S7

Chromosome conformation capture techniques

In a non-dividing cell at interphase, each chromosome is a roughly globular mass of folded chromatin that occupies a particular 'territory' in the nucleus. Chromosome conformation capture (CCC) assays aim to determine which sites in this mass of chromatin are in close proximity to each other and in this way can reveal details about chromosome structure that cannot be determined simply from isolated DNA. These techniques have confirmed physical associations between promoters and distant control regions, for example, and have uncovered conserved structural features in chromosomes such as chromatin loops and the topologically associating domains (TADs) described in Section 8.2.

One early use of a CCC technique was to confirm that expression of the β-globin gene in mouse embryonic liver was due to chromosome looping that brought the β-globin gene promoter into close proximity with control sites in the locus control region (LCR) some 60–70 kb away (see Fig. 8.21). The interaction frequencies of DNA sub-regions on average 20 kilobases long in a 200-kb DNA fragment containing the β-globin gene and the LCR were compared in liver and in brain tissue (which does not express the β-globin gene), from mouse embryos. In the liver, a DNA region containing the LCR showed a higher frequency of interaction with a region containing the β-globin gene promoter and enhancer than with contiguous regions, and this higher-frequency interaction was missing in the brain.

Since the first CCC technique was devised in 2002, various versions have been developed that can pick up greater numbers of interactions. The most recent development is the Hi-C technique, which was used to generate the interaction maps shown in Fig. 8.8. This version of CCC takes advantage of the most recent DNA sequencing methods to map all-to-all interactions across large regions of chromosome and even whole genomes. Fig. S7.1 illustrates the basic principles underlying all CCC techniques.

The results of Hi-C, which can produce billions of fragments over a whole genome, are conventionally displayed as a distance matrix or 'heat map' of interaction frequencies, in which the interaction between two sites on the chromosome is depicted as a coloured pixel, with the depth of colour reflecting the frequency of interaction (Fig. S7.2). In the heat map shown here, for example, which depicts one of the TADs adjoining the mouse Hoxa locus (see Fig. 8.8), the interaction frequency of two sequences located at a and b along the linear chromosome map is indicated by the colour of the pixel at c. Chromosomal regions in which there are a higher frequency of interactions between sites than would be expected from their linear distances from each other along the chromosome show up as distinct areas, from which the typical 'mountain-peak' depiction of a TAD is derived.

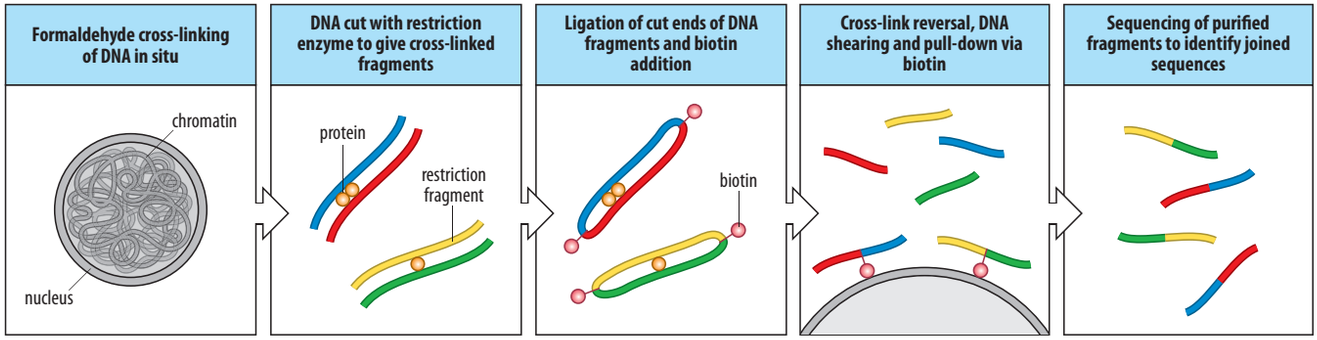

Fig. S7.1 The principles of chromosome conformation capture. Chromatin is first chemically cross-linked, or fixed, *in situ* in living cells or in isolated nuclei to capture the folded conformation of the chromatin within the chromosomes. Fixation with formaldehyde links proteins to nearby proteins and to DNA, and so stabilizes and captures interactions between proteins that bring two sites in the DNA together. The fixed chromatin is then treated with a restriction enzyme that will cut the DNA at specific sequences so that many DNA fragments are produced. After cutting, sites of interaction are represented by two fragments of DNA held together by the cross-links and the interacting proteins. The ends of these fragments are then enzymatically ligated to each other, in some cases incorporating a biotin label to aid subsequent isolation; cross-links and proteins are removed; and the DNA is purified to give DNA fragments that contain the two sites that were in close physical proximity in the nucleus. The DNA fragments are then identified by direct sequencing or by other methods, and the interacting sites identified by reference to the genome sequence. As CCC techniques rely on restriction enzymes that cut the DNA at known sequences and identification of fragments by their DNA sequence, they can only be applied to genomes or regions of genomes whose DNA sequence is known.

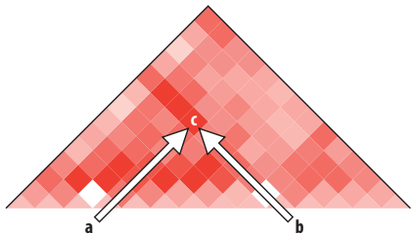

Fig. S7.2 A typical distance matrix or 'heat-map' of interactions along a chromosome identified by Hi-C. The interaction frequency of two sequences located at a and b along the linear chromosome map is indicated by the colour of the pixel at c.

Further reading

Davies, J.O.J., Oudelaar, A.M., Higgs D.R., Hughes, J.R.: **How best to identify chromosomal interactions: a comparison of approaches.** *Nat. Meth.* 2017, **14** 125–134.

Dekker, J., Rippe, K., Dekker, M., Kleckner, N.: **Capturing chromosome conformation.** *Science* 2002, **295** 1306–1210.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y.Shen, I., Hu, M., Liu, J.S., Ren, B.: **Topological domains in mammalian genomes identified by analysis of chromatin interactions.** *Nature* 2012, **485** 376–380.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al*. **Comprehensive mapping of long-range interactions reveals folding principles of the human genome.** *Science* 2009, **326** 289–293.

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., de Laat, W.: **Looping and interaction between hypersensitive sites in the active β-globin locus.** *Mol. Cell* 2002, **10** 1453–1465.