

# Multinomic 6-base sequencing enhances the performance of early colorectal cancer detection from cell-free DNA

Ritesh Prasad, Anshu Khanna, Arun Kumar, James Spillner, Rishu Sethi, Nick Harding, Alex Hill, Emma Usher, Casper Lumby, Jean Taysseaud, Michael Wilson, Robert Crawford, Tom Charalambous, Robert Osborne, Shunhai Babubharanathan, Paul O'Keefe



## Introduction

Early detection of colorectal cancer (CRC) has the potential to improve treatment outcomes and survival rates. Liquid biopsy for profiling of cell free DNA (cfDNA) in blood holds huge promise for early CRC detection in otherwise asymptomatic patients.

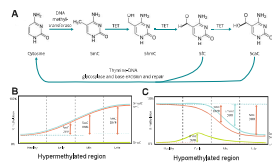
Epigenetic biomarkers have already been shown to significantly contribute to cancer detection in liquid biopsies, but traditional DNA methylation sequencing correlates two cytosine modifications, 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC), with different and opposing biological functions. Discrimination of these two states could therefore be crucial for increasing the amount of functional information for CRC detection.

We therefore applied duet evoc, a bimodal technology that provides the 6-base genome (the complete genetic sequence whilst simultaneously distinguishing 5mC and 5hmC), to cfDNA obtained from a cohort of 32 healthy volunteers and 37 patients with CRC at stages I and IV. Through machine learning approaches, we built classifiers to differentiate between cfDNA from patients with stage I CRC and individuals without cancer using features based 5mC alone, 5hmC alone, both 5mC and 5hmC, or the combined 5mC/5hmC signal (mo5C), as would be generated by traditional epigenetic technologies.

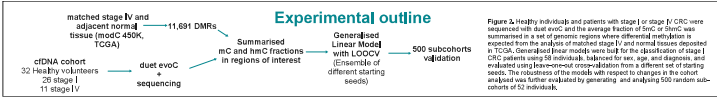
Our findings reveal that, compared to traditional approaches, 5mC and 5hmC behave synergistically for the detection of early stage CRC, enhancing diagnostic accuracy (AUC = 0.95). Notably, most regions with an increase in 5hmC in stage I CRC correlate with decreased in 5mC in stage IV CRC, suggesting that 5hmC could be an early marker of demethylation that occurs further down the cancer development trajectory.

Applying duet evoc 6-base sequencing to larger clinical cohorts, and across different indications, will evaluate the potential of 6-base data to improve the earlier detection, diagnosis, and treatment of other diseases.

## 5hmC as a potential marker for early CRC detection

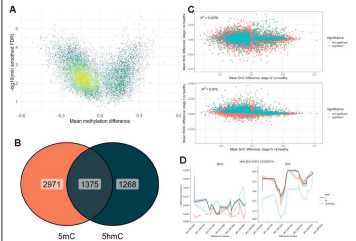


**Figure 1.** (A) Schematic of the enzymatic steps involved in cytosine methylation and demethylation. DNA methyltransferases catalyse the addition of a methyl group to a cytosine base in a CpG context. The methyl group is oxidised by TET enzymes to 5mC, and downstream oxidised derivatives 5fC and 5caC, before thymine DNA glycosylase mediates the removal of 5mC or 5fC, with subsequent replacement with unmethylated cytosine by gap-filling events. (B) Schematic representation of the increased 5mC levels in hypermethylation or CpG loss of 5mC leading to hypomethylation in a region of the genome. (C) Represents a region with 5mC hypomethylation in late-stage cancer, where 5mC accumulates during disease progression, and these regions, changes in 5mC are proportionally equivalent and would therefore be expected to have similar power as biomarkers to distinguish between disease stages and healthy individuals. (D) Represents a region with hypomethylation in late-stage disease. Changes in 5mC are proportionally larger at the early stage than changes in 5mC or mo5C. At later stages, as demethylation completes, 5mC and mo5C become proportionally larger. Changes in mo5C are masked by the correlation of 5mC and 5hmC and only become distinguishable at mid-to-late-stage.



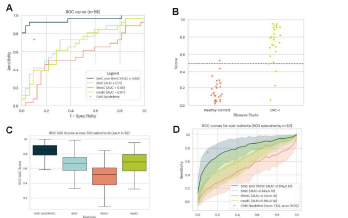
**Figure 2.** Healthy individuals and patients with stage I or stage IV CRC were recruited with their 5mC and the average fraction of 5fC or 5caC, with summarised a set of genetic regions under differential methylation. A subset of these regions was used for the analysis of matched stage IV and normal tissues deposited in TCGA. Generalised linear models were built for the quantification of stage I CRC patients using 58 individuals, balanced for sex, age, and diagnosis, and evaluated using leave-one-out cross-validation from a different set of testing sets. The robustness of the models with respect to changes in the cohort analysed was further evaluated by generating and analysing 500 random sub-cohorts of 52 individuals.

## Loss of methylation in late stage CRC correlates with gains of hmC in stage I



**Figure 3.** (A) Venn diagram for DMRs between stage IV CRC tissue and adjacent matched normal, using data from TCGA. Colours indicate density of points, with yellow used to indicate a higher density. Demethylation is a prominent feature in late stage CRC tissue compared to gain of methylation. The vast majority of DMRs identified in tissues could be reproduced from stage IV cfDNA samples (10,657 out of 11,691). (B) Venn diagram of regions with statistically significant ( $p < 0.05$ ) differences in 5mC and 5hmC in stage I plasma, with 1,202(54) of these DMRs did not show statistically significant differences in 5mC, suggesting that measuring 5hmC in stage I plasma provides complementary information to that obtained from 5mC. (C) Scatter plot comparing 5mC (upper panel) and 5hmC (lower panel) differences measured in stage I and stage IV plasma. Statistical significance ( $p < 0.05$ ) at stage I is indicated by blue colour. The majority (71.2%, 1,646(54)) of DMRs with statistically significant changes in 5hmC in stage I and 5mC in stage IV plasma had an increase in 5hmC at stage I and a decrease in 5mC at stage IV. (D) Example of a genetic region where loss of 5mC in late stage CRC is preceded by gain of 5hmC in early stage. Trace plots of 5hmC (left) and 5mC (right) fractions.

## Combination of 5mC and 5hmC features provides greater discriminatory power for stage I CRC detection



**Figure 4.** (A) Receiver operating characteristic curves for mo5C only, 5mC only, 5hmC only, and combined 5mC and 5hmC models, generated over an ensemble of 25 random seeds. Combining 5mC and 5hmC information substantially improves performance. (B) Sample scores for healthy controls and CRC patients from the combined 5mC and 5hmC model. A threshold of 0.5 is indicated with the dotted line. The score of the only false positive is close to the threshold, while that of the false negatives are very close to the threshold. (C) Box plot showing AUCs across 500 subcohorts of 52 individuals for mo5C only, 5mC only, 5hmC only, and combined 5mC and 5hmC models. While 5mC and 5hmC models can also reach high performance, mo5C consistently performs best on the cohort analysed, combined 5mC and 5hmC models achieve higher performance more consistently. (D) Average ROC curves for 500 subcohorts of 52 individuals, error bands correspond to one standard deviation. On average, 5mC only and 5hmC models perform indistinguishably from each other, while combined 5mC and 5hmC models perform consistently better.

## Conclusion

This study leveraged the duet evoc solution to distinguish between 5mC and 5hmC, allowing for the analysis of sequencing data using mo5C, 5mC-only, 5hmC-only, and both 5mC and 5hmC models. This flexible approach provides a powerful tool for discovery, where different indications may exhibit different levels of signal from 5mC, 5hmC, and genetics, building on the discovery step. It is then possible to scale an assay that focuses specifically on the combination of methylation and genetics that yield the required signal. In conclusion, distinguishing between 5mC and 5hmC enhanced performance in early-stage CRC detection, showing improvement when compared to mo5C models. Our data supports the use of sequencing methods that distinguish 5mC and 5hmC in early detection. We feel there may be merit in applying such an approach in future studies with larger clinical cohorts, across different cancers, to more fully evaluate the potential to improve early detection performance.

## References

1. Fullgrabe J, et al. Simultaneous sequencing of genetic and epigenetic bases in DNA. Nat Biotechnol. 2023 Oct 4;41(10):1457-1464.
2. Purohit P, et al. 5-methylcytosine and 5-hydroxymethylcytosine are synergistic biomarkers for early detection of colorectal cancer. bioRxiv. 2024.10.30.621152.

