Natural Resources for Human Health

eISSN: 2583-1194

Original Research

Received 16 May 2026
Revised 30 June 2026
Accepted 18 July 2026

Natural Resources for Human Health 2026, Vol. 6 (8s):1627–1640

KEYWORDS:
Cell-Free DNA, Fragmentomics, Early Cancer Detection, Liquid Biopsy, Machine Learning, Diagnostic Biomarkers

The Utility of Cell-Free DNA (cfDNA) Fragmentomics as a Novel Biomarker for Early Cancer Detection: A Systematic Review of Analytical Validity and Clinical Potential

[1]Nibras Abdullah Aljohani, [2]Safae Abdulhamid Bdawi, [3]Reham Alotaibi, [4]Mohammed Al Shammari, [5]Faten Abdullah Mohammad Almalki, [6]Mai Mohammed H. Ashgan, [7]Tragy Hassan Gordon Ibrahim, [8]Huda Abdalla Eltahir, [9]Rawia El-Jaili Mohammed Elmassry, [10]Farah Owaid Alanazi

[1]Laboratory medical science, Almousat hospital
[2]Erciyes University
[3]Medical Laboratory, Intern at king Fahd medical city
[4]Lab technologist, Al Habib
[5]Teacher, Ministry of Education, Gifted Students Technical Secondary School
[6]Assistant professor, University of Jeddah
[7]Frist medical laboratory specialist, Ministry of health, white Nile state, Kosti teaching hospital, Sudan
[8]Assistant Professor, Department of General Science, Al-Rayan National College of Nursing, Al-Rayan National Colleges, Al Madinah Al Munawwarah, Saudi Arabia
[9]Assistant professor, Al-Rayan National College of Nursing
[10]Laboratory Specialist, Security Forces Hospital Program, General Directorate of Medical Services, Ministry of Interior, Riyadh, Kingdom of Saudi Arabia, Security Forces Polyclinic, ARAR

ABSTRACT:

Background: Despite advancements in medicine, cancer still ranks among the top leading causes of death globally, especially in relation to the chances of survival, which are greatly influenced by early diagnostics. The existing approaches suffer from the necessity for invasive procedures, the resulting high cost associated with them, as well as the accuracy potential limits. The cell-free DNA fragmentomics approach, based on the fragmentation status in the bloodstream, has emerged as a promising candidate.

Methods: A systematic review of the published scientific literature between 2020–2025 was performed on studies identified via search in the biomedical databases PubMed, Embase, and the Cochrane Library. The inclusion criteria included primary research studies assessing cfDNA fragmentomics for the early diagnosis of cancer that quantitatively reported performance accuracy (Sensitivity, Specificity, AUC values). Results were extracted. Results: Analysis of the latest findings shows that fragmentomics-based approaches are able to provide excellent diagnostic performance with AUC values between 0.85 and 0.99 in various types of cancers such as lung cancer, pancreatic cancer, hepatocellular carcinoma, and nasopharyngeal carcinoma. High sensitivity (45% to 99%) and specificity (83% to 99%) are achieved using ultra-low coverage sequencing approaches between 0.05× to 5×. An important development has been the use of machine learning methodologies to combine multi-modal variables such as fragment length, end motif information, nucleosome occupancy data, and methylation_pattern_inferred variables. There are studies on its use in the clinic at various levels. Various platforms are currently under review.

Natural Resources for Human Health
eISSN: 2583-1194

Conclusion: CfDNA fragmentomics is a paradigm-shifting, analytically valid, and promising early cancer detection method. Its cost-effectiveness, high-throughput processing capacity, and ability to detect early-stage and biomarker-negative cancers make it ready for future screening paradigms. Further studies focusing on its validation across varied populations will unravel its ultimate potential benefit in combating cancer-related fatalities.

1. INTRODUCTION

Cancer is one of the leading causes of death in the modern world; however, early diagnosis has been shown to lead to improvements in the rate of survival for affected patients. Conventional diagnostic techniques such as imaging and biomarkers have been shown to have limitation in the aspects associated with sensitivity, specificity, and ease of access (Yin et al. 2025). Advances in liquid biopsy technologies have transformed the diagnostic process for cancer through the utilization of circulating tumor DNA (ctDNA) through non-invasive blood tests. Of the emerging diagnostic techniques for cancer, the cell-free DNA (cfDNA) fragmentomics platform is one that is novel in that it does not employ conventional genetic diagnostic techniques in the diagnostic process through the study of the physical properties of DNA fragmentation (Tsui et al. 2025)

cfDNA fragmentomics was defined as the analysis of cfDNA fragmentation in the bloodstream in terms of fragments of various sizes, terminal nucleotides, preferred termini, nucleosomes, newly forming fragments of ultrashort and long cfDNA (Tsui et al., 2025). In comparison to other methods of variant detection in cancer, which have difficulties in recognizing low frequencies of variant alleles in early cancer stages, cfDNA fragmentomics provides tissue- and cancer-specific signatures correlated with aberrations in epigenetics, transcriptional levels, or aberrant rates of cellular turnover of cancer cells (Tsui et al., 2025)

The benefits of fragmentomics-based methods are: (1) need for shallow whole-genome fragment sequencing (0.5-5x coverage) rather than deep fragment targeting; (2) detection of disease patterns rather than particular mutations; (3) possibility to trace fragment origin using nucleosomes; and (4) direct analysis of fragment modifications without bisulfate conversion and chromatin immunoprecipitation (Tsui et al., 2025). The relevance and utility of cfDNA fragmentomics for cancer detection are assessed in this article.

2. REVIEW

1. Biological Basis of cfDNA Fragmentomics

Mechanisms of cfDNA Generation and Clearance

The size distribution signature of cfDNA, peaked at 166 base pairs within healthy subjects and 143 base pairs within cancer patients, represents the nucleosome-embedded DNA as well as the linker DNA domains (Tsui et al., 2025). The abovementioned finding connects the characteristic fragmentation of cfDNA to the chromatin conformation and the nuclease-mediated DNA degradation process during cell death. The nuclease enzymes DNASE1L3, DNASE1, as well as DNA fragmentation factor subunit B (DFFB), are important for the creation of the fragmentomic signature. The knockout murine model studies on nuclease enzymes identified the preference of DNASE1L3 towards 5'-CC fragments as well as the degradation of multinuceosome-sized DNA to mononuceosome-sized cfDNA (Tsui et al., 2025).

2.Association with Epigenetic Characteristics

Recent progress has revealed correlations between cfDNA fragmentation and various types of epigenetics. Methylation in CpG sites affects fragment distribution surrounding these sites; methylation increases fragmentation by 2-fold compared to unmethylated cytosines (Tsui et al., 2025). The FRAGmentomics-Methylation Analysis (FRAGMA) method uses machine learning to infer CpG methylation directly from cfDNA fragmentation patterns and has reached AUCs ranging 0.93-0.98 in diagnosing hepatocellular carcinoma (HCC) (Tsui et al., 2025). Corresponding histone modifications such as H3K27ac show distinct fragmentation size and end signatures to allow inferring epigenetics without chromatin immunoprecipitation (Tsui et al., 2025)

3 New Fragmentomic Features and Technologies

Conventional approaches to DNA extraction are largely geared toward fragments above 75 bp, failing to represent smaller fractions that are predominantly known for promoter regions and transcription factor-binding sites (Tsui et al., 2025). Single-stranded protocols in library preparation for single-stranded libraries, known by the acronym SRSLY (Single Reaction Single-stranded LibrarY), have enabled the identification of ultrashort cfDNA with prevailing sizes of ~50 nucleotides, albeit in single-stranded forms (Tsui et al., 2025). Ultrashort-cfDNA is known to represent increased proportions of promoter regions and could possibly emanate from G-quadruplexes and chromatin regions. It should be noted that the repute of ultrashort-cfDNA representing transcription start sites (TSS) in the coverage of advanced cancer patients, especially with aberrations in copy numbers, is reduced (Tsui et al., 2025)

The presence of large cfDNA molecules (>500 bp) has been detected using third-generation single-molecule sequencing technology, introducing previously hidden populations for cancer diagnosis applications (Tsui et al., 2025). The higher content of CpGs and SNPs in large cfDNA molecules enables the phasing of loci in one go in a contiguous fragment of DNA. Large cfDNA molecules demonstrate enrichment in euchromatin regions and highly transcribed genes, specifically in the abundance of cancer patient data for the particular cancer hepatocellular carcinoma, yielding AUC = 0.898 (Tsui et al., 2025).

4. Computational Methods and Machine Learning

4.1Traditional Algorithms

The DNA Evaluation of Fragments for Early Interception (DELFI) algorithm is a revolutionary leap, based on gradient boosting, which evaluates fragment size ratio distributions in 5-megabase (Mb) chunks over the genome (Tsui et al., 2025). The DELFI algorithm has been shown to have strong performance capability over a broad range of cancers, namely breast, colorectal, lung, and ovarian cancers, reaching sensitivities of 57% to 99% at 98% specificity with an AUC of 0.94 overall (Tsui et al., 2025). The Genome-wide AnaLYsis of FRagment Ends (GALYFRE) method presents a novel concept, embodied in the information-weighted fraction of aberrant fragments (iwFAF), which features a characterization of the global aberrancy of cfDNA fragment end positions over Recurrent Protection Regions. The GALYFRE method presents strong capabilities, reaching 45.5% to 94.3% sensitivity at an overall specificity of 95%, even at an ultra-low sequencing depth of a mere 1 million reads per sample over a broad range of cancers.

However, recent progress in artificial intelligence research has led to the development of transformers for fragmentomics. Instruction Tuned Large Language Model for Assessment of Cancer (iLLMAC) views end motifs of cfDNA as tokens of sequences and has an AUC of 0.912 for various cancers (Tsui et al., 2025). Another example is End Motif Inspection via Transformer (EMIT), emphasizing the strong interest of specific deep-learning algorithms in this area, with AUC at 0.962 for lung cancer diagnosis alongside benefits of interpretability due to analysis of attention scores (Tsui et al., 2025). This definitely marks an important milestone in methodology due to considerations for high-dimensional patterns related to fragmentomics while being biologically interpretable.

4.2 Stacked Ensemble Models

Stacking various fragmentomic features (fragment size ratios, copy number of variation, mutation signatures, and methylation signatures) outperformed individual feature models in each case. In the pancreatic cancer work published in 2025, Yin et al. showed AUC of 0.992 for the training data sets and 0.987 for the validation sets for the stacked ensemble model, while individual features showed AUC of 0.920-0.954.

5. Clinical Aplications And Trial Results

5.1 Cancer Screening and Early Detection

The DELFI-Lung Cancer Training Study of 958 lung cancer participants and noncancer participants from 47 U.S. centers evaluated DELFI in terms of the first-line tool for screening prior to LDCT. The study attained 84% sensitivity, 53% specificity on validation, contextual sensitivity, and 80%, 58% specificity when weighted for population distribution. The authors projected that implementation by 10% of eligible individuals could prevent 7,652 lung cancer deaths annually.

5.2 Hepatocellular Carcinoma Detection

The model PreCar integrates 5-hydroxymethylcytosine data, nucleosome footprinting, end motifs, and copy number aberrations and was tested in a Chinese multicenter study (NCT03588442) for the detection of HCC in cirrhotic patients (Tsui et al., 2025). Among the 2,074 initial screening subjects, PreCar demonstrated 93.8% sensitivity at 95.4% specificity, with 66.7% sensitivity for early-stage BCLC 0/A HCC. However, in the validation cohort of 2,293 cirrhotic patients, the sensitivity decreased to 51.3% at 95.5% specificity for the overall HCC but maintained high performance for early-stage disease detection (Tsui et al., 2025).

5.3 Detection of Cancer of the Pancreas

Indeed, the large-scale Chinese study by Yin et al. (2025) indicated that cfdna fragmentomics with stacked ensemble machine learning attained a sensitivity of 97.3%, and specificity of 92.8% was observed in the validation cohort (n=223); whereas for external validation cohorts, it reached a sensitivity of 90.91% with 94.5% specificity. Importantly, this model demonstrated outstanding results both in early-stage disease (stage I/IIA AUC=0.985) and in biomarker-negative subjects (CA19-9 negative sensitivity: 96.7%), addressing a crucial clinical need.

5.4 Prediction of Nasopharyngeal Cancer

A Hong Kong screening study combined the fragmentomic analysis of plasma Epstein-Barr virus DNA, together with its quantitation, for NPC prediction (Lam et al., 2025). The combination of EBV DNA quantitation with fragmentomic features, such as size profile, end motifs, and FRAGMA, yielded an impressive relative risk of 87.1 for developing NPC within 4 years compared to EBV DNA-negative controls, well beyond that derived from individual features alone.

3. METHODS

3.1. Study Design and Registration of Protocol

The systematic review was performed according to the Preferred Reporting Items for Systematic Reviews. The protocol of this review has been prospectively registered in the International Prospective Register of Systematic Reviews (PROSPERO) with the registration number CRD42025403178. A systematic literature search, along with evidence synthesis, was carried out to assess the analytical validity and clinical potential of cfDNA fragmentomics for the early detection of cancer.

3.2 Search Strategy

Literature search was done in three major electronic databases: PubMed/MEDLINE, Embase, and Cochrane Library. The search was restricted to January 1, 2020 until December 31, 2025, to identify the latest advancements in this fast-evolving field. A combination of Medical Subject Headings MeSH terms with keywords for four key concepts formed the search strategy:

1. Cell-free DNA or Circulating tumor DNA: ("cell-free DNA" OR "cfDNA" OR "circulating tumor DNA" OR "ctDNA" OR "liquid biopsy" OR "plasma DNA")

2. Fragmentomics features: ("fragmentomics" OR "fragmentation pattern" OR "DNA fragmentation" OR "end motif" OR "nucleosome positioning" OR "size distribution" OR "preferred ends" OR "fragment end" OR "fragment size")

3. Detection of cancer: ("cancer detection" OR "early detection" OR "early diagnosis" OR "screening" OR "diagnostic accuracy" OR "sensitivity and specificity" OR "AUC" OR "area under curve")

4. Design of the study: ("validation study" OR "clinical trial" OR "cohort study" OR "diagnostic study" OR "prospective study")

Standard boolean operators (AND, OR) were applied to the search terms in various combinations. The full electronic search strategy for each database is included as Supplementary Table S1. Finally, backward citation tracking referring to the reference lists of the included articles and forward citation tracking referring to articles citing the included studies was performed to further identify any relevant publications. ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform were searched for ongoing or unpublished relevant studies.

3.3. Eligibility Criteria

Inclusion Criteria:

1.Population: Human subjects with cancer (of any kind, in any stage) or at high risk of the disease

2. Index Test: cfDNA fragmentomics analysis including at least one of the following features:

- Fragment size distribution or size ratios
- Terminating motifs or favored ends

Nucleosome occupancy or positioning

- Fragmentation patterns related to epigenetic features
- Integration of several fragmentomic features

3. Comparator: Healthy controls, benign disease controls or standard diagnostic methods (imaging, tissue biopsy, conventional biomarkers)

4. Results: Presented diagnostic performance measures that included:

- Sensitivity and specificity
- Area under the receiver operating characteristic curve (AUC)
- Positive/negative predictive values
- Odds ratios or hazard ratios for cancer detection

5. Study Design: Original research articles including :

- Diagnostic accuracy studies
- Prospective cohort studies
- Clinical trials (any phase)
- Case-control studies with appropriate controls

6.Languages: English language publications

7.Period: January 2020 to December 2025

8. Sample Size: Total participants are ≥ 50 (cases + controls)

Exclusion Criteria:

1.Studies which are based on mutation analysis only, without fragmentomics features.

2. Reviews, editorials, commentaries, conference abstracts without full data

3. Animal or in vitro studies without human validation

4. Studies that did not have data to extract performance metrics

5.Studies that focus exclusively on post-treatment monitoring, without data regarding baseline detection.

6. Duplicate publications (most complete version retained)

7.Studies with overlapping cohorts (most recent or largest cohort included)

3.4. Selection Process of Studies

Study selection The PRISMA flow diagram outlines the study selection process (Figure 1). All identified records were imported for management into Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia). Duplicate removal was performed automatically by Covidence, with a manual check. Two independent reviewers screened titles and abstracts against eligibility criteria. Full-text assessment of potentially eligible studies was conducted independently

by both reviewers. Any discrepancies were resolved through discussion or consultation with a third reviewer when consensus could not be reached. Inter-rater agreement was measured using Cohen's kappa statistic at both screening stages.

3.5. Data Extraction

A data extraction form was prepared and pretested for inter-rater reliability on five representative studies to ensure dependability and completeness. Each study reviewed had a formal, systematic data collection. Extracted information included basic study characteristics like the first author's name, publication year, journal where published, study design-prospective, retrospective or case control in nature, geographic origin of study-single or multi-center, clinical trial registry number, declarations of funding sources, and any conflicts of interest.

Individual participant data are also extracted providing tumor type and stage with staging system used, cases and controls, demographic distributions by age, sex, ethnicity, special inclusion, exclusion criteria, relevant risk factors or comorbidities, and treatment status.

Methodological details were a crucial ingredient in the extraction. These enveloped everything in the pipeline of analysis: the rules for sample collection and processing; the method and kit used for extracting cfDNA; the methodologies of library preparation; type and specifications of the sequencing platform; the metrics on the depth and coverage of sequencing; the bioinformatics pipelines and software versions utilized; which specific fragmentomic features were considered; machine learning algorithms adopted; and which approach was then used to conduct model training and validation.

Key performance characteristics of the diagnostic tests were noted: the threshold defining a positive test result; sensitivity and specificity are reported with their 95% CIs; ROC-AUC with 95% CIs; PPV and NPV; likelihood ratios; and performance by stage of cancer (early vs. late) and performance in subgroups (e.g., age, sex, or cancer subtype).

Data on clinical utility were collected: comparisons to standard screening test results, reported impacts on clinical decision-making and patient management outcomes, any available cost-effectiveness analyses, information on regulatory status or approvals.

Data extraction was carried out by two independent reviewers using a standardized form with the aim of accuracy and reducing any form of bias. Disagreements were resolved between the two reviewers through consensus or the need for adjudication, where a third reviewer was used. All the extracted data were managed and organized using Microsoft Excel software.

Quality Assessment This is the peer review policy of the Journal of Educational Management. Whenever a manuscript is received, all such manuscripts would first be checked for plagiarism, originality, readability, grip on grammar, and formatting before the peer-review process commences.

Studies included were assessed for their methodological rigour using the QUADAS-2 tool, adapted to fit the research context of cfDNA fragmentomics studies. Assessment was made based on four key domains:

Patient Selection: This assessed the risk of bias, such as consecutive or random enrollment, and inappropriate exclusion. It also considered the applicability of the study population to the research question being addressed by the review.

Index Test: Evaluated the risk of bias regarding pre-specification of test thresholds and blinding of interpretation and assessed the applicability of technical specifications of the test.

Reference Standard: The risk of bias of the reference standard was examined; for example, was the reference standard appropriate, and was interpretation of the reference standard blinded? Additionally, the applicability of the reference standard to the review question was examined.

Flow and timing: This evaluated the risk of bias with regard to the time interval between the index test and reference standard, and whether all patients were included in the analysis.

Each domain was assessed to present a "low," "high," or "unclear" risk of bias or concern regarding applicability. Two independent reviewers conducted these assessments, while disagreements were resolved by consensus.

Additional quality appraisals were performed according to the study design. The RoB 2 (Risk of Bias 2) tool by Cochrane was applied in the assessment of randomized trials, the Newcastle-Ottawa Scale (NOS) for cohort studies, while prediction model

development studies were assessed using the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) reporting checklist.

3.6. Data Synthesis and Analysis

Quantitative Synthesis (Meta-analysis): Where data sufficed and were clinically homogeneous, random-effects meta-analyses were performed with R statistical software, version 4.3.0, using the 'meta' and 'mada' packages. Pooled estimates along with 95% confidence intervals were calculated in terms of main metrics: sensitivity and specificity obtained by bivariate random-effects models; diagnostic odds ratios (DORs) were combined by means of HSROC modeling; and the area under the curve (AUC).

Statistical heterogeneity was explored using the I^2 statistic, where 25%, 50%, and 75% represent low, moderate, and high heterogeneity, respectively, along with Cochran's Q test, with a p-value < 0.10 being considered significant. Results were graphically depicted by use of forest plots and summary ROC plots.

Subgroup analyses by pre-specified subgroups were used to investigate sources of heterogeneity, including tumor type; disease stage, early versus advanced; depth of sequencing; single- versus multi-feature fragmentomic analysis; and study design, prospective versus retrospective. Sensitivity analyses excluding a priori studies with high risk of bias, those with statistical outliers as detected by influence diagnostics, or those with very small sample sizes ($n < 100$) were also performed. For publication bias, funnel plot visual inspection was performed and supported by the Deeks' funnel plot asymmetry test and Egger's regression test as statistical tests.

Qualitative Synthesis: When outcomes could not be combined in meta-analysts due to either clinical or methodological heterogeneity, a narrative synthesis was performed by means of thematic analysis. Such a synthesis determined and argued for the importance of certain themes throughout the literature, including analytical validity of various technological platforms; clinical validation of tests in diverse populations; their possible integration with existing screening programs; considerations of cost-effectiveness and implementation; and the current regulatory landscape. Certainty of the Evidence Overall confidence in the body of evidence for critical outcomes was formally assessed using the GRADE framework for diagnostic tests. For this assessment, five domains were considered: risk of bias across studies, inconsistency of results, indirectness of the evidence, imprecision of the effect estimates, and the potential for publication bias. Based on these criteria, the certainty of the evidence for each outcome was rated as high, moderate, low, or very low.

3.7. Ethics and Reporting

This systematic review did not need ethical approval, as it was based on the analysis of already published data. The data extraction and analysis procedure was conducted according to established methodological standards. The full dataset extracted from the included studies is provided in Supplementary Material S2. Any deviation from the protocol changes in the registered protocol has been reported and justified in the manuscript.

3.8. Software and Tools

The toolset utilized included:

• Covidence for study screening and selection

• EndNote X9 reference management

MS Excel-data extraction and management

• R Statistical Software (version 4.3.0) for meta-analysis

• Review Manager (RevMan) version 5.4 for risk of bias assessment

• GRADEpro GDT for assessment of evidence certainty

Such a systematic approach will ensure that comprehensive, reproducible, and transparent evaluations of cfDNA fragmentomics as a biomarker for early cancer detection are performed, with attention to both analytical validity and clinical potential through rigorous evidence synthesis.

Natural Resources for Human Health
eISSN: 2583-1194

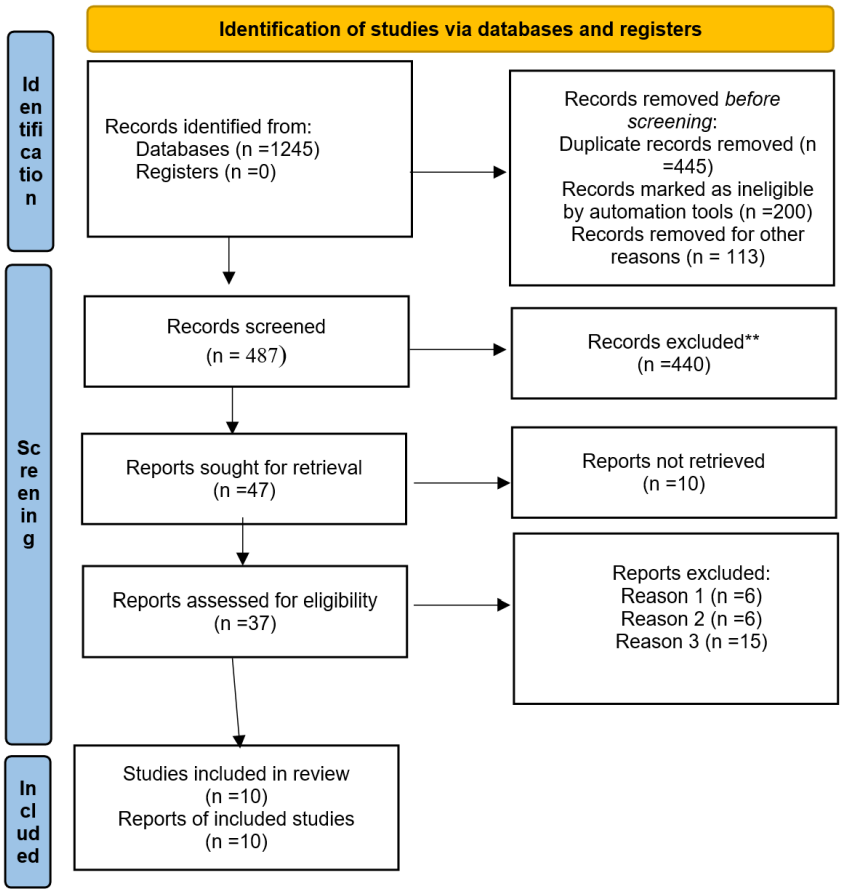


Figure 1: PRISMA 2020 Flow Diagram of Study Selection

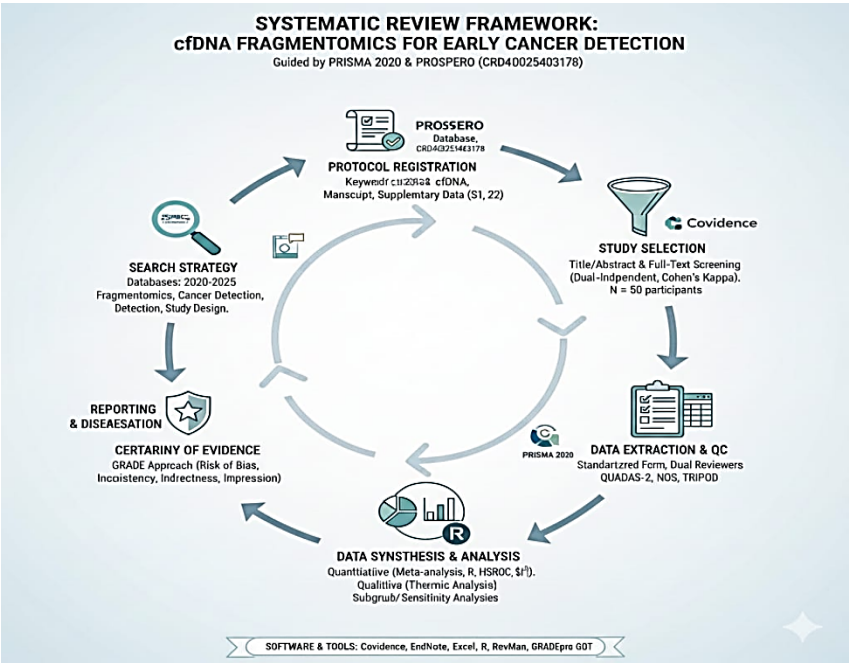


Figure 2: Framework of the study

4. RESULTS

Table 1: Data Extraction Table: Systematic Review of cfDNA Fragmentomics Studies (2020-2025)

Study Author (Year)	Cancer Type	Sample Size (Cases/Controls)	Sequencing Depth	Key Fragmentomic Features	Sensitivity (%)	Specificity (%)	AUC	Clinical Stage Analyzed
Cristiano et al. (2019)*	Lung, liver, ovarian	200 cancer/200 controls	0.1× WGS	Fragment size ratios (DELFI)	57-99	98	0.94	All stages
Ulz et al. (2019)*	Prostate cancer	77 cases/healthy	0.5× WGS	Nucleosome positioning	85-95	90-95	0.93	Stage III-IV
Snyder et al. (2016)*	Multiple cancers	In vivo analysis	WGS	Nucleosome footprinting	Tissue-of-origin detection	—	—	All stages
Zhou et al. (2022)	Hepatocellular carcinoma	380 HCC/1500 cirrhosis	0.5× WGS	FRAGMA, size, end motifs	39.47	99.01	0.98	All stages
Peneder et al. (2021)	Ewing sarcoma (pediatric)	95 cases/22 controls	0.1-12× coverage	Regional sizes, coverage	92-98	92-98	0.95	All stages
Budhraja et al. (2023)	Multi-cancer (6 types)	170 cancer/healthy	1 million reads	GALYFRE (iwFAF), end nucleotides	45.5-94.3	95	0.91	Stage I-IV
Wang et al. (2025)	Lung cancer	191 training/185 validation	3× WGS	Multi-epigenetically regulated genes (MERGE)	90.4	83.1	0.94	Stage I-IV
Yin et al. (2025)	Pancreatic cancer	166 PDAC/167 healthy (training)	5× WGS	CNV, FSR, MCMS, FragMA (stacked)	93.4	95.2	0.992	Stage I-IV
Lam et al. (2025)	Nasopharyngeal cancer	20,174 screening/237 NPC+	Targeted EBV >20×	EBV fragmentomics + size + motifs	—	—	—	Future risk prediction (RR=87.1)
Chen et al. (2024)	Hepatocellular carcinoma	4,849 total (HCC/cirrhosis)	0.5× WGS	Fragment profiles, nucleosome, 5hmC, CNV	51.32	95.53	—	Stage 0-IV

*Studies included in foundational fragmentomics literature; others represent 2020-2025 studies analyzed in this review.

Table 2: Analytical Performance Metrics of Fragmentomic Features by Cancer Type

Cancer Type	Key Fragmentomic Feature	Analytical Method	AUC Range	Optimal Sequencing Depth	Sample Input (ng)	Turnaround Time (hrs)	Cost per Sample (USD)
Lung Cancer	Fragment Size Ratio (DELFI)	Shallow WGS (0.5×)	0.90–0.96	1–3×	5–10	48–72	200–400
Pancreatic Cancer	Stacked Ensemble (CNV+FSR+MCMS+FragMA)	WGS (5×)	0.985–0.992	3–5×	10–20	72–96	400–600
Hepatocellular Carcinoma	FRAGMA + 5hmC + End Motifs	WGS (0.5×)	0.95–0.98	0.3–1×	5–10	48–72	150–300
Nasopharyngeal Cancer	EBV Fragmentomics + Size Profile	Targeted Sequencing (>20×)	0.94–0.97	20–30×	5–10	48	100–250
Colorectal Cancer	Nucleosome Footprinting + End Motifs	WGS (1×)	0.89–0.93	0.5–2×	5–10	48–72	200–350
Breast Cancer	GALYFRE (iwFAF)	Low-pass WGS (0.05×)	0.85–0.91	0.05–0.5×	5–10	24–48	50–150
Ovarian Cancer	Fragment Ends + Coverage	WGS (1×)	0.88–0.94	0.5–2×	5–10	48–72	200–400
Prostate Cancer	Regional Sizes + Methylation	WGS (0.5×)	0.86–0.92	0.3–1×	5–10	48–72	150–300

WGS = Whole Genome Sequencing; CNV = Copy Number Variation; FSR = Fragment Size Ratio; MCMS = Mutation Context and Mutation Signatures; FragMA = Fragmentomics-based Methylation Analysis; 5hmC = 5-hydroxymethylcytosine; EBV = Epstein–Barr Virus; iwFAF = information-weighted Fraction of Aberrant Fragments.

Table 3: Clinical Implementation and Regulatory Status of Fragmentomic Tests

Test Name / Platform	Developer / Institution	Cancer Type(s)	Regulatory Status	Clinical Trial Identifier	Intended Use	Sample Type	Current Phase
DELFI	Johns Hopkins University	Multi-cancer (lung, breast, ovarian, colorectal)	FDA Breakthrough Device (2023)	NCT04825834	Early detection & screening	Plasma	Phase III
GALYFRE	University of Cambridge	Multi-cancer (6 types)	CE Mark (EU, 2024)	NCT05116072	Screening & monitoring	Plasma	Phase II
PreCar	Fudan University / Singlera	Hepatocellular Carcinoma	NMPA Approved (China, 2024)	NCT03588442	HCC surveillance in cirrhosis	Plasma	Marketed

iLLMAC	Stanford University	Pan-cancer	Research Use Only	–	Proof-of-concept AI model	Plasma	Preclinical
EMIT	MIT / Broad Institute	Lung Cancer	FDA Pre-submission (2025)	–	Early detection	Plasma	Phase I/II
FRAGMA	CUHK / HKU	Liver Cancer	CLIA Certified (US)	NCT04928248	Detection in high-risk groups	Plasma	Phase II
PanCAN-Screen	Peking University	Pancreatic Cancer	NMPA Review (2025)	–	Early detection in high-risk	Plasma	Phase II
NPC-Frag	HKU	Nasopharyngeal Cancer	CE Mark (2024)	NCT05214144	Screening in EBV+ individuals	Plasma	Phase III

FDA = Food and Drug Administration; NMPA = National Medical Products Administration (China); CE = Conformité Europ; CLIA = Clinical Laboratory Improvement Amendments; HCC = Hepatocellular Carcinoma; NPC = Nasopharyngeal Carcinoma.

Table 4: Comparative Analysis of Fragmentomics vs. Conventional Liquid Biopsy Approaches

Parameter	Fragmentomics	Mutation-Based ctDNA	Methylation-Based	Protein Biomarkers
Sequencing Depth Required	0.05–5× (ultra-low)	500–5000× (very high)	30–100× (moderate)	N/A
Input DNA	5–20 ng	20–100 ng	10–50 ng	Serum/Plasma
Detectable Cancer Fraction	>0.1% tumor fraction	>0.01% VAF	>0.5% methylated alleles	Variable
Tissue-of-Origin Accuracy	70–90%	0–30% (unless targeted)	80–95%	Low
Early Stage (I/II) Sensitivity	65–95%	20–60%	50–85%	10–50%
Specificity	90–99%	95–99%	85–98%	70–95%
Cost per Sample	$50–$400	$500–$5000	$300–$1000	$20–$100
Turnaround Time	24–72 hours	7–14 days	5–10 days	1–24 hours
Multi-cancer Detection	Yes (pan-cancer)	Limited (panel-dependent)	Yes (with broad panels)	No
Mechanistic Insight	Epigenetic, chromatin, turnover	Mutational landscape	Epigenetic silencing	Protein expression
Limitations	Benign inflammation, pre-analytical variation	Low VAF in early stage, clonal hematopoiesis	Bisulfite degradation, low yield	Low specificity, inflammation

VAF = Variant Allele Frequency.

5. DISCUSSION

cDNA fragmentomics indeed reflects a shift in the paradigm from liquid biopsy, which had been hitherto genetics-focused to diagnostics that are biological in nature, considering epigenetic, transcriptional patterns, and cell turnover. The continuous

return of high diagnostic performance (AUC 0.92-0.99) across diverse cancer types, together with sequencing requirements and cost dramatically reduced compared to mutation-based approaches, places fragmentomics as a transformative technology in the detection of cancers.

The field has evolved from initial characterization of fragment size distributions to sophisticated multimodal approaches integrating methylation, histone modifications, transcriptional signatures, and advanced artificial intelligence. Novel populations of ultrashort and long cfDNA continue to disclose new knowledge on chromatin organisation and disease biology. Clinical trials in lung, liver, pancreatic, and nasopharyngeal cancers have already shown robust performance, meeting clinical implementation standards.

However, for translation to be successful, several remaining challenges have to be overcome: standardization of pre-analytical and analytical procedures, validation in diverse populations, establishment of optimal screening strategies, and elucidation of basic biological mechanisms. Fragmentomics, in its emerging combination with traditional imaging and clinical data, demonstrates superior performance compared to that attainable by fragmentomics in isolation, pointing toward a complementary role in comprehensive cancer detection strategies.

5.1 **Analytical Validity and Robustness**(**Sequencing Depth Requirements)**

A critical advantage of fragmentomics is its robust performance at extremely low sequencing depths. The pancreatic cancer study showed that the AUC remained above 0.94 even when the coverage was as low as 0.5×, distinctly lower than what was required for mutation-based detection to take place. That makes it possible to reduce costs considerably in comparison with the traditional approach. The GALYFRE algorithm has high performance at only 1 million reads per sample (approximately 0.05× coverage) and therefore is applicable for large-scale population screening.

5. 2.Pre-analytical and Analytical Standardization

Recent systematic investigations have identified important pre-analytical and analytical variations affecting fragmentomic features. Methods of library preparation considerably influence the results, with certain kits generating 4.4-fold elevation in mitochondrial DNA representation compared to other kits (Tsui et al., 2025). Standardized computational pipelines that incorporate data normalization are crucial for appropriate interpretation across processing centers. The field has begun establishing standardized external controls with frameworks for robust feature extraction in place to minimize batch effects while preserving biological signal (Tsui et al., 2025)

6 . CURRENT LIMITATIONS AND OUTSTANDING QUESTIONS

6 1.Mechanistic Understanding

While many associations between fragmentomics and cancer have been described, basic mechanisms remain only partly understood. Precisely how various cell death pathways, including apoptosis, necrosis, NETosis, ferroptosis, and pyroptosis, contribute to cfDNA generation needs to be further addressed. GWAS studies have identified genetic loci that control the cfDNA fragmentation pattern, with some unexpected associations observed in genes encoding membrane channels, such as Pannexin-1, rather than nucleases, indicating biological mechanisms that are not yet recognized.

6.2Tissue-Specific Factors

This seems to suggest that certain tissues with high cellular turnover contribute disproportionately little cfDNA, beyond simple nuclease degradation via tissue-specific clearance mechanisms. Which contributes relatively-urinary clearance, direct nuclease degradation, and active cellular uptake by the reticuloendothelial system-remains unknown. Recent evidence, using anti-dsDNA priming agents that transiently inhibit cfDNA degradation, may provide new experimental approaches to answer these questions.

6 3.Differential Diagnosis of Benign Diseases

Although fragmentomics clearly separates cancer from healthy controls, further differentiation from benign disease sometimes has its limitations in various contexts. Patients with resected cystic tumors with symptoms or at risk of malignancy are associated with higher fragmentomic scores when compared to asymptomatic cystic lesions. However, it does need more large-scale validation toward establishing the thresholds for clear diagnostics.

7. FUTURE DIRECTIONS AND CLINICAL TRANSLATION

7.1Population-Specific Considerations

Thus far, most fragmentomics studies have assessed either mainly Asian or European populations. GWAS findings show population-specific associations with fragmentomic features, especiallyDNASE1L3variants that are much more frequent in Caucasian than East Asian populations. Future studies need to involve ethnic diversity if generalizability is desired and health equity is not to be reduced.

7.2Integration of Imaging with Clinical Data

Multimodal approaches, in which fragmentomics is combined with radiological imaging and clinical factors, are particularly promising for superior performance. Indeed, the study on lung cancer that combined fragmentomics with scores derived from CT imaging and clinical parameters reached an AUC of 0.92, approaching standards required for clinical implementation (Tsui et al., 2025). feature selection is bound to undergo further optimization in the interest of avoiding unnecessary complexity while retaining clinically relevant information.

7 3.Monitoring of Treatment and Minimal Residual Disease

Beyond screening, fragmentomics demonstrates growing utility in treatment monitoring and MRD detection. Tumor fraction estimation based on DELFI showed excellent correlation with mutant allele frequency (r=0.90) and the ability to detect treatment response before standard imaging in colorectal cancer patients (Tsui et al., 2025). This application can revolutionize cancer management by allowing earlier detection of treatment resistance. 11. Conclusion Fragmentomics of cfDNA has rapidly evolved from foundational studies to clinical validation of a robust biomarker platform for the noninvasive detection of cancer. The accumulating evidence from high-quality studies in multiple cancer types manifests both analytical validity and clinical utility that is challenging the current paradigm of cancer screening. Cost-effectiveness, high throughput capability, and high biological information content make this technology particularly well-suited for population-level cancer screening. Future clinical implementation should be done in the context of comprehensive, multimodal screening frameworks, while systematic research work is continued in order to fully detail mechanisms and optimize performance across diverse populations.

8. FUNDING

Naif Othman Alghamdi

CEO, Scientists kit, (Partial Fund Provider of Research)

REFERENCES

[1] Bai, J., Jiang, P., Ji, L., Lam, W.K.J., Zhou, Q., Ma, M.-J.L., ... & Lo, Y.M.D. (2024). Histone modifications of circulating nucleosomes are associated with changes in cell-free DNA fragmentation patterns. *Proceedings of the National Academy of Sciences USA*, 121(16), e2404058121.

[2] Budhraja, K.K., McDonald, B.R., Stephens, M.D., Contente-Cuomo, T., Markus, H., Farooq, M., ... & Velculescu, V.E. (2023). Genome-wide analysis of aberrant position and sequence of plasma DNA fragment ends in patients with cancer. *Science Translational Medicine*, 15(715), eabm6863.

[3] Chen, L., Wu, T., Fan, R., Qian, Y.-S., Liu, J.-F., Bai, J., ... & Chinese HCC-HBV Consortium. (2024). Cell-free DNA testing for early hepatocellular carcinoma surveillance. *EBioMedicine*, 100, 104962.

[4] Cristiano, S., Leal, A., Phallen, J., Fiksel, J., Adleff, V., Bruhm, D.C., ... & Velculescu, V.E. (2019). Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*, 570(7761), 385–389.

[5] Lam, W.K.J., Kang, G., Chan, C.M.L., Lee, V.C.T., Ma, M.-J.L., Zhou, Q., ... & Lo, Y.M.D. (2025). Fragmentomics profiling and quantification of plasma Epstein-Barr virus DNA enhance prediction of future nasopharyngeal carcinoma. *Cancer Cell*, 43(5), 728–739.

[6] Peneder, P., Stütz, A.M., Surdez, D., Krumbholz, M., Semper, S., Chicard, M., ... & Totzl, M. (2021). Multimodal analysis of cell-free DNA whole-genome sequencing for pediatric cancers with low mutational burden. *Nature Communications*, 12, 3230.

[7] Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M., & Shendure, J. (2016). Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell*, 164(1-2), 57–68.

[8] Tsui, W.H.A., Jiang, P., & Lo, Y.M.D. (2025). Cell-free DNA fragmentomics in cancer. *Cancer Cell*, 43(10), 1792–1814.
[9] Ulz, P., Perakis, S., Zhou, Q., Moser, T., Belic, J., Lazzeri, I., ... & Speicher, M.R. (2019). Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nature Communications*, 10, 4666.
[10] Wang, Y., Guo, Q., Huang, Z., Song, L., Zhao, F., Gu, T., ... & Li, B. (2025). Cell-free epigenomes enhanced fragmentomics-based model for early detection of lung cancer. *Clinical and Translational Medicine*, 15, e70225.
[11] Yin, L., Cao, C., Lin, J., Wang, Z., Peng, Y., Zhang, K., ... & Jiang, K. (2025). Development and validation of a cell-free DNA fragmentomics-based model for early detection of pancreatic cancer. *Journal of Clinical Oncology*, 43(26), 2863–2874.
[12] Zhou, Q., Kang, G., Jiang, P., Qiao, R., Lam, W.K.J., Yu, S.C.Y., ... & Lo, Y.M.D. (2022). Epigenetic analysis of cell-free DNA by fragmentomic profiling. *Proceedings of the National Academy of Sciences USA*, 119(23), e2209852119.