

Tumor-agnostic multimodal cell-free DNA analysis improves detection and enables longitudinal disease monitoring in esophageal adenocarcinoma

Alexandra Bartolomucci

Research Institute of the McGill University Health Centre

Lisa-Monique Edward

University Health Network

Sarah Tadhg Ferrier

Research Institute of the McGill University Health Centre

Jeffrey P. Bruce

University Health Network

Dor D. Abelman

University Health Network

Monyse Nobrega

Research Institute of the McGill University Health Centre

Kwang Bo Joung

Research Institute of the McGill University Health Centre

Arnavaz Danesh

University Health Network

Stephenie D. Prokopec

University Health Network

Luis Santos Castro

McGill University Health Centre

Abirami P. Sharma

McGill University Health Centre

Rawan Sakalla

McGill University Health Centre

Wotan Zeng

McGill University Health Centre

Carolina Martini

Research Institute of the McGill University Health Centre

Kyle Dickinson

Research Institute of the McGill University Health Centre

Nicholas Bertos

Research Institute of the McGill University Health Centre

Lorenzo Ferri

Research Institute of the McGill University Health Centre

Jonathan Cools-Lartigue

Research Institute of the McGill University Health Centre

Trevor J. Pugh

University Health Network

Julia V. Burnier

julia.burnier@mcgill.ca

Research Institute of the McGill University Health Centre

Research Article

Keywords: cell-free DNA, circulating tumor DNA, liquid biopsy, esophageal adenocarcinoma, fragmentomics

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11210292/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Esophageal adenocarcinoma (EAC) is an aggressive malignancy for which liquid biopsy–based clinical monitoring using cell-free DNA (cfDNA) remains challenging due to low tumor cfDNA shedding, few recurrent mutations, and limited access to tissue required for tumor-informed assays. Moreover, genome-wide cfDNA analysis remains largely unexplored in prospective longitudinal EAC cohorts.

Methods

Here, we developed a tumor-agnostic multimodal cfDNA framework integrating genomic and fragmentomic features. We applied this framework to 137 prospectively collected longitudinal plasma samples from 49 patients using deep targeted sequencing, personalized digital PCR, and shallow whole-genome sequencing (sWGS). Using sWGS-derived cfDNA features, disease-status and 6-month mortality classifiers were developed using elastic-net logistic regression and evaluated in held-out test cohorts.

Results

At baseline, targeted sequencing detected tumor cfDNA in only 52% of patients, while integrating targeted mutation analyses with genome-wide copy-number alterations and fragmentomic features increased detection to 85%. Longitudinal profiling detected tumor cfDNA in all patients with plasma available at clinical relapse. In held-out test cohorts, the classifiers achieved AUCs of 0.82 for disease status and 0.81 for 6-month mortality risk.

Conclusions

Together, these findings establish that integrating complementary mutation, copy-number, and fragmentomic signals improves cfDNA detection in EAC, while machine-learning analysis of features from a single low-cost, tumor-agnostic sWGS assay provides disease and prognostic information. These results support sWGS as a scalable strategy for longitudinal liquid-biopsy assessment in EAC, with potential applicability to other low-shedding cancers.

Background

Esophageal cancer is the seventh most common cause of cancer-related death worldwide (1). Esophageal adenocarcinoma (EAC) is the most prevalent type of esophageal cancer in North America, with incidence rates rising sharply over the last three decades (1). The five-year survival rate of EAC remains low at less than 20%, which can be attributed to patients commonly presenting with advanced disease that frequently exhibits initial or acquired resistance to chemotherapy treatments (2). Complete

response after neoadjuvant chemotherapy is rare, accounting for less than 15% of EAC cases (3). Moreover, up to 50% of EAC patients will eventually develop disease recurrence (4). The current assessment of treatment response and disease recurrence relies primarily on 1) radiologic imaging, which is conducted infrequently and lacks sensitivity for low tumor burden, and 2) endoscopic surveillance, an invasive procedure ill-suited for real-time longitudinal monitoring. Overall, there is a dire clinical need for minimally invasive biomarkers that can support prognostication, guide the selection and escalation or de-escalation of systemic therapy, enable dynamic monitoring of treatment response and resistance, and facilitate the detection of minimal residual disease and recurrence in EAC.

Liquid biopsy, referring to the analysis of circulating biomarkers from patient biofluids, has emerged as an effective technique for the minimally invasive surveillance of cancer. Circulating tumor DNA (ctDNA) from liquid biopsy samples has seen particular success as a biomarker for cancer diagnostics, prognostication, monitoring of treatment response, and detecting minimal residual disease (MRD) in multiple tumor types (5). Recent FDA approvals for the use of ctDNA for the clinical management of lung, breast, ovarian, and colorectal cancers exemplify the rapid development of the field as well as the clinical utility of ctDNA (5). However, ctDNA use in EAC remains relatively under-explored, likely due to these tumors being low ctDNA shedding, lacking recurrent hotspot mutations, and being highly driven by structural variants, which limits the sensitivity of ctDNA assays based on gene panels (6–9). In fact, despite the aggressive nature of the disease and the propensity towards late-stage diagnosis, prior studies using tumor-agnostic targeted panel-based liquid biopsy approaches have reported ctDNA detection rates of only approximately 38–63% (10–12). Tumor-informed approaches that track patient-specific alterations can achieve higher assay sensitivity (13, 14). However, these approaches require access to adequate amounts of tumor tissue, prior tumor sequencing, and the development of individualized assays before plasma analysis can be performed. This can greatly increase turnaround time and workflow complexity and may limit implementation when tumor tissue is unavailable. Moreover, targeted approaches only examine a predefined set of genomic regions and may therefore be affected by low ctDNA shedding, spatial tumor heterogeneity, clonal evolution, and the absence of suitable trackable variants. These limitations highlight the need for scalable tumor-agnostic liquid biopsy strategies capable of detecting cancer-derived signals without reliance on tumor-informed mutation profiling.

Over the past few years, shallow whole-genome sequencing (sWGS) of cell-free DNA (cfDNA) has emerged as a powerful, cost-effective approach to detect genome-wide tumor signals. sWGS enables estimation of tumor fraction from large-scale copy number alterations and from fragmentomics—the analysis of cfDNA fragmentation patterns, including fragment sizes, genomic positioning, nucleosome footprints, and end motifs (5). Cancer-associated cell death processes, together with the altered epigenetic, transcriptomic, and genomic landscapes of tumor cells, generate distinctive fragmentomic signatures in the cfDNA of cancer patients (15). Importantly, both copy number and fragmentomic analyses can be derived from sWGS data, making this approach a rapid, inexpensive, and clinically scalable alternative to traditional mutation-based liquid biopsy assays. However, genome-wide cfDNA fragmentomics remains largely unexplored in EAC. To our knowledge, only one prior multi-cancer study

has evaluated fragmentomic features in patients with EAC, and no study has comprehensively integrated mutation, copy-number, and fragmentomic signals in a prospective longitudinal EAC cohort (16). Such genome-wide profiling may be particularly valuable for improving detection sensitivity in EAC, a genomically complex cancer characterized by chromosomal instability, structural variation, extensive copy-number alterations, and few recurrent hotspot mutations.

Here, we introduce a tumor-agnostic multimodal cfDNA framework integrating mutation, copy-number, and fragmentomic analyses to improve detection sensitivity and enable longitudinal disease monitoring in EAC. Applying this approach to 137 serial plasma samples from 49 patients, we assessed targeted sequencing (TS), droplet digital PCR (ddPCR), and multiple genome-wide features derived from sWGS. While these complementary approaches increased overall detection, the greatest breadth of clinically relevant information was obtained from sWGS-derived copy-number and fragmentomic features, which supported scalable, dynamic treatment monitoring and survival risk stratification without requiring prior tumor sequencing or personalized mutation tracking.

Methods

Sex as a biological variable

Both male and female participants were included in this study. As expected for esophageal adenocarcinoma, the cohort was predominantly male (86%). The study was designed to evaluate genomic and cell-free DNA features associated with disease detection, monitoring, and outcomes and was not designed to assess sex-specific differences.

Patient cohort

Following informed consent, 49 patients were enrolled in this study at the Montreal General Hospital Thoracic and Upper Gastrointestinal Clinic. Longitudinal blood samples were obtained from patients at diagnosis (baseline), follow-up appointments, surgery, and post-treatment, between June 2022 and April 2025. Data from cancer-free controls ($n = 26$) was obtained from the CHARM consortium, accessible in the European Genome-Phenome Archive (EGA) under the accession number EGAS00001006539 (17).

Blood collection and processing

10 mL blood samples were collected from patients from a peripheral vein or central line in PAXgene blood ccfDNA tubes (BD Biosciences #768165). Samples were centrifuged at 2000g for 15 min to isolate plasma, followed by a second spin at 2000g for 10 min to remove any remaining cells, as previously described by our group (18). Buffy coat was isolated after the first round of centrifugation as a source of white blood cells (WBCs). Samples were stored at -80°C until DNA isolation.

DNA extraction

Genomic DNA from WBCs was extracted using the QIAamp DNA Mini Kit (Qiagen #51304). cfDNA was isolated from plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen #55114) modified by omitting carrier RNA and adding extra elution steps. Samples were briefly vortexed and centrifuged before quantification using 1 µL of DNA. DNA was quantified using the Qubit fluorometer dsDNA high sensitivity kit (Invitrogen # Q32854) following the manufacturer's protocol.

DNA library construction and sequencing

For targeted sequencing (TS), 150 nanograms (ng) of buffy coat DNA or 20 ng of cell-free DNA (cfDNA) were used for preparation of targeted sequencing library. Buffy coat DNA was sheared to a targeted average fragment size of 250 bp using a M220 Focused-Ultrasonicator (Covaris). Pre-capture libraries were prepared from the sheared DNA using KAPA DNA HyperPrep Kit (Roche Cat#07962363001) with IDT xGen Duplex Seq Adapter – Tech Access (IDT #1080799). Target capture with xGen™ Pan-Cancer Hyb Panel (IDT #1056204) was performed using IDT xGEN Hybridization and Wash Kit (IDT #1080584) and IDT xGEN Universal Blockers – TS mis (IDT #1075475) according to manufacturer's instruction. Prepared libraries were balanced, pooled and loaded on an Illumina NovaSeq 6000 Sequencing System (Illumina #20012850) and sequenced at 2x151 cycles. For TS samples, the mean on-target base coverage was 4,144x (range: 855x-6,432x) across WBC control samples and 2,771x (range: 691x-6,388x) across plasma cfDNA samples. For shallow whole-genome sequencing (sWGS), the pre-capture libraries were sequenced on an Illumina NextSeq 2000 Sequencing System (Illumina #20038897) and sequenced at 2x100 cycles. The mean base coverage across sWGS samples was 0.2x (range: 0.04x-0.4x). All sequencing coverages (TS, sWGS) can be found in **Supplementary Table S8**.

Targeted sequencing mutation analysis

Consensus bam files were produced from FASTQ files using ConsensusCruncher, as previously described (19, 20). Briefly, variant detection was performed individually on single-strand consensus sequences (SSCS), duplex consensus sequences (DCS) and all unique bam files, then merged. Germline variants were identified using the GATK's (v3.8) HaplotypeCaller and Genotype GVCFs following germline variant calling best practices (21). Germline variant significance was assessed using CPSR (v0.6.1) and PCGR (v0.9.1). Variants with a CPSR pathogenicity score of 5 were manually reviewed for predicted functional consequence and ClinVar classification. Variants consistent with pathogenic loss of function were classified as germline pathogenic variants, whereas variants with benign, conflicting, or insufficient evidence were excluded. Somatic single nucleotide variants (SNVs) and short insertions/deletions (INDELs) were called using MuTect (v1.1.5), MuTect2 (v3.8), Pindel (v0.2.5b8), SomaticSniper (v1.0.5.0), Strelka (v2.9.10), VarDict (v1.7.0), and VarScan (v2.4.2). An ensemble approach was applied, such that only SNVs identified by four or more tools, INDELs identified by three or more tools, or variants identified by MuTect2 with variant allele frequencies (VAFs) less than 0.1 were carried forward for downstream

analyses. Germline variants identified by matched WBC targeted sequencing were excluded. Finally, variants were annotated using VEP (ensembl-vep release 98) and vcf2maf (v.1.6.17). Population allele frequencies were obtained from ExAC and gnomAD, and variants with a population allele frequency (popAF) > 0.001 were flagged as common. Clinical significance was annotated using ClinVar, and OncoKB (oncokb-annotator v.3.4).

Targeted sequencing-based somatic variant prioritization for personalized ddPCR assay development

Deep TS of plasma cfDNA was used to identify candidate somatic mutations for downstream personalized ddPCR development. To systematically prioritize high-confidence, biologically relevant mutations for ddPCR monitoring, only somatic variants meeting predefined biological and technical confidence criteria were selected. This approach was necessary given the tumor-agnostic experimental design and the large quantity of candidate variants identified. Variants were eligible for selection if they met at least one of the following criteria: (i) protein-altering coding mutations annotated as oncogenic in OncoKB and occurring in genes recurrently altered in EAC (*ACVR2A*, *APC*, *ARID1A*, *CDKN2A*, *CTNNB1*, *EGFR*, *FBXW7*, *KRAS*, *NOTCH1*, *PIK3CA*, *SMAD4*, or *TP53*); (ii) variants supported at their highest confidence by ≥ 3 independent annotation or variant-calling sources; or (iii) variants occurring in recurrently altered EAC genes with high plasma variant allele frequency (VAF > 15%), consistent with a clonal tumor-derived signal. Samples were considered positive by TS if at least one high-confidence plasma somatic mutation was detected. Where multiple eligible variants were identified in a patient, one mutation was prioritized for personalized ddPCR assay development. Mutations selected for personalized ddPCR assay development are listed in **Supplementary Table S9**.

ddPCR assays and PCR conditions

Primers, probes, and synthetic DNA (gBlock gene fragments as positive and negative controls) were purchased from IDT. Primer and probe sequences can be found below in **Supplementary Table S9**. ddPCRs were performed as per the manufacturer's protocol, using 10 μ L ddPCR Supermix for probes (No UTP) (Bio-Rad #1863024) or 5 μ L ddPCR Multiplex Supermix (Bio-Rad # 12005910), 900 nM forward/reverse primers (IDT), 250 nM FAM/HEX Affinity Plus qPCR probes (IDT), up to 13 μ L of DNA, and nuclease-free water (IDT # 11-04-02-01). For each reaction, the mix was added to the cartridge, followed by 70 μ L of Droplet Generation Oil (Bio-Rad #1863005). Droplets were generated using the QX200 Droplet Generator (Bio-Rad) and transferred to a 96-well PCR plate (Bio-Rad #12001925). PCR reactions were run as follows: 10 min at 95°C, followed by 50 cycles of 30 s at 95°C, 1 min at 55–62°C (optimized for each primer set, listed in **Supplementary Table S9**) and 30 s at 72°C, and finally, 5 min at 98°C. A ramp rate of 2°C/s was used. The plate was read using the QX200 Droplet Reader (Bio-Rad) and data was analyzed using QuantaSoft software. All samples were performed in duplicate. Samples with < 10,000 droplets were excluded from analysis. Baseline, surgical and/or metastatic patient samples were

used to determine positivity by ddPCR. Follow-up samples were only tested if patients showed initial ddPCR positivity.

Copy-Number Alteration Analysis

Copy number and tumor burden prediction were performed using ichorCNA v.0.3.2 (22). Read counts were first collected using hmmcopy_utils readCounter, using a window size of 1M and quality threshold of 20. The resulting WIG file was used as input to the ichorCNA algorithm using default parameters and the default panel of normals (22). For all ichorCNA analyses, only an estimated tumor fraction (TF) > 0.03 was considered positive, as per the published ichorCNA limit of detection.

Fragment size and score analysis

Global fragment size distributions were calculated using Picard CollectInsertSizeMetrics (v2.6.0). The short-to-normal fragment ratio was defined as the number of fragments < 150 bp divided by the number of fragments between 150–220 bp. Fragment scores were calculated as previously described by Vessies et al. (17, 23). Fragment size and score features were Z-score normalized relative to cancer-free controls. Positivity was defined using feature-specific dynamic thresholds derived from the control distribution, set at 10% above the maximum or 10% below the minimum control Z-score in the direction of abnormality for each feature.

Fragment ratio analysis

Genome-wide fragmentation profiles were produced using DELFI, as described by Cristiano et al. (24). Briefly, the ratio of short (90–150 bp) to long (151–220 bp) cfDNA fragments was calculated across 5 Mb genomic bins. Samples were considered positive by DELFI when the Pearson's correlation coefficient of genome-wide 5-Mb fragment ratios with the cancer-free reference profile was < 0.5.

Nucleosome footprint analysis

Nucleosome footprinting analysis was performed as outlined by Vanderstichele et al. (25). Cancer-free control nucleosome peaks were obtained from Snyder et al. (26). Deduplicated BAM files were converted to BED files with fragment starts and ends mapped using Bedtools (v2.27.1). Fragment starts and ends were mapped to the nearest nucleosome peak and only fragments located within 900 bp of a peak were retained for further analyses. Fragment frequencies were calculated as a function of distance from the nearest nucleosome peak and Z-score normalized relative to controls. Positivity was defined using dynamic thresholds derived from the control distribution for two predefined regions: nucleosome centres (– 10 to + 10 bp from nearest peak), where reduced depletion relative to controls was considered abnormal, and nucleosome ends (– 93 to – 73 bp and + 73 to + 93 bp from nearest peak), where increased depletion relative to controls was considered abnormal. Samples were classified as positive when the median Z-score within either region exceeded a dynamic threshold defined as 10% beyond the maximum or minimum control Z-score in the direction of abnormality.

Nucleosome accessibility analysis

Nucleosome accessibility was determined using the Griffin framework, as described by Doebley et al. (27) Esophageal carcinoma-specific chromatin accessibility sites were obtained from the TCGA's ATAC-Seq data (28). Accessibility scores were calculated as the mean corrected coverage across the central $\pm$ 30 bp region surrounding ESCA-specific sites and sign-inverted such that lower central coverage produced higher scores. Samples were classified as positive when the ESCA Griffin Z-score exceeded the 90th percentile of cancer-free control scores.

Fragment end-motif analysis

Fragment end-motif signatures were calculated by extracting 4 bp 5' end motifs for all cfDNA fragments and computing the frequencies of the 256 tetranucleotide motifs (29). DeconstructSigs (v1.8.0) was used to deconvolute the end-motif profiles onto the reference signature matrix using non-negative regression (30). Signature weights were Z-score normalized relative to cancer-free controls. Positivity was defined using signature-specific dynamic thresholds derived from the control distribution, set at 10% above the maximum or 10% below the minimum control Z-score in the direction of abnormality for each signature.

Data integration and genome-wide cfDNA classifier

We developed a ML framework to integrate CNA and fragmentomic features derived from sWGS, adapted from Abelman et al., to evaluate whether combinations of genome-wide cfDNA signals could improve longitudinal disease classification and short-term mortality risk stratification in EAC (31). Specifically, features extracted for the ML analysis included TF estimates from ichorCNA, fragment length metrics (median insert sizes, fragmentation scores, proportion of fragments < 150 bp, and short:normal fragment ratios), genome-wide fragmentation patterns (DELFI correlations), nucleosome positioning features, tissue-of-origin signals (Griffin-derived chromatin accessibility scores for esophageal carcinoma), and end-motif frequencies (F-profiles I and VI). Most fragmentomic features were expressed as Z-scores relative to cancer-free controls, whereas ichorCNA TF and DELFI correlation were included as raw features. We defined a curated set of candidate feature combinations, including both single-feature models and multi-feature models integrating complementary signals. A complete description of the evaluated feature sets is provided in **Supplementary Table S3**. All features were centered and scaled during model training. Binary classifiers were trained using elastic-net regularized logistic regression implemented using the glmnet package (v4.1-8) within the caret ML framework. Hyperparameters were optimized across a grid of: alpha (elastic-net mixing parameter: 0, 0.25, 0.5, 0.75, 1) and lambda (regularization strength, spanning 50 log-scaled values from 0.001 to 10). For single-feature models, standard logistic regression was used without penalization.

Model optimization was performed using a nested CV framework. To minimize data leakage arising from repeated longitudinal sampling, patients were first randomly stratified into a training cohort (70% of patients) and an independent held-out test cohort (30% of patients), such that no samples from the same patient were included in both sets. Within the training cohort, model development was performed using nested CV with patient-level resampling. The outer five-fold CV loop was used to estimate internal

model performance, while the inner five-fold CV loop was used for hyperparameter tuning within each outer training fold. All resampling was performed at the patient level to prevent samples from the same patient appearing in both training and validation folds.

Out-of-fold predictions from the outer CV loop were used to generate pooled performance estimates and ROC curves. Classifier performance was assessed using AUC, sensitivity, specificity, balanced accuracy, positive predictive value (PPV), negative predictive value (NPV), F1 score, Brier score, and partial AUC in the high-specificity region (pAUC90). Performance metrics were averaged across outer folds, and CI for pooled ROC curves were calculated using the DeLong method.

Following nested CV, the best-performing feature combinations were selected based on cross-validated performance metrics, including balanced accuracy and partial AUC in the high-specificity region. Selected models were then refit on the full training dataset to generate final model coefficients. Decision thresholds were determined in the training data only using the Youden index, which identifies the threshold that maximizes the sum of sensitivity and specificity. These fixed thresholds were then applied unchanged to the independent held-out test cohort, which was not used during model training or hyperparameter tuning. Performance metrics reported for the test cohort were therefore based on thresholds derived exclusively from the training data.

Statistical Analysis

Comparisons of ctDNA levels and fragmentomic metrics between clinical stages were performed using the non-parametric Mann-Whitney U test in GraphPad Prism (v.10.6.1). Associations between ddPCR positivity and clinical features were evaluated using Fisher's exact test. OS analyses were performed in R (v.4.5.2) using the survival and survminer packages, with Kaplan-Meier curves and log-rank tests used for unadjusted comparisons. For baseline sWGS-derived cfDNA features, separate multivariable Cox proportional hazards models were used to evaluate associations with OS after adjustment for clinical stage, grouped as stage IV versus stages I–III. Each model included one baseline cfDNA feature and clinical stage.

For the 6-month landmark mortality analysis, survival time was measured from each blood sample collection and administratively censored at 6 months. The association between the binary short-term mortality risk classifier and OS was evaluated using Cox proportional hazards regression implemented with the survival package, with patient-level clustering to account for repeated longitudinal samples. The multivariable model included classifier status and clinical stage, grouped as stage IV versus stages I–III because of the small test cohort, limited number of events, and only one stage I patient. Adjusted HR and 95% CI are reported. The proportional hazards assumption was assessed using Schoenfeld residuals. All statistical tests were two-sided, and P values < 0.05 were considered statistically significant.

Results

Patient cohort

We analyzed cfDNA from 49 prospective patients with EAC collected at diagnosis and longitudinally during routine clinical care. Plasma samples were obtained at multiple time points before, during, and after systemic treatment and surgical intervention, enabling evaluation of cfDNA dynamics across the disease course. Clinical characteristics of the cohort are summarized in **Supplementary Table S1**. The majority of patients were men (86%) with a median age of 68 years (range 49–84), reflecting the known demographic distribution of EAC. Primary tumors were located in the distal esophagus in 53% of patients, at the gastro-esophageal junction (GEJ) in 43%, and in the mid-esophagus in 4%. Tumor stages ranged from I to IVB, and most patients (82%) were treated with curative intent, while 18% received palliative therapy. First-line treatment reflected stage at diagnosis, with most patients receiving chemotherapy and surgery (53%). Other treatment approaches included chemoimmunotherapy with or without surgery (12%), chemoradiation (4%), endoscopic resection alone (8%), or single-modality chemotherapy or radiation (10%).

A total of 137 blood samples were analyzed using a multimodal cfDNA profiling framework integrating deep TS, TS-informed ddPCR, and sWGS (Fig. 1). This approach enabled simultaneous assessment of mutation-based ctDNA detection, genome-wide copy-number alterations (CNAs), and fragmentomic features derived from cfDNA fragmentation patterns. Deep TS was performed on 82 cfDNA samples and sWGS was conducted on 132 cfDNA samples from the 49 patients. Mutation-specific ddPCR analysis was performed on 39 cfDNA samples from 25 patients, using one selected mutation per patient. A detailed overview of all analytical assays performed across sampling timepoints is provided in **Supplementary Figure S1**. At least one analyzed cfDNA sample was available for all 49 patients. Serial sample analysis, defined as the analysis of more than one cfDNA sample from the same patient, was available for 34/49 patients (69%), with a median of 2 samples analyzed per patient (range 1–7). Clinical follow up was available for a mean of 17.8 months from the first blood collection (range 2.1–36.2, median 18.7) and 11.3 months from the final blood collection (range 0–36.2, median 9.0), with clinical follow up censored at the date of last known clinical contact or death.

Tumor-agnostic targeted sequencing enables personalized ddPCR monitoring in a subset of patients

We first evaluated plasma somatic variant detection in the cohort using tumor-agnostic deep TS. Matched white blood cell (WBC) sequencing was performed to exclude germline and clonal hematopoiesis variants. Analysis of WBC sequencing revealed incidental pathogenic germline variants in 3/49 patients (6%), comprising frameshift deletions in *ATM* and *BRCA2*. Large-scale genomic studies have reported pathogenic germline variants in approximately 9% of EAC patients, with enrichment in *TP53* wild-type tumors (32). Consistent with these observations, none of the patients harboring germline pathogenic variants in our cohort had detectable somatic *TP53* mutations in cfDNA. While not the primary focus of this study, these findings highlight the potential added value of liquid biopsy

approaches that incorporate matched leukocyte sequencing for the identification of clinically relevant germline alterations.

Tumor-agnostic TS of plasma cfDNA initially identified a large number of candidate somatic variants across the cohort (Fig. 2B). To identify likely tumor-derived alterations and prioritize biologically relevant targets for downstream personalized ddPCR monitoring, variants were filtered using predefined technical and biological criteria (Fig. 2A; see **Methods**). Figure 2B summarizes the progressive reduction from all candidate variants to a subset of high-confidence variant calls. Following this prioritization, at least one high-confidence plasma somatic variant was identified in 26/49 patients (53%) across all sequenced timepoints. Baseline plasma samples were available for 46 of the 49 enrolled patients, and ctDNA was detected *via* high-confidence TS in 24/46 patients (52%) at baseline (Fig. 2C).

From the high-confidence prioritized variants, one mutation per patient was selected for personalized ddPCR assay development. One patient was excluded from ddPCR assay development due to insufficient DNA, leaving 25 patients evaluable for personalized ddPCR monitoring. The mutations selected for ddPCR assay design are displayed in Fig. 2D. Notably, five of the selected mutations occurred in *TP53*, the most frequently mutated gene in EAC, and all five were detectable by ddPCR, consistent with the high prevalence and clonal nature of *TP53* alterations in this disease (9). Among the 25 evaluable patients, selected mutations were detectable by ddPCR at baseline in only 9/25 cases (36%, Fig. 2C). Detection rates by ddPCR were strongly associated with clinical stage (Fisher's exact test, $p = 0.0357$), with ctDNA identified in 0% (0/2) of stage I, 17% (2/12) of stage III, and 64% (7/11) of stage IV patients. Patients with stage IV disease also showed significantly higher mutant copy numbers compared to patients with localized disease (Mann-Whitney U test, $p = 0.0087$) (Fig. 2E). Baseline ddPCR positivity was also associated with reduced overall survival (OS) (log-rank $p = 0.015$) (Fig. 2F). Overall, tumor-agnostic TS enabled personalized ddPCR monitoring in a subset of patients. However, the low baseline ddPCR detection rate, particularly in localized disease, motivated evaluation of complementary genome-wide cfDNA features that do not rely on mutation-specific tracking.

Genome-wide copy-number analysis identifies complementary tumor-associated signals

We next evaluated whether genome-wide cfDNA profiling using sWGS could detect complementary tumor-derived signals through CNAs. Using the ichorCNA tool, we estimated tumor fraction from sWGS data by identifying large-scale genomic copy-number changes characteristic of cancer genomes. At baseline, 14/46 (30%) patients were positive for ctDNA by ichorCNA (tumor fraction > 3%), including 6 cases that were negative by both TS and ddPCR (Fig. 3A). ichorCNA positivity was strongly associated with clinical stage, with all positive cases occurring in stage III or IV disease, consistent with higher tumor burden in advanced-stage patients (Fig. 3B).

Importantly, ichorCNA-based ctDNA detection was also observed in a subset of samples collected after treatment, including at recurrence, in two patients for whom personalized ddPCR assays could not be developed because no suitable plasma mutation was identified for tracking (Fig. 5D). These findings

suggest that genome-wide CNA analysis can detect tumor-derived cfDNA even when somatic mutation tracking is not feasible, supporting the clinical utility of sWGS as a complementary modality for longitudinal monitoring in low-shedding tumors such as EAC without the need for tumor tissue.

Fragmentomic analysis reveals additional tumor-associated cfDNA signals

In addition to large-scale CNAs, tumor-derived cfDNA exhibits distinctive fragmentation patterns reflecting differences in chromatin organization, nuclease activity, and cell death processes between tumor and normal cells (15). We therefore investigated whether fragmentomic features derived from sWGS data could further improve detection of tumor-associated cfDNA signals to complement the targeted and CNA-based approaches described above.

Several fragmentomic metrics were computed from sWGS data, including median insert size, proportion of short fragments (< 150 bp), short-to-normal fragment ratio, fragment score, genome-wide fragmentation profiles (DELFI), nucleosome positioning patterns, chromatin accessibility signatures using Griffin, and fragment end-motif signatures. For all features except DELFI, sample-level values were converted to Z-scores relative to cancer-free controls and ctDNA positivity was defined using feature-specific thresholds derived from the control distributions (**see Methods**). For DELFI, samples were considered positive when the correlation coefficient of genome-wide 5-Mb fragment ratios with the cancer-free reference profile was < 0.5 .

Multiple baseline fragmentomic features were significantly associated with clinical stage at diagnosis, with EAC patients with more advanced disease showing greater deviation from stage I patients or cancer-free control profiles. In particular, stage IV patients had significantly shorter cfDNA fragments as compared to those with stage I disease (Fig. 3C, Mann-Whitney U test, $p = 0.0479$), and correspondingly higher fragment scores, greater proportions of short fragments, and higher short:normal fragment ratios (**Supplementary Figure S2B-D**). Additionally, when genome-wide fragmentation profiles were assessed by calculating the ratio of short (90–150 bp) to long (151–220 bp) cfDNA fragments across 5 Mb bins, DELFI correlation coefficients were significantly lower in patients with stage IV disease than in those with stage I or stage III disease (Mann-Whitney U test; stage IV versus stage I, $p = 0.0499$; stage IV versus stage III, $p = 0.0104$), indicating greater deviation from the cancer-free fragmentation profile in advanced disease (Fig. 3D). Correlation coefficients < 0.5 were observed only in patients with stage III or IV disease.

Next, we assessed the end motifs (5' 4-mer) of cfDNA fragments across our cohort, as these patterns have been shown to reflect nuclease activity and cell death mechanisms, and thus may provide insight into the underlying cancer biology of EAC patients. Indeed, a significant increase in F-profile VI weight Z-scores was observed in stage IV patients as compared to stage I (Fig. 3E, Mann-Whitney U test, $p = 0.0050$) or stage III patients (Mann-Whitney U test, $p = 0.0207$). Additionally, a significant decrease in F-profile I weight Z-scores was observed in stage IV or III patients as compared to stage I patients

(**Supplementary Figure S2F**, Mann-Whitney U test, $p = 0.0004$ and $p = 0.0003$, respectively). Together, these opposing shifts demonstrate stage-associated alterations in cfDNA end-motif composition, with the greatest deviations observed in advanced EAC. The decrease in F-profile I and increase in F-profile VI are consistent with patterns previously reported in hepatocellular carcinoma, although these end-motif profiles have not, to our knowledge, been previously characterized in EAC (30).

Finally, to further characterize cfDNA fragmentation, we mapped read starts (fragment-ends) relative to published nucleosome positions derived from healthy plasma (26). In cancer-free controls and EAC patients with early (stage I) disease, an M-shaped distribution of fragment-ends with peaks ± 83 bp from the nucleosome center was observed, corresponding to the length of DNA wrapped around a single nucleosome plus linker DNA (~ 167 bp) (Fig. 3F). In contrast, EAC patients with stage III and IV disease exhibited a reduced enrichment of read starts at nucleosome edges and an increased frequency within nucleosome centers as compared to controls (Fig. 3F). These findings indicate altered fragment-end positioning compared to reference nucleosome coordinates in advanced EAC. We next used the Griffin tool to evaluate cfDNA coverage patterns around predefined esophageal carcinoma-associated chromatin-accessible sites derived from TCGA ATAC-seq data. Griffin-derived scores were higher in patients with stage IV disease, although the difference was not statistically significant (**Supplementary Figure S2E**).

Taken together, these findings demonstrate that sWGS-derived fragmentomic features capture multiple dimensions of tumor-associated cfDNA biology that were associated with clinical stage. Fragmentomic abnormalities were also observed in baseline samples without detectable mutations by TS or personalized ddPCR and without CNAs detected by ichorCNA, supporting the integration of fragmentomic analysis into multimodal liquid biopsy strategies for EAC (Fig. 3A).

sWGS-derived CNA and fragmentomic features are associated with survival in EAC

We next investigated whether genome-wide cfDNA features derived from sWGS were associated with clinical outcomes. Several of the genome-wide cfDNA features identified in baseline samples were significantly associated with OS (Fig. 4). Indeed, patients who were positive for ctDNA by ichorCNA at baseline had significantly reduced OS as compared to those who were negative (Fig. 4B, log-rank $p = 0.015$). Similarly, aberrant fragmentomic profiles at baseline were also associated with worse survival outcomes. Specifically, larger ratio of short-to-normal fragments, altered nucleosome distances, an increase in frequency of F-profile VI, increased fragment score, and a larger proportion of fragments less than 150 bp in length in baseline samples were all significantly associated with worse OS in EAC patients (Figs. 4C-E, **Supplementary Figure S3 C-D**, log-rank $p = 0.00079, < 0.0001, 0.027, 0.032, 0.029$, respectively). In Cox proportional hazards models adjusted for clinical stage, nucleosome-distance positivity remained significantly associated with worse OS (adjusted hazard ratio (HR), 3.36; 95% confidence interval (CI), 1.21–9.30; $P = 0.0197$), while associations for the remaining features were attenuated after adjustment for stage (**Supplementary Table S2**). Overall, these findings demonstrate

that sWGS-based CNA and fragmentomic alterations are associated with survival and support their further evaluation as prognostic biomarkers, while also highlighting the potential advantages of tumor-agnostic approaches in low ctDNA-shedding tumors such as EAC.

Multimodal cfDNA profiling increases tumor-associated signal detection in EAC

Given that targeted mutation-based, copy-number, and fragmentomic analyses each showed clinically relevant associations with clinical stage and survival, we next evaluated whether integrating these complementary signals could improve overall detection sensitivity across the samples. Concordance between the assays was partial, with only a subset of patients positive across all approaches at baseline (Fig. 5A-B). This partial concordance likely indicates that each modality captures distinct aspects of tumor-derived cfDNA biology. In particular, genome-wide CNA and/or fragmentomic abnormalities were identified in 15 patients without detectable signals through TS.

When all cfDNA features were considered together, tumor-associated signals were detected in 39/46 (85%) of patients at baseline, compared to only 24/46 (52%) by TS alone (Fig. 5A, 5C). Thus, multimodal profiling broadened baseline detection beyond targeted approaches alone. In the post-treatment setting, ddPCR, CNA, and fragmentomic analyses collectively identified tumor-associated signals in all three patients with plasma available at the time of clinical recurrence (Fig. 5D).

Importantly, these analyses were performed using tumor-agnostic approaches, without requiring prior tumor sequencing or tissue-informed assay development. Although combining complementary modalities substantially increased overall detection, sWGS-derived CNA and fragmentomic features provided the broadest source of clinically relevant information. These features therefore formed the basis of the longitudinal disease-monitoring and prognostic models evaluated in the subsequent sections. Overall, these findings support multimodal cfDNA profiling for improved tumor-associated signal detection while highlighting sWGS as an inexpensive, scalable, and tumor-agnostic strategy for EAC monitoring.

sWGS-derived cfDNA classifiers enable longitudinal disease monitoring in EAC

Whereas the preceding analyses evaluated prognostic associations between individual baseline cfDNA features and survival, in the following sections we used supervised multivariable modeling to assess sample-level classification of disease status and short-term mortality risk. Since individual genome-wide cfDNA features showed promising performance when analyzed independently, we next evaluated whether integrating multiple features could improve classification of a clinically defined disease-status endpoint. We developed a machine learning (ML) framework to integrate CNA and fragmentomic features derived from our sWGS data, adapted from Abelman et al., to address the unmet need for sensitive disease biomarkers in EAC (31).

For disease-status modeling, the outcome represented a clinically defined composite endpoint based on imaging, endoscopy, and/or biopsy findings at the time of blood collection, rather than a single narrowly standardized biological reference. Samples collected when active disease was present were classified as positive, whereas samples obtained in the absence of clinical evidence of disease were classified as negative. If disease status was uncertain, samples were excluded. Both baseline and longitudinal samples were included to reflect the heterogeneity of real-world disease monitoring. For each task, samples were partitioned at the patient level into training and test cohorts, with 35 patients in the training cohort (n = 93 samples, 27 negative, 66 positive) and 14 patients in the test cohort (n = 35 samples, 10 negative, 25 positive). Model development was performed using nested cross-validation (CV), followed by independent evaluation in the held-out test cohort.

To determine which cfDNA signals contributed most to classification performance, we evaluated models built from individual features as well as combinations of fragmentomic and copy-number features, to test whether integrating complementary signals improved classification performance. A complete list of models tested, as well as all model performances in the training/testing cohorts can be found in **Supplementary Tables S3-S5**. Among the evaluated feature combinations, several demonstrated encouraging performance in the independent held-out test cohort (Fig. 6, **Supplementary Figures S4-S5**). The full integrated model incorporating all sWGS-derived features ("fragmentomics max") achieved the highest test area under the curve (AUC, 0.84), with 64% sensitivity and 90% specificity at the training-derived Youden threshold (Fig. 6A-B). Likewise, the model combining fragment size metrics ("fragmentomics core") with end-motif features (F-profiles I and VI) also performed strongly, achieving a comparable test AUC of 0.82, although with a more balanced sensitivity and specificity ("fragmentomics core + end motifs", 76% and 80%, respectively). End motif F-profile I alone showed a slightly lower overall AUC (0.79) but achieved the highest sensitivity among the selected models (88%), albeit with lower specificity (70%). When applied longitudinally across the cohort, the fragmentomics core plus end-motif classifier demonstrated separation between samples collected in the presence and absence of clinical disease and captured dynamic changes across treatment, surveillance, recurrence, and survival events (Fig. 6C). Together, these findings suggest that genome-wide cfDNA features can support longitudinal disease monitoring in EAC, with different feature combinations offering distinct trade-offs between sensitivity and specificity depending on the intended clinical application.

ML models identify patients at high risk of short-term mortality

Given the aggressive clinical course of EAC, we next evaluated whether genome-wide cfDNA features could identify patients at high risk of short-term mortality. Using OS within 6 months of sample collection as the endpoint, we trained a survival risk classifier using the same ML framework and candidate feature combinations. Samples were split at the patient level using a 70:30 training-to-test ratio, yielding 34 training patients (n = 87 samples, 68 negative samples, 19 positive samples) and 14 test patients (n = 41 samples, 35 negative samples, 6 positive samples), with no overlapping patients

between sets. A complete list of model performances in the training/testing cohorts can be found in **Supplementary Tables S6-S7**.

In the independent held-out test cohort, several feature combinations showed encouraging performance for 6-month mortality classification, with AUCs ranging from 0.81 to 0.90 (Fig. 7, **Supplementary Figures S6-S7**). The full integrated feature set (“fragmentomics max”) and the combination of ichorCNA tumor fraction with fragment size metrics (“purity + length core”) provided the most balanced operating characteristics, each achieving 83% sensitivity and 86% specificity with a test AUC of 0.81 (Fig. 7A-B). Although some individual features also performed well, namely Griffin ESCA (AUC 0.9) and end motif F-profile VI (AUC 0.86) alone, they showed less balanced sensitivity and specificity at their corresponding operating thresholds. Thus, integrated models generally showed the most balanced classification performance, consistent with the complementary biological information captured across genome-wide cfDNA features.

To evaluate the clinical relevance of the classifier over time, we performed a 6-month landmark survival analysis in the held-out test cohort using the fixed training-derived threshold from the full integrated “fragmentomics max” model. Samples classified as positive showed significantly worse OS than those classified as negative (Fig. 7C). In a Cox proportional hazards model adjusted for clinical stage and incorporating patient-level clustering to account for repeated longitudinal samples, classifier positivity remained significantly associated with increased mortality risk (adjusted HR, 20.7; 95% CI, 2.3–183.0; $P = 0.0065$). Together, these results demonstrate that tumor-agnostic sWGS-derived cfDNA features can identify patients at elevated risk of short-term mortality and support further evaluation of genome-wide cfDNA profiling for prognostic risk stratification in EAC.

Discussion

Here, we present a comprehensive multimodal cfDNA analysis of 137 longitudinal plasma samples from 49 patients with EAC. By comparing targeted mutation-based analyses with genome-wide copy-number and fragmentomic profiling, we demonstrate that these approaches capture complementary features of tumor-associated cfDNA biology and provide clinically meaningful disease biomarkers in EAC. Combining TS, ddPCR, CNA, and fragmentomic data increased baseline tumor-associated cfDNA detection to 85% using a fully tumor-agnostic framework and identified disease-associated signals in all patients with plasma available at the time of clinical recurrence. Importantly, the most clinically informative signals were derived solely from sWGS. These genome-wide features formed the basis of the disease-monitoring and short-term mortality-risk models, supporting the potential of low-pass sWGS as a scalable strategy for EAC liquid biopsy without requiring prior tumor sequencing or patient-specific assay development.

TS detected high-confidence plasma mutations in 52% of patients at baseline, consistent with previously reported detection rates for tumor-agnostic targeted-panel approaches in EAC (10–12). TS also enabled the development of personalized ddPCR assays in a subset of patients, providing mutation-specific

quantification and prognostic information. However, only 9/25 patients with selected trackable mutations were positive by ddPCR at baseline, with detection occurring predominantly in advanced-stage disease. This illustrates the practical limitations of personalized mutation tracking when ctDNA abundance is low. In addition, suitable trackable alterations could not be identified for all patients, further restricting the proportion of the cohort eligible for personalized ddPCR monitoring. Thus, while TS and ddPCR provided valuable complementary mutation information, they were not uniformly applicable across the cohort.

The performance of genome-wide CNA and fragmentomic features as biomarkers in the cohort may reflect the underlying biology of EAC. EAC is characterized by chromosomal instability, structural variation, and extensive copy-number remodeling, while lacking a small set of universally recurrent hotspot mutations (7–9). Whereas a targeted sequencing panel interrogates only predefined genomic regions, sWGS can capture broad genomic alterations and downstream fragmentation patterns across the genome. This biological distinction may help explain why genome-wide cfDNA abnormalities were identified in patients without detectable targeted mutation-based signals. Within our cohort, multiple sWGS-derived features were associated with clinical stage at diagnosis, suggesting that these signals reflect disease burden and biology. Additionally, a larger ratio of short-to-normal fragments, altered nucleosome distances, an increase in frequency of F-profile VI, increased fragment score, and a larger proportion of fragments less than 150 bp in length in baseline samples were all significantly associated with worse OS in KM analyses in EAC patients. This lends support for the use of fragmentomic features as prognostic biomarkers in this disease. Overall, the absence of detectable plasma mutations does not necessarily indicate the absence of tumor-associated cfDNA biology. In low-shedding tumors such as EAC, genome-wide cfDNA features may capture broader disease-associated processes, including tumor burden, altered cell death, chromatin organization, inflammation, and host response, that are not reducible to single-variant detection. While these signals may not exclusively originate from tumor cells, their association with clinical stage, recurrence, and survival suggests that they remain biologically and clinically informative. This supports the value of fragmentomic approaches as complementary biomarkers that extend beyond genetic measures of tumor burden alone.

Given the strong independent performance of several genome-wide cfDNA features, we next applied ML approaches to evaluate whether integrating these signals could improve classification performance. In the held-out patient-level test cohort, genome-wide models showed encouraging performance for both clinical disease-status classification and short-term mortality risk stratification. For disease-status classification, while the full integrated model incorporating all sWGS-derived features achieved the highest test AUC, a more compact model combining fragment size metrics with end-motif features achieved a comparable AUC with a more balanced sensitivity and specificity. This is an important finding, as it suggests that parsimonious combinations of complementary cfDNA signals may offer more favorable operating characteristics than larger, more redundant feature sets in modest-sized longitudinal cohorts. Importantly, the disease-status classifier was trained against a clinically defined composite endpoint based on imaging, endoscopy, and biopsy findings rather than a single narrowly standardized biological reference. Although this introduces some label heterogeneity, it also reflects the translational

reality of longitudinal disease monitoring in routine clinical practice. For 6-month mortality classification, integrated models generally provided the most balanced sensitivity and specificity with competitive AUCs. Notably, the full integrated model and the more compact purity and fragment-length model performed similarly, further suggesting that larger feature sets may not necessarily improve performance over parsimonious combinations of complementary cfDNA signals. Some individual features also performed strongly on selected metrics, indicating that the optimal feature set may depend on whether sensitivity, specificity, or overall discrimination is prioritized. Moreover, positivity by the full integrated classifier remained significantly associated with increased mortality risk after adjustment for clinical stage and patient-level clustering for repeated longitudinal samples. However, the wide CI reflects the small test cohort and limited number of mortality events, and the estimated effect size should therefore be interpreted cautiously. Together, these findings suggest that distinct genome-wide cfDNA features may capture different aspects of aggressive disease biology and support the value of combining orthogonal genome-wide cfDNA signals while also emphasizing the importance of parsimonious model design.

Recently, the EMERALD study has demonstrated the clinical utility of monitoring alternative circulating biomarkers, namely microRNAs, for the early detection of EAC (33). These findings, together with our results, suggest that future approaches may benefit from multi-omic data integration from multiple complementary liquid biopsy analytes for increased assay sensitivity and biological significance. In addition, a recent study showed promising results for the analysis of salivary cfDNA fragmentomics features for the detection of gastric cancer (34). This suggests that incorporating non-blood biofluids into future studies could further improve the sensitivity and specificity of liquid biopsy assays in EAC.

This study had several limitations. First, the cohort size was modest, limiting formal calibration assessment, and external validation was not available, which may reduce the generalizability of our findings. Additionally, there were technical differences between the processing of EAC patient samples and the cancer-free control samples used for normalization, including differences in sequencing platform, read length, and blood collection protocols, which may have introduced variability into some sWGS-derived fragmentomic measures. Although many of the observed signals were biologically plausible and clinically correlated within the EAC cohort, future studies using prospectively matched controls and harmonized pre-analytical workflows will be important to confirm the robustness of these features.

Conclusions

Overall, this study demonstrates that multimodal cfDNA profiling improves tumor-associated signal detection in EAC and, importantly, reveals the substantial clinical information contained within tumor-agnostic sWGS data. From a single low-pass genome-wide dataset, CNA and fragmentomic features supported disease detection, longitudinal monitoring, and short-term mortality-risk stratification without requiring prior tumor sequencing or personalized mutation tracking. These findings position sWGS-

derived genome-wide cfDNA profiling as a promising, scalable strategy for liquid-biopsy monitoring in EAC and provide a foundation for validation in larger prospective cohorts.

Abbreviations

AUC

area under the receiver operating characteristic curve

cfDNA

cell-free DNA

CI

confidence interval

CNA

copy-number alteration

ctDNA

circulating tumor DNA

CV

cross-validation

DCS

duplex consensus sequence

ddPCR

droplet digital PCR

EAC

esophageal adenocarcinoma

EGA

European Genome-Phenome Archive

GEJ

gastroesophageal junction

HR

hazard ratio

INDEL

insertion or deletion

ML

machine learning

MRD

minimal residual disease

NPV

negative predictive value

OS

overall survival

PPV

positive predictive value
ROC
receiver operating characteristic
SNV
single-nucleotide variant
SSCS
single-strand consensus sequence
sWGS
shallow whole-genome sequencing
TF
tumor fraction
TS
targeted sequencing
VAF
variant allele frequency
WBC
white blood cell.

Declarations

Ethics approval and consent to participate

All patients provided informed consent in accordance with ethics protocols #2007 – 856 and #2022–8457 approved by the Research Institute McGill University Health Centre Research Ethics Board.

Consent for publication

Informed consent was obtained from all participants in this study.

Competing interests

The authors declare that they have no competing interests.

Funding

This research is funded by a Canadian Cancer Society Emerging Scholar Award (Québec) (grant #707153) to JVB, Fonds de Recherche du Québec en Santé (to JVB [#312831] and AB [#330312]). TJP is supported by the Canada Research Chair in Translational Genomics and receives funding from the Gattuso-Slaight Personalized Cancer Medicine Fund at the Princess Margaret Cancer Foundation as well as an OICR Senior Investigator Award.

Author Contribution

Conceptualization, TJP, JCL, and JVB; Resources, LF, JCL, TJP, and JVB; Data curation, AB, STF, and LSC; Software, LME, JPB, DDA, AD, and SDP; Formal analysis, AB, LME, STF, JPB, and DDA; Supervision, LF, JCL, TJP, and JVB; Funding acquisition, JCL, TJP, and JVB; Validation, AB and MN; Investigation, AB, STF, KBJ, WZ, APS, and RS; Visualization, AB and LME; Methodology, AB, JPB, TJP, and JVB; Writing original draft, AB and JVB; Project administration, CM, KD, and NB; Writing review and editing, all authors.

Acknowledgement

We would like to thank the patients for their generous participation in this study. This study was conducted with the support of the Ontario Institute for Cancer Research's Genomics Program (genomics.oicr.on.ca) through funding provided by the Government of Ontario. Figure 1 was made in part using biorender.com.

Data Availability

All sequencing files are deposited at the European Genome-Phenome Archive (EGA) under the accession numbers EGAS50000001998 (sWGS dataset) and EGAS50000001999 (TS dataset). Code necessary to reproduce the core analyses is available without restrictions at <https://github.com/pughlab/pipeline-suite> (TS, fragmentomics pipelines) and <https://github.com/pughlab/EAC-ML-pipeline> (ML pipeline).

References

1. Sung H, Filho AM, Laversanne M, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *CA: A Cancer. J Clin.* 2026;76(4):e70090.
2. Myers AL, Lin L, Nancarrow DJ, Wang Z, Ferrer-Torres D, Thomas DG, et al. IGFBP2 modulates the chemoresistant phenotype in esophageal adenocarcinoma. *Oncotarget.* 2015;6(28):25897–916.
3. Noorani A, Bornschein J, Lynch AG, Secrier M, Achilleos A, Eldridge M, et al. A comparative analysis of whole genome sequencing of esophageal adenocarcinoma pre- and post-chemotherapy. *Genome Res.* 2017;27(6):902–12.
4. Gu J, Zhang J, Zheng L, Ajani JA, Wu X, Ye Y. Serum miR-331-3p predicts tumor recurrence in esophageal adenocarcinoma. *Sci Rep.* 2018;8(1):14006.
5. Bartolomucci A, Nobrega M, Ferrier T, Dickinson K, Kaorey N, Nadeau A, et al. Circulating tumor DNA to monitor treatment response in solid tumors and advance precision oncology. *NPJ Precis Oncol.* 2025;9(1):84.
6. Azad TD, Chaudhuri AA, Fang P, Qiao Y, Esfahani MS, Chabon JJ, et al. Circulating Tumor DNA Analysis for Detection of Minimal Residual Disease After Chemoradiotherapy for Localized

- Esophageal Cancer. *Gastroenterology*. 2020;158(3):494–e5056.
7. Frankell AM, Jammula S, Li X, Contino G, Killcoyne S, Abbas S, et al. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet*. 2019;51(3):506–16.
 8. Stachler MD, Taylor-Weiner A, Peng S, McKenna A, Agoston AT, Odze RD, et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet*. 2015;47(9):1047–55.
 9. TCGA. Integrated genomic characterization of oesophageal carcinoma. *Nature*. 2017;541(7636):169–75.
 10. Bonazzi VF, Aoude LG, Brosda S, Lonie JM, Patel K, Bradford JJ, et al. ctDNA as a biomarker of progression in oesophageal adenocarcinoma. *ESMO Open*. 2022;7(3):100452.
 11. van den Ende T, van der Pol Y, Creemers A, Moldovan N, Boers D, van Berge Henegouwen MI, et al. Genome-wide and panel-based cell-free DNA characterization of patients with resectable esophageal adenocarcinoma. *J Pathol*. 2023;261(3):286–97.
 12. Kelly RJ, Landon BV, Zaidi AH, Singh D, Canzoniero JV, Balan A, et al. Neoadjuvant nivolumab or nivolumab plus LAG-3 inhibitor relatlimab in resectable esophageal/gastroesophageal junction cancer: a phase Ib trial and ctDNA analyses. *Nat Med*. 2024;30(4):1023–34.
 13. Mehta R, Rivero-Hinojosa S, Dayyani F, Ferguson J, Shariff B, Aushev VN, et al. Circulating tumor DNA informs clinical practice in patients with recurrent/metastatic gastroesophageal cancers. *Cancer*. 2026;132(1):e70242.
 14. Verschoor YL, van de Haar J, van den Berg JG, van Sandick JW, Kodach LL, van Dieren JM, et al. Neoadjuvant atezolizumab plus chemotherapy in gastric and gastroesophageal junction adenocarcinoma: the phase 2 PANDA trial. *Nat Med*. 2024;30(2):519–30.
 15. Tsui WHA, Jiang P, Lo YMD. Cell-free DNA fragmentomics in cancer. *Cancer Cell*. 2025;43(10):1792–814.
 16. Moldovan N, van der Pol Y, van den Ende T, Boers D, Verkuijlen S, Creemers A, et al. Multi-modal cell-free DNA genomic and fragmentomic patterns enhance cancer survival and recurrence analysis. *Cell Rep Med*. 2024;5(1):101349.
 17. Wong D, Luo P, Oldfield LE, Gong H, Brunga L, Rabinowicz R, et al. Early Cancer Detection in Li-Fraumeni Syndrome with Cell-Free DNA. *Cancer Discov*. 2024;14(1):104–19.
 18. Ferrier ST, Tsering T, Sadeghi N, Zeitouni A, Burnier JV. Blood and saliva-derived ctDNA is a marker of residual disease after treatment and correlates with recurrence in human papillomavirus-associated head and neck cancer. *Cancer Med*. 2023;12(15):15777–87.
 19. Wang TT, Abelson S, Zou J, Li T, Zhao Z, Dick JE, et al. High efficiency error suppression for accurate detection of low-frequency variants. *Nucleic Acids Res*. 2019;47(15):e87–e.
 20. Lheureux S, Prokopec SD, Oldfield LE, Gonzalez-Ochoa E, Bruce JP, Wong D, et al. Identifying Mechanisms of Resistance by Circulating Tumor DNA in EVOLVE, a Phase II Trial of Cediranib Plus Olaparib for Ovarian Cancer at Time of PARP Inhibitor Progression. *Clin Cancer Res*. 2023;29(18):3706–16.

21. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491–8.
22. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun.* 2017;8(1):1324.
23. Vessies DCL, Schuurbijs MMF, van der Noort V, Schouten I, Linders TC, Lanfermeijer M, et al. Combining variant detection and fragment length analysis improves detection of minimal residual disease in postsurgery circulating tumour DNA of stage II-IIIa NSCLC patients. *Mol Oncol.* 2022;16(14):2719–32.
24. 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):385–9.
25. Vanderstichele A, Busschaert P, Landolfo C, Olbrecht S, Coosemans A, Froyman W, et al. Nucleosome footprinting in plasma cell-free DNA for the pre-surgical diagnosis of ovarian cancer. *NPJ Genom Med.* 2022;7(1):30.
26. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell.* 2016;164(1–2):57–68.
27. Doebley AL, Ko M, Liao H, Cruikshank AE, Santos K, Kikawa C, et al. A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. *Nat Commun.* 2022;13(1):7475.
28. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. *Science.* 2018;362(6413):eaav1898.
29. Jiang P, Sun K, Peng W, Cheng SH, Ni M, Yeung PC, et al. Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer, Pregnancy, and Transplantation. *Cancer Discov.* 2020;10(5):664–73.
30. Zhou Z, Ma M-JL, Chan RWY, Lam WKJ, Peng W, Gai W et al. Fragmentation landscape of cell-free DNA revealed by deconvolutional analysis of end motifs. *Proceedings of the National Academy of Sciences.* 2023;120(17):e2220982120.
31. Abelman DD, Eagles J, Wong A, Shah S, Bruce JP, Pedersen S et al. Cell-free DNA Whole Genome Sequencing for Non-Invasive Minimal Residual Disease Detection in Multiple Myeloma. *medRxiv.* 2025:2025.10.24.25338566.
32. Lee M, Eng G, Handte-Reinecker A, Deshpande VS, Yilmaz OH, Gala MK. Germline Determinants of Esophageal Adenocarcinoma. *Gastroenterology.* 2023;165(5):1276–e97.
33. Miyoshi J, Mannucci A, Scarpa M, Gao F, Toden S, Whitsett T, et al. Liquid biopsy to identify Barrett's oesophagus, dysplasia and oesophageal adenocarcinoma: the EMERALD multicentre study. *Gut.* 2025;74(2):169–81.
34. Swarup N, Cheng J, Choi I, Heo YJ, Kordi M, Aziz M, et al. Multi-faceted attributes of salivary cell-free DNA as liquid biopsy biomarkers for gastric cancer detection. *Biomark Res.* 2023;11(1):90.

Figures

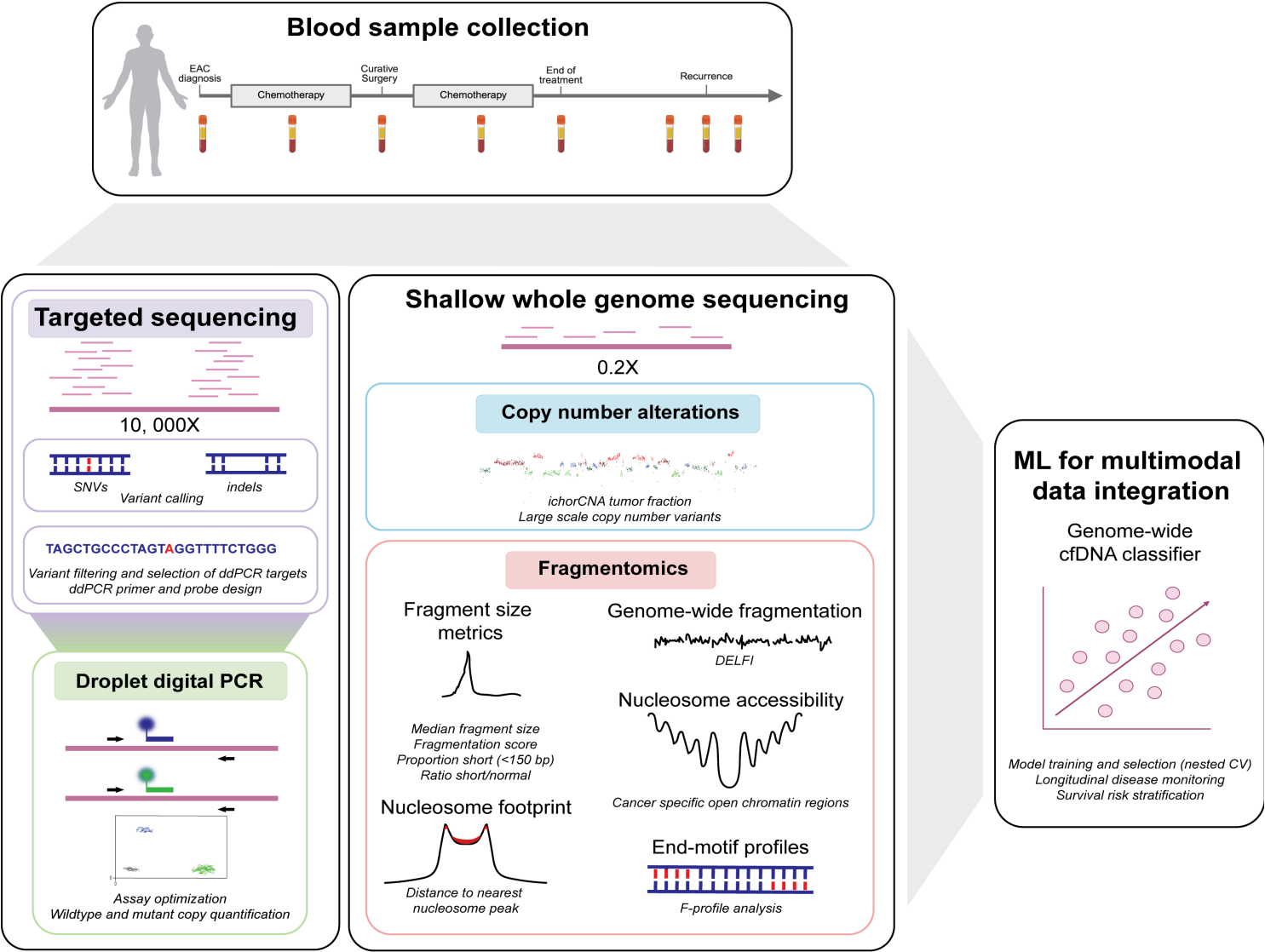


Figure 1

Multimodal cfDNA profiling framework. Schematic of the multimodal cfDNA analysis framework. Longitudinal plasma samples were collected and analyzed using deep targeted sequencing (TS), droplet digital (dd)PCR, and shallow whole genome sequencing (sWGS). TS allowed for the detection of single nucleotide variants (SNVs) and indels, and for the selection of targets for ddPCR assays. sWGS enabled the assessment of copy-number alterations (CNAs) and fragmentomic features. sWGS-derived copy-number and fragmentomic features were integrated using machine learning (ML) to develop genome-wide cfDNA classifiers for longitudinal disease monitoring and short-term mortality-risk stratification.

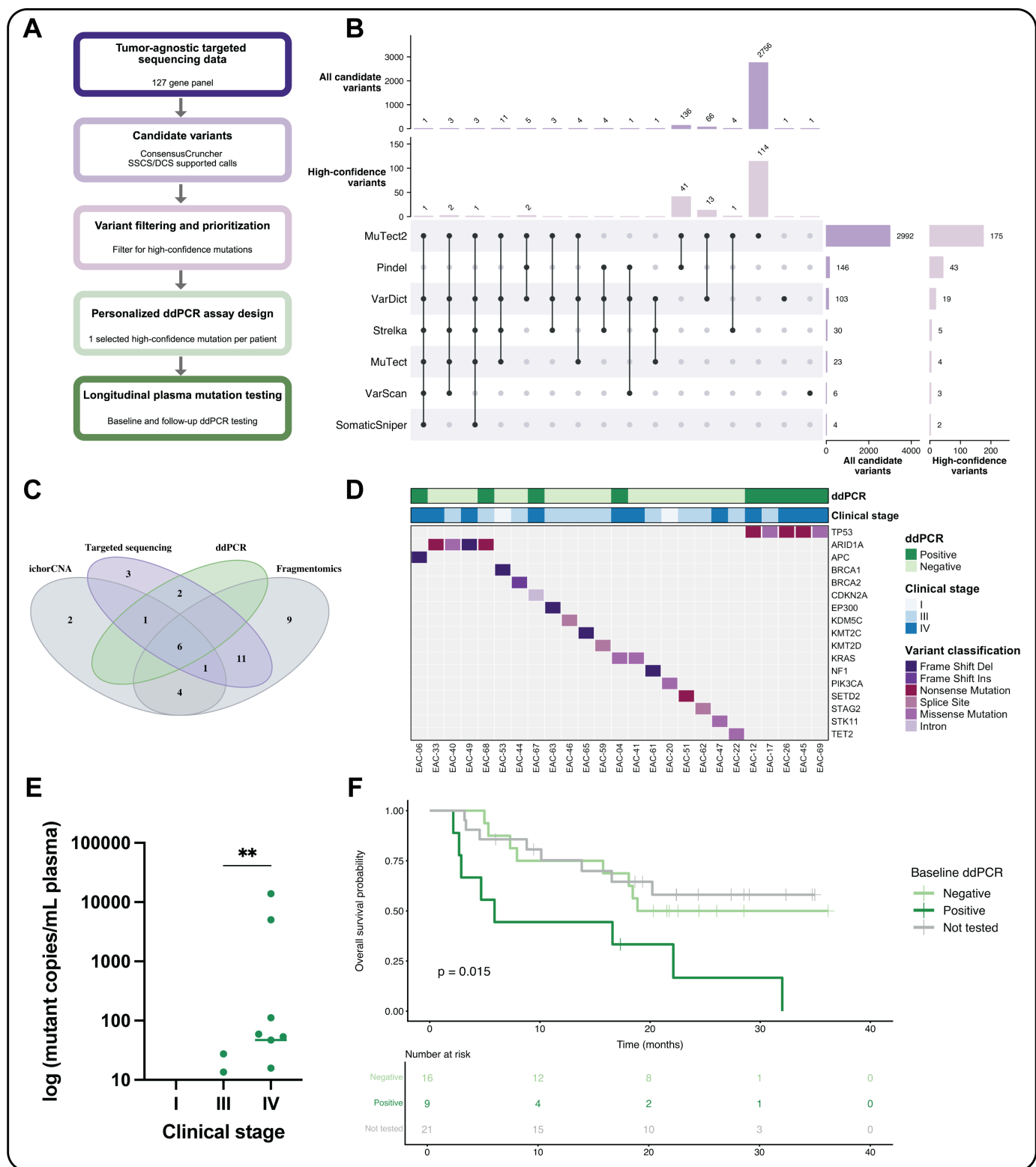


Figure 2

Tumor-agnostic targeted sequencing enables personalized ddPCR monitoring in a subset of patients. (A) Workflow used to prioritize candidate plasma somatic variants identified by tumor-agnostic TS for personalized ddPCR assay development. Candidate variants were assessed using predefined technical and biological criteria, including OncoKB oncogenic annotation, occurrence in genes recurrently altered in EAC, support from multiple annotation or variant-calling sources, and plasma variant allele frequency.

One prioritized mutation per patient was selected for personalized ddPCR assay development. **(B)** UpSet plot showing the number and overlap of candidate somatic variants identified by MuTect2, Pindel, VarDict, Strelka, MuTect, VarScan, and SomaticSniper. Vertical bars indicate the number of variants identified by each indicated combination of callers for all candidate variants (upper panel) and the high-confidence subset (lower panel). Horizontal bars indicate the total number of variants identified by each individual caller. **(C)** Venn diagram showing the overlap of positivity in different multimodal analyses at baseline sampling in the cohort. **(D)** Somatic mutations identified by TS and selected for ddPCR assay development. Each column represents a patient for whom one mutation was selected for ddPCR assay development, and each row represents the gene harboring that mutation. Mutation classes are indicated by color. Top annotations show baseline ddPCR status and clinical stage. **(E)** Plasma mutant copy number measured by ddPCR in baseline-positive samples and stratified by clinical stage. Patients with stage IV disease exhibited significantly higher mutant copy numbers than patients with localized disease. **(F)** Kaplan-Meier curves of OS stratified by baseline ddPCR ctDNA detection status. Baseline ctDNA positivity was associated with worse OS. $**p<0.01$.

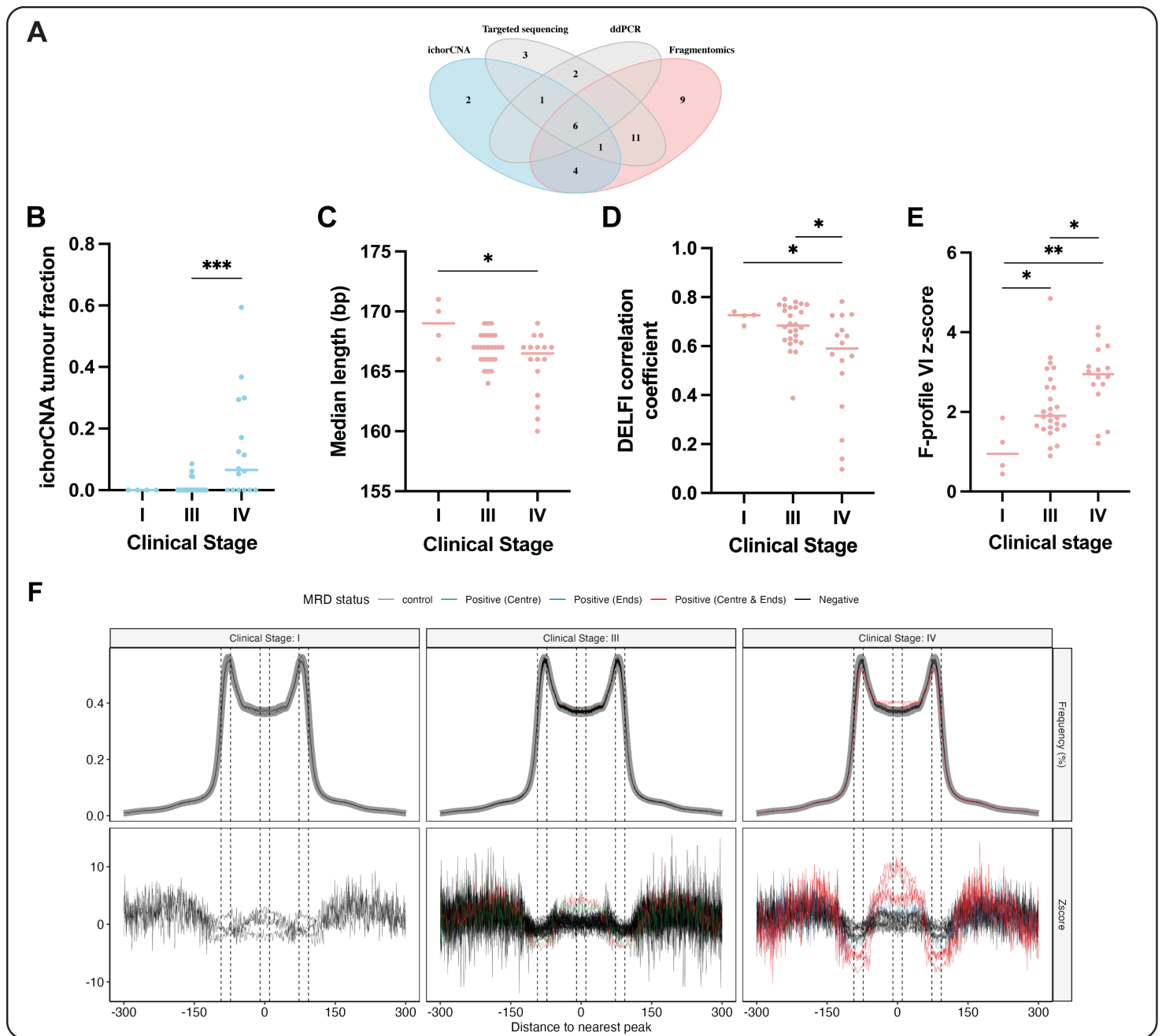


Figure 3

Baseline CNA and fragmentomic analyses are associated with clinical staging at diagnosis. (A) Venn diagram showing the overlap of positivity in different multimodal analyses at baseline sampling in the cohort. (B) Baseline tumor fraction (TF) estimates from ichorCNA stratified by clinical stage at diagnosis. Significant differences in TF were observed between stage III and IV patients (Mann-Whitney U test, $p = 0.0002$). (C) Median cfDNA fragment length across clinical stages. Stage IV patients exhibited significantly shorter cfDNA fragments compared with stage I patients. (D) Genome-wide fragment ratio correlation coefficients across clinical stages. Lower correlation values indicate greater deviation from the cancer-free reference fragmentation profile. DELFI correlation coefficients were significantly lower in stage IV compared with stage I and stage III patients. (E) Fragment-end motif analysis showing Z-scores

of F-profile VI weights across clinical stages. Increased frequencies of F-profile VI were observed in advanced-stage disease, with significant differences between stage IV and stages I and III. **(F)** Nucleosome positioning analysis of baseline cfDNA fragments. Frequency distribution (top panel) and Z-scores (bottom panel), of the distance between fragment-ends (read starts) and the closest nucleosome peak are shown in patients with stage I (left), stage III (middle), and stage IV (right) disease. Curves are coloured by alteration pattern: green (enrichment at nucleosome centres), blue (depletion at nucleosome ends), red (combined enrichment at centres and depletion at ends), and black (no alterations). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

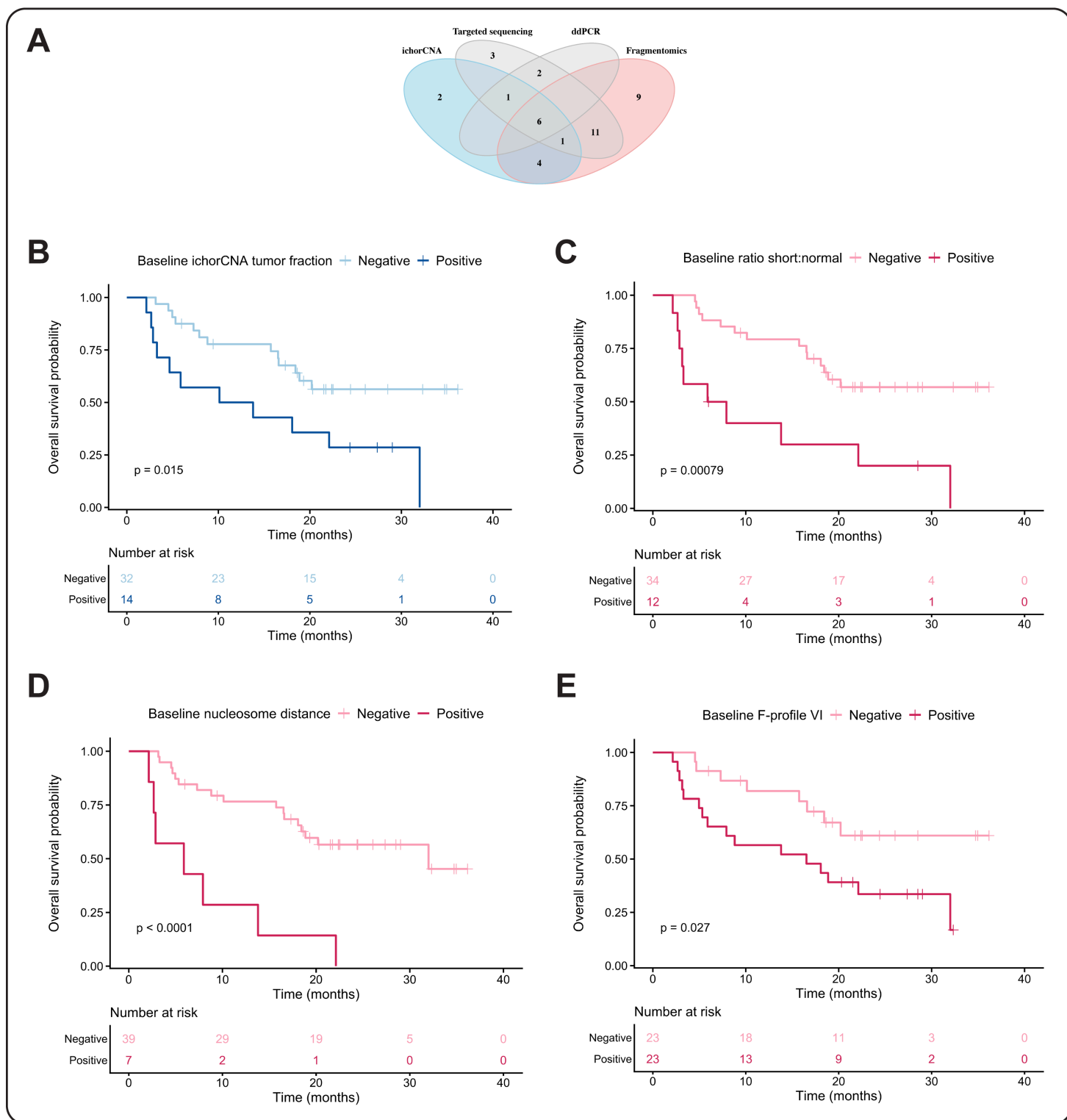


Figure 4

Baseline sWGS metrics are associated with OS. (A) Venn diagram showing the overlap of positivity in different multimodal analyses at baseline sampling in the cohort. Kaplan-Meier curves of OS stratified by (B) baseline ichorCNA, (C) baseline ratio of short-to-normal fragments, (D) baseline nucleosome distance, and (E) baseline F-profile VI.

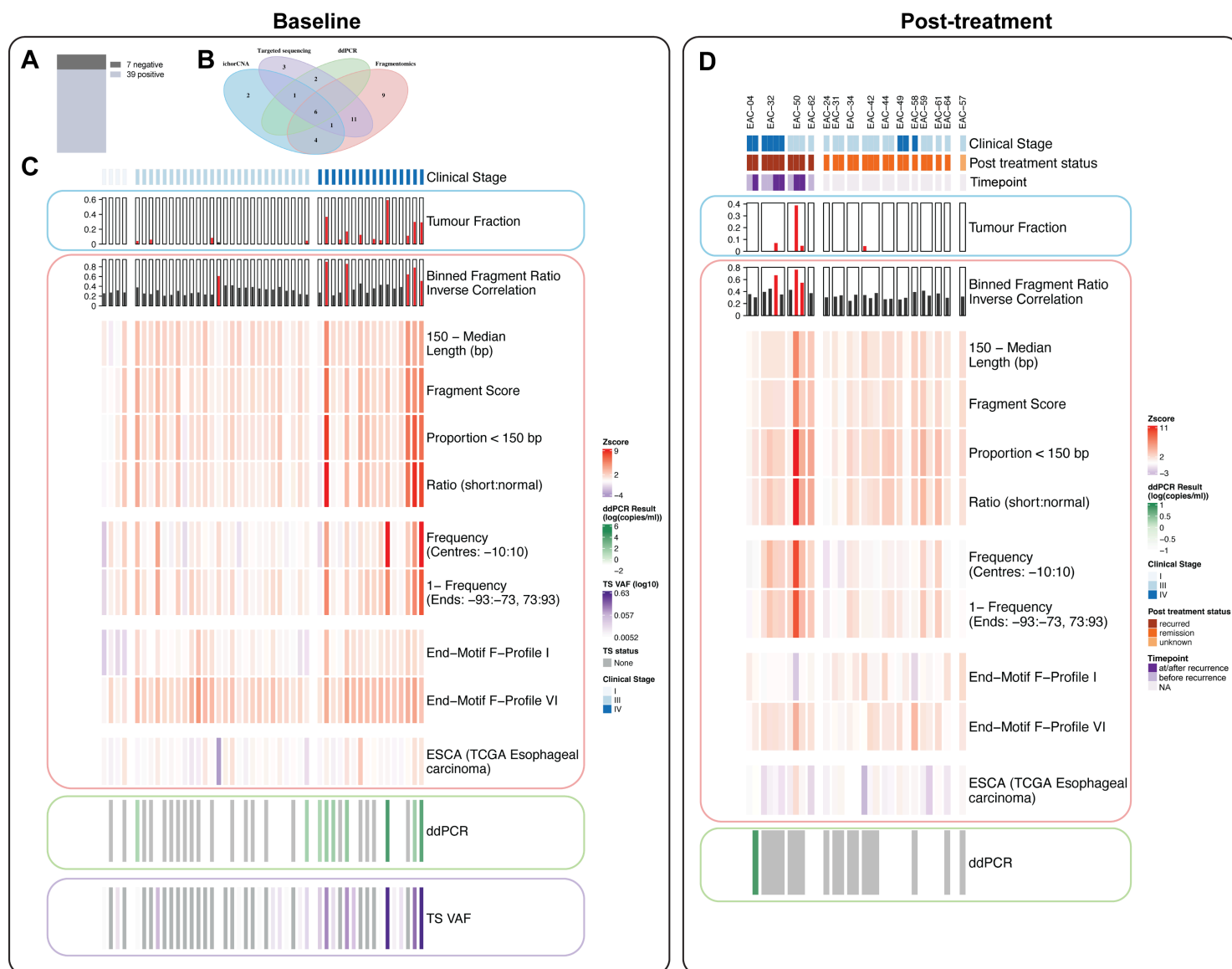


Figure 5

Integration of mutation, copy-number, and fragmentomic signals improves cfDNA detection in EAC. (A)

Proportion of patients with detectable tumor-derived cfDNA signals at baseline using any available analytical approach. **(B)** Venn diagram showing the overlap of positivity in different multimodal analyses at baseline sampling in the cohort. **(C)** Baseline multimodal data integration. Each column represents a patient's baseline sample, ordered by clinical features. Rows display integrated cfDNA-derived features including tumor fraction (ichorCNA), genome-wide fragmentation metrics (fragment size distribution, fragment score, proportion <150 bp, short-to-normal ratio), DELFI fragmentation profiles, nucleosome positioning, ESCA-associated chromatin-accessibility scores, and end-motif features, alongside mutation-based measurements (targeted sequencing variant allele frequency (VAF), and ddPCR mutant copies/mL plasma). Z-scores are normalized to healthy blood controls. This integrated view highlights concordant and complementary ctDNA signals across modalities at diagnosis. **(D)** Post-treatment multimodal data integration. Each column represents a post-treatment sample, ordered by time from diagnosis to blood sampling. The same multimodal features as in (C) are shown. Annotation tracks

include clinical stage, post-treatment status (whether the patient recurred or entered remission), and if the sample was taken pre or post recurrence. The persistence or re-emergence of ctDNA-associated signals across cfDNA features in patients with recurrence is evident, as compared to largely suppressed signals in EAC patients in remission.

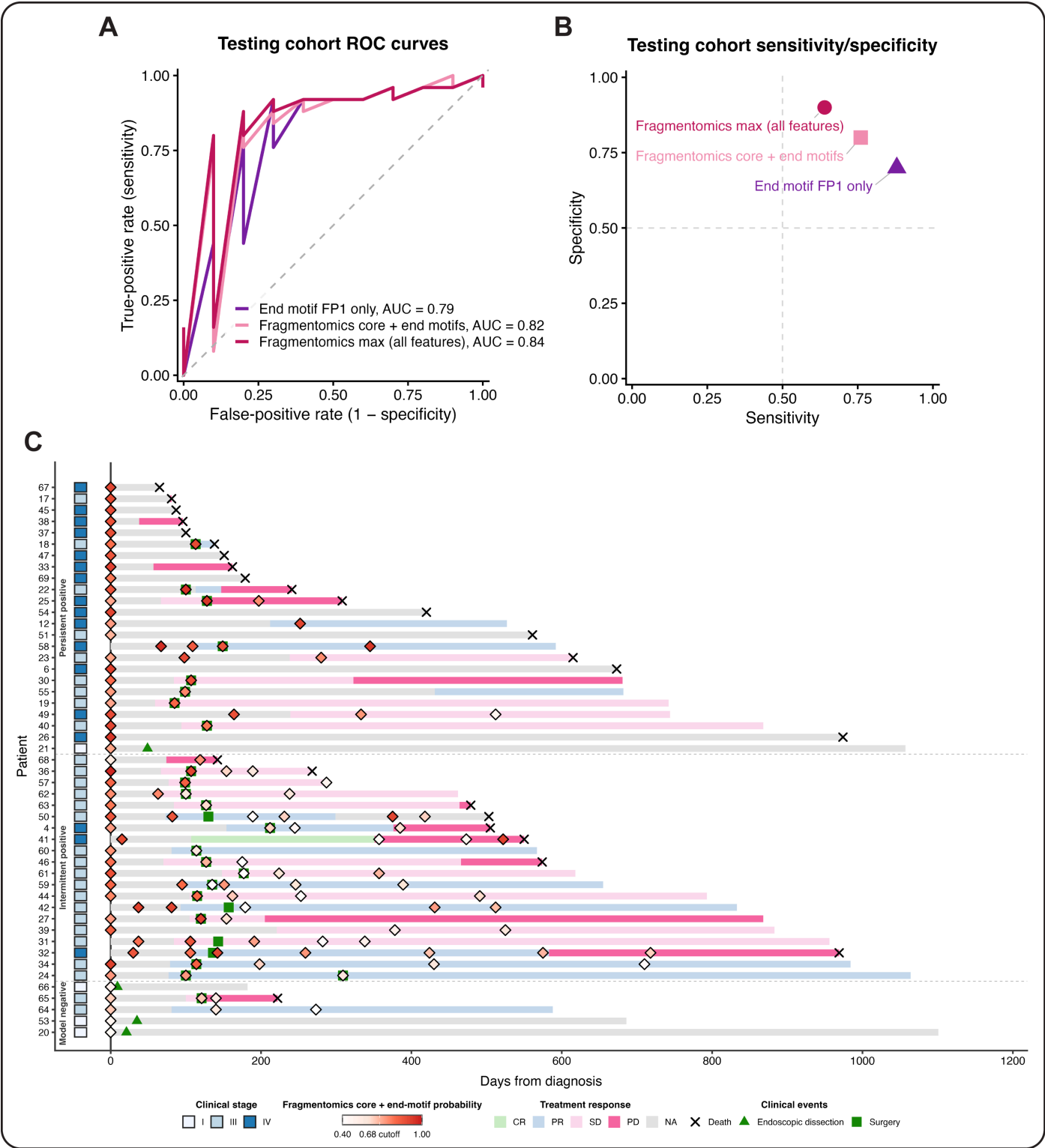


Figure 6

Genome-wide cfDNA classifier enhances longitudinal monitoring in EAC. (A) Receiver operating characteristic (ROC) curves for the top-performing genome-wide models in the independent test cohort for classification of clinical disease status. **(B)** Sensitivity and specificity of selected genome-wide models in the independent held-out test cohort at the training-derived Youden index threshold. Operating thresholds were determined in the training cohort using the Youden index and then applied unchanged to the test cohort. **(C)** Longitudinal swimmer plot showing clinical stage, treatment response according to Response Evaluation Criteria in Solid Tumors (RECIST) 1.1, surgical and endoscopic interventions, survival or follow-up duration, and cfDNA-derived disease-classifier probabilities over time. Diamonds represent individual plasma samples, with colour intensity indicating the predicted probability of clinical disease generated by the “fragmentomics core + end-motif” model. Using the model probability cutoff of 0.68, patients were grouped as persistently positive ($\geq 75\%$ of evaluable samples positive), intermittently positive ($>0\%$ to $<75\%$ positive), or model negative (0% positive). Within each group, patients were ordered from shortest to longest survival or last visible follow-up. Crosses indicate death; green triangles indicate endoscopic dissection; and green squares indicate surgery. CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; NA, not assessed.

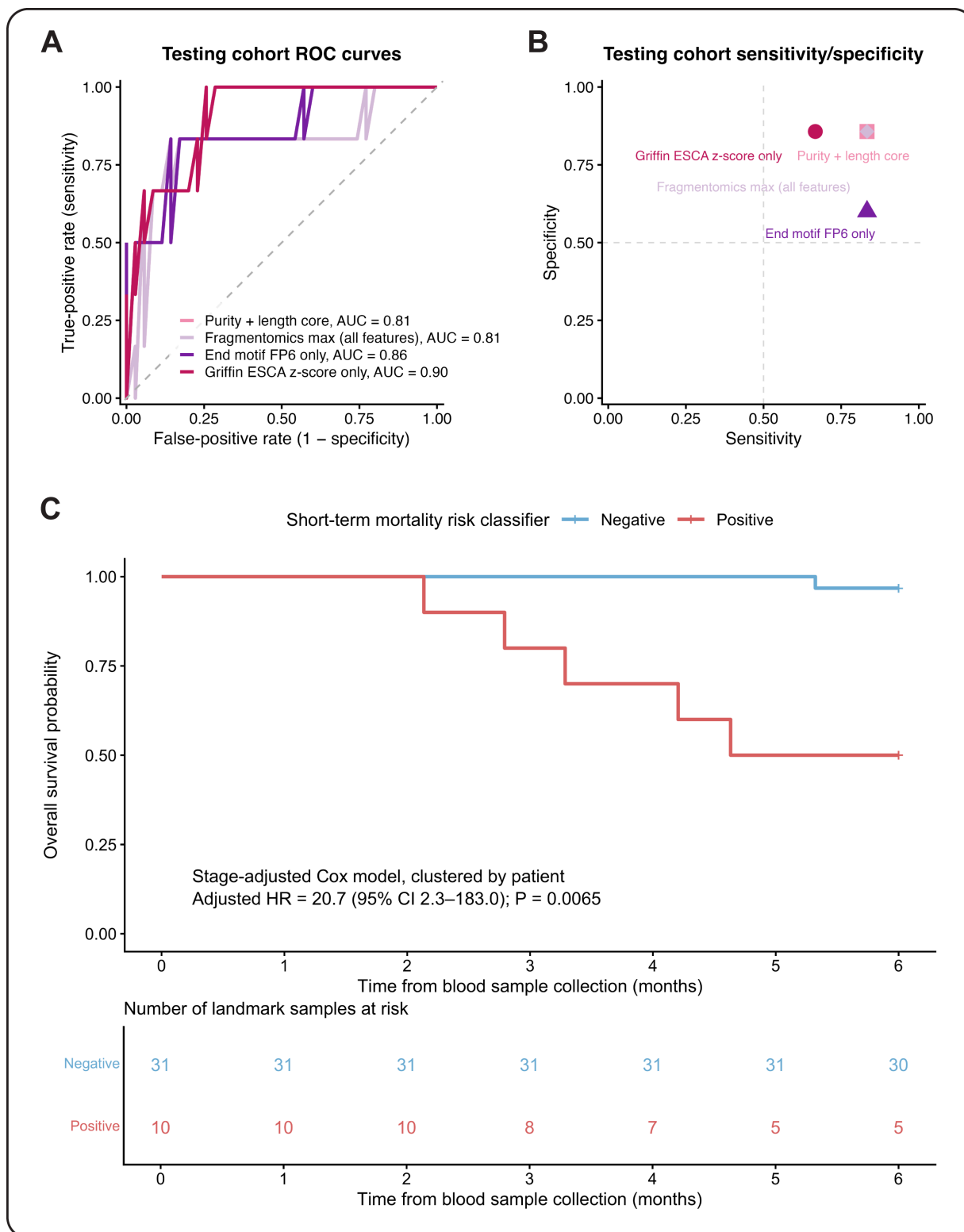


Figure 7

Genome-wide cfDNA classifier for short-term mortality risk stratification. (A) ROC curves for the top-performing genome-wide models in the independent test cohort for classification of 6-month mortality risk. (B) Sensitivity and specificity of selected genome-wide models in the independent held-out test cohort at the training-derived Youden index threshold. Operating thresholds were determined in the training cohort using the Youden index and then applied unchanged to the test cohort. (C) Unadjusted

Kaplan-Meier estimates of OS within 6 months of blood sample collection in the independent held-out test cohort, stratified by the binary classification generated by the full integrated “fragmentomics max” model using the fixed training-derived threshold. Survival time was measured from each blood sample collection and administratively censored at 6 months. The adjusted HR, 95% CI, and P value were estimated using a multivariable Cox proportional hazards model adjusted for clinical stage and incorporating patient-level clustering to account for repeated longitudinal landmark samples from the same patient. Because the test cohort contained only one patient with stage I disease and a limited number of mortality events, clinical stage was grouped as stage IV (metastatic) versus stages I–III (non-metastatic) to reduce model instability and limit overfitting. The risk table reports the number of landmark samples at risk.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryfiguresEACctDNAfinal2.pdf](#)
- [SupplementaryTableS1.docx](#)
- [SupplementaryTableS2.docx](#)
- [SupplementaryTableS3.docx](#)
- [SupplementaryTableS4.xlsx](#)
- [SupplementaryTableS5.xlsx](#)
- [SupplementaryTableS6.xlsx](#)
- [SupplementaryTableS7.xlsx](#)
- [SupplementaryTableS8.xlsx](#)
- [SupplementaryTableS9.docx](#)