



OPEN ACCESS

EDITED BY

Lisha Mou,
Shenzhen Second People's Hospital, China

REVIEWED BY

Ling Xiao,
Harvard Medical School, United States
Ping Han,
Baylor College of Medicine, United States
Yuan Liu,
University of Illinois at Urbana-Champaign,
United States
Jun Yin,
CVS Health, United States

*CORRESPONDENCE

Yingli Sun
✉ scaroll99@hotmail.com

†These authors have contributed equally to this work

RECEIVED 22 October 2024

ACCEPTED 05 December 2024

PUBLISHED 23 December 2024

CITATION

Hai L, Jiang Z, Zhang H and Sun Y (2024)
From multi-omics to predictive biomarker: AI in tumor microenvironment.
Front. Immunol. 15:1514977.
doi: 10.3389/fimmu.2024.1514977

COPYRIGHT

© 2024 Hai, Jiang, Zhang and Sun. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

From multi-omics to predictive biomarker: AI in tumor microenvironment

Luo Hai ^{1,2}, Ziming Jiang^{3†}, Haoxuan Zhang^{3†}
and Yingli Sun^{1,2,4*}

¹Central Laboratory, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China, ²Shenzhen Key Laboratory of Epigenetics and Precision Medicine for Cancers, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China, ³Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, ⁴Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

In recent years, tumors have emerged as a major global health threat. An increasing number of studies indicate that the production, development, metastasis, and elimination of tumor cells are closely related to the tumor microenvironment (TME). Advances in artificial intelligence (AI) algorithms, particularly in large language models, have rapidly propelled research in the medical field. This review focuses on the current state and strategies of applying AI algorithms to tumor metabolism studies and explores expression differences between tumor cells and normal cells. The analysis is conducted from the perspectives of metabolomics and interactions within the TME, further examining the roles of various cytokines. This review describes the potential approaches through which AI algorithms can facilitate tumor metabolic studies, which offers a valuable perspective for a deeper understanding of the pathological mechanisms of tumors.

KEYWORDS

tumor cells, tumor microenvironment, metabolomics, interactions, artificial intelligence

1 Introduction

The tumor is characterized by the uncontrolled growth of abnormal cells, invasion of adjacent parts of the body beyond normal boundaries, or metastasis to other organs, ultimately leading to death (1). According to the World Health Organization, malignant tumors have become one of the major threats to human health (2). Every year, millions of people die from cancer, with lung cancer, colorectal cancer, liver cancer, breast cancer, and cervical cancer being the most common types (3). During the development and progression of tumors, tumor cells are exposed to harsh conditions (4). To survive and sustain growth, cells must adapt to the environment. Cellular metabolic reprogramming is a mechanism by which cells promote cell proliferation and growth by altering metabolic patterns to meet

energy needs. Metabolic reprogramming not only helps tumor cells resist external stress but also endows them with new functions, e.g. immune suppression and evasion (5, 6). However, it is still difficult to discover the signaling pathways and mechanisms that control the metabolic reprogramming of tumor cells and immune cells.

To explore the above challenges, research on tumor metabolism has gradually entered metabolomics. Metabolomics is the collection of metabolites, or low molecule chemicals involved in metabolism, and can directly reflect the functional readouts of biochemical reactions, providing insight into many aspects of cell physiology. Metabolomics has been applied in many aspects of tumor research, including tumor pathology discovery, biomarker discovery, and treatment efficacy evaluation (7, 8). Changes in the metabolic spectrum reflect the process of tumor occurrence and development. Another important area is the exploration of personalized treatment strategies, namely the recognition of personalized tumor biomarkers (9, 10). Moreover, tumor metabolomics analysis can integrate biomolecular information, kinetic data, and other omics data to further study the activity of metabolites within tumor cells and track deep changes in metabolic pathways (11).

With technological advancements, artificial intelligence (AI) has increasingly been applied in tumor metabolomics research. AI technology offers significant advantages in handling large-scale data, uncovering complex biological networks, and improving research accuracy. Specifically, AI can be utilized in various research stages, including data preprocessing, feature extraction, pattern recognition, and data integration. For example, deep learning algorithms can identify feature peaks in mass spectrometry data, increasing the accuracy of metabolite identification. Machine learning methods can perform integrative analysis of multi-omics data, helping to reveal intricate biological networks and pathways. By integrating AI technology, researchers

have achieved a more comprehensive and in-depth understanding of tumor metabolism processes. These technologies not only increase the efficiency and accuracy of data analysis but also offer new perspectives for personalized medicine and precision therapy. As AI technology continues to advance, further breakthroughs in tumor metabolism research are anticipated, driving significant progress in biomedical research.

2 Tumor microenvironment

Tumors are caused by the accumulation of genetic mutations and global epigenetic changes in chromatin that regulate gene expression (12). Genetic alterations in tumor suppressor genes or oncogenes can lead to dramatic changes in gene expression leading to cancer (13). Epigenetic modifications of chromatin, including DNA methylation, histone modifications, nucleosome positioning, and non-coding RNAs, can regulate DNA access to transcription factors and other cis-regulatory elements, thereby affecting gene expression (14). Genetic and epigenetic factors complement each other to drive tumor initiation and progression. Figures 1A, B show a graphical representation of various patient characteristics (e.g., gender, age, and dietary), environment (e.g., water, air, and stress), tumor-intrinsic factors, and extrinsic factors affecting the cancer cells and thereby regulate the tumor microenvironment (TME) (13). Both intrinsic and extrinsic factors of tumors regulate the immune response in the TME. Genomic mutations, chromatin modifiers, and non-coding RNAs, among other intrinsic factors of cancer cells, regulate tumorigenesis, metastasis, and immunogenicity. Epigenetic modifications (such as DNA methylation and histone acetylation) regulate gene expression (15). Non-coding RNAs (including long non-coding RNAs, microRNAs, and circular non-coding RNAs) regulate gene

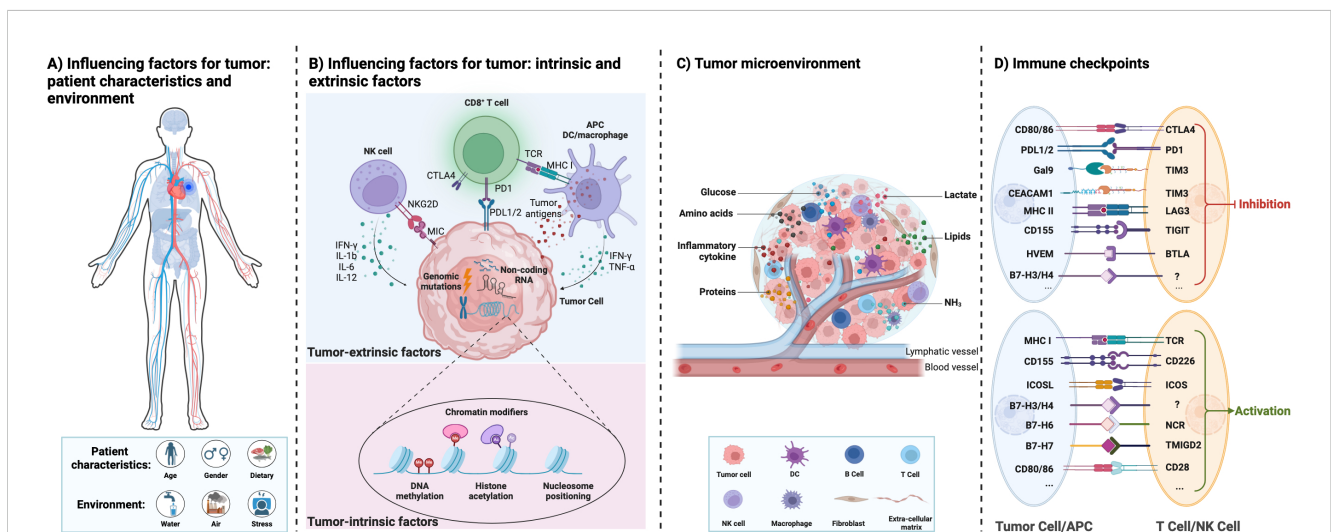


FIGURE 1 Interactions between the influencing factors of tumor, tumor microenvironment, and tumor immunity. (A) The different influencing factors of tumor: patient characteristics and environment. (B) The different influencing factors of tumor: tumor intrinsic and extrinsic factors. (C) The tumor microenvironment with different cell types and their interactions directly affects cancer progression. They provide tumors with abundant nutrients (including glucose, amino acids, lipids, proteins, etc.) and an immunosuppressive environment for tumor growth. (D) Immune checkpoint proteins between immune cells and tumor cells are summarized.

transcription and mRNA stability (16). Other intrinsic mechanisms of cancer cells include the expression of immunosuppressive cytokines to evade anti-tumor immunity, expression of immunosuppressive molecules such as PD-L1 and PD-L2, and inhibition of antigen processing and presentation mechanisms and tumor-associated antigens (17). Extrinsic factors of cancer cells include tumor-infiltrating immune cells, fibroblasts, stromal cells, and endothelial cells (18). Extrinsic factors also include secreted factors such as cytokines, chemokines, metabolites, growth factors, and immune checkpoint molecules (19). Tumor-associated antigens presented by antigen-presenting cells such as macrophages and dendritic cells can activate CD8⁺ T cells and thereby induce effective anti-tumor immunity. However, immune checkpoint molecules expressed by cancer cells regulate the inflammatory state of tumors and suppress inflammation (Figure 1D) (20).

Tumor cells interact with normal cells and other factors to create TME (21–23). TME contains blood vessels, immune cells, stromal cells, fibroblasts, signaling molecules, and extracellular matrix (Figure 1C) (20). The TME plays an important role in the occurrence and development of tumors. Mostly, the microenvironment of early-stage tumors tends to exert anti-tumor effects, while the microenvironment of late-stage tumors tends to worsen conditions (24). Moreover, the metabolism of tumor cells is complex and heterogeneous, involving the metabolism of glucose, lactate, pyruvate, glutamine, and fatty acids (25). Tumor cells adapt to changes through remodeling in different TME, including the production of ATP bioenergy, oxygen balance, and nutrient absorption (26). Damaghi et al. (24) found through their study of ductal carcinoma *in situ* cells that the harsh TME promotes the Warburg effect through transcriptional reprogramming. Due to the potential complex metabolic interactions between the TME and tumor cells, studying the relationship between the TME and metabolism can help elucidate the remodeling mechanism of tumor tissue and the metabolic changes in tumor cells.

The metabolic characteristics and preferences of tumor cells constantly change during the development of tumors. Meanwhile, the TME is a complex and constantly changing entity whose composition varies depending on the type and location of the tumor (27). The communication between the tumor cells and TME results in specific patterns of tumor cell growth and development and has spawned various ways to help tumor cells evade immune surveillance (28). Fibroblasts regulate the production of extracellular matrix and tumor signaling molecules, promoting tumor growth, invasion, and metastasis (29). Tumor cells and specific cells of the TME avoid or inhibit immune responses by inhibiting the proliferation of helper and cytotoxic T cells or by promoting the recruitment of immune suppressive regulatory T cells (Treg) and bone marrow-derived suppressive cells (MDSC) mediated by inflammation (30). Chekulayev et al. (31) discovered the “reverse Warburg effect” when studying the crosstalk between colon cancer cells and stromal cells. Rossi et al. (32) reported that microbial metabolites can be important regulatory factors in the TME, regulating inflammation, proliferation, and cell death in a positive or negative manner. In pancreatic

adenocarcinoma, stromal-associated pancreatic stellate cells secrete alanine, which provides carbon and nitrogen for tumor cell proliferation (33).

3 Tumor metabolism

Metabolites are extremely sensitive to internal signals and external stimuli, which means that metabolomics has the potential to become a probe of biological phenotypes, revealing what is happening inside cells. Metabolomics can systematically identify and quantify all metabolites in biological samples at high throughput, providing key information about the state of cancer that other omics technologies cannot. Metabolic perturbations can lead to characteristic metabolic phenotypes that can be used for early cancer diagnosis, surveillance, and as targets for cancer therapy. Studying tumor metabolism involves analyzing a series of biochemical reactions in tumor cells, which involve various types of reactants and their related enzymes (34–36). Therefore, studying changes in metabolism-related substances is the foundation of most tumor metabolism research.

Analytical methods to study metabolomics include the use of classical chemical analysis (37, 38), nuclear magnetic resonance (NMR) (39–41), and mass spectrometry (MS) (42–44). These methods have the advantage of high identification and quantitative accuracy when studying biochemical reactions related to tumor metabolism. However, there are also some disadvantages. For example, NMR spectroscopy requires high purity samples, which limits its applicability in analyzing the fine structure of metabolites (45). MS has been fueled by combining with highly efficient separation techniques such as gas chromatography or liquid chromatography, which improve the resolution of analytes by increasing sensitivity and specificity (46). Although these combinations allow the detection of more analytes with high sensitivity and specificity, MS-based techniques still face several challenges, such as the chromatographic separation of isomers, the elimination of exogenous contaminants, the lack of a complete reference database, and the identification of unknown metabolites (47).

Advances in single-cell sequencing, spatial transcriptomics, proteomics, and artificial intelligence have propelled tumor metabolomics into a new era. AI significantly contributes to elucidating tumor metabolism mechanisms, identifying diagnostic and prognostic biomarkers, and facilitating clinical applications. For example, Zhao et al. (48) successfully demonstrated the utility of machine learning algorithms in enhancing diagnostic accuracy for early-stage esophageal squamous cell carcinoma (ESCC). The study employed Uniform Manifold Approximation and Projection (UMAP) and Hierarchical Clustering Analysis (HCA) to categorize tissue samples into distinct groups. Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) revealed significant metabolic differences between ESCC tumors and normal mucosa. Additionally, Random Forest analysis was utilized to identify critical metabolic biomarkers, such as glutamate, which effectively distinguished early-stage ESCC from normal tissues. Support Vector Machine (SVM) models were further applied to develop

simplified metabolite panels, achieving an area under the curve (AUC) of 0.984 in serum samples. These findings illustrate that AI algorithms have become integral to nearly every aspect of tumor metabolism research (48). The emergence of advanced algorithms has significantly enhanced our understanding of tumorigenesis. For instance, Deng et al. (49) developed an explainable deep learning algorithm, DeepMSPProfiler, which effectively removes batch effects by systematically excluding batch-related information through hidden layers. Building on this algorithm, the integration of metabolomics and methylation data in the study verified the associations between the PLA and UGT gene families and disease-specific metabolites. Su et al. (50) revealed the drug resistance pathway in melanoma cells through single-cell multi-omics analysis, providing a key and unique tool for addressing the signaling pathway function and metabolic changes in the adaptive development of drug resistance in tumor cells. Sun et al. (51) used spatial metabolomics to locate and analyze metabolites of different metabolic pathways in the tissues of 256 esophageal cancer patients, revealing the molecular level of tumor occurrence from metabolites to enzymes and providing a new perspective for understanding the metabolic reprogramming of tumors. Jin et al. (52) established a relationship between metabolic characteristics and oncogenic mutations of receptor tyrosine kinase by integrating metabolomics and transcriptomics, providing a basis for metabolic targeted therapy of specific tumor genotypes.

Almost all malignant tumors exhibit uncontrolled cell proliferation (53). To support cell growth, tumor cells must adaptively adjust their metabolism to meet their material and energy needs (54). Otto Warburg and colleagues first discovered that tumor cells tend to convert oxidative metabolism into fermentation metabolism (55). In normal cells, most of the pyruvate formed by glycolysis enters the tricarboxylic acid cycle and is oxidized through oxidative phosphorylation (OXPHOS) (56). In contrast, in tumor cells, most pyruvate is converted into lactate through fermentation, a phenomenon known as the “Warburg effect” (57, 58). Through competitive uptake of glucose, cancer cells gain a survival advantage over normal cells through metabolic adaptations in oxygen-limited conditions. Moreover, cancer cells need to use their nutrient inputs as cellular building blocks. Thus, cancer cells switch from OXPHOS (nutrient-consuming) to glycolysis (biomass-building) pathways to support their rapid, uncontrolled proliferation. Hu et al. (59) found that the overall metabolic status of different tumors was similar, such as the upregulation of nucleotide biosynthesis and glycolysis. However, the expression of specific metabolic pathways, such as OXPHOS, is heterogeneous among different tumors (60).

4 Metabolic modulation of cell-mediated immunity

With the deepening of research, researchers have discovered that tumors are essentially a metabolic disorder, in which several major metabolic pathways are altered to accommodate increased proliferation of tumor cells, and the reprogrammed metabolic pathways include glucose, amino acids, lipids, and other

metabolism. The metabolic reprogramming not only meets the nutritional or energy needs of tumor cells but also affects the function of immune cells (61). Understanding the metabolic modulation of tumor immunity may provide therapeutic insights into immunotherapy resistance and facilitate the development of new strategies for tumor therapy.

4.1 Glycolysis metabolism

Glycolysis (2 mol ATP/mol glucose) is a less efficient pathway for ATP production than mitochondrial OXPHOS (36 mol ATP/mol glucose). However, the rate of glycolysis is 10~100 times faster than the rate of mitochondrial TCA cycle and OXPHOS. Therefore, the amount of ATP produced by the two metabolic pathways of glucose is similar at the same time. In tumor cells, glucose is metabolized by glycolysis, producing lactate and nicotinamide adenine dinucleotides under aerobic conditions. The low-yielding but high-rate ATP production mode is more conducive to the competition of tumor cells for nutrients and meets their energy demand for rapid growth. Tumor cells take up a large amount of glucose from the environment with a strong competitive advantage, ensuring their energy supply and self-growth, depriving immune cells of glucose utilization, and inhibiting their tumor cell-killing effect. In addition, glucose metabolism including glycolysis, pentose phosphate pathway, hexosamine pathway, and glycogen synthesis are reprogrammed in tumor cells (62). As shown in Figure 2, pyruvate kinase M2 (PKM2), which is highly expressed in tumors and promotes the Warburg effect, is upstream of the decision point between glycolytic and oxidative metabolism. PKM2 converts phosphoenolpyruvate (PEP) into pyruvate, which can be metabolized either to lactate or acetyl-CoA (63). Moreover, PKM2 also functions as a transcriptional coactivator, PKM2 interacts with HIF-1 α in the nucleus and binds to the PD-L1 promoter region, enhancing the expression of PD-L1 in tumor cells (64). The interaction between the immune checkpoint PD-L1 and B7-H3 activates the classic aerobic glycolytic pathway PI3K-AKT mTOR in tumor cells (65). Moreover, due to glucose deprivation and downregulation of mTOR activity, the glycolytic signaling pathway PI3K-AKT mTOR is inhibited in T cells (66). In addition, the CTLA-4 pathway competitively inhibits CD28-mediated co-stimulation and reduces Akt phosphorylation and activation, thereby impairing T-cell glucose metabolism and mitochondrial remodeling (67). Moreover, the interaction between immune checkpoints and their ligands (such as PD-1/PD-L1 and CTLA-4/CD86) further participates in the metabolic reprogramming of tumor cells and immune cells (68). In summary, the synergistic effect of tumor metabolism regulation and inhibition of immune checkpoints in the TME can activate the host anti-tumor immune response.

Compared with normal cells, tumor cells utilize a large amount of glucose to produce lactate under aerobic conditions and provide the amino acids and intermediate metabolites of the pentose phosphate required for tumor cell proliferation. The accumulation of lactate in the TME has profound effects on immune cells. In Figure 3A, tumor cells consume large amounts of glucose, which is associated with poor CD4⁺ and CD8⁺ T cell

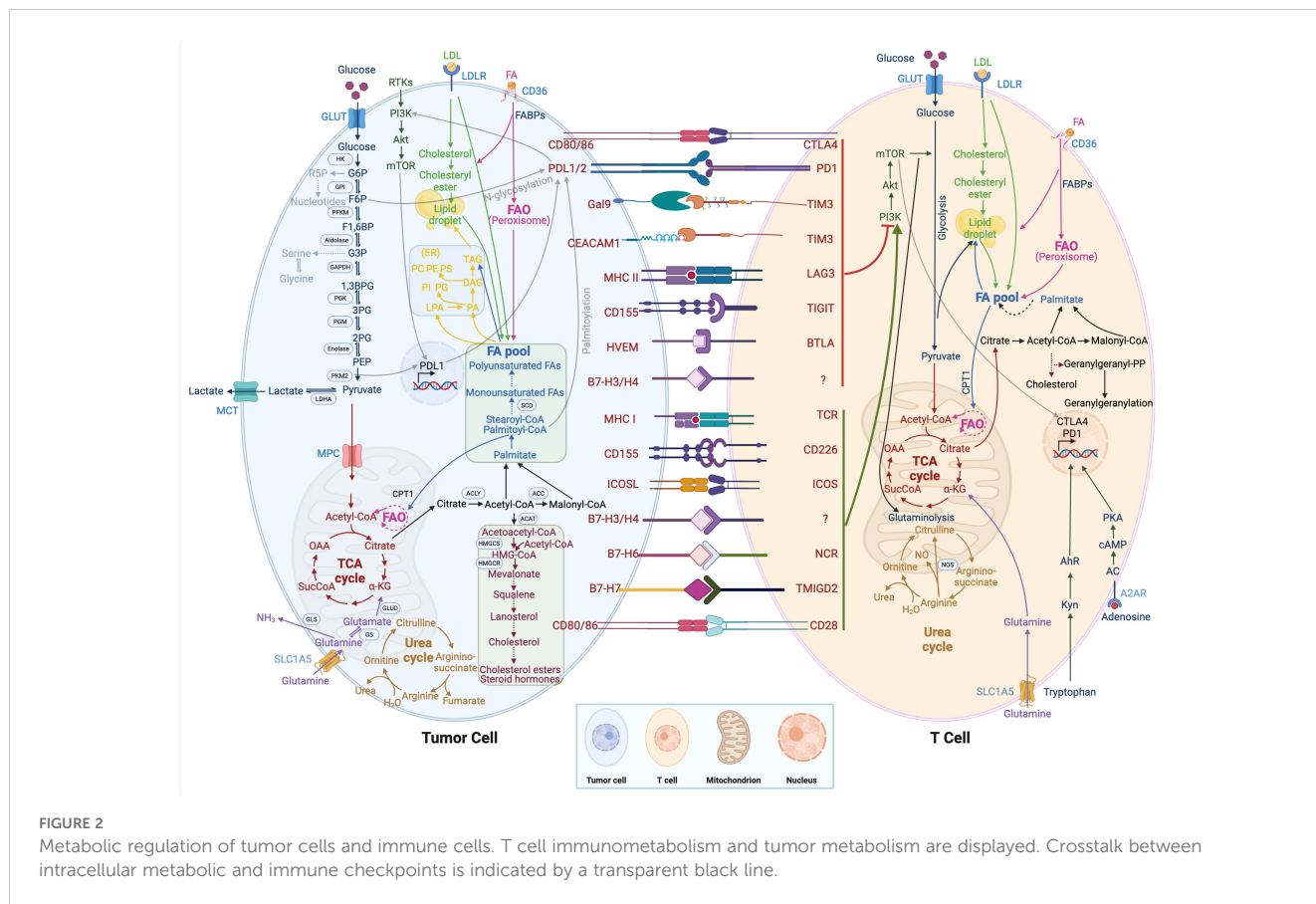


FIGURE 2 Metabolic regulation of tumor cells and immune cells. T cell immunometabolism and tumor metabolism are displayed. Crosstalk between intracellular metabolic and immune checkpoints is indicated by a transparent black line.

infiltration (69). High glucose depletion leads to the production and secretion of lactate into the TME, where lactate acts in an immunosuppressive manner, reduces the cytolytic activity of natural killer (NK) cells, and enhances PD-1 expression and the immunosuppressive capacity of Treg. In addition, lactate increases the frequency of MDSC in tumors and spleen and induces “M2-like” polarization in tumor-associated macrophages. Additionally, the accumulation of lactate in tumors leads to a pH decrease and hinders the ability of dendritic cells (DC) to recognize and present antigens, as well as the stability of antigen MHC-I complexes (Figure 4) (20). Acidification also reduces the ability of mannan receptors (MR) to bind to antigens, inhibits glycolysis, and promotes monocyte differentiation into monocyte-derived dendritic cells. Moreover, lactate in tumors inhibits toll-like receptor 3 (TLR3) and stimulator of interferon genes (STING) leading to interferon- γ reduction, accelerates antigen degradation, and impairs cross-presentation. Pilon-Thomas et al. (70) reported that neutralizing the acidity of TME with bicarbonate can increase T cell infiltration and improve response to immunotherapy in immune checkpoint inhibition and adoptive cell transfer therapy.

4.2 Amino acid metabolism

Amino acids are essential for maintaining the rapid proliferation of tumor cells. Besides as substrates for protein synthesis, amino acids play an important role in energy

production, lipid and nucleic acid synthesis, and cellular redox homeostasis maintenance. The increased demand for amino acids results in tumor cells becoming strongly dependent on exogenous amino acids or reprogramming of amino acid metabolism. Alterations in amino acid metabolism can be used not only as clinical biomarkers of cancer progression but also as therapeutic targets.

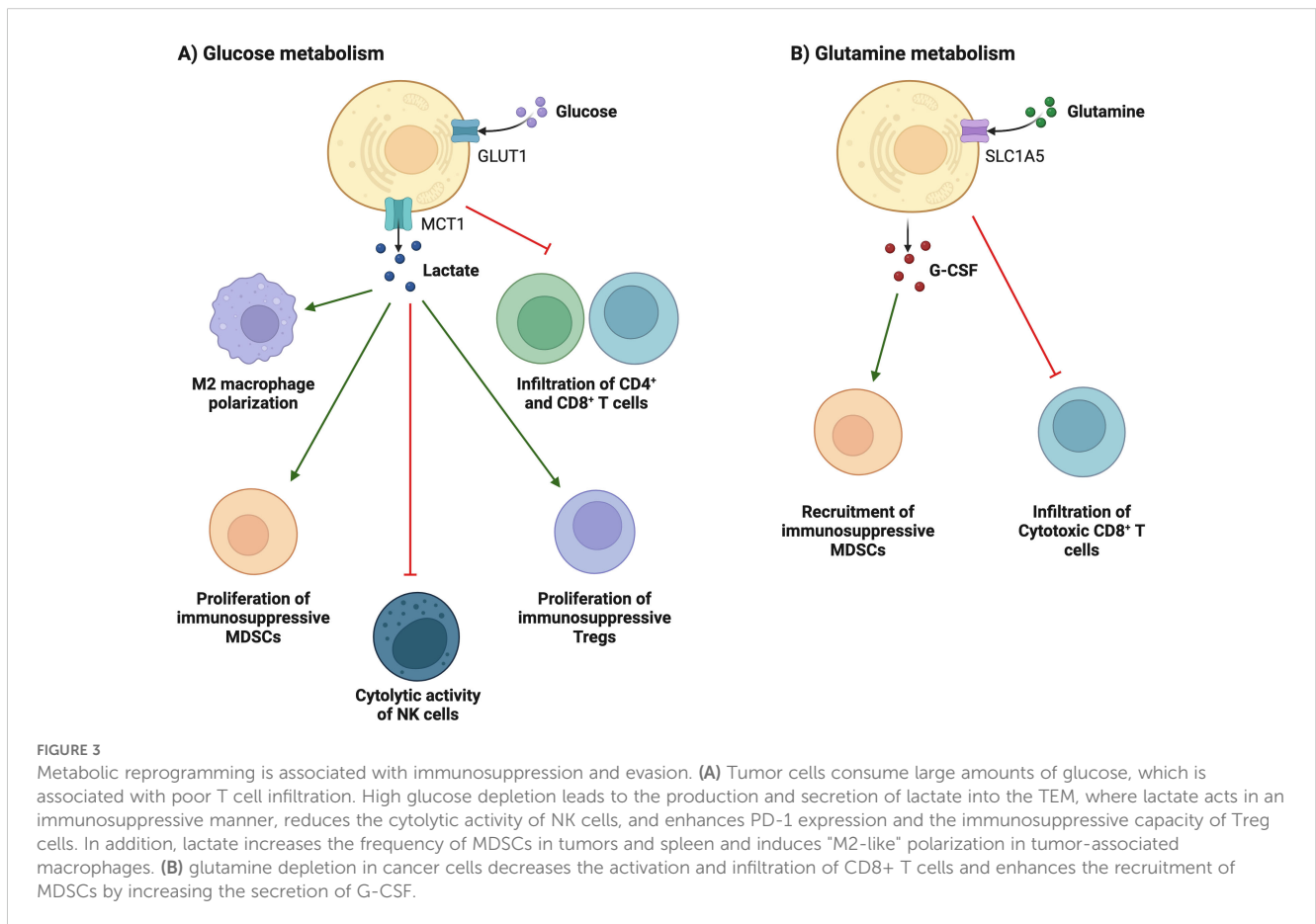
Glutamine is an important metabolic fuel that helps meet the high demand for energy, biosynthetic precursors, and reducing agents in rapidly proliferating cancer cells (71). In Figure 2, glutamine is transported to the cytoplasm through the solute carrier family 1, member 5 (SLC1A5) (72), and is converted to glutamate in the mitochondria through a deamination reaction catalyzed by glutaminase. Glutamate is converted to the TCA cycle intermediate α -ketoglutarate (α -KG) by glutamate dehydrogenase (73). α -KG is a critical metabolite that serves in both ATP production and in replenishing TCA cycle intermediates (74). Glutamine metabolism is essential for developing effector T cells (75). The overexpression of inhibitory receptors (such as CTLA-4 or PD-1) can inhibit the upregulation of glucose and glutamine metabolism after TCR involvement and co-stimulation (76, 77). The interaction between PD-1 and PD-L1 or PD-L2 can inhibit the PI3K-Akt mTOR pathway (78) and disrupt T cell metabolic reprogramming, including glutaminolysis (79). Huang et al. (80) reported that SLC1A5 overexpression can stimulate the growth and survival of colon cancer cells. In Figure 3B, glutamine depletion in cancer cells decreases the activation and infiltration of CD8⁺ T cells

and enhances the recruitment of MDSC by increasing the secretion of granulocyte-colony stimulating factor (69).

Arginine is a versatile amino acid. In addition to being a building block in protein synthesis, it is a precursor of nitric oxide (NO), creatine, and polyamines. The arginine metabolic pathway is shown in Figure 2, and there are two main metabolic pathways: Arginine is metabolized into citrulline and NO by nitric oxide synthase, and into ornithine and urea by arginase. Ornithine is an important resource for putrescine, which is a key precursor for polyamines. NO plays a diverse role in tumorigenesis and metastasis, and the promotion or inhibition of tumors depends on sensitivity to NO, exposure time, and NO concentration (81). Low concentrations of NO (<100 nmol/L) may promote tumor development by inhibiting apoptosis and stimulating endothelial cell proliferation. In contrast, high concentrations of NO (400-1000 nmol/L) would disrupt the cell cycle and accelerate cellular apoptosis (82). Because tumor cells rely on arginine in the TME, depriving tumors of arginine has emerged as a novel anti-tumor therapy and has shown encouraging efficacy in clinical trials against certain types of cancer. However, simply depriving arginine not only starves tumor cells but also impairs the anti-tumor immune response. Geiger et al. (83) reported that the reduction of intracellular arginine negatively affected the survival of T cells. When arginine is completely depleted in the medium, the cell

cycle of T cells is arrested in the G₀-G₁ phase. Arginine supplementation, while enhancing the anti-tumor immune response, may support tumor growth. Both deprivation and supplementation alter arginine levels in the TME, which influences all cells. To achieve tumor-specific arginine restriction without influencing immune cells, an in-depth study of differences in arginine metabolism between tumor cells and immune cells is needed.

Tryptophan, an essential amino acid, plays a vital role in protein biosynthesis and serves as a precursor for the synthesis of a variety of important bioactive compounds. As shown in Figure 2, tryptophan is first converted to N-formyl kynurenine, which is deformylated by arylformamidase to kynurenine (Kyn) (84). Kyn then activates the transcription factor aryl hydrocarbon receptor (AhR). Activated AhR could upregulate levels of PD-1 in T cells, inhibiting the anti-tumor immune response (85). Tryptophan metabolism and its key enzymes affect a variety of cell biology functions, including immune response, cell proliferation, and migration, through interactions with downstream molecules or pathways. Le Naour et al. (86) found that tryptophan is associated with restricting immune response in the TME and is often upregulated in human tumors. Targeted therapy based on tryptophan metabolism provides a new and potentially advantageous therapeutic strategy for cancer.



4.3 Lipid metabolism

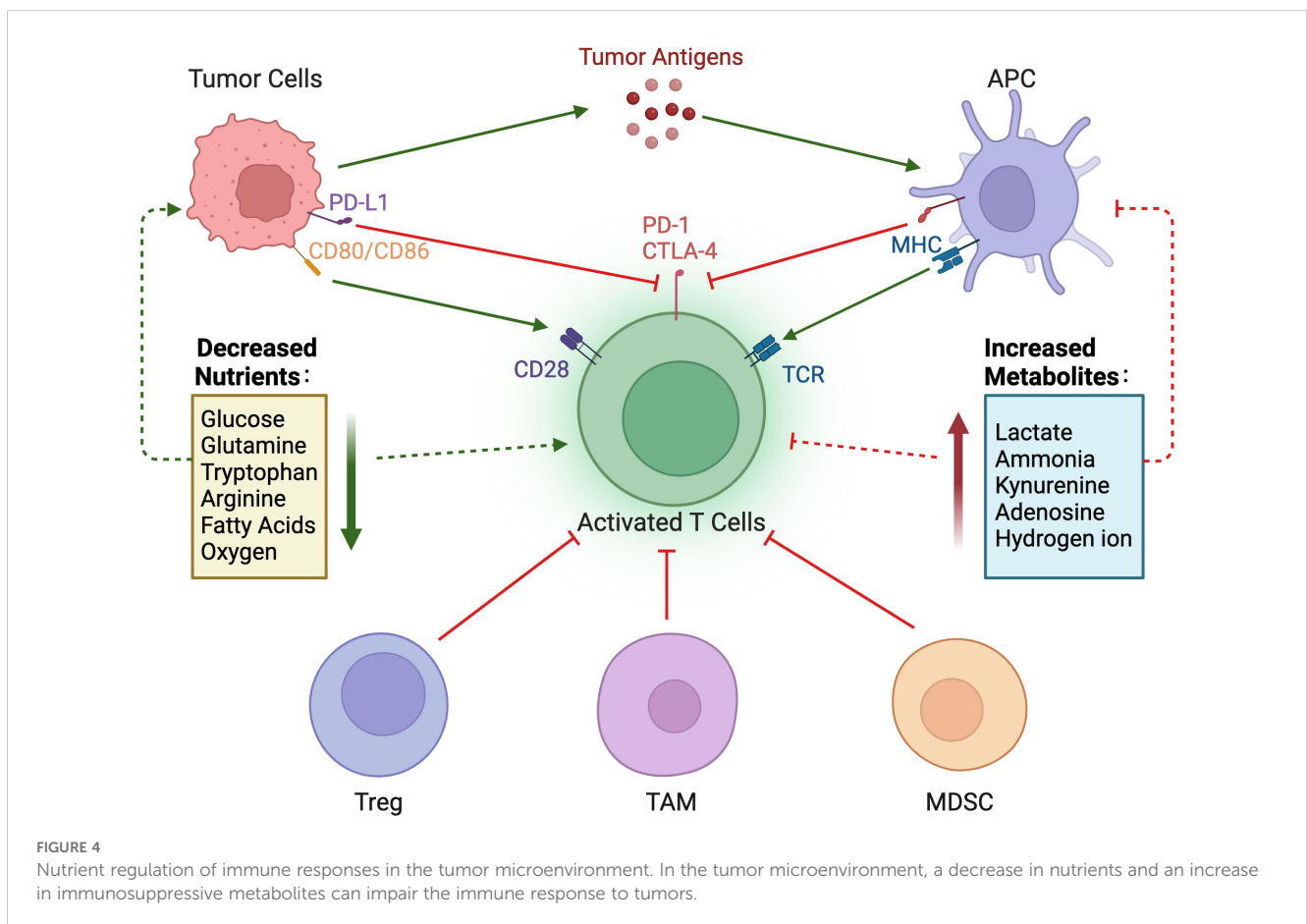
Lipids are essential nutrients for cells, the main structural components of cell membranes, signal molecules, and energy providers. The most common lipids include fatty acids, triglycerides, sphingolipids, phospholipids, and cholesterol (87). In addition, abnormal fatty acid metabolism is related to the occurrence and development of tumors, such as liver cancer, gastric cancer, ovarian cancer, and cervical cancer (88–90).

The sterol regulatory element binding protein pathway in lipid metabolism controls lipid synthesis in cells (91). In Figure 2, *De novo* synthesis of lipids can promote high expression of PD-L1 and tumor immune suppression through palmitoylation of the immune checkpoint PD-L1 (92). In addition, PD-1 stimulates the AMP-activated protein kinase activity, inhibits glycolysis, and promotes fatty acid oxidation by upregulating activated CD4⁺ T cells, thereby inhibiting the development of effector T cells (93). The lipid transporter protein CD36 mediates intracellular fatty acid uptake and lipid droplet growth (94). CD36-mediated fatty acid uptake in tumor-infiltrating CD8⁺ T cells activates lipid peroxidation and iron apoptosis, while CD36 deficiency inhibits tumor growth (95). In addition, CD36 on the surface of CD8⁺ TILs takes up oxidized low-density lipoprotein and induces lipid peroxidation, thereby promoting CD8⁺ T cell dysfunction in tumors (96). Cholesterol

metabolism regulates the antitumor activity of CD8⁺ T cells. Inhibition of acetyl-CoA acetyltransferase-1 increases plasma membrane cholesterol levels and enhances the effector and proliferation function of CD8⁺ cells (97).

5 Metabolic modulation of humoral immunity

The phenotype and function of T cells in the TME have been extensively studied, but the role of immune metabolism in B cell function, differentiation, and its impact on tumor immunity remains elusive, perhaps partially due to their range of actions and heterogeneity. B cell activation triggered by antigen binding to the B cell receptor (BCR) triggers naive B cell proliferation and differentiation into plasma cells that produce specific antibodies. This process requires increased energetic activity to supply sufficient energy and raw material. Upon receiving antigenic stimulation, the metabolic requirements of B cells undergo significant changes. Activated B cells rapidly increase glycolysis, the TCA cycle, and oxidative metabolism, whereas naive B cells exhibit metabolic quiescence (98). Upon activation of the BCR, B cells tend to aerobic glycolysis and produce more lactate. Antigen-stimulated B cells undergo glycolysis mainly before the S phase.



Once in the S phase, B cells skew to the pentose phosphate pathway, possibly to provide ribose-5-phosphate for nucleotide synthesis and NADPH for redox homeostasis (99). Additionally, plasma cells utilize amino acid metabolism, mitochondrial respiration, and the TCA cycle to maintain antibody secretion (100). Glycolysis is increased in regulatory B cells and their differentiation is associated with hypoxic environments, but their metabolic profile remains incompletely understood (101).

In the TME, B lymphocytes can exhibit anti-tumor or pro-tumor characteristics depending on localization, tumor type, TME, and antibody isotype (102). For example, B lymphocytes in hepatocellular carcinoma can express anti-tumor cytokines, e.g., IFN- γ , IL-12, TRAIL, and granzyme B (103). In contrast, B cells can also secrete anti-inflammatory molecules, e.g., TGF- β , IL-10, and IL-35, and release immunosuppressive factors, such as PD-L1, Tim-1, and FasL (104–106).

In B-cell lymphoma, the redox state of tumor cells is imbalanced due to rapid proliferation. To avoid cell damage and apoptosis caused by abnormal oxidative stress, the pentose phosphate pathway is selectively activated by serine/threonine protein phosphatase type-2A to produce NADPH to maintain redox homeostasis. There are “metabolic checkpoints” during the development of B cells to avoid overactivation and malignant transformation. For example, the transcription factors paired box 5 and Ikaros family zinc finger 1 in B cells can limit intracellular energy metabolism. This process reduces glucose uptake and ATP synthesis, limiting cell proliferation by restricting energy metabolism (107). The use of agonists of paired box 5 and Ikaros family zinc finger 1 increases the expression of glucose feedback sensor thioredoxin interacting protein and cannabinoid receptor 2 restricting the glucose uptake and limiting the B cells glycolysis, in combination with glucocorticoids, and has a potential therapeutic effect in B-cell lymphoma. Additionally, Feist et al. (108) reported that the aspartate aminotransferase glutamic-oxaloacetic transaminase 2/signal transducer and activator of transcription-3/p53 signaling pathway is essential for lymphocyte malignant proliferation. Xiong et al. (109) reported that MYC-phosphate cytidylyltransferase 1 choline- α induced abnormal choline metabolism and hindered necrosis in B-cell lymphoma. These studies reveal the potential clinical application prospects of using exogenous metabolic modulators to limit the proliferation of malignant tumors. In summary, tumor metabolism affects tumor therapy by altering the TME and resetting immune cells. Tumor metabolomics is becoming a new research hotspot.

6 Advanced artificial intelligence methods in tumor metabolomics

Tumor metabolism and the TME are complex, dynamic systems that play pivotal roles in tumor progression and immune evasion. Traditional methods for studying tumor metabolism, such as chemical screening, gene knockout, and mass spectrometry imaging, have provided valuable insights but also face limitations in terms of operational complexity and personalized metabolic analysis (110). To overcome these challenges, AI has emerged as

a powerful tool, revolutionizing the field of metabolomics by enabling more robust and comprehensive data analysis.

The study of tumor metabolism requires the integration of various omics technologies, including metabolomics, epigenomics, and proteomics, with a primary focus on metabolomics. Metabolomics is a technique used to analyze small molecule compounds and can be divided into four categories based on research objectives: metabolite target analysis, metabolite fingerprint analysis, metabolite profile analysis, and metabolomics analysis. With advancements in sequencing technology and AI, researchers are developing more robust methods for metabolite analysis and quantification via data-driven approaches (Figure 5) (111).

AI encompasses a range of tools and mathematical methods that enable computer systems to perform tasks and make decisions that typically require human intelligence. AI technologies can be classified into several main categories. Unsupervised learning involves modeling the underlying structure of unlabeled data via techniques such as principal component analysis (PCA) (112), t-SNE (113), and autoencoders (112). On the other hand, supervised learning learns from labeled datasets to focus on predictive tasks. Deep learning, which includes multilayered neural networks such as artificial neural networks (ANNs) (112), automatically learns hierarchical features from raw data and has greatly supported metabolomics research. Additionally, natural language processing (NLP) models such as BERT and GPT have introduced new approaches to metabolomics research (114, 115). These AI advancements aid biologists and medical researchers in various aspects of omics studies, including experimental design, data preprocessing, metabolite annotation, data analysis, multi-omics integration, and clinical applications. This study summarizes the main steps in metabolomics research and the development of AI technologies (Table 1).

6.1 Data preprocessing

After acquiring mass spectrometry data, several processing steps are typically performed before the data can be annotated. These steps include format conversion, peak detection, peak alignment, calculating peak areas, and filtering to eliminate redundancies (111, 146).

Several deep learning algorithms are particularly suitable for preprocessing mass spectrometry data. For example, convolutional neural networks (CNNs) excel at capturing local features by detecting patterns such as peaks and characteristic signals through convolution operations (43, 112, 147). Furthermore, CNNs achieve scale invariance through pooling operations, enabling the model to recognize the same features at different scales, thus enhancing robustness and generalization (148). Melnikov et al. (111) developed the “peakonly” algorithm, utilizing CNNs to enhance peak detection and integration in LC-MS data for metabolomics, achieving high precision and flexibility in managing noisy peaks. Similarly, research by Risum A.B (117), Kantz (119), and Lim (120) demonstrates the application of CNNs for tasks such as peak identification and alignment. Improved RNN

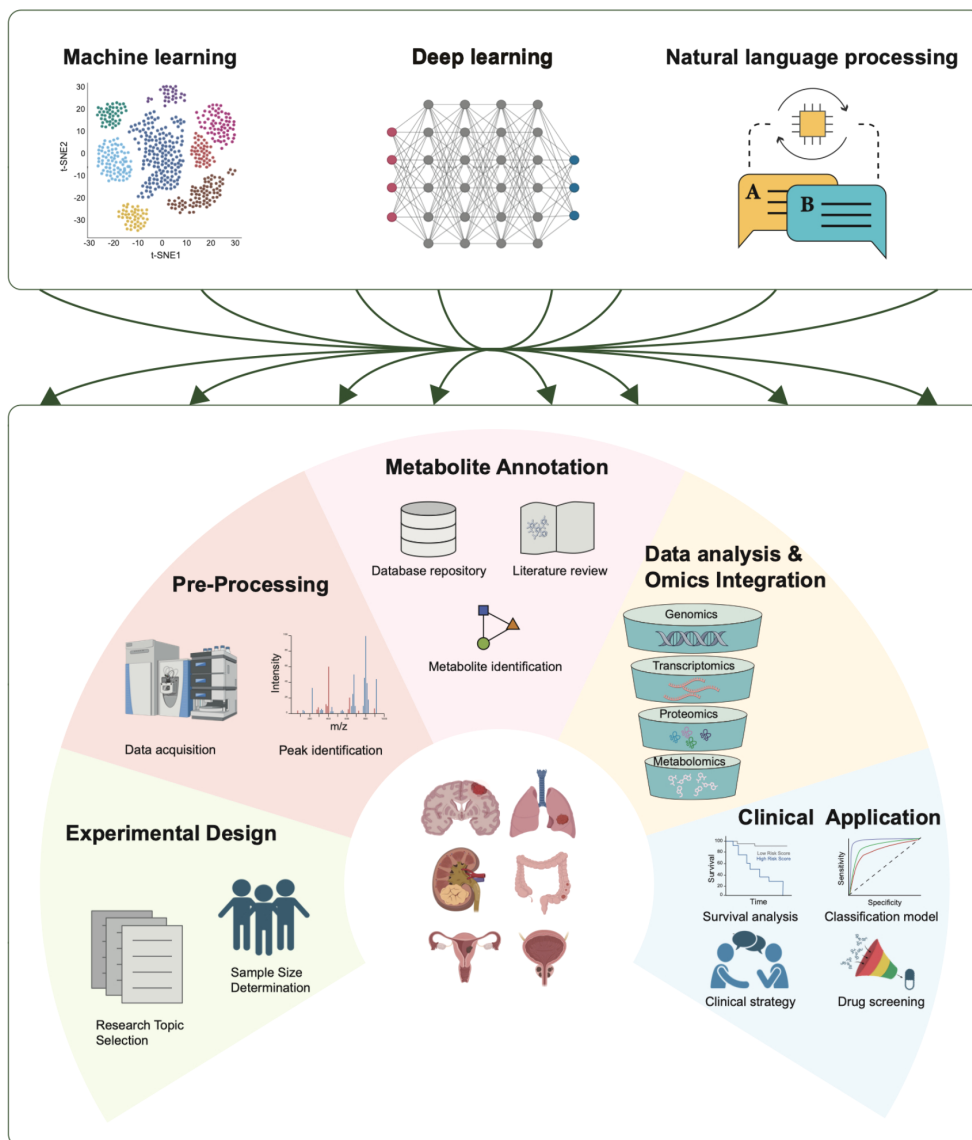


FIGURE 5 Applications of AI in tumor metabolomics. AI technologies play crucial roles in various stages of tumor metabolomics research, including experimental design, data preprocessing, metabolite annotation, data analysis and omics integration, and clinical applications. Three AI technologies—machine learning, deep learning, and natural language processing—contribute significantly to each stage.

TABLE 1 Available ML algorithms of metabolomics.

Tool/Algorithm Name	Data Processing Stage	Omics Type	Functionality	Method Category	Specific algorithm	Publication Year	Reference
	Sample collection & Pre-Processing	NMR	Spectrum Reconstruction & Peak Intensity Calibration	ML/DL	LSTM, DNN	2019	(116)
	Sample collection & Pre-Processing	GC-MS	Peak Detection, Peak Classification	ML/DL	CNN	2019	(117)

(Continued)

TABLE 1 Continued

Tool/Algorithm Name	Data Processing Stage	Omics Type	Functionality	Method Category	Specific algorithm	Publication Year	Reference
konly	Sample collection & Pre-Processing	LC-MS	Peak Detection, Peak Integration	ML/DL	CNN	2020	(111)
ChromAlignNet	Sample collection & Pre-Processing	GC-MS	Peak Alignment, Peak Filtering	ML/DL	LSTM (Long Short-Term Memory)	2019	(118)
	Sample collection & Pre-Processing	LC-MS	Peak Detection, Artifact Elimination	ML/DL	CNN	2019	(119)
	Metabolite Annotation	Raman spectra	Component Identification	ML/DL	CNN	2018	(120)
	Metabolite Annotation	FTIR, MS	Structure Identification, Functional Group Prediction	ML/DL	Autoencoder, MLP (Multi-Layer Perceptron)	2020	(121)
	Metabolite Annotation	LC-MS	Spectral Prediction, Fragmentation Prediction	ML/DL	ANN	2016	(122)
NEIMS	Metabolite Annotation	GC-MS	Spectral Prediction	ML/DL	MLP	2019	(123)
DeepCCS	Metabolite Annotation	IMS	CCS Value Prediction	ML/DL	CNN	2019	(124)
DarkChem	Metabolite Annotation	MS	m/z and CCS Value Prediction, Structure Prediction	ML/DL	VAE (Variational Autoencoder)	2020	(125)
MSHub	Sample collection & Pre-Processing	GC-MS	Auto-deconvolution, Fragmentation Pattern Analysis, Data Sharing, Molecular Networking, m/z and Retention Time Shift Analysis	ML/DL	One-Layer Neural Network	2021	(126)
Combat	Data Analysis/ Omics Integration	MS	Batch Correction	ML/DL	Empirical Bayes Framework	2007	(127)
MNN	Data Analysis/ Omics Integration	MS	Batch Correction	ML/DL	Mutual Nearest Neighbor (MNN)	2018	(128)
Spec2Vec	Data Analysis/ Omics Integration	LC-MS	Spectral Similarity Calculation, Library Matching, Molecular Networking	LLM/NLP	Word2Vec	2021	(129)
METLIN	Metabolite Annotation	LC-MS	Compound Identification	ML/DL	Input-Output Kernel Regression Model	2018	(130)
MassBank	Metabolite Annotation	GC-MS	Compound Identification, Data Repository	ML/DL	Weighted Cosine Correlation	2010	(131)
SIRIUS 4	Metabolite Annotation	LC-MS	Isotope Pattern Scoring, Fragmentation Tree Computation, Element Detection, Structure Prediction	ML/DL	DNN/ Bayesian Network Scoring	2019	(132)
CANOPUS	Metabolite Annotation	LC-MS	Compound Class Annotation	ML/DL	DNN (Deep Neural Network)	2020	(133)

(Continued)

TABLE 1 Continued

Tool/Algorithm Name	Data Processing Stage	Omics Type	Functionality	Method Category	Specific algorithm	Publication Year	Reference
MS2Mol	Metabolite Annotation	LC-MS	<i>De Novo</i> Structure Prediction, Unknown Metabolite Identification	LLM/NLP	Transformer	2023	(134)
Mmvec	Data Analysis/Omics Integration	–	Microbe-Metabolite Association	LLM/NLP	Word2Vec	2013	(135)
MetGem	Metabolite Annotation	LC-MS	Molecular Network Visualization, Spectral Similarity Calculation	ML/DL	t-SNE	2018	(113)
iMet	Metabolite Annotation	LC-MS	Structural Annotation, Novel Metabolite Identification	ML/DL	Radan forest	2017	(136)
MassGenie	Metabolite Annotation	LC-MS	<i>De Novo</i> Structure Prediction, Molecular Identification	LLM/NLP	Transformer	2021	(114)
MIST (Metabolite Inference with Spectrum Transformers)	Metabolite Annotation	LC-MS	Small Molecule Structure Elucidation	LLM/NLP	Transformer	2022	(137)
RT-Transformer	Metabolite Annotation	LC-MS	Metabolite Identification	LLM/NLP	Transformer	2024	(138)
BERT-m7G	Data Analysis/Omics Integration	–	Identification of RNA Modifications	LLM/NLP	Transformer	2021	(139)
Bert2Ome	Data Analysis/Omics Integration	–	Identification of RNA Modifications	LLM/NLP	Transformer/CNN	2023	(140)
IDSL_MINT	Metabolite Annotation	LC-MS	Annotation of Untargeted Metabolomics and Exposomics Data	LLM/NLP	Transformer/CNN	2024	(141)
MSNovelist	Metabolite Annotation	LC-MS	<i>De Novo</i> Structure Prediction, Small Molecule Structure Elucidation, Fingerprint Prediction	ML/DL	MLP	2022	(142)
MS2Prop	Metabolite Annotation	LC-MS	Property Prediction, Compound Identification	LLM/NLP	Transformer/CNN	2022	(143)
Mass2SMILES	Metabolite Annotation	LC-MS	Chemical Structure Prediction, Functional Group Prediction, Molecular Formula Prediction	LLM/NLP	Transformer/CNN	2023	(144)
MS2DeepScore	Metabolite Annotation	LC-MS	Structural Similarity Prediction, Spectral Similarity Calculation, Spectral Library Matching	ML/DL	Siamese neural network	2021	(145)

structures, like long short-term memory (LSTM) networks and gated recurrent units (GRUs), can capture long-range dependencies, aiding in analyzing temporal correlations in mass spectrometry data (149). For instance, Li and colleagues developed ChromAlignNet based on LSTM, a deep learning model that significantly improves peak alignment in GC-MS data, outperforming existing methods and requiring no user input for reference chromatograms and parameters (118). Metabolomics spectral and mass spectrometry data, essentially continuous two-dimensional data, can be encoded for processing by large language models. However, there currently lacks a dedicated natural language processing model for this preprocessing task, which warrants attention.

6.2 Metabolite annotation

Metabolite annotation involves translating spectral patterns into matching chemical molecules. The most common approach involves matching mass spectra (GC-MS's electron ionization (EI) and LC-MS's tandem MS) with reference library spectra of known molecules (150). This step is currently the most widely applied machine learning application in metabolomics, as AI can enhance both matching and reference library generation.

The main steps in spectral matching involve comparing the query spectrum with the library spectrum and calculating similarity scores. Machine learning models, such as the DNN-based SIRIUS 4 (132), the MLP-based MSNovelist (142), and the CNN-based

DeepCCs (124), are widely used for spectral prediction, fragmentation prediction, and structure identification. Notably, large language models (LLMs), a type of deep learning AI model, excel in various tasks, particularly natural language processing (NLP) (151). LLMs, such as OpenAI's GPT-X and Google's BERT, are composed of neural networks with numerous parameters and trained on vast amounts of unlabeled data via self-supervised or semi-supervised learning (151). These models are also used in protein structure prediction and drug screening (151).

LLMs' capabilities in handling complex data, pattern recognition, and reading scientific literature make them suitable for spectral matching and metabolite annotation. For example, Fakouri Baygi et al. (141) developed IDSL_MINT, a cheminformatics deep learning framework based on transformer models, which predicts molecular fingerprint descriptors from MS/MS fragmentation spectra and facilitates the training of predictive models from various mass spectral libraries. RT-Transformer is another deep neural network model that uses graph attention networks and 1D-Transformer to predict retention times via various chromatographic methods, showing competitive performance and excellent scalability, thus increasing the accuracy of metabolite identification in liquid chromatography (138).

Although still in the early exploration stages and limited by the number of tokens in semantic libraries, LLMs present a promising solution for handling larger databases and various forms of metabolomics annotation sources.

6.3 Multi-omics joint analysis

We are in an era of data explosion, where the advancement of sequencing technologies has accumulated an unimaginable volume of data over the past few decades. This includes genomics, transcriptomics, epigenomics, proteomics, and microbiomics. Several successful examples have demonstrated the powerful effect of integrating multi-omics in biological and medical research. For instance, Chakraborty et al. (152) integrated ChIP-Seq and RNA-Seq data to study head and neck squamous cell carcinoma (HNSCC), finding that tumor-specific histones H3K4me3 and H3K27ac are associated with transcriptional changes in HNSCC driver genes, such as EGFR, FGFR1, and FOXA1. Additionally, Vaske et al. (153) used the path recognition algorithm (PARADIGM) integrated with genomic model data to infer the activity of patient-specific biological pathways from multi-omics data, better identifying changes in tumor-related metabolic pathways in glioblastoma multiforme (GBM) and breast cancer datasets. Mo et al. (154) hypothesized that different molecular phenotypes can be predicted through a set of orthogonal latent variables, which represent different molecular driving factors. They proposed a new framework that uses generalized linear regression to construct a joint model of categorical and numerical variables (both continuous and discrete) from integrated genomics, epigenomics, and transcriptomics data (154). In another study, integrating metabolomics and transcriptomics revealed molecular perturbations underlying prostate cancer. The metabolite sphingosine demonstrated high specificity and sensitivity for

distinguishing prostate cancer from benign prostatic hyperplasia. Downstream of sphingosine, impaired sphingosine-1-phosphate receptor 2 signaling represents a loss of tumor suppressor function and a potential key oncogenic pathway for therapeutic targeting (155).

The combined effect of multi-omics analysis exceeds the sum of its parts, thus necessitating the integration and joint analysis of various omics data. However, different omics data have distinct analysis workflows and information densities. For example, transcriptomics data provide gene expression levels characterized by their quantitative nature, while proteomics and metabolomics data are obtained through mass spectrometry, which is more complex and diverse (156).

This integration presents several challenges, requiring robust computational methods to handle these challenges and effectively extract inter-omics relationships from heterogeneous, large-scale, and noisy biological data generated across different platforms, technologies, tissues, and species (149, 156). AI is a promising approach to address these challenges.

In metabolomics, AI technologies can be applied in multiple areas, including data preprocessing, feature extraction, pattern recognition, and data integration. For instance, deep learning algorithms can identify feature peaks in mass spectrometry data, enhancing the accuracy of metabolite identification (111, 117, 149). Additionally, machine learning methods can be used for the integrative analysis of multi-omics data, helping to uncover complex biological networks and pathways (156).

The analysis of metabolomic data and the integration of multi-omics typically involve steps such as correcting batch effects, inferring networks, reducing data dimensionality, and recognizing patterns. Batch effects, which result from differences in experimental conditions and personnel, can mask true biological variations, thereby impacting the accuracy and reliability of analyses (157).

AI is especially suited for exploring complex patterns and interpreting non-linear effects within such data. AI techniques, by learning the hidden structures and relationships in large, high-dimensional datasets, help researchers uncover complex biological signals that might be overlooked by traditional statistical methods (151). This capability is crucial for effectively integrating diverse omics data, providing a comprehensive view of biological systems. ComBat, an algorithm operating within an empirical Bayesian framework, is commonly used for batch correction (127). It adjusts batch effects by estimating and modifying batch-specific parameters to align them closer to pooled estimates, harmonizing data across different batches while preserving biological signals. Another sophisticated method for batch correction is the Mutual Nearest Neighbors (MNN) technique (128), which uses shared nearest neighbor information to correct batch discrepancies. By aligning nearest neighbors across batches, MNN effectively harmonizes datasets.

In terms of data correlation network, MMvec, inspired by word2vec, uses matrix factorization to analyze the co-occurrence patterns of microbes and metabolites, offering insights such as the metabolites produced by microbes in specific diseases (135). Similarly, Similarity Network Fusion (SNF) integrates multiple

omics datasets by creating separate networks for each data type and merging them using non-linear fusion techniques based on message-passing theory (158). This integration deepens as the networks converge through successive iterations. Network-Based Multi-group Data Integration (NetICS) provides a strategy for integrating diverse group data in tumor gene sequencing (159). It utilizes network diffusion models on directed functional interaction networks to predict the impacts of genetic, epigenetic, and miRNA variations on downstream genes and proteins.

Multi-omics datasets encompass a wide range of data types and sizes, from gene or metabolite abundances in hundreds of samples to additional dimensions like sample origin and clinical data in single-cell studies, posing the 'curse of dimensionality' (160). Moreover, the high correlation among variables can introduce multicollinearity, complicating the identification of significant biological markers. Specific machine learning methods for metabolomics, such as PCA (112), t-SNE (161), and NMF (162), facilitate data reduction and visualization. However, deep learning and large language models surpass these methods by offering superior capabilities for feature learning and contextual understanding, thus expanding potential applications in complex biological data analysis.

6.4 Biomarker research

Metabolomics not only provides crucial insights into metabolic pathways and physiological states but also plays a significant role in disease diagnosis, drug development, and nutritional research. However, the development of biomarkers faces considerable challenges, as traditional methods are time-consuming, costly, and often lack accuracy (163, 164). The rapid advancement of AI technologies, particularly machine learning and deep learning, has significantly increased the efficiency and accuracy of biomarker development. Machine learning and deep learning algorithms, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) (112), and ANNs (112), excel in handling complex and high-dimensional data, capturing patterns and relationships that traditional methods might overlook. Compared with conventional models, the incorporation of these technologies has improved the predictive capabilities of biomarkers.

For instance, in a study on gastric cancer diagnosis and prognosis prediction, Chen et al. (165) utilized a machine learning random forest model to analyze plasma samples from 702 participants across multiple centers. They developed a diagnostic model comprising ten metabolites, achieving a sensitivity of 0.905, significantly surpassing the sensitivity of traditional cancer protein marker methods (less than 0.40) (165). Similarly, in a study by Kuwabara et al. (166) liquid chromatography-mass spectrometry (LC-MS) and the alternative decision tree (ADTree) algorithm were used to analyze 2602 saliva samples for colorectal cancer detection. Their model achieved an AUC value of 0.870 in distinguishing colorectal cancer from healthy controls, whereas traditional detection methods typically have AUC values around 0.70 (166).

The methods for establishing predictive and diagnostic models in metabolomics primarily include traditional machine learning

algorithms and deep learning algorithms (167, 168). Among traditional machine learning algorithms, logistic regression (LR) is widely used for binary classification problems due to its simplicity, ease of interpretation, and fast computation (169). It has shown good performance in disease prediction based on metabolites. For example, in a study on endometrial cancer (EC) by Bahado-Singh et al. (170) a logistic regression model was developed using multiple metabolites and demographic characteristics. This model combined C14:2, phosphatidylcholine with acyl-alkyl residue sum C38:1 (PCae C38:1), and 3-hydroxybutyric acid, achieving an AUC (95% CI) of 0.826 (0.706-0.946), with a sensitivity of 82.6% and specificity of 70.8% (170). However, logistic regression is less capable of modeling complex non-linear relationships and may underperform when handling high-dimensional data with complex interactions.

Random forest (RF), on the other hand, improves model robustness and resistance to overfitting by constructing multiple decision trees and averaging their results, making it suitable for high-dimensional data processing. For instance, in a study on colorectal cancer by Telleria et al., RF and logistic regression models were used to develop an accurate predictive model based on several metabolites. This model combined hemoglobin (Hgb), bilirubin E,E, lactosyl-N-palmitoyl-sphingosine, glycocholate sulfate, and STLVT, achieving an accuracy of 91.67% (95% CI 0.7753-0.9825), with a sensitivity of 0.7 and specificity of 1 (171).

Support vector machine (SVM) constructs hyperplanes for classification, making it suitable for small samples and high-dimensional data with good generalization ability. In a study on breast cancer by An et al., an SVM model was developed based on 47 metabolites. This model achieved high accuracy in breast cancer prediction (AUC = 1), with an AUC of 0.794 for breast cancer vs healthy controls (HC), and 0.879 for benign vs HC in the testing cohort (172).

Bayesian methods, based on Bayes' theorem, can handle uncertainty in data and are suitable for small sample learning, although they require high computational complexity and rely on prior information. In a study on early lung cancer by Xie et al., a Naive Bayes algorithm was used to develop a predictive model based on six plasma metabolites. This model significantly distinguished early lung cancer patients from healthy individuals, achieving an AUC of 0.989, with a sensitivity of 98.1% and specificity of 100.0%. The study also identified the top five important metabolites as potential biomarkers for early lung cancer screening (173).

CNN uses combinations of convolutional layers, pooling layers, and fully connected layers to extract spatial features, making them suitable for image and time-series data processing. In metabolomics research, CNNs have demonstrated excellent feature extraction and prediction performance, although they require large amounts of data and high computational resources. In a study on cholangiocarcinoma and pancreatic adenocarcinoma by Urman et al., a neural network (NN) algorithm was used to develop predictive models based on multiple lipid and protein biomarkers. This model differentiated between benign strictures and cholangiocarcinoma patients with an AUC of 0.984, a sensitivity of 94.1%, and a specificity of 92.3%. The same method was also used

TABLE 2 AI-based metabolomics studies for disease prediction and diagnosis.

Disease Category	Discovery Cohort Size	Validation Cohort Size	Metabolomics Technique	Feature Count	Feature Types	Modeling Method	AUC Value	Sensitivity	Specificity	Reference
Cancer cachexia	192	–	NMR	15	Mixed Metabs	RF, LR	0.991	0.9464	–	(176)
Endometrial cancer	116	47	NMR	181	Mixed Metabs	LR	0.826	0.826	0.708	(170)
Colorectal cancer	130	344	Targeted Metabolomics	3	Amino Acids	LR	0.955	0.968	0.833	(170)
Colorectal cancer	60	60	UPLC-MS/MS	25	Mixed Metabs	RF, LR	0.917	0.7	1	(171)
Breast cancer	114	75	NMR	3	Lipoprotein	LASSO	0.999, 0.892	–	–	(177)
Breast cancer	75	–	LC-MS	19	Lipids	SVM	0.946	0.954	0.916	(178)
Colorectal cancer	82	79	UHPLC-HRMS	10	Lipids	OPLS-DA, SVM	0.94	–	–	(179)
Gastric cardia adenocarcinoma	276	588	UPLC-MS/MS	25	Mixed Metabs	RF, LR, SVM, Cox regression	0.869-0.900	–	–	(180)
Rectal cancer	106	–	UHPLC-QTOF-MS	8	Mixed Metabs	ANOVA, PLS-DA	0.54-0.67	–	–	(181)
Lung cancer	131	–	LC-MS	241	Mixed Metabs	LR	0.886-0.934	–	–	(182)
Colorectal adenomas, Colorectal cancer	71	81	LC-MS/MS	8	Proteins	LR	–	53%-66%	0.95	(182)
Hepatocellular carcinoma, chronic liver disease	110	–	LC-MS	12	Mixed Metabs	RF	0.84	0.848	0.924	(183)
Lung cancer	110	43	LC-MS/MS	6	Mixed Metabs	NB	0.989	0.981	1	(173)
Non-small cell lung cancer	360	201	UPLC-MS/MS	8	Mixed Metabs	RF	0.981 (diagnosis), 0.954 (efficacy)	–	–	(184)
Breast cancer	282	–	SELDI-TOF-MS	3	Proteins	SVM	–	0.9645	0.9487	(185)
Breast cancer	91	20 (healthy controls)	LC-MS	1269	Mixed Metabs	RF	1	1	1	(186)
Primary central nervous system lymphoma	68	34	UHPLC-MS/MS	14	Amino Acids	LR	0.83	–	–	(187)
Colorectal cancer	514	–	LC-MS/MS	5	Mixed Metabs	LR	0.89, 0.90	–	–	(188)
Gastric cancer	72	29	LC-TOF-MS	8	Lipids	SVM	0.915	0.69	–	(189)
Non-small cell lung cancer	112	123	SELDI-TOF-MS	3	Proteins	SVM	–	0.9656	0.9479	(190)
Intrahepatic Cholangiocarcinoma	100	100	LC-MS/MS	10	Lipids	LR	0.993	–	–	(191)

(Continued)

TABLE 2 Continued

Disease Category	Discovery Cohort Size	Validation Cohort Size	Metabolomics Technique	Feature Count	Feature Types	Modeling Method	AUC Value	Sensitivity	Specificity	Reference
Gastric cancer	181	66	UPLC-QTOF-MS	6	Bacterial Genera	RF	0.85	–	–	(192)
Chronic Pancreatitis	160	502	GC-MS	8	Mixed Metabs	NB	0.85	–	–	(193)
Non-small cell lung cancer	250	250	UHPLC-QTOF-MS	24	Mixed Metabs	LR	0.91, 0.94	–	–	(194)
Oral Squamous Cell Carcinoma	68	–	GC-MS	3	Mixed Metabs	RF	0.91	–	–	(195)
Pancreatic Cancer	55	16	GC-MS/MS	16	Mixed Metabs	LR	–	0.907	0.895	(196)
Non-Small Cell Lung Cancer	251	103	LC-MS/MS	7	Mixed Metabs	LR	0.912	–	–	(197)
Breast Cancer	113	99	LC-IT-MS	31	Nucleosides	SVM	–	0.8767	0.899	(198)
Breast Cancer	55	55	UHPLC/Q-TOF-MS	16	Mixed Metabs	LASSO	0.973	–	–	(199)
Cervical Cancer, CIN	69	–	UPLC-QTOF-MS	28	Mixed Metabs	LR	>0.80	–	–	(200)
Maternal Pregnancy Smoking	894	–	High-resolution metabolomics (HMR)	7	Cotinine and Hydroxycotinine	LR	0.81	–	–	(201)
Prostate cancer	247	139	GC-MS	22538	Mixed Metabs	LR	0.99	–	–	(202)
Pancreatic cancer	82	82	LC-MS	3	Mixed Metabs	LASSO	0.784	–	–	(203)
Papillary Thyroid Carcinoma	108	116	SELDI-TOF-MS	3	Proteins	SVM	–	0.9515	0.9397	(204)
Thyroid Nodules	78	–	Untargeted Metabolomics	15	Mixed Metabs	DNN	0.945	–	–	(205)
Oral, Breast, Pancreatic cancer	215	–	CE-TOF-MS	57	Mixed Metabs	LR	0.865 (oral), 0.973 (breast), 0.993 (pancreatic)	–	–	(206)
Rectal cancer	106	–	LC-MS	57	Mixed Metabs	PLS-DA, LR	0.88, 0.81, 0.84	0.88, 0.81, 0.84	–	(207)
Type 2 diabetes, Coronary artery disease	1538	2521	LC-MS	111	Mixed Metabs	PLS-DA, LR	–	0.73, 0.70	–	(208)
Colorectal cancer	282	291	GC-MS	29	Mixed Metabs	LR	0.996	0.993	0.938	(209)
Lung cancer	51	–	SPME-GC-MS	19	Mixed Metabs	RF	0.94	0.885	–	(210)

NB, naive bayes; LASSO, LASSO regression; LR, logistic regression; SVM, sSupport vector machine; RF, random forest.

to distinguish pancreatic adenocarcinoma patients from control groups, achieving an AUC of 0.98, a sensitivity of 88%, and a specificity of 100% (174). Furthermore, the study evaluated the performance of other machine-learning algorithms. The authors used a Bayesian variant of the general linear model (BGLM) and the C5.0 decision tree algorithm on the same data. Both C5.0 and BGLM demonstrated good performance in feature selection and prediction, although their predictive power was slightly lower than that of the NN algorithm.

ANN simulates the structure of human brain neurons and learns complex features and patterns through multiple layers of neurons and weight adjustments (116). ANN has strong feature extraction capabilities and is suitable for processing large-scale data, but it requires extensive data for training and high computational resources, and the model interpretability is poor. In a study on oral cancer by Monedeiro et al., an ANN model was developed based on nine relevant volatile organic compounds (VOCs). These compounds included 1-octen-3-ol, hexanoic acid, E-2-octenal, heptanoic acid, octanoic acid, E-2-nonenal, nonanoic acid, 2,4-decadienal, and 9-undecenoic acid. The model's performance was assessed using 10-fold cross-validation and receiver operating characteristic curves, achieving an overall accuracy of 90%, with 100% sensitivity and specificity for oral cancer cases (175).

In Table 2, we summarize the recent applications of different AI technologies in establishing predictive or diagnostic models through metabolomics. Various predictive models have achieved very high performance in their respective cohorts. It is important to note that most studies have employed cross-validation with separated cohorts and lack independent external validation cohorts. Most cohorts consist of around a hundred samples, which may lead to overfitting when using AI models to build predictive models, potentially resulting in poor validation performance in real-world external cohorts (167). The main reasons for the lack of external validation cohorts are data acquisition difficulties and high costs. Establishing comprehensive and interconnected channels and large integrated databases through AI technology may be a feasible approach in the future.

7 Discussion

This review highlights the current state and strategies for applying AI algorithms to tumor metabolism studies. By exploring metabolic differences between tumor and normal cells, particularly from the perspectives of metabolomics and TME interactions, this study demonstrates the significant advantages of AI in data preprocessing, feature extraction, pattern recognition, and data integration. These technologies not only enhance the efficiency and accuracy of data analysis but also offer new perspectives for personalized medicine and precision therapy.

Compared to existing studies, this review emphasizes the unique advantages of LLMs in tumor metabolism research. LLMs can handle complex data, identify patterns, and utilize their strengths in natural language processing to improve metabolite annotation and multi-omics integration (151). This innovation

provides a more comprehensive understanding and higher analytical precision in tumor research.

Despite the significant potential of AI technology in tumor metabolism research, there are some limitations. First, the training of AI models relies on large-scale, high-quality datasets, but the acquisition and standardization of such data remain challenging (151). Second, the complexity and black-box nature of AI algorithms may lead to difficulties in interpreting results, necessitating further algorithmic improvements and transparency. Additionally, the complexity and heterogeneity of the TME remain a research challenge, requiring more effective simulation and study of this complex system (211).

To address these limitations, future research should focus on several key areas. Firstly, further refinement of AI algorithms, especially in improving data quality and sample size, is essential to enhance their reliability in practical applications. Secondly, broadening the application of AI in the integration of multi-omics data will allow for the exploration of synergistic effects between different omics data and the revelation of more complex biological networks. Thirdly, deepening the study of the interaction between the TME and metabolism, using AI technology to simulate complex biological systems, will advance tumor research toward more refined and personalized directions.

The application of AI in tumor metabolism research discussed in this review holds significant theoretical importance and shows great promise in practical applications. AI technology can significantly improve the efficiency of tumor diagnosis and treatment, particularly in personalized and precision medicine (212). With continuous advancements in AI technology, future breakthroughs in clinical applications will further drive the development of biomedical research.

Author contributions

LH: Conceptualization, Funding acquisition, Investigation, Writing – original draft, Writing – review & editing. ZJ: Formal Analysis, Investigation, Resources, Writing – original draft. HZ: Investigation, Methodology, Validation, Writing – original draft. YS: Conceptualization, Funding acquisition, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (32270633, 22005343), the Shenzhen Science and Technology Program (ZDSYS20220606101604009, KCXFZ20201221173008022), the Cooperation Fund of CHCAMS and SZCH (CFA202201010), the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academic of Medical Sciences and Peking Union Medical College, Shenzhen (E010124001, SZ2020ZD004).

Acknowledgments

Some icons or graphic elements in all of our Figures were adapted from BioRender.com (2024), and final schematic illustrations were created and integrated by our original design.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Jain B, Pandey S, ScienceDirect. *Understanding Cancer: From Basics to Therapeutics*. London, England: Academic Press (2022).
- Liu JJ, Chen Q, Feng LZ, Liu Z. Nanomedicine for tumor microenvironment modulation and cancer treatment enhancement. *Nano Today*. (2018) 21:55–73. doi: 10.1016/j.nantod.2018.06.008
- Higginson IJ, Costantini M. Dying with cancer, living well with advanced cancer. *Eur J Cancer*. (2008) 44:1414–24. doi: 10.1016/j.ejca.2008.02.024
- Zhang Z, Ding C, Sun T, Wang L, Chen C. Tumor therapy strategies based on microenvironment-specific responsive nanomaterials. *Adv Health Mater*. (2023) 12:e2300153. doi: 10.1002/adhm.202300153
- Wang Z, Zhao F, Xu C, Zhang Q, Ren H, Huang X, et al. Metabolic reprogramming in skin wound healing. *Burns Trauma*. (2024) 12:tkad047. doi: 10.1093/burnst/tkad047
- Rice CM, Davies LC, Subleski JJ, Maio N, Gonzalez-Cotto M, Andrews C, et al. Tumour-elicited neutrophils engage mitochondrial metabolism to circumvent nutrient limitations and maintain immune suppression. *Nat Commun*. (2018) 9:5099. doi: 10.1038/s41467-018-07505-2
- Zhang AH, Sun H, Qiu S, Wang XJ. Metabolomics in noninvasive breast cancer. *Clin Chim Acta*. (2013) 424:3–7. doi: 10.1016/j.cca.2013.05.003
- Lin C, Tian Q, Guo S, Xie D, Cai Y, Wang Z, et al. Metabolomics for clinical biomarker discovery and therapeutic target identification. *Molecules*. (2024) 29:2198. doi: 10.3390/molecules29102198
- Pang H, Hu Z. Metabolomics in drug research and development: the recent advances in technologies and applications. *Acta Pharm Sin B*. (2023) 13:3238–51. doi: 10.1016/j.apsb.2023.05.021
- Gaca-Tabaszewska M, Bogusiewicz J, Bojko B. Metabolomic and lipidomic profiling of gliomas—a new direction in personalized therapies. *Cancers (Basel)*. (2022) 14:5041. doi: 10.3390/cancers14205041
- Taunk K, Jajula S, Bhavsar PP, Choudhari M, Bhanuse S, Tamhankar A, et al. The process of metabolomics in cancer research: current trends, challenges and future perspectives. *Mol Cell Biochem*. (2024). doi: 10.1007/s11010-024-05041-w
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013
- NooNepalle SKR, Karabon L, Chiappinelli KB, Villagra A. Editorial: genetic and epigenetic control of immune responses. *Front Immunol*. (2021) 12:775101. doi: 10.3389/fimmu.2021.775101
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. (2012) 13:484–92. doi: 10.1038/nrg3230
- He S, Jiang F. Editorial: the interconnection between epigenetic modifications and the tumor microenvironment. *Front Oncol*. (2023) 13:1166676. doi: 10.3389/fonc.2023.1166676
- Yang F, Yang Y, Qiu Y, Tang L, Xie L, Guan X. Long non-coding RNAs as regulators for targeting breast cancer stem cells and tumor immune microenvironment: biological properties and therapeutic potential. *Cancers (Basel)*. (2024) 16:290. doi: 10.3390/cancers16020290
- Zou W, Wolchok JD, Chen L. Pd-L1 (B7-H1) and pd-1 pathway blockade for cancer therapy: mechanisms, response biomarkers, and combinations. *Sci Transl Med*. (2016) 8:328rv4. doi: 10.1126/scitranslmed.aad7118
- Kim SJ, Khadka D, Seo JH. Interplay between solid tumors and tumor microenvironment. *Front Immunol*. (2022) 13:882718. doi: 10.3389/fimmu.2022.882718
- Agrawal B. New therapeutic targets for cancer: the interplay between immune and metabolic checkpoints and gut microbiota. *Clin Transl Med*. (2019) 8:23. doi: 10.1186/s40169-019-0241-x

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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.

- Wang Y, Wang Y, Ren Y, Zhang Q, Yi P, Cheng C. Metabolic modulation of immune checkpoints and novel therapeutic strategies in cancer. *Semin Cancer Biol*. (2022) 86:542–65. doi: 10.1016/j.semcancer.2022.02.010
- Langsten KL, Kim JH, Sarver AL, Dewhirst M, Modiano JF. Comparative approach to the temporo-spatial organization of the tumor microenvironment. *Front Oncol*. (2019) 9:1185. doi: 10.3389/fonc.2019.01185
- Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol*. (2020) 30:R921–R5. doi: 10.1016/j.cub.2020.06.081
- Ribatti D, ScienceDirect. *Tumor Microenvironment Regulation of Tumor Expansion*. Waltham, Massachusetts: Elsevier (2021).
- Damaghi M, West J, Robertson-Tessi M, Xu L, Ferrall-Fairbanks MC, Stewart PA, et al. The harsh microenvironment in early breast cancer selects for a warburg phenotype. *Proc Natl Acad Sci U.S.A.* (2021) 118:e2011342118. doi: 10.1073/pnas.2011342118
- Martinez-Outschoorn UE, Peiris-Pages M, Pestell RG, Sotgia F, Lisanti MP. Cancer metabolism: A therapeutic perspective. *Nat Rev Clin Oncol*. (2017) 14:113. doi: 10.1038/nrclinonc.2017.1
- Yuan Z, Li Y, Zhang S, Wang X, Dou H, Yu X, et al. Extracellular matrix remodeling in tumor progression and immune escape: from mechanisms to treatments. *Mol Cancer*. (2023) 22:48. doi: 10.1186/s12943-023-01744-8
- de Visser KE, Joyce JA. The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell*. (2023) 41:374–403. doi: 10.1016/j.ccell.2023.02.016
- Chen C, Wang Z, Ding Y, Qin Y. Tumor microenvironment-mediated immune evasion in hepatocellular carcinoma. *Front Immunol*. (2023) 14:1133308. doi: 10.3389/fimmu.2023.1133308
- Baghban R, Roshangar L, Jahanban-Esfahlan R, Seidi K, Ebrahimi-Kalan A, Jaymand M, et al. Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Commun Signal*. (2020) 18:59. doi: 10.1186/s12964-020-0530-4
- Seager RJ, Hajal C, Spill F, Kamm RD, Zaman MH. Dynamic interplay between tumour, stroma and immune system can drive or prevent tumour progression. *Converg Sci Phys Oncol*. (2017) 3:034002. doi: 10.1088/2057-1739/aa7e86
- Chekulayev V, Mado K, Shevchuk I, Koit A, Kaldma A, Klepinin A, et al. Metabolic remodeling in human colorectal cancer and surrounding tissues: alterations in regulation of mitochondrial respiration and metabolic fluxes. *Biochem Biophys Res*. (2015) 4:111–25. doi: 10.1016/j.bbrep.2015.08.020
- Rossi T, Vergara D, Fanini F, Maffia M, Bravaccini S, Pirini F. Microbiota-derived metabolites in tumor progression and metastasis. *Int J Mol Sci*. (2020) 21:5786. doi: 10.3390/ijms21165786
- Sousa CM, Biancur DE, Wang X, Halbrook CJ, Sherman MH, Zhang L, et al. Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion. *Nature*. (2016) 536:479–83. doi: 10.1038/nature19084
- Meng Y, Guo D, Lin L, Zhao H, Xu W, Luo S, et al. Glycolytic enzyme pfkl governs lipolysis by promoting lipid droplet-mitochondria tethering to enhance beta-oxidation and tumor cell proliferation. *Nat Metab*. (2024) 6:1092–107. doi: 10.1038/s42255-024-01047-2
- Linder SJ, Bernasocchi T, Martinez-Pastor B, Sullivan KD, Galbraith MD, Lewis CA, et al. Inhibition of the proline metabolism rate-limiting enzyme P5cs allows proliferation of glutamine-restricted cancer cells. *Nat Metab*. (2023) 5:2131–47. doi: 10.1038/s42255-023-00919-3
- Khan F, Lin Y, Ali H, Pang L, Dunterman M, Hsu WH, et al. Lactate dehydrogenase a regulates tumor-macrophage symbiosis to promote glioblastoma progression. *Nat Commun*. (2024) 15:1987. doi: 10.1038/s41467-024-46193-z

37. Pebriana RB, Sanchez-Lopez E, Giera M. (Pre)Clinical metabolomics analysis. *Methods Mol Biol.* (2025) 2855:3–19. doi: 10.1007/978-1-0716-4116-3_1
38. Castelli FA, Rosati G, Moguet C, Fuentes C, Marrugo-Ramirez J, Lefebvre T, et al. Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests. *Anal Bioanal Chem.* (2022) 414:759–89. doi: 10.1007/s00216-021-03586-z
39. Letertre MPM, Dervilly G, Giraudeau P. Combined nuclear magnetic resonance spectroscopy and mass spectrometry approaches for metabolomics. *Anal Chem.* (2021) 93:500–18. doi: 10.1021/acs.analchem.0c04371
40. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet.* (2015) 8:192–206. doi: 10.1161/CIRCGENETICS.114.000216
41. Wurtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: A primer on -omic technologies. *Am J Epidemiol.* (2017) 186:1084–96. doi: 10.1093/aje/kwx016
42. Perez de Souza L, Alseikh S, Scossa F, Fernie AR. Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. *Nat Methods.* (2021) 18:733–46. doi: 10.1038/s41592-021-01116-4
43. Liebal UW, Phan ANT, Sudhakar M, Raman K, Blank LM. Machine learning applications for mass spectrometry-based metabolomics. *Metabolites.* (2020) 10:243. doi: 10.3390/metabo10060243
44. Bauermeister A, Mannochio-Russo H, Costa-Lotufo LV, Jarmusch AK, Dorrestein PC. Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol.* (2022) 20:143–60. doi: 10.1038/s41579-021-00621-9
45. Giera M, Yanes O, Siuzdak G. Metabolite discovery: biochemistry's scientific driver. *Cell Metab.* (2022) 34:21–34. doi: 10.1016/j.cmet.2021.11.005
46. Alseikh S, Aharoni A, Brotman Y, Contrepois K, D'Auria J, Ewald J, et al. Mass spectrometry-based metabolomics: A guide for annotation, quantification and best reporting practices. *Nat Methods.* (2021) 18:747–56. doi: 10.1038/s41592-021-01197-1
47. Collins SL, Koo I, Peters JM, Smith PB, Patterson AD. Current challenges and recent developments in mass spectrometry-based metabolomics. *Annu Rev Anal Chem (Palo Alto Calif).* (2021) 14:467–87. doi: 10.1146/annurev-anchem-091620-015205
48. Zhao Y, Ma C, Cai R, Xin L, Li Y, Ke L, et al. Nmr and ms reveal characteristic metabolome atlas and optimize esophageal squamous cell carcinoma early detection. *Nat Commun.* (2024) 15:2463. doi: 10.1038/s41467-024-46837-0
49. Deng Y, Yao Y, Wang Y, Yu T, Cai W, Zhou D, et al. An end-to-end deep learning method for mass spectrometry data analysis to reveal disease-specific metabolic profiles. *Nat Commun.* (2024) 15:7136. doi: 10.1038/s41467-024-51433-3
50. Su Y, Ko ME, Cheng H, Zhu R, Xue M, Wang J, et al. Multi-omic single-cell snapshots reveal multiple independent trajectories to drug tolerance in a melanoma cell line. *Nat Commun.* (2020) 11:2345. doi: 10.1038/s41467-020-15956-9
51. Sun C, Li T, Song X, Huang L, Zang Q, Xu J, et al. Spatially resolved metabolomics to discover tumor-associated metabolic alterations. *Proc Natl Acad Sci U.S.A.* (2019) 116:52–7. doi: 10.1073/pnas.1808950116
52. Jin N, Bi A, Lan X, Xu J, Wang X, Liu Y, et al. Identification of metabolic vulnerabilities of receptor tyrosine kinases-driven cancer. *Nat Commun.* (2019) 10:2701. doi: 10.1038/s41467-019-10427-2
53. Brown JS, Amend SR, Austin RH, Gatenby RA, Hammarlund EU, Pienta KJ. Updating the definition of cancer. *Mol Cancer Res.* (2023) 21:1142–7. doi: 10.1158/1541-7786.MCR-23-0411
54. Nong S, Han X, Xiang Y, Qian Y, Wei Y, Zhang T, et al. Metabolic reprogramming in cancer: mechanisms and therapeutics. *MedComm.* (2023) 4:e218. doi: 10.1002/mco.2.218
55. Otto Warburg FW, Negelein E. The metabolism of tumors in the body. *J Gen Physiol.* (1927) 8:519–30. doi: 10.1085/jgp.8.6.519
56. Warburg O. The chemical constitution of respiration ferment. *Science.* (1928) 68:437–43. doi: 10.1126/science.68.1767.437
57. Warburg O. On respiratory impairment in cancer cells. *Science.* (1956) 124:269–70. doi: 10.1126/science.124.3215.26
58. Warburg O. On the origin of cancer cells. *Science.* (1956) 123:309–14. doi: 10.1126/science.123.3191.309
59. Hu J, Locasale JW, Bielas JH, O'Sullivan J, Sheahan K, Cantley LC, et al. Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat Biotechnol.* (2013) 31:522–9. doi: 10.1038/nbt.2530
60. Kim J, DeBerardinis RJ. Mechanisms and implications of metabolic heterogeneity in cancer. *Cell Metab.* (2019) 30:434–46. doi: 10.1016/j.cmet.2019.08.013
61. Martinez-Reyes J, Chandel NS. Cancer metabolism: looking forward. *Nat Rev Cancer.* (2021) 21:669–80. doi: 10.1038/s41568-021-00378-6
62. Han J, Li Q, Chen Y, Yang Y. Recent metabolomics analysis in tumor metabolism reprogramming. *Front Mol Biosci.* (2021) 8:763902. doi: 10.3389/fmolb.2021.763902
63. Toller-Kawahisa JE, Hiroki CH, Silva CMS, Nascimento DC, Publio GA, Martins TV, et al. The metabolic function of pyruvate kinase M2 regulates reactive oxygen species production and microbial killing by neutrophils. *Nat Commun.* (2023) 14:4280. doi: 10.1038/s41467-023-40021-6
64. Jemal M, Getinet M, Amare GA, Tegegne BA, Baylie T, Mengistu EF, et al. Non-metabolic enzyme function of pyruvate kinase M2 in breast cancer. *Front Oncol.* (2024) 14:1450325. doi: 10.3389/fonc.2024.1450325
65. Zhang H, Zhu M, Zhao A, Shi T, Xi Q. B7-H3 regulates anti-tumor immunity and promotes tumor development in colorectal cancer. *Biochim Biophys Acta Rev Cancer.* (2024) 1879:189031. doi: 10.1016/j.bbcan.2023.189031
66. Cao J, Liao S, Zeng F, Liao Q, Luo G, Zhou Y. Effects of altered glycolysis levels on cd8(+) T cell activation and function. *Cell Death Dis.* (2023) 14:407. doi: 10.1038/s41419-023-05937-3
67. Zhang Y, Qin N, Wang X, Liang R, Liu Q, Geng R, et al. Glycogen metabolism-mediated intercellular communication in the tumor microenvironment influences liver cancer prognosis. *Oncol Res.* (2024) 32:563–76. doi: 10.32604/or.2023.029697
68. Pandey P, Khan F, Upadhyay TK, Maqsood R. Review to understand the crosstalk between immunotherapy and tumor metabolism. *Molecules.* (2023) 28:862. doi: 10.3390/molecules28020862
69. Zhang H, Li S, Wang D, Liu S, Xiao T, Gu W, et al. Metabolic reprogramming and immune evasion: the interplay in the tumor microenvironment. *biomark Res.* (2024) 12:96. doi: 10.1186/s40364-024-00646-1
70. Pilon-Thomas S, Kodumudi KN, El-Kenawi AE, Russell S, Weber AM, Luddy K, et al. Neutralization of tumor acidity improves antitumor responses to immunotherapy. *Cancer Res.* (2016) 76:1381–90. doi: 10.1158/0008-5472.CAN-15-1743
71. Counihan JL, Grossman EA, Nomura DK. Cancer metabolism: current understanding and therapies. *Chem Rev.* (2018) 118:6893–923. doi: 10.1021/acs.chemrev.7b00775
72. Scalise M, Pochini L, Console L, Losso MA, Indiveri C. The human slc1a5 (Asct2) amino acid transporter: from function to structure and role in cell biology. *Front Cell Dev Biol.* (2018) 6:96. doi: 10.3389/fcell.2018.00096
73. Spinelli JB, Yoon H, Ringel AE, Jeanfavre S, Clish CB, Haigis MC. Metabolic recycling of ammonia via glutamate dehydrogenase supports breast cancer biomass. *Science.* (2017) 358:941–6. doi: 10.1126/science.aam9305
74. Wu J, Liu N, Chen J, Tao Q, Li Q, Li J, et al. The tricarboxylic acid cycle metabolites for cancer: friend or enemy. *Res (Wash D C).* (2024) 7:351. doi: 10.34133/research.0351
75. Han C, Ge M, Ho PC, Zhang L. Fueling T-cell antitumor immunity: amino acid metabolism revisited. *Cancer Immunol Res.* (2021) 9:1373–82. doi: 10.1158/2326-6066.CIR-21-0459
76. Parry RV, Chemnitz JM, Frauwrith KA, Lanfranco AR, Braunstein I, Kobayashi SV, et al. CtlA-4 and pd-1 receptors inhibit T-cell activation by distinct mechanisms. *Mol Cell Biol.* (2005) 25:9543–53. doi: 10.1128/MCB.25.21.9543-9553.2005
77. Boussiotis VA, Chatterjee P, Li L. Biochemical signaling of pd-1 on T cells and its functional implications. *Cancer J.* (2014) 20:265–71. doi: 10.1097/PPO.0000000000000059
78. Freereman AJ, Johnson AR, Sacks GN, Milner JJ, Kirk EL, Troester MA, et al. Metabolic reprogramming of macrophages: glucose transporter 1 (Glut1)-mediated glucose metabolism drives a proinflammatory phenotype. *J Biol Chem.* (2014) 289:7884–96. doi: 10.1074/jbc.M113.522037
79. Palaskas NJ, Garcia JD, Shirazi R, Shin DS, Puig-Saus C, Braas D, et al. Global alteration of T-lymphocyte metabolism by pd-L1 checkpoint involves a block of *de novo* nucleoside phosphate synthesis. *Cell Discovery.* (2019) 5:62. doi: 10.1038/s41421-019-0130-x
80. Huang F, Zhao Y, Zhao J, Wu S, Jiang Y, Ma H, et al. Upregulated slc1a5 promotes cell growth and survival in colorectal cancer. *Int J Clin Exp Pathol.* (2014) 7:6006–14.
81. Hu Y, Xiang J, Su L, Tang X. The regulation of nitric oxide in tumor progression and therapy. *J Int Med Res.* (2020) 48:300060520905985. doi: 10.1177/0300060520905985
82. Choudhari SK, Chaudhary M, Bagde S, Gadail AR, Joshi V. Nitric oxide and cancer: A review. *World J Surg Oncol.* (2013) 11:118. doi: 10.1186/1477-7819-11-118
83. Geiger R, Rieckmann JC, Wolf T, Basso C, Feng Y, Fuhrer T, et al. L-arginine modulates T cell metabolism and enhances survival and anti-tumor activity. *Cell.* (2016) 167:829–42 e13. doi: 10.1016/j.cell.2016.09.031
84. Munn DH, Mellor AL. Indoleamine 2,3 dioxygenase and metabolic control of immune responses. *Trends Immunol.* (2013) 34:137–43. doi: 10.1016/j.it.2012.10.001
85. Liu Y, Liang X, Dong W, Fang Y, Lv J, Zhang T, et al. Tumor-repopulating cells induce pd-1 expression in cd8(+) T cells by transferring kynurenine and ahr activation. *Cancer Cell.* (2018) 33:480–94 e7. doi: 10.1016/j.ccell.2018.02.005
86. Le Naour J, Galluzzi L, Zitvogel L, Kroemer G, Vacchelli E. Trial watch: ido inhibitors in cancer therapy. *Oncoimmunology.* (2020) 9:1777625. doi: 10.1080/2162402X.2020.1777625
87. Matsushita Y, Nakagawa H, Koike K. Lipid metabolism in oncology: why it matters, how to research, and how to treat. *Cancers (Basel).* (2021) 13:474. doi: 10.3390/cancers13030474
88. Yang P, Su C, Luo X, Zeng H, Zhao L, Wei L, et al. Dietary oleic acid-induced cd36 promotes cervical cancer cell growth and metastasis via up-regulation src/erk pathway. *Cancer Lett.* (2018) 438:76–85. doi: 10.1016/j.canlet.2018.09.006
89. Pan J, Fan Z, Wang Z, Dai Q, Xiang Z, Yuan F, et al. Cd36 mediates palmitate acid-induced metastasis of gastric cancer via akt/gsk-3beta/beta-catenin pathway. *J Exp Clin Cancer Res.* (2019) 38:52. doi: 10.1186/s13046-019-1049-7

90. Ladanyi A, Mukherjee A, Kenny HA, Johnson A, Mitra AK, Sundaresan S, et al. Adipocyte-induced cd36 expression drives ovarian cancer progression and metastasis. *Oncogene*. (2018) 37:2285–301. doi: 10.1038/s41388-017-0093-z
91. Cheng C, Geng F, Cheng X, Guo D. Lipid metabolism reprogramming and its potential targets in cancer. *Cancer Commun (Lond)*. (2018) 38:27. doi: 10.1186/s40880-018-0301-4
92. Yang Y, Hsu JM, Sun L, Chan LC, Li CW, Hsu JL, et al. Palmitoylation stabilizes pd-L1 to promote breast tumor growth. *Cell Res*. (2019) 29:83–6. doi: 10.1038/s41422-018-0124-5
93. Ma S, Ming Y, Wu J, Cui G. Cellular metabolism regulates the differentiation and function of T-cell subsets. *Cell Mol Immunol*. (2024) 21:419–35. doi: 10.1038/s41423-024-01148-8
94. Hao JW, Wang J, Guo H, Zhao YY, Sun HH, Li YF, et al. Cd36 facilitates fatty acid uptake by dynamic palmitoylation-regulated endocytosis. *Nat Commun*. (2020) 11:4765. doi: 10.1038/s41467-020-18565-8
95. Ma X, Xiao L, Liu L, Ye L, Su P, Bi E, et al. Cd36-mediated ferroptosis dampens intratumoral cd8(+) T cell effector function and impairs their antitumor ability. *Cell Metab*. (2021) 33:1001–12 e5. doi: 10.1016/j.cmet.2021.02.015
96. Tang Y, Chen Z, Zuo Q, Kang Y. Regulation of cd8+ T cells by lipid metabolism in cancer progression. *Cell Mol Immunol*. (2024) 21:1215–30. doi: 10.1038/s41423-024-01224-z
97. Yang W, Bai Y, Xiong Y, Zhang J, Chen S, Zheng X, et al. Potentiating the antitumor response of cd8(+) T cells by modulating cholesterol metabolism. *Nature*. (2016) 531:651–5. doi: 10.1038/nature17412
98. Kunisawa J. Metabolic changes during B cell differentiation for the production of intestinal iga antibody. *Cell Mol Life Sci*. (2017) 74:1503–9. doi: 10.1007/s00018-016-2414-8
99. Fu Y, Wang L, Yu B, Xu D, Chu Y. Immunometabolism shapes B cell fate and functions. *Immunology*. (2022) 166:444–57. doi: 10.1111/imm.13499
100. Lam WY, Jash A, Yao CH, D'Souza L, Wong R, Nunley RM, et al. Metabolic and transcriptional modules independently diversify plasma cell lifespan and function. *Cell Rep*. (2018) 24:2479–92 e6. doi: 10.1016/j.celrep.2018.07.084
101. Qian T, Hong J, Wang L, Wang Z, Lu Z, Li Y, et al. Regulation of cd11b by hif-1alpha and the stat3 signaling pathway contributes to the immunosuppressive function of B cells in inflammatory bowel disease. *Mol Immunol*. (2019) 111:162–71. doi: 10.1016/j.molimm.2019.04.005
102. Zhong Z, Nan K, Weng M, Yue Y, Zhou W, Wang Z, et al. Pro- and anti- effects of immunoglobulin a- producing B cell in tumors and its triggers. *Front Immunol*. (2021) 12:765044. doi: 10.3389/fimmu.2021.765044
103. Shi JY, Gao Q, Wang ZC, Zhou J, Wang XY, Min ZH, et al. Margin-infiltrating cd20(+) B cells display an atypical memory phenotype and correlate with favorable prognosis in hepatocellular carcinoma. *Clin Cancer Res*. (2013) 19:5994–6005. doi: 10.1158/1078-0432.CCR-12-3497
104. Xiao S, Bod L, Pochet N, Kota SB, Hu D, Madi A, et al. Checkpoint receptor tigit expressed on tim-1(+) B cells regulates tissue inflammation. *Cell Rep*. (2020) 32:107892. doi: 10.1016/j.celrep.2020.107892
105. Shalpour S, Lin XJ, Bastian IN, Brain J, Burt AD, Aksenov AA, et al. Inflammation-induced iga+ Cells dismantle anti-liver cancer immunity. *Nature*. (2017) 551:340–5. doi: 10.1038/nature24302
106. Shalpour S, Font-Burgada J, Di Caro G, Zhong Z, Sanchez-Lopez E, Dhar D, et al. Immunosuppressive plasma cells impede T-cell-dependent immunogenic chemotherapy. *Nature*. (2015) 521:94–8. doi: 10.1038/nature14395
107. Chan LN, Chen Z, Braas D, Lee JW, Xiao G, Geng H, et al. Metabolic gatekeeper function of B-lymphoid transcription factors. *Nature*. (2017) 542:479–83. doi: 10.1038/nature21076
108. Feist M, Schwarzfischer P, Heinrich P, Sun X, Kemper J, von Bonin F, et al. Cooperative stat/nf-kappab signaling regulates lymphoma metabolic reprogramming and aberrant got2 expression. *Nat Commun*. (2018) 9:1514. doi: 10.1038/s41467-018-03803-x
109. Xiong J, Wang L, Fei XC, Jiang XF, Zheng Z, Zhao Y, et al. Myc is a positive regulator of choline metabolism and impedes mitophagy-dependent necroptosis in diffuse large B-cell lymphoma. *Blood Cancer J*. (2017) 7:e0. doi: 10.1038/bcj.2017.61
110. Luo T, Loira-Pastoriza C, Patil HP, Ucakar B, Muccioli GG, Bosquillon C, et al. Pegylation of paclitaxel largely improves its safety and anti-tumor efficacy following pulmonary delivery in a mouse model of lung carcinoma. *J Control Release*. (2016) 239:62–71. doi: 10.1016/j.jconrel.2016.08.008
111. Melnikov AD, Tsentelovich YP, Yanshole VV. Deep learning for the precise peak detection in high-resolution LC-MS data. *Anal Chem*. (2020) 92:588–92. doi: 10.1021/acs.analchem.9b04811
112. Ringnér M. What is principal component analysis? *Nat Biotechnol*. (2008) 26:303–4. doi: 10.1038/nbt0308-303
113. Olivon F, Elie N, Grelier G, Roussi F, Litaudon M, Touboul D. Metgem software for the generation of molecular networks based on the T-sne algorithm. *Anal Chem*. (2018) 90:13900–8. doi: 10.1021/acs.analchem.8b03099
114. Shrivastava AD, Swainston N, Samanta S, Roberts I, Wright Muelas M, Kell DB. Massgenie: A transformer-based deep learning method for identifying small molecules from their mass spectra. *Biomolecules*. (2021) 11:1793. doi: 10.3390/biom11121793
115. Brown TB, Mann B, Ryder N, Subbiah M, Amodei D. Language models are few-shot learners. *ArXiv*. (2020) 2005:14165. doi: 10.48550/arXiv.2005.14165
116. Qu X, Huang Y, Lu H, Qiu T, Guo D, Agback T, et al. Accelerated nuclear magnetic resonance spectroscopy with deep learning. *Angew Chem Int Ed Engl*. (2020) 59:10297–300. doi: 10.1002/anie.201908162
117. Risum AB, Bro R. Using deep learning to evaluate peaks in chromatographic data. *Talanta*. (2019) 204:255–60. doi: 10.1016/j.talanta.2019.05.053
118. Li M, Wang XR. Peak alignment of gas chromatography-mass spectrometry data with deep learning. *J Chromatogr A*. (2019) 1604:460476. doi: 10.1016/j.chroma.2019.460476
119. Kantz ED, Tiwari S, Watrous JD, Cheng S, Jain M. Deep neural networks for classification of LC-MS spectral peaks. *Anal Chem*. (2019) 91:12407–13. doi: 10.1021/acs.analchem.9b02983
120. Lim J, Wong J, Wong MX, Tan LHE, Chieu HL, Choo D, et al. Chemical structure elucidation from mass spectrometry by matching substructures. *ArXiv*. (2018) 1811:07886. doi: 10.48550/arXiv.1811.07886
121. Fine JA, Rajasekar AA, Jethava KP, Chopra G. Spectral deep learning for prediction and prospective validation of functional groups. *Chem Sci*. (2020) 11:4618–30. doi: 10.1039/c9sc06240h
122. Allen F, Pon A, Greiner R, Wishart D. Computational prediction of electron ionization mass spectra to assist in gc/ms compound identification. *Anal Chem*. (2016) 88:7689–97. doi: 10.1021/acs.analchem.6b01622
123. Wei JN, Belanger D, Adams RP, Sculley D. Rapid prediction of electron-ionization mass spectrometry using neural networks. *ACS Cent Sci*. (2019) 5:700–8. doi: 10.1021/acscentsci.9b00085
124. Plante P-L, Francovic-Fontaine É, May JC, McLean JA, Baker ES, Laviolette F, et al. Predicting ion mobility collision cross-sections using a deep neural network: deepccs. *Anal Chem*. (2019) 91:5191–9. doi: 10.1021/acs.analchem.8b05821
125. Colby SM, Nuñez JR, Hodas NO, Corley CD, Renslow RR. Deep learning to generate in silico chemical property libraries and candidate molecules for small molecule identification in complex samples. *Anal Chem*. (2020) 92:1720–9. doi: 10.1021/acs.analchem.9b02348
126. Aksenov AA, Laponogov I, Zhang Z, Doran SLF, Belluono I, Veselkov D, et al. Auto-deconvolution and molecular networking of gas chromatography-mass spectrometry data. *Nat Biotechnol*. (2021) 39:169–73. doi: 10.1038/s41587-020-0700-3
127. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*. (2007) 8:118–27. doi: 10.1093/biostatistics/kjx037
128. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol*. (2018) 36:421–7. doi: 10.1038/nbt.4091
129. Huber F, Ridder L, Verhoeven S, Spaaks JH, Diblen F, Rogers S, et al. Spec2vec: improved mass spectral similarity scoring through learning of structural relationships. *PLoS Comput Biol*. (2021) 17:e1008724. doi: 10.1371/journal.pcbi.1008724
130. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, Palermo A, Warth B, Hermann G, et al. Metlin: A technology platform for identifying knowns and unknowns. *Anal Chem*. (2018) 90:3156–64. doi: 10.1021/acs.analchem.7b04424
131. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, et al. Massbank: A public repository for sharing mass spectral data for life sciences. *J Mass Spectrom*. (2010) 45:703–14. doi: 10.1002/jms.1777
132. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, et al. Sirius 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods*. (2019) 16:299–302. doi: 10.1038/s41592-019-0344-8
133. Dührkop K, Nothias L-F, Fleischauer M, Reher R, Ludwig M, Hoffmann MA, et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat Biotechnol*. (2021) 39:462–71. doi: 10.1038/s41587-020-0740-8
134. Butler T, Frandsen A, Lighthouse R, Bargh B, Kerby T, West K, et al. MS2Mol: A transformer model for illuminating dark chemical space from mass spectra. *ChemRxiv* (2023). doi: 10.26434/chemrxiv-2023-vsmxp-v4
135. Mikolov T, Sutskever I, Chen K, Corrado Gs, Dean J. Distributed representations of words and phrases and their compositionality. *Adv Neural Inf Process Syst*. (2013) 26:1–9. doi: 10.48550/arXiv.1310.4546
136. Aguilar-Mogas A, Sales-Pardo M, Navarro M, Guimera R, Yanes O. Imet: A network-based computational tool to assist in the annotation of metabolites from tandem mass spectra. *Anal Chem*. (2017) 89:3474–82. doi: 10.1021/acs.analchem.6b04512
137. Goldman S, Wohlwend J, Stražar M, Haroush G, Xavier R, Coley C. Annotating metabolite mass spectra with domain-inspired chemical formula transformers. *Nat Mach Intell*. (2023) 5:1–15. doi: 10.1038/s42256-023-00708-3
138. Xue J, Wang B, Ji H, Li W. Rt-transformer: retention time prediction for metabolite annotation to assist in metabolite identification. *Bioinformatics*. (2024) 40:btac084. doi: 10.1093/bioinformatics/btac084
139. Zhang L, Qin X, Liu M, Liu G, Ren Y. Bert-M7g: A transformer architecture based on bert and stacking ensemble to identify RNA N7-methylguanosine sites from sequence information. *Comput Math Methods Med*. (2021) 2021:7764764. doi: 10.1155/2021/7764764

140. Soylu N, Sefer E. Bert2ome: prediction of 2'-O-methylation modifications from RNA sequence by transformer architecture based on bert. *Trans Comput Biol Bioinf.* (2023) 20:2177–89. doi: 10.1109/TCBB.2023.3237769
141. Fakouri Baygi S, Barupal D, Idsl_Mint: A deep learning framework to predict molecular fingerprints from mass spectra. *J Cheminformatics.* (2024) 16:8. doi: 10.1186/s13321-024-00804-5
142. Stravs M, Dührkop K, Böcker S, Zamboni N. Msnovelist: *de novo* structure generation from mass spectra. *Nat Methods.* (2022) 19:865–70. doi: 10.1038/s41592-022-01486-3
143. Voronov G, Frandsen A, Bargh B, Healey D, Lighthouse R, Kind T, et al. MS2Prop: A machine learning model that directly predicts chemical properties from mass spectrometry data for novel compounds. *bioRxiv.* (2022). doi: 10.1101/2022.10.09.511482
144. Elser D, Huber F, Gaquerel E. Mass2SMILES: deep learning based fast prediction of structures and functional groups directly from high-resolution MS/MS spectra. *bioRxiv.* (2023). doi: 10.1101/2023.07.06.547963
145. Huber F, van der Burg S, van der Hooft JJJ, Ridder L. Ms2deepscore: A novel deep learning similarity measure to compare tandem mass spectra. *J Cheminformatics.* (2021) 13:84. doi: 10.1186/s13321-021-00558-4
146. Tianwei Y, Jones DP. Improving peak detection in high-resolution lc/ms metabolomics data using preexisting knowledge and machine learning approach. *Bioinformatics.* (2014) 20:2941–8. doi: 10.1093/bioinformatics/btu430
147. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539
148. Elizar E, Zulkifley MA, Muharar R, Zaman MHM, Mustaza SM. A review on multiscale-deep-learning applications. *Sensors (Basel).* (2022) 22:7384. doi: 10.3390/s22197384
149. Pomyen Y, Wanichthanarak K, Pongsombat P, Fahrman J, Grapov D, Khoomrung S. Deep metabolome: applications of deep learning in metabolomics. *Comput Struct Biotechnol J.* (2020) 18:2818–25. doi: 10.1016/j.csbj.2020.09.033
150. Neumann S, Böcker S. Computational mass spectrometry for metabolomics: identification of metabolites and small molecules. *Anal Bioanal Chem.* (2010) 398:2779–88. doi: 10.1007/s00216-010-4142-5
151. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt J-N, Laleh NG, et al. The future landscape of large language models in medicine. *Commun Med (Lond).* (2023) 3:141. doi: 10.1038/s43856-023-00370-1
152. Chakraborty S, Hosen MI, Ahmed M, Shekhar HU. Onco-multi-omics approach: A new frontier in cancer research. *BioMed Res Int.* (2018) 2018:9836256. doi: 10.1155/2018/9836256
153. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using paradigm. *Bioinformatics.* (2010) 26:i237–45. doi: 10.1093/bioinformatics/btq182
154. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc Natl Acad Sci.* (2013) 110:4245–50. doi: 10.1073/pnas.1208949110
155. Ren S, Shao Y, Zhao X, Hong CS, Wang F, Lu X, et al. Integration of metabolomics and transcriptomics reveals major metabolic pathways and potential biomarker involved in prostate cancer. *Mol Cell Proteomics.* (2016) 15:154–63. doi: 10.1074/mcp.M115.052381
156. Coler E, Chen W, Melnik A, Morton J, Aksenov A. Metabolomics in the era of artificial intelligence. *Microbiota Host.* (2024) 2:e230017. doi: 10.1530/MAH-23-0017
157. Yu Y, Zhang N, Mai Y, Ren L, Chen Q, Cao Z, et al. Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method. *Genome Biol.* (2023) 24:201. doi: 10.1186/s13059-023-03047-z
158. Dimitrakopoulos C, Hindupur SK, Häfliger L, Behr J, Montazeri H, Hall MN, et al. Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics.* (2018) 34:2441–8. doi: 10.1093/bioinformatics/bty148
159. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights.* (2020) 14:1177932219899051. doi: 10.1177/1177932219899051
160. Imoto Y, Nakamura T, Escobar EG, Yoshiwaki M, Kojima Y, Yabuta Y, et al. Resolution of the curse of dimensionality in single-cell RNA sequencing data analysis. *Life Sci Alliance.* (2022) 5:e202201591. doi: 10.26508/lsa.202201591
161. Kobak D, Berens P. The art of using T-sne for single-cell transcriptomics. *Nat Commun.* (2019) 10:5416. doi: 10.1038/s41467-019-13056-x
162. Abe K, Hirayama M, Ohno K, Shimamura T. Hierarchical non-negative matrix factorization using clinical information for microbial communities. *BMC Genomics.* (2021) 22:104. doi: 10.1186/s12864-021-07401-y
163. Gromova M, Vaggelas A, Dallmann G, Seimetz D. Biomarkers: opportunities and challenges for drug development in the current regulatory landscape. *biomark Insights.* (2020) 15:1177271920974652. doi: 10.1177/1177271920974652
164. Babrak LM, Menetski J, Rebhan M, Nisato G, Zinggeler M, Brasier N, et al. Traditional and digital biomarkers: two worlds apart? *Digit Biomark.* (2019) 3:92–102. doi: 10.1159/000502000
165. Chen Y, Wang B, Zhao Y, Shao X, Wang M, Ma F, et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. *Nat Commun.* (2024) 15:1657. doi: 10.1038/s41467-024-46043-y
166. Kuwabara H, Katsumata K, Iwabuchi A, Udo R, Tago T, Kasahara K, et al. Salivary metabolomics with machine learning for colorectal cancer detection. *Cancer Sci.* (2022) 113:3234–43. doi: 10.1111/cas.15472
167. Chekroud AM, Hawrilenko M, Loho H, Bondar J, Gueorgieva R, Hasan A, et al. Illusory generalizability of clinical prediction models. *Science.* (2024) 383:164–7. doi: 10.1126/science.adg8538
168. Poirion OB, Jing Z, Chaudhary K, Huang S, Garmire LX. Deepprog: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome Med.* (2021) 13:112. doi: 10.1186/s13073-021-00930-x
169. Xi B, Gu H, Baniyasi H, Raftery D. Statistical analysis and modeling of mass spectrometry-based metabolomics data. *Methods Mol Biol.* (2014) 1198:333–53. doi: 10.1007/978-1-4939-1258-2_22
170. Bahado-Singh RO, Lugade A, Field J, Al-Wahab Z, Han B, Mandal R, et al. Metabolomic prediction of endometrial cancer. *Metabolomics.* (2017) 14:6. doi: 10.1007/s11306-017-1290-z
171. Telleria O, Alboniga OE, Clos-Garcia M, Nafria-Jimenez B, Cubiella J, Bujanda L, et al. A comprehensive metabolomics analysis of fecal samples from advanced adenoma and colorectal cancer patients. *Metabolites.* (2022) 12:550. doi: 10.3390/metabo12060550
172. An R, Yu H, Wang Y, Lu J, Gao Y, Xie X, et al. Integrative analysis of plasma metabolomics and proteomics reveals the metabolic landscape of breast cancer. *Cancer Metab.* (2022) 10:13. doi: 10.1186/s40170-022-00289-6
173. Xie Y, Meng W-Y, Li R-Z, Wang Y-W, Qian X, Chan C, et al. Early lung cancer diagnostic biomarker discovery by machine learning methods. *Transl Oncol.* (2021) 14:100907. doi: 10.1016/j.tranon.2020.100907
174. Urman JM, Herranz JM, Uriarte I, Rullán M, Oyón D, González B, et al. Pilot multi-omic analysis of human bile from benign and Malignant biliary strictures: A machine-learning approach. *Cancers (Basel).* (2020) 12:1644. doi: 10.3390/cancers12061644
175. Monedeiro F, Monedeiro-Milanowski M, Zmysłowski H, De Martinis BS, Buszewski B. Evaluation of salivary voc profile composition directed towards oral cancer and oral lesion assessment. *Clin Oral Investig.* (2021) 25:4415–30. doi: 10.1007/s00784-020-03754-y
176. Yang Q-J, Zhao J-R, Hao J, Li B, Huo Y, Han Y-L, et al. Serum and urine metabolomics study reveals a distinct diagnostic model for cancer cachexia. *J Cachexia Sarcopenia Muscle.* (2018) 9:71–85. doi: 10.1002/jcsm.12246
177. Jin Y, Fan S, Jiang W, Zhang J, Yang L, Xiao J, et al. Two effective models based on comprehensive lipidomics and metabolomics can distinguish hcs, and tnbc versus non-tnbc. *Proteomics Clin Appl.* (2023) 17:e2200042. doi: 10.1002/prca.202200042
178. Silva AAR, Cardoso MR, de Oliveira DC, Godoy P, Talarico MCR, Gutiérrez JM, et al. Plasma metabolome signatures to predict responsiveness to neoadjuvant chemotherapy in breast cancer. *Cancers (Basel).* (2024) 16:2473. doi: 10.3390/cancers16132473
179. Wang X, Zhu Y, Zhou H, Huang Z, Chen H, Zhang J, et al. Integrated analysis of serum untargeted metabolomics and targeted bile acid metabolomics for identification of diagnostic biomarkers for colorectal cancer. *Nan Fang Yi Ke Da Xue Xue Bao.* (2023) 43:443–53. doi: 10.12122/j.issn.1673-4254.2023.03.15
180. Wei MX, Yang Z, Wang PP, Zhao XK, Song X, Xu RH, et al. Novel metabolic biomarker for early detection and diagnosis to the patients with gastric cardia adenocarcinoma. *Cancer Med.* (2024) 13:e7015. doi: 10.1002/cam4.7015
181. Lv J, Jia H, Mo M, Yuan J, Wu Z, Zhang S, et al. Changes of serum metabolites levels during neoadjuvant chemoradiation and prediction of the pathological response in locally advanced rectal cancer. *Metabolomics.* (2022) 18:99. doi: 10.1007/s11306-022-01959-8
182. You L, Fan Y, Liu X, Shao S, Guo L, Noreldeen HAA, et al. Liquid chromatography-mass spectrometry-based tissue metabolic profiling reveals major metabolic pathway alterations and potential biomarkers of lung cancer. *J Proteome Res.* (2020) 19:3750–60. doi: 10.1021/acs.jproteome.0c00285
183. Hershberger CE, Rodarte AI, Siddiqi S, Moro A, Acevedo-Moreno L-A, Brown JM, et al. Salivary metabolites are promising non-invasive biomarkers of hepatocellular carcinoma and chronic liver disease. *Liver Cancer Int.* (2021) 2:33–44. doi: 10.1002/lci.225
184. Sun R, Fei F, Wang M, Jiang J, Yang G, Yang N, et al. Integration of metabolomics and machine learning revealed tryptophan metabolites are sensitive biomarkers of pemetrexed efficacy in non-small cell lung cancer. *Cancer Med.* (2023) 12:19245–59. doi: 10.1002/cam4.6446
185. Fan Y, Wang J, Yang Y, Liu Q, Fan Y, Yu J, et al. Detection and identification of potential biomarkers of breast cancer. *J Cancer Res Clin Oncol.* (2010) 136:1243–54. doi: 10.1007/s00432-010-0775-1
186. Jové M, Collado R, Quiles JL, Ramirez-Tortosa M-C, Sol J, Ruiz-Sanjuan M, et al. A plasma metabolomic signature discloses human breast cancer. *Oncotarget.* (2017) 8:19522–33. doi: 10.18632/oncotarget.14521
187. Ma J, Chen K, Ding Y, Li X, Tang Q, Jin B, et al. High-throughput quantitation of amino acids and acylcarnitine in cerebrospinal fluid: identification of pcnsl biomarkers and potential metabolic messengers. *Front Mol Biosci.* (2023) 10:1257079. doi: 10.3389/fmolb.2023.1257079
188. Zhang L, Zheng J, Ismond KP, MacKay S, LeVatte M, Constable J, et al. Identification of urinary biomarkers of colorectal cancer: towards the development of a

- colorectal screening test in limited resource settings. *Cancer biomark.* (2023) 36:17–30. doi: 10.3233/CBM-220034
189. Matsumoto T, Fukuzawa M, Itoi T, Sugimoto M, Aizawa Y, Sunamura M, et al. Targeted metabolomic profiling of plasma samples in gastric cancer by liquid chromatography-mass spectrometry. *Digestion.* (2023) 104:97–108. doi: 10.1159/000526864
190. Yang Y, Zhao S, Fan Y, Zhao F, Liu Q, Hu W, et al. Detection and identification of potential biomarkers of non-small cell lung cancer. *Technol Cancer Res Treat.* (2009) 8:455–66. doi: 10.1177/153303460900800607
191. Li J, Chen Q, Guo L, Li J, Jin B, Wu X, et al. *In situ* detecting lipids as potential biomarkers for the diagnosis and prognosis of intrahepatic cholangiocarcinoma. *Cancer Manag Res.* (2022) 14:2903–12. doi: 10.2147/CMAR.S357000
192. Xu S, Xiang C, Wu J, Teng Y, Wu Z, Wang R, et al. Tongue coating bacteria as a potential stable biomarker for gastric cancer independent of lifestyle. *Dig Dis Sci.* (2021) 66:2964–80. doi: 10.1007/s10620-020-06637-0
193. Adam MG, Beyer G, Christiansen N, Kamlage B, Pilarsky C, Distler M, et al. Identification and validation of a multivariable prediction model based on blood plasma and serum metabolomics for the distinction of chronic pancreatitis subjects from non-pancreas disease control subjects. *Gut.* (2021) 70:2150–8. doi: 10.1136/gutjnl-2020-320723
194. Kim JO, Balshaw R, Trevena C, Banerji S, Murphy L, Dawe D, et al. Data-driven identification of plasma metabolite clusters and metabolites of interest for potential detection of early-stage non-small cell lung cancer cases versus cancer-free controls. *Cancer Metab.* (2022) 10:16. doi: 10.1186/s40170-022-00294-9
195. da Costa NL, de Sá Alves M, de Sá Rodrigues N, Bandeira CM, Oliveira Alves MG, Mendes MA, et al. Finding the combination of multiple biomarkers to diagnose oral squamous cell carcinoma - a data mining approach. *Comput Biol Med.* (2022) 143:105296. doi: 10.1016/j.compbiomed.2022.105296
196. Hirata Y, Kobayashi T, Nishiumi S, Yamanaka K, Nakagawa T, Fujigaki S, et al. Identification of highly sensitive biomarkers that can aid the early detection of pancreatic cancer using GC/MS/MS-based targeted metabolomics. *Clin Chim Acta.* (2017) 468:98–104. doi: 10.1016/j.cca.2017.02.011
197. Tian Y, Wang Z, Liu X, Duan J, Feng G, Yin Y, et al. Prediction of chemotherapeutic efficacy in non-small cell lung cancer by serum metabolomic profiling. *Clin Cancer Res.* (2018) 24:2100–9. doi: 10.1158/1078-0432.CCR-17-2855
198. Frickenschmidt A, Frohlich H, Bullinger D, Zell A, Laufer S, Gleiter CH, et al. Metabolomics in cancer diagnosis: mass spectrometry-based profiling of urinary nucleosides from breast cancer patients. *Biomarkers.* (2008) 13:435–49. doi: 10.1080/13547500802012858
199. Da Cunha PA, Nitusca D, Canto LMD, Varghese RS, Resson HW, Willey S, et al. Metabolomic analysis of plasma from breast cancer patients using ultra-high-performance liquid chromatography coupled with mass spectrometry: an untargeted study. *Metabolites.* (2022) 12:447. doi: 10.3390/metabo12050447
200. Khan I, Nam M, Kwon M, Seo S-S, Jung S, Han JS, et al. LC/MS-based polar metabolite profiling identified unique biomarker signatures for cervical cancer and cervical intraepithelial neoplasia using global and targeted metabolomics. *Cancers (Basel).* (2019) 11:511. doi: 10.3390/cancers11040511
201. He D, Huang X, Arah OA, Walker DI, Jones DP, Ritz B, et al. A prediction model for classifying maternal pregnancy smoking using california state birth certificate information. *Paediatr Perinat Epidemiol.* (2024) 38:102–10. doi: 10.1111/ppe.13021
202. Badmos S, Noriega-Landa E, Holbrook KL, Quaye GE, Su X, Gao Q, et al. Urinary volatile organic compounds in prostate cancer biopsy pathologic risk stratification using logistic regression and multivariate analysis models. *Am J Cancer Res.* (2024) 14:192–209. doi: 10.62347/RXSH1223
203. Borgmästars E, Jacobson S, Simm M, Johansson M, Billing O, Lundin C, et al. Metabolomics for early pancreatic cancer detection in plasma samples from a Swedish prospective population-based biobank. *J Gastrointest Oncol.* (2024) 15:755–67. doi: 10.21037/jgo-23-930
204. Fan Y, Shi L, Liu Q, Dong R, Zhang Q, Yang S, et al. Discovery and identification of potential biomarkers of papillary thyroid carcinoma. *Mol Cancer.* (2009) 8:79. doi: 10.1186/1476-4598-8-79
205. D'Andréa G, Jing L, Peyrottes I, Guignon J-M, Graslén F, Lindenthal S, et al. Pilot study on the use of untargeted metabolomic fingerprinting of liquid-cytology fluids as a diagnostic tool of Malignancy for thyroid nodules. *Metabolites.* (2023) 13:782. doi: 10.3390/metabo13070782
206. Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics.* (2010) 6:78–95. doi: 10.1007/s11306-009-0178-y
207. Jia H, Shen X, Guan Y, Xu M, Tu J, Mo M, et al. Predicting the pathological response to neoadjuvant chemoradiation using untargeted metabolomics in locally advanced rectal cancer. *Radiother Oncol.* (2018) 128:548–56. doi: 10.1016/j.radonc.2018.06.022
208. Smith E, Ericson U, Hellstrand S, Orho-Melander M, Nilsson PM, Fernandez C, et al. A healthy dietary metabolic signature is associated with a lower risk for type 2 diabetes and coronary artery disease. *BMC Med.* (2022) 20:122. doi: 10.1186/s12916-022-02326-z
209. Nishiumi S, Kobayashi T, Kawana S, Unno Y, Sakai T, Okamoto K, et al. Investigations in the possibility of early detection of colorectal cancer by gas chromatography/triple-quadrupole mass spectrometry. *Oncotarget.* (2017) 8:17115–26. doi: 10.18632/oncotarget.15081
210. Koureas M, Kirgou P, Amoutzias G, Hadjichristodoulou C, Gourgoulis K, Tsakalof A. Target analysis of volatile organic compounds in exhaled breath for lung cancer discrimination from other pulmonary diseases and healthy persons. *Metabolites.* (2020) 10:317. doi: 10.3390/metabo10080317
211. Ge R, Wang Z, Cheng L. Tumor microenvironment heterogeneity an important mediator of prostate cancer progression and therapeutic resistance. *NPJ Precis Oncol.* (2022) 6:31. doi: 10.1038/s41698-022-00272-w
212. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. *Cancer Discovery.* (2021) 11:900–15. doi: 10.1158/2159-8290.CD-21-0090