Cell Genomics

CellPress
OPEN ACCESS

Review

Circulating DNA fragmentomics and cancer screening

A.R. Thierry[1,*]
[1]IRCM, Institut de Recherche en Cancérologie de Montpellier, INSERM U1194, Université de Montpellier, and ICM, Institut régional du Cancer de Montpellier, Montpellier 34298, France
*Correspondence: alain.thierry@inserm.fr
https://doi.org/10.1016/j.xgen.2022.100242

SUMMARY

The high fragmentation of nuclear circulating DNA (cirDNA) relies on chromatin organization and protection or packaging within mononucleosomes, the smallest and the most stabilized structure in the bloodstream. The detection of differing size patterns, termed fragmentomics, exploits information about the nucleosomal packing of DNA. Fragmentomics not only implies size pattern characterization but also considers the positioning and occupancy of nucleosomes, which result in cirDNA fragments being protected and persisting in the circulation. Fragmentomics can determine tissue of origin and distinguish cancer-derived cirDNA. The screening power of fragmentomics has been considerably strengthened in the omics era, as shown in the ongoing development of sophisticated technologies assisted by machine learning. Fragmentomics can thus be regarded as a strategy for characterizing cancer within individuals and offers an alternative or a synergistic supplement to mutation searches, methylation, or nucleosome positioning. As such, it offers potential for improving diagnostics and cancer screening.

INTRODUCTION

Fragmentomics is a new field studying circulating DNA (cirDNA), especially with regard to cancer screening. Ivanov et al.[1] first used the term to refer to the study of cirDNA fragment size profiles. The term fragmentomics was initially suggested by Zamyatnin[2] referring to the study of the structure and functions of a set of molecular fragments; calling the whole set of biomolecule fragments the ''fragmentome.'' For some proteins, hundreds of fragments have been studied in detail, identifying functions related to specific protein fragment structures. With respect to the application of cirDNAs to oncological studies, cirDNA sequences needed to be investigated for potential applications as a companion test, particularly for targeted therapy.[3–8] Subsequently, the detection and quantification of mutant cirDNA led to its use in the detection of residual disease, clonal heterogeneity/treatment resistance, and surveillance for tumor recurrence.[5,9,10]

When cirDNA is detected in the blood, the DNA molecule is highly fragmented[11–14] with a characteristic size profile.[15–18] This offers the potential to determine cirDNA tissue-of-origin, and to contribute to the development of a pan-cancer screening or early detection test. cirDNA-associated fragmentomics can involve non-sequencing technologies, sequencing, or a mixture of both methods. The characterization of cancer cirDNA topology has benefited from work performed on cirDNA derived from healthy subjects (especially during the development of a non-invasive prenatal test [NIPT]), notably, Denis Lo's group's studies of the size distribution of fetal cirDNA, initially performed using qPCR,[19,20] then with whole-genome sequencing.[15,21–27] Similarly, non-omic (qPCR, AFM, capillary electrophoresis, etc.)[11–13,28] and subsequently omic methods[14,16,17,29–32] identified differences in the fragment size distribution of cirDNA deriving from healthy individuals and cancer patients.

cirDNA SHOWS A CHROMATIN ORGANIZATION PATTERN

Non-omics observations

A fraction of cirDNA is composed of extracellular DNA released during cell death, in particular during necrosis, phagocytosis, or apoptosis.[33–36] DNA fragments longer than 10,000 base pairs (bp) are likely to originate from necrotic cells, whereas DNA fragments shorter than 1,000 bp, particularly of 180 bp or multiples of this size, are reminiscent of the oligonucleosomal DNA ladder observed in apoptotic cells.[37–39] DNA nucleosomal fragmentation is a hallmark of apoptosis. Thus, analyses of the size of the DNA fragments may enable the source of cirDNA to be discerned.

Early studies of cirDNA utilized gel electrophoresis,[38] and described cirDNA as consisting of fragments of approximatively 100–500 bp.[37,38,40] Lo's team[19] analyzed the size distributions of maternal and fetal DNA in maternal plasma using a qPCR assay with primer sets that amplify sequences from 105 to 798 bp, enabling more precise size profiling. They suggested that ''most of the ctDNA molecules were in the range of 145–201 bp,''[19] which is approximately the size of a nucleosome. However, a nested qPCR system method (N-qPCR) showed[11–13,28] that a significant proportion of cirDNA is smaller than 100 bp and that targeting 60–100 bp sequences optimizes cirDNA quantification.[12,13] This observation enhanced the recovery of ctDNA, for clinical implementation.[41–43] Note that the cirDNA fragment distribution may also be determined with methods, such as atomic force microscopy,[41,44] microfluidic

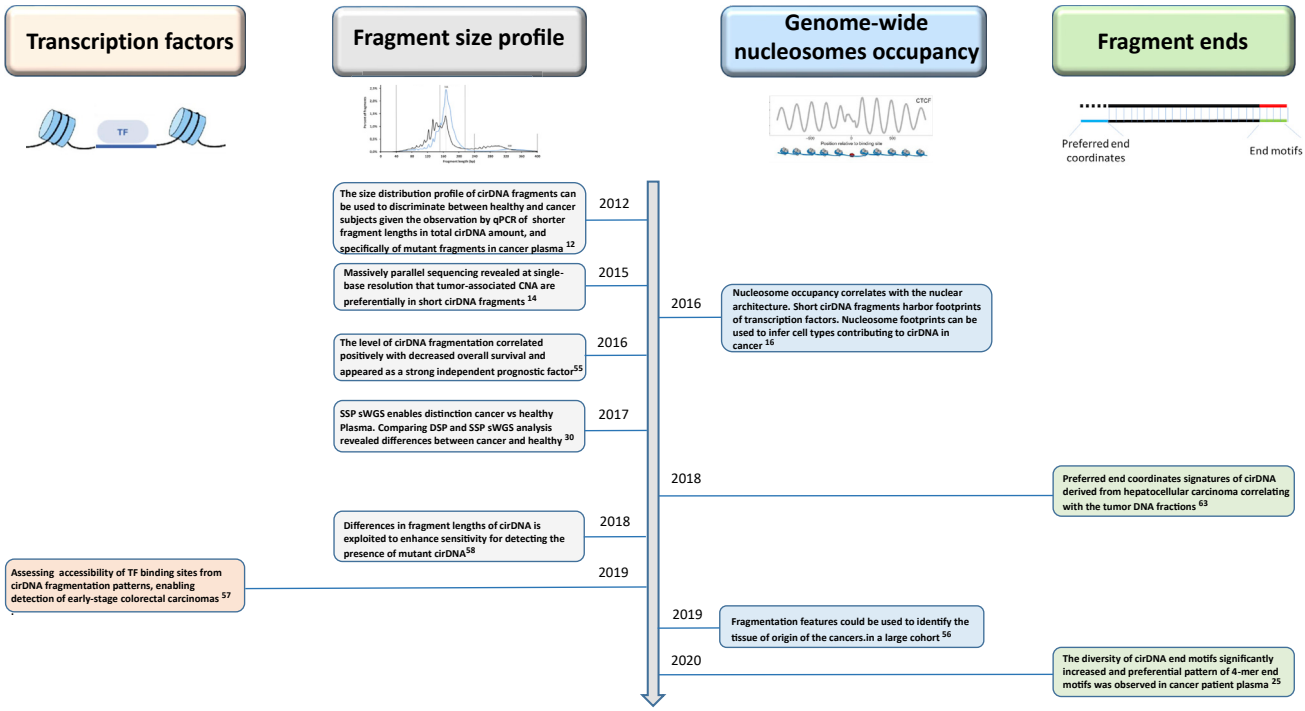

Figure 1. Milestone studies revealing quantitative differences in the cirDNA size distribution between cancer patients and healthy subjects

CNA, copy number alterations; qPCR, quantitative polymerase chain reaction; sWGS, shallow whole-genome sequencing; SSP, single-strand DNA library preparation; DSP, double-strand DNA library preparation; TF, transcription factor.

single-molecule spectroscopy,[45] or viscoelastic lift force capillary technology.[46] This variety of options may offer further possibilities for comparing cirDNA plasma extracts or origins.

The mononucleosomes as the preponderant structures associating cirDNA

Depending on tissue-of-origin and physiological or pathological conditions (including apoptosis, necrosis, or NETosis), only a small number of the structures are stable enough to be detectable in blood as cirDNA. Those structures are: mono-nucleosomes, the smallest and most stable structure in the blood stream; tightly packed long double-stranded DNA (dsDNA); and, to a lesser extent, di-nucleosomes. Minor structures associated with cirDNA include: short-sized (<70 bp) transcription factor-binding dsDNA; long- and short-sized DNA-associated micro-particles; apoptotic particles; short-sized lipo-proteo-nucleic complexes; and, cells or cell-part associations.

Two opposing theories of cirDNA mechanisms of release have been proposed. One points to the prominence of mononucleosomes as an indication that apoptosis is the principal release mechanism.[38,47] The alternative theory proposes that short-sized nucleosomal structures ensue from the progressive nuclease degradation of longer cirDNA, which derive from micro-particle-containing DNA, necrosis, phagocytosis, or active release resulting from lymphocytes or neutrophils (NETs).[48,49] The data to date on cirDNA fragment size profiling suggest a dynamic process that relies on both the rapid rate of chromatin decondensation and inter-nucleosomal DNA fragmentation, and the slower rate of intra-nucleosomal fragmentation by nuclease activities. While this full process appears highly homogeneous in healthy individuals, it varies in cancer[14,30,50] and in pregnancy,[15,19,27] as discussed later.

Despite the range of topological information provided by whole-genome sequencing (WGS)-based sizing, this analytical approach can only detect cirDNA fragment lengths below ~1,000 bp. Using qPCR or capillary electrophoresis, studies[50–53] report a significant fraction (10%–25%) of fragments in the 1,000–8,000 bp range in healthy individuals. By combining double-stranded DNA library preparation-based sequencing (DSP-S), single-stranded DNA library preparation-based sequencing (SSP-S), and N-qPCR analysis, cirDNA associated in mono-nucleosomes, di-nucleosomes, and chromatin of higher molecular size (>1,000 bp) was estimated as 67.5%–80%, 9.4%–11.5%, and 8.5%–21.0%, respectively.

Omics observations

While various analytical non-omics methods have revealed fragment sizes corresponding to mono-nucleosomes[47,54]; subsequently, sequencing allowed for in-depth observations that greatly enhanced fragmentomics. Cancer screening relies on precise characterization, high reproducibility, and stringent standardization of control parameters. WGS has increased the analytical power and resolution of fragmentomics. Several milestone reports[12,14,16,25,55–59] illustrated in Figure 1 have paved the way for developing a cirDNA-based cancer screening test, further empowered by bioinformatics advances.[17,60–64]

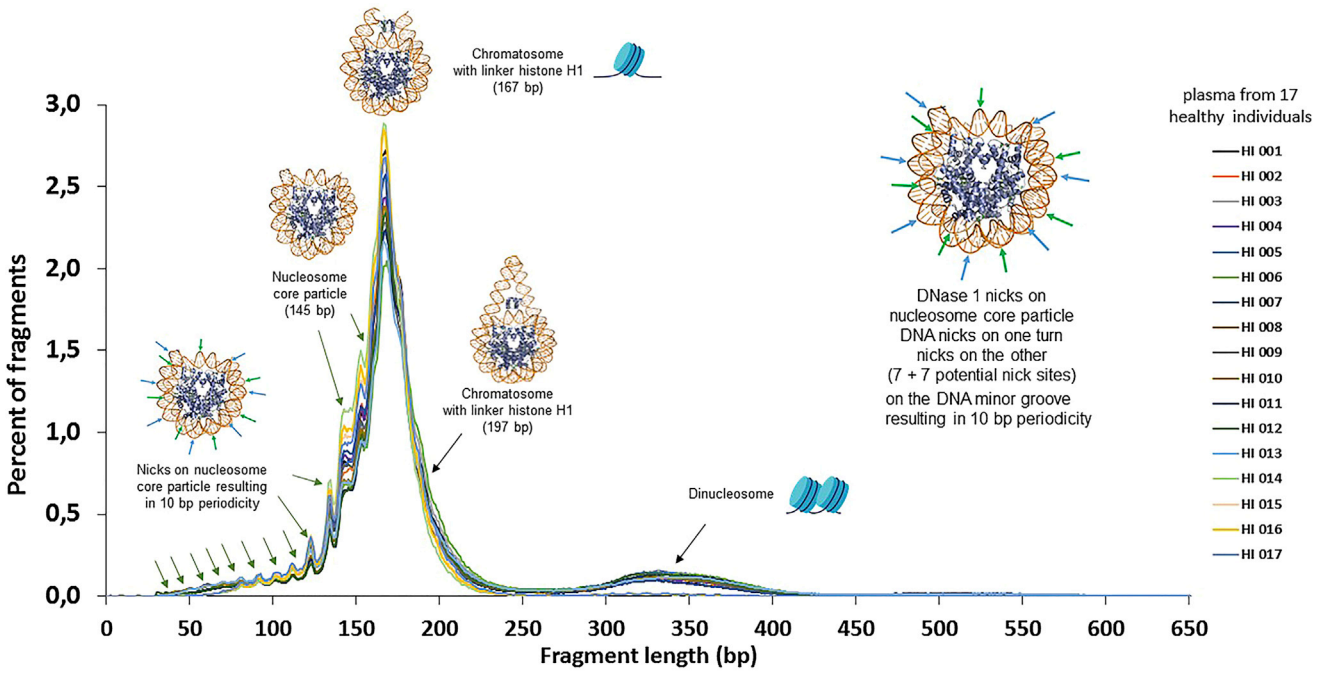

Figure 2. Consistency of healthy individual cirDNA fragment size profiles

cirDNA plasma extracts of 17 healthy individuals were subjected to sWGS from DSP. The frequency of each read is plotted and expressed in percentages for each plasma. The 17 size profiles were generated from four different sets of plasma DNA (n = 7, 4, 3, and 3) at four different times (n = 7, 4, 3, and 3) and with two different platforms and different bioinformatics systems (n = 3), and are represented as curves of various superimposed colors. Overall, this results in a major population associated with chromatosomes/mononucleosomes (large peak between 141 and 181 bp), and a minor population associated with dinucleosomes (small peak between 301 and 381 bp). No reads above the background are detectable between 661 bp and the limit of ~1,000 bp (the WGS readout limit). Subpeaks with a periodicity of ~10 bp detectable between 40 and 177 bp are the result of exposition to nuclease activity at the DNA molecule's minor groove at the surface of the chromatosome/mononucleosome structures. Arrows indicate the position of the minor groove at which nicks may occur, and also offer a schematic view of the nucleosome structures associated with cirDNA, from which cirDNA fragment length can be inferred. Also highlighted are the structures most frequently associated with cirDNA: the chromatosome with a 166 bp DNA fragment (~3%) and the nucleosome core particle devoid of H1 containing 147–160 bp fragments (0.8% to ~1.5%) are the structures which most frequently associate cirDNA.[17] The figure is adapted from the data from a recent report.[17] Images of the crystal structure of chromatosome and nucleosome, shown at 3.5 Å resolution, were obtained from the NIPDB databank (4QLC and 5ONW, respectively). NIPDB, Nucleic Acid-Protein Interaction Database, https://npidb.belozersky.msu.ru.

cirDNA fragment size profile characteristics

Nuclear cirDNA fragmentation is dependent on chromatin organization.[1,65] cirDNA sizes typically correspond to the fragment sizes of mostly mono-nucleosomes, a low fraction of di-nucleosomes, with traces of tri-nucleosomes. cirDNA exhibits a ~10 bp periodicity footprint detectable down to 31 bp[1,15–17,66] (Figure 2). This is indicative of nucleosome-derived degradation, likely ascribed to cleavage, occurring in nucleotides that remain accessible at every helical turn where DNA wraps around the core and is farthest from the histone core surface.[1,15,16] Yu et al.[15] first revealed that the size profile of maternally derived cirDNA exhibits a ~10 bp periodicity below 143 bp. Overall, the majority of detectable cirDNA in the blood exhibits a footprint resulting from the association between DNA and the nucleosome structure, which provides stability for cirDNA.[1,15,16,18,24,30,50] While a small proportion of cirDNA fragments are associated with di-nucleosomes they are identified by the ~10 bp periodicity footprint within the 300–360 bp range, reflecting the conformation of DNA wrapped about a di-nucleosome.

The detected cirDNA fragments exhibiting the highest frequency are derived from trimmed mono-nucleosome, also called the chromatosome. The chromatosome, theoretically condensing 166 bp length DNA, comprises a histone octamer ((H2A-H2B)2 (H3-H4)2) and the histone monomer linker H1 (Figure 2). Since there are only 2%–3% of cirDNA with 166 bp, the cirDNA molecule appears substantially nicked, revealing nuclease activity on the nucleosomal structure.[17] Consequently, the nucleosome core particle structure, which is lacking the histone monomer linker H1 and condenses merely 147 bp DNA (the mononucleosome), is also present in a high proportion of the cirDNA structural forms (Figure 2). sWGS (shallow or low pass WGS) identifies cirDNA fragments where a single nick is present in both strands within the same area, resulting in a nucleosomal footprints of smaller cirDNA-sized fragments down to 30–40 bp.

The majority of cirDNA molecules shorter than non-degraded nucleosomes (<167 bp) are derived either from double-stranded nicks at the extremity of the nucleosome or at one of the 14 points at the nucleosome's surface where DNA is exposed on the minor groove DNA structure (Figure 2). It is also possible that double-stranded breaks can occur at two separate positions of the nucleosome surface (Figure 2). The rarity of observations of cirDNA fragments shorter than 41 bp is due to limitations in technical

sensitivity. The smaller the size of the DNA molecule, the weaker the forces binding it to the nucleosome, and the lower the frequency of short fragments. This may be caused by DNA peeling off the nucleosome edge,[17] followed by rapid degradation.

Size profile determined from DSP-S vs. SSP-S

The conventional DSP-S method detects double-stranded breaks in the DNA molecule. In contrast, SSP-S size profiling reveals nicks on both strands. Since SSP implies DNA denaturation and strand separation prior to sequencing, it can result in an unrealistic quantification of single-stranded cirDNA fragment length. Thus, data collected via DSP-S and SSP-S sizing offer clues as to the position of the cirDNA molecule on proteo-nucleo complexes, as well as the means by which a DNA molecule is protected from degradation within the blood.

SSP-S revealed a higher proportion of shorter cirDNA fragments (30–145 bp range) compared with DSP-S. This is because of the fact that when a DNA molecule has only a single nick on one of its two strands it remains wrapped around the nucleosome and, consequently, does not peel away.[17,30] Detailed analysis of cirDNA sizing by SSP-S also showed an ~10 nucleotide (nt) periodicity footprint detectable to 31 nt within the 31–166 nt size range. In this case, the two strands exposed at the nucleosome surface are shifted by 3 bp, with a stagger of 3′.[16,17,30,50] Therefore, the size of molecules detected by DSP-S and SSP-S revealed a shift of 3 bp; this observation confirms both the ~10 bp periodicity and the fact that cirDNA are wrapped around the nucleosome.[15,16]

SSP-S enables the detection of the shorter cirDNA fragments resulting from nicks dynamically occurring in blood in one or both DNA strands packed in the mono-nucleosome. While overall size distribution values obtained from DSP-S differ slightly from data obtained by N-qPCR, those determined by both SSP-S and N-qPCR agree because the denaturation step employs single-stranded DNA as an initial template[30] in both methods. Nevertheless, given that the same peak of ~166 bp was observed in both SSP-S and DSP-S, we hypothesize that a substantial, but small, fraction (2%–3%) of the cirDNA fragments within this size range are free of nicks in at least one strand.

Homogeneity of healthy individual cirDNA size profile

The reproducibility of the cirDNA size profile from healthy plasmas is evident from superimposed profile curves (Figure 2).[17] In both the major peak and all subpeaks, the maximal frequency and size varied by no more than 2%–20% and 1 or 2 bp, respectively (Figure 2).[17] We therefore postulate that (1) in healthy subjects, the post cell release DNA degradation dynamic is consistent; (2) this results in consistent specific chromatin structures in the blood, mainly mono-nucleosome/chromatosomes, and a low proportion of di-nucleosomes; (3) the nicks/breaks on those structures rely on dsDNA molecule position/packaging, and on nuclease accessibility; (4) this homogeneity may be due to the homogeneity of the cirDNA cell-of-origin, especially with respect to immune cells, as in healthy individuals[17]; and (5) jagged cirDNA on nucleosomal structures are preponderant,[50] with their proportion on mono-nucleosome structures depending on their length. Length impinges on the electrostatic interaction between DNA and nucleosomes, which impacts cirDNA fragment maintenance.

Transcription factor association pattern

Shendure's group[16] provided evidence that, besides binding to DNA wrapped around a histone octamer, cirDNA may also bind to transcription factors, typically proteins that protect DNA from nucleases in the blood.[16,62] Maps of genome-wide *in vivo* nucleosome occupancy revealed the presence of shorter (35–80 bp) fragments associated with cleavage adjacent to transcription factor-binding sites, and harboring DNA motifs associated with specific transcription factors.[16]

DIFFERENCES BETWEEN CANCER AND HEALTHY INDIVIDUALS REVEALED BY cirDNA FRAGMENTOMICS

The comparison of cirDNA fragment lengths from both healthy people and individuals with cancer was first carried out with transmission electron microscopy.[36] Because of contamination from intracellular DNA deriving from white blood cells in the DNA plasma extract (more than 50% of fragments were over 1,000 bp), it was not possible to perform a quantitative assessment of the differences in cirDNA between individuals with cancer and healthy subjects.[36] This was also the case in other studies.[4,7,8,38,67,68]

Similarly, cirDNA mutation detection has been performed as a companion test with a flow cytometry-assisted qPCR method. In this test, small fragments from cancer samples are enriched, but only in quantities ranging from 0.01% to 1.7% of the total cirDNA molecules.[6] Although these studies could not provide a quantitative assessment or a defined size range difference, they were the first to suggest (albeit without adequate demonstration) that cancer-derived fragments are generally shorter than control-derived fragments. qPCR was applied to differentially quantify shorter (100–150 bp) or longer (200–380 bp) length cirDNA fragments relative to the DNA contained in a mono-nucleosome. In studies of two different cancer types, Ellinger and co-workers reported both a lower[69] and a higher[70] cirDNA integrity compared with that of healthy subjects. In light of these findings, divergent conclusions have been made, with cirDNA integrity being described as lower,[11,13,69,71] equivalent,[72,73] or higher[70,74,75] in individuals with cancer. These discrepancies are mainly due to reported confounding pre-analytical (plasma isolation, volume sampling, freezing, etc.) and analytical factors (qPCR targeted sequences, detection method, standard curves, etc.)[74,76]

On the basis of early findings, we postulated that "the size distribution profile of cirDNA fragments can be used to discriminate between healthy and cancer patients."[12] Employing a N-qPCR assay, which amplified nine sequence sizes from 60 bp upward, the number of fragments in the 60–145 bp range was ~3-fold higher in cancer-patient-derived cirDNA than in that of healthy individuals.[13,17] Few cancer-derived cirDNA fragments were over 400 bp (~1%), in contrast to healthy subjects with ~24% larger cirDNA fragments.[11,13,17] In addition, fragments containing a specific mutant oncogene sequence were more fragmented (especially below 138 bp) than the fragments containing the same counterpart wild-type sequence.[28]

Early studies disagreed on the size profiles of cirDNA in cancer patients. This was resolved by standardizing pre-analytics.[67,76–80] First, mutant cell-free DNA (cfDNA) fragments

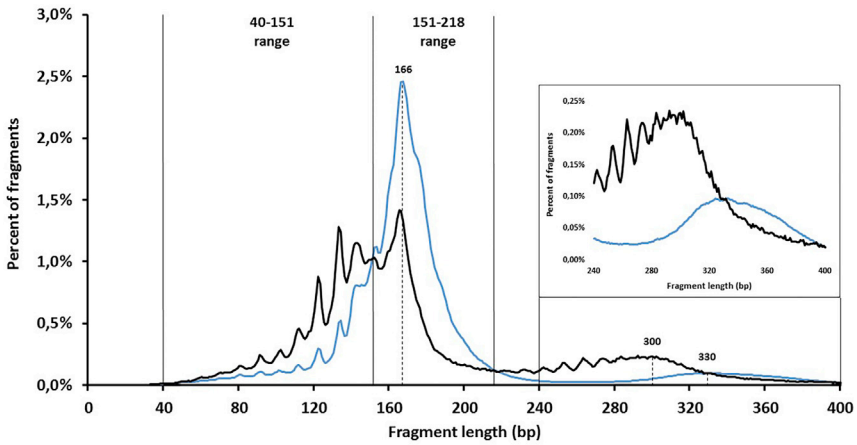

Figure 3. Compared with that of healthy individuals (blue), the cirDNA size profile of cancer patients trends toward shorter fragments (black)

The figure presents an illustrated comparison of the fragmentome of a metastatic colorectal cancer (mCRC) patient with a mutation allele frequency (MAF) of 68.5% (black line) and that of the mean of 17 healthy individuals (blue line) as determined by sWGS from DSP. Vertical lines indicate fragment lengths where the mean size profile curve of cirDNA from healthy individuals crosses that of the cancer patient cfDNA. Inset: zoom on the 240–480 bp size range corresponding to the DNA fragments associated with dinucleosomes. Compared with the healthy mean, the cancer patient size profile has an increased number of fragments in the 40–151 and 218–320 bp ranges, a decreased number of fragments in the 151–218 and 320–440 bp ranges. In addition, the cancer patient cirDNA size profile specifically shows (1) fragment levels plateauing between 145 and 156 bp; (2) lower fragment levels at 166 bp length; and (3) shorter corresponding peak at the dinucleosome (300 vs. 330 bp). The figure is adapted from data reported previously.[17]

were shown to be shorter than the corresponding wild-type sequence.[12,13,28] Thus, differences in size profile can be accurately observed with sequencing at fragment-level resolution (Figure 3). Comparing the frequency of sWGS reads of fragment size at 1 bp, resolution on the cirDNA size profile from plasma from subjects with cancer vs. healthy subjects, Thierry and Sanchez[29] defined a list of parametrics to distinguish cancer vs. healthy individual cirDNA extracts. Overall, cancer patient cirDNA shows subtle but reliable differences relative to healthy subjects (Figure 3). In individuals with cancer, cirDNA differences include a higher number of fragments below 150 bp, as well as a lower number of fragments between 151 and 218 bp; with dinucleosome corresponding fragments peaking at ~300 vs. ~330 bp. In addition, N-qPCR in combination with sWGS identified a lower proportion of fragments >1,000 bp in cancer patients, compared with healthy people.[17] Whereas sWGS provides accurate and precise size determination of cirDNA fragments below ~1,000 bp, cirDNA of higher molecular weight has diagnostic utility. For instance, the dynamic degradation of high-molecular-weight DNA or chromatin in the bloodstream, due to NETs, can be associated with cancer.[48] However, the quantification of such long fragments is currently subject to bias, by factors discussed above. Further research on long cirDNA fragments is required to more fully decipher their diagnostic value.

More precise parametrics enabling us to highly distinguish between cancer patients and healthy individuals based on cirDNA size profile have been determined as being frequency at a single fragment length, frequency of unique reads at single-base resolution, the difference of frequency of two single fragment lengths at 1 bp resolution, the size range fraction frequency, or the difference of size range ratio[17,29] (Figure 2). The higher the composition of mutant or malignant cell-derived cirDNA content (quantified as mutation allele frequency [MAF]), the greater the differences between the cirDNA size profiles of the plasma of metastatic colorectal[17,29] and hepatocellular[14] cancer patients and healthy individuals. However, this correlation has not been observed to be directly proportional.

ADVANCED OMICS APPROACHES UNDER DEVELOPMENT TO SCREEN CANCER

Chromatin organization along the genome is not random, and may be specific to cell gene regulation and cell types[81,82] (Figure 2). Consequently, fragment length is not the only parameter that needs to be investigated for the purpose of differentiating cirDNA fragments of different origins. Other analytical signals from the cirDNA fragmentome include a non-random fragmentation pattern, transcription factor occupancy, DNA GC content, and end motif pattern.

Genome-wide map of nucleosome occupancy

Deep sequencing of cirDNA allowed the identification of the genome-wide map of the nucleosome occupancy, confirmed the chromatin/nucleosome occupancy related patterns, with short cirDNA fragments revealing the footprint of transcription factor occupancy[16] (Figure 1). This method can infer which cell types contribute to cirDNA in pathological states, particularly cancer (Table 1). Almost all short cirDNA fragments are associated with nucleosomes with mutant alleles commonly occurring as shorter fragments.[62] Bioinformatics isolation of a specific subset of fragment lengths[62] aids in the detection of mutant cirDNA. This approach has been exploited[58,83] (Figure 1) to enhance sensitivity in detecting the presence of cirDNA with *in vitro* and *in silico* methods to enrich cirDNA in fragment sizes between 90 and 150 bp. In a large cohort, this size selection improved the identification of cirDNA from patients with glioma, renal, and pancreatic cancer.[58,83]

Nucleosome depletion regions

From initial observations with micrococcal nuclease digestion, in which nucleosome patterns at promoters were identified,[90] Ulz et al.[57,91] showed that analysis of nucleosome-depleted regions (NDR) provides functional information about cells releasing their DNA into the blood. In particular, fragmentomic patterns from healthy donors reflected the expression signature of hematopoietic cells. Machine learning analyses of the plasma from individuals with cancer revealed cancer driver genes associated with somatic

Table 1. Fragmentomics-based clinical assays toward early detection of cancer

	Methods	Analytical approach	Tested population	Characteristics summary	Reference
Large-scale chromatin organization	DELFI: genome-wide cell-free DNA fragmentation	sequencing at 5 Mb resolution	multiple early-stage cancers	needs deep sequencing; large retrospective cohort (n = 236); discriminate various cancer types; potential confirmed by subsequent report	Cristiano et al.[56] Mathios et al.[84]
Local genomic bins	orientation-aware cirDNA fragmentation analysis (OCF)	sequencing at 1 kb resolution	early-stage hepatocellular carcinoma	small size cohort retrieved from the EGA (n = 90); informs tissue of origin compared with pregnant women and organ transplantation recipients	Sun et al.[85]
	DNA fragmentation hotspot	sequencing at 1 kb resolution	early-stage hepatocellular carcinoma	small size cohort retrieved from public datasets (n = 60); informs some elements of tissue of origin; fine-scale fragmentation at open chromatin regions	Zhou et al.[86]
	TFBS: assessing TF-binding site accessibility	sequencing at 1 kb resolution	early-stage colorectal cancer	large retrospective cohort (n = 213); more advanced method following fragment coverage near transcription-start sites	Ulz et al.[57]
Fragment level	preferred end	sequencing at fragment level (<1,000 bp) resolution	early-stage hepatocellular carcinoma	need deep sequencing; small size cohort (n = 90); informs tissue of origin as compared with cirrhotic patients; no sensitivity or specificity level reported	Jiang et al.[59]
	end motif and diversity score	sequencing at fragment level (<1,000 bp) resolution	early-stage hepatocellular carcinoma	needs deep sequencing; small size cohort (n = 34); global and local analysis; no sensitivity or specificity level reported	Jiang et al.[25]
	jagged end	sequencing at fragment level (<1,000 bp) resolution	early-stage hepatocellular carcinoma	small size cohort (n = 34); followed 2008 observation on protruding forms of double-stranded DNA in plasma	Suzuki et al.[87] Jiang et al.[59]
	LIQUORICE algorithm: epigenetic signatures based on cirDNA fragmentation patterns	sequencing at multiple resolutions	Ewing sarcoma and pediatric sarcomas.	small size cohort (95 patients with Ewing sarcoma and 31 patients with other pediatric sarcomas); no reported patient stage; no sensitivity or specificity reported; assistance of machine learning classifiers	Peneder et al.[88]
	MITEST algorithm: nuclear and mitochondrial origin and signatures based on cirDNA fragmentation distribution	N-qPCR	multiple early- and late-stage cancers	large size cohort (N = 983); assistance of machine learning classifiers	Tanos et al.[89]

This table illustrates most of the various technologies employed to date with observed data. However, it does not attempt to compare their respective efficacy, since in most studies machine learning algorithms are not indicated, and some studies have carefully selected samples that may have their biases. For instance, AUC (area under curves) values cannot be compared precluding assessment of the performance values. OCF, orientation-aware plasma cell-free fragmentation signals; TF, transcription factor; TFBS, transcription factor-binding sites; AUROC, area under receiver operating characteristic curve; MDS, end motif diversity score; N-qPCR, nested qPCR; European genome archives, EGA.

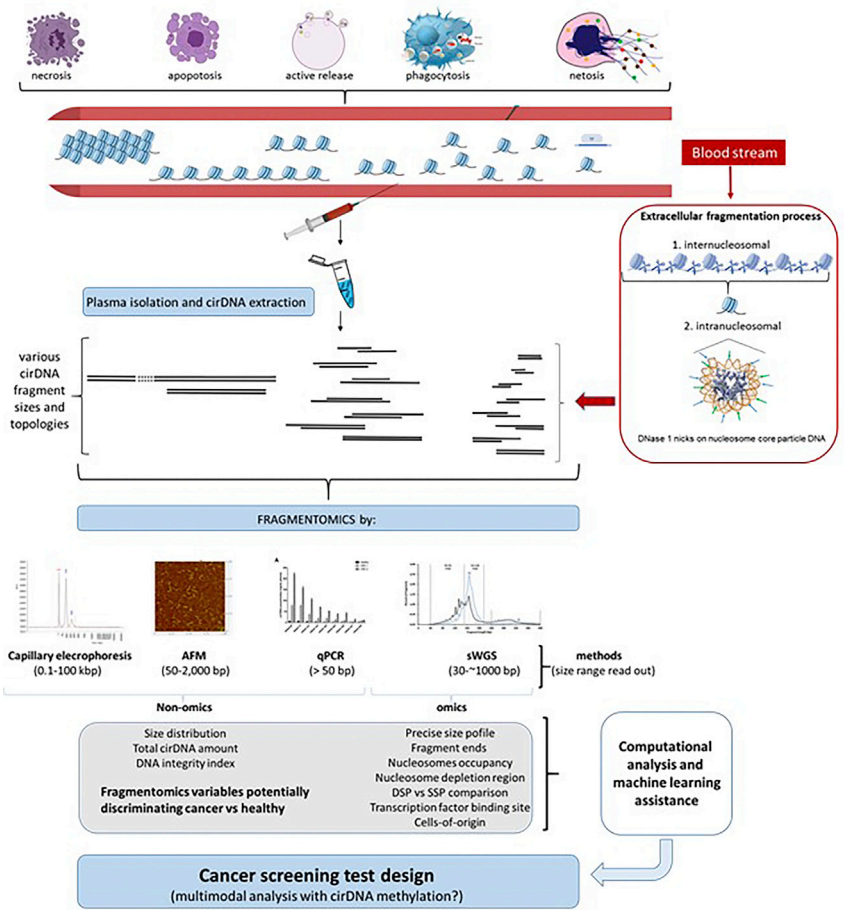

Figure 4. cirDNA fragmentation and fragmentomics variables enabling discrimination between cancer and healthy individuals toward a cancer screening test

qPCR, quantitative polymerase chain reaction; sWGS, shallow whole-genome sequencing; SSP, single-strand DNA library preparation; DSP, double-strand DNA library preparation; AFM, atomic force microscopy.

copy number gains.[92] Thus, NDR analysis at transcription start sites and exon junctions could offer a quantitative and cost-efficient approach to track tumor cirDNA dynamics and progression.

DELFi genome-wide cirDNA fragmentation

While the first machine learning-assisted cancer screening test using cirDNA was reported in 2016,[91] DELFI (DNA evaluation of fragments for early interception) is the first test to employ fragmentomics (Figure 1). DELFI exploits information regarding fragment length in 5 Mb windows along the genome in bins from the ratio of short (100–150 bp) over long (151–220 bp) fragments; similar to the parametric revealed by N-qPCR.[12,13,29] This test, along with mutation-based cirDNA analyses,[56] clinical risk factors, and carcinogen embryonic antigen levels, enabled the detection of cirDNA in 91% of subjects and in 94% of subjects with cancer across all stages and subtypes, including 91% of stage I/II and 96% of stage III/IV, at 80% specificity.[84] This study confirmed that fragmentomics analysis utilizing transcription factor binding sites to distinguished between individuals with small cell lung cancer and those with non-small cell lung cancer (AUC = 0.98).[84]

cirDNA fragment ends

cirDNA fragment ends from hepatocellular carcinoma (HCC) typically map to specific genomic coordinates, which can be quantitatively assessed, and are correlated with the amounts of mutant- or liver-derived cirDNA (Figure 1)[25] This shows promise as a cost-effective approach for improved theragnostic or potential cancer screening. In providing the proof of concept that end motifs may characterize cancer-derived cirDNA, Lo's group confirmed the importance of analyzing cirDNA fragment ends. In addition, the CCCA motif is less abundant in HCC patients,[25,63] and the profile of cirDNA end motifs may depend on tissue-of-origin. On the basis of these observations, it has been proposed that cirDNA may be used to identify biomarkers in other clinical areas, such as transplantation monitoring[25,85] and NIPT.[63]

Combining sWGS analysis from DSP-S and SSP-S

SSP-S has previously been employed in the field of paleontology to generate high-resolution genomes of ancient and generally highly degraded (short) DNA.[93] It reveals a higher proportion of short fragments (<168 nt) relative to DSP sequencing.[17,29,93,94] Note, SSP-S analysis better harmonized the fragment size distribution determined by sequencing and N-qPCR analysis compared with DSP-S.[30] SSP-S revealed a higher proportion of shorter cirDNA fragments in the ~40–152 nt range in cancer patients compared with cirDNA derived from healthy individuals[17,29,30] (Figure 1). Since SSP-S-based size profiles provide enrichment of fragment number of short size, SSP-S might therefore be better suited to cancer screening.[17,29,30] In addition, SSP-S detects fragment size due to different DNA strand breaks.

Combining DSP-S and SSP-S, Sanchez et al. [17] inferred that the DNA molecule packed on the chromatosome is double stranded and exhibits various structures due to nuclease activities. These structures are various: intact/blunt dsDNA, dsDNA with one or more nicks in one strand, and dsDNA with one or more nicks in both strands, the latter DNA ends are mostly jagged, implying that they bear single-stranded protruding ends[17] (Figure 4). Moreover, SSP-S provides more sequence information than DSP-S, since it retains the exact 5′ and 3′ protruded endpoints of each input DNA fragment, improving the performance of methods based on cirDNA fragment end motifs.[25]

Combining qPCR and sWGS fragmentomics

The combination of sWGS from DSP and SSP with N-qPCR encompasses the full cirDNA fragment size distribution.[30,50] This combinatory approach accurately revealed differences between cancer and healthy cirDNA. For example, differences were identified when comparing the cirDNA frequency either at a specific single size with 1 nt resolution, or at selected size ranges, or at selected size fraction ranges ratio below ~1,000 bp and over 1,000 bp.[50] The lower and upper size limits of fragment detection by conventional sequencing is estimated as 20–30 bp and ~1,000 bp. By contrast, analysis combining N-qPCR and shallow WGS revealed that 10%–20% of fragments were over 1,000 bp in healthy individuals, whereas in individuals with cancer only approximately 1% of fragments were that size.

Mapping tumor-specific transcription factor binding sites

cirDNA transcription factor profiling along with nucleosome occupancy levels of expressed and silent housekeeping genes can inform upon copy number alterations in cancer. The expression of genes affected by such genome modifications typically differs in particular genes in instances of cancer. Applying machine learning-assisted gene classification to the data identified a hematopoietic expression signature in the plasma of healthy individuals. In contrast, metastatic cancer patients' plasma expressed cancer driver genes exhibiting somatic copy number gains. This strategy was validated by mapping tumor-specific transcription factor binding sites (TFBS) in more than a thousand cirDNA samples from both healthy individuals and cancer patients.[57] The ability to map tumor-specific transcription factor binding *in vivo* from blood samples aids in clinical analysis of the noncoding genome. Utilizing data from cirDNA fragmentation patterns and the accessibility of TFBS, Ulz et al. distinguished patient-specific and tumor-specific patterns with a high level of accuracy, permitting the detection of early-stage cancer.[57]

cirDNA jagged end analysis

The ability to detect jagged ends[59] also enables the detection of cancer[86] (Figure 4). Analyses of double-stranded plasma cirDNA that carries single-stranded protruding 5′ ends identified an uneven proportion of C and G cirDNA fragments at 5′ and 3′ ends and also revealed 5′ protruding forms of dsDNA in plasma.[87] The presence of jagged ends was inferred by directly comparing DSP-S- and SSP-S-derived size profiles.[17,30,50] A higher percentage of jagged ends was observed in HCC cirDNA fragments compared with control individuals.[59] Further development of a CC-tag technology based on the use of methylated cytosine for the end repair reaction, given the usually unmethylated non-CpG cytosine in the human genome, was later reported. This technological advance helped in the identification of urinary cfDNA from bladder cancer patient plasma has been identified by the analysis of jagged ends using bisulfite paired-end sequencing (termed Jag-seq).[86]

Computational analysis and machine learning assistance applied to fragmentomics

cirDNA fragmentomics generally produced counts of cirDNA from sequencing at low depth (termed low-pass or shallow sequencing, 2–10×). This contrasts with high depth (~50,000×) sequencing of small regions (2 Mb), which are used for detecting mutations in a "liquid biopsy" approach. Computational analysis is based on sequencing conventional workflow including (1) quality control and reads pre-processing; (2) alignment and visualization; and (3) statistical analysis and interpretation.[95] Reference-based or reference-free deconvolution algorithms enable the recovery of the original signal from a mixture of signals, and appears to help in mitigating tumor heterogeneity issues.

WGS analysis applied when searching for disease-relevant patterns requires unbiased methods. Machine learning enables the identification/classification of features that can be relevant for the general screening population in an assay with many measurable variables. Thus, classification from a large cohort of patients with and without early-stage cancer can be implemented by computational approaches assisted with learning associations between cirDNA profiles and cancer status.[96] Various machine learning methods, such as artificial neural networks, Bayesian networks, support vector machines, and decision trees were developed for the application of predictive models. In the case of cancer screening, in assay-based methods, which rely on cirDNA analysis, the decision tree approach appears to be the most frequently used.[56,61,89]

OTHER STATE-OF-THE-ART HIGH-THROUGHPUT APPROACHES

Fragmentomics-based technology may prove useful for screening plasma from cancer patients (Table 1). An exciting development is the utilization of the fragmentation pattern to characterize open chromatin regions (OCR).[85] This technology provides a means to determine the tissue-of-origin for specific cirDNA fragments. The bioinformatics pipeline, OCRDetector, enables the detection of OCRs by combining cirDNA sequencing coverage with the number of DNA fragments spanning a 120 bp genomic window, minus the number of fragments with an endpoint within that same window.[97]

Thousands of unique plasma extrachromosomal circular DNAs (eccDNAs) in normal and cancer subjects have been reported.[98] These microDNAs, first discovered in cell nuclei, originate from the chromosomal genome and exhibit longer sequence lengths (30–60% > 250 bp) than linear cirDNA, as determined by NGS. We see promise in the exploration of extending cirDNA fragmentomics to include eecDNAs for diagnosis, the scrutiny of gene regulation, and better understanding intercellular communication.

cirDNA fragmentation hotspots may characterize early-stage cancer patients. Hypo-fragmentation in specific genomic regions have identified hotspots in early-stage HCC. These findings suggest the possibility of detecting gene-regulatory aberrations in early-stage cancers. Finally, we see promise in incorporating the 3D genome status in analysis of cirDNA co-fragmentation patterns deriving from two spatially close regions inside the nucleus with the fragmentation evaluation of epigenetics from cfDNA sequencing[61] method.

THE NUCLEASIC ACTIVITY

cirDNA fragmentation is a dynamic process[30] and depends upon nuclease activity within the circulatory system. This

influences analytical signatures, such as structure, fragmentation pattern, length, and cirDNA fragment end sequence.[23,24,61,99] DNASE1L3, DNASE1, and DNA fragmentation factor B all have a role in the fragmentation process.[23,63,99] A model of the cirDNA fragmentation process postulates that DNA fragmentation factor B and DNASE1L3, as well as apoptotic nucleases, intracellularly generate DNA fragment plasma, and that plasma DNASE1L3 is responsible for extracellular nuclease activity. While A-end enriched cirDNA are predominantly produced intracellularly, DNASE1L3 generates C-end enriched cirDNA in plasma.[99] It is assumed that the nucleosome is then further degraded concomitantly by DNASE1 and by proteolysis to the point of complete digestion. A wider perspective on cirDNA fragmentation has been proposed[49]: in addition to an apoptotic mechanism of release, mononucleosomes in circulation may derive from nucleases through the dynamic degradation of chromatin or high-molecular-weight DNA deriving from various biological processes, in particular from NETs. Despite the fact that, subsequent to induced apoptosis, mono and di-N-associated fragments have the same pattern as cirDNA found in the blood, this similarity alone is an insufficient reason to consider all cirDNA the product of apoptosis. Two NETs markers, myeloperoxidase and neutrophil elastase, have a catalytic capacity on HMW DNA, indicating that NETs have auto-catabolic properties.[100] The suggested primary cirDNA cells-of-origin are megakaryocytic and neutrophilic.[101] This was identified through cell-free chromatin immunoprecipitation followed by sequencing (cfChIP-seq) in individual who are healthy and those that have cancer. This experiment showed that low sequencing depth (sWGS) can obtain systemic and genome-wide information from blood samples, opening a range of opportunities for diagnosis and the interrogation of physiological and pathological processes generating cirDNAs.[101]

NETs are constitutively formed in an unproblematic and unobtrusive manner, but their formation can be elevated during the progression of cancer,[48,102] where they manifest as an inflammatory disease. N-qPCR or capillary electrophoresis can detect HMW cirDNA over 1,000 bp, especially between 2,000 and 10,000 bp,[50] which may originate from necrosis and NETosis (Figure 4). However, cancer-derived cirDNA have a low frequency of HMW cirDNA compared with cirDNA derived from healthy individuals.[30,50]

Fragments below 140 bp, in particular below 70 bp,[14,30,50,56,60] points to a difference in the cirDNA fragmentation process probably due to a difference in nuclease activity. Such phenomena have, to date, been poorly described. Despite this, we speculate that this fragmentation process may be due to the specific genetic and epigenetic features found in the DNA of malignant cells. Alternatively, this might be explained by the tumor microenvironment, since tumors are composed of malignant, stromal, endothelial and immunological cells, within a specific milieu (characterized by ROS, hypoxia, minerals, pH, nutrients, etc.). Despite much study needed, overall, fragmentomics has enabled the identification that cancer derived cirDNA exhibit lower level of HMW cirDNA, and the higher frequency of short cirDNA relative to healthy individuals.

LIMITATIONS OF FRAGMENTOMICS TOWARD A ROUTINE SCREENING TEST

For clinical use, cirDNA fragmentomics is a nascent approach within the emerging field of cirDNA markers. As with any emerging biomarker, numerous hurdles remain.

Technical/analytical

Pre-analytics

As noted, size profile differences between cirDNA derived from tumor vs. healthy tissue are significant but small. A potential confounder includes the potential contamination of the cirDNA extract by blood cell genomic DNA due to improper blood sample handling and storage before plasma isolation.[76] High-molecular-weight DNA released from the genome of blood cells following degradation is subject to nuclease activity, and results in the detection of DNA fragments whose sizes correspond to their packaging in tri- or quadri-nucleosomes (of 440–500 and 600–700 bp ranges, respectively), which are not observed following sWGS analysis of a nearly pure plasma DNA extract.[50,56] To estimate this confounder, the proportion of fragments less than 260 bp should not be below 70% of the total cirDNA.[76] Quality control may be also performed with Agilent capillary electrophoresis or using sWGS.

Computational

While bioinformatics has considerably improved the analytical power of sequencing/WGS approaches, unfortunately it also presents potential limitations. First, the NGS-based methodology developed for the assessment of fragmentomics calls for the use of significant computational resources, high-throughput NGS platforms, complex bioinformatics analysis, and high-speed data processing; these requirements can significantly augment the cost and time needed to produce the analytical data. The provision/implementation of multimodal analysis from cirDNA is currently challenging, given its reliance on standard high-throughput sequencing, which requires specific protocols for the investigation of different classes of biomarkers. Multiple factors make the identification of fragmentome characteristics particularly challenging, including mapping and sequencing bias, inadequate read depth, GC content, PCR amplification, the choice of reference genome, and k-mer composition; these elements all can result in a misleading result from the computational model.[64,97] However, methods of correcting and normalizing coverage and size-derived technical artifacts have been described.[103,104] In addition, computational problems can arise in the detection of specific fragmentome characteristics, such as Jagged ends or OCRs. Since typical sequencing protocols take advantage of several DNA repair steps to maximize the amount of analyzable DNA, information as to the state of the DNA ends may be lost.

Controlled-access repositories

There is currently no publicly available centralized database with uniformly processed cirDNA datasets. In addition, "batch effects" due to the quality of reads, length, and mapping approach affect mapping locations of paired-end short-read sequencing, and consequently the downstream computational inference and data analysis. There has been a recent attempt to propose

a comprehensive database to host cirDNA WGS datasets across different pathological conditions.[105]

Biological limitations

With the field of fragmentomics still in its infancy, the lack of known biological confounders limits the implementation of most technological approaches. One biological limitation to the use of cirDNA as a tool for pan-cancer screening is the level of detection sensitivity,[50,106–108] irrespective of the genetic, epigenetic, or fragmentomics approaches used. Early detection of cancer before any signs of illness is a major objective. Therefore, cirDNA relies on its capacity to track tumor cirDNA from small tumors. Numerous studies have reported difficulties in detecting cirDNA derived from malignant cells in early-stage cancer, especially in tumors of low volume.[6,8,109] This is, first of all, due to the cirDNA amount being limited by the tumor size; and, second, by the cirDNA MAF often being low, making it difficult to distinguish cirDNA bearing genetic or epigenetic alterations relative to the preponderance of wild-type cirDNA. Tumors larger than 10–15 mm in diameter could probably be predictably detected by analyzing mutant cirDNA (cir-mutDNA). However, cirDNA with a MAF below 0.1%[110] is below the sensitivity level of most omics technologies.

The practical application of sWGS-based fragmentomics is currently limited in early-stage cancer screening. For instance, the sensitivity of most cirDNA-based methods in detecting stage I or II cancer has been low (<30%).[56,110,111] This is only true, however, when relying solely on cir-mutDNA. Our group speculates that cirDNA fragment size may also depend on cirDNA derived from the tumor microenvironment, which may produce cells with cancer-specific cirDNA fragmentation characteristics. Biological factors due to the presence of tumor cells might affect the fragmentation of a fraction of the wild-type cirDNA.[89] For instance (1) the nuclease activity of tumor microenvironment cells, such as the endothelial, stromal, or immune cells, might be modified compared with normal cells; and (2) cirDNA deriving from the degradation of cancer-induced NETs from stimulated neutrophils at the surface or inside a tumor might have a specific fragmentation process.

Necrosis is associated with tumor development, as are apoptosis and phagocytosis (Figure 4), which are in turn associated with defense mechanisms. Necrosis causes the destruction of malignant as well as tumor microenvironment cells, and of adjacent, non-tumor tissues.[39] In addition, tumorigenesis and tumor progression elicit a strong immune response, and lymphocytic cells in the tumor microenvironment may be a source of cirDNA release.[16,62] Thus, the tumor microenvironment, especially the immune cells, are now considered a major source of cirDNA in healthy individuals.[16,101,112] Understanding these processes could help determine the respective ratio of the proportion of cirDNA deriving from a malignant or a non-malignant tumor microenvironment or from germinal cells. We can therefore postulate, first, that a comparison of the mutation or variant allele frequency (MAF or VAF) determined from tumor tissue cannot be compared with that of cirDNA; and, second, that any comparison of cancer patients with the same RECIST criteria with respect to the value of the MAF or methylation level as determined from cirDNA must be done with caution, since immune response is specific to each patient. This highlights how the cell-of-origin may confound analysis of cirDNA.

Strategies that rely on machine learning classification, which is often used to analyze the data, can be biased by the quality of the control cohort. Thus, cautious standardization of the pre-analytical and analytical procedures for both the control cohort and the tested subjects are required. While the growing role of machine learning in therapeutics and diagnostics shows promise,[88] it also has limitations, such as the requirement for a large number of samples to train, batch effects, and classification bias due to imbalance in training dataset.[95]

High reproducibility of the size profiles (Figure 2) is thus important since even a slight variation in a tested individual's cirDNA may distinguish a cancer-based cirDNA from the cirDNA size profile of a healthy subject. Such reproducibility therefore needs to be tested against individual epidemiological factors that might potentially generate bias. Finally, while a large independent prospective cohort offers advantages relative to retrospective studies, cancer diagnostic power needs to be evaluated through cohort cross-validation.

Mechanisms behind cirDNA fragmentomics

To date, knowledge of the mechanisms behind cirDNA fragmentomics remains limited. Numerous potential factors have a direct or indirect role in cirDNA fragmentation, including epigenetic factors, such as DNA methylation or histone modifications. The impact of epigenetics on cirDNA fragmentation would appear to be strong, but remains poorly delineated. Understanding the mechanisms underlying cirDNA fragmentomics, may contribute to our understanding of the biology of non-cancer diseases.

Currently, the most important questions arising from the observation of cirDNA fragmentation concern its origin in gene regulation. Thus far, research has been done only on a small number of genes, notably silenced genes or housekeeping genes. However, this might lead to bias, in particular with respect to the maintenance of gene expression levels. Furthermore, the extent to which gene regulation effects the cirDNA fragmentation pattern may be impacted by physiological confounders, such as an individual's physical activity, drug treatment, cancer comorbidity, ethnicity, menstruation, chronobiology, and other factors. Furthermore, as investigated by Lo's[24,99] and Watanabe's[113] groups, over- or underexpression of nuclease activity may affect cirDNA fragmentation. Numerous studies have reported that cancer or other pathologies show an association of NET formation and cirDNA levels, supporting that cirDNA are NET byproducts.[49,100,102] Finally, cirDNA fragmentomics lacks a gold standard experimental model, especially a cell culture model, which would allow for the pooling of data obtained by various teams.

Comparing methods performance

Although evidence suggests a diagnostic potential of some characteristics of fragmentomes, any comparison of their performance tends to be difficult to assess when directly comparing them to healthy controls. Thus, their screening power is currently more descriptive than quantitative. The observations from replicated studies cannot currently be pooled, as most observations cannot be validated, given the variation in experimental models used by different labs. So far, only Velculescu's group has

confirmed its striking earlier observations on genome-wide fragmentation.[56,84] They used both the short:long ratio and select copy number variation events as a feature in a principal-component analysis prior to machine learning-based gradient boosted trees. The establishment of a cirDNA fragmentomics database to be used by all researchers in the fragmentomics field, as described,[105] would be invaluable. Efforts toward a better ranking test performance are likely in the coming years and will help to evolve fragmentomics.[88,89,114]

Current observational analyses to correlate cirDNA with assessments of survival to evaluate cancer diagnostic performance may result in bias, since this could vary over calendar time and across populations.[115] Relative survival estimates can be affected by lead-time bias, due to any delays between the diagnosis of cancer via screening and cancer diagnosis due to symptoms.[115] The impact of lead time on survival has not yet been evaluated with reference to the development of a fragmentomics assisted cancer screening test.

Limitations in the context of cancer screening requirements

A fragmentomics-based screening test remains far from a practical reality. An ideal cancer screening test would discriminate perfectly between individuals with and without cancer. In reality, all screening tests sometimes give both false-negative and false-positive results.

To adequately judge a screening test's performance, selecting the appropriate clinical study target population is crucial. Various options may be taken into consideration, but all should be opted for when evaluating a cancer screening test: healthy vs. at-risk vs. high-risk populations, symptomatic patients vs. patients with indeterminate physical or radiographic findings. Indeed, along with a screening test's results, it is necessary to conduct a blind study and associate the results with previously diagnosed individuals and at-risk populations. As indicated above, no clearly defined standardized control cohort has yet been established. Any cancer screening test must have positive results confirmed with other techniques.[107] In addition, sensitivity in the detection of early-stage cancers (stage I or II) is needed. Another obstacle to establishing the value of a screening test's performance is the lack of a gold standard universal screening test. Clinical workflows are needed, especially in cases of a positive result, with respect to: (1) its interpretation; (2) guiding its evaluation toward diagnostic resolution; and (3) monitoring patients with no cancer diagnosis. Furthermore, the question arises whether it would be preferable to initially target specific cancer types to compare a fragmentomics-based test with specific biomarkers currently used in the routine management care of that cancer type.

Finally, cirDNA fragmentomics studies should be extended with respect to non-cancer pathologies, in particular inflammatory diseases, which potentially may produce cancer screening test non-specificity.[48]

Conclusion

Fragmentomics constitutes a potentially powerful tool for the assessment of cirDNA, although it remains in its infancy. While fragmentomics may be limited in application as a stand-alone cancer screening test, a multianalyte/multimodal approach might provide the reliability required to exploit fragmentomics signatures (Figure 4). A combination of the detection of mutant cirDNA, based on the number of informative reads sequenced across multiple patient-specific loci, and signal-enrichment approaches might be used to screen individuals for a specific cancer type.[116] The combination with an approach unrelated to fragmentomics, such as the determination of genetic alterations, cellular or protein cancer markers, or methylation-based methods,[109,111,117,118] should be effective. Methylation is a strong candidate as a cancer screening analyte, since changes to interactions between DNA methylation, chromatin structure, and gene transcription can cause the repositioning of nucleosomes and carcinogenesis.[111,119,120] In addition, advances in cirDNA quantification, identification of the structure- or origin-based indicators, as determined by advanced qPCR methods, might be added to an omics approach.[30,50,89,121] Since physiological changes, such as tissue injury in human disease, could individually impact cirDNA fragmentation, fragmentomics could also guide clinicians during patient follow-up.

We speculate that fragmentomics would help in the monitoring of cancer patient treatment following early detection of the disease. Apart from cirDNA obtained from the circulation/blood samples, we foresee the possibility of applying fragmentomics to extracellular DNA derived from other sources, depending on the specific cancer in question: urine[26] for prostate or bladder cancer, LCR for brain cancer, pleural effusion supernatant or saliva for lung cancer, for example.[122] Given that the field of fragmentomics remains at an early stage, many significant issues must be resolved before envisaging its clinical implementation. Nevertheless, current technological innovations promise a strategy with strong potential, and we are convinced that recent advances in this field will help to overcome most of its current limitations.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2022.100242.

ACKNOWLEDGMENTS

The author thanks Cynthia Sanchez, Florence Frayssinoux, and Ekaterina Pisareva (IRCM, Montpellier) for excellent technical assistance; Charles Marcaillou and Mehdi Bensassi (Integragen SA, Paris); and Cormac Mc Carthy (Mc Carthy Consultant, Montpellier) for English editing. This study was partially supported by SIRIC Montpellier Cancer Grant INCa_Inserm_DGOS_12553 and AR Thierry by Inserm.

DECLARATION OF INTERESTS

The author has two patents related to this work: A.R. Thierry and C. Sanchez. "Method for screening a subject for a cancer," THIERRY17392NC, no. 17306721.6, December 7, 2017. US2021172019 (A1) 2021-06-10. A.R. Thierry. "Method for screening a subject for a cancer." August 1, 2019, EP 19 30 6003. US2022290244 (A1) 2022-09-15.

REFERENCES

1. Ivanov, M., Baranova, A., Butler, T., Spellman, P., and Mileyko, V. (2015). Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. BMC Genom. *16*, S1. https://doi.org/10.1186/1471-2164-16-S13-S1.

2. Zamyatnin, A.A. (2009). Fragmentomics of natural peptide structures. Biochemistry. *74*, 1575–1585. https://doi.org/10.1134/S0006297909130100.

3. Diaz, L.A., and Bardelli, A. (2014). Liquid biopsies: genotyping circulating tumor DNA. J. Clin. Oncol. *32*, 579–586. https://doi.org/10.1200/JCO.2012.45.2011.

4. Otandault, A., Anker, P., Al Amir Dache, Z., Guillaumon, V., Meddeb, R., Pastor, B., Pisareva, E., Sanchez, C., Tanos, R., Tousch, G., et al. (2019). Recent advances in circulating nucleic acids in oncology. Ann. Oncol. *30*, 374–384. https://doi.org/10.1093/annonc/mdz031.

5. Wan, J.C.M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J.D., Caldas, C., Pacey, S., Baird, R., and Rosenfeld, N. (2017). Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat. Rev. Cancer *17*, 223–238. https://doi.org/10.1038/nrc.2017.7.

6. Diehl, F., Li, M., Dressman, D., He, Y., Shen, D., Szabo, S., Diaz, L.A., Goodman, S.N., David, K.A., Juhl, H., et al. (2005). Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc. Natl. Acad. Sci. USA *102*, 16368–16373. https://doi.org/10.1073/pnas.0507904102.

7. Bronkhorst, A.J., Ungerer, V., Diehl, F., Anker, P., Dor, Y., Fleischhacker, M., Gahan, P.B., Hui, L., Holdenrieder, S., and Thierry, A.R. (2020). Towards systematic nomenclature for cell-free DNA. Hum. Genet. *140*, 565–578. https://doi.org/10.1007/s00439-020-02227-2.

8. Thierry, A.R., El Messaoudi, S., Gahan, P.B., Anker, P., and Stroun, M. (2016). Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev. *35*, 347–376. https://doi.org/10.1007/s10555-016-9629-x.

9. Tie, J., Cohen, J.D., Wang, Y., Christie, M., Simons, K., Lee, M., Wong, R., Kosmider, S., Ananda, S., McKendrick, J., et al. (2019). Circulating tumor DNA analyses as markers of recurrence risk and benefit of adjuvant therapy for stage III colon cancer. JAMA Oncol. *5*, 1710–1717. https://doi.org/10.1001/jamaoncol.2019.3616.

10. Thierry, A.R., Pastor, B., Jiang, Z.-Q., Katsiampoura, A.D., Parseghian, C., Loree, J.M., Overman, M.J., Sanchez, C., Messaoudi, S.E., Ychou, M., and Kopetz, S. (2017). Circulating DNA demonstrates convergent evolution and common resistance mechanisms during treatment of colorectal cancer. Clin. Cancer Res. *23*, 4578–4591. https://doi.org/10.1158/1078-0432.CCR-17-0232.

11. Thierry, A.R., Mouliere, F., Gongora, C., Ollier, J., Robert, B., Ychou, M., Del Rio, M., and Molina, F. (2010). Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. Nucleic Acids Res. *38*, 6159–6175. https://doi.org/10.1093/nar/gkq421.

12. Thierry, A.R., and Molina, F.; CNRS (2012). Analytical Methods for Cell Free Nucleic Acids and Applications.

13. Mouliere, F., Robert, B., Arnau Peyrotte, E., Del Rio, M., Ychou, M., Molina, F., Gongora, C., and Thierry, A.R. (2011). High fragmentation characterizes tumour-derived circulating DNA. PLoS One *6*, e23418. https://doi.org/10.1371/journal.pone.0023418.

14. Jiang, P., Chan, C.W.M., Chan, K.C.A., Cheng, S.H., Wong, J., Wong, V.W.-S., Wong, G.L.H., Chan, S.L., Mok, T.S.K., Chan, H.L.Y., et al. (2015). Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc. Natl. Acad. Sci. USA *112*, E1317–E1325. https://doi.org/10.1073/pnas.1500076112.

15. Yu, S.C.Y., Lee, S.W.Y., Jiang, P., Leung, T.Y., Chan, K.C.A., Chiu, R.W.K., and Lo, Y.M.D. (2013). High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing. Clin. Chem. *59*, 1228–1237. https://doi.org/10.1373/clinchem.2013.203679.

16. Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M., and Shendure, J. (2016). Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell *164*, 57–68. https://doi.org/10.1016/j.cell.2015.11.050.

17. Sanchez, C., Roch, B., Mazard, T., Blache, P., Dache, Z.A.A., Pastor, B., Pisareva, E., Tanos, R., and Thierry, A.R. (2021). Circulating nuclear DNA structural features, origins, and complete size profile revealed by fragmentomics. JCI Insight *6*, e144561. https://doi.org/10.1172/jci.insight.144561.

18. Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L., and Quake, S.R. (2010). Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing. Clin. Chem. *56*, 1279–1286. https://doi.org/10.1373/clinchem.2010.144188.

19. Chan, K.C.A., Zhang, J., Hui, A.B.Y., Wong, N., Lau, T.K., Leung, T.N., Lo, K.-W., Huang, D.W.S., and Lo, Y.M.D. (2004). Size distributions of maternal and fetal DNA in maternal plasma. Clin. Chem. *50*, 88–92. https://doi.org/10.1373/clinchem.2003.024893.

20. Lo, Y.M.D., Tein, M.S., Lau, T.K., Haines, C.J., Leung, T.N., Poon, P.M., Wainscoat, J.S., Johnson, P.J., Chang, A.M., and Hjelm, N.M. (1998). Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am. J. Hum. Genet. *62*, 768–775. https://doi.org/10.1086/301800.

21. Chiu, R.W.K., Cantor, C.R., and Lo, Y.M.D. (2009). Non-invasive prenatal diagnosis by single molecule counting technologies. Trends Genet. *25*, 324–331. https://doi.org/10.1016/j.tig.2009.05.004.

22. Lo, Y.M.D., Chan, K.C.A., Sun, H., Chen, E.Z., Jiang, P., Lun, F.M.F., Zheng, Y.W., Leung, T.Y., Lau, T.K., Cantor, C.R., and Chiu, R.W.K. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci. Transl. Med. *2*, 61ra91. https://doi.org/10.1126/scitranslmed.3001720.

23. Serpas, L., Chan, R.W.Y., Jiang, P., Ni, M., Sun, K., Rashidfarrokhi, A., Soni, C., Sisirak, V., Lee, W.-S., Cheng, S.H., et al. (2019). *Dnase1l3* deletion causes aberrations in length and end-motif frequencies in plasma DNA. Proc. Natl. Acad. Sci. USA *116*, 641–649. https://doi.org/10.1073/pnas.1815031116.

24. Han, D.S.C., and Lo, Y.M.D. (2021). The Nexus of cfDNA and nuclease biology. Trends Genet. *37*, 758–770. https://doi.org/10.1016/j.tig.2021.04.005.

25. Jiang, P., Sun, K., Peng, W., Cheng, S.H., Ni, M., Yeung, P.C., Heung, M.M.S., Xie, T., Shang, H., Zhou, Z., et al. (2020). Plasma DNA end motif profiling as a fragmentomic marker in cancer, pregnancy and transplantation. Cancer Discov. CD-19-0622. https://doi.org/10.1158/2159-8290.CD-19-0622.

26. Tsui, N.B.Y., Jiang, P., Chow, K.C.K., Su, X., Leung, T.Y., Sun, H., Chan, K.C.A., Chiu, R.W.K., and Lo, Y.M.D. (2012). High resolution size analysis of fetal DNA in the urine of pregnant women by paired-end massively parallel sequencing. PLoS One *7*, e48319. https://doi.org/10.1371/journal.pone.0048319.

27. Lo, Y.M.D., and Poon, L.L.M. (2003). The ins and outs of fetal DNA in maternal plasma. Lancet *361*, 193–194. https://doi.org/10.1016/S0140-6736(03)12319-7.

28. Mouliere, F., El Messaoudi, S., Gongora, C., Guedj, A.-S., Robert, B., Del Rio, M., Molina, F., Lamy, P.-J., Lopez-Crapez, E., Mathonnet, M., et al. (2013). Circulating cell-free DNA from colorectal cancer patients may reveal high KRAS or BRAF mutation load. Transl. Oncol. *6*, 319–328. https://doi.org/10.1593/tlo.12445.

29. Thierry, A.R., and Sanchez, C.; INSERM (2017). Method for Screening a Subject for a Cancer.

30. Sanchez, C., Snyder, M.W., Tanos, R., Shendure, J., and Thierry, A.R. (2018). New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis. NPJ Genom. Med. *3*, 31. https://doi.org/10.1038/s41525-018-0069-0.

31. Quake, S. (2012). Sizing up cell-free DNA. Clin. Chem. *58*, 489–490. https://doi.org/10.1373/clinchem.2011.174250.

32. Guo, J., Ma, K., Bao, H., Ma, X., Xu, Y., Wu, X., Shao, Y.W., Jiang, M., and Huang, J. (2020). Quantitative characterization of tumor cell-free DNA shortening. BMC Genom. *21*, 473. https://doi.org/10.1186/s12864-020-06848-9.

33. Stroun, M., Anker, P., Lyautey, J., Lederrey, C., and Maurice, P.A. (1987). Isolation and characterization of DNA from the plasma of cancer patients.

Eur. J. Cancer Clin. Oncol. *23*, 707–712. https://doi.org/10.1016/0277-5379(87)90266-5.

34. Boender, P.J., Heijtink, R.A., and Hellings, J.-A. (1989). Nucleosomal fragments in serum may directly reflect cell-mediated cytotoxic activity in vivo. Clin. Immunol. Immunopathol. *53*, 87–98. https://doi.org/10.1016/0090-1229(89)90104-9.

35. Rumore, P.M., and Steinman, C.R. (1990). Endogenous circulating DNA in systemic lupus erythematosus. Occurrence as multimeric complexes bound to histone. J. Clin. Invest. *86*, 69–74. https://doi.org/10.1172/JCI114716.

36. Giacona, M.B., Ruben, G.C., Iczkowski, K.A., Roos, T.B., Porter, D.M., and Sorenson, G.D. (1998). Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. Pancreas *17*, 89–97. https://doi.org/10.1097/00006676-199807000-00012.

37. Heitzer, E., Auinger, L., and Speicher, M.R. (2020). Cell-free DNA and apoptosis: how dead cells inform about the living. Trends Mol. Med. *26*, 519–528. https://doi.org/10.1016/j.molmed.2020.01.012.

38. Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F.O., Hesch, R.D., and Knippers, R. (2001). DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. *61*, 1659–1665.

39. Rostami, A., Lambie, M., Yu, C.W., Stambolic, V., Waldron, J.N., and Bratman, S.V. (2020). Senescence, necrosis, and apoptosis govern circulating cell-free DNA release kinetics. Cell Rep. *31*, 107830. https://doi.org/10.1016/j.celrep.2020.107830.

40. Fleischhacker, M., and Schmidt, B. (2007). Circulating nucleic acids (CNAs) and cancer—a survey. Biochim. Biophys. Acta *1775*, 181–232. https://doi.org/10.1016/j.bbcan.2006.10.001.

41. Mouliere, F., El Messaoudi, S., Pang, D., Dritschilo, A., and Thierry, A.R. (2014). Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. Mol. Oncol. *8*, 927–941. https://doi.org/10.1016/j.molonc.2014.02.005.

42. Zonta, E., Nizard, P., and Taly, V. (2015). Assessment of DNA integrity, applications for cancer research. In Advances in Clinical Chemistry (Elsevier), pp. 197–246. https://doi.org/10.1016/bs.acc.2015.03.002.

43. Andersen, R.F., Spindler, K.-L.G., Brandslund, I., Jakobsen, A., and Pallisgaard, N. (2015). Improved sensitivity of circulating tumor DNA measurement using short PCR amplicons. Clin. Chim. Acta *439*, 97–101. https://doi.org/10.1016/j.cca.2014.10.011.

44. Pang, D., Thierry, A.R., and Dritschilo, A. (2015). DNA studies using atomic force microscopy: capabilities for measurement of short DNA fragments. Front. Mol. Biosci. *2*, 1. https://doi.org/10.3389/fmolb.2015.00001.

45. Liu, K.J., Brock, M.V., Shih, I.-M., and Wang, T.-H. (2010). Decoding circulating nucleic acids in human serum using microfluidic single molecule spectroscopy. J. Am. Chem. Soc. *132*, 5793–5798. https://doi.org/10.1021/ja100342q.

46. Andriamanampisoa, C.-L., Bancaud, A., Boutonnet-Rodat, A., Didelot, A., Fabre, J., Fina, F., Garlan, F., Garrigou, S., Gaudy, C., Ginot, F., et al. (2018). 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. *90*, 3766–3774. https://doi.org/10.1021/acs.analchem.7b04034.

47. Van den Ackerveken, P., Lobbens, A., Turatsinze, J.-V., Solis-Mezarino, V., Völker-Albert, M., Imhof, A., and Herzog, M. (2021). A novel proteomics approach to epigenetic profiling of circulating nucleosomes. Sci. Rep. *11*, 7256. https://doi.org/10.1038/s41598-021-86630-3.

48. Pastor, B., Abraham, J.-D., Pisareva, E., Sanchez, C., Kudriavstev, A., Tanos, R., Mirandola, A., Mihalovičová, L., Pezzella, V., Adenis, A., et al. (2022). Association of neutrophil extracellular traps with the production of circulating DNA in patients with colorectal cancer. iScience *25*, 103826. https://doi.org/10.1016/j.isci.2022.103826.

49. Thierry, A.R., and Roch, B. (2020). Neutrophil extracellular traps and by-products play a key role in COVID-19: pathogenesis, risk factors, and therapy. J. Clin. Med. *9*, 2942. https://doi.org/10.3390/jcm9092942.

50. Sanchez, C., Roch, B., Mazard, T., Blache, P., Dache, Z.A.A., Pastor, B., Pisareva, E., Tanos, R., and Thierry, A.R. (2021). Circulating nuclear DNA structural features, origins, and complete size profile revealed by fragmentomics. JCI Insight *6*, e144561. https://doi.org/10.1172/jci.insight.144561.

51. Fernando, M.R., Jiang, C., Krzyzanowski, G.D., and Ryan, W.L. (2018). Analysis of human blood plasma cell-free DNA fragment size distribution using EvaGreen chemistry based droplet digital PCR assays. Clin. Chim. Acta *483*, 39–47. https://doi.org/10.1016/j.cca.2018.04.017.

52. Bronkhorst, A.J., Ungerer, V., and Holdenrieder, S. (2019). Comparison of methods for the quantification of cell-free DNA isolated from cell culture supernatant. Tumour. Biol. *41*, 101042831986636. https://doi.org/10.1177/1010428319866369.

53. Maggi, E.C., Gravina, S., Cheng, H., Piperdi, B., Yuan, Z., Dong, X., Libutti, S.K., Vijg, J., and Montagna, C. (2018). Development of a method to implement whole-genome bisulfite sequencing of cfDNA from cancer patients and a mouse tumor model. Front. Genet. *9*, 6. https://doi.org/10.3389/fgene.2018.00006.

54. Holdenrieder, S., Stieber, P., Bodenmüller, H., Busch, M., Pawel, J., Schalhorn, A., Nagel, D., and Seidel, D. (2006). Circulating nucleosomes in serum. Ann. N. Y. Acad. Sci. *945*, 93–102. https://doi.org/10.1111/j.1749-6632.2001.tb03869.x.

55. El Messaoudi, S., Mouliere, F., Du Manoir, S., Bascoul-Mollevi, C., Gillet, B., Nouaille, M., Fiess, C., Crapez, E., Bibeau, F., Theillet, C., et al. (2016). Circulating DNA as a strong multimarker prognostic tool for metastatic colorectal cancer patient management care. Clin. Cancer Res. *22*, 3067–3077. https://doi.org/10.1158/1078-0432.CCR-15-0297.

56. Cristiano, S., Leal, A., Phallen, J., Fiksel, J., Adleff, V., Bruhm, D.C., Jensen, S.Ø., Medina, J.E., Hruban, C., White, J.R., et al. (2019). Genome-wide cell-free DNA fragmentation in patients with cancer. Nature *570*, 385–389. https://doi.org/10.1038/s41586-019-1272-6.

57. Ulz, P., Perakis, S., Zhou, Q., Moser, T., Belic, J., Lazzeri, I., Wölfler, A., Zebisch, A., Gerger, A., Pristauz, G., et al. (2019). Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. Nat. Commun. *10*, 4666. https://doi.org/10.1038/s41467-019-12714-4.

58. Mouliere, F., Chandrananda, D., Piskorz, A.M., Moore, E.K., Morris, J., Ahlborn, L.B., Mair, R., Goranova, T., Marass, F., Heider, K., et al. (2018). Enhanced detection of circulating tumor DNA by fragment size analysis. Sci. Transl. Med. *10*, eaat4921. https://doi.org/10.1126/scitranslmed.aat4921.

59. Jiang, P., Xie, T., Ding, S.C., Zhou, Z., Cheng, S.H., Chan, R.W.Y., Lee, W.-S., Peng, W., Wong, J., Wong, V.W.S., et al. (2020). Detection and characterization of jagged ends of double-stranded DNA in plasma. Genome Res. *30*, 1144–1153. https://doi.org/10.1101/gr.261396.120.

60. Shi, J., Zhang, R., Li, J., and Zhang, R. (2020). Size profile of cell-free DNA: a beacon guiding the practice and innovation of clinical testing. Theranostics *10*, 4737–4748. https://doi.org/10.7150/thno.42565.

61. Liu, Y. (2022). At the dawn: cell-free DNA fragmentomics and gene regulation. Br. J. Cancer *126*, 379–390. https://doi.org/10.1038/s41416-021-01635-z.

62. Underhill, H.R., Kitzman, J.O., Hellwig, S., Welker, N.C., Daza, R., Baker, D.N., Gligorich, K.M., Rostomily, R.C., Bronner, M.P., and Shendure, J. (2016). Fragment length of circulating tumor DNA. PLoS Genet. *12*, e1006162. https://doi.org/10.1371/journal.pgen.1006162.

63. Jiang, P., Sun, K., Tong, Y.K., Cheng, S.H., Cheng, T.H.T., Heung, M.M.S., Wong, J., Wong, V.W.S., Chan, H.L.Y., Chan, K.C.A., et al. (2018). Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA *115*, E10925–E10933. https://doi.org/10.1073/pnas.1814616115.

64. Chang, A., Mzava, O., Lenz, J.S., Cheng, A.P., Burnham, P., Motley, S.T., Bennett, C., Connelly, J.T., Dadhania, D.M., Suthanthiran, M., et al. (2021). Measurement biases distort cell-free DNA fragmentation profiles and define the sensitivity of metagenomic cell-free DNA sequencing assays. Clin. Chem. *68*, 163–171. https://doi.org/10.1093/clinchem/hvab142.

65. Chandrananda, D., Thorne, N.P., and Bahlo, M. (2015). High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. BMC Med. Genom. *8*, 29. https://doi.org/10.1186/s12920-015-0107-z.

66. De Vlaminck, I., Valantine, H.A., Snyder, T.M., Strehl, C., Cohen, G., Luikart, H., Neff, N.F., Okamoto, J., Bernstein, D., Weisshaar, D., et al. (2014). Circulating cell-free DNA enables noninvasive diagnosis of heart transplant rejection. Sci. Transl. Med. *6*, 241ra77. https://doi.org/10.1126/scitranslmed.3007803.

67. Meddeb, R., Dache, Z.A.A., Thezenas, S., Otandault, A., Tanos, R., Pastor, B., Sanchez, C., Azzi, J., Tousch, G., Azan, S., et al. (2019). Quantifying circulating cell-free DNA in humans. Sci. Rep. *9*, 5220. https://doi.org/10.1038/s41598-019-41593-4.

68. Pérez-Barrios, C., Nieto-Alcolado, I., Torrente, M., Jiménez-Sánchez, C., Calvo, V., Gutierrez-Sanz, L., Palka, M., Donoso-Navarro, E., Provencio, M., and Romero, A. (2016). Comparison of methods for circulating cell-free DNA isolation using blood from cancer patients: impact on biomarker testing. Transl. Lung Cancer Res. *5*, 665–672. https://doi.org/10.21037/tlcr.2016.12.03.

69. Ellinger, J., Bastian, P.J., Haan, K.I., Heukamp, L.C., Buettner, R., Fimmers, R., Mueller, S.C., and von Ruecker, A. (2008). Noncancerous PTGS2 DNA fragments of apoptotic origin in sera of prostate cancer patients qualify as diagnostic and prognostic indicators. Int. J. Cancer *122*, 138–143. https://doi.org/10.1002/ijc.23057.

70. Hauser, S., Zahalka, T., Ellinger, J., Fechner, G., Heukamp, L.C., VON Ruecker, A., Müller, S.C., and Bastian, P.J. (2010). Cell-free circulating DNA: diagnostic value in patients with renal cell cancer. Anticancer Res. *30*, 2785–2789.

71. Wang, B.G., Huang, H.-Y., Chen, Y.-C., Bristow, R.E., Kassauei, K., Cheng, C.-C., Roden, R., Sokoll, L.J., Chan, D.W., and Shih, I.-M. (2003). Increased plasma DNA integrity in cancer patients. Cancer Res. *63*, 3966–3968.

72. Mamon, H., Hader, C., Li, J., Wang, L., Kulke, M., Amicarelli, G., Shehi, E., Adlerstein, D., Roper, K., Killion, L., et al. (2008). Preferential amplification of apoptotic DNA from plasma: potential for enhancing detection of minor DNA alterations in circulating DNA. Clin. Chem. *54*, 1582–1584. https://doi.org/10.1373/clinchem.2008.104612.

73. Schmidt, B., Weickmann, S., Witt, C., and Fleischhacker, M. (2008). Integrity of cell-free plasma DNA in patients with lung cancer and nonmalignant lung disease. Ann. N. Y. Acad. Sci. *1137*, 207–213. https://doi.org/10.1196/annals.1448.034.

74. Umetani, N., Giuliano, A.E., Hiramatsu, S.H., Amersi, F., Nakagawa, T., Martino, S., and Hoon, D.S.B. (2006). Prediction of breast tumor progression by integrity of free circulating DNA in serum. J. Clin. Oncol. *24*, 4270–4276. https://doi.org/10.1200/JCO.2006.05.9493.

75. Deligezer, U., Eralp, Y., Akisik, E.E., Akisik, E.Z., Saip, P., Topuz, E., and Dalay, N. (2008). Size distribution of circulating cell-free DNA in sera of breast cancer patients in the course of adjuvant chemotherapy. Clin. Chem. Lab. Med. *46*, 311–317. https://doi.org/10.1515/CCLM.2008.080.

76. Meddeb, R., Pisareva, E., and Thierry, A.R. (2019). Guidelines for the preanalytical conditions for analyzing circulating cell-free DNA. Clin. Chem. *65*, 623–633. https://doi.org/10.1373/clinchem.2018.298323.

77. El Messaoudi, S., Rolet, F., Mouliere, F., and Thierry, A.R. (2013). Circulating cell free DNA: preanalytical considerations. Clin. Chim. Acta *424*, 222–230. https://doi.org/10.1016/j.cca.2013.05.022.

78. Chiu, R.W., Poon, L.L., Lau, T.K., Leung, T.N., Wong, E.M., and Lo, Y.M. (2001). Effects of blood-processing protocols on fetal and total DNA quantification in maternal plasma. Clin. Chem. *47*, 1607–1613.

79. Chiu, R.W.K., and Lo, Y.M.D. (2002). Application of fetal DNA in maternal plasma for noninvasive prenatal diagnosis. Expert Rev. Mol. Diagn. *2*, 32–40. https://doi.org/10.1586/14737159.2.1.32.

80. Jiang, P., Peng, X., Su, X., Sun, K., Yu, S.C.Y., Chu, W.I., Leung, T.Y., Sun, H., Chiu, R.W.K., Lo, Y.M.D., and Chan, K.C.A. (2016). FetalQuantSD: accurate quantification of fetal DNA fraction by shallow-depth sequencing of maternal plasma DNA. NPJ Genom. Med. *1*, 16013. https://doi.org/10.1038/npjgenmed.2016.13.

81. Roychowdhury, T., and Abyzov, A. (2019). Chromatin organization modulates the origin of heritable structural variations in human genome. Nucleic Acids Res. *47*, 2766–2777. https://doi.org/10.1093/nar/gkz103.

82. Audit, B., Zaghloul, L., Baker, A., Arneodo, A., Chen, C.-L., d'Aubenton-Carafa, Y., and Thermes, C. (2013). Megabase replication domains along the human genome: relation to chromatin structure and genome organisation. In Epigenetics: Development and Disease Subcellular Biochemistry, T.K. Kundu, ed. (Springer Netherlands), pp. 57–80. https://doi.org/10.1007/978-94-007-4525-4_3.

83. Hellwig, S., Nix, D.A., Gligorich, K.M., O'Shea, J.M., Thomas, A., Fuertes, C.L., Bhetariya, P.J., Marth, G.T., Bronner, M.P., and Underhill, H.R. (2018). Automated size selection for short cell-free DNA fragments enriches for circulating tumor DNA and improves error correction during next generation sequencing. PLoS One *13*, e0197333. https://doi.org/10.1371/journal.pone.0197333.

84. Mathios, D., Johansen, J.S., Cristiano, S., Medina, J.E., Phallen, J., Larsen, K.R., Bruhm, D.C., Niknafs, N., Ferreira, L., Adleff, V., et al. (2021). Detection and characterization of lung cancer using cell-free DNA fragmentomes. Nat. Commun. *12*, 5060. https://doi.org/10.1038/s41467-021-24994-w.

85. Sun, K., Jiang, P., Cheng, S.H., Cheng, T.H.T., Wong, J., Wong, V.W.S., Ng, S.S.M., Ma, B.B.Y., Leung, T.Y., Chan, S.L., et al. (2019). Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. Genome Res. *29*, 418–427. https://doi.org/10.1101/gr.242719.118.

86. Zhou, Z., Cheng, S.H., Ding, S.C., Heung, M.M.S., Xie, T., Cheng, T.H.T., Lam, W.K.J., Peng, W., Teoh, J.Y.C., Chiu, P.K.F., et al. (2021). Jagged ends of urinary cell-free DNA: characterization and feasibility assessment in bladder cancer detection. Clin. Chem. *67*, 621–630. https://doi.org/10.1093/clinchem/hvaa325.

87. Suzuki, N., Kamataki, A., Yamaki, J., and Homma, Y. (2008). Characterization of circulating DNA in healthy human plasma. Clin. Chim. Acta *387*, 55–58. https://doi.org/10.1016/j.cca.2007.09.001.

88. Peneder, P., Stütz, A.M., Surdez, D., Krumbholz, M., Semper, S., Chicard, M., Sheffield, N.C., Pierron, G., Lapouble, E., Tötzl, M., et al. (2021). Multimodal analysis of cell-free DNA whole-genome sequencing for pediatric cancers with low mutational burden. Nat. Commun. *12*, 3230. https://doi.org/10.1038/s41467-021-23445-w.

89. Tanos, R., Tosato, G., Otandault, A., Al Amir Dache, Z., Pique Lasorsa, L., Tousch, G., El Messaoudi, S., Meddeb, R., Diab Assaf, M., Ychou, M., et al. (2020). Machine learning-assisted evaluation of circulating DNA quantitative analysis for cancer screening. Adv. Sci. *7*, 2000486. https://doi.org/10.1002/advs.202000486.

90. Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature *474*, 516–520. https://doi.org/10.1038/nature10002.

91. Ulz, P., Thallinger, G.G., Auer, M., Graf, R., Kashofer, K., Jahn, S.W., Abete, L., Pristauz, G., Petru, E., Geigl, J.B., et al. (2016). Inferring expressed genes by whole-genome sequencing of plasma DNA. Nat. Genet. *48*, 1273–1278. https://doi.org/10.1038/ng.3648.

92. Zhu, G., Guo, Y.A., Ho, D., Poon, P., Poh, Z.W., Wong, P.M., Gan, A., Chang, M.M., Kleftogiannis, D., Lau, Y.T., et al. (2021). Tissue-specific cell-free DNA degradation quantifies circulating tumor DNA burden. Nat. Commun. *12*, 2229. https://doi.org/10.1038/s41467-021-22463-y.

93. Burnham, P., Kim, M.S., Agbor-Enoh, S., Luikart, H., Valantine, H.A., Khush, K.K., and De Vlaminck, I. (2016). Single-stranded DNA library

preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. Sci. Rep. *6*, 27859. https://doi.org/10.1038/srep27859.

94. Troll, C.J., Kapp, J., Rao, V., Harkins, K.M., Cole, C., Naughton, C., Morgan, J.M., Shapiro, B., and Green, R.E. (2019). A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. BMC Genom. *20*, 1023. https://doi.org/10.1186/s12864-019-6355-0.
95. Sharma, M., Verma, R.K., Kumar, S., and Kumar, V. (2022). Computational challenges in detection of cancer using cell-free DNA methylation. Comput. Struct. Biotechnol. J. *20*, 26–39. https://doi.org/10.1016/j.csbj.2021.12.001.
96. Wan, N., Weinberg, D., Liu, T.-Y., Niehaus, K., Ariazi, E.A., Delubac, D., Kannan, A., White, B., Bailey, M., Bertin, M., et al. (2019). Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. BMC Cancer *19*, 832. https://doi.org/10.1186/s12885-019-6003-8.
97. Wang, J., Chen, L., Zhang, X., Tong, Y., and Zheng, T. (2021). OCRDetector: accurately detecting open chromatin regions via plasma cell-free DNA sequencing data. Int. J. Mol. Sci. *22*, 5802. https://doi.org/10.3390/ijms22115802.
98. Kumar, P., Dillon, L.W., Shibata, Y., Jazaeri, A.A., Jones, D.R., and Dutta, A. (2017). Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. Mol. Cancer Res. *15*, 1197–1205. https://doi.org/10.1158/1541-7786.MCR-17-0095.
99. Han, D.S.C., Ni, M., Chan, R.W.Y., Chan, V.W.H., Lui, K.O., Chiu, R.W.K., and Lo, Y.M.D. (2020). The biology of cell-free DNA fragmentation and the roles of DNASE1, DNASE1L3, and DFFB. Am. J. Hum. Genet. *106*, 202–214. https://doi.org/10.1016/j.ajhg.2020.01.008.
100. Pisareva, E., Mihalovičová, L., Pastor, B., Kudriavstev, A., Mirandola, A., Mazard, T., Badiou, S., Maus, U., Ostermann, L., Weinmann-Menke, J., et al. (2022). Neutrophil extracellular traps have auto-catabolic activity and produce mononucleosome-associated circulating DNA. Molecular Biology. https://doi.org/10.1101/2022.09.01.506266.
101. Sadeh, R., Sharkia, I., Fialkoff, G., Rahat, A., Gutin, J., Chappleboim, A., Nitzan, M., Fox-Fisher, I., Neiman, D., Meler, G., et al. (2021). Author Correction: ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. Nat. Biotechnol. *39*, 642. https://doi.org/10.1038/s41587-021-00831-9.
102. Daniel, C., Leppkes, M., Muñoz, L.E., Schley, G., Schett, G., and Herrmann, M. (2019). Extracellular DNA traps in inflammation, injury and healing. Nat. Rev. Nephrol. *15*, 559–575. https://doi.org/10.1038/s41581-019-0163-2.
103. Benjamini, Y., and Speed, T.P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. *40*, e72. https://doi.org/10.1093/nar/gks001.
104. Chandrananda, D., Thorne, N.P., Ganesamoorthy, D., Bruno, D.L., Benjamini, Y., Speed, T.P., Slater, H.R., and Bahlo, M. (2014). Investigating and correcting plasma DNA sequencing coverage bias to enhance aneuploidy discovery. PLoS One *9*, e86993. https://doi.org/10.1371/journal.pone.0086993.
105. Zheng, H., Zhu, M.S., and Liu, Y. (2021). FinaleDB: a browser and database of cell-free DNA fragmentation patterns. Bioinformatics *37*, 2502–2503. https://doi.org/10.1093/bioinformatics/btaa999.
106. Aravanis, A.M., Lee, M., and Klausner, R.D. (2017). Next-Generation sequencing of circulating tumor DNA for early cancer detection. Cell *168*, 571–574. https://doi.org/10.1016/j.cell.2017.01.030.
107. Tanos, R., and Thierry, A.R. (2018). Clinical relevance of liquid biopsy for cancer screening. Transl. Cancer Res. *7*, S105–S129. https://doi.org/10.21037/tcr.2018.01.31.
108. Fiala, C., and Diamandis, E.P. (2018). Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection. BMC Med. *16*, 166. https://doi.org/10.1186/s12916-018-1157-9.
109. Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., Anagnostou, V., Fiksel, J., Cristiano, S., Papp, E., et al. (2017). Direct detection of early-stage cancers using circulating tumor DNA. Sci. Transl. Med. *9*, eaan2415. https://doi.org/10.1126/scitranslmed.aan2415.
110. Pons-Belda, O.D., Fernandez-Uriarte, A., and Diamandis, E.P. (2021). Can circulating tumor DNA support a successful screening test for early cancer detection? The grail paradigm. Diagnostics *11*, 2171. https://doi.org/10.3390/diagnostics11122171.
111. Klein, E.A., Richards, D., Cohn, A., Tummala, M., Lapham, R., Cosgrove, D., Chung, G., Clement, J., Gao, J., Hunkapiller, N., et al. (2021). Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann. Oncol. *32*, 1167–1177. https://doi.org/10.1016/j.annonc.2021.05.806.
112. Moss, J., Magenheim, J., Neiman, D., Zemmour, H., Loyfer, N., Korach, A., Samet, Y., Maoz, M., Druid, H., Arner, P., et al. (2018). Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat. Commun. *9*, 5068. https://doi.org/10.1038/s41467-018-07466-6.
113. Watanabe, T., Takada, S., and Mizuta, R. (2019). Cell-free DNA in blood circulation is generated by DNase1L3 and caspase-activated DNase. Biochem. Biophys. Res. Commun. *516*, 790–795. https://doi.org/10.1016/j.bbrc.2019.06.069.
114. Ma, X., Chen, Y., Tang, W., Bao, H., Mo, S., Liu, R., Wu, S., Bao, H., Li, Y., Zhang, L., et al. (2021). Multi-dimensional fragmentomic assay for ultrasensitive early detection of colorectal advanced adenoma and adenocarcinoma. J. Hematol. Oncol. *14*, 175. https://doi.org/10.1186/s13045-021-01189-w.
115. Syriopoulou, E., Gasparini, A., Humphreys, K., and Andersson, T.M.-L. (2022). Assessing lead time bias due to mammography screening on estimates of loss in life expectancy. Breast Cancer Res. *24*, 15. https://doi.org/10.1186/s13058-022-01505-3.
116. Wan, J.C.M., Heider, K., Gale, D., Murphy, S., Fisher, E., Mouliere, F., Ruiz-Valdepenas, A., Santonja, A., Morris, J., Chandrananda, D., et al. (2020). ctDNA monitoring using patient-specific sequencing and integration of variant reads. Sci. Transl. Med. *12*, eaaz8084. https://doi.org/10.1126/scitranslmed.aaz8084.
117. Cohen, J.D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., Douville, C., Javed, A.A., Wong, F., Mattox, A., et al. (2018). Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science *359*, 926–930. https://doi.org/10.1126/science.aar3247.
118. Chen, X., Gole, J., Gore, A., He, Q., Lu, M., Min, J., Yuan, Z., Yang, X., Jiang, Y., Zhang, T., et al. (2020). Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. Nat. Commun. *11*, 3475. https://doi.org/10.1038/s41467-020-17316-z.
119. Jensen, S.Ø., Øgaard, N., Ørntoft, M.B.W., Rasmussen, M.H., Bramsen, J.B., Kristensen, H., Mouritzen, P., Madsen, M.R., Madsen, A.H., Sunesen, K.G., et al. (2019). Novel DNA methylation biomarkers show high sensitivity and specificity for blood-based detection of colorectal cancer—a clinical biomarker discovery and validation study. Clin. Epigenetics *11*, 158. https://doi.org/10.1186/s13148-019-0757-3.
120. CCGA Consortium; Oxnard, G.R., Klein, E.A., Swanton, C., Seiden, M.V., Liu, M.C., Oxnard, G.R., Klein, E.A., Smith, D., Richards, D., et al. (2020). Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann. Oncol. *31*, 745–759. https://doi.org/10.1016/j.annonc.2020.02.011.
121. R Thierry, A. (2019). Method for screening a subject for a cancer WO - 11.02.2021 PCT/EP2020/071637.
122. Tivey, A., Church, M., Rothwell, D., Dive, C., and Cook, N. (2022). Circulating tumour DNA — looking beyond the blood. Nat. Rev. Clin. Oncol. *19*, 600–612. https://doi.org/10.1038/s41571-022-00660-y.