The SChISM study: Circulating cell-free DNA size profiles as predictors of progression in advanced carcinoma treated with immune-checkpoint inhibitors

Linh Nguyen Phuong, Frederic Fina, Laurent Greillier, Pascale Tomasini, Jean-Laurent Deville, Romain Zakrasjek, Lucie Della-Negra, Audrey Boutonnet, Frédéric Ginot, Jean-Charles Garcia, et al.

The SChISM study: Circulating cell-free DNA size profiles as predictors of progression in advanced carcinoma treated with immune-checkpoint inhibitors

Linh Nguyen Phuong[1,2], Frédéric Fina[3], Laurent Greillier[1,2,4], Pascale Tomasini[4], Jean-Laurent Deville[5], Romain Zakrajsek[1,2], Lucie Della-Negra[1,2], Audrey Boutonnet[6], Frédéric Ginot[6], Jean-Charles Garcia[6], Sébastien Benzekry[1,2,*], Sébastien Salas[1,2,5,*]

[1] COMPutational pharmacology and clinical Oncology, Centre Inria d'Université Côte d'Azur, Marseille, France

[2] Cancer Research Centre of Marseille, Insitut Paoli-Calmettes, Inserm UMR1068, CNRS UMR7258, Aix Marseille University UM105, Marseille, France

[3] ID-Solutions oncology, Marseille, France

[4] Multidisciplinary Oncology & Therapeutic Innovations Department, Assistance Publique-Hôpitaux de Marseille, Nord Hospital, Aix Marseille University, Marseille, France

[5] Assistance Publique-Hôpitaux de Marseille, Timone Hospital, Aix Marseille University, Marseille, France

[6] Adelis Technologies, Labège, France

* = joint senior authors

ABSTRACT

Background

Circulating cell-free DNA (cfDNA) offers a promising noninvasive way to predict resistance to immune-checkpoint inhibitors (ICI), for which robust biomarkers are still lacking.

Methods

The SChISM (Size CfDNA Immunotherapy Signature Monitoring) proof-of-concept study (NCT05083494) collected baseline plasmatic cfDNA size profiles from 126 ICI-treated advanced carcinomas, quantified using the innovative, patented and standardized BIABooster device (Adelis). Fragmentome-derived variables and standard clinical variables (including neutrophils-to-lymphocyte ratio, NLR) were analyzed for univariable associations with early progression (EP, progression at first imaging) and progression-free survival (PFS). Multivariable analysis was carried through both unsupervised and supervised learning. Twenty-six variable selection methods combined with 11 models were benchmarked to derive a multivariable predictive model relying on a minimal subset of variables.

Results

Higher cfDNA concentration and high quantities of short fragments (111-240 base pairs (bp)) were associated with poor response and reduced PFS, unlike long fragments (> 300 bp). The proportion of fragments longer than 1650 bp exhibited the strongest association, with non-EP odds ratio = 0.39 [95% CI: 0.25-0.62] and PFS hazard ratio = 0.54 [95% CI: 0.42-0.68]. Unsupervised learning identified four patient clusters significantly associated with EP ($p = 0.004$, Pearson's Chi-squared test) and PFS ($p = 0.001$, log-rank test). The multivariable machine learning analysis identified a subset of nine variables that shown greater performances in a logistic regression model ($AUC_{signature} = 88.5 \pm 3.3\%$, EP positive predictive value $PPV_{signature} = 69.4 \pm 7.49\%$) compared to single marker ($AUC_{R_{[>1650]}} = 73.6 \pm 3.70\%$, $PPV_{R_{[>1650]}} = 55.8 \pm 7.46\%$, $AUC_{NLR} = 68.9 \pm 5.02\%$, $PPV_{NLR} = 52.6 \pm 8.75\%$).

Conclusions

cfDNA size profiles significantly associated with progression and PFS during ICI, outperforming the routinely used markers.

Trial registration: (NCT05083494), date of registration: 2021-10-19.

Keywords: cell-free DNA; size profile; fragmentomics; cancer; biomarker; prediction of immunotherapy resistance

What is already known on this topic: There are currently no standardized biomarkers with sufficient accuracy to reliably predict resistance or response to immune checkpoint inhibitors. Commonly used markers such as PD-L1 expression and the neutrophil-to-lymphocyte ratio (NLR) show limited and inconsistent predictive performance across cancer types (PD-L1 positive predictive value (PPV): 34%; NLR PPV: 53%).

What this study adds: Long cfDNA fragments (>1650 bp) are better associated with ICI response and PFS than NLR and PD-L1. A nine-variable signature combining cfDNA size-profile features with clinical factors reached 69.5±7.5% PPV for progression prediction.

How this study might affect research, practice or policy: The BIABooster device enables fast, non-invasive cfDNA profiling from minimal plasma without DNA extraction or sequencing.

INTRODUCTION

Circulating cell-free DNA (cfDNA) sampling through “liquid biopsy” has emerged as a non-invasive, systemic biomarker for cancer monitoring. cfDNA consists of DNA fragments released into the bloodstream by blood (1)—white blood cells, erythrocyte progenitors, vascular endothelial cells (2)—and tumor cells through apoptosis, necrosis, lysis, or active secretion (3,4). With a short half-life (15 minutes-2 hours (5)),

cfDNA reflects real-time snapshot of system activity. In cancer, its plasmatic concentration rises and fluctuates with tumor burden, metastatic spread, and treatment dynamics (6,7). Specifically, both the genetic attributes and, more recently, the fragmentation patterns of cfDNA have shown promise in monitoring immunotherapy, since traditional biomarkers such as Programmed Death-Ligand 1 (PD-L1) protein expression and Tumor Mutation Burden (TMB) often fail to accurately predict treatment resistance in advanced cancers, lack standardization, and vary across tumor types and patient populations (8,9).

Most cfDNA work in immuno-oncology has focused on circulating tumor-derived DNA (ctDNA) sequencing (10) to detect somatic genomic alterations (mutations, TMB, and microsatellite instability) used as predictors of immunotherapy outcome. However, even in highly mutated cancers, approximately 20% of patients lack detectable ctDNA alterations (11). Because wild-type cfDNA is predominantly shed by immune cells (2,3), growing interest has shifted toward global cfDNA features. In 2015, Ivanov et al. showed that fragment patterns differed across cancer patients and tissue of origin (12). Several studies have since investigated various fragmentation features, such as end motifs, nucleosome footprints, and fragment sizes, with enrichment in shorter fragments (90-150 bp) among cancer patients (13). Fragment-size analysis thus offers a non-invasive approach for evaluating treatment response independently of molecular targets, histology, or treatment modality (14).

The ongoing SChISM (Size cfDNA Immunotherapy Signature Monitoring) proof-of-concept study introduces a novel approach leveraging an innovative, patented, standardized, reproducible, and low-cost method for cfDNA quantification (BIABooster (15,16), Adelis technologies, France). This technology, based on the micro-laboratory for DNA analysis and separation (μLAS) principle, enables highly sensitive detection

and accurate size profiling directly from small plasma volumes (~100 µL) without DNA extraction, representing significant operational advantages for therapeutic monitoring. This approach offers a major advantage for plasma cfDNA, which is scarce and has variable fragment sizes, especially in cancer. Its strong correlation with digital PCR measurements (15), and its capacity to detect biologically meaningful differences in cfDNA concentration and short-fragment abundance in carcinoma for early cancer detection(17), support strong analytical validation for this platform and its relevance for liquid-biopsy applications.

SChISM aims to determine whether cfDNA fragment-size profiles can identify resistance to immune checkpoint inhibitors (ICI). Specifically, we evaluated the predictive and prognostic performance of pre-treatment cfDNA size profiles in routine-care patients to anticipate early progression (EP) and progression-free survival (PFS) following immunotherapy across four carcinomas: non-small cell lung cancer (NSCLC), head and neck squamous cell carcinoma (HNSCC), urothelial carcinoma (UC), and clear cell renal cell carcinoma (ccRCC).

METHODS

Study design and patient population

SChISM is a prospective, multicenter, collaborative, non-interventional clinical study conducted across departments of the Assistance Publique–Hôpitaux de Marseille (AP-HM). cfDNA analysis was performed using BiaBooster™ technologies to investigate predictive and prognostic value in 128 patients with metastatic and/or recurrent carcinomas (NSCLC, HNSCC, ccRCC, and UC), initiating treatment with ICIs as standard care between April 2021 and July 2023. Follow-up extended until radiological progression under ICI therapy or 12 months of treatment. Blood was collected

immediately before first ICI-infusion. The final analytic cohort included baseline data from patients meeting inclusion criteria with adequate follow-up.

Clinical and biological data

Clinical and biological data were prospectively collected, including age, sex, Eastern Cooperative Oncology Group (ECOG) performance status, tumor type, and disease stage at immunotherapy initiation. Treatment response was assessed every three months by contrast-enhanced computed tomography (CT) scans, interpreted using iRECIST criteria (18). Sum of largest diameters (SLD) was computed from these CT scans. Dates of radiological progression, last follow-up, or death were recorded for survival analyses. Routine laboratory parameters collected at baseline included complete blood count, serum lactate dehydrogenase (LDH), serum creatinine, and liver function tests (AST, ALT, ALP, and bilirubin). LDH and neutrophil-to-lymphocyte ratio (NLR) were included as widely studied nonspecific biomarkers with prognostic and predictive relevance in immunotherapy (19,20). PD-L1 expression—combined positive score (CPS) and tumor proportion score (TPS)—, was collected when available in routine practice. Data were entered into a secure anonymized database and monitored regularly.

Endpoints

EP was defined as confirmed progression at the first radiological assessment (3 months after treatment initiation). PFS was defined as the interval between first ICI infusion and progression, death, or last follow-up. Progression date corresponded to radiological confirmation or last available follow-up.

Regulatory and Ethical Aspects

The study complied with the Declaration of Helsinki, Good Clinical Practice, and French regulations. It was classified as Category 2 research under the Jardé Law and

approved by the national ethics committee, with registration through the French Ministry of Health. AP-HM, as sponsor, ensured data protection regulations (GDPR) compliance through secure anonymized electronic Case Report Forms. All participants provided written informed consent. The study is registered as NCT05083494.

Blood collection and cfDNA quantification

The certified AP-HM Biobank (CRB, NFS 96-900, ISO 9001:2015) managed plasma processing and storage. Peripheral blood (Roche cfDNA Collection Tube) was centrifuged at 1,600×g for 10 minutes at room temperature. Plasma supernatant was collected into a 15 mL conical tube and subjected to a second centrifugation at 4,500×g for 10 minutes to remove residual debris. Plasma was transferred into a new 15 mL tube and aliquoted in 300 to 500 µL volumes, then stored at –80°C. cfDNA analysis was performed at Adelis (Labège, France) using BIABooster™ technology (15,16).

BIABooster workflow

Experiments were carried out with a G7100A CE system (Agilent Technologies, Germany) equipped with a Zetalif fluorescence detector (Adelis, France) and a BIABooster capillary device (Adelis, # 16-BB-DNA/11, France). DNA was concentrated at the junction of two capillaries of different diameters using dual hydrodynamic and electrokinetic actuation, allowing the removal of salts and proteins, which enables in-line purification and size-dependent migration. Gel electrophoresis quantified fragments, providing cfDNA size distribution with 10 fg/µL sensitivity (16).

Plasma was pretreated with lysis buffer (56°C, 2h, 900rpm) to release cfDNA from vesicles and protein complexes. One µL was injected into the BIABooster device for cfDNA analysis. Migration curves plotted fluorescence intensity against time (Figure 1A). Analytics software converted these into concentration and fragment size using a DNA ladder. The device gives reliable size and concentration measurements for DNA

fragments between 75 and 1650 bp; fragments outside this range were reported as relative arbitrary units (r.a.u.) due to unreliable size estimation, but still provided a relative quantitative indication of fragment abundance.

The analytical method's robustness was previously assessed through repeatability and reproducibility study (15). Reproducibility was evaluated by varying the instrument, capillary device, buffer and ladder batches, time, and operator, yielding standard deviations of 1-1.8 bp for the mono- and dinucleosomal peaks and a 12% coefficient of variation for cfDNA concentration.

A cfDNA curve is defined as the concentration of fragments (pg/µL) according to the fragment size (bp). Twelve quantitative variables were derived from each curve, corresponding to nucleosomal multiples (Figure 1B):

- P_1 and P_2 (bp): the first and second peak's position, corresponding to the most frequent size of the fragments originating from mono- and dinucleosomes, respectively.
- $P_2 - P_1$ (bp): the difference between P_2 and P_1 in bp.
- HP (pg/µL): the height of the first peak.
- HW (pg/µL): the left half-width of the first peak at mid-height. This variable was designed to capture the expected larger fragmentation of patients compared to healthy individuals (13).
- C_{TOT} (pg/µL): the global concentration defined as the area under the cfDNA curve between 75 and 1650 bp computed by the trapeze method.
- The relative concentration in pg/µL of cfDNA fragments measuring between x_i and x_j bp, to the total concentration C_{TOT}, defined as the following absolute concentration

$$R_{[x_i,x_j]} = \frac{C_{[x_i,x_j]}}{C_{TOT}}$$

for $[x_i, x_j] \in \{[75,111], [111,240], [240,370], [370,580], [580,1650]\}$.

- $R_{<75}$ and $R_{>1650}$ (r.a.u.): the relative quantity of cfDNA fragments of less than 75 bp and greater than 1650 bp, respectively, to the total concentration C_{TOT}.

An optimization algorithm was used to identify the optimal size-profile segmentation for EP classification, leading to the identification of two novel cfDNA fragment-size intervals, $[160,220]$ and $[280,320]$ bp. For these intervals, curves were first power transformed (0.15) to enhance detection of inter-patient differences.

Preprocessing

Two patients were excluded due to extreme pre-treatment cfDNA concentration value ($|C_{TOT} - \mu| > 1.5 \cdot \sigma$, with μ the mean and σ the standard deviation of cfDNA concentration). Variables following a log-normal distribution (including NLR, continuous TPS, HP, C_{TOT}, $R_{<75}$, $R_{[75,111]}$, $R_{[160,220]}$, $R_{[280,320]}$, $R_{[580,1650]}$, $R_{>1650}$, and HW) were log-transformed.

Statistical univariable analysis

Patients with missing values were removed and continuous features were z-transformed. All statistical computations used *R* version *4.4.0*. Pearson's Chi-squared and Welch t-tests compared respectively categorical and continuous features with EP status. Associations with EP and PFS were assessed with logistic (*stats::glm 4.4.0*) and Cox regression (*survival::coxph 3.8.3*), first univariable (UV) before then adjusted (MV) for clinical confounding features: age, sex, ECOG status, tumor type, and NLR. Survival was displayed with Kaplan-Meier curves (*survival::survfit 3.8.3*). Continuous features were optimally dichotomized using *survminer::surv_cutpoint 0.5.0*, maximizing the log-rank statistic while ensuring a minimum of 20% of patients per group. Pearson correlation analysis was computed between all continuous variables (*corr::correlate 0.4.5*).

Multivariable machine learning

Unsupervised learning

Unsupervised hierarchical clustering (*ComplexHeatmap* 2.22.0) was performed using Euclidian distance and Ward linkage to cluster both patients and features (cfDNA features, NLR, LDH, TPS, and age).

Supervised learning

Multivariable supervised learning was performed incorporating control variables that demonstrated statistical significance (NLR, ECOG, tumor type, sex), in conjunction with cfDNA variables. This comprehensive dataset was initially subjected to a decorrelation process iteratively removing variables with largest variance inflation factor (VIF), until none had VIF ≥ 5 (21). Missing values (2.8%) were imputed using the median and continuous variables were z-transformed. Categorical variables were encoded into binary variables: tumor type was converted into NSCLC or not, and ECOG status was categorized into 0-1 or ≥2. These steps were trained on the in-bag samples and applied to the out-of-bag samples in the machine learning pipelines described below.

To identify a minimal feature set predictive of EP, we conducted a benchmark of 27 feature selection methods, each combined with 11 machine learning classifiers (including linear, tree-based, and neural network architectures), within an optimism-correcting bootstrapping framework (22). Each combination was assessed both for the stability of the selected subset (23) and predictive performance. This was performed over 100 bootstrap resamples, using the package *ROOFS* with *Python* version *3.10*, accessible at https://gitlab.inria.fr/compo/roofs (24). Predictive performances were evaluated for accuracy, specificity, sensitivity, positive predictive value (PPV), and

negative predictive value (NPV), all estimated using the .632+ bootstrap correction method to account for optimism bias (25).

RESULTS

Patient characteristics

The data comprised 126 patients with advanced/metastatic carcinoma: NSCLC (n=68), UC (n=10), HNSCC (n=34), and ccRCC (n=14). Mean age was 64 years, with 73% of male (Table 1). Patients received first-line (n=100), second-line (n=25), or third-line immunotherapy (n=1): nivolumab (n=38), pembrolizumab (n=87), or atezolizumab (n=1). Concomitant chemotherapy was administered in 48 patients (43 NSCLC, 5 HNSCC, all first-line), targeted therapy in 11 (9 ccRCC, 1 UC, 1 HNSCC), and dual checkpoint blockade (nivolumab/ipilimumab) in 2 ccRCC patients.

PD-L1 CPS was available in 56% of HNSCC patients and categorized as <1, 1-19, and ≥20%[1]. TPS was available in 91% of the NSCLC patients and categorized as <1%, 1-49, and ≥50% (27). Pre-treatment LDH was available for 56%, SLD for 92%, and NLR for 87% of the cohort.

Median PFS was 9.53 months [95% confidence interval (CI): 6.8-NA] (Table 1), with median follow-up of 16 months [95% CI: 13.6-21.5]. Progression occurred in 69 patients (55%), including 44 patients (34.9%) progressing within three months. PFS varied by tumor type (Figure S1). EP proportion was 70% in UC, 59% in HNSCC, and 21% in NSCLC/ccRCC patients (Figure S1). One pseudoprogression was reclassified as non-EP after subsequent stability. Median OS was 14.4 months [95% CI: 12–NA] (Figure S2).

cfDNA size profiles characteristics

The baseline cfDNA size distribution revealed a median primary peak at 157 bp (range: 132-160) (Table S1, Figure S3), and a secondary peak at 307 bp (range: 274-326), reflecting mono- and dinucleosomes (Figure 1B-C). A median 58% (range: 28-85) of fragments were 111-240 bp. Considerable inter-patient variability was observed across all fragment sizes (Figure 1C). cfDNA concentration was similar between EP and non-EP across fragments below 580 bp, while fragments between 580 and 1650 bp represented 8% (range: 1-17) of the total cfDNA in EP versus 11% (range: 2-36) in non-EP patients (Figure S3). This difference was modest on size profiles (Figure 1B) because of the concentration scale but apparent on fluorescence-time curves (Figure 1A). High-molecular weight (HMW, >1650 bp) fragments averaged 0.031 r.a.u. (range: 0.005-0.086) in EP versus 0.05 r.a.u. (range: 0.006-0.165) in non-EP patients. Although low in abundance, signal quality was confirmed through DNA10K validation in a 24-sample sub-analysis (Figure S4).

cfDNA fragmentome is associated with progression and PFS

Higher cfDNA concentration (C_{TOT}) was associated with EP (AUC=0.65, Table 2) and patients with $C_{TOT} > 14$ pg/µL experienced shorter PFS (C-index=0.62, Figure 2A). In MV, C_{TOT} remained significantly associated with EP (OR 2.0 [95% CI: 1.1–3.4], $p = 0.006$) and PFS (HR 1.8 [95% CI: 1.3–2.3], $p < 0.0001$).

Short fragments $R_{[160,220]}$ were associated with EP in both UV and MV analyses (MV OR 2.1 [95% CI: 1.2-3.7], $p = 0.009$, AUC=0.7, Table 2, Figure 2A), and were strongly associated with shorter PFS in both UV and MV analyses (MV HR 1.8 [95% CI: 1.4-2.4], $p < 0.0001$, C-Index=0.66). Dinucleosomal fragments ($R_{[280,320]}$) showed similar results (MV OR 2.1 [95% CI: 1.2-3.8], $p = 0.01$, AUC=0.67; MV HR 1.7 [95% CI: 1.3-2.3], $p = 0.0001$, C-Index=0.64).

298 Conversely, longer fragment features (>370 bp) were associated with favorable
299 outcomes. $R_{>1650}$ achieved strongest performance (AUC=0.73 and C-index=0.69).
300 Higher quantity of these longest fragments was significantly associated with non-EP
301 (t-test: $p < 0.0001$) and longer PFS (log-rank test: $p < 0.0001$) (Table 2, Figure 2A),
302 both in UV and MV analyses (MV OR 0.28 [95% CI: 0.13-0.6], $p = 0.0009$; MV HR
303 0.44 [95% CI: 0.33-0.58], $p < 0.0001$). $R_{[580,1650]}$ showed similar trends (AUC=0.65, C-
304 Index=0.64). An increase of 1% in the relative amount of $R_{[580,1650]}$ corresponded to a
305 ~50% reduction in odds and hazards of progression. The difference between first and
306 second peak's position showed the second-best performance (AUC=0.72, C-
307 index=0.67). It was associated with non-EP (t-test: $p < 0.0001$) and longer PFS (log-
308 rank test: $p < 0.0001$) (Table 2, Figure 2A), both in UV and MV analyses (MV OR 0.36
309 [95% CI: 0.18-0.73], $p = 0.005$; MV HR 0.6 [95% CI: 0.44-0.83], $p = 0.002$). P_2 was
310 associated with non-EP (AUC=0.64; MV OR 0.46 [95% CI: 0.25-0.85], $p = 0.01$)
311 analyses, and a longer PFS (MV HR 0.59 [95% CI: 0.44-0.8], $p = 0.0006$) in UV and
312 MV analyses (Table 2, Figure 2A). Among feature ratios, $\frac{R_{>1650}}{R_{[160,,280]}}$ achieved the best
313 performance for association with EP (AUC=0.74) and PFS (C-index=0.69), equivalent
314 to our single best variable $R_{>1650}$.
315 Among clinical and biological variables, ECOG status, sex, and NLR were significantly
316 associated with EP and PFS (Table 2). NLR demonstrated was strongly associated
317 with EP (AUC=0.69; UV OR 2.4 [95% CI: 1.5-4], $p = 0.0006$) and shorter PFS (C-
318 index=0.63; UV OR 1.6 [95% CI: 1.2-2.2], $p = 0.002$, Figure S5). ECOG status ≥2 had
319 a markedly higher risk of EP (UV OR 6.0, 95% CI: 2.1–17, $p = 0.0009$) and poor PFS
320 (UV HR 3.3 [95% CI: 1.8–6.2], $p = 0.0001$).
321 Among patients with available NLR data (87%), $R_{>1650}$ and $\mathrm{P}_2 - P_1$ identified additional
322 11 out of 37 (30%) EP patients and 18 out of 73 (25%) non-EP patients missed by NLR

(Figure 2B). The cfDNA variables had weak correlations with clinical predictors (Figure 2C), suggesting independent predictive value. Notably, correlations with NLR were all very low (max $|r| = 0.16$). The cfDNA variables were correlated with each other; $R_{[>1650]}$ demonstrated high correlation with short fragments ($R_{[111,240]}$, $R_{[160,180]}$, $R_{[280,320]}$), suggesting that HMW fragments represent the reciprocal of short-fragment depletion. However, the superior explanatory power of $R_{[>1650]}$ over the short fragment variables in the regression models suggests that $R_{[>1650]}$ provides additional information relevant to EP status.

Unsupervised learning reveals patient clusters associated with progression and PFS

Unsupervised hierarchical clustering (Figure 3A) identified three main feature clusters characterized by: (1) the total cfDNA concentration (C_{TOT}) and mono- and dinucleosomal features ($R_{[111,240]}$, $R_{[160,220]}$, $R_{[280,320]}$, and HP), grouped with LDH, (2) very short fragments (HW, $R_{<75}$ and $\mathrm{R}_{[75,111]}$), and (3) long fragments ($R_{[240,370]}$, $R_{[370,580]}$, $\mathrm{R}_{[580,1650]}$, and $\mathrm{R}_{>1650}$) and the first two peaks' positions (P_1, P_2, and $P_2 - P_1$), grouped with age, PD-L1 TPS, and NLR. Notably, conventional variables, particularly NLR, exhibited distinct variation patterns from cfDNA features.

Four patient clusters (A-D) emerged with distinct fragmentome profiles. Clustering was unsupervised yet significantly associated with EP (Pearson's Chi-squared test $p = 0.004$) and PFS (log-rank test $p = 0.001$, Figure 3B). More specifically, clusters B (n=45) and C (n=13) were associated with good prognosis (84% and 69% of non-EP, respectively, compared to 65% overall, median PFS not reached). On the other hand, clusters A (n=20) and D (n=48) were associated with EP (55% and 46% of EP, respectively, compared to 35% overall, median PFS 2.87 and 5.2 months, compared to 9.5 months overall). Cluster A exhibited elevated cfDNA concentrations and

predominance of short fragments (cluster 1) over long ones (cluster 3), while cluster D—while having a similar PFS distribution (Figure 3B)—was characterized by intermediate concentrations and heterogeneity in fragment size, with patients exhibiting large concentrations of long fragments (cluster 3). Cluster B showed low cfDNA concentrations, and a predominance of very short (<111 bp, cluster 2) and long fragments (>240 bp, cluster 3). Cluster C showed high very-short-fragment abundance (cluster 2), correlating with left-shifted first peak on the distribution (weak values of P_1). Interestingly, tumor types dispersed across all clusters.

These results suggest that MV modelling of cfDNA variables could improve EP prediction.

Multivariable modeling improves prediction of EP

Subsequently, a machine learning pipeline was conducted to derive a multivariable predictive model based on a minimal set of variables. The variables P_2 and P_1 were excluded since the combined $P_2 - P_1$ had been constructed and found to have superior performances in the UV analysis. Given correlations among cfDNA variables (Figure 2C), we applied iterative VIF-based reduction (see Methods), removing five variables (HP, C_{TOT}, $R_{[111,240]}$, $R_{[160,220]}$, $R_{[370,580]}$), leading to a subset of 12 weakly correlated variables (Figure 3C) referred to hereafter as the full dataset: NLR, sex, ECOG, tumor type, $P_2 - P_1$, HW, $R_{[>75]}$, $R_{[75,111]}$, $R_{[280,320]}$, $R_{[240,370]}$, $R_{[580,1650]}$, and $R_{[>1650]}$.

Among 27 variable-selection methods, rigorously benchmarked for both performance and bootstrapped-stability, the optimal method identified retaining variables significant in both t-test and MV logistic regression (Benjamini-Hochberg-adjusted $p < 0.05$). The final signature dataset included NLR, sex, ECOG, tumor type, $P_2 - P_1$, $R_{[75,111]}$, $R_{[280,320]}$, $R_{[580,1650]}$, and $R_{[>1650]}$. Logistic regression achieved the best combination of

both optimism-corrected AUC (88.5%) and stability (0.63), with simultaneously lower bias than more complex models.

For benchmark comparisons, optimism-corrected metrics were computed on NLR and $R_{[>1650]}$ in UV settings, their combination, the signature dataset, and the full dataset (Figure 3D). All metrics were derived applying the same bootstrapped pipelines using the *ROOFS* package (24).

$R_{[>1650]}$ outperformed NLR across all metrics ($AUC_{R_{[>1650]}} = 73.6 \pm 3.70\%$ vs. $AUC_{NLR} = 68.9 \pm 5.02\%$, $PPV_{R_{[>1650]}} = 55.8 \pm 7.46\%$ vs. $PPV_{NLR} = 52.6 \pm 8.75\%$, Figure 3D). Moreover, multivariable combinations outperformed both NLR and $R_{[>1650]}$, with performance improving as more variables were included. Interestingly, the signature dataset achieved metrics equivalent to those of the full dataset ($AUC_{signature} = 88.5 \pm 3.3\%$, $AUC_{full} = 88.5 \pm 3.37\%$ $ACC_{signature} = 79.2 \pm 4.31\%$ VS $ACC_{full} = 79.9 \pm 4.39\%$, $NPV_{signature} = 87 \pm 5.35\%$ VS $NPV_{full} = 86.6 \pm 4.99\%$). The full dataset achieved a higher PPV than the signature one, although the latter exhibited a narrower CI ($PPV_{signature} = 69.4 \pm 7.49\%$ VS $PPV_{full} = 72.5 \pm 9.91\%$).

Overall, this establishes a compact nine-feature signature integrating clinical, biological, and cfDNA fragmentome variables, markedly improving prediction of EP under ICI therapy.

DISCUSSION

In this prospective proof-of-concept study, we show that cfDNA fragment-size profiles measured with BIABooster (15) provide innovative, non-invasive, pre-treatment biomarkers to predict EP and PFS under ICI in advanced carcinomas. Single size-

profile variables—especially the proportion of HMW fragments—outperformed conventional clinical and biological features (e.g., NLR(28,29), PD-L1 TPS/CPS, TMB(30)). Indeed, the only FDA-approved biomarkers (TMB and PD-L1) exhibit pooled AUCs of 69% and 65% and pooled PPVs of 42% and 34% (30), respectively, and we found $AUC_{NLR} = 68.9 \pm 5.02\%$, $PPV_{NLR} = 52.6 \pm 8.75\%$. In contrast, HMW fragments achieved $AUC_{R_{[>1650]}} = 73.6 \pm 3.70\%$ and $PPV_{R_{[>1650]}} = 55.8 \pm 7.46\%$. These associations remained significant after adjustment for established confounders (including NLR), for association with both EP and PFS. Interestingly, the model identified both EP and non-EP missed by NLR but captured by fragmentomic variables, demonstrating their biological complementarity. These findings suggest that cfDNA size profiles provide information orthogonal to systemic inflammation captured by NLR. To go beyond univariable associations, we developed a supervised MV model based on a robust variable-selection framework evaluating 27 feature-selection methods and 11 classifiers. Using optimism-corrected bootstrapping and explicit stability estimation (23), we derived a robust nine-variable signature combining clinical factors, NLR, and fragmentomic variables. This MV model markedly improved EP prediction compared with any single variable, consistently outperforming NLR and HMW fragments alone and their combination. The signature reached $AUC_{signature} = 88.5 \pm 3.3\%$ and $PPV_{signature} = 69.4 \pm 7.49\%$, clearly superior to previously published MV models (e.g., AUC~0.64-0.67 (31,32), although these studies had a 6-months horizon for progression). The signature also outperformed previous studies using pre- and on-treatment cfDNA variables (mainly ctDNA ones) that achieved AUCs ranging from 0.70 to 0.82 (33–35). These results underscore the robustness and reproducibility of a composite biomarker rather than dependence on numerous correlated variables.

Our findings suggest that cfDNA size profiles capture biologically relevant information not reflected in current clinical or pathological markers. Correlations between cfDNA features and LDH, PD-L1, NLR, or tumor burden (SLD) were minimal, supporting their role as independent indicators of immune response.

Previous research on cfDNA has mainly focused on total cfDNA concentration or short fragments abundance (< 200 bp) (36). Although high total cfDNA has been associated with poor PFS and OS(37,38), size-based variables have mainly been applied to optimize ctDNA sequencing sensitivity, exploiting the enrichment of short fragments (90–150 bp) (13) in cancer patients compared with the typical 166 bp peak in healthy individuals (39). They have also been explored for chemotherapy response prediction, where shorter fragments have shown predictive value(36). Fragment size has also shown value in chemotherapy response prediction. In contrast, the SChISM study highlighted a distinct pattern for immunotherapy: longer fragments—particularly HMW fragments ($R_{>1650}$)—best discriminated progressors from responders, and outperformed mononucleosomal fragments ($R_{[111,240]}$, $R_{[160,220]}$, HP, P_1, $P_2 - P_1$) or total cfDNA levels (C_{TOT}). Different ratios of long to short fragments were explored, in line with the concept of cfDNA integrity (40) (cfDI), which computes ratios on specific genetic locus, and is typically assessed for diagnostic or prognostic purposes. However, relying on the single cfDNA feature $R_{>1650}$ consistently yielded better or equivalent performance than either ratio.

The biological basis for the enrichment of HMW fragments in responders is not yet fully understood. Possible mechanisms include immune-cell activation (41) or tumor-cell death (42), processes known to release longer DNA fragments through necrosis, NETosis (42), phagocytosis, or chromosomal-instability–associated secretion (43). Although some of these processes are generally associated with poor prognosis (44),

NETosis, for example, can contribute to antitumor activity (44). Progressors in our study exhibited shorter dinucleosome-derived fragments ($P_2 < 302$ bp) and reduced spacing between nucleosomal peaks ($P_2 - P_1 < 154$ bp), spacing tighter nucleosome spacing, altered nuclease activity (45), or hypomethylation (3). Overall, increased fragmentation at baseline appears to predict rapid progression and shorter PFS.

Beyond correlation, cfDNA may play an active immunomodulatory role. Long fragments could activate the cGAS–STING pathway and enhance interferon-mediated immune activity (46,47). This hypothesis remains to be confirmed in complementary biological studies. Therefore, lower cfDNA fragmentation at baseline induced by systemic DNase treatment (48,49), in combination with ICI therapy, may facilitate more effective tumor eradication. Thus, our findings not only propose a novel predictive biomarker but also provide mechanistic hypotheses regarding ICI resistance.

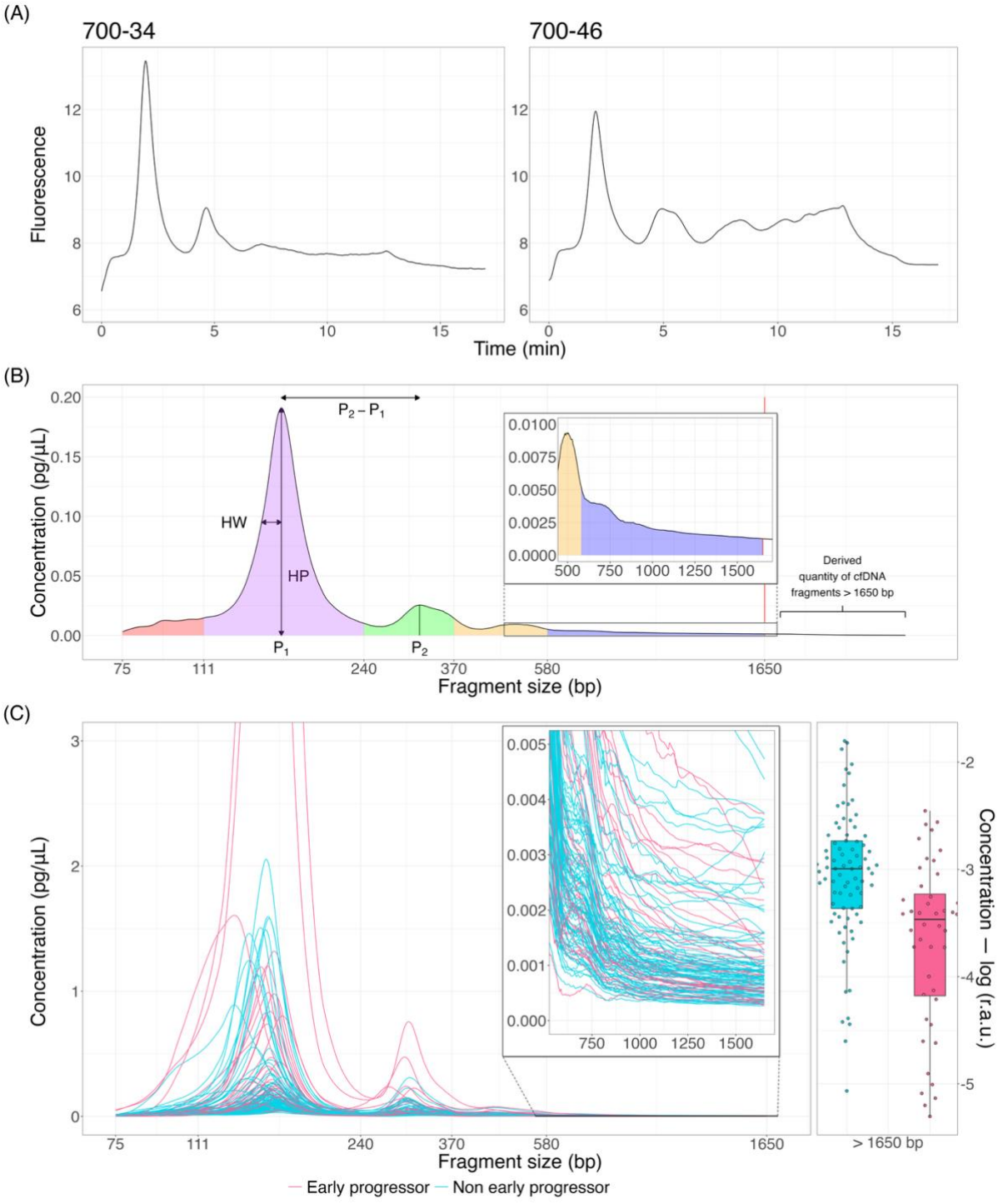
CfDNA size profile offers practical advantages for clinical implementation. Because it contains fragments from both tumor and wild-type cells, global cfDNA is independent of tumor genotype, making it applicable across genetically heterogeneous tumors. It also avoids reliance on tissue biopsies, enabling non-invasive serial monitoring. Using the BIABooster device, cfDNA size and concentration can be directly measured from minimal plasma volumes without DNA extraction, allowing rapid and standardized implementation in clinical laboratories. Combined with their strong prognostic value, these features support the development of tools to optimize early therapeutic decisions—such as avoiding unnecessary continuation of ICI in patients likely to progress, thereby reducing toxicity and cost. If validated in homogeneous cohorts with higher PPV, these markers could also help avoid initiating ineffective treatment in poor responders.

However, while our findings are statistically robust, they require external validation in independent and disease- or treatment-specific cohorts. Moreover, size ranges were arbitrary set according to technology limits and as nucleosomal-length bounds. More optimized intervals could lead to better performance. Furthermore, the analysis is limited to a single pre-treatment timepoint. To address this, we plan to integrate both baseline and on-treatment cfDNA size profiles, collecting plasma samples before each immunotherapy cycle, into a mechanistic learning framework to provide dynamic biomarkers of response (50).

In conclusion, our approach, based on a standardized, low-cost technology (BIABooster), showed that robust signals can be derived from global size profiling without requiring sequencing-based methods. The integration of either the single HMW variable or the predictive multivariable model into longitudinal frameworks may pave the way for precision immuno-oncology strategies and real-time adaptive treatment management. External validation in independent and multi-centric cohorts is warranted to move towards application at bedside.

FIGURES

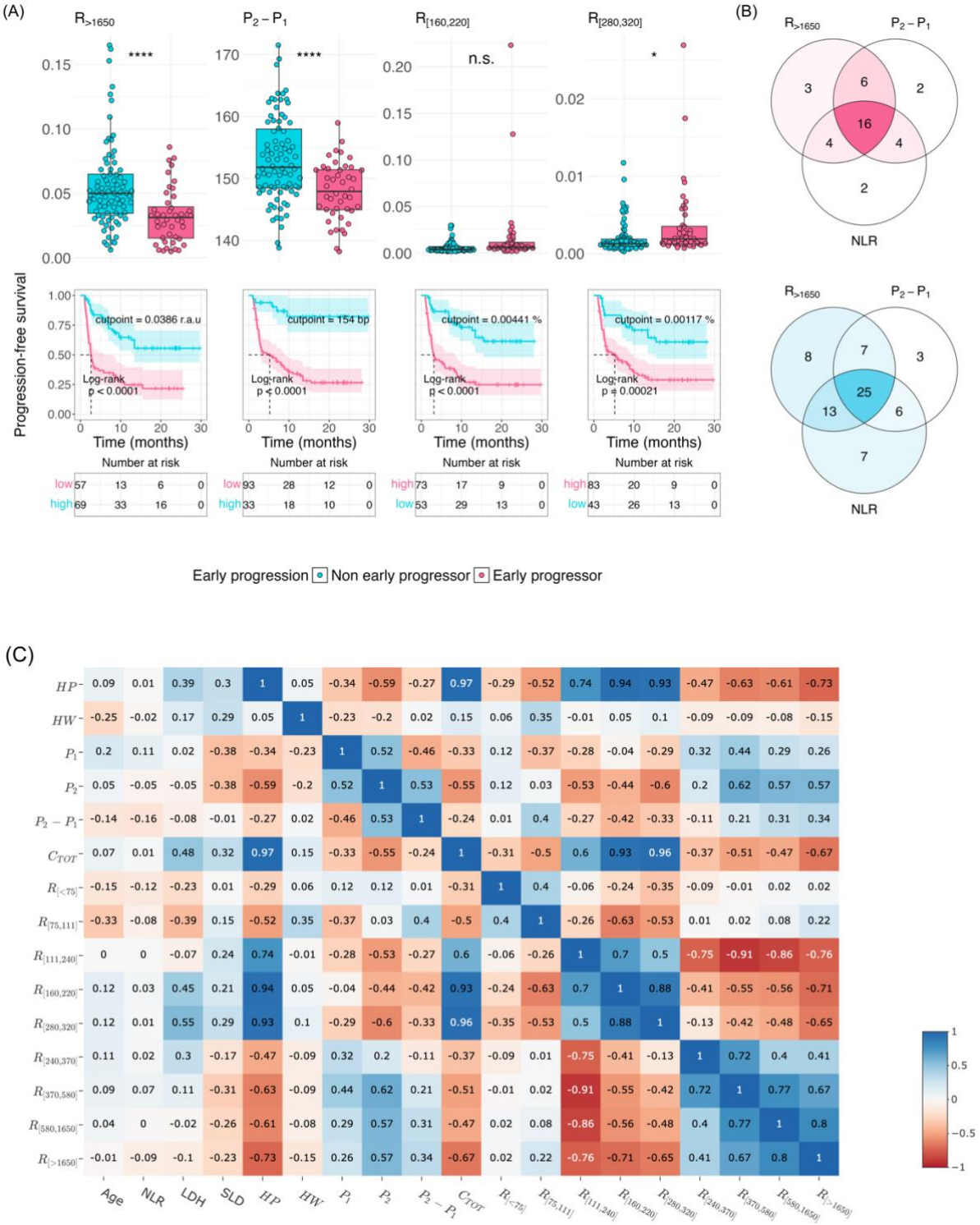
Figure 1: Cancer patients exhibit different cfDNA size profiles under ICI treatment



(A) Fluorescence vs. time curve, distinguishing patients with low vs. high levels of high-molecular-weight DNA (700-34 vs. 700-46). (B) Pre-treatment cfDNA size profile of the patient 700-46, derived from the fluorescence/time plot on the right panel of (A). Eleven of the extracted cfDNA variables are displayed: - the first, second peak' locations (P_1 and P_2) in base pairs (bp) and their difference ($P_2 - P_1$); - the height of the first peak

494 (HP); - the half-width of the left side of the first peak HW; - the concentration (area
495 under the curve) of cfDNA from different size ranges: $R_{[75,111]}$: red area; $R_{[111,240]}$:
496 purple area; $R_{[240,370]}$: green area; $R_{[370,580]}$: yellow area; $R_{[580,1650]}$: purple area.; the
497 quantity of cfDNA greater than 1650 bp (C) CfDNA data of the 126 patients, colored
498 by outcome: early-progressor (pink), non early-progressor (blue). Left: pre-treatment
499 cfDNA size profiles. Zoom: focus on size distributions between 580 and 1650 bp Right:
500 pre-treatment distribution of $R_{>1650}$.

501 ***Figure 2: Higher ratio of long cfDNA fragments is associated with extended***
502 ***progression-free survival and no early progression under treatment***



503

504 (A) Distribution of the best cfDNA variables ($R_{>1650}$, $P_2 - P_1$, $R_{[580,1650]}$, and C_{TOT}) by

505 early progression status and progression-free survival. Boxplots show cfDNA variable

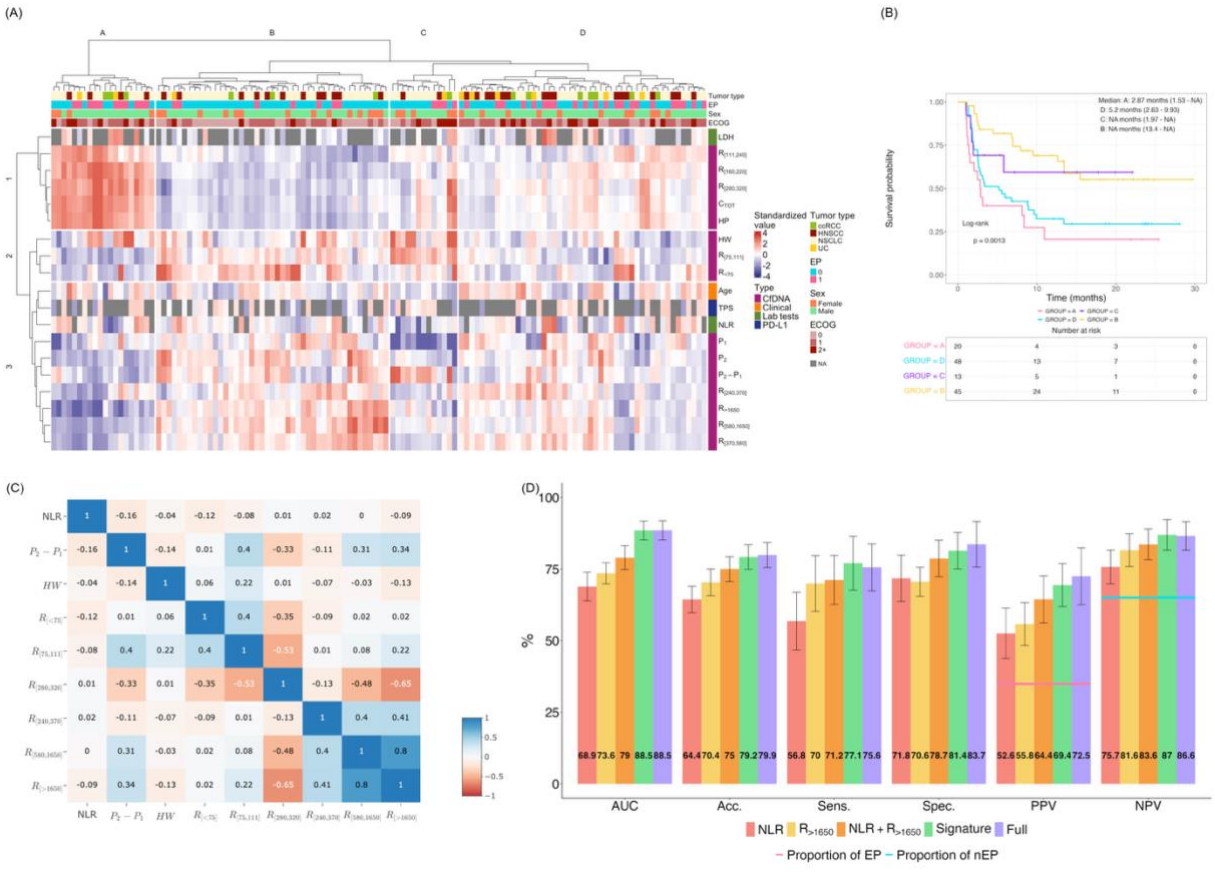
506 distributions stratified by early progression status and display the Welch two sample t-

507 test significance. Kaplan-Meier curves display progression-free survival stratified by

cfDNA levels, display the log-rank test p-value and the optimal threshold computed to stratify patients between long and short progression-free survival (see Methods). (B) Pearson correlation was computed between the eleven cfDNA variables and five conventional features: age, neutrophil-to-lymphocyte ratio (NLR), lactate dehydrogenase (LDH), and sum of the largest target diameters (SLD) ($n_{SLD} = 116$, $n_{NLR} = 110$, $n_{LDH} = 71$). (C) Venn diagrams were generated based on the counts of true early progressors (pink) and true responders (blue) identified as the two best cfDNA variables ($R_{>1650}$, $\mathrm{P}_2 - P_1$) in the statistical analyses and NLR.

(Significance: ****: p-value < 0.0001; ***: p-value < 0.001; **: p-value < 0.01; *: p-value < 0.05; n.s.: non-significant, > 0.05)

Figure 3: Patients are clustered according to their fragment size distribution and combining cfDNA variables increases predictive performance



(A) Two-way hierarchical clustering was computed to cluster patients (columns) and

features (rows) simultaneously. The resulting dendrogram and clustered data are presented as a heatmap, where color intensity represents standardized feature values and the hierarchical relationships are shown by the accompanying dendrograms.

The three main feature groups were labeled from 1 to 3, while the three main patient groups were labeled from A to D. (B) Kaplan-Meier curves display progression-free survival stratified by patient clusters computed through unsupervised hierarchical clustering. The plot displays the log-rank test p-value. (C) Pearson correlation plot of the nine continuous variables of the full dataset. (D) Area under the receiver operating curve (AUC), accuracy (Acc.), sensitivity (Sens.), specificity (Spec.), positive predive value (PPV) and negative predictive value (NPV) of the different combinations of variables: NLR, $R_{[>1650]}$, NLR + $R_{[>1650]}$, signature, and full dataset. Each performance metric was computed 100 times on out-of-bag datasets. The reported values represent the mean of the 100 iterations, while error bars indicate the mean ± 1 standard deviation. The true rate of EP is indicated in pink within the PPV bars, and the true rate of non-EP is indicated in blue within the NPV bars.

(ccRCC: clear cell renal cell carcinoma; ECOG: Eastern Cooperative Oncology Group status; EP: early progression; HNSCC: head and neck squamous cell carcinoma; LDH: lactate dehydrogenase; NLR: neutrophil to lymphocyte ratio; NSCLC: non-small cell lung cancer; PD-L1: programmed cell death protein 1; TPS: tumor proportion score; UC: urothelial carcinoma)

TABLES

Table 1: Clinical characteristics of the SChISM patients

Variable	Overall N = 126[1]	Pathology				p-value[2]
		UC N = 10[1]	HNSCC N = 34[1]	ccRCC N = 14[1]	NSCLC N = 68[1]	
Sex						0,6
Male	92 (73%)	9 (90%)	26 (76%)	10 (71%)	47 (69%)	
Female	34 (27%)	1 (10%)	8 (24%)	4 (29%)	21 (31%)	
Age (years)	64 (11)	69 (8)	64 (10)	59 (14)	63 (10)	0,2
Treatment[3]						
Atezolizumab	1 (0,8%)	0 (0%)	0 (0%)	0 (0%)	1 (1,5%)	
Nivolumab	32 (25%)	0 (0%)	25 (74%)	1 (7,1%)	6 (8,8%)	
Nivolumab + Ipilimumab	2 (1,6%)	0 (0%)	0 (0%)	2 (14%)	0 (0%)	
Nivolumab + targeted therapy	4 (3,2%)	0 (0%)	0 (0%)	4 (29%)	0 (0%)	
Pembrolizumab	32 (25%)	9 (90%)	3 (8,8%)	2 (14%)	18 (26%)	
Pembrolizumab + chemotherapy	48 (38%)	0 (0%)	5 (15%)	0 (0%)	43 (63%)	
Pembrolizumab + targeted therapy	7 (5,6%)	1 (10%)	1 (2,9%)	5 (36%)	0 (0%)	
Treatment line						<0,001
1	100 (79%)	3 (30%)	23 (68%)	14 (100%)	60 (88%)	
2	25 (20%)	7 (70%)	11 (32%)	0 (0%)	7 (10%)	
3	1 (0,8%)	0 (0%)	0 (0%)	0 (0%)	1 (1,5%)	
CPS (%)						0,6
[1, 20[	8 (40%)	0 (0%)	8 (42%)	0 (-%)	0 (-%)	
< 1	5 (25%)	0 (0%)	5 (26%)	0 (-%)	0 (-%)	
>= 20	7 (35%)	1 (100%)	6 (32%)	0 (-%)	0 (-%)	
Missing	106	9	15	14	68	
TPS (%)						>0,9
[1, 50[	25 (40%)	0 (-%)	0 (-%)	0 (-%)	25 (40%)	
< 1	14 (23%)	0 (-%)	0 (-%)	0 (-%)	14 (23%)	
>= 50	23 (37%)	0 (-%)	0 (-%)	0 (-%)	23 (37%)	
Missing	64	10	34	14	6	
NLR	4,5 (3,8)	3,6 (2,3)	7,3 (5,4)	2,5 (1,3)	3,6 (2,3)	<0,001
Missing	12	0	2	1	9	
LDH (UI/L)	238 (147)	221 (91)	236 (195)	252 (92)	243 (88)	0,4
Missing	56	0	1	3	52	
SLD (mm)	63 (50)	50 (46)	40 (22)	92 (62)	70 (53)	0,016
Missing	7	0	5	0	2	
Mutations[4]						
KRAS	26 (62%)	0 (0%)	0 (0%)	0 (-%)	26 (65%)	
TP53	25 (66%)	1 (100%)	1 (100%)	0 (-%)	23 (64%)	
Early progression[5]						<0,001
0	82 (65%)	3 (30%)	14 (41%)	11 (79%)	54 (79%)	
1	44 (35%)	7 (70%)	20 (59%)	3 (21%)	14 (21%)	
PFS (median)	9,5 (6,8, —)	2,8 (2,4, 5,6)	2,6 (1,8, —)	9,6 (6,0, —)	— (13, —)	<0,001

[1] n (%); Mean (SD)
[2] Fisher's exact test; Kruskal-Wallis rank sum test; NA
[3] Immunotherapy combination and potential associated therapy
[4] Other hidden mutations are: EGFR (3), BRAF (5), CTNNB1 (1), ERBB4 (1), FGFR 3 (1), P13 KINASE (1), PDGFRA (1), PIK3CA (2), PTEN (4), RET (1), STK11 (1), TTF1 (1)
[5] 1, progression before 3 months of immunotherapy I 0, otherwise

(ccRCC: clear cell renal cell carcinoma; CPS: combined positive score; HNSCC: head and neck squamous cell carcinoma; KRAS: Kirsten rat sarcoma viral oncogene; LDH: lactate dehydrogenase; NLR: neutrophil to lymphocyte ratio; NSCLC: non-small cell lung cancer; SLD: sum of the largest diameters of target lesions; TPS: tumor proportion score; TP53: transformation-related protein 53; UC: urothelial carcinoma)

Table 2: Logistic and Cox regression results

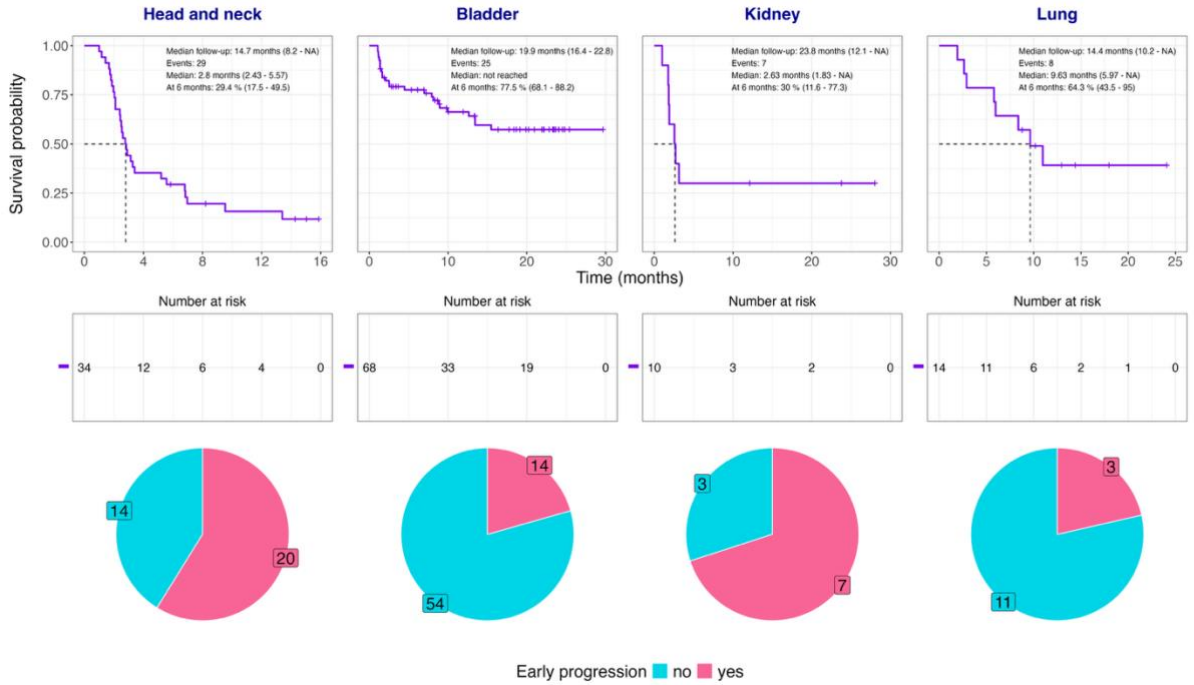
ALL PATIENTS: CLINICAL AND BIOLOGICAL VARIABLES												
			EARLY PROGRESSION					PROGRESSION-FREE SURVIVAL				
VARIABLE	LEVEL	N	AUC	OR UV	SIGNIF UV	OR MV	SIGNIF MV	C INDEX	HR UV	SIGNIF UV	HR MV	SIGNIF MV
CFDNA METRICS												
$R_{[>1650]}$		126	0.73	0.39 (0.25 — 0.62)	****	0.28 (0.13 — 0.6)	***	0.69	0.54 (0.42 — 0.68)	****	0.44 (0.33 — 0.58)	****
$P_2 - P_1$		126	0.72	0.38 (0.23 — 0.62)	***	0.36 (0.18 — 0.73)	**	0.67	0.55 (0.42 — 0.72)	****	0.6 (0.44 — 0.83)	**
$R_{[160,220]}$		126	0.70	2.1 (1.3 — 3.2)	**	2.1 (1.2 — 3.7)	**	0.66	1.7 (1.4 — 2.1)	****	1.8 (1.4 — 2.4)	****
$R_{[280,320]}$		126	0.67	1.8 (1.2 — 2.7)	**	2.1 (1.2 — 3.8)	*	0.64	1.6 (1.3 — 2)	****	1.7 (1.3 — 2.3)	***
$R_{[580,1650]}$		126	0.65	0.51 (0.32 — 0.82)	**	0.48 (0.23 — 0.99)	*	0.64	0.55 (0.4 — 0.76)	***	0.52 (0.36 — 0.76)	***
P_2		126	0.64	0.56 (0.36 — 0.86)	**	0.46 (0.25 — 0.85)	*	0.63	0.6 (0.46 — 0.79)	***	0.59 (0.44 — 0.8)	***
HP		126	0.66	1.7 (1.1 — 2.5)	**	1.9 (1.1 — 3.3)	*	0.63	1.6 (1.2 — 2)	***	1.8 (1.4 — 2.3)	****
C_{TOT}		126	0.65	1.7 (1.1 — 2.5)	**	2 (1.1 — 3.4)	*	0.62	1.5 (1.2 — 1.9)	***	1.8 (1.3 — 2.3)	****
$R_{[111,240]}$		126	0.60	1.5 (1 — 2.2)	*	1.7 (0.94 — 3)	n.s.	0.59	1.5 (1.2 — 2)	**	1.7 (1.3 — 2.4)	***
$R_{[75,111]}$		126	0.61	0.68 (0.46 — 0.99)	*	0.76 (0.44 — 1.3)	n.s.	0.59	0.75 (0.58 — 0.97)	*	0.78 (0.59 — 1)	n.s.
$R_{[370,580]}$		126	0.57	0.74 (0.51 — 1.1)	n.s.	0.5 (0.27 — 0.94)	*	0.57	0.73 (0.56 — 0.94)	*	0.6 (0.44 — 0.82)	**
$R_{[240,370]}$		126	0.51	1 (0.69 — 1.4)	n.s.	0.89 (0.52 — 1.5)	n.s.	0.51	0.92 (0.71 — 1.2)	n.s.	0.73 (0.53 — 0.99)	*
CLINICAL AND BIOLOGICAL VARIABLES												
Neutrophil-to-lymphocyte ratio		114	0.69	2.4 (1.5 — 4)	***			0.63	1.6 (1.2 — 2.2)	**		
ECOG	0 (reference)	126										
	2_sup			6 (2.1 — 17)	***			0.62	3.4 (1.8 — 6.3)	****		
Pathology	NSCLC (reference)	126										
	HNSCC			5.5 (2.2 — 14)	***			0.60	3.8 (2.2 — 6.6)	****		
	UC			9 (2.1 — 39)	**			0.60	2.9 (1.2 — 6.7)	*		
Sex	male (reference)	126										
	female			0.12 (0.034 — 0.42)	***			0.59	0.28 (0.14 — 0.59)	***		

(AUC: area under the receiving operator curve; ECOG: Eastern Cooperative Oncology Group status; HNSCC: head and neck squamous cell carcinoma; HR: hazard ratio; MV: multivariable; NSCLC: non-small cell lung cancer; OR: odds ratio; UV: univariable) (Significance: ****: p-value < 0.0001; ***: p-value < 0.001; **: p-value < 0.01; *: p-value < 0.05; n.s.: non-significant, > 0.05)

The categorical features present results for each of their levels comparing to the reference level.

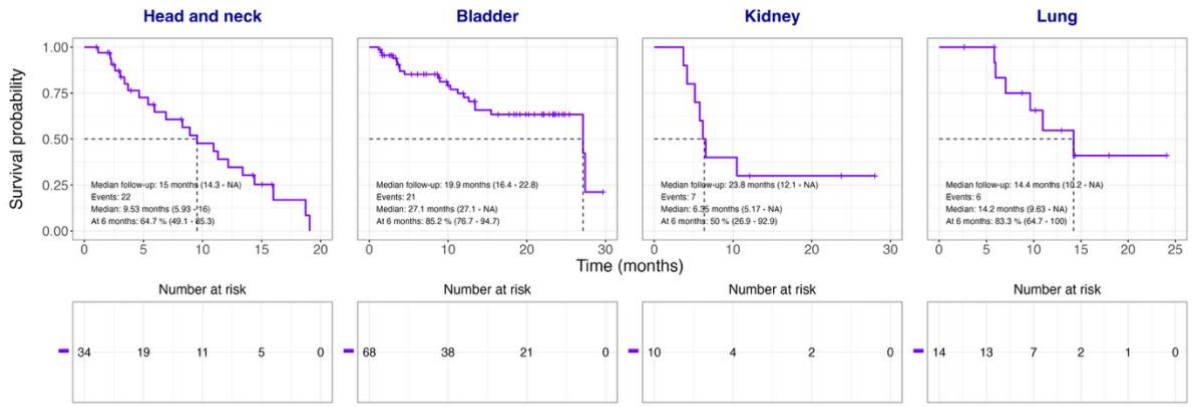
SUPPLEMENTARY

Figure S1: Time-to-progression distributions vary across pathological subpopulations



Survival curves (purple line) of the progression-free survival, estimated by Kaplan-Meier method, across all pathological subpopulations. Dotted lines indicate the median time of progression. Pie charts display the early-progression distribution across the same subpopulations.

Figure S2: Time-to-death distributions across pathological subpopulations



570

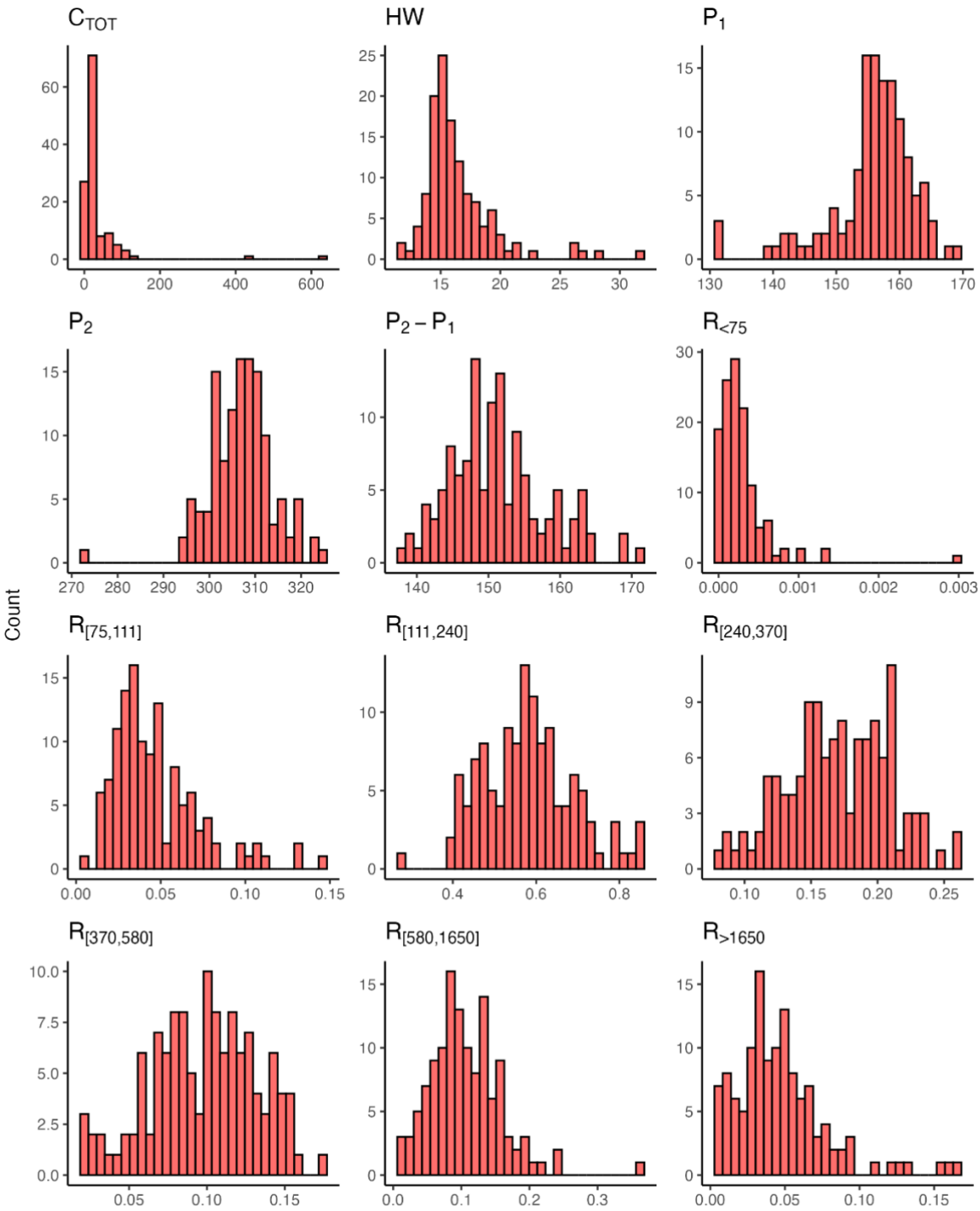
571 Survival curves (purple line) of the overall survival, estimated by Kaplan-Meier method,

572 across different pathological subpopulations. Dotted lines indicate the median time of

573 progression

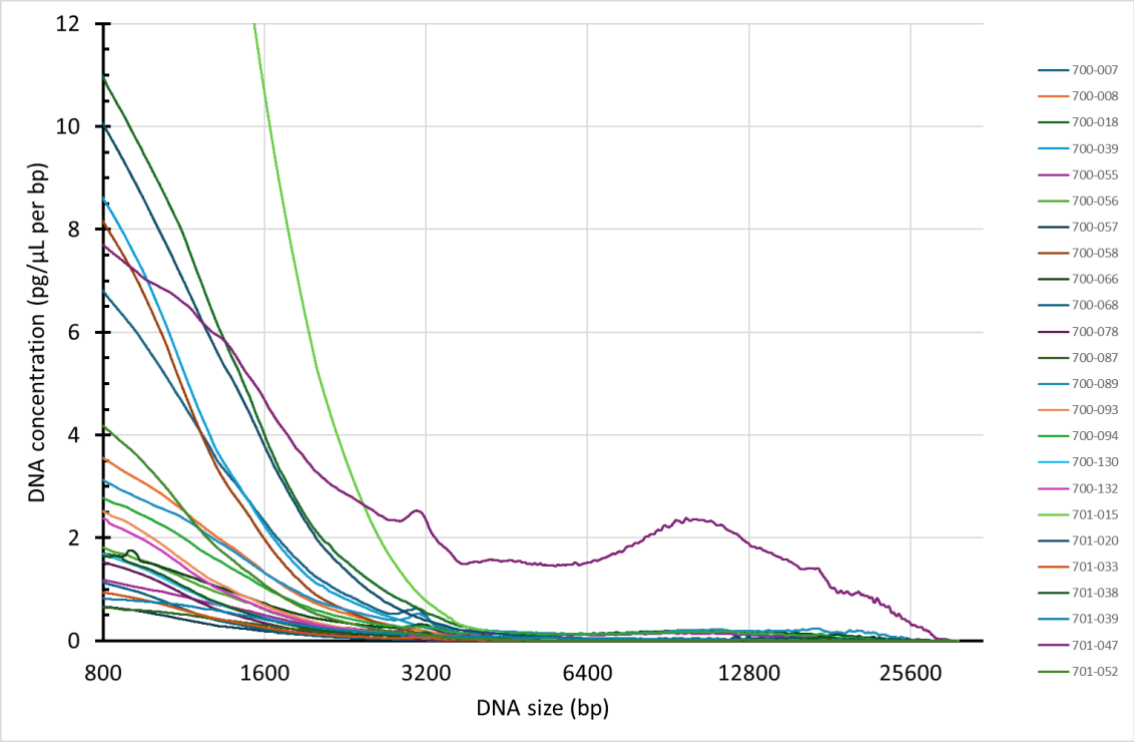
574

575 ***Figure S3: Distribution of cfDNA size profile variables***



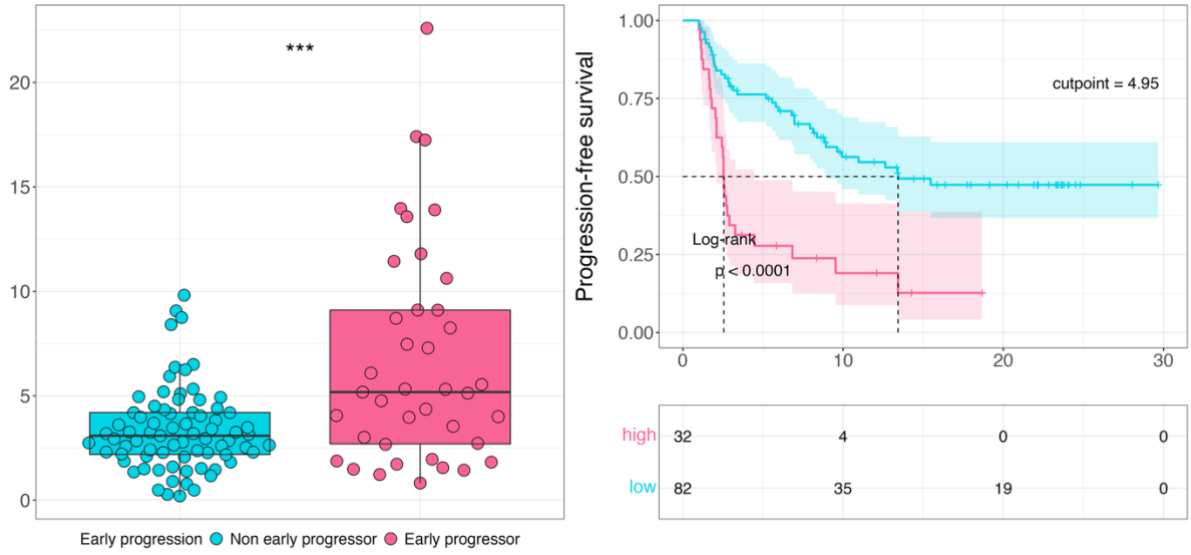
576

577 ***Figure S4: Analysis of fragments greater than 1 kilobase of 24 samples***



Twenty-four samples were selected to analyze further the large DNA fragments; 20 of them were selected to contain high-molecular-weight based on the standard cfDNA analysis. For these samples, DNA was extracted from 4 mL of plasma using IDXtract-MAG kits (ID-Solutions), and DNA was eluted in 100 µL of TE buffer. Five µL of samples were diluted twice in TE, and then analyzed with a BIABooster device using the DNA10K method, which analyzes purified DNA samples up to 10 kb (ADELIS; https://www.adelis-tech.com/wp-content/uploads/2025/09/Product-sheet-DNA-10K-KIT-BIABOOSTER-CE.pdf).

Figure S5: Higher neutrophil-to-lymphocyte ratio is associated with shorter progression-free survival and early progression under treatment



589

590 Neutrophil-to-lymphocyte ratio distribution is displayed using boxplots stratified by
591 early progression status and Kaplan-Meier progression-free survival curves.
592 Boxplots display the Welch two sample t-test significance. Kaplan-Meier plots display
593 the log-rank test p-value and the optimal threshold computed to stratify patients
594 between long and short progression-free survival (see Methods).
595 (Significance: ****: p-value < 0.0001; ***: p-value < 0.001; **: p-value < 0.01; *: p-value
596 < 0.05; n.s.: non-significant, > 0.05)

597 ***Table s1: Empirical statistics of cfDNA variables***

Variable	Overall, N = 126[1]	Early-progression		p-value[2]
		Early-progressors N = 44[1]	Non early-progressors N = 82[1]	
C_{TOT}	16 [6-637]	18 [7-637]	14 [6-109]	0,006
HW	15.61 [11.84-31.82]	15.98 [12.82-28.12]	15.50 [11.84-31.82]	0,4
P_1	157 [132-169]	157 [132-169]	156 [132-168]	0,3
P_2	307 [274-326]	306 [274-319]	308 [295-326]	0,010
$P_2 - P_1$	151 [138-172]	148 [138-159]	152 [139-172]	<0,001
$R_{<75}$	0.0002 [0.0000-0.0030]	0.0002 [0.0000-0.0014]	0.0002 [0.0000-0.0030]	0,8
$R_{[75.111]}$	0.040 [0.006-0.147]	0.034 [0.006-0.129]	0.044 [0.013-0.147]	0,041
$R_{[111.240]}$	0.58 [0.28-0.85]	0.60 [0.41-0.85]	0.57 [0.28-0.82]	0,055
$R_{[240.370]}$	0.17 [0.08-0.26]	0.17 [0.08-0.26]	0.17 [0.09-0.26]	0,8
$R_{[370.580]}$	0.10 [0.02-0.17]	0.09 [0.02-0.15]	0.10 [0.03-0.17]	0,2
$R_{[580.1650]}$	0.10 [0.01-0.36]	0.08 [0.01-0.17]	0.11 [0.02-0.36]	0,005
$R_{>1650}$	0.043 [0.005-0.165]	0.031 [0.005-0.086]	0.050 [0.006-0.165]	<0,001

[1] Median [Min-Max]
[2] Wilcoxon rank sum test

DECLARATIONS

Ethics approval and consent to participate

The study complied with the Declaration of Helsinki, Good Clinical Practice, and French regulations. It was classified as Category 2 research under the Jardé Law and approved by the national ethics committee, with registration through the French Ministry of Health. AP-HM, as sponsor, ensured data protection regulations (GDPR) compliance through secure anonymized electronic Case Report Forms. All participants provided written informed consent. The study is registered as NCT05083494.

Consent for publication

Not applicable.

Availability of data and material

The data used in this study is publicly available at https://zenodo.org/records/17854973. The repository contains baseline clinical data, and longitudinal tumor size and size-based cfDNA concentration measurements from advanced and/or metastatic carcinoma patients under immune checkpoint inhibitors from the SChISM clinical study.

All software packages used in this study, along with their version information, are reported in the Methods section.

Competing interests

All authors have contributed substantially to this work and have approved the final manuscript. A. Boutonnet, F. Ginot, J.-G. Garcia are respectively employee, CTO and CEO of the Adelis Technologies company that develops the BIABooster technology employed to quantify the cfDNA fragment size concentrations. F. Fina holds shares in the Adelis Technologies company.

Funding

This work received support from the French government under the France 2030 investment plan, as part of the Initiative d'Excellence d'Aix-Marseille Université - A*MIDEX (AMX-19-IET-001 & AMX-21-IET-017).

Authors' contributions

Conceptualization: SS, FF, SB

Methodology: LNP, SS, SB

Data curation: LNP

Formal analysis: LNP, RZ, LDN, SB

Investigation: FG, AB

Resources (patient inclusion): SS, LG, PT, JLD

Supervision: SS, SB, FF, JCG

Project administration: SS

Writing – original draft: LNP, SB, SS

Writing – review & editing: all authors

REFERENCES

1. Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun. 2018 Nov 29;9(1):5068.

2. Mattox AK, Douville C, Wang Y, Popoli M, Ptak J, Silliman N, et al. The Origin of Highly Elevated Cell-Free DNA in Healthy Individuals and Patients with Pancreatic, Colorectal, Lung, or Ovarian Cancer. Cancer Discov. 2023 Oct 5;13(10):2166–79.

3. Heitzer E, Auinger L, Speicher MR. Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living. Trends in Molecular Medicine. 2020 May;26(5):519–28.

4. Hu Z, Chen H, Long Y, Li P, Gu Y. The main sources of circulating cell-free DNA: Apoptosis, necrosis and active secretion. Crit Rev Oncol Hematol. 2021 Jan 1;157:103166.

5. Khier S, Lohan L. Kinetics of circulating cell-free DNA for biomedical applications: critical appraisal of the literature. Future Science OA. 2018 Apr;4(4):FSO295.

6. Esposito A, Bardelli A, Criscitiello C, Colombo N, Gelao L, Fumagalli L, et al. Monitoring tumor-derived cell-free DNA in patients with solid tumors: clinical perspectives and research opportunities. Cancer Treat Rev. 2014 June;40(5):648–55.

7. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev. 2016;35(3):347–76.

8. Gandara DR, Agarwal N, Gupta S, Klempner SJ, Andrews MC, Mahipal A, et al. Tumor mutational burden and survival on immune checkpoint inhibition in >8000 patients across 24 cancer types. J Immunother Cancer. 2025 Feb 6;13(2):e010311.

9. Ngo P, Cooper WA, Wade S, Fong KM, Canfell K, Karikios D, et al. Why PD-L1 expression varies between studies of lung cancer: results from a Bayesian meta-analysis. Sci Rep. 2025 Feb 4;15(1):4166.

10. Ding SC, Lo YMD. Cell-Free DNA Fragmentomics in Liquid Biopsy. Diagnostics. 2022 Apr;12(4):978.

11. Stadler JC, Belloum Y, Deitert B, Sementsov M, Heidrich I, Gebhardt C, et al. Current and Future Clinical Applications of ctDNA in Immuno-Oncology. Cancer Research. 2022 Feb 1;82(3):349–58.

12. Ivanov M, Baranova A, Butler T, Spellman P, Mileyko V. Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. BMC Genom. 2015 Dec 16;16(Suppl 13):S1.

13. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med. 2018 Nov 7;10(466):eaat4921.

14. Qi T, Pan M, Shi H, Wang L, Bai Y, Ge Q. Cell-Free DNA Fragmentomics: The Novel Promising Biomarker. Int J Mol Sci. 2023 Jan 12;24(2):1503.

15. Andriamanampisoa CL, Bancaud A, Boutonnet-Rodat A, Didelot A, Fabre J, Fina F, et al. BIABooster: Online DNA Concentration and Size Profiling with a Limit of Detection of 10 fg/μL and Application to High-Sensitivity Characterization of Circulating Cell-Free DNA. Anal Chem. 2018 Mar 20;90(6):3766–74.

16. Boutonnet A, Pradines A, Mano M, Kreczman-Brun M, Mazières J, Favre G, et al. Size and Concentration of Cell-Free DNA Measured Directly from Blood Plasma, without Prior DNA Extraction. Anal Chem. 2023 June 20;95(24):9263–70.

17. Modesto A, Gibert S, Delmas C, Tougeron D, Paumier A, Taillez A, et al. Sizing and concentration analysis of cfDNA using Biabooster technology: Results from a prospective plasma-based collection of 77 patients with locally advanced unresectable esophageal cancer. International Journal of Cancer [Internet]. [cited 2025 Dec 3];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.70011

18. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol. 2017 Mar;18(3):e143–52.

19. Bagley SJ, Kothari S, Aggarwal C, Bauml JM, Alley EW, Evans TL, et al. Pretreatment neutrophil-to-lymphocyte ratio as a marker of outcomes in nivolumab-treated patients with advanced non-small-cell lung cancer. Lung Cancer. 2017 Apr;106:1–7.

20. Mezquita L, Auclin E, Ferrara R, Charrier M, Remon J, Planchard D, et al. Association of the Lung Immune Prognostic Index With Immune Checkpoint Inhibitor Outcomes in Patients With Advanced Non–Small Cell Lung Cancer. JAMA Oncology. 2018 Mar 1;4(3):351–7.

21. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An introduction to statistical learning [Internet]. Springer; 2013 [cited 2024 Aug 21]. Available from: https://biblio.cerist.dz/index.php/hrbdonf5214/ouvrages/00000000062104800001_2.pdf

22. Efron B. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. Journal of the American Statistical Association. 1983 June 1;78(382):316–31.

23. Nogueira S, Sechidis K, Brown G. On the Stability of Feature Selection Algorithms. Journal of Machine Learning Research. 2018;18(174):1–54.

24. Bakhmach A, Dufossé P, Vaglio A, Monville F, Greillier L, Barlési F, et al. ROOFS: RObust biOmarker Feature Selection [Internet]. 2026 [cited 2026 Feb 2]. Available from: https://inria.hal.science/hal-05241230

25. Efron B, Tibshirani R. Improvements on Cross-Validation: The .632+ Bootstrap Method. Journal of the American Statistical Association. 1997;92(438):548–60.

26. Burtness B, Harrington KJ, Greil R, Soulières D, Tahara M, de Castro G, et al. Pembrolizumab alone or with chemotherapy versus cetuximab with chemotherapy for recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-048): a randomised, open-label, phase 3 study. Lancet. 2019 Nov 23;394(10212):1915–28.

27. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. New England Journal of Medicine. 2016 Nov 10;375(19):1823–33.

28. Diem S, Schmid S, Krapf M, Flatz L, Born D, Jochum W, et al. Neutrophil-to-Lymphocyte ratio (NLR) and Platelet-to-Lymphocyte ratio (PLR) as prognostic markers in patients with non-small cell lung cancer (NSCLC) treated with nivolumab. Lung Cancer. 2017 Sept;111:176–81.

29. Mosca M, Nigro MC, Pagani R, De Giglio A, Di Federico A. Neutrophil-to-Lymphocyte Ratio (NLR) in NSCLC, Gastrointestinal, and Other Solid Tumors: Immunotherapy and Beyond. Biomolecules. 2023 Dec 18;13(12):1803.

30. Lu S, Stein JE, Rimm DL, Wang DW, Bell JM, Johnson DB, et al. Comparison of Biomarker Modalities for Predicting Response to PD-1/PD-L1 Checkpoint Blockade: A Systematic Review and Meta-analysis. JAMA Oncology. 2019 Aug 1;5(8):1195–204.

31. Yoo SK, Fitzgerald CW, Cho BA, Fitzgerald BG, Han C, Koh ES, et al. Prediction of checkpoint inhibitor immunotherapy efficacy for cancer using routine blood tests and clinical data. Nat Med. 2025 Mar;31(3):869–80.

32. Captier N, Lerousseau M, Orlhac F, Hovhannisyan-Baghdasarian N, Luporsi M, Woff E, et al. Integration of clinical, pathological, radiological, and transcriptomic data improves prediction for first-line immunotherapy outcome in metastatic non-small cell lung cancer. Nat Commun. 2025 Jan 12;16(1):614.

33. Nabet BY, Esfahani MS, Moding EJ, Hamilton EG, Chabon JJ, Rizvi H, et al. Noninvasive Early Identification of Therapeutic Benefit from Immune Checkpoint Inhibition. Cell. 2020 Oct 15;183(2):363-376.e13.

34. Anagnostou V, Ho C, Nicholas G, Juergens RA, Sacher A, Fung AS, et al. ctDNA response after pembrolizumab in non-small cell lung cancer: phase 2 adaptive trial results. Nat Med. 2023 Oct;29(10):2559–69.

35. Zhang Q, Luo J, Wu S, Si H, Gao C, Xu W, et al. Prognostic and Predictive Impact of Circulating Tumor DNA in Patients with Advanced Cancers Treated with Immune Checkpoint Blockade. Cancer Discovery. 2020 Dec 1;10(12):1842–53.

36. Lapin M, Oltedal S, Tjensvoll K, Buhl T, Smaaland R, Garresori H, et al. Fragment size and level of cell-free DNA provide prognostic information in patients with advanced pancreatic cancer. J Transl Med. 2018 Nov 6;16:300.

37. Assaf ZJF, Zou W, Fine AD, Socinski MA, Young A, Lipson D, et al. A longitudinal circulating tumor DNA-based model associated with survival in metastatic non-small-cell lung cancer. Nat Med. 2023 Mar 16;1–10.

38. Del Re M, Crucitta S, Paolieri F, Cucchiara F, Verzoni E, Bloise F, et al. The amount of DNA combined with TP53 mutations in liquid biopsy is associated with clinical outcome of renal cancer patients treated with immunotherapy and VEGFR-TKIs. J Transl Med. 2022 Aug 16;20:371.

39. Jiang P, Chan CWM, Chan KCA, Cheng SH, Wong J, Wong VWS, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci U S A. 2015 Mar 17;112(11):E1317–25.

40. Wang BG, Huang HY, Chen YC, Bristow RE, Kassauei K, Cheng CC, et al. Increased plasma DNA integrity in cancer patients. Cancer Res. 2003 July 15;63(14):3966–8.

41. Thierry AR, Pisareva E. A New Paradigm of the Origins of Circulating DNA in Patients with Cancer. Cancer Discovery. 2023 Oct 5;13(10):2122–4.

42. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA Fragments in the Blood Plasma of Cancer Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells1. Cancer Res. 2001 Feb 2;61(4):1659–65.

43. Grabuschnig S, Bronkhorst AJ, Holdenrieder S, Rosales Rodriguez I, Schliep KP, Schwendenwein D, et al. Putative Origins of Cell-Free DNA in Humans: A Review of Active and Passive Nucleic Acid Release Mechanisms. Int J Mol Sci. 2020 Oct 29;21(21):8062.

44. Liu Y, Liu L. The pro-tumor effect and the anti-tumor effect of neutrophils extracellular traps. BST. 2019 Dec 31;13(6):469–75.

45. Lo YMD, Han DSC, Jiang P, Chiu RWK. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. Science. 2021 Apr 9;372(6538):eaaw3616.

46. Sicard G, Fina F, Fanciullino R, Barlesi F, Ciccolini J. Like a Rolling Stone: Sting-Cgas Pathway and Cell-Free DNA as Biomarkers for Combinatorial Immunotherapy. Pharmaceutics. 2020 Aug;12(8):758.

47. Luecke S, Holleufer A, Christensen MH, Jønsson KL, Boni GA, Sørensen LK, et al. cGAS is activated by DNA in a length-dependent manner. EMBO reports. 2017 Oct;18(10):1707–15.

48. Trejo-Becerril C, Pérez-Cardenas E, Gutiérrez-Díaz B, De La Cruz-Sigüenza D, Taja-Chayeb L, González-Ballesteros M, et al. Antitumor Effects of Systemic DNAse I and Proteases in an In Vivo Model. Integr Cancer Ther. 2016 Dec;15(4):NP35–43.

49. Han DSC, Ni M, Chan RWY, Chan VWH, Lui KO, Chiu RWK, et al. The Biology of Cell-free DNA Fragmentation and the Roles of DNASE1, DNASE1L3, and DFFB. Am J Hum Genet. 2020 Feb 6;106(2):202–14.

50. Benzekry S. Artificial Intelligence and Mechanistic Modeling for Clinical Decision Making in Oncology. Clinical Pharmacology & Therapeutics. 2020;108(3):471–86.