

1. Introduction

DNA methylation plays a central role in gene regulation, but conventional epigenetic sequencing approaches either combine suppressive 5-methylcytosine (5mC) and activating 5-hydroxymethylcytosine (5hmC) into a single modified-cytosine signal or omit 5hmC entirely. This can obscure biologically meaningful changes because 5mC and 5hmC have distinct associations with regulatory activity and may change independently or in opposing directions relative to the healthy/control condition.

duet 6-base sequencing overcomes these limitations by independently resolving genetic sequence, 5mC and 5hmC from the same DNA molecules in a single assay. **6-base sequencing** enables more complete detection and functional interpretation of differential methylation, whilst expanding the feature space available to study health and disease.

Figure 1: The duet 6-base workflow¹ ligates a hairpin to DNA before a copy step preserves genetic information. 5mC and not 5hmC are copied onto the copy strand before protection and deamination steps. After conventional short read sequencing duet software resolves original and copied strands to produce 6-base data including A, T, G, C, suppressive 5mC and activating 5hmC.

2. 6-base sequencing resolves biologically distinct signals

5mC and 5hmC play complementary roles in gene regulation. Promoter-associated 5mC is generally linked to repression, whereas 5hmC is enriched at active enhancers and transcribed genes. Resolving both modifications independently therefore provides additional context for interpreting regulatory activity

Figure 2: Lung and liver tissue samples were prepared with **duet** and 6-base sequenced to >30x depth. Differential methylation analysis using **modality XPLR** software detected significant DMRs (q<0.05 dashed coral line), revealing widespread repressive promoter 5mC hypomethylation and activating enhancer 5hmC hypermethylation in liver tissue compared to lung tissue.

3. 6-base differential methylation analysis reveals hidden changes in regulatory regions

Figure 3A: Resolving 5mC and 5hmC independently identified 97% more promoter DMRs than modC analysis and revealed whether differential signals were associated with 5mC loss, 5hmC gain or changes in both modifications.

Figure 3B: At enhancer regions, restricting analysis to 5mC-only missed approximately 33% of the DMRs detected through independent 5mC and 5hmC analysis, demonstrating the additional regulatory information contributed by 5hmC

4. 6-base sequencing reveals hidden biology

Example liver biomarkers include crucial proteins involved in the removal of free plasma haemoglobin (HP - Haptoglobin) and triglyceride metabolism (APOA5 - Apolipoprotein A5). 6-base analysis reveals significant repressive and activating regulatory signal that would be hidden in modC or 5mC-only analyses.

Figure 4A: At the *HP* promoter, liver tissue showed a significant loss of 5mC (coral) and opposing significant gain of 5hmC (green), consistent with activation. The corresponding modC signal is not significant as a consequence of summing 5mC and 5hmC, therefore masking a key regulatory change and contributing to the modC deficit observed in Figure 3A.

Figure 4B: At the *GH11F16733* enhancer, significant differential methylation gain (green) is observed for 5hmC, consistent with activation. The modC and 5mC signals are not significant, indicating the need for 6-base analysis to discover epigenetic changes in key enhancer regulatory regions, and contributing to the 5mC-only deficit observed in Figure 3B.

5. Epigenetic accuracy drives biomarker discovery

A high-confidence reference set of 9,368 promoter DMRs was identified from original methylation counts. Using the same underlying data, ten datasets were simulated for each of three 5mC accuracy profiles (Table 1). 5mC counts were retained according to sensitivity, while C and 5hmC counts were misclassified as 5mC according to 1-specificity. Promoter DMRs identified by modality XPLR (q < 0.05) were then compared with the reference set.

Table 1: 5mC accuracy profiles

Simulated accuracy profile	5mC sensitivity	5mC specificity
duet 6-base mosaic	98.7%	99.85%
Comparator 1	94.7%	99.1%
Comparator 2	90.7%	96.5%

Figure 5A: Higher 5mC accuracy recovers more reference DMRs, increasing biomarker discovery. Lower-accuracy profiles missed biologically relevant liver biomarkers including *KLF15* and *MTTP*.

Figure 5B: Reduced 5mC accuracy increased false DMR discovery. Simulated datasets with lower sensitivity and specificity produced up to 2x more false-positive DMRs (y-axis) than duet 6-base mosaic.

6. Conclusions

- Independent analysis of 5mC and 5hmC reveals regulatory changes obscured by modC and 5mC-only approaches.
- 6-base analysis detected 97% more lung vs liver promoter DMRs than modC analysis, while 5mC-only analysis missed ~33% of DMRs driven by 5hmC changes.
- Higher CpG-level accuracy improves DMR recall and precision, supporting more complete and reliable biomarker discovery.

Together, these findings show that accurate, independent measurement of 5mC and 5hmC provides a more complete and biologically informative view of differential methylation.

7. References

1. Fullgrabe, et al. Simultaneous sequencing of genetic and epigenetic bases in DNA. Nat Biotechnol. 2023 Oct;41(10):1437-1464.

bioModal