



Article

# Integration of Cell-Free DNA End Motifs and Fragment Lengths Can Identify Active Genes in Liquid Biopsies

Christoffer Trier Maansson <sup>1,2,3</sup> , Louise Skov Thomsen <sup>1</sup>, Peter Meldgaard <sup>4</sup> , Anders Lade Nielsen <sup>3</sup> and Boe Sandahl Sorensen <sup>1,2,\*</sup>

<sup>1</sup> Department of Clinical Biochemistry, Aarhus University Hospital, 8200 Aarhus, Denmark; ctm@clin.au.dk (C.T.M.)

<sup>2</sup> Department of Clinical Medicine, Aarhus University, 8000 Aarhus, Denmark

<sup>3</sup> Department of Biomedicine, Aarhus University, 8000 Aarhus, Denmark; aln@biomed.au.dk

<sup>4</sup> Department of Oncology, Aarhus University Hospital, 8200 Aarhus, Denmark; petemeld@rm.dk

\* Correspondence: boesoere@rm.dk; Tel.: +45-61715248

**Abstract:** Multiple studies have shown that cell-free DNA (cfDNA) from cancer patients differ in both fragment length and fragment end motif (FEM) from healthy individuals, yet there is a lack of understanding of how the two factors combined are associated with cancer and gene transcription. In this study, we conducted cfDNA fragmentomics evaluations using plasma from lung cancer patients ( $n = 12$ ) and healthy individuals ( $n = 7$ ). A personal gene expression profile was established from plasma using H3K36me3 cell-free chromatin immunoprecipitation sequencing (cfChIP-seq). The genes with the highest expression displayed an enrichment of short cfDNA fragments (median = 19.99%, IQR: 16.94–27.13%,  $p < 0.0001$ ) compared to the genes with low expression. Furthermore, distinct GC-rich FEMs were enriched after cfChIP. Combining the frequency of short cfDNA fragments with the presence of distinct FEMs resulted in an even further enrichment of the most expressed genes (median = 37.85%, IQR: 30.10–39.49%,  $p < 0.0001$ ). An in vitro size selection of <150 bp cfDNA could isolate cfDNA representing active genes and the size-selection enrichment correlated with the cfChIP-seq enrichment (Spearman  $r$  range: 0.499–0.882,  $p < 0.0001$ ). This study expands the knowledge regarding cfDNA fragmentomics and sheds new light on how gene activity is associated with both cfDNA fragment lengths and distinct FEMs.

**Keywords:** liquid biopsies; fragmentomics; cell-free chromatin immunoprecipitation; gene expression; histone modifications; fragment end motifs



**Citation:** Maansson, C.T.; Thomsen, L.S.; Meldgaard, P.; Nielsen, A.L.; Sorensen, B.S. Integration of Cell-Free DNA End Motifs and Fragment Lengths Can Identify Active Genes in Liquid Biopsies. *Int. J. Mol. Sci.* **2024**, *25*, 1243. <https://doi.org/10.3390/ijms25021243>

Academic Editors: Sandra V. Fernandez and Carmela Paolillo

Received: 13 December 2023

Revised: 15 January 2024

Accepted: 17 January 2024

Published: 19 January 2024



**Copyright:** © 2024 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/>).

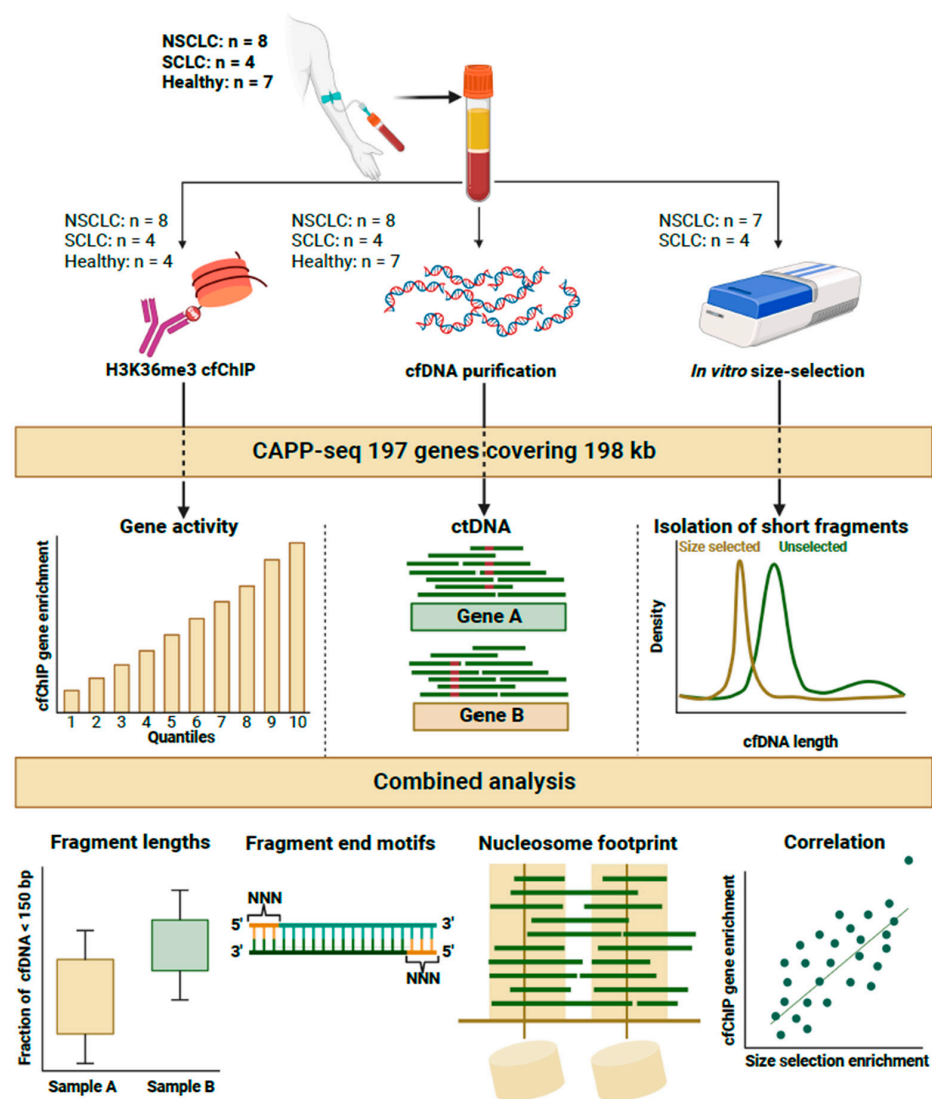
## 1. Introduction

Cell-free DNA (cfDNA) in liquid biopsies has been studied extensively in recent years to improve the sensitivity of methods for detecting somatic variants [1–4]. Plasma cfDNA originating from tumors is known as circulating tumor DNA (ctDNA), and mutations in ctDNA are important biomarkers for cancer therapies [5–7]. However, the use of liquid biopsies is limited by a relatively low ctDNA fraction, below 0.01% in some cases [8]. ctDNA is most frequently identified by somatic mutations, but other features of ctDNA have been studied to increase the sensitivity of methods for detecting ctDNA in early-stage cancers [9–11] and minimal residual disease [4,12], for which the ctDNA content is low. These features are related to the fragmentation of cfDNA before circulatory release. First, short cfDNA fragments (<150 bp) contain a larger fraction of mutated fragments than the total cfDNA pool [10,13–16]. Little is known about why ctDNA is shorter than cfDNA originating from noncancer cells, although the underlying chromatin structure [17–19] and the level of the endonuclease, DNASE1L3 [20], are known to affect cfDNA fragmentation. Second, the levels of distinct cfDNA fragment end motifs (FEMs) differ between cancer patients and healthy individuals [9,21,22]. These observations have been used to develop

machine learning models to discriminate between plasma samples from healthy individuals and cancer patients [9,21].

In addition, the fragmentation of cfDNA near transcription start sites (TSSs) and transcription-factor binding sites (TFBSs) has been coupled to gene activity [17–19,23,24]. With promoter regions in an open chromatin state, active genes are more accessible to endonucleases, leading to lower cfDNA sequencing coverage and a more diverse fragment length profile [23]. Previously, the cell-free chromatin immunoprecipitation (cfChIP) technique was demonstrated to identify active genes in cfDNA. cfChIP offers a new opportunity to identify active genes in liquid biopsies by the immunoprecipitation of trimethylated histone 3 lysine 36 (H3K36me3) nucleosomes from plasma [25–28].

cfDNA fragmentomics is a growing field in liquid biopsies; however, no comprehensive explanation has been proposed for the relationship between the length of cfDNA or FEMs and gene activity. Next-generation sequencing (NGS) of cfDNA applied to cfChIP has given us an opportunity to create a personal gene expression profile based on liquid biopsies. In the present paper, we took advantage of the ability to obtain information on gene activity and used this information to determine the correlation of gene activity with cfDNA features such as FEM and fragment length (Figure 1). We hereby demonstrated that the combination of cfDNA fragment length and FEM data is superior for predicting the expression patterns of corresponding genes in lung cancer patients and healthy individuals.



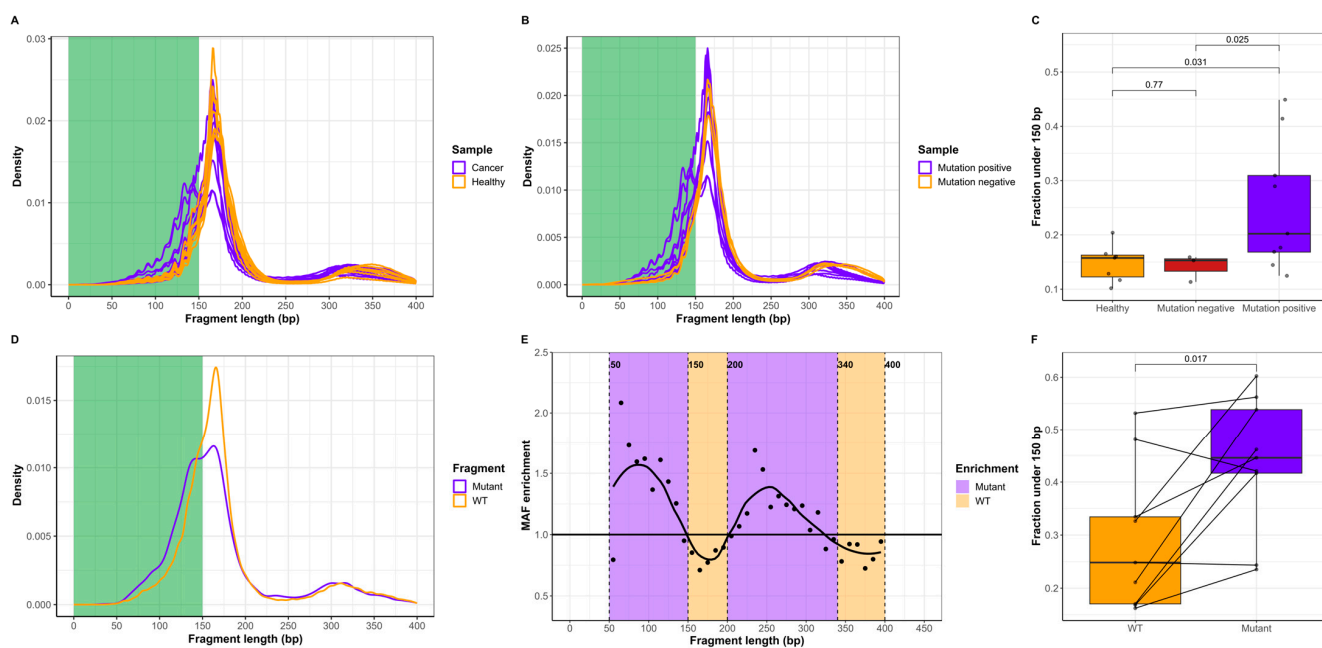
**Figure 1.** Experimental setup of the study. Plasma samples from 12 lung cancer patients and 7 healthy individuals were subjected to input cfDNA purification (middle), H3K36me3 cfChIP (left), or in vitro

size selection (**right**). These samples were sequenced via Cancer Personalized Profiling by deep sequencing (CAPP-seq) 197 genes. cfChIP samples were used to estimate gene activity for each individual. The input cfDNA samples were used to detect mutations in cell-free DNA. In vitro size selection was used to investigate the properties of short cfDNA fragments. The data from each type of sample were analyzed in relation to fragment length, fragment end motifs (FEMs), nucleosomal footprint, and enrichment correlation.

## 2. Results

### 2.1. ctDNA Is Shorter Than cfDNA of Noncancer Origin

We performed Cancer Personalized Profiling by deep sequencing (CAPP-seq) cfDNA from twelve stage IV lung cancer patients and seven healthy individuals (Figure 1). The cancer patients had shorter cfDNA than the healthy individuals (Figure 2A). Among the 12 cancer patients, ctDNA was detected in 9/12 patients based on the detection of mutations via CAPP-seq. An analysis of all the cfDNA in mutation-positive cancer patients demonstrated that the length of cfDNA was shorter than that of the cfDNA from mutation-negative patients (Figure 2B). Quantifying the fraction of short cfDNA (<150 bp) revealed similar results for mutation-negative cancer patients and healthy individuals, whereas mutation-positive patients had a greater fraction of short cfDNA fragments than both healthy individuals ( $p = 0.031$ ) and mutation-negative patients ( $p = 0.025$ ) (Figure 2C).



**Figure 2.** cfDNA fragment lengths in cancer patients and healthy individuals. **(A)** The cfDNA fragment size distribution in lung cancer patients ( $n = 12$ , purple) and healthy individuals ( $n = 7$ , orange). Short cfDNA (0–150 bp) is marked in green. **(B)** cfDNA fragment size distribution in mutation-positive ( $n = 9$ , purple) and mutation-negative ( $n = 3$ , orange) lung cancer patients. **(C)** Fraction of cfDNA shorter than 150 bp in healthy individuals ( $n = 7$ , orange), mutation-negative ( $n = 3$ , red) and mutation-positive ( $n = 9$ , purple) lung cancer patients. Groups were compared using an unpaired  $t$ -test. **(D)** Size distribution of mutated and WT cfDNA fragments in the 9 patient samples containing ctDNA mutations. **(E)** The relative mutant/WT cfDNA fraction in bins (10 bp) from 50 to 400 bp. Each point represents the relative mutant/WT cfDNA enrichment compared to the molecular allele fraction (MAF) of the unselected cfDNA pool. Purple indicates cfDNA fragment lengths with mutant enrichment, and orange represents WT cfDNA enrichment. **(F)** Pairwise comparison of the fractions under 150 bp for WT cfDNA ( $n = 9$ , orange) and mutated fragments ( $n = 9$ , purple). Groups were compared using a paired  $t$ -test.

We hypothesize that ctDNA of known cancer cell origin is shorter than cfDNA originating from noncancer cells from the same individual. We combined cfDNA fragments

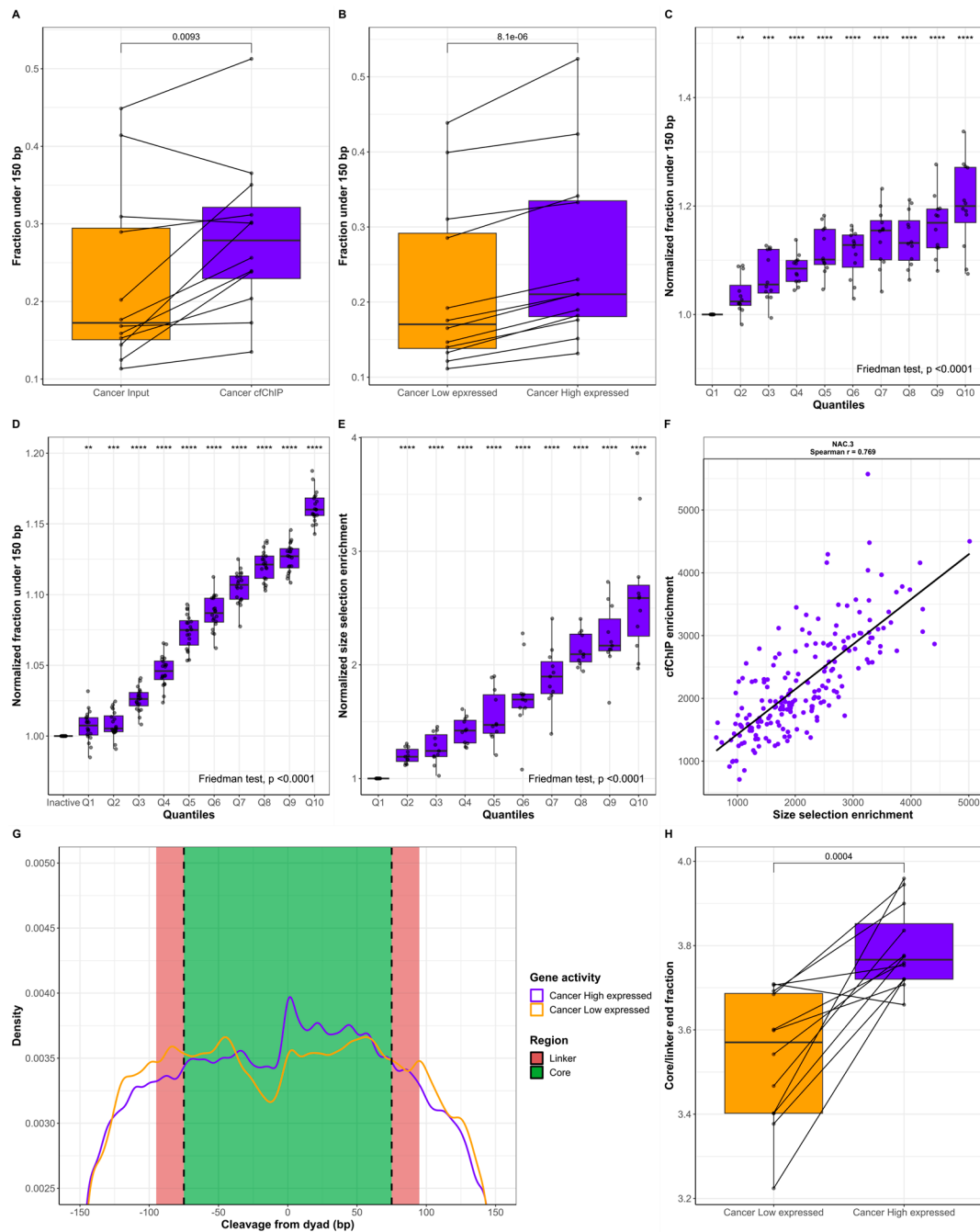
from the nine patients with mutation-positive plasma samples and compared the fragment length of the mutated ctDNA with that of the wild-type (WT) cfDNA fragments at the same genomic position in these samples (Figure 2D, Supplementary Figure S1). This revealed an increase in mutated fragments with a length less than 150 bp and with a length of approximately 300 bp compared to the corresponding cfDNA. We used *in silico* size-selection in bins of 10 bp from 50 to 400 bp to evaluate the effect of cfDNA fragment length on the mutant/WT DNA ratio. For each bin, the mutant/WT DNA ratio was normalized to the ratio in the original combined sample (Figure 2E). This revealed the enrichment of mutant ctDNA from 50 to 150 bp and from 200 to 340 bp, whereas WT cfDNA was most prevalent from 150 to 200 bp and from 340 to 400 bp. A pairwise comparison of mutated ctDNA and WT cfDNA also revealed that the abundance of mutated fragments below 150 bp was greater than that of WT cfDNA in the same genomic region (Figure 2F). Collectively, the results in Figure 2 demonstrate that the length of ctDNA with verified mutations is shorter than WT cfDNA from the same samples and that cancer patients have a greater fraction of short cfDNA than healthy individuals.

### *2.2. cfDNA Originating from Genes with High Expression Is Shorter Than cfDNA from Genes with Low Expression*

Next, we performed cfChIP-seq on 12 cancer patients. The cfChIP-seq of H3K36me3-containing nucleosomes resulted in the enrichment of cfDNA fragments from genes with active transcription [25–30]. A comparison of the cfDNA fragment lengths in cfChIP samples with those in input cfDNA samples from cancer patients revealed the enrichment of shorter cfDNA fragments (Supplementary Figure S2A). Similarly, with a cutoff of 150 bp, we also demonstrated that the fraction under 150 bp was larger in the cfChIP samples than in the matching input samples (Figure 3A,  $p = 0.0093$ ). These results suggest that increased gene activity contributes to the circulatory release of shorter cfDNA fragments. To verify that cfChIP did not result in the artificial trimming of cfDNA fragments, we established two gene sets for each patient representing highly expressed and genes with low levels of expression. We used these gene sets to filter reads from the input cfDNA sample and analyzed the cfDNA fragment lengths of the two gene groups (Supplementary Figure S2B). Quantifying the fraction under 150 bp in the two gene sets revealed that highly expressed genes were systematically represented with shorter cfDNA fragments compared to genes with low expression (Figure 3B,  $p = 8.1 \times 10^{-6}$ ). Results from healthy individuals ( $n = 4$ ) revealed a similar pattern (Supplementary Figure S3), implying that the shortening of cfDNA from active genes is not cancer-specific but is related to cfDNA biology in all individuals. cfChIP data could be obtained from four of the seven healthy individuals as the remaining three healthy individuals had insufficient cfChIP yields for an NGS analysis.

Next, we grouped all genes in the CAPP-seq panel into 10 quantiles (Q) based on cfChIP enrichment for each patient ( $n = 12$ ). In each quantile, the fraction of input cfDNA less than 150 bp was estimated and normalized to Q1, which contained the genes with the lowest cfChIP-based activity (Figure 3C). A stepwise increase in gene activity (Q1 to Q10) resulted in a greater fraction of short fragments (Friedman's ANOVA:  $p < 0.0001$ ). Q10 had a median increase in the <150 bp fraction of 19.99% [IQR: 16.94–27.13%,  $p < 0.0001$ ] compared to that of Q1. These findings were also observed in the validation dataset comprising of healthy individuals (Figure 3D). Here, Q10 had a median increase of 16.01% [IQR: 15.60–16.83%,  $p < 0.0001$ ] in the <150 bp fraction compared to inactive genes. To validate these findings, we performed both *in vitro* size selection and cfChIP-seq on 11 lung cancer plasma samples (Figure 1). Compared with the input samples, the size selection resulted in a large increase in the <150 bp cfDNA fraction (Supplementary Figure S4). Following sequencing, the gene enrichment with size selection was estimated using the same gene enrichment applied to the cfChIP samples [28]. The genes were divided into quantiles Q1 to Q10 based on cfChIP enrichment, and the normalized size-selection enrichment in each quantile is displayed in Figure 3E. An increase in gene activity (Q1 to Q10) resulted in greater *in vitro* size-selection enrichment (Friedman's ANOVA:  $p < 0.0001$ ). The

Q10 median increase in vitro size-selection enrichment was 158.84% (IQR: 125.29–170.11%,  $p < 0.0001$ ) compared to Q1. A correlation of cfChIP gene enrichment with in vitro size-selection gene enrichment revealed a positive correlation (Spearman  $r$  range: 0.499–0.882,  $p < 2.2 \times 10^{-16}$ ; Figure 3F, Supplementary Figure S5). This finding indicates that the in vitro size selection of plasma cfDNA can be used to predict gene activity from a liquid biopsy. In silico size selection at different fragment length cutoffs also resulted in a positive correlation with cfChIP (Supplementary Figure S6). The highest correlation was observed for the fraction of fragments less than 160 bp (Spearman  $r$  median: 0.403, range: 0.158–0.528), whereas the correlation was reduced for cutoffs above 160 bp.



**Figure 3.** Fragment length analysis of cfDNA in active genes. (A) Pairwise comparison of the fraction under 150 bp for input ( $n = 12$ , orange) and cfChIP ( $n = 12$ , purple) samples. (B) Pairwise comparison of the fraction under 150 bp for genes with low expression ( $n = 12$ , orange) and high expression ( $n = 12$ , purple). (C) Normalized fraction under 150 bp for genes in the cfChIP quantiles Q1 to Q10 for lung



cancer patients ( $n = 12$ ). (D) Normalized fraction under 150 bp for cfDNA in healthy individuals. The fragments are grouped as inactive or in quantiles Q1 to Q10. (E) Normalized in vitro size-selection enrichment for genes in the cfChIP quantiles Q1 to Q10 for lung cancer patients ( $n = 11$ ). For (C,E), each quantile is normalized to the fraction in Q1, whereas quantiles in (D) are normalized to inactive genes. (F) Representative example of cfChIP enrichment compared to size-selection enrichment. NAC.3 represents the median Spearman's correlation between the 11 samples. (G) Distribution of fragment end cleavage relative to the nucleosome dyad in lung cancer patients ( $n = 12$ ). Purple represents highly expressed genes and orange represents genes with low levels of expression. (H) Pairwise comparison of the core/linker end fraction for genes with low expression ( $n = 12$ , orange) and high expression ( $n = 12$ , purple). For (A,B,H), groups were compared using a paired *t*-test. \*\*:  $p \leq 0.01$ , \*\*\*:  $p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ .

We hypothesize that the underlying chromatin structure could explain the differences in fragment lengths between active and inactive genes. To investigate this possibility, we estimated the cleavage of cfDNA relative to nucleosome centers which were determined in a previous study [24]. Figure 3G displays the frequency of fragment ends in the core and linker regions of a nucleosome for highly expressed genes and genes with low levels of expression. Compared with the genes with low levels of expression, the highly expressed genes have an increased core/linker end fraction ( $p = 0.0004$ , Figure 3H). This finding demonstrates that gene activity results in changes in chromatin structure which are reflected in cfDNA fragment end distributions and fragment lengths.

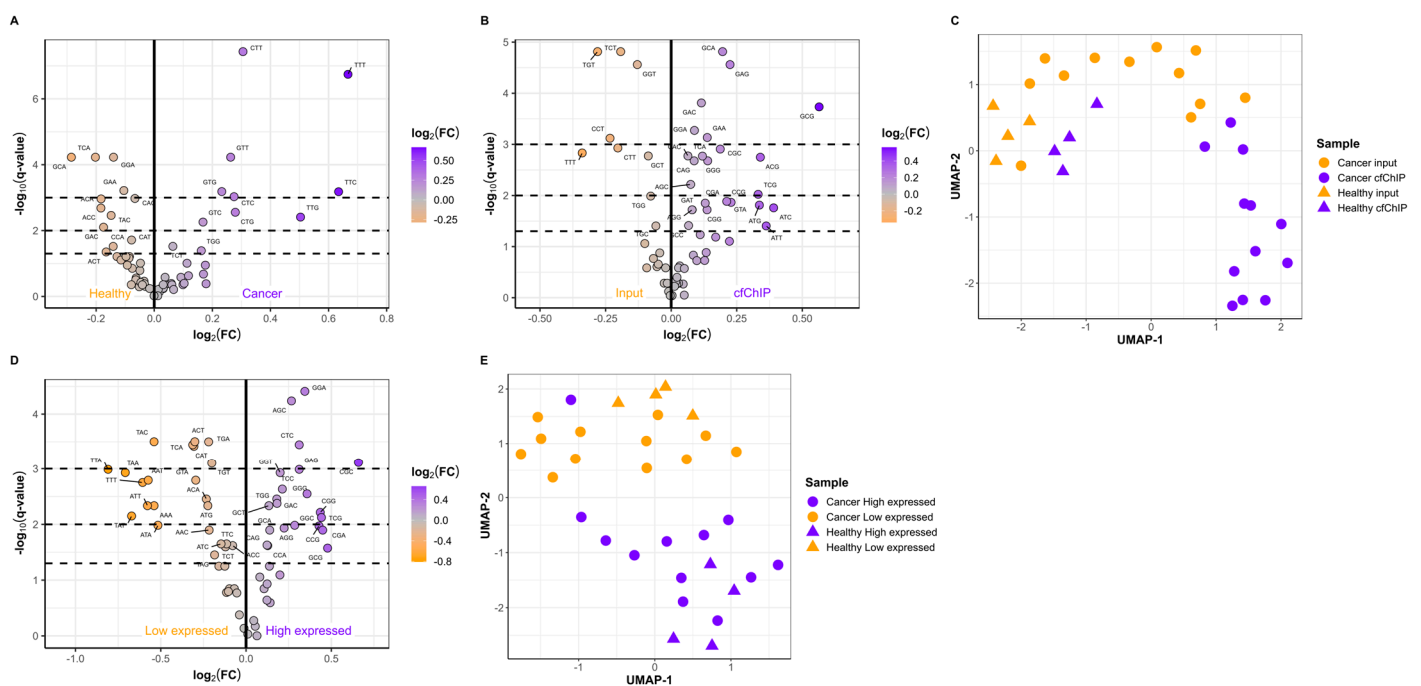
### 2.3. The Active Genes Have Distinct cfDNA FEMs

A comparison of the cfDNA fragment end motifs in the input samples from healthy individuals ( $n = 7$ ) and cancer patients ( $n = 12$ ) revealed differences between the two types of samples (Figure 4A). In addition, we compared the FEM frequencies between paired input and cfChIP samples. These findings demonstrated that 24 FEMs were enriched in the cfChIP samples from cancer patients (Figure 4B), healthy individuals (Supplementary Figure S7A), and all individuals combined (Supplementary Figure S7B). Uniform Manifold Approximation and Projection (UMAP) embedding based on FEM frequencies in input and cfChIP samples from both cancer patients and healthy individuals resulted in distinct clusters (Figure 4C). Samples from healthy individuals were grouped together but were also separated into input and samples containing active genes based on cfChIP enrichment. Moreover, cancer input and cancer cfChIP samples were also separated into two clusters. This finding is in accordance with our results that cfChIP and input samples have distinct FEM frequencies (Figure 4B). The genes associated with low expression and high expression also had a distinct enrichment of FEMs (Figure 4D, Supplementary Figure S7C,D). The FEMs enriched in the cfChIP samples were generally GC-rich compared to those enriched in input samples (Supplementary Figure S8). Similarly, compared with genes with low expression, genes with high expression were also enriched in GC-rich FEMs (Supplementary Figure S8). Interestingly, a significant overlap in enriched FEMs for cfChIP samples and highly expressed genes was observed ( $p = 5.08 \times 10^{-6}$ ). UMAP embedding FEM frequencies in genes with low and high expression separated the two gene sets into two groups (Figure 4E). The two clusters represent genes with low and high expression, respectively, indicating that cfDNA fragments from genes with low expression have distinct FEMs compared to cfDNA originating from active genes.

### 2.4. Short Fragments Have Distinct FEMs and Originate from Active Genes

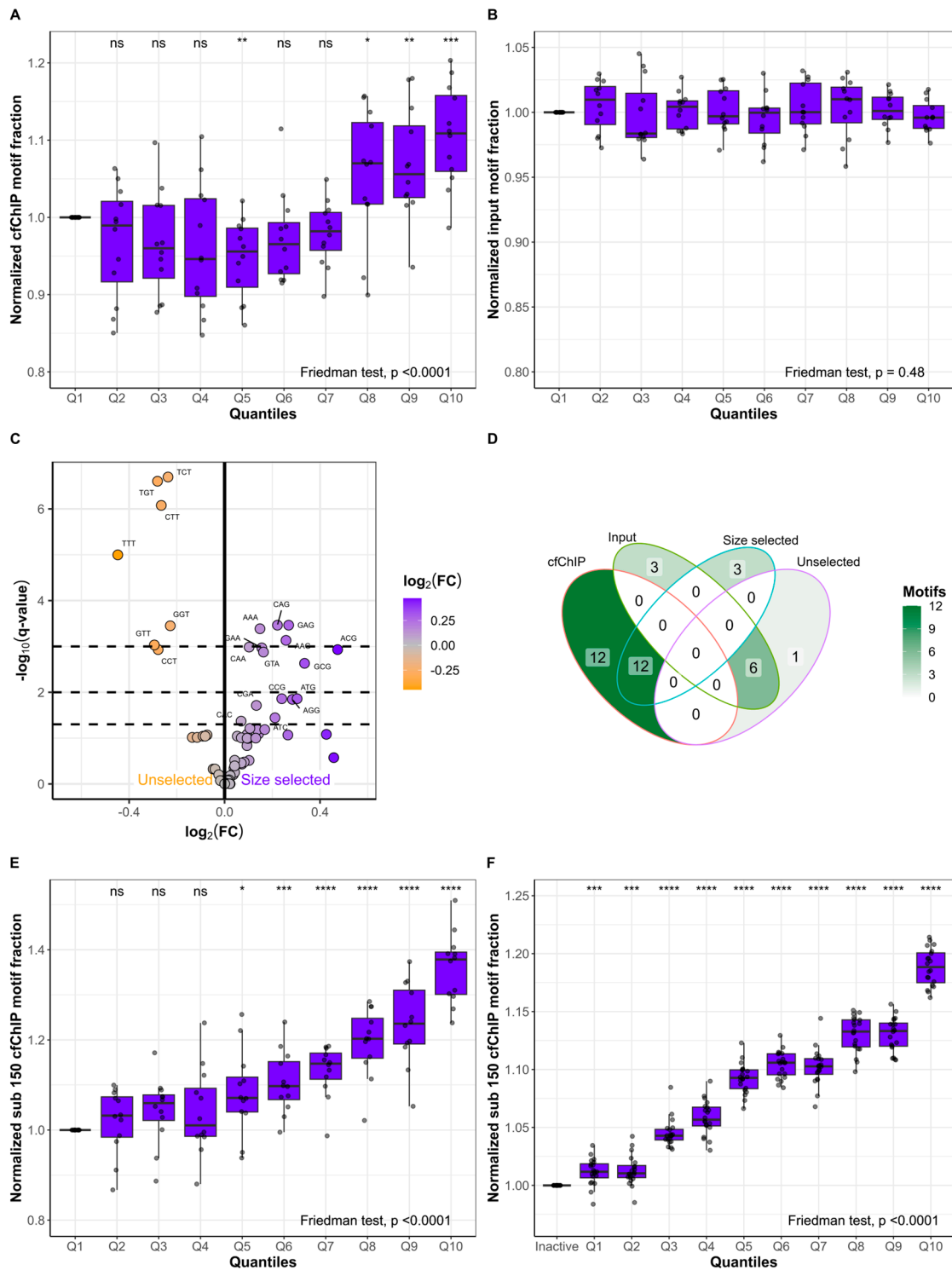
To determine whether subsets of the 64 possible 3 bp long FEMs of cfDNA could predict gene expression, we defined the 24 FEMs enriched in the 12 lung cancer cfChIP samples as "cfChIP motifs" (Figure 4B). The nine FEMs enriched in the input samples were defined as "input motifs". For each lung cancer patient, we grouped all genes based on cfChIP quantiles and quantified the fraction of cfDNA fragments with cfChIP and input motifs (Figure 5A,B). The fraction of fragments with cfChIP motifs increased from Q1 to Q10 (Friedman's ANOVA:  $p < 0.0001$ ). Q10 had a median increase in the fraction of

fragments with cfChIP motifs of 10.87% [IQR: 5.95–15.78%,  $p < 0.001$ ] compared to that of Q1 (Figure 5A). In contrast, the fraction of fragments with input motifs was consistent from Q1 to Q10 (Friedman’s ANOVA:  $p = 0.48$ ; Figure 5B). Similarly, we defined high-expression and low-expression motifs based on the FEMs enriched in highly expressed genes and genes with low levels of expression, respectively (Figure 4D). The fraction of input cfDNA fragments with motifs from highly expressed genes increased from Q1 to Q10 (Friedman’s ANOVA:  $p < 0.0001$ , Supplementary Figure S9A). Correspondingly, the fraction of fragments with motifs from genes associated with low expression decreased from Q1 to Q10 (Friedman’s ANOVA:  $p < 0.0001$ ; Supplementary Figure S9B). Given that distinct FEMs were identified in active genes and that active genes are enriched in short cfDNA fragments, we investigated the FEMs in size-selected samples compared to unselected samples (Figure 5C). Many of the FEMs enriched in the size-selected samples were GC-rich, similar to the cfChIP motifs, and 12/15 (80%) FEMs enriched in the size-selected samples were also enriched in the cfChIP samples ( $p = 0.00018$ ; Figure 5D). These results indicate that cfDNA originating from active genes contains more short fragments (<150 bp) and a specific subset of FEMs. We combined this knowledge to quantify the fraction of input cfDNA fragments shorter than 150 bp and with cfChIP motifs in gene quantiles based on cfChIP enrichment (Figure 5E). A stepwise increase was observed from Q1 to Q10 (Friedman’s ANOVA:  $p < 0.0001$ ). This indicates that increases in gene activity affect cfDNA fragmentation, resulting in short cfDNA molecules with distinct FEMs. Q10 had a median increase in the fraction of fragments less than 150 bp and with cfChIP motifs of 37.85% [IQR: 30.10–39.49%,  $p < 0.0001$ ] compared to Q1. Combining both fragment lengths and FEMs resulted in greater enrichment in Q10 than fragment lengths alone (median increase = 16.78%, IQR: 8.60–18.20%,  $p < 0.0001$ ; Figure 3C) or cfChIP motifs alone (median increase = 20.35%, IQR: 16.87–32.40%,  $p < 0.0001$ ; Figure 5A). Similar results were observed in the validation dataset comprising healthy individuals (Figure 5F). Here, the median increase in Q10 compared to that in inactive genes was 18.84% [IQR: 17.49–20.07%,  $p < 0.0001$ ], which was also greater than when evaluating fragment lengths alone (median increase = 2.16%, IQR: 1.18–3.51%,  $p < 0.0001$ ; Figure 3D) or FEMs alone (median increase = 15.76%, IQR: 14.40–16.48%,  $p < 0.0001$ ; Supplementary Figure S10).



**Figure 4.** Active and inactive genes can be distinguished based on fragment end motifs (FEMs). (A) Differences in FEM frequencies between cancer patients ( $n = 12$ ) and healthy individuals ( $n = 7$ ). (B) Differences in FEMs in paired input and cfChIP cancer samples ( $n = 12$ ). (C) Uniform Manifold

Approximation and Projection (UMAP) based on FEM frequencies in the input and cfChIP samples. (D) Differences in FEMs for genes expressed at low and high levels in cancer input samples ( $n = 12$ ). (E) UMAP based on the FEM frequencies in genes expressed at low and high levels. For (A,B), and (D), FEMs with a  $q$ -value less than 0.05 are labeled with the motif. Dashed lines indicate  $q = 0.05$ ,  $q = 0.01$ , and  $q = 0.001$ . For (C,E), cancer patients are represented as circles, and healthy individuals are represented as triangles. The colors indicate the type of sample.



**Figure 5.** Fragment end motifs from short fragments predict gene expression. For each lung cancer patient ( $n = 12$ ), the genes in the cfChIP quantiles were determined. For each quantile, the fraction of fragments with fragment end motifs enriched in cfChIP (A) or input (B) samples was calculated.



(C) Differences in fragment end motifs in paired unselected and size-selected cancer samples ( $n = 11$ ). (D) Venn diagram representing the overlap in significant fragment end motifs in the analyses of cfChIP compared to input samples and unselected compared to size-selected samples. (E) Normalized fraction of fragments less than 150 bp with cfChIP motifs for genes in cfChIP quantiles Q1 to Q10 for lung cancer patients ( $n = 12$ ). (F) Normalized fraction of fragments less than 150 bp with cfChIP motifs for genes in the validation dataset. The fragments are grouped as inactive or in quantiles Q1 to Q10. For (A,B,E), each quantile is normalized to the fraction in Q1, whereas quantiles in (F) are normalized to inactive genes. ns:  $p > 0.05$ , \*:  $p \leq 0.05$ , \*\*:  $p \leq 0.01$ , \*\*\*:  $p \leq 0.001$ , \*\*\*\*:  $p \leq 0.0001$ .

### 3. Discussion

In this study, we conducted cfChIP-seq on plasma from non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) patients to generate a personal gene expression profile for each individual. In a previous study, we described how cfChIP-seq on these patients can be used to study gene expression differences between the two types of lung cancers [28]. We demonstrated how cfChIP-seq on NSCLC patients resulted in the enrichment of genes known to be upregulated in NSCLC, such as *EGFR* and *RIN3*, whereas SCLC patients had increased gene enrichment of, e.g., *KIF19* and *SMAD4*, which are upregulated in SCLC tumors. This highlights the fact that cfChIP-seq results reflect tumor gene expression and the fragmentomic features identified in this study arise from the tumor cells.

Previous studies used cfDNA fragment end motif frequencies in machine learning algorithms for early cancer detection and treatment response evaluation [9,21,22,31]. In this study, we focused on the effect of gene activity on FEMs. In a recent study [32], the authors linked FEM frequencies to the methylation status of cfDNA fragments, which indirectly reflects the gene activity in the cell of origin. The authors demonstrated how FEM frequencies vary at different CpG sites, reflecting the methylation status, and they used this knowledge to perform a tissue of origin analyses. It is likely that the FEMs identified in our study that are connected to gene activity are also affected by the methylation of the gene. Future studies are required to expand the body of knowledge concerning the associations between gene activity, methylation status, and cfDNA FEMs.

The ability of cfDNA fragmentomics to infer gene expression has been developed in recent years. These studies have focused primarily on the cfDNA fragmentation profiles surrounding TSS [19,23], open chromatin regions [17], and TFBSs [18,33]. Given that the H3K36me3 modification, which was used in our cfChIP assays, is located primarily in the gene body of active genes [29,30], and that the applied NGS panel is not focused around the TSS [28], our results preferentially reflected the cfDNA fragmentation within the gene body. Nevertheless, we observed that increased gene activity results in shorter cfDNA fragments, which is in accordance with the findings of a previous study [23]. Our results, based on cfChIP-seq, were verified using an in vitro size-selection enrichment of cfDNA, as the results from these two approaches were found to be correlated. Combining FEMs with fragment lengths increased the association with cfDNA fragments originating from active genes, indicating that increased gene activity results in both shorter fragments and distinct fragment end motifs. FEMs are a new field in liquid biopsy research, and multiple factors, including nuclease activity [20,34,35] and methylation [32], can affect FEMs. Future whole genome sequencing (WGS)-based studies of FEMs and cfChIP in combination with in vitro cfDNA size selection could further validate our findings and provide comprehensive insights into the relationship between cfDNA fragmentomics and underlying gene activity.

cfDNA fragment size distribution differs between cancer patients and healthy individuals [4,15,36], which was also observed in this study. Previous studies have identified how cancer patients have increased amounts of short fragments [4,15] and high molecular fragments caused by cell necrosis [37]. In addition, we evaluated the lengths of the mutated ctDNA fragments and compared them to WT cfDNA fragments in the same patient. A limitation of these analyses is that only 9 of the 12 lung cancer patients were mutation-positive and the number of ctDNA mutations was low (median = 3, range 0–6). However, we are

able to demonstrate an enrichment of mutated ctDNA in short cfDNA fragments. This result is in alignment with a previous study [10].

A limitation of our study is the relatively few individuals included (twelve cancer patients and seven healthy individuals) and the size of our gene panel used for targeted NGS. In addition, only stage IV lung cancer patients with a high ctDNA fraction were analyzed. As a result, our findings need to be repeated in larger patient cohorts and expanded to include other types of cancer, patients of different ethnicities and ages, as well as the full spectrum of cancer stages. Likewise, it would be interesting to perform these experiments using a broader NGS approach, such as WGS, to generalize the information obtained from cfDNA fragmentomics. Despite this, we were able to replicate our findings in cfDNA from healthy individuals using WGS data from a previous study [24] in which the expected gene activity was based on data from peripheral blood mononuclear cells (PBMCs). These findings suggest that our results are applicable to other types of NGS methods and that the fragmentomic features identified in our study can be generalized to other samples. Future studies will demonstrate how gene expression based on liquid biopsies can be used to monitor tumor development and identify treatment responses using longitudinal plasma samples from the same individuals.

#### 4. Materials and Methods

##### 4.1. Plasma Samples

Plasma samples from twelve stage IV lung cancer patients diagnosed with either NSCLC ( $n = 8$ ) or SCLC ( $n = 4$ ) and from seven healthy individuals were used. Peripheral blood was drawn in 10 mL EDTA tubes from each individual and centrifuged at room temperature within 2 h at  $1400 \times g$  for 15 min. Plasma was aliquoted and stored at  $-80^\circ\text{C}$ . The plasma was split and subjected to (1) input cfDNA purification, (2) cfChIP, or (3) cfDNA purification and in vitro size selection (Figure 1). One NSCLC patient had insufficient remaining plasma for in vitro size selection, and therefore, 11/12 lung cancer patients were subjected to in vitro size selection. All three types of samples were subjected to NGS via CAPP-seq [2], as described in detail below. The amount of plasma and the resulting number of reads for each sample are presented in Supplementary Table S1.

##### 4.2. Cell-Free Chromatin Immunoprecipitation (cfChIP) Enrichment

cfChIP-seq was performed as described by Maansson et al. [28] (Figure 1, left). H3K36me3 cfChIP enrichment was calculated for each gene, and the results were used as a surrogate for gene activity [25,27]. In brief, the plasma was cleared of circulating antibodies using empty protein A/G magnetic beads (ThermoFisher Scientific, 88802, Waltham, MA, USA). The cleared plasma was added to protein A/G magnetic beads coated with anti-H3K36me3 antibodies (Abcam 9050, Cambridge, United Kingdom) and incubated overnight at  $4^\circ\text{C}$ . The beads were then washed, and the precipitated DNA was eluted. The cfChIP sample and the input plasma cfDNA were purified using an Apostle MiniMax High Efficiency cfDNA Isolation Kit (Beckman Coulter, Indianapolis, IN, USA) and subjected to CAPP-seq. Deduplicated cfChIP BAM files were used for a H3K36me3 enrichment analysis. For each gene, the number of reads was normalized to the total number of reads and the NGS coverage of that particular gene [28].

##### 4.3. In Vitro Size Selection

A PippinHT (Sage Science, Beverly, MA, USA) instrument was used to perform the in vitro size-selection of cfDNA (Figure 1, right). First, 20  $\mu\text{L}$  of purified cfDNA from each plasma sample was added to a 3% agarose gel cassette. The range mode of the system was set to collect cfDNA within 95–230 bp with a pause at 152 bp, resulting in the collection of two cfDNA fractions. The two fractions corresponded to cfDNA fragments with lengths of 95–152 bp and 152–230 bp, respectively. The cfDNA fragments were eluted in 30  $\mu\text{L}$  of electrophoresis buffer. The fraction with short fragments (95–152 bp) was subjected to CAPP-seq.

#### 4.4. CAPP-Seq

CAPP-seq is based on hybridization-capture NGS [2]. The sequencing library was prepared using the AVENIO ctDNA surveillance kit (Roche Sequencing Solutions, Mannheim, Germany), covering 198 kb in 197 hypermutated genes [38]. The gene fragments were sequenced using a NextSeq 500 (Illumina, San Diego, CA, USA), and the data were analyzed using the AVENIO Oncology Analysis Software (v. 2.0.0). Deduplicated BAM files based on unique molecular identifiers (UMIs) were used for a cfChIP enrichment analysis, whereas position-deduplicated BAM files were used to estimate fragment lengths and FEMs.

#### 4.5. ctDNA Detection

Mutations in cfDNA were identified using the AVENIO Oncology Analysis software v. 2.0.0 (Roche Sequencing Solutions, Mannheim, Germany), as described previously [39]. The excluded variants included synonymous variants and variants present in the Exome Aggregation Consortium (ExaC) with an allele frequency > 0.10%, the 1000 Genomes Project, or the Single Nucleotide Polymorphism database (dbSNP Common). The included variants needed to be present in the Catalog of Somatic Mutations in Cancers (COSMIC) database, The Cancer Genome Atlas (TCGA), or the Loci of Interest list. Mutated cfDNA fragments were isolated as described previously [12], and fragments covering the same genomic position without the mutation were defined as WT fragments.

#### 4.6. Fragment Length and FEM Analyses

Position-deduplicated BAM files were used for fragment length and FEM analyses. Fragment lengths were estimated using the start and end positions of paired-end sequencing reads. Reads with no fragment length information or a fragment length of 0 were excluded from the analysis. Reads containing N nucleotides in either of the fragment termini were removed. The frequencies of 5'-3' 3-mers at each end of a fragment were determined. The frequencies at both ends of the fragments were combined, and the fraction of each possible 3-mer ( $n = 4^3 = 64$ ) for each sample was calculated. The FEM frequencies for all 64 possible motifs were used for a Uniform Manifold Approximation and Projection (UMAP) analysis.

#### 4.7. High/Low-Expressed Genes and cfChIP Quantiles

Gene expression was estimated using cfChIP-seq for each individual, and the 15 most enriched genes were considered to be highly expressed, whereas the 15 least enriched genes were considered to be expressed at low levels. Furthermore, genes were grouped in cfChIP quantiles for each patient. Each quantile contained 19 or 20 genes such that Q1 contained the genes with the lowest enrichment and Q10 contained the most enriched genes. Reads from cfDNA input samples in either high/low-expressed genes or cfChIP quantiles were filtered, and fragment lengths and FEMs were analyzed.

#### 4.8. Nucleosomal Positioning Analysis

For each cfDNA fragment in high- or low-expressed genes, we estimated the distance from the nearest nucleosome center to the cfDNA fragment ends as described previously [15]. Information regarding nucleosome positions was based on the CH01 nucleosome track from Snyder et al. [24]. This track is based on the pooled whole-genome sequencing of cfDNA from multiple healthy individuals. We identified input cfDNA fragments overlapping with nucleosome centers and calculated the distance from each nucleosome center to the start and end of the fragment. For fragments overlapping with multiple nucleosome centers, only the nearest nucleosome to each start and end was considered. We defined the nucleosome core as  $-75$  to  $+75$  bp relative to the nucleosome dyad and the linker region as  $-75$  to  $-95$  or  $+75$  or  $+95$  bp relative to the nucleosome dyad. Based on these regions, we calculated a core/linker end fraction as the number of fragment ends in the core region relative to the number of fragment ends in the linker region.

#### 4.9. Validation Data

cfDNA WGS data published by Snyder et al. [24] were used to validate the findings in this study. The BH01 (GSM1833219) cfDNA-seq dataset representing the deep WGS of an unknown number of healthy individuals was down sampled to 20 0.9X WGS samples. This dataset is referred to as the “validation dataset”. Furthermore, RNA-seq data from PBMCs (GSE107011,  $n = 13$ ) [40] were used to determine gene expression patterns in healthy individuals. Reads in the validation dataset were grouped into gene sets based on PBMC gene activity. Genes with no expression ( $\log_2(\text{TPM} + 1) = 0$  in all samples) were classified as “inactive” ( $n = 2262$ ). The remaining genes were grouped into quantiles with 1747 to 1751 genes in each quantile. Q1 contained the genes with the lowest average  $\log_2(\text{TPM} + 1)$ , and Q10 contained the genes with the highest average  $\log_2(\text{TPM} + 1)$ .

#### 4.10. Statistical Analysis

All the data analyses were performed in R version 4.3.0. BAM files were loaded into R using the chromstaR [41] package. cfDNA fragments were analyzed using the GenomicRanges [42] Biostrings [43], and Rsamtools [44] packages. Plots and graphs were created using the dplyr [45], ggpubr [46], and ggplot2 [47] packages. Differences in both fragment lengths and FEMs between the input and cfChIP samples were evaluated with a paired *t*-test. A comparison of highly expressed genes and genes with low levels of expression from the same individual was also performed with a paired *t*-test. Differences between cancer patients and healthy individuals as well as between mutation-positive and mutation-negative patients were tested using an unpaired *t*-test. Global differences in fragmentation for cfChIP quantiles were tested using Friedman’s ANOVA, and if  $p < 0.05$ , Q2–Q10 were compared individually to Q1 using paired *t*-tests. For the FEM analysis, *p*-values were adjusted for multiple testing using the false discovery rate (FDR) [48], and *q*-values  $< 0.05$  were considered significant. Significant overlap between FEM sets was determined using a hypergeometric test, where  $p < 0.05$  was considered to indicate statistical significance.

### 5. Conclusions

In this study, we identified key fragmentomic features related to cfDNA from active genes. Based on the cfChIP-seq of cfDNA, which is a novel technique for measuring gene activity in liquid biopsies, we created a personal gene expression profile and used this information to detect differences in fragmentomics between highly expressed genes and genes with low levels of expression. Here, we demonstrated that active transcription is associated with shorter cfDNA fragments and more GC-rich FEMs. Interestingly, combining the frequency of short fragments and distinct FEMs resulted in greater enrichment in highly expressed genes than evaluating fragment lengths or FEMs separately. This was also observed in the validation dataset from healthy individuals. This study provides novel insights into the biology behind cfDNA fragmentomics and demonstrates that FEMs are an important new characteristic of cfDNA which can increase the utility of liquid biopsies for determining gene activity under various pathological conditions.

**Supplementary Materials:** The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms25021243/s1>.

**Author Contributions:** Conceptualization, C.T.M., A.L.N. and B.S.S.; data curation, C.T.M., L.S.T. and P.M.; formal analysis, C.T.M.; funding acquisition, A.L.N. and B.S.S.; investigation, C.T.M. and B.S.S.; methodology, C.T.M. and L.S.T.; project administration, B.S.S.; resources, P.M., A.L.N. and B.S.S.; software, C.T.M.; supervision, P.M., A.L.N. and B.S.S.; validation, C.T.M.; visualization, C.T.M.; writing—original draft, C.T.M., A.L.N. and B.S.S.; writing—review and editing, C.T.M., L.S.T., P.M., A.L.N. and B.S.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the following grants: Vilhelm Pedersen and Hustrus Legat, Dagmar Marshalls Mindelegat, Fabrikant Einar Willumsens Mindelegat, Marie og Børge Krogs Fond, P.A. Messerschmidt og Hustrus Fond, Familien Erichsens Familiefond, and Kræftfonden.



**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki and approved by the Central Denmark Region Committee on Biomedical Research Ethics (no. 1-10-72-83-14, 2014-09).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** This study used the following publicly available datasets: RNA-seq of PBMCs (GSE107011), WGS of cfDNA in healthy individuals (GSM1833219; BH01), and nucleosome position information based on the cfDNA WGS of healthy individuals (GSE71378; CH01). The cfChIP-seq data are from a previous study from our group [28]. The remaining data associated with this study are available upon request from the corresponding author. The data are not publicly available due to privacy concerns.

**Acknowledgments:** We thank Tina Fuglsang Daugaard and Birgit Westh Mortensen for excellent technical assistance. We thank all of the plasma donors included in the project.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Diehl, F.; Schmidt, K.; Choti, M.A.; Romans, K.; Goodman, S.; Li, M.; Thornton, K.; Agrawal, N.; Sokoll, L.; Szabo, S.A.; et al. Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **2008**, *14*, 985–990. [[CrossRef](#)] [[PubMed](#)]
2. Newman, A.M.; Bratman, S.V.; To, J.; Wynne, J.F.; Eclow, N.C.; Modlin, L.A.; Liu, C.L.; Neal, J.W.; Wakelee, H.A.; Merritt, R.E.; et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **2014**, *20*, 548–554. [[CrossRef](#)]
3. Oxnard, G.R.; Paweletz, C.P.; Kuang, Y.; Mach, S.L.; O’Connell, A.; Messineo, M.M.; Luke, J.J.; Butaney, M.; Kirschmeier, P.; Jackman, D.M.; et al. Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. *Clin. Cancer Res.* **2014**, *20*, 1698–1705. [[CrossRef](#)]
4. 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](#)]
5. Xi, L.; Pham, T.H.; Payabyab, E.C.; Sherry, R.M.; Rosenberg, S.A.; Raffeld, M. Circulating Tumor DNA as an Early Indicator of Response to T-cell Transfer Immunotherapy in Metastatic Melanoma. *Clin. Cancer Res.* **2016**, *22*, 5480–5486. [[CrossRef](#)] [[PubMed](#)]
6. Weber, S.; van der Leest, P.; Donker, H.C.; Schlange, T.; Timens, W.; Tamminga, M.; Hasenleithner, S.O.; Graf, R.; Moser, T.; Spiegl, B.; et al. Dynamic Changes of Circulating Tumor DNA Predict Clinical Outcome in Patients With Advanced Non-Small-Cell Lung Cancer Treated With Immune Checkpoint Inhibitors. *JCO Precis. Oncol.* **2021**, *5*, 1540–1553. [[CrossRef](#)] [[PubMed](#)]
7. Tie, J.; Kinde, I.; Wang, Y.; Wong, H.L.; Roebert, J.; Christie, M.; Tacey, M.; Wong, R.; Singh, M.; Karapetis, C.S.; et al. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Ann. Oncol.* **2015**, *26*, 1715–1722. [[CrossRef](#)]
8. Gale, D.; Heider, K.; Ruiz-Valdepenas, A.; Hackinger, S.; Perry, M.; Marsico, G.; Rundell, V.; Wulff, J.; Sharma, G.; Knock, H.; et al. Residual ctDNA after treatment predicts early relapse in patients with early-stage non-small cell lung cancer. *Ann. Oncol.* **2022**, *33*, 500–510. [[CrossRef](#)]
9. Budhraj, K.K.; McDonald, B.R.; Stephens, M.D.; Contente-Cuomo, T.; Markus, H.; Farooq, M.; Favaro, P.F.; Connor, S.; Byron, S.A.; Egan, J.B.; et al. Genome-wide analysis of aberrant position and sequence of plasma DNA fragment ends in patients with cancer. *Sci. Transl. Med.* **2023**, *15*, eabm6863. [[CrossRef](#)]
10. Chabon, J.J.; Hamilton, E.G.; Kurtz, D.M.; Esfahani, M.S.; Moding, E.J.; Stehr, H.; Schroers-Martin, J.; Nabet, B.Y.; Chen, B.; Chaudhuri, A.A.; et al. Integrating genomic features for non-invasive early lung cancer detection. *Nature* **2020**, *580*, 245–251. [[CrossRef](#)]
11. Kim, Y.J.; Jeon, H.; Jeon, S.; Lee, S.H.; Kim, C.; Ahn, J.H.; Um, H.; Woo, Y.J.; Jeong, S.H.; Kim, Y.; et al. A method for early diagnosis of lung cancer from tumor originated DNA fragments using plasma cfDNA methylome and fragmentome profiles. *Mol. Cell. Probes* **2022**, *66*, 101873. [[CrossRef](#)] [[PubMed](#)]
12. Vessies, D.C.L.; Schuurbijs, M.M.F.; van der Noort, V.; Schouten, I.; Linders, T.C.; Lanfermeijer, M.; Ramkisoensing, K.L.; Hartemink, K.J.; Monkhorst, K.; van den Heuvel, M.M.; et al. Combining variant detection and fragment length analysis improves detection of minimal residual disease in postsurgery circulating tumour DNA of stage II-III NSCLC patients. *Mol. Oncol.* **2022**, *16*, 2719–2732. [[CrossRef](#)] [[PubMed](#)]
13. Mouliere, F.; Robert, B.; Arnau Peyrotte, E.; Del Rio, M.; Ychou, M.; Molina, F.; Gongora, C.; Thierry, A.R. High fragmentation characterizes tumour-derived circulating DNA. *PLoS ONE* **2011**, *6*, e23418. [[CrossRef](#)] [[PubMed](#)]
14. Ishida, Y.; Takano, S.; Maekawa, S.; Yamaguchi, T.; Yoshida, T.; Kobayashi, S.; Iwamoto, F.; Kuno, T.; Hayakawa, H.; Matsuda, S.; et al. Fractionated small cell-free DNA increases possibility to detect cancer-related gene mutations in advanced colorectal cancer. *JGH Open* **2020**, *4*, 978–986. [[CrossRef](#)]
15. Markus, H.; Chandrananda, D.; Moore, E.; Mouliere, F.; Morris, J.; Brenton, J.D.; Smith, C.G.; Rosenfeld, N. Refined characterization of circulating tumor DNA through biological feature integration. *Sci. Rep.* **2022**, *12*, 1928. [[CrossRef](#)]



16. Jiang, P.; Chan, C.W.; Chan, K.C.; Cheng, S.H.; Wong, J.; Wong, V.W.; Wong, G.L.; Chan, S.L.; Mok, T.S.; Chan, H.L.; et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1317–E1325. [[CrossRef](#)] [[PubMed](#)]
17. 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](#)]
18. Ulz, P.; Perakis, S.; Zhou, Q.; Moser, T.; Belic, J.; Lazzeri, I.; Wölfler, 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](#)]
19. 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](#)]
20. 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](#)]
21. Guo, W.; Chen, X.; Liu, R.; Liang, N.; Ma, Q.; Bao, H.; Xu, X.; Wu, X.; Yang, S.; Shao, Y.; et al. Sensitive detection of stage I lung adenocarcinoma using plasma cell-free DNA breakpoint motif profiling. *EBioMedicine* **2022**, *81*, 104131. [[CrossRef](#)] [[PubMed](#)]
22. Wang, Y.; Fan, X.; Bao, H.; Xia, F.; Wan, J.; Shen, L.; Wang, Y.; Zhang, H.; Wei, Y.; Wu, X.; et al. Utility of Circulating Free DNA Fragmentomics in the Prediction of Pathological Response after Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin. Chem.* **2023**, *69*, 88–99. [[CrossRef](#)] [[PubMed](#)]
23. 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](#)]
24. 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](#)]
25. Månsson, 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](#)] [[PubMed](#)]
26. 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](#)] [[PubMed](#)]
27. Vad-Nielsen, J.; Meldgaard, P.; Sorensen, B.S.; Nielsen, A.L. Cell-free Chromatin Immunoprecipitation (cfChIP) from blood plasma can determine gene-expression in tumors from non-small-cell lung cancer patients. *Lung Cancer* **2020**, *147*, 244–251. [[CrossRef](#)]
28. Trier Maansson, C.; Meldgaard, P.; Stougaard, M.; Nielsen, A.L.; Sorensen, B.S. Cell-free chromatin immunoprecipitation can determine tumor gene expression in lung cancer patients. *Mol. Oncol.* **2023**, *17*, 722–736. [[CrossRef](#)]
29. Huang, C.; Zhu, B. Roles of H3K36-specific histone methyltransferases in transcription: Antagonizing silencing and safeguarding transcription fidelity. *Biophys. Rep.* **2018**, *4*, 170–177. [[CrossRef](#)]
30. Lerner, A.M.; Hepperla, A.J.; Keele, G.R.; Meriesh, H.A.; Yumerefendi, H.; Restrepo, D.; Zimmerman, S.; Bear, J.E.; Kuhlman, B.; Davis, I.J.; et al. An optogenetic switch for the Set2 methyltransferase provides evidence for transcription-dependent and -independent dynamics of H3K36 methylation. *Genome Res.* **2020**, *30*, 1605–1617. [[CrossRef](#)]
31. Zhou, H.; Zhang, X.; Liu, Q.; Yang, J.; Bai, J.; Yin, M.; Cao, D.; Zhang, Q.; Zheng, L. Can circulating cell free DNA be a promising marker in ovarian cancer?—A genome-scale profiling study in a single institution. *J. Ovarian Res.* **2023**, *16*, 11. [[CrossRef](#)]
32. Zhou, Q.; Kang, G.; Jiang, P.; Qiao, R.; Lam, W.K.J.; Yu, S.C.Y.; Ma, M.L.; Ji, L.; Cheng, S.H.; Gai, W.; et al. Epigenetic analysis of cell-free DNA by fragmentomic profiling. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2209852119. [[CrossRef](#)] [[PubMed](#)]
33. Mathios, D.; Johansen, J.S.; Cristiano, S.; Medina, J.E.; Phallen, J.; Larsen, K.R.; Bruhm, D.C.; Niknafs, N.; Ferreira, L.; Adleff, V.; et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nat. Commun.* **2021**, *12*, 5060. [[CrossRef](#)] [[PubMed](#)]
34. 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. Dnase113 deletion causes aberrations in length and end-motif frequencies in plasma DNA. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 641–649. [[CrossRef](#)] [[PubMed](#)]
35. 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](#)]
36. 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](#)] [[PubMed](#)]
37. Marotta, V.; Cennamo, M.; La Civita, E.; Vitale, M.; Terracciano, D. Cell-Free DNA Analysis within the Challenges of Thyroid Cancer Management. *Cancers* **2022**, *14*, 5370. [[CrossRef](#)]
38. Verma, S.; Moore, M.W.; Ringler, R.; Ghosal, A.; Horvath, K.; Naef, T.; Anvari, S.; Cotter, P.D.; Gunn, S. Analytical performance evaluation of a commercial next generation sequencing liquid biopsy platform using plasma ctDNA, reference standards, and synthetic serial dilution samples derived from normal plasma. *BMC Cancer* **2020**, *20*, 945. [[CrossRef](#)]

39. Stensgaard, S.; Thomsen, A.; Helstrup, S.; Meldgaard, P.; Sorensen, B.S. Blood tumor mutational burden and dynamic changes in circulating tumor DNA predict response to pembrolizumab treatment in advanced non-small cell lung cancer. *Transl. Lung Cancer Res.* **2023**, *12*, 971–984. [[CrossRef](#)]
40. Monaco, G.; Lee, B.; Xu, W.; Mustafah, S.; Hwang, Y.Y.; Carré, C.; Burdin, N.; Visan, L.; Ceccarelli, M.; Poidinger, M.; et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep.* **2019**, *26*, 1627–1640.e1627. [[CrossRef](#)]
41. Taudt, A.; Nguyen, M.A.; Heinig, M.; Johannes, F.; Colomé-Tatché, M. chromstaR: Tracking combinatorial chromatin state dynamics in space and time. *bioRxiv* **2016**, 038612. [[CrossRef](#)]
42. Lawrence, M.; Huber, W.; Pagès, H.; Aboyoun, P.; Carlson, M.; Gentleman, R.; Morgan, M.T.; Carey, V.J. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **2013**, *9*, e1003118. [[CrossRef](#)] [[PubMed](#)]
43. Pagès, H.; Aboyoun, P.; Gentleman, R.; DebRoy, S. *Biostrings: Efficient Manipulation of Biological Strings*, R package version 2.68.1; Bioconductor: Boston, MA, USA, 2023.
44. Morgan, M.; Pagès, H.; Obenchain, V.; Hayden, N. *Rsamtools: Binary Alignment (BAM), FASTA, Variant Call (BCF), and Tabix File Import*, R package version 2.16.0; Bioconductor: Boston, MA, USA, 2023.
45. Wickham, H.; François, R.; Henry, L.; Müller, K.; Vaughan, D. *dplyr: A Grammar of Data Manipulation*, R package version 1.1.2; CRAN: Vienna, Austria, 2023.
46. Kassambara, A. *ggpubr: 'ggplot2' Based Publication Ready Plots*, R package version 0.6.0; CRAN: Vienna, Austria, 2023.
47. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, R package version 3.4.2; CRAN: Vienna, Austria, 2023.
48. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.