PNAS COMMENTARY OPEN ACCESS

Efficient, scalable, and near-nucleotide-resolution profiling of protein occupancy in the genome with deaminases

Lei Chang[a] and Bing Ren[a,b,c,1]

Transcription factors (TFs) bind to specific DNA sequences (*cis*-regulatory elements), often in cooperation with other DNA-binding proteins and cofactors, to establish the transcriptional program in each cell that is responsible for cell identity, function, and adaptability (1). Knowing where TF–DNA interactions take place throughout the genome is important because aberrations in TF binding have been linked to developmental disorders, cancers, and complex diseases such as diabetes and neurodegeneration. In this issue of PNAS, He et al. introduce an innovative approach, named FOODIE, which enables near base-resolution analysis of transcription factor occupancy genome wide at cellular level. This tool substantially enhances our ability to dissect gene regulatory networks in diverse cell types of multicellular organisms.

TF binding analysis in multicellular organisms is challenging due to several factors. First, TF occupancy is inherently dynamic and context-dependent, varying by cell type, developmental stage, and environmental conditions, demanding analysis of TF binding in primary tissues at single-cell resolution. Second, TFs recognize relatively short sequence motifs, as such the presence of a motif alone is not strongly predictive of actual binding, necessitating empirical measurements. Third, TF function frequently relies on cooperative interactions with other factors, further complicating the analysis of their binding patterns. Finally, the genome can encode thousands of TFs, and current methodologies lack the scalability, resolution, and throughput required to systematically map their occupancy across distinct cell populations.

Conventional methods (Fig. 1) to map TF binding, such as chromatin immunoprecipitation followed by sequencing (ChIP-seq) (2, 3), ChIP-exo (4), or Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (5), depend on antibodies that vary in quality and availability, limiting their generalizability. Alternatively, analysis of TF binding can also be achieved through detection of their footprints on DNA using high-throughput DNA sequencing platforms because the TF-bound sequences are resistant to nuclease cleavage or transposon insertions. Employing this principle, DNase-seq (6) and ATAC-seq (7) have been successfully used to map TF footprints (8–10). However, these assays typically require large numbers of cells and deep DNA sequencing, limiting their application to bulk samples and precluding the resolution of combinatorial TF-binding events on the same DNA molecules.

Recent advances in single-molecule footprinting (SMF) leverage exogenous DNA-modifying enzymes to detect TF footprints on individual DNA molecules (11). Because SMF can reveal both occupied and unoccupied DNA, it offers the advantage to determine the frequency at which a TF is bound to DNA in a cell population. To date, three classes of SMF approaches have been reported, as detailed below (Fig. 1).

The first class employs cytosine methyltransferases to methylate cytosines on the exposed DNA sequences. One such enzyme, M.CviPI, methylates cytosines at GpC dinucleotides. Binding of a TF to DNA shields the cytosines within the binding sites from methylation, leaving them detectable as thymines after bisulfite conversion, while the cytosines outside the TF binding sites are detected as cytosines (12, 13). When combined with third-generation, single-molecular long-reads DNA sequencing technologies, this approach can yield haplotype-resolved patterns of DNA binding by multiple transcription factors (14). However, the resolution of this approach is constrained by the uneven distribution of GpC sites within the TF binding sites as well as the endogenous cytosine methylation.

The second class of SMF methods makes use of the N6 adenine methyltransferases, which methylates the exposed adenines, yielding m6A that can be detected via third-generation sequencing. Methods such as Fiber-seq (15), SAMOSA (16), and DiMeLo-seq (17) use this strategy to map TF footprints and nucleosome positioning along the DNA in cells. Because adenines are abundant in the genome, these methods achieve near-base resolution. Still, their widespread adoption is limited by the relatively high cost of third-generation sequencing and the challenges in reliably detecting m6A.

In the current issue of PNAS, He et al. (18) introduce a third class of SMF methods, which utilizes double-strand DNA deaminases to convert cytosines to uracils, which can be detected as thymines after PCR and next-generation sequencing. Cytosines within TF-bound sites remain unconverted, enabling the detection of footprints at near-nucleotide resolution (Fig. 1*A*). Moreover, because this approach is compatible with widely used next-generation sequencing platforms, it promises lower cost, higher scalability, and suitability for single-cell analyses (Fig. 1*B*).

To find highly active double-stranded DNA deaminases suitable for SMF, He et al. screened various bacterial enzymes and identified two candidates—DddB and MGYPDa829—that

Author affiliations: [a]Department of Cellular and Molecular Medicine, University of California, San Diego School of Medicine, La Jolla, CA 92093; [b]Center for Epigenomics, University of California, San Diego School of Medicine, La Jolla, CA 92093; and [c]Moores Cancer Center, University of California, San Diego School of Medicine, La Jolla, CA 92093

Author contributions: L.C. and B.R. wrote the paper.

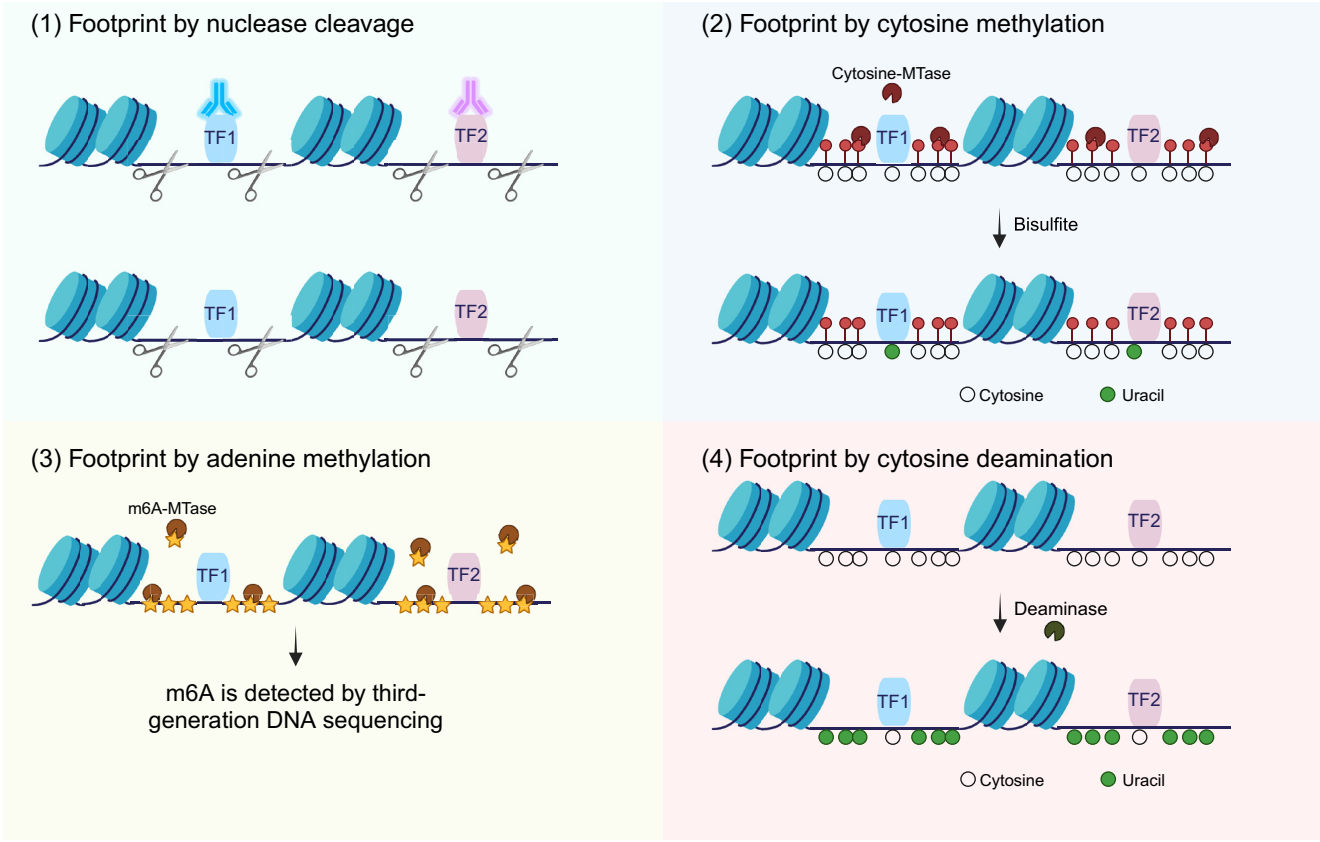
Competing interest statement: B.R. is a cofounder of Epigenome Technologies, Inc. and has equity in Arima Genomics, Inc. The remaining authors declare no competing interests.

See companion article, "Genome-wide single-cell and single-molecule footprinting of transcription factors with deaminase," 10.1073/pnas.2423270121.

[1]To whom correspondence may be addressed. Email: biren@health.ucsd.edu.

Published January 27, 2025.

A



B

Category	Representative methods	Resolution	Bulk version	Single-cell version	NGS platform	TGS platform
Nuclease cleavage	DNase-seq, ChIP-exo ATAC-seq, CUT&Run	Near motif	√		√	
Cytosine methylation	Single Molecule Footprinting (SMF), NOMe–seq	>7bp	√		√	√
Adenine methylation	Fiber-seq, SAMOSA, DiMeLo-seq	Near base	√			√
Cytosine deamination	FOODIE, DAF-seq	Near base	√	√	√	√

Fig. 1. (*A*) Schematic representation of four approaches of TF footprint profiling. (*B*) Comparison of different TF footprinting methods in terms of genomic resolution, cellular resolution, and compatibility with different types of sequencing platforms. NGS, Nextgen sequencing platforms; TGS, third-generation sequencing platforms.

displayed the necessary enzymatic activity and even sequence coverage. However, MGYPDa829's tendency to progressively convert cytosines into uracils complicated the detection of cooperatively bound TFs on adjacent sites. As a result, DddB was chosen for the development of FOODIE. Because TFs typically bind to open, accessible chromatin, the authors further integrated FOODIE with Tn5 transposase-based fragmentation to isolate these regions from cells for TF footprint analysis. This streamlined protocol is robust, easy to implement with relatively small cell numbers, and is amendable for single-cell analyses.

He et al. demonstrated that FOODIE accurately detected TF footprints in a human lymphoblastoid cell line GM12878 and offered major improvements in sensitivity and resolution over previous TF footprinting methods. They confidently assigned the footprints from this cell type to 79 known TFs based on existing ChIP-seq data, and further characterized the distribution of the TF binding sites relative to each other in the context of promoters and enhancers. The authors also developed single-cell FOODIE (scFOODIE) to enable TF footprint analysis in heterogeneous cell populations. As proof of principle, the authors demonstrated that scFOODIE was able to resolve TF binding patterns in a mix of four human cell lines (GM12878, K562, HEK293, and HeLa), where cell-type-specific open chromatin signals can be used to classify the cells. They then applied scFOODIE to mouse hippocampus, successfully profiling TF footprints in 11,200 cells from eight major brain cell types.

FOODIE analysis revealed how TFs coordinate transcription of genes expressed at the same time (referred to as correlated gene modules, CGMs) (19). The results suggest that genes with essential cellular functions rely on TF binding

at promoters for robustness, while tissue-specific genes depend more on enhancers for greater tunability. The authors also used FOODIE to study the pairwise TF binding patterns on DNA, which can be either positively correlated (cooperativity) or negatively correlated (exclusivity) (20). The authors devised a computational algorithm to quantify TF cooperativity from FOODIE data and found that more TFs exhibit positive cooperativity than exclusivity. The positive cooperativity between TFs support concentration-dependent DNA binding by multiple factors, whereas the negative cooperativity assures fast bidirectional gene regulation with short response time when TF concentrations in the nuclei suddenly change. The ability to detect both positive and negative cooperativity between TFs will greatly improve our understanding of gene regulatory networks in multicellular organisms.

> **"In PNAS, He et al. introduce an innovative approach, named FOODIE, which enables near base-resolution analysis of transcription factor occupancy genome wide at cellular level."**

Independently, Swanson et al. (21) in a preprint described a similar deaminase-based SMF method named Deaminase-Assisted single-molecule chromatin Fiber sequencing (DAF-seq). Unlike FOODIE, this method enables the mapping of TF footprints at select genomic loci with near single-nucleotide precision on a third-generation sequencing platform. FOODIE, compared to other methods for detecting TF footprints, have several key advantages. It is straightforward to implement, adaptable to high-throughput workflows, and cost-effective. Moreover, it can be extended to single-cell resolution, to enable detection of distinct cell types within heterogeneous tissues and assess TF footprints genome-wide in each of the cellular populations. By applying this method across a diverse range of tissues, it will be possible to uncover patterns of combinatorial TF binding and gene regulatory modules, ultimately deepening our understanding of how transcriptional regulation drives development, shapes evolution of species, and contributes to human diseases.

In summary, FOODIE represents a transformative approach that could greatly accelerate our understanding of gene regulatory programs in multicellular organisms. Additionally, by profiling TF footprints in clinical samples, it would allow the identification of disease-specific regulatory signatures. On the other hand, FOODIE alone cannot yet confidently identify the full spectrum of transcription factors bound throughout the genome, because our current catalog of TFs with known sequence specificity is still incomplete, and many transcription factors share similar DNA binding motifs. Combining FOODIE with other TF binding profiling tools such as ChIP-seq, Cut&Run, or Cut&Tag would address these limitations and lead to more complete understanding of the gene regulatory networks.

ACKNOWLEDGMENTS. Research in B.R.'s lab is supported by NIH grants 1UM1HG009402, 1U19MH114831, 1U01MH121282, 1R01AG066018, R01AG067153, U01DA052769, 1UM1HG011585, RF1MH124612, 1R56AG069107, R01EY031663, 1U01HG012059, R24AG073198, RF1MH128838, R41MH128993, UM1MH130994, and 1U54AG079758.

1. S. A. Lambert *et al.*, The human transcription factors. *Cell* **172**, 650–665 (2018).
2. G. Robertson *et al.*, Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
3. D. S. Johnson, A. Mortazavi, R. M. Myers, B. Wold, Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
4. H. S. Rhee, B. F. Pugh, Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
5. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).
6. J. R. Hesselberth *et al.*, Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
7. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
8. Z. Li *et al.*, Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).
9. M. Bentsen *et al.*, ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).
10. Y. Hu *et al.*, Single-cell multi-scale footprinting reveals the modular organization of DNA regulatory elements. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.03.28.533945 (Accessed 7 December 2024).
11. A. R. Krebs, Studying transcription factor function in the genome at molecular resolution. *Trends Genet.* **37**, 798–806 (2021).
12. C. Sönmezer *et al.*, Molecular co-occupancy identifies transcription factor binding cooperativity in vivo. *Mol. Cell* **81**, 255–267.e6 (2021).
13. F. D. Lay, T. K. Kelly, P. A. Jones, "Nucleosome occupancy and methylome sequencing (NOMe-seq)" in *DNA Methylation Protocols*, J. Tost, Ed. (Springer New York, New York, NY, 2018), pp. 267–284.
14. S. Battaglia *et al.*, Long-range phasing of dynamic, tissue-specific and allele-specific regulatory elements. *Nat. Genet.* **54**, 1504–1513 (2022).
15. A. B. Stergachis, B. M. Debo, E. Haugen, L. S. Churchman, J. A. Stamatoyannopoulos, Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
16. N. J. Abdulhay *et al.*, Massively multiplex single-molecule oligonucleosome footprinting. *Elife* **9**, e59404 (2020).
17. N. Altemose *et al.*, DiMeLo-seq: A long-read, single-molecule method for mapping protein–DNA interactions genome wide. *Nat. Methods* **19**, 711–723 (2022).
18. R. He *et al.*, Genome-wide single-cell and single-molecule footprinting of transcription factors with deaminase. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2423270121 (2024).
19. A. Dixit *et al.*, Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016).
20. A. Jolma *et al.*, DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**, 384–388 (2015).
21. E. G. Swanson *et al.*, Deaminase-assisted single-molecule and single-cell chromatin fiber sequencing. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.11.06.622310 (Accessed 7 December 2024).