

Africa usually coincides with the end of the wet season in this region, when annual grass seeds are in abundance. It will be worth investigating whether migratory birds in the Southern Hemisphere also influence the redistribution of plant communities during global warming. Likewise, exploring the long-distance dispersal of seeds of aquatic plants, such as seagrasses¹⁰ by water birds, is another area for future research that might benefit from González-Varo and colleagues' methods.

This study provides a great example of how migratory birds might assist plant redistribution to new locations that would normally be difficult for them to reach on their own, and which might offer a suitable climate. As the planet warms, understanding how such biological mechanisms reorganize plant communities complements the information available from climate-projection models, which offer predictions of future species distributions.

Barnabas H. Daru is in the Department of Life Sciences, Texas A&M University-Corpus Christi, Corpus Christi, Texas 78412, USA. e-mail: barnabas.daru@tamucc.edu

- Jordano, P., García, C., Godoy, J. A. & García-Castaño, J. L. *Proc. Natl Acad. Sci. USA* **104**, 3278–3282 (2007).
- González-Varo, J. P. *et al.* *Nature* **595**, 75–79 (2021).
- Viana, D. S., Santamaría, L. & Figuerola, J. *Trends Ecol. Evol.* **31**, 763–775 (2016).
- Lehouck, V., Spanhove, T. & Lens, L. *Plant Ecol. Evol.* **144**, 96–100 (2011).
- Shikang, S., Fuqin, W. & Yuehua, W. *Sci. Rep.* **5**, 11615 (2015).
- Jordano, P. in *Seeds: The Ecology of Regeneration in Plant Communities 3rd edn* (ed. Gallagher, R. S.) 18–61 (CAB Int., 2014).
- Dingle, H. *Migration: The Biology of Life on the Move* (Oxford Univ. Press, 2014).
- Bairlein, F. *Science* **354**, 547–548 (2016).
- Daru, B. H., Farooq, H., Antonelli, A. & Faurby, S. *Nature Commun.* **11**, 2115 (2020).
- Rock, B. M. & Daru, B. H. *Front. Mar. Sci.* **8**, 608867 (2021).

The author declares no competing interests. This article was published online on 23 June 2021.

Chromosome biology

Base-pair view of gene and enhancer interactions

Anne van Schoonhoven & Ralph Stadhouders

A technique reveals how folded chromosomal DNA interacts in the nucleus, providing information at the level of single base pairs. The achievement offers an unprecedented level of detail about how gene activity is regulated. **See p.125**

How can 2 metres of DNA fit into a nucleus that has an average diameter of only 10 micrometres? Almost all the cells in our body face this storage conundrum, which has intrigued scientists for decades. Moreover, this compaction tour de force folds DNA in the nucleus in a way that is far from random. The pattern of DNA folding is important for many processes that involve our genome, including the regulation of expression of our approximately 20,000 genes. On page 125, Hua *et al.*¹ describe a method they have developed to monitor 3D genome architecture. This information can pinpoint genomic interactions at the level of single base pairs of DNA. It suggests new ways of thinking about how gene expression is controlled, and opens up exciting possibilities for future research.

Humans and other organisms have evolved complex mechanisms to precisely regulate gene expression. Different types of cell express different sets of genes, and these expression patterns might depend on a cell's function, or arise in response to environmental cues, such as viral infection. Central to the control of gene expression are short regulatory

sequences of DNA, termed enhancers, which are highly abundant in our genomes. According to current estimates², there are up to 810,000 enhancers across the human genome.

Enhancers are bound by the 'bookkeepers' of gene expression: DNA-binding proteins called transcription factors, which bind to short motifs of DNA sequences corresponding

“This level of detail will enable high-resolution dissection of processes involving gene regulation.”

to 6–12 base pairs³. Enhancers can be located far from the gene(s) that they regulate, and how they stimulate gene expression is a major topic of research⁴. The current leading model is that enhancers and genes are brought into closer spatial proximity by specific patterns of DNA folding, enabling transcription factors to stimulate gene expression despite large intervening genomic distances between an

enhancer and a particular gene^{5–7}.

Study of the 3D organization of genomes has been revolutionized by an approach called chromosome conformation capture (3C), which enables researchers to infer the frequencies of interactions between different DNA regions⁸. Such approaches indicate that enhancer–gene interactions occur preferentially in 'insulated' genomic neighbourhoods in the nucleus called topologically associating domains (TADs)⁹. Most TADs are formed by the cooperative action of a DNA-binding protein termed CTCF and a ring-shaped protein complex called cohesin, which is a type of molecular motor that drives a process known as loop extrusion¹⁰. In this process, cohesin engages DNA and extrudes it, in a similar way to how threading yarn through the eye of a needle forms a loop (Fig. 1). This extrusion continues until cohesin encounters DNA bound to CTCF, which forms a 'roadblock' for loop extrusion, stopping it.

TADs are thought to 'trap' genes and enhancers by thwarting DNA interactions across TAD borders, thereby increasing the probability that matching enhancer–gene pairs find each other. However, until now, 3C technology has been unable to define the nature of the physical contacts between genes and enhancers on the base-pair scale – this would be on a par with the precision with which interactions between DNA and the key transcription factors influencing gene expression have been determined. Hua *et al.* now close this resolution gap by developing a version of 3C that the authors call Micro-Capture-C (MCC).

Building on their previously developed version of 3C methodology¹¹, the authors made key technical refinements that strikingly improved the resolution of the DNA interactions that could be identified. Like all 3C techniques, MCC captures interactions through chemical crosslinking, which generates bonds between interacting regions of DNA. The crosslinked DNA is then cut into smaller fragments, after which the interactions are captured by gluing together (ligating) interacting DNA strands that are close to each other in the nuclear space.

For the pair of molecular 'scissors' that cuts DNA into small fragments, MCC uses the enzyme micrococcal nuclease (MNase), which fragments DNA in a mainly random fashion, independently of DNA sequences. This enables the generation of much smaller DNA fragments than those obtained using sequence-specific enzymes for DNA digestion. The approach helps to increase the resolution – as previously shown for another version of 3C technology¹². Crucially, Hua and colleagues show that DNA fragmentation by MNase does not have any major biases in terms of the DNA that is digested, with a minor preference for less-condensed DNA (characteristic of regions containing genes being

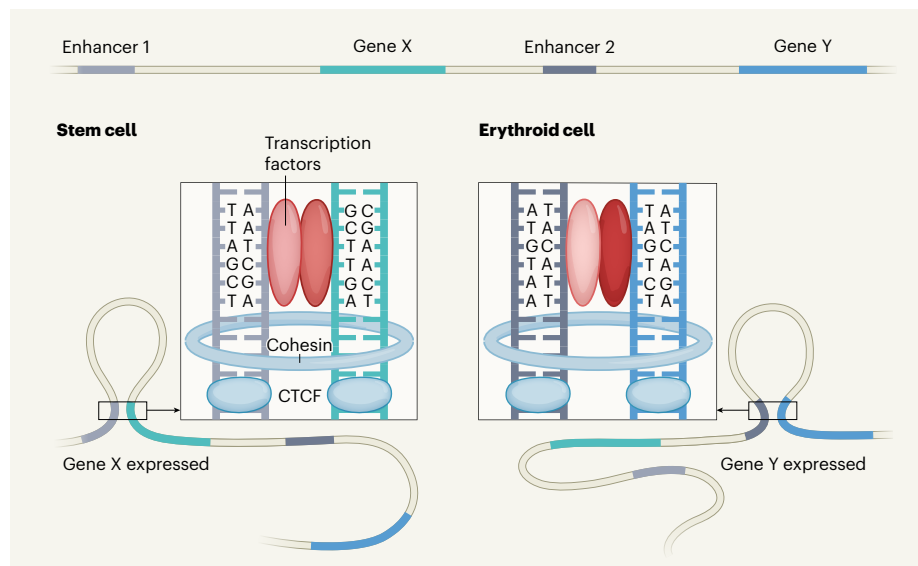


Figure 1 | Monitoring genomic folding. Hua *et al.*¹ have developed a new version of a method termed chromosome conformation capture (3C), called Micro-Capture-C (MCC). This method can identify regions of interacting DNA that are far apart in the linear genome sequence, such as enhancers (regulatory sequences that can promote gene expression) and genes. MCC can pinpoint interactions between base pairs of DNA from different parts of the genome, which is substantially more precise than was previously possible for other types of 3C. The authors used MCC to study stem cells and erythroid cells (precursor red blood cells) in mice. Their findings provide evidence for distinct base-pair patterns of gene and enhancer interactions in different cell types. The results are consistent with a model in which DNA from distant genomic locations is brought into close proximity by a ring-shaped protein complex called cohesin, which helps to generate DNA loops¹⁰. The DNA-binding protein CTCF organizes these loops into ‘insulated’ genomic neighbourhoods, within which interactions occur¹⁰. The ability to identify sites of DNA interactions at the level of individual bases – adenine (A), thymine (T), guanine (G) or cytosine (C) – sheds light on how DNA-binding proteins called transcription factors control gene expression.

expressed) over more-condensed DNA.

The DNA fragments corresponding to the ligated regions of interacting DNA are short and the full sequence of the fragment can be determined, which means that the exact position of the ligation junction is known for each captured interaction. MCC therefore enables the base pairs exactly at the ligation junction to be identified as the interacting regions. This offers a huge leap forward in terms of resolution. Hua and colleagues’ approach also enables DNA-binding-protein ‘footprints’ (the DNA sites to which such proteins bind) to be detected because DNA that is bound to proteins is protected from digestion by MNase.

In addition to introducing us to this exciting technique, the authors immediately put MCC to use in investigating several fundamental aspects of 3D genome folding using embryonic stem cells and precursor red blood cells from mice. Remarkably, interactions between enhancers, genes and CTCF-binding sites occurred as highly localized signals (sharp peaks) in the data for DNA-interaction sites, rather than as broader regions of interaction, as is typical for earlier forms of 3C. Consistent with previous observations¹³, such discrete interactions involving genes almost always (around 87% of the time) occurred in TADs. The precise contacts revealed by MCC were often cell-type-specific, and were associated with the binding of transcription factors that

are important for shaping the identity of particular cell lineages. If the authors mutated a transcription-factor binding site at the centre of an enhancer–gene interaction site, this resulted in localized loss of an interaction detected by MCC and reduced expression of the gene, compared with cells in which the transcription-factor binding site was intact. These findings suggest that transcription factors are responsible for maintaining highly specific 3D genome-folding patterns that are involved in transcriptional control.

The genomic locations of the binding sites for CTCF are mostly the same in different cell types. This raises the question of how loop extrusion orchestrated by CTCF and cohesin contributes to tissue-specific DNA interactions. Hua *et al.* report that contacts between CTCF-bound DNA regions were increased when the intervening regions of DNA had greater numbers of actively transcribed genes and enhancers. The authors demonstrate that both cohesin and a protein called Nipbl, which can load cohesin onto DNA, were enriched at active genes and their enhancers, compared with their presence at less active genes and their enhancers. These data support a model in which cell-type-specific loading of cohesin onto active genes and enhancers aids loop extrusion towards cell-type-invariant CTCF ‘roadblock’ sites – an idea that fits well with previous observations^{14–16} that cohesin

aids enhancer–gene interactions.

Although, at first glance, the individual technological innovations in the MCC method might not seem revolutionary, when combined, they offer something the field has long been waiting for: a way to precisely detect which DNA bases mediate long-range genomic interactions. This level of detail will enable high-resolution dissection of processes involving gene regulation, including those found in complex genomic regions containing multiple genes and regulatory elements, and where enhancer–gene interactions occur over short ranges (less than 20 kilobases of DNA). Although Hua and colleagues’ method does not allow genome-scale analyses, the approach might be adapted to make this possible in the future. Moreover, the base-pair resolution that MCC offers makes it an attractive tool for investigating how regulatory proteins set up and maintain 3D genome architecture.

MCC is also ideally suited to the search for links between disease-associated genetic-sequence variants in regulatory elements and their target genes. Given that such variation can disrupt the binding of transcription factors and often has subtle effects on gene expression, the quantitative nature and footprinting capacity of MCC would be extremely valuable for investigations of this kind.

Anne van Schoonhoven and Ralph

Stadhouders are in the Departments of Pulmonary Medicine and of Cell Biology, Erasmus University Medical Center, Rotterdam 3015 GD, the Netherlands. e-mail: r.stadhouders@erasmusmc.nl

- Hua, P. *et al.* *Nature* **595**, 125–129 (2021).
- The ENCODE Project Consortium *et al.* *Nature* **583**, 699–710 (2020).
- Lambert, S. A. *et al.* *Cell* **172**, 650–665 (2018).
- de Laat, W. & Duboulet, D. *Nature* **502**, 499–506 (2013).
- Stadhouders, R., Filion, G. J. & Graf, T. *Nature* **569**, 345–354 (2019).
- Furlong, E. E. M. & Levine, M. *Science* **361**, 1341–1345 (2018).
- Schoenfelder, S. & Fraser, P. *Nature Rev. Genet.* **20**, 437–455 (2019).
- de Wit, E. & de Laat, W. *Genes Dev.* **26**, 11–24 (2012).
- Bonev, B. & Cavalli, G. *Nature Rev. Genet.* **17**, 661–678 (2016).
- Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. *Cold Spring Harb. Symp. Quant. Biol.* **82**, 45–55 (2017).
- Davies, J. O. J. *et al.* *Nature Methods* **13**, 74–80 (2016).
- Krietenstein, N. *et al.* *Mol. Cell* **78**, 554–565 (2020).
- Downen, J. M. *et al.* *Cell* **159**, 374–387 (2014).
- Kagey, M. H. *et al.* *Nature* **467**, 430–435 (2010).
- Chien, R. *et al.* *J. Biol. Chem.* **286**, 17870–17878 (2011).
- Liu, N. Q. *et al.* *Nature Genet.* **53**, 100–109 (2021).

The authors declare no competing interests.

This article was published online on 9 June 2021.