cfDNA Fragmentation Analysis in ALS and Control Plasma Samples

ML Mansego

Contents

Introduction

Cell-free DNA (cfDNA) in plasma originates predominantly from apoptotic and necrotic cellular processes. The resulting fragmentation pattern reflects nucleosomal architecture and chromatin accessibility, providing valuable insights into tissue-specific cell death and disease-associated molecular changes. In the context of amyotrophic lateral sclerosis (ALS), cfDNA has emerged as a potential non-invasive biomarker under active investigation. Fragmentation characteristics—particularly insert size distributions—may capture signals related to disease-specific cell turnover, inflammatory processes, or epigenomic alterations.

The present analysis investigates differences in cfDNA fragmentation patterns between ALS patients and healthy controls, based on insert size distributions derived from EM-seq data. Fragment size profiles are quantified using insert size metrics obtained through Picard tools and compared across groups to identify disease-associated patterns.

Metadata Loading

The sample metadata, including subject ID, condition (ALS or control), age, gender, and ALS phenotype (bulbar or spinal, for patients only), is imported and used to stratify insert size data.

Insert Size Analysis

Extraction and Preprocessing

Insert size data were extracted from Picard output files. Only fragment lengths 600 bp were retained for biological relevance.

Nucleosome Peak Abundance Analysis

In this section, peak cfDNA fragment sizes associated with mono-, di-, and trinucleosomes were identified within predefined size ranges. For each nucleosome category, the fragment size showing the highest mean abundance was selected separately for ALS and control groups.

To assess potential differences in peak fragment abundance between conditions, the Wilcoxon rank-sum test was applied to each size range. A summary table was generated, presenting the mean and standard deviation of fragment counts per group, along with the corresponding p-values from the statistical tests.

This analysis was performed to explore whether specific nucleosome-associated cfDNA fragments differ in abundance between ALS and controls, which may reflect alterations in chromatin structure or cell death dynamics.

```
# This script identifies peak nucleosome fragment sizes in different ranges
# and compares their abundance between conditions (ALS vs Control)

## Define nucleosome size ranges
# We analyze three nucleosome configurations:
ranges <- list(
  "Mono" = c(120, 170),    # Mononucleosomes (core particle + short linker)
  "Di"   = c(280, 350),    # Dinucleosomes
  "Tri"  = c(430, 600)     # Trinucleosomes
)

## Find peak abundances for each condition and range

peaks_by_condition <- lapply(names(ranges), function(r) {
  fragment_summary %>%
    filter(insert_size >= ranges[[r]][1], insert_size <= ranges[[r]][2]) %>%
    group_by(Condition) %>%
    slice_max(mean_count, n = 1, with_ties = FALSE) %>%
    mutate(Nucleosome_Type = r) %>%
    select(Condition, Nucleosome_Type, insert_size, mean_count, sd_count)
}) %>% bind_rows()

## 3. Test (Wilcoxon)
stat_results <- lapply(names(ranges), function(nuc) {
  range <- ranges[[nuc]]
  test <- fragment_summary %>%
    filter(insert_size >= range[1], insert_size <= range[2]) %>%
    wilcox.test(mean_count ~ Condition, data = ., exact = FALSE)
  data.frame(Nucleosome_Type = nuc, p_value = test$p.value)
}) %>% bind_rows() %>%
  mutate(p_value = ifelse(p_value < 0.001, "< 0.001", round(p_value, 3)))

## 4. summary by nucleosome and condition
```

```r
table_data <- peaks_by_condition %>%
  mutate(Count_SD = paste0(round(mean_count, 2), " (", round(sd_count, 2), ")")) %>%
  select(Nucleosome_Type, Condition, Count_SD) %>%
  pivot_wider(names_from = Condition, values_from = Count_SD) %>%
  left_join(stat_results, by = "Nucleosome_Type") %>%
  select(Nucleosome_Type, ALS, Control, p_value)

## 5. table
table_data %>%
  kable("latex",
        booktabs = TRUE,
        col.names = c("Nucleosome Type", "Bulbar", "Spinal", "p-value"),
        caption = "Table 2. Summary of peak nucleosome-associated cfDNA fragment counts by condition.")
  kableExtra::kable_styling(latex_options = c("hold_position", "striped"))
```

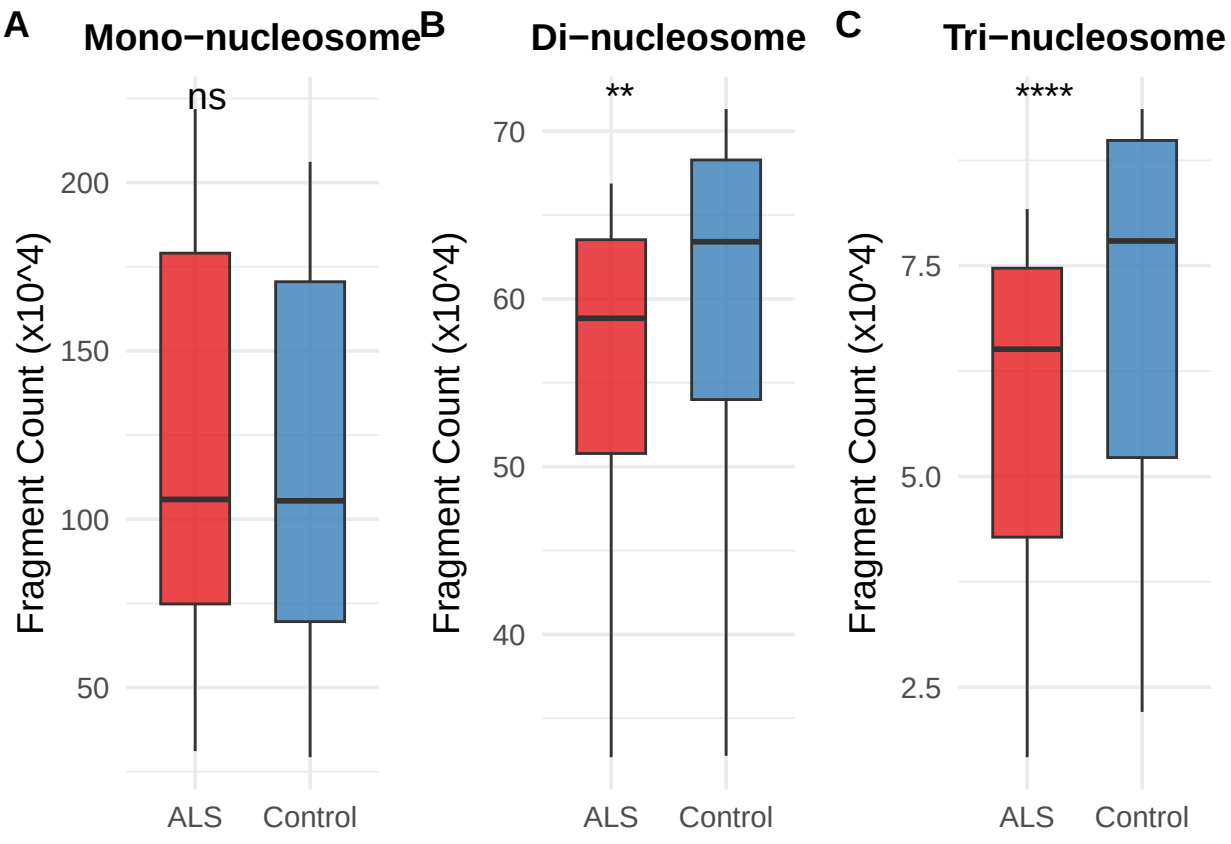
Table 1: Table 2. Summary of peak nucleosome-associated cfDNA fragment counts by condition.

Nucleosome Type	Bulbar	Spinal	p-value
Mono	2218747.25 (491914.27)	2061863.62 (426511.49)	0.574
Di	668719.12 (72993.38)	713160.62 (123962.95)	0.002
Tri	81709.33 (11782.98)	93580.25 (29880.54)	< 0.001

Mean and standard deviation of fragment counts are shown for the most abundant fragment size within each nucleosome range (Mono-, Di-, and Trinucleosomes) in ALS and control groups. P-values were obtained using the Wilcoxon rank-sum test.

```r
##  Boxplot visualization

fragment_plot_data <- fragment_summary %>%
  filter(insert_size >= 120 & insert_size <= 600) %>%
  mutate(Nucleosome_Type = case_when(
    insert_size >= 120 & insert_size <= 170 ~ "Mono",
    insert_size >= 280 & insert_size <= 350 ~ "Di",
    insert_size >= 430 & insert_size <= 600 ~ "Tri",
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(Nucleosome_Type))


fragment_plot_data_scaled <- fragment_plot_data %>%
  mutate(mean_count_scaled = mean_count / 10000)

plot_nuc_type_scaled <- function(nuc_type) {
  df <- fragment_plot_data_scaled %>% filter(Nucleosome_Type == nuc_type)

  ggplot(df, aes(x = Condition, y = mean_count_scaled, fill = Condition)) +
    geom_boxplot(width = 0.6, alpha = 0.8) +
    stat_compare_means(method = "wilcox.test", label = "p.signif", size = 5) +
    labs(
      title = paste0(nuc_type, "-nucleosome"),
      x = NULL,
```

```
      y = "Fragment Count (x10^4)"
    ) +
    scale_fill_brewer(palette = "Set1") +
    theme_minimal(base_size = 14) +
    theme(
      plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
      legend.position = "none",
      plot.margin = margin(10, 10, 10, 10)
    )
}


p_mono <- plot_nuc_type_scaled("Mono")
p_di   <- plot_nuc_type_scaled("Di")
p_tri  <- plot_nuc_type_scaled("Tri")


final_plot <- ggarrange(p_mono, p_di, p_tri,
                        ncol = 3,
                        labels = c("A", "B", "C"),
                        align = "v")

print(final_plot)
```




```
ggsave("figures/Nucleosome_Peak_Abundance_ALSvsCT.png", final_plot,
       width = 8, height = 5, dpi = 300, bg = "white")
```

Figure 1. Nucleosome-associated cfDNA fragment abundance across conditions. Boxplots show the distribution of cfDNA fragment counts within mono- (A), di- (B), and tri-nucleosome (C) size ranges. Comparisons are shown between ALS and control groups. Fragment counts are scaled by 10 for visualization. Statistical comparisons were performed using the Wilcoxon rank-sum test; significance levels are indicated as follows: $p < 0.05$ (), $p < 0.01$ (), $p < 0.001$ (), and $p < 0.0001$ (****).

Visualization of Fragment Size Distribution

The mean insert size distribution per group is shown below. Shaded ribbons represent ±1 standard error. Vertical dashed lines indicate mono- and di-nucleosome boundaries. Additionally, yellow and blue shaded areas highlight the short (110–150 bp) and long (=>180 bp) fragment ranges, respectively.

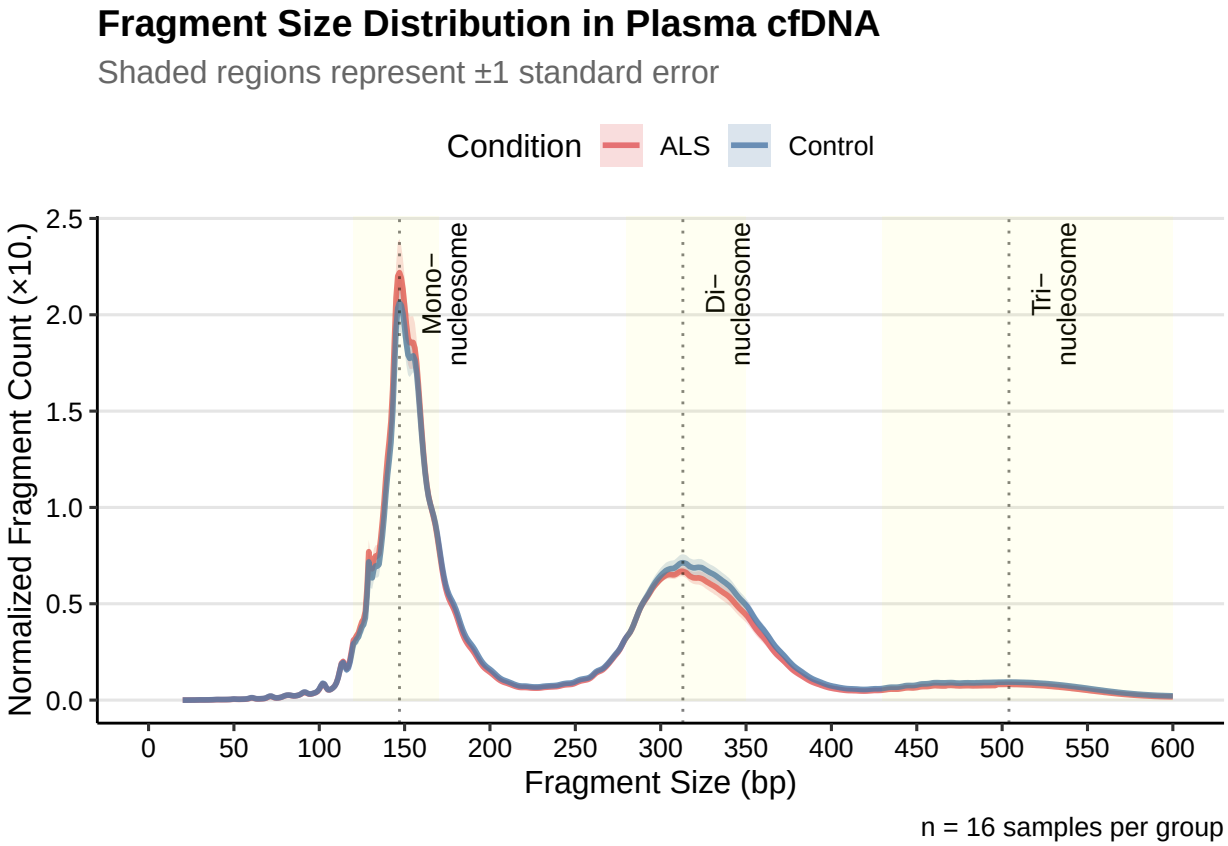


Figure 1: Mean fragment size distribution by group

Short-to-Long Fragment Ratio Analysis

Fragment size distributions in cfDNA are shaped by nucleosomal protection and degradation dynamics. Short fragments (~110–150 bp) typically originate from apoptotic cleavage, while longer fragments (~200–600 bp) may reflect alternative modes of cell death or reduced chromatin degradation. The relative abundance of these fragment populations can provide insight into underlying biological processes.

Previous studies have proposed the short-to-long fragment ratio as a potential biomarker in cfDNA-based analyses. Therefore, in this section, the abundance of short and long fragments was quantified for each sample, and the corresponding short-to-long fragment ratio was calculated. All values were normalized per million fragments and compared between ALS and control groups.

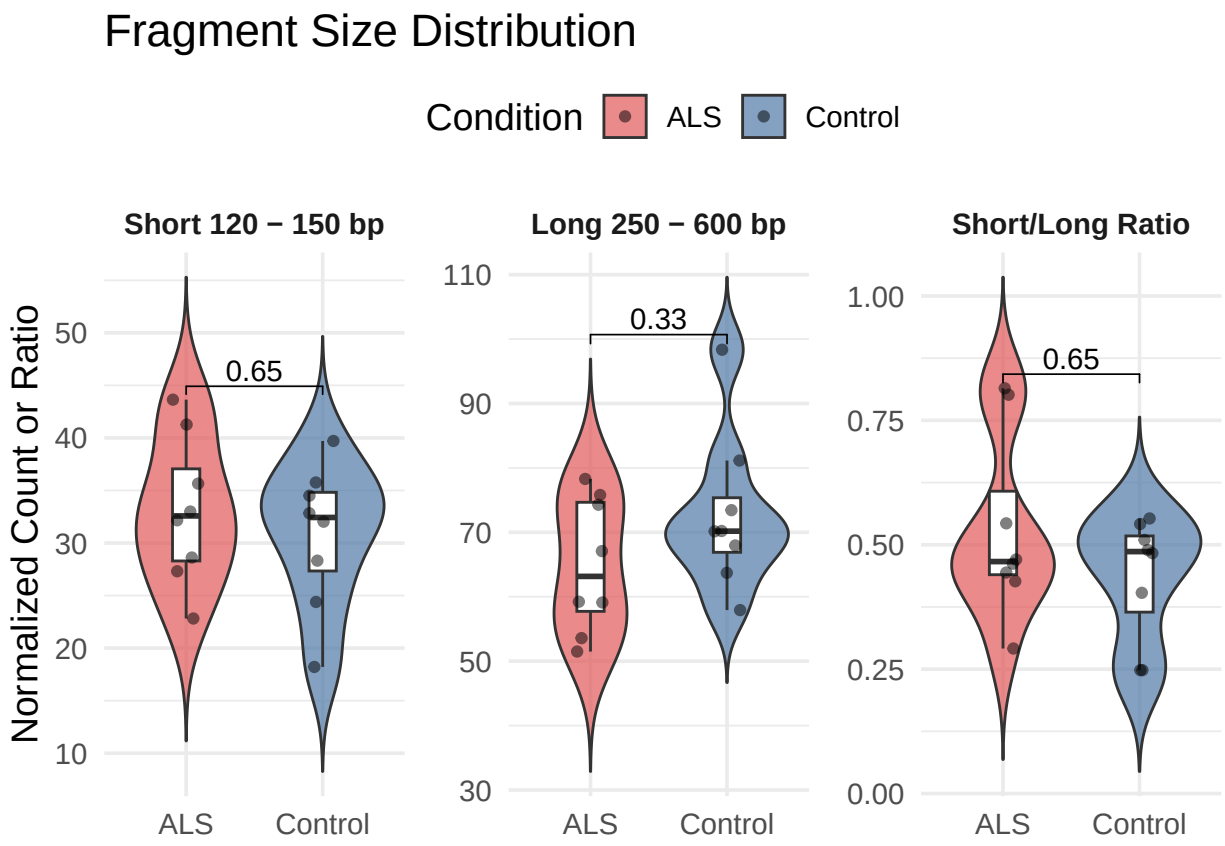


Figure 2: Short-to-long fragment ratio by condition

Distribution of Short-to-Long Ratios ALS samples showed a non-significant trend toward higher abundance of short fragments and reduced levels of long fragments compared to controls. Consequently, the short-to-long ratio appeared slightly elevated in the ALS group. While these tendencies are biologically plausible—potentially reflecting increased apoptotic cfDNA release—they did not reach statistical significance and should be interpreted with caution.

Fragmentation Differences Between Bulbar- and Spinal-Onset ALS

Within the ALS group, patients can present distinct onset phenotypes: bulbar or spinal. These phenotypes may reflect different neurodegenerative mechanisms and progression patterns. Here, we examine whether cfDNA fragmentation differs between bulbar- and spinal-onset ALS patients.

Nucleosome Peak Abundance Analysis (Bulbar vs Spinal onset ALS)

```
# Nucleosome Peak Abundance Analysis
# This script identifies peak nucleosome fragment sizes in different ranges
# and compares their abundance between conditions (ALS vs Control)

## Define nucleosome size ranges
# We analyze three nucleosome configurations:
ranges <- list(
  "Mono" = c(120, 170),    # Mononucleosomes (core particle + short linker)
  "Di"   = c(280, 350),    # Dinucleosomes
  "Tri"  = c(430, 600)     # Trinucleosomes
)

## Find peak abundances for each condition and range

peaks_by_pheno <- lapply(names(ranges), function(r) {
  fragment_pheno_summary %>%
    filter(insert_size >= ranges[[r]][1], insert_size <= ranges[[r]][2]) %>%
    group_by(ALS_Phenotype) %>%
    slice_max(mean_count, n = 1, with_ties = FALSE) %>%
    mutate(Nucleosome_Type = r) %>%
    select(ALS_Phenotype, Nucleosome_Type, insert_size, mean_count, sd_count)
}) %>% bind_rows()

## 3. Test (Wilcoxon)
stat_results_ALS <- lapply(names(ranges), function(nuc) {
  range <- ranges[[nuc]]
  test <- fragment_pheno_summary %>%
    filter(insert_size >= range[1], insert_size <= range[2]) %>%
    wilcox.test(mean_count ~ ALS_Phenotype, data = ., exact = FALSE)
  data.frame(Nucleosome_Type = nuc, p_value = test$p.value)
}) %>% bind_rows() %>%
  mutate(p_value = ifelse(p_value < 0.001, "< 0.001", round(p_value, 3)))

## 4. summary by nucleosome and condition
table_data_ALS <- peaks_by_pheno %>%
  mutate(Count_SD = paste0(round(mean_count, 2), " (", round(sd_count, 2), ")")) %>%
  select(Nucleosome_Type, ALS_Phenotype, Count_SD) %>%
```

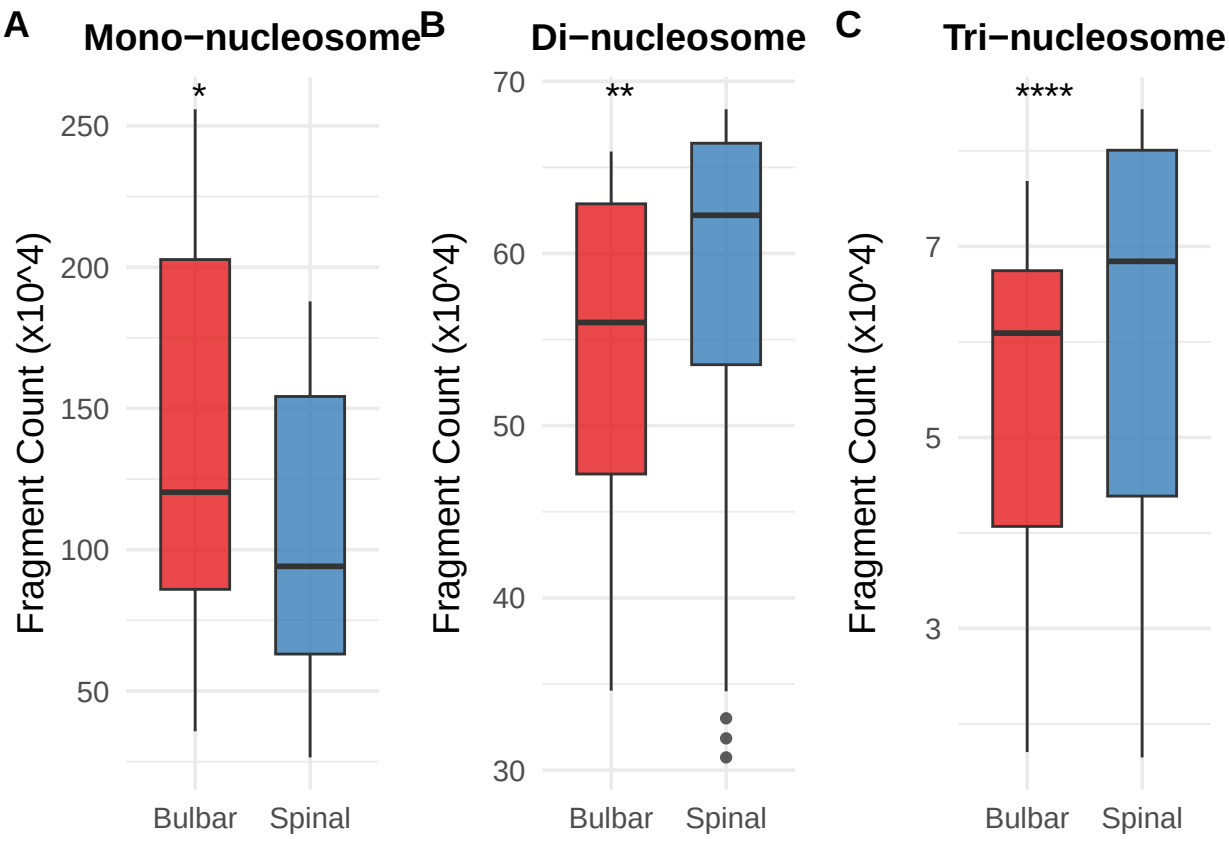
```r
  pivot_wider(names_from = ALS_Phenotype, values_from = Count_SD) %>%
  left_join(stat_results_ALS, by = "Nucleosome_Type") %>%
  select(Nucleosome_Type, Bulbar, Spinal, p_value)

## 5. table
table_data_ALS %>%
  kable("latex",
        booktabs = TRUE,
        col.names = c("Nucleosome Type", "Bulbar", "Spinal", "p-value"),
        caption = "Table 1. Summary of peak nucleosome-associated cfDNA fragment counts by ALS phenotype
  kableExtra::kable_styling(latex_options = c("hold_position", "striped"))
```

Table 2: Table 1. Summary of peak nucleosome-associated cfDNA fragment counts by ALS phenotype.

Nucleosome Type	Bulbar	Spinal	p-value
Mono	2558797 (446407.41)	1878697.5 (238710.56)	0.014
Di	659191.75 (58622.55)	683759.75 (82170.24)	0.001
Tri	76839 (13295.02)	84348.75 (11858.59)	< 0.001

```r
##  Boxplot visualization

fragment_plot_data_ALS <- fragment_pheno_summary %>%
  filter(insert_size >= 120 & insert_size <= 600) %>%
  mutate(Nucleosome_Type = case_when(
    insert_size >= 120 & insert_size <= 170 ~ "Mono",
    insert_size >= 280 & insert_size <= 350 ~ "Di",
    insert_size >= 430 & insert_size <= 600 ~ "Tri",
    TRUE ~ NA_character_
  )) %>%
  filter(!is.na(Nucleosome_Type))

fragment_plot_data_scaled_ALS <- fragment_plot_data_ALS %>%
  mutate(mean_count_scaled = mean_count / 10000)

plot_nuc_type_scaled <- function(nuc_type) {
  df <- fragment_plot_data_scaled_ALS %>% filter(Nucleosome_Type == nuc_type)

  ggplot(df, aes(x = ALS_Phenotype, y = mean_count_scaled, fill = ALS_Phenotype)) +
    geom_boxplot(width = 0.6, alpha = 0.8) +
    stat_compare_means(method = "wilcox.test", label = "p.signif", size = 5) +
    labs(
      title = paste0(nuc_type, "-nucleosome"),
      x = NULL,
      y = "Fragment Count (x10^4)"
    ) +
    scale_fill_brewer(palette = "Set1") +
    theme_minimal(base_size = 14) +
    theme(
      plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
      legend.position = "none",
      plot.margin = margin(10, 10, 10, 10)
```

```
    )
}

p_mono <- plot_nuc_type_scaled("Mono")
p_di   <- plot_nuc_type_scaled("Di")
p_tri  <- plot_nuc_type_scaled("Tri")

final_plot_ALS <- ggarrange(p_mono, p_di, p_tri,
                            ncol = 3,
                            labels = c("A", "B", "C"),
                            align = "v")

print(final_plot_ALS)
```




```
ggsave("figures/Nucleosome_Peak_Abundance_BvsS.png", final_plot_ALS,
       width = 8, height = 5, dpi = 300, bg = "white")
```

Figure 2. Nucleosome-associated cfDNA fragment abundance across conditions. Boxplots show the distribution of cfDNA fragment counts within mono- (A), di- (B), and tri-nucleosome (C) size ranges. Comparisons are shown between Bulbar and Spinal.Fragment counts are scaled by 10 for visualization. Statistical comparisons were performed using the Wilcoxon rank-sum test; significance levels are indicated as follows: $p <$ 0.05 (), $p < 0.01$ (), $p < 0.001$ (), and $p < 0.0001$ (****).

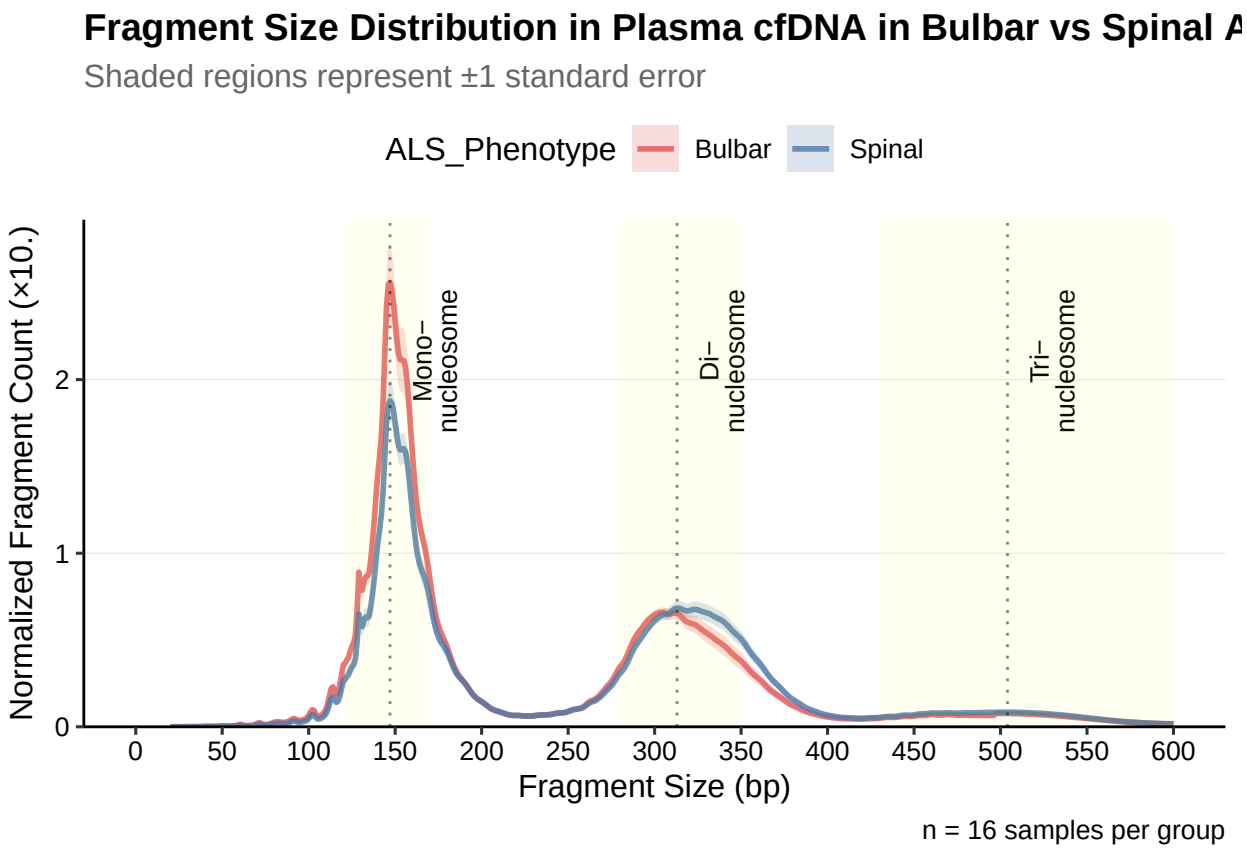


Figure 3: Fragment size distribution in bulbar- vs spinal-onset ALS

Fragment Size Distribution by ALS Phenotype

Short-to-Long Fragment Ratio Analysis

As described previous sections, the short-to-long fragment ratio was also evaluated within the ALS group to compare ALS phenotypes, aiming to explore potential differences in fragmentation patterns associated with clinical presentation.

In the comparison between ALS phenotypes, bulbar-onset samples showed a significantly higher short-to-long fragment ratio compared to spinal-onset cases ($p = 0.029$). This difference was driven by both an increased abundance of short fragments ($p = 0.057$) and a decreased abundance of long fragments ($p = 0.029$) in the bulbar group.

These results suggest a potential shift in cfDNA fragmentation profiles associated with ALS clinical subtype. However, given the limited sample size ($n = 4$ per group), findings should be interpreted with caution and require validation in larger cohorts.

Summary and Interpretation

Nucleosomal cfDNA fragmentation patterns were assessed in ALS patients and healthy controls. No significant differences were observed in mononucleosomal fragment abundance between groups; however, ALS samples showed significantly lower levels of di- and tri-nucleosomal fragments, suggesting enhanced cfDNA fragmentation or altered chromatin processing in ALS.

In line with these findings, ALS samples also exhibited a non-significant trend toward a higher abundance of short fragments and a reduced abundance of long fragments, resulting in an increased short-to-long fragment ratio. Although this difference was not statistically significant, it is consistent with a shift toward apoptotic cfDNA release in ALS.

Within the ALS cohort, the comparison between bulbar- and spinal-onset phenotypes revealed more pronounced differences in cfDNA fragmentation. Bulbar-onset patients exhibited significantly lower levels of mono-, di-, and tri-nucleosomal fragments compared to spinal-onset cases (e.g., mononucleosomes: $p = 0.014$), along with higher short-to-long fragment ratios. These findings suggest potential phenotype-specific differences in cfDNA release or degradation dynamics that may reflect distinct pathophysiological mechanisms. However, due to the limited sample size ($n = 4$ per group), these results should be interpreted with caution and confirmed in larger cohorts.

Conclusion

This study provides a preliminary characterization of cfDNA fragmentation patterns in ALS plasma samples using EM-seq data. Although no significant differences were detected between ALS and control groups in overall fragmentation metrics, a selective reduction in di- and tri-nucleosomal fragments was observed in ALS samples, suggesting increased cfDNA degradation or altered chromatin processing.

Within the ALS cohort, phenotype-specific differences were more evident: bulbar-onset patients showed significantly lower nucleosome-associated fragment abundance and higher short-to-long ratios compared to spinal-onset cases. These findings point to potential differences in cfDNA release dynamics linked to disease phenotype. Overall, the results support further investigation in larger, independent cohorts to evaluate the robustness and biomarker potential of cfDNA fragmentation features in ALS.