UCLA Computational Medicine

Cell-free DNA fragmentation signatures for disease state prediction in amyotrophic lateral sclerosis

GALEN HEUER, Fei-man Hsu, Christa Caggiano, Noah Zaitlen, Matteo Pellegrini

Abstract

Cell-free DNA (cfDNA) is released into bodily fluids after cell death. During this process, DNA is preferentially cleaved at linker regions between nucleosome cores, leading to distinct fragmentation patterns which reflect a cell's transcriptional accessibility. This has prompted investigation into cfDNA fragmentation as a disease biomarker, notably in cancer. However, this method's application in neurodegenerative conditions such as amyotrophic lateral sclerosis (ALS) remains unexplored. Here, we analyze whole-genome bisulfite sequenced cfDNA extracted from blood plasma of two independent cohorts of ALS patients and healthy controls. We observe a distinct difference in cfDNA fragment length between patients and controls in both cohorts, with ALS patients exhibiting shorter fragments than their healthy counterparts on the average (p=0.0001). Furthermore, we quantify decreased periodicity of fragment length probability density in ALS patients. Our work lends insight into cell death biology and cfDNA degradation in neurodegenerative processes, with the potential to inform diagnosis and patient outcome.

Background

ALS is a rare neurodegenerative condition characterized by death of motor neurons. Currently, there is no single test or procedure used by physicians to confirm this disease. As a result, it is possible that patients will progress beyond the point of lifesaving treatment while awaiting an established diagnosis. One avenue being explored as a potential biomarker is cell-free DNA, which consists of short fragments of DNA released into bodily fluids after apoptosis and necrosis. Previous work has demonstrated differential cfDNA fragmentation patterns in circulating tumor DNA when compared to healthy DNA, but this method has not been explored for neurodegenerative disease.

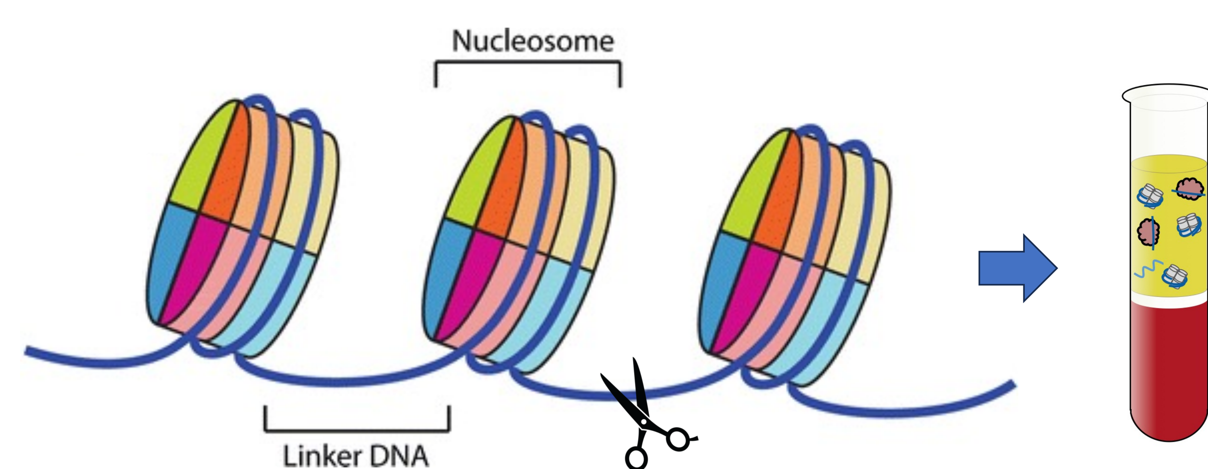


Figure 1a: CfDNA fragmentation. *Once outside the cell membrane, DNA is preferentially cleaved at linker regions between nucleosomes. Fragment length therefore reflects chromosome organization and nucleosome spacing, which we infer by cfDNA sampling and sequencing.*

Because cfDNA fragmentation patterns allow inference into nucleosome spacing and therefore gene expression (**Fig. 1a**), the process by which these short fragments are degraded has the potential to elucidate underlying cell death biology in neurodegeneration. Our aim is to be able to use this fragmentation profile as a biomarker to differentiate between ALS patients and healthy controls, and to find some relationship between fragmentation and ALS progression that could help explain heterogeneity in patient outcomes.

Methods

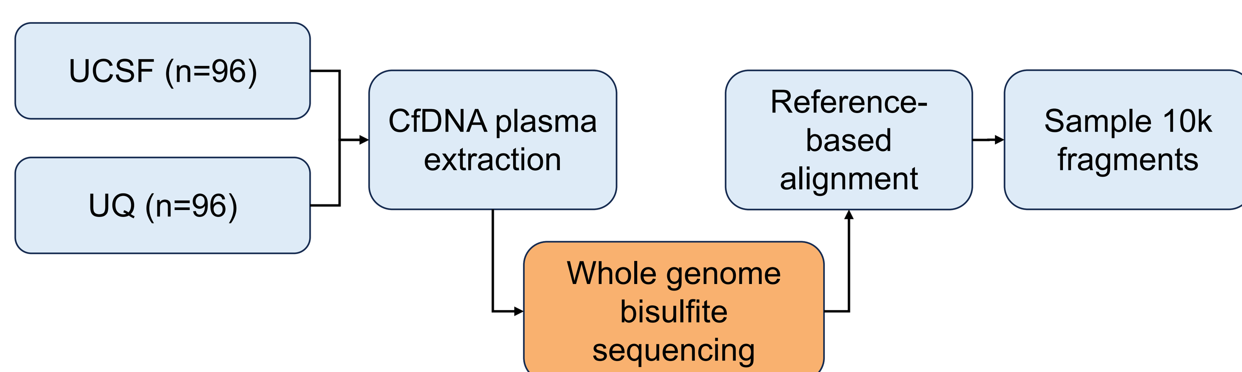
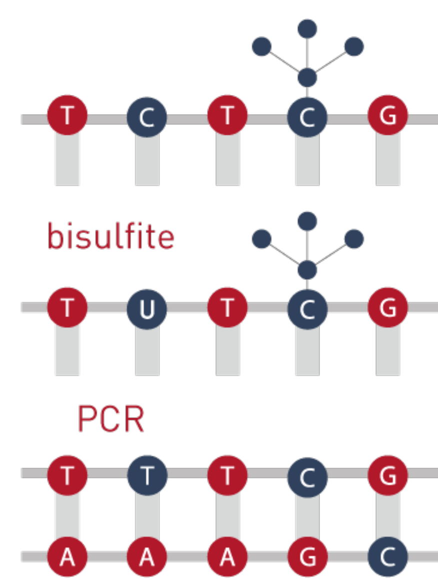


Figure 1b: Study workflow. *Each cohort contained consisted of roughly equal numbers of case and control samples.*

Methods (cont.)



Here, 192 cfDNA samples are sequenced using whole genome bisulfite sequencing (**Fig. 1b**) to allow for differential methylation analysis in development of a multimodal biomarker. To our knowledge, fragmentation analysis has not been attempted on bisulfite-converted (**Fig. 1c**) cfDNA. It is unknown how this processing affects fragment degradation.

Figure 1c (left): Bisulfite conversion. *DNA molecules are treated with sodium bisulfite, converting unmethylated cytosines to uracils.*

Results

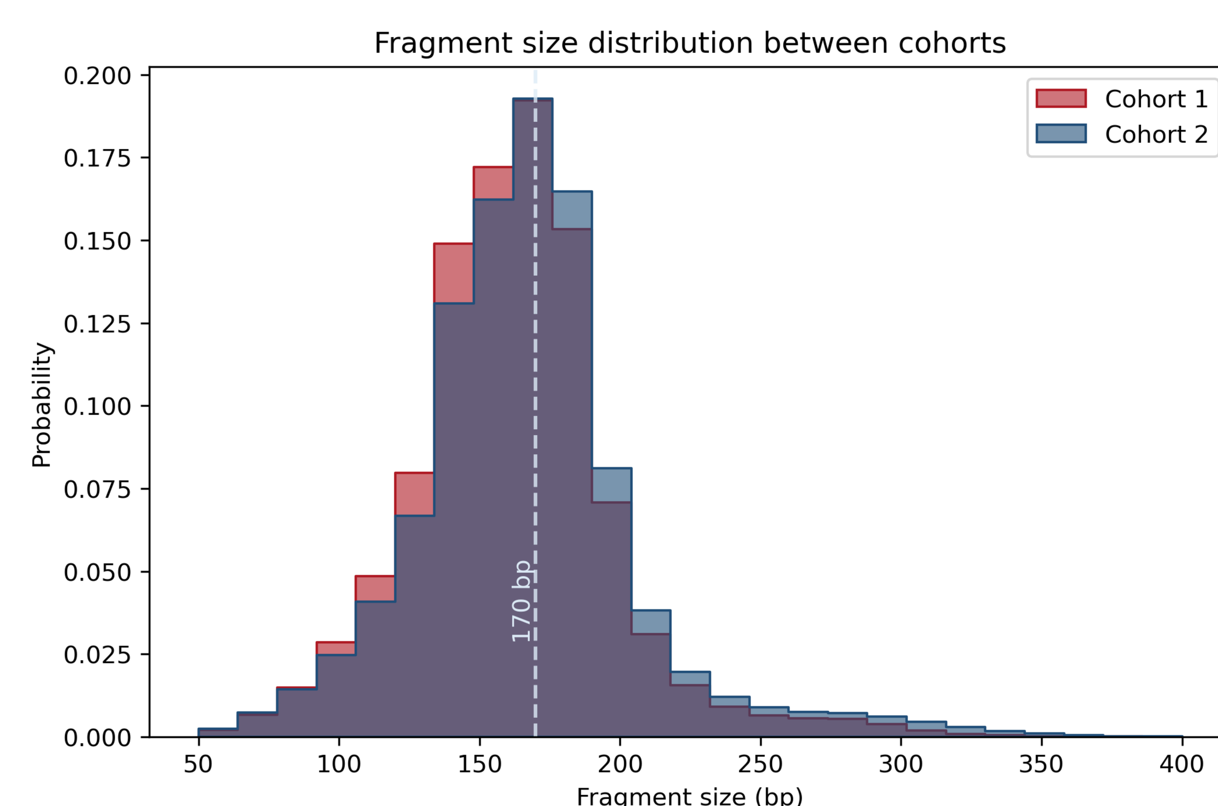


Figure 2: Fragment size distribution between cohorts. *Despite bisulfite enzymatic processing, cfDNA fragment length still peaks at the expected length of ~170 bp (approximate size of one nucleosome). This distribution is approximately the same in both cohorts.*

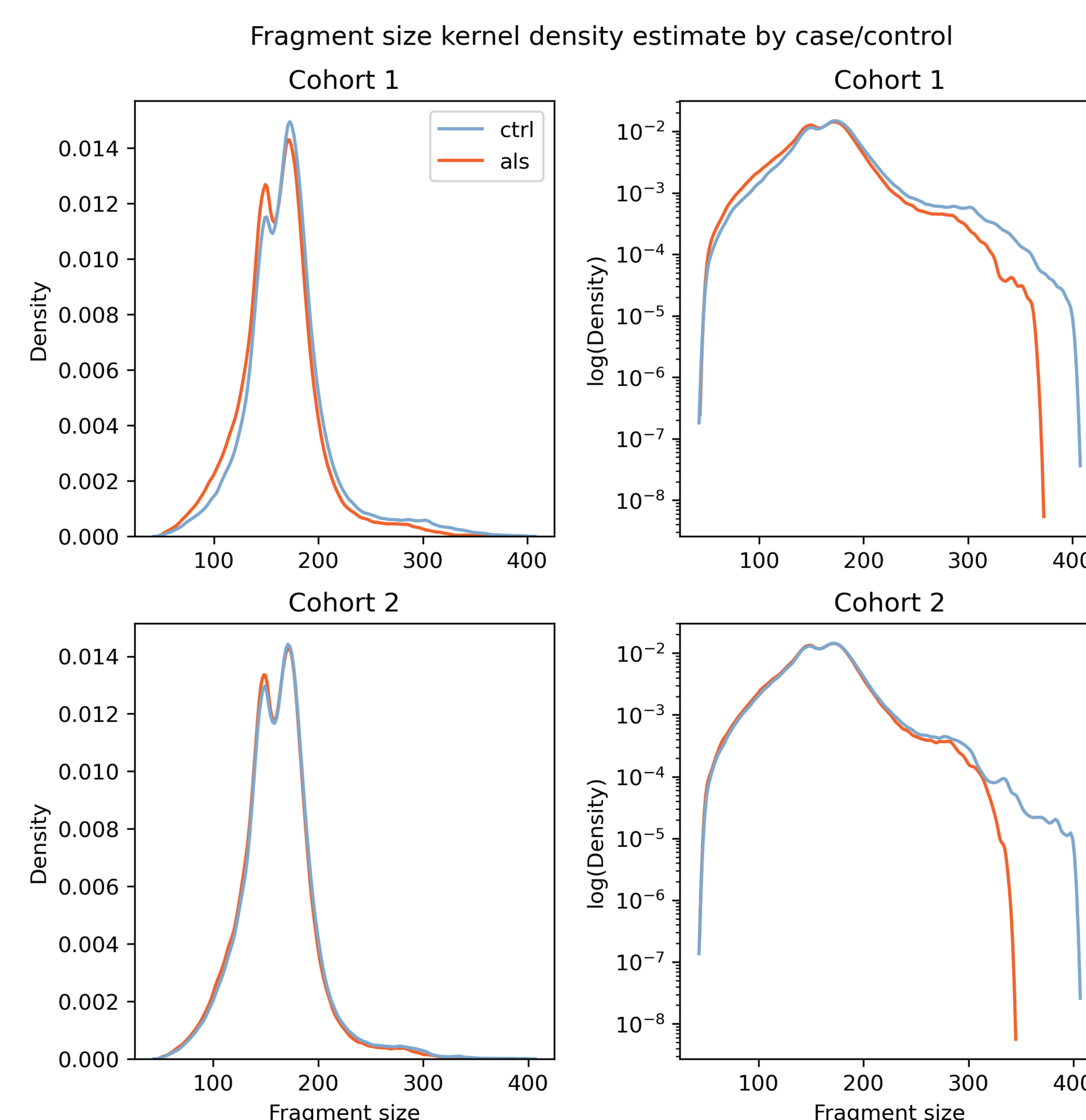


Figure 3: Fragment length kernel density estimate (KDE). *KDE is a smoothed visualization of probability density along a continuous spectrum, in this case the likelihood of a fragment being anywhere from 50-400bp. Visually, we see a higher peak around 150bp in ALS cases than controls (left), as well as a shorter right tail representing fewer large outliers (right).*

Results (cont.)

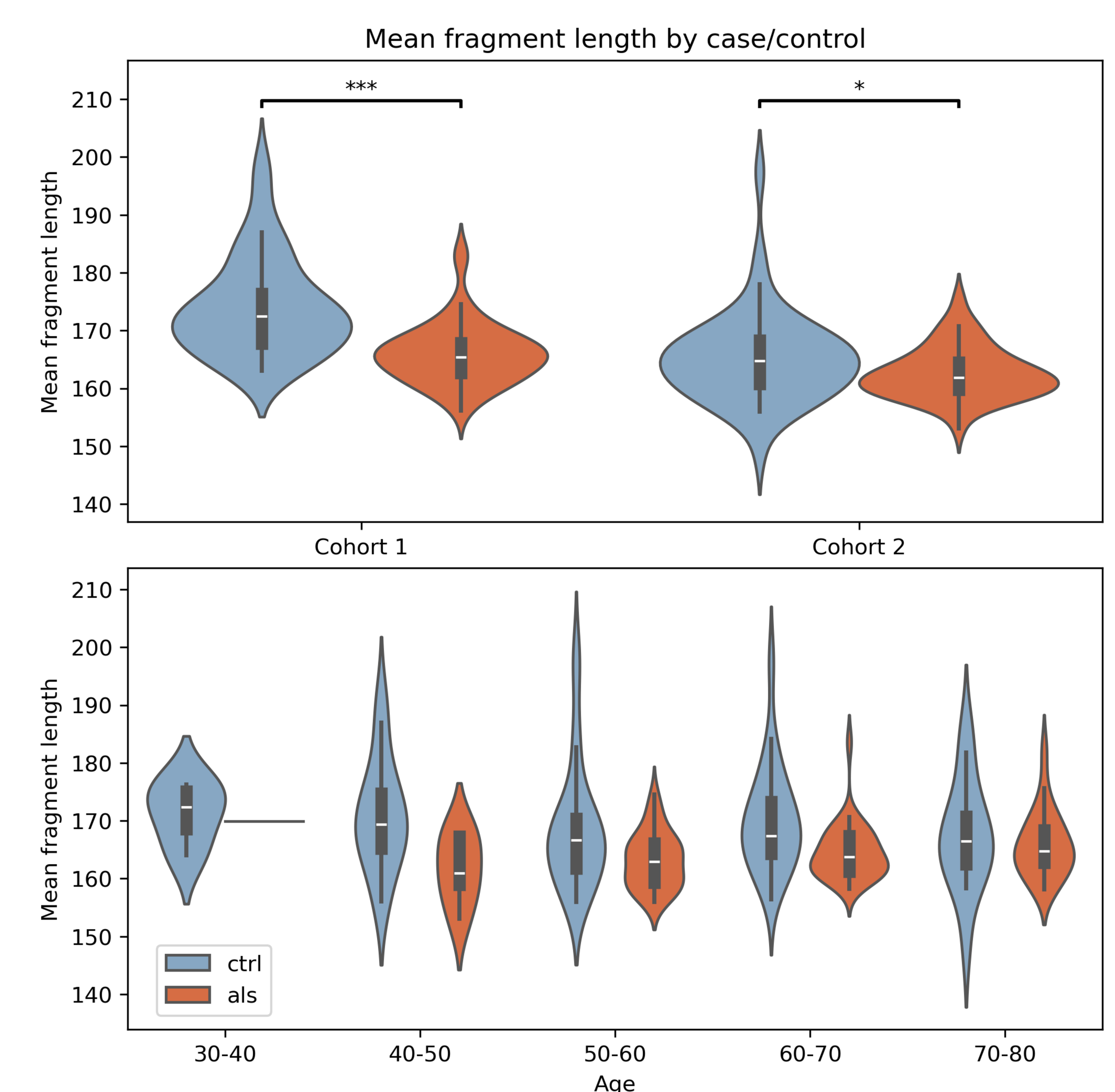


Figure 4: Mean fragment length for ALS patients vs. controls. *In both cohorts (top), ALS patients had significantly shorter mean fragment length. when adding age, sex, and ethnicity as covariates (p<0.001 and p=0.047 for cohorts 1 & 2 respectively). This relationship also holds across age groups (bottom).*

Discussion

1. Fragment size distribution exhibits a mononucleosomal peak.

Because cfDNA is preferentially cleaved between nucleosomes after release, we expect a fragment size peak at approximately 170 base pairs, or the amount of DNA needed to wrap around one nucleosome. This aligns with the fragment size distribution of samples from both independent cohorts (**Fig. 2**) despite sodium bisulfite processing. However, we do not observe a secondary dinucleosome peak around 340 bp, nor any samples from either cohort that contain fragments with lengths greater than 500 bp. This could be due to the enzymatic processing and molecule degradation involved in bisulfite sequencing.

2. Fragment length kernel density distribution differs between cases and controls.

In addition to the mononucleosome peak at ~170 bp, we observe a secondary peak at ~150 bp, which appears to be more pronounced in cases than controls (**Fig. 3**). We also observe a longer tail of larger fragment sizes in healthy samples than controls. These features will be used to predict patient status by fitting a Gaussian mixture model to our data.

3. ALS patients exhibit shorter fragments than their healthy counterparts, across cohorts and age groups.

To account for heterogeneity between sample cohorts, we added age, sex, ethnicity, and geography into our regression for hypothesis testing. We found that when combining cohorts, ALS patients had significantly shorter average fragment length than healthy controls (p<0.001). When examining each cohort separately (**Fig. 4**), the difference between cases and controls in the second cohort was less pronounced (p=0.047) than that of the first cohort (p<0.001), but still significant. This could be due to the underlying processes involved in disease biology, corresponding with how circulating tumor DNA in cancer is shorter than healthy DNA.