Metabolomics and Plasma Biomarkers in Gastric Cancer Diagnostics: A Survey

www.surveyx.cn

Abstract

Gastric cancer (GC) remains a significant global health challenge due to its high incidence and mortality rates, often exacerbated by late-stage diagnoses. This survey paper explores the potential of metabolomics and plasma biomarkers in revolutionizing GC diagnostics by facilitating early detection and personalized treatment strategies. Metabolomics provides a comprehensive analysis of metabolic alterations associated with tumorigenesis, enabling the identification of specific biomarkers that serve as early indicators of GC. The integration of advanced mass spectrometry and computational techniques within metabolomics has been pivotal in revealing endogenous metabolic pathways, thereby aiding in the discovery of novel diagnostic markers. Plasma biomarkers offer a non-invasive alternative for diagnostic screening, enhancing patient outcomes by minimizing the discomfort and risks associated with traditional invasive methods. The survey highlights the transformative potential of integrating metabolomics with other omics data, such as genomics and proteomics, to provide a holistic view of the molecular underpinnings of GC. This multi-omics approach facilitates the discovery of novel therapeutic targets and enhances the precision of treatment strategies. Despite the challenges posed by data complexity and high-dimensionality, advancements in machine learning and AI integration hold promise for improving the accuracy and scalability of metabolomic analyses. As the field evolves, the application of metabolomics in GC diagnostics is expected to expand, offering new opportunities for early intervention and more effective management of the disease, ultimately improving patient outcomes and reducing the global burden of gastric cancer.

1 Introduction

1.1 Global Impact of Gastric Cancer

Gastric cancer (GC) is a major global health concern, ranking as the fifth most common cancer, with over 1 million new cases and approximately 700,000 deaths reported in 2020 [1]. Its high incidence and mortality rates highlight the urgent need for improved diagnostic and therapeutic strategies, particularly given the poor prognosis associated with late-stage diagnoses [2]. As the second leading cause of cancer-related mortality worldwide, GC's aggressive nature necessitates a thorough understanding of epidemiological trends and risk factors to inform effective prevention and management approaches [3, 4, 5].

1.2 Potential of Metabolomics in Gastric Cancer

Metabolomics has emerged as a pivotal field for the early detection and diagnosis of gastric cancer, elucidating the metabolic alterations linked to tumorigenesis. By analyzing comprehensive metabolic profiles from biological samples, metabolomics facilitates the identification of specific biomarkers that can serve as early indicators of GC [6]. This approach not only enhances our understanding of the biological processes underlying cancer development but also bridges existing knowledge gaps

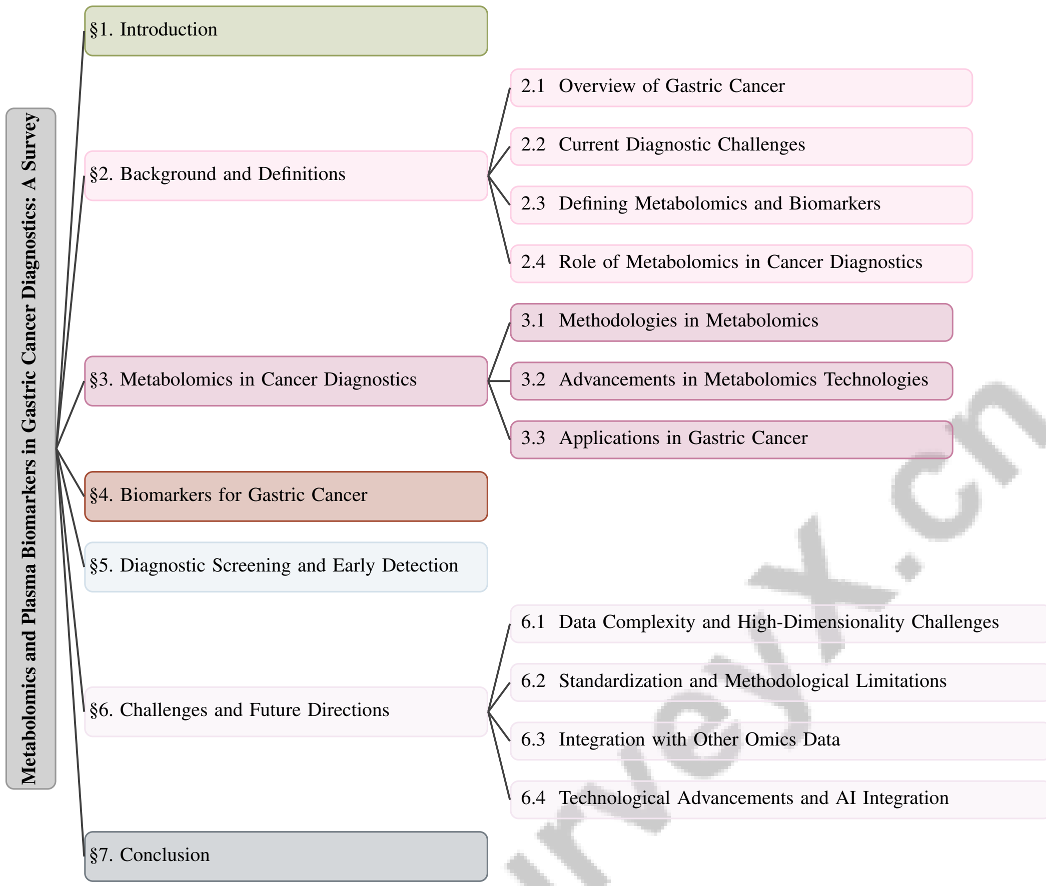

Figure 1: chapter structure

[7]. The integration of advanced mass spectrometry techniques has been critical in uncovering the endogenous metabolic pathways involved in GC, aiding in the discovery of novel diagnostic markers [8]. As the global incidence of gastric cancer rises, innovative diagnostic strategies are essential for improving early detection rates and patient prognoses [9]. Clinical applications of metabolomics are set to transform cancer diagnostics by providing a comprehensive view of metabolic changes, which are vital for developing targeted therapeutic interventions [10]. By addressing current diagnostic deficiencies, metabolomics stands out as a promising tool for early diagnosis and management of gastric cancer, ultimately aiming to reduce associated mortality rates [11].

1.3 Importance of Plasma Biomarkers

Plasma biomarkers are crucial for the non-invasive diagnostic screening of gastric cancer, offering a viable alternative to traditional invasive methods. They enable the detection of specific metabolic alterations in the bloodstream indicative of GC, facilitating early diagnosis and enhancing patient outcomes. This non-invasive approach minimizes patient discomfort and risk while allowing for repeated sampling to monitor disease progression and therapeutic response. The necessity for serological tests to identify precancerous lesions is underscored, highlighting the potential of plasma biomarkers to meet this critical diagnostic need [5]. By revealing systemic metabolic changes associated with gastric cancer, plasma biomarkers improve early detection capabilities, contributing to timely and targeted interventions.

1.4 Structure of the Survey

The survey is organized to provide an in-depth exploration of metabolomics and plasma biomarkers in gastric cancer diagnostics. It begins with the **Introduction**, discussing the global impact of gastric cancer and the transformative potential of metabolomics for early detection. The significance of

AI-generated, for reference only.

plasma biomarkers in non-invasive diagnostic screening is also emphasized. Following this, the **Background and Definitions** section outlines gastric cancer's epidemiology, current diagnostic challenges, and definitions of key concepts such as metabolomics and biomarkers, elucidating metabolomics' specific role in cancer diagnostics. The third section, **Metabolomics in Cancer Diagnostics**, focuses on methodologies and technological advancements in metabolomics, particularly for gastric cancer applications. The fourth section, **Biomarkers for Gastric Cancer**, examines various biomarkers, highlighting recent advances and challenges in discovery efforts. **Diagnostic Screening and Early Detection** evaluates metabolomics' role in enhancing screening methods and its impact on patient outcomes. The penultimate section, **Challenges and Future Directions**, identifies current challenges in utilizing metabolomics for gastric cancer diagnostics and outlines future research directions, including data complexity, standardization, integration with other omics data, and technological advancements. Finally, the **Conclusion** summarizes the key points discussed, emphasizing the transformative potential of metabolomics and plasma biomarkers in gastric cancer diagnostics.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Overview of Gastric Cancer

Gastric cancer (GC) is a multifaceted malignancy with significant global health implications, arising from a combination of genetic, environmental, and lifestyle factors [2]. It is one of the most lethal cancers, primarily due to late-stage diagnosis and the limited efficacy of treatments. GC is categorized into cardia and non-cardia types, each with distinct epidemiological patterns and risk factors, including Helicobacter pylori infection, dietary influences, and genetic predispositions [5]. The typically asymptomatic nature of GC leads to late-stage detection, negatively impacting survival rates [2]. Locally advanced gastric cancer (LAGC) accounts for approximately two-thirds of cases, with neoadjuvant chemotherapy (NACT) being the standard treatment [1]. Emerging research on gut microbiota diversity and circular RNAs (circRNAs) offers insights into GC risk and progression, potentially informing diagnostic and therapeutic approaches [9, 12]. Chronic gastritis (CG), a precursor to GC, is defined in traditional Chinese medicine (TCM) as Cold ZHENG (Asthenic Cold) and Hot ZHENG (Excess Hot) [13]. Biomarker development for GC is hindered by ambiguous definitions, impeding diagnostic and therapeutic advancements [14]. While genome-wide association studies (GWAS) focus on common variants, innovative approaches are needed to address the 'missing heritability' in complex diseases like GC by evaluating the cumulative impact of single-nucleotide variants (SNVs) [15]. Current treatments, including chemotherapy and surgery, have yet to significantly improve survival rates [3].

2.2 Current Diagnostic Challenges

The diagnostic landscape of gastric cancer is complicated by the high-dimensional nature of metabolomic data, where the number of covariates often exceeds samples, complicating hypothesis testing and estimation [16]. Traditional metabolite selection methods often overlook interdependencies among metabolites in high-dimensional LC-MS data, missing crucial metabolic interactions [17]. Furthermore, inadequate integration of metabolomics data with biological outcomes hampers the identification of active metabolites as potential GC biomarkers [18]. Challenges also include the comparability of numerical data and its biological interpretation [19], and accurate metabolite annotation and pathway enrichment often treated separately, leading to inaccuracies [20]. Managing data and metadata in large, multi-site clinical studies adds complexity [21]. Proper imputation of missing values is crucial for accurate analyses [22], and the complexity of metabolomic data management hinders the identification of significant patterns related to health outcomes [4]. Validation of biomarkers remains problematic, with many failing in validation studies, limiting their diagnostic utility [3]. Accurate mass spectra annotation remains challenging, as current methods struggle to map spectral signatures to molecular identities, complicating biomarker discovery [15]. These challenges highlight the need for innovative methodologies to improve the accuracy and reliability of GC diagnostics.

2.3 Defining Metabolomics and Biomarkers

Metabolomics, a key component of systems biology, offers a comprehensive framework for understanding metabolic networks and pathways in biological processes and diseases, including cancer [6]. Utilizing techniques such as ultra-high-performance liquid chromatography coupled with mass spectrometry (UHPLC-MS/MS) and flow injection mass spectrometry (FI-MS/MS), metabolomics facilitates both targeted and untargeted analyses of small molecule metabolites in biological samples [23]. It plays a crucial role in cancer diagnostics by identifying metabolic alterations that characterize disease states, revealing potential biomarkers indicative of pathological conditions [10]. Biomarkers, as measurable indicators of biological states, are vital in cancer diagnostics and therapeutics, encompassing genetic, epigenetic, and metabolic changes to enhance disease detection, monitoring, and management [14]. In GC, biomarkers serve as non-invasive serological markers, facilitating early intervention and personalized treatment strategies [11]. Integrating metabolomics with other omics data, such as genomics and transcriptomics, enhances biomarker discovery by revealing phenotype-specific associations that improve diagnostic accuracy [24]. Advancements in computational approaches, including machine learning and network-guided feature selection, further enhance biomarker diagnostic capabilities by optimizing predictive power and clinical applicability [16]. However, conventional cross-validation methods often yield over-optimistic estimates of biomarker performance due to trial heterogeneity [20]. Addressing these challenges through innovative methodologies and integrative approaches promises to advance cancer diagnostics, ultimately improving patient outcomes [4]. As metabolomics evolves, its role in biomarker discovery and validation is expected to expand, providing insights into molecular mechanisms underlying cancer progression and therapy resistance [21].

2.4 Role of Metabolomics in Cancer Diagnostics

Metabolomics offers a unique advantage over other omics approaches by providing a comprehensive analysis of metabolites, the small molecule intermediates and end products of cellular processes. This allows metabolomics to deliver a direct snapshot of the physiological state of cells and tissues, revealing biochemical alterations associated with cancer [7]. Unlike genomics or transcriptomics, which capture genetic or transcriptional changes, metabolomics directly reflects the functional outcomes of cellular activities, serving as a more immediate indicator of disease states [25]. Integrating metabolomics with other omics data, such as genomics and proteomics, enhances biomarker discovery by elucidating complex interactions and pathways involved in cancer progression [26]. This cross-omics approach provides a holistic view of the molecular underpinnings of cancer, facilitating the identification of novel therapeutic targets and diagnostic markers [27]. Advanced computational techniques, including machine learning algorithms such as support vector machines (SVM), random forests (RF), and deep learning (DL), are employed to analyze metabolomics data, improving the accuracy and predictive power of cancer diagnostics [28]. Innovative methodologies, such as the Random Projection Algorithm (RPA) for feature dimensionality reduction, address challenges posed by the high-dimensional nature of metabolomics data [29]. Techniques like Lasso-based variable selection methods and automated NMR spectral profiling systems further enhance the precision of metabolite identification and quantification, solidifying metabolomics' role in cancer diagnostics. Metabolomics uncovers dysregulated metabolic pathways in cancer cells, providing insights into tumor metabolism and potential therapeutic intervention points. By focusing on the metabolome, researchers can identify metabolic signatures specific to different cancer types, aiding early detection and personalized treatment strategies. Scalable cloud-based infrastructures have improved the processing efficiency of metabolomics data, contributing to the robustness and scalability of analyses [30]. As metabolomics evolves, its integration with other omics approaches and advancements in computational analysis are likely to enhance its role in cancer diagnostics, offering new opportunities for improving patient outcomes.

3 Metabolomics in Cancer Diagnostics

Metabolomics has become integral to cancer diagnostics, offering insights into the biochemical changes inherent in malignancies. This section delves into the methodologies used in metabolomics, key to understanding metabolic profiles and their implications in cancer detection and treatment. By examining these techniques, advancements in diagnostic accuracy and therapeutic strategies can be appreciated. As illustrated in **??**, the hierarchical structure of metabolomics in cancer diagnostics highlights key methodologies, advancements in technologies, and specific applications

Category	Feature	Method
Methodologies in Metabolomics	Statistical Approaches	BAYESIL[31], CTGT[32], LZSR[33], DPPCA[18]
	Analytical Techniques	T&M[23]
	Data Handling and Management	RPA[29]
	Advanced Modeling	ESP[34]
Advancements in Metabolomics Technologies	Pathway and Biomarker Analysis	PF[35], PUMA[36]
	Data Management and Integrity	MeKDDaM[28], PMSF[19]
	Biomarker Optimization	maxTPR[37]
Applications in Gastric Cancer	Metabolite and Biomarker Analysis	M-SAGA[38], MVS[25], GERBIL[21]
	Data Integration and Scalability	LP[39], iMD4GC[1]
	Molecular Interaction Analysis	IntLIM[24]

Table 1: This table presents a comprehensive summary of the methodologies, technological advancements, and specific applications in metabolomics, particularly focusing on gastric cancer diagnostics. It categorizes various approaches, such as statistical methods, analytical techniques, and data management strategies, highlighting their contributions to enhancing diagnostic accuracy and therapeutic strategies in cancer research.

in gastric cancer. The chart categorizes essential techniques, computational enhancements, and technological innovations, demonstrating their roles in improving diagnostic accuracy and therapeutic strategies. Additionally, it underscores the significance of metabolomics in identifying biomarkers and understanding molecular mechanisms in gastric cancer. Table 1 provides a detailed overview of the methodologies, advancements, and applications in metabolomics, emphasizing their role in cancer diagnostics and treatment strategies. Furthermore, Table 2 offers a comparative analysis of prominent metabolomics methodologies, focusing on their applications and enhancements in cancer diagnostics. The following subsection focuses on metabolomics methodologies, highlighting crucial techniques like mass spectrometry and nuclear magnetic resonance spectroscopy, essential for analyzing metabolic profiles in cancer research.

3.1 Methodologies in Metabolomics

Metabolomics utilizes sophisticated methodologies for analyzing metabolic profiles, with mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy being predominant [6]. MS, particularly in untargeted studies, is crucial for detecting and quantifying a vast array of metabolites, providing insights into various disease states, including cancer [8]. Techniques like liquid chromatography-mass spectrometry (LC-MS) enable detailed metabolic profiling, facilitating biomarker identification [40]. NMR spectroscopy offers a non-destructive analysis method, delivering detailed molecular structure and dynamics information, with platforms like BAYESIL enhancing metabolite identification and quantification [31].

Advancements in computational biology have significantly enhanced metabolomics capabilities. The integration of statistical modeling with metabolomics methodologies, such as the MetSizeR method, aids in sample size estimation in studies lacking pilot data [18]. Machine learning techniques, including the Random Projection Algorithm (RPA), reduce high-dimensional metabolomics data while maintaining integrity, improving classification accuracy [29]. Additionally, linear models evaluating gene expression and metabolite level interactions across phenotypes further illustrate computational methods' role in metabolomics [24].

As illustrated in Figure 2, the hierarchical organization of methodologies in metabolomics highlights key spectrometry techniques, computational advancements, and data management platforms. Each category is crucial for advancing the field and improving the analysis and interpretation of metabolic data. Advanced methodologies like 'Closed Testing with Globaltest' (CTGT) enhance metabolomics data interpretation robustness by controlling the family-wise error rate in pathway analysis [32]. Probabilistic approaches, such as Pathway Activity Likelihood Analysis (PUMA), use generative models to infer pathway activities and annotate metabolites based on mass measurements from untargeted metabolomics [36]. Platforms like LabPipe streamline metabolomics data management across various sampling approaches, addressing large-scale studies' complexities [39].

The evolution of targeted and non-targeted metabolomics approaches continues, each offering distinct advantages in metabolite detection and analysis [23]. LZSR provides a consistent framework for identifying biomarkers, enhancing discovery reliability [33]. The ESP model refines spectral predictions through an ensemble approach, weighing MLP and GNN model outputs based on ranking performances [34]. Technological advancements in mass spectrometry and computational biology remain crucial for future discoveries, enabling precise analyses of metabolic alterations associated with diseases such as cancer [7]. GERBIL exemplifies computational advancements integration, employing a multi-agent system for data collection and an encoder-evaluator-decoder paradigm to optimize biomarker subsets, revolutionizing discovery [21]. As these methodologies advance, they promise to enhance metabolomics' diagnostic and therapeutic potential in cancer diagnostics significantly.

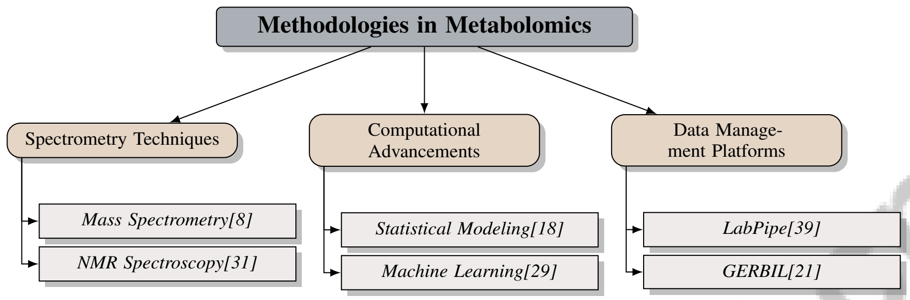

Figure 2: This figure shows the hierarchical organization of methodologies in metabolomics, highlighting key spectrometry techniques, computational advancements, and data management platforms. Each category is crucial for advancing the field and improving the analysis and interpretation of metabolic data.

3.2 Advancements in Metabolomics Technologies

Recent advancements in metabolomics technologies have markedly improved analysis precision, accuracy, and efficiency, enhancing their application in cancer diagnostics. Notable innovations include structured frameworks categorizing metabolomics measurements into levels, facilitating data comparison and interpretation [40]. LabPipe, an extensible informatics platform, offers a semi-automated data collection and management process, minimizing manual handling and errors crucial for large-scale studies [39].

Machine learning integration has further advanced metabolomics. Transformer-based architectures, like PathFormer, improve prediction accuracy by incorporating pathway information into the attention mechanism, enhancing biomarker identification reproducibility [35]. The MeKDDaM framework provides a systematic approach to managing complex metabolomics data, ensuring alignment with overall analytical objectives for robust outcomes [28].

Cloud-based infrastructures have significantly reduced processing times, exemplified by reducing BATMAN processing time from 4 days to approximately 10 minutes on large clusters, facilitating rapid analyses [30]. The PUMA method has shown considerable improvements in pathway activity prediction and metabolite annotation, offering biologically meaningful insights diverging from traditional methods [36].

Developments in statistical methods for high-dimensional data, as discussed by Antonelli et al., underscore the importance of flexible frameworks for analyzing complex metabolomics datasets [17]. Probabilistic approaches evaluating metabolite consistency over time provide novel means of assessing sample integrity, crucial for accurate metabolomic analysis [19].

These technological advancements enhance cancer diagnostics' accuracy and efficiency, paving the way for biomarker discovery and a deeper understanding of disease mechanisms, ultimately contributing to improved patient outcomes. Integrating these technologies with innovative biomarker combination methods, as shown by improved performance in maximizing true positive rates at specified false positive rates, highlights their potential in clinical applications [37].

3.3 Applications in Gastric Cancer

Metabolomics is crucial in gastric cancer (GC) diagnostics, providing insights into metabolic alterations associated with tumorigenesis, essential for early detection and treatment. Advanced methodologies like the MultiVarSel approach have significantly improved relevant metabolite selection in LC-MS data, offering a robust alternative to traditional methods [25]. This enhancement in metabolite selection is vital for identifying metabolic signatures specific to gastric cancer, facilitating early detection and personalized treatment strategies [8].

Metabolomics' application in GC diagnostics is further demonstrated by computational frameworks like iMD4GC, which excel in predicting treatment responses and survival outcomes, surpassing existing methods [1]. Moreover, integrating metabolomics with other omics data, as exemplified by the IntLIM method, allows comprehensive analysis of gene-metabolite interactions, enhancing understanding of molecular mechanisms underlying gastric cancer [24].

Machine learning-driven approaches identify potential biomarkers from large datasets, leveraging computational algorithms to reveal patterns and associations not apparent through conventional methods [38]. The role of circRNAs as potential biomarkers in GC progression highlights their substantial involvement in tumorigenesis and potential as therapeutic targets [12]. This underscores the importance of integrating metabolomics with genomic data to discover novel biomarkers for diagnosing and treating gastric cancer.

Additionally, metabolomics has been instrumental in analyzing gut microbiota's influence on gastric cancer, identifying significant bacterial genera associated with the disease and contributing to understanding microbial interactions in cancer progression [9]. Platforms like LabPipe in studies involving multiple clinics and breathomics data collection across different sites demonstrate metabolomics' scalability and applicability in clinical settings [39].

Experiments on real-world datasets, including Gastric Cancer, Epstein-Barr virus-associated Gastric Cancer, and Intestinal Metaplasia, using a random forest model for evaluation, validate metabolomics' efficacy in gastric cancer diagnostics [21]. Metabolomics' application in gastric cancer diagnostics enhances metabolic alterations detection and provides a framework for integrating diverse biological data, advancing cancer diagnostics and treatment strategies' precision and efficacy [4].

Feature	Mass Spectrometry	NMR Spectroscopy	MetSizeR
Detection Method	Untargeted Studies	Non-destructive Analysis	Sample Size Estimation
Computational Enhancement	Lc-MS Profiling	Bayesil Platform	Statistical Modeling
Specific Application	Biomarker Identification	Molecular Structure	Pilot Data Studies

Table 2: This table compares three key methodologies utilized in metabolomics for cancer diagnostics: Mass Spectrometry, NMR Spectroscopy, and MetSizeR. It highlights their detection methods, computational enhancements, and specific applications, underscoring their roles in biomarker identification, molecular structure analysis, and sample size estimation in studies. The comparison provides insights into the strengths and specific uses of each method in the context of cancer research.

4 Biomarkers for Gastric Cancer

4.1 Types of Biomarkers in Gastric Cancer

Identifying biomarkers in gastric cancer is crucial for improving diagnostic precision and tailoring therapeutic strategies. These biomarkers, encompassing genetic, epigenetic, proteomic, and metabolomic categories, offer unique insights into the disease's pathophysiology. As illustrated in Figure 3, biomarkers are classified into distinct categories: genetic, epigenetic, and proteomic/metabolomic types, each accompanied by specific examples and references. Genetic biomarkers, notably single nucleotide variants (SNVs), are pivotal for understanding hereditary components and genetic predispositions in gastric cancer [15]. Epigenetic markers, such as blood-derived long non-coding RNAs (lncRNAs), present promising non-invasive diagnostic options, reflecting gene expression changes and tumor progression [41]. Additionally, biomarkers associated with Cold/Hot ZHENG chronic gastritis (CG) highlight the immune and metabolic complexities in gastric cancer pathogenesis [13].

Proteomic biomarkers elucidate altered protein expression in gastric cancer, aiding therapeutic target identification. Integrative models like iCARH enhance the detection of metabolic biomarkers, infer disrupted pathways, and reveal significant omic associations, enriching our understanding of gastric cancer's molecular environment [42]. Metabolomic biomarkers, identified using mass spectrometry and NMR spectroscopy, reveal metabolic alterations in gastric cancer, offering early detection and personalized treatment opportunities [6]. Techniques like the Random Projection Algorithm optimize data dimensionality and classification accuracy, advancing biomarker discovery [29].

Innovative approaches, including tongue image analysis for color and texture features, expand the scope of biomarker discovery in gastric cancer [43]. Biomarkers are classified into diagnostic, monitoring, pharmacodynamic/response, predictive, prognostic, safety, and susceptibility/risk types, providing a structured clinical application framework [14]. Advances in integrating diverse data modalities through models like IGCN, which assigns attention coefficients to each data modality per patient, promise to enhance biomarker interpretability and identification, advancing gastric cancer diagnostics [44].

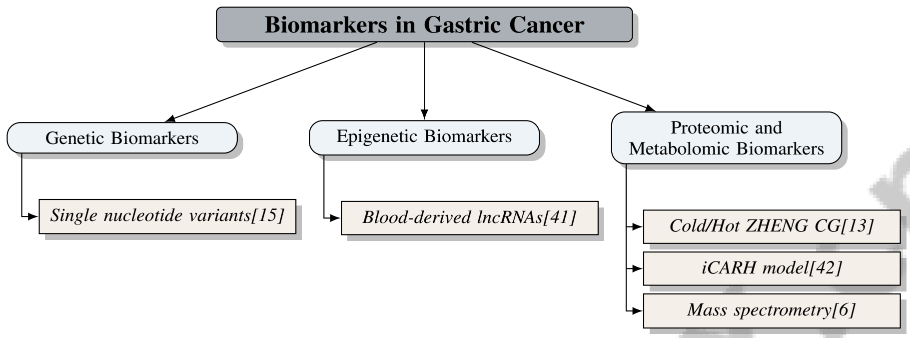

Figure 3: This figure illustrates the classification of biomarkers in gastric cancer, categorizing them into genetic, epigenetic, and proteomic/metabolomic types, each with specific examples and references.

4.2 Recent Advances in Biomarker Discovery

Recent progress in gastric cancer biomarker discovery is marked by the integration of advanced computational methods and high-dimensional data analyses, significantly enhancing biomarker identification and validation. PathFormer exemplifies this advancement, showcasing superior prediction accuracy and reproducibility in biomarker detection, underscoring the role of machine learning in refining biomarker discovery [35]. Statistical learning approaches, particularly those enforcing sparsity, are beneficial for analyzing high-dimensional metabolomics data, aiding in the identification of clinically relevant biomarkers [45]. This approach distinguishes predictive from prognostic biomarkers, refining patient stratification based on clinical outcomes [46]. Network-guided biomarker discovery highlights the need for robust validation protocols, emphasizing additional datasets to confirm biomarker reliability [16].

In genetic biomarker research, the S-space BBT method consistently surpasses other techniques in detection power, enhancing joint SNV analysis capabilities [15]. These methodological advancements are vital for uncovering genetic predispositions in gastric cancer, informing personalized treatment strategies. Current research identifies numerous biomarkers for early diagnosis and monitoring, significantly improving patient outcomes [11]. These discoveries advance gastric cancer diagnostic precision, paving the way for targeted therapeutic interventions. Continuous refinement of computational models and validation strategies is essential for translating biomarker discoveries into clinical practice, ultimately improving gastric cancer management and prognosis.

4.3 Challenges in Biomarker Identification

Biomarker discovery and validation in gastric cancer diagnostics face significant challenges due to biological system complexity and methodological limitations. A major issue is distinguishing predictive from prognostic biomarkers, as current methods often inadequately separate these categories, complicating clinical decision-making [46]. This ambiguity can lead to misinterpretations of biomarker utility, hindering personalized treatment strategies. Many studies are constrained by small sample sizes and demographic biases, limiting generalizability across diverse populations [47]. Biological system complexity further complicates these challenges, introducing variability that can obscure meaningful biomarker-clinical outcome associations. Robust validation protocols are necessary to ensure identified biomarkers' reliability and applicability in clinical settings.

Machine learning integration offers promise in addressing these challenges by enhancing diagnostic accuracy with fewer biomarkers, improving interpretability, and reducing spurious correlation risks

[48]. However, developing and applying these advanced computational methods requires careful model selection and validation to avoid overfitting, ensuring findings are statistically and clinically relevant. A comprehensive strategy is essential to effectively tackle biomarker identification challenges. This includes integrating innovative methodologies—such as generative AI for efficient data embedding and automated training data preparation—with robust validation practices to ensure identified biomarkers' reliability and applicability in personalized medicine [49, 21, 48, 16]. This integrated approach will facilitate the successful translation of biomarker discoveries into clinical tools, enhancing early detection and management of gastric cancer, ultimately improving patient outcomes.

5 Diagnostic Screening and Early Detection

5.1 Current Screening Methods

Endoscopic examination and histological analysis remain the gold standards for gastric cancer (GC) detection but are limited by invasiveness and high costs, making them impractical for widespread screening [50]. Non-invasive alternatives, including serological tests and imaging, are being explored to improve early detection rates, though they often lack the necessary sensitivity and specificity, leading to false positives or negatives that complicate clinical decision-making. Recent developments in microbial analysis, particularly fecal microbial markers, show promise in distinguishing cancerous from non-cancerous states, as demonstrated in colorectal cancer studies [50]. This approach, though nascent in GC, highlights the potential of microbiome profiles as biomarkers for early detection. Integrating these innovative methods with traditional screening could substantially enhance the accuracy and efficiency of GC diagnostics, enabling timely interventions and improved patient outcomes.

5.2 Role of Metabolomics in Enhancing Screening

Metabolomics is transforming GC screening by identifying metabolic changes linked to the disease, offering a non-invasive alternative to traditional methods like endoscopy [9, 51]. Advanced computational techniques, such as the Random Projection Algorithm (RPA), improve predictive accuracy in assessing GC risk by optimizing data dimensionality reduction while maintaining classification accuracy [29]. The LZSR method enhances biomarker discovery reliability in metabolomics, independent of data scale [33]. Frameworks like GERBIL automate biomarker identification, outperforming traditional methods and enhancing metabolomics scalability in clinical settings [21]. Algorithms like ADPA, which adjust processing parameters based on real-time data, further improve large-scale machine learning tasks, leading to more accurate screening outcomes [52]. Integrating metabolomics with genomics and microbiomics provides a comprehensive understanding of GC's molecular basis, aiding the discovery of novel diagnostic markers and therapeutic targets [9]. This holistic approach not only improves screening precision but also supports personalized treatment strategies, enhancing patient outcomes and reducing GC mortality rates.

5.3 Impact on Patient Outcomes

Incorporating metabolomics into GC diagnostics significantly enhances patient outcomes by enabling early intervention and personalized treatment strategies. As illustrated in Figure 4, metabolomics identifies specific metabolic signatures indicative of early disease, which is crucial for timely therapeutic interventions that can improve survival rates and reduce GC mortality [6, 4]. This figure highlights the role of metabolomics in facilitating early intervention and personalized treatment, as well as the enhancement of predictive models through advanced computational techniques. Machine learning algorithms utilizing the Random Projection Algorithm improve classification accuracy for high-risk patients, allowing for targeted interventions based on individual metabolic profiles [29]. This personalized approach not only boosts treatment efficacy but also minimizes adverse effects by aligning therapies with patients' unique metabolic characteristics. Furthermore, combining metabolomics with genomics and proteomics offers a holistic view of the molecular mechanisms underlying GC, facilitating the discovery of new therapeutic targets and refining treatment strategies [24]. This comprehensive approach enables the development of tailored therapies that address dysregulated metabolic pathways in GC, thereby improving treatment outcomes and patient quality of life. As metabolomics technologies advance, their role in GC diagnostics is expected to expand, offering new

AI-generated, for reference only.

opportunities for early intervention and effective disease management, ultimately improving patient outcomes and alleviating the global health burden posed by GC [11].

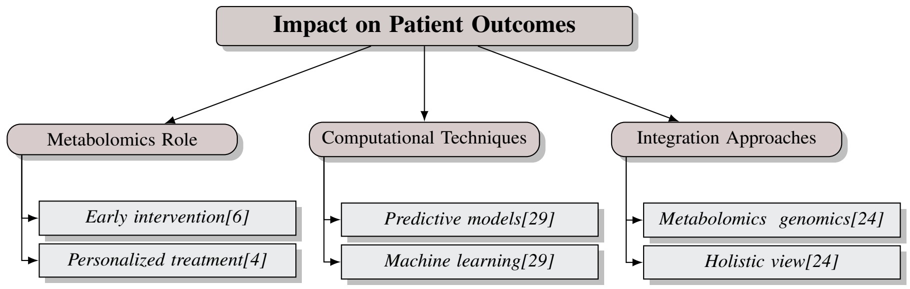

Figure 4: This figure illustrates the impact of metabolomics on patient outcomes in gastric cancer (GC) diagnostics, highlighting its role in early intervention and personalized treatment, the enhancement of predictive models through computational techniques, and the integration with genomics for a holistic understanding of GC.

6 Challenges and Future Directions

6.1 Data Complexity and High-Dimensionality Challenges

Metabolomics encounters significant challenges due to the complexity and high-dimensionality of data, making metabolite identification and quantification difficult. The 'small n, large p' problem, with limited samples compared to the vast number of metabolites, often results in unreliable analyses, necessitating careful feature selection to avoid redundancy and overfitting [6]. Computational challenges include instability in feature selection and ensuring statistical significance of detected modules [16]. Variability and noise in metabolomics data, coupled with high interobserver variability, further complicate analysis [49]. Existing scaling methods may distort results, but the LZSR method mitigates this by ensuring scale independence, aiding consistent biomarker identification [33].

Models like Dynamic Probabilistic Principal Components Analysis (DPPCA) face computational demands, particularly in estimating stochastic volatility parameters, limiting practical applications [18]. Integrative graph convolutional networks (IGCN), while promising, rely on bulk tissue samples, which may obscure cellular heterogeneity [44]. Probabilistic modeling offers nuanced insights into sample fitness but is computationally intensive, with significant hurdles in variable selection and metabolite interdependencies [19, 42]. High-quality training data and the complexity of biological systems pose additional challenges [21].

Addressing these challenges is vital for clinical metabolomics integration, particularly in validating new biomarkers. Future research should focus on robust statistical methods and hybrid approaches to enhance metabolomics data reliability in cancer diagnostics. Techniques like Leave-One-Study-Out (LOSO) can mitigate overfitting by accurately reflecting differences between trials [20].

6.2 Standardization and Methodological Limitations

Metabolomics faces standardization challenges and methodological limitations, primarily due to the lack of standardized protocols for sample collection and preparation, affecting data quality and reproducibility [10]. Variability in sample handling complicates result comparisons and biomarker validation [14]. Methodological limitations arise from data complexity, intercorrelations among metabolites, and missing data, complicating analysis and interpretation [17]. Methods like MetSizeR are limited by binary group comparison assumptions, restricting applicability to complex designs [53]. The ESP method's performance depends on candidate set quality and size, highlighting the need for robust standardization [34].

As illustrated in Figure 5, the figure presents a comprehensive overview of the key challenges and approaches in metabolomics, emphasizing standardization issues, methodological limitations, and statistical approaches. Each category within the figure delineates specific challenges or methods, thereby underscoring the complexity and ongoing efforts to improve metabolomics research. Statistical approaches, like Xu et al.'s pathway testing method, are effective at low significance levels but falter if thresholds are not low enough [32]. Zhao et al.'s adaptive algorithm dynamically adjusts data processing parameters, potentially improving metabolomics studies' adaptability and efficiency [52]. Challenges persist in extreme sparsity or when assumptions about underlying distributions are violated, affecting hypothesis testing robustness [54]. These issues, along with biomarker validation complexities and misclassification risks, underscore the need for precise and standardized approaches in metabolomics research [14]. Addressing these through standardized protocols and advanced analytical techniques is crucial for enhancing metabolomics' reliability and clinical applicability in cancer diagnostics.

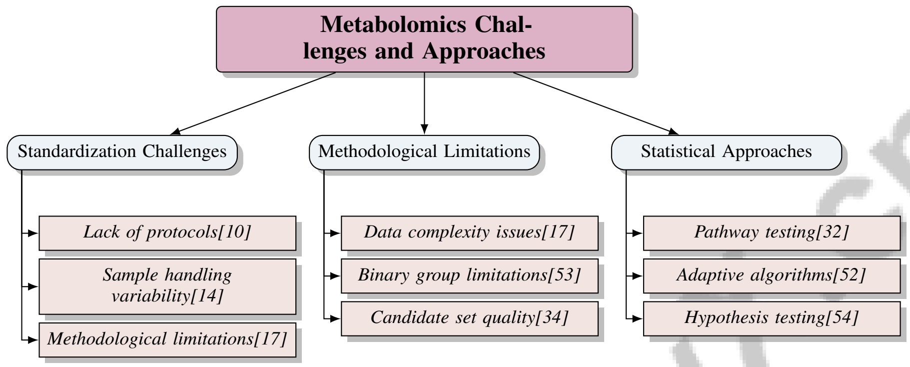

Figure 5: This figure illustrates the key challenges and approaches in metabolomics, highlighting standardization issues, methodological limitations, and statistical approaches. Each category includes specific challenges or methods, underscoring the complexity and ongoing efforts to improve metabolomics research.

6.3 Integration with Other Omics Data

Integrating metabolomics with genomics, transcriptomics, and proteomics provides a comprehensive diagnostic approach, enhancing the understanding of gastric cancer's molecular basis. This multi-omics strategy leverages each omics layer's strengths, offering a holistic view of cancer progression and aiding novel biomarker and therapeutic target identification. Despite its potential, optimal data integration methods and further exploration of metabolomic variations across populations are needed [6].

Future research should optimize diverse data type integration—such as tissue microarray, gene expression, proteomics, and metabolomics—to improve diagnostic model accuracy and reliability. This integration is vital for refining computational pathology workflows, allowing better classification, grouping, and segmentation of heterogeneous data sources. Addressing biomarker identification challenges in translational medicine requires considering clinical trial heterogeneity, impacting diagnostic model performance. Innovative evaluation strategies like leave-one-study-out (LOSO) cross-validation can help account for variability and improve personalized medicine's robustness [49, 20]. This includes refining computational models to better account for systematic errors and exploring their applicability across various biological contexts. Developing standardized methodologies for data integration ensures consistency and reproducibility in multi-omics studies. Advanced machine learning techniques can facilitate integration by reducing human intervention and enhancing automation capabilities. Exploring additional radiomics features and expanding datasets could improve model accuracy and applicability to other cancer types, broadening integration efforts.

Tackling data heterogeneity challenges and improving clinical biomarkers' generalizability could focus on enhancing statistical models that exploit trial heterogeneity. This would enable robust biomarker validation and facilitate multi-omics findings' clinical translation. Validating scale-invariant methods across diverse datasets can enhance metabolomics integration with other omics data, promoting consistent biomarker discovery across biological contexts. Traditional scaling methods can impact biomarker selection and biological interpretation, potentially leading to misleading results. Employing robust, scale-invariant approaches, like logistic zero-sum regression and weighted scaling techniques, ensures metabolite profiles accurately reflect biological conditions, improving metabolomics studies' reliability. Incorporating advanced statistical models that account for experimental design and variable interdependencies can further refine biomarker identification and elucidate complex biological relationships [19, 33, 42, 55].

Metabolomics integration with other omics data significantly enhances gastric cancer diagnostics by detailing the disease's molecular landscape, including identifying specific metabolite biomarkers reflecting cancer-associated metabolic disturbances. This comprehensive approach facilitates deeper exploration of underlying biological mechanisms, improving diagnostic accuracy and potentially guiding personalized treatment strategies [10, 8]. Such integration enhances diagnostic tool precision and paves the way for personalized treatment strategies, significantly improving patient outcomes.

6.4 Technological Advancements and AI Integration

Advanced technologies and artificial intelligence (AI) integration are set to revolutionize metabolomics, particularly in gastric cancer diagnostics. Future research should develop robust methodologies to withstand model misspecifications and extend applicability to other omics data, enhancing metabolomics analyses' reliability and versatility [56]. Improving metabolite identification techniques and expanding mass spectrometry (MS) applications to diverse biological systems and diseases are crucial for broadening metabolomics' scope [56].

Enhancing detection limits for lower concentration metabolites and expanding spectral libraries to include more biofluids are essential for increasing metabolomic analyses' sensitivity and specificity. Integrating metabolomics with other omics data provides a comprehensive understanding of metabolites' roles in biological systems, facilitating novel therapeutic target and diagnostic marker identification [7]. Leveraging new technologies for activity screening will further enhance functional insights from metabolomics studies.

Optimizing AI models for greater scalability and exploring additional biological datasets are pivotal for advancing disease diagnosis and biomarker identification [35]. Enhancing AI model interpretability clarifies identified biomarkers' biological significance, bridging computational predictions and clinical applications. Refining feature selection techniques and integrating additional data types improve AI-driven analyses' robustness, enabling broader applications across biological contexts [57].

Refining machine learning (ML) algorithms for better interpretability and exploring novel ML approaches will address existing metabolomics gaps, enhancing diverse data type integration and facilitating personalized medicine [48]. Developing non-invasive biomarkers and translating molecular classifications into clinical applications are essential for advancing personalized treatment strategies and improving patient outcomes [11]. Future graph neural networks (GNN) work suggests potential technological advancements in improving annotation tasks, indicating ongoing AI methodology evolution within metabolomics [27].

Extending MetSizeR to accommodate different experimental designs and exploring alternative sample size estimation methods, like bootstrap sampling, will enhance metabolomics studies' statistical rigor [53]. Future research should focus on integrating multimodal data, improving model robustness and fairness, and establishing guidelines for ethical AI deployment in clinical settings [51]. Additionally, elucidating circRNA biogenesis and function mechanisms, exploring their roles in the tumor microenvironment, and developing standardized methods for their clinical application as biomarkers and therapeutic targets remain critical areas for further investigation [12]. These technological advancements and AI integrations promise to transform gastric cancer diagnostics, offering more accurate, efficient, and personalized disease management approaches.

7 Conclusion

The advancement of metabolomics and plasma biomarkers marks a pivotal shift in the diagnostic landscape of gastric cancer, enabling earlier detection and the development of personalized treatment strategies. Through comprehensive metabolic profiling, metabolomics uncovers specific alterations that serve as early indicators of gastric cancer, thereby facilitating timely therapeutic interventions that can significantly enhance survival rates and reduce mortality. The integration of sophisticated computational techniques, such as machine learning algorithms, further refines diagnostic accuracy by identifying high-risk patients and tailoring interventions to their unique metabolic profiles. This personalized approach not only optimizes treatment efficacy but also minimizes adverse effects by aligning therapies with individual metabolic characteristics. Moreover, the convergence of metabolomics with other omics disciplines, including genomics and proteomics, offers a holistic view of the molecular underpinnings of gastric cancer. This comprehensive integration aids in the discovery of novel therapeutic targets and the refinement of treatment strategies, ultimately fostering the development of tailored therapies that address the specific metabolic disruptions in gastric cancer. As technologies in metabolomics continue to evolve, their application in gastric cancer diagnostics is expected to expand, opening new avenues for early intervention and effective disease management. By enhancing the precision of diagnostic screening and establishing the foundation for personalized treatment approaches, metabolomics holds the promise of significantly improving patient outcomes and alleviating the global health impact of gastric cancer.

AI-generated, for reference only.

References

[1] Fengtao Zhou, Yingxue Xu, Yanfen Cui, Shenyan Zhang, Yun Zhu, Weiyang He, Jiguang Wang, Xin Wang, Ronald Chan, Louis Ho Shing Lau, Chu Han, Dafu Zhang, Zhenhui Li, and Hao Chen. imd4gc: Incomplete multimodal data integration to advance precise treatment response prediction and survival analysis for gastric cancer, 2024.

[2] Lei Yang, Xiangji Ying, Shuo Liu, Guoqing Lyu, Zekuan Xu, Xi Zhang, Huichao Li, Qingyu Li, Ning Wang, and Jiafu Ji. Gastric cancer: Epidemiology, risk factors and prevention strategies. *Chinese Journal of Cancer Research*, 32(6):695, 2020.

[3] Zhitong Chen, Li Lin, Xiaoqian Cheng, Eda Gjika, and Michael Keidar. Treatment of gastric cancer cells with non-thermal atmospheric plasma generated in water, 2016.

[4] Claudia Burz, Vlad Pop, Ciprian Silaghi, Iulia Lupan, and Gabriel Samasca. Prognosis and treatment of gastric cancer: a 2024 update. *Cancers*, 16(9):1708, 2024.

[5] Cavatorta Ottavia, Scida Serena, Miraglia Chiara, Barchi Alberto, Nouvenne Antonio, Leandro Gioacchino, Meschi Tiziana, Luigi de'Angelis Gian, et al. Epidemiology of gastric cancer and risk factors. *Acta Bio Medica: Atenei Parmensis*, 89(Suppl 8):82, 2018.

[6] Aline Klassen, Andréa Tedesco Faccio, Gisele André Baptista Canuto, Pedro Luis Rocha da Cruz, Henrique Caracho Ribeiro, Marina Franco Maggi Tavares, and Alessandra Sussulini. Metabolomics: definitions and significance in systems biology. *Metabolomics: from fundamentals to clinical applications*, pages 3–17, 2017.

[7] Markus M Rinschen, Julijana Ivanisevic, Martin Giera, and Gary Siuzdak. Identification of bioactive metabolites using activity metabolomics. *Nature reviews Molecular cell biology*, 20(6):353–367, 2019.

[8] Qiang Yang, Ai-hua Zhang, Jian-hua Miao, Hui Sun, Ying Han, Guang-li Yan, Fang-fang Wu, and Xi-jun Wang. Metabolomics biotechnology, applications, and future trends: a systematic review. *RSC advances*, 9(64):37245–37257, 2019.

[9] Aadhith Shankarnarayanan, Dheeman Gangopadhyay, and Ayman Alzaatreh. Multivariate analysis of gut microbiota composition and prevalence of gastric cancer, 2024.

[10] Emily Grace Armitage and Michal Ciborowski. Applications of metabolomics in cancer studies. *Metabolomics: From fundamentals to clinical applications*, pages 209–234, 2017.

[11] Tasuku Matsuoka and Masakazu Yashiro. Biomarkers of gastric cancer: Current topics and future perspective. *World journal of gastroenterology*, 24(26):2818, 2018.

[12] Rong Li, Jiajia Jiang, Hui Shi, Hui Qian, Xu Zhang, and Wenrong Xu. Circrna: a rising star in gastric cancer. *Cellular and Molecular Life Sciences*, 77:1661–1680, 2020.

[13] Boyang Wang, Pan Chen, Peng Zhang, and Shao Li. A network-based biomarkers discovery of cold/hot zheng chronic gastritis and cold/hot herbs of formulae, 2023.

[14] Robert M Califf. Biomarker definitions and their applications. *Experimental biology and medicine*, 243(3):213–221, 2018.

[15] Jin-Xiong Lv, Shikui Tu, and Lei Xu. A comparative study of joint-snvs analysis methods and detection of susceptibility genes for gastric cancer in korean population, 2017.

[16] Chloé-Agathe Azencott. Network-guided biomarker discovery, 2016.

[17] Joseph Antonelli, Brian Claggett, Mir Henglin, Jeramie D. Watrous, Kim A. Lehmann, Pavel Hushcha, Olga Demler, Samia Mora, Teemu Niiranen, Alexandre C. Pereira, Mohit Jain, and Susan Cheng. Statistical methods and workflow for analyzing human metabolomics data, 2018.

[18] Gift Nyamundanda, Isobel Claire Gormley, and Lorraine Brennan. A dynamic probabilistic principal components model for the analysis of longitudinal metabolomic data, 2013.

[19] Barry M. Slaff, Shane T. Jensen, and Aalim M. Weljie. Probabilistic approach for evaluating metabolite sample integrity, 2015.

[20] Yichen Lu, Jane Fridlyand, Tiffany Tang, Ting Qi, Noah Simon, and Ning Leng. The future will be different than today: Model evaluation considerations when developing translational clinical biomarker, 2021.

[21] Wangyang Ying, Dongjie Wang, Xuanming Hu, Ji Qiu, Jin Park, and Yanjie Fu. Revolutionizing biomarker discovery: Leveraging generative ai for bio-knowledge-embedded continuous space exploration, 2024.

[22] Sarah Schmell, Falk Zakrzewski, Walter de Back, Martin Weigert, Uwe Schmidt, Torsten Wenke, Silke Zeugner, Robert Mantey, Christian Sperling, Ingo Roeder, Pia Hoenscheid, Daniela Aust, and Gustavo Baretton. An interpretable automated detection system for fish-based her2 oncogene amplification testing in histo-pathological routine images of breast and gastric cancer diagnostics, 2020.

[23] Anton Ribbenstedt, Haizea Ziarrusta, and Jonathan P Benskin. Development, characterization and comparisons of targeted and non-targeted metabolomics methods. *PLoS One*, 13(11):e0207082, 2018.

[24] Jalal K. Siddiqui, Elizabeth Baskin, Mingrui Liu, Carmen Z. Cantemir-Stone, Bofei Zhang, Russell Bonneville, Joseph P. McElroy, Kevin R. Coombes, and Ewy A. Mathé. Intlim: Integration using linear models of metabolomics and gene expression data, 2018.

[25] M. Perrot-Dockès, C. Lévy-Leduc, J. Chiquet, L. Sansonnet, M. Brégère, M. P. Étienne, S. Robin, and G. Genta-Jouve. A multivariate variable selection approach for analyzing lc-ms metabolomics data, 2017.

[26] Chen Zhao, Kuan-Jui Su, Chong Wu, Xuewei Cao, Qiuying Sha, Wu Li, Zhe Luo, Tian Qin, Chuan Qiu, Lan Juan Zhao, Anqi Liu, Lindong Jiang, Xiao Zhang, Hui Shen, Weihua Zhou, and Hong-Wen Deng. Multi-view variational autoencoder for missing value imputation in untargeted metabolomics, 2024.

[27] Hao Zhu, Liping Liu, and Soha Hassoun. Using graph neural networks for mass spectrometry prediction, 2020.

[28] Ahmed BaniMustafa and Nigel Hardy. Applications of a novel knowledge discovery and data mining process model for metabolomics, 2019.

[29] Seyedehnafiseh Mirniaharikandehei, Morteza Heidari, Gopichandh Danala, Sivaramakrishnan Lakshmivarahan, and Bin Zheng. Applying a random projection algorithm to optimize machine learning model for predicting peritoneal metastasis in gastric cancer patients using ct images, 2020.

[30] Jianliang Gao, Noureddin Sadawi, Ibrahim Karaman, Jake T M Pearce, Pablo Moreno, Anders Larsson, Marco Capuccini, Paul Elliott, Jeremy K Nicholson, Timothy M D Ebbels, and Robert Glen. Metabolomics in the cloud: Scaling computational tools to big data, 2019.

[31] Siamak Ravanbakhsh, Philip Liu, Trent Bjorndahl, Rupasri Mandal, Jason R. Grant, Michael Wilson, Roman Eisner, Igor Sinelnikov, Xiaoyu Hu, Claudio Luchinat, Russell Greiner, and David S. Wishart. Accurate, fully-automated nmr spectral profiling for metabolomics, 2014.

[32] Ningning Xu, Aldo Solari, and Jelle Goeman. Pathway testing in metabolomics with globaltest, allowing post hoc choice of pathways, 2021.

[33] Helena U. Zacharias, Thorsten Rehberg, Sebastian Mehrl, Daniel Richtmann, Tilo Wettig, Peter J. Oefner, Rainer Spang, Wolfram Gronwald, and Michael Altenbuchinger. Scale-invariant biomarker discovery in urine and plasma metabolite fingerprints, 2017.

[34] Xinmeng Li, Hao Zhu, Li ping Liu, and Soha Hassoun. Ensemble spectral prediction (esp) model for metabolite annotation, 2022.

[35] Zehao Dong, Qihang Zhao, Philip R. O. Payne, Michael A Province, Carlos Cruchaga, Muhan Zhang, Tianyu Zhao, Yixin Chen, and Fuhai Li. Highly accurate disease diagnosis and highly reproducible biomarker identification with pathformer, 2024.

[36] Ramtin Hosseini, Neda Hassanpour, Li-Ping Liu, and Soha Hassoun. Pathway-activity likelihood analysis and metabolite annotation for untargeted metabolomics using probabilistic modeling, 2020.

[37] Allison Meisner, Marco Carone, Margaret S. Pepe, and Kathleen F. Kerr. Combining biomarkers by maximizing the true positive rate for a fixed false positive rate, 2019.

[38] Ahmed BaniMustafa and Nigel Hardy. Computer-aided data mining: Automating a novel knowledge discovery and data mining process model for metabolomics, 2019.

[39] Bo Zhao, Luke Bryant, Michael Wilde, Rebecca Cordell, Dahlia Salman, Dorota Ruszkiewicz, Wadah Ibrahim, Amisha Singapuri, Tim Coats, Erol Gaillard, Caroline Beardsmore, Toru Suzuki, Leong Ng, Neil Greening, Paul Thomas, Paul S. Monks, Christopher Brightling, Salman Siddiqui, and Robert C. Free. Labpipe: an extensible informatics platform to streamline management of metabolomics data and metadata, 2019.

[40] Age K. Smilde and Thomas Hankemeier. Numerical representations of metabolic systems, 2019.

[41] Cedric Badowski, Bing He, and Lana X Garmire. Blood-derived lncrnas as potential biomarkers for early cancer diagnosis: The good, the bad and the beauty, 2021.

[42] Takoua Jendoubi and Timothy M. D. Ebbels. Integrative analysis of time course metabolic data and biomarker discovery, 2018.

[43] Elham Gholami, Seyed Reza Kamel Tabbakh, and Maryam Kheirabadi. Proposing method to increase the detection accuracy of stomach cancer based on colour and lint features of tongue using cnn and svm, 2020.

[44] Cagri Ozdemir, Mohammad Al Olaimat, Yashu Vashishath, Serdar Bozdag, and Alzheimer's Disease Neuroimaging Initiative. Igcn: Integrative graph convolution networks for patient level insights and biomarker discovery in multi-omics integration, 2024.

[45] Brian L. Claggett, Joseph Antonelli, Mir Henglin, Jeramie D. Watrous, Kim A. Lehmann, Gabriel Musso, Andrew Correia, Sivani Jonnalagadda, Olga V. Demler, Ramachandran S. Vasan, Martin G. Larson, Mohit Jain, and Susan Cheng. Quantitative comparison of statistical methods for analyzing human metabolomics data, 2018.

[46] Konstantinos Sechidis, Emily Turner, Paul D. Metcalfe, James Weatherall, and Gavin Brown. Ranking biomarkers through mutual information, 2016.

[47] Ahmad Hussein, Mukesh Prasad, and Ali Braytee. Explainable ai methods for multi-omics analysis: A survey, 2024.

[48] Divyagna Bavikadi, Ayushi Agarwal, Shashank Ganta, Yunro Chung, Lusheng Song, Ji Qiu, and Paulo Shakarian. Machine learning driven biomarker selection for medical diagnosis, 2024.

[49] Thomas J. Fuchs and Joachim M. Buhmann. Computational pathology: Challenges and promises for tissue analysis, 2015.

[50] Jian Jiang. Miidl: a python package for microbial biomarkers identification powered by interpretable deep learning, 2021.

[51] Hamideh Kerdegari, Kyle Higgins, Dennis Veselkov, Ivan Laponogov, Inese Polaka, Miguel Coimbra, Junior Andrea Pescino, Marcis Leja, Mario Dinis-Ribeiro, Tania Fleitas Kanonnikoff, and Kirill Veselkov. Foundational models for pathology and endoscopy images: Application for gastric inflammation, 2024.

[52] Zhi Zhao, John Zobolas, Manuela Zucknick, and Tero Aittokallio. Tutorial on survival modeling with applications to omics data, 2024.

[53] Gift Nyamundanda, Isobel Claire Gormley, Yue Fan, William M Gallagher, and Lorraine Brennan. Metsizer: selecting the optimal sample size for metabolomic studies using an analysis based approach, 2013.

[54] Rong Ma, T. Tony Cai, and Hongzhe Li. Global and simultaneous hypothesis testing for high-dimensional logistic regression models, 2020.

[55] Biplab Biswas, Nishith Kumar, Md Aminul Hoque, and Md Ashad Alam. Weighted scaling approach for metabolomics data analysis, 2022.

[56] Scott Powers, Trevor Hastie, and Robert Tibshirani. Customized training with an application to mass spectrometric imaging of cancer tissue, 2016.

[57] Charanpal Dhanjal and Stéphan Clémençon. On recent advances in supervised ranking for metabolite profiling, 2014.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.