HCMMD: systematic evaluation of metabolites in body fluids as liquid biopsy biomarker for human cancers

Xun Dong¹*, Yaoyao Qu¹*, Tongtong Sheng²*, Yuanming Fan¹, Silu Chen², Qinbo Yuan³, Gaoxiang Ma¹,4,5, Yuqiu Ge⁶

¹State Key Laboratory of Natural Medicines, School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing, China
²School of Public Health, Nanjing Medical University, Nanjing, China
³Department of Urology, Wuxi Fifth People’s Hospital, Wuxi, China
⁴The Clinical Metabolomics Center, China Pharmaceutical University, Nanjing, China
⁵Department of Oncology, Pukou Hospital of Chinese Medicine affiliated to China Pharmaceutical University, Nanjing, China
⁶Department of Public Health and Preventive Medicine, Wuxi School of Medicine, Jiangnan University, Wuxi, China

*Equal contribution

Correspondence to: Qinbo Yuan, Gaoxiang Ma, Yuqiu Ge; email: qinbo_yuan@126.com, https://orcid.org/0000-0003-3408-0865; gaoxiang_ma@163.com, https://orcid.org/0000-0002-1774-6445; yuqiu_ge0816@jiangnan.edu.cn

Keywords: cancer, liquid biopsy, metabolites, diagnostic biomarker, database

Received: October 27, 2023 Accepted: January 3, 2024 Published: April 26, 2024

Copyright: © 2024 Dong et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Metabolomics is a rapidly expanding field in systems biology used to measure alterations of metabolites and identify metabolic biomarkers in response to disease processes. The discovery of metabolic biomarkers can improve early diagnosis, prognostic prediction, and therapeutic intervention for cancers. However, there are currently no databases that provide a comprehensive evaluation of the relationship between metabolites and cancer processes. In this review, we summarize reported metabolites in body fluids across pan-cancers and characterize their clinical applications in liquid biopsy. We conducted a search for metabolic biomarkers using the keywords (“metabolomics” OR “metabolite”) AND “cancer” in PubMed. Of the 22,254 articles retrieved, 792 were deemed potentially relevant for further review. Ultimately, we included data from 573,300 samples and 17,083 metabolic biomarkers. We collected information on cancer types, sample size, the human metabolome database (HMDB) ID, metabolic pathway, area under the curve (AUC), sensitivity and specificity of metabolites, sample source, detection method, and clinical features were collected. Finally, we developed a user-friendly online database, the Human Cancer Metabolic Markers Database (HCMMD), which allows users to query, browse, and download metabolite information. In conclusion, HCMMD provides an important resource to assist researchers in reviewing metabolic biomarkers for diagnosis and progression of cancers.

INTRODUCTION

Cancer is a major global public health problem, with a significant impact on mortality worldwide. In 2020, an estimated 10 million people died of cancer globally [1]. Early diagnosis is crucial for effective cancer management and better prognosis. However, approximately 50% of cancers are diagnosed at advanced stages [2, 3]. Effective treatment of advanced cancer often involves the use of modern systemic and targeted drugs, which can be costly and may have limited efficacy [4]. Early cancer detection has been
shown to provide substantial health benefits, including increased survival rates and reduced morbidity [2]. Although several blood-based biomarkers, such as carcinoembryonic antigen (CEA) and prostate-specific antigen (PSA), have been used for cancer screening in the past few decades, their sensitivity and specificity have been found to be unsatisfactory, limiting their effectiveness [5]. Therefore, there is a pressing need to identify biomarkers that exhibit high sensitivity and specificity for the early detection of cancer.

Metabolomics, which involves the comprehensive analysis of small molecule metabolites in cells, tissues, or whole organisms, has undergone rapid technological evolution in the past two decades [6–8]. By measuring downstream chemical phenotypes of genomic, transcriptomic, and proteomic variability, metabolomics can provide a more comprehensive understanding of the biological system [6, 9, 10]. Research has shown that metabolites play a crucial role in various diseases such as obesity, diabetes, cardiovascular disease, respiratory conditions, and cancer [6, 11]. Metabolomics has emerged as an accurate and non-invasive diagnostic tool, accompanied by the development of novel and sensitive measurement techniques [12, 13]. The uncontrolled proliferation of tumor cells requires metabolic regulation [14–16], and metabolic reprogramming is a hallmark of malignancy [17]. In recent years, several highly sensitive and specific metabolic biomarkers have been identified in liquid biopsy studies. For instance, Sreekumar et al. reported that sarcosine had a diagnostic value with an AUC of 0.69 (95% CI: 0.55, 0.84) for prostate cancer [18]. Soga et al. discovered that serum γ-glutamyl dipeptides had an AUC of 0.76 for hepatocellular carcinoma [19]. Tyrosine and glutamine-leucine in serum had an AUC of 0.98 for the diagnosis of colorectal cancer [20]. N¹, N¹²-diacylspermine in serum had an AUC of 0.65 (95% CI, 0.59 to 0.72) for the diagnosis of non–small-cell lung cancer [21]. The AUC value of creatine nucleoside in urine was 0.79 for differential diagnosis between adrenocortical carcinoma and benign adrenal tumors [22].

This study aimed to comprehensively evaluate the role of metabolites in cancers. Cancer-related metabolites were searched from the PubMed database. The collected information includes cancer types, sample size, HMDB ID, metabolic pathway, area under the curve (AUC), sensitivity, specificity, sample source, detection method, and clinical features. Importantly, a user-friendly online database was developed, named Human Cancer Metabolic Markers Database (HCMMD), to assist users in querying, browsing, and downloading information about the cancer-related metabolites.

Advances in metabolomics

Metabolomic analysis is a technique used to analyze the type and content of small molecule metabolites in biological samples [23]. Four major technologies are commonly used for metabolomics: gas chromatography mass spectrometry (GC-MS) [24], liquid chromatography mass spectrometry (LC-MS) [25], capillary electrophoresis mass spectrometry (CE-MS) [26], and nuclear magnetic resonance spectroscopy (NMR) [27]. These techniques can assess changes in metabolic processes and provide a summary of alterations at the DNA, RNA, and protein levels [28]. Metabolomics has been used to reveal the mechanisms of basal metabolic processes in diseases [29, 30], and in some cases, it may be the most sensitive method for identifying the pathological state of cancer patients, as even small changes in gene or protein expression can lead to remarkable changes in protein activity and metabolite levels [31–33]. Metabolomics can provide an effective method for screening for cancer, guiding treatment strategies, assessing efficacy, and tracking cancer progression [34–37]. Additionally, it can help to identify therapeutic targets and promote drug discovery [38, 39].

Metabolism-related biomarkers for cancers

Database establishment

To search for cancer-related metabolites from PubMed, we used the keywords “metabolomics” OR “metabolite” AND “cancer”. The eligible data included 22,254 articles published before July 30, 2022. We selected literature that focused on human tumors, samples from liquid biopsies, and individual biomarkers with diagnostic, prognostic, and predictive value. After filtering the results, we identified 792 studies with a total of 573,300 samples and 17,083 metabolic biomarkers. The included studies covered 24 types of cancer derived from 92 subtypes. We recorded information such as cancer type, sample size, HMDB ID, metabolic pathway, AUC, sensitivity, specificity, sample source, detection method, and clinical features. Figure 1 summarizes the basic information and diagnostic value of metabolic biomarkers in different cancers. Importantly, we created a new online database, called the HCMMD, which allows users to explore and analyze cancer-related metabolic biomarkers (Figure 2).

Digestive system tumors

Digestive system cancer encompasses oral cancer, esophageal cancer, gastric cancer, colorectal cancer, liver cancer, pancreatic cancer, and gallbladder cancer.
The five-year survival rates for liver cancer and pancreatic cancer are only 20% and 11%, respectively [40]. A total of 315 articles were collected, covering 7 cancer types and 23 subtypes, with sample sizes ranging from 9 to 3,109 [20, 41]. The reports included 6,731 diagnostic biomarkers, 64 progressive biomarkers, and 163 prognostic biomarkers. Of these biomarkers, 3,609 were from serum samples, 1,619 from plasma samples, 974 from urine samples, and 460 from saliva samples. Among the biomarkers, 1,763 belonged to amino acids and 246 belonged to bile acids. Several studies have shown that bile acids are closely related to digestive system tumors [23, 42, 43]. Of the biomarkers, 79 had an AUC ≥0.95, and 420 had an AUC ≥0.80.

To explore the underlying pathogenesis of digestive system cancer, metabolite pathway enrichment analysis was conducted (Supplementary Figure 1). As shown in Figure 3, the four most closely related pathways for digestive system cancer were glycine serine and threonine metabolism, arginine biosynthesis, alanine aspartate and glutamate metabolism, and valine leucine and isoleucine biosynthesis.

Reproductive system tumors

Reproductive system cancer comprises breast cancer, prostate cancer, endometrial cancer, cervical cancer, and ovarian cancer. Among all cancers, new cases of reproductive system cancer rank first [40]. In this review, we collected 238 articles covering 5 cancer types and 14 subtypes. The minimum sample size was 12 [44], and the maximum sample size was 6,114 [45]. These studies included 3,701 diagnostic biomarkers, 73 progression biomarkers, and 429 prognostic biomarkers. Of these metabolic markers, 1,462 biomarkers were from serum samples, 1,715 were from plasma samples, and 750 were from urine samples. Among the metabolic markers, 1,044 belonged to amino acids. A total of 213 diagnostic markers had an AUC ≥0.80.

Metabolite pathway enrichment analysis of diagnostic metabolites was performed for reproductive system cancers (Supplementary Figure 2). The four most closely related pathways for reproductive system cancer were glycine serine and threonine metabolism, alanine aspartate and glutamate metabolism, arginine and proline metabolism, and valine leucine and isoleucine biosynthesis. Figure 4 shows the three most significant pathways involved in reproductive system cancer.

Respiratory system tumors

Respiratory system cancer includes lung cancer and throat cancer. Lung cancer is the leading cause of cancer-related morbidity and mortality worldwide, with a five-year relative survival rate of only 22% [1, 40]. In this review, we collected 102 publications on respiratory system tumors, including 2 cancer types and 8 cancer subtypes. The sample size ranged from 14 to 1,196 [46, 47]. These publications included 1,976 diagnostic biomarkers and 204 prognostic biomarkers. Among these biomarkers, 1,335 biomarkers came from serum samples, and 467 biomarkers came from plasma samples. In addition, metabolites from breath samples were used to diagnose lung cancer in seven studies [46, 48–53]. These studies identified 162 diagnostic markers with an AUC greater than or equal to 0.80. We also
enriched the metabolic pathways of diagnostic metabolites in respiratory system cancers (Supplementary Figure 3). The main relevant pathways were aminoacyl-tRNA biosynthesis, arginine biosynthesis, glycine serine and threonine metabolism, and glyoxylate and dicarboxylate metabolism.
Urinary system tumors

Urinary system cancer includes bladder cancer, kidney cancer, and ureteral cancer. Bladder cancer is known for its high recurrence rate worldwide [1, 40]. In this review, we collected 77 publications on urinary system tumors, encompassing 3 cancer types and 17 cancer subtypes. The sample size varied

Figure 3. The four most related metabolic pathways in digestive system tumors. The metabolic process of metabolites and the proteins involved in their regulation. Differentially expressed metabolites (red), unchanged metabolites (grey), regulated proteins (yellow), metabolic pathway names (blue). Abbreviations: ACY1: aminoacylase; ADSL: adenylosuccinate lyase; ADSS1: adenylosuccinate synthase 1; AGST: alanine-glyoxylate transaminase; ARG1: arginase 1; AS: asparagine synthase; ASL: argininosuccinate lyase; ASS1: argininosuccinate synthase 1; GCAT: glycine C-acetyltransferase; GLS: glutaminase; GLT1: glutamate synthase; GLUT1: glucose transporter 1; glyA: glycine hydroxymethyltransferase; GOT1L1: glutamic-oxaloacetic transaminase 1 like 1; ilvE: branched-chain amino acid aminotransferase; IL4I1: interleukin 4 induced 1; OTC: ornithine transcarbamylase; PC: pyruvate carboxylase; SDS: threonine ammonia-lyase; SCL1A5: solute carrier family 1; TDH: threonine 3-dehydrogenase.
from 12 to 2,610 [54, 55]. The publications included 1,779 diagnostic biomarkers, 135 progression biomarkers, and 44 prognostic biomarkers. Urine samples were the primary diagnostic tool for urinary system cancers, with 51 articles including 1,483 biomarkers from urine used to detect these cancers. Additionally, there were 382 biomarkers from serum samples and 93 biomarkers from plasma samples. We identified 128 diagnostic markers with an AUC of 0.80 or higher. The enrichment analysis of metabolic pathways of diagnostic metabolites in urinary system cancer is presented in Supplementary Figure 4, showing that these metabolites were enriched in aminoacyl-tRNA biosynthesis, steroid hormone biosynthesis, glycine serine and threonine metabolism, and arginine biosynthesis.

**Nervous system tumors**

Nervous system cancer, including brain cancer and other types of nervous system cancer, is one of the deadliest cancers, with 18,600 people in the United States dying from the disease in 2021 [40, 56]. Six cancer subtypes were reviewed in 16 articles, with sample sizes ranging from 17 to 220 [57, 58]. The studies included 223 diagnostic biomarkers, 31 progression biomarkers, and 13 prognostic biomarkers. Biomarkers were derived from serum samples (92 biomarkers), plasma samples (118 biomarkers), and cerebrospinal fluid samples (38 biomarkers). Of these studies, 24 diagnostic markers had an AUC of 0.80 or higher. A metabolic pathway enrichment analysis of diagnostic metabolites in nervous system cancers was conducted, and the four most relevant pathways were aminoacyl-tRNA biosynthesis, arginine biosynthesis, alanine aspartate and glutamate metabolism, and glyoxylate and dicarboxylate metabolism (Supplementary Figure 5).

**Other tumors**

In addition to the aforementioned common cancers, abnormal metabolites have also been observed in other types of cancer. We collected 65 articles for the diagnosis of other cancers, including six cancer types (thyroid cancer, myeloma, leukemia, lymphoma, skin cancer, and osteosarcoma) and 24 cancer subtypes. The sample sizes ranged from 10 to 846 [59, 60]. Those studies included 1,393 diagnostic biomarkers, 27 progression biomarkers, and 97 prognostic biomarkers. Biomarkers were mainly derived from three types of samples: 725 biomarkers in serum samples, 586 biomarkers in plasma samples, and 105 biomarkers in urine samples. Those publications contained a total of 66 diagnostic markers with an AUC of 0.80 or higher.

**DISCUSSION**

Cancer is one of the major threats to human health because of its high morbidity and mortality rates [61]. Highly specific and sensitive diagnostic or prognostic biomarkers can improve the efficiency of treatment and prolong the survival of patients [2]. Metabolomics has many exciting opportunities to promote the treatment of cancer [62]. For example, metabolomics combined with other “omics” can uncover valuable drug targets [63–65]. Metabolomics also has the potential to influence cancer screening and diagnosis. Since many studies have identified biomarkers in body fluids with high diagnostic value for human cancers [66, 67], Zhou et al. reported that 4-Dodecylbenzenesulfonic acid, PC (30:1) and PC (44:5) were downregulated in the serum of colorectal adenoma patients compared to healthy subjects, with an AUC of 1.00 [68]. Plasma levels of beta-sitosterol were upregulated in pancreatic cancer patients compared to healthy individuals with an AUC value of 0.99 [69]. Plasma of hexadecasphinganine with an AUC value of 0.99 in the diagnosis of gastric cancer [70]. Serum levels of hypoxanthine were upregulated in patients with lung adenocarcinoma compared to normal controls with an AUC value of 0.99 [71]. Jové et al. found that hexanoic acid in the plasma had an AUC value of 1.00 for breast cancer diagnosis [72]. Metabolic pathway enrichment analysis is a good method to discover potential pathogenesis of different systemic cancers. Glycine serine and threonine metabolism and arginine biosynthesis were enriched in each system of cancer. These two metabolic pathways may provide inspiration for future cancer research.

Although many liquid biopsy biomarkers with high diagnostic value in human cancers have been reported, there are still difficulties and challenges in the clinical application of these metabolites. First, there is a lack of multi-center, large-scale studies to validate the clinical feasibility and reproducibility of metabolic markers [73]. Second, in order to incorporate biomarker assays into the clinical workflow, supporting assay resources, staff logistics, and technical education are needed, which can be costly in the clinic [73, 74]. Third, there are huge fluctuations in the concentration of metabolites in vivo, as well as a fragmented distribution of specialized small molecules in the body [75]. In addition, metabolomics is diverse and chemically complex, and varies in different tumor lesions. For example, L-alanine is significantly downregulated in pancreatic cancer but significantly upregulated in colorectal cancer, which adds great difficulty for tumor screening [7, 47, 76]. This article has some innovations that need to be clarified. First, previous studies mainly focused on biopsy markers of a single tumor and lacked a summary of diagnostic data on multiple tumor biopsies. This article summarizes
diagnostic markers covering 24 tumor types. Second, our database contains more detailed information, such as AUC, accuracy, specificity, HMDB ID, metabolic pathway, sample source and so on. In addition, pathway analysis demonstrated that glycine, serine and threonine metabolism and arginine biosynthesis metabolic pathways were enriched in multiple cancer systems, suggesting that these two metabolic pathways play an important role in cancer diagnosis and treatment.

CONCLUSION

With the development of standardized protocols, the measurement of metabolomics has become cheaper and more convenient. Metabolomics plays an increasingly important role in cancers, alongside other diagnostic and prognostic tests in the clinic. To provide an important resource for users to query, browse, and download information on cancer-related metabolites, we have established a user-friendly website.

Abbreviations

HMDB: human metabolome database; AUC: area under curve; HCMMD: Human Cancer Metabolic Markers Database; CEA: carcinoembryonic antigen; PSA: prostate-specific antigen; GC-MS: gas chromatography mass spectrometry; LC-MS: liquid chromatography mass spectrometry; CE-MS: capillary electrophoresis mass spectrometry; NMR: nuclear magnetic resonance spectroscopy.

AUTHOR CONTRIBUTIONS

XD drafted the manuscript and collected the literatures. XD and YMF prepared Figures, Tables and website. GXM and YQG designed the study and revised the manuscript. All authors approved the final version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This work was partly supported by the National Natural Science Foundation of China (82204146).

REFERENCES


production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. Sci Adv. 2020; 6:eaa5781. https://doi.org/10.1126/sciadv.aay5781 PMID: 31998842


PMID: 25261994


PMID: 27746051


PMID: 35331175


PMID: 32278036


PMID: 33777793


PMID: 34674559


PMID: 28076849


PMID: 34732219


PMID: 31092915

https://doi.org/10.1021/acs.jproteome.0c00990
PMID:33617266
Supplementary Figures

Supplementary Figure 1. Top 25 metabolic pathways for diagnostic metabolite enrichment analysis in digestive system tumors.
Supplementary Figure 2. Top 25 metabolic pathways for diagnostic metabolite enrichment analysis in reproductive system tumors.
Supplementary Figure 3. Top 25 metabolic pathways for diagnostic metabolite enrichment analysis in respiratory system tumors.
Supplementary Figure 4. Top 25 metabolic pathways for diagnostic metabolite enrichment analysis in urinary system tumors.
Supplementary Figure 5. Top 25 metabolic pathways for diagnostic metabolite enrichment analysis in nervous system tumors.