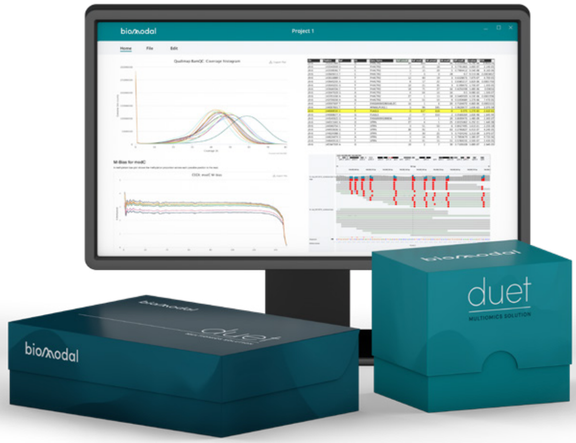
The 6-base genome: a breakthrough technology for maximising multiomic data from cell-free DNA

Your single solution for multimodal data from liquid biopsy samples



Liquid biopsy is a powerful tool in oncology for the non-invasive detection of disease biomarkers, treatment selection, and disease recurrence monitoring. Remarkable technology advances are enabling liquid biopsy application in research and precision oncology, however several challenges remain.

It is increasingly clear that multiomic data are important for optimal liquid biopsy test performance and discovery power[1,2]. Genetic sequencing has been the primary application to obtain information from these samples. Traditional sequencing data only provides information on actionable somatic mutations, and the detection of those loci in circulating cell-free DNA (cfDNA) has remained challenging.

Combining genetic sequence with DNA methylation data is used to explore the relationship between genetic variants and changes in DNA methylation. Adding methylation information can differentiate cancer from non-cancer samples, as 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) profiles have been shown to be predictive of early-stage cancer. By interrogating 5mC and 5hmC, you can explore beyond the genome by examining the methylation status of control elements, differentially methylated regions, and other variant associated methylation states at single-base resolution. The ability to differentiate 5mC from 5hmC offers the potential to discover novel and more sensitive biomarkers in liquid biopsy.

cfDNA samples are limited. A standard blood draw yields an average of only 10 ng of cfDNA, of which circulating tumour DNA (ctDNA), is a small minority. With limited input amounts, splitting samples into two workflows for separate genetic and epigenetic analyses is often not possible. For biomarker discovery, the ideal solution would provide complete genetic and epigenetic information from a single low input sample.

To overcome these challenges and get the most biologically relevant information from scarce cfDNA samples, we introduce duet multiomics solution evoC. duet evoC is an innovative technology that integrates complete genetic sequence data with differentiated 5mC and 5hmC at single-base resolution from low nanogram amounts of cfDNA.

Figure 1: duet evoC provides multimodal data without compromised accuracy. The fully enzymatic workflow provides complete and accurate 6-base genome information from the same cfDNA sample with reference quality SNP calling without ambiguity in C-to-T mutation calls and high sensitivity and specificity whilst distinguishing 5mC from 5hmC.

Accurate SNP calls

To validate the performance of the duet evoC solution on cfDNA samples, 5 ng of cfDNA was extracted from plasma taken from 88 patients with colorectal cancer (CRC) and healthy volunteers, and sequencing libraries were prepared using duet evoC followed by sequencing and analysis. duet evoC delivers SNP calling accuracy greater than or equal to current genetic and epigenetic sequencing technologies and provides methylation state information (5mC vs. 5hmC) at high sensitivity and specificity (Figure 2).

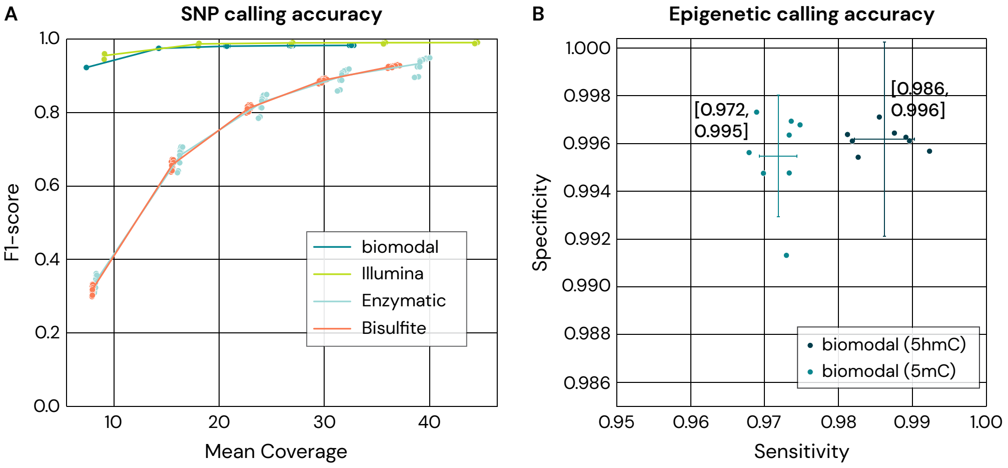


Figure 2: duet evoC provides complete, accurate genome and methylome information in a single sample. (A) SNP calling accuracy was assessed on Genome in a Bottle (GIAB) genomic DNA (gDNA) reference materials. duet evoC samples deliver higher or equivalent performance compared to Illumina WGS or common epigenetic sequencing approaches. (B) Accurate 5mC and 5hmC information is calculated from 5mC, 5hmC, and unmodified controls.

Determine cfDNA fragment size

cfDNA from healthy and cancerous tissues is known to enter the bloodstream through apoptosis or necrosis during cell death. Nucleosomal complexes partially protect cfDNA from nuclease degradation, seen in the fragment-length profile. Lengths tend to follow a tight distribution ranging between 150-200 bp and multiples thereof, with the majority of cfDNA being of the mononucleosomal variety. In ctDNA many fragments are ~20bp shorter which has led to the use of fragment length size distributions as features in liquid biopsy applications.

duet evoC can be used to infer the length of original fragments that are smaller than read length. In 88 cfDNA samples, including 25 healthy and 53 CRC stage I-IV, fragment size data obtained with duet evoC aligned with published observations of shorter mononucleosomal fragments in later stage cancer[3] (Figure 3). Fragment profiles combined with other genetic and epigenetic features may more accurately detect disease.

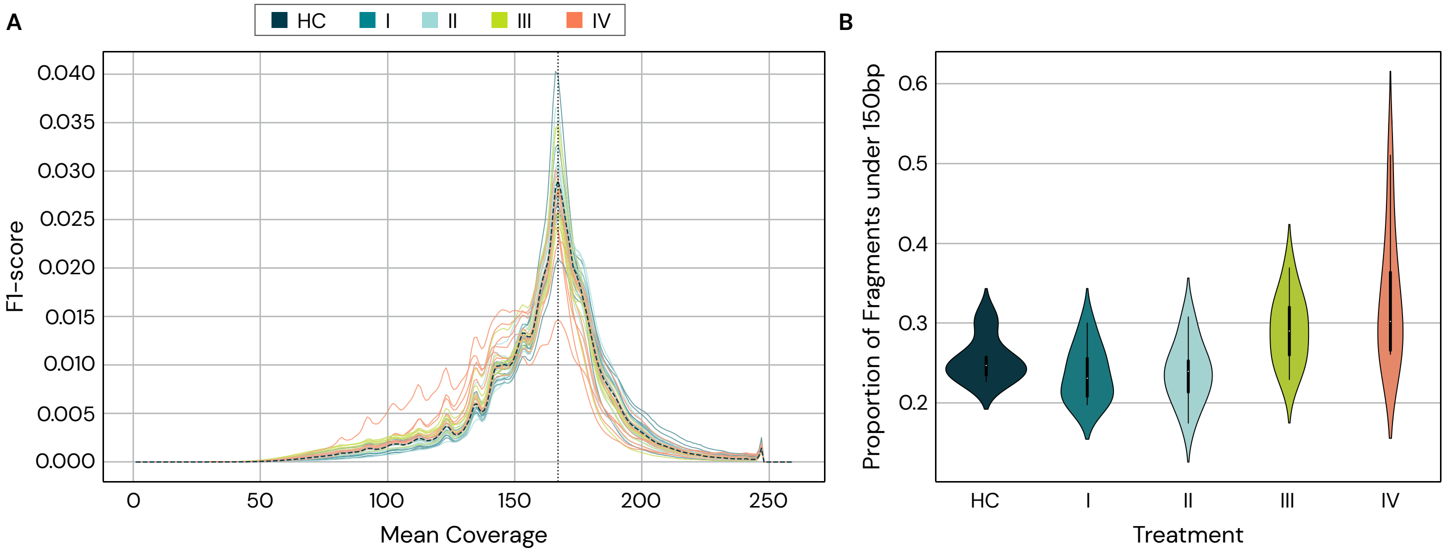


Figure 3: Fragment size distribution of cfDNA fragments derived from duet evoC data from a cohort of CRC and healthy samples. Included are traces (A) and violin plots (B) for healthy (black), CRC stage I (dark teal), stage II (light teal), stage III (green) and stage IV (coral).

See variant and CNV information

From the same CRC cfDNA cohort, variants and copy number information were extracted from duet evoC data, with expected trends of increased variants and deviations from diploidy in later stage cancer samples (Figures 4 and 5). With duet evoC there is no loss of genetic information, providing a complete picture of genetic aberrations in your cfDNA samples.

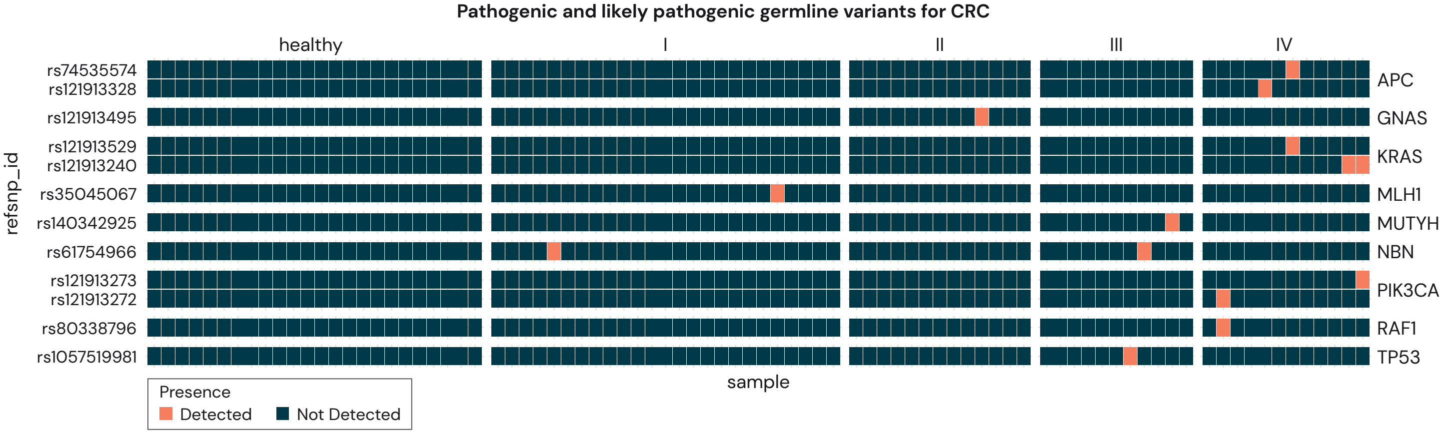


Figure 4: Pathogenic or likely pathogenic CRC variants identified using duet evoC data from healthy or stage I-IV cfDNA samples. Each column represents a single cfDNA sample with rows representing variant loci.

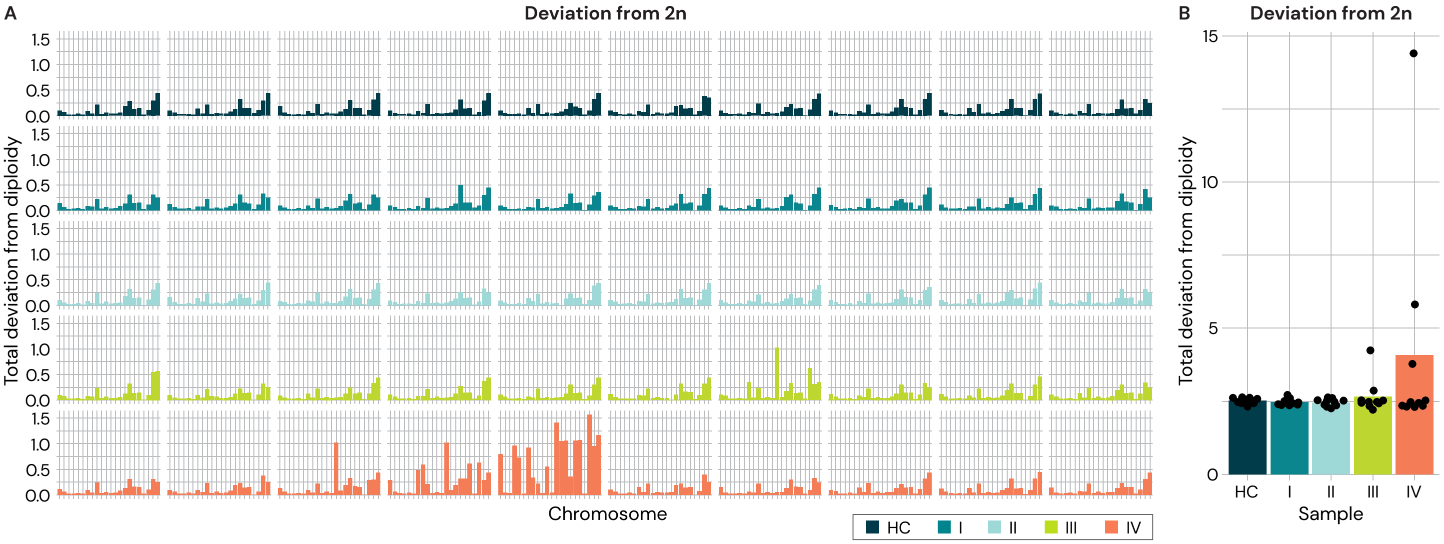


Figure 5: Copy number information extracted from cfDNA analysed with duet multiomics data. Deviation from diploidy is plotted across all chromosomes in all samples (A) and summarized across healthy and stages I to IV. (B) Data analysed with CNVkit[4].

See differential methylation and hydroxymethylation states

Comparing stage I-IV cfDNA samples from healthy individuals allows identification of differentially methylated and hydroxymethylated regions (Figure 6). Dynamics of methylation and hydroxymethylation are apparent in the different profiles of the volcano plots, revealing additional biomarkers that would not be visible without duet evoC data, enabling functional interpretation based on 5mC and 5hmC status at individual loci.

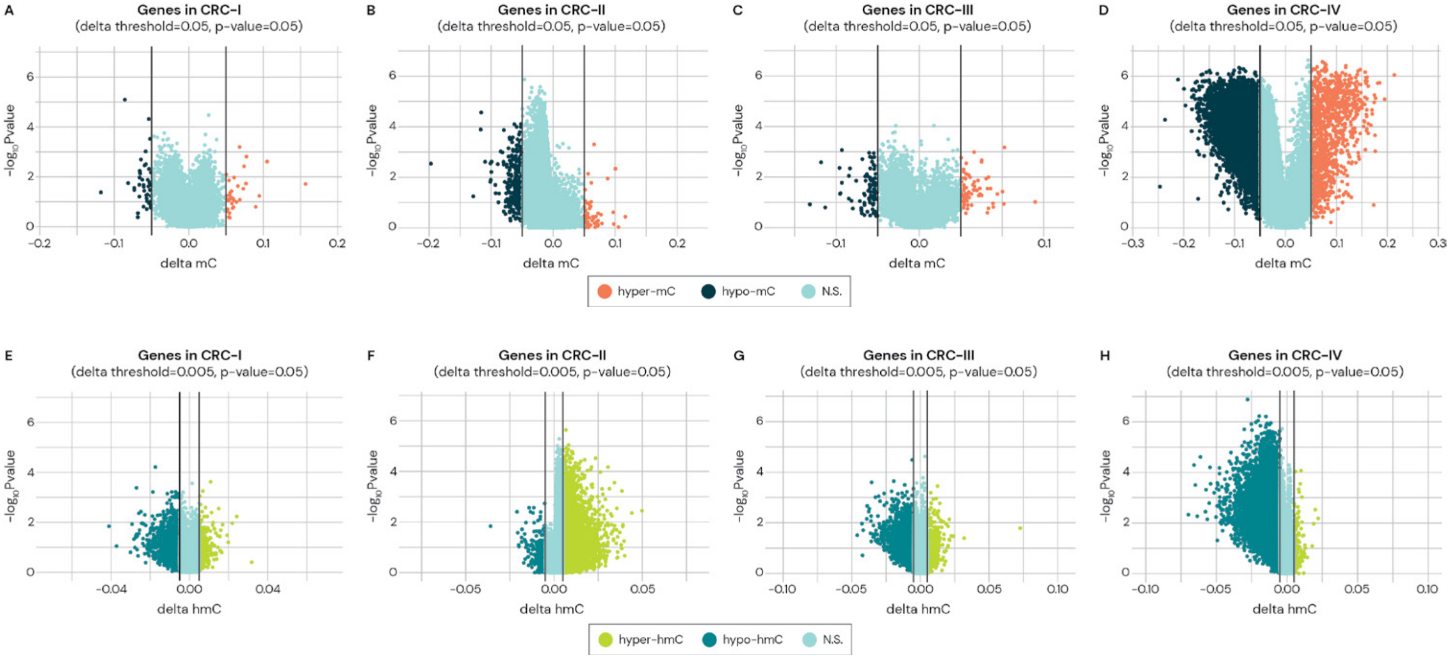


Figure 6: Volcano plots representing either 5mC or 5hmC differentially methylated regions (DMRs) calculated from duet evoC data. Plotted are changes in methylation on the x axis against the significance of those changes on the y axis. Data is plotted for methylation and hydroxymethylation across genes

Discover novel variant-associated methylation

Variant-associated methylation directly links a genetic variant with methylation variation. The genetic variant may be germline or somatic, identified between samples or between different alleles within a sample. The unique ability to accurately determine genetic and epigenetic variation on the same read is a powerful feature of duet evoC data and represents another biomarker that can be used to classify and understand disease. Figure 7 shows allelic variant-associated methylation where a G/A germline variant appears to favour methylation/ hydroxymethylation respectively in the NDUFS2 gene, which has demonstrated links to cancer and age-related diseases[5]. Given the opposing function of the two methylated marks, the data provides insight into how gene expression may be controlled.

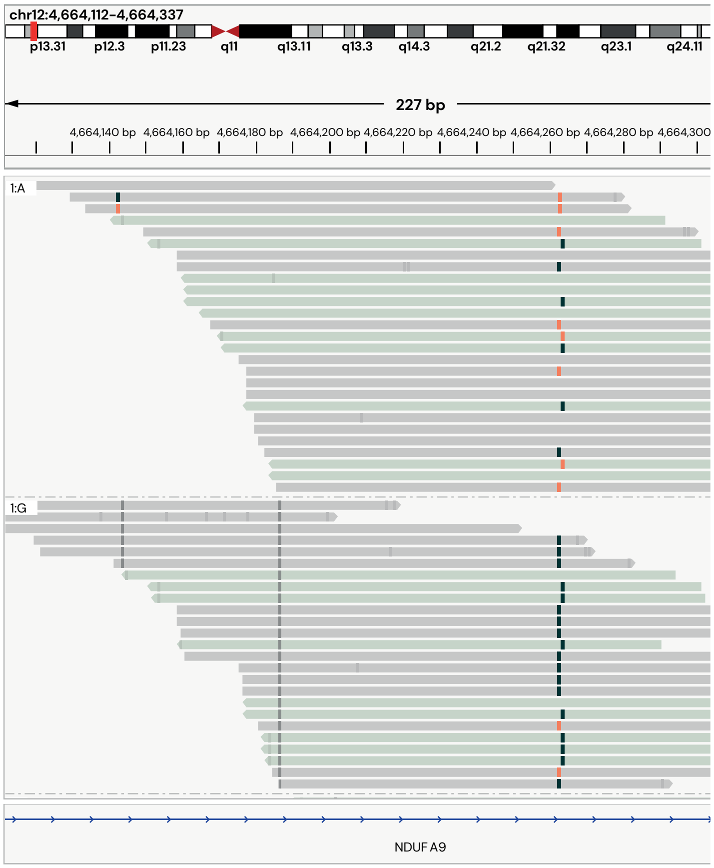
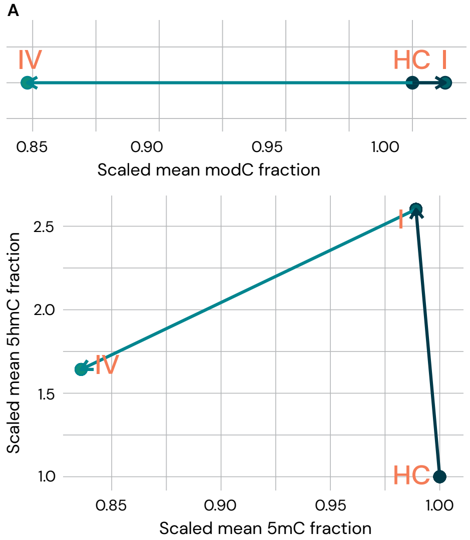


Figure 7: Variant-associated methylation identified in a cfDNA sample from a healthy donor. An allelic variant is identified, and on the same read shows methylation (black) is associated with the G allele and hydroxymethylation (red) with the A allele in the NDUFS2 gene.

Discover novel, inaccessible biomarkers

Enhancers are regulatory regions that interact with multiple distal genes and enhance expression when activated. An enhancer can be active or repressed but can also be primed, signifying that it is ready to transition from a repressed to an active state and initiate gene expression. duet evoC data can accurately determine whether an enhancer is repressed, active, or primed.

The duet evoC 6-base sequencing provides new data upon which to build novel biomarkers in cfDNA samples. Analysis of cfDNA from a CRC cohort identified enhancers with methylation signatures that varied between stage IV and healthy samples (Figure 8). Using the ratio of 5mC and 5hmC methylation data, clear signals for repressed enhancers in healthy samples, active enhancers in stage IV samples, and primed enhancers in stage I samples were observed. When methylation is presented in one dimension there is no ability to discriminate between healthy and early-stage cancer. Discrimination between healthy and stage I samples would not have been possible without obtaining individual signals for 5mC and 5hmC.



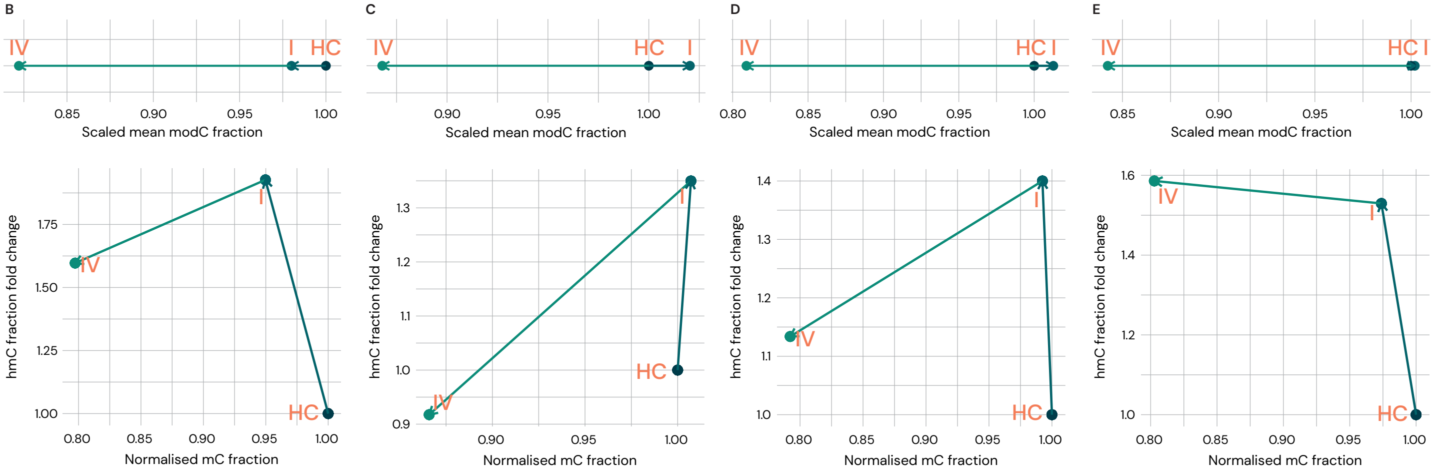


Figure 8: Differential enhancers identified between healthy control (HC) and stage IV CRC from single cfDNA samples (A). The addition of 5hmC provides two-dimensional information and allows discrimination of early-stage cancer from healthy. duet evoC identified novel biomarkers across multiple enhancers, all of which were discriminatory in cancer vs. healthy samples (B–E).

Conclusion

duet evoC 6-base sequence data provides new information for novel biomarker discovery in cfDNA samples. With duet evoC as the foundational assay of your liquid biopsy research, you can understand the interplay between genetics and epigenetics within a single DNA sample. With the 6-base genome, you can explore how methylation and hydroxymethylation differ between healthy and disease states and gain understanding of the mechanisms of epigenetic gene regulation that would not be possible with other methods. Discover more sensitive multimodal biomarkers for disease, build more accurate multimodal disease classifiers and detect biological changes in patient samples earlier.

duet evoC software and assay specifications

Input requirements	cfDNA 5-50ng gDNA 10-80ng
Assay turnaround time	Less than 18 hours
Assay hands on time	Less than 10 hours
Shelf-life	Minimum 3 months
Environment	On site HPC
Installation time	Commercial cloud: fast On site HPC: dependent on institution
User interface	Command Line Interface
Output files	QC reports, BAM, BedMethyl, VCF, Quant, ASM

Ordering information

Catalogue number	Product name	Product description
6101	duet multiomics solution evoC: 8x reaction	Pre-sequencing workflow + post sequencing software for 8 reactions
4101	UDI: 8x reaction	UDIs for 8 reactions
4104	UDI: 96x reaction	UDIs for 96 reactions

1. Wang, Pei, Qianqian Song, Jie Ren, Weilong Zhang, Yuting Wang, Lin Zhou, Dongmei Wang, *et al.* Simultaneous Analysis of Mutations and Methylations in Circulating Cell-Free DNA for Hepatocellular Carcinoma Detection. *Science Translational Medicine 14*, no. 672 (November 23, 2022): eabp8704.
2. Kim, Seung-Tae, Victoria M. Raymond, Joon Oh Park, Elena Zotenko, Young Suk Park, Matthew Schultz, Won Ki Kang, *et al.* Abstract 916: Combined Genomic and Epigenomic Assessment of Cell-Free Circulating Tumor DNA (ctDNA) Improves Assay Sensitivity in Early-Stage Colorectal Cancer (CRC). *Cancer Research 79*, no. 13_Supplement (1 July 2019): 916.
3. Mouliere, F. *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med. 10*, eaat4921 (2018).
4. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Comput. Biol. 12*, e1004873 (2016).
5. Rajarajacholan, U. K., Thalappilly, S. & Riabowol, K. The ING1a Tumor Suppressor Regulates Endocytosis to Induce Cellular Senescence Via the Rb-E2F Pathway. *PLOS Biol. 11*, e1001502 (2013).

Footnote

Publications demonstrate the enhancers detailed here regulate the genes MED8, ST3GAL3, CEP152, SHC4, DSG3, FCGR2C and RASSF9.

Disclaimer

The duet multiomics solution is for research use only.

biomodal
Chesterford Research Park
Cambridge, UK
CB10 1XL

+44 (0) 1223 800 700
biomodal.com
info@biomodal.com