Integrated 6-base cfDNA profiling reveals complementary genetic, epigenetic and fragmentomic biomarkers, enabling the selection of the most powerful biomarker combinations for cancer detection

Fabio Puddu[1], **Annelie Johansson**[1], Angela Simeone[1], Mike Stubbington[1]
1. biomodal Ltd, The Trinity Building, Chesterford Research Park, Cambridge, UK

biomodal

1. Introduction

Cell-free DNA (cfDNA) contains complementary genetic, epigenetic and fragmentomic signals that can inform cancer detection. However, these biomarkers are typically measured using separate assays, increasing sample input requirements, analytical complexity, and inter-assay variability.

duet 6-base mosaic and **duet evoC** can simultaneously profile 5mC, 5hmC, genetic variants and fragmentomic features from a single cfDNA sample[1]. Using cfDNA in plasma from healthy controls and Stage I–IV colorectal cancer (CRC) patients, we derived multiple molecular features using **duet evoC** and evaluated their individual classification performance. Integrating modality-specific signal consistently improved CRC detection, demonstrating the value of multimodal liquid biopsy with 6-base sequencing.

2. Study design & analysis

CRC cohort: 92 individuals; 32 healthy donors, 60 CRC patients.[2]

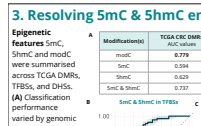
Features:
(i) **epigenetic signals** (5mC, 5hmC, modC) at CRC-associated TCGA DMR (differentially methylated regions)[3], TFBS (transcription factor binding sites)[3] and DHSs (DNase I hypersensitive sites)[4].
(ii) **genetic signals** represented by SBS96 mutational patterns[5]
(iii) **fragmentomic signals** including fragment size ratios[6], end motif frequencies[7], and nucleosome occupancy profiles[8].

Multimodal integration: Modality-specific classifiers were trained and evaluated using leave-one-out cross-validation followed by weighted integration of predictions to identify optimised biomarker combinations for CRC classification.

3. Resolving 5mC & 5hmC enhances CRC detection

Epigenetic features 5mC, 5hmC and modC were summarised across TCGA DMRs, TFBSs, and DHSs. **(A)** Classification performance varied by genomic context and modification. **(B)** The top-performing epigenetic model combined **5mC & 5hmC at TFBSs** (AUC = 0.844).

Modification(s)	TCGA CRC DMRs AUC values	TFBS regions AUC values	DHS regions AUC values
modC	**0.779**	0.759	0.637
5mC	0.594	0.779	0.657
5hmC	0.629	0.685	0.698
5mC & 5hmC	0.737	**0.844**	**0.728**



(C) Frequently selected 5mC and 5hmC features included TFBSs associated with CRC-relevant transcription factors such as TCF7L1, HMGA1, RUNX1, KDM2B, GATA4. **(D)** High sensitivity indicates strong detection of CRC, with moderate specificity for healthy controls.

Together, these results highlight that 5mC & 5hmC profiling at regulatory regions provides a informative and biologically relevant biomarkers for cfDNA-based CRC detection.

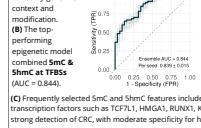
4. Mutational profiles distinguish later-stage CRC

Genetic features (SBS96 mutational profiles) were derived from high-confidence somatic variants identified using the biomodal duet software. Later-stage CRC samples with higher tumour fractions showed distinct mutational profiles, while early-stage CRC more closely resembled controls.

(A) PCA shows separation enriched for high-tumour fraction Stage III/IV samples (filled circles). **(B)** The SBS96 classifier achieved an AUC of 0.620 **(C)** Top substitutions included C[T>C]G, G[C>T]A, and A[T>C]G. **(D)** High sensitivity but low specificity indicates limited discrimination of healthy controls. Overall, genetic features alone showed modest predictive power but may provide complementary information when integrated with other modalities.

5. Methylation-aware end motifs strengthen CRC classification

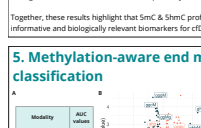
Modality	AUC values
Fragment Size Ratios (regional)	0.715
6-base End Motif Frequencies	0.708



Fragmentomics features. 6-base sequencing enabled estimation of regional fragment size ratios (FSR) and quantification of 6-base end-motif frequencies (EMF), incorporating cytosine modification states: unmodified (c), modC (C), 5mC (M), and 5hmC (H).

(A) Both fragmentomics features distinguished CRC from healthy controls, with FSR showing the highest performance (AUC=0.715), followed by 6-base EMF (AUC=0.708). **(B)** Top 6-base EMF features included canonical 4-base motifs (gcgg, atgc, attg) and modification-aware motifs (gggM, gHgH, ggMg) highlighting contributions from 5mC and 5hmC. **(C)** FSR showed high sensitivity with moderate specificity. **(D)** The 6-base EMF classifier achieved high sensitivity, but lower specificity, reflecting better performance for the identification of cancer than controls.

6. Nucleosome profiling reveals CRC-associated TF activity



Nucleosome profile features were derived from 6-base fragment-level coverage at TFBSs, providing a read-out of nucleosome occupancy and chromatin accessibility; lower central coverage indicates higher accessibility.

(A) TFBS nucleosome profiles distinguished CRC from healthy controls (AUC = 0.657). **(B)** Top features included CRC-associated transcription factors WT1, HMGA1, FOSL1, and KDM2B. **(C)** The classifier had high sensitivity with moderate specificity. **(D)** FOSL1 showed reduced central coverage in Stage IV CRC versus control, consistent with increased accessibility and accompanied by changes in methylation, reduced 5mC and 5hmC, that are expected to drive the opening of chromatin. Together, nucleosome profiling provides a complementary fragmentomic readout of CRC-associated regulatory changes in cfDNA with intriguing complementarity with 5mC and 5hmC changes that would be invisible with a traditional modC readout.

7. Multimodal cfDNA integration improves CRC classification

Epigenetic, genetic and fragmentomic classifiers were combined in a **late-fusion framework** using weighted prediction probabilities. Together, these results show that combining complementary 6-base cfDNA signals improves CRC classification beyond individual modalities.

(A) Weights optimisation improved performance (AUC 0.821 → 0.890). **(B)** Key contributors included modC in TCGA DMRs, 5mC & 5hmC in TFBSs and DHSs, 6-base EMF and FSR. **(C)** Multimodal integration produced stepwise gains in classification performance. **(D)** The multimodal model correctly reclassifies 3 CRC samples missed by the best modC model.

8. Conclusions

6-base sequencing data enables simultaneous interrogation of epigenetic, genetic and fragmentomic biomarkers from a single cfDNA sample. Each modality captures distinct complementary aspects of tumour biology, with epigenetic features providing the strongest individual signal. Integrating modalities consistently improves CRC classification, achieving an AUC of 0.890. Together, these results demonstrate the potential of a unified multi-modal workflow to maximize liquid biopsy performance while reducing assay complexity.

9. References

1. Fullgrabe J, et al. Simultaneous sequencing of genetic and epigenetic bases in DNA. Nat Biotechnol. 41(9):1457-1464 (2023).
2. Puddu F, et al. Genetic, epigenetic and 5-hydroxymethylcytosine are synergistic biomarkers for early detection of colorectal cancer. Commun Med. 4(1):2 (2024).
3. Doebley AL, Ko M, Liao H, et al. A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. Nat Commun. 13:7475 (2022).
4. Meuleman W, et al. Index and biological spectrum of human DNase I hypersensitive sites. Nature. 584:244-251. doi:10.1038/s41586-020-2559-3 (2020).
5. Islam S, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genomics. 2(11) (2022).
6. Mouliere F, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med. 10(466) (2018).
7. Jiang P, et al. Plasma DNA end-motif profiling as a fragmentomic marker in cancer, pregnancy, and transplantation. Cancer Discov. 10(5):664-673 (2020).
8. Doebley AL, et al. A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. Nat Commun. 3(1):7475 (2022).

biomodal