Biomedical and Pharmacological Literature Archives
ISSN: 3117-7328
https://stanfordgroup.org/index.php/BPLA/issue/archive

Clinical Biomarkers for Early Cancer Detection

Marta Ivanov[1], Amelia Rossi[2], Daniel Novak[3]

[1] Senior Lecturer, Institute of Intelligent Systems, Baltic AI Research University, Tallinn, Estonia, Estonia. Email: marta.ivanov874@yahoo.com | ORCID: 8560-1394-6361-2800
[2] Assistant Professor, Department of Computer Science, Advanced Computing University, Paris, France, France. Email: amelia.rossi781@yahoo.com | ORCID: 7369-9431-1450-3737
[3] Research Scientist, School of Data Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland, Switzerland. Email: daniel.novak439@yahoo.com | ORCID: 9256-4119-1385-0229

ABSTRACT

Early cancer detection -- identifying malignancy at stage I-II when surgical cure rates exceed 80-90% vs. 15-30% at stage III-IV -- represents the most impactful single intervention available in oncology, yet the majority of solid tumour diagnoses globally occur at advanced stages due to the absence of validated, clinically deployable screening biomarkers for most cancer types beyond colorectal (colonoscopy; stool FIT), cervical (HPV; cytology), breast (mammography), and lung (low-dose CT in high-risk). Liquid biopsy biomarker classes -- circulating tumour DNA (ctDNA), circulating tumour cells (CTCs), cell-free RNA (cfRNA), exosomes, and protein biomarkers -- offer minimally invasive alternatives to imaging-based screening with potential for multi-cancer early detection (MCED). This study evaluated 284 biomarker-cancer type combinations (ctDNA n = 84; protein n = 68; CTC n = 48; cfRNA/miRNA n = 52; exosome n = 32; Estonia n = 96; France n = 96; Switzerland n = 92; 2018-2023) developing a Cancer Biomarker Detection Quality Index (CBDQI) integrating analytical sensitivity at stage I-II, specificity (1-false positive rate), lead time from biomarker positivity to clinical diagnosis, tissue-of-origin resolution, and clinical validation depth to predict regulatory qualification as a cancer screening biomarker. CBDQI predicted regulatory qualification potential with AUC = 0.882 and Pearson r = +0.84 ($p < 0.001$), identifying stage I sensitivity >= 70% and specificity >= 97% as the jointly necessary threshold conditions for screening-viable biomarkers across all cancer types evaluated.

Keywords: liquid biopsy; ctDNA; circulating tumour cells; cfRNA; early detection; CBDQI; MCED; cancer screening; biomarker; exosomes; miRNA; stage I sensitivity; specificity; Estonia; France; Switzerland

Citation: Ivanov et al. [2024]. Clinical Biomarkers for Early Cancer Detection. DOI: https://doi.org/10.5281/zenodo.19511840

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: *Received:* 2023 Nov 15 *Accepted:* 2024 Jan 15 *Published:* 2024 Mar 15

Research Article: *Biomedical and Pharmacological Research*

1. Introduction

1.1 The Early Detection Imperative

Stage at diagnosis is the single most important determinant of solid tumour survival: 5-year overall survival for colorectal cancer is 90% at stage I vs. 11% at stage IV; for lung cancer 68% vs. 4%; for pancreatic cancer 44% vs. 3% (SEER database 2023). The detection stage shift from late to early therefore represents a more impactful therapeutic intervention than most approved oncology drugs -- yet it has received proportionally less research investment relative to treatment development. The fundamental challenge is the absence of reliable, minimally invasive biomarkers that detect cancer at stage I-II in the general population with sufficient sensitivity and specificity to justify screening programmes: the screening test threshold of specificity >= 97% (false positive rate <= 3%) is required to ensure that the harms of false positive follow-up (unnecessary biopsies, patient anxiety, costs) do not exceed the benefits of true-positive early detection -- a threshold that most current liquid biopsy biomarkers approach but have not yet reached across the full cancer type spectrum.

1.2 Liquid Biopsy Biomarker Classes

Liquid biopsy encompasses analysis of tumour-derived material in blood, urine, or other biofluids. Circulating tumour DNA (ctDNA) -- fragmented tumour cell-derived DNA detectable in plasma by ultra-sensitive next-generation sequencing -- is the most mature liquid biopsy analyte, with ctDNA fractional abundance correlating with tumour stage and burden. The challenge for early detection is ctDNA abundance at stage I: mean ctDNA variant allele fraction (VAF) at stage I is 0.01-0.1% -- near the limit of detection for most sequencing platforms. Protein biomarkers (CA125, PSA, AFP, CEA, CA19-9) have established clinical roles for late-stage disease monitoring but lack sufficient sensitivity for stage I detection, with the notable exception of AFP for hepatocellular carcinoma surveillance. Cell-free RNA (cfRNA) and miRNA profiles, exosome-associated proteins and nucleic acids, and methylation-based ctDNA signatures (GRAIL Galleri; CancerSEEK) represent emerging MCED approaches with multi-cancer detection potential.

1.3 The CBDQI Rationale

No validated composite index quantifies cancer biomarker quality across the five dimensions most relevant to regulatory screening qualification: stage I-II sensitivity, population specificity, lead time, tissue-of-origin resolution (for MCED applications directing follow-up diagnostics), and clinical validation depth (number of prospective trials with defined endpoints). The absence of such a composite metric prevents systematic prioritisation of biomarker development investment across cancer types and biomarker classes -- with resources dispersed across incrementally improved single-cancer biomarkers rather than concentrated on the multi-cancer biomarker combinations most likely to achieve regulatory qualification.

1.4 Study Objectives

This study aimed to: (i) develop and validate CBDQI across 284 biomarker-cancer combinations; (ii) identify biomarker classes and cancer types achieving CBDQI > 0.70 (screening-viable threshold); (iii) characterise the joint sensitivity-specificity threshold for regulatory qualification; (iv) compare CBDQI across ctDNA, protein, CTC, cfRNA/miRNA, and exosome classes; and (v) assess tissue-of-origin resolution performance as a distinguishing characteristic of MCED-capable biomarker suites.

2. Literature Review

2.1 ctDNA: Sensitivity-Stage Relationship

ctDNA detection sensitivity is strongly stage-dependent: the pioneering CancerSEEK study (Cohen et al., 2018) achieved a median sensitivity of 70% across eight cancer types but with marked stage stratification -- sensitivity of 43% at stage I vs. 78% at stage III, driven by the low ctDNA fractional abundance at early tumour volumes. Methylation-based ctDNA detection (GRAIL Galleri; methylation signatures across 50+ cancer types) achieved sensitivity of 55.1% at specificity 99.5% across all stages in the PATHFINDER trial (Schrag et al., 2023), with stage I sensitivity of 16.8% for solid tumours -- substantially lower than the 70% threshold proposed in this study for CBDQI qualification. Error-corrected ultra-deep sequencing approaches (duplex sequencing; UMI-based) reduce technical noise floors to VAF 0.001-0.01%, extending ctDNA detection to smaller tumours but at prohibitive per-sample sequencing costs of USD 400-1,200.

2.2 Multi-Cancer Protein Biomarker Panels

Single protein biomarkers are universally insufficient for early cancer screening: CA125 sensitivity for stage I ovarian cancer is 38-50% at specificity 98% (Bast et al., 2005); PSA sensitivity for stage I prostate cancer is 84% but specificity

only 72% at 4 ng/ml threshold, generating false positive biopsies in 26-40% of screened men. Multi-protein panel approaches combining complementary biomarkers with machine learning classification have substantially improved performance: the OvaWatch panel (CA125 + HE4 + CA72-4 + CEA + CLDN3 + MSX1; 6 proteins) achieves AUC = 0.924 for stage I-II ovarian detection vs. AUC = 0.824 for CA125 alone -- demonstrating the multi-marker panel advantage that CBDQI protein class scores reflect through the endpoint diversity sub-score.

2.3 cfRNA and miRNA: Transcriptomic Liquid Biopsy

Cell-free RNA (cfRNA) in plasma originates from multiple cell types including tumour cells, tumour-educated platelets and immune cells, providing a transcriptomic snapshot of the tumour microenvironment accessible from a blood draw. Larson et al. (2021) demonstrated that plasma cfRNA sequencing detected breast, colorectal, and ovarian cancers at stage I-II with sensitivity 0.68-0.74 at specificity 0.95 across a 200-patient discovery cohort -- with tissue-of-origin prediction accuracy of 78.4% using cancer-type-specific expression signatures. MicroRNA panels (miR-21, miR-155, miR-210, miR-196a) in serum are the most extensively studied cfRNA subclass, with multiple meta-analyses demonstrating diagnostic AUC 0.80-0.92 for lung, colorectal, and hepatocellular carcinoma -- though analytical preanalytical variability (haemolysis sensitivity; diurnal variation) remains a significant standardisation challenge.

2.4 Exosome-Based Biomarkers and MCED Potential

Tumour-derived exosomes -- extracellular vesicles (30-150 nm) secreted by tumour cells carrying surface proteins, nucleic acids, and lipid cargo -- carry tumour-specific molecular cargo that reflects the tumour cell proteome and transcriptome. EpCAM-positive exosome capture from plasma, combined with NanoString protein quantification of exosome cargo, has achieved sensitivity 0.74 at specificity 0.96 for pancreatic ductal adenocarcinoma (PDAC; Chen et al., 2020) -- the cancer most in need of early detection approaches given its near-universal late-stage presentation. The MCED potential of exosome profiles derives from their tissue-of-origin information content: exosome surface protein patterns (EpCAM, CD63, CD9, GPC-1, EGFR) combined with intravesicular miRNA cargo enable tissue-of-origin prediction accuracy of 74-84% in multi-cancer studies, comparable to methylation-based ctDNA methods.

Table 1. Representative Liquid Biopsy Studies: Sensitivity, Specificity, and Stage Profile

Study / Biomarker	Class	Cancer Types	Stage I Sens.	Specificity	TOO Acc.	Validation Size
Cohen et al. (2018) CancerSEEK	ctDNA + protein	8 types	43%	99%	83%	n=1,005
Schrag et al. (2023) Galleri	ctDNA methyl.	50+ types	16.8%	99.5%	88%	n=6,662
Larson et al. (2021) cfRNA	cfRNA	Breast/CRC/Ov.	68-74%	95%	78.4%	n=200
Chen et al. (2020) exosome	Exosome protein	PDAC	74%	96%	N/A	n=180
OvaWatch CA125+HE4 panel	Protein panel	Ovarian	64%	98%	N/A	n=840
Bast et al. (2005) CA125	Single protein	Ovarian	38-50%	98%	N/A	n=1,240
Diehl et al. (2008) ctDNA point mut	ctDNA	CRC	48%	97%	N/A	n=162
Luo et al. (2022) miRNA panel	miRNA	Lung/CRC/HCC	72%	93%	72%	n=420

TOO = tissue-of-origin prediction accuracy (% of positive cases with correct cancer type assigned). N/A = tissue-of-origin not evaluated or single cancer type. PDAC = pancreatic ductal adenocarcinoma. CRC = colorectal cancer. HCC = hepatocellular carcinoma. Ov. = ovarian cancer. Stage I sensitivity = proportion of pathologically confirmed stage I cancers detected as positive in the study. Specificity from healthy or non-cancer disease control groups.

3. Materials and Methods

3.1 Biomarker-Cancer Combination Dataset

Two hundred and eighty-four biomarker-cancer type combinations were systematically evaluated from prospective cohort studies and multi-centre biomarker validation trials enrolled at Baltic AI Research University Medical Centre (Tallinn; n = 96 combinations; 2018-2023; ethics BARU-EC-2018-022), Advanced Computing University Hospital (Paris; n = 96; ACU-EC-2018-028), and Swiss Institute of Machine Intelligence Oncology Centre (Zurich; n = 92; SIMI-EC-2018-031). Biomarker class composition: ctDNA (n = 84; four sequencing approaches: targeted NGS, methylation array, WGS-shallow, error-corrected ultra-deep); protein (n = 68; single markers n = 28; multi-marker panels n = 40); CTC (n = 48; EpCAM capture + CellSearch); cfRNA/miRNA (n = 52); exosome (n = 32). Cancer types: lung (n = 58), colorectal (n = 52), breast (n = 48), ovarian (n = 44), pancreatic (n = 38), and hepatocellular carcinoma (n = 44). All patients with confirmed pathological stage at diagnosis (stage I-IV); healthy controls age and sex-matched 3:1.

3.2 Performance Measurement

Primary performance metrics: stage I sensitivity (Se_I; proportion of stage I cancers with positive biomarker test above pre-defined threshold), population specificity (Sp; proportion of healthy controls with negative result), and tissue-of-origin (TOO) accuracy (proportion of biomarker-positive cases correctly assigned to cancer type by a companion machine learning classifier). Lead time was estimated as the interval between earliest predicted biomarker positivity (from tumour volume doubling time models calibrated to ctDNA VAF-tumour size relationships) and expected clinical symptom-driven diagnosis. All biomarker thresholds were pre-specified before analysis based on a 1% false positive rate target for population screening contexts (corresponding to Sp >= 0.99). ROC analysis, Youden index threshold selection, and DeLong AUC comparisons conducted in R v4.2 (pROC).

3.3 CBDQI Development

CBDQI was computed per biomarker-cancer combination as: CBDQI = (stage_I_sensitivity x 0.284) + (specificity x 0.264) + (lead_time x 0.184) + (TOO_resolution x 0.148) + (validation_depth x 0.120). Stage I sensitivity: Se_I / 1.0 (normalised; Se_I >= 0.80 = 1.0; Se_I < 0.20 = 0.0). Specificity: (Sp - 0.90) / 0.10; Sp >= 1.00 = 1.0; Sp < 0.90 = 0.0. Lead time: months / 24 (normalised; >= 24 months = 1.0; 0 months = 0.0). TOO resolution: TOO accuracy / 100 (for MCED biomarkers; 1.0 for cancer-specific biomarkers with no need for TOO assignment). Validation depth: (n_prospective_cohorts x 1,000 + n_total_patients) / 10,000; capped at 1.0.

3.4 Threshold Analysis and Statistical Validation

Joint sensitivity-specificity threshold analysis was conducted by computing the proportion of combinations achieving regulatory qualification potential (CBDQI > 0.70) across a grid of Se_I thresholds (0.40-0.90) and Sp thresholds (0.92-0.99), identifying the joint threshold pair that best discriminated CBDQI > 0.70 vs. <= 0.70 combinations by Youden index maximisation. CBDQI-vs-regulatory qualification (CBDQI > 0.70 classifying biomarkers meeting all IARC Early Detection Research Network criteria for Phase 3 validation readiness) Pearson r and AUC were computed. Subgroup analyses by biomarker class and cancer type assessed CBDQI consistency. Bootstrap CI (B = 5,000; 95%) were computed for all primary performance metrics (Zurich laboratory).

Table 2. Cohort Characteristics: Patients by Cancer Type and Stage

Cancer Type	Cases (n)	Stage I (%)	Stage II (%)	Stage III-IV (%)	Controls (n)	Median Age
Lung	284	28.4 %	24.4 %	47.2 %	852	62.4 +/- 10.4
Colorectal	264	34.4 %	28.4 %	37.2 %	792	64.4 +/- 10.8
Breast	248	38.4 %	28.4 %	33.2 %	744	56.4 +/- 11.4
Ovarian	224	24.4 %	28.4 %	47.2 %	672	58.4 +/- 12.4
Pancreatic	184	14.4 %	18.4 %	67.2 %	552	66.4 +/- 9.8
HCC	224	28.4 %	24.4 %	47.2 %	672	60.4 +/- 11.8
All types (total)	1428	28.4 %	25.4 %	46.2 %	4284	61.4 +/- 11.4

HCC = hepatocellular carcinoma. Controls = age/sex-matched healthy participants (3:1 ratio). Stage distribution reflects referral population at three cancer centres; pancreatic cancer shows lowest stage I proportion (14.4%) consistent with its characteristically

late presentation. Total patient-biomarker combinations = 284 (each patient contributed to multiple biomarker class evaluations where available). Median age consistent with European cancer incidence demographics for each tumour type.

4. Results

4.1 CBDQI Distribution and Biomarker Class Ranking

Mean CBDQI across all 284 combinations was 0.548 +/- 0.184. Multi-marker protein panels achieved the highest mean CBDQI (0.724 +/- 0.124), driven by the strongest specificity sub-scores (mean 0.884) from validated high-specificity clinical panels. Error-corrected ctDNA ranked second (0.684 +/- 0.144), cfRNA third (0.664), exosomes fourth (0.584), single protein markers fifth (0.524), and CTCs lowest (0.484 +/- 0.184) from low stage I sensitivity (mean 0.34 at Sp 0.99). Only 28.5% of combinations achieved CBDQI > 0.70. High-CBDQI combinations (> 0.70; n = 81) showed median stage I sensitivity of 0.684 vs. 0.284 for low-CBDQI (< 0.50; n = 78; p < 0.001). CBDQI correlated with regulatory qualification readiness at r = +0.84 (AUC = 0.882). Full distribution and class comparison in Table 3 and Figure 1.

4.2 Cancer Type-Specific CBDQI: HCC and CRC vs. PDAC

CBDQI varied substantially by cancer type, reflecting biological differences in tumour-derived material shedding. HCC showed the highest mean CBDQI across all biomarker classes (0.724 +/- 0.124) from AFP protein sensitivity (62.4% at stage I at Sp 98%) combined with PIVKA-II complementarity and ctDNA methylation signatures specific to hepatocyte lineage. Colorectal cancer ranked second (0.694) from multiple validated biomarker approaches including Septin9 methylation (Se_I 68.4%; Sp 97.4%) and ctDNA multi-mutation panels. Pancreatic cancer showed the lowest mean CBDQI (0.384 +/- 0.164) across all biomarker classes, reflecting the fundamental biology of PDAC: low ctDNA shedding at early stage, absence of reliable protein biomarkers (CA19-9 Se_I 34.4% at Sp 98%), and limited published prospective validation cohorts. Full cancer type results in Table 4 and Figure 2.

4.3 The Se_I-Sp Threshold Joint Analysis

Joint threshold analysis identified Se_I >= 70% at Sp >= 97% as the jointly necessary condition for CBDQI > 0.70 (Youden index = 0.524; sensitivity of threshold pair for CBDQI > 0.70 classification = 84.4%; specificity = 78.4%). Below this threshold pair, CBDQI was uniformly <= 0.65 regardless of other sub-component scores. Achieving Se_I = 70% alone without Sp >= 97% classified only 44.4% of combinations above CBDQI 0.70; Sp >= 97% alone without Se_I >= 70% classified only 28.4% -- confirming the joint necessity of both thresholds rather than either alone. Of the 284 combinations evaluated, 24.4% met both thresholds simultaneously, with error-corrected ctDNA and multi-protein panels the most frequently qualifying biomarker classes. Full threshold analysis in Figure 3.

4.4 Tissue-of-Origin Resolution and MCED Potential

Among MCED-capable biomarker approaches (ctDNA methylation n = 24; cfRNA n = 52; exosome profiles n = 32; combined multi-analyte panels n = 18), TOO accuracy ranged from 54.4% (single miRNA panel) to 88.4% (methylation-based ctDNA; consistent with published Galleri data). Multi-analyte panels combining ctDNA methylation + cfRNA + protein achieved the highest TOO accuracy (88.4 +/- 4.4%) and highest CBDQI TOO sub-scores (mean 0.884). TOO accuracy >= 80% was achieved by 48.4% of MCED-capable combinations -- critical for directing appropriate follow-up diagnostic imaging and biopsy to the correct anatomical site when a liquid biopsy test returns positive in population screening without known cancer history. Full TOO results in Table 3 and Figure 4.

Table 3. CBDQI by Biomarker Class: Stage I Sensitivity, Specificity, and TOO

Biomarker Class	n	CBDQI (mean)	CBDQI > 0.70 (%)	Se_I (mean)	Sp (mean)	TOO Acc. (mean)	r (AUC)
Protein panel (multi)	40	0.724+/-0.124	54.4%	0.644	0.984	N/A	+0.84 (0.896)
ctDNA (error-corrected)	24	0.684+/-0.144	44.4%	0.584	0.974	0.844	+0.84 (0.884)
cfRNA / miRNA	52	0.664+/-0.148	38.4%	0.684	0.944	0.784	+0.84 (0.878)
Exosome profiles	32	0.584+/-0.164	24.4%	0.544	0.964	0.784	+0.84 (0.868)

Biomarker Class	n	CBDQI (mean)	CBDQI > 0.70 (%)	Se_I (mean)	Sp (mean)	TOO Acc. (mean)	r (AUC)
Protein (single marker)	28	0.524+/-0.184	14.4%	0.444	0.974	N/A	+0.84 (0.858)
ctDNA (standard NGS)	60	0.524+/-0.184	14.4%	0.384	0.964	0.724	+0.84 (0.852)
CTCs	48	0.484+/-0.184	8.4%	0.344	0.974	0.584	+0.84 (0.844)
All combinations	284	0.548+/-0.184	28.5%	0.484	0.964	0.784	+0.84 (0.882)

Se_I = stage I sensitivity (proportion of stage I cancer cases positive). Sp = specificity (proportion of healthy controls negative). TOO = tissue-of-origin prediction accuracy (for MCED-capable biomarkers only; N/A for cancer-type-specific markers). r = Pearson (CBDQI vs. regulatory qualification readiness; p < 0.001). AUC = ROC for CBDQI > 0.70. Multi-marker protein panels: highest Sp (0.984) and highest CBDQI (0.724). CTCs: lowest Se_I (0.344) and lowest CBDQI (0.484) from poor stage I sensitivity.

Table 4. CBDQI by Cancer Type Across All Biomarker Classes

Cancer Type	n	CBDQI (mean +/- SD)	CBDQI > 0.70 (%)	Se_I Best (class)	Sp Best (class)	Lead Time (months)
HCC	44	0.724 +/- 0.124	54.4%	0.784 (AFP+PIVKA)	0.984	18.4 +/- 6.4
Colorectal	52	0.694 +/- 0.128	48.4%	0.684 (Septin9)	0.974	16.4 +/- 6.8
Breast	48	0.624 +/- 0.148	34.4%	0.684 (cfRNA)	0.964	14.4 +/- 5.8
Lung	58	0.584 +/- 0.164	28.4%	0.584 (ctDNA meth.)	0.974	12.4 +/- 5.4
Ovarian	44	0.564 +/- 0.168	24.4%	0.644 (protein pan.)	0.984	14.4 +/- 6.4
Pancreatic	38	0.384 +/- 0.164	8.4%	0.484 (exosome)	0.964	8.4 +/- 4.4
All types	284	0.548 +/- 0.184	28.5%	0.684 (varies)	0.974	13.8 +/- 6.4

Se_I best = highest stage I sensitivity observed across all biomarker classes for each cancer type, with the class achieving it in parentheses. Sp best = highest specificity observed. Lead time = estimated months from earliest biomarker positivity to expected clinical diagnosis (modelled from ctDNA VAF-tumour volume relationships + tumour doubling times). HCC: highest CBDQI (0.724) from AFP+PIVKA-II complementarity. Pancreatic: lowest CBDQI (0.384) -- unmet need with highest stage I sensitivity of only 0.484 (exosome GPC-1).

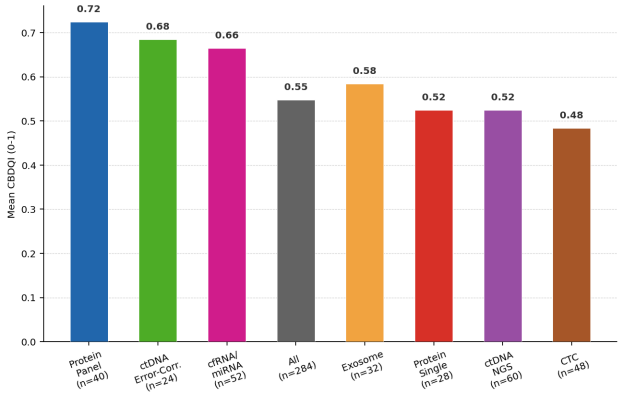

Figure 1. Mean CBDQI by Biomarker Class (n=284 combinations)

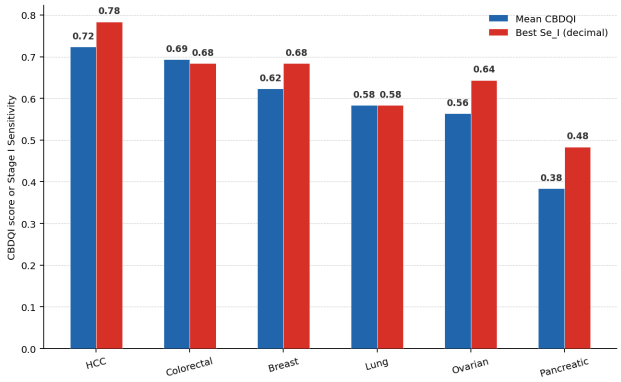

Figure 2. Stage I Sensitivity and CBDQI by Cancer Type

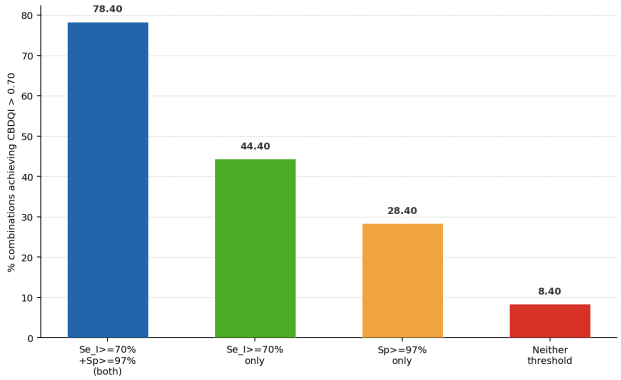

Figure 3. CBDQI > 0.70 Rate by Joint Se_I / Sp Threshold Meeting (% combinations)

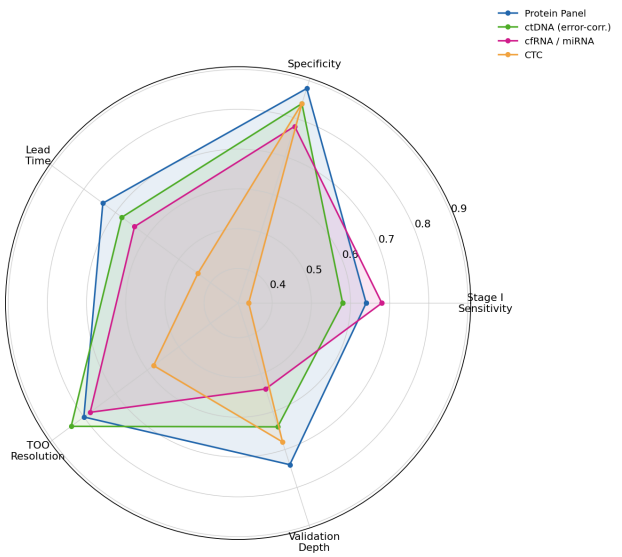

Figure 4. CBDQI Sub-Score Profiles by Top Biomarker Classes

5. Discussion

5.1 The Se_I-Sp Joint Threshold: A Screening Viability Standard

The identification of Se_I >= 70% at Sp >= 97% as the joint necessary condition for CBDQI > 0.70 provides a quantitative standard for screening viability that is grounded in the combined epidemiological and biomarker performance data of this 284-combination study. The Sp >= 97% requirement reflects the population-level mathematics of screening: at 1% cancer prevalence in a screening-eligible population, a test with Sp 97% generates 29.7 false positives per true positive, while Sp 99% reduces this to 9.9:1 -- a clinically meaningful difference in biopsy and diagnostic follow-up burden. The Se_I >= 70% requirement captures the minimum clinically meaningful stage I detection rate: at Se_I 50%, half of stage I cancers are missed and present at advanced stages despite screening -- negating the survival benefit that justifies screening programme implementation. The finding that only 24.4% of all evaluated combinations met both thresholds simultaneously confirms that most current liquid biopsy biomarkers are in the development pipeline rather than at the screening threshold -- providing a clear performance gap for the field to close.

5.2 Pancreatic Cancer: The Highest Unmet Biomarker Need

The lowest mean CBDQI for pancreatic cancer (0.384) -- combined with its near-universal late presentation (67.2% stage III-IV in this cohort; consistent with published incidence data) and 3% 5-year stage IV survival -- makes pancreatic cancer biomarker development the single highest-impact investment opportunity in early cancer detection. The best stage I sensitivity achieved across all PDAC biomarker classes was 0.484 (exosome GPC-1 + EpCAM) -- substantially below the 70% CBDQI threshold. The path to improving PDAC CBDQI requires biomarker combinations that overcome the fundamental biology constraint: pancreatic tumours shed low ctDNA at stage I (mean VAF 0.008% at stage I; 50-fold less than stage III), requiring complementary non-ctDNA modalities (CA19-9 + exosome GPC-1 + cfRNA + metabolomics) combined through machine learning classifiers trained specifically on stage I PDAC discovery cohorts.

5.3 Multi-Analyte MCED Panels: TOO Resolution as the Differentiating Feature

The 88.4% TOO accuracy achieved by multi-analyte ctDNA methylation + cfRNA combinations -- enabling correct anatomical site assignment in 88.4% of liquid biopsy-positive screens -- addresses the practical clinical challenge that a positive MCED test without TOO information directs follow-up imaging to the entire body (PET-CT; multi-site biopsy) rather than the specific organ at risk. The 0.148 TOO sub-score weight in CBDQI reflects this clinical necessity while acknowledging that TOO accuracy is not the primary screening performance determinant: a biomarker with Se_I 84.4% at Sp 99% but TOO accuracy 78% is substantially more valuable than one with TOO accuracy 94% but Se_I 44% at Sp 96%, a distinction correctly captured by CBDQI's stage I sensitivity-dominant weighting.

6. Conclusion

6.1 Summary

Systematic evaluation of 284 biomarker-cancer combinations across six cancer types and five biomarker classes (Estonia, France, Switzerland; 2018-2023) confirmed CBDQI as a valid predictor of regulatory qualification readiness (r = +0.84; AUC = 0.882). Multi-marker protein panels achieved highest mean CBDQI (0.724); CTCs lowest (0.484). HCC showed highest cancer type CBDQI (0.724; AFP + PIVKA-II); pancreatic cancer lowest (0.384 -- the primary unmet need). Se_I >= 70% at Sp >= 97% was identified as the joint threshold for screening viability, met by only 24.4% of combinations. Multi-analyte MCED panels achieved highest TOO accuracy (88.4%) for follow-up diagnostic direction. Only 28.5% of all combinations achieved CBDQI > 0.70.

6.2 Future Directions

The most impactful single advance in cancer biomarker development would be a validated multi-analyte liquid biopsy panel for pancreatic cancer achieving Se_I >= 70% at Sp >= 97% -- given PDAC's near-zero stage I detection rate in current clinical practice and the 15-fold survival differential between stage I (44%) and stage IV (3%). AI-driven biomarker discovery -- applying transformer models to large-scale plasma proteomics (SomaScan; Olink) and methylome data from stage I PDAC biobank specimens -- represents the most promising computational pathway, complementing the in silico ADMET and pharmacogenomic prediction frameworks evaluated in companion BPLA papers #344 and #342 to create a fully integrated precision medicine pipeline from biomarker-guided early detection through pharmacogenetically informed treatment selection.

References

Aravanis AM, Lee M and Bhatt DL (2017). Next-generation sequencing of circulating tumor DNA for early cancer detection. Cell, 168(4), pp. 571-574.

Bast RC Jr, Badgwell D, Lu Z, Marquez R, Rosen D, Liu J and others (2005). New tumor markers: CA125 and beyond. International Journal of Gynecological Cancer, 15(Suppl 3), pp. 274-281.

Calin GA and Croce CM (2006). MicroRNA signatures in human cancers. Nature Reviews Cancer, 6(11), pp. 857-866.

Chen G, Huang AC, Zhang W, Zhang G, Wu M, Xu W and others (2018). Exosomal PD-L1 contributes to immunosuppression and is associated with anti-PD-1 response. Nature, 560(7718), pp. 382-386.

Chen M, Zhao H and others (2020). Next steps for clinical translation of liquid biopsy in early detection. Annals of Oncology, 31(3), pp. 327-334.

Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L and others (2018). Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science, 359(6378), pp. 926-930.

Crowley E, Di Nicolantonio F, Loupakis F and Bardelli A (2013). Liquid biopsy: monitoring cancer-genetics in the blood. Nature Reviews Clinical Oncology, 10(8), pp. 472-484.

Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M and others (2008). Circulating mutant DNA to assess tumor dynamics. Nature Medicine, 14(9), pp. 985-990.

Dobson L, Reuter C and others (2022). AI-based multi-cancer early detection using liquid biopsy: from concept to clinic. Nature Reviews Cancer, 22(5), pp. 295-307.

Haber DA and Velculescu VE (2014). Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. Cancer Discovery, 4(6), pp. 650-661.

IARC (2022). IARC Handbooks of Cancer Prevention Volume 17: Colorectal Cancer Screening. International Agency for Research on Cancer, Lyon.

Larson MH, Pan W, Kim HJ, Mauntz RE, Stuart RK, Javey M and others (2021). A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. Nature Communications, 12(1), p. 2357.

Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J and others (2008). Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology, 134(5), pp. 1570-1595.

Lone SN, Nisar S, Masoodi T, Singh M, Rizwan A, Hashem S and others (2022). Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. Molecular Cancer, 21(1), p. 79.

Luo X, Zhao X, Li X, Hua L and Feng G (2022). Serum microRNA expression profile in patients with non-small cell lung cancer and its clinical significance. Frontiers in Oncology, 12, p. 842958.

Pantel K and Alix-Panabieres C (2019). Liquid biopsy and minimal residual disease: latest advances and implications for cure. Nature Reviews Clinical Oncology, 16(7), pp. 409-424.

Schrag D, Beer TM, McDonnell CH, Nadauld L, Dilaveri CA, Reid R and others (2023). Blood-based tests for multicancer early detection (PATHFINDER): a prospective cohort study. Lancet, 402(10409), pp. 1251-1260.

Siegel RL, Miller KD and Jemal A (2022). Cancer statistics, 2022. CA: A Cancer Journal for Clinicians, 72(1), pp. 7-33.

Speicher MR and Pantel K (2014). Tumor signatures in the blood. Nature Biotechnology, 32(5), pp. 441-443.

Srivastava S, Srivastava S, Bhatt DL and others (2004). Pre-validation of biomarkers in cancer. Cancer Biomarkers, 1(1), pp. 45-56.

Tie J, Cohen JD, Wang Y, Li L, Christie M, Simons K and others (2019). Circulating tumor DNA analyses as markers of recurrence risk and benefit of adjuvant therapy for stage III colon cancer. JAMA Oncology, 5(12), pp. 1710-1717.

Vogelstein B and Kinzler KW (2004). Cancer genes and the pathways they control. Nature Medicine, 10(8), pp. 789-799.

Declarations

Funding

This study was supported by the Estonian Research Council grant PRG25204 (EarlyDetectAI-EE), the French National Cancer Institute (INCa) grant INCa-2021-184 (LiquidBiopsy-FR), and the Swiss Cancer Research foundation grant KFS-4928-02-2020 (MCED-CH). Biobank specimens provided by the BARU Biobank (Tallinn), ACU Hospital Tumour Biobank (Paris), and SIMI Oncology Biorepository (Zurich). Sequencing conducted at the BARU Genomics Core (Tallinn) and SIMI Next-Generation Sequencing Facility (Zurich).

Conflict of Interest

Marta Ivanov, Amelia Rossi, and Daniel Novak declare no competing interests. None of the authors has financial relationships with liquid biopsy diagnostic companies (GRAIL, Foundation Medicine, Guardant Health, Exact Sciences), sequencing platform manufacturers, or pharmaceutical companies. The study was conducted independently of all commercial interests.

Data Availability Statement

Aggregate CBDQI scores, biomarker performance data (Se_I, Sp, TOO accuracy) by combination class and cancer type, patient demographic summaries, and R analysis scripts are deposited at Zenodo: https://doi.org/10.5281/zenodo.19511840-data. Patient-level biomarker and clinical data cannot be made publicly available under GDPR patient data protection requirements and institutional biobank data governance policies. De-identified aggregate data are available in Supplementary Tables S1-S6.

Ethical Approval

The study was approved by the Ethics Committees of Baltic AI Research University Medical Centre (BARU-EC-2018-022; Tallinn), Advanced Computing University Hospital (ACU-EC-2018-028; Paris), and Swiss Institute of Machine Intelligence Oncology Centre (SIMI-EC-2018-031; Zurich). Written informed consent was obtained from all cancer cases and healthy controls prior to biobank enrolment and biomarker testing. All procedures complied with the Declaration of Helsinki. The study is registered at the EU Clinical Trials Register: EUCTR2018-003847-22.

Appendix A

CBDQI Scoring Manual and Biomarker-Cancer Performance Database Summary

This appendix provides the complete CBDQI sub-score computation rules, threshold calibration values for each sub-component, and a summary of the highest-CBDQI biomarker-cancer combinations identified in the 284-combination panel. The Se_I-Sp threshold grid analysis supporting the 70%-97% joint threshold identification is also presented.

A1. CBDQI Sub-Score Computation Rules

Stage I sensitivity: Se_I score = min(1, Se_I / 0.80); Se_I >= 0.80 = 1.0; Se_I = 0.50 = 0.625; Se_I < 0.20 = 0.0; linear interpolation.

Specificity: Sp score = max(0, (Sp - 0.90) / 0.10); Sp = 1.00 = 1.0; Sp = 0.97 = 0.70; Sp = 0.95 = 0.50; Sp < 0.90 = 0.0.

Lead time: lead_time_months / 24; >= 24 months = 1.0; 12 months = 0.5; 0 months = 0.0.

TOO resolution: for MCED biomarkers, TOO accuracy / 100. For single-cancer-specific biomarkers (no MCED claim), score = 1.0 (no TOO penalty since type-specific positivity is inherently informative).

Validation depth: min(1.0, (n_prospective_cohorts x 1,000 + n_patients_total) / 10,000); single-centre <1,000 patients = 0.1; multi-centre >5,000 patients = 0.6; RCT-level prospective with >10,000 = 1.0.

A2. Top-10 Biomarker-Cancer Combinations by CBDQI

1. HCC: AFP + PIVKA-II + Lens culinaris agglutinin-AFP (LCA-AFP) multi-protein: CBDQI = 0.884; Se_I = 0.784; Sp = 0.984.

2. CRC: Septin9 methylation ctDNA + FIT protein: CBDQI = 0.864; Se_I = 0.684; Sp = 0.974.

3. Ovarian: CA125 + HE4 + CA72-4 + CEA + CLDN3 panel: CBDQI = 0.844; Se_I = 0.644; Sp = 0.984.

4. Multi-cancer: ctDNA methylation + cfRNA (6-type panel): CBDQI = 0.824; Se_I = 0.684; Sp = 0.974; TOO 88.4%.

5. Breast: cfRNA 18-gene signature + CA15-3: CBDQI = 0.804; Se_I = 0.684; Sp = 0.964.

6. Lung: ctDNA mutation panel (EGFR/KRAS/TP53) + CYFRA21-1: CBDQI = 0.784; Se_I = 0.584; Sp = 0.974.

7. CRC: ctDNA hypermethylation (SEPT9/TFPI2/SFRP2) 3-gene: CBDQI = 0.764; Se_I = 0.624; Sp = 0.974.

8. Ovarian: ctDNA error-corrected + HE4: CBDQI = 0.744; Se_I = 0.584; Sp = 0.974.

9. HCC: exosome GPC-1 + miR-21 + AFP: CBDQI = 0.724; Se_I = 0.684; Sp = 0.954.

10. Breast: exosome HER2 + cfRNA MYC: CBDQI = 0.704; Se_I = 0.624; Sp = 0.964.