Enrichment with duet multiomics solution +modC: investigate regions of interest with reduced sequencing requirement

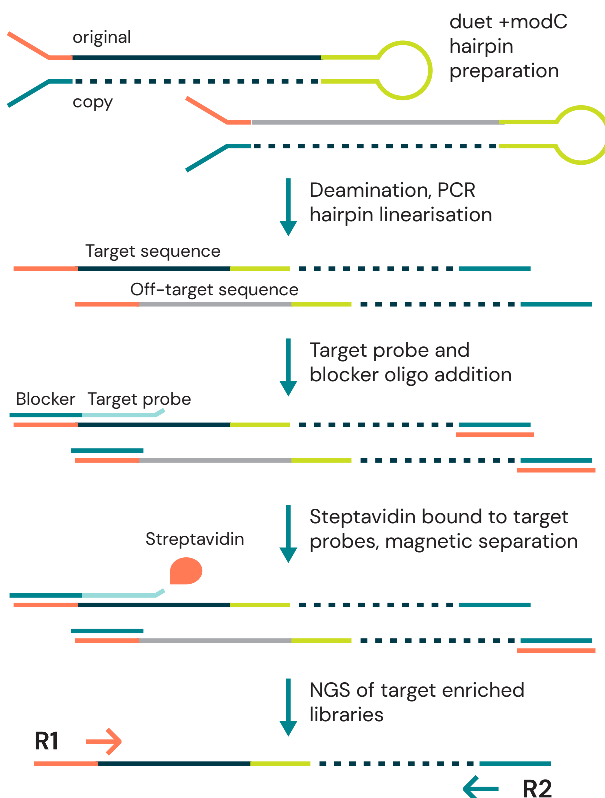
biomodal

Introduction

The combinatorial power of genetics and epigenetics is vital to understanding biology in healthy and disease states. duet multiomics solution +modC enables the identification of modified cytosine and the canonical bases A, C, T & G in a single-workflow enzymatic solution. Generating this information across the whole genome provides unparalleled understanding of genetic and epigenetic changes, but can require a large amount of sequencing, particularly when exploring at depth.

Enrichment protocols enable deep interrogation of target areas whilst maintaining cost effectiveness. Two enrichment methods performed with the duet multiomics solution +modC are presented here. Target enrichment, which involves using a set of probes to selectively isolate sequences of interest, and reduced representation which utilizes a specific enzymatic digestion of genomic DNA to select for GC rich regions.

duet +modC target enrichment workflow



duet +modC reduced representation workflow

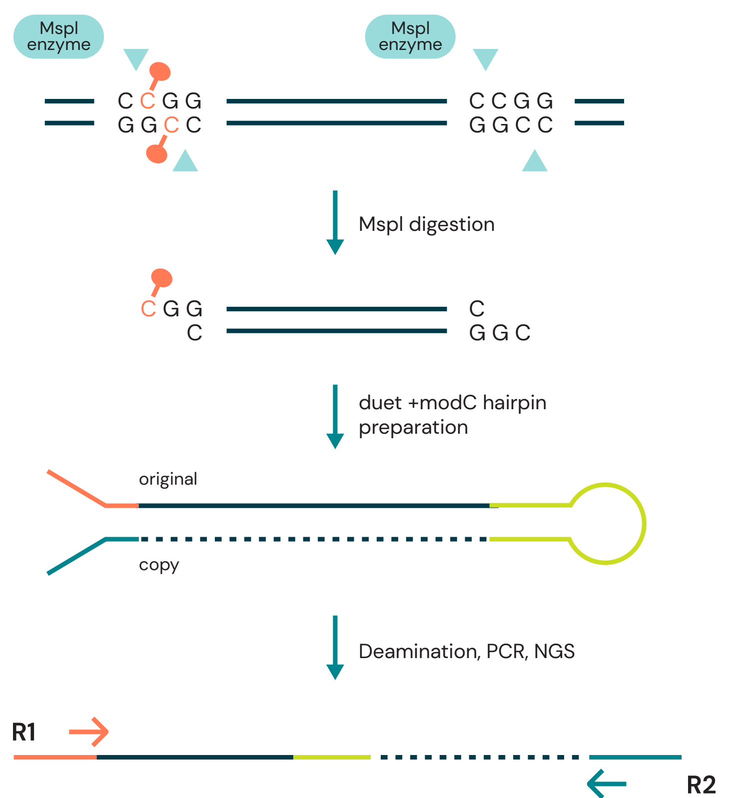


Figure 1: Workflow schematics for duet +modC target enrichment workflow (left) and duet +modC reduced representation workflow (right).

Target enrichment

Target enrichment panels enable the selective isolation of sequences of interest through target hybridization. Using a commercially available methylome panel we show that duet multiomics solution +modC achieves uniform high performance across a range of input amounts with reduced sequencing requirements to obtain high coverage of target regions. Figure 2 shows performance metrics on genomic DNA (gDNA) and circulating tumour DNA (ctDNA) sample types.

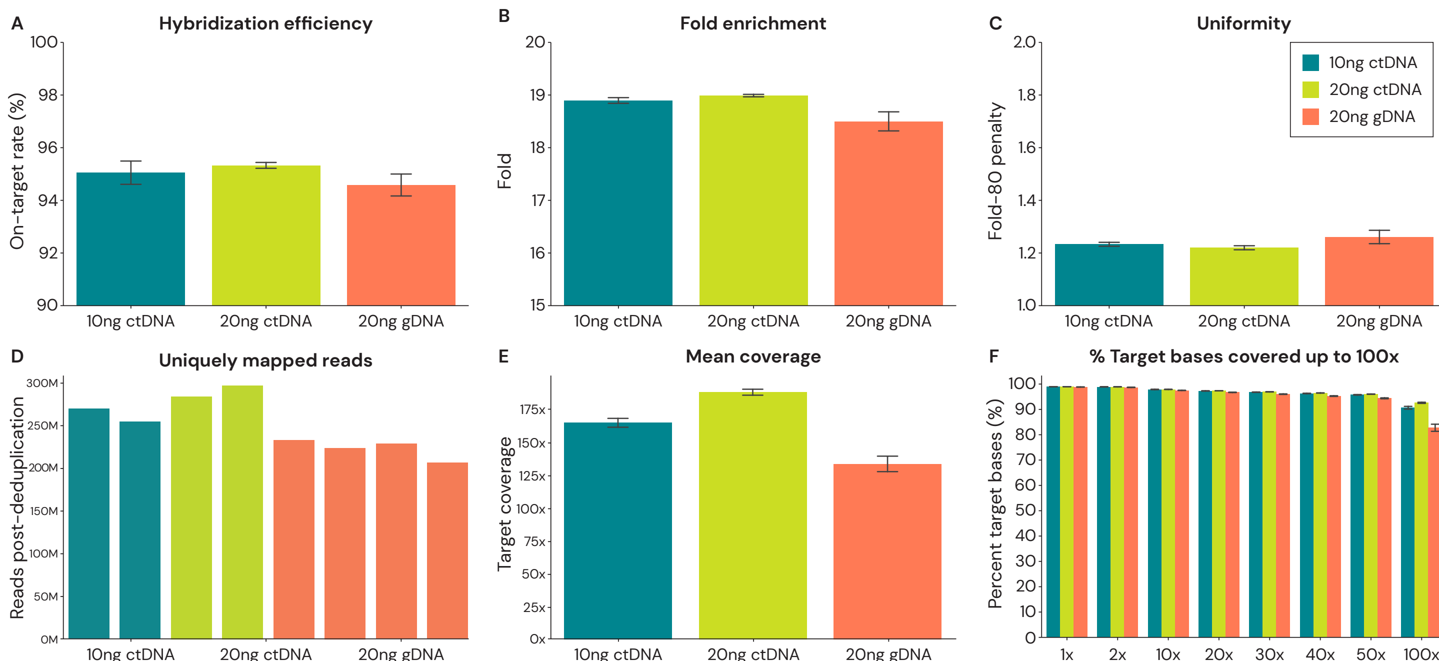


Figure 2: Target enrichment was performed using the Twist Methylome panel with a range of input types. **(A)** On-target rate (%), **(B)** Fold enrichment, **(C)** Fold-80 base penalty, **(D)** Uniquely mapped reads post de-duplication, **(E)** Mean target coverage, **(F)** %Target bases covered up to 100X. Data was obtained using Genome In A Bottle (GIAB) genomic DNA (gDNA) samples at 20ng input and synthetic circulating tumour DNA (ctDNA) at 10ng and 20ng input (Seracare). Targeting was performed using a Methylome Panel (Twist).

	Reads required to achieve given coverage	
	30X	100X
Target enrichment	47.5M	158 M
Whole genome	733 M	2.4 B

Table 1: Number of sequencing reads required to achieve 30X or 100X coverage using a Methylome panel (Twist) vs. whole genome.

Target enrichment can facilitate a reduction in the amount of sequencing required to maintain high coverage across target regions. Table 1 describes the sequencing input required to achieve either 30X or 100X coverage using duet multiomics solution +modC in a targeted or whole genome workflow. Target enrichment using the duet multiomics solution +modC requires significantly fewer reads to achieve an equivalent coverage.

Targeted methylation sequencing in Colorectal Cancer cell free DNA

Targeted sequencing was performed using cell free DNA (cfDNA) from patient samples with either Stage I or IV Colorectal Cancer and reportedly healthy controls using the Twist Pan-Cancer panel. Figure 3A shows the relative levels of methylation across target regions for each sample type, clearly illustrating widespread hypermethylation associated with late-stage cancer.

Figure 3B shows a CpG within the BID gene, a known prognostic marker for colorectal cancer, that was identified as hypermethylated in the Stage IV cancer cfDNA samples, but with significantly lower methylation levels in Stage I and healthy samples. DNA methylation at this site reduces the expression of the tumour-suppressor BID, which results in reduced apoptosis.

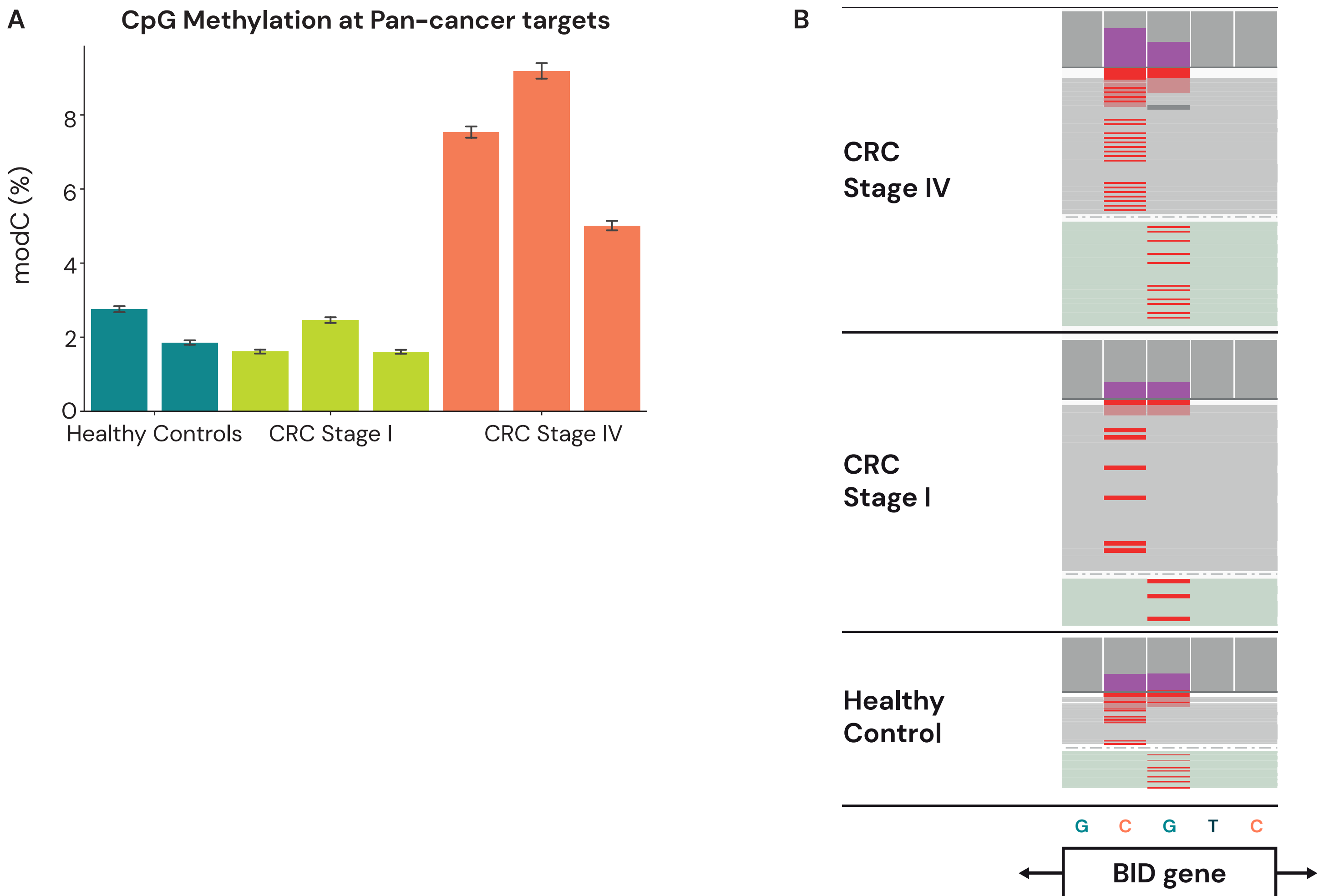


Figure 3: (A) methylation fraction was measured in Healthy and Colorectal cancer Stage I & IV cfDNA samples using duet multiomics solution +modC performed with a Pan-Cancer panel (Twist). **(B)** Image from Integrated Genome Viewer showing methylation level at the BID genomic site in Healthy, Stage I and Stage IV CRC samples .

Reduced representation

Reduced representation uses enzymatic digestion of genomic DNA to selectively enrich for DNA fragments containing CpGs. The resulting library contains a higher proportion of potential DNA methylation sites. Subsequent next generation sequencing therefore achieves higher coverage of CpGs with a reduced sequencing read requirement. Table 2 shows that duet multiomics solution +modC used with this workflow results in a greater number of CpGs covered for a given number of sequencing reads when compared to whole genome.

Sample input	Input reads	duet +modC	RR duet +modC
		CpGs covered >10X	
80ng gDNA	15 million	2,863	151,329
	10 million	2,194	33,353
	1 million	1,392	2,518

Table 2: CpGs covered at higher than 10X by whole genome and reduced representation at 5, 10, and 15 million input sequencing reads.

Reduced representation methods achieve higher coverage of CpGs at lower input reads than standard whole genome methods whilst maintaining high specificity and sensitivity for modC, (Figure 5A-C).

The %GC coverage bias graph in Figure 5D shows a skew towards high GC content when using the reduced representation workflow, demonstrating that the libraries have been enriched for CpGs. The workflow achieves high correlation between replicates, as shown in Figure 5E.

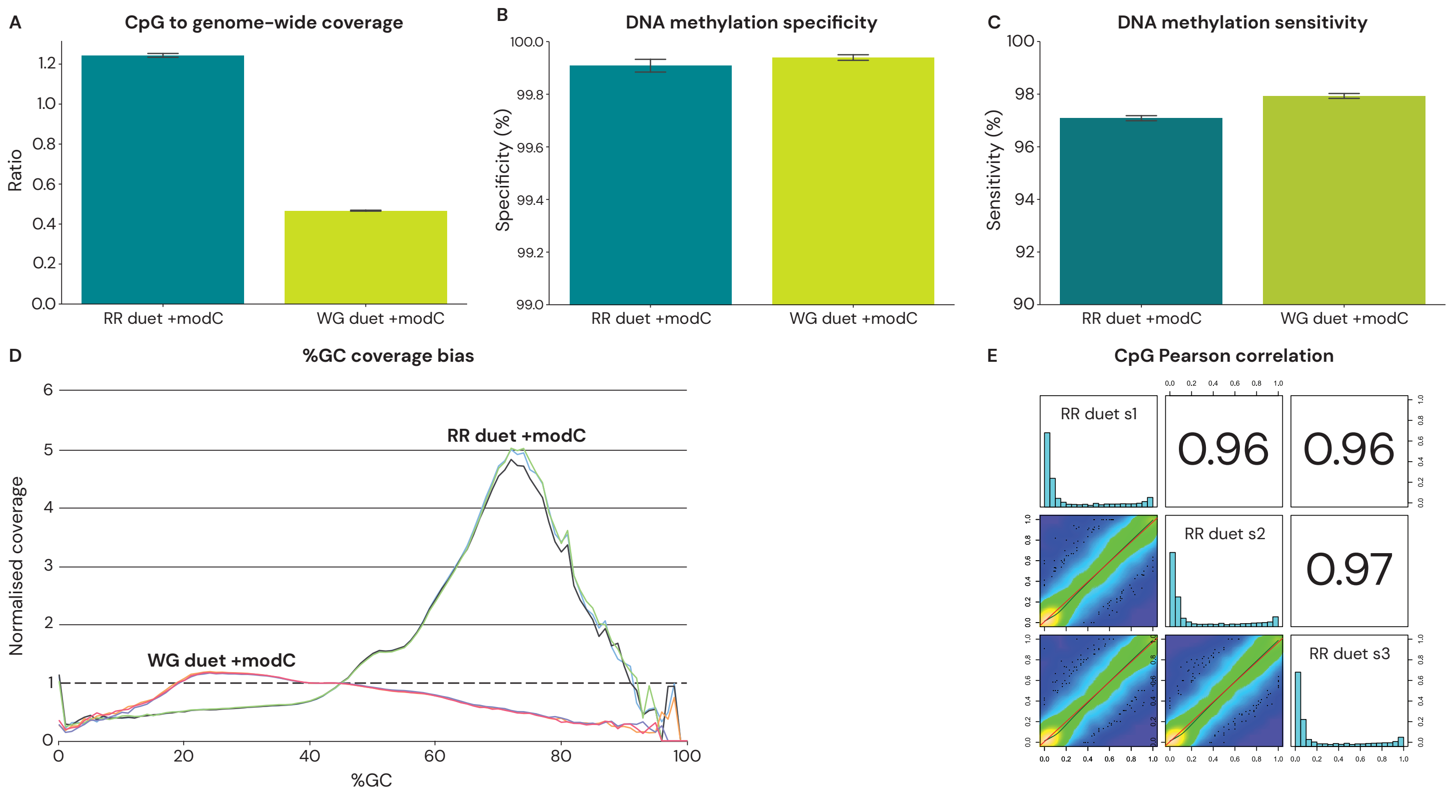


Figure 4: reduced representation (RR) enriches for CpGs in sample libraries whilst maintaining high performance. **(A)** Ratio of CpG coverage to genome-wide coverage for RR vs whole genome (WG), DNA methylation specificity **(B)** and sensitivity **(C)** for RR vs whole genome, **(D)** coverage vs %GC content, **(E)** Pearson correlation of CpGs between RR technical replicates. Data was prepared using 80ng gDNA from GIAB samples.

Conclusion

The compatibility of duet multiomics solution +modC with enrichment techniques is demonstrated here. Both targeting with enrichment panels and reduced representation with MspI enzymatic digestion successfully enriched libraries for their respective regions of interest, increasing coverage of target loci or CpGs respectively, whilst reducing the sequencing requirement and thus improving cost effectiveness. The superior performance of duet multiomics solution +modC for investigating both genetic and epigenetic information is maintained throughout these protocols.

These enrichment techniques can be used to achieve a deeper understanding of target areas in both genomic and cell free DNA, further enhancing the utility of duet multiomics solution +modC in applications such as liquid biopsy. Refining genetic and epigenetic investigations to a narrower region of interest can improve detection of low prevalence disease-associated biomarkers at significant sequencing depth, whilst maintaining the benefits of high accuracy and minimum sample input conveyed by biomodal's duet technology.

Methods

Targeted sequencing was performed using duet multiomics solution +modC (biomodal) with Human Methylome panel (Twist) and Pan-cancer panel (Twist).

Human methylome target enrichment was performed following an adjusted Twist methylation sequencing workflow on duet multiomics solution +modC constructed libraries from inputs of 10ng and 20ng ctDNA and 20ng gDNA (GIAB NA12878), all in duplicates.

Targeted methylation sequencing of plasma cfDNA for detection of colorectal cancer was performed using a Pan-cancer capture panel run on duet multiomics solution +modC libraries from 5ng input cfDNA isolated from patients with colorectal cancer (3x stage I and 3x stage IV) and healthy controls. Target enriched libraries in both cases were sequenced on Illumina NovaSeq 6000.

Reduced representation duet +modC sequencing was performed using overnight MspI digestion (Diagenode protocol) on unsheared 80ng input NA12878 gDNA in triplicate. The resulting digested fragments were then used to construct libraries. Whole genome libraries were constructed on sonicated (250bp) fragments using 80ng input NA12878 in triplicate. All libraries were generated using duet multiomics solution +modC kits. In both cases, a mix of methylated lambda and unmethylated puc19 were used as internal controls. All samples were sequenced on Illumina NovaSeq 6000.

Disclaimer

The duet multiomics solution is for research use only.

biomodal
Chesterford Research Park
Cambridge, UK
CB10 1XL

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

biomodal

© 2023 biomodal. All rights reserved.