CHARACTERIZATION OF CIRCULATING TUMOR DNA FEATURES AND EXPLORATORY EVALUATION OF THE SPOT-MAS ASSAY FOR NON-METASTATIC COLORECTAL CANCER DETECTION

Pham Truong Giang[1],Tran Bao Long[1,2], Pham Cam Phuong[3]
Tran Ngoc Dung[2], Trieu Van Truong[4], Tran Le Son[5]
Nguyen Hoai Nghia[5] and Luong Thi Lan Anh[1,2,✉]

[1]Hanoi Medical University
[2]Hanoi Medical University Hospital
[3]Bach Mai Hospital
[4]Phenikaa University Hospital
[5]Medical Genetics Institute

Colorectal cancer (CRC) is the fifth leading cancer and causes death among all types of cancers in Vietnam. Early detection of colorectal cancer remains challenging, as existing screening methods are either invasive or have limited sensitivity. Circulating tumor DNA (ctDNA) analysis from liquid biopsy offers a promising non-invasive screening methods. Plasma samples from 50 patients with non-metastatic CRC (stages I-III) and 50 healthy controls were analyzed. Multiple cfDNA features, including targeted and genome-wide methylation, copy number alterations (CNAs), and fragment length profiles, were extracted. These cfDNA features were subsequently analyzed using the previously developed SPOT-MAS machine learning model to evaluate the performance of multi-feature ctDNA integration for colorectal cancer detection. CRC samples exhibited widespread genome-wide hypomethylation, targeted hypermethylation of cancer-associated regions (SEPT9), subchromosomal CNAs, and a higher ratio of short cfDNA fragments (≤150 bp) compared with controls. Using integrated multi-feature analysis, SPOT-MAS achieved a sensitivity of 60% and a specificity of 94% for CRC detection. Our results confirm the presence of several known cfDNA features associated with CRC in plasma sample for non-metastatic patients. The preliminary evaluation using the SPOT-MAS framework suggests that integrating multiple cfDNA characteristics may support CRC detection from early stage. However, larger cohorts are required to further validate these finding.

Keywords: Colorectal cancer, ctDNA, cancer screening.

I. INTRODUCTION

Colorectal cancer (CRC) is one of the most common cancers in developed countries, with an estimated 1.9 million new cases worldwide in 2022 and approximately 900,000 deaths each year globally.[1] As with other cancers, patients with CRC have a significantly better prognosis when the disease is diagnosed at an early stage compared with advanced stages. However, only about 40% of CRC cases are detected at an early stage.[2] Therefore, early detection plays a critical role in improving survival outcomes for these patients.

Colonoscopy remains the gold standard for CRC screening and diagnosis. Nevertheless, this procedure is invasive, time-consuming, costly, and may cause discomfort and procedure-related complications, which

Corresponding author: Luong Thi Lan Anh
Hanoi Medical University
Email: luongthilananh@hmu.edu.vn
Received: 02/02/2026
Accepted: 27/03/2026

can limit patient compliance. Non-invasive screening methods have therefore attracted considerable attention, but most of them show limited detection performance. Previously, several non-invasive screening approaches, such as fecal immunochemical tests and blood-based biochemical markers, were introduced to improve population-level screening uptake.[3,4] However, these methods are often limited by suboptimal sensitivity and specificity. These limitations may negatively affect patient adherence to screening recommendations, highlighting the need for a non-invasive screening approach with improved accuracy for early CRC detection.

Liquid biopsy has emerged as a promising advancement in cancer research by overcoming the limitations of tissue biopsy, which is invasive and difficult to repeat. Liquid biopsy enables tumor profiling through the analysis of circulating tumor DNA (ctDNA) in blood samples. ctDNA consists of short DNA fragments released into the bloodstream, mainly as a result of apoptosis and necrosis of cancer cells. As a result, ctDNA carries both genetic and epigenetic alterations specific to the tumor, making it a representative surrogate of tumor DNA for cancer detection and prognosis.

However, detecting ctDNA in blood remains challenging because ctDNA typically accounts for only about 0.1% of total circulating cell-free DNA, and its abundance is even lower in early-stage disease. Therefore, highly sensitive and advanced technologies are required. Recent advances in sequencing technologies, together with machine learning approaches, have improved the ability to detect tumor-derived signals from very low levels of ctDNA. In recent years, research in this field has shifted from focusing on tumor-specific mutations to developing genome-wide ctDNA features, including DNA methylation changes, copy number variations, and cfDNA fragmentation patterns. Several studies have described these features and demonstrated their potential to distinguish cancer patients from healthy individuals.[5,6]

However, many previous studies have investigated individual cfDNA biomarkers separately. Although these molecular signals have been reported in various cancers, their combined presence and behavior in plasma samples from patients with non-metastatic colorectal cancer remain less well characterized in exploratory cohorts using low-coverage sequencing approaches. The extent to which multiple cfDNA features can be simultaneously observed in plasma samples from patients with non-metastatic colorectal cancer remains incompletely characterized, particularly in small exploratory cohorts using low-coverage sequencing approaches. In this study, we aimed to characterize several circulating cfDNA features-including DNA methylation patterns, copy number alterations, and fragment size distributions-in patients with non-metastatic colorectal cancer compared with healthy controls. In addition, we performed an exploratory evaluation of the previously published SPOT-MAS model to assess whether these combined cfDNA signals could support the detection of colorectal cancer in our study cohort

II. MATERIAL AND METHODS

1. Study population

Inclusion criteria

- The colorectal cancer (CRC) group included 50 patients with histologically confirmed CRC, classified as stage I, II, or III without distant metastasis according to the TNM system, and who had not received any prior treatment.

Patients were recruited from Bach Mai Hospital, Hanoi Medical University Hospital, and Medical Genetics Institute.

- The control group included 50 healthy individuals with no diagnosis or history of cancer, confirmed through clinical examination.

Exclusion criteria

- Patients with two primary cancers or recurrent CRC were excluded from the CRC group.

- For the healthy control group, individuals with clinical symptoms suspicious for CRC, abnormal findings on colonoscopy, pregnant women, or those who had received a blood transfusion within the previous three months were excluded.

2. Methods

Study design

Analytical cross-sectional study

No formal sample size calculation was performed because this study was designed as an exploratory pilot investigation.

Study process

Sampling and cfDNA extraction

Peripheral blood samples (10 mL) were collected in EDTA Vacutainer tubes (Becton Dickinson), stored at 4°C, and processed by two-step centrifugation to separate plasma. cfDNA was extracted from plasma using the MagMAX Cell-Free DNA Kit (Thermo Fisher Scientific, USA) on the automated KingFisher Flex Magnetic 96DW system (Thermo Fisher Scientific, USA). Extracted cfDNA was then subjected to bisulfite conversion using the EZ DNA Methylation-Gold Kit (Zymo research, USA), ensuring high conversion efficiency even with low DNA input.

Library preparation and bisulfite sequencing

Bisulfite-treated cfDNA was ligated with adapters and prepared for sequencing using the xGen Methyl-Seq Library Prep Kit (IDT, USA). Libraries were divided into two fractions: a targeted fraction enriched for 450 cancer-related CpG-rich regions using xGEN® Lockdown Reagents, and the flow-through fraction after hybrid capture, representing genome-wide cfDNA. Both library fractions were amplified, purified, and sequenced in paired-end mode (2 × 100 bp) on the MGI DNB-G400 platform, with an average sequencing depth of approximately 20 million reads per sample. Sequencing quality was evaluated based on mapping rate, duplicate rate, and bisulfite conversion efficiency. The mapping rate was consistently above 95%, duplication rate was below 20%, and bisulfite conversion efficiency exceeded 98%, indicating adequate quality for downstream analysis Samples that did not meet predefined quality thresholds were excluded from downstream analysis.

cfDNA sequencing data analysis

Raw sequencing data were processed through demultiplexing, read pairing, adapter trimming and alignment to the bisulfite-converted human reference genome (hg19). The following cfDNA features were extracted:

- Methylation status of 450 targeted regions

- Genome-wide methylation using 1 Mb bins

- Genome-wide copy number alterations (CNAs)

- cfDNA fragment length features based on the ratio of fragments ≤150 bp and >150 bp using a 5 Mb bin

Machine learning model and statistical analysis

- A previously published machine learning framework, SPOT-MAS, was used to evaluate the ability of combined cfDNA features to discriminate colorectal cancer patients from

healthy individuals. The SPOT-MAS model integrates multiple cfDNA features, including targeted and genome-wide methylation patterns, fragment size profiles, and copy number alterations. In this study, SPOT-MAS was applied as a pre-trained model to evaluate its performance in our dataset. No model retraining was performed using the current cohort in order to avoid potential overfitting given the limited sample size.

- Statistical analyses were performed using R software. Group comparisons were conducted using the Wilcoxon or Mann-Whitney U test with multiple-testing correction. Model performance was evaluated using sensitivity and specificity.

3. Ethics Approval

This study was approved by the Institutional Review Board (Approval No. 192/HĐĐĐ-ĐHYD, February 21, 2022). All participants were informed about the study objectives and provided voluntary consent to participate. Participant confidentiality was strictly maintained, and all data were used solely for research purposes.

III. RESULTS

1. Participant characteristics

A total of 50 patients with colorectal cancer confirmed by oncology specialists and 50 healthy volunteers without a cancer diagnosis or history of cancer were enrolled in the study. In the colorectal cancer group, females accounted for 40% and males for 60% of participants. The mean age of patients with colorectal cancer was 56.2 years old, ranging from 32 to 86 years old. The two groups were not formally matched for age or sex due to the exploratory and cross-sectional nature of this study; the difference in mean age between groups (56.2 vs. 48.3 years) represents a potential confounding factor that should be addressed in future matched-cohort validation studies. Among the 50 CRC patients, 12% were classified as stage I, 44% as stage II, and 44% as stage III.

2. Differential cfDNA features

Methylation status of 450 targeted regions

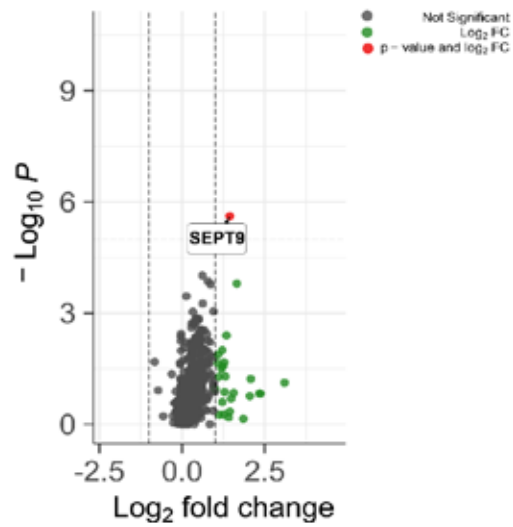

Figure 1.Target methylation density difference on 450 regions

Volcano plot (Figure 1) illustrates differential DNA methylation levels of 450 targeted regions in circulating cell-free DNA (cfDNA) between colorectal cancer (CRC) patients and healthy controls. The x-axis represents $\log_2$ fold change in methylation levels, while the y-axis represents $-\log_{10}(P)$. Differentially methylated regions are color-coded as follows: grey, not significant; green, regions meeting the fold change criterion only. Regions with $\log_2 FC \geq 1$ and p-value < 0.05 are highlighted in red. The ***SEPT9*** region shows significant hypermethylation in colorectal cancer.

Genome-wide methylation

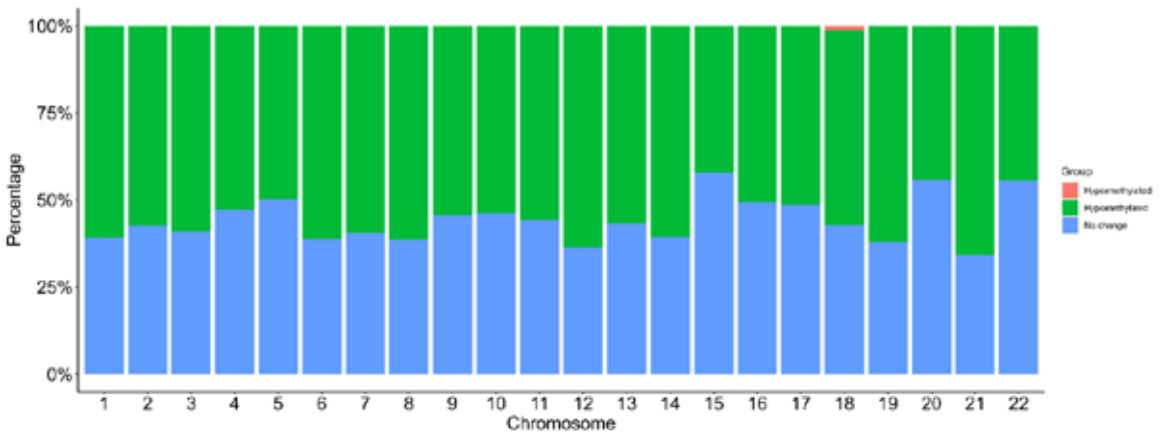

Figure 2. Methylation percentage on the whole genome compared to healthy-control samples

The chromosome-wise distribution of differentially methylated regions in circulating cell-free DNA (cfDNA) was examined across 22 autosomes (2734 bins), comparing colorectal cancer (CRC) patients with healthy controls (Figure 2). Overall, hypomethylated regions accounted for the largest proportion of differentially methylated sites across nearly all autosomes, representing approximately half or more of the bins per chromosome. In contrast, bins showing no significant methylation change constituted a substantial but variable proportion across chromosomes, while hypermethylated bins were rare and detected only at low frequencies. No single chromosome exhibited a marked enrichment of hypermethylation or hypomethylation relative to others, indicating a relatively uniform distribution of methylation alterations across the genome. These findings suggest that cfDNA methylation differences in CRC are predominantly driven by widespread hypomethylation.

Genome-wide copy number alterations (CNAs)

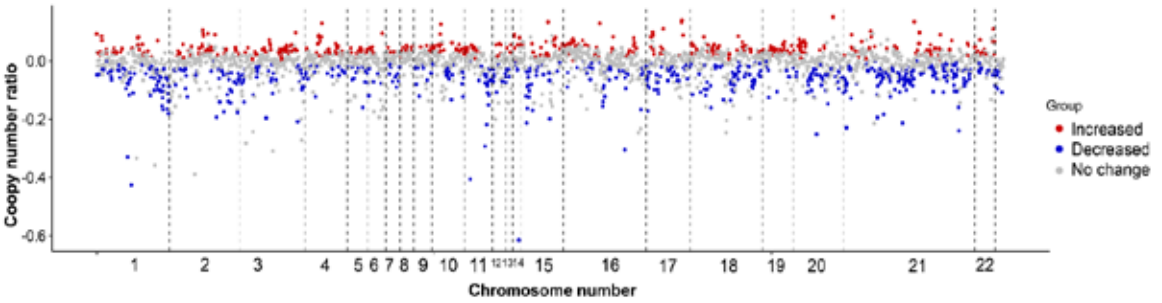

Figure 3. Copy number changes on whole genome compared to healthy-control samples

Copy number variation (CNV) analysis of circulating cell-free DNA showed that most regions across all chromosomes had copy number ratios close to zero, indicating generally small changes (Figure 3). Regions with copy number loss were more frequent and showed

larger decreases than regions with copy number gain. In contrast, copy number gains were usually mild. No whole-chromosome gains or losses were observed. These results indicate that CNV alterations in colorectal cancer cfDNA mainly occur in localized genomic regions rather than affecting entire chromosomes.

cfDNA fragment length

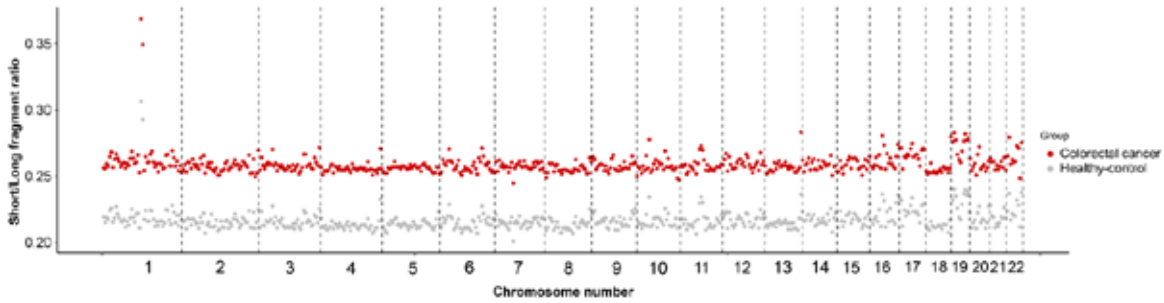

Figure 4. Ratio of short fragments to long fragments compared to healthy-control samples

The ratio of short (≤150 bp) and long (>150 bp) cfDNA fragments was higher in colorectal cancer patients than in healthy controls across all chromosomes. Cancer samples showed a stable genome-wide elevation in fragment ratios (~0.25-0.27), whereas controls remained lower (~0.21-0.22). No strong chromosome-specific effects were observed, indicating a global shift toward shorter cfDNA fragments in colorectal cancer. These findings suggest a genome-wide enrichment of short cfDNA fragments associated with colorectal cancer.

3. Diagnostic Performance of SPOT-MAS

Table 2. Performance of SPOT-MAS in colorectal cancer detection

CRC	ctDNA	
	ctDNA positive	ctDNA negative
Patients	30	20
Healthy	3	47
Sensitivity	60% (30/50)	
Specificity	94% (47/50)	

The SPOT-MAS assay was applied to plasma samples from all participants to evaluate the diagnostic performance of multimodal analysis of ctDNA profiles for CRC detection. Among 50 CRC patients, 30 true-positive and 20 false-negative cases were detected. Among 50 healthy controls, 3 false-positive and 47 true-negative cases were identified. The overall sensitivity and specificity were 60% (95% CI: 45-73%) and 94% (95% CI: 83-98%), respectively.

IV. DISCUSSION

Cell-free DNA is commonly used to study tumor-related changes in a non-invasive way. In this study, we combined cfDNA methylation, copy number variation (CNV), and fragment size analyses to examine molecular alterations associated with CRC. Methylation has been identified as one of the most promising cfDNA biomarkers for CRC detection, and methylated cfDNA markers have demonstrated strong

diagnostic performance in multiple cohorts. Hypomethylation in cfDNA reflects both the presence of tumor-derived DNA and changes in nucleosomal organization.[8] These changes are associated with reduced methylation levels and increased DNA fragmentation in cancer cfDNA. In our analysis, most bins were hypomethylated in CRC compared with healthy controls. This pattern was observed across all autosomal chromosomes and did not appear to be driven by specific chromosomes. Similar genome-wide hypomethylation has been reported in previous CRC studies using both tissue and plasma samples.[9] The widespread occurrence of hypomethylation indicates the loss of epigenetic regulation and increased genomic instability during colorectal tumor development.

Despite this global hypomethylation background, the volcano plot highlighted a small number of significantly hypermethylated regions, most notably *SEPT9*. This finding is consistent with earlier studies showing that CRC is characterized by promoter-specific hypermethylation of certain cancer-related genes.[10] *SEPT9* hypermethylation has been validated as a blood-based biomarker for CRC and is already used in clinical screening tests. The clear detection of this signal suggests that specific regional hypermethylation events can be captured even when the proportion of tumor-derived DNA in plasma is low.

CNV analysis revealed low-level copy number gains and losses affecting subchromosomal regions, with no evidence of whole-chromosome aneuploidy. This pattern is consistent with previous plasma cfDNA sequencing studies in CRC, which reported recurrent amplifications (e.g., 1q, 8q) and deletions (e.g., 18q) in relevant genomic regions.[11] These CNVs often involve genes in key pathways such as cell cycle regulation and WNT signaling. The absence of whole-chromosome CNVs may reflect the low tumor fraction in cfDNA, particularly in early-stage or non-metastatic cases. It may also be influenced by the limited sensitivity of targeted panels compared with whole-genome approaches.

Importantly, CRC samples showed a consistently higher ratio of short (≤150 bp) to long (>150 bp) cfDNA fragments across all chromosomes. Shorter cfDNA fragments have been associated with tumor-derived DNA and altered chromatin structure in cancer cells, which suggest that cfDNA fragmentation might be linked to epigenetic features.[5] Hypomethylated and open chromatin regions tend to contribute a higher proportion of short DNA fragments. Genome-wide fragmentation profiles were altered in cancer patients across cancer types, and elevated short fragment ratios have been linked to higher tumor fraction and greater sensitivity for detection when integrated with other cfDNA features. Together, the consistent patterns of genome-wide hypomethylation, subchromosomal CNVs, and increased short cfDNA fragments indicate that epigenomic and structural alterations co-occur in tumor-derived cfDNA from CRC patients. These observations are broadly consistent with previous reports describing characteristic ctDNA signals in colorectal cancer.

Table 3. Comparison of ctDNA-based cancer screening methods

Model	SPOT-MAS (This study)	GRAIL (Liu et al)	DELFI (Cristiano et al.)	CANCER SEEK
ctDNA features	Methylation (targeted & genome-wide), fragment size, CNV	Fragmentation	Methylation	Mutations
Sensitivity	60%	67,3%	70%	66%
Specificity	94%	99%	98%	99%

Several non-invasive early cancer detection models based on ctDNA biomarkers have been developed for multi-cancer screening, including colorectal cancer (CRC). Most approaches follow two main steps: detection of ctDNA signals in plasma and prediction of cancer presence or tissue of origin using machine learning. In this study, we applied the SPOT-MAS workflow for early CRC detection, which integrates multiple cfDNA features, including targeted and genome-wide methylation, cfDNA fragment size, and copy number alterations. Rather than developing a new prediction model, our analysis aimed to examine whether these cfDNA signals could be detected in our cohort and whether the previously described SPOT-MAS analytical framework could capture these patterns. Compared with other ctDNA-based screening models (Table 3), SPOT-MAS shows comparable performance while using a different analytical strategy. The GRAIL model, which relies on cfDNA methylation profiling with deep sequencing, reported a similar overall sensitivity for CRC but showed limited performance in early-stage disease, particularly stage I.[12] In contrast, SPOT-MAS combines multiple cfDNA features using low-coverage sequencing, enabling broader biological signal capture at reduced sequencing depth. The DELFI model focuses on genome-wide cfDNA fragmentation patterns and reported higher sensitivity for CRC; however, its evaluation was based on a relatively small CRC cohort, and the complexity of its analytical framework requires further large-scale validation.[13] CancerSEEK, which is based on the detection of somatic mutations, also demonstrated comparable sensitivity and high specificity. Nevertheless, mutation-based approaches face challenges in early-stage disease due to low ctDNA abundance, high sequencing cost, and potential interference from clonal hematopoiesis.

Overall, early cancer detection using ctDNA remains challenging, mainly due to the low abundance of tumor-derived DNA in plasma, particularly in early-stage disease. Many existing approaches improve sensitivity by increasing sequencing depth, which substantially raises cost and limits scalability. In contrast, the SPOT-MAS approach focuses on integrating multiple complementary cfDNA features, including methylation, copy number alterations, and fragment size patterns, using low-coverage sequencing.[7]circulating tumor DNA (ctDNA This strategy allows broader capture of tumor-associated signals without increasing sequencing depth. Such an approach may support the potential application of cfDNA-based screening strategies while maintaining a balance between sequencing cost and biological signal detection.

This study has several limitations that should

be considered. First, the sample size was small, which may limit the generalizability of the findings. We attribute this to the strict selection criteria for early-stage and non-metastatic CRC, which is when cancer detection confers significant clinical benefits. Larger cohorts are needed to further validate the performance of the proposed approach. Second, although multiple cfDNA features were integrated, the relative contribution of each feature to model performance was not systematically assessed. In addition, potential confounding factors such as age, gender, and non-cancer-related biological variation were not extensively evaluated. Finally, follow-up data were not available for individuals in the healthy control group due to limited resources. As a result, we could not determine whether any control participants developed colorectal cancer during a defined follow-up period. Future studies with longitudinal follow-up would be valuable to further assess the long-term predictive value of cfDNA-based screening approaches.

V. CONCLUSION

In this study, we characterized multiple ctDNA features in patients with non-metastatic colorectal cancer using a liquid biopsy approach, including genome-wide hypomethylation, hypermethylation of cancer-related regions, copy number alterations, and cfDNA fragmentation. These patterns were consistent with previously reported molecular characteristics of tumor-derived cfDNA. When the previously developed SPOT-MAS model was applied to our cohort, moderate sensitivity and high specificity were observed. Given the limited sample size, these results should be interpreted as preliminary. Larger cohorts will be required to further evaluate the clinical utility of multi-feature cfDNA analysis for colorectal cancer detection.

ACKNOWLEDGEMENT

We thank the staff of Bach Mai Hospital, Hanoi Medical University Hospital, and the Institute of Medical Genetics for their support in patient recruitment and sample collection. We also acknowledge the technical assistance provided by the laboratory staff during cfDNA processing and sequencing.

REFERENCES

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2024; 74(3): 229-263. doi:10.3322/caac.21834.

2. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*. 2023; 73(3): 233-254. doi:10.3322/caac.21772.

3. Hauptman N, Glavač D. Colorectal Cancer Blood-Based Biomarkers. *Gastroenterol Res Pract*. 2017; 2017: 2195361. doi:10.1155/2017/2195361.

4. Imperiale TF, Ransohoff DF, Itzkowitz SH, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med*. 2014; 370(14): 1287-1297. doi:10.1056/NEJMoa1311194.

5. Lo YMD, Han DSC, Jiang P, Chiu RWK. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science*. 2021; 372(6538): eaaw3616. doi:10.1126/science.aaw3616.

6. H L, Q Z, W W, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. *Science translational medicine*. 2020; 12(524). doi:10.1126/scitranslmed.aax7533.

7. Nguyen VTC, Nguyen TH, Doan NNT,

et al. Multimodal analysis of methylomics and fragmentomics in plasma cell-free DNA for multi-cancer early detection and localization. *eLife*. 2023; 12. doi:10.7554/eLife.89083.

8. Noë M, Mathios D, Annapragada AV, et al. DNA methylation and gene expression as determinants of genome-wide cell-free DNA fragmentation. *Nat Commun*. 2024; 15(1): 6690. doi:10.1038/s41467-024-50850-8.

9. Chan KCA, Jiang P, Chan CWM, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc Natl Acad Sci U S A*. 2013; 110(47): 18761-18768. doi:10.1073/pnas.1313995110.

10. Fang Q, Yuan Z, Hu H, Zhang W, Wang G, Wang X. Genome-wide discovery of circulating cell-free DNA methylation biomarkers for colorectal cancer detection. *Clin Epigenetics*. 2023; 15(1): 119. doi:10.1186/s13148-023-01518-5.

11. Li J, Dittmar RL, Xia S, et al. Cell-free DNA copy number variations in plasma from colorectal cancer patients. *Mol Oncol*. 2017; 11(8): 1099-1111. doi:10.1002/1878-0261.12077.

12. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol*. 2020; 31(6): 745-759. doi:10.1016/j.annonc.2020.02.011.

13. Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*. 2019; 570(7761): 7761. doi:10.1038/s41586-019-1272-6.