EDITORIAL

5625 Serrated lesions: A challenging enemy

Trovato A, Turshudzhyan A, Tadros M

REVIEW

5630 Liver disorders in COVID-19, nutritional approaches and the use of phytochemicals


5666 Recent advances in blood-based and artificial intelligence-enhanced approaches for gastrointestinal cancer diagnosis

Li LS, Guo XY, Sun K

5682 Liver disease and COVID-19: The link with oxidative stress, antioxidants and nutrition

Ristic-Medic D, Petrovic S, Arsic A, Vacic V

MINIREVIEWS

5700 Updates in diagnosis and management of pancreatic cysts

Lee LS

5715 Artificial intelligence for hepatitis evaluation


5727 Machine perfusion of the liver: Putting the puzzle pieces together

Boteon YL, Martins PN, Muiesan P, Schlegel A

ORIGINAL ARTICLE

Retrospective Cohort Study

5737 MTNR1B polymorphisms with CDKN2A and MGMT methylation status are associated with poor prognosis of colorectal cancer in Taiwan

Lee CC, Kuo YC, Hu JM, Chang PK, Sun CA, Yang T, Li CW, Chen CY, Lin FH, Hsu CH, Chou YC

Retrospective Study

5753 Validation of conventional non-invasive fibrosis scoring systems in patients with metabolic associated fatty liver disease

Wu YL, Kumar R, Wang MF, Singh M, Huang JF, Zhu YY, Lin S

Observational Study

5764 Secular decreasing trends in gastric cancer incidence in Taiwan: A population-based cancer registry study

Lin YT, Chiang CJ, Yang YW, Huang SP, You SL
META-ANALYSIS

5775 Dietary intake in patients with chronic pancreatitis: A systematic review and meta-analysis

LETTER TO THE EDITOR

5793 Relationship between nonalcoholic fatty liver disease and chronic kidney disease could start in childhood
ABOUT COVER

Editorial Board Member of World Journal of Gastroenterology, Konstantinos Triantafyllou, MD, PhD, FEBGH, Professor, Hepatogastroenterology Unit, 2nd Department of Propaedeutic Internal Medicine, Medical School, National and Kapodistrian University of Athens, Attikon University General Hospital, 1, Rimini Street, Athens 12462, Greece. ktriant@med.uoa.gr

AIMS AND SCOPE

The primary aim of World Journal of Gastroenterology (WJG, World J Gastroenterol) is to provide scholars and readers from various fields of gastroenterology and hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online. WJG mainly publishes articles reporting research results and findings obtained in the field of gastroenterology and hepatology and covering a wide range of topics including gastroenterology, hepatology, gastrointestinal endoscopy, gastrointestinal surgery, gastrointestinal oncology, and pediatric gastroenterology.

INDEXING/ABSTRACTING

The WJG is now indexed in Current Contents®/Clinical Medicine, Science Citation Index Expanded (also known as SciSearch®), Journal Citation Reports®, Index Medicus, MEDLINE, PubMed, PubMed Central, and Scopus. The 2021 edition of Journal Citation Report® cites the 2020 impact factor (IF) for WJG as 5.742; Journal Citation Indicator: 0.79; IF without journal self cites: 5.590; 5-year IF: 5.044; Ranking: 28 among 92 journals in gastroenterology and hepatology; and Quartile category: Q2. The WJG’s CiteScore for 2020 is 6.9 and Scopus CiteScore rank 2020: Gastroenterology is 19/136.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: Ji-Hong Liu; Production Department Director: Yu-Jie Ma; Editorial Office Director: Ze-Mao Gong.

NAME OF JOURNAL
World Journal of Gastroenterology

ISSN
ISSN 1007-9327 (print) ISSN 2219-2840 (online)

LAUNCH DATE
October 1, 1995

FREQUENCY
Weekly

EDITORS-IN-CHIEF
Andrzej S Tarnawski, Subrata Ghosh

EDITORIAL BOARD MEMBERS
http://www.wjgnet.com/1007-9327/editorialboard.htm

PUBLICATION DATE
September 14, 2021

COPYRIGHT
© 2021 Baishideng Publishing Group Inc
Recent advances in blood-based and artificial intelligence-enhanced approaches for gastrointestinal cancer diagnosis

Li-Shi Li, Xiang-Yu Guo, Kun Sun

Abstract

Gastrointestinal (GI) cancers are among the most common cancer types and leading causes of cancer-related deaths worldwide. There is a tremendous clinical need for effective early diagnosis for better healthcare of GI cancer patients. In this article, we provide a short overview of the recent advances in GI cancer diagnosis. In the first part, we discuss the applications of blood-based biomarkers, such as plasma circulating cell-free DNA, circulating tumor cells, extracellular vesicles, and circulating cell-free RNA, for cancer liquid biopsies. In the second part, we review the current trends of artificial intelligence (AI) for pathology image and tissue biopsy analysis for GI cancer, as well as deep learning-based approaches for purity assessment of tissue biopsies. We further provide our opinions on the future directions in blood-based and AI-enhanced approaches for GI cancer diagnosis, and we think that these fields will have more intensive integrations with clinical needs in the near future.

Key Words: Liquid biopsy; Plasma cell-free DNA; Circulating tumor cell; Extracellular vesicle; Exosome; Artificial intelligence

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Recent studies have discovered a variety of blood-based biomarkers with great potential in improving the diagnosis and surveillance of gastrointestinal (GI) cancers. In this article, we review the latest advances in the diagnosis of various GI cancers.
INTRODUCTION

Gastrointestinal (GI) cancers, which include tumors from the colon, rectum, stomach, pancreas, esophagus, anus, gallbladder, liver, and bile duct, are among the most common cancer types and leading causes of cancer-related deaths worldwide[1]. Colorectal cancer (CRC), in particular, is the third most common cancer in men (9%) and women (8%) in the United States[1-3]. Most patients with GI cancers are in the advanced stage upon diagnosis and have relatively poor prognosis outcome. For example, the overall 5-year survival rate for patients with hepatocellular carcinoma (HCC) is around 20%[3], while the survival rate for early stage patients is as high as 70%[1,3]. This is because various efficient treatment strategies, such as surgical resection and organ transplantation, are feasible for early-stage patients only. Hence, early diagnosis is of high clinical significance for better healthcare of cancer patients.

Pathological analysis of the tumor tissue is currently the “gold standard” for clinical diagnosis of cancers[4,5]. In particular, image-based computed tomography (CT) colonography, optical colonoscopy, and endoscopic biopsy are mostly widely used in the diagnosis of GI cancers. Recent studies have demonstrated that blood-based approaches, also known as liquid biopsy, offer substantial advantages over conventional tissue biopsy-based diagnostic methods, including minimally invasive nature, cost-efficiency, clinical convenience for real-time monitoring, and the potential to promote higher patient compliance. Major analytes in current liquid biopsy studies include circulating cell-free DNA (cfDNA), circulating tumor cells (CTCs), as well as extracellular vesicles (EVs) that transport the biomolecules from tumor cells, such as lipids, proteins, DNA fragment, and RNAs[6-9]. Moreover, advances and integration of artificial intelligence (AI) have brought revolutionary changes to the clinical and translational cancer studies[10-12], especially in cancer diagnosis. The identification and utilization of disease-specific biomarkers, and development of exquisite diagnostic algorithms are both highly valuable for early diagnosis of cancers.

In this article, we provide an overview of current approaches for GI cancer diagnosis, especially in CRC and HCC. We focus on blood-based biomarkers, as well as AI algorithms for image-based tissue biopsies. We review the identification of biomarkers and development of diagnostic methods, and provide our opinions on future directions in early diagnosis of GI cancers.

BLOOD-BASED LIQUID BIOPSY APPROACHES

With the advantages of being highly efficient, minimally invasive, and cost-effective, liquid biopsy has become an emerging technique for early diagnosis of cancers. Tumor-derived cell-free DNA in peripheral blood provide a surrogate for researchers/clinicians to investigate the genomic and epigenomic landscape as well as dynamics during treatment of malignant tumors[13,14]. Conventionally, liquid biopsy approaches cover the analysis of various circulating components, including cfDNA, cfRNA, CTCs, and tumor-associated EVs (Figure 1)[15], and significant progress has been made in these areas in the past few years (Table 1).

**Circulating cell-free DNA**

CfDNA molecules in plasma originate from various tissues in human body. In cancer...
<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Biomarker type</th>
<th>Biomarkers</th>
<th>No. of patients / controls</th>
<th>Accuracy</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC</td>
<td>cfDNA copy number</td>
<td>HER2 gene</td>
<td>60/30</td>
<td>Sn = 73.3%, Sp = 93.3%</td>
<td>[19]</td>
</tr>
<tr>
<td>GC</td>
<td>cfDNA methylation</td>
<td>BARHL2</td>
<td>128/30</td>
<td>Sn = 90%, Sp = 100%</td>
<td>[104]</td>
</tr>
<tr>
<td>GC</td>
<td>CTC</td>
<td>CTC level</td>
<td>116/31</td>
<td>Sn = 85.3%, Sp = 90.3%</td>
<td>[49]</td>
</tr>
<tr>
<td>GC</td>
<td>Exosomal IncRNA</td>
<td>IncRNA-UIGC1</td>
<td>10/5</td>
<td>AUC = 0.876</td>
<td>[105]</td>
</tr>
<tr>
<td>GC</td>
<td>Exosomal IncRNA</td>
<td>IncRNA-pcsk2-2:1</td>
<td>63/29</td>
<td>Sn = 84%, Sp = 86.5%</td>
<td>[77]</td>
</tr>
<tr>
<td>GC</td>
<td>Exosomal IncRNA</td>
<td>IncRNA-GNAQ6:1</td>
<td>43/27</td>
<td>AUC = 0.732</td>
<td>[71]</td>
</tr>
<tr>
<td>GC</td>
<td>Exosomal IncRNA</td>
<td>IncRNA-HOTTIP</td>
<td>126/120</td>
<td>AUC = 0.827</td>
<td>[106]</td>
</tr>
<tr>
<td>GC</td>
<td>Exosomal miRNA</td>
<td>miR-1246</td>
<td>117/82</td>
<td>AUC = 0.843</td>
<td>[107]</td>
</tr>
<tr>
<td>GC</td>
<td>Serum cfRNA</td>
<td>miR-10b-5p, miR-195-5p, miR-185-5p, miR-296-5p, miR-132-3p, and miR-20a-3p</td>
<td>441/233</td>
<td>AUC = 0.702</td>
<td>[68]</td>
</tr>
<tr>
<td>GC</td>
<td>Serum cfRNA</td>
<td>miR-21</td>
<td>50/50</td>
<td>Sn = 88.4%, Sp = 79.6%</td>
<td>[81]</td>
</tr>
<tr>
<td>GC</td>
<td>Serum cfRNA</td>
<td>miR-30a-5p, miR-659-5p and miR-3917</td>
<td>354</td>
<td>AUC = 0.82</td>
<td>[82]</td>
</tr>
<tr>
<td>GC</td>
<td>Serum cfRNA</td>
<td>BIALT5-AS1</td>
<td>107/87</td>
<td>AUC = 0.816</td>
<td>[108]</td>
</tr>
<tr>
<td>CRC</td>
<td>CTC mRNA</td>
<td>ECT2 gene</td>
<td>90/151</td>
<td>AUC = 0.821</td>
<td>[61]</td>
</tr>
<tr>
<td>CRC</td>
<td>Serum cfRNA</td>
<td>miR-21, miR-29a, miR-125b</td>
<td>160/77</td>
<td>AUC = 0.827</td>
<td>[96]</td>
</tr>
<tr>
<td>CRC</td>
<td>Serum cfRNA</td>
<td>miR-30a-5p</td>
<td>138/60</td>
<td>Sn = 77.5%, Sp = 78.3%</td>
<td>[109]</td>
</tr>
<tr>
<td>CRC</td>
<td>cfDNA methylation</td>
<td>SFRP2</td>
<td>62/55</td>
<td>Sn = 69.4%, Sp = 87.3%</td>
<td>[21]</td>
</tr>
<tr>
<td>CRC</td>
<td>CTC protein</td>
<td>CD133+CD54+CD44+ Protein</td>
<td>10/10</td>
<td>Sn = 88.2%, Sp = 92.4%</td>
<td>[62]</td>
</tr>
<tr>
<td>CRC</td>
<td>Exosomal miRNA</td>
<td>miR-27a, miR-130a</td>
<td>40/40</td>
<td>AUC = 0.773</td>
<td>[110]</td>
</tr>
<tr>
<td>CRC</td>
<td>Exosomal miRNA</td>
<td>miR-125a-3p</td>
<td>50/50</td>
<td>AUC = 0.685</td>
<td>[111]</td>
</tr>
<tr>
<td>CRC</td>
<td>Exosomal protein</td>
<td>CPNE3</td>
<td>92/32</td>
<td>Sn = 67.5%, Sp = 84.4%</td>
<td>[112]</td>
</tr>
<tr>
<td>PC</td>
<td>CTC protein</td>
<td>Vimentin</td>
<td>100/30</td>
<td>AUC = 0.968</td>
<td>[58]</td>
</tr>
<tr>
<td>PC</td>
<td>cfDNA methylation</td>
<td>BNC1, ADAMTS1</td>
<td>39/95</td>
<td>Sn = 94.8%, Sp = 91.6%</td>
<td>[23]</td>
</tr>
<tr>
<td>PC</td>
<td>cfDNA methylation</td>
<td>CD01</td>
<td>160</td>
<td>Sn = 95%</td>
<td>[24]</td>
</tr>
<tr>
<td>PC</td>
<td>CTC</td>
<td>CTC level</td>
<td>126</td>
<td>Sn = 100%, Sp = 88.6%</td>
<td>[113]</td>
</tr>
<tr>
<td>PC</td>
<td>Exosomal IncRNA</td>
<td>CLDN1, TIMP1, MAL2, MARCH2, ITIH2, HIST1H2BK, KRT19 and FGA</td>
<td>284/117</td>
<td>AUC = 0.931</td>
<td>[79]</td>
</tr>
<tr>
<td>PC</td>
<td>Exosomal miRNA</td>
<td>miR-21</td>
<td>27</td>
<td>AUC = 0.9</td>
<td>[76]</td>
</tr>
<tr>
<td>PC</td>
<td>Exosomal miRNA</td>
<td>miR-155</td>
<td>27</td>
<td>AUC = 0.89</td>
<td>[76]</td>
</tr>
<tr>
<td>PC</td>
<td>Exosomal miRNA</td>
<td>miR-451a</td>
<td>56</td>
<td>Sn = 69.2%, Sp = 70.8%</td>
<td>[114]</td>
</tr>
<tr>
<td>PC</td>
<td>Exosomal miRNA</td>
<td>miR-133a</td>
<td>110/64</td>
<td>Sn = 90.6%, Sp = 87.2%, AUC = 0.893</td>
<td>[115]</td>
</tr>
<tr>
<td>HCC</td>
<td>cfDNA methylation</td>
<td>cfDNA 5hmC</td>
<td>1204/958</td>
<td>AUC = 0.846</td>
<td>[30]</td>
</tr>
<tr>
<td>HCC</td>
<td>Exosomal miRNA</td>
<td>miR-10b-5p</td>
<td>90/28</td>
<td>AUC = 0.934</td>
<td>[75]</td>
</tr>
<tr>
<td>HCC</td>
<td>Exosomal miRNA</td>
<td>hnRNPH1</td>
<td>88/68</td>
<td>Sn = 85.2%, Sp = 76.5%</td>
<td>[80]</td>
</tr>
<tr>
<td>HCC</td>
<td>Plasma cfRNA</td>
<td>hsa_circ_000976, hsa_circ_0007750 and hsa_circ_0139897</td>
<td>158/53</td>
<td>AUC = 0.863</td>
<td>[85]</td>
</tr>
<tr>
<td>HCC</td>
<td>Serum cfRNA</td>
<td>miR-143</td>
<td>131/122</td>
<td>Sn = 80.3%, Sp = 82.4%</td>
<td>[87]</td>
</tr>
<tr>
<td>HCC</td>
<td>Serum cfRNA</td>
<td>miR-132, miR-212</td>
<td>80/42</td>
<td>Sn = 93.75%, Sp = 63.44%</td>
<td>[116]</td>
</tr>
</tbody>
</table>

GC: Gastric cancer; CRC: Colorectal cancer; PC: Pancreatic cancer; HCC: Hepatocellular carcinoma; Sn: Sensitivity; Sp: Specificity; AUC: Area under the
patients, the dying tumor cells release their DNA, which contains the tumor-specific signatures, including somatic mutations, copy number aberrations (CNAs), and altered DNA methylation profiles, into the circulation (which is known as ctDNA). However, tumor-derived cfDNA usually account for a small fraction of the total cfDNA, and the concentration is even lower in early-stage cancer patients, which largely limits the accuracy of most cfDNA-based liquid biopsy assays. To this end, in recent studies, researcher have explored the application of extra-deep sequencing and discovered a broad range of tumor-related signals in cfDNA to further improve their performance in both diagnosis and prognosis monitoring.

Conventional genomic and epigenomic alterations: Somatic mutations are one of the most widely studied tumor-associated genomic alterations. In a proof-of-concept study, through extra-high sequencing of cfDNA in HCC patients, researchers are able to non-invasively resolve the mutation landscape of the tumor with high accuracy\[^16\], which would be informative in guiding the therapeutic strategies for cancer patients. The somatic mutations in cfDNA are widely utilized as biomarkers for early diagnosis and prognosis monitoring. For example, in pancreatic cancer (PC), the overall survival of patients with detectable KRAS mutation in cfDNA was significantly shorter than that of patients without such mutations\[^17\]. Notably, the somatic mutations detected in cfDNA may not solely come from the tumors. In fact, recent studies have proven that clonal hematopoiesis also introduces somatic mutations which are detectable in cfDNA. Hence, it is highly recommended to analyze the blood cells in parallel to cfDNA in somatic mutation-oriented studies.

DNA modifications, mostly 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), are abundant epigenetic regulators in the human genome. These modifications are known to affect gene expression, and suffer from gross abnormalities in almost all kinds of tumors. In cancer patients, such modifications are detectable in cfDNA and could be used to point to the tissue origin of the tumor. For example, Chan et al\[^20\] analyzed the CNA and 5mC profiles of cfDNA simultaneously using whole-
Emerging cfDNA fragmentomic features: Plasma cell-free DNA is known as naturally fragmented, small molecules. Recent studies have revealed that the tumor-derived DNA possess a remarkable difference in fragmentation characteristics compared to the background DNA (which mostly comes from the hematopoietic system)[31], and demonstrated the clinical significance of these features in the diagnosis of various GI cancers.

CFDNA molecules show a dominant size peak at 166 bp and most of them are shorter than 200 bp, and this phenomenon is believed to correlate with the nucleosome structure[32]. In cancer patients, tumor-derived cfDNA is even shorter than the background DNA[33]. Hence, it is doable to enrich the tumor-derived cfDNA through selection of shorter cfDNA molecules in principle. Indeed, Mouliere et al[34] demonstrated the enrichment of tumor-derived cfDNA in fragment sizes between 90 and 150 bp in a cohort of 344 plasma samples from various cancer patients. They also showed that using an integrated analysis of cfDNA at different size ranges, they could classify CRC patients from healthy individuals with a high specificity of 94%[34].

Besides size pattern, recent studies further showed that the relationship between the fragmentation pattern and nucleosome structure could point to the tissue origin of cfDNA[35,36]. In particular, cfDNA of different origins tends to have different preferences on fragment ends. Jiang et al[16] compared the genome-wide fragment end patterns of cfDNA in one HCC patient and one non-cancerous control; they reported that many genomic loci show significantly higher chances to serve as fragment ends (which they called “preferred-ends”) in the HCC patient, and the usage of such genomic loci could differentiate HCC patients from non-cancerous controls with a favorable accuracy. Moreover, the same team further showed that the nucleotide frequencies (or motifs) at the cfDNA end, which are believed to correlate with nuclease cutting preference, also change in various cancers[37,38]. In fact, they found a nuclease, DNASIL3, which plays key roles in the apoptotic fragmentation process and shows strong sequence preferences when cutting the DNA[39]. Interestingly, this gene is significantly down-regulated in many cancers and thus affects the end motif pattern of the tumor-derived DNA[37]. The discovery of these novel biomarkers largely extends our knowledgebase on both the biology of cfDNA and cancer biomarkers; these novel biomarkers also promise the development of less complex and cheaper assays than those based on the conventional genomic/epigenomic biomarkers. However, most current studies are still in proof-of-principle stage, thus large-scale validation studies are essential to assure the performances before clinical investigations.
Circulating tumor cells

CTCs are released into the bloodstream of patients through the blood vessels and lymphatic vessels. When the primary cancer cells acquire an increased propensity for migration, invasiveness, and resistance, these tumor cells will invade through the tissue surrounding their site of origin and keep invading until they enter the blood or the lymphatic system, and become the CTCs[40,41]. Although most of CTCs will die within hours, only a small part of them can potentially form a new cancerous lesion at a distant site and grow into a new metastasis[42,43].

There are many commercial kits for CTC isolation and enrichments. CTC isolation and enrichment are mostly based on their physical and immunological properties. The physical features of CTCs, such as the size, density, and electric charge, are utilized to enrich CTCs from blood. For example, ISET (isolation by size of epithelial tumor cells) system and a microfluidic chip developed by Haber’ lab filter CTCs from blood by the size of CTCs[44-46], CAP (centrifugal affinity plate) system isolates CTCs by the density gradient centrifugation[47], and ApoStream system enriches CTCs by dielectrophoretic field-flow fractionation[48,49]. Enriching CTCs based on the immunological properties of CTCs depends on the antigens expressed on the surface of epithelial cells, including epithelial cell adhesion molecule (EpCAM) and cytokeratin (CK). CellSearch is the first and only FDA-approved test for capturing CTCs that express EpCAM by using immunomagnetic beads[30,51]. There also are some methods that combine epithelial marker immunofluorescence staining with automated microscopy, such as NYONE[52]. At the protein level, fluorescence in situ hybridization (FISH) has been utilized for identifying chromosomal rearrangements in CTCs[53]. Immunocytochemistry (ICC) is the major technique that is used by several commercial kits for isolating CTCs, including ApoStream and OcoQuick. Flow cytometry is another protein analysis method employed for CTC isolation. The analysis of genetic information of CTCs is mainly based on PCR and next-generation sequencing (NGS), such as RNA-based digital PCR and single-cell sequencing[54,55]. Due to the low concentration and fragile property of CTCs and limitations of current methods, combination and further optimization may achieve better enrichment and detection for CTCs.

The clinical application of CTCs depends on the number of CTCs and the information about protein expression and gene mutation in CTCs. Kang et al[56] showed that for individuals with two or more CTCs per 7.5 mL of blood, 97.1% of them were GC patients in their cohort. The increased number of CTCs after treatment was also correlated with early recurrence[27]. Zhang et al[27] showed that in GC patients, the patients with five or more CTCs per 7.5 mL in postoperative blood had a significantly shorter overall survival and disease-free survival compared to those with less CTCs. The protein expression patterns in CTCs are also an important biomarker for cancer diagnosis. Around 50% GC patients with CK’CD44+ CTCs were diagnosed with distant metastases[57]. Wei et al[58] found that, in 76% of PDAC patients, vimentin was upregulated in CTCs, and a high level of vimentin in CTCs was associated with a shorter recurrence-free survival. Similarly, PDAC patients with CTCs that expressed high levels of mucin 1 (MUC-1) and cancer stem cell markers, such as ALDH, CD133, and CD44, were associated with a worse overall survival and tumor recurrence[59,60]. Chen et al[61] found up-regulated expression of ECT2 in CTCs from patients with CRC. Fang et al[62] found that CRC patients with liver metastasis showed a higher percentage of ‘CD133’CD54’ CD44+ cell subpopulations. Moreover, the upregulated genes in CTCs, such as SHH, SMO, POU5F1B, and ALCAM, were demonstrated as potential markers for therapeutic response and prognosis prediction in PC patients[63]. In HCC patients, Zhou et al[64] discovered that the presence of EpCAM mRNA+ CTCs was associated with early recurrence when combined with Treg/CD4+ T cell ratio.

Exosomes

Exosomes are membrane-bound extracellular nanovesicles (EVs) released by many types of cells, having a size ranging from 30 to 150 nm. Exosomes have been found in various bodily fluids, including plasma, urine, and malignant ascites[65]. Exosomes are involved in mechanisms of cell-cell communication in physiological and pathological tissues through transporting biomolecules, such as lipids, proteins, mRNA, microRNAs, and DNA fragments, and the intercellular transfer of signal components[66,67]. Tumor-associated exosomes are larger than those derived from healthy cells and are enriched with different types of mediators of tumorigenesis, including proteins, growth factors, immunomodulatory molecules, and nucleic acids that arise from the cytoplasm of donor cancer cells[68].
In the last decade, a significant number of studies have focused their attention on the exosome-mediated cross-talk which can promote the activation of signaling pathways and reprogram the functions of recipient cells through their cargo transfer that occurs between cancer and normal cells, especially in the tumor microenvironment[65,69-71].

Several examples of exosome biomarkers, such as proteins, microRNAs, and long non-coding RNAs (lncRNAs), have been shown to have the potential for the diagnosis and prognosis of specific types of cancer. For example, GPC-1 was identified as a diagnostic exosome marker for PC and CRC. CD9 and CD147 were found to be upregulated in exosomes derived from serum of CRC patients[72]. Previously, seven miRNAs including let-7a, miR-21, miR-23a, miR-155, miR-223, miR-1246, and miR-1229 were reported to be highly expressed in the serum exosomes of CRC patients and were significantly decreased after surgical resection[73]. MiR-103 was reported to be delivered from hepatoma cell-derived exosomes into endothelial cells, and induced metastasis[74]. In HCC patients, serum exosomal miR-10b-5p was found to be upregulated and could be used to distinguish HCC patients from healthy controls with an AUC of 0.934[75]. Nakamura et al[76] found that exosomal miR-21 and miR-155 were significantly upregulated in PDAC patients compared to chronic pancreatitis patients. In GC patients, exosomal lncRNA pcssk2-2:1 and lncRNA GNAQ-6:1 were significantly downregulated and could be served as biomarkers for early diagnosis for GC[77,78].

Yu et al[79] developed a panel of eight exosomal RNAs, which could distinguish PDAC patients from chronic pancreatitis patients with an AUC of 0.931. The exosomal mRNA hnRNPH1 was found to have a high expression level in HCC patients and could be used to distinguish HCC patients from chronic hepatitis B patients with an AUC of 0.865[80]. Furthermore, lncRNAs, such as LncRNA-ARSR, Lnc-son2et, and LncRNA-h19, were showed to be associated with tumor progression[72,81].

Circulating cell-free tumor RNAs

Circulating cell-free RNAs (cfRNAs) are also present in the blood circulation and have the potential to serve as cancer biomarkers for cancer diagnosis. CfRNAs may be released into the blood through mechanisms of cell death or exosome-mediated signaling by living cells.

Historically, the study of cfRNA has focused on a small group of known cancer-related mRNA, miRNAs, and lncRNAs[76-78,82]. For instance, miR-21 showed an increased expression level in GC patients, which could serve as a classification biomarker with an 88.4% sensitivity and 79.6% specificity[83]. Huang et al[84] designed a panel of six miRNAs that were upregulated in the serum of GC patients, which showed an AUC of 0.702 in distinguishing GC patients from healthy controls. Shimura et al[85] found that three miRNAs (miR-30a-5p, miR-659-3p, and miR-3917) were overexpressed in GC patients who had peritoneal metastasis and could be used to distinguish patients with or without peritoneal metastasis with an AUC of 0.82. Yamada et al[86] showed that the expression of three serum miRNAs, including miR-21, miR-29a, and miR-125b, could discriminate CRC patients from healthy ones with an AUC of 0.827. In HCC patients, serum miR-143 was downregulated and could distinguish HCC patients from healthy controls with an AUC of 0.831[87]. Another study showed that a panel of three plasma circRNAs were found to distinguish HCC patients from healthy ones with an AUC of 0.863[88].

AI-ENHANCED ALGORITHMS FOR IMAGE-BASED AND TISSUE-BASED PATHOLOGY ANALYSIS

The early diagnosis and prognosis of a cancer type have become a necessity in cancer research. Image-based and tissue biopsy-based pathology analyses are considered to be the “gold standard” for their accuracy and effectiveness for cancer screening.

In recent years, AI has become an emerging field with the advances of computational power and massive amounts of learning data. Over the last two decades, different machine learning (ML) algorithms, such as support vector machines (SVM), random forest (RF), and artificial neural networks (ANNs), have been widely used in the field of medicine[89-91]. More recently, convolutional neural networks (CNN), a novel algorithm, show its great potential in processing and interpreting radiological and pathological images (Figure 2)[92,93]. Previous studies demonstrated that a multilayer CNN with pre-trained GoogleNet Inception v3 architecture on skin cancer level prediction could achieve comparable results to dermatologists[94]. Similarly, another image-based deep learning model proposed by Kermany et al[10] in analyzing optical
The effective combination of artificial intelligence and big data can provide efficient and accurate clinical-decision support for physicians. CT: Computed tomography; MRI: Magnetic resonance imaging. Coherence tomography (OCT) images of the retina is on par with the ophthalmologist, with an accuracy of 96.6%, sensitivity of 97.8%, and specificity of 97.4% on the four-class classification problem. Most importantly, both studies utilized transfer learning when training the model to save the training time, which is also a widely used technique in training an ML model in medical images due to the lack of training data in most cases.\[^{95}\]

Traditionally, the most effective and direct method of screening colorectal carcinoma is endoscopy, particularly colonoscopy. Abdominal computerized tomography (CT) or abdominal magnetic resonance imaging (MRI) are employed when more precise and additional information is needed about the lesion to help assess cancer and look for any signs of spread. However, these methods are usually labor-intensive and time-consuming and require professional knowledge. In the era of big data, the number of colonoscopy examinations and radiology screening is enormously increasing.\[^{96}\] To this end, the deep learning-based radiology and digital pathology workflow could assist the radiologists and pathologists to diagnose diseases in various aspects, such as tumor classification, gland segment, tumor microenvironment analysis, and prognostic survival prediction, to ease their ever-growing workload and improve their efficiency.\[^{92}\] More recently, Luo et al.\[^{97}\] built a real-time Gastrointestinal Artificial Intelligence Diagnostic System (GRAIDS) for the diagnosis of upper GI cancer through analysis of imaging data from clinical endoscopies. The GRAIDS achieved a high diagnostic accuracy in detecting upper GI cancers and had a similar predicted sensitivity as expert endoscopies and outperformed the non-expert endoscopies. Besides, they performed a multi-center study and further validated the system from the other five primary care hospitals, thus indicating the general applicability of GRAIDS.
Apart from AI application in colonoscopy, CT, and MRI images, the digital pathology also drew much attention from researchers, in particular adopting deep learning on the whole slide images (WSIs)\cite{92,98-100}. However, the WSIs are multi-gigabyte images, present high morphological variance, and contain various types of artifacts. Such circumstances prevent CNN directly from applying to the high-resolution images. On the other hand, typically small labeled datasets in pathological images impede the generality of the deep learning model\cite{101}. Kermany et al\cite{102} developed a multiple-instance learning-based deep learning system that only used the reported annotated datasets to tackle the limited number of labeled images and the AUCs were all above 0.98 when testing on multiple cancer types. Joseph et al\cite{96} introduced a novel machine learning-based system on prediction of CRC outcome for whole digitized haematoxylin and eosin (HE) stained histopathology slides in learning heterogeneous and discriminative contents. They demonstrated the effectiveness of the method and presented a detailed analysis of its different elements which corroborate its ability to extract and learn salient, discriminative, and clinically meaningful content.

At present, pathological analysis of tissue biopsies is still the most widely approach in cancer diagnosis. Recent studies have shown that combination of molecular data and deep learning approach could predict the cancerous status and tissue origin of pan-cancer biopsies with high accuracy. For instance, Sun et al\cite{103} developed GeneCT, which showed an overall accuracy of 98.2% in differentiating tumor samples from adjacent normal samples, and an overall accuracy of 98.6% in predicting the tissue origins of the biopsies. In addition, as tissue biopsy is one of the most widely materials in cancer studies, and the purity of the biopsies is crucial for appropriate project design and correct interpretation of the data. To this end, Fan et al\cite{12} showed that the low prediction accuracy of GeneCT in some datasets suggested impurity of the samples. The performance of GeneCT was further evaluated using metastatic tumor samples, which is a common scenario in GI cancers. As a result, GeneCT shows potential as an accurate classification tool for tissue biopsy-based diagnosis, as well as an easy-to-use quality-control tool for cancer studies.

**CONCLUSION**

GI malignancies, especially CRC and HCC, are common cancers that lead to death of patients with cancers worldwide. Liquid biopsy plays an important role in early-stage GI cancer diagnosis as a noninvasive tumor detection strategy. Biomarkers including cfDNA, CTCs, and exosomes of liquid biopsy are widely used for the detection of tumor-associated molecules. The advantages of cfDNA, such as high amount, being relatively stable and easy for extraction and preservation, mostly representing the information from the major tumor clone, and the significant features including fragment sizes, the proportion of fragment, the end-motif, and genomic alterations, make it the preferential and accurate biomarker for the cancer diagnosis. CTCs are also important liquid biomarkers because they not only contain direct molecular information from the tumor clone but also can be cultured in vitro. However, CTCs are very rare and require fresh blood for isolation.

The potential of liquid biopsy in clinical practice is enormous in the future. However, there are several challenges that need to be resolved for the utility of liquid biopsy in clinical diagnosis. First, standardization should be established for the sample collection and storage conditions, and biomarker molecules and the suitable detection and analysis strategy depend on sample type. Second, tumor-associated molecule detection techniques require high sensitivity for early tumor detection and high specificity for early-stage tumor screening. At the same time, the low content of tumor biomarkers and the varied interference factors in the circulation make the high requirements for the high sensitivity and specificity of detection strategies. Reducing purification steps could shorten the time of sample treatment and reduce the cost, and optimization of the throughput platform could help to increase both the sensitivity and specificity of the biomarkers.

AI based on machine learning and deep learning is widely used in all walks of life. Recently, rapid developments have been made in the utility of AI in medicine areas and showed promising results in terms of accuracy for cancer diagnosis. However, the scale and quality of the training and validation datasets of most of the studies were relatively limited to apply this technique in clinical practice. Moreover, it also requires the external cross-validation, especially in tumor classification. Therefore, future studies with a larger number of datasets with high-quality annotations and external
cross-validation are required for routine practice-level validation.

ACKNOWLEDGEMENTS

We would like to thank Ms. Qi Wang from Shenzhen Bay Laboratory for her technical assistance.

REFERENCES


Li LS et al. Recent advances in gastrointestinal cancer diagnosis


---

**References:**

- Li LS et al. Recent advances in gastrointestinal cancer diagnosis

---

**Note:** The references are cited in the text according to the APA 6th edition style. The full list of references is not provided here, but it would include all the articles mentioned in the text.


Li LS et al. Recent advances in gastrointestinal cancer diagnosis

DOI: 10.1002/jtc.32647


Recent advances in gastrointestinal cancer diagnosis


