BIOLOGICAL TECHNIQUES

Mapping the diploid genome, one cell at a time

Attempts to develop a method for 3D genome reconstruction of single cells have been frustrated by the inability to distinguish between chromosome homologs. A novel Hi-C workflow uses haplotype imputation to map the nuclear organization of single diploid cells.

Blake A. Caldwell and Marisa S. Bartolomei

In the 15 years since the first paper describing chromosome-conformation capture (3C) was published, the technique has evolved from a PCR-driven assay of two interacting loci to modern Hi-C platforms capable of testing all possible pairwise interactions across the genome[1,2]. Coincident with this development, single-cell 'omics' has yielded tremendous insight into the cellular heterogeneity present in biological systems[3]. However, the application of single-cell technologies to the study of chromatin architecture has always faced a 'diploid problem': how to resolve paired Hi-C sequencing reads as representing the maternal or the paternal homologous chromosome? In a recent paper in *Science*, Tan et al.[4] report a novel single-cell chromatin-conformation-capture method (termed Dip-C) using haplotype imputation to reconstruct the 3D genomes of single diploid human cells.

Nuclear organization serves several critical functions in eukaryotic genome regulation[5]. By compartmentalizing chromosomes into discrete regions of active and inactive chromatin, cells are able to preserve genomic integrity while exerting tight control over nuclear processes such as transcription, DNA repair, and replication. Hierarchical clustering of chromatin into topologically associated domains has also been proposed to function in cell-identity maintenance, thus spatially restricting DNA into lineage-specific promoter–enhancer interactions. Finally, chromatin looping provides an additional measure of epigenetic control through which cells respond to environmental stimuli, such as in memory formation and circadian entrainment[6,7].

To study nuclear organization, researchers use proximity-ligation-based 3C and Hi-C methods to obtain precise spatial information on interacting DNA regions. These measurements, however, represent the population mean of bulk samples and do not capture information on 3D genome arrangement and allelic disparities visible only at the single-cell level. Recent work using Oligopaint fluorescent probes has yielded considerable insight into intercellular differences in chromosome territories and nuclear organization, but these studies lack the spatial resolution of sequencing-based approaches[8]. Despite progress in the application of single-cell Hi-C for 3D genome reconstruction of haploid cell lines, the inability to distinguish between chromosome haplotypes has been a major barrier in the development of single-cell proximity-ligation-based techniques[9].

Tan et al. overcame this limitation by using a modified single-cell Hi-C methodology to reconstruct the 3D genomes of diploid cells (Fig. 1). By implementing a transposon-based whole-genome-amplification method (multiplex end-tagging amplification, or META), they improved upon standard Hi-C sensitivity to preserve ligated 'contacts' while limiting the introduction of artifactual paired-end chimeras. To resolve the haplotypes of sequenced reads, the researchers next developed an imputation algorithm based on two statistically validated properties of interchromosomal contacts: first, chromosome homologs normally contact different partners; second, the probability of additional contacts between two interacting chromosomes diminishes over increasing genomic distances. Using this algorithm, the researchers defined high-probability interchromosomal-contact neighborhoods to impute haplotype on the basis of allele-specific single-nucleotide polymorphisms present in the genome. From these 2D haplotype maps, the researchers were then able to reconstruct 3D maps of chromosome localization in the nucleus.

The resulting Dip-C datasets profile nuclear organization across all genomic scales. At the macro level, Tan and colleagues were able to visualize the organization of chromosomes into 'fractal globules' of spatially clustered chromatin, each of which segregated into distinct regions within the nucleus. As expected, these chromatin territories exhibited varying radial preferences: gene-rich regions localized to the nuclear center, and gene-poor regions localized to the nuclear periphery. At higher resolution, the researchers visualized CCCTC-binding-factor loops and found that their organization is highly heterogeneous among cells. Notably, many of these features would be concealed at the bulk sequencing level.

The researchers next applied the Dip-C methodology to study known structural differences between maternal and paternal alleles. They first examined differences in the 3D arrangement of genes in the *H19*/*IGF2* genomic-imprinting cluster. Here, transcription of *H19* from the maternal allele or *IGF2* from the paternal allele is regulated by parent-of-origin-specific DNA methylation patterns[10]. In agreement with this mechanism, looping between *IGF2* and the shared distal enhancer region was disrupted on the maternal allele, although the degree of looping exhibited clear intercellular heterogeneity. The researchers then probed differences in X-chromosome organization resulting from random X inactivation. Although the active and inactive X chromosomes were difficult to distinguish on the basis of compaction alone, the researchers found that the active chromosome could be identified according to its compartmentalization of heterochromatic and euchromatic regions, and the inactive chromosome could be identified according to its formation of long-distance 'superloops'[11].

Having validated their methodology, the researchers next set out to test its utility in probing cell-type-specific interactions. One of the most valuable applications of single-cell technologies is the deconvolution of heterogeneity present in complex tissues and mixed cell populations. For example, the brain has been estimated to be composed of 10,000 different neuronal subtypes, a number further complicated by the intrasubtype differences in gene expression that contribute to learning and behavior[12]. Although early efforts sought to parse this information on the basis of single-cell transcriptomics, recent data have shown that intercellular epigenetic

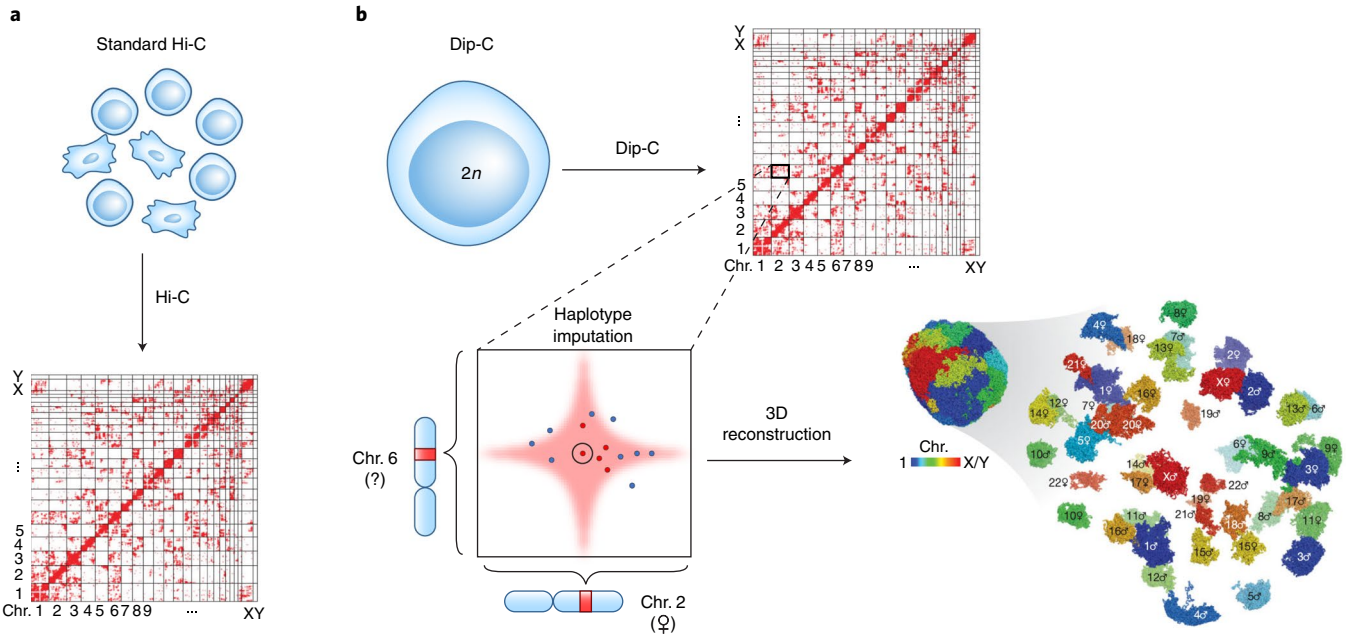

Fig. 1 | 3D genome technologies. a, In standard Hi-C, populations of cells are subjected to chromatin-conformation capture and high-throughput sequencing to measure mean pairwise interactions across the entire genome. Chr., chromosome. **b**, The modified Dip-C methodology enables whole-genome chromatin-conformation capture and 3D genome reconstruction of single diploid cells. Using 2D Hi-C contact maps, the researchers performed haplotype imputation to infer whether a given contact read represented the maternal or paternal chromosome. Around each interchromosomal contact ('reference', black circle), contact neighborhoods (red superellipses) were defined in which there was a high probability of additional contacts between interacting chromosomes. Haplotypes were inferred according to the presence of contacts containing allele-specific single-nucleotide polymorphisms (red dots) within these neighborhoods. After haplotype imputation, 3D genome reconstruction was performed to model nuclear organization within each cell. Images from ref. [4] are reprinted with permission from AAAS.

differences may be a more reliable predictor of cellular identity[13,14]. To test whether differences in 3D genome structures can be used to identify cell types from a mixed population, Tan and colleagues performed Dip-C on peripheral blood mononuclear cells. On the basis of single-cell chromatin compartmentalization alone, unsupervised principal component analysis successfully identified four cell-type clusters present in the cell samples, thus highlighting the technology's potential as a powerful new tool in applied single-cell omics.

Beyond single-cell-profiling experiments, the ability of Dip-C to identify cell-type-specific DNA loops and chromatin structure should be useful in addressing several longstanding questions in biology. For example, symmetry breaking in the eight-cell embryo represents the earliest stage of lineage commitment during mammalian embryonic development, but the origins of that asymmetry remain mysterious[15]. With Dip-C profiling, it would be interesting to test whether previously uncharacterized intercellular differences in chromatin structure might promote segregation into embryonic and extraembryonic lineages. Similarly, Dip-C would be an effective tool for the study of dynamic 3D genome rearrangements present in transition states, such as those observed during meiotic crossover or reprogramming of induced pluripotent stem cells[16,17].

As the appreciation of nuclear organization has evolved from a 'simple packaging tool' to an 'essential regulator of genome integrity and cellular identity', researchers have been constrained by a lack of suitable techniques to probe intercellular differences in 3D genome organization. With Dip-C, Tan and colleagues have circumvented the limitations inherent to proximity-ligation-based methods to capture the 3D genomes of single diploid cells. Although their current results largely reinforce known principles of nuclear organization, it will be exciting to apply Dip-C to the study of complex tissues and developmental time points previously unsuited to standard Hi-C profiling. ❐

Blake A. Caldwell* and Marisa S. Bartolomei*

Epigenetics Institute, Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

**e-mail: blakec@pennmedicine.upenn.edu; bartolom@pennmedicine.upenn.edu*

Published online: 29 October 2018

https://doi.org/10.1038/s41594-018-0149-y

References

1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. *Science* **295**, 1306–1311 (2002).
2. Lieberman-Aiden, E. et al. *Science* **326**, 289–293 (2009).
3. Wang, D. & Bodovitz, S. *Trends Biotechnol.* **28**, 281–290 (2010).
4. Tan, L., Xing, D., Chang, C.-H., Li, H. & Xie, X. S. *Science* **361**, 924–928 (2018).
5. Bonev, B. & Cavalli, G. *Nat. Rev. Genet.* **17**, 661–678 (2016).
6. Watson, L. A. & Tsai, L.-H. *Curr. Opin. Neurobiol.* **43**, 48–55 (2017).
7. Kim, Y. H. et al. *Science* **359**, 1274–1277 (2018).
8. Beliveau, B. J. et al. *Nat. Commun.* **6**, 7147 (2015).
9. Stevens, T. J. et al. *Nature* **544**, 59–64 (2017).
10. Tremblay, K. D., Saam, J. R., Ingram, R. S., Tilghman, S. M. & Bartolomei, M. S. *Nat. Genet.* **9**, 407–413 (1995).
11. Rao, S. S. P. et al. *Cell* **159**, 1665–1680 (2014).
12. Wichterle, H., Gifford, D. & Mazzoni, E. *Science* **341**, 726–727 (2013).
13. Lake, B. B. et al. *Science* **352**, 1586–1590 (2016).
14. Luo, C. et al. *Science* **357**, 600–604 (2017).
15. Rossant, J. & Tam, P. P. *Dev. Cell* **7**, 155–164 (2004).
16. Muller, H. et al. *Mol. Syst. Biol.* **14**, e8293 (2018).
17. Krijger, P. H. et al. *Cell Stem Cell* **18**, 597–610 (2016).

Competing interests

The authors declare no competing interests.