Wang B, Ji H *et al.* RT-transformer: retention time prediction for metabolite annotation to assist in metabolite identification. *Bioinformatics* 2024;**40**:btac084. <https://doi.org/10.1093/bioinformatics/btac084>
86. Young A, Wang B, Röst H. Massformer: tandem mass spectrum prediction for small molecules using graph transformers. arXiv preprint arXiv:211104824. 2021.
87. Voronov G, Frandsen A, Bargh B *et al.* MS2Prop: a machine learning model that directly predicts chemical properties from mass spectrometry data for novel compounds. *bioRxiv* 2022;**10**. (2022.10):09.511482.
88. Baygi SF, Barupal DK. IDSL_MINT: a deep learning framework to predict molecular fingerprints from mass spectra. *J Chem* 2024;**16**:8. <https://doi.org/10.1186/s13321-024-00804-5>
89. Goldman SL. *Machine Learning Methods for Discovering Metabolite Structures from Mass Spectra*. Cambridge, MA, USA: Massachusetts Institute of Technology, 2024.
90. Würf V, Köhler N, Molnar F *et al.* LipiDetective a deep learning model for the identification of molecular lipid species in tandem mass spectra. *bioRxiv* 2024. 2024.10. 07.617094.
91. Butler T, Frandsen A, Lighthouse R *et al.* MS2Mol: a transformer model for illuminating dark chemical space from mass spectra. *ChemRxiv* 2023. <https://doi.org/10.26434/chemrxiv-2023-vsmvx-v3>
92. Liu Z, Sun Y, Li Y *et al.* An explainable graph neural framework to identify cancer-associated intratumoral microbial communities. *Adv Sci (Weinh)* 2024;**11**:e2403393. <https://doi.org/10.1002/adv.202403393>
93. Greathouse KL, Choudhury A. Precision nutrition and the gut microbiome: harnessing AI to revolutionize cancer prevention and therapy. *Cell Host Microbe* 2025;**33**:766–76. <https://doi.org/10.1016/j.chom.2025.05.011>
94. Mechttersheimer D, Ding W, Xu X *et al.* IMPACT: interpretable microbial phenotype analysis via microbial characteristic traits. *Bioinformatics* 2024;**41**:btac702. <https://doi.org/10.1093/bioinformatics/btac702>
95. Wang J, Cong Y, Tang B *et al.* Integrative analysis of multi-omics data and gut microbiota composition reveals prognostic subtypes and predicts immunotherapy response in colorectal cancer using

- machine learning. *Sci Rep* 2025;**15**:25268. <https://doi.org/10.1038/s41598-025-08915-1>
96. Han J, Zhang H, Ning K. Techniques for learning and transferring knowledge for microbiome-based classification and prediction: review and assessment. *Brief Bioinform* 2025;**26**:bbaf015. <https://doi.org/10.1093/bib/bbaf015>
 97. Zhang H, Zhang Y, Kang Z *et al*. MGM as a large-scale pretrained foundation model for microbiome analyses in diverse contexts. *bioRxiv* 2025. 2024.12. 30.630825.
 98. Zhang H, Kang Z, Zhang Y *et al*. Towards a generative paradigm for large-scale microbiome analysis by generative language model. *bioRxiv* 2025:2025.01.15.633278. <https://doi.org/10.1101/2025.01.15.633278>
 99. Lv Z, Lin Y, Yan R *et al*. TransSurv: transformer-based survival analysis model integrating histopathological images and genomic data for colorectal cancer. *IEEE/ACM Trans Comput Biol Bioinform* 2023;**20**:3411–20. <https://doi.org/10.1109/TCBB.2022.3199244>
 100. Gao F, Ding J, Gai B *et al*. Interpretable multimodal fusion model for bridged histology and genomics survival prediction in pancreatic cancer. *Adv Sci (Weinh)* 2025;**12**:e2407060. <https://doi.org/10.1002/adv.202407060>
 101. Juyal D, Padigela H, Shah C *et al*. Pluto: pathology-universal transformer. arXiv preprint arXiv:240507905. 2024.
 102. Ding T, Wagner SJ, Song AH *et al*. Multimodal whole slide foundation model for pathology. arXiv preprint arXiv:241119666. 2024.
 103. Hamamci IE, Er S, Wang C *et al*. Developing generalist foundation models from a multimodal dataset for 3d computed tomography. arXiv preprint arXiv:240317834. 2024.
 104. Xiang J, Wang X, Zhang X *et al*. A vision-language foundation model for precision oncology. *Nature* 2025;**638**:769–78. <https://doi.org/10.1038/s41586-024-08378-w>
 105. Vorontsov E, Bozkurt A, Casson A *et al*. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nat Med* 2024;**30**:2924–35. <https://doi.org/10.1038/s41591-024-03141-0>
 106. Chen J, Hu Z, Sun S *et al*. Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. arXiv preprint arXiv:220400300. 2022.
 107. Shen H, Liu J, Hu J *et al*. Generative pretraining from large-scale transcriptomes for single-cell deciphering. *iScience* 2023;**26**:106536. <https://doi.org/10.1016/j.isci.2023.106536>
 108. Luo Y, Yang K, Hong M *et al*. Molfm: a multimodal molecular foundation model. arXiv preprint arXiv:230709484. 2023.
 109. Chen Y, Zou J. GenePT: a simple but effective foundation model for genes and cells built from ChatGPT. *bioRxiv* 2024:2023.10.16.562533. <https://doi.org/10.1101/2023.10.16.562533>
 110. Zhou Z, Ji Y, Li W *et al*. Dnabert-2: efficient foundation model and benchmark for multi-species genome. arXiv preprint arXiv:230615006. 2023.
 111. Luo Y, Zhang J, Fan S *et al*. Biomedgpt: open multimodal generative pre-trained transformer for biomedicine. arXiv preprint arXiv:230809442. 2023.
 112. Zhang S, Xu Y, Usuyama N *et al*. Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv preprint arXiv:230300915. 2023.
 113. Yang H, Sheng Y, Jiang Y *et al*. Subtype-former: a deep learning approach for cancer subtype discovery with multi-omics data. arXiv preprint arXiv:220714639. 2022.
 114. Chang TG, Cao Y, Sfreddo HJ *et al*. LORIS robustly predicts patient outcomes with immune checkpoint blockade therapy using common clinical, pathologic and genomic features. *Nat Cancer* 2024;**5**:1158–75. <https://doi.org/10.1038/s43018-024-00772-7>
 115. Theus A, Barkmann F, Wissel D *et al*. CancerFoundation: a single-cell RNA sequencing foundation model to decipher drug resistance in cancer. *bioRxiv* 2024. 2024.11. 01.621087.
 116. Zhang J, Ubas AA, de Borja R *et al*. Tahoe-100m: a giga-scale single-cell perturbation atlas for context-dependent gene function and cellular modeling. *BioRxiv* 2025. 2025.02. 20.639398.
 117. Bitencourt-Ferreira G, Villarreal MA, Quiroga R *et al*. Exploring scoring function space: developing computational models for drug discovery. *Curr Med Chem* 2024;**31**:2361–77. <https://doi.org/10.2174/0929867330666230321103731>
 118. Kamimoto K, Stringa B, Hoffmann CM *et al*. Dissecting cell identity via network inference and *in silico* gene perturbation. *Nature* 2023;**614**:742–51. <https://doi.org/10.1038/s41586-022-05688-9>
 119. Muneer A, Waqas M, Saad MB *et al*. From classical machine learning to emerging foundation models: review on multimodal data integration for cancer research. *Artif Intell Rev* 2026;**59**:119. <https://doi.org/10.1007/s10462-026-11522-9>
 120. Jumper J, Evans R, Pritzel A *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. <https://doi.org/10.1038/s41586-021-03819-2>
 121. Evans R, O'Neill M, Pritzel A *et al*. Protein complex prediction with AlphaFold-multimer. *bioRxiv* 2021. 2021.10. 04.463034.
 122. Schulte-Sasse R, Budach S, Hniss D *et al*. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. *Nat Mach Intell* 2021;**3**:513–26. <https://doi.org/10.1038/s42256-021-00325-y>
 123. Gong J, Hao M, Cheng X *et al*. xTrimoGene: an efficient and scalable representation learner for single-cell RNA-seq data. *Adv Neural Inf Proces Syst* 2023;**36**:69391–403.
 124. Chen D, Wang X, Zhu H *et al*. Predicting anticancer synergistic drug combinations based on multi-task learning. *BMC Bioinformatics* 2023;**24**:448. <https://doi.org/10.1186/s12859-023-05524-5>
 125. Ocana A, Pandiella A, Privat C *et al*. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. *Biomark Res* 2025;**13**:45. <https://doi.org/10.1186/s40364-025-00758-2>
 126. Buntz B. Genentech's lab in the loop aims to tap the power of quantity for quality drug discovery. *Drug Discovery and Development*, Cranbury, NJ, USA: Drug Discovery and Development (MJH Life Sciences); 2025. 2025/08/21.
 127. Regev A. *Lab in a Loop: AI to Transform Drug Discovery and Development*, Santa Clara, CA, USA: NVIDIA Corporation; 2024. 2024/03/20.
 128. Cui H, Tejada-Lapueta A, Brbic M *et al*. Towards multimodal foundation models in molecular cell biology. *Nature* 2025;**640**:623–33. <https://doi.org/10.1038/s41586-025-08710-y>
 129. Niehues JM, Quirke P, West NP *et al*. Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: a retrospective multi-centric study. *Cell Rep Med* 2023;**4**:100980. <https://doi.org/10.1016/j.xcr.2023.100980>
 130. Wei H, Zhang X, Zhou Z *et al*. Hybrid model for predicting microsatellite instability in colorectal cancer using hematoxylin & eosin-stained images and clinical features. *Front Oncol* 2025;**15**:1580195. <https://doi.org/10.3389/fonc.2025.1580195>
 131. Guittton T, Allaupe P, Rabilloud N *et al*. Artificial intelligence in predicting microsatellite instability and KRAS, BRAF mutations from whole-slide images in colorectal cancer: a systematic review. *Diagnostics* 2024;**14**:99. <https://doi.org/10.3390/diagnostics14010099>

132. Wang Z, Wu Y, Li Y *et al*. Pathological omics prediction of early and advanced colon cancer based on artificial intelligence model. *Discov Oncol* 2025;**16**:1330. <https://doi.org/10.1007/s12672-025-03119-5>
133. Rieke N, Hancox J, Li W *et al*. The future of digital health with federated learning. *NPJ Digit Med* 2020;**3**:119. <https://doi.org/10.1038/s41746-020-00323-1>
134. Tonekaboni S, Joshi S, McCradden MD *et al*. What clinicians want: Contextualizing explainable machine learning for clinical end use. In: Finale Doshi-Velez, Jim Fackler, Ken Jung *et al*. (eds). *Machine Learning for Healthcare Conference*. Cambridge, MA, USA: PMLR, 2019, 331–46.
135. Holzinger A, Langs G, Denk H *et al*. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019;**9**:e1312. <https://doi.org/10.1002/widm.1312>
136. Collins GS, Reitsma JB, Altman DG *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *J Br Surg* 2015;**102**:148–58. <https://doi.org/10.1002/bjs.9736>
137. Rajkomar A, Hardt M, Howell MD *et al*. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med* 2018;**169**:866–72. <https://doi.org/10.7326/M18-1990>
138. Nguyen E, Poli M, Faizi M *et al*. HyenaDNA: long-range genomic sequence modeling at single nucleotide resolution. *ArXiv*. 2023;**36**:43177–201.
139. Vishniakov K, Viswanathan K, Medvedev A *et al*. Genomic foundationless models: pretraining does not promise performance. *bioRxiv* 2024:2024.12.18.628606. <https://doi.org/10.1101/2024.12.18.628606>
140. De Lange M, Aljundi R, Masana M *et al*. A continual learning survey: defying forgetting in classification tasks. *IEEE Trans Pattern Anal Mach Intell* 2021;**44**:3366–85.
141. Schölkopf B, Locatello F, Bauer S *et al*. Toward causal representation learning. *Proc IEEE* 2021;**109**:612–34. <https://doi.org/10.1109/JPROC.2021.3058954>
142. Laubenbacher R, Niarakis A, Helikar T *et al*. Building digital twins of the human immune system: toward a roadmap. *NPJ Digit Med* 2022;**5**:64. <https://doi.org/10.1038/s41746-022-00610-z>