The 6-base genome reveals novel 5mC and 5hmC biomarkers enabling the detection of neurological disease from sub-nanogram quantities of CSF cfDNA

Mark Consugar[1], Annelie Joahnsson[1], Aurelie Modat[1], Rebecca Elwood[1], Tom Charlesworth[1], Angela Simeone[1], Mike Stubbington[1], Robert J Osborne[1], Maria Starvaggi Cucuzza[1,2,3], Maria Needhamsen[1,2], Maja Jagodic[1,2,3]

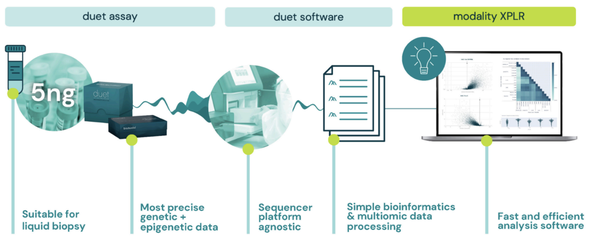
1. biomodal Ltd, The Trinity Building, Chesterford Research Park, Cambridge, UK
2. Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
3. Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden
4. Center for Neurology, Academic Specialist Center, Stockholm, Sweden

1. Introduction

5-methylcytosine (5mC) and 5-hydroxyethylcytosine (5hmC) are epigenetic modifications of cytosine, they have distinct tissue-specific distributions and can act as tissue fingerprints. While DNA methylation analyses have historically focused on 5mC, it is now recognised that 5hmC is a biologically distinct epigenetic mark associated with gene activation, in contrast to the generally repressive role of 5mC. 5hmC is particularly abundant in neuronal tissues, making it a valuable marker of neuronal identity and gene regulatory activity.

5hmC levels, particularly when combined with 5mC measurements, are a predictor of gene expression, even from cfDNA. Simultaneous measurement of 5mC and 5hmC therefore enables deeper understanding of tissue identity, gene regulation and disease-associated biology.

The duet technology uses a hairpin approach to provide simultaneous, high accuracy genetics and epigenetics, including distinguishing 5hmC to 5mC. from a single low input sample.



2. 5hmC is enriched in neuronal tissues

5hmC levels vary substantially between tissues, reflecting tissue-specific biology. Analysis of human tissues shows that brain tissue contains markedly elevated levels of 5hmC (up to ~20%) relative to leukocyte, liver, lung and prostate tissue (<10%) (**Figure 1**).

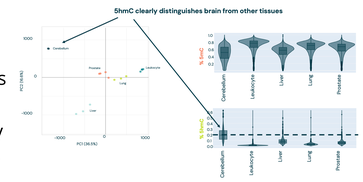


Figure 1. 5hmC is enriched in neuronal tissues

These observations extend beyond tissue-level differences to cell type. Within the brain, 5mC and 5hmC patterns vary between neurons, microglia and oligodendrocytes, revealing cell-type-specific epigenetic that are obscured in conventional 5modC measurement, where the signal is conflated (**Figure 2**). Notably neurons contain ~40% 5hmC where other cell types contain ~0-20% (5).

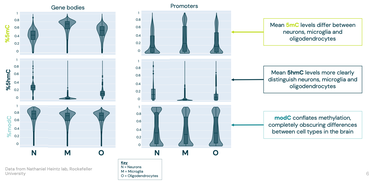


Figure 2. 5hmC reveals cell-type-specific biology within the brain.

Together, these findings demonstrate that 5hmC is both a biomarker of neuronal identity and a source of biologically meaningful information that can be leveraged to study neurological biology using both bulk tissue and sorted cells.

3. 6-base data reflects gene activity and enables inference of gene expression

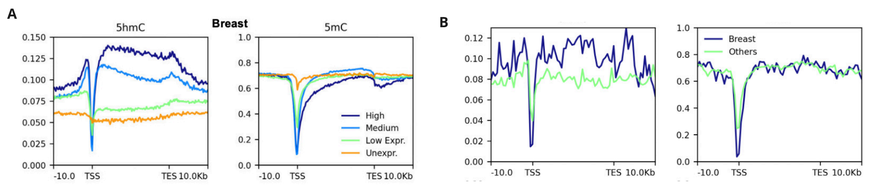


Figure 3. (A) The relationship between 5hmC and 5mC and gene expression in breast tissue using matched 6-base and RNA-seq data from the same sample. High, medium, low, and unexpressed genes using thresholds were grouped and the 5mC and 5hmC level plotted across scaled regions centred on the gene body. 5hmC is most abundant in the gene bodies of highly expressed genes and high gene expression is associated with low promoter 5mC and 5hmC. **(B)** Comparison of 5mC and 5hmC levels specifically expressed in breast tissue compared to those specifically expressed in other tissues.

Unlike conventional methylation approaches that measure a conflated modC signal, the duet technology independently resolves 5mC and 5hmC, preserving the distinct biological information carried by each modification. Highly expressed genes exhibit elevated gene-body 5hmC and reduced 5mC (**Figure 3A**). The separation between different levels of expression is far more distinct for 5hmC than for 5mC, even with a lower dynamic range. Together these observations suggest that the ability to distinguish 5mC from 5hmC could provide significant value in predicting functional genomic readouts like gene expression over traditional modC methylation sequencing data. These opposing patterns would be obscured if the two modifications were combined as modC.

Furthermore, tissue-specific genes more clearly show elevated gene-body 5hmC relative to genes expressed in other tissues than 5mC (**Figure 3B**), supporting 5hmC as a powerful biomarker of tissue-specific gene activity. These observations are consistent with recent work that demonstrated combined 5mC and 5hmC measurements can be used to infer gene expression in tissue from cfDNA (1).

4. duet mosaic is optimised for 6-base analysis of cfDNA

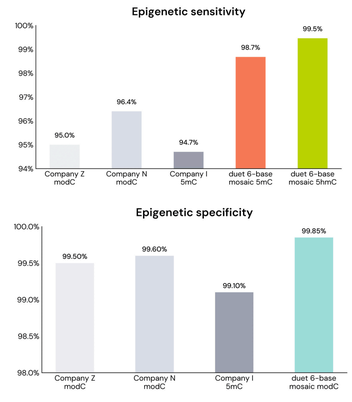


Figure 4. duet mosaic methylation sensitivity (A) and specificity (B) compared to other epigenetic sequencing technologies. For duet mosaic, sensitivity and specificity are assessed using methylated pUC19 control and unmethylated lambda controls that are included in the every kit. 5hmC sensitivity is calculated using the included short oligo controls. For other sequencing technologies publicly available data was used (3). duet mosaic has a 6-9 fold lower number of false positives than company I whilst achieving higher sensitivity to recover the real methylation signal.

cfDNA is a challenging sample type due to its low abundance and fragmented nature. duet mosaic uses a single-stranded ligation workflow optimised for cfDNA, improving recovery of short and damaged fragments while yielding high accuracy 6-base information. These improvements lead to market leading sensitivity and specificity (**Figure 4).** The high accuracy and efficient capture of fragments including single stranded DNA lead to ultra low LoD for detection of genetic and epigenetic biomarkers (**Table 1**). Combined with the ability to detect 5hmC, which is elevated in the brain, duet mosaic provides greater power to detect neurological disease from cfDNA, opening new opportunities liquid biopsy applications.

	LoD95
Genetic	0.561%
Epigenetic (hypermethylation)	6.65 ppm
Epigenetic (hypomethylation)	12.34 ppm

Table 1: Methylation (5mC) and single variant genetic LoD95 derived from contrived spike in samples of known VAF or methylated/ unmethylated concentrations

5. 6-base sequencing reveals underlying biology from CSF cfDNA

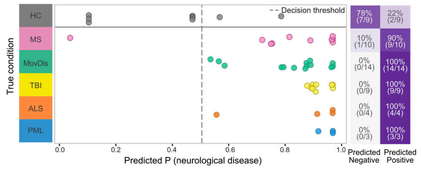


Figure 5. (A) 5mC and 5hmC biomarkers accurately distinguish neurological diseases from Healthy Control, even from picogram levels of CSF cfDNA

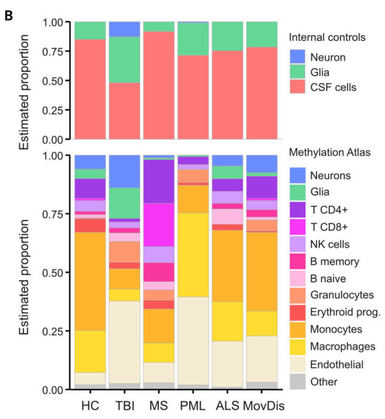


Figure 5. (B) Cell-type deconvolution of 6-base CSF cfDNA data reveals disease-specific biology including elevated contributions from CD8+ cells in MS and endothelial cells in TBI

Through a collaboration with the Jagodic group at the Karolinska Institutet (2), the duet technology was used to generate 6-base data from ultra low input (as little at 100pg) cerebrospinal fluid (CSF) cfDNA from individuals with traumatic brain injury (TBI), multiple sclerosis (MS), progressive multifocal leukoencephalopathy (PML), amyotrophic lateral sclerosis (ALS), movement disorders (MovDis) and healthy controls.

Combined 5mC and 5hmC biomarkers enabled accurate discrimination of neurological disease from healthy controls (**Figure 5A**).

Beyond disease detection, analysis of the 6-base CSF cfDNA profiles provided biologically meaningful insight into the cellular processes underlying each condition. Cell-type deconvolution revealed increased Central Nervous System (CNS) contributions in TBI, including neuronal and glial cells whilst MS samples showed enrichment for immune-cell signatures, particularly CD8+ T cells (**Figure 5B**). Pathway analysis of differential 5mC and 5hmC confirmed synaptic functions in MS and revealed enrichment of sensory-neuron-related pathways in TBI, which may reflect disease-specific consequences of trauma dynamics.

These findings demonstrate the potential of 6-base data to detect neurological disease in a liquid biopsy setting. 6-base data provided both sensitive disease detection and insight into the underlying cellular processes driving disease pathology (2).

6. Conclusions

DNA methylation is more than just 5mC, 5hmC is a biologically distinct biomarker that marks activating enhancers and actively transcribed genes, in opposition to the repressive role of 5mC. 5hmC is also a powerful marker of tissue and cell identity and is markedly elevated in neuronal tissues. By independently resolving 5mC and 5hmC, 6-base sequencing enables functional genomic insight beyond conventional methylation analysis and has particular utility in neurological applications. The duet technology has proven utility in liquid biopsy applications (1,2,4) and the recently launched duet mosaic products further improve these capabilities. In neuronal liquid biopsy recent data demonstrates the ability of 6-base biomarkers to not only detect neurological disease from CSF cfDNA but also extract functional biological insight into disease pathology. Taken together the duet platform is a powerful tool in neurological research and opens up liquid biopsy applications in neurology.

7. References

(1) Sowalsky AG, et al. Integrative dual ctDNA 5mC/5hmC methylomics and clonal reconstruction infer tumor transcription and resistance phenotypes in metastatic prostate cancer. Cell Reports Medicine. 2025.
(2) Starvaggi Cucuzza C et al., Advancing neurological liquid biopsy via six-base sequencing of cerebrospinal fluid cell-free DNA. bioRxiv. 2026. doi:10.64898/2026.09.06.749692.
(3) https://www.illumina.com/science/genomics-research/articles/5-base-solution.html
(4) Puddu F, et al. 5-methylcytosine and 5-hydroxymethylcytosine are synergistic biomarkers for early detection of colorectal cancer. Commun Med. 6:15 (2026).
(5) Data kindly shared by Nathaniel Heintz laboratory, The Rockefeller University.