81663_Auto_Edited-check.docx

Novel biomarkers for early detection of gastric cancer

Matsuoka T *et al*. Novel biomarkers of GC

Abstract

Gastric cancer (GC) remains a leading cause of cancer-related death worldwide. Less than half of GC cases are diagnosed at an advanced stage due to its lack of early symptoms. GC is a heterogeneous disease associated with a number of genetic and somatic mutations. Early detection and effective monitoring of tumor progression are essential for reducing GC disease burden and mortality. The current widespread use of semi-invasive endoscopic methods and radiologic approaches has increased the number of treatable cancers: However, these approaches are invasive, costly, and time-consuming. Thus, novel molecular noninvasive tests that detect GC alterations seem to be more sensitive and specific compared to the current methods. Recent technological advances have enabled the detection of blood-based biomarkers that could be used as diagnostic indicators and for monitoring postsurgical minimal residual disease. These biomarkers include circulating DNA, RNA, extracellular vesicles, and proteins, and their clinical applications are currently being investigated. The identification of ideal diagnostic markers for GC that have high sensitivity and specificity would improve survival rates and contribute to the advancement of precision medicine. This review provides an overview of current topics regarding the novel, recently developed diagnostic markers for GC.

Core Tip: Gastric cancer (GC) is a major unsolved clinical problem. It is the fifth most [35] diagnosed cancer in the world, and the fourth most frequent cause of cancer-related deaths worldwide. The outcome of advanced GC remains tremendously poor despite the use of surgery and adjuvant therapy. Therefore, the detection new biomarkers for early diagnosis will be potential therapeutic strategies for improving survival of GC

patients. The aim of this review is to summarizes the current status and approaches for novel biomarkers, which could be useful for early diagnosis.

INTRODUCTION

Gastric cancer (GC) is the fifth most common cancer and the fourth most common cause of cancer-related death worldwide[1]. The overall survival rate of individuals with GC remains poor, because it is often diagnosed at an advanced stage or even as metastasis. One strategy for lowering the burden of GC is early screening and diagnosis. Even though upper gastrointestinal endoscopy has been well established as the standard for GC screening, capturing a concealed tumor in the stomach requires higher sensitivity and specificity. Upper gastrointestinal endoscopies also involve high costs with regard to cost and human resources, requiring experienced endoscopists and patient understanding.

To achieve large-scale GC screening among healthy individuals in the future, a more general and cost-effective approach is needed. Cancer biomarkers (tumor markers) are molecules indicative of the presence and progression of cancer; as such, they play a crucial role in cancer diagnoses and treatment selection[2]. Their applications are rapidly expanding owning to advances in genetic analysis technology and molecular targeting. There are at present no biomarkers with sufficient sensitivity and specificity for the diagnosis of GC that could be applied in clinical settings. Biomarkers are needed for every stage of GC to improve its clinical course. The emergence of liquid biopsy technology, which enables the detection of specific molecular information of solid tumors from body fluids, has provided major changes in cancer diagnosis and treatment[3]. There are many studies on biomarkers for GC screening. However, many of these candidate biomarkers identified thus far are preferentially expressed at advanced stages, and thus unsuitable for early detection. Moreover, a comprehensive review focusing on early detection of GC is still lacking. Herein, we review novel candidates of biomarkers focusing on early detection of GC comprehensively, from the

laboratory to the clinical prevalence and the future prospects. This information will contribute to further research on GC biomarkers and their clinical applications.

METHODS

A non-systematic review was performed based on an electronic search through the medical literature using PubMed and Google Scholar. The keywords "biomarker", "gastric cancer", "early detection", "diagnostics", "liquid biopsy" were used. Review articles and guidelines investigating the values of biomarker for early detection of GC from gastroenterology, oncology and genetics were included in this review. When more than one guideline concerning the same subject was available, the most up-to-date one was selected. Only full articles in the English language published in the last ten years were considered for further review. Great importance was also given to "Clinical study" and "Review" articles dealing with the topic. The exclusion criteria comprised duplicated articles, studies absent of diagnostic outcomes. Case reports, correspondences, letters, and non-human research were not included.

LIQUID BIOPSY

Various diverse molecules in blood (serum/plasma) have been said to be correlated with GC carcinogenesis. Among them, DNA, proteins, and noncoding RNA (ncRNA) molecules have been focused. These molecules allow the dependable finding of new noninvasive biomarkers for GC in plasma/serum. Liquid biopsy has shown good potential in detecting circulating tumor cells (CTCs) from any bodily fluid, including peripheral blood, urine and cerebrospinal fluid, ascites, pleural effusion, *etc.*, and includes a genomic, proteomic and metabolite assessment. It is a quick, easy and inexpensive method with minimal invasiveness, making it widely accepted by patients, without major side effects[4]. Liquid biopsy (using mainly peripheral blood) can be used as a screening method to detect early GC (EGC) and minimal residual disease (MRD) after surgery and can create an appropriate treatment plan in GC patients.

CTCs/circulating tumor DNA

CTCs defined as tumor cells that have been detached from the primary tumor or metastases and shed in patient's bloodstream, characterize an easily attainable sample of cancer specimens. As a liquid biopsy, emerging technologies for CTC isolation have enabled the research on the pathology of CTCs and have prompted the clinical uses of CTCs in the diagnosis of malignancies[5]. Although many researchers have come to the same conclusion that CTCs positivity was associated with prognosis, few researches were found for the early diagnosis, especially in the last several years.

Circulating cell-free DNA (cfDNA) consists of fragments of extracellular DNA that can be found in the blood[6]. cfDNA changes across different pathological states have been successfully exploited as noninvasive biomarkers for liquid biopsies. Particularly, the cfDNA that is identified from tumors and shows tumor-specific mutations is named circulating tumor DNA (ctDNA)[7]. The analysis of ctDNA has been predominantly of interest over a decade as its remarkable stability in body fluids. Besides, accumulating evidences showed that ctDNA is useful for early incidence during disease progress after surgical procedure. Further improvement of the ctDNA detection system is progressing. The cfDNA-based liquid biopsy for early diagnosis is mostly restrained by the low signals shed from early-stage cancer, which directly influence the sensitivity of sequencing analyses.

Recent paper has described that the circulating biomarkers, such as cfDNA was highly expressed in stage I GC patients compared with those in benign gastric disease patients and health controls, suggesting the potential for their use for the early detection of GC[8,9]. The human epidermal growth factor receptor 2 (HER2) gene copy number detected in ctDNA was higher in GC patients compared to healthy control. Elucidation of HER2 copy number variation could distinguish healthy individuals and patients with GCs with 58% sensitivity and 98% specificity [area under curve (AUC) = 0.707, 95% confidence interval (CI): 0.593-0.821][10]. ctDNA sequencing identifies fibroblast growth factor receptor 2 (FGFR2) amplification missed by tissue testing in patients with advanced GC, and these patients may respond to FGFR inhibition. ctDNA can detect

plasma *FGFR2* alterations, occasionally at a higher frequency than tissue testing, identify patients and these patients may respond to FGFR inhibition[11].

Current cancer biomarkers can be categorized into two groups: (1) Biomarkers generally used for the detection of specific cancer types; and (2) Biomarkers to be aimed to decide personalized medicine, such as companion diagnostics. Even when surgical procedure is successfully carried out, detecting MRD after surgery is still challenging for doctors, and it is difficult to detect MRD in time using radiologic imaging or tissue biopsy. Recent studies presented that ctDNA-based testing were able to detect MRD several weeks earlier than imaging[12]. Another study monitoring ctDNA in postoperative blood has verified that, the median duration from positive ctDNA detection to relapse was 4.05 mo[13]. Moreover, the postoperative positive ctDNA was significantly correlated with tumor recurrence within 12 mo after surgery, suggesting that postoperative ctDNA monitoring possess clinical application value in the prediction of postoperative GC recurrence[13]. In a similar, ctDNA positivity preceded radiographic recurrence by a median of 6 mo, suggesting that, ctDNA is a sensitive and specific biomarker for finding patients at high risk for relapse after curative treatment[14]. ctDNA-positive patients were at higher risk of relapse and exhibited worse outcome compared to the ctDNA-negative group, implying that ctDNA monitoring in the post-treatment setting is predicted to offer knowledge about necessity for further therapy[15]. The clinical usefulness of ctDNA as biomarker for longitudinal disease monitoring and identification of MRD in GC was assessed by ultra-deep sequencing methods. The occurrence of relapse was 100% (7 of 7), 100% (5 of 5), and 94% (16 of 17), respectively in patients with measurable ctDNA at the post-operative, post-adjuvant chemotherapy, and consequent longitudinal time spots. To date, novel high throughput sequencing techniques such as next-generation sequencing (NGS) or whole exosome sequence can identify mutations in multiple genetic regions[16]. Recent study conducted surveillance for recurrent disease after curative surgical resection analyzed by whole genome sequence. For 1630 patients with GC and esophageal adenocarcinomas, detection of MRD after successful surgical procedure is intensely

correlated with an enhanced risk of relapse by utilizing a Guardant 360, ctDNA-NGS assay[17]. These studies elucidate that risk stratification of patients for adjuvant chemotherapy escalation/de-escalation could be predicted by incorporation of ctDNA. ctDNA-guided treatment strategy in post-surgery surveillance for patients with GC is a novel emerging strategy that will likely replace the current treatment decision, as outlined in Figure 1.

NC RNA

Diverse ncRNAs, such as microRNA (miRNA), circular RNA (circRNA), and long ncRNA (lncRNA), are implicated in GC carcinogenesis[18]. MiRNAs are small ncRNA molecules of approximately 18-22 nucleotides that regulate cell differentiation, proliferation, apoptosis and gene regulation[19]. Currently, due to the stability and specificity of expression in circulation, accumulating evidences suggest that miRNAs can serve as novel biomarkers with a prospective clinical significance tool for patient with GC. MiRNAs are very stable in blood and can be identified by various assays such as reverse transcription quantitative real-time polymerase chain reaction, and NGS. Recent studies challenged to verify the importance of various miRNAs in human plasma/serum as diagnostic markers for early-stage GC, including miR-17, miR-25, and miR-133b[20], miR-425-5p, miR-1180-3p, miR-122-5p, miR-24-3p, and miR-4632-5p[21]. A recent distinctive expression study of miRNA profiles utilizing two datasets from plasma samples identified hsa-miR-320a, hsa-miR-1260b, and hsa-miR-6515.5p as potential biomarkers for primary diagnosis of GC, with an AUC higher than 0.95 in all cases. Furthermore, hsa-miR-320a and hsa-miR-1260b was more stable in all the discovery and validation datasets[22]. Similarly, a recent study developed an EGC index to differentiate EGC from non-cancer controls based on the serum levels of four miRNAs (miR-4257, miR-6785-5p, miR-187-5p, and miR-5739). This index provided a sensitivity of 0.996 and a specificity of 0.953, with an AUC of 0.998 for the detection of EGC in a large sample size study (n = 1417)[23]. These results suggest that miRNAs could act as promising biomarkers for the early detection of GC. Nevertheless, further

research is needed to optimize their detection methods and reproducibility of accrued results to meet the rigorous requirements for clinical setting.

LncRNAs are the transcripts of more than 200 nucleotides that have been found to participate widely in various pathological processes of organisms[24]. In colon cancer, LncRNA small nucleolar RNA host gene 11 has been presented as a promising biomarker for early detection of this malignancy[25]. Meanwhile, hepatocellular carcinoma upregulated lncRNA and ZNFX1 antisense RNA 1 have been shown to distinguish GC patients from healthy subjects and proposed as novel diagnostic biomarkers of GC[26]. LncRNAs promoter of CDKN1A antisense DNA damage activated RNA, FOXD2 adjacent opposite strand RNA 1, and SWI/SN F related, matrix associated, actin dependent regulator of chromatin subfamily c member 2 in plasma may serve as novel biomarkers for early detection of GC. The combined these three lncRNAs offer a high sufficient diagnostic sensitivity[27]. LncRNA B3GALT5 antisense RNA 1 was shown to be better than carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) and may serve as a diagnostic biomarker of GC[28]. Distinct levels of serum C5orf66 antisense RNA 1 (C5orf66-AS1) were shown between patients with GC and gastritis. A risk of gastric dysplasia and GC negatively associated with serum expression of C5orf66-AS1, implying that C5orf66-AS1 a potential biomarker for predicting EGC[29]. Combined lncRNAs have been displayed to improve biomarker values for early detection of GC. LncRNA HLA Complex P5 (HCP5) is a kind of RNA gene without protein-coding. The AUC of serum lncRNA HCP5 was 0.818 (95%CI: 0.757-0.880, 80% sensitivity, and 70% specificity) in distinguishing between GC and healthy controls. Besides, the highest AUC of 0.870 (95%CI: 0.819-0.921) with 81% sensitivity and 79% specificity were revealed by the combination of three diagnoses, HCP5, CEA, and CA19-9[30]. Thus, the results above indicate that lncRNAs possess the diagnostic potential for early diagnosis of GC.

CircRNAs are also generally categorized as a ncRNA and physically vigorous nucleic acid molecules that construct closed loop RNA and require polyadenylated tails unlike messenger RNAs[31]. CircRNAs are more stable in tissues, have a superior life period

and are more resistant to Ribonuclease R compared with linear RNAs[32]. In a recent, the rapid development of high-throughput sequencing has promoted the detection of circRNAs and their roles in physiological and pathological processes, including GC carcinogenesis, have been evaluated. A recent study found circRNA Has_circ_0000745 was downregulated in GC tissues and in plasma from patients with GC. Therefore, the use of these circRNAs has great potential as a new biomarker for diagnosis of GC[33]. The expression of hsa_circ_0001789 in 24 paired of cancerous and paracancerous plasma specimens were examined. A down-regulated expression of hsa_circ_0001789 was detected in GC plasma samples in comparison to the healthy controls, indicating that hsa_circ_0001789 has the potential to be a novel GC diagnostic biomarker[34]. CircPTPN22 is an outstanding diagnostic marker for GC. Utilizing receiver operator characteristic (ROC) analysis of circPTPN22 expression in GC patients, gastritis patients, and healthy controls, showed that its AUC was higher than that of traditional tumor markers CEA and CA19-9, and the AUC reached 0.892 after combined use[35]. 8-circRNA biomarker panel (hsa_circ_0001013, hsa_circ_0052001, hsa_circ_0034398, circ_0006089, hsa_circ_0002019, hsa_circ_0007380, hsa_circ_0008768, hsa_circ_0045602) discriminated patients with GC from normal control tissues, with an AUC value of 0.87 (95%CI: 0.82-0.93, sensitivity 78.3%, specificity 78.3%). Interestingly, 7 of 8 circRNA markers significantly lower expressed in serum specimens after surgery. Furthermore, 8-circRNA panel showed a greater specificity for the identification of patients with GC *vs* all other gastrointestinal cancers, such as colorectal cancer, pancreatic cancer, esophageal cancer, and hepatocellular carcinoma[36]. Studies using ncRNAs in screening/diagnostics of GC in the past 5 years are collected in Table 1.

Exosome

Exosomes are lipid bilayer-enclosed extracellular vehicles (EVs) approximately 30-150 nm in size, consist of a lipid bilayer interspersed with various membrane proteins, and which contain a variety of nucleic acids, proteins, and lipids[37]. Exosomes are secreted continuously by all cells and offer distinct prospects for clinically relevant diagnostic

content, including DNA, RNA, and proteins[37]. Exosomes have been said to be associated with driving key features of cancer activity, including facilitation of tumor cell proliferation, suppression of the immune response, stimulation of tumor cell migration, initiation of angiogenesis, and formation of metastases[38]. Recently, growing evidences have established the latent diagnostic value of EVs or circulating exosomes for early diagnosis of GC. Due to the easiness to isolate exosomes compared with CTCs and cfDNA in malignant tumors, an growing number of studies will be highlighted on exosomes in the diagnosis of EGC in the next several decades. Exosomal lncUEGC1 exhibited an AUC of 0.8760 and 0.8406 in distinguishing EGC patients from normal subjects and those with atrophic gastritis, respectively, implying that exosomal lncUEGC1 had higher diagnostic reliability compared with CEA[39]. Circulating exosomal lncRNA proprotein convertase subtilisin/kexin type 2-2:1, G protein subunit alpha q-6:1 and GC-associated lncRNA1 (lncRNA-GC1) have been reported as better biomarkers for distinguishing GC patients from healthy people, compared with CEA, CA19-9, and CA72-4[40-42]. Similarly, serum four exosoms, miR-92b-3p, miR-146b-5p, miR-9-5p, and miR-let-7g-5p have been shown to be latent biomarkers for early detection of GC[43]. Likewise, the ROC showed that serum exosomal miR-590-5p showed a good classifier with an AUC of 0.810 (95%CI: 0.751-0.860) with a sensitivity of 63.7% and specificity of 86.0%, suggesting that serum exosomal miR-590-5p expression may be a noninvasive diagnostic biomarker of GC[44]. MiR-195-5p and miR-211-5p in GC patients' plasma exosomes have shown to be upregulated compared to healthy controls. The AUC values of plasma exosomal 2-miRNAs signature was 0.820, which was more meaningful to discriminate the GC patients from the healthy samples, compared to the AUC values obtained for combined CEA, CA19-9[45]. circ_0065149 levels in plasma exosomes from EGC patients were significantly lower than those from healthy individuals. Thus, circ_0065149 revealed distinguished diagnostic accuracy, with an AUC of 0.64, sensitivity of 48.7%, and specificity of 90.2%[46]. Piwi-interacting RNAs (piRNAs) are another class of small ncRNAs that are 26-31 nucleotides in length, and quite comparable in size to miRNAs. PiRNAs are distinctly different from miRNA

counterparts because they lack the sequence conservation present in miRNAs. As a source for liquid biopsy, piRNAs have similar properties to miRNAs and stabile in the blood flow and resistant to ribonuclease mediated degradation. Therefore, piRNAs can be used as effective biomarkers for the early detection of GC. Serum exosomal piR-019308, piR-004918 and piR-018569 can be utilized for the diagnosis of GC with the AUC value of 0.820, 0.754 and 0.732, respectively. Diagnostic value of these piRNAs were much better than that of CEA, CA19-9 and alpha-fetoprotein[47]. Collectively, circulating exosome has clinical potentials in the early diagnosis of GC, which may lead to the identification of cancer subtypes. Exosome investigated as potential diagnostic GC markers are listed in Table 2.

Collectively, this section highlights the developments in the ncRNAs as a source for the development of non-invasive diagnostic biomarkers for GC patients. NcRNAs appear to have a crucial potential to be implemented into biomarker development for early detection of GC. On the other hand, a variety of circulating biomarkers have been reported in different research team. The clinical value of all these factors is not still defined. Some discrepancies among different studies can be caused by the heterogeneity of GC. Thus, a more precise patient inclusion criteria or enlarging of the cohort size could overcome the differences in ncRNAs expression caused by the intrinsic heterogeneity of the disease. More multicenter, greater, longer-term studies are warranted to apply liquid biopsies in clinical settings, in terms of early diagnosis of GC. We systematically summarized the ncRNAs as biomarker for early detection of GC in Figure 2.

DNA METHYLATION

DNA methylation is an epigenetic mechanism which lead to oncogenesis in GC through activation of oncogenes and silencing of tumor-suppressor genes[48]. In a recent, a variety of studies demonstrated that aberrant epigenetic regulations, such as DNA methylation, histone posttranslational modifications, ncRNAs and chromatin remodeling, have a pivotal influence on the pathogenesis of GC. DNA methylation is

well-investigated epigenetic change among the epigenetic modifications[49]. Nowadays, the importance of aberrant DNA methylation is identified by liquid biopsy, including plasma/serum and gastric juice. Tumor suppressor genes methylation in peripheral blood has been examined most broadly. Aberrant gene methylation identified in body fluids could be a useful biomarker for the early detection of GC, likewise SEPT9 in CRC. Numerous methylated gene promoters detected in ctDNA have been proposed for GC detection as a latent diagnostic marker[50]. Among them, aberrant methylation of Reprimo, RUNX3, RASSF1A, SFRP2, PCDH10, H19, and MALAT1 show high sensitivity and specificity, suggesting that they have superiority in the early detection of GC[51-55]. A recent study elucidated the association of aberrant DNA methylation with the risk of GC. Hypermethylation of SOCS3 promoter in blood extensively facilitated GC risk[56]. Genome-wide hypomethylation commonly occurs in repeated components that are normally hypermethylated to keep genomic stability. Long interspersed nucleotide element 1, Alu repetitive elements and human endogenous retroviruses are the main components of scattered repeated sequences[57]. Although a number of studies examined the perspective of DNA methylation as a biomarker in GC, the methylation levels in those studies were mainly shown from the tissues. Thus, novel aberrantly methylated genes whose roles in GC should be investigated, especially those detected by non-invasive methods.

DNA methylation is major epigenetic changes when pathogens invade to gastric mucosa, which allows the infection to persist and promotes the development of GC[58]. *Helicobacter pylori* (*H. pylori*) is one of the most important pathogens for GC. Previous studies have demonstrated that Alu and Satα hypomethylation was significantly correlated with *H. pylori* infection in patients with GC[59]. Although numerous studies were examined to elucidate the mechanism of epigenetic changes induced by *H. pylori* infection, useful biomarker for early detection in GC is still not applied in clinical setting. Most recent report demonstrated that FGFR4 activation was interacted with signal transducer and activator of transcription 3 thorough steroid receptor coactivator

in response to *H. pylori* infection, suggesting that FGFR4 could be a candidate of biomarker for early detection of GC[60].

NOVEL BIOMAKER IDENTIFIED FROM MULTI-OMICS DATA

Previous classifications of GC have quite restricted achievement in advocating the advancement of subtype-specific treatment applies due to the heterogeneity of GC and their incapacity to detect latent molecular targets. With NGS, omics technologies have offered beneficial tools to examine the molecular basis of GC. Multiple omics-based studies have been used to evaluate the biofluids of GC patients, which resulted in numerous novel biomarkers due to rapid progress in machine learning approaches[61]. A number of mutations, gene alteration, protein abundance differences, epigenetic mutations, and metabolite concentrations which is linked to GC heterogeneity and staging, leading to improve our understanding of GC, have been successfully identified. NGS has high throughput and can screen unknown variants. A variety of methods are applying NGS to target panel, namely Tagged-Amplicon deep sequencing, Safe-Sequencing System, and Cancer Personalized Profiling by deep sequencing. The developments of NGS enable many applications in both DNA and RNA sequencing (RNA-seq), such as whole-genome, whole-exome, and targeted sequencing of DNA, and total RNA, mRNA, and small RNA[62].

Proteins contribute to determine the identity of a cell[63]. Cell function can be affected by aberrant protein expression, or altered post-translational arrangements. Proteomic signatures can afford complementary information for patient stratification to genetic signature. Besides, proteins can be a useful tool as a biomarker for malignances[63]. The development of proteomics technologies, such as mass-spectrometry (MS), has made it easier to identify protein biomarkers applied to tumor diagnosis. Liquid chromatography-MS (LC-MS) of digested proteins conducted with high-resolution instruments enable to quantitate thousands of proteins from complex biological specimens. Study using LC-MS/MS combined with TMT labeling, found that 11 proteins differentially expressed in plasma from EGC patients and healthy controls

identified from the total of 2040 proteins. This model presented that the altered proteins found by plasma proteomics could provide a resource of potential biomarkers for diagnosis of EGC[64]. Isobaric tags for relative and absolute quantification (iTRAQ) are said to be the most widely used method for high-throughput protein quantification. The study using iTRAQ-labeled quantitative proteomics revealed that NAD(P)-dependent steroid dehydrogenase-like (NSDHL) and neutral cholesterol ester hydrolase 1 (NCEH1), key enzymes in cholesterol metabolism, were aberrantly expressed in patients with GC. Moreover, the combined analysis of NSDHL and NCEH1 showed high sensitivity compared with a single marker in detecting GC, indicating that NSDHL and NCEH1 may be adequate for novel biomarkers in the screening[65]. Identically, thioredoxin reductase 1 was aberrantly expressed in GC by using Itraq-labelling and/or LC-MS, supporting that these proteins are an important target for early determining and follow-up monitoring for GC[66]. Integration of phosphoproteome data with other types of omics data from young diffuse-type GC patients provided links of somatic mutations to phosphorylation changes in signaling pathways, association of mRNA-protein abundance correlations with survivals, and four subtypes of young diffuse-type GC patients characterized by genomic and proteomic signatures, which may contribute to have a deep insight into diffuse-type GCs[67]. More recent study using proteomics methods of serum based on high-resolution MS with ion mobility separation followed by multivariate statistics and network analysis presented that a 29-protein marker panel containing, integrin β6 and glutathione peroxidase, and ten serum markers specific for GC (ITIH1, FZD6, DPP10, SERPINA4, AKAP12, S100A9, POTEF, CACNB1, CRP KIAA1328), which independent of *H. pylori* infection[68]. Multi-omic analysis characterizing the intricate relationships between the intragastric microbiome and gastric mucosal gene expression revealed higher abundances in several bacterial taxa (*Pasteurellaceae*, *Enterococcaceae*, *Helicobacteraceae*, *Gemellales* and *Neisseriaceae*) and genes (*SOCS3* and *ITM2A*) were shown in GC patients. In turn, *Lachnospiraceae* was closely associated with the expression of *Ubiquitin D*, which regulates mitosis and cell cycle time in GC[69]. Besides, lower abundances of B cell signatures in GC compared to

gastritis may suggest a previously unidentified immune evasion process in gastric carcinogenesis. These results suggest that a combined analysis of microbiome and gene expression may help to devise a diagnostic kit that screens for high-risk patients of GC who require surveillance endoscopy.

STOMACH SPECIFIC BIOMARKER

The measurement of a biomarker in a gastric juice sample should be specific for early-stage GC than the same biomarker measured in plasma/serum because it is obtained directly from the area of the lesion, which would lead to avoid dilution of the biomarkers and their lack of specificity[70]. Because DNA is easily degraded due to gastric juice acidity, the main part of biomarkers identified in gastric juice are ncRNAs and some proteins. A significantly increased miR-135b-3p and miR-199a-3p, and a decreased miR-451a levels in gastric juice-derived EVs from GC patients was shown compared to non-cancer individuals[71]. LncRNAs in gastric juice and urine have also been reported as biomarkers of GC and urinary system cancer, respectively[72]. On the other hand, a recent study revealed that the levels of SNCG in gastric juice detected by enzyme-linked immunosorbent assay were significantly higher in the GC group compared with the control group[73]. Analysis using gastric juice-derived exosomal DNA samples presented that BARHL2 methylation had an area under the curve of 0.923 with 90% sensitivity and 100% specificity which suggested the feasibility of the application of BARHL2 methylation for screening for GC biomarker[74]. Although a number of studies elucidating the value of gastric juice as a biomarker for early detection of GC were carried out, the limitation is the tolerance of the patients as a usual examination due to the discomfort of gastroscopy.

Serological examinations, which are noninvasive and relatively convenient. can offer efficient indicators of gastrointestinal diseases. Markers of gastric function such as pepsinogen (PG)-I, PG-II, PG-I/PG-II ratio, gastrin 17, anti-*H. pylori* antibodies are used to cooperatively diagnose chronic atrophic gastritis (CAG)[75]. A recent meta-analysis has presented the use of serum pepsinogens revealed significant diagnostic accuracy for

the early detection of gastric CAG or GC[76]. Among the various cut-off values investigation, PG rate ≤ 3 and PG-I ≤ 70 ng/mL have been proposed for CAG and GC prediction. Although there are numerous studies for evaluating PG as biomarker for early detection of GC, the potential use of pepsinogens in GC prevention is still not well-established.

DISCUSSION AND FUTURE PERSPECTIVE

Biomarkers are biological molecules implicated in the initiation and progression of cancer, and determination of their levels can clarify abnormal molecular changes in patients[77]. Morphological detection of diseased gastric epithelial cells using upper gastrointestinal endoscopy has always been the standard method for diagnosis of GC. However, the false-positive rates for endoscopic screening in the first round are said to be 14.9%. Notably, endoscopy is easily missed during screening, even when it is performed by qualified endoscopists. In this review, we describe the comprehensive and systematic search for and development of blood-based biomarkers for early detection of GC. Identification of highly sensitive biomarkers for GC screening might render upper gastrointestinal endoscopy unnecessary. Biomarker has assisted in the advancement of new DNA, RNA, and protein-based cancer biomarkers that can be detected from noninvasive, sensitive, specific, and cost-effective body fluids. These include the OncoBEAM™ RAS CRC liquid biopsy assay approved by the European Commission as an *in vitro* diagnostic tool for colorectal cancer therapy, and this has led to the early detection of emergent resistance to anti-epidermal growth factor receptor treatment and the assessment of genomic profile during treatment in clinical settings[78]. Such assays may also enable us to abandon approved but useless treatments for patients resistant to molecular targeted agents or immunosuppressants. For instance, with this technology the future patients with stage II or III GC with a low risk of postoperative recurrence may be simply monitored under active surveillance without adjuvant chemotherapy and possibly be started to treat from when MRD is detected (Figure 1). Although ctDNA for detection of MRD can be an emerging clinical

biomarker for disease monitoring in GC, applications of ctDNA in clinical practice require more comprehensive examination in prospective trials.

However, several technical factors are hindering the use of liquid biopsy biomarkers in clinical use. A major challenge is the very low levels of peripheral DNA/RNA[79]. In 2022, ESMO published recommendations on the use of ctDNA in genomic testing and clinical practice. Based on its review of the different aspects of ctDNA, the ESMO Precision Medicine Working Group provides some outlook on the future of ctDNA examination. Considering deficient sensitivity in clinical practice for gene fusions and copy number variations and advancing tests to discriminate accurate results for more progressive genotyping are included in their recommendations[80]. To overcome these issues, several upgraded liquid biopsy approaches for cancer diagnosis are under development. It is likely that precancerous and cancerous lesions can be detected by combinations of biomarkers with superior sensitivity and specificity, fulfilling an unmet need in clinical practice. Combining biomarkers, such as ncRNAs, can enhance the sensitivity and specificity for detecting EGC. However, it would be difficult to identify the ideal combination of other biomarkers that can enhance the overall detection performance. Correlations between different types of biomarkers like the association pattern to guide the choice of panels can be effectively identified using biostatistical approaches. For example, a serum biomarker miRNA panel consisting of 12 miRNAs was developed for risk assessment in patients with GC. With the use of a multitarget miRNA assay in more than 5000 individuals in Singapore and Korea, this panel showed an AUC of 0.848 and 87% sensitivity and 68.4% specificity values to distinguish GC patients from healthy controls. Moreover, upgrading of the AUC to 0.884 and specificity to 69.4% for GC detection were achieved by the combination of this 12-miRNA panel with patient age, *H. pylori* serology, and the plasminogen I:II ratio[81]. Combinations of ctDNA mutations with other biomarkers, such as protein or methylation have been examined toward the goal of improving the overall sensitivity. It has been demonstrated that the combination of ctDNA and protein biomarkers can markedly increase the sensitivity[82]. Genome-wide methylation and protein marker profiles have

been validated for a combination of *mSEPT9*, *mRNF180* and CA724 for the detection of GC, and the combination of all three markers detected 68.6% of GC cases[83].

Another important strategy for improving the sensitivity and specificity of GC detection is the use of innovative bioinformatic algorithms. Machine learning involves various computational techniques that are used to simplify large numbers of measurements into lower-dimensional outputs that are more explainable[84]. In the case of GC, a combination of artificial intelligence (AI)-based endoscopy and miR148a methylation has been evaluated for the diagnosis of gastric indefinite dysplasia[85]. Interestingly, a recent study using the microbial compositions of the Cancer Genome Atlas (TCGA) of 18116 human blood samples across 33 cancer types and stochastic gradient-boosting machine learning models effectively discriminated cancer-free individuals, patients with cancer, and patients with multiple types of cancer[86]. A new concept of cfDNA is also expected to be applied to cancer diagnoses in the near future.

Unfortunately, there is a lack of standardized technique for isolation, amplification and detection in the use liquid biopsies for early diagnoses in clinical settings. One of the most meaningful accomplishments regarding biomarkers for early cancer detection is the emergence of high-throughput technological platforms such as NGS, which allows comprehensive quantification of whole genome or transcriptome. The cost of NGS has greatly declined over the past several decades, which has allowed the sequencing of a larger number of clinical biospecimens, as well as the ability to generate datasets with higher depth of coverage. The opportunity to generate higher resolution data for low expressing transcripts improves the ability to detect rare cancer-derived transcripts in circulation. A study aiming to found the latent miRNA for predicting GC by AI in the Gene Expression Omnibus (GEO) datasets, particularly with several states of the machine learning methods and the Boruta algorithm, demonstrated that the use of hsa-miR-1343-3p could predict GC with the AUC of 100% (sensitivity 100%, specificity 100%, ROC 100%)[87]. In a recent study using high-throughput RNA-seq, the expression profiles of circRNAs in serum EV between GC patients and healthy individuals were examined. The results demonstrated that the circRNA, Chr10q11,

Chr1p11, and Chr7q11 were upregulated in GC; the AUC of the three combined circRNAs was 0.839 (95%CI: 0.772-0.893) with 73.2% sensitivity and 84.1% specificity. The construction of a circRNA-miRNA-mRNA network predicted that the three identified circRNAs interact with 13 miRNAs and 91 mRNAs. These findings suggest that panels of EV circRNAs may provide new insights into the early diagnosis of GC[88]. With the use of publicly available gene expression profiles from the TCGA and GEO datasets, a recent study explored the circRNA-miRNA-mRNA interaction axis by constructing a competing endogenous (ce)RNA network. On the basis of the hub RNAs included in ceRNA, the screening of the model for predicting the mRNA signature and miRNA signature by LASSO regression analysis found that both five mRNA-based signatures (CTF1, FKBP5, RNF128, GSTM2 and ADAMTS1) and five miRNA-based signatures (miR-145-5p, miR-615-3p, miR-6507-5p, miR-937-3p and miR-99a-3p) had superior capacity to predict the diagnosis of GC patients with the AUC of 0.9975[89]. Similarly, a study using regulatory networks for circRNA-miRNA-immune-related mRNA validated in the new GEO and TCGA datasets identified various ceRNA (circRNA-miRNA-immune-related mRNA) regulatory networks in GC constructed by hsa_circ_0050102–hsa-miR-4537-NRAS-Tgd cells, hsa_circ_0001013-hsa-miR-485-3p-MAP2K1-Tgd cells, hsa_circ_0003763-hsa-miR-145-5p-FGF10-StromaScore, hsa_circ_0001789-hsa-miR-1269b-MET-adipocytes, hsa_circ_0040573-hsa-miR-3686-RAC1-Tgd cells, and hsa_circ_0006089-hsa-miR-5584-3p-LYN-neurons, among which FGF10, MET, NRAS, RAC1, MAP2K1, and LYN showed promising diagnostic significance for GC patients[90]. In a study based solely on bioinformatics models constructed a circRNA-miRNA-mRNA network based on three circRNAs, 43 miRNAs, and 119 mRNAs, and upregulations of hsa_circ_0001013, hsa_circ_0007376, and hsa_circ_0043947 were observed, which may offer a new pathway for mechanical examinations and provide possible biomarkers for early detection of GC[91]. Likewise, in a bioinformatics circRNA-miRNA-gene interaction network analysis, two circRNAs (CircCEACAM5 and CircCOL1A1) cooperated with GC correlated miRNAs and their host genes were suggested as the pivotal diagnostic biomarker for GC[92]. The

emergence of single-cell sequencing technologies has markedly improved the resolution of low-input sequencing. The recent introduction of single-cell DNA/RNA-seq provides an opportunity to detect cell types and cell status. For example, a single-cell expression atlas of gastric premalignant gastric lesions and EGC identified the expression signatures of multiple cell types across different lesions, as well as a panel of six high-confidence markers that could serve as specific biomarkers for early diagnosis of GC[93]. A weighted gene co-expression network analysis (WGCNA) aims to: (1) Identify co-expressed modules in multiple biological samples based on correlations between different gene expression profiles; and (2) Explore the relationship between gene networks and phenotypes of interest, as well as the hub genes in the network[94]. Performing a WGCNA can prevent the false-negative and false-positive results of previous biological methods. WGCNA data from the GEO database revealed hub genes involved in GC and identified four functional genes (*ITGAX*, *CCL14*, *ADHFE1*, and *HOXB13*) that were differentially expressed in tumor and normal gastric samples. Another study applying a WGCNA to screen potential biomarkers for EGC using RNA-seq and clinical data from the TCGA database identified five hub genes, *MS4A1*, *THBS2*, *VCAN*, *PDGFRB*, and *KCNA3*[94]. These technological developments will allow more researchers to conduct comprehensive assessment of molecular profiles in the human circulatory system.

Systematic studies applying the molecular profiling of ncRNAs for cancer biomarker have been developed. Several investigations have used tissue profiling data to identify ncRNA as a potential biomarker, leading to the evaluation of these cancer-associated ncRNA signatures in blood cohorts. Multiple tissue sequencing datasets were recently used to identify highly overexpressed miRNAs in GC tissues. Risk probabilities using quantitative reverse transcriptase polymerase chain reaction data for three miRNAs (miR-18a, miR-181b, and miR-335) were validated and showed much better diagnostic performance than the conventional tumor markers CEA and CA19-9 currently in use. Moreover, the three-miRNA signature discriminated between patients with stage I GC and cancer-free individuals. More importantly, using the miRNA signature was

significantly more cost-effective relative to the current clinical practice of endoscopic screening, suggesting the robustness of a GC associated miRNA signature in multiple cohorts[95].

More multicenter studies and prospective evaluations in large clinical trials are necessary for the integrations of such biomarkers into GC screening platforms supporting the daily clinical treatments of GC patients. Several clinical trials have explored the use of a liquid biopsy for early diagnosis of setting for patients with GC (Table 3). However, all of the trials were either observational, case-control, or cohort trials; no randomized control trial assessing liquid biopsy have been conducted to date. Hence, incorporating liquid biopsies into clinical practice for GC may be difficult.

Regarding the development of a circulating non-invasive biomarkers for early detection of GC, specificity for stomach would provide an additional confidence and suggested robustness of GC biomarkers. Biomarkers that are highly specific to GC will reduce the false positive rates, and thus are helpful in clinical practice. The detection of dysregulated genes that are highly expressed specifically in distinct organ could make it easier to find the locations of cancer. For example, KRAS mutations are rarely identified in healthy individuals, which makes these mutations highly cancer-specific. Nevertheless, detecting KRAS mutations in the circulation does not help identify the location of cancer as KRAS mutations can occur in multiple cancer types. Using candidates for GC specific driver genes, such as CDH1 and RhoA, may be possible to increase the specificity as diagnostic biomarkers of GC[96]. Growing evidences suggest that some EVs released from specific organs retain tissue-specific molecular features. It was recently reported that tumor cell-specific molecules, *i.e.*, circulating exosome-derived lncRNA-GC1 can be used as a biomarker to detect EGC and monitor disease progression[42]. Interestingly, a clinical study (ClinicalTrials.gov identifier NCT05397548) investigated whether circulating exosomal lncRNA-GC1 could support the early-detection and progression monitoring of GC. It is thus possible to measure ncRNA content of cancer-specific exosomes to further increase cancer specificity as biomarkers for GC diagnosis. The status of acquired DNA methylation can be tissue-

specific and maintained during carcinogenesis, and examination of this status could allow the discovery of the tumor of origin[97], suggesting that biomarkers based on DNA methylation could have the additional potential to provide key information to guide follow-up clinical decision making. Multiple cancer types have been identified *via* high-depth, targeted bisulfite sequencing analysis of a large panel of methylation sites that have signals specific to certain organs; these panels include DEEPGEN™, CancerSEEK, PanSEER, cfMeDIP-se,q GRAIL, IvyGene®, and stMCEDs. Using these panels, excellent rates of detection in individuals with cancer were obtained[98]. However, the specificity of GC in individuals with early disease have been limited. A cost-effective experimental assay, called cell-free DNA Methylome Sequencing (cfMethyl-Seq) profiles the genome-wide methylation of cfDNA, offering > 12 × enrichment over whole-genome bisulfite sequencing in CpG islands[99]. By integrating four marker types, namely cancer-specific hypermethylation markers, tissue hypermethylation markers, cancer-specific hypomethylation markers types, and tissue hypomethylation markers, a cfMethyl-Seq model showed an AUR of 0.974 (95%CI: 0.926-0.998), yielding a sensitivity of 80.7% (95%CI: 68.6%-90.7%) at 97.9% specificity in multiple organs, including those with GC.

Collectively, the combination of multiple parameters and multiple analytes obtained by emerging machine learning methods will help address the challenges of EGC detection. Overcoming the limitation of cancer biomarker translation into clinical practice will require the development of high throughput, multiplexing techniques that can be easily incorporated into clinical settings as well as large-scale prospective research and interventional clinical trials with standardized specimen processing and data collection procedures.

CONCLUSION

Increased understanding of the molecular genetics and epigenetics of GC has led to the progress of molecular marker assays for GC screening. Liquid biopsy is emerging as a promising future screening technique for clinical testing owing to the use of a single sample and ease of specimen collection. The advancement of new technologies with

improved sensitivity, such as methylation sequencing and multi-omics analysis, and rapid, global detection capacities that allowed the identification of novel potential biomarkers for early detection of GC.

81663_Auto_Edited-check.docx

ORIGINALITY REPORT

28%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	391 words —	5%
2	www.nature.com Internet	241 words —	3%
3	molecular-cancer.biomedcentral.com Internet	124 words —	2%
4	www.frontiersin.org Internet	98 words —	1%
5	www.cancercommun.com Internet	73 words —	1%
6	Cuiliu Liu, Xiaoqiang Xiang, Shuangqing Han, Hannah Ying Lim et al. "Blood-based liquid biopsy: Insights into early detection and clinical management of lung cancer", Cancer Letters, 2022 Crossref	70 words —	1%
7	Aneta L. Zygulska, Piotr Pierzchalski. "Novel Diagnostic Biomarkers in Colorectal Cancer", International Journal of Molecular Sciences, 2022 Crossref	65 words —	1%
8	link.springer.com Internet	63 words —	1%

9   Yunqing Zeng, Huimin Rong, Jianwei Xu, Ruyue Cao, Shuhua Li, Yanjing Gao, Baoquan Cheng, Tao Zhou. "DNA Methylation: An Important Biomarker and Therapeutic Target for Gastric Cancer", Frontiers in Genetics, 2022
Crossref
59 words — 1%

10 Hongpeng Lu, Zhixin Zhang, Xin Yuan, Haojun Song, Peifei Li. "The role of circular RNA hsa_circ_0001789 as a diagnostic biomarker in gastric carcinoma", Scandinavian Journal of Gastroenterology, 2022
Crossref
56 words — 1%

11 mdpi-res.com
Internet
54 words — 1%

12 Ke Xiao, Shirong Li, Juan Ding, Zhen Wang, Ding Wang, Xiangting Cao, Yi Zhang, Zhaogang Dong. "Expression and clinical value of circRNAs in serum extracellular vesicles for gastric cancer", Frontiers in Oncology, 2022
Crossref
46 words — 1%

13 Xiang-Tian Yu, Ming Chen, Jingyi Guo, Jing Zhang, Tao Zeng. "Noninvasive detection and interpretation of gastrointestinal diseases by collaborative serum metabolite and magnetically controlled capsule endoscopy", Computational and Structural Biotechnology Journal, 2022
Crossref
41 words — 1%

14 www.mdpi.com
Internet
41 words — 1%

15 Barbara Cardinali, Roberta Tasso, Patrizia Piccioli, Maria Chiara Ciferri, Rodolfo Quarto, Lucia Del Mastro. "Circulating miRNAs in Breast Cancer Diagnosis and Prognosis", Cancers, 2022
Crossref
38 words — 1%

16 doaj.org
Internet

38 words — 1%

17 scholarworks.bwise.kr
Internet

38 words — 1%

18 cancerci.biomedcentral.com
Internet

35 words — < 1%

19 Shahid Aziz, Faisal Rasheed, Rabaab Zahra, Simone König. "Gastric Cancer Pre-Stage Detection and Early Diagnosis of Gastritis Using Serum Protein Signatures", Molecules, 2022
Crossref

34 words — < 1%

20 www.cell.com
Internet

34 words — < 1%

21 Rebecca C. Fitzgerald, Antonis C. Antoniou, Ljiljana Fruk, Nitzan Rosenfeld. "The future of early cancer detection", Nature Medicine, 2022
Crossref

33 words — < 1%

22 Jian Yang, Yuhua Gong, Vincent K. Lam, Yan Shi et al. "Deep sequencing of circulating tumor DNA detects molecular residual disease and predicts recurrence in gastric cancer", Cell Death & Disease, 2020
Crossref

30 words — < 1%

23 Diana S.C. Han, Y.M. Dennis Lo. "The Nexus of cfDNA and Nuclease Biology", Trends in Genetics, 2021
Crossref

29 words — < 1%

24 Wenyu Chen, Haijiao Yan, Xiaodong Li, Kele Ge, Jun Wu. "Circulating tumor DNA detection and its

28 words — < 1%

application status in gastric cancer: a narrative review", Translational Cancer Research, 2021
Crossref

25   Zohreh Rezaei, Javad Ranjbaran, Hossein Safarpour, Samira Nomiri et al. "Identification of early diagnostic biomarkers via WGCNA in gastric cancer", Biomedicine & Pharmacotherapy, 2022
Crossref

25 words — < 1%

26   Ruoyue Tan, Guanghui Zhang, Ruochen Liu, Jianbing Hou, Zhen Dong, Chaowei Deng, Sicheng Wan, Xiaodong Lai, Hongjuan Cui. "Identification of Early Diagnostic and Prognostic Biomarkers via WGCNA in Stomach Adenocarcinoma", Frontiers in Oncology, 2021
Crossref

23 words — < 1%

27   pubmed.ncbi.nlm.nih.gov
Internet

23 words — < 1%

28   www.dovepress.com
Internet

23 words — < 1%

29   www.researchgate.net
Internet

23 words — < 1%

30   Gleb O. Skryabin, Svetlana V. Vinokurova, Sergey A. Galetsky, Danila S. Elkin et al. "Isolation and Characterization of Extracellular Vesicles from Gastric Juice", Cancers, 2022
Crossref

21 words — < 1%

31   iv.iiarjournals.org
Internet

21 words — < 1%

32   Bin Zhou, Zhe Zhou, Yuling Chen, Haiteng Deng, Yunlong Cai, Xiaolong Rao, Yuxin Yin, Long Rong.

20 words — < 1%

"Plasma proteomics-based identification of novel biomarkers in early gastric cancer", Clinical Biochemistry, 2020
Crossref

33   veri.larvol.com
      Internet
                                                    20 words — < 1%

34   Geoffrey Yuet Mun Wong, Connie Diakos, Thomas J. Hugh, Mark P. Molloy. "Proteomic Profiling and Biomarker Discovery in Colorectal Liver Metastases", International Journal of Molecular Sciences, 2022
      Crossref
                                                    18 words — < 1%

35   pdffox.com
      Internet
                                                    18 words — < 1%

36   Srinivasulu Yerukala Sathipati, Ming-Ju Tsai, Tonia Carter, Patrick Allaire, Sanjay K. Shukla, Afshin Beheshti, Shinn-Ying Ho. "Survival estimation in patients with stomach and esophageal carcinoma using miRNA expression profiles", Computational and Structural Biotechnology Journal, 2022
      Crossref
                                                    17 words — < 1%

37   jamanetwork.com
      Internet
                                                    16 words — < 1%

38   Jianbiao Xu, Jianlin Song, Tongmin Wang, Wenchuan Zhu, Liangyu Zuo, Jinzhi Wu, Jianhui Guo, Xiaochun Yang. "A combination of methylation and protein markers is capable of detecting gastric cancerdetection by combined markers", Epigenomics, 2021
      Crossref
                                                    14 words — < 1%

39   Tiantian Jiang, Lin Mei, Xiao Yang, Tingkai Sun, Zhidong Wang, Yuanyuan Ji. "Biomarkers of gastric cancer: current advancement", Heliyon, 2022
      Crossref
                                                    14 words — < 1%