

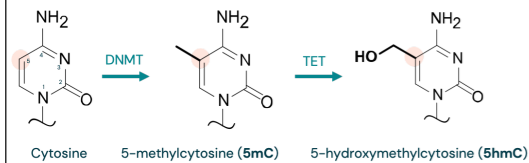
# Simultaneous single cell sequencing of genetic and epigenetic bases

Aldo Cia-Uitz<sup>1</sup>, Chenfu Shi<sup>1</sup>, Iris Valent<sup>1</sup>, Jack M Monahan<sup>1</sup>, Jinfeng Chen<sup>1</sup>, Julia Vivian<sup>1</sup>, Minna Taipale<sup>1</sup>, Páidí Creed<sup>1</sup>, Paula Golder<sup>1</sup>, Paula Kokko-Gonzales<sup>1</sup>, Rafael Tavares<sup>2</sup>, Shankar Balasubramanian<sup>2</sup>, Walraj S. Gosal<sup>1</sup>

bioModal

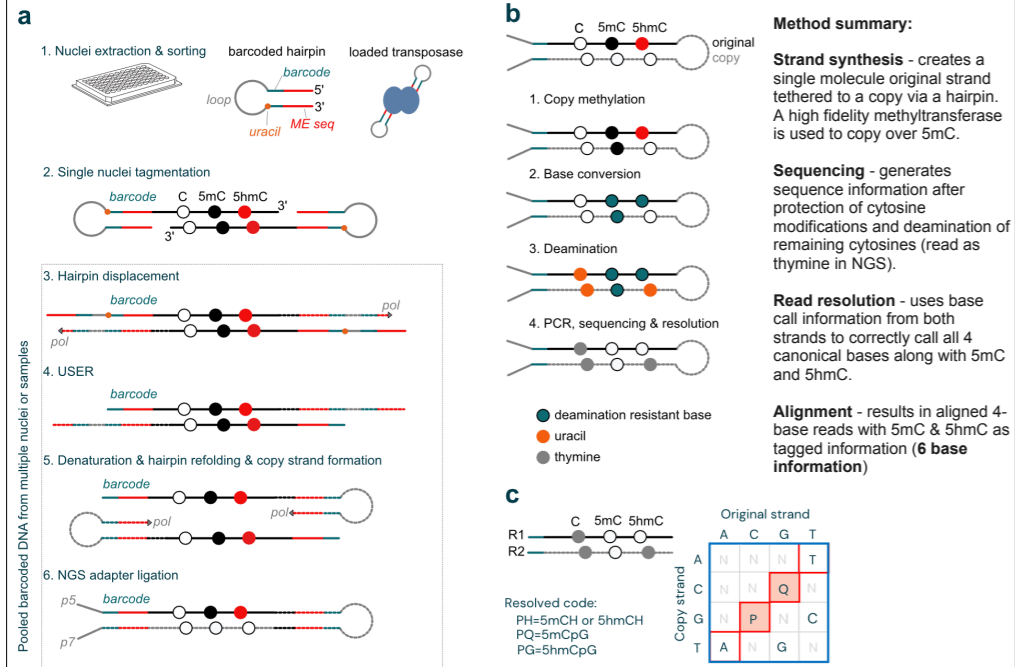
## 1. Introduction

A body of evidence indicates that epigenetic modifications in DNA comprises a fundamental pathway by which genes can be silenced or activated, determining cell fate and function. In mammalian genomes, the fifth carbon of cytosine is one major target for epigenetic modification (Figure 1). Currently, reading the two most prominent modifications 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), simultaneously at high resolution remains largely elusive at the single cell level, making the function and relationship between these modifications difficult to precisely ascertain. Here we present a single nuclei workflow, enabled by barcoding with Tn5, to simultaneously determine 5mC and 5hmC at the single molecule level (Figure 2).



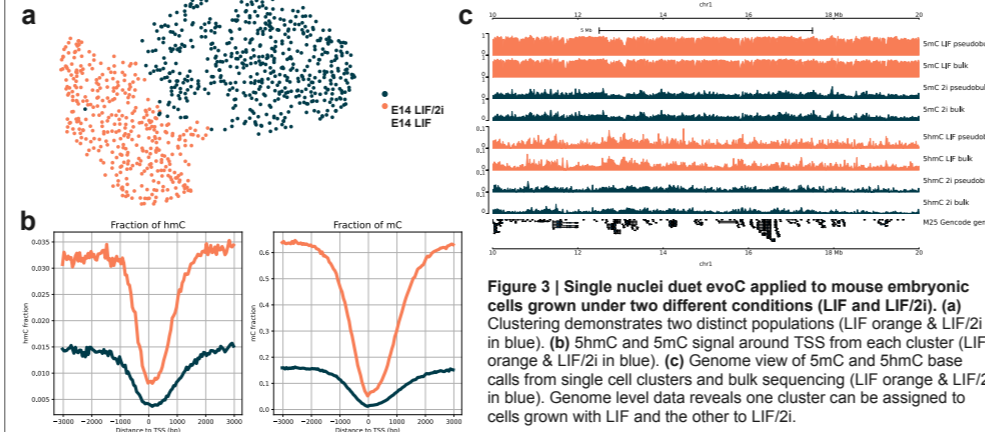
**Figure 1 | Cytosine modifications at the fifth carbon is a major target of epigenetic modification in mammalian genomes.** 5-methylcytosine (5mC) is associated with gene silencing and patterns of this modification are altered in diseases such as cancer. In contrast, 5-hydroxymethylcytosine (5hmC) has only recently been suggested to play a role in gene regulation and is generally considered a marker of active tissue-specific genes and enhancers.

## 2. Single nuclei duet evoC: reading genetics & epigenetics simultaneously



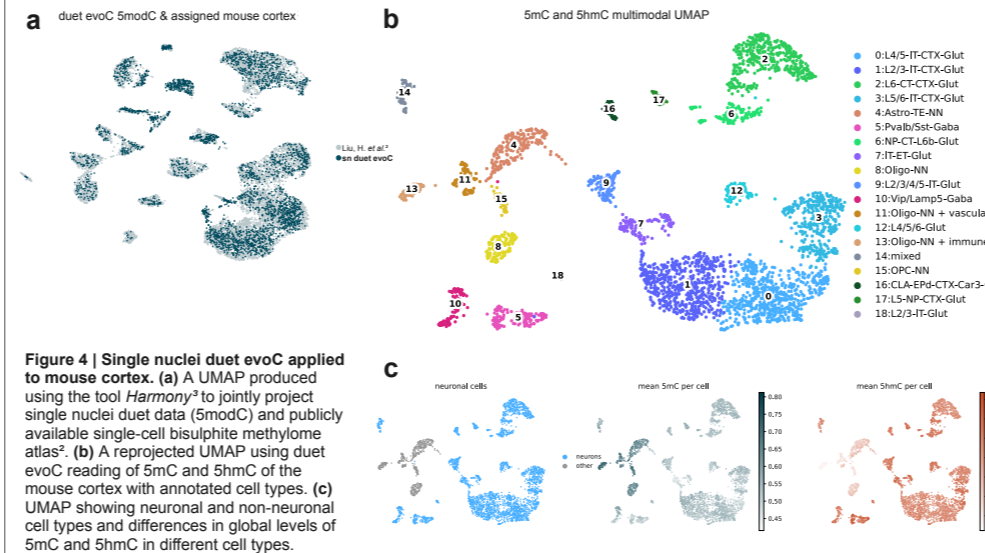
**Figure 2 | Single nuclei duet evoC is enabled by a Tn5 barcoding workflow and allows simultaneous calling of 5mC and 5hmC.** (a) Nuclei sorted into wells are tagmented after nucleosomal disruption using Tn5 loaded with a hairpin containing a cell-specific barcode, and DNA from multiple nuclei is pooled. A duet library is constructed using steps 3-6. (b) 5mC and 5hmC can be determined using enzymatic conversion. Barcoded duet libraries undergo copy methylation only at 5mC (using DNMT5), base conversion (using TET) and deamination (using the ssDNA deaminase A3A and the UvrD helicase). (c) Reads from paired-end NGS are resolved using a lookup table into a single 4-base read with annotated 5mC and 5hmC base calls.

## 3. 5mC & 5hmC can be read at a single cell level



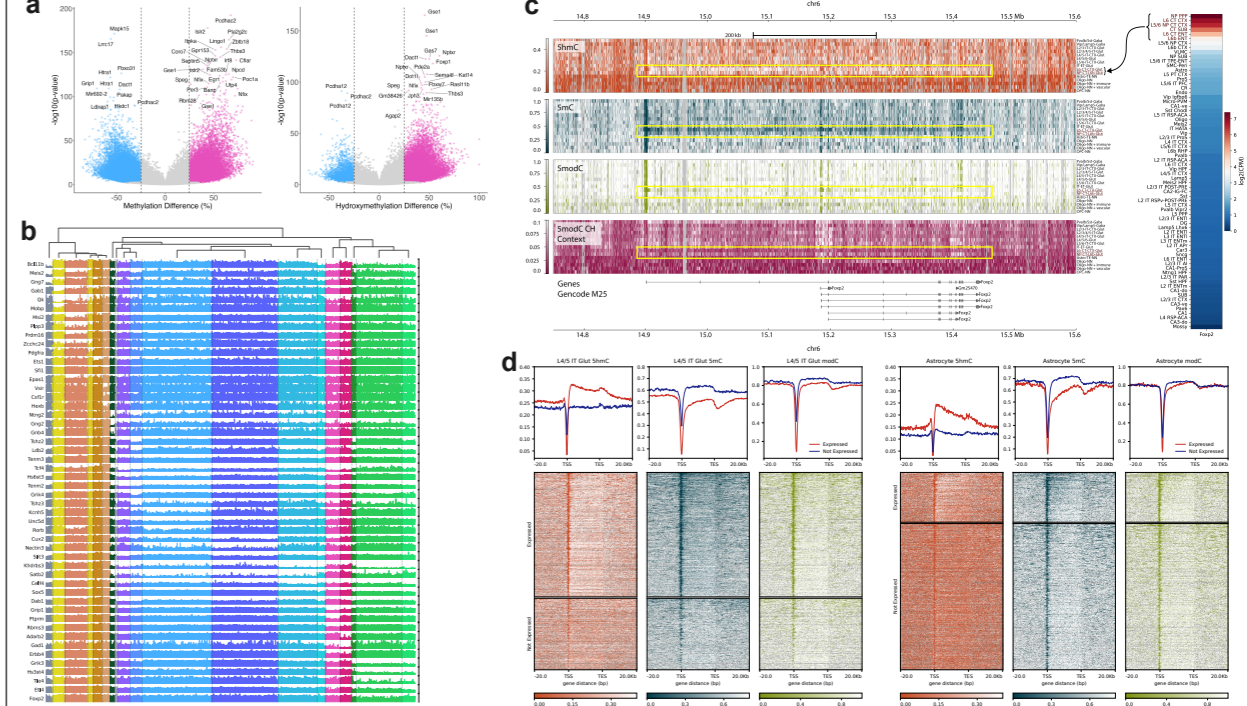
To demonstrate that duet evoC is able to read single cell cytosine modifications, we applied the method to mouse embryonic stem cells, E14, grown under two different conditions (Figure 3). These conditions affect cytosine modifications. We show the method can produce a multimodal UMAP based on 5hmC and 5mC, with two clusters as expected. The profile for 5mC and 5hmC from each pseudo-cluster matches bulk sequencing of cells grown under LIF and LIF/2i conditions.

## 4. 5mC & 5hmC can be simultaneously determined in complex tissue



We applied single nuclei duet evoC to cells from mouse cortex. We first collapsed the 5mC and 5hmC into a joint signal (5mC) and integrated our data with a comprehensively annotated bisulphite atlas<sup>2</sup> using Harmony<sup>3</sup> (Figure 4a) The cell type assignments were then transferred to duet 5mC and 5hmC data revealing 19 distinct clusters of cells, and the resulting UMAP is presented in Figure 4b. This represents the most complete map of both 5mC and 5hmC at single cell level of the mouse cortex to date. As a first validation that this UMAP represents known cell-type specific epigenetic information, we show that neuronal cells in this UMAP are particularly enriched in 5hmC as is expected from the literature (Figure 4c).

## 5. Low 5mC & high 5hmC is a feature of cell-specific gene expression



Using our assigned UMAP containing the complete predominate cytosine modification profiles for each cell and cell type, we first show that a number of genes are differentially methylated or hydroxymethylated (Figure 5a), and certain genes have a pattern of hypomethylation across gene bodies that are cell-type specific (Figure 5b). To examine this in more detail, we highlight one gene from this dataset, *Foxp2*, which is highly expressed in certain glutamatergic neurons such as L6-CT-CTX-Glut and NP-CT-L6b-Glut (Figure 5c)<sup>4</sup>. We show that for these cell types, there is a pattern of low 5mC (hypomethylation) and high levels of 5hmC (hyperhydroxymethylation) across the gene body for these two cell types in our dataset. However, this signal is attenuated when observed without discrimination for which cytosine modification is present (e.g. 5mC), due to the opposing effect of 5mC and 5hmC. 5mC in CH context is also less clear. To see if this is a global phenomenon in the mouse cortex, we examine patterns across expressed and non-expressed genes. Here, we find that indeed the pattern of reduced 5mC and increased levels of 5hmC are present at expressed genes for different cell types (both neuronal and non-neuronal cell types are shown above). The pattern of 5mC is again attenuated in comparison.

## 6. Conclusions

Here we present duet multimomics solution evoC - an enzymatic method that reads the four canonical bases in DNA together with complete epigenetic information encoded in DNA - as applied to single cells. Applying this method to nuclei isolated from mouse cortex we generate a UMAP for mouse cortex and looked at genome wide methylome and hydroxymethylome patterns across the genome for different cell types at a resolution not previously achieved, even by recent attempts<sup>5</sup>. We observe that cell-type specific genes are marked by low methylation and higher hydroxymethylation across the gene body. Since current methods using bisulphite often do not make the distinction between these two cytosine modifications, dynamic changes in cell-specific genes are radically reduced by a combined 5mC signal. This demonstrates the power of reading all six bases as a new lens to examine the dynamic information encoded in DNA.

## 7. References

1. Simultaneous sequencing of genetic and epigenetic bases in DNA, Fullgrave and Gosal et al., Nature Biotechnology (2023) (duet multimomics solution technology paper).
2. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain, Liu, H. et al., Nature (2023).
3. Fast, sensitive, and accurate integration of single cell data with Harmony, Korsunsky, I. et al., Nature Methods (2019).
4. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation, Zichen, Y. et al., Cell (2021).
5. Simultaneous single-cell analysis of 5mC and 5hmC with SIMPLE-seq, Bai, D. et al., Nature Biotechnology (2024)



View online