

DIAGNOSTIC PERFORMANCE OF NOVEL MULTIMODAL BIOMARKERS FOR EARLY CANCER DETECTION IN INDIA: AN EXPLORATORY DIAGNOSTIC ACCURACY STUDY

Dr. E. Praveen¹, Dr. Chittathur Vignesh², Dr. Srikanth Kodam³, Dr. R. Govindarajan⁴

¹Associate Professor, Department of Pathology, Rajalakshmi Medical College Hospital & Research Institute, The Tamil Nadu Dr. M.G.R. Medical University, Ponnalur, Tamil Nadu 602117, India, Email: praveen.easwar@gmail.com

²Associate Professor, Department of Pathology, Sri Lalithambigai Medical College and Hospital, Dr. M.G.R. Educational and Research Institute, Service Road, Maduravoyal, Adayalampattu, Chennai, Tamil Nadu 600095, India, Email: vikkyajj@gmail.com

³Associate Professor, Department of Pathology, Panimalar Medical College Hospital & Research Institute, Varadharajapuram, The Tamil Nadu Dr. M.G.R. Medical University, Poonamallee, Tamil Nadu 600123, India, Email: drsrikanthkodam@gmail.com

⁴Professor & Head, Department of Pathology, Shri Sathya Sai Medical College & Research Institute, Sri Balaji Vidyapeeth University (DU), Chennai Campus, Ammapettai, Nellikuppam Post, Chengalpattu District, Tamil Nadu 603108, India, Email: gjhospital@yahoo.com

*Corresponding author: Dr. R. Govindarajan | Email: gjhospital@yahoo.com

ABSTRACT

Background: Cancer screening coverage remains low in India, while blood- and body-fluid-based biomarkers could potentially complement organ-specific screening. This study evaluated the diagnostic performance of multiple novel biomarker domains for distinguishing cancer from non-cancer states in an Indian cohort, with a specific exploratory analysis of stage I-II disease.

Materials and Methods: An exploratory diagnostic accuracy study included 60 participants from 14 Indian states. Department of Pathology, Shri Sathya Sai Medical College & Research Institute, Sri Balaji Vidyapeeth University (DU), Chennai Campus, Ammapettai, Nellikuppam Post, Chengalpattu District, Tamil Nadu, India.

Study period was between July 2024- June 2026 Final diagnostic categories were cancer (n=15), high-risk (n=7), benign (n=22), and no evidence of cancer (n=16). Peripheral venous blood, saliva, and first-void urine were collected under a standardized pre-analytical SOP with barcoding and time-temperature monitoring. Nine systemic biomarker domains were evaluated: DNA methylation, fragmentomics, cell-free RNA (cfRNA), extracellular vesicles (EV), circulating tumour cells (CTC), tumour-educated platelets (TEP), protein biomarkers, autoantibodies, and metabolomics. Salivary and urinary HPV markers were assessed in relevant site-specific subsets. Molecular and cellular assays incorporated sequencing, RT-qPCR, immunoassay, microfluidic/cell-based enrichment, mass spectrometry, and methylation-based platforms. Receiver operating characteristic analysis compared cancer with all non-cancer categories; a separate exploratory analysis compared stage I-II cancers with non-cancer participants.

Results: The cohort had a mean age of 55.0 ± 7.3 years; 32 (53.3%) were male, and all 60 participants (100%) had no recorded prior cancer-screening history. Six of 15 cancers (40.0%) were stage I-II. Biomarker distributions differed significantly across diagnostic categories for all nine systemic domains (Kruskal-Wallis $p < 0.001$). For cancer versus non-cancer, AUCs ranged from 0.919 for TEP to 0.933 for CTC. Methylation, CTC, and metabolomics each achieved 100% specificity at their exploratory Youden-derived thresholds, with 86.7% sensitivity. In the stage I-II analysis, AUCs were lower, ranging from 0.798 to 0.833, with wide confidence intervals because only six early-stage cancers were available. All nine systemic biomarkers showed strong positive correlations with cancer stage (Spearman ρ 0.963-0.979, all $p < 0.001$). Tissue-of-origin classification was concordant with the recorded primary site category in all 15 cancer cases.

Conclusion: Multiple circulating biomarker domains showed strong discrimination between cancer and non-cancer states in this exploratory Indian cohort, but performance attenuated for stage I-II disease. The findings support multimodal biomarker development while also demonstrating the need for larger prospective, screening-based Indian validation cohorts with standardized assay methods, blinded reference standards, and external validation before clinical implementation.

KEYWORDS: early cancer detection; biomarkers; liquid biopsy; methylation; fragmentomics; cfRNA; circulating tumour cells; tumour-educated platelets; India

INTRODUCTION

India faces a rising cancer burden, with national registry-based estimates reporting more than 1.46 million incident cancers in 2022 and a continuing increase projected thereafter [1]. Yet early detection remains uneven. NFHS-5 analyses have shown very low uptake of oral, breast, and cervical cancer screening in the eligible population, underscoring the implementation gap between policy and routine practice [2]. This gap is especially

relevant where access to imaging, specialist examination, pathology, and follow-up is geographically fragmented.

Liquid-biopsy approaches are attractive because they can sample tumour-associated molecular signals through minimally invasive specimens and can, in principle, support repeated testing [3]. Cell-free DNA methylation has emerged as one of the most mature multi-cancer strategies [4]. Independent validation of methylation-based multi-cancer early detection has demonstrated high specificity but a pronounced stage gradient, with much lower sensitivity in stage I than in advanced disease [5]. Fragmentomic analysis provides complementary information from genome-wide cell-free DNA fragmentation patterns [6], and lung-cancer studies have shown that fragmentome-based classifiers can detect clinically relevant disease across stages [7].

Other biomarker classes may capture distinct dimensions of tumour biology. Cell-free RNA can reflect early transcriptional alterations [8]. Extracellular vesicles contain proteins and nucleic acids released from tumour and host cells [9]. Circulating tumour cells are biologically linked to dissemination and may identify clinically significant disease, although their rarity remains a technical challenge [10]. Tumour-educated platelet RNA profiles have also shown pan-cancer diagnostic potential [11]. Multi-analyte strategies that combine molecular alterations with circulating proteins have further demonstrated the feasibility of integrating heterogeneous signals for cancer detection [12]. Autoantibodies can amplify immune recognition of tumour-associated antigens [13], while metabolomic signatures may capture downstream metabolic reprogramming [14]. In India, saliva is particularly relevant for oral cancer detection because collection is inexpensive and non-invasive [15], whereas self-collected HPV testing could expand cervical-screening reach in women who are underserved by facility-based programmes [16].

The present study therefore examined the diagnostic behaviour of a broad multimodal biomarker panel in an Indian cohort spanning cancer, high-risk, benign, and no-evidence categories. The primary objective was to estimate the discrimination of individual systemic biomarker domains for cancer versus non-cancer. Secondary objectives were to assess performance for stage I-II cancer, examine stage-related biomarker gradients, describe tissue-of-origin concordance, and summarize site-specific saliva and urinary HPV findings.

MATERIALS AND METHODS

Study design and setting

This was an exploratory diagnostic accuracy study conducted through the Department of Pathology, Shri Sathya Sai Medical College & Research Institute, Sri Balaji Vidyapeeth University (DU), Chennai Campus, Ammapettai, Nellikuppam Post, Chengalpattu District, Tamil Nadu, India. Study period was between July 2024–June 2026. Reporting was structured with reference to the STARD 2015 principles for diagnostic accuracy studies [17]. The study included 60 participants representing 14 Indian states.

Study population and diagnostic classification

Participants were classified into four final diagnostic categories: cancer (n=15), high-risk (n=7), benign (n=22), and no evidence of cancer (n=16). All 60 participants had no recorded prior cancer screening. The cancer group included oral (n=3), cervical (n=3), breast (n=3), lung (n=2), stomach (n=1), liver (n=1), colorectal (n=1), and bladder (n=1) primary sites. Cancer stage distribution was stage I in 2 cases, stage II in 4, stage III in 5, and stage IV in 4. Histology was recorded as confirmed for all 15 cancers. Histological confirmation was not recorded for the non-cancer categories, which were retained according to their final study classification.

Biomarker assessment

Nine systemic biomarker domains were evaluated: methylation, fragmentomics, cfRNA, EV, CTC, TEP, protein, autoantibody, and metabolomics. Methylation, fragmentomics, cfRNA, EV, TEP, protein, autoantibody, and metabolomics were represented by continuous recorded study scores; CTC was represented as a count. Salivary biomarker values were available for participants evaluated for oral lesions, and urinary HPV results were available for cervical evaluation and selected no-evidence participants.

Specimen collection and pre-analytical processing

Peripheral venous blood (10 mL) was collected into EDTA tubes. Plasma was separated within 2 hours using two-step centrifugation ($1,600 \times g$ for 10 minutes at 4°C , followed by $16,000 \times g$ for 10 minutes at 4°C), aliquoted, and stored at -80°C until analysis. Saliva was collected by passive drool, centrifuged at $10,000 \times g$ for 10 minutes, and the supernatant was stored at -80°C . First-void urine was collected with preservative, centrifuged, and stored at -80°C . All specimens were barcoded, and collection, processing, storage, and transfer were documented using time-temperature logs under a standardized pre-analytical SOP.

Molecular and cellular biomarker assays

DNA methylation: Cell-free DNA was extracted from plasma using the QIAamp Circulating Nucleic Acid Kit and subjected to bisulfite conversion with the EZ DNA Methylation-Gold kit. Methylation profiling was performed using targeted methylation sequencing on an Illumina NovaSeq 6000 with a custom pan-cancer methylation panel or, where genome-wide profiling was used, the Illumina MethylationEPIC array.

Fragmentomics: Low-pass whole-genome sequencing at approximately $0.5\text{--}1\times$ coverage was performed on an Illumina NovaSeq 6000. Fragment-size distribution, fragment-end motifs, and genomic coverage features were derived using a validated bioinformatics pipeline.

Cell-free RNA: Plasma cfRNA was extracted using the miRNeasy Serum/Plasma Kit. Small-RNA sequencing was performed on an Illumina NextSeq 2000, or selected microRNAs were quantified by RT-qPCR, with normalization to spike-in controls.

Extracellular vesicles: EVs were isolated using size-exclusion chromatography (qEV) or ultracentrifugation. Vesicle preparations were characterized by nanoparticle tracking analysis (NanoSight), transmission electron microscopy, and Western blotting for CD9, CD63, and TSG101. EV cargo was subsequently evaluated by mass spectrometry or RNA sequencing, as appropriate.

Circulating tumour cells: CTCs were enriched using the CellSearch system or a microfluidic enrichment chip and enumerated by immunofluorescence using the CK+/DAPI+/CD45⁻ phenotype. Molecular characterization, where performed, used targeted sequencing.

Tumour-educated platelets: Platelets were isolated from platelet-rich plasma, followed by RNA extraction and platelet RNA sequencing on an Illumina NovaSeq platform. Tumour-educated platelet signatures were interpreted using a locked classifier.

Protein biomarkers and autoantibodies: Protein panels were measured using multiplex immunoassays based on Luminex xMAP or Olink proximity extension technology. Autoantibodies were assessed by ELISA or protein microarray, with Western blot used for validation where required.

Metabolomics: Metabolic profiling incorporated LC-MS/MS using reverse-phase and hydrophilic interaction liquid chromatography, GC-MS for suitable polar metabolites, and NMR for targeted analyses. Data preprocessing and feature analysis used XCMS and MetaboAnalyst-based workflows.

Site-specific saliva and urinary HPV assays

Salivary biomarkers: Salivary IL-6, IL-8, and MMP-9 were measured using multiplex ELISA. Salivary RNA targets were quantified by RT-qPCR, and protein measurements were normalized to total salivary protein.

Urinary HPV testing: First-void urine was processed for HPV DNA detection by PCR-based testing, including Cobas HPV where applicable, with genotyping. Host and viral methylation markers were assessed by bisulfite pyrosequencing or methylation-specific PCR.

Reference standard and outcomes

The primary reference classification was the final study diagnosis. The principal diagnostic comparison treated cancer as the target condition and combined high-risk, benign, and no-evidence categories as non-cancer. The early-stage exploratory comparison treated stage I-II cancers as the target condition and excluded stage III-IV cancers. Tissue-of-origin prediction was compared with the recorded primary-site category, with stomach and colorectal cancers mapped to the gastrointestinal category.

Statistical analysis

Continuous variables were summarized as mean $\pm$ standard deviation, and categorical variables as number and percentage. Differences in biomarker distributions across the four diagnostic categories were tested using the Kruskal-Wallis test because of small and unequal group sizes. Diagnostic performance was assessed using receiver operating characteristic curves and area under the curve (AUC). Ninety-five percent confidence intervals for AUC were estimated by non-parametric bootstrap resampling. Exploratory cut-offs were selected using the maximum Youden J statistic, and sensitivity and specificity were calculated at those thresholds. Among cancer cases, ordinal stage (I-IV) was correlated with each systemic biomarker using Spearman rank correlation. All tests were two-sided and $p < 0.05$ was considered statistically significant. Because this was an exploratory study with a small sample, p -values were interpreted alongside effect size and confidence intervals rather than as stand-alone evidence.

RESULTS

Participant characteristics

The 60 participants had a mean age of 55.0 ± 7.3 years and mean BMI of 25.2 ± 3.2 kg/m². Thirty-two (53.3%) were male, and the same 32 participants (53.3%) had a recorded smoking history, 26 (43.3%) had smokeless-tobacco exposure, and 13 (21.7%) reported alcohol use. Thirty-two (53.3%) were symptomatic at presentation. In the recorded cohort, smoking history co-segregated completely with sex, accounting for the identical sex and smoking counts across diagnostic categories. All 60 participants (100%) had no recorded prior cancer-screening history. Diagnostic categories were cancer in 15 (25.0%), high-risk in 7 (11.7%), benign in 22 (36.7%), and no evidence of cancer in 16 (26.7%). Baseline characteristics differed across the four categories, reflecting the heterogeneous clinical composition of the cohort (Table 1).

Table 1. Baseline characteristics of the study participants by diagnostic category

Characteristic	Overall (n=60)	Cancer (n=15)	High-risk (n=7)	Benign (n=22)	No evidence (n=16)	p value
Age, years	55.0 $\pm$ 7.3	55.3 $\pm$ 7.6	53.3 $\pm$ 6.1	59.9 $\pm$ 5.5	48.6 $\pm$ 4.1	<0.001
BMI, kg/m ²	25.2 $\pm$ 3.2	24.7 $\pm$ 3.2	25.4 $\pm$ 2.5	23.1 $\pm$ 2.3	28.6 $\pm$ 1.1	<0.001
Male sex	32 (53.3%)	9 (60.0%)	3 (42.9%)	20 (90.9%)	0 (0.0%)	<0.001

Smoking history	32 (53.3%)	9 (60.0%)	3 (42.9%)	20 (90.9%)	0 (0.0%)	<0.001
Smokeless-tobacco exposure	26 (43.3%)	8 (53.3%)	3 (42.9%)	15 (68.2%)	0 (0.0%)	<0.001
Alcohol use	13 (21.7%)	4 (26.7%)	0 (0.0%)	9 (40.9%)	0 (0.0%)	0.010
Symptomatic at presentation	32 (53.3%)	9 (60.0%)	3 (42.9%)	20 (90.9%)	0 (0.0%)	<0.001
Prior cancer screening, yes	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	Not estimable

Values are mean $\pm$ SD or n (%). Kruskal-Wallis test was used for continuous variables; chi-square test was used for categorical variables. No inferential test was performed for prior cancer screening because all 60 participants had the same recorded status. Smoking history co-segregated completely with sex in this cohort: all 32 male participants had a recorded smoking history and all 28 female participants did not. These comparisons describe cohort heterogeneity and are not causal analyses.

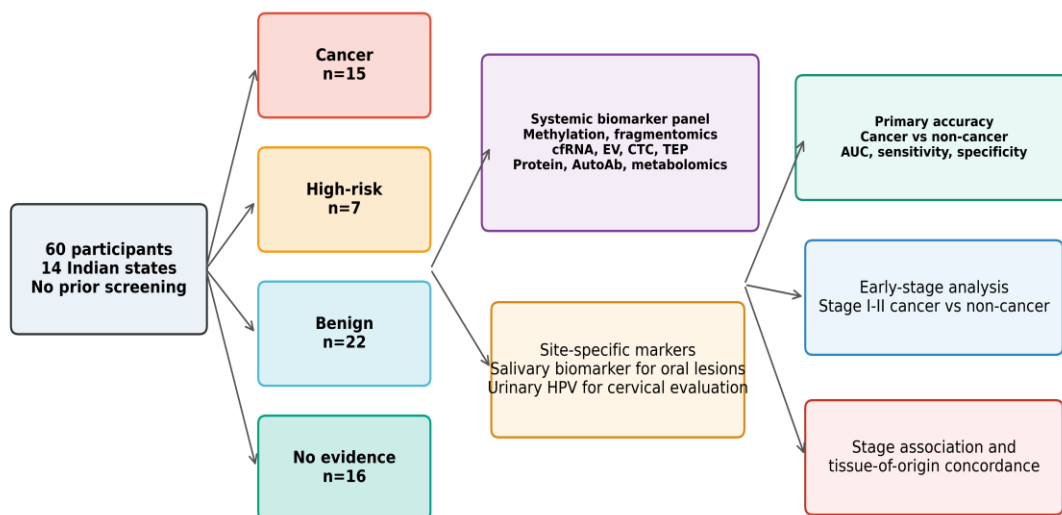


Figure 1. Analytical framework of the exploratory multimodal biomarker study.

The figure summarizes the recorded participant groups, systemic and site-specific biomarker domains, and the prespecified analytic comparisons used in this manuscript.

Biomarker distributions across diagnostic categories

All nine systemic biomarker domains differed significantly across the four diagnostic categories (all Kruskal-Wallis $p < 0.001$). For the score-based markers, mean values were consistently highest in cancer and lowest in the no-evidence group. Mean methylation score, for example, was 67.9 ± 16.5 in cancer compared with 22.0 ± 2.4 in the no-evidence group. CTC count averaged 2.3 ± 1.7 in cancer and was zero in all non-cancer categories (Table 2). The score separation across categories is visualized in Figure 2.

Table 2. Distribution of systemic biomarker measurements across diagnostic categories

Biomarker	Cancer	High-risk	Benign	No evidence	Kruskal-Wallis H	p value
Methylation	67.9 ± 16.5	43.9 ± 9.3	46.0 ± 6.1	22.0 ± 2.4	43.98	<0.001
Fragmentomics	64.7 ± 16.4	41.7 ± 9.4	44.0 ± 6.1	20.0 ± 2.4	43.92	<0.001
cfRNA	60.7 ± 15.7	39.3 ± 8.8	41.0 ± 5.9	18.0 ± 2.4	43.81	<0.001
EV	62.5 ± 15.9	40.7 ± 9.4	43.0 ± 6.1	19.0 ± 2.4	43.87	<0.001

CTC	2.3 ± 1.7	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	48.05	<0.001
TEP	62.9 ± 18.3	39.7 ± 9.4	41.9 ± 6.4	18.0 ± 2.4	43.47	<0.001
Protein	64.0 ± 14.9	42.9 ± 9.3	45.0 ± 6.1	21.0 ± 2.4	43.92	<0.001
AutoAb	58.8 ± 14.5	38.3 ± 8.8	40.0 ± 6.1	17.0 ± 2.4	43.69	<0.001
Metabolomics	64.3 ± 15.2	41.9 ± 9.3	44.0 ± 6.1	20.0 ± 2.4	44.04	<0.001

Values are mean ± SD on the recorded study scale, except CTC, which is reported as an enumerated cell count.

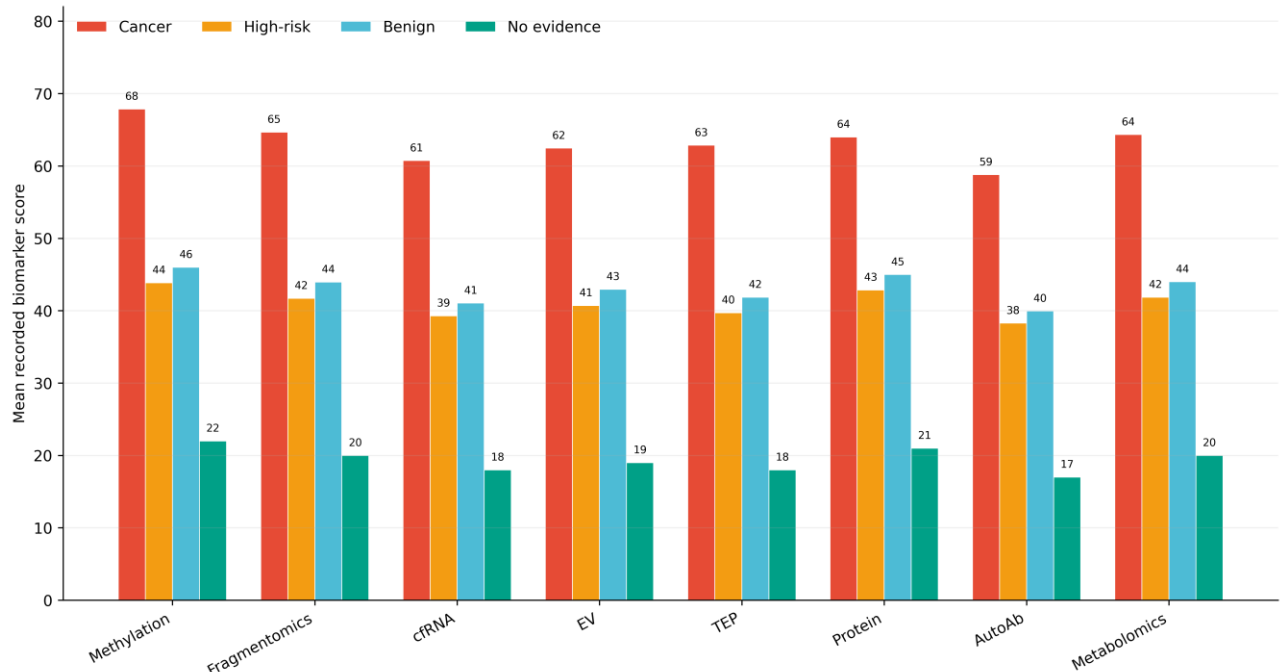


Figure 2. Mean score-based biomarker measurements across the four diagnostic categories.

CTC is omitted from this figure because it is recorded as a count on a different scale. Numerical values are displayed above each bar.

Diagnostic accuracy for cancer versus non-cancer

All systemic biomarkers showed strong discrimination for cancer versus the combined non-cancer group. AUCs were closely clustered between 0.919 and 0.933. CTC had the numerically highest AUC at 0.933 (95% CI 0.833-1.000), followed by metabolomics at 0.927 (95% CI 0.816-1.000) and methylation at 0.926 (95% CI 0.812-1.000). These small numerical differences were not subjected to pairwise AUC comparison and should not be interpreted as a ranking of definitive clinical superiority. At exploratory Youden-derived thresholds, methylation, CTC, and metabolomics each yielded sensitivity of 86.7% with specificity of 100.0%. Fragmentomics, cfRNA, EV, and protein each showed 86.7% sensitivity with 97.8% specificity (Table 3).

Table 3. Receiver operating characteristic performance of systemic biomarkers

Biomarker	AUC, all cancer (95% CI)	Exploratory cut-off	Sensitivity	Specificity	AUC, stage I-II (95% CI)	Early-stage sensitivity	Early-stage specificity
Methylation	0.926 (0.812-1.000)	56.0	86.7%	100.0%	0.815 (0.585-1.000)	66.7%	100.0%
Fragmentomics	0.925 (0.821-1.000)	53.0	86.7%	97.8%	0.813 (0.578-1.000)	66.7%	97.8%
cfRNA	0.924 (0.815-	50.0	86.7%	97.8%	0.809 (0.567-	66.7%	97.8%

	1.000)				1.000)		
EV	0.924 (0.813- 1.000)	52.0	86.7%	97.8%	0.811 (0.570- 1.000)	66.7%	97.8%
CTC	0.933 (0.833- 1.000)	1.0	86.7%	100.0%	0.833 (0.667- 1.000)	66.7%	100.0%
TEP	0.919 (0.803- 0.999)	49.0	86.7%	93.3%	0.798 (0.554- 0.985)	66.7%	93.3%
Protein	0.925 (0.807- 1.000)	54.0	86.7%	97.8%	0.813 (0.576- 1.000)	66.7%	97.8%
AutoAb	0.922 (0.808- 1.000)	48.0	86.7%	95.6%	0.806 (0.556- 1.000)	66.7%	95.6%
Metabolomics	0.927 (0.816- 1.000)	54.0	86.7%	100.0%	0.817 (0.580- 1.000)	66.7%	100.0%

Cancer analysis: 15 cancer versus 45 non-cancer participants. Early-stage analysis: 6 stage I-II cancers versus 45 non-cancer participants; stage III-IV cancers were excluded. Cut-offs were selected by Youden J within the same sample and are therefore exploratory, not validated clinical thresholds. Bootstrap 95% CIs are shown for AUC.

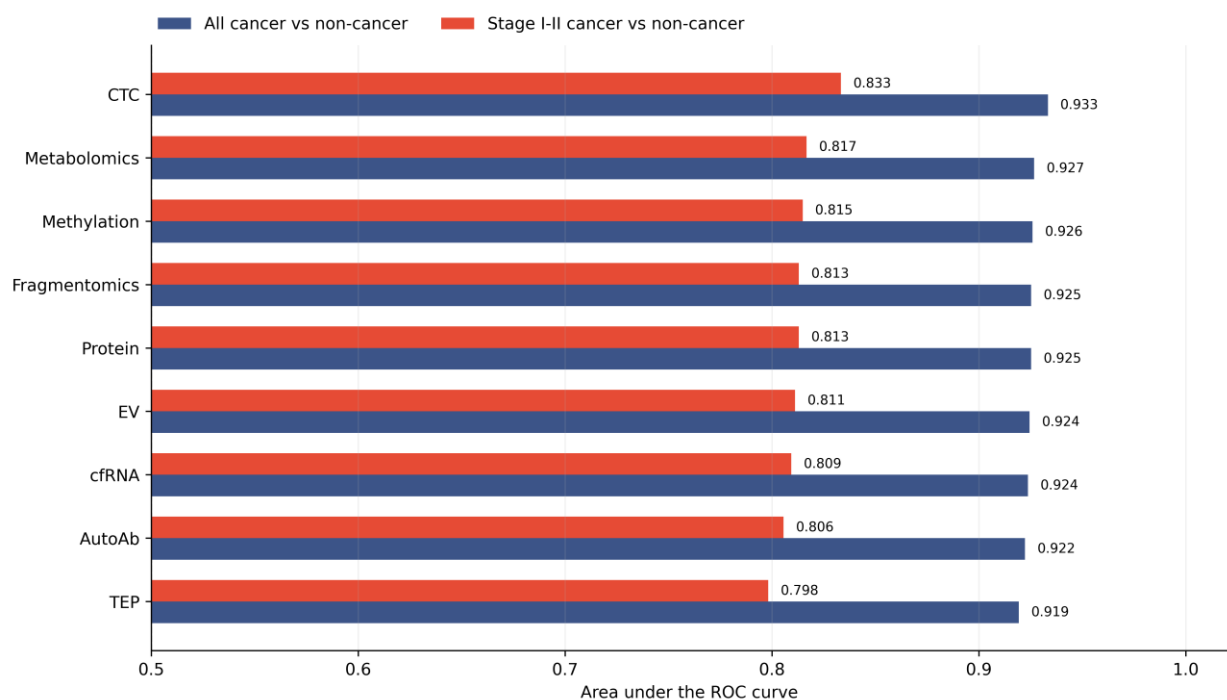


Figure 3. AUC comparison for all-cancer detection and stage I-II cancer detection.

The decrease in AUC for stage I-II disease illustrates the expected challenge of detecting lower tumour-burden states. Early-stage confidence intervals are wide because only six stage I-II cancers were present.

Stage-related biomarker gradients

Among the 15 cancer cases, all systemic biomarker measurements increased with advancing stage. Spearman correlations between ordinal stage and biomarker values ranged from $\rho=0.963$ to $\rho=0.979$, with all $p<0.001$. The strongest correlation was observed for CTC ($\rho=0.979$). For methylation, mean score increased from 36.5 ± 3.5 in stage I to 85.5 ± 2.1 in stage IV. Similar monotonic gradients were evident for fragmentomics, cfRNA, EV, TEP, protein, autoantibody, and metabolomics (Table 4 and Figure 4).

Table 4. Stage-specific systemic biomarker measurements among histologically confirmed cancers

Biomarker	Stage I (n=2)	Stage II (n=4)	Stage III (n=5)	Stage IV (n=4)	Spearman rho	p value
Methylation	36.5 ± 3.5	58.5 ± 2.1	73.8 ± 3.5	85.5 ± 2.1	0.963	<0.001
Fragmentomics	33.5 ± 3.5	55.5 ± 2.1	70.4 ± 3.0	82.2 ± 1.7	0.963	<0.001
cfRNA	30.5 ± 3.5	52.5 ± 2.1	65.8 ± 2.6	77.8 ± 1.9	0.964	<0.001
EV	32.5 ± 3.5	53.2 ± 1.5	67.8 ± 2.3	80.0 ± 1.4	0.965	<0.001
CTC	0.0 ± 0.0	1.0 ± 0.0	2.4 ± 0.5	4.5 ± 0.6	0.979	<0.001
TEP	30.5 ± 4.9	50.5 ± 1.3	69.2 ± 3.2	83.5 ± 2.1	0.963	<0.001
Protein	35.0 ± 4.2	56.0 ± 1.8	69.6 ± 2.7	79.5 ± 2.1	0.963	<0.001
AutoAb	30.5 ± 2.1	51.5 ± 2.6	63.6 ± 2.1	74.2 ± 1.7	0.963	<0.001
Metabolomics	35.0 ± 2.8	56.0 ± 1.4	69.8 ± 1.9	80.5 ± 1.3	0.964	<0.001

Values are mean ± SD. Stage was treated as an ordinal variable for Spearman correlation. Small stage-specific sample sizes make these estimates exploratory.

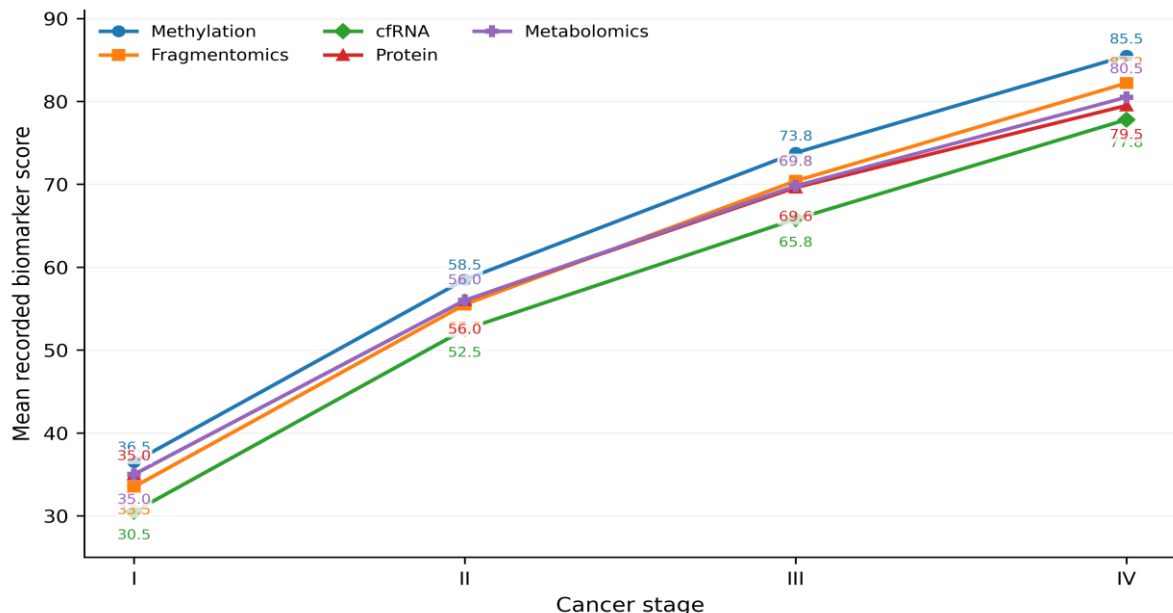


Figure 4. Stage-wise progression of selected score-based biomarkers among cancer cases.

Point labels show mean recorded scores within each stage. Only representative score-based biomarkers are plotted to maintain legibility; the complete stage-specific results appear in Table 4.

Tissue-of-origin and site-specific markers

Tissue-of-origin prediction was concordant with the recorded primary-site category in all 15 cancer cases when stomach and colorectal sites were mapped to a gastrointestinal category. Because classifier development details and external validation were not available, this 100% concordance should be considered descriptive rather than an independent estimate of generalizable accuracy.

Salivary biomarker values were recorded for 10 participants with oral lesions. Mean salivary values were 82.0 ± 6.0 in oral cancer ($n=3$), 62.5 ± 2.1 in the high-risk oral group ($n=2$), and 54.6 ± 4.3 in benign oral lesions ($n=5$). Urinary HPV results were available in 11 participants: all three cervical cancer cases and both high-risk cervical cases were positive, whereas one benign cervical case and five no-evidence participants were negative. These small site-specific subsets are hypothesis-generating and were not used to claim population-level screening accuracy.

DISCUSSION

Principal findings

This exploratory study shows a coherent multimodal signal across nine systemic biomarker domains. Cancer cases had markedly higher values than no-evidence participants, while high-risk and benign groups occupied

intermediate ranges. Overall cancer discrimination was strong across all domains, with AUC values around 0.92-0.93. However, a more clinically demanding analysis, restricted to stage I-II cancers, reduced AUCs to approximately 0.80-0.83. This stage-related attenuation is central to interpretation. A biomarker can distinguish established cancer from non-cancer very well and still be substantially less sensitive at the point where screening is intended to intervene.

Interpretation in relation to contemporary biomarker research

The stage gradient observed here is biologically consistent with large methylation-based studies. Liu and colleagues demonstrated that cfDNA methylation can detect and localize multiple cancers at high specificity, but sensitivity rose substantially with stage [4]. Klein and colleagues similarly reported stage I sensitivity far below stage IV sensitivity in an independent validation cohort [5]. Our own early-stage analysis follows the same direction, although the present sample is far smaller and cannot provide comparable population-level estimates.

Fragmentomics is conceptually attractive because it does not depend solely on the presence of a small number of tumour-specific mutations. Genome-wide fragmentation patterns have shown cancer-associated signals across tumour types [6], and clinically oriented lung studies have demonstrated the value of integrating fragmentomic patterns with other risk information [7]. In the present cohort, fragmentomics had an AUC of 0.925 for all cancer and 0.813 for stage I-II disease, again highlighting both promise and stage dependence.

cfRNA, EV, CTC, and TEP measurements represent complementary biology rather than interchangeable tests. cfRNA can capture transcriptional activity and has shown early-stage classification potential when sensitive sequencing methods are used [8]. EV cargo can carry tumour- and host-derived proteins and nucleic acids [9]. More recently, a plasma extracellular-vesicle protein signature demonstrated high performance across retrospective multi-cancer samples and prospective early lung-cancer evaluation, indicating continuing technical maturation of this field [18]. CTCs may identify biologically aggressive disease but are uncommon at very early stages [10], a limitation mirrored by the absence of CTCs in both stage I cancers in the present cohort. TEP profiling has demonstrated pan-cancer and tissue-classification potential through platelet RNA signatures [11].

The strong performance of proteins, autoantibodies, and metabolomics in this study also supports a multimodal view of early detection. CancerSEEK established the principle that combining circulating proteins with cfDNA can improve multi-cancer detection [12]. Autoantibodies may provide signal amplification because host immunity can respond to low-abundance tumour antigens [13], whereas metabolomics captures downstream pathway changes that may be orthogonal to genomic markers [14]. These layers could be especially useful when ctDNA abundance is low, but platform harmonization and independent validation remain essential.

Relevance to India

The Indian context changes the practical meaning of biomarker innovation. The national cancer burden is large and increasing [1], yet screening uptake remains low [2]. A future biomarker strategy therefore has to be evaluated not only by AUC but also by cost per person screened, specimen stability, laboratory throughput, false-positive referral burden, availability of confirmatory imaging and pathology, and the feasibility of serving rural and semi-urban populations. A technically sophisticated test that generates referrals without diagnostic capacity could shift rather than solve the access problem.

Site-specific markers may offer pragmatic entry points. Saliva is particularly attractive for oral-cancer pathways because collection is simple, repeatable, and anatomically close to many lesions [15]. Recent evidence syntheses have identified several salivary candidates with encouraging diagnostic performance but also substantial heterogeneity and a need for larger pragmatic validation studies [19]. For cervical screening, Indian studies of HPV self-sampling have shown high agreement with clinician-collected samples and strong acceptability [16], making urine- or self-collected HPV approaches relevant where pelvic examination is a barrier.

Clinical caution: diagnostic accuracy is not equivalent to screening benefit

The distinction between test accuracy and population benefit has become especially important after publication of the randomized NHS-Galleri trial. In more than 142,000 participants, adding a multicancer early-detection blood test to usual care did not significantly reduce the primary endpoint of stage III-IV cancer incidence across 12 prespecified cancers after three annual screening rounds, although stage IV findings were directionally different in a secondary endpoint [20]. This trial demonstrates why promising case-control AUCs, including those in the present study, should not be interpreted as proof that a biomarker programme will reduce late-stage disease or mortality. Prospective screening trials with clinical outcomes remain necessary.

Strengths and limitations

A strength of this study is the simultaneous examination of multiple biomarker classes across cancer, high-risk, benign, and no-evidence categories rather than a simple cancer-versus-healthy comparison. The inclusion of site-specific saliva and urinary HPV variables also reflects modalities that are relevant to Indian cancer-control priorities. Analytically, the manuscript separates overall cancer detection from the more demanding stage I-II task and presents confidence intervals rather than point estimates alone.

Several limitations are substantial. First, the sample size was only 60, with 15 cancers and only six stage I-II cancers, producing wide early-stage confidence intervals. Second, exploratory cut-offs were derived and evaluated in the same cohort, so sensitivity and specificity are optimistically biased until externally validated. Third, all cancer cases were histologically confirmed, whereas non-cancer categories did not have uniform

histologic verification, introducing potential differential-verification bias. Fourth, some biomarker domains could be measured through more than one validated analytical route, creating potential platform-related heterogeneity despite standardized pre-analytical handling. Fifth, baseline demographic and exposure profiles differed sharply among diagnostic groups, so confounding cannot be excluded. Sixth, the tissue-of-origin result should be interpreted cautiously because the classifier-development process was not independently evaluated. Finally, this was not a population-screening cohort; the findings therefore estimate diagnostic discrimination within the study sample, not screening effectiveness in asymptomatic Indian adults.

Implications for future research

The next step should be a prospective, multicentre Indian study that recruits participants before definitive diagnosis and deliberately enriches stage I-II disease without creating case-control spectrum bias. Pre-analytics should be standardized across centres, biomarker thresholds should be locked before validation, laboratory readers should be blinded to the reference diagnosis, and all positive and a representative sample of negative participants should receive a sufficiently rigorous reference standard. A training cohort should be separated from an independent validation cohort. In addition to AUC, future work should report positive predictive value at realistic screening prevalence, number needed to investigate, interval cancers, diagnostic-resolution time, cost per cancer detected, stage shift, and patient-reported burden.

CONCLUSION

This exploratory Indian study found strong cancer-versus-non-cancer discrimination across methylation, fragmentomics, cfRNA, EV, CTC, TEP, protein, autoantibody, and metabolomic biomarkers. The common reduction in performance for stage I-II disease is a clinically important finding and argues against relying on overall AUC alone. Multimodal biomarker strategies remain promising for India, particularly if they can be integrated with oral and cervical screening pathways, but clinical adoption should await standardized laboratory methods, blinded external validation, and prospective evidence that testing improves the stage and timeliness of cancer diagnosis without creating disproportionate false-positive burden.

DECLARATIONS

Funding: NIL

Conflict of interest: none

Data availability: De-identified participant-level data underlying the analyses are available from the corresponding author subject to institutional requirements and applicable data-governance policies.

Author contributions: Dr. R. Govindarajan: conceptualization, clinical interpretation, pathology oversight, manuscript review and approval.

REFERENCES

1. Sathishkumar K, Chaturvedi M, Das P, Stephen S, Mathur P. Cancer incidence estimates for 2022 & projection for 2025: Result from National Cancer Registry Programme, India. *Indian J Med Res.* 2022 Oct-Nov;156(4&5):598-607. doi:10.4103/ijmr.ijmr_1821_22.
2. Patel KK, Chakravarti P, Chaturvedi P, Dikshit R, Budukh A. The fifth round of the National Family Health Survey of India 2019 to 2021 reported low screening uptake alarming to strengthen the implementation of early detection services of the cervix, breast and oral cancer. *Int J Cancer.* 2022;150(10):1734-1736. doi:10.1002/ijc.33958.
3. Zhang Q, Zhang X, Xie P, Zhang W. Liquid biopsy: an arsenal for tumour screening and early diagnosis. *Cancer Treat Rev.* 2024;129:102774. doi:10.1016/j.ctrv.2024.102774.
4. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV; CCGA Consortium. 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.
5. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol.* 2021;32(9):1167-1177. doi:10.1016/j.annonc.2021.05.806.
6. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature.* 2019;570(7761):385-389. doi:10.1038/s41586-019-1272-6.
7. Mathios D, Johansen JS, Cristiano S, Medina JE, Phallen J, Larsen KR, et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nat Commun.* 2021;12(1):5060. doi:10.1038/s41467-021-24994-w.
8. Wang J, Huang J, Hu Y, Guo Q, Zhang S, Tian J, et al. Terminal modifications independent cell-free RNA sequencing enables sensitive early cancer detection and classification. *Nat Commun.* 2024;15(1):156. doi:10.1038/s41467-023-44461-y.
9. De Giorgis V, Barberis E, Manfredi M. Extracellular vesicles proteins for early cancer diagnosis: from omics to biomarkers. *Semin Cancer Biol.* 2024;104-105:18-31. doi:10.1016/j.semcancer.2024.07.003.
10. Lawrence R, Watters M, Davies CR, Pantel K, Lu YJ. Circulating tumour cells for early detection of clinically relevant cancer. *Nat Rev Clin Oncol.* 2023;20(7):487-500. doi:10.1038/s41571-023-00781-y.

11. Best MG, Sol N, Kooi I, Tannous J, Westerman BA, Rustenburg F, et al. RNA-Seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell*. 2015;28(5):666-676. doi:10.1016/j.ccell.2015.09.018.
12. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018;359(6378):926-930. doi:10.1126/science.aar3247.
13. Montero-Calle A, Garranzo-Asensio M, Moreno-Casbas MT, Campuzano S, Barderas R. Autoantibodies in cancer: a systematic review of their clinical role in the most prevalent cancers. *Front Immunol*. 2024;15:1455602. doi:10.3389/fimmu.2024.1455602.
14. Bernhardt GV, Bernhardt KL, Pinto JRT, Vashe A. Metabolomics in cancer detection: a review of techniques, biomarkers, and clinical utility. *Biomedicine (Taipei)*. 2025;15(3):1-23. doi:10.37796/2211-8039.1665.
15. Vats R, Yadav P, Bano A, Wadhwa S, Bhardwaj R. Salivary biomarkers in non-invasive oral cancer diagnostics: a comprehensive review. *J Appl Oral Sci*. 2024;32:e20240151. doi:10.1590/1678-7757-2024-0151.
16. Joshi S, Kumar BK, Ramshankar V, Maitra N, Palit SP, Ravindran S, et al. Performance of HPV self-sample collected by a novel kit in comparison with clinician collected sample for cervical cancer screening. *Asian Pac J Cancer Prev*. 2024;25(12):4211-4216. doi:10.31557/APJCP.2024.25.12.4211.
17. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527. doi:10.1136/bmj.h5527.
18. Lobb RJ, Zhou Q, Fielding D, Visan KS, Wuethrich A, Wang J, et al. Early-stage multi-cancer detection through a plasma extracellular vesicle protein signature. *Cell Rep Med*. 2026;7(4):102694. doi:10.1016/j.xcrm.2026.102694.
19. Khijmatgar S, Yong J, Rübsamen N, Lorusso F, Rai P, Cenzato N, et al. Salivary biomarkers for early detection of oral squamous cell carcinoma and head/neck squamous cell carcinoma: a systematic review and network meta-analysis. *Jpn Dent Sci Rev*. 2024;60:32-39. doi:10.1016/j.jdsr.2023.10.003.
20. Sasieni P, Johnson P, Round T, Warwick J, Jones H, Dolly S, et al. Effect of screening with multicancer early-detection test on late-stage cancer diagnosis. *N Engl J Med*. 2026 Sep 22. doi:10.1056/NEJMoa2505723. Online ahead of print.