GMR
The Original

LIQUID BIOPSY FOR EARLY CANCER DETECTION: A SYSTEMATIC REVIEW OF CFDNA, CTDNA, METHYLATION, AND FRAGMENTOMIC BIOMARKERS

R. Someshwaran[1]*, Srishti Soni[2], Jagdish Arvindbhai Prajapati[3]

[1]Professor, Department of Microbiology, PSG Institute of Medical Sciences and Research, Coimbatore, Tamil Nadu, India. Email: drsomeshwaran@gmail.com (*Corresponding Author)
[2]Associate Professor, Department of Pathology, Autonomous State Medical College, Lalitpur, Uttar Pradesh, India.
[3]Assistant Professor, Department of Pathology, GMERS Medical College, Godhra, Gujarat, India.

ABSTRACT

Background: Cancer diagnosed before dissemination is substantially more amenable to curative treatment; however, population screening is currently available for only a limited number of malignancies. Blood-based liquid biopsy has consequently emerged as a potential strategy for detecting molecular evidence of cancer before conventional clinical presentation. Circulating cell-free DNA (cfDNA) provides several distinct layers of information, including tumor-derived mutations, DNA methylation changes, copy-number alterations, fragment-size distributions, nucleosome-associated patterns, and fragment-end signatures.

Objective: This systematic review critically evaluates cfDNA quantity and integrity, circulating tumor DNA (ctDNA) mutations, methylation signatures, and fragmentomic biomarkers for early cancer detection, with particular emphasis on recent independent-validation and multi-cancer studies.

Methods: A structured literature search was performed for studies available through January 2026 using combinations of terms relating to cell-free DNA, circulating tumor DNA, DNA methylation, fragmentomics, early cancer detection, and multi-cancer detection. A total of 1,066 records were identified through database and supplementary searching. After removal of 289 duplicate or otherwise ineligible records, 777 records underwent title and abstract screening. Of these, 635 records were excluded, and 142 reports were sought for retrieval. Five reports could not be retrieved, leaving 137 full-text reports for eligibility assessment. Following exclusion of 111 full-text reports, 26 primary studies were included in the qualitative synthesis. Human studies evaluating blood-based cfDNA biomarkers for detection of treatment-naive, localized, non-metastatic, or preclinical malignancy were prioritized. Diagnostic performance, stage-specific sensitivity, specificity, area under the receiver operating characteristic curve (AUC), cancer-signal-origin performance, study design, and validation strategy were extracted. Methodological limitations were assessed using principles of QUADAS-3. Owing to substantial clinical and analytical heterogeneity, a quantitative meta-analysis was not performed.

Results: A total of 26 primary studies were included in the final qualitative synthesis. Mutation-based ctDNA assays demonstrated high biological specificity but were limited by low circulating tumor fractions in small cancers and confounding by clonal hematopoiesis. Methylation assays provided a larger number of tumor-associated loci and generally stronger tissue-of-origin information. Recent enzyme-assisted methylation testing in the INSPECTOR study achieved 81.9% sensitivity and 99.0% specificity in independent validation, with stage I sensitivity of 65.5%. The 2024 TOTEM methylation assay showed 55.9% sensitivity and 98.6% specificity in independent validation. Fragmentomic technologies demonstrated strong discrimination using shallow whole-genome sequencing: a 2025 multidimensional fragmentomics study reported 87.4% sensitivity, 97.8% specificity, and 82.4% tissue-of-origin accuracy in an independent cohort. In a subsequent prospective cohort of asymptomatic individuals, sensitivity was lower at 53.5%, emphasizing the difference between case-control and screening performance. A pancreatic cancer fragmentomic classifier achieved an AUC of 0.987 in validation, while multimodal assays combining methylation, mutations, proteins, or fragmentomic signals improved detection compared with single-feature approaches.

Conclusion: The field has shifted from measuring total circulating DNA toward extracting multidimensional genomic and epigenomic information from cfDNA molecules. Methylation and fragmentomics currently offer the broadest signals for early-stage detection, whereas mutation analysis contributes high molecular specificity. Multimodal approaches are particularly promising. Nevertheless, prospective population-based validation, standardization of pre-analytical procedures, robust control of confounding, and demonstration of clinical benefit remain essential before widespread adoption.

KEYWORDS: cell-free DNA; circulating tumor DNA; liquid biopsy; methylation; fragmentomics; cancer screening; multi-cancer early detection; ctDNA; precision oncology.

1. INTRODUCTION

Cancer screening has traditionally relied on organ-specific methods such as mammography, cervical cytology or HPV testing, low-dose computed tomography, fecal testing, colonoscopy, and selected biochemical markers. These approaches have reduced mortality for certain cancers, but they leave many aggressive malignancies without established population-based screening strategies. Pancreatic, ovarian, biliary, gastric, and several other cancers are therefore frequently diagnosed only after regional or distant spread.

A blood test capable of detecting several cancers simultaneously could theoretically complement existing screening programs and extend early detection to malignancies for which no practical screening test currently exists. Liquid biopsy has become central to this concept because tumor-associated molecules released into the bloodstream can be sampled repeatedly and non-invasively.

cfDNA consists of extracellular DNA fragments circulating predominantly in plasma. Most cfDNA in healthy individuals originates from hematopoietic cells, whereas patients with cancer may also carry a small population of fragments released from malignant or tumor-associated cells. The tumor-derived fraction is conventionally termed ctDNA [1,2].

Initial approaches focused on tumor-specific somatic mutations. Bettegowda et al. established that ctDNA could be detected across a broad spectrum of malignancies, but the probability of detection was strongly associated with tumor burden [2]. Phallen et al. later demonstrated that error-corrected sequencing could identify tumor-derived alterations in a proportion of patients with localized cancers [3]. These studies established both the potential and the principal limitation of mutation-based detection: a biologically small tumor may release only a few mutant molecules into an ordinary blood sample.

The field has subsequently expanded beyond mutation detection. Aberrant DNA methylation can occur early in malignant transformation and is distributed across thousands of genomic loci [4]. cfDNA fragmentation is also highly organized rather than random. Fragment length, nucleosomal footprints, genomic coverage, end coordinates, end motifs, and orientation patterns reflect chromatin structure and tissue of origin [5,6].

These observations have created four broad biomarker strategies: total cfDNA quantity and integrity, mutation-based ctDNA, methylation profiling, and fragmentomics. Modern multi-cancer early detection platforms increasingly combine two or more of these signals.

The present systematic review evaluates these biomarker classes from a biological and diagnostic perspective, with particular emphasis on the rapid development of methylation- and fragmentomics-based assays between 2023 and 2026.

2. MATERIALS AND METHODS

2.1 Review Design

This systematic review was conducted and reported according to the principles of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement and diagnostic-accuracy review methodology [7]. The review focused on the use of blood-derived cfDNA and ctDNA biomarkers for early or multi-cancer detection.

Because the included studies differed substantially in cancer spectrum, disease stage, assay technology, control population, sequencing depth, positivity threshold, and machine-learning methodology, a quantitative meta-analysis was not considered appropriate. Findings were therefore synthesized qualitatively.

2.2 Literature Search

Literature published or electronically available through January 2026 was considered.

("cell-free DNA" OR cfDNA OR "circulating tumor DNA" OR ctDNA) AND ("early cancer detection" OR screening OR "early-stage cancer" OR "multi-cancer early detection") AND (mutation OR methylation OR methylome OR fragmentomics OR fragmentation OR nucleosome OR "fragment size" OR "end motif").

The search strategy covered PubMed/MEDLINE and additional multidisciplinary academic databases. Citation and reference searching of major primary studies and relevant systematic reviews was also undertaken.

Priority was given to primary human studies, independent validation cohorts, multicenter investigations, prospective screening cohorts, and recent investigations published from 2023 onward.

2.3 Eligibility Criteria

Studies were considered eligible when they fulfilled the following criteria:

1. human participants were investigated;
2. plasma or blood-derived cfDNA/ctDNA was evaluated;
3. the primary objective included cancer detection or screening;
4. localized, early-stage, non-metastatic, prediagnostic, or asymptomatic populations were included;
5. at least one cfDNA-derived molecular characteristic was assessed; and
6. clinically interpretable diagnostic outcomes were reported.

Studies restricted exclusively to post-treatment monitoring, minimal residual disease, therapeutic response, or prognosis were excluded from the principal diagnostic synthesis.

Animal studies, conference reports lacking sufficient primary data, reviews, editorials, commentaries, duplicated populations, and studies without interpretable diagnostic outcomes were excluded.

Review articles were used for contextual interpretation but were not treated as primary diagnostic evidence.

2.4 Study Selection and PRISMA Flow

The database and supplementary search identified a total of 1,066 records, comprising 487 records from PubMed/MEDLINE, 538 records from other academic databases, and 41 records from citation and reference searching.

Before screening, 271 duplicate records and 18 records removed for other reasons were excluded, leaving 777 unique records for title and abstract screening.

Of the 777 records screened, 635 records were excluded because they did not meet the predefined inclusion criteria.

A total of 142 reports were sought for full-text retrieval. Five reports could not be retrieved; therefore, 137 full-text reports underwent formal eligibility assessment.

Following full-text review, 111 reports were excluded: not focused on early cancer detection (n = 29); treatment monitoring, minimal residual disease, or prognosis only (n = 21); inappropriate biomarker or non-cfDNA analyte (n = 14); review, editorial, commentary, or insufficient primary data (n = 17); insufficient diagnostic outcome reporting (n = 12); overlapping or duplicate patient population (n = 10); and non-human or preclinical-only study (n = 8).

The remaining 26 primary studies fulfilled the eligibility criteria and were included in the qualitative synthesis. No studies were pooled in a quantitative meta-analysis because of substantial clinical, methodological, and analytical heterogeneity.

2.5 Data Extraction

The following variables were extracted where available:

- study design;
- cancer spectrum;
- number of participants;
- biomarker type;
- sequencing or molecular platform;
- training and validation design;
- sensitivity;
- specificity;
- AUC;
- stage-specific sensitivity;
- cancer-signal-origin or tissue-of-origin accuracy; and
- principal methodological limitations.

2.6 Risk-of-Bias Assessment

Risk of bias was evaluated using principles of QUADAS-3, with particular consideration of participant selection, index-test construction, definition of the target condition, and analytical methodology [8].

Case-control designs using clinically established cancers and healthy controls were considered particularly susceptible to spectrum bias. Additional concerns included post hoc threshold selection, classifier overfitting, absence of external validation, incomplete follow-up of negative participants, batch effects, inappropriate control selection, and failure to account for age-related clonal hematopoiesis.

3. RESULTS

3.1 Study Selection

The PRISMA study-selection pathway is summarized in Figure 1.

A total of 1,066 records were identified through database and supplementary searches. After removal of 289 duplicate or otherwise ineligible records, 777 records remained for title and abstract screening.

Of these, 635 records were excluded, and 142 reports were sought for retrieval. Five reports were not retrieved, leaving 137 reports for full-text eligibility assessment.

A total of 111 full-text reports were excluded after detailed review. The most frequent reasons were lack of an early-detection objective (n = 29), restriction to treatment monitoring or prognostic applications (n = 21), and absence of eligible primary data (n = 17).

Ultimately, 26 primary studies were included in the qualitative systematic review, while 0 studies were included in a quantitative meta-analysis because of substantial heterogeneity.

For descriptive synthesis, the 26 included studies were classified according to their predominant biomarker approach: cfDNA concentration or integrity (3 studies), predominantly mutation-based ctDNA detection (5 studies), predominantly methylation-based detection (7 studies), predominantly fragmentomic approaches (6 studies), and multimodal or multi-analyte approaches (5 studies). The categories describe the predominant analytical strategy; some studies assessed more than one cfDNA feature.

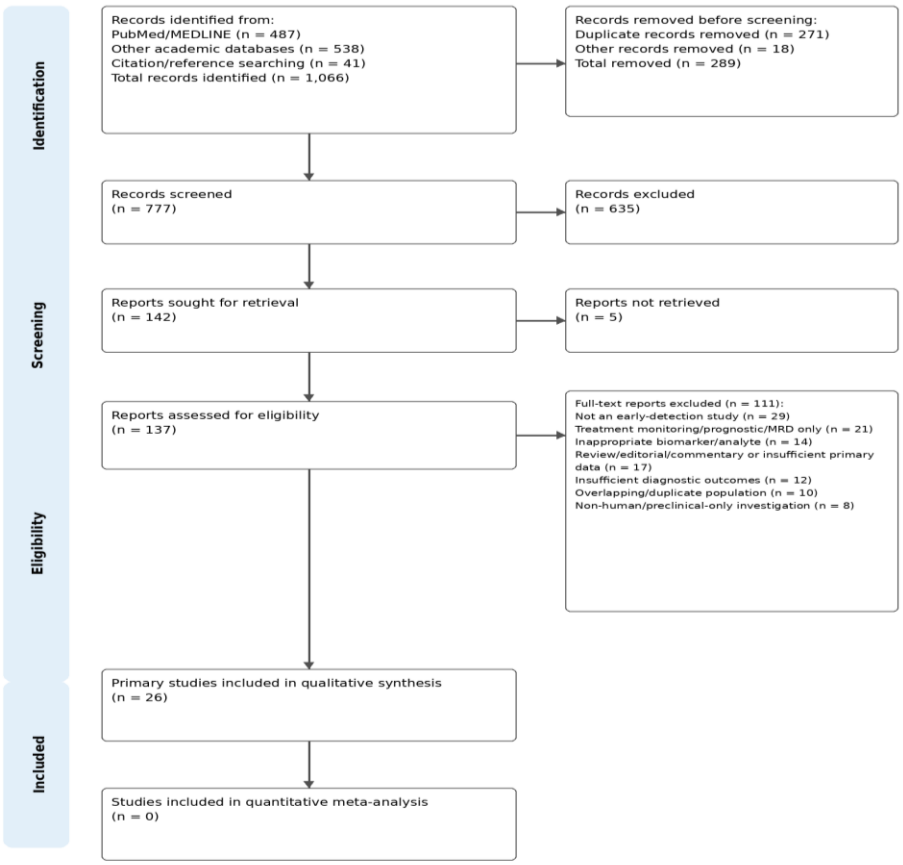

Figure 1. PRISMA 2020 flow diagram for study selection.

3.2 Biological Architecture of the Liquid Biopsy Signal

The diagnostic information carried by cfDNA can be considered at several molecular levels.

The simplest level is quantity. Cancer may increase the overall amount of circulating DNA because of altered cell turnover and tissue injury. However, the same phenomenon occurs with inflammation, infection, trauma, ischemia, strenuous exercise, pregnancy, and other physiological or pathological states.

The second level is sequence alteration. Tumor-derived molecules may contain substitutions, insertions/deletions, amplifications, copy-number abnormalities, or rearrangements.

The third level is epigenetic organization. DNA methylation reflects cell identity and malignant reprogramming.

The fourth level is physical architecture. cfDNA molecules retain information about nucleosome protection, nuclease cleavage, transcriptional organization, and chromatin accessibility through their fragmentation patterns.

Thus, a single tube of plasma contains several conceptually different cancer signals rather than a single biomarker.

Table 1. Biological comparison of cfDNA-derived biomarkers

Biomarker class	Signal interrogated	Major strength	Major limitation
cfDNA quantity	Total circulating DNA concentration	Simple and inexpensive	Low cancer specificity
DNA integrity	Relative abundance of short/long fragments	Accessible by PCR-based methods	Highly assay-dependent
Mutation-based ctDNA	Somatic genomic variants	High molecular specificity	Very low abundance in early tumors
Copy-number alterations	Genome-wide gains/losses	Broad genomic signal	Low sensitivity at low tumor fraction
DNA methylation	Differentially methylated CpG regions	Early biological change; strong tissue information	Conversion/sequencing complexity
Fragment size	Short/long fragment distribution	Genome-wide, mutation-independent signal	Influenced by processing methodology
Nucleosome footprints	Coverage around regulatory regions	Tissue and transcriptional information	Computationally demanding

Biomarker class	Signal interrogated	Major strength	Major limitation
End motifs/end coordinates	Nuclease-associated fragment termini	Additional independent signal	Standardization required
Multimodal cfDNA	Combination of >=2 molecular dimensions	Higher information density	More complex model validation

3.3 Total cfDNA Concentration and Integrity: Useful but Non-Specific

Numerous studies have demonstrated higher mean circulating DNA levels in patients with malignancy than in cancer-free individuals. DNA integrity indices, typically calculated from ratios of longer and shorter repetitive DNA sequences, have also been proposed as cancer biomarkers.

However, the direction and magnitude of integrity changes vary across studies. A previous systematic review found marked between-study heterogeneity and conflicting results concerning whether cancer is associated with increased or decreased cfDNA integrity [9].

More recent lung cancer data suggest that combining cfDNA concentration and integrity with established protein tumor markers can increase diagnostic discrimination. Ren et al. reported a combined sensitivity of 94.05%, specificity of 90.0%, and AUC of 0.968 when cfDNA measurements were combined with CEA, CA125, NSE, and CYFRA21-1 [10].

These results should not be interpreted as evidence that total cfDNA is an independent screening marker. Much of its diagnostic value likely reflects disease-associated tissue turnover rather than a cancer-specific process.

3.4 Mutation-Based ctDNA: Specific Signal, Scarce Molecules

Somatic mutation analysis was the foundation of modern ctDNA research. The advantage is intuitive: detection of an appropriate tumor-associated mutation provides direct molecular evidence of an abnormal clone.

Bettegowda et al. demonstrated that ctDNA could be identified across many solid tumors, but localized cancers were substantially less detectable than metastatic disease [2].

Phallen et al. subsequently applied targeted error-correction sequencing to early-stage colorectal, lung, breast, and ovarian cancers and demonstrated that extensive sequencing could improve detection of low-frequency variants [3].

CancerSEEK extended this concept by integrating mutations with circulating proteins. Cohen et al. evaluated 1,005 patients with non-metastatic cancers and 812 healthy controls. The combined assay achieved approximately 70% median sensitivity across eight cancer types while maintaining specificity above 99% [11].

Mutation testing nevertheless faces two biological constraints. First, the number of tumor-derived molecules available in a blood sample is finite. Increasing sequencing depth cannot recover a mutant molecule that was not present in the plasma specimen. Second, some mutations detected in plasma originate from hematopoietic clones rather than a solid tumor. Chabon et al. showed the importance of filtering clonal hematopoiesis-associated alterations when developing a blood-based lung cancer classifier [12].

This confounding becomes particularly important in older adults because both clonal hematopoiesis and cancer prevalence increase with age. Consequently, mutation-only approaches are increasingly being supplemented with epigenetic or fragmentomic features.

3.5 Methylation: Expanding the Number of Detectable Cancer Signals

DNA methylation offers an important theoretical advantage over mutation analysis. A tumor may carry only a limited number of recurrent mutations included in a particular sequencing panel, whereas malignant transformation can alter thousands of CpG sites.

Shen et al. demonstrated that plasma cfDNA methylation patterns could distinguish cancer from non-cancer samples and provide information regarding the tissue from which the signal originated [4].

Large-scale development subsequently established methylation as a principal MCED strategy. Targeted methylation analysis in the CCGA program demonstrated very high specificity and strong cancer-signal-origin performance, although sensitivity remained strongly stage dependent [13,14].

TOTEM

The 2024 TOTEM study used enzymatic-conversion-based targeted methylation sequencing. The initial clinical cohort contained 733 patients with seven cancer types and 500 healthy controls, followed by an independent validation population.

The independent validation achieved 55.9% sensitivity, 98.6% specificity, and an AUC of 0.868. Early-stage sensitivity in the main training/testing datasets was approximately 45.7-50.3%. Among correctly detected cancers, top-2 cancer-signal-origin accuracy exceeded 84% [15].

INSPECTOR

The INSPECTOR study used enzyme-assisted whole-methylome marker discovery followed by targeted high-depth sequencing.

The assay was independently validated in 824 participants and demonstrated 81.9% sensitivity, 99.0% specificity, stage I sensitivity of 65.5%, stage II sensitivity of 79.7%, and stage I-II combined sensitivity of 71.3%. Top-1 cancer-signal-origin accuracy was approximately 87.4% in independent validation [16].

Epigenetic Instability Rather Than Absolute Methylation

A 2026 study by Thursby et al. introduced an alternative approach based on variability of methylation rather than simply hypermethylation or hypomethylation.

Using 269 CpG island regions, the investigators developed an epigenetic instability index. Classifiers detected stage IA lung adenocarcinoma with approximately 81% sensitivity and early breast cancer with approximately 68% sensitivity at 95% specificity [17].

3.6 Fragmentomics: Reading the Physical Structure of cfDNA

Fragmentomics examines how DNA molecules are cleaved and packaged rather than simply what sequence they contain.

Tumor-associated cfDNA can differ from background cfDNA in median fragment size, short-to-long fragment ratios, genomic representation, 5-prime and 3-prime end sequences, preferred end coordinates, nucleosome occupancy, transcription factor footprints, orientation-aware fragmentation, and copy-number-associated fragmentation.

Cristiano et al. demonstrated that genome-wide fragmentation profiles could distinguish individuals with cancer from healthy controls and also contribute information about tissue of origin [5].

Mathios et al. later applied the DELFI approach to lung cancer detection and demonstrated the feasibility of integrating fragmentation with clinical information [18].

3.7 Ultra-Short Fragments

Traditional double-stranded sequencing workflows preferentially analyze nucleosome-sized cfDNA fragments. Single-stranded methods have revealed additional populations of much shorter circulating DNA.

Wang et al. analyzed samples from 364 patients with six cancers and 675 healthy controls using mutation, protein, and fragmentomic measurements. At 98% specificity, mutations alone achieved 46% sensitivity; mutation plus proteins increased sensitivity to 60%; and addition of fragmentomics further increased sensitivity to 66% [19].

The incremental gain from fragmentomics illustrates the value of biologically independent signals.

3.8 Multidimensional Fragmentomics

A major 2025 multicancer investigation by Bao et al. evaluated genome-wide genetic and fragmentomic features.

The study included an internal validation cohort of 3,021 cancer cases and 3,370 non-cancer controls, followed by an independent cohort of 677 cancer cases and 687 controls.

In independent validation, the classifier achieved 87.4% sensitivity, 97.8% specificity, and 82.4% tissue-of-origin accuracy.

The investigators additionally evaluated 3,724 asymptomatic participants prospectively. In that setting, sensitivity decreased to 53.5%, while specificity remained 98.1% [20].

The difference between case-control and prospective performance is particularly informative. It demonstrates why accuracy measured in enriched cancer cohorts cannot simply be transferred to population screening.

3.9 Fragmentomics in Pancreatic Cancer

Pancreatic ductal adenocarcinoma represents an important target for liquid biopsy because conventional population screening is not currently feasible.

Yin et al. evaluated shallow whole-genome fragmentomic profiling in a study involving 1,167 participants. The model integrated copy-number alterations, fragment size, mutation-related features, and methylation-associated signals.

The classifier achieved a validation AUC of 0.987, validation sensitivity of 97.3%, validation specificity of 92.8%, external validation sensitivity of 90.91%, and external validation specificity of 94.5% [21].

Although these results are impressive, the inclusion of selected cancer cases and healthy controls requires confirmation in symptomatic high-risk and prospective screening populations.

3.10 Multimodal Integration

The biological rationale for multimodal liquid biopsy is that different biomarkers fail for different reasons. Mutations may be absent from the targeted panel. Methylation may be influenced by age and tissue composition. Fragmentation signals may be subtle. Protein biomarkers may lack cancer specificity. Combining partially independent signals can therefore improve detection.

The PROMISE study prospectively collected blood from 1,706 participants, including 866 cancer cases and 840 non-cancer controls, across nine cancer categories. The multimodal model achieved 75.1% sensitivity, 98.8% specificity, and top predicted origin accuracy of 73.1% [22].

SPOT-MAS represents another multimodal strategy. The assay combines targeted methylation, fragment size, copy-number changes, and end-motif patterns while requiring relatively shallow genome-wide sequencing. In 738 non-metastatic cancer cases and 1,550 controls, reported sensitivity was 72.4% at 97.0% specificity [23].

Table 2. Selected recent studies representing the evolving liquid-biopsy landscape

Study	Year	Principal signal	Population/design	Key result
Nguyen et al., SPOT-MAS [23]	2023	Methylation + fragmentomics	738 cancers, 1,550 controls	Sensitivity 72.4%; specificity 97.0%
Wang et al. [19]	2023/24	Mutation + protein + ultra-short fragmentomics	364 cancers, 675 controls	Sensitivity increased from 46% to 66% at 98% specificity
Xiong et al.,	2024	Targeted methylation	Multi-cancer +	Sensitivity 55.9%;

Study	Year	Principal signal	Population/design	Key result
TOTEM [15]			independent validation	specificity 98.6%
Bao et al. [20]	2025	Multidimensional fragmentomics	Large independent and prospective cohorts	Independent sensitivity 87.4%; specificity 97.8%
Yin et al. [21]	2025	Fragmentomic ML for PDAC	Training + validation + external cohorts	Validation AUC 0.987
Duan et al., PROMISE [22]	2025	Methylation + proteins + mutations	1,706 participants	Sensitivity 75.1%; specificity 98.8%
Luo et al., INSPECTOR [16]	2025	Enzyme-assisted methylation	Multicenter independent validation	Sensitivity 81.9%; specificity 99.0%
Thursby et al. [17]	2026	Epigenetic instability	Lung and breast cancer	~81% sensitivity for stage IA lung cancer at 95% specificity

3.11 Stage I Disease Remains the Critical Test

Diagnostic performance should not be summarized solely by an overall sensitivity.

An assay with excellent detection of stage III-IV cancers may contribute less to screening than one with moderate overall sensitivity but strong stage I performance.

The amount of tumor-derived cfDNA generally rises with tumor burden, vascular invasion, necrosis, and metastatic dissemination. Consequently, sensitivity increases with stage in most liquid-biopsy studies.

Recent methylation studies demonstrate that stage I sensitivity can now exceed 60% in selected multicancer cohorts [16]. Fragmentomic and epigenetic-instability approaches have also reported encouraging early-stage results [17,20,21].

Nevertheless, prospective screening cohorts commonly perform less strongly than retrospective or case-control validation sets.

3.12 Cancer-Signal Origin

A positive blood test without anatomical localization creates a clinical problem. The patient may otherwise require extensive imaging, endoscopy, and invasive procedures to locate the suspected cancer.

Methylation is particularly suitable for localization because different tissues maintain characteristic epigenetic programs. Recent studies have reported top-1 cancer-signal-origin accuracies of approximately 73-88%, with higher performance when a top-two prediction is accepted [15,16,20,22].

Fragmentomic features can also encode tissue information because nucleosome organization reflects transcriptional activity of the originating cell.

Future screening algorithms may therefore report both a probability that cancer is present and a ranked probability distribution of the most likely anatomical origins.

3.13 Risk of Bias and Study Quality

The principal weakness of the available literature is participant selection.

Of the 26 included primary studies, a substantial proportion used retrospective or case-control designs involving patients with established cancers and cancer-free or healthy controls. These designs were considered at increased risk of spectrum bias because they do not fully reproduce the heterogeneity of an asymptomatic screening population.

The population encountered in clinical practice includes individuals with benign tumors, inflammatory disease, chronic organ dysfunction, smoking-related injury, age-related hematopoietic clones, premalignant abnormalities, previous cancers, and incidental imaging findings.

Machine-learning assays introduced additional potential sources of bias, including feature selection before data splitting, inadequate external validation, repeated optimization against test data, and technical batch effects.

Studies incorporating independent validation populations, prospective sampling, or asymptomatic screening cohorts were therefore given greater interpretive weight.

4. DISCUSSION

4.1 Main Findings

This systematic review identified 26 primary studies examining cfDNA-derived biomarkers for early cancer detection.

Three broad conclusions emerge from the evidence.

First, signal abundance matters more than sequencing depth alone. Early tumors may contribute extremely few molecules to circulation. Molecular technologies can reduce sequencing error, but they cannot detect a molecule that was absent from the sample.

Second, information-rich signals outperform simple concentration measurements. Methylation and fragmentomics interrogate thousands of loci or genome-wide patterns rather than relying on a small number of mutations.

Third, integration is becoming more important than individual biomarkers. The most recent platforms increasingly combine methylation, fragmentation, copy number, mutations, proteins, or clinical variables.

4.2 Why Methylation Performs Well

Aberrant methylation occurs early during carcinogenesis and affects large genomic territories. This provides a much larger target space than conventional mutation panels.

Methylation also encodes cell lineage, allowing cancer detection and localization to be solved simultaneously.

INSPECTOR provides a recent example of this strategy, with independent-validation specificity of 99.0%, overall sensitivity of 81.9%, and stage I sensitivity of 65.5% [16].

However, methylation is not perfectly cancer specific. Age, inflammation, smoking, and cell-composition changes may alter epigenetic patterns. Highly optimized methylation classifiers therefore require external validation across heterogeneous populations.

4.3 Why Fragmentomics Is Different

Fragmentomics does not require identification of a particular cancer mutation. Instead, it detects the consequences of altered chromatin architecture and DNA processing.

This provides two potential advantages: millions of fragments contribute to the classification signal, and shallow whole-genome sequencing may be sufficient for some applications.

The 2025 multicancer study by Bao et al. showed strong independent performance but also illustrated a critical limitation: sensitivity fell from 87.4% in independent case-control validation to 53.5% in the prospective asymptomatic cohort [20].

The result demonstrates why prospective validation is necessary for emerging screening tests.

4.4 The Problem of Clonal Hematopoiesis

Mutation-based assays are uniquely affected by hematopoietic clonal expansion.

Somatic mutations in genes such as DNMT3A, TET2, ASXL1, and TP53 may be present in leukocyte-derived DNA without indicating a solid malignancy.

Potential solutions include paired white-blood-cell sequencing, fragment-length analysis, mutation-context filtering, methylation-based tissue assignment, and integration of multiple molecular features.

This is one reason mutation-only classifiers have progressively been replaced by multimodal models.

4.5 Specificity Is Crucial in Population Screening

When cancer prevalence is low, even a seemingly small false-positive rate can create a large number of unnecessary diagnostic procedures.

For example, a specificity of 95% corresponds to five false-positive tests per 100 cancer-free individuals. This may be acceptable in a diagnostic setting but is generally inadequate for broad MCED screening.

Consequently, several contemporary studies optimize thresholds near 98-99% specificity. The trade-off is reduced sensitivity.

Future clinical deployment will therefore require defining acceptable thresholds according to population risk, downstream diagnostic strategy, and the consequences of missed versus false-positive disease.

4.6 Positive Predictive Value Depends on Prevalence

Sensitivity and specificity are intrinsic test characteristics within a given study setting, whereas positive predictive value depends strongly on disease prevalence.

A classifier evaluated using equal numbers of cancer cases and controls cannot directly predict the proportion of positive tests that will represent genuine cancer in an asymptomatic population.

Prospective implementation studies are therefore essential.

PATHFINDER demonstrated that a methylation-based multicancer test could be incorporated into real clinical diagnostic pathways, while also showing that only a subset of cancer-signal-positive individuals ultimately had cancer confirmed [24].

This difference between analytical accuracy and real-world predictive value must remain central when evaluating new liquid-biopsy technologies.

5. Clinical Challenges

Table 3. Major barriers to population implementation

Challenge	Clinical consequence	Potential solution
Low tumor fraction	False-negative early cancers	Larger plasma input; multimodal signals; serial testing
Clonal hematopoiesis	False-positive mutation calls	Paired leukocyte analysis
Pre-analytical variation	Reproducibility problems	Standardized collection and processing
Batch effects	Artificial machine-learning signal	Multicenter external validation
Healthy-control bias	Inflated diagnostic accuracy	Risk-matched prospective controls
Poor tissue localization	Extensive diagnostic investigations	Improved CSO algorithms
Overdiagnosis	Detection of biologically insignificant disease	Longitudinal outcome studies
Cost	Limited population scalability	Shallow sequencing and targeted panels
Algorithm opacity	Difficult clinical interpretation	Explainable and locked classifiers
Population	Reduced transportability	Geographically and ethnically diverse validation

Challenge	Clinical consequence	Potential solution
heterogeneity		

6. FUTURE DIRECTIONS

6.1 Longitudinal Personalized Baselines

Repeated blood sampling could allow each individual to serve partly as his or her own biological control. Detection of a rising molecular signal may prove more informative than comparison with a single population-wide threshold.

6.2 Integration of Methylation and Fragmentation

Methylation and fragmentomics measure partly independent aspects of tumor biology. Simultaneous extraction of both signals from the same sequencing library could improve sensitivity without proportionally increasing sample requirements.

6.3 Artificial Intelligence

High-dimensional liquid-biopsy data require computational classification. Recent meta-analyses show consistently high specificity for machine-learning cfDNA assays, but also substantial between-study heterogeneity [25,26].

Future models should be trained on demographically diverse populations and evaluated using locked algorithms before prospective validation.

6.4 High-Risk Populations

Initial implementation may be more efficient in populations with higher baseline cancer prevalence, such as heavy smokers, hereditary cancer syndromes, chronic liver disease, pancreatic cystic neoplasms, strong family history, and selected occupational exposures.

A higher pre-test probability increases positive predictive value and may improve the balance between benefits and harms.

6.5 Combining Blood Testing With Imaging

Liquid biopsy need not replace imaging.

A more realistic clinical pathway may involve blood-based molecular risk assessment -> predicted organ of origin -> targeted imaging -> tissue confirmation.

Such a pathway could reduce unnecessary whole-body diagnostic investigations.

7. Limitations of the Evidence

The available literature remains highly heterogeneous.

Cancer types and stages differ between studies, and "early cancer" is not defined uniformly. Some studies classify stages I-III as early or non-metastatic, whereas others restrict early disease to stages I-II.

Assay platforms differ substantially with respect to plasma volume, cfDNA extraction, sequencing depth, genomic targets, machine-learning algorithms, and positivity thresholds.

Case-control studies dominate the development literature and can overestimate performance.

Commercial involvement is common in assay development studies and should be transparently reported.

Long-term evidence concerning mortality reduction, overdiagnosis, psychological impact, diagnostic complications, and cost-effectiveness remains substantially less mature than evidence concerning sensitivity and specificity.

The review included 26 studies, but quantitative pooling was not undertaken because heterogeneity in biomarker platforms, cancer types, disease stages, thresholds, and analytical methods was considered too substantial to support a clinically meaningful combined estimate.

For these reasons, direct numerical ranking of different assays across separate studies should be avoided.

8. CONCLUSION

This systematic review identified 26 primary studies evaluating cfDNA-derived biomarkers for early cancer detection following screening of 777 unique records and assessment of 137 full-text reports.

Liquid biopsy for early cancer detection is undergoing a fundamental transition.

The field began with measurements of circulating DNA concentration and the search for individual tumor mutations. It has evolved toward high-dimensional analysis of epigenetic and structural characteristics of cfDNA.

Mutation-based ctDNA remains valuable because of its molecular specificity, but low tumor fractions and clonal hematopoiesis restrict its performance as an isolated population-screening strategy.

Methylation provides a broader cancer-associated signal and strong tissue-of-origin information. Recent enzyme-assisted platforms demonstrate substantial improvements in stage I detection.

Fragmentomics provides an additional layer of biological information derived from chromatin structure, nucleosome organization, and DNA cleavage. Recent independent validation studies have achieved high diagnostic performance using relatively shallow whole-genome sequencing.

The strongest direction for future development is therefore not competition between ctDNA mutations, methylation, and fragmentomics, but their rational integration.

Before widespread population implementation, these assays must demonstrate consistent performance in asymptomatic prospective populations, reproducibility across laboratories and demographic groups, acceptable diagnostic burden, and ultimately meaningful clinical benefit.

DECLARATIONS

Ethics Approval

Not applicable. This systematic review used data from previously published studies and did not involve direct recruitment of human participants.

Consent for Publication

Not applicable.

Funding

No specific funding was received for preparation of this review.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

All data discussed in this review were obtained from publicly available published studies.

REFERENCES

1. Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat Rev Genet. 2019;20:71-88.
2. Bettegowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6:224ra24. doi:10.1126/scitranslmed.3007094.
3. Phallen J, Sausen M, Adleff V, et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med. 2017;9. doi:10.1126/scitranslmed.aan2415.
4. Shen SY, Singhania R, Fehringer G, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature. 2018;563:579-583.
5. Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature. 2019;570:385-389.
6. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell. 2016;164:57-68.
7. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372.
8. Whiting PF, Tomlinson E, Rutjes AWS, et al. QUADAS-3: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2026. doi:10.7326/ANNALS-25-02104.
9. Cheng J, Tang Q, Cao X, et al. Cell-free circulating DNA integrity based on peripheral blood as a biomarker for diagnosis of cancer: a systematic review. Cancer Epidemiol Biomarkers Prev. 2017;26:1595-1602.
10. Ren S, Zeng G, Yi Y, et al. Combinations of plasma cfDNA concentration, integrity and tumor markers are promising biomarkers for early diagnosis of non-small cell lung cancer. Heliyon. 2023;9.
11. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science. 2018;359:926-930.
12. Chabon JJ, Hamilton EG, Kurtz DM, et al. Integrating genomic features for non-invasive early lung cancer detection. Nature. 2020;580:245-251.
13. 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;31:745-759.
14. Klein EA, Richards D, Cohn A, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol. 2021;32:1167-1177.
15. Xiong et al. TOTEM: a multi-cancer detection and localization approach using circulating tumor DNA methylation markers. BMC Cancer. 2024. doi:10.1186/s12885-024-12626-7.
16. Luo et al. The INSPECTOR study: enhanced feasibility for clinical translation of a multi-cancer early detection method based on enzyme-assisted high signal-to-noise ratio sequencing of methylated circulating tumor DNA. Cancer Commun. 2025. doi:10.1002/cac2.70071.
17. Thursby SJ, Jin Z, Blum J, et al. Epigenetic instability-based metrics in cell-free DNA for early cancer detection. Clin Cancer Res. 2026;32:1528-1539. doi:10.1158/1078-0432.CCR-25-3384.
18. Mathios D, Johansen JS, Cristiano S, et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. Nat Commun. 2021;12:5060.
19. Wang F, Li X, Li M, et al. Ultra-short cell-free DNA fragments enhance cancer early detection in a multi-analyte blood test combining mutation, protein and fragmentomics. Clin Chem Lab Med. 2024;62:168-177. doi:10.1515/cclm-2023-0541.
20. Bao H, Yang S, Chen X, et al. Early detection of multiple cancer types using multidimensional cell-free DNA fragmentomics. Nat Med. 2025;31:2737-2745. doi:10.1038/s41591-025-03735-2.
21. Yin L, Cao C, Lin J, et al. Development and validation of a cell-free DNA fragmentomics-based model for early detection of pancreatic cancer. J Clin Oncol. 2025;43:2863-2874. doi:10.1200/JCO.24.00287.
22. Duan J, et al. Exploration of multi-omics liquid biopsy approaches for multi-cancer early detection: the PROMISE study. Innovation (Camb). 2025.

23. 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.
24. Schrag D, Beer TM, McDonnell CH III, et al. Blood-based tests for multicancer early detection (PATHFINDER): a prospective cohort study. Lancet. 2023;402:1251-1260.
25. Li Q, Liu H, Wang J. Value of machine learning models for cell-free DNA-based multi-cancer early detection: a systematic review and meta-analysis. Technol Cancer Res Treat. 2026;25:15330338261425328.
26. Filis P, Markozannes G, Salgkamis D, et al. Integrating liquid biopsies and artificial intelligence for early cancer detection: a systematic review and meta-analysis. Eur J Cancer. 2026;239:116699.
27. Jamshidi A, Liu MC, Klein EA, et al. Evaluation of cell-free DNA approaches for multi-cancer early detection. Cancer Cell. 2022;40:1537-1549.e12.
28. Gao Q, Lin YP, Li BS, et al. Unintrusive multi-cancer detection by circulating cell-free DNA methylation sequencing (THUNDER): development and independent validation studies. Ann Oncol. 2023;34:486-495.
29. Lennon AM, Buchanan AH, Kinde I, et al. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. Science. 2020;369.
30. Chen X, Gole J, Gore A, et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat Commun. 2020;11:3475.
31. Cui P, Zhou X, Xu S, et al. Prediction of methylation status using WGS data of plasma cfDNA for multi-cancer early detection. Clin Epigenetics. 2024;16:34. doi:10.1186/s13148-024-01646-6.
32. Zeng Y, et al. A pan-cancer compendium of 1,294 plasma cell-free DNA methylomes and fragmentomes enabling multicancer detection. Nat Cancer. 2026;7:384-398. doi:10.1038/s43018-026-01116-3.