Refining liquid biopsy with the 6-base genome: Generating more information from cell free DNA

biomodal

Fabio Puddu, **Tom Charlesworth**, Robert Crawford, Nick Harding, Riccha Sethi, Jamie Scotcher, Annelie Johansson, Ermira Lleshi, Aurelie Modat, Michael S Wilson, Páidí Creed

1. Introduction

Liquid biopsy for profiling of cell free DNA (cfDNA) in blood holds huge promise to transform how we experience and manage cancer by early detection and identification of residual disease and subtype. However, a standard blood draw yields an average of only 10 ng of cfDNA, of which DNA derived from the tumour is a small minority.

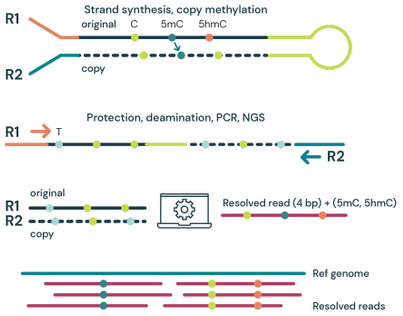
Genetic and methylation data together have been shown to be more powerful for the detection of early cancer than either alone. Constrained to measuring four states of information, existing NGS-based technologies sacrifice genetic information for methylation calling.

State	Standard Sequencing Protocol	Protocol with C→T deamination
1	A	A
2	C/mC/hmC	mC/hmC
3	G	G
4	T	C/T

duet multiomics solution evoC is a new sequencing technology that simultaneously derives all four genetic bases without ambiguity in C or T calls alongside distinguishing 5-methylcytosine and 5 hydroxymethylcytosine (**6-base** data) in a single read from a single DNA molecule [1]. The technology consists of pre-sequencing library prep and post-sequencing analysis pipeline, providing single-base resolution of genetics and epigenetics at high accuracy

2. duet multiomics solution evoC

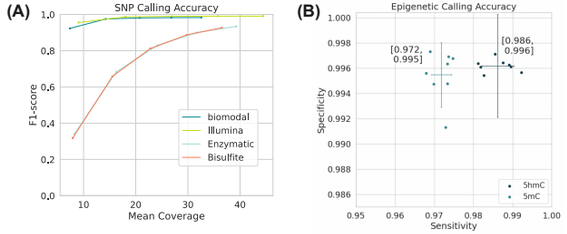
1. **Strand synthesis** - creates a single molecule with a direct copy of the original information tethered together with a hairpin. The copy strand is without cytosine modifications initially, but importantly, utilises a high fidelity methyltransferase to copy over only 5mC from the original to the copy strand.
2. **Sequencing paired-end read**-generates sequence information after protection of cytosine modifications followed by deamination of all remaining cytosines (read as thymine in NGS).
3. **Read resolution**- aligns original and copy strands to correctly call all 4 canonical bases along with 5mC and 5hmC.
4. **Aligned** (4 base) reads with 5mC & 5hmC are tagged (**6 base information**)



duet multiomics solution evoC is a 6-base calling technology that reads all four canonical bases plus 5mC and 5hmC.

3. Accurate genetic and epigenetic data

Complete accurate genome and methylome information from duet evoC. Genomic information is provided in full at higher or equivalent accuracy as other genetic and epigenetic methods (Figure 3A).



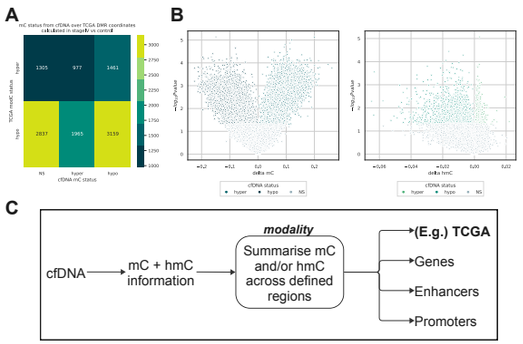
Simultaneously full methylome information is provided at high accuracy for both 5mC and 5hmC (Figure 3B). Figure **(A)** uses data generated on the Genome-in-a-bottle reference materials [2]

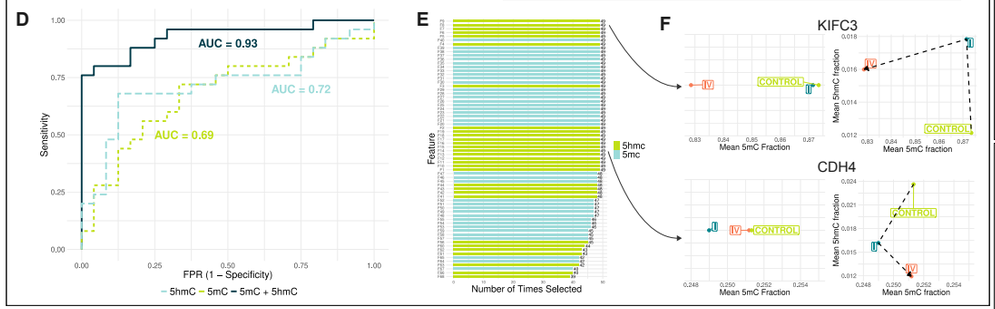
4. Accurate detection of genetic variants in a CRC patient cohort

cfDNA is thought to enter the bloodstream through apoptosis or necrosis, with cfDNA from healthy and cancer tissues released into the blood of cancer patients. To assess the ability of *duet evoC* to generate multimodal information from liquid biopsy samples, we obtained and sequenced cfDNA from 87 individuals, ranging from healthy volunteers to stage IV colorectal cancer (CRC) patients. **(A)** Somatic variants were called using Mutect2 on *duet evoC* in tumour-only mode and the presence of pathogenic or likely pathogenic variants associated with CRC is shown. An increase in the prevalence of these variants from stage I to stage IV patients mirrors an increase in the amounts of ctDNA **(B)**, as estimated by ichorCNA [3].

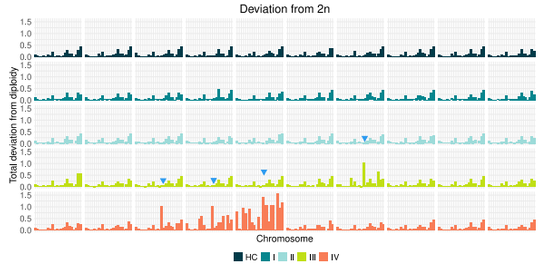
5. Accurate epigenetic information for CRC detection

(A) DMRs identified from stage IV CRC tissue in the TCGA-COAD cohort [4] using Infinium Human Methylation 450K arrays were reproduced from 5mC levels in cfDNA obtained from stage IV CRC patients using *duet evoC* (12 CRC; 24 healthy volunteers). Cell-free DMRs where 5mC was greater in stage IV were defined as hypermethylated, and vice versa ($p < 0.05$; t-test). **(B)** Within these DMRs from the TCGA-COAD cohort, *duet evoC* captured differential (hydroxy)methylation in cfDNA from stage IV CRC patients. **(C)** Analysing cfDNA with *duet evoC* produces 6-base readouts that can be summarised across regions with ease using *modality*, part of the *duet* analysis suite. **(D)** ROC curves demonstrating that combining 5mC and 5hmC features in cfDNA improved separation of stage I CRC patients from healthy volunteers, when compared with using 5mC or 5hmC alone. Candidate features were defined as 5mC and 5hmC fractions of regions defined as stage IV tissue DMRs in TCGA-COAD, calculated from cfDNA from 24 stage I CRC patients and 25 healthy volunteers. Using *glmnet* [5], generalised linear models were trained on either 5mC or 5hmC features, or both (5mc + 5hmC), and evaluated using a leave-one-out cross-validation (LOOCV) approach. **(E)** Candidate features ranked by the number of times they were selected in each train/test split during LOOCV of the (5mC + 5hmC) model. **(F)** Analysing 5hmC improved the ability to distinguish between CRC stages in genomic regions with subtle 5mC differences. 5hmC (but not 5mC) fractions of regions overlapping KIFC3 and CDH4 were selected as features in every LOOCV split. Averages were taken across the stage I CRC patients and healthy volunteers from (D), and cfDNA from 12 additional stage IV CRC patients.



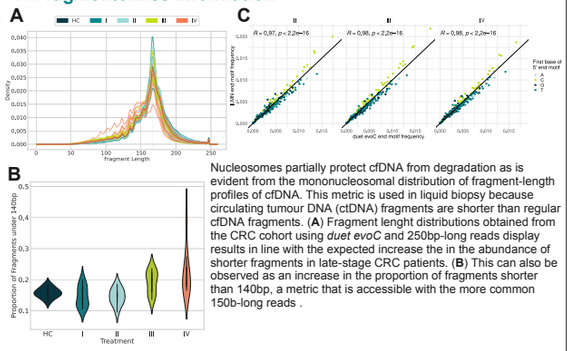


6. Copy Number Variation



Copy number information extracted with CNVkit [6] from 10 cfDNA samples for each stage shows an increasing prevalence of chromosomal aberrations in later stages of CRC. Note that some chromosomes appear to be consistently rearranged in several samples (blue arrows)

7. Fragmentomics information



Nucleosomes partially protect cfDNA from degradation as is evident from the mononucleosomal distribution of fragment-length profiles of cfDNA. This metric is used in liquid biopsy because circulating tumour DNA (ctDNA) fragments are shorter than regular cfDNA fragments. **(A)** Fragment lenght distributions obtained from the CRC cohort using *duet evoC* and 250bp-long reads display results in line with the expected increase the in the abundance of shorter fragments in late-stage CRC patients. **(B)** This can also be observed as an increase in the proportion of fragments shorter than 140bp, a metric that is accessible with the more common 150b-long reads .

(C) High correlation between end motif frequencies obtained from duet evoC reads and regular Illumina reads for the same sample indicates that *duet evoC* can also be used to accurately measure this cfDNA feature. A sample for each of the indicated stages is shown.

8. Variant-associated methylation

Simultaneous sequencing of genetic and epigenetic information allows phasing of the methylation/hydroxymethylation signal with heterozygous variants, and detection of variant-associated methylation across approximately 300 bp around the variant of interest. In this panel an example from the NDUFA9 gene is shown, where the G allele is associated with near complete methylation, while the A allele displays

9. Conclusions

We have presented data illustrating the potential of duet evoC for liquid biopsy. With duet evoC it is possible to obtain multi-modal information, including SNPs, methylation, hydroxymethylation, fragmentomics, copy number variation and novel 2 dimensional biomarkers, from a single low-input sample of cfDNA. Further we have demonstrated improved ability to differentiate between healthy and stage IV CRC using combined methylation and hydroxymethylation features.

duet evoC, which provides 6-base sequencing, is available to order now as a product from biomodal.

10. References

1. *Simultaneous sequencing of genetic and epigenetic bases in DNA*, Füllgrabe and Gosal et al., Nature Biotechnology (2023) **(duet multiomics solution technology paper)**
2. Zook et al., Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci Data* **3** (2016)
3. Adalsteinsson, V.A., Ha, G., Freeman, S.S. *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* **8**, 1324 (2017)
4. *TCGA-COAD*, https://portal.gdc.cancer.gov/projects/TCGA-COAD
5. Tay JK, Narasimhan B, Hastie T. "Elastic Net Regularization Paths for All Generalized Linear Models." *Journal of Statistical Software*, **106**(1), 1–31 (2023)
6. Talevich et al., CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Computational Biology* (2016)

biomodal

View online

biomodal.com