Genome-wide Mapping Eukaryotic Open Chromatin Regions with Double-Stranded DNA Deaminase

Wenyang Dong[1,#], Jiankun Zhang[1,2,3,#], Leyi Dai[1,2,4,#], Zhi Wang[1,2,3], Yuxuan Pang[1,2,4], Runsheng He[1,2], Jun Ren[1], Yijun Liu[1], Yu Tian[1], Xiaoliang Sunney Xie[1,2,*]

Affiliations

[1] Changping Laboratory; Beijing, 102206, China.

[2] Beijing Advanced Innovation Center for Genomics (ICG) & Biomedical Pioneering Innovation Center (BIOPIC), Peking University; Beijing 100871, China.

[3] School of Life Sciences, Peking University; Beijing, 100871, China.

[4] Academy for Advanced Interdisciplinary Studies, Peking University; Beijing, 100871, China.

# Authors contributed equally.

* Corresponding author. Email: sunneyxie@biopic.pku.edu.cn

Abstract

A variety of methods such as ATAC-seq, DNase-seq and MNase-seq have been developed to map open chromatin or nucleosome positioning. However, these techniques involve enzymatic cleavage of DNA, which cannot be easily carried out with other genome-wide DNA profiling simultaneously. While methyltransferase-based methods like NOMe-seq and Fiber-seq have been used for chromatin profiling free of DNA cleavage, their processing requires bisulfite sequencing or amplification-free third generation sequencing. We recently reported the use of double-stranded DNA (dsDNA) deaminases for footprinting of transcription factors. Here, we present a variant of FOODIE, chromatin accessibility by footprinting with deaminase (caFOODIE), for detecting open chromatin regions in eukaryotic nucleus. caFOODIE exhibits high degree of consistency with ATAC-seq and DNase-seq. Notably, after

caFOODIE deamination, genomic DNA remains unfragmented, and the cytosine conversion is preserved after amplification. The versatility of this technique offers potential for easy integration with other genomic profiling methods, such as 3D genome.

Introduction

In eukaryotic cells, DNA is packaged into chromatin within the nucleus, with nucleosome organization playing a critical role in regulating transcription. Chromatin accessibility is a reflection of areas of the genome bound by sequence-specific transcription factors and their co-factors, which are often associated with gene activation and repression (1). To map open chromatin regions along the genome, several enzymatic cleavage-based methods, such as DNase-seq (2), MNase-seq (3), and ATAC-seq (4), have been established. Methods based on enzymatic modifying-based methods have also been developed. These include NOMe-seq, which uses cytosine methyltransferase M.CviPI and requires a follow-up bisulfite-conversion treatment to detect methylation (3). More recently, DNA N6-adenine methyltransferase has been used for chromatin structure profiling, enabling footprinting at high enzyme concentrations (Fiber-seq) and open chromatin mapping at lower concentrations (AdMTase-seq) (5). However, 6mA modifications in Fiber-seq are not preserved during PCR amplification, posing a technical limitation.

Deaminases are part of a superfamily of enzymes that are widely distributed across eukaryotic and prokaryotic organisms, catalyzing the deamination of DNA (single- and double-stranded), RNA, nucleosides, and other nucleotide derivatives (6, 7). Deaminase-involved processes are important for many biological functions, such as immunity, metabolism, and evolution (8, 9). Single-stranded DNA (ssDNA) deaminases, including enzymes that act on adenosine, cytidine, and guanine, have been utilized in various biotechnological applications, such as DNA cleavage-free genome-editing techniques (e.g., base editors) (10, 11), or mapping epigenetic modifications in DNA and RNA (12, 13). One such example is the Cabernet technique based on the

APOBEC deaminase, which enables single-cell bisulfite-free 5mC and 5hmC sequencing (14).

Double-stranded cytosine deaminases have been recently exploited for use in mitochondrial DNA editing (7), or one-step enzymatic 5mC detection on dsDNA (9). Instead of developing tools for gene editing, we aimed to repurpose dsDNA deaminase for studying chromatin and transcription. Previously, we reported FOODIE, a novel technique enabling genome-wide identification of human transcription factor binding sites with near single-base resolution on a single-molecule and single-cell basis (15).

In this study, we present a chromatin interrogation method, caFOODIE, that directly identifies open chromatin on DNA molecules. Detecting chromatin accessibility with caFOODIE is DNA cleavage-free and can be detected after PCR amplification, which holds the promise for integration with other genomic methods in a single assay.

Results

caFOODIE detects open chromatin regions

In the caFOODIE workflow (Fig. 1A), permeabilized nuclei are treated with DddA, a dsDNA deaminase that catalyzes cytosine (C) to uracil (U) conversion without DNA fragmentation. Following library preparation and sequencing, converted sites can be precisely identified. The cytosine deamination of genomic DNA by DddA acts only on cytosine preceded by thymine (TC) (Fig. 1B), consistent with a previous report (7).

C to U conversion by caFOODIE produces reliable data with a low level of genome-wide cytosine conversion (Fig. 1C), which occurred preferably in open chromatin regions (Figure 1D). Conversion ratios exhibit a strong enrichment near transcription start sites (TSS) of genes, with higher ratios observed for genes with elevated expression levels. Similarly, the conversion ratios are higher at DNase hypersensitive sites (DHS), indicating that caFOODIE deamination is enriched in regions of open chromatin.

The low levels of cytosine conversion required by caFOODIE relative to methods like FOODIE allow a broad range of conversion ratios that can be tuned to maximize detection sensitivity of open chromatin regions. In the experiments described here, reads were typically 150bp (with next generation sequencing), and a positive read in an open chromatin region was defined to have at least 20% TC conversion.

Comparison with Other Open Chromatin Detection Methods

We compared caFOODIE to established methods for detection of open chromatin, including ATAC-seq and DNase-seq. Like ATAC-seq and DNase-seq, caFOODIE revealed clear enrichment of signals around transcription start sites (TSSs) (Fig. 2A). Comparison of signal tracks from caFOODIE, ATAC-seq, and DNase-seq showed a clear pattern of caFOODIE peaks (Fig. 2B) and strong concordance with ATAC-seq, and DNase-seq, identifying accessible regions with similar resolution.

To further validate the performance of caFOODIE, we assessed its reproducibility and concordance with other chromatin accessibility methods through Pearson correlation analysis (Fig. 2C). caFOODIE demonstrated strong reproducibility across biological replicates (R=0.8), highlighting its robustness and reliability. caFOODIE exhibited moderate correlation with DNase-seq (R = 0.6) and ATAC-seq (R = 0.49), while DNase-seq and ATAC-seq exhibit high correlation with each other (R = 0.83). These results suggest that caFOODIE captures open chromatin regions in common with ATAC-seq and DNase-seq, while potentially harboring method-specific patterns. A similar phenomenon has been reported in NOMe-seq, which has low correlation (R = 0.2) with ATAC-seq or DNase-seq (16).

Outlook

caFOODIE achieves high reproducibility and results consistent with established chromatin accessibility techniques, supporting its utility as a reliable and complementary approach for genome-wide open chromatin profiling. Further,

caFOODIE preserves genomic DNA integrity and enables detection of converted cytosines after amplification, which offers unique advantages for integration with long-read sequencing technologies like PacBio or Nanopore sequencing, even with few (or single) input cells. This makes it particularly suitable for profiling chromatin dynamics in challenging or low-abundance samples, such as biopsies or progenitor cells. caFOODIE provides an opportunity to study long-range co-accessibility and chromatin interactions across multiple genomic regions in rare cell types. Combined with Hi-C, this method could provide insights into the interplay between 3D genome organization and chromatin accessibility. The integration of these technologies will enable more comprehensive analyses of gene regulation, chromatin structure and interactions.

Acknowledgments

This project is financially supported by Changping Laboratory, the Ministry of Science and Technology of China, Beijing Advanced Innovation Center for Genomics at Peking University and Peking-Tsinghua Center for Life Sciences.

Competing interests

W. D., R. H., Z. W., and X. S. X. are coinventors on patent application WO/2024/065721 that includes discoveries described in this manuscript. The remaining authors declare no competing interests.

Figures

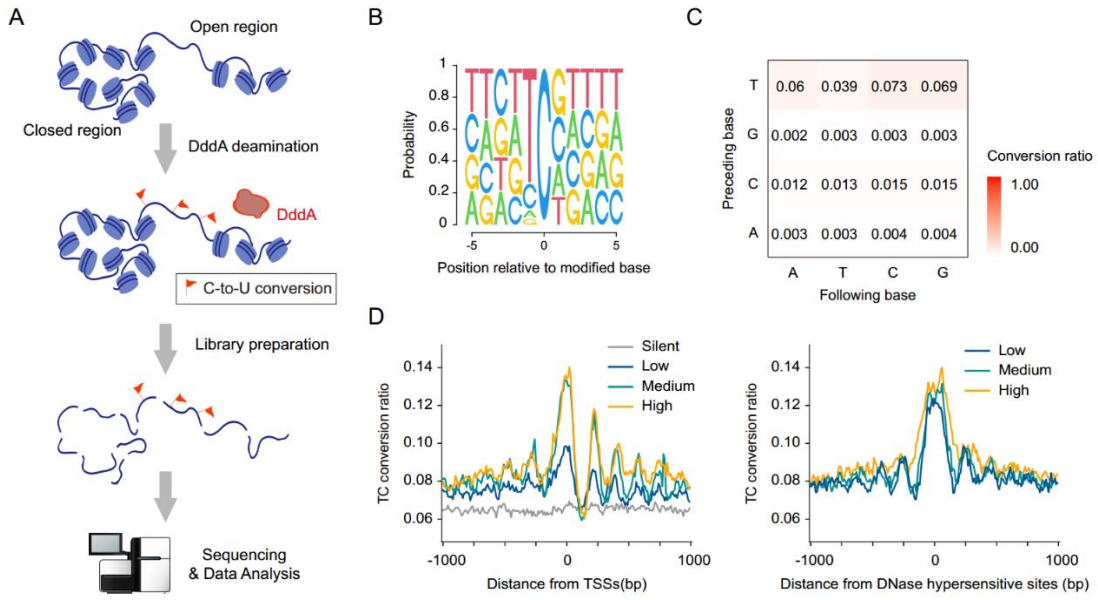

Figure 1: caFOODIE workflow and the corresponding cytosine conversion of genomic DNA. (A) Schematic representation of the caFOODIE workflow. The double-stranded DNA deaminase DddA catalyzes cytosine (C) to uracil (U) conversion. After deamination, genomic DNA remains intact and is processed for library preparation, sequencing, and analysis. (B) The Sequence logo of nucleotide frequency at positions relative to C nucleotides converted by DddA on genomic DNA. (C) DddA's C to U conversion efficiency on genomic DNA. The x-axis is the base following, and the y-axis base proceeding the cytidine. (D) Conversion ratios around transcription start sites (TSS, left) and DNase hypersensitive sites (DHS, right) stratified by gene expression levels or chromatin accessibility. Higher TC conversion ratios are observed near TSS and DHS regions, particularly in active genes or open chromatin.

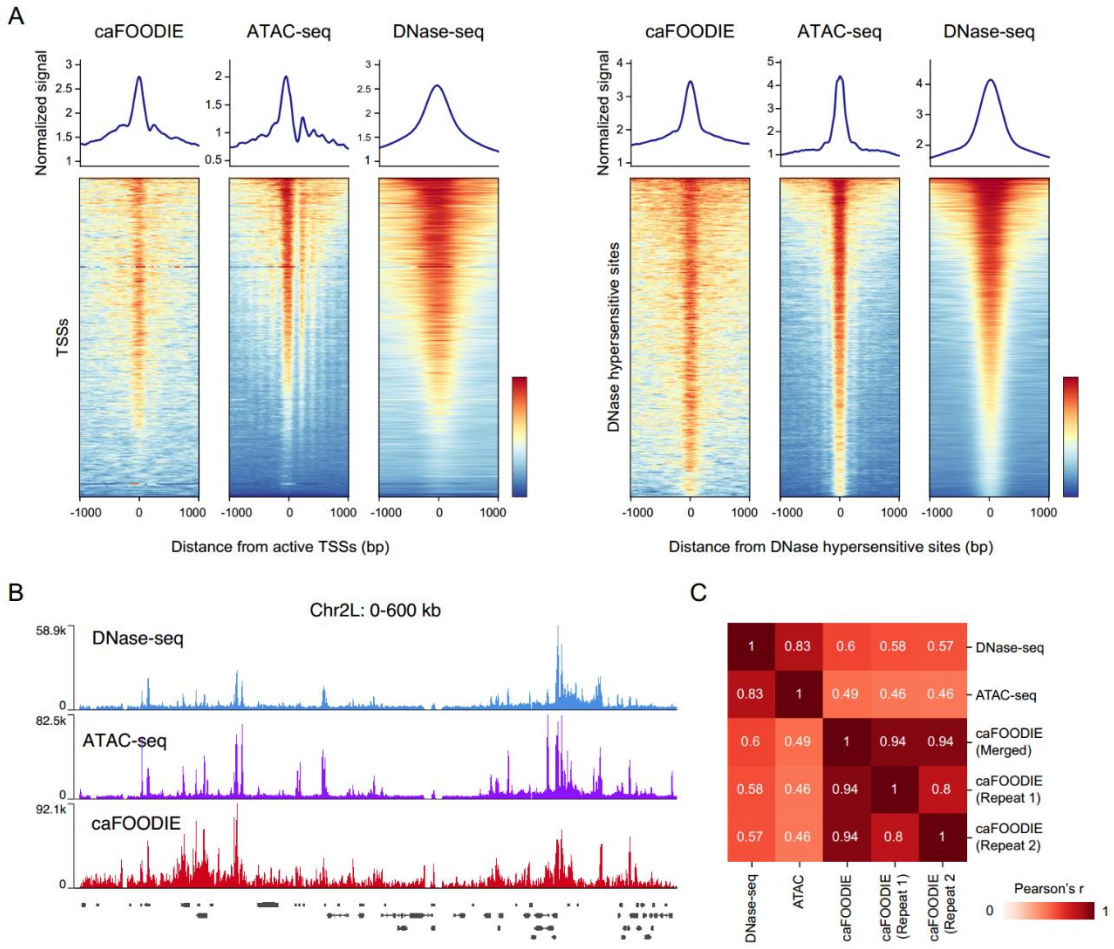

Figure 2. Comparison of caFOODIE with ATAC-seq and DNase-seq for chromatin accessibility detection. (A) Comparison of signals of reads detected by methods of caFOODIE, ATAC-seq, and DNase-seq around transcription start sites (TSS, left) and DNase hypersensitive sites (DHS, right). caFOODIE signals are enriched at both TSS and DHS regions, displaying patterns highly similar to those of ATAC-seq and DNase-seq. (B) Genome browser tracks for DNase-seq (blue), ATAC-seq (purple), and caFOODIE (red) signals across a representative genomic region (Chr2L: 0–600 kb). caFOODIE exhibits strong concordance with ATAC-seq and DNase-seq. (C) Heatmap of Pearson correlation coefficients between caFOODIE, ATAC-seq, and DNase-seq signals.

Materials and Methods

Cell culture and preparation

Drosophila Schneider 2 (S2) cells were gifted from Dr. Yan Song at Peking University. S2 cells were cultured in Schneider's Drosophila Medium (Gibco, cat. no. 21720024) supplemented with 10% fetal bovine serum (FBS) (Cellmax, cat. no. SA301.02) and 1% Penicillin-Streptomycin (Pen-Strep) (Gibco, Thermo Fisher Scientific, cat. no. 15140122) at 25°C without CO_2. Upon use, S2 cells were collected by centrifuge at 300g for 5 min and resuspended in 1×PBS. When fixation (0.5% PFA) is needed, 3 million S2 cells were resuspended in 500μL 0.5% PFA for 10 minutes. Fixed cells were collected by centrifuge at 600g for 5 min, resuspend in 300μL 1% BSA/PBS for 5 min to quench PFA, and lastly washed three times with 1% BSA-PBS.

Open chromatin labelling through deamination

A total of 500,000 0.5% PFA fixed S2 cells were lysed with 150μL cold lysis buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.5 mM DTT and 0.1% IGEPAL-CA630) by incubating on ice for 5 min, then collected by centrifuge at 600g for 5 min. Lysed nuclei were resuspended in 50μL 1× deamination buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.5 mM DTT) containing a titration of DddA enzyme. The incubation reaction was performed at 37°C for 30 minutes with gentle rotating. After deamination, nuclei were washed twice with PBS. After that, genomic DNA extraction. A negative control of nuclei without deamination can be set before genomic DNA extraction. For genomic DNA extraction, the nuclei were collected by centrifuge at 600g for 5 min, and resuspend in 20μL 0.1%SDS-PBS with 2μL proteinase K (P8107S, NEB). After overnight incubation at 55°C with lid-heat on, the cleanup reaction was performed with a Zymo DNA Clean and Concentrator-5 Kit (D4014, Zymo). Library preparation with eluted genomic DNA can be performed using TruePrep DNA Library Prep Kit (TD501, Vazyme) following the manual. In general, a genome-wide TC

conversion ratio at around 5% after deamination by DddA at certain titration is ideal for open chromatin labelling.

Data Analysis

Raw sequencing reads were aligned to the dm6 reference genome using biscuit (version 1.5.0). PCR duplicates were counted and removed using dupsifter (version 1.0.0). The C-to-T conversions were then extracted using biscuit tools with default parameters. To obtain open chromatin signals, we used a custom Python script to selected reads with >20% converted TC sites. The bigwig files were generated using bamCoverage (deeptools, version 3.5.1) for downstream analysis and visualization.

Published ATAC-seq data of Drosophila S2 cells were downloaded from the Gene Expression Omnibus under no. GSE103177. Bulk RNA-seq data were taken from the Gene Expression Omnibus under no. GSE145320. DNase-seq data were taken from the online flybrain resource (http://compbio.med.harvard.edu/flychromatin/data.html). LiftOver was used to convert genomic coordinates from dm3 to dm6.

References

1. Jiang C & Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10(3):161-172.
2. Hesselberth JR, *et al.* (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* 6(4):283-289.
3. Johnson SM, Tan FJ, McCullough HL, Riordan DP, & Fire AZ (2006) Flexibility and constraint in the nucleosome core landscape of Caenorhabditis elegans chromatin. *Genome Res* 16(12):1505-1516.
4. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, & Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10(12):1213-1218.
5. Stergachis AB, Debo BM, Haugen E, Churchman LS, & Stamatoyannopoulos JA (2020) Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* 368(6498):1449-1454.
6. Iyer LM, Zhang D, Rogozin IB, & Aravind L (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic

nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res* 39(22):9473-9497.

7. Mok BY, *et al.* (2020) A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* 583(7817):631-637.
8. Huang J, *et al.* (2024) Discovery of deaminase functions by structure-based protein clustering. *Cell* 187(16):4426-4428.
9. Vaisvila R, *et al.* (2024) Discovery of cytosine deaminases enables base-resolution methylome mapping using a single enzyme. *Mol Cell* 84(5):854-866 e857.
10. Komor AC, Kim YB, Packer MS, Zuris JA, & Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533(7603):420-424.
11. Nishida K, *et al.* (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353(6305).
12. Schutsky EK, *et al.* (2018) Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nat Biotechnol*.
13. Xiao YL, *et al.* (2023) Transcriptome-wide profiling and quantification of N(6)-methyladenosine by enzyme-assisted adenosine deamination. *Nat Biotechnol* 41(7):993-1003.
14. Cao Y, *et al.* (2023) Single-cell bisulfite-free 5mC and 5hmC sequencing with high sensitivity and scalability. *Proc Natl Acad Sci U S A* 120(49):e2310367120.
15. He R, *et al.* (2024) Genome-wide single-cell and single-molecule footprinting of transcription factors with deaminase. *Proc Natl Acad Sci U S A* 121(52):e2423270121.
16. Nordstrom KJV, *et al.* (2019) Unique and assay specific features of NOMe-, ATAC- and DNase I-seq data. *Nucleic Acids Res* 47(20):10580-10596.