

# 6-base multi-modal data is a powerful liquid biopsy platform for cancer detection, enabling the discovery of epigenetic, genetic and fragmentomics biomarkers in a single workflow

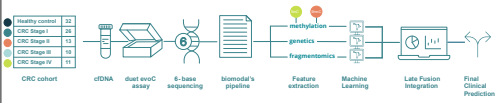
Fabio Puddu<sup>1</sup>, Annelie Johansson<sup>1</sup>, Cillian Nolan<sup>1</sup>, Robert Blanshard<sup>1</sup>, Edyta Bocian-Canepa<sup>1</sup>, Tom Charlesworth<sup>1</sup>, Angela Simeone<sup>1</sup>, Mike Stubbington<sup>1</sup>  
 1. biomodal Ltd, The Trinity Building, Chesterford Research Park, Cambridge, UK



## 1. Introduction

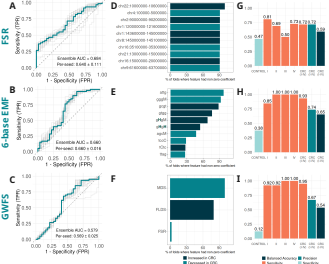
**Cell-free DNA (cfDNA)** contains complementary genetic, epigenetic and fragmentomics signals relevant for cancer detection but these biomarkers are typically assessed using separate assays, increasing sample input requirements, analytical complexity and inter-assay variability. We evaluated 6-base sequencing with **duet evoC**, which simultaneously profiles 5mC, 5hmC, genetic variants and fragmentomics features from a single cfDNA sample<sup>1</sup>. Using plasma from healthy controls and Stage I-IV colorectal cancer (CRC) patients, we assessed the classification performances of each modality and the benefit of integrating modality-specific predictions. Multi-modal integration consistently improved CRC detection, demonstrating the value of a unified liquid biopsy workflow

## 2. Study design & analysis



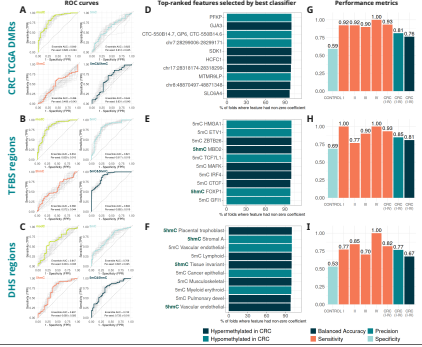
**CRC cfDNA cohort<sup>2</sup>**: 32 healthy controls; 26 CRC patients (26 - I; 13 - II; 10 - III; 11 - IV). **Epigenetic features** were 5mC, 5hmC and collapsed methylation signal (modC) at CRC-associated TCGA differentially methylated regions (CRC TCGA DMRs), transcription factor binding sites (TFBSs)<sup>3</sup> and DNase I hypersensitive sites (DHSs)<sup>4</sup>. **Genetic information** was represented by SBS96 mutational signatures<sup>5</sup>. **Fragmentomic features** included fragment size ratios, end motif frequencies and genome-wide fragmentation statistics. **Classifiers** were developed using leave-one-out cross-validation (LOOCV) and subsequently combined through weighted integration to generate an optimized multi-modal predictor.

## 5. duet evoC enables fragmentomics analysis, including methylation-aware 6-base end motif profiling



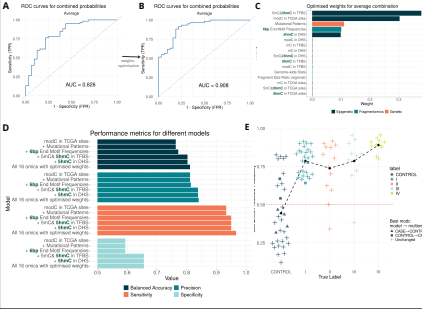
**Fragmentomics features** include regional fragment size ratios (FSR), 6-base End Motif Frequencies (EMF) (a, c, t, g, modC [C], 5mC [M], 5hmC [H]) and genome-wide fragmentation statistics (GWFS), including sample-level Motif Diversity Score (MDS), Fragment Length Diversity Score (FLDS), calculated as normalised Shannon entropy of 45-145 bp fragments, and genome-wide Fragment Size Ratio<sup>6,7</sup>. These features showed independent discriminatory power, with **(A)** regional FSR (AUC=0.684) as the strongest model for discriminating CRC from healthy samples, followed by **(B)** 6-base EMF models (AUC=0.660) and **(C)** GWFS models (AUC=0.579). **D-F** show the top features, where 6-base EMF classifiers **(E)** included both 4-base (atg, gcgt, gttgg) and 6-base end motifs containing modified cytosines (ggm, ghm, ghH), highlighting the added predictive value of epigenetic modifications. Highest sensitivity was observed in 6-base EMF **(G)** and genome-wide fragmentomics statistics **(H)**, whereas highest specificity was observed for regional FSR **(I)**.

## 3. Resolving 5mC & 5hmC enhances CRC detection across all stages

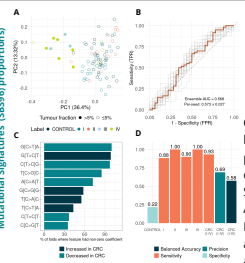


**Epigenetic features** (5mC, 5hmC, and modC) were summarised across TCGA DMRs, TFBS, and DHS regions. Across these contexts, epigenetic signals consistently distinguished CRC from healthy samples, with performance dependent on genomic context and modification type. The best-performing models were observed for combined 5mC & 5hmC at TFBS (AUC = 0.893) **(B)**, while modC achieved the highest performance for TCGA DMRs (AUC = 0.840) **(A)**, likely reflecting that these DMR regions were originally identified using modC data. For TFBS and DHS, combined 5mC & 5hmC models consistently outperformed individual signals **(B-C)**. Top features mapped to CRC-relevant genes (PFKP, GP6, SLC6A4) **(D)** and regulatory elements (HMGAI1, ETV1, TCF7L1) **(E)**, supporting enrichment in functionally active regions. Epigenetic classifiers showed consistently high sensitivity for cancer samples across stages, with strongest performance in later stages and maintained signal in early-stage disease **(G-I)**. Together, these results highlight 5mC/5hmC profiling in regulatory regions as a strong and biologically meaningful signal for cfDNA-based CRC detection, particularly where mutation signal is limited.

## 6. Integrating the 6-base derived modalities improves CRC classification performance



To assess the complementary value of the 16 individual biomarker classes enabled by 6-base sequencing, the epigenetic, genetic and fragmentomic CRC classifiers were integrated into a unified multi-modal, late-fusion framework. Epigenetic features alone provided strong classification performance, with additional gains achieved by incorporating genetic and fragmentomic signals, reflecting their complementary biological contributions. Optimising modality contributions (weights) further improved performance (AUC 0.826 → 0.908) **(A-B)**, enabling the model to prioritise the most informative features, including 5mC & 5hmC in TFBS, modC in TCGA, Mutational Signatures, 6-Base EMF and 5hmC in DHS **(C)**. Multi-modal integration gave stepwise performance gains **(D-E)**, and rescued Stage I samples incorrectly classified as CONTROL using the best modC model **(E)**, demonstrating proof of concept for how integrating 6-base modalities translates into meaningful improvements in classification.



## 4. Mutational signatures enable late-stage CRC classification, but is limited in low tumour fraction samples

**Genetic features** as SBS96 proportions were derived from high-confidence somatic variants identified using the biomodal duet pipeline (VAF ≤ 0.35, depth ≥ 15X, and population allele frequency ≤ 1 × 10<sup>-4</sup> in gnomAD). High proportions were observed in late-stage (III/IV) CRC samples with high tumour fractions, but remained low otherwise, indicating genetic cfDNA biomarkers are more informative at advanced cancer stages. A PCA of the SBS96 mutational profiles **(A)** shows a subset of Stage III/IV samples as outliers. Mutational signatures achieve an AUC of 0.568 **(B)** for the classification of CRC vs Control with top substitutions G[C>T]A, G[T>C]T, and C[T>C]G **(C)**. Performance metrics show high sensitivities for late stage CRC (II-IV), moderate precision and balanced accuracy across stages, but low specificity for Controls **(D)**. This suggests limited discriminative power for genetic information alone in cfDNA, but it becomes an important contributor when combined with other omics.

## 7. Conclusions

**6-base sequencing** with **duet evoC** enables simultaneous interrogation of epigenetic, genetic and fragmentomic biomarkers from a single cfDNA sample. Each modality captures distinct aspects of tumour biology, with epigenetic features providing the strongest individual signal. Importantly, combining modalities consistently improves classification performance, achieving an AUC of 0.91 for CRC detection. These results demonstrate the potential of a unified multi-omic workflow to maximize liquid biopsy performance while reducing assay complexity.

## 8. References

- Fullgraber J, et al. Simultaneous sequencing of genetic and epigenetic bases in DNA. Nat Biotechnol. 2023 Oct;41(10):1457-1464.
- Puddu F, et al. 5-methylcytosine and 5-hydroxymethylcytosine are synergistic biomarkers for early detection of colorectal cancer. Commun Med. 15 (2025).
- Dowling AL, Kuo M, Liao H, et al. A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. Nat Commun. 13: 7475 (2022).
- Meuleman W, et al. In-depth biological spectrum of human DNase I hypersensitive sites. Nature. 2020;584:244-251. doi: 10.1038/s41586-020-2559-3.
- Idam S, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genomics. 2, 11 (2022).
- Mojir P, et al. Enhanced detection of circulating tumour DNA by fragment size analysis. Sci Transl Med. 2018 Nov 7;10(466).
- Jiang P, et al. Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer. Pregnancy, and Transplantation Cancer Discov. 2020; 10 (10): 664-673.

