



HAL
open science

Computational modeling approaches for circulating cell-free DNA in oncology

Linh Nguyen Phuong, Sébastien Salas, Sébastien Benzekry

► **To cite this version:**

Linh Nguyen Phuong, Sébastien Salas, Sébastien Benzekry. Computational modeling approaches for circulating cell-free DNA in oncology. 2024. hal-04481689v1

HAL Id: hal-04481689

<https://hal.science/hal-04481689v1>

Preprint submitted on 28 Feb 2024 (v1), last revised 1 Oct 2024 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Computational modeling approaches for circulating cell-free DNA in oncology

Linh Nguyen Phuong¹, Sébastien Salas^{1,2} and Sébastien Benzekry¹

¹ COMPUTational pharmacology and clinical Oncology Department, Inria Sophia Antipolis-Méditerranée, Cancer Research Centre of Marseille, Inserm UMR1068, CNRS UMR7258, Aix Marseille University UM105, Marseille, France

² Assistance Publique-Hôpitaux de Marseille, Timone Hospital, Aix Marseille University, Marseille, France

Social media handles:

Authors: @SBenzekry

Institutions: @aphm_actu, @inria_sophia, @crcm_marseille

ABSTRACT

Liquid biopsy has emerged as a powerful tool for cancer early diagnosis, prognosis, and treatment monitoring across a wide range of cancer types. The non-invasive collection of blood markers enables real-time insights into the disease biology. Cell-free circulating DNA (cfDNA) offers a potential window into various biological and genetic processes, especially circulating tumor DNA directly originated from tumor cells.

Considering the attributes of cfDNA data, their inherent complexity, and the ease of collecting them over time, employing statistical modeling analyses appears necessary to extract relevant information. This review explores the diverse modeling approaches used to analyze cfDNA in oncology, emphasizing its role in oncology. After an overview of the current knowledge of cfDNA biology, the use of statistical analysis, machine learning, and non-linear mixed effects models is detailed for their application in clinical studies, both in processing cfDNA data at a particular time point and in capturing their temporal dynamics.

Overall, this review provides a comprehensive overview of the diverse modeling approaches applied to cfDNA in oncology, with a focus on dynamic approaches.

Keywords: cell-free DNA, circulating tumor DNA, fragmentomics, cancer, computational modeling, biomarker, early detection, prognosis, treatment monitoring

1 INTRODUCTION

2 Cell-free DNA (cfDNA) consists of fragments of encapsulated DNA released in body
3 fluids such as blood, urine or cerebrospinal fluid¹ (Figure 1A), allowing easy access
4 to

5 the genetic background of hardly-reachable tissues. Originating from events like
6 necrosis, apoptosis² or lysis³, the understanding of cfDNA biology, particularly in
7 oncology, is a current research focus. At bedside, applications are relevant for
8 diagnosis, prognosis and treatment monitoring. They also allow for adapting and
9 personalizing the therapeutic strategy.

10 The interest in studying cfDNA lies in its ability to encapsulate and share biological
11 and genetic information. Notably, circulating-tumor DNA (ctDNA), which is a part of
12 cfDNA, comprises fragments originated from tumor cells (primary tumor, circulating
13 tumor cells and micro- or macro-metastases) and may contain specific tumor
14 aberrations⁴. Consequently, cfDNA analyses can offer insights into the disease
15 biology, tumor genetic mutations and tumor growth.

16 Furthermore, cfDNA offers the benefit of being collected through liquid biopsies⁵
17 (Figure 1B), less invasive than tissue biopsies. This provides frequent biological
18 markers enabling real-time monitoring of the disease. Additionally, cfDNA is
19 representative of genetic information coming from all parts of the body, including
20 primary or metastatic tumor sites. This feature can help for the detection of new sites
21 and serve as a surrogate for tissue biopsies in case the primary site has not been
22 identified. Furthermore, liquid biopsies provide the unique ability to capture intra-
23 tumoral genetic heterogeneity⁶. This intricate genetic landscape is often missed by
24 tissue biopsies, which only sample a small portion of the tumor. Eventually, liquid
25 biopsies offer the valuable advantage of detecting genetic and genomic adaptations
26 during the treatment course.

27 The sequencing and quantification of cfDNA over time generated extensive datasets,
28 offering valuable insights into the biological dynamics between tumors and their
29 microenvironment. These data include the detectability or level of aberrations in
30 specific genes, fragment size profiles, ratio between short and long fragments, as
31 well as the description of base pair patterns (Figure 1C). To manage these data,
32 studies have turned to mathematical and computational modeling of cfDNA (CM-

33 cfDNA), comprising statistical, machine learning (ML) and mechanistic modeling. The
34 latter refers to simplifying real-world physical and biological concepts throughout
35 mathematical systems and computer programs. These approaches allow data
36 mining, predictive modeling and *in silico* simulations that have become necessary as
37 the data complexity continues to grow. Since the early 21st century, CM has played a
38 role in understanding individual variability in drug responses, paving the way to
39 personalized medicine. The emergence of CM-cfDNA around 2015, aligning with the
40 increasing popularity of cfDNA, constituted 24% of tumoral cfDNA studies in 2022
41 and raised to 28% in 2023 (“modeling” OR “computational modeling” OR “machine
42 learning” OR “survival analysis”) AND “tumoral cfDNA” / “tumoral cfDNA” PubMed
43 search, 249/883 entries in 2023). CM-cfDNA embraces three primary modeling
44 approaches: survival analysis, investigating time-to-event data (such as progression);
45 ML, which exhibits a substantial surge over the past two decades and contributed
46 significantly to medical decision-making (e.g., diagnosis and prediction of treatment
47 response); and mechanistic modeling, employing mathematical systems to describe
48 interactions between tumors, drugs, and biological systems.

49 Within the last eight years, CM-cfDNA has demonstrated efficacy in oncology for
50 personalized early detection, cancer subtype classification, prognosis, treatment
51 monitoring and prediction of response in multiple cancer types, including melanoma⁷,
52 lung⁸, breast⁹ and colorectal¹⁰ cancers (Figure 1D).

53 **BIOLOGY OF cfDNA AND ctDNA**

54 The history of cfDNA began with its discovery in 1948 by Mendel and Métais¹¹, who
55 first described the presence of nucleic acids in the blood. Later in 1977, it was
56 observed that patients with various types of cancer exhibited elevated levels of
57 cfDNA¹². A dozen years later, Stroun et al. demonstrated the presence of cancer-
58 derived DNA fragments in the blood of cancer patients, encouraging the exploration
59 of new blood-borne markers but also in other body fluids^{13,14}.

60 As previously mentioned, the biology of cfDNA is not fully understood, although it
61 appears that cfDNA may originate from apoptosis, necrosis, active secretion of
62 exosomes¹⁵ or the hematopoietic system¹⁶. The half-life of cfDNA ranges from 15
63 minutes to 2 hours¹⁷, allowing for a representative monitoring of the actual released
64 amounts over time. Nevertheless, cfDNA quantities and genetic aberrations provide
65 access to biological characteristics.

66

67 ***CfDNA quantification and sequencing assays***

68 Initially, numerous studies have explored cfDNA genotyping to quantify fragments,
69 detect specific tumor aberrations or discover new ones. These methods fall into two
70 categories; targeted and non-targeted approaches¹⁸. Targeted approaches rely on a
71 set of predefined genes, mainly associated to the patient's pathology. They comprise
72 digital droplet Polymerase Chain Reaction (PCR), Q-PCR, amplification-refractory
73 mutation system, BEAMing-PCR, tagged-amplicon deep sequencing and cancer
74 personalized profiling by deep-sequencing¹⁹. They present good sensitivity and better
75 specificity than non-targeted approaches¹⁸. Non-target approaches assess the entire
76 genome (array-comparative genomic hybridization, whole exome sequencing, whole
77 genome sequencing). While these methods have the potential to identify new genetic
78 aberrations through genome-wide screening, they require larger cfDNA amounts.

79 These quantification and sequencing methodologies have enabled the extraction of
80 relevant cfDNA characteristics under specific medical conditions.

81

82 ***CfDNA concentration***

83 For instance, the proportion of cfDNA coming from a single cancer depends on the
84 tumor's size. For a 100 grammes tumor (approximately 3×10^{10} tumor cells), 3.3%
85 of its DNA is released daily into the bloodstream²⁰. The presence of metastases and
86 circulating tumor cells also impacts on these quantities. In healthy individuals, cfDNA
87 concentrations typically range from 0 to 100 ng/ml of blood, with an average of 30
88 ng/ml, whereas the concentration of cfDNA in the blood of cancer patients varies
89 from 0 to 1,000 ng/ml, with an average of 180 ng/ml¹². The cancer stage also impacts
90 on the proportion of detected ctDNA within cfDNA, being approximately two times
91 smaller in stage I patients compared to stage III patients²¹.

92 However, elevated cfDNA levels, while not exclusive to cancer, can also result from
93 conditions like pro-inflammatory or auto-immune diseases, cirrhosis, hepatitis²², or
94 systemic lupus²³. Furthermore, quantities of cfDNA may also be influenced by
95 pregnancy. Studies have detected maternal and fetal cfDNA in the mother's blood,
96 with a rapid decrease of cfDNA levels after birth (mean ~ 16.3 min)²⁴. Lastly, intense
97 physical activity may increase cfDNA amounts due to inflammation²⁵. Consequently,
98 the origin of cfDNA could introduce a potential bias in the results.

99

100 ***CfDNA fragmentomics***

101 Another category of research, called fragmentomics, focuses on the study of the
102 fragments of cfDNA. This notion has been introduced in 2015²⁶. The authors showed
103 that fragment patterns differed across cancer patients and tissue of origin. Several
104 studies have since investigated various fragmentation features, such as fragment
105 sizes, end motifs, breakpoint motifs, jagged ends or nucleosome footprints (Figure
106 1C).

107 First, it was revealed that fragment sizes are relative to the clinical condition of a
108 patient. Healthy individuals typically exhibit cfDNA size distributions ranging from 130
109 to 200 base pairs (bp), with a peak at 166 bp²⁷. This phenomenon seems to be linked
110 to the nucleosome footprint (nucleosome positions)²⁸. This specific length
111 corresponds to the DNA fragments wrapped around a nucleosome core (~ 147 bp)
112 added to a linker fragment of ~ 20 bp²⁹. These fragments appear to originate from a
113 double process of apoptosis-induced proliferation and proliferation-induced
114 apoptosis². In contrast, longer fragments such as 10,000 bp fragments appear to
115 originate from necrosis or phagocytosis of necrotic tumor cells by macrophages³⁰,

116 and are much less present in the plasma². Additionally, cancer patients appear to
117 have an enrichment in shorter fragments (90-150 bp). Subsequently, focusing on
118 these fragments could enhance the detection of ctDNA³¹.

119 Second, some end motifs appear more frequently in healthy individuals, whereas
120 cancer patients present distinct and more variable end motifs³², which can be
121 relevant to find the tissue of origin. The proportion of jagged ends (uneven DNA
122 extremities, Figure 1C) was found to vary between tumor and wild-type fragments. It
123 appeared to increase in hepatocellular patients³³ and decrease in bladder cancer
124 patients³⁴, compared with healthy individuals.

125 Last, breakpoint motifs also vary between healthy individuals to cancer patients. The
126 proportion of AATTGC motifs is larger in cancer patients, whereas the GCAGTA,
127 GCACTT and CTCAAA motifs proportions are smaller³⁵.

128

129 ***CfDNA mutations and ctDNA***

130 A large body of literature examined mutations that can be detected on cfDNA
131 fragments using sequencing techniques. When these mutations are tumor specific,
132 the afferent cfDNA is then considered as part of ctDNA. These studies historically
133 preceded fragmentomics but examine data at a smaller scale, i.e. molecular
134 alterations detected by sequencing techniques (Figure 1B-C).

135 The detection of ctDNA quantities can be influenced by the primary tumor type. For
136 instance, ctDNA is detected in most patients with metastatic bladder, colorectal,
137 gastroesophageal or ovarian cancer²¹. Conversely, cancers such as prostate or
138 thyroid cancers appear to have few or undetectable ctDNA²¹. Detectability can also
139 vary depending on the investigated body fluid. In brain cancer, plasma ctDNA
140 typically accounts for less than 1% of cfDNA due to the blood brain barrier.
141 Therefore, cerebrospinal fluid or urine collection has proven effectiveness for
142 identifying ctDNA³⁶.

143 The analyze of cfDNA has shown power to monitor, e.g., the mutation level of
144 oncogene addicted cancers such as endothelial growth factor receptor (EGFR) or
145 Kirsten rat sarcoma viral oncogene homologue (KRAS) positive lung cancers, in the
146 adaptation of treatment lines.

147 Blood EGFR mutations seem to match with tissue EGFR mutations, suggesting a
148 transition from tissue sampling to liquid biopsy³⁷. For example, anti-EGFR gefitinib
149 has shown efficacy in blood EGFR-positive NSCLC patients in first line ³⁷. Similarly,
150 renal cancer patients carrying at least one cfDNA TP53 mutation had shorter
151 progression-free survival³⁸. In colorectal cancer, patients initially treated with an anti-
152 EGFR antibody, the on-treatment monitoring of genetic alterations in, e.g., the KRAS
153 or EGFR genes could reveal the emergence of resistance, enabling the early
154 adaptation of treatment³⁹. Finally, the detection of more than four KRAS mutant
155 copies per milliliter in peripheral blood of pancreatic ductal adenocarcinoma has been
156 linked with shorter PFS (HR=3.4 (1.2-9.7)) at pre- and post- resection, as well as in
157 portal venous blood (HR=4.6 (1.6-13.3))⁴⁰.

158 All this genetic and biological variability, together with the data type diversity calls for
159 the use of CM-cfDNA to develop new tools for personalized diagnosis, treatment
160 setup and monitoring (Figure 1D).

161

162 **CFDNA COMPUTATIONAL MODELING FOR DIAGNOSIS**

163 One of the primary goals of CM-cfDNA is to detect tumors before clinical symptoms.
164 Some studies aimed to discern changes in cfDNA characteristics to discriminate
165 pathological to healthy individuals (Table 1) or to distinguish between different tumor
166 subtypes. Moreover, identification of the primary tumor site is a major challenge. In
167 the plasma, ctDNA can originate from any body part, making specific ctDNA
168 mutations essential for providing clues about the primary tumor location.
169 Consequently, multiple studies conducted pan-cancer analyses.

170

171 ***CtDNA-based machine learning***

172 Classical cfDNA studies primarily employ conventional statistical hypothesis testing
173 to identify dependencies between cfDNA markers and categorical outcomes, such as
174 diagnosis, or cancer type/subtype. Additionally, ML models (Figure 2A), especially
175 supervised classification methods, have been increasingly employed in the last
176 decade for early cancer detection⁴¹. These statistical analysis techniques enable
177 learning from an initial patient dataset to predict diagnosis for new patients.

178 Logistic regression (LR) models and predicts the probability of a binary data, seeking
179 a linear relationship between the log-odds of the event occurring and variables.
180 Support vector machines (SVM) (Figure 2A) aim to find the hyperplane that
181 maximally separates individuals into two or more classes. Decision trees are a series
182 of interconnected binary choices, enabling regression and classification. Random
183 forests (RF) learn from multiple decision trees trained on different features subset.
184 Finally, deep learning, a subset of machine learning based on neural networks is
185 mostly used for complex features patterns and relationships and consists of
186 interconnected artificial neurons that evaluate a weighted sum of inputs and pass the
187 results to the next-layer neurons through a nonlinear activation function.

188 ML models have enabled the differentiation between lung adenocarcinoma, lung
189 squamous carcinoma and squamous cell lung cancer, leveraging copy number
190 profiling of cfDNA⁴². This distinction was achieved comparing five classifiers: RF,
191 SVM, LR with ridge, elastic-net (EN) or least absolute shrinkage and selection
192 operator (LASSO) regularizations. Another study used SVM to select discriminative
193 differentially methylated blocks for early lung cancer detection⁴³, detecting 52–81% of

194 the stages IA to III patients, with a specificity (true negative rate) of 96% (95%
195 confidence interval (CI) 93–98%). Liu et al.⁴¹ reviewed the most relevant and recent
196 early detection ML-based studies, noting four CM-cfDNA studies employing linear
197 models, possibly with elastic-net (EN) or least absolute shrinkage and selection
198 operator (LASSO) regularization. One identified elevated plasmatic cfDNA levels in
199 oral cancer patients, compared to control subjects⁴⁴. Three CM-cfDNA studies
200 employed SVM, one investigating a significant mammalian DNA epigenetic
201 modification (5-hydroxymethylcytosine) for pancreatic ductal adenocarcinoma early
202 detection⁴⁵. Five studies used RF, including one exploring 5-hydroxymethylcytosine
203 patterns to discriminate among seven cancer types, achieving 87.5% and 92%
204 accuracy for two datasets⁴⁶. Another one identified new biomarkers in the plasma
205 cfDNA methylome profiling to diagnose and locate gastrointestinal cancers, resulting
206 in an area under the curve of 0.96 ± 0.04 (mean \pm standard deviation), 0.89 ± 0.06 ,
207 0.91 ± 0.07 for hepatocellular carcinoma, colorectal cancer and pancreatic cancer
208 respectively⁴⁷.

209 Deep learning has been less frequently used due to small sample size of these types
210 of studies. Nevertheless, Liu et al.⁴¹ cited one study using it for early stage lung
211 cancer detection⁴⁸. They employed a convolutional neural network with two-
212 dimensional grids representing the sequenced reads. Each column was representing
213 a read and exclusively colored according to the row corresponding to its respective
214 base (A, C, G, T, or N if unrecognized). The algorithm then detected base changes,
215 such as deletions, mutations or insertions, focusing on distinguishing artifacts from
216 genuine cancer mutations.

217 Moser et al.⁴⁹ cited over twenty diagnosis studies employing ML models or
218 combinations of them. A common issue involves somatic mutations of non-cancerous
219 origin, increasing the number of false positives. For instance, age can induce the
220 development of somatic mutations during clonal hematopoiesis, resulting in
221 misleading results. Chabon et al.⁵⁰ developed a classification framework (5-nearest
222 neighbor, 3-nearest neighbor, naïve Bayes, LR and decision tree) to distinguish
223 tumor from clonal hematopoiesis mutations and matched them with risk-matched
224 controls. They demonstrated that clonal hematopoiesis mutations tended to occur in
225 longer cfDNA fragments.

226 Eventually, unsupervised learning (Figure 2A) has also been employed to classify
227 individuals into cancer subtypes. Luo et al. used hierarchical clustering to distinguish
228 colorectal cancer patients from normal subjects according to methylation markers,
229 and finally classify patients into two subgroups with different overall survival (OS)¹⁰.

230

231 ***Fragmentomics-based machine learning***

232 Chen et al. used LR to differentiate hepatocellular carcinoma from liver cirrhosis and
233 healthy controls using four cfDNA fragmentome features: genome-wide 5-
234 hydroxymethylcytosine, nucleosome footprint, 5' end motif and fragmentation profiles.
235 This led to a 95.4% sensitivity and a 97.8% specificity in the test set⁵¹. Similarly, Duo
236 et al. compared LR, deep learning and extreme gradient boosting to early detect lung
237 adenocarcinoma, using the 6bp breakpoint motif, defined as « the 3bp extensions to
238 both directions of the aligned cfDNA 5' », achieving a 92.5% sensitivity and a 90.0%
239 specificity in the external validation cohort³⁵. Ma et al. compared five ML algorithms
240 (generalized linear model, deep learning, RF, gradient and extreme gradient
241 boosting), integrating fragment size ratio and distribution, end and breakpoint motif,
242 and copy number variation⁵². They reached a 94.8% specificity and 98% sensitivity to
243 distinguish healthy individuals from early-stage colorectal adenocarcinoma.

244 An important and influential work has been performed by Cristiano et al.⁵³. They
245 employed the cfDNA integrity index, defined as the ratio of short fragments (100-150
246 bp) to long fragments (150-200 bp), across 504 genome bins, to classify healthy
247 individuals and cancer patients (with seven different pathologies). They highlighted
248 distinct variations in fragment size profiles across different genome regions of cancer
249 patients. The features were then integrated into a stochastic gradient tree boosting
250 framework. Samples were split according to a 10-fold cross-validation repeated 10
251 times, with features selected at each of the ten steps on the inner-fold training
252 dataset. At each iteration, the model was estimated on the training dataset and
253 evaluated on the test dataset. Predictions were made based on the average of
254 predictions over the hundred steps. With a 95% specificity, they detected 80% of the
255 cancer patients. Expanding the framework, they identified tumor tissue origin with
256 90% specificity, 61% accuracy and reaching 75% accuracy when looking at the top
257 two predictions.

258 Mathios et. al implemented a comparable ML approach based on similar
259 fragmentation features for lung cancer detection and stage identification in high-risk
260 symptomatic patients⁵⁴. Initially, they reduced dimensionality of the fragmentation
261 features by selecting principal components explaining 90% of the fragmentation
262 variance. Subsequently, they used a LASSO-penalized LR to assess the
263 fragmentation components along with 39 chromosomal arms Z-scores (number of
264 standard deviations from the mean of the mapped read fraction). With 10 replicates
265 of a 5-fold cross-validation, they defined a score able to detect 94% of the cancer
266 patients with 80% specificity, in a population with 91% early-stage (I-II) cancer
267 patients.

268 Mouliere et al. also developed CM-cfDNA ML (RF and LR)³¹. They used the
269 proportion of fragments in multiple size ranges, ratios of some of these proportions,
270 and 10 bp periodicity oscillation amplitudes occurring before 150 bp to detect cancer
271 patients, even for pathologies presenting weaker levels of ctDNA. They reached an
272 area under the receiver operating characteristic curve (AUC) of 0.891 for cancers
273 with low amounts of ctDNA (pancreatic, renal and glioma) with RF, having selected 4
274 features among 9.

275 Unsupervised learning is less commonly encountered in the literature. Renaud et al.
276 performed such analysis to detect the presence of cancer fragments in the overall
277 cfDNA thanks to fragments lengths, measured by shallow whole genome
278 sequencing⁵⁵. The aim was to decompose the matrix of the fragment size profiles
279 through the non-negative matrix factorization (NMF) method:

$$sample_{n \times m} = weights_{n \times k} \times signatures_{k \times m},$$

280 where:

281 - each row i of $sample_{n \times m}$ represents a sample; each column j represents a
282 fragment size. One matrix cell indicates the relative fragments frequency with length
283 represented in the column j , for the sample i .

284 - $signatures_{k \times m}$ represents the signature matrix. Each row corresponds to a typical
285 profile of fragment lengths, according to the source of the fragments. The
286 hyperparameter k is the number of sources used in the factorization and must be set.
287 As an example, $k = 2$ sources could represent typical healthy and cancer
288 distributions of the fragment lengths.

289 - $weights_{n \times k}$ are the weights associated to each signature profile for each sample.
290 They initially calibrated the number of sources to $k = 2$ and inferred signatures by the
291 NMF method. They found correlations between the weights and ctDNA ratios (variant
292 allele fractions) ($r = 0.75$). Eventually, NMF was applied to a cohort of healthy and
293 pathological individuals with diverse cancer types⁵³. The weights were integrated into
294 an SVM framework to detect cancer patients. By increasing the number of sources to
295 30, they reached an AUC of 0.95.

296

297 ***Mechanistic modeling for annual screening***

298 Using longitudinal data, Avanzini et al. developed a mechanistic model of ctDNA
299 shedding during apoptosis linked to the tumor size evolution over time to determine
300 the optimal screening time for early lung cancer detection⁵⁶. They modeled the
301 expected number of ctDNA haploid genomic equivalent (hGE) circulating in the
302 bloodstream for a tumor with size M as a Poisson-distributed random variable, with
303 mean:

$$C = M \times \frac{d \times q_d}{\epsilon + r}$$

304 where:

305 - d is the tumor death rate per day. $d = b - r$ where b is the cell division rate per day
306 and r is the net tumor growth rate per day.

307 - q_d is the mean shedding rate of a cell death. On average, $q_d \approx 0.1$ hGE per cell
308 death.

309 - ϵ is the ctDNA elimination rate per day.

310 Using mechanistic modeling and considering various sources of biological and
311 technical errors, they could predict the expected tumor detection size. They
312 demonstrated that ctDNA-based annual screening had a median detection size of 2.0
313 to 2.3 cm of diameter, against 3.5 cm for usual annual screening, highlighting the
314 optimal marker-frequency combination for cancer screening.

315 Taken together, there has been in recent years a dwealth of studies and results
316 related to CM-cfDNA-based early cancer detection, mainly focusing on cfDNA
317 aberrations. Drug and medical regulatory organizations have begun to recommend
318 the use of cfDNA. In 2020, the FoundationOne Liquid CDx test⁵⁷, analyzing cfDNA-

319 based genes, has been approved by the Food and Drug Administration. In 2022, the
320 European Society for Medical Oncology Precision Medicine Working Group
321 published an article warranting the use of ctDNA as an adjunctive diagnostic tool⁵⁸.

322 In summary, ML is increasingly integrated to find associations between fragments
323 features and diagnosis outcomes such as cancer detection, stage or histological
324 type. Fewer researchers modeled cfDNA kinetics by integrating biological
325 mechanisms to prove the effectiveness of these fragments as novel biological
326 markers. Finally, limitations include the necessity to evaluate prospectively these
327 methods on asymptomatic individuals at risk for cancer⁵⁹, in order to generalize their
328 applicability in real-world conditions and detect cancer before symptoms arise.

329

330 **CFDNA COMPUTATIONAL MODELING FOR PROGNOSIS AND** 331 **TREATMENT PREDICTION**

332 Many studies used CM-cfDNA to monitor tumor size and predict therapeutic
333 responses (Table 2). Their aim was to identify new signatures enabling early
334 treatment adjustments and preventing adverse events.

335 Most of them relied on baseline markers (cfDNA/ctDNA concentration, ctDNA
336 positivity, fragmentomics). These markers are typically assessed at one specific time,
337 either following surgical resection to establish associations with time to relapse, or
338 just before treatment initiation to determine correlations with imaging-evaluated
339 treatment response. Sometimes, these studies also rely on biological markers
340 collected at various time points, including at some or all treatment cycles. They
341 provide longitudinal datasets including absolute values and/or relative changes from
342 baseline data. These datasets enable comprehensive analysis of cfDNA dynamics,
343 aiming to discover new patterns associated with time to relapse, disease
344 progression, treatment response, or mortality.

345 These studies first employed classical statistical analysis (e.g., survival analysis) to
346 establish connections between cfDNA measurements and relapse or response to
347 treatment. However, computational methodologies have evolved to ML, non-linear
348 mixed effects models (NLME) and mechanistic modeling.

349

350 ***Classical survival analysis***

351 Survival analysis (Figure 2B) analyzes the duration before a specific event, such as
352 progression or death, and specifically accounts for censored data (unreached event).
353 The main methods include Kaplan-Meier (KM) estimation and univariable
354 /multivariable Cox proportional-hazards regression (CPHR).

355 In the fragmentome field, Lapin et al. demonstrated an association of fragment sizes
356 smaller than 147 bp and high cfDNA levels pre-treatment with shorter PFS and OS,
357 using Kaplan-Meier estimation in advanced pancreatic cancer patients⁶⁰.
358 Multivariable CPHR demonstrated that cfDNA levels could predict PFS (Hazard Ratio
359 (HR): 3.05, 95% CI: 1.40-6.65) and OS (HR: 2.24, 95% CI: 1.09-4.59).

360 Moding et al.⁶¹ monitored ctDNA molecular residual disease in advanced non-small
361 cell lung cancer (NSCLC) patients, collecting ctDNA immediately after
362 chemoradiation therapy (CRT), later followed or not by consolidation with immune-
363 checkpoint inhibition (ICI). A second ctDNA sample was collected early during ICI
364 treatment for the immunotherapy-treated arm. Undetectable ctDNA before ICI
365 treatment correlated with good prognosis, irrespective of ICI treatment. Detectable
366 ctDNA early on ICI also correlated with a shorter progression-free survival (PFS).
367 Analyzing ctDNA changes over time demonstrated that increased ctDNA levels were
368 associated with worse prognosis compared to a decrease.

369 Powles et al.⁶² performed CPHR to predict atezolizumab response in urothelial
370 carcinoma. They collected plasma onset and early on treatment (first day of the first
371 and third cycles). They revealed significant disease-free survival differences
372 according to ctDNA changes over time. Patients with ctDNA clearance appeared to
373 have three to four times lower relapse risks, according to univariable, stratified and
374 multivariable CPHRs.

375

376 ***Longitudinal dynamics modeling***

377 Few studies have explored dynamical CM-ctDNA (Figure 2C). One study developed
378 of a mechanistic modeling of the joint ctDNA–tumor size evolution over time, to
379 assess atezolizumab response in NSCLC and melanoma patients⁶³. The authors
380 used a bi-exponential system to independently describe both the log10-transformed
381 number of mutant molecules per mL and the sum of the longest diameters (SLD) of
382 lesions, assessed by the RECIST 1.1 criteria. Parameters included the model-
383 estimated value at the first time point, the growth rate and the decay rate. The
384 estimated ctDNA growth rate showed high correlation with the estimated SLD growth
385 rate. The final joint system is:

$$SLD(t) = SLD_0 (e^{-k_{sT} \cdot t} + e^{k_{gT} \cdot t} - 1)$$
$$ctDNA(t) = ctDNA_0 (e^{-\zeta \cdot k_{sT} \cdot t} + e^{k_g \cdot t} - 1),$$

386 where:

387 - SLD_0 and $ctDNA_0$ are the baseline values of SLD and $ctDNA$ respectively.

388 - k_{gT} is the growth rate of the tumor size.

389 - k_{gT} is the decay rate of the tumor size.

390 - k_g is the decay rate of the ctDNA level.

391 ζ is the coefficient linking the tumor size growth rate to the ctDNA growth rate.

392 This system fitted well the ctDNA and tumor kinetics over time, even when negatively
393 correlated (one increasing while the other decreasing). Thus, they highlighted the
394 mechanistic link between tumor growth and ctDNA release for patients under ICI.

395 Janssen et al. employed NLME (Figure 2C) for ctDNA biomarkers analysis to predict
396 early treatment responses⁶⁴. They used a NLME model to describe the dynamics of
397 EGFR mutations in ctDNA from NSCLC patients treated with erlotinib or gefitinib. In
398 this model, the concentration of three mutations was modeled by a zero-order growth
399 model, chosen between baseline, turnover and first-order growth models. The first
400 model described both L858R or exon19del (driver) mutations concentrations, while
401 the second one described T790M mutation concentrations. The equation writes:

$$\frac{dy}{dt} = k_{in} - k_{out} \cdot y(t) \cdot R(t) \#(1)$$

402 where $R(t) = e^{-\lambda t}$ for driver mutations, and $R(t) = e^{-\lambda \cdot y(t)}$ for the T790M mutation
403 concentration. Here:

404 - $y(t)$ is the change in either L858R or exon19del over time.

405 - k_{in} represents the zero-order increase in mutations concentrations.

406 - k_{out} represents the drug-driven decrease in mutations concentrations.

407 - $R(t)$ accounts for the time-dependent resistance development where λ is the
408 progression rate.

409 Another parameter was estimated to consider the baseline mutations concentration,
410 unavailable in this study. The growth model was fitted to observed concentrations to
411 identify each parameter. Subsequently, the predicted time-course of mutations
412 concentrations were compared to the observed ones, revealing that a zero-order
413 increase and a first-order elimination model (equation 1) best approximated the
414 actual concentrations.

415 These predicted ctDNA values were integrated into parametric survival models to
416 predict PFS. Among exponential, Weibull and Gompertz hazard models, Weibull was
417 found to be the more efficient. Considering all the predictors of disease progression

418 at random timepoints post-treatment initiation, including relative changes from
419 baseline and absolute values of the three mutations concentrations, only the relative
420 change in driver mutations was statistically significant for stratifying responders and
421 non-responders ($p = 0.001$, likelihood-ratio test). This significance was validated
422 using stratified KM curves. Consequently, patients with a predicted relative change
423 from baseline greater than zero (median value) experienced a shorter disease
424 progression.

425 Prior to this study, Khan et al. sought to model the carcinoembryonic antigen
426 dynamics, which is proportional to the total number of tumor cells⁶⁵, to assess
427 cetuximab response in colorectal cancer patients. They used the following equation
428 for the tumor burden:

$$N(t) = n_s e^{-\lambda_s t} + n_r e^{\lambda_r t},$$

429 where the cells number $N(t)$ is divided into a population of treatment sensitive cells
430 n_s , dying under treatment at rate λ_s , and a population of treatment resistant cells n_r ,
431 growing at rate λ_r .

432 CfDNA mutant frequencies was modeled using a single exponential growth model.
433 Rates were estimated for each patient with at least three time points. The cfDNA
434 model well described the real dynamics ($R^2 = 0.979$). Particularly, the cfDNA relapse
435 rate was shown to be correlated with the tumor one. By comparing relapse rates with
436 RECIST v1.1 criteria, the model could precisely predict the time to relapse.

437 Fragmentomics data are emerging as prognosis markers but remain underused for
438 treatment monitoring. Some studies have investigated the cfDNA integrity index
439 variations in breast cancer during neoadjuvant or adjuvant chemotherapy. Only two
440 studies were found searching for “fragmentomics” AND “chemotherapy” on PubMed,
441 and only one searching for “fragmentomics” AND “immune-checkpoint inhibitors” (no
442 results for “fragmentomics” AND “immunotherapy”). The latter investigated cfDNA
443 fragmentation profiles in lung carcinoma and diffuse B cell carcinoma patients⁶⁶.
444 They calculated an « expression inference from cfDNA-sequencing » score, based
445 on a statistic of expression levels changes of genes before ICI-treatment and after ~
446 4 weeks. This score allowed to identify both patients with durable clinical benefit of
447 ICI and shorter PFS (HR: 11.38, Wald test: $p = 0.006$). They also developed a
448 mechanistic model of the nucleosome accessibility at transcription start sites regions.
449 The model established a connection between this accessibility and fragmentation

450 profiles and expression levels. The model was used to perform simulations and
451 explore the parameters influencing the detection of a specific gene expression within
452 cfDNA.

453

454 ***Longitudinal ML modeling***

455 ML leverages longitudinal data by merging the features from each time point,
456 comprising absolute and relative changes over time and integrating them into ML
457 models to classify patients as either responders or non-responders, employing
458 supervised or semi-supervised classification methods.

459 Assaf et al.⁶⁷ conducted such a study with 466 NSCLC patients. They developed a
460 ML framework to predict immunotherapy response using longitudinal ctDNA data.
461 CtDNA was collected at baseline (before treatment) and at day 1 of cycle 2 (C2D1),
462 and cycle 3 (C3D1). The models integrated 19 ctDNA levels metrics and 59 relative
463 ctDNA changes from baseline. Three models were compared: only baseline features,
464 baseline + C2D1 and baseline + C2D1 + C3D1. The latter was selected for OS
465 prediction, as it exhibited the highest C-index. Then, the ctDNA features were
466 combined with baseline clinical features, including ECOG status, metastases count,
467 age group, sex, smoking history, PD-L1 status, and the sum of lesion diameters.
468 Baseline and C2D1 tumor size was also considered. They employed an Elastic Net
469 (EN) approach with leave-one-out cross-validation (LOOCV) conducted through a 10-
470 nested cross-validation process. Features were retained if they were selected in over
471 50% of cross-validation iterations and if the gain metric was positive according to the
472 next-door analysis. This analysis involves fitting the same model after removing one
473 predictor and comparing the error rate to the one of the full model. Consequently, the
474 feature set was ultimately reduced to five features, including the global cfDNA
475 concentration at C3D1. The final model categorized patients into three groups: high
476 risk (progressive disease), intermediate risk (stable disease) and low risk
477 (responders), and KM curves demonstrated significant risk stratification in both
478 training and testing datasets.

479 Similar to the previous study but in a smaller cohort of 94 NSCLC patients treated
480 with atezolizumab or docetaxel, Zou et al. applied LOOCV LASSO-penalized
481 regression, linking ctDNA metrics (collected at baseline, C2D1 and C3D1) to OS⁶⁸.

482 The model highlighted the C3D1 median number of mutant molecules per mL as the
483 most important predictor for OS.

484

485 **CONCLUSION**

486 Recent studies have delved into the use of cfDNA as an innovative biological marker
487 for cancer detection and treatment monitoring in the last years. Most of these studies
488 focused on early detection and demonstrated sufficient level of evidence to prompt
489 regulators to recommend cfDNA as a complementary diagnostic tool^{57,58}. There is,
490 however, a pressing need for standardizing ctDNA detection and cfDNA
491 quantification methods. Furthermore, over the past two years, approximately one in
492 four studies employed CM-cfDNA as a support for precision medicine. While the
493 majority concentrated on classical survival and classification analyses, there is a
494 growing need to model biological mechanisms over time to improve precision, which
495 has been addressed by only few studies.

496 To enhance the reliability of findings and assess the methods' reproducibility, it is
497 also crucial to conduct studies on larger datasets, with validation on external cohorts.
498 Currently, most assays train and test their models on data from symptomatic or
499 diagnosed patients, or even high stage patients. Additional studies that specifically
500 target and validate markers on data from asymptomatic individuals are warranted.

501 In contrast to its use for diagnosis, CM-cfDNA for patient prognosis and treatment
502 monitoring is less common. Nevertheless, it is particularly well suited for such
503 purposes. Specifically, mechanistic modeling parameterized on experimental data
504 facilitates the better understanding of biological mechanisms behind cfDNA release
505 in body fluids. Combining mechanistic modeling with ML methods and survival
506 analysis (mechanistic learning⁶⁹) holds the premise to develop genuinely informative
507 biologico-computational markers and associated predictive tools.

508 Most research efforts have focused on ctDNA analysis, targeting mutations of a
509 specific cancer and requiring a proper understanding of pathology-associated genetic
510 aberrations. But ctDNA is only detectable in less than 80%⁷⁰ of the cancer patients,
511 making it worthwhile to further explore the use of alternative cfDNA characteristics,
512 such as global concentration, methylation, fragment size profiles and end motifs, in
513 the quest for pan-cancer biomarkers.

514 To this regard, optimizing data acquisition, improving data relevance, and diversifying
515 data types will be the first keys to finally better adapt and improve CM-cfDNA.

516 **FIGURE LEGENDS**

517 ***Figure 1: Cell-free DNA data: a new biological tool for on oncology***

518 A. Cell-free DNA (cfDNA) are fragments of encapsulated data released in the human
519 body fluids, such as blood, urine, cerebrospinal liquid. Some of these fragments are
520 originated from tumoral cells, which can be primary, metastatic or circulating tumor
521 cells. Plasmatic cfDNA is the most analyzed because of its ease of collection.

522 B. After the blood collection, cfDNA is extracted and amplified, usually by
523 Polymerase Chain Reaction (PCR) methods. Fragments are then sequenced by
524 various methods, targeted ones (which target specific genes) and non-targeted
525 approaches (sequencing the whole genome) to provide information at smaller case of
526 the molecular alterations.

527 C. PCR and sequencing processes yield diverse cfDNA data. a) Global concentration
528 is the first main quantitative feature describing cfDNA. b) Fragmentomics study a
529 wide range of data, focusing on the fragment sizes and patterns. It provides a profile
530 of fragment size distribution, enabling the extraction of quantities of fragments of
531 various sizes. The sequencing of the fragments also provides nucleotide base
532 patterns from end motifs, jagged ends, and breakpoint motifs. Additionally,
533 fragmentomics work on nucleosome footprints and cfDNA integrity index (cfDI),
534 calculated as the ratio between short and long fragments at a same locus. c) At a
535 smaller case, ctDNA mutations are predominantly analyzed, using features such as
536 ctDNA positivity (number of mutations detected greater than x mutations), ctDNA
537 concentration (copies number per milliliter), or variant allele frequency.

538 D) These cfDNA data are incorporated into computational modeling frameworks to
539 identify associations with clinical outcomes. This enables the establishment of cfDNA
540 as new biological marker in cancer research, serving for diagnosis, prognosis, and
541 prediction.

542

543 ***Figure 2: Cell-free DNA data: a new biological tool for oncology***

544 A) i) To early diagnose, evaluate the cancer type or stage, make prognosis or predict
545 response to treatment, studies compute machine learning methods. CfDNA data are
546 collected at different moments of the cancer progression according to the outcome to
547 predict. Data can be collected over time, but they are not considered as time-
548 dependent during modeling. Machine learning methods can be divided into two major
549 groups: ii) the unsupervised learning and iii) the supervised learning models. ii) First
550 ones are built to discriminate k groups within the complete set of individuals or
551 reduce dimensionality of the features space. The idea is to find individuals that are
552 closed into the space of features. A typical unsupervised learning method is the
553 hierarchical clustering, which build a hierarchy of individuals groups. Another one is
554 the non-negative matrix factorization, which decomposed a matrix of non-negative
555 elements into two matrices, for example by factorizing a matrix of cfDNA size profiles
556 into a coefficient's matrix and a matrix of size profile signatures. iii) Supervised
557 learning methods learn outcome's individuals on a train set to then predict outcomes
558 of a new cohort of patients. Most common supervised algorithms in cfDNA modeling
559 are the logistic regression, support vector machines (SVM), decision trees and
560 random forest. Neural networks are used mostly in the case of complex patterns and
561 relationships between features.

562 B) i) Classical survival modeling gathers technical tools that enable the modeling of a
563 duration until the occurrence of an event. Progression and death are the main events
564 modeled in the medical domain. Thus, individuals may be censored as the event
565 never occur during the study's time, due to the track loss of the patient, or the end of
566 follow-up by the study. In those cases, the event of progression or death is not
567 observed: patients are referred to as censored.

568 ii) A usual nonparametric estimation is the Kaplan-Meier one, which allows to
569 visualize and check hypothesis about the ability of a variable to discriminate long to
570 short survival.

571 iii) The Cox proportional-hazard regression is a widely used method for the analyze
572 of censored time data in survival modeling. This method assumes that the effect of
573 predictor variables on the hazard rate remains constant over time. Cox regression
574 helps to identify significant features as machine learning regressions do, estimate the
575 hazard ratios, which indicate the proportional changes in the hazard for one unit

576 change in a predictor variable. Additionally, it may generate survival curves (survival
577 probability over time for different levels of the feature).

578 C) i) Longitudinal data may be modelled as time-dependent data, to follow the
579 evolution of cfDNA kinetics during treatment. ii) Mechanistic modeling integrates
580 biological hypothesis and fundamental principles, known to induce the observed
581 kinetics, into a dynamic system. The models are then parameterized on experimental
582 data thanks to non-linear mixed effect models, which allows a better understanding of
583 the biological mechanisms and the validation of hypothesis.

584

585

586 **TABLES**587 **Table 1: Summary of cfDNA computational modeling studies for early**
588 **diagnosis, organ, stage and histological classification**

Source	Cancer	Modeling	Marker	Purpose
46	Pancancer	RF	ctDNA aberrations	Cancer type classification
71	Colorectal	LASSO – LR		Early diagnosis
72	NSCLC	Linear regression		
50		5nn – 3nn naïve Bayes – LR – Decision tree		
48	Lung	CNN		
45	Pancreas	RF – SVM – EN LR		Stage classification
42	NSCLC	LASSO – Ridge – EN LR		Histological classification
44	Oral	LR	cfDNA quantification	Early diagnosis
56	Lung	Mechanistic modeling	cfDNA concentration / ctDNA mutations	Early diagnosis
54	Lung	PCA – LR	Fragmentome	Early diagnosis
31	Pancancer	RF – LR		
51	Hepatocellular carcinoma	LR		
35	Lung adenocarcinoma	LR – deep learning, gradient and extreme gradient boosting		
52	Colorectal	Generalized linear model, deep learning, RF, extreme		

	adenocarcinoma	gradient boosting		
53	Pancancer	Gradient-tree boosting		Early diagnosis and cancer type classification
73	Colorectal	LASSO LR	Methylome	Early diagnosis
10		Hierarchical clustering		
43	Lung	SVM		
74	SCLC	PCA		Histological classification
47	Gastrointestinal	RF		

589 NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer; RF: random forest;
590 LASSO: least absolute shrinkage and selection operator; LR: logistic regression; X -
591 nn: X nearest neighbors; CNN: convolutional neural network; SVM: support vector
592 machine; EN: elastic-net; PCA: principal component analysis; ctDNA: circulating
593 tumoral DNA; cfDNA: cell-free DNA.

594 **Table 2: Summary of cfDNA modeling assays for treatment monitoring**

Source	Cancer	Treatment	Modeling	Marker	Baseline / Longitudinal
76	NSCLC	(Durvalumab \pm tremelimumab) + platinum-based chemotherapy	CPH	ctDNA aberrations	Baseline
53	Pancancer	Anti-EGFR or anti-ERBB2	KM		
77	Hepatocellular carcinoma	Atezolizumab + bevacizumab	KM – CPH		
8	NSCLC	Atezolizumab or docetaxel			
78	Pancancer	ICI			
79	Colorectal	Surgery or chemotherapy			
80	Melanoma	(Pembrolizumab or nivolumab) \pm ipilimumab	KM – CPH – LR		

7	Melanoma	Ipilimumab	Descriptive statistics		
81	Pancancer	ICI	KM		
61, 82, 83, 84	NSCLC	///	KM – CPH		
62	Urothelial carcinoma	Atezolizumab			
85	Melanoma	Nivolumab ± ipilimumab			
86	Pancancer	Durvalumab ± tremelimumab			
87	NSCLC	ICI		KM – CPH – Bayesian probit model	
88	Pancancer	Pembrolizumab	KM – CPH – LR		Longitudinal
63	NSCLC / Melanoma	ICI ± cobimetinib	NLME		
68	NSCLC	Atezolizumab or docetaxel	LASSO linear model		
64	NSCLC	Erlotinib / gefitinib	Mechanistic modeling / NLME		
65	Colorectal cancer	Cetuximab	Mechanistic modeling		
67	NSCLC	(Atezolizumab ± bevacizumab) + carboplatin + paclitaxel	EN linear regression	cfDNA concentration / ctDNA mutations	
38	Renal	(Ipilimumab + nivolumab) or anti-VEGFR-TKIs	KM – CPH – LR	cfDNA concentration	Baseline
89	NSCLC	TKI or pembrolizumab-CT or CT	KM – CPH		Longitudinal
66	Lung adenocarcino	PD-(L)1 ICI	KM – CPH – LR Mechanistic	Fragmentomics	Longitudinal

	ma & B cell carcinoma		modeling		
--	--------------------------	--	----------	--	--

595 NSCLC: non-small cell lung cancer; CPH: Cox proportional hazards (model); KM:
596 Kaplan-Meier (estimation); LR: logistic regression; NLME: nonlinear mixed effects
597 (model); ctDNA: circulating tumoral DNA; cfDNA: cell-free DNA; ICI: immune
598 checkpoint inhibitors; EGFR: epidermal growth factor receptor; ERBB2: erythroblastic
599 oncogene B 2; VEGFR: vascular EGFR; TKI: tyrosine kinase inhibitors.

600

601 **ACKNOWLEDGEMENTS**

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

605 Figures were partially created with BioRender.com.

606 The authors declare no conflict of interest.

607

608 **REFERENCES**

- 609 1. Dasari A, Morris VK, Allegra CJ, et al. ctDNA applications and integration in
610 colorectal cancer: an NCI Colon and Rectal–Anal Task Forces whitepaper. *Nat Rev Clin*
611 *Oncol.* 2020;17(12):757-770. doi:10.1038/s41571-020-0392-0
- 612 2. Heitzer E, Auinger L, Speicher MR. Cell-Free DNA and Apoptosis: How Dead Cells
613 Inform About the Living. *Trends in Molecular Medicine.* 2020;26(5):519-528.
614 doi:10.1016/j.molmed.2020.01.012
- 615 3. Hu Z, Chen H, Long Y, Li P, Gu Y. The main sources of circulating cell-free DNA:
616 Apoptosis, necrosis and active secretion. *Crit Rev Oncol Hematol.* 2021;157:103166.
617 doi:10.1016/j.critrevonc.2020.103166
- 618 4. Goebel G, Zitt M, Zitt M, Müller HM. Circulating Nucleic Acids in Plasma or Serum
619 (CNAPS) as Prognostic and Predictive Markers in Patients with Solid Neoplasias. *Dis*
620 *Markers.* 2005;21(3):105-120. doi:10.1155/2005/218759
- 621 5. Thompson JR, Menon SP. Liquid Biopsies and Cancer Immunotherapy. *Cancer J.*
622 2018;24(2):78-83. doi:10.1097/PPO.0000000000000307
- 623 6. Coto-Llerena M, Benjak A, Gallon J, et al. Circulating Cell-Free DNA Captures the
624 Intratumor Heterogeneity in Multinodular Hepatocellular Carcinoma. *JCO Precis Oncol.*
625 2022;6:e2100335. doi:10.1200/PO.21.00335
- 626 7. Lipson EJ, Velculescu VE, Pritchard TS, et al. Circulating tumor DNA analysis as a
627 real-time method for monitoring tumor burden in melanoma patients undergoing treatment
628 with immune checkpoint blockade. *J Immunother Cancer.* 2014;2(1):42. doi:10.1186/s40425-
629 014-0042-0

- 630 8. Gandara DR, Paul SM, Kowanetz M, et al. Blood-based tumor mutational burden as a
631 predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab.
632 *Nat Med.* 2018;24(9):1441-1448. doi:10.1038/s41591-018-0134-3
- 633 9. Magbanua MJM, Swigart LB, Ahmed Z, et al. Clinical significance and biology of
634 circulating tumor DNA in high-risk early-stage HER2-negative breast cancer receiving
635 neoadjuvant chemotherapy. *Cancer Cell.* 2023;41(6):1091-1102.e4.
636 doi:10.1016/j.ccell.2023.04.008
- 637 10. Luo H, Zhao Q, Wei W, et al. Circulating tumor DNA methylation profiles enable
638 early diagnosis, prognosis prediction, and screening for colorectal cancer. *Sci Transl Med.*
639 2020;12(524):eaax7533. doi:10.1126/scitranslmed.aax7533
- 640 11. Mandel P, Metais P. [Nuclear Acids In Human Blood Plasma]. *C R Seances Soc Biol*
641 *Fil.* 1948;142(3-4):241-243.
- 642 12. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer
643 patients and the effect of therapy. *Cancer Res.* 1977;37(3):646-650.
- 644 13. Stroun M, Anker P, Lyautey J, Lederrey C, Maurice PA. Isolation and characterization
645 of DNA from the plasma of cancer patients. *Eur J Cancer Clin Oncol.* 1987;23(6):707-712.
646 doi:10.1016/0277-5379(87)90266-5
- 647 14. Stroun M, Anker P, Maurice P, Lyautey J, Lederrey C, Beljanski M. Neoplastic
648 characteristics of the DNA found in the plasma of cancer patients. *Oncology.* 1989;46(5):318-
649 322. doi:10.1159/000226740
- 650 15. Thakur BK, Zhang H, Becker A, et al. Double-stranded DNA in exosomes: a novel
651 biomarker in cancer detection. *Cell Res.* 2014;24(6):766-769. doi:10.1038/cr.2014.44
- 652 16. Lui YY, Chik KW, Chiu RW, Ho CY, Lam CW, Lo YD. Predominant Hematopoietic
653 Origin of Cell-free DNA in Plasma and Serum after Sex-mismatched Bone Marrow
654 Transplantation. *Clin Chem.* 2002;48(3):421-427. doi:10.1093/clinchem/48.3.421
- 655 17. Khier S, Lohan L. Kinetics of circulating cell-free DNA for biomedical applications:
656 critical appraisal of the literature. *Future Science OA.* 2018;4(4):FSO295. doi:10.4155/fsoa-
657 2017-0140
- 658 18. Alix-Panabières C, Pantel K. Clinical Applications of Circulating Tumor Cells and
659 Circulating Tumor DNA as Liquid Biopsy. *Cancer Discov.* 2016;6(5):479-491.
660 doi:10.1158/2159-8290.CD-15-1483

- 661 19. Nikanjam M, Kato S, Kurzrock R. Liquid biopsy: current technology and clinical
662 applications. *J Hematol Oncol*. 2022;15(1):131. doi:10.1186/s13045-022-01351-y
- 663 20. Diehl F, Li M, Dressman D, et al. Detection and quantification of mutations in the
664 plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A*. 2005;102(45):16368-
665 16373. doi:10.1073/pnas.0507904102
- 666 21. Bettgowda C, Sausen M, Leary RJ, et al. Detection of Circulating Tumor DNA in
667 Early- and Late-Stage Human Malignancies. *Sci Transl Med*. 2014;6(224):224ra24.
668 doi:10.1126/scitranslmed.3007094
- 669 22. Shapiro B, Chakrabarty M, Cohn EM, Leon SA. Determination of circulating DNA
670 levels in patients with benign or malignant gastrointestinal disease. *Cancer*.
671 1983;51(11):2116-2120. doi:10.1002/1097-0142(19830601)51:11<2116::aid-
672 cncr2820511127>3.0.co;2-s
- 673 23. Raptis L, Menard HA. Quantitation and characterization of plasma DNA in normals
674 and patients with systemic lupus erythematosus. *J Clin Invest*. 1980;66(6):1391-1399.
- 675 24. Lo YM, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM. Rapid clearance of fetal
676 DNA from maternal plasma. *Am J Hum Genet*. 1999;64(1):218-224.
- 677 25. Breitbach S, Tug S, Simon P. Circulating Cell-Free DNA. *Sports Med*.
678 2012;42(7):565-586. doi:10.2165/11631380-000000000-00000
- 679 26. Ivanov M, Baranova A, Butler T, Spellman P, Mileyko V. Non-random fragmentation
680 patterns in circulating cell-free DNA reflect epigenetic regulation. *BMC Genom*.
681 2015;16(Suppl 13):S1. doi:10.1186/1471-2164-16-S13-S1
- 682 27. Jiang P, Chan CWM, Chan KCA, et al. Lengthening and shortening of plasma DNA in
683 hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A*. 2015;112(11):E1317-E1325.
684 doi:10.1073/pnas.1500076112
- 685 28. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an
686 In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*. 2016;164(1):57-68.
687 doi:10.1016/j.cell.2015.11.050
- 688 29. Lo YMD, Chan KCA, Sun H, et al. Maternal Plasma DNA Sequencing Reveals the
689 Genome-Wide Genetic and Mutational Profile of the Fetus. *Sci Transl Med*. 2010;2(61).
690 doi:10.1126/scitranslmed.3001720

- 691 30. Jahr S, Hentze H, Englisch S, et al. DNA Fragments in the Blood Plasma of Cancer
692 Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells1.
693 *Cancer Res.* 2001;61(4):1659-1665.
- 694 31. Mouliere F, Chandrananda D, Piskorz AM, et al. Enhanced detection of circulating
695 tumor DNA by fragment size analysis. *Sci Transl Med.* 2018;10(466):eaat4921.
696 doi:10.1126/scitranslmed.aat4921
- 697 32. Jiang P, Sun K, Peng W, et al. Plasma DNA End-Motif Profiling as a Fragmentomic
698 Marker in Cancer, Pregnancy, and Transplantation. *Cancer Discov.* 2020;10(5):664-673.
699 doi:10.1158/2159-8290.CD-19-0622
- 700 33. Jiang P, Xie T, Ding SC, et al. Detection and characterization of jagged ends of
701 double-stranded DNA in plasma. *Genome Res.* 2020;30(8):1144-1153.
702 doi:10.1101/gr.261396.120
- 703 34. Avgeris M, Marmarinos A, Gourgiotis D, Scorilas A. Jagged Ends of Cell-Free DNA:
704 Rebranding Fragmentomics in Modern Liquid Biopsy Diagnostics. *Clinical Chem.*
705 2021;67(4):576-578. doi:10.1093/clinchem/hvab036
- 706 35. Guo W, Chen X, Liu R, et al. Sensitive detection of stage I lung adenocarcinoma using
707 plasma cell-free DNA breakpoint motif profiling. *eBioMedicine.* 2022;81:104131.
708 doi:10.1016/j.ebiom.2022.104131
- 709 36. Wadden J, Ravi K, John V, Babila CM, Koschmann C. Cell-Free Tumor DNA (cf-
710 tDNA) Liquid Biopsy: Current Methods and Use in Brain Tumor Immunotherapy. *Front*
711 *Immunol.* 2022;13:882452. doi:10.3389/fimmu.2022.882452
- 712 37. Douillard JY, Ostoros G, Cobo M, et al. First-line gefitinib in Caucasian EGFR
713 mutation-positive NSCLC patients: a phase-IV, open-label, single-arm study. *Br J Cancer.*
714 2014;110(1):55-62. doi:10.1038/bjc.2013.721
- 715 38. Del Re M, Crucitta S, Paolieri F, et al. The amount of DNA combined with TP53
716 mutations in liquid biopsy is associated with clinical outcome of renal cancer patients treated
717 with immunotherapy and VEGFR-TKIs. *J Transl Med.* 2022;20(1):371. doi:10.1186/s12967-
718 022-03557-7
- 719 39. Siravegna G, Mussolin B, Buscarino M, et al. Monitoring clonal evolution and
720 resistance to EGFR blockade in the blood of metastatic colorectal cancer patients. *Nat Med.*
721 2015;21(7):795-801. doi:10.1038/nm.3870

- 722 40. Nitschke C, Markmann B, Walter P, et al. Peripheral and Portal Venous KRAS ctDNA
723 Detection as Independent Prognostic Markers of Early Tumor Recurrence in Pancreatic
724 Ductal Adenocarcinoma. *Clinical Chemistry*. 2023;69(3):295-307.
725 doi:10.1093/clinchem/hvac214
- 726 41. Liu L, Chen X, Petinrin OO, et al. Machine Learning Protocols in Early Cancer
727 Detection Based on Liquid Biopsy: A Survey. *Life (Basel)*. 2021;11(7):638.
728 doi:10.3390/life11070638
- 729 42. Raman L, Van der Linden M, Van der Eecken K, et al. Shallow whole-genome
730 sequencing of plasma cell-free DNA accurately differentiates small from non-small cell lung
731 carcinoma. *Genome Med*. 2020;12:35. doi:10.1186/s13073-020-00735-4
- 732 43. Liang N, Li B, Jia Z, et al. Ultrasensitive detection of circulating tumour DNA via
733 deep methylation sequencing aided by machine learning. *Nat Biomed Eng*. 2021;5(6):586-
734 599. doi:10.1038/s41551-021-00746-5
- 735 44. Lin LH, Chang KW, Kao SY, Cheng HW, Liu CJ. Increased Plasma Circulating Cell-
736 Free DNA Could Be a Potential Marker for Oral Cancer. *Int J Mol Sci*. 2018;19(11):3303.
737 doi:10.3390/ijms19113303
- 738 45. Guler GD, Ning Y, Ku CJ, et al. Detection of early stage pancreatic cancer using 5-
739 hydroxymethylcytosine signatures in circulating cell free DNA. *Nat Commun*.
740 2020;11(1):5270. doi:10.1038/s41467-020-18965-w
- 741 46. Song CX, Yin S, Ma L, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA
742 provide information about tumor types and stages. *Cell Res*. 2017;27(10):1231-1242.
743 doi:10.1038/cr.2017.106
- 744 47. Wang Y, Zheng J, Li Z, et al. Development of a novel liquid biopsy test to diagnose
745 and locate gastrointestinal cancers. *J Clin Oncol*. 2020;38(15_suppl):1557-1557.
746 doi:10.1200/JCO.2020.38.15_suppl.1557
- 747 48. Kothen-Hill ST, Zviran A, Schulman R, et al. Deep learning mutation prediction
748 enables early stage lung cancer detection in liquid biopsy. *International Conference on*
749 *Learning Representations*. February 15, 2018.
- 750 49. Moser T, Kühberger S, Lazzeri I, Vlachos G, Heitzer E. Bridging biological cfDNA
751 features and machine learning approaches. *Trends Genet*. 2023;39(4):285-307.
752 doi:10.1016/j.tig.2023.01.004

- 753 50. Chabon JJ, Hamilton EG, Kurtz DM, et al. Integrating genomic features for
754 noninvasive early lung cancer detection. *Nature*. 2020;580(7802):245-251.
755 doi:10.1038/s41586-020-2140-0
- 756 51. Chen L, Abou-Alfa GK, Zheng B, et al. Genome-scale profiling of circulating cell-free
757 DNA signatures for early detection of hepatocellular carcinoma in cirrhotic patients. *Cell Res*.
758 2021;31(5):589-592. doi:10.1038/s41422-020-00457-7
- 759 52. Ma X, Chen Y, Tang W, et al. Multi-dimensional fragmentomic assay for
760 ultrasensitive early detection of colorectal advanced adenoma and adenocarcinoma. *J*
761 *Hematol Oncol*. 2021;14:175. doi:10.1186/s13045-021-01189-w
- 762 53. Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in
763 patients with cancer. *Nature*. 2019;570(7761):385-389. doi:10.1038/s41586-019-1272-6
- 764 54. Mathios D, Johansen JS, Cristiano S, et al. Detection and characterization of lung
765 cancer using cell-free DNA fragmentomes. *Nat Commun*. 2021;12(1):5060.
766 doi:10.1038/s41467-021-24994-w
- 767 55. Renaud G, Nørgaard M, Lindberg J, et al. Unsupervised detection of fragment length
768 signatures of circulating tumor DNA using non-negative matrix factorization. *eLife*.
769 11:e71569. doi:10.7554/eLife.71569
- 770 56. Avanzini S, Kurtz DM, Chabon JJ, et al. A mathematical model of ctDNA shedding
771 predicts tumor detection size. *Sci Adv*. 2020;6(50):eabc4308. doi:10.1126/sciadv.abc4308
- 772 57. Woodhouse R, Li M, Hughes J, et al. Clinical and analytical validation of
773 FoundationOne Liquid CDx, a novel 324-Gene cfDNA-based comprehensive genomic
774 profiling assay for cancers of solid tumor origin. *PLoS One*. 2020;15(9):e0237802.
775 doi:10.1371/journal.pone.0237802
- 776 58. Pascual J, Attard G, Bidard FC, et al. ESMO recommendations on the use of
777 circulating tumour DNA assays for patients with cancer: a report from the ESMO Precision
778 Medicine Working Group. *Ann Oncol*. 2022;33(8):750-768.
779 doi:10.1016/j.annonc.2022.05.520
- 780 59. LeeVan E, Pinsky P. Predictive Performance of Cell-Free Nucleic Acid-Based Multi-
781 Cancer Early Detection Tests: A Systematic Review. *Clinical Chem*. Published online
782 October 4, 2023:hvad134. doi:10.1093/clinchem/hvad134

- 783 60. Lapin M, Oltedal S, Tjensvoll K, et al. Fragment size and level of cell-free DNA
784 provide prognostic information in patients with advanced pancreatic cancer. *J Transl Med.*
785 2018;16:300. doi:10.1186/s12967-018-1677-2
- 786 61. Moding EJ, Liu Y, Nabet BY, et al. Circulating Tumor DNA Dynamics Predict
787 Benefit from Consolidation Immunotherapy in Locally Advanced Non-Small Cell Lung
788 Cancer. *Nat Cancer.* 2020;1(2):176-183. doi:10.1038/s43018-019-0011-0
- 789 62. Powles T, Assaf ZJ, Davarpanah N, et al. ctDNA guiding adjuvant immunotherapy in
790 urothelial carcinoma. *Nature.* 2021;595(7867):432-437. doi:10.1038/s41586-021-03642-9
- 791 63. Ribba B, Roller A, Helms HJ, Stern M, Bleul C. Circulating tumor DNA:
792 Opportunities and challenges for pharmacometric approaches. *Front Pharmacol.* 2023;13.
793 Accessed October 26, 2023. <https://www.frontiersin.org/articles/10.3389/fphar.2022.1058220>
- 794 64. Janssen JM, Verheijen RB, van Duijl TT, et al. Longitudinal nonlinear mixed effects
795 modeling of EGFR mutations in ctDNA as predictor of disease progression in treatment of
796 EGFR- mutant non- small cell lung cancer. *Clin Transl Sci.* 2022;15(8):1916-1925.
797 doi:10.1111/cts.13300
- 798 65. Khan KH, Cunningham D, Werner B, et al. Longitudinal liquid biopsy and
799 mathematical modelling of clonal evolution forecast waiting time to treatment failure in the
800 PROSPECT-C phase II colorectal cancer clinical trial. *Cancer Discov.* 2018;8(10):1270-1285.
801 doi:10.1158/2159-8290.CD-17-0891
- 802 66. Esfahani MS, Hamilton EG, Mehrmohamadi M, et al. Inferring gene expression from
803 cell-free DNA fragmentation profiles. *Nat Biotechnol.* 2022;40(4):585-597.
804 doi:10.1038/s41587-022-01222-4
- 805 67. Assaf ZJF, Zou W, Fine AD, et al. A longitudinal circulating tumor DNA-based model
806 associated with survival in metastatic non-small-cell lung cancer. *Nat Med.* Published online
807 March 16, 2023:1-10. doi:10.1038/s41591-023-02226-6
- 808 68. Zou W, Yaung SJ, Fuhlbrück F, et al. ctDNA Predicts Overall Survival in Patients
809 With NSCLC Treated With PD-L1 Blockade or With Chemotherapy. *JCO Precis Oncol.*
810 2021;(5):827-838. doi:10.1200/PO.21.00057
- 811 69. Ciccolini J, Barbolosi D, André N, Barlesi F, Benzekry S. Mechanistic Learning for
812 Combinatorial Strategies With Immuno-oncology Drugs: Can Model-Informed Designs Help
813 Investigators? *JCO Precis Oncol.* 2020;(4):486-491. doi:10.1200/PO.19.00381

- 814 70. Nagasaka M, Uddin MH, Al-Hallak MN, et al. Liquid biopsy for therapy monitoring
815 in early-stage non-small cell lung cancer. *Mol Cancer*. 2021;20:82. doi:10.1186/s12943-021-
816 01371-1
- 817 71. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable
818 cancers with a multi-analyte blood test. *Science*. 2018;359(6378):926-930.
819 doi:10.1126/science.aar3247
- 820 72. Jamal-Hanjani M, Wilson GA, Horswell S, et al. Detection of ubiquitous and
821 heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung
822 cancer. *Annals of Oncology*. 2016;27(5):862-867. doi:10.1093/annonc/mdw037
- 823 73. Wu X, Zhang Y, Hu T, et al. A novel cell-free DNA methylation-based model
824 improves the early detection of colorectal cancer. *Mol Oncol*. 2021;15(10):2702.
825 doi:10.1002/1878-0261.12942
- 826 74. Chemi F, Pearce SP, Clipson A, et al. cfDNA methylome profiling for detection and
827 subtyping of small cell lung cancers. *Nat Cancer*. 2022;3(10):1260-1270.
828 doi:10.1038/s43018-022-00415-9
- 829 75. Weiss GJ, Beck J, Braun DP, et al. Tumor Cell-Free DNA Copy Number Instability
830 Predicts Therapeutic Response to Immunotherapy. *Clin Cancer Res*. 2017;23(17):5074-5081.
831 doi:10.1158/1078-0432.CCR-17-0231
- 832 76. Si H, Kuziora M, Quinn KJ, et al. A Blood-based Assay for Assessment of Tumor
833 Mutational Burden in First-line Metastatic NSCLC Treatment: Results from the MYSTIC
834 Study. *Clin Cancer Res*. 2021;27(6):1631-1640. doi:10.1158/1078-0432.CCR-20-3771
- 835 77. Matsumae T, Kodama T, Myojin Y, et al. Circulating Cell-Free DNA Profiling
836 Predicts the Therapeutic Outcome in Advanced Hepatocellular Carcinoma Patients Treated
837 with Combination Immunotherapy. *Cancers*. 2022;14(14):3367.
838 doi:10.3390/cancers14143367
- 839 78. Khagi Y, Goodman AM, Daniels GA, et al. Hypermutated Circulating Tumor DNA:
840 Correlation with Response to Checkpoint Inhibitor-Based Immunotherapy. *Clinical Cancer*
841 *Research*. 2017;23(19):5729-5736. doi:10.1158/1078-0432.CCR-17-1439
- 842 79. Luo H, Zhao Q, Wei W, et al. Circulating tumor DNA methylation profiles enable
843 early diagnosis, prognosis prediction, and screening for colorectal cancer. *Sci Transl Med*.
844 2020;12(524):eaax7533. doi:10.1126/scitranslmed.aax7533

- 845 80. Lee JH, Long GV, Boyd S, et al. Circulating tumour DNA predicts response to anti-
846 PD1 antibodies in metastatic melanoma. *Ann Oncol.* 2017;28(5):1130-1136.
847 doi:10.1093/annonc/mdx026
- 848 81. Cabel L, Riva F, Servois V, et al. Circulating tumor DNA changes for early
849 monitoring of anti-PD1 immunotherapy: a proof-of-concept study. *Ann Oncol.*
850 2017;28(8):1996-2001. doi:10.1093/annonc/mdx212
- 851 82. Ricciuti B, Jones G, Severgnini M, et al. Early plasma circulating tumor DNA
852 (ctDNA) changes predict response to first-line pembrolizumab-based therapy in non-small
853 cell lung cancer (NSCLC). *J Immunother Cancer.* 2021;9(3):e001504. doi:10.1136/jitc-2020-
854 001504
- 855 83. Goldberg SB, Narayan A, Kole AJ, et al. Early Assessment of Lung Cancer
856 Immunotherapy Response via Circulating Tumor DNA. *Clin Cancer Res.* 2018;24(8):1872-
857 1880. doi:10.1158/1078-0432.CCR-17-1341
- 858 84. Anagnostou V, Ho C, Nicholas G, et al. ctDNA response after pembrolizumab in non-
859 small cell lung cancer: phase 2 adaptive trial results. *Nat Med.* 2023;29(10):2559-2569.
860 doi:10.1038/s41591-023-02598-9
- 861 85. Herbreteau G, Vallée A, Knol AC, et al. Circulating Tumor DNA Early Kinetics
862 Predict Response of Metastatic Melanoma to Anti-PD1 Immunotherapy: Validation Study.
863 *Cancers (Basel).* 2021;13(8):1826. doi:10.3390/cancers13081826
- 864 86. Zhang Q, Luo J, Wu S, et al. Prognostic and Predictive Impact of Circulating Tumor
865 DNA in Patients with Advanced Cancers Treated with Immune Checkpoint Blockade. *Cancer*
866 *Discov.* 2020;10(12):1842-1853. doi:10.1158/2159-8290.CD-20-0047
- 867 87. Nabet BY, Esfahani MS, Moding EJ, et al. Noninvasive Early Identification of
868 Therapeutic Benefit from Immune Checkpoint Inhibition. *Cell.* 2020;183(2):363-376.e13.
869 doi:10.1016/j.cell.2020.09.001
- 870 88. Bratman SV, Yang SYC, Iafolla MAJ, et al. Personalized circulating tumor DNA
871 analysis as a predictive biomarker in solid tumor patients treated with pembrolizumab. *Nat*
872 *Cancer.* 2020;1(9):873-881. doi:10.1038/s43018-020-0096-5
- 873 89. Gristina V, Barraco N, La Mantia M, et al. Clinical Potential of Circulating Cell-Free
874 DNA (cfDNA) for Longitudinally Monitoring Clinical Outcomes in the First-Line Setting of
875 Non-Small-Cell Lung Cancer (NSCLC): A Real-World Prospective Study. *Cancers (Basel).*
876 2022;14(23):6013. doi:10.3390/cancers14236013

