diagnostics

MDPI

Review

Tracing the Origin of Cell-Free DNA Molecules through Tissue-Specific Epigenetic Signatures

Angela Oberhofer, Abel J. Bronkhorst, Carsten Uhlig, Vida Ungerer and Stefan Holdenrieder *

Munich Biomarker Research Center, Institute of Laboratory Medicine, German Heart Center, Technical University Munich, Lazarettstraße 36, D-80636 Munich, Germany; oberhofer@dhm.mhn.de (A.O.); bronkhorst@dhm.mhn.de (A.J.B.); uhlig@dhm.mhn.de (C.U.); ungerer@dhm.mhn.de (V.U.)
* Correspondence: holdenrieder@dhm.mhn.de

Abstract: All cell and tissue types constantly release DNA fragments into human body fluids by various mechanisms including programmed cell death, accidental cell degradation and active extrusion. Particularly, cell-free DNA (cfDNA) in plasma or serum has been utilized for minimally invasive molecular diagnostics. Disease onset or pathological conditions that lead to increased cell death alter the contribution of different tissues to the total pool of cfDNA. Because cfDNA molecules retain cell-type specific epigenetic features, it is possible to infer tissue-of-origin from epigenetic characteristics. Recent research efforts demonstrated that analysis of, e.g., methylation patterns, nucleosome occupancy, and fragmentomics determined the cell- or tissue-of-origin of individual cfDNA molecules. This novel tissue-of origin-analysis enables 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 towards a wide range of pathologies and early diagnosis. In this review, we summarize the currently available tissue-of-origin approaches and point out the next steps towards clinical implementation.

Keywords: cell-free DNA; epigenetics; liquid biopsy; tissue-of-origin

check for updates

Citation: Oberhofer, A.; Bronkhorst, A.J.; Uhlig, C.; Ungerer, V.; Holdenrieder, S. Tracing the Origin of Cell-Free DNA Molecules through Tissue-Specific Epigenetic Signatures. *Diagnostics* **2022**, *12*, 1834. https://doi.org/10.3390/diagnostics12081834

Academic Editor: Paola Gazzaniga

Received: 14 June 2022
Accepted: 25 July 2022
Published: 29 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Liquid biopsy is a minimally invasive diagnostic approach in which a variety of biomarkers present in body fluids (mostly plasma and serum), including nucleic acids, are analyzed in order to detect diseases, e.g., cancer. The analysis of circulating nucleic acids already complements 'classical' solid biopsy (i.e., tissue biopsy) by providing a more comprehensive picture of the progression and heterogeneity of cancer, the response of tumors to therapy, the presence of minimal residual disease, and has recently been suggested as promising screening tool for early cancer diagnosis. Liquid biopsy has significantly advanced prenatal testing for genetic disorders and monitoring of graft rejection. Due to its minimally invasiveness, it offers the possibility to obtain serial snapshots during disease progression and is well suited for monitoring of individual therapy response. So far, non-invasive prenatal testing (NIPT) [1], detection of circulating tumor DNA (ctDNA) in the plasma of cancer patients [2], and detection of donor-derived DNA in the plasma of transplantation recipients [3] have been clinically implemented. These analyses are all based on genetic differences (fetal and maternal DNA; donor and recipient DNA in graft patients) or mutations (cancer), limiting the approach to diseases involving genetic aberrations. Additionally, detection of the low number of mutated cfDNA molecules poses a great analytical challenge and lies beyond the current limit of detection at early cancer stages. Thus, screening tests for early detection of many types of cancer and various pathologies are still lacking and are urgently needed. One way to fill this diagnostic gap might be to utilize more general features carried by all cfDNA molecules, such as epigenetic characteristics. This might broaden their diagnostic applications and expand them to early diagnosis of a wide range of pathologies.

DNA fragments from virtually all cell and tissue types are constantly released into various human body fluids, e.g., plasma, serum, cerebrospinal fluid, and urine [4–6]. Exogenous DNA (e.g., bacterial or viral DNA) can also be found in the different body fluids [7–11] in addition to host DNA (genomic and mitochondrial DNA). The so-called cell-free DNA (cfDNA) in plasma is very short-lived (half-time between 15 min and 2.5 h [12,13]) and is released into the circulation by different cellular pathways including different sorts of cell death, regular cellular turnover and upon pathologies [11,14–16]. Thereby, liquid biopsy provides a current snapshot of what cell or tissue types contribute to the plasma DNA pool at the time of blood draw. In diseased tissues, more cell death has been observed [17,18], and consequently, more cfDNA molecules from that particular cell type(s) are released into the blood compared with healthy individuals. Enzymatic digestion is usually involved in DNA release and, e.g., in plasma, cfDNA predominantly exists as short double-stranded DNA fragments in the size between 100 and 200 base pairs (bp). However, considerably longer fragments (almost 24,000 bp) have been detected recently [19]. It is widely accepted that cfDNA molecules retain the cell-type specific epigenetic features and it has been shown that methylation or fragmentation patterns of cfDNA molecules are cell-type and tissue-specific [20–23], as genes are differentially regulated in distinct cell types and different release pathways are employed by distinct cell types. Furthermore, fragment length and fragment end motifs of cfDNA molecules are non-random and depend on the cellular release pathway, the involved enzymes, and regulatory state of the releasing tissue [15,24–26]. Therefore, methylation or fragmentation patterns of cfDNA molecules represent epigenetic features that are well-applicable to infer tissue-of-origin of individual cfDNA molecules and thereby determine the contribution of individual cell types or tissues to the total plasma cfDNA pool. During the last years, research efforts have focused on identifying cell type-specific methylation or fragmentation patterns that can be employed to trace the origin of individual cfDNA molecules (Figure 1). Various groups developed methods to assign cfDNA molecules to their cell- or 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 [20,21,23,27–29]. Establishing methylation or fragmentation pattern atlases of distinct cell types is a prerequisite for developing sensitive epigenetic-based tissue-of-origin analyses. Deciphering these patterns and attributing them to the corresponding cell- or tissue-of-origin opens up novel ways for non-invasive early diagnostic tests. The focus on more general cfDNA characteristics 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 analysis.

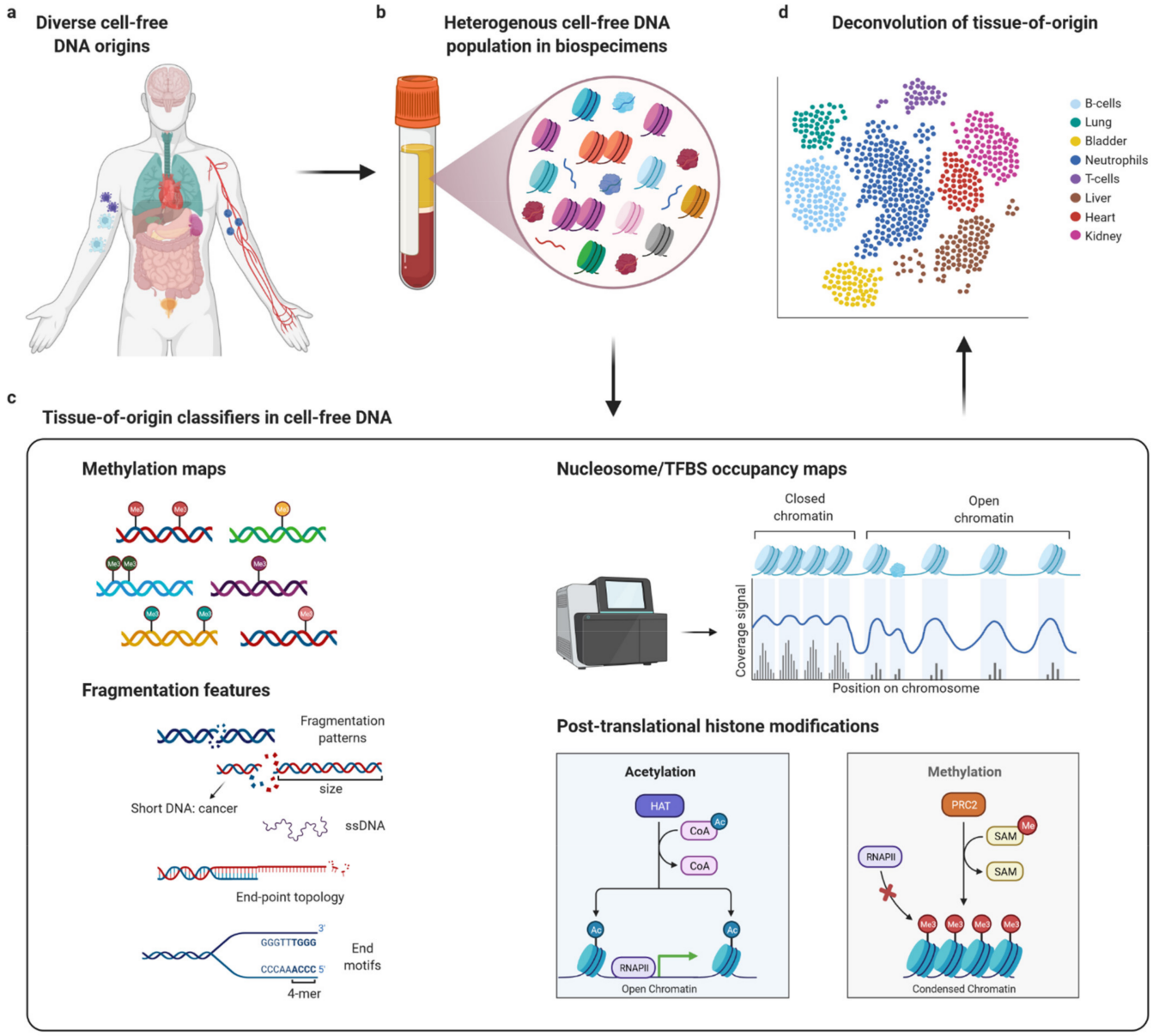

Figure 1. Tissue-of-origin analysis of cell-free DNA. (**a**) Different organs and various cell types release cell-free DNA (cfDNA) into blood plasma. (**b**) This clinical biospecimen represents a highly heterogeneous mixture of cfDNA molecules, often complicating the analytical differentiation between different cfDNA subtypes. (**c**) Multiple different epigenetic characteristics can be employed for tissue-of-origin analysis such as unique methylation patterns, fragmentation profiles and fragment end-points, transcription-factor binding sites occupancy, nucleosome positioning, as well as post-translational histone modifications. (**d**) Analysis of these features poses an analytical challenge, but various approaches developed recently enable determination of tissue-of-origin of individual cfDNA molecules, facilitating localization of tumors or tissue damage in specific regions such as the heart or liver.

2. Epigenetic-Based Biomarkers in Liquid Biopsy

Gene transcription is a tightly regulated process that involves multiple epigenetic regulatory mechanisms, such as DNA methylation, DNA compaction in accessible and non-accessible chromatin regions through binding of histones, and transcription factor (TF) binding. Transcriptional programs differ significantly between cell types and are altered upon tumorigenesis or onset of pathology. Investigating the differences in these epigenetic patterns in circulating nucleic acids might enable the early diagnosis of a wide range of

pathologies. In the following section, we summarize currently developed approaches that infer tissue-of-origin from cfDNA molecules based on different epigenetic characteristics (see Table 1 for an overview of important approaches). Handling of plasma DNA molecules necessitates very high caution with respect to DNA isolation and treatment procedures to account for the very low amounts of available DNA (very low input DNA) and its already fragmented nature. Therefore, existing protocols for the analysis of DNA from, e.g., tissue need to be adapted to the special needs of circulating nucleic acids from blood. We will also include information on cfDNA-adapted protocols within this review.

Table 1. Overview of key studies that performed tissue-of-origin analysis on cfDNA using various approaches based on methylation patterns, nucleosome positioning patterns, TFBS occupancy, histone modifications, and fragmentomics. The data literature search was performed with the PubMed NCBI database. All deconvolution methods listed in this table are reference-based and employ a classification, unless stated otherwise.

Epigenetic Feature	Method	Approach	Disease	Deconvolution Method	References
Methylation	CpG islands analysis	WGBS	HCC, NIPT, Transplant	QP	[21]
Methylation	Analysis of adjacent CpG sites	Bisulfite amplicon-seq	PDAC, CRC, Diabetes, Transplant, MS, TBI, IBD	Read-specific binary classification	[17,22]
Methylation	Methylation haplotype block analysis	scRRBS, WGBS	CRC, LCP	QP, Random forest, feature extension "haplotype blocks"	[30]
Methylation	Analysis of differentially methylated regions (DMRs) + ctDNA abundance	CfMeDIP-seq	PDAC, AML, lung and breast cancer, CRC, RCC, bladder cancer	Limma, binomial GLM (GLMnet)	[31]
Methylation	Cell-type methylation atlas	Microarray	Sepsis, islet transplantation, CRC, lung, breast, prostate cancer, CUP	NNLS	[32]
Methylation	CancerDetector	Microarray, WGBS	Liver cancer	Maximizing log-likelihood model (grid search)	[33]
Methylation	Cell-type methylation atlas Analysis of blocks of homogenously methylated CpG sites	Deep WGBS	COVID-19	dynamically programmed probabilistic Bayes model, NNLS, wgbstools	[34]
Methylation	MCED test validation	Bisulfite amplicon-seq	12 cancer types	Ensemble logistic regression, Perceptron	[35]

Table 1. *Cont.*

Epigenetic Feature	Method	Approach	Disease	Deconvolution Method	References
Nucleosome occupancy/ TFBS occupancy	Windowed protection score (L-WPS/S-WPS) analysis of long/short fragments	Deep WGS	Small-cell lung cancer, squamous cell lung cancer, colorectal adenocarcinoma, HCC, ductal carcinoma in situ breast cancer	Windowed approach, Fast Fourier transformation, no classification	[20]
Nucleosome occupancy	Nucleosome-depleted region (NDR) analysis	WGS	Breast cancer	ABSOLUTE [36], estimation of tumor purity and ploidy	[37]
TFBS occupancy	Accessibility score analysis	sWGS	Prostate adenocarcinoma, breast cancer, colon adenocarcinoma	Logistic regression	[29]
Histone modifications	Analysis of activating histone modifications	cfChIP-seq	Colorectal carcinoma, diverse liver diseases, AMI	Robust linear regression (rlm R-package)	[28]
Fragment size	Fragment size distribution analysis	sWGS + in vitro size selection	High-grade serous ovarian cancer	Logistic regression, random forest	[23]
Orientation of fragments	Orientation-aware plasma DNA fragmentation analysis (OCF)	WGS	Pregnancy, transplant, HCC, CRC, lung cancer	Nucleosome-depletion signal used to calculate OCF value, no classification	[38]
Fragment size	DNA evaluation of fragments for early interception (DELFI)	sWGS + genome-wide fragmentation pattern analysis	Breast, colorectal, lung, ovarian, pancreatic, gastric, and bile duct cancer	Multi-classifying approach, stochastic gradient boosting model	[27]
Fragment end motif	Motif diversity score (MDS) analysis using an adopted normalized Shannon entropy	WGS	HCC, pregnancy, liver transplantation, CRC, lung cancer; head and neck squamous cell, and nasopharyngeal carcinoma	SVM, logistic regression	[39]
Fragment size	Promoter fragmentation entropy (PFE) analysis using a modified Shannon index	Epigenetic expression inference from cfDNA-seq (EPIC-seq)	Non-small-cell lung cancer, diffuse large B cell lymphoma	Dirichlet-multinomial model, logistic regression	[40]

Abbreviations: TFBS: transcription factor binding site; HCC: hepatocellular carcinoma; NIPT: non-invasive prenatal testing; PDAC: pancreatic ductal adenocarcinoma; CRC: colorectal cancer; MS: multiple sclerosis; TBI: traumatic brain injury; IBD: inflammatory bowel disease; LCP: lung cancer primary; AML: acute myeloid leukemia; RCC: renal cell carcinoma; AMI: acute myocardial infarction; MCEP: multi-cancer early prediction; WGBS: whole genome bisulfite sequencing; scRRBS: single-cell reduced representation bisulfte sequencing; cfMeDIP-seq: cell-free methylated DNA immunoprecipitation and sequencing; cfChIP-seq: cell-free chromatin immunoprecipitation and sequencing; GLM: generalized linear model; NNLS: non-negative least squares; QP: quadratic programming; SVM: support vector machine.

2.1. Methylation-Based Tissue-of-Origin Analysis

The transfer of a methyl group onto the C5 position of cytosine residues results in the formation of 5-methylcytosine (5 mC) and occurs almost exclusively on cytosines with an adjacent guanosine, which are termed CpG sites. This DNA modification inhibits gene expression by hindering the binding of TFs to DNA or by inhibiting the recruitment of proteins involved in gene expression. Thus, methylated promoters are usually a sign of a silenced gene with no or low gene expression. Transcriptional programs are tightly regulated by cell-specific DNA methylation patterns and these methylation marks reflect the gene expression profile of the respective cell type. Methylation patterns are conserved among cells of the same cell type, are highly stable under physiological or pathological conditions, and are unique to each cell type. In cancer cells, transcriptional programs are altered and lead to different methylation patterns. Promoter regions of tumor suppressor genes are generally hypermethylated (i.e., silenced gene expression), whereas promoters of cancer driver genes are hypomethylated (i.e., activated gene expression) [41,42].

There are numerous different techniques to identify methylated cytosines. Generally, it can be differentiated between genome-wide, reduced genome and targeted approaches; the order mentioned represents increasing cost effectiveness. For a complete and detailed description of all available DNA methylation detection approaches, please refer to the recent review by Galardi et al. [43]. In order to detect 5 mCs, the DNA needs to be treated either by (i) restriction enzyme digestion, (ii) bisulfite treatment, (iii) affinity enrichment of methylated DNA or (iv) a combined treatment of enzymatic and chemical modifications. The gold standard for genome-wide methylation analysis is bisulfite conversion [44,45]. For this purpose, unmethylated cytosine residues are converted into uracil residues, while methylated cytosines remain unmodified. During PCR, uracil residues are replaced by thymine residues and subsequent analysis observes cytosine residues as methylated cytosines, while thymine residues represent unmethylated cytosines in the original DNA sample. Whole genome sequencing of bisulfite converted (WGBS) DNA samples enabled genome-wide DNA methylation profiling and was employed to establish genome-wide methylation maps. However, the harsh conditions of bisulfite conversion in combination with the already fragmented population of cfDNA molecules is problematic and might result in significant sample loss [46]. A bisulfite conversion-based approach adjusted for low input DNA is the post bisulfite adaptor tagging (PBAT) technique, which is a PCR-free method that reduces bisulfite induced DNA degradation by attaching adaptors to DNA after bisulfite treatment [47,48]. An adaptation of the PBAT technique for the analysis of methylation status at single cell level by single cell bisulfite sequencing (scBS-seq) [49,50] together with the PBAT method might be employable for cfDNA analyses.

Other genome-wide approaches combining the use of enzymatic and chemical modifications are enzymatic methyl-sequencing (EM-seq) [51,52] and ten-eleven translocation (TET)-assisted pyridine borane sequencing (TAPS) [53]. These 5 mC (and 5 hmC, see last part of chapter) mapping techniques are bisulfite-free and might be an alternative for plasma samples. Recently, a cfDNA-adapted protocol for TAPS was published [54].

An alternative genome-wide method for mapping methylation is immunoprecipitation of methylated DNA (MeDIP) with 5-methylcytidine-specific antibodies and subsequent sequencing of captured genomic regions [55]. This sensitive bisulfite-free workflow is especially applicable for low input samples. The original MeDIP protocol was optimized for even lower input [56], such as cfDNA from plasma, and enables detection of large-scale DNA methylation changes that are enriched for tumor-specific patterns. Immunoprecipitation of methylated DNA can further be achieved via proteins containing a methyl-binding domain (MBD). This technique followed by sequencing of captured DNA is called MBD- or MethylCap-seq [57,58] for which a cfDNA-optimized protocol exists [59].

A reduced genome approach for DNA methylation analysis is reduced representation bisulfite sequencing (RRBS) [60]. Here, CpG-rich regions across the genome are enriched via restriction enzyme digestion followed by bisulfite conversion and sequencing. This reduced genome approach is more cost-effective compared to whole genome sequencing (WGS).

This technique was also optimized for cfDNA samples (cfRRBS) [61,62]. Microarrays are another option for reduced genome methylation profiling. Commercially available microarrays facilitate DNA methylation analysis of specific, pre-selected regions of interest. This reduced cost technique is ideal for clinical applications, but prior target knowledge is required. Microarrays have been extensively employed for methylation mapping and establishing reference methylome atlases [32].

Targeted approaches for methylation mapping include targeted bisulfite amplicon sequencing (targeted BS-seq) [63,64]. Target selection can be performed by PCR amplification or probe hybridization capture. For targeted BS-seq, DNA is bisulfite converted and amplified using specific and validated primers. This technique was also optimized for plasma samples [65].

For tissue-of-origin analysis, genomic regions are identified that exhibit hyper- or hypomethylation in one cell or tissue type compared with others. These differentially methylated regions (DMRs) are subsequently employed as markers for tissue-of-origin analysis. An early study demonstrated that differentially methylated regions can be employed to detect prostate cancer [66]. In a next step, differential methylation across distinct cell or tissue types can be utilized to trace the cell- or tissue-of-origin of a single cfDNA molecule, even of different cell types within a particular tissue [67]. The basis for methylation-based tissue-of-origin analysis are high resolution methylation maps of multiple different reference cell types or tissues [68–71]. Several projects were initiated that generated high resolution methylation or epigenetic maps that are mostly available as open access datasets, such as the Roadmap Epigenomics Project [68], the ENCODE Project [72,73], the International Human Epigenome Consortium (IHEC) [74], the Cancer Genome Atlas (TCGA; https://www.cancer.gov/tcga, accessed on 11 February 2022), and the Gene Expression Omnibus (GEO) [75]. Tissue-of-origin analysis requires the development of qualified algorithms for deconvolution of sequencing data with reference methylation profiles of different tissues to determine the cell- or tissue-of-origin of cfDNA molecules and to estimate the major tissue contributors to the cfDNA pool. WGBS of plasma samples and analysis of tumor-associated hypomethylation in combination with tumor-associated copy number aberrations (CNAs) enabled the detection of several non-metastatic cancer types with a sensitivity and specificity of 87 and 88%, respectively [76]. Further, WGBS was employed to generate cfDNA methylation profiles that were used to infer relative contributions of four different tissues using a deconvolution approach [21]. For this purpose, several high resolution patterns of multiple tissue types were employed as reference methylomes [68–70]. The authors identified two types of methylation markers: (i) type I marker is a genomic locus that shows a methylation level in one of the tissues that is significantly different from those in other tissues and (ii) type II marker is a genomic locus with high variability in methylation densities. This approach demonstrated that $\geq$70% of the cfDNA pool originated from white blood cells (i.e., neutrophils and lymphocytes) in conditions in which source tissue differed genetically from host tissue (i.e., pregnancy, transplantation, cancer) and showed that the methylation deconvolution approach is able to identify the tissue-of-origin of CNAs [21]. Moreover, tissue-of-origin analysis was successfully performed with urinary cfDNA as well [77]. The authors found a high variation of proportional contribution from different tissues (i.e., neutrophils, T-cells, B-cells, urothelium, and kidney) to the urine DNA pool [77].

Expanding the analysis window from one CpG site to a number of adjacent CpGs led to enhanced sensitivity and reduced background of methylation analysis [22]. The authors were able to demonstrate that tissue-specific methylation patterns in cfDNA could be used to detect tissue cell death with a high level of specificity and sensitivity in multiple human pathologies. This approach was based on targeted sequencing of specific markers, which reduced sequencing costs immensely, and 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 [22]. A systematic search of highly coordinated methylation by a combined method approach of WGBS, RRBS, and methylation

arrays identified thousands of tightly coupled CpG sites, termed methylation haplotype blocks [30]. By defining the methylation haplotype load (MHL) as a metric that is capable of quantitatively distinguishing blocks that have the same average methylation levels but various degrees of coordinated methylation, the authors were able to directly compare different regions across multiple data sets and observed a reduction in perfectly coupled CpG pairs in cancer patients. The study showed that this methylation haplotype blocks approach can be employed for quantitative estimation of tumor load and tissue-of-origin mapping of cfDNA in patients with lung or colorectal cancer [30]. The development of probabilistic models such as CancerLocator or CancerDetector enabled researchers to infer the relative proportions and tissue-of-origin of tumor-derived cfDNA from genome-wide DNA methylation data or joint methylation states of multiple adjacent CpG sites. This allowed the detection of cancer with high sensitivity and specificity [33,78].

Comparative methylome analysis identified several genomic loci that are unmethylated specifically in cardiomyocytes or hepatocytes, respectively, and could be utilized to quantify cardiomyocyte- or hepatocyte-derived DNA in cfDNA to detect acute cardiomyocyte or hepatocyte death [17,18]. This method is based on the analysis of a limited number of genomic loci, which reduces turnaround time tremendously and might complement existing biomarkers such as troponins or liver enzymes.

The development of a sensitive, immunoprecipitation-based protocol to analyze the methylome of small quantities of cfDNA (cfMeDIP) marked an important step in methylation pattern analysis of low-input plasma DNA samples that are already of fragmented nature [31]. Specifically, the authors optimized an existing low-input MeDIP-seq protocol using exogenous *Enterobacteria phage* λ DNA (filler DNA) for DNA inputs as low as 1 to 10 ng. The work demonstrated that cfMeDIP-seq was able to detect large-scale DNA methylation changes that were enriched for tumor-specific patterns and robustly detect and classify cancer in plasma samples from several tumor types [31].

Tissue-of-origin analysis strongly relies on reference methylomes of key tissues. Since tissues mostly represent mixtures of distinct cell types, Moss and colleagues established a reference atlas of 25 human tissues and cell types covering major organs and cells involved in common diseases [32]. They employed methylation microarrays and demonstrated that plasma methylation patterns could be used to accurately identify cell type-specific cfDNA in healthy and pathological conditions including islet transplantation, sepsis, and cancer of unknown primary [32]. Further, they quantified the major cell types contributing to cfDNA in healthy individuals with 55% originating from white blood cells, 30% from erythrocyte progenitors, 10% deriving from vascular endothelial cells and 1% was hepatocyte-derived [32]. An in silico approach using automated machine-learning conducted differential methylation analysis on available microarray methylomes to establish low number biosignatures (i.e., several differentially methylated genes) for breast cancer, osteoarthritis, and diabetes mellitus and achieved high performance [79]. In order to detect distant metastases and potentially advance early detection, cell type-specific cfDNA methylation markers were recently used to identify collateral tissue damage in cancer [80]. In this work, elevated levels of hepatocyte-derived cfDNA were detected in the plasma of patients with liver metastases vs. cancer patients without liver metastases. Moreover, patients with brain metastases were identified by increased levels of neuron-, oligodendrocyte-, and astrocyte-derived cfDNA [80].

Recently, the single-molecule real-time (SMRT) long-read sequencing technology by Pacific Biosciences was modified to accurately detect 5 mC modifications [81]. This method uses a convolutional neural network to analyze the sequence context and pulse signals associated with DNA polymerase kinetics for accurate detection of 5 mC modifications and might be employed as 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 [34]. 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 hypermethylated loci are rare and enriched

for CpG islands, polycomb targets and CTCF binding sites [34]. The authors developed a computational machine-learning suite to represent, compress, visualize and analyze WGBS data (available at: https://github.com/nloyfer/wgbs_tools, accessed on 7 February 2022) [34]. The Circulating Cell-free Genome Atlas study aimed at establishing a blood-based multi-cancer early detection (MCED) test utilizing cfDNA targeted methylation-based sequencing in combination with machine-learning in order to detect cancer signals across multiple cancer types and predict cancer signal origin with high accuracy [35,65,82]. The clinical validation study concluded that the MCED test demonstrated high specificity and accuracy of cancer signal origin and the test detected cancer signals across many different types of cancers. These results support that this minimally invasive MCED test complements and expands existing single-cancer screening tests [35,65,82].

For methylation patterns, there are currently two bioinformatics approaches that are employed: (i) the reference-based approach [21,22,30,32,34], and (ii) the reference-independent approach [31,83,84]. Both methods rely on WGBS, RRBS, or microarray methylation data to detect DNA methylation at single-nucleotide resolution genome-wide or at thousands of sites in the case of microarrays. For the reference-based approach, differentially methylated regions or positions (DMRs or DMPs) that are unique to a specific cell type are identified (i.e., methylated in a specific cell type and unmethylated in all other cell types or vice versa). Quadratic programming (QP; also referred to as constrained projection(CP)) is a well-established methylation-based bioinformatic approach that infers proportions of cell types present in the reference DNA methylation database using a constrained projection via least-squares minimization [67]. Other reference-based algorithms are CIBERSORT, an advanced Support Vector Regression for penalized multivariate regression after estimating the regression weights, or robust partial correlation (RPC) [85,86]. For these algorithms to perform well, high quality methylation atlases [32,34] based on samples comprising of preferably one cell type are required to select appropriate markers/loci that are robust enough to differentiate between distinct cell types in a mixture of unknown cell types. There is currently a trend towards broader sequencing techniques called 'third generation liquid biopsy' [87,88]. Techniques such as WGBS lead to higher sensitivity by using many markers—also called "features". Therefore, feature selection is becoming increasingly important to reduce the number of markers, which in turn facilitates the interpretation of the data. Feature selection can be guided by databases such as the NCBI Gene Expression Omnibus [75] to learn which loci are reliable [32]. Other groups, such as Loyfer et al., exploited the inherent sequential data structure of the genome by grouping features that were adjacent and co-varied by cell type using a dynamically programmed probabilistic Bayes model [34]. Their aim was to obtain homogeneous blocks of methylated or unmethylated markers to (i) perform dimensionality reduction and (ii) increase the robustness of the markers. The blocks were then selected by the number of markers per block to reduce 7.2 million blocks to 2.1 million blocks. Based on this data set, Loyfer et al. filtered out further characteristics that they needed for their research goals. They arrived at a minimum of about a thousand blocks for their atlas [34]. The reference-based approach is limited to tissue and cell types that can be purified sufficiently. Furthermore, a reference in an atlas needs to be measurable and the approach does not account for influences from other cell types on the expression profile of the investigated cell type. On the other hand, quantification of cell type contributions at the single sample level is possible and the algorithms are relatively assumption-free. Reference-free algorithms include EWASher [89], RefFreeEWAS [90], ReFACTOR [91], the removing unwanted variation (RUV) framework [92], surrogate variable analysis (SVA) [93], independent surrogate variable analysis (ISVA) [94], or are based on non-negative matrix factorization (NMF) [95,96]. The reference-independent algorithms do not require reference methylation profiles or prior knowledge of cell types, and are therefore applicable to any tissue type, and account for cell-cell interactions. However, the performance of reference-independent algorithms strongly depends on whether the model assumptions are valid and sample-specific estimates of cell type fractions are not possible.

In addition to the extensively scrutinized 5 mC modification, 5-hydroxymethylcytosine (5 hmC) is an intermediate DNA modification that influences biological processes ranging from development to pathogenesis [97,98]. The ten-eleven translocation (TET) family dioxygenases convert 5 mC into 5 hmC and this DNA modification is generally thought to reflect gene activation on permissive chromatin [99]. It is particularly located at enhancers, gene bodies and promoters and changes in 5 hmC reflect alterations in gene expression levels [100,101]. 5 hmC additionally displays a tissue-specific mass distribution [102,103] and decreased levels of 5 hmC are often observed in many solid tumors compared with corresponding healthy tissues [104]. Several groups demonstrated that 5 hmC signatures in cfDNA can be utilized to detect cancer type and stage [105–107]. One study demonstrated that pancreatic ductal adenocarcinoma tissue-derived hyper-hydroxymethylated genes can separate non-cancer cfDNA from PDAC cfDNA samples [105]; making hydroxymethylation patterns another promising approach for early detection and tissue-of-origin analysis.

Taken together, deep sequencing methods such as WGBS and the development of deconvolution algorithms paved the way for methylation-based tissue-of-origin analysis of cfDNA to detect various cancer types and numerous other pathologies. Employment of targeted approaches and methodological advancements such as immunoprecipitation of methylated DNA (MeDIP) reduced the costs significantly, which is a prerequisite for clinical application. However, for methylation-based tissue-of-origin analysis to become routine application the turnaround time and workflow need to be optimized and larger validation studies are required.

2.2. Nucleosome Positioning-Based Tissue-of-Origin Analysis

DNA is packaged into chromatin by several histone proteins in order to fit within the nucleus. The degree of DNA compaction regulates its accessibility and gene expression. The histone protein octamer with ~147 bp of DNA wrapped around it is termed a nucleosome and represents the basic unit of DNA compaction. Another histone protein—H1—binds to ~20 bp of linker DNA adjacent of the nucleosome core and the ~167 bp DNA with the histone protein octamer and H1 is called chromatosome. Nucleosomes are spaced with varying distances to each other along the DNA strand and DNA regions tightly occupied by nucleosomes are rather inaccessible, whereas DNA regions with a wider spacing of nucleosomes are more accessible for DNA binding proteins. By hindering proteins involved in gene regulation and transcription from binding, nucleosomes regulate DNA accessibility and transcription. Additionally, for TF binding and gene expression it is essential that nucleosomes move along the DNA or are removed for chromatin opening at specific genomic regions [108–110]. Gene expression is unique for every cell type, differing considerably between distinct cell types and tissues, and consequently nucleosome positioning varies significantly between different cell types [20]. Upon release into a body fluid, nuclear DNA is cleaved by distinct enzymes dependent on cell type and release mechanism [25]. Open chromatin regions lacking nucleosomes are poorly protected against digestion compared with nucleosome-occupied DNA regions. Thus, nucleosome-bound regions are expected to be found more frequently in the plasma DNA pool than nucleosome-depleted regions. Transcription-prone regions with low nucleosome occupancy are underrepresented in sequencing data of cfDNA, whereas higher coverage suggests lower expression levels of this genomic region. Nucleosome positioning is consequently reflected by sequence read-density across the genome and can be employed to extract information about gene expression and thereby cell identity [111], enabling tissue-of-origin analysis via cfDNA read frequency pattern.

Several accessibility assays exist to characterize nucleosome positioning and examine nucleosome architecture and gene regulation [112]. Most techniques employ enzymatic digestion or mechanical shearing of DNA of interest and subsequent analysis of resulting fragments. DNase-seq involves enzymatic digestion by DNase I that preferentially digests open regions of chromatin, leaving behind mostly nucleosome-bound DNA regions that are subsequently sequenced [113]. Seldomly found regions in DNase-seq are called DNase

I hypersensitive sites (DHSs) and reflect active regulatory regions. Micrococcal nuclease (MNase) is another endo-exonuclease that preferentially digests unprotected DNA (i.e., accessible linker DNA between nucleosomes) and mostly leaves nucleosome-occupied DNA intact. MNase assays followed by sequencing were performed to identify nucleosome occupied regions of DNA [114], showing pronounced nucleosome-depleted regions (NDRs) directly upstream of the transcription start site (TSS) at promoters of highly expressed genes [108,115,116]. Another NDR was found at transcription termination sites (TTSs) of highly expressed genes [115]. Pronounced NDRs are not detected at genes with no or low transcription activity. Many nucleosome position maps were generated via the MNase assay. Formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) utilizes crosslinking of chromatin-interacting proteins to DNA by formaldehyde, chromatin shearing and subsequent phenol-chloroform extraction to separate DNA unbound by proteins (aqueous layer) from protein-DNA crosslinks (organic layer) [117]. An assay for transposase accessibility sequencing (ATAC-seq) employs hyperactive Tn5 transposase to insert sequencing adapters at accessible regions of the genome that will subsequently be sequenced [118,119]. Thereby, resulting sequencing reads represent regions of increased accessibility and can be used to map nucleosome positions.

To determine the nucleosome positions by analysis of plasma DNA, no experimental enzymatic digestion is necessary due to the fragmented nature of cfDNA. Nucleases that are present intracellularly and in the blood degrade preferentially DNA regions that are not bound by nucleosomes or other proteins. Because nucleosome-bound DNA is protected better against degradation, (deep) sequencing of the total plasma DNA pool primarily yields reads of nucleosome-bound regions. By aligning the resulting sequencing reads of a plasma sample to the reference genome, nucleosome positions can be inferred from and information on the cell identity can be extracted.

One study built a genome-wide map of in vivo nucleosome occupancy of cfDNA based on deep sequencing of total cfDNA and demonstrated that the resulting nucleosome spacing pattern could be utilized to perform tissue-of-origin analysis [20]. Specifically, they defined a metric—the windowed protection score (WPS)—to determine the nucleosome occupancy at a given genomic coordinate. For this purpose, they looked at 120 bp windows and quantified how many fragments (120–180 bp) ended in that window and how many fragments completely spanned it. By subtracting the number of fragments ending in that window from the number of fragments completely covering that window, they yielded the L-WPS of any genomic coordinate. A higher WPS is correlated with many fragments spanning that specific window and few fragments ending in that window, indicating that this DNA region is occupied by a nucleosome or other protein. Vice versa, a lower L-WPS corresponds to a low number of fragments spanning the window and a higher number of fragments ending in that window, suggesting that this region is not protected by a nucleosome. The metric yielded a graph that spanned the entire genome and that was analyzed for local maxima that represented regions occupied by nucleosomes, resulting in approx. 10 million peaks. To verify this hypothesis that WPS peaks represented regions occupied by nucleosomes, the authors used ChIP-seq data from the ENCODE database [120] to adjust the filtering to only receive high confidence peaks. The fragment length was the main aspect for inferring nucleosome positions in Snyder et al. and resulted in a specific pattern that was distinct for different cell types and genomic regions [20]. Additionally, single-stranded libraries revealed shorter DNA fragments that directly footprinted TF occupancy (for details see Section 2.3). Those epigenetic footprints matched hematopoietic lineage in healthy individuals, while additional contributions were observed in cancer patients, often aligning with the cancer type [20].

Another group reported that gene expression could be predicted with plasma DNA coverage in promoter regions [37]. The authors developed a nucleosome promoter analysis utilizing machine-learning and WGS datasets to determine the expression status of a gene. For this purpose, they used coverage at the TSS to infer gene expression from nucleosome occupancy [37]. They defined two ranges: (i) 2 kbp centered around the TSS and (ii) −150 bp

to 50 bp with respect to the TSS. Applying kernel density estimation to the normalized read coverage in these two ranges led to the distinction between two clusters. These two clusters could be classified with an accuracy of 0.91 using a support vector machine [121] of the most and least expressed 100 genes [37]. This approach identified two discrete regions at TSSs where nucleosome occupancy resulted in different read depth coverage patterns for expressed and silent genes, allowing classification of expressed cancer driver genes in regions with somatic copy number gains in patients with metastatic cancer [37].

A compilation of published in vivo nucleosome positioning datasets in a database called NucPosDB provides a comprehensive overview on available datasets and includes datasets of sequenced cfDNA that reflect nucleosome positioning in situ in the cells of origins [122]. Moreover, a list of computational tools for the analysis of nucleosome positioning or cfDNA experiments can be found at NucPosDB and the database contains theoretical algorithms for the prediction of nucleosome positioning preferences from DNA sequence [122]. Different approaches employing nucleosome footprinting for tissue-of-origin analysis of cfDNA fragments determined varying contributions of white blood cells and other organs to the plasma DNA pool [20,21,123–125].

The combination of EM-seq to study methylation patterns in combination with analyzing nucleosome occupancy in a single assay (cell-free DNA-based nucleosome occupancy and methylation profiling: cfNOMe) might advance tissue-of-origin analysis in plasma samples [126].

A better understanding of the underlying release mechanisms and cfDNA biology is pivotal to unravel the wealth of valuable information encoded in nucleosome positions.

2.3. Tissue-of-Origin Analysis by Inferring Transcription Factor Binding Sites

TFs are essential factors in regulating gene expression by promoting or blocking the recruitment of RNA polymerase to specific genes. Thereby they fine-tune the expression of their target genes and are key players in development and differentiation [127]. By failing to suppress cancer driver genes or silencing tumor suppressor genes, dysfunctional TFs can lead to carcinogenesis. As transcription differs tremendously between distinct cell types, distinct TFs are involved in gene expression depending on the cell type. Identifying occupied transcription factor binding sites (TFBS) within plasma DNA might be a way for tissue-of-origin analysis, allowing a more dynamic snapshot on rapidly changing gene expression in response to cancer or other pathologies.

To determine accessible DNA regions, the afore-mentioned techniques for studying accessible chromatin/nucleosome occupancy can be employed (see Section 2.2). In the case of cfDNA, enzymatic digestion is not needed due to the fragmented state of cfDNA. Sequencing cfDNA and aligning the obtained reads to the human genome results in sequence read density across the genome. Regions with low read density are suggestive of DNA lacking TFs or nucleosomes, while regions with high read density indicate TFs or nucleosomes bound to that region. However, due to the smaller size of TFs, TFBS occupancy is more challenging to determine experimentally. Consequently, few studies have to date attempted to infer TFBS occupancy from cfDNA fragmentation patterns.

One study employed single-stranded library preparation of cfDNA samples to recover short fragments and used deep sequencing to investigate if it is possible to utilize TFBS occupancy for tissue-of-origin analysis. The authors observed that short cfDNA fragments (35–80 bp) directly footprint the in vivo occupancy of TFBSs of several TFs [20]. To demonstrate this, they determined the occupancy of a TFBS by defining the short fragment windowed protection score (S-WPS) at a given genomic coordinate. The S-WPS is the number of short DNA fragments (35–80 bp) completely spanning a 16 bp window centered at a given genomic coordinate subtracted by the number of fragments with an endpoint within that window. By focusing on short cfDNA fragments, they were able to infer additional contributing tissues to the plasma DNA pool in non-healthy states [20]. The bioinformatic analysis occurred identically to the L-WPS analysis. Ulz et al. employed publicly available ATAC-seq data and WGS data from plasma cfDNA to infer accessibil-

ity of hundreds of TFBSs from cfDNA fragmentation patterns [29]. To determine TFBS occupancy, they developed a metric called accessibility score that measured the strength of TF phasing at the TFBS, reflecting the strength of the TF binding. [29]. This score was based on the read coverage quantifying the amplitudes close to TFBSs and then applying signal filtering with the Savitzky-Golay algorithm [128] to smoothen the graph, making the amplitudes easier to measure and detect. 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 were used as the accessibility score. A logistic regression was applied based on 504 markers that were pre-selected. A cross-validation approach was used to train and test the algorithm. By quantifying the amplitudes at TFBSs, the authors were able to identify nucleosome-depleted regions (NDRs) and thereby distinguish between control and cancer samples and even different subtypes of cancers. Applying this approach, the authors were able to profile numerous individual TFs and objectively compare TF binding events in plasma samples, allowing subclassification of tumor entities and TFBS plasticity during disease progression [29].

2.4. Tissue-of-Origin Analysis Utilizing Histone Modifications

Histone proteins involved in DNA compaction can be post-translationally modified in multiple ways, including methylation, acetylation, phosphorylation, adenosine diphosphate (ADP) ribosylation, ubiquitylation, sumoylation, formylation, and hydroxylation of specific histone amino acids [129,130]. These various modifications change the interaction between DNA and nuclear proteins (i.e., by compelling or attracting DNA from or to histone proteins, respectively) or modify the binding affinity of chromatin remodelers or the transcription machinery to the nucleosome. Thereby, they regulate chromatin accessibility and gene transcription and can be grouped into four major functions: activating, repressing, heterogeneous, and bivalent modifications. Methylation and acetylation are among the most well-studied histone modifications. The addition of an acetyl group influences chromatin compaction by neutralizing the basic charge at unmodified lysine residues and weakens the electrostatic interaction between the negatively charged DNA and histones [131]. Histone acetylation has further been implicated in regulation of the intracellular pH [132]. Histone acetylation is largely associated with active transcription, particularly when located at enhancers, promoters, and the gene body [133]. Histone methylation, on the other side, retains the basic character of the modified histone and represents a subtler modification [134]. Trimethylation of histone H3 at lysine residues 9 and 27 (H3K9me3 and H3K27me3) for instance has a repressive effect on transcription and marks regions of closed chromatin [129], while (tri-) methylation of histone H3 at lysine residues 4 and 36 (H3K4me1/2/3, and H3K36me1/2/3) and acetylation of histone H4 at lysine 16 (H4K16ac) are considered activating marks and hint at open chromatin regions where active transcription occurs [129]. Histone modification patterns mirror recent and transient changes in chromatin regulation and RNA polymerase activity, often occurring at the onset of pathologies when cells attempt to adapt to altered conditions by changing their transcriptional profiles. Accessible and active promoters, enhancers and gene bodies of actively transcribed genes are characterized by the presence of specific combinations of histone modifications [135–140]. Hence, histone modifications are indicative of altered transcriptional programs in response to pathologies. Specifically, changes in histone modification patterns have been extensively linked to cancer [134]. Hyperacetylation, particularly at proto-oncogenes, may activate gene expression. Conversely, hypoacetylation of tumor suppressor genes (often localized at promoters) might silence gene expression. Post-translational methylation of histones represents a functionally very complex regulation mechanism and both elevations and reductions in histone methylation have been associated with, e.g., cancer.

The most commonly used technique for profiling protein-DNA interaction is chromatin immunoprecipitation (ChIP) [141,142]. First, crosslinking is performed to covalently bind the lysines of interacting proteins with local DNA. Second, cell lysis and subsequent

shearing by sonication are carried out, followed by incubation with an antibody targeting the protein of interest or the specific histone modification of interest and finally the interacting DNA region is pulled down, usually via beads attached to the secondary antibody, crosslinks are reversed, protein is digested and sequencing is performed. One drawback of this method is, however, the high input needed to produce a high signal-to-noise ratio. Adaptations to the original protocol reduced the input need (μChIP) [143]. Employing MNase digestion instead of sonication (native ChIP) further decreased the DNA input required [144]. However, applying ChIP to very low-input plasma samples remains technically challenging. Recently, the chromatin immunoprecipitation protocol was modified specifically for low-input cfDNA from plasma samples (cfChIP) [28].

cfChIP followed by sequencing (cfChIP-seq) demonstrated that plasma nucleosomes maintain the epigenetic information of the cell they originated from and that cfChIP-seq recapitulates the original genomic distribution of histone modifications related to active transcription. Antibodies specific for different histone modifications, which were immobilized on paramagnetic beads, were directly incubated with plasma samples. Subsequently, on-bead adapter ligation was performed before DNA isolation. Four antibodies specific for active transcription (accessible/active promoters: H3K4me3 or H3K4me2; accessible enhancers: H3K4me2; gene bodies of actively transcribed genes: H3K36me3) were employed in this study that included plasma samples from healthy individuals, patients with different liver diseases, and patients with metastatic colorectal carcinoma. The authors showed that 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 [28]. Individual tailored bioinformatics procedures were applied for each of the four antibodies used in this study. The analysis showed that bone marrow megakaryocytes were identified as major contributors to the plasma DNA pool in healthy individuals and pathology-related changes in hepatocytes chromatin were found in patients with liver diseases.

Colorectal cancer was detected with high sensitivity using cfChIP-seq and could identify subgroups of patients with distinct cancer-related transcriptional programs [28]. Another study employing blood plasma cfChIP revealed 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 [145]. This focus on tumor-specific transcriptional activity of genes harboring somatic mutations will help to gain more insights on the relevance of mutations in, e.g., therapy resistance mechanisms.

So far, however, only activating histone modifications for tissue-of-origin analysis have been studied. Prospectively, it will be of great interest to investigate repressive and heterogeneous modifications as well. It might also be helpful to look into combinations of distinct histone modifications, e.g., immunoprecipitate a specific histone acetylation prior to a specific histone methylation. Hereby, it is essential to employ validated antibodies (e.g., by IHEC) of sufficient quality and to streamline the different existing protocols to achieve standardization and comparability of studies performed by different groups. Importantly, cfChIP-seq provides information on transient changes in gene expression altered upon various pathologies and during disease progression.

2.5. Tissue-of-Origin Analysis Based on Fragmentomics

Nucleosome occupancy and DNA compaction degree are not the only determinants of cfDNA fragment size, but the release pathway and nucleases present in blood also play major roles [146,147]. Depending on the cell type, release mechanism (e.g., apoptosis or necrosis), and DNA condensation mode, nuclear DNA is cleaved by distinct nucleases, resulting in varying cfDNA fragments that are characteristic for each cell type [11]. DNA fragments bound to proteins (typically histones or TFs among other proteins such as albumin, HDL, etc.) are preferentially found in bodily fluids, while naked DNA is mostly digested [148]. Commonly, the majority of cfDNA molecules exhibit the size of DNA

wrapped around one nucleosome plus DNA linker (i.e., ~167 bp) [20,23,27]. A series of additional peaks with ~10 bp periodicity below ~143 bp putatively correspond to the helical pitch and binding sites of DNA to the nucleosome core [20]. Size distribution of cfDNA fragments is altered by pathologies, e.g., cfDNA from cancer patients was found to be slightly shorter than cfDNA from healthy individuals (147 vs. 167 bp) [23,27], which is possibly explained by a different binding strength of the external histone H1. Generally, the lengths of cancer-derived cfDNA fragments have a tendency to be more variable than non-cancer DNA [23]. In addition to mono-, di-, and trinucleosome-sized cfDNA fragments predominantly found in plasma samples, considerably longer cfDNA fragments were detected, particularly a ~3 kbp cfDNA population (ranging between 1–6 kbp with an average size of 3 kbp) initially thought to originate from randomly lysed cells that might represent genomic DNA contamination [149–153]. Currently, many groups hold the view that this ~3 kbp cfDNA population is actively extruded by live cells [154–159] and it might represent an artifact of incomplete size separation with an underlying DNA laddering pattern of at least seven nucleosomes or more [160]. Depending on the nuclease(s) involved in the release pathway, the cfDNA fragments exhibit distinct fragment end motifs (i.e., four-base motif at both ends of a cfDNA fragment) [39] with "CCCA" being the most common 4-mer end motif in plasma DNA fragments from healthy individuals [161]. DNASE1L3 was suggested to generate predominantly end motifs starting with "CC" [24,162], DNA fragmentation factor subunit β (DFFB) was found to be responsible for 5′ "A" end motifs [25], and DNASE1 activity was associated with generating "T" end motifs [25]. Nuclease activity is thought to be altered upon pathogenesis and might be utilized for tissue-of-origin analysis.

In particular, cfDNA size distribution, preferred ends (i.e., genomic coordinates that are found at cfDNA ends more often than others), end motifs of cfDNA, end orientation of cfDNA, and topology of ends (i.e., double-stranded vs. single-stranded) are informative about tissue-of-origin and numerous studies demonstrated the feasibility of fragmentomics-based liquid biopsies to detect cancer and other pathologies [23,26,27,38]. These fragmentomics features can be studied by sequencing plasma samples and analyzing the fragmentomics feature of interest. However, to determine fragmentomics features of cfDNA molecules accurately, it is essential to preserve the charactistics of plasma DNA molecules by a preanalytical routine that does not alter fragmentation and by a gentle and suitable library preparation procedure (e.g., optimized single-stranded library preparation or a library preparation method without end repair). Thus, the cfDNA fragmentomics features can be employed to distinguish between healthy and diseased cells. Pathologies and injuries 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 pathogenic tissue. Tissue-of-origin analysis based on fragmentation patterns have been shown to 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.

An early study on plasma cfDNA fragmentation patterns demonstrated that fragmentation occurs non-randomly and cfDNA retained characteristics previously found in genome-wide analysis of chromatin structure and are concordant with corresponding cell-line derived patterns [163]. The same authors further investigated fragmentation patterns of cfDNA from cerebrospinal fluid (CFS) from glioma patients and found a distinct fragmentation pattern of cfDNA in CFS [164]. Another research group developed an approach to detect the total pool of cancer-associated somatic mutations in plasma and studied the cfDNA end characteristics in patients with hepatocellular carcinoma and chronic hepatitis [26]. The authors found tumor-associated cfDNA preferred end-coordinates at certain genomic coordinates that point to cfDNA derived from transplanted liver, hepatocellular carcinoma, or the placenta. The number of cfDNA molecules with end signatures for tumor or liver correlated with the amounts of tumor- or liver-derived cfDNA in plasma [26]. The comprehensive analysis of cfDNA fragmentation patterns in plasma samples from patients

with different cancer types and healthy individuals identified differences in the size distribution of tumor-derived and non-cancer DNA fragments [23], namely mutant ctDNA was more fragmented than non-mutant cfDNA. This observation led the authors to develop an approach that selectively analyzes short cfDNA fragments in the size of 90 to 150 bp with a machine-learning algorithm, achieving greater sensitivity to detect tumor DNA from multiple cancer types in plasma [23]. Extending the fragmentation pattern analysis to additionally determining cfDNA fragment orientation (i.e., upstream or downstream fragment end profile) can identify short linker DNA and tissue-specific open chromatin regions, allowing determination of relative contributions of various tissues to the plasma DNA pool [38].

Evaluating fragmentation patterns across the genome at megabase level (the authors looked at ~500 windows of 5 Mbp size each) to observe large-scale fragmentation patterns was the aim of a different approach termed DNA evaluation of fragments for early interception (DELFI) [27]. The authors used shallow coverage WGS data and developed a machine-learning model that incorporated genome-wide fragmentation features and examined fragmentation patterns from healthy and cancer samples. For this classifying approach, the authors utilized the gbm (stochastic gradient boosting) machine-learning model and performed a 10-fold cross-validation. They determined the ratio of short (100–150 bp) and long fragments (151–220 bp) for these 500 Mbp windows and normalized for GC content via a LOESS smoother [165]. They simulated the limit of detection based on the fraction of ctDNA in the plasma DNA pool and attempted to improve performance by adding copy number changes, chromosomal arm changes or mitochondrial DNA, but the genomic profile as defined originally performed better or equal. They observed that profiles of healthy individuals reflected nucleosomal patterns of white blood cells and patients with cancer exhibited altered fragmentation patterns. Applying the DELFI algorithm to samples from different cancer types could identify the tissue-of-origin of cancers in three-quarters of examined samples. Combining DELFI with mutation-based cfDNA analysis detected 91% of patients with cancer [27].

Studying the diversity of plasma DNA end motifs (i.e., the first 4-nucleotide sequence on each 5′ end of the Watson and Crick strand) and determining the motif diversity score (MDS) identified differences in the end motifs between patients with different cancers (including hepatocellular carcinoma) and healthy subjects [39]. Additionally, it was observed that plasma DNA molecules from liver, hepatocellular carcinoma, placenta, hematopoietic cells bore characteristic plasma DNA end motifs that could be utilized for tissue-of-origin analysis [39]. For instance, the end motif "CCCA" was less frequently found in samples from patients with hepatocellular carcinoma than from healthy individuals. The authors found a downregulation of DNASE1L3 in hepatocellular carcinoma cells and hypothesized that decreased DNASE1L3 activity was responsible for decrease of "CCCA"-bearing cfDNA fragments in plasma of patients with hepatocellular carcinoma.

Moreover, cfDNA fragment end sequence patterns were investigated and utilized for tissue-of-origin analysis [166]. This work focused on the diversity of bases at the ends of cfDNA fragments (i.e., cfDNA termini) and defined a quantitative metric referred to as the fragment end integrated analysis (FrEIA) score to objectively compare fragment ends. For comparison, it was focused on the first mapped 5′ trinucleotide, the first mapped 5′ mononucleotide and the last mapped 3′ nucleotide [166]. These nucleotides were categorized into the first 5′ trinucleotide, the first 5′ nucleotide and the first and last 5′ nucleotide pair. Fractions for the fragment categories were calculated. Using these fractions, the FrES entropy (i.e., information content) was calculated by applying the normalized Shannon entropy and the Gini index. The FrEIA score developed in that study was defined as the ratio of increased trinucleotides and decreased trinucleotides in cancer multiplied by the previously mentioned normalized entropy. The machine-learning model for cancer classification was cross-validated 10-fold and included hyper-tuning and applying several different models resulted in the selection of a support vector machine (SVM) [121] as the best model. In addition, unsupervised machine-learning was employed using t-sne

for clustering via k-means. The limit of detection was determined by simulating down-sampled populations [166]. By compiling a genome-wide catalogue of cfDNA fragment end sequence patterns of a large cohort of cancer patients, the authors demonstrated that fragment-end sequence and diversity were altered in 18 distinct cancer types. Furthermore, they were able to classify cancer samples from controls at low tumor content [166].

Most recently, a novel approach for analyzing expression based on cfDNA fragmentomics was developed that measured fragment length diversity to infer RNA expression levels at individual genes [40]. The working hypothesis was that cfDNA fragments from active promoters would display more random cleavage than fragments from inactive promoters due to the difference in nucleosome occupancy. To test this, targeted sequencing of promoters of genes of interest was performed and fragment length diversity was determined by promoter fragmentation entropy (PFE). The developed prediction model was based on two features: (i) promoter fragmentation entropy (PFE) that measured the diversity in fragment length distribution at a TSS compared to control genes and (ii) NDR defined as normalized counts per million of 2 kbp around the TSS. For classification of cancer vs. control and subtype classification, the authors employed logistic regression with regularization (elastic net; between 50–150 features per model). They performed cross-validation and preanalytic checks and in silico simulation for determining the limit of detection. Employing this analysis method to plasma samples could classify subtypes of lung carcinoma and diffuse large B cell lymphoma. Further, it was possible to correlate gene expression profiles with clinical response in serial blood samples from patients with PD-(L)1 immune checkpoint inhibitors treatment [40].

It has become increasingly clear during the last years that another cfDNA population exists in plasma that might exhibit diagnostic potential: ultrashort single-stranded cfDNA fragments up to 100 bp [20,167–170]. These fragments are often omitted by conventional library preparation due to the adapter ligation process that needs double strands and/or applied size selection procedures. An early study employing single-stranded library preparation and subsequent sequencing demonstrated that short cfDNA molecules directly footprint TFBS occupancy [20]. Recent efforts optimizing different DNA extraction and library preparation procedures independently demonstrated the presence of this novel population of cfDNA molecules that might inform on gene regulatory regions and DNA secondary structures [168–170].

Overall, cfDNA fragments hold an abundance of information that can be utilized for tissue-of-origin analysis (reviewed in detail in [171]) with broad potential clinical applications after further validation.

2.6. Bioinformatic Analysis of Epigenetic Features of cfDNA

The basis for a rigorous and comprehensive bioinformatic tissue-of-origin analysis is the generation of high-quality data from samples collected with well-defined preanalytical and analytical procedures. Sequencing data is quality-checked, filtered and then further processed in different ways prior to downstream analysis for deconvolution of contributing cell or tissue types. We will briefly discuss the general underlying bioinformatic procedures that have to-date been utilized for tissue-of-origin analysis based on the different epigenetic features of cfDNA molecules.

In most cases, the determination of the tissue-of-origin is done either by finding the right mixture of cfDNA from the corresponding tissues by deconvolution or by classifying whether the cfDNA contains the signature of a tissue or a disease. Deconvolution is performed using either label-free methods or reference-based methods. The reference-based methods rely on tissue references and use them to find the correct tissue admixture. The most common machine-learning methods used for cfDNA classification are random forest, logistic regression and support vector machines.

With the advancement of biotechnological methods, it is now possible to perform deconvolution and classification using many features instead of focusing on a few selected mutations. This is a paradigm shift and means that the way data from liquid biopsies is

processed has to be adapted. The result is an enormous amount of data points per sample that need to be processed in a way to find the right information. This part of the process is called featurization or feature engineering. It is often based on the epigenetic feature measured using, for example, a window approach to combine adjacent CpG sites into cfDNA methylation data. The main goal is generally to extract the information appropriate to the research question and thus reduce the number of data points per sample. This process is called feature selection or dimensionality reduction. Once the data have been selected, the method of deconvolution or machine-learning is applied.

Understanding biases, such as preanalytical factors and the protocol used, becomes increasingly important as most methods use references as ground truth for comparison. When ground truths differ, interpretation of the data becomes much more difficult and can lead to incorrect assumptions.

An interesting factor for cfDNA is the available epigenetic features that can be used and differ from the use of cfDNA mutations previously. In most cases, a single epigenetic feature type, such as e.g., methylation data, is analyzed, with the exception of a few papers such as Siejka-Zielinska et al., 2021, which used methylation and fragmentation data [54]. Other research groups have used a single epigenetic trait but used a different trait to verify the gathered information, such as Tang et al., 2017, using methylation data and then verifying against miRNA expression data [172].

Overall, a wide variety of tissue-of-origin analyses based on distinct epigenetic features have been developed and their great potential for screening tests have been demonstrated. However, extensive bioinformatic expertise is needed for the development and application of the different sophisticated machine-learning algorithms. Numerous comprehensive databases that are publicly available await to be further analyzed (e.g., Roadmap Epigenomics Project [68], the ENCODE Project [72,73], the International Human Epigenome Consortium (IHEC) [74], the Cancer Genome Atlas (TCGA; https://www.cancer.gov/tcga, accessed on 11 February 2022), the Gene Expression Omnibus (GEO) [75], the Gene Transcription Regulation Database (GTRD; http://gtrd.biouml.org, accessed on 8 June 2022), the Genome-Tissue Expression (GTEx), the BLUEPRINT Epigenome (https://www.blueprint-epigenome.eu/, accessed on 8 June 2022)). Combining different approaches based on epigenetic features of cfDNA might advance minimally invasive liquid biopsies.

3. Conclusions and Future Perspectives

Liquid biopsies based on the analysis of epigenetic cfDNA features significantly advanced the scope of this minimally invasive approach towards a more sensitive early detection of multiple types of cancer and detection of various pathologies beyond cancer (e.g., autoimmune disorders, organ pathologies, systemic inflammation). Analysis of methylation patterns, nucleosome footprints, histone modifications, and the emerging field of fragmentomics offers a yet unprecedented potential for clinical implementation of liquid biopsies, for example as screening methods for pathologies lacking reliable tests or sensitive and serial monitoring of therapy success. However, larger validation studies, additional insights on the release mechanism as well as cfDNA biology, guidelines for standardized preanalytical procedures, and harmonized bioinformatic pipelines are urgently required prior to broad applicability of epigenetic-based liquid biopsy approaches.

Tissue-of-origin approaches utilizing epigenetic characteristics of circulating nucleic acids are only at the early stages of development. It is not yet possible to evaluate the full potential of minimallyinvasive liquid biopsy tests for early detection and serial monitoring. As research on liquid biopsies continues to evolve, it will become clear which approach(es) or combination of approaches are best suited for detection of distinct pathologies. In addition, technical optimization of tissue-of-origin techniques will be performed as research continues. In particular, possible influences of preanalytical and analytical biases on tissue-of-origin analysis need to be further elucidated and minimized. Equally importantly, a comparative validation of the appropriate bioinformatic approach needs to be achieved.

Subsequently, focus should be placed on advancing the most promising methods in order to maximize the benefit to patients. The prerequisite for clinical application is the ability of an approach to reliably stratify between healthy individuals and the patient cohort. For this purpose, all of the proof-of-concept studies presented in this review need to be further validated with larger cohorts in appropriately powered clinical trials and the workflows need to be standardized to evaluate their performance. Clinical applicability and utility for different settings (i.e., for disease identification and organ/cell type localization) need to be rigorously tested. Despite of the large amount of work that still needs to be performed, tissue-of-origin approaches based on epigenetic characteristics of plasma DNA molecules hold great promise for precision and targeted medicine.

Taken together, shifting research interest to epigenetic characteristics of circulating nucleic acids considerably boosted the performance of liquid biopsy particularly for detection of pathologies besides cancer and lays the ground for early and precision diagnostics of a broad range of diseases.

Author Contributions: A.O. and C.U. wrote the first draft of the manuscript. A.J.B. edited the manuscript and conceptualized the figure. V.U. conceptualized and created the figure. S.H. edited the manuscript and conceptualized the figure. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Figure 1 was created with Biorender.com, accessed on 4 May 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lo, Y.M.; Corbetta, N.; Chamberlain, P.F.; Rai, V.; Sargent, I.L.; Redman, C.W.; Wainscoat, J.S. Presence of fetal DNA in maternal plasma and serum. *Lancet* **1997**, *350*, 485–487. [CrossRef]
2. 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*, 1035–1037. [CrossRef] [PubMed]
3. Lo, Y.M.; Tein, M.S.; Pang, C.C.; Yeung, C.K.; Tong, K.L.; Hjelm, N.M. Presence of donor-specific DNA in plasma of kidney and liver-transplant recipients. *Lancet* **1998**, *351*, 1329–1330. [CrossRef]
4. Hui, L.; Maron, J.L.; Gahan, P.B. Other Body Fluids as Non-invasive Sources of Cell-Free DNA/RNA. In *Circulating Nucleic Acids in Early Diagnosis, Prognosis and Treatment Monitoring. Advances in Predictive, Preventive and Personalised Medicine*; Gahan, P., Ed.; Springer: Dordrecht, The Netherlands, 2015; Volume 5.
5. Mandel, P.; Metais, P. Nuclear Acids In Human Blood Plasma. *C. R. Seances Soc. Biol. Fil.* **1948**, *142*, 241–243. [PubMed]
6. Bronkhorst, A.J.; Ungerer, V.; Holdenrieder, S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol. Detect. Quantif.* **2019**, *17*, 100087. [CrossRef]
7. Kohler, C.; Radpour, R.; Barekati, Z.; Asadollahi, R.; Bitzer, J.; Wight, E.; Burki, N.; Diesch, C.; Holzgreve, W.; Zhong, X.Y. Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors. *Mol. Cancer* **2009**, *8*, 105. [CrossRef]
8. Zachariah, R.R.; 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*, 843–850. [CrossRef]
9. Kowarsky, M.; Camunas-Soler, J.; Kertesz, M.; De Vlaminck, I.; Koh, W.; Pan, W.; Martin, L.; Neff, N.F.; Okamoto, J.; Wong, R.J.; 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*, 9623–9628. [CrossRef]
10. Long, Y.; Zhang, Y.; Gong, Y.; Sun, R.; Su, L.; Lin, X.; Shen, A.; Zhou, J.; Caiji, Z.; Wang, X.; et al. Diagnosis of Sepsis with Cell-free DNA by Next-Generation Sequencing Technology in ICU Patients. *Arch. Med. Res.* **2016**, *47*, 365–371. [CrossRef] [PubMed]
11. Grabuschnig, S.; Bronkhorst, A.J.; Holdenrieder, S.; Rosales Rodriguez, I.; Schliep, K.P.; Schwendenwein, D.; Ungerer, V.; Sensen, C.W. 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. [CrossRef]
12. To, E.W.; Chan, K.C.; Leung, S.F.; Chan, L.Y.; To, K.F.; Chan, A.T.; Johnson, P.J.; Lo, Y.M. Rapid clearance of plasma Epstein-Barr virus DNA after surgical treatment of nasopharyngeal carcinoma. *Clin. Cancer Res.* **2003**, *9*, 3254–3259. [PubMed]

13. 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*, 142–148. [CrossRef] [PubMed]
14. Delgado, P.O.; Alves, B.C.; de Sousa Gehrke, F.; Kuniyoshi, R.K.; Wroclavski, M.L.; Del Giglio, A.; Fonseca, F.L. Characterization of cell-free circulating DNA in plasma in patients with prostate cancer. *Tumor Biol.* **2013**, *34*, 983–986. [CrossRef] [PubMed]
15. Jahr, S.; Hentze, H.; Englisch, S.; Hardt, D.; Fackelmayer, F.O.; Hesch, R.D.; Knippers, R. DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **2001**, *61*, 1659–1665.
16. Sai, S.; Ichikawa, D.; Tomita, H.; Ikoma, D.; Tani, N.; Ikoma, H.; Kikuchi, S.; Fujiwara, H.; Ueda, Y.; Otsuji, E. Quantification of plasma cell-free DNA in patients with gastric cancer. *Anticancer Res.* **2007**, *27*, 2747–2751.
17. Lehmann-Werman, R.; Magenheim, J.; Moss, J.; Neiman, D.; Abraham, O.; Piyanzin, S.; Zemmour, H.; Fox, I.; Dor, T.; Grompe, M.; et al. Monitoring liver damage using hepatocyte-specific methylation markers in cell-free circulating DNA. *JCI Insight* **2018**, *3*, e120687. [CrossRef]
18. Zemmour, H.; Planer, D.; Magenheim, J.; Moss, J.; Neiman, D.; Gilon, D.; Korach, A.; Glaser, B.; Shemer, R.; Landesberg, G.; et al. Non-invasive detection of human cardiomyocyte death using methylation patterns of circulating DNA. *Nat. Commun.* **2018**, *9*, 1443. [CrossRef]
19. Yu, S.C.Y.; Jiang, P.; Peng, W.; Cheng, S.H.; Cheung, Y.T.T.; Tse, O.Y.O.; Shang, H.; Poon, L.C.; Leung, T.Y.; Chan, K.C.A.; 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. [CrossRef]
20. Snyder, M.W.; Kircher, M.; Hill, A.J.; Daza, R.M.; Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **2016**, *164*, 57–68. [CrossRef]
21. Sun, K.; Jiang, P.; Chan, K.C.; Wong, J.; Cheng, Y.K.; Liang, R.H.; Chan, W.K.; Ma, E.S.; Chan, S.L.; Cheng, S.H.; 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–E5512. [CrossRef]
22. Lehmann-Werman, R.; Neiman, D.; Zemmour, H.; Moss, J.; Magenheim, J.; Vaknin-Dembinsky, A.; Rubertsson, S.; Nellgard, B.; Blennow, K.; Zetterberg, H.; et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1826–E1834. [CrossRef] [PubMed]
23. 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. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **2018**, *10*, eaat4921. [CrossRef] [PubMed]
24. Chandrananda, D.; Thorne, N.P.; Bahlo, M. High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med. Genom.* **2015**, *8*, 29. [CrossRef] [PubMed]
25. Han, D.S.C.; Ni, M.; Chan, R.W.Y.; Chan, V.W.H.; Lui, K.O.; Chiu, R.W.K.; Lo, Y.M.D. The Biology of Cell-free DNA Fragmentation and the Roles of DNASE1, DNASE1L3, and DFFB. *Am. J. Hum. Genet.* **2020**, *106*, 202–214. [CrossRef]
26. 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. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E10925–E10933. [CrossRef]
27. Cristiano, S.; Leal, A.; Phallen, J.; Fiksel, J.; Adleff, V.; Bruhm, D.C.; Jensen, S.O.; Medina, J.E.; Hruban, C.; White, J.R.; et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **2019**, *570*, 385–389. [CrossRef]
28. Sadeh, R.; Sharkia, I.; Fialkoff, G.; Rahat, A.; Gutin, J.; Chappleboim, A.; Nitzan, M.; Fox-Fisher, I.; Neiman, D.; Meler, G.; et al. ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. *Nat. Biotechnol.* **2021**, *39*, 586–598. [CrossRef]
29. Ulz, P.; Perakis, S.; Zhou, Q.; Moser, T.; Belic, J.; Lazzeri, I.; Wolfler, A.; Zebisch, A.; Gerger, A.; Pristauz, G.; et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nat. Commun.* **2019**, *10*, 4666. [CrossRef] [PubMed]
30. Guo, S.; Diep, D.; Plongthongkum, N.; Fung, H.L.; 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*, 635–642. [CrossRef]
31. Shen, S.Y.; Singhania, R.; Fehringer, G.; Chakravarthy, A.; Roehrl, M.H.A.; Chadwick, D.; Zuzarte, P.C.; Borgida, A.; Wang, T.T.; Li, T.; et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **2018**, *563*, 579–583. [CrossRef] [PubMed]
32. Moss, J.; Magenheim, J.; Neiman, D.; Zemmour, H.; Loyfer, N.; Korach, A.; Samet, Y.; Maoz, M.; Druid, H.; Arner, P.; et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **2018**, *9*, 5068. [CrossRef] [PubMed]
33. Li, W.; Li, Q.; Kang, S.; Same, M.; Zhou, Y.; Sun, C.; Liu, C.C.; Matsuoka, L.; Sher, L.; Wong, W.H.; 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. [CrossRef] [PubMed]
34. Loyfer, N.; Magenheim, J.; Peretz, A.; Cann, G.; Bredno, J.; Klochendler, A.; Fox-Fisher, I.; Shabi-Porat, S.; Hecht, M.; Pelet, T.; 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**. [CrossRef]

35. Klein, E.A.; Richards, D.; Cohn, A.; Tummala, M.; Lapham, R.; Cosgrove, D.; Chung, G.; Clement, J.; Gao, J.; Hunkapiller, N.; et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann. Oncol.* **2021**, *32*, 1167–1177. [CrossRef]
36. Carter, S.L.; Cibulskis, K.; Helman, E.; McKenna, A.; Shen, H.; Zack, T.; Laird, P.W.; Onofrio, R.C.; Winckler, W.; Weir, B.A.; et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **2012**, *30*, 413–421. [CrossRef]
37. 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. Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **2016**, *48*, 1273–1278. [CrossRef]
38. 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. Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. *Genome Res.* **2019**, *29*, 418–427. [CrossRef]
39. 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. Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer, Pregnancy, and Transplantation. *Cancer Discov.* **2020**, *10*, 664–673. [CrossRef] [PubMed]
40. Esfahani, M.S.; Hamilton, E.G.; Mehrmohamadi, M.; Nabet, B.Y.; Alig, S.K.; King, D.A.; Steen, C.B.; Macaulay, C.W.; Schultz, A.; Nesselbush, M.C.; et al. Inferring gene expression from cell-free DNA fragmentation profiles. *Nat. Biotechnol.* **2022**, *40*, 585–597. [CrossRef] [PubMed]
41. Estecio, M.R.; Issa, J.P. Dissecting DNA hypermethylation in cancer. *FEBS Lett.* **2011**, *585*, 2078–2086. [CrossRef]
42. Ross, J.P.; Rand, K.N.; Molloy, P.L. Hypomethylation of repeated DNA sequences in cancer. *Epigenomics* **2010**, *2*, 245–269. [CrossRef]
43. Galardi, F.; Luca, F.; Romagnoli, D.; Biagioni, C.; Moretti, E.; Biganzoli, L.; Leo, A.D.; Migliaccio, I.; Malorni, L.; Benelli, M. Cell-Free DNA-Methylation-Based Methods and Applications in Oncology. *Biomolecules* **2020**, *10*, 1677. [CrossRef] [PubMed]
44. Frommer, M.; McDonald, L.E.; Millar, D.S.; Collis, C.M.; Watt, F.; Grigg, G.W.; Molloy, P.L.; Paul, C.L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827–1831. [CrossRef] [PubMed]
45. Clark, S.J.; Harrison, J.; Paul, C.L.; Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **1994**, *22*, 2990–2997. [CrossRef] [PubMed]
46. Grunau, C.; Clark, S.J.; Rosenthal, A. Bisulfite genomic sequencing: Systematic investigation of critical experimental parameters. *Nucleic Acids Res.* **2001**, *29*, E65. [CrossRef] [PubMed]
47. Miura, F.; Ito, T. Post-Bisulfite Adaptor Tagging for PCR-Free Whole-Genome Bisulfite Sequencing. *Methods Mol. Biol.* **2018**, *1708*, 123–136. [CrossRef] [PubMed]
48. Miura, F.; Shibata, Y.; Miura, M.; Sangatsuda, Y.; Hisano, O.; Araki, H.; Ito, T. Highly efficient single-stranded DNA ligation technique improves low-input whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* **2019**, *47*, e85. [CrossRef] [PubMed]
49. Clark, S.J.; Smallwood, S.A.; Lee, H.J.; Krueger, F.; Reik, W.; Kelsey, G. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc.* **2017**, *12*, 534–547. [CrossRef]
50. Smallwood, S.A.; Lee, H.J.; Angermueller, C.; Krueger, F.; Saadeh, H.; Peat, J.; Andrews, S.R.; Stegle, O.; Reik, W.; Kelsey, G. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **2014**, *11*, 817–820. [CrossRef] [PubMed]
51. Vaisvila, R.; Ponnaluri, V.K.C.; Sun, Z.; Langhorst, B.W.; Saleh, L.; Guan, S.; Dai, N.; Campbell, M.A.; Sexton, B.S.; Marks, K.; et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res.* **2021**, *31*, 1280–1289. [CrossRef]
52. Williams, L.; Bei, Y.; Church, H.E.; Dai, N.; Dimalanta, E.T.; Ettwiller, L.M.; Evans, T.C.; Langhorst, B.W.; Borgaro, J.G.; Guan, S.; et al. Enzymatic Methyl-Seq: The Next Generation of Methylome Analysis. Available online: https://international.neb.com/tools-and-resources/feature-articles/enzymatic-methyl-seq-the-next-generation-of-methylome-analysis (accessed on 18 February 2022).
53. Liu, Y.; Siejka-Zielinska, P.; Velikova, G.; Bi, Y.; Yuan, F.; Tomkova, M.; Bai, C.; Chen, L.; Schuster-Bockler, B.; Song, C.X. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* **2019**, *37*, 424–429. [CrossRef] [PubMed]
54. Siejka-Zielinska, P.; Cheng, J.; Jackson, F.; Liu, Y.; Soonawalla, Z.; Reddy, S.; Silva, M.; Puta, L.; McCain, M.V.; Culver, E.L.; et al. Cell-free DNA TAPS provides multimodal information for early cancer detection. *Sci. Adv.* **2021**, *7*, eabh0534. [CrossRef] [PubMed]
55. Taiwo, O.; Wilson, G.A.; Morris, T.; Seisenberger, S.; Reik, W.; Pearce, D.; Beck, S.; Butcher, L.M. Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat. Protoc.* **2012**, *7*, 617–636. [CrossRef] [PubMed]
56. Shen, S.Y.; Burgener, J.M.; Bratman, S.V.; De Carvalho, D.D. Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA. *Nat. Protoc.* **2019**, *14*, 2749–2780. [CrossRef] [PubMed]
57. Brinkman, A.B.; Simmer, F.; Ma, K.; Kaan, A.; Zhu, J.; Stunnenberg, H.G. Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* **2010**, *52*, 232–236. [CrossRef]
58. Cross, S.H.; Charlton, J.A.; Nan, X.; Bird, A.P. Purification of CpG islands using a methylated DNA binding column. *Nat. Genet.* **1994**, *6*, 236–244. [CrossRef] [PubMed]

59. Huang, J.; Soupir, A.C.; Wang, L. Cell-free DNA methylome profiling by MBD-seq with ultra-low input. *Epigenetics* **2022**, *17*, 239–252. [CrossRef] [PubMed]
60. Meissner, A.; Gnirke, A.; Bell, G.W.; Ramsahoye, B.; Lander, E.S.; Jaenisch, R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **2005**, *33*, 5868–5877. [CrossRef]
61. De Koker, A.; Van Paemel, R.; De Wilde, B.; De Preter, K.; Callewaert, N. A versatile method for circulating cell-free DNA methylome profiling by reduced representation bisulfite sequencing. *bioRxiv* **2019**. [CrossRef]
62. Van Paemel, R.; De Koker, A.; Caggiano, C.; Morlion, A.; Mestdagh, P.; De Wilde, B.; Vandesompele, J.; De Preter, K. Genome-wide study of the effect of blood collection tubes on the cell-free DNA methylome. *Epigenetics* **2021**, *16*, 797–807. [CrossRef]
63. Holmila, R.; Sklias, A.; Muller, D.C.; Degli Esposti, D.; Guilloreau, P.; McKay, J.; Sangrajrang, S.; Srivatanakul, P.; Hainaut, P.; Merle, P.; et al. Targeted deep sequencing of plasma circulating cell-free DNA reveals Vimentin and Fibulin 1 as potential epigenetic biomarkers for hepatocellular carcinoma. *PLoS ONE* **2017**, *12*, e0174265. [CrossRef] [PubMed]
64. Widschwendter, M.; Evans, I.; Jones, A.; Ghazali, S.; Reisel, D.; Ryan, A.; Gentry-Maharaj, A.; Zikan, M.; Cibula, D.; Eichner, J.; et al. Methylation patterns in serum DNA for early identification of disseminated breast cancer. *Genome Med.* **2017**, *9*, 115. [CrossRef] [PubMed]
65. Liu, M.C.; Oxnard, G.R.; Klein, E.A.; Swanton, C.; Seiden, M.V.; Consortium, C. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* **2020**, *31*, 745–759. [CrossRef] [PubMed]
66. Ellinger, J.; Haan, K.; Heukamp, L.C.; Kahl, P.; Buttner, R.; Muller, S.C.; von Ruecker, A.; Bastian, P.J. CpG island hypermethylation in cell-free serum DNA identifies patients with localized prostate cancer. *Prostate* **2008**, *68*, 42–49. [CrossRef] [PubMed]
67. Houseman, E.A.; Accomando, W.P.; Koestler, D.C.; Christensen, B.C.; Marsit, C.J.; Nelson, H.H.; Wiencke, J.K.; Kelsey, K.T. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **2012**, *13*, 86. [CrossRef]
68. Roadmap Epigenomics, C.; Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J.; et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317–330. [CrossRef]
69. Hodges, E.; Molaro, A.; Dos Santos, C.O.; Thekkat, P.; Song, Q.; Uren, P.J.; Park, J.; Butler, J.; Rafii, S.; McCombie, W.R.; et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell* **2011**, *44*, 17–28. [CrossRef] [PubMed]
70. Lun, F.M.; Chiu, R.W.; Sun, K.; Leung, T.Y.; Jiang, P.; Chan, K.C.; Sun, H.; Lo, Y.M. Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. *Clin. Chem.* **2013**, *59*, 1583–1594. [CrossRef]
71. Fernandez, A.F.; Assenov, Y.; Martin-Subero, J.I.; Balint, B.; Siebert, R.; Taniguchi, H.; Yamamoto, H.; Hidalgo, M.; Tan, A.C.; Galm, O.; et al. A DNA methylation fingerprint of 1628 human samples. *Genome Res.* **2012**, *22*, 407–419. [CrossRef]
72. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [CrossRef] [PubMed]
73. Davis, C.A.; Hitz, B.C.; Sloan, C.A.; Chan, E.T.; Davidson, J.M.; Gabdank, I.; Hilton, J.A.; Jain, K.; Baymuradov, U.K.; Narayanan, A.K.; et al. The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **2018**, *46*, D794–D801. [CrossRef] [PubMed]
74. Bujold, D.; Morais, D.A.L.; Gauthier, C.; Cote, C.; Caron, M.; Kwan, T.; Chen, K.C.; Laperle, J.; Markovits, A.N.; Pastinen, T.; et al. The International Human Epigenome Consortium Data Portal. *Cell Syst.* **2016**, *3*, 496–499.e2. [CrossRef] [PubMed]
75. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, *30*, 207–210. [CrossRef] [PubMed]
76. Chan, K.C.; Jiang, P.; Chan, C.W.; Sun, K.; Wong, J.; Hui, E.P.; Chan, S.L.; Chan, W.C.; Hui, D.S.; Ng, S.S.; 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*, 18761–18768. [CrossRef] [PubMed]
77. Cheng, T.H.T.; Jiang, P.; Tam, J.C.W.; Sun, X.; Lee, W.S.; Yu, S.C.Y.; Teoh, J.Y.C.; Chiu, P.K.F.; Ng, C.F.; Chow, K.M.; et al. Genomewide bisulfite sequencing reveals the origin and time-dependent fragmentation of urinary cfDNA. *Clin. Biochem.* **2017**, *50*, 496–501. [CrossRef] [PubMed]
78. Kang, S.; Li, Q.; Chen, Q.; Zhou, Y.; Park, S.; Lee, G.; Grimes, B.; Krysan, K.; Yu, M.; Wang, W.; et al. CancerLocator: Non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol.* **2017**, *18*, 53. [CrossRef]
79. Karaglani, M.; Panagopoulou, M.; Baltsavia, I.; Apalaki, P.; Theodosiou, T.; Iliopoulos, I.; Tsamardinos, I.; Chatzaki, E. Tissue-Specific Methylation Biosignatures for Monitoring Diseases: An In Silico Approach. *Int. J. Mol. Sci.* **2022**, *23*, 2959. [CrossRef]
80. Lubotzky, A.; Zemmour, H.; Neiman, D.; Gotkine, M.; Loyfer, N.; Piyanzin, S.; Ochana, B.L.; Lehmann-Werman, R.; Cohen, D.; Moss, J.; et al. Liquid biopsy reveals collateral tissue damage in cancer. *JCI Insight* **2022**, *7*, e153559. [CrossRef]
81. Tse, O.Y.O.; Jiang, P.; Cheng, S.H.; Peng, W.; Shang, H.; Wong, J.; Chan, S.L.; Poon, L.C.Y.; Leung, T.Y.; Chan, K.C.A.; et al. Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2019768118. [CrossRef] [PubMed]
82. Chen, X.; Dong, Z.; Hubbell, E.; Kurtzman, K.N.; Oxnard, G.R.; Venn, O.; Melton, C.; Clarke, C.A.; Shaknovich, R.; Ma, T.; et al. Prognostic Significance of Blood-Based Multi-cancer Detection in Plasma Cell-Free DNA. *Clin. Cancer Res.* **2021**, *27*, 4221–4229. [CrossRef] [PubMed]
83. Teschendorff, A.E.; Zheng, S.C. Cell-type deconvolution in epigenome-wide association studies: A review and recommendations. *Epigenomics* **2017**, *9*, 757–768. [CrossRef] [PubMed]

84. Titus, A.J.; Gallimore, R.M.; Salas, L.A.; Christensen, B.C. Cell-type deconvolution from DNA methylation: A review of recent applications. *Hum. Mol. Genet.* **2017**, *26*, R216–R224. [CrossRef] [PubMed]
85. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.G.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [CrossRef] [PubMed]
86. Teschendorff, A.E.; Breeze, C.E.; Zheng, S.C.; Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinform.* **2017**, *18*, 105. [CrossRef] [PubMed]
87. Barefoot, M.E.; Loyfer, N.; Kiliti, A.J.; McDeed, A.P.; Kaplan, T.; Wellstein, A. Detection of Cell Types Contributing to Cancer From Circulating, Cell-Free Methylated DNA. *Front. Genet.* **2021**, *12*, 671057. [CrossRef]
88. Im, Y.R.; Tsui, D.W.Y.; Diaz, L.A., Jr.; Wan, J.C.M. Next-Generation Liquid Biopsies: Embracing Data Science in Oncology. *Trends Cancer* **2021**, *7*, 283–292. [CrossRef]
89. Zou, J.; Lippert, C.; Heckerman, D.; Aryee, M.; Listgarten, J. Epigenome-wide association studies without the need for cell-type composition. *Nat. Methods* **2014**, *11*, 309–311. [CrossRef] [PubMed]
90. Houseman, E.A.; Molitor, J.; Marsit, C.J. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* **2014**, *30*, 1431–1439. [CrossRef]
91. Rahmani, E.; Zaitlen, N.; Baran, Y.; Eng, C.; Hu, D.; Galanter, J.; Oh, S.; Burchard, E.G.; Eskin, E.; Zou, J.; et al. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat. Methods* **2016**, *13*, 443–445. [CrossRef]
92. Gagnon-Bartsch, J.A.; Speed, T.P. Using control genes to correct for unwanted variation in microarray data. *Biostatistics* **2012**, *13*, 539–552. [CrossRef]
93. Leek, J.T.; Johnson, W.E.; Parker, H.S.; Jaffe, A.E.; Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **2012**, *28*, 882–883. [CrossRef] [PubMed]
94. Teschendorff, A.E.; Zhuang, J.; Widschwendter, M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics* **2011**, *27*, 1496–1505. [CrossRef] [PubMed]
95. Moffitt, R.A.; Marayati, R.; Flate, E.L.; Volmar, K.E.; Loeza, S.G.; Hoadley, K.A.; Rashid, N.U.; Williams, L.A.; Eaton, S.C.; Chung, A.H.; et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **2015**, *47*, 1168–1178. [CrossRef] [PubMed]
96. Repsilber, D.; Kern, S.; Telaar, A.; Walzl, G.; Black, G.F.; Selbig, J.; Parida, S.K.; Kaufmann, S.H.; Jacobsen, M. Biomarker discovery in heterogeneous tissue samples -taking the in-silico deconfounding approach. *BMC Bioinform.* **2010**, *11*, 27. [CrossRef] [PubMed]
97. Shen, L.; Song, C.X.; He, C.; Zhang, Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev. Biochem.* **2014**, *83*, 585–614. [CrossRef] [PubMed]
98. Vasanthakumar, A.; Godley, L.A. 5-hydroxymethylcytosine in cancer: Significance in diagnosis and therapy. *Cancer Genet.* **2015**, *208*, 167–177. [CrossRef] [PubMed]
99. Mellen, M.; Ayata, P.; Dewell, S.; Kriaucionis, S.; Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **2012**, *151*, 1417–1430. [CrossRef]
100. Han, D.; Lu, X.; Shih, A.H.; Nie, J.; You, Q.; Xu, M.M.; Melnick, A.M.; Levine, R.L.; He, C. A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Mol. Cell* **2016**, *63*, 711–719. [CrossRef]
101. Konstandin, N.; Bultmann, S.; Szwagierczak, A.; Dufour, A.; Ksienzyk, B.; Schneider, F.; Herold, T.; Mulaw, M.; Kakadia, P.M.; Schneider, S.; 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*, 1649–1652. [CrossRef]
102. Globisch, D.; Munzel, M.; Muller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Bruckl, T.; Biel, M.; Carell, T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **2010**, *5*, e15367. [CrossRef]
103. Kriaucionis, S.; Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **2009**, *324*, 929–930. [CrossRef] [PubMed]
104. Jin, S.G.; Jiang, Y.; Qiu, R.; Rauch, T.A.; Wang, Y.; Schackert, G.; Krex, D.; Lu, Q.; Pfeifer, G.P. 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res.* **2011**, *71*, 7360–7365. [CrossRef] [PubMed]
105. Guler, G.D.; Ning, Y.; Ku, C.J.; Phillips, T.; McCarthy, E.; Ellison, C.K.; Bergamaschi, A.; Collin, F.; Lloyd, P.; Scott, A.; et al. Detection of early stage pancreatic cancer using 5-hydroxymethylcytosine signatures in circulating cell free DNA. *Nat. Commun.* **2020**, *11*, 5270. [CrossRef] [PubMed]
106. Li, W.; Zhang, X.; Lu, X.; You, L.; Song, Y.; Luo, Z.; Zhang, J.; Nie, J.; Zheng, W.; Xu, D.; et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res.* **2017**, *27*, 1243–1257. [CrossRef] [PubMed]
107. Song, C.X.; Yin, S.; Ma, L.; Wheeler, A.; Chen, Y.; Zhang, Y.; Liu, B.; Xiong, J.; Zhang, W.; Hu, J.; et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res.* **2017**, *27*, 1231–1242. [CrossRef] [PubMed]
108. Schones, D.E.; Cui, K.; Cuddapah, S.; Roh, T.Y.; Barski, A.; Wang, Z.; Wei, G.; Zhao, K. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **2008**, *132*, 887–898. [CrossRef]
109. Li, M.; Hada, A.; Sen, P.; Olufemi, L.; Hall, M.A.; Smith, B.Y.; Forth, S.; McKnight, J.N.; Patel, A.; Bowman, G.D.; et al. Dynamic regulation of transcription factors by nucleosome remodeling. *eLife* **2015**, *4*, e06249. [CrossRef]
110. Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thastrom, A.; Field, Y.; Moore, I.K.; Wang, J.P.; Widom, J. A genomic code for nucleosome positioning. *Nature* **2006**, *442*, 772–778. [CrossRef]

111. Radman-Livaja, M.; Rando, O.J. Nucleosome positioning: How is it established, and why does it matter? *Dev. Biol.* **2010**, *339*, 258–266. [CrossRef] [PubMed]
112. Klein, D.C.; Hainer, S.J. Genomic methods in profiling DNA accessibility and factor localization. *Chromosome Res.* **2020**, *28*, 69–85. [CrossRef]
113. He, H.H.; Meyer, C.A.; Hu, S.S.; Chen, M.W.; Zang, C.; Liu, Y.; Rao, P.K.; Fei, T.; Xu, H.; Long, H.; et al. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **2014**, *11*, 73–78. [CrossRef] [PubMed]
114. Tsompana, M.; Buck, M.J. Chromatin accessibility: A window into the genome. *Epigenetics Chromatin* **2014**, *7*, 33. [CrossRef] [PubMed]
115. Valouev, A.; Johnson, S.M.; Boyd, S.D.; Smith, C.L.; Fire, A.Z.; Sidow, A. Determinants of nucleosome organization in primary human cells. *Nature* **2011**, *474*, 516–520. [CrossRef] [PubMed]
116. Venkatesh, S.; Workman, J.L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 178–189. [CrossRef] [PubMed]
117. Rodriguez-Gil, A.; Riedlinger, T.; Ritter, O.; Saul, V.V.; Schmitz, M.L. Formaldehyde-assisted Isolation of Regulatory Elements to Measure Chromatin Accessibility in Mammalian Cells. *J. Vis. Exp.* **2018**, *134*, 57272. [CrossRef]
118. Buenrostro, J.D.; Wu, B.; Litzenburger, U.M.; Ruff, D.; Gonzales, M.L.; Snyder, M.P.; Chang, H.Y.; Greenleaf, W.J. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **2015**, *523*, 486–490. [CrossRef]
119. Buenrostro, J.D.; Giresi, P.G.; Zaba, L.C.; Chang, H.Y.; Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **2013**, *10*, 1213–1218. [CrossRef]
120. The ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **2004**, *306*, 636–640. [CrossRef]
121. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]
122. Shtumpf, M.; Piroeva, K.V.; Agrawal, S.P.; Jacob, D.R.; Teif, V.B. NucPosDB: A database of nucleosome positioning in vivo and nucleosomics of cell-free DNA. *Chromosoma* **2022**, *131*, 19–28. [CrossRef] [PubMed]
123. Heitzer, E.; Haque, I.S.; Roberts, C.E.S.; Speicher, M.R. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **2019**, *20*, 71–88. [CrossRef]
124. Koh, W.; Pan, W.; Gawad, C.; Fan, H.C.; Kerchner, G.A.; Wyss-Coray, T.; Blumenfeld, Y.J.; El-Sayed, Y.Y.; Quake, S.R. Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7361–7366. [CrossRef] [PubMed]
125. Wu, D.C.; Lambowitz, A.M. Facile single-stranded DNA sequencing of human plasma DNA via thermostable group II intron reverse transcriptase template switching. *Sci. Rep.* **2017**, *7*, 8421. [CrossRef] [PubMed]
126. Erger, F.; Norling, D.; Borchert, D.; Leenen, E.; Habbig, S.; Wiesener, M.S.; Bartram, M.P.; Wenzel, A.; Becker, C.; Toliat, M.R.; et al. cfNOMe-A single assay for comprehensive epigenetic analyses of cell-free DNA. *Genome Med.* **2020**, *12*, 54. [CrossRef] [PubMed]
127. Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2018**, *172*, 650–665. [CrossRef]
128. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
129. Zhao, Z.; Shilatifard, A. Epigenetic modifications of histones in cancer. *Genome Biol.* **2019**, *20*, 245. [CrossRef]
130. Audia, J.E.; Campbell, R.M. Histone Modifications and Cancer. *Cold Spring Harb. Perspect. Biol.* **2016**, *8*, a019521. [CrossRef]
131. Kouzarides, T. Chromatin modifications and their function. *Cell* **2007**, *128*, 693–705. [CrossRef]
132. McBrian, M.A.; Behbahan, I.S.; Ferrari, R.; Su, T.; Huang, T.W.; Li, K.; Hong, C.S.; Christofk, H.R.; Vogelauer, M.; Seligson, D.B.; et al. Histone acetylation regulates intracellular pH. *Mol. Cell* **2013**, *49*, 310–321. [CrossRef]
133. Di Cerbo, V.; Schneider, R. Cancers with wrong HATs: The impact of acetylation. *Brief. Funct. Genom.* **2013**, *12*, 231–243. [CrossRef] [PubMed]
134. Bannister, A.J.; Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **2011**, *21*, 381–395. [CrossRef] [PubMed]
135. Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.Y.; Schones, D.E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-resolution profiling of histone methylations in the human genome. *Cell* **2007**, *129*, 823–837. [CrossRef] [PubMed]
136. Calo, E.; Wysocka, J. Modification of enhancer chromatin: What, how, and why? *Mol. Cell* **2013**, *49*, 825–837. [CrossRef]
137. Heintzman, N.D.; Hon, G.C.; Hawkins, R.D.; Kheradpour, P.; Stark, A.; Harp, L.F.; Ye, Z.; Lee, L.K.; Stuart, R.K.; Ching, C.W.; et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **2009**, *459*, 108–112. [CrossRef]
138. Heintzman, N.D.; Stuart, R.K.; Hon, G.; Fu, Y.; Ching, C.W.; Hawkins, R.D.; Barrera, L.O.; Van Calcar, S.; Qu, C.; Ching, K.A.; et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **2007**, *39*, 311–318. [CrossRef] [PubMed]
139. Lawrence, M.; Daujat, S.; Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **2016**, *32*, 42–56. [CrossRef] [PubMed]
140. Lenhard, B.; Sandelin, A.; Carninci, P. Metazoan promoters: Emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **2012**, *13*, 233–245. [CrossRef] [PubMed]

141. Gilmour, D.S.; Lis, J.T. Detecting protein-DNA interactions in vivo: Distribution of RNA polymerase on specific bacterial genes. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 4275–4279. [CrossRef]
142. Solomon, M.J.; Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 6470–6474. [CrossRef]
143. Dahl, J.A.; Jung, I.; Aanes, H.; Greggains, G.D.; Manaf, A.; Lerdrup, M.; Li, G.; Kuan, S.; Li, B.; Lee, A.Y.; et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **2016**, *537*, 548–552. [CrossRef] [PubMed]
144. O'Neill, L.P.; Turner, B.M. Immunoprecipitation of native chromatin: NChIP. *Methods* **2003**, *31*, 76–82. [CrossRef]
145. Mansson, C.T.; Vad-Nielsen, J.; Meldgaard, P.; Nielsen, A.L.; Sorensen, B.S. EGFR transcription in non-small-cell lung cancer tumours can be revealed in ctDNA by cell-free chromatin immunoprecipitation (cfChIP). *Mol. Oncol.* **2021**, *15*, 2868–2876. [CrossRef]
146. Chan, K.C.; Jiang, P.; Sun, K.; Cheng, Y.K.; Tong, Y.K.; Cheng, S.H.; Wong, A.I.; Hudecova, I.; Leung, T.Y.; Chiu, R.W.; 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–E8168. [CrossRef]
147. Sun, K.; Jiang, P.; Wong, A.I.C.; Cheng, Y.K.Y.; Cheng, S.H.; Zhang, H.; Chan, K.C.A.; Leung, T.Y.; Chiu, R.W.K.; Lo, Y.M.D. 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–E5114. [CrossRef] [PubMed]
148. Holdenrieder, S.; Stieber, P.; Chan, L.Y.; Geiger, S.; Kremer, A.; Nagel, D.; Lo, Y.M. Cell-free DNA in serum and plasma: Comparison of ELISA and quantitative PCR. *Clin. Chem.* **2005**, *51*, 1544–1546. [CrossRef] [PubMed]
149. Alcaide, M.; Cheung, M.; Hillman, J.; Rassekh, S.R.; Deyell, R.J.; Batist, G.; Karsan, A.; Wyatt, A.W.; Johnson, N.; Scott, D.W.; et al. Evaluating the quantity, quality and size distribution of cell-free DNA by multiplex droplet digital PCR. *Sci. Rep.* **2020**, *10*, 12564. [CrossRef]
150. Augustus, E.; Van Casteren, K.; Sorber, L.; van Dam, P.; Roeyen, G.; Peeters, M.; Vorsters, A.; Wouters, A.; Raskin, J.; Rolfo, C.; et al. The art of obtaining a high yield of cell-free DNA from urine. *PLoS ONE* **2020**, *15*, e0231058. [CrossRef]
151. Garcia-Murillas, I.; Turner, N.C. Assessing HER2 Amplification in Plasma cfDNA. *Methods Mol. Biol.* **2018**, *1768*, 161–172. [CrossRef]
152. Heider, K.; Wan, J.C.M.; Hall, J.; Belic, J.; Boyle, S.; Hudecova, I.; Gale, D.; Cooper, W.N.; Corrie, P.G.; Brenton, J.D.; et al. Detection of ctDNA from Dried Blood Spots after DNA Size Selection. *Clin. Chem.* **2020**, *66*, 697–705. [CrossRef]
153. Lee, E.Y.; Lee, E.J.; Yoon, H.; Lee, D.H.; Kim, K.H. Comparison of Four Commercial Kits for Isolation of Urinary Cell-Free DNA and Sample Storage Conditions. *Diagnostics* **2020**, *10*, 234. [CrossRef] [PubMed]
154. Aucamp, J.; Bronkhorst, A.J.; Peters, D.L.; Van Dyk, H.C.; Van der Westhuizen, F.H.; Pretorius, P.J. Kinetic analysis, size profiling, and bioenergetic association of DNA released by selected cell lines in vitro. *Cell. Mol. Life Sci.* **2017**, *74*, 2689–2707. [CrossRef]
155. Bronkhorst, A.J.; Wentzel, J.F.; Aucamp, J.; van Dyk, E.; du Plessis, L.; Pretorius, P.J. Characterization of the cell-free DNA released by cultured cancer cells. *Biochim. Biophys. Acta* **2016**, *1863*, 157–165. [CrossRef]
156. Bronkhorst, A.J.; Wentzel, J.F.; Ungerer, V.; Peters, D.L.; Aucamp, J.; de Villiers, E.P.; Holdenrieder, S.; Pretorius, P.J. Sequence analysis of cell-free DNA derived from cultured human bone osteosarcoma (143B) cells. *Tumor Biol.* **2018**, *40*, 1010428318801190. [CrossRef] [PubMed]
157. Jeppesen, D.K.; Fenix, A.M.; Franklin, J.L.; Higginbotham, J.N.; Zhang, Q.; Zimmerman, L.J.; Liebler, D.C.; Ping, J.; Liu, Q.; Evans, R.; et al. Reassessment of Exosome Composition. *Cell* **2019**, *177*, 428–445.e18. [CrossRef] [PubMed]
158. Panagopoulou, M.; Karaglani, M.; Balgkouranidou, I.; Pantazi, C.; Kolios, G.; Kakolyris, S.; Chatzaki, E. Circulating cell-free DNA release in vitro: Kinetics, size profiling, and cancer-related gene methylation. *J. Cell. Physiol.* **2019**, *234*, 14079–14089. [CrossRef] [PubMed]
159. Wang, W.; Kong, P.; Ma, G.; Li, L.; Zhu, J.; Xia, T.; Xie, H.; Zhou, W.; Wang, S. Characterization of the release and biological significance of cell-free DNA from breast cancer cell lines. *Oncotarget* **2017**, *8*, 43180–43191. [CrossRef] [PubMed]
160. Ungerer, V.; Bronkhorst, A.J.; Van den Ackerveken, P.; Herzog, M.; Holdenrieder, S. Serial profiling of cell-free DNA and nucleosome histone modifications in cell cultures. *Sci. Rep.* **2021**, *11*, 9460. [CrossRef] [PubMed]
161. 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. Dnase1l3 deletion causes aberrations in length and end-motif frequencies in plasma DNA. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 641–649. [CrossRef] [PubMed]
162. Suzuki, N.; Kamataki, A.; Yamaki, J.; Homma, Y. Characterization of circulating DNA in healthy human plasma. *Clin. Chim. Acta* **2008**, *387*, 55–58. [CrossRef]
163. 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* (Suppl. 13), S1. [CrossRef] [PubMed]
164. Mouliere, F.; Mair, R.; Chandrananda, D.; Marass, F.; Smith, C.G.; Su, J.; Morris, J.; Watts, C.; Brindle, K.M.; Rosenfeld, N. Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients. *EMBO Mol. Med.* **2018**, *10*, e9323. [CrossRef] [PubMed]
165. Cleveland, W.S.; Devlin, S.J. Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **1988**, *83*, 596–610. [CrossRef]
166. Moldovan, N.; van der Pol, Y.; van den Ende, T.; Boers, D.; Verkuijlen, S.; Creemers, A.; Ramaker, J.; Vu, T.; Fransen, M.F.; Pegtel, M.; et al. Genome-wide cell-free DNA termini in patients with cancer. *medRxiv* **2021**. [CrossRef]

167. Burnham, P.; Kim, M.S.; Agbor-Enoh, S.; Luikart, H.; Valantine, H.A.; Khush, K.K.; De Vlaminck, I. Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci. Rep.* **2016**, *6*, 27859. [CrossRef]
168. Cheng, J.; Morselli, M.; Huang, W.; Heo, Y.J.; Pinheiro-Ferreira, T.; Li, F.; Wei, F.; Chia, D.; Kim, Y.; He, H.; et al. Plasma contains ultrashort single-stranded DNA in addition to nucleosomal cell-free DNA. *iScience* **2022**, *25*, 104554. [CrossRef]
169. Hisano, O.; Ito, T.; Miura, F. Short single-stranded DNAs with putative non-canonical structures comprise a new class of plasma cell-free DNA. *BMC Biol.* **2021**, *19*, 225. [CrossRef]
170. Hudecova, I.; Smith, C.G.; Hansel-Hertsch, R.; Chilamakuri, C.S.; Morris, J.A.; Vijayaraghavan, A.; Heider, K.; Chandrananda, D.; Cooper, W.N.; Gale, D.; et al. Characteristics, origin, and potential for cancer diagnostics of ultrashort plasma cell-free DNA. *Genome Res.* **2022**, *32*, 215–227. [CrossRef]
171. Lo, Y.M.D.; Han, D.S.C.; Jiang, P.; Chiu, R.W.K. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science* **2021**, *372*, eaaw3616. [CrossRef]
172. Tang, B.; Zhou, Y.; Wang, C.M.; Huang, T.H.; Jin, V.X. Integration of DNA methylation and gene transcription across nineteen cell types reveals cell type-specific and genomic region-dependent regulatory patterns. *Sci. Rep.* **2017**, *7*, 3626. [CrossRef]