Review

Angela Oberhofer, Abel Jacobus Bronkhorst, Vida Ungerer and Stefan Holdenrieder*

Profiling disease and tissue-specific epigenetic signatures in cell-free DNA

https://doi.org/10.1515/labmed-2022-0031
Received March 8, 2022; accepted July 6, 2022;
published online August 2, 2022

Abstract: Programmed cell death, accidental cell degradation and active extrusion constantly lead to the release of DNA fragments into human body fluids from virtually all cell and tissue types. It is widely accepted that these cell-free DNA (cfDNA) molecules retain the cell-type specific genetic and epigenetic features. Particularly, cfDNA in plasma or serum has been utilized for molecular diagnostics. The current clinically implemented liquid biopsy approaches are mostly based on detecting genetic differences in cfDNA molecules from healthy and diseased cells. Their diagnostic potential is limited to pathologies involving genetic alterations, by the low proportion of cfDNA molecules carrying the mutation(s) relative to the total cfDNA pool, and by the detection limit of employed techniques. Recently, research efforts turned to epigenetic features of cfDNA molecules and found that the tissue-of-origin of individual cfDNA molecules can be inferred from epigenetic characteristics. Analysis of, e.g., methylation patterns, nucleosome or transcription factor binding site occupancies, fragment size distribution or fragment end motifs, and histone modifications determined the cell or tissue-of-origin of individual cfDNA molecules. With this tissue-of origin-analysis, it is possible to estimate the contributions of different tissues to the total cfDNA pool in body fluids and find tissues with increased cell death (pathologic condition), expanding the portfolio of liquid biopsies beyond genetics and towards a wide range of pathologies, such as autoimmune disorders, cardiovascular diseases, and inflammation, among many others. In this review, we give an overview on the status of tissue-of-origin approaches and focus on what is needed to exploit the full potential of liquid biopsies towards minimally invasive screening methods with broad clinical applications.

Keywords: cell-free DNA; circulating tumor DNA; epigenetics; fragmentomics; liquid biopsy; tissue-of-origin.

***Corresponding author: Stefan Holdenrieder**, Institute of Laboratory Medicine, German Heart Centre, Technical University Munich, Lazarettstr. 36, 80636 Munich, Germany,
E-mail: holdenrieder@dhm.mhn.de
Angela Oberhofer, Abel Jacobus Bronkhorst and Vida Ungerer, Institute of Laboratory Medicine, German Heart Centre, Technical University Munich, Munich, Germany

Introduction

DNA fragments from different origins are constantly released into various human body fluids, e.g., serum, plasma, urine, and cerebrospinal fluid [1–3]. Not only host DNA (genomic and mitochondrial DNA) can be found in the different body fluids, but exogenous DNA (e.g., bacterial or viral DNA) is also present [4–8]. The so-called cell-free DNA (cfDNA) in plasma is very short-lived, its half-time ranges between 15 min and 2.5 h [9, 10], and has been the subject of intense research efforts in order to develop minimally invasive methods for obtaining diagnostic information for e.g., early diagnosis, therapy monitoring, or detection of minimal residual disease in cancer. Specifically, non-invasive prenatal testing (NIPT) [11], detection of circulating tumor DNA (ctDNA) in the plasma of cancer patients for early diagnosis [12], and detection of donor-derived DNA in the plasma of transplantation recipients [13] have been clinically implemented so far. These analyses have in common that they are based on genetic differences (fetal and maternal DNA; donor and recipient DNA in graft patients) or mutations (cancer) that are only present in the minority of cfDNA molecules, limiting especially the sensitivity of those approaches by the technical limit of detection and to diseases involving genetic alterations. Early diagnosis of cancer is especially hampered by this limitation.

More recently, research groups focused on epigenetic characteristics of cfDNA molecules – that is the majority of cfDNA molecules – to develop complementary assays to detect diverse pathologies in the human body and expanding the detection to pathologies without genetic aberrations. It has, for example, been shown that methylation or fragmentation patterns of cfDNA molecules are cell-type and tissue-specific [14–17], as genes are differentially regulated in distinct cell types and different release

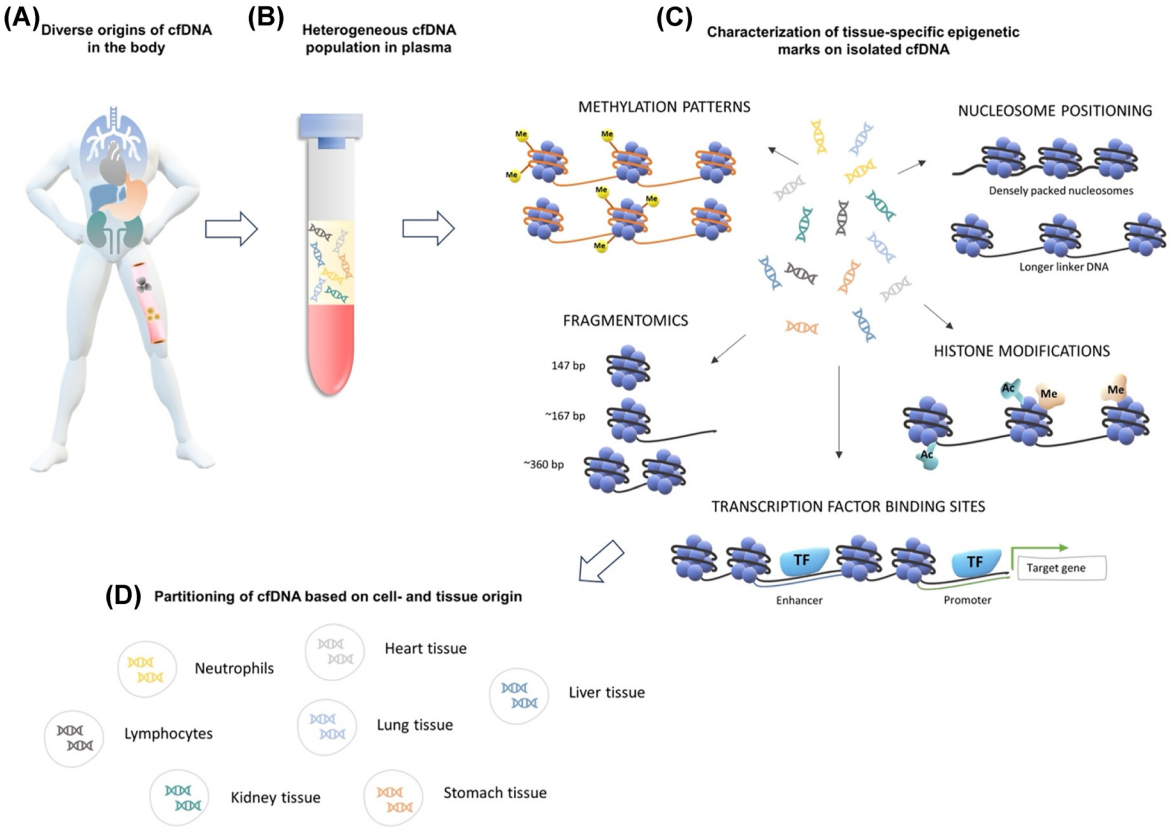

Figure 1: Determining the tissue-of-origin of cell-free DNA.
(A) Cell-free DNA (cfDNA) can be released into blood plasma from various cell types and different organs. (B) This results in a clinical biospecimen with a highly heterogeneous mixture of cfDNA molecules, often complicating the analytical differentiation between different cfDNA subtypes. (C) However, there are many tissue-specific physico-chemical and epigenetic features of cfDNA that can be characterized, including unique methylation patterns, fragmentation profiles and fragment end-points, transcription-factor binding sites, nucleosome positioning and occupancy, as well as post-translational histone modifications. (D) Comprehensive characterization of these tissue-specific markers may enable the determination of the tissue-of-origin of different cfDNA molecules. This may in turn, for example, facilitate the minimally-invasive localization of tumors or identification of tissue damage in specific regions such as the heart.

pathways are employed by distinct cell types. Deciphering these patterns and attributing them to the corresponding cell or tissue-of-origin opens up novel ways for noninvasive early diagnosis (Figure 1).

In plasma, cfDNA predominantly prevails as short double-stranded DNA fragments in the size ranging between 100 and 200 base pairs (bp). However, considerably longer fragments up to 30 kbp have been detected [18]. DNA is released into the circulation by different cellular pathways including different sorts of cell death, regular cellular turnover and active secretion upon pathologies [8, 19–21]. In diseased tissues, more cell death has been observed, and consequently, more cfDNA molecules from that particular cell type(s) are released into the blood compared to a healthy individual.

Lengths and fragment end motifs of cfDNA molecules are nonrandom and depend on the cellular release pathway and regulatory state of the releasing tissue [20, 22–24]. Investigating cfDNA fragmentomics is another way to pinpoint the origin of a cfDNA molecule. Different read-depth coverage patterns hint at transcription factor binding sites or expression status of a specific gene and aid in determining the tissue-of-origin.

Chromatin immunoprecipitation (ChIP) of specific histone modifications and subsequent sequencing of nucleosome-bound cfDNA yields genomic parts associated with transcriptionally active regions that inform on dynamic changes in transcriptional programs in the cells of origin that are pathology-related.

Numerous methods evolved during the last years to assign cfDNA molecules to their cell/tissue-of-origin based on epigenetic features such as DNA methylation patterns, nucleosome footprinting, transcription factor binding sites, fragmentation patterns, and histone modifications among others [14, 15, 17, 25–27]. The focus on the haystack (features found in the majority of cfDNA molecules) instead of the

needle (few mutated sites in only a handful of tumor-derived DNA molecules) holds promise to develop multiple powerful diagnostic tools in the future. In this review, we will discuss the currently available toolbox for liquid biopsies and the early detection of diseases based on the tissue-of-origin.

Mutation patterns

Genetic alterations (e.g., single nucleotide polymorphisms and copy number variations) are often involved in the onset of cancer [28]. These genetic differences can be detected by analyzing/sequencing the DNA from tumor tissue (solid biopsy), circulating tumor cells (CTCs) or ctDNA in plasma released into the circulation by tumor tissue. The first liquid biopsy approaches were based on genetic differences between healthy and cancerous tissue and therefore mostly detected a specific cancer type. Several mutation-based liquid biopsy assays have been approved by the FDA so far (e.g., EGFR mutations in non-small cell lung cancer and PIK3CA mutations in breast cancer) [29, 30]. These approaches are particularly powerful for selection and stratification of targeted therapies. However, mutation-based liquid biopsies are limited to pathologies involving genetic aberrations. For a long time, many clinically utilized organ-specific biomarkers have been employed for the diagnosis of various diseases, including elevated protein levels in different organs (e.g., troponin levels in the heart, transaminases in the liver, prostate-specific antigen in the prostate). Similarly, tumor markers have been applied for detection and monitoring of cancer disease. However, these biomarkers are not deployable for early diagnosis or detection of minimal residual disease.

A multi-analyte blood test (CancerSEEK) – combining mutational analysis in ctDNA with measuring protein levels in plasma – aimed at the earlier detection with higher sensitivity and specificity of pancreatic ductal adenocarcinoma [31]. This approach might be expanded for the screening of many cancer types (reviewed in [32]). To expand liquid biopsies additionally towards pathologies lacking genetic aberrations, the focus turned to epigenetic features of cfDNA molecules. Analysis of epigenetic features carried by the majority of cfDNA molecules and inferring the tissue-of-origin might enhance the performance of liquid biopsy tools in the future.

Methylation patterns

DNA methylation is an epigenetic mechanism where a methyl group is transferred onto the C5 position of the cytosine to form 5-methylcytosine (5mC). 5mC methylation occurs almost exclusively on cytosines with a neighboring guanosine – so-called CpG sites. This modification of the DNA in promoter regions regulates gene expression by inhibiting the binding of transcription factor(s) to DNA or by inhibiting the recruitment of proteins involved in gene repression. Consequently, methylated promoter regions usually correlate negatively with gene expression. Gene expression programs are tightly regulated by cell-specific DNA methylation patterns and these methylation marks correlate with the gene-expression profile of the respective cell type. Methylation patterns are unique to each cell type, are conserved among cells of the same cell type within an individual and among different individuals, and are highly stable under physiologic or pathologic conditions. Methylation patterns are changed in tumor cells, commonly leading to hypermethylation (i.e., silencing) of tumor suppressor genes and hypomethylation (i.e., activation of transcription) of cancer driver genes. The differentially methylated regions can be employed to detect prostate cancer, for example [33]. The methylation pattern can further be utilized to trace the cell or tissue-of-origin of a single cfDNA molecule, even of different cell types within a particular tissue [34].

With the advent of next-generation sequencing (NGS) technology in combination with bisulfite conversion, it became possible to determine the methylation state of each CpG in the entire human genome and compile high-resolution methylation maps of multiple tissue types [35–38]. This resulted in several projects that generated high resolution methylation or epigenetic maps, many of which are available as open access datasets, such as the Roadmap Epigenomics Project (35), the ENCODE Project [39, 40], the International Human Epigenome Consortium (IHEC) [41], and the Cancer Genome Atlas (TCGA; https://www.cancer.gov/tcga). The development of qualified algorithms for deconvolution of sequencing data with reference methylation profiles of different tissues enabled the determination of the cell or tissue-of-origin of cfDNA molecules and the estimation of major tissue contributors to the cfDNA pool. Analysis of tumor-associated hypomethylation in plasma cfDNA in combination with tumor-associated copy number aberrations (CNAs) using whole-genome bisulfite sequencing (WGBS) enabled the detection of several nonmetastatic cancer types with a sensitivity and specificity of 87 and 88%, respectively [42]. Sun et al. inferred relative contributions of four different tissues using deconvolution of cfDNA methylation profiles from WGBS [15]. This approach for mapping the tissue origins of plasma cfDNA using deep sequencing

demonstrated the tissue origins of cfDNA in conditions in which source tissue differs genetically from host tissue (i.e., pregnancy, transplantation, cancer). Additionally, it has been demonstrated that deconvolution of cfDNA methylation patterns revealed the tissue-of-origin in urinary cfDNA as well [43].

Superior sensitivity and reduced background in detecting tissue-specific signatures in cfDNA was gained by extending the analysis window from a single CpG site to a number of adjacent CpG haplotypes based on sparse genome coverage of methylation arrays [16]. These authors developed a method of detecting tissue-specific cell death in humans based on tissue-specific methylome databases to identify cell-type specific DNA methylation patterns. This approach was able to show origins of cfDNA in pathologies such as β-cell death in diabetes, brain cell death in multiple sclerosis and head trauma without genetically distinguishable tissue [16]. Single-cell reduced representation bisulfite sequencing (RRBS) was employed to identify regions of highly coordinated methylation (methylation haplotype blocks) that can be utilized in quantitative estimation of tumor load and tissue-of-origin mapping of cfDNA in patients with lung or colorectal cancer [44]. Identification of genomic loci that are unmethylated specifically in cardiomyocytes or hepatocytes, respectively, might serve as biomarkers for the detection of cardiac disease or liver damage [45, 46].

Immunoprecipitation of methylated cfDNA combined with high-throughput bisulfite-free sequencing (cfMeDIP-seq) enabled methylome analysis of small quantities of cfDNA and facilitated sensitive tumor detection and classification [47]. The enrichment of methylated cfDNA molecules marked an important step in methylation pattern analysis because it circumvents DNA degradation-prone bisulfite conversion.

Several probabilistic models such as CancerLocator or CancerDetector infer proportions and tissue-of-origin of tumor-derived cfDNA using genome-wide DNA methylation data or joint methylation states of multiple adjacent CpG sites to detect cancer with high sensitivity and specificity [48, 49]. With real plasma cfDNA samples, CancerLocator and CancerDetector reached a sensitivity of 74.4 and 94.8%, respectively, when specificity was 100%. The Pearson's correlation coefficient for both models were 0.975 and 0.9974, respectively.

To improve tissue-of-origin analysis, it is crucial to establish reference methylomes of key cell types and not only tissues that mostly represent mixtures of cell types. For this purpose, Moss et al. generated a reference atlas of 25 human tissues and cell types covering major organs and cells involved in common diseases and demonstrated that plasma methylation patterns can be used to accurately identify cell type-specific cfDNA in healthy and pathological conditions [50].

As an alternative approach, Tse et al. achieved the accurate detection of 5mC using single-molecule real-time sequencing (SMRT-seq) using the long-read sequencing technology PacBio [51]. This method takes advantage of the sequence context and pulse signals associated with DNA polymerase kinetics analyzed by a convolutional neural network to accurately detect 5mC modifications and might serve as an alternative technique for methylation pattern analysis. Most recently, a human methylome atlas based on deep WGBS and 39 cell types sorted from healthy tissue samples was completed [52].

Loci uniquely unmethylated in a specific cell type are often located at transcriptional enhancers and contain DNA binding sites for tissue-specific transcriptional regulators, whereas uniquely hyper-methylated loci are rare and enriched for CpG islands, polycomb targets and CTCF binding sites [52]. The authors developed a computational machine learning suite to represent, compress, visualize and analyze WGBS data (available at: https://github.com/nloyfer/wgbs_tools) [52]. One clinical validation study based on targeted methylation analysis of cfDNA [53, 54] reported an overall sensitivity of 51.5% and an overall specificity of 99.5% for cancer type prediction [55].

Besides the widely studied 5mC modification, 5-hydroxymethylcytosine (5hmC) affects a wide range of biological processes from development to pathogenesis [56, 57]. 5hmC is converted from 5mC by the ten-eleven translocation (TET) family dioxygenases and is generally thought to reflect gene activation on permissive chromatin [58]. This intermediate DNA modification is enriched in enhancers, gene bodies and promoters and changes in 5hmC correlate with changes in gene expression levels [59, 60]. 5hmC additionally displays a tissue-specific mass distribution [61, 62] and decreased levels of 5hmC are often observed in many solid tumors compared to corresponding healthy tissues [63]. Several studies have utilized 5hmC signatures in cfDNA to detect cancer type and stage [64–66]. One study demonstrated that PDAC tissue-derived hyper-hydroxymethylated genes can separate non-cancer cfDNA from PDAC cfDNA samples [64]; making hydroxymethylation patterns another promising approach for tissue-of-origin analysis.

Overall, WGBS-based methods and deconvolution algorithms paved the way for methylation pattern analysis to determine the tissue-of-origin of cfDNA. On the one hand, those approaches were utilized to put together high-resolution methylation profiles of different reference tissues or cell types and on the other hand were already

able to sensitively detect some cancer types and several other pathologies. With the development of bisulfite-free techniques such as cfMeDIP or the investigation of the hydroxymethylome, the potential of epigenetic-based liquid biopsies might be exploited further.

Nucleosome footprinting

In addition to DNA methylation, gene expression is also regulated by the spacing of nucleosomes that are responsible for condensing DNA into chromatin. The basic unit of DNA compaction is the nucleosome, which is composed of a histone protein octamer with ~147 bp of DNA wrapped around it. The fifth histone H1 is located outside the nucleosome core and binds ~20 bp of linker DNA. Nucleosomes are repeated throughout the entire genome, each separated by unwrapped linker DNA of varying lengths. Nucleosomes govern the accessibility of DNA by occluding proteins involved in gene regulation and transcription from binding. On the other hand, chromatin opening and movement of nucleosome from specific genome regions is essential for binding of transcription factors. The gene expression of distinct cell types differs significantly and consequently, nucleosome positions vary considerably between distinct cell types [14].

During the release of cfDNA into a body fluid, the nuclear DNA is cleaved by different enzymes depending on the release mechanism. Nucleosome-bound DNA is better protected against cleavage than linker DNA or open chromatin regions lacking nucleosomes. Consequently, nucleosome-bound regions are expected to generate more reads when sequencing the plasma DNA pool, whereas nucleosome-depleted regions will be underrepresented in sequencing reads. As a result, uncovered DNA sites represent transcription-prone regions, whereas higher coverage indicates lower expression levels of this genomic region. Thus sequence read-density across the genome reflects nucleosome positioning, which is in turn informative about gene expression and is highly related to cell identity [67] and can be employed to trace the tissue-of-origin of cfDNA molecules.

Micrococcal nuclease (MNase) is an endo-exonuclease that preferentially digests unprotected DNA (i.e., accessible linker DNA between nucleosomes) and mostly leaves nucleosome-protected DNA intact. MNase was utilized to identify nucleosome occupied regions of DNA [68]. Most nucleosome mapping experiments and established nucleosome position maps until now are based on MNase assays. MNase treatment followed by sequencing showed that promoters of highly expressed genes contain pronounced nucleosome-depleted regions (NDRs) directly upstream of the transcriptional start site (TSS) [69–71]. Additionally, transcription termination sites (TTSs) were found to be nucleosome depleted as well [70]. In contrast, promoters of less expressed or silent genes are mostly occupied by nucleosomes and pronounced NDRs are not detected at sites with low or no transcription.

Deep sequencing of total cfDNA extended the available information on NDRs and yielded a genome-wide map of *in vivo* nucleosome occupancy [14]. The authors observed that short cfDNA fragments directly footprinted transcription factor occupancy and found that the nucleosome spacing pattern could be used to infer the tissue origin of cfDNA. In healthy individuals, these epigenetic footprints match hematopoietic lineage, whereas additional contributions are detected in cancer patients, often aligning with the cancer type [14]. Ulz et al. reported that plasma DNA coverage in the promoters could be used to predict the expression of genes [72]. WGS of plasma and nucleosome promoter analysis using machine learning for gene classification identified two discrete regions at TSSs where nucleosome occupancy results in different read depth coverage patterns for expressed and silent genes [72]. This allowed the authors to classify expressed cancer driver genes in regions with somatic copy number gains in patients with metastatic cancer [72].

A systematic nucleosomics database (NucPosDB) is now available that curates published nucleosome positioning datasets *in vivo* [73]. It also includes datasets of sequenced cfDNA that reflect nucleosome positioning *in situ* in the cells of origins [73]. Additionally, NucPosDB lists computational tools for the analysis of nucleosome positioning or cfDNA experiments and contains theoretical algorithms for the prediction of nucleosome positioning preferences from DNA sequence.

However, studies that investigated the origin of plasma cfDNA in a genome-wide and tissue-wide manner came to varying results [14, 15, 74–76]. They all found that white blood cells contributed most to the plasma cfDNA pool but with varying percentages. Further, the contributions of other organs (e.g., liver, heart, lung tissue) are still poorly understood. Overall, cfDNA sequence read coverage holds abundant information about nucleosome positioning and a sensitive tissue-of-origin analysis might be performed using this approach for multiple pathologies.

Transcription factor binding sites

Transcription factors (TFs) play a pivotal role in the regulation of gene expression. They fine-tune the expression of

their target genes and are key players in development and differentiation [77]. Deregulated TFs are involved in the emergence of cancer by not suppressing the expression of cancer driver genes correctly or by mistakenly silencing tumor suppressor genes. Distinct TFs govern gene expression in different cell types. Identifying occupied transcription factor binding sites (TFBS) within the plasma DNA will open up an additional perspective for a more dynamic analysis of gene expression and tissue-of-origin.

So far, two studies dealt with inferring TFBSs from cfDNA fragmentation patterns. Snyder et al. observed that short cfDNA fragments (35–80 bp) directly footprint the *in vivo* occupancy of TFBSs by CTCF and seven other TFs [14]. Specifically, they used single-stranded libraries that included significantly more short cfDNA molecules than double-stranded libraries and determined whether a TFBS was occupied or not by calculating the windowed protection score (WPS) at a given genomic coordinate. The WPS is the number of DNA fragments completely spanning a 16 bp window centered at a given genomic coordinate minus the number of fragments with an endpoint within that same window [14]. With this approach, they were able to infer additional contributing tissues in non-healthy states.

In contrast, another study included hundreds of TFs with 1,000 TFBSs per TF from publicly available ATAC-seq data and from plasma cfDNA WGS data to infer accessibility of TFBSs from cfDNA fragmentation patterns. For this purpose, they developed an accessibility score that measures the strength of nucleosome phasing at the TFBS, reflecting the strength of the TF binding [27]. Briefly, the raw coverage signal was split by Savitzky-Golay filtering into a high and low frequency signal. The rank differences (i.e., overall *z*-scores) in this high frequency signal between tumor and healthy samples with defined thresholds for TFBS accessibility differences was used as the accessibility score. With this approach, the authors were able to profile numerous individual TFs and objectively compare TF binding events in plasma samples. Thereby, a dynamic view on TF activity was possible which allowed subclassification of tumor entities and TFBS plasticity during disease progression [27].

Taken together, cfDNA fragmentation patterns do not only give information about nucleosome occupancy, but also offer molecular insights on TFBS occupancy, allowing analysis of gene expression and determination of the tissue-of-origin of different cfDNA molecules. This minimally-invasive method might substantially add to a more dynamic picture on recent transcriptional events during disease progression and therapy monitoring.

Fragmentomics

It has become increasingly clear that cfDNA size distribution is nonrandom and dictated by the cell origin, the release pathway, and nucleases present in the blood [78, 79]. The generation of plasma DNA is related to cell death (e.g., apoptosis and necrosis among others) and most probably involves enzymatic degradation processes [8]. Differences in tissue expression of involved nuclease enzymes might influence the end motif profiles of resultant cfDNA released by the corresponding tissue and offer insights on the tissue-of-origin. Protein-bound DNA fragments, typically associated with histones or TFs, preferentially survive digestion and are released into the blood, while naked DNA is digested and is not detectable in cfDNA analysis.

Commonly, the majority of cfDNA molecule sizes peak at the DNA length wrapped around one nucleosome (i.e., 167 bp) [14, 17, 25]. A series of additional peaks with ~10 bp periodicity starting from ~143 bp downward putatively correspond to the helical pitch and binding sites of DNA to the nucleosome core [14]. It has also been shown that fragment size of cfDNA is influenced by pathologies, e.g., cfDNA from cancer patients tends to be a bit shorter than cfDNA from healthy individuals (147 vs. 167 bp) [17, 25]. Generally, the lengths of cancer-derived cfDNA fragments tend to be more variable than non-cancer DNA and mutant ctDNA is more fragmented than nonmutant cfDNA [17]. Cancer, pathologies or injuries all have in common that they induce increased cell death in the affected tissue, leading to elevated levels of cfDNA molecules to the plasma DNA pool from the corresponding tissue. Tissue-of-origin analysis based on fragmentation patterns might reveal elevated contributions to the plasma DNA pool and enable sensitive detection of pathologies lacking genetic differences, such as myocardial infarction, stroke and autoimmune disorders.

Ivanov et al. investigated cfDNA fragmentation patterns for the first time and demonstrated that cfDNA carries a nonrandom fragmentation pattern and retains characteristics previously found in genome-wide analysis of chromatin structure and are concordant with corresponding cell-line derived patterns [80], demonstrating the potential of fragmentation pattern analysis of cfDNA as novel diagnostic biomarkers. An early study investigated cfDNA end characteristics in addition to cancer-associated somatic mutations [24]. The authors demonstrated the existence of tumor-associated cfDNA preferred end-coordinates; specifically, they found distinct preferred end coordinates of cfDNA derived from transplanted liver,

hepatocellular carcinoma, or the placenta. Quantitative assessment of cfDNA molecules bearing respective groups of end signatures correlated with the amounts of tumor- or liver-derived cfDNA in plasma [24].

The establishment of a pan-cancer catalog of ctDNA fragmentation features identified characteristic differences in the size distribution of tumor-derived and noncancer DNA fragments [17], where mutant ctDNA is generally more fragmented than nonmutant cfDNA. These observations guided the way for designing a machine-learning-based method of tumor DNA detection with greater sensitivity that is based on *in vitro* selection of short cfDNA fragments (90–150 bp), shallow whole-genome sequencing (sWGS), and a machine-learning algorithm that is able to detect multiple cancer types in plasma [17]. Fragmentation pattern analysis has also been applied to cfDNA from cerebrospinal fluid (CFS) from glioma patients, demonstrating that the fragmentation pattern of cfDNA in CFS is different from that in plasma [81].

Sun et al. developed an orientation-aware plasma DNA fragmentation analysis that included – in addition to fragmentation pattern analysis measured by sequence coverage imbalance – the orientation of cfDNA fragments (i.e., upstream or downstream fragment end profile) [82]. This enabled the authors to determine not only nucleosome positions, but also identify short linker DNA and tissue-specific open chromatin regions. This, in turn, allowed the measurement of the relative contribution of various tissues toward the plasma DNA pool for tissue-of-origin analysis.

A different approach named DNA evaluation of fragments for early interception (DELFI) considered large-scale cfDNA fragmentation patterns at megabase level in patients with cancer and observed altered fragmentation patterns in cancer samples [25]. A machine learning model that incorporated genome-wide fragmentation features was able to identify the tissue-of-origin of distinct cancers to a limited number of sites in 75% of cases. Combining this approach with mutation-based cfDNA analyses detected 91% of patients with cancer [25].

Recently, a genome-wide catalogue of cfDNA fragment end sequence patterns of a large cohort of cancer patients was published [83]. Specifically, the work focused on the diversity of bases at the end of cfDNA fragments (cfDNA termini) and converted the fragment end sequences into a quantitative metric – the Fragment End Integrated Analysis (FrEIA) score. The authors showed that fragment-end sequence and diversity were altered in 18 different cancer types using low-coverage whole genome sequencing and were able to classify cancer samples from controls at low tumor content.

Overall, several different approaches have been developed during the last years that inferred cell or tissue-of-origin from cfDNA fragment size distribution, fragment preferred end coordinates, and fragment end sequence patterns (also reviewed in Ref. [84]). All of these techniques rely on whole-genome sequencing and sophisticated machine learning models and were validated in proof-of-concept studies with a limited number of samples, yet with promising results regarding early detection of cancer and several other pathologies.

Histone modifications

Histone modifications alter the interaction between DNA and nuclear proteins and thereby regulate chromatin accessibility and gene transcription. Histone proteins can be post-translationally modified in multiple ways, including methylation, acetylation, and ubiquitination [85]. For example, the histone modifications H4K16ac, H3K4me1/2/3, and H3K36me1/2/3 represent activating marks and are considered hallmarks of open chromatin where active transcription takes places, whereas the histone modifications H3K9me3 and H3K27me3 have a repressive effect on transcription and mark regions of closed chromatin [85].

Histone modification patterns reflect recent events related to chromatin regulation and activity of RNA polymerase. Different combinations of such histone modifications mark accessible/active promoters, enhancers and gene bodies of actively transcribed genes [86–91]. Therefore, histone modifications can give insights on changes in transcriptional programs in cells upon pathologies, as transcription is often altered at the onset of disease. Recently, chromatin immunoprecipitation and sequencing of cell-free nucleosomes from human plasma (cfChIP-seq) demonstrated that plasma nucleosomes retain the epigenetic information of their cells of origin and that cfChIP-seq recapitulates the original genomic distribution of modifications associated with active transcription [26].

Briefly, chromatin immunoprecipitation was performed with ChIP antibodies (specific for different histone modifications) immobilized on paramagnetic beads directly in plasma, followed by on-bead adaptor ligation before DNA isolation. The authors employed four antibodies specifically marking accessible/active promoters (H3K4me3 or H3K4me2), enhancers (H3K4me2) and gene bodies of actively transcribed genes (H3K36me3). CfChIP-seq allowed genome-wide unbiased analysis and was capable of determining the tissue-of-origin and detecting differences in patient- and disease-specific

transcriptional programs (including cancer-specific signatures) by generating biologically relevant reduced representation of the genome.

Another study demonstrated that H3K36me3 cfChIP followed by droplet digital PCR can be used to identify tumor-specific transcriptional activity of the mutated *EGFR-L858R* allele in non-small cell lung cancer [92]. Specifically, blood plasma cfChIP results revealed active transcription of *EGFR-L858R* in NSCLC tumors. This focus on tumor-specific transcriptional activity of genes harboring somatic mutations will help to gain more insights on the revelance of mutations in, e.g., therapy resistance mechanisms.

In contrast to stable genomic alterations or DNA methylation, cfChIP-seq offers insights into transient changes in gene expression altered upon various pathologies and during disease progression. However, its clinical utility remains to be demonstrated.

Summary and outlook

Tissue-of-origin analyses developed during the last couple of years by numerous groups expanded cfDNA research from focusing on stable genetic aberrations exhibited by a very limited number of cfDNA molecules towards epigenetic characteristics observed on a higher number of molecules in the plasma cfDNA pool. Many proof-of-concept studies were published recently, demonstrating the potential of these novel approaches. Most methods are based on targeted approaches or enrichment of targets through antibody- or chemical affinity-based techniques, leading to lower sequencing costs by reduced sequencing depth, higher sensitivity, and lower background.

The continuous development of methods, for example the entirely enzymatic conversion for mapping 5mC and 5hmC methylation [93, 94] instead of bisulfite conversion, may pave the way for detecting methylation patterns more sensitively. Measuring two distinct epigenetic cfDNA features in a single assay might enhance the value of liquid biopsies further. Recently, Erger and colleagues presented an approach for measuring cytosine methylation and nucleosome occupancy simultaneously [95, 96]. This might be of special interest, as most pathologies are not caused by a single risk factor, but by multiple distinct risk factors.

Lastly, combination of different liquid biopsy approaches significantly increased sensitivity and specificity; for instance, analyzing mutations in cfDNA and levels of circulating proteins in plasma in the same sample [97]. Fragmentation profile-based tissue-of-origin analysis combined with mutational cfDNA analysis, for example, elevated the sensitivity of the combined approach in patients with cancer from 75% (fragmentation profile analysis only) to 91% [25]. This clearly emphasizes the potential of liquid biopsy diagnostics by combining different approaches, but simultaneously demonstrates the need for further development.

However, several limitations still need to be overcome. First, despite the availability of many proof-of-principle studies, validation of the methods in larger multicenter studies is still lacking. Second, the significant influence of the biology of DNA release from different cells and preanalytics on cfDNA yield as well as size distribution deserves further attention and streamlining to assure high quality of analyses [98]. The community needs to define guidelines for handling cfDNA samples from blood draw until analysis. Lastly, the presence of high quantities of non-tumor cfDNA in cancer patients necessitates an extensive bioinformatic expertise. There is a great need for well-documented and well-maintained open-source bioinformatic packages for the analysis of cfDNA epigenetic features. Streamlining bioinformatics pipelines will achieve comparability between different studies that is not given so far. Different machine-learning models need to be tested on independent datasets to be better evaluated. Further, it will be important to integrate multiple cfDNA features from one dataset. Finally, open access to raw cfDNA NGS data is urgently needed. First efforts in solving these issues were started recently by publication of e.g. cfDNApipe, which is a comprehensive quality control and analysis pipeline for high-throughput cfDNA sequencing data [99], Finale DB [100] – a browser and database of cfDNA fragmentation patterns, or CFEA database [101], a comprehensive public database with cfDNA-based epigenome profiles for 27 human diseases. A reproducible bioinformatics pipeline has uniformly processed these data [101].

The substantial advancements in cfDNA analysis during the last decade extended the focus from genetic alterations to epigenetic cfDNA characteristics. With this shift towards more general marks on a higher number of cfDNA molecules, it was possible to determine the cell or tissue-of-origin of individual cfDNA molecules and estimate contributions of different tissues to the plasma DNA. This greatly expanded potential clinical applications of liquid biopsies in early detection of cancer and other pathologies, as well as therapy and minimal residual disease monitoring. Mutation-based liquid biopsy is already clinically implemented as companion diagnostics for numerous cancer types, aiding selection of targeted therapy and detection of therapy resistance; and epigenetic-based liquid biopsies might enable earlier detection of pathologies and minimal residual disease monitoring.

Importantly, this opens the door to more comprehensive tissue diagnostics beyond genetics, making it possible to detect all kinds of different organ pathologies as well as systematic inflammation processes. Combining tissue-of-origin analysis with mutational analysis might further boost its performance and expand its application to yet uncovered aspects of pathology. Given the many distinct approaches developed so far, it will be interesting to see which method or combined approach will make its way into broad application after further development and validation in larger studies.

In the future, it will be of great interest to investigate the interaction of cfDNA with circulating cells (e.g., neutrophils, platelets) and gain basic insights on cfDNA biology and function. Taken together, tissue-of-origin analyses in plasma samples based on epigenetic features hold great promise to add to a versatile diagnostic toolbox. Thereby, the needle in the haystack is as relevant as the haystack itself.

Research funding: None declared.
Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.
Competing interests: Authors state no conflict of interest.
Informed consent: Not applicable.
Ethical approval: Not applicable.

References

1. Hui L, Maron JL, Gahan PB. Other body fluids as non-invasive sources of cell-free DNA/RNA. In: Gahan PB, editor. Circulating nucleic acids in early diagnosis, prognosis and treatment monitoring advances in predictive, preventive and personalised medicine. Dordrecht: Springer; 2015, vol 5.
2. Mandel P, Metais P. Nuclear acids in human blood plasma. C R Seances Soc Biol Fil 1948;142:241–3.
3. Bronkhorst AJ, Ungerer V, Holdenrieder S. The emerging role of cell-free DNA as a molecular marker for cancer management. Biomol Detect Quantification 2019;17:100087.
4. Kohler C, Radpour R, Barekati Z, Asadollahi R, Bitzer J, Wight E, et al. Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors. Mol Cancer 2009;8:105.
5. Zachariah RR, Schmid S, Buerki N, Radpour R, Holzgreve W, Zhong X. Levels of circulating cell-free nuclear and mitochondrial DNA in benign and malignant ovarian tumors. Obstet Gynecol 2008;112:843–50.
6. Kowarsky M, Camunas-Soler J, Kertesz M, De Vlaminck I, Koh W, Pan W, et al. Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. Proc Natl Acad Sci USA 2017;114: 9623–8.
7. Long Y, Zhang Y, Gong Y, Sun R, Su L, Lin X, et al. Diagnosis of sepsis with cell-free DNA by next-generation sequencing technology in ICU patients. Arch Med Res 2016;47:365–71.
8. Grabuschnig S, Bronkhorst AJ, Holdenrieder S, Rosales Rodriguez I, Schliep KP, Schwendenwein D, et al. Putative origins of cell-free DNA in humans: a review of active and passive nucleic acid release mechanisms. Int J Mol Sci 2020; 21:8062.
9. To EW, Chan KC, Leung SF, Chan LY, To KF, Chan AT, et al. Rapid clearance of plasma Epstein-Barr virus DNA after surgical treatment of nasopharyngeal carcinoma. Clin Cancer Res 2003; 9:3254–9.
10. Yao W, Mei C, Nan X, Hui L. Evaluation and comparison of in vitro degradation kinetics of DNA in serum, urine and saliva: a qualitative study. Gene 2016;590:142–8.
11. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. Lancet 1997;350:485–7.
12. Nawroz H, Koch W, Anker P, Stroun M, Sidransky D. Microsatellite alterations in serum DNA of head and neck cancer patients. Nat Med 1996;2:1035–7.
13. Lo YM, Tein MS, Pang CC, Yeung CK, Tong KL, Hjelm NM. Presence of donor-specific DNA in plasma of kidney and liver-transplant recipients. Lancet 1998;351:1329–30.
14. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell 2016;164:57–68.
15. Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci USA 2015;112: E5503–12.
16. Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci USA 2016;113: E1826–34.
17. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med 2018;10:eaat4921.
18. Yu SCY, Jiang P, Peng W, Cheng SH, Cheung YTT, Tse OYO, et al. Single-molecule sequencing reveals a large population of long cell-free DNA molecules in maternal plasma. Proc Natl Acad Sci USA 2021;118:e2114937118.
19. Delgado PO, Alves BC, Gehrke Fde S, Kuniyoshi RK, Wroclavski ML, Del Giglio A, et al. Characterization of cell-free circulating DNA in plasma in patients with prostate cancer. Tumor Biol 2013;34:983–6.
20. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 2001;61:1659–65.
21. Sai S, Ichikawa D, Tomita H, Ikoma D, Tani N, Ikoma H, et al. Quantification of plasma cell-free DNA in patients with gastric cancer. Anticancer Res 2007;27:2747–51.
22. Chandrananda D, Thorne NP, Bahlo M. High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. BMC Med Genom 2015;8:29.
23. Han DSC, Ni M, Chan RWY, Chan VWH, Lui KO, Chiu RWK, et al. The biology of cell-free DNA fragmentation and the roles of

DNASE1, DNASE1L3, and DFFB. Am J Hum Genet 2020;106:202–14.
24. Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, Heung MMS, et al. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. Proc Natl Acad Sci USA 2018;115:E10925–33.
25. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature 2019;570:385–9.
26. Sadeh R, Sharkia I, Fialkoff G, Rahat A, Gutin J, Chappleboim A, et al. ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. Nat Biotechnol 2021;39:586–98.
27. Ulz P, Perakis S, Zhou Q, Moser T, Belic J, Lazzeri I, et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. Nat Commun 2019;10:4666.
28. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009;458:719–24.
29. Sacher AG, Paweletz C, Dahlberg SE, Alden RS, O'Connell A, Feeney N, et al. Prospective validation of rapid plasma genotyping for the detection of EGFR and KRAS mutations in advanced lung cancer. JAMA Oncol 2016;2:1014–22.
30. Andre F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, et al. Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer. N Engl J Med 2019;380:1929–40.
31. Cohen JD, Javed AA, Thoburn C, Wong F, Tie J, Gibbs P, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. Proc Natl Acad Sci USA 2017;114:10202–7.
32. Holdenrieder S, Ungerer V, Oberhofer A, Bronkhorst AJ. Pan-cancer screening by circulating tumor DNA (ctDNA) – recent breakthroughs and chronic pitfalls. J Lab Med 2022;46:247–53.
33. Ellinger J, Haan K, Heukamp LC, Kahl P, Buttner R, Muller SC, et al. CpG island hypermethylation in cell-free serum DNA identifies patients with localized prostate cancer. Prostate 2008;68:42–9.
34. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinf 2012;13:86.
35. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–30.
36. Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell 2011;44:17–28.
37. Lun FM, Chiu RW, Sun K, Leung TY, Jiang P, Chan KC, et al. Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. Clin Chem 2013;59:1583–94.
38. Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, et al. A DNA methylation fingerprint of 1628 human samples. Genome Res 2012;22:407–19.
39. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57–74.
40. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res 2018;46:D794–801.
41. Bujold D, Morais DAL, Gauthier C, Cote C, Caron M, Kwan T, et al. The international human epigenome consortium data portal. Cell Syst 2016;3:496–9.e2.
42. Chan KC, Jiang P, Chan CW, Sun K, Wong J, Hui EP, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. Proc Natl Acad Sci USA 2013;110:18761–8.
43. Cheng THT, Jiang P, Tam JCW, Sun X, Lee WS, Yu SCY, et al. Genomewide bisulfite sequencing reveals the origin and time-dependent fragmentation of urinary cfDNA. Clin Biochem 2017;50:496–501.
44. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. Nat Genet 2017;49:635–42.
45. Lehmann-Werman R, Magenheim J, Moss J, Neiman D, Abraham O, Piyanzin S, et al. Monitoring liver damage using hepatocyte-specific methylation markers in cell-free circulating DNA. JCI Insight 2018;3:e120687.
46. Zemmour H, Planer D, Magenheim J, Moss J, Neiman D, Gilon D, et al. Non-invasive detection of human cardiomyocyte death using methylation patterns of circulating DNA. Nat Commun 2018;9:1443.
47. Shen SY, Singhania R, Fehringer G, Chakravarthy A, Roehrl MHA, Chadwick D, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature 2018;563:579–83.
48. Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, et al. CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol 2017;18:53.
49. Li W, Li Q, Kang S, Same M, Zhou Y, Sun C, et al. CancerDetector: ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. Nucleic Acids Res 2018;46:e89.
50. Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat Commun 2018;9:5068.
51. Tse OYO, Jiang P, Cheng SH, Peng W, Shang H, Wong J, et al. Genome-wide detection of cytosine methylation by single molecule real-time sequencing. Proc Natl Acad Sci USA 2021;118:e2019768118.
52. Loyfer N, Magenheim J, Peretz A, Cann G, Bredno J, Klochendler A, et al. A human DNA methylation atlas reveals principles of cell type-specific methylation and identifies thousands of cell type-specific regulatory elements. bioRxiv 2022. https://doi.org/10.1101/2022.01.24.477547.
53. Chen X, Dong Z, Hubbell E, Kurtzman KN, Oxnard GR, Venn O, et al. Prognostic significance of blood-based multi-cancer detection in plasma cell-free DNA. Clin Cancer Res 2021;27:4221–9.
54. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Consortium C. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol 2020;31:745–59.
55. Klein EA, Richards D, Cohn A, Tummala M, Lapham R, Cosgrove D, et al. Clinical validation of a targeted methylation-

based multi-cancer early detection test using an independent validation set. Ann Oncol 2021;32:1167–77.
56. Shen L, Song CX, He C, Zhang Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. Annu Rev Biochem 2014;83:585–614.
57. Vasanthakumar A, Godley LA. 5-hydroxymethylcytosine in cancer: significance in diagnosis and therapy. Cancer Genet 2015;208:167–77.
58. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5 hmC enriched within active genes and accessible chromatin in the nervous system. Cell 2012;151: 1417–30.
59. Han D, Lu X, Shih AH, Nie J, You Q, Xu MM, et al. A highly sensitive and robust method for genome-wide 5 hmC profiling of rare cell populations. Mol Cell 2016;63:711–9.
60. Konstandin N, Bultmann S, Szwagierczak A, Dufour A, Ksienzyk B, Schneider F, et al. Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. Leukemia 2011;25:1649–52.
61. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 2010;5:e15367.
62. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 2009;324:929–30.
63. Jin SG, Jiang Y, Qiu R, Rauch TA, Wang Y, Schackert G, et al. 5-hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer Res 2011;71:7360–5.
64. Guler GD, Ning Y, Ku CJ, Phillips T, McCarthy E, Ellison CK, et al. Detection of early stage pancreatic cancer using 5-hydroxymethylcytosine signatures in circulating cell free DNA. Nat Commun 2020;11:5270.
65. Li W, Zhang X, Lu X, You L, Song Y, Luo Z, et al. 5-hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. Cell Res 2017;27: 1243–57.
66. Song CX, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, et al. 5-hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. Cell Res 2017;27: 1231–42.
67. Radman-Livaja M, Rando OJ. Nucleosome positioning: how is it established, and why does it matter? Dev Biol 2010;339: 258–66.
68. Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. Epigenet Chromatin 2014;7:33.
69. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, et al. Dynamic regulation of nucleosome positioning in the human genome. Cell 2008;132:887–98.
70. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. Determinants of nucleosome organization in primary human cells. Nature 2011;474:516–20.
71. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. Nat Rev Mol Cell Biol 2015;16:178–89.
72. Ulz P, Thallinger GG, Auer M, Graf R, Kashofer K, Jahn SW, et al. Inferring expressed genes by whole-genome sequencing of plasma DNA. Nat Genet 2016;48:1273–8.
73. Shtumpf M, Piroeva KV, Agrawal SP, Jacob DR, Teif VB. NucPosDB: a database of nucleosome positioning in vivo and nucleosomics of cell-free DNA. Chromosoma 2022;131:19–28.
74. Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat Rev Genet 2019;20:71–88.
75. Koh W, Pan W, Gawad C, Fan HC, Kerchner GA, Wyss-Coray T, et al. Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. Proc Natl Acad Sci USA 2014;111: 7361–6.
76. Wu DC, Lambowitz AM. Facile single-stranded DNA sequencing of human plasma DNA via thermostable group II intron reverse transcriptase template switching. Sci Rep 2017;7:8421.
77. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human transcription factors. Cell 2018;172:650–65.
78. Chan KC, Jiang P, Sun K, Cheng YK, Tong YK, Cheng SH, et al. Second generation noninvasive fetal genome analysis reveals de novo mutations, single-base parental inheritance, and preferred DNA ends. Proc Natl Acad Sci USA 2016;113:E8159–68.
79. Sun K, Jiang P, Wong AIC, Cheng YKY, Cheng SH, Zhang H, et al. Size-tagged preferred ends in maternal plasma DNA shed light on the production mechanism and show utility in noninvasive prenatal testing. Proc Natl Acad Sci USA 2018;115:E5106–14.
80. Ivanov M, Baranova A, Butler T, Spellman P, Mileyko V. Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. BMC Genom 2015; 16(13 Suppl):S1.
81. Mouliere F, Mair R, Chandrananda D, Marass F, Smith CG, Su J, et al. Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients. EMBO Mol Med 2018;10:e9323.
82. Sun K, Jiang P, Cheng SH, Cheng THT, Wong J, Wong VWS, et al. Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. Genome Res 2019;29:418–27.
83. Moldovan N, van der Pol Y, van den Ende T, Boers D, Verkuijlen S, Creemers A, et al. Genome-wide cell-free DNA termini in patients with cancer. medRxiv 2021. https://doi.org/10.1101/2021.09.30.21264176 [Epub ahead of print].
84. Lo YMD, Han DSC, Jiang P, Chiu RWK. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. Science 2021:372:eaaw3616.
85. Zhao Z, Shilatifard A. Epigenetic modifications of histones in cancer. Genome Biol 2019;20:245.
86. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007;129:823–37.
87. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell 2013;49:825–37.
88. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 2009;459: 108–12.
89. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 2007;39:311–8.
90. Lawrence M, Daujat S, Schneider R. Lateral thinking: how histone modifications regulate gene expression. Trends Genet 2016;32:42–56.

91. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat Rev Genet 2012;13:233–45.
92. Mansson CT, Vad-Nielsen J, Meldgaard P, Nielsen AL, Sorensen BS. EGFR transcription in non-small-cell lung cancer tumours can be revealed in ctDNA by cell-free chromatin immunoprecipitation (cfChIP). Mol Oncol 2021;15:2868–76.
93. Vaisvila R, Ponnaluri VKC, Sun Z, Langhorst BW, Saleh L, Guan S, et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. Genome Res 2021 Jun 17. https://doi.org/10.1101/gr.266551.120 [Epub ahead of print].
94. Williams L, Bei Y, Church HE, Dai N, Dimalanta ET, Ettwiller LM, et al. Enzymatic methyl-seq: the next generation of methylome analysis; 2019. Available from: https://international.neb.com/tools-and-resources/feature-articles/enzymatic-methyl-seq-the-next-generation-of-methylome-analysis.
95. Erger F, Norling D, Borchert D, Leenen E, Habbig S, Wiesener MS, et al. cfNOMe – a single assay for comprehensive epigenetic analyses of cell-free DNA. Genome Med 2020;12:54.
96. Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 2012;22:2497–506.
97. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 2018;359:926–30.
98. Ungerer V, Bronkhorst AJ, Holdenrieder S. Preanalytical variables that affect the outcome of cell-free DNA measurements. Crit Rev Clin Lab Sci 2020;57:484–507.
99. Zhang W, Wei L, Huang J, Zhong B, Li J, Xu H, et al. cfDNApipe: a comprehensive quality control and analysis pipeline for cell-free DNA high-throughput sequencing data. Bioinformatics 2021 May 27. https://doi.org/10.1093/bioinformatics/btab413 [Epub ahead of print].
100. Zheng H, Zhu MS, Liu Y. FinaleDB: a browser and database of cell-free DNA fragmentation patterns. Bioinformatics 2021;37:2502–3.
101. Yu F, Li K, Li S, Liu J, Zhang Y, Zhou M, et al. CFEA: a cell-free epigenome atlas in human diseases. Nucleic Acids Res 2020;48:D40–4.