

Using the 6-base genome for full multiomic analysis of cfDNA through combined methylation and fragmentomic analysis to enhance classification of clinical cancer cfDNA samples

Mark Consugar¹, Fabio Puddu, Thao Huynh¹, Fabio Puddu¹, Annelie Johansson¹, Ermira Lleshi¹, Robert Crawford¹, Tom Charlesworth¹, Steven Ciaramaglia¹, Robert J Osborne¹

¹ biomodal Ltd, The Trinity Building, Chesterford Research Park, Cambridge, UK.

1. Introduction

Sensitive detection of circulating tumour DNA (ctDNA) within cell-free DNA (cfDNA) is essential for early cancer detection, disease monitoring, and assessment of treatment response. However, ctDNA typically represents only a small fraction of total cfDNA, limiting the performance of conventional single biomarker-based approaches. Fragmentomics has emerged as a powerful complementary strategy, with fragment size, end-motif frequencies, and nucleosome positioning patterns shown to improve ctDNA detection sensitivity and specificity. Recently, fragmentomics-based approaches have shown promising results for ctDNA detection in liquid biopsy samples, with studies demonstrating that analysis of fragment size distributions across genomic regions, the frequency of 5' end motifs, and nucleosome positioning patterns near functional genomic sites can substantially enhance sensitivity and specificity.

The advent of 6-base sequencing with duet evoC, which distinguishes 5-methylcytosine (5mC) from 5-hydroxymethylcytosine (5hmC), has further expanded the analytical landscape.

- The duet evoC solution provides complete genetic, epigenetic and fragmentomic information in a single workflow from a single sample, enabling integrated multiomic biomarker extraction and classification.
- 5mC and 5hmC extends the spectrum of fragment end motifs and provides orthogonal data layers to fragment size and nucleosome positioning analysis.
- These multidimensional epigenetic signatures hold the potential to markedly improve the resolution and accuracy of ctDNA detection.

Here, we assess the contribution of combined fragmentomic and epigenetic features to ctDNA detection in cfDNA from healthy individuals and colorectal cancer (CRC) patients.

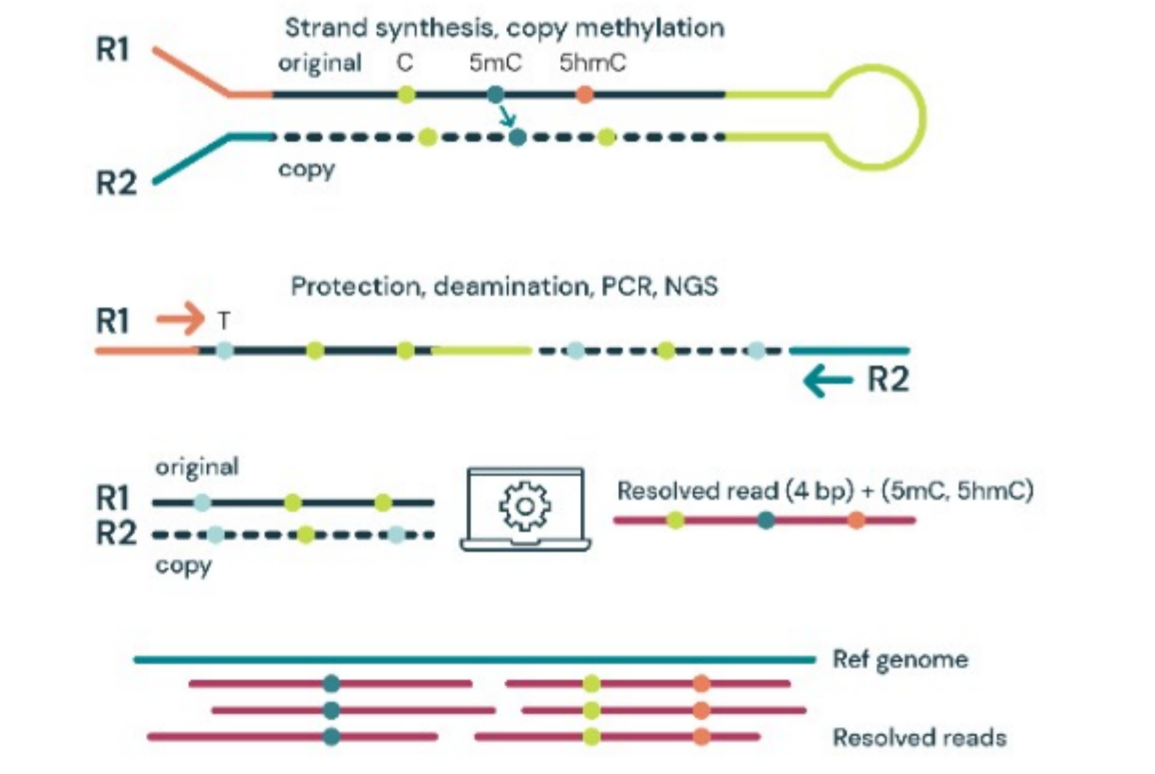
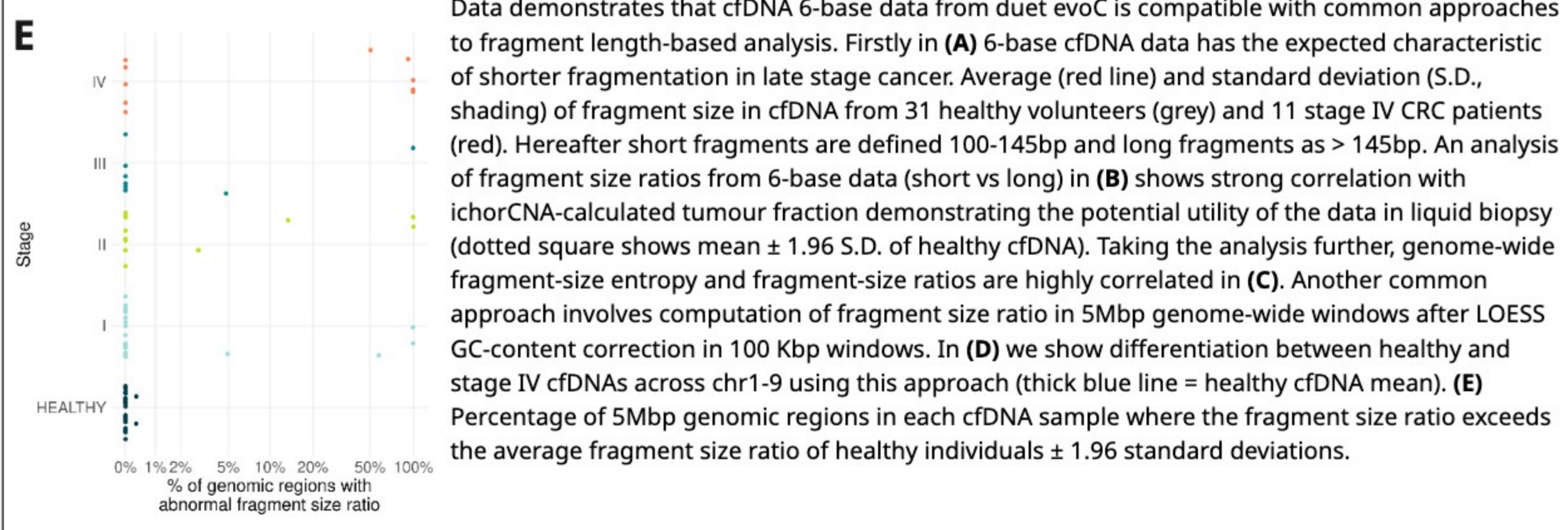
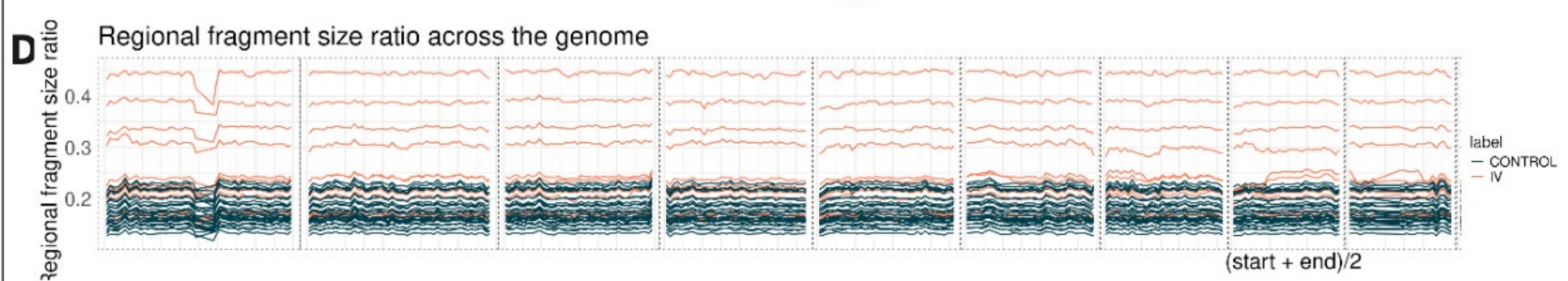
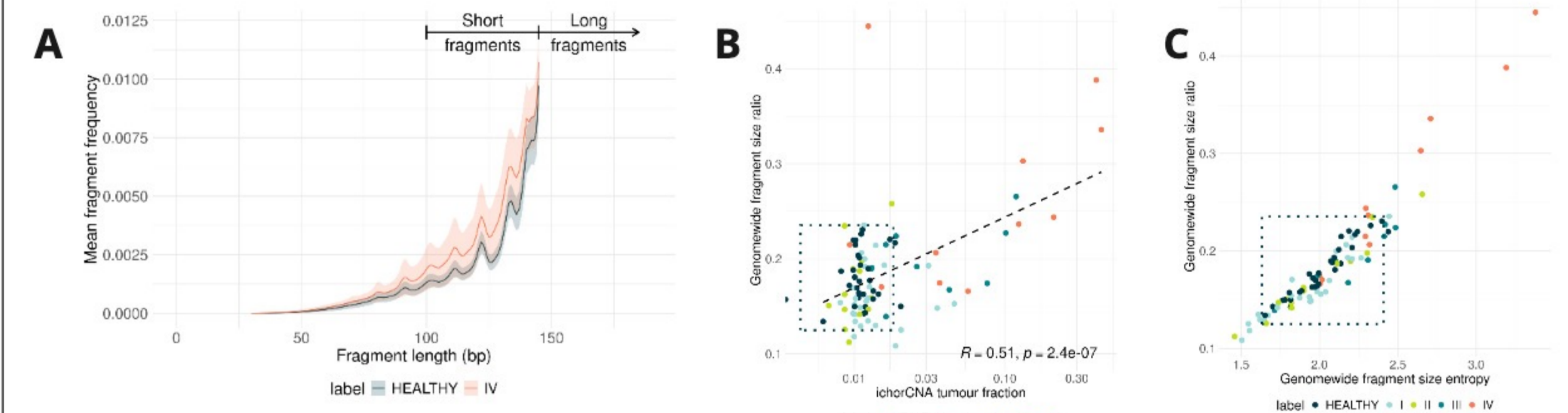


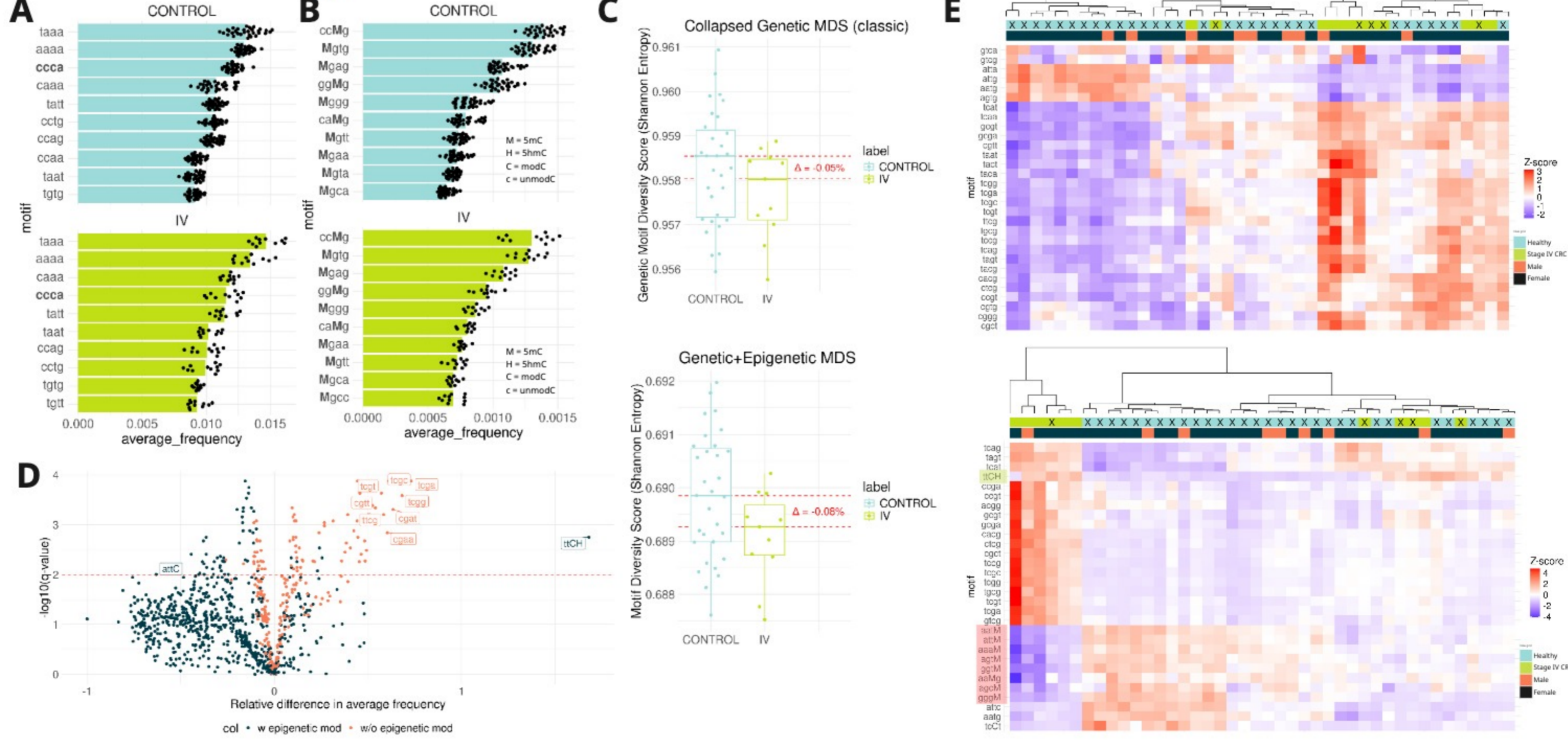
Figure 1. duet evoC is a 6-base sequencing technology that reads all four canonical bases plus 5mC and 5hmC¹.

2. 6-base data enables fragment length analysis



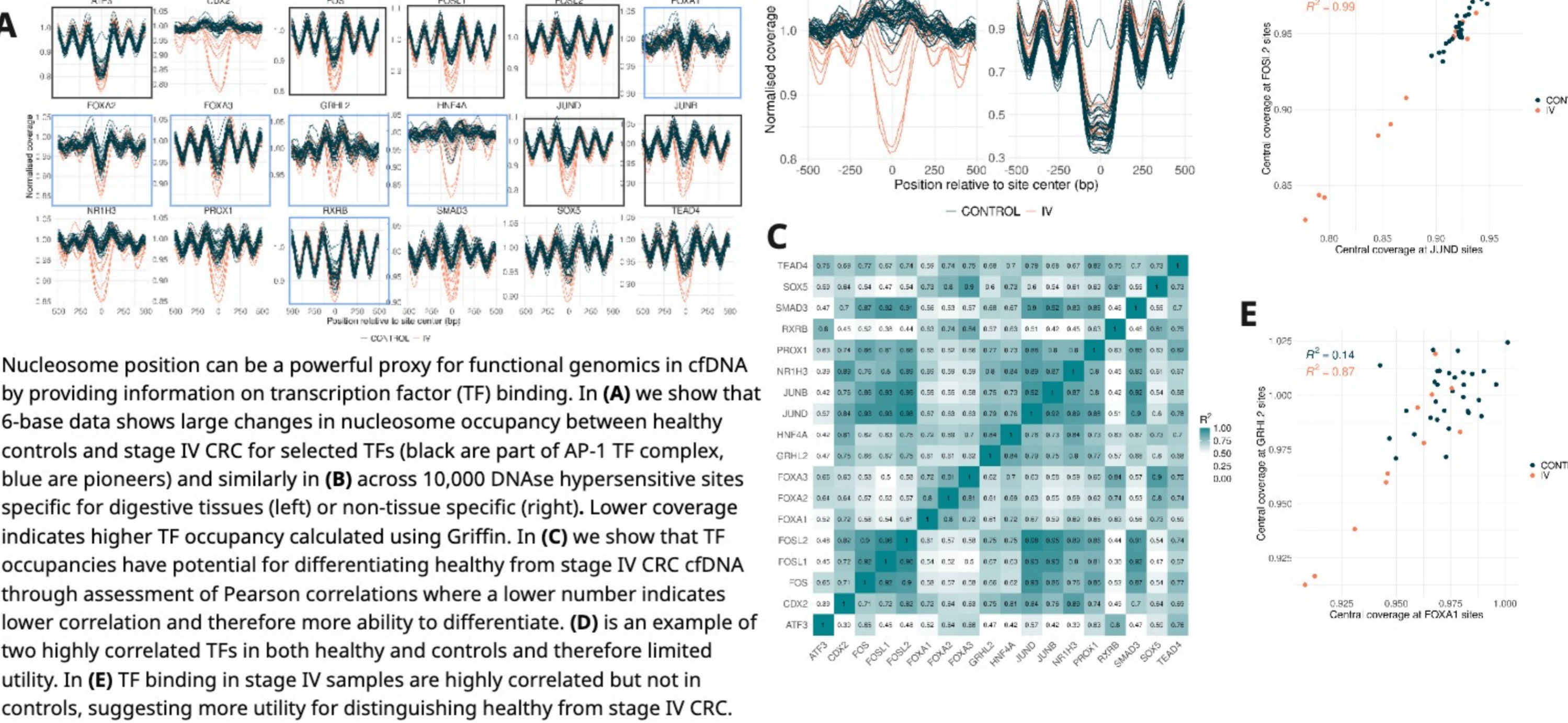
Data demonstrates that cfDNA 6-base data from duet evoC is compatible with common approaches to fragment length-based analysis. Firstly in (A) 6-base cfDNA data has the expected characteristic of shorter fragmentation in late stage cancer. Average (red line) and standard deviation (S.D., shading) of fragment size in cfDNA from 31 healthy volunteers (grey) and 11 stage IV CRC patients (red). Hereafter short fragments are defined 100-145bp and long fragments as > 145bp. An analysis of fragment size ratios from 6-base data (short vs long) in (B) shows strong correlation with IchorCNA-calculated tumour fraction demonstrating the potential utility of the data in liquid biopsy (dotted square shows mean \pm 1.96 S.D. of healthy cfDNA). Taking the analysis further, genome-wide fragment-size entropy and fragment-size ratios are highly correlated in (C). Another common approach involves computation of fragment size ratio in 5Mbp genome-wide windows after LOESS GC-content correction in 100 Kbp windows. In (D) we show differentiation between healthy and stage IV cfDNAs across chr1-9 using this approach (thick blue line = healthy cfDNA mean). (E) Percentage of 5Mbp genomic regions in each cfDNA sample where the fragment size ratio exceeds the average fragment size ratio of healthy individuals \pm 1.96 standard deviations.

3. 6-base end motifs are more powerful biomarkers than conventional genetic end motifs



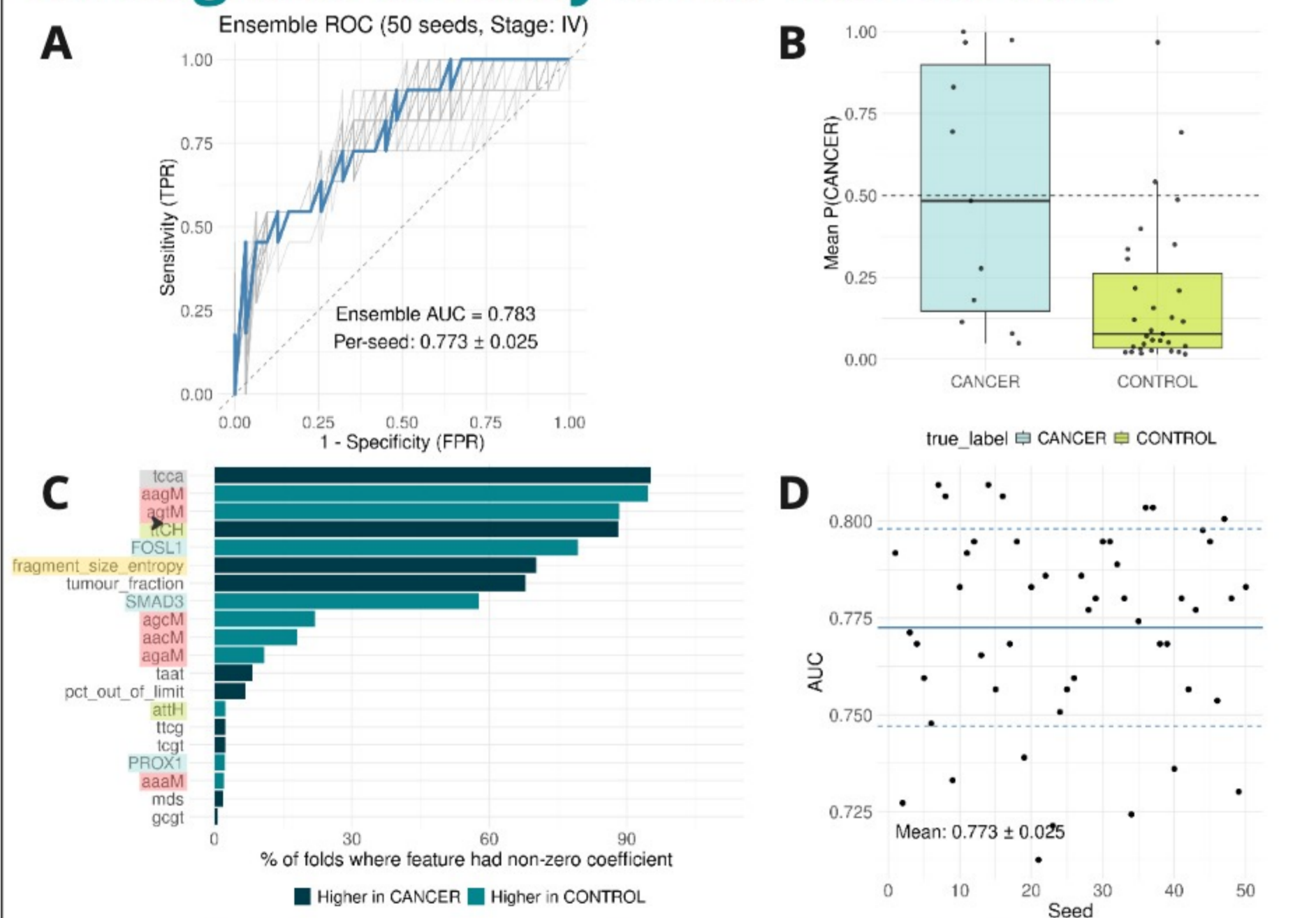
6-base supports analysis of the most common genetic end motifs in healthy and stage IV cfDNA (A) and through its embedded base-resolution 5mC and 5hmC information enables extension to epigenetic end motifs (B) with the most notable changes in ccMg and ggMg epigenetic motifs. In (C) motif diversity scores (normalised Shannon entropy) are calculated for genetic (top) or combined genetic and epigenetic (bottom) motifs with better separation observed for the latter, demonstrating the value of combined complete genetic and epigenetic information in a single dataset. (D) Both genetic and epigenetic motifs show significant (above the dotted line) differences in frequencies between healthy and stage IV CRC cfDNA. (E) Clustered heatmap of the top 30 discriminatory genetic (top) or genetic and epigenetic (bottom) end-motifs show clearer clustering when all 6-bases are used and include end motifs containing every cytosine state: unmodified, 5mC (coral highlighted) and 5hmC (green highlighted). Features were selected by Random Forest importance (Mean Decrease in Accuracy, 1000 trees). Columns represent samples and rows represent motifs, with hierarchical clustering (Ward's D2 method) applied to both axes. Tile colour indicates Z-scored motif frequency (blue: below mean; red: above mean). Annotation bars above the heatmap indicate sample group (coloured by diagnosis: control or Stage IV cancer) and sex. "X" marks samples with tumour fraction below 5%.

4. 6-base data is fully compatible with conventional nucleosome positioning analysis



Nucleosome position can be a powerful proxy for functional genomics in cfDNA by providing information on transcription factor (TF) binding. In (A) we show that 6-base data shows large changes in nucleosome occupancy between healthy controls and stage IV CRC for selected TFs (black are part of AP-1 TF complex, blue are pioneers) and similarly in (B) across 10,000 DNase hypersensitive sites specific for digestive tissues (left) or non-tissue specific (right). Lower coverage indicates higher TF occupancy calculated using Griffin. In (C) we show that TF occupancies have potential for differentiating healthy from stage IV CRC cfDNA through assessment of Pearson correlations where a lower number indicates lower correlation and therefore more ability to differentiate. (D) is an example of two highly correlated TFs in both healthy and controls and therefore limited utility. In (E) TF binding in stage IV samples are highly correlated but not in controls, suggesting more utility for distinguishing healthy from stage IV CRC.

5. 6-base fragmentomic features distinguish healthy from CRC cfDNA



The combination of all fragment length, genetic and epigenetic end motif and nucleosome positioning features were used to build classifiers to distinguish between healthy and stage IV CRC cfDNA. The classifier demonstrates high performance across all samples with an AUC of 0.783 (A) and that performance is stable across all seeds tested (D), and clear differentiation when we assess per-sample prediction probabilities (B). The most frequently selected features (C) demonstrate that the full complement of 6-base fragmentomic features are useful for distinguishing healthy from cancer with 5mC and 5hmC-containing end motifs (coral and green respectively) some of the most important. Also highlighted are genetic end motifs (orange) and nucleosomal positioning (light teal) biomarkers. Classification was performed using L1-regularised logistic regression (LASSO, glmnet) within a leave-one-out cross-validation (LOOCV) framework. To avoid data leakage, feature pre-filtering was conducted independently within each LOO fold: motif and transcription factor features were ranked by Fisher's discriminant ratio (between-group variance / within-group variance), retaining the top 25% of each category. Regularisation parameter λ was selected via stratified inner cross-validation. To account for stochasticity in fold assignment, the full LOOCV procedure was repeated across 25 random seeds.

6. Conclusions

- We demonstrate that the duet evoC technology enables all conventional fragmentomic biomarkers whilst revealing others that are entirely novel to the 6-base genome.
- In real world clinical cfDNA samples from healthy volunteers or those with stage IV CRC we show that these fragmentomic features are able to distinguish between cancer and healthy cfDNA.
- Classification approaches clearly show that 6-base fragmentomic biomarkers are some of the most commonly selected and therefore the most powerful.

In posters #123 and #7844 we present data showing the power of fragment-based classification approaches and combining genetic with 5mC and 5hmC features to enhance liquid biopsy performance. In combination with the data presented here they reinforce the ability of the 6-base genome, provided through duet evoC, to extract the most complete and powerful set of biomarkers that can improve early detection, therapy selection and detection of recurrence of cancer in a liquid biopsy setting.

7. References

1. Füllgrabe J. et al., Simultaneous sequencing of genetic and epigenetic bases in DNA. *Nat Biotechnol.* 2023 Oct;41(10):1457-1464.
2. Puddu F. et al., 5-methylcytosine and 5-hydroxymethylcytosine are synergistic biomarkers for early detection of colorectal cancer. *Commun Med* 6, 15 (2026).
3. Mouliere F., et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med.* 2018 Nov 7;10(466)
4. Jiang P, et al., Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer, Pregnancy, and Transplantation. *Cancer Discov* (2020) 10 (5): 664-673.
5. Doebley, et al., A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. *Nat Commun* 13, 7475 (2022)

