An evaluation of 3C-based methods to capture DNA interactions

Marieke Simonis[1,2], Jurgen Kooren[1,2] & Wouter de Laat[1]

The shape of the genome is thought to play an important part in the coordination of transcription and other DNA-metabolic processes. Chromosome conformation capture (3C) technology allows us to analyze the folding of chromatin in the native cellular state at a resolution beyond that provided by current microscopy techniques. It has been used, for example, to demonstrate that regulatory DNA elements communicate with distant target genes through direct physical interactions that loop out the intervening chromatin fiber. Here we discuss the intricacies of 3C and new 3C-based methods including the 4C, 5C and ChIP-loop assay.

3C technology was originally developed to study the conformation of a complete chromosome in yeast[1] and was subsequently adapted to investigate the folding of complex gene loci in mammalian cells[2]. It has now become a standard research tool for studying the relationship between nuclear organization and transcription in the native cellular state. Other technologies based on the 3C principle have been developed that aim to increase the throughput. 4C technology allows for an unbiased genome-wide screen for interactions with a locus of choice, whereas 5C technology permits parallel analysis of interactions between many selected DNA fragments. ChIP-loop methodology combines 3C with chromatin immunoprecipitation to analyze interactions between specific protein-bound DNA sequences. Detailed protocols that should help researchers setting up 3C[3–5] and 5C[6] technology in their own laboratory and an excellent review[7] explaining the controls necessary for correct interpretation of 3C results have been published. Here we present a detailed 4C procedure as a **Supplementary Protocol** online.

Common principles

In brief, the 3C procedure involves five experimental steps (**Fig. 1**). First, cells are fixed with formaldehyde, which cross-links proteins to other proteins and to DNA segments that are in close proximity in the nuclear space. Second, the cross-linked chromatin is digested with an excess of restriction enzyme, separating cross-linked from non-cross-linked DNA fragments. Third, DNA ends are ligated under conditions that favor junctions between cross-linked DNA fragments. Fourth, cross-links are reversed. Finally, ligation events between selected pairs of restriction fragments are quantified by PCR, using primers specific for the fragments being studied.

The technique allows the identification of physical interactions between distant DNA segments and of chromatin loops that are formed as a consequence of these interactions, for example between transcriptional regulatory elements and distant target genes[2,8–11]. 3C technology is particularly suited to study the conformation of genomic regions that range in size roughly from five to several hundred kilobases (kb) in size. To our knowledge, the smallest region studied so far by 3C technology spans 6,700 base pairs (bp)[12], whereas the largest region analyzed spans ~600 kb[13]. It is important to note that because of the flexibility of the chromatin fiber, DNA segments on the same fiber are engaged in random collisions, with a frequency inversely proportional to the genomic distance between them. Therefore, the mere detection of a ligation product does not necessarily reveal a specific interaction. To ascertain that an interaction is specific requires the demonstration that two DNA sites interact more frequently with each other than with neighboring DNA sequences. Thus, 3C tech-

[1]Department of Cell Biology and Genetics, Erasmus MC, PO Box 2040, 3000 CA, Rotterdam, The Netherlands. [2]These authors contributed equally to this work. Correspondence should be addressed to W.d.L. (w.delaat@erasmusmc.nl).

PUBLISHED ONLINE 30 OCTOBER 2007; DOI:10.1038/NMETH1114

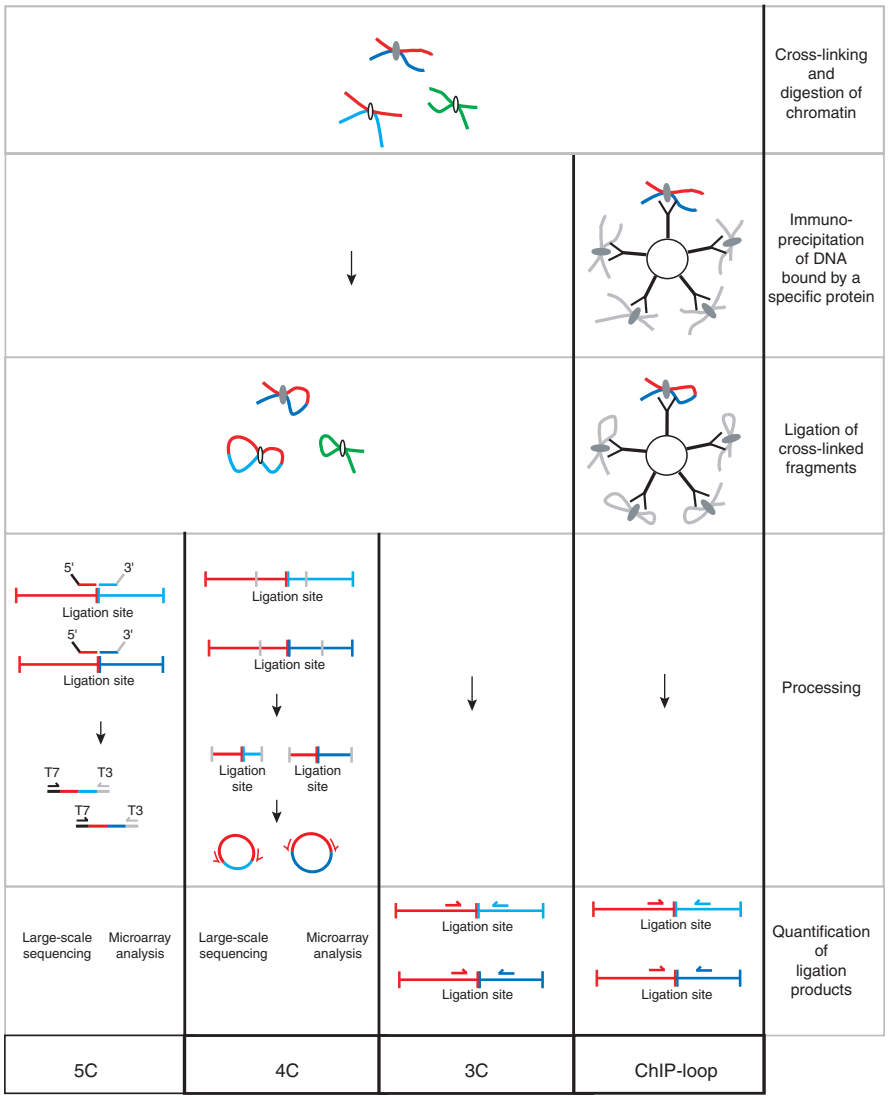

Figure 1 | Schematic representation of 3C-based methods. In 3C, 4C and 5C methods, DNA interactions are captured by formaldehyde treatment, DNA digestion with a restriction enzyme and ligation of cross-linked fragments, and then ligation frequencies are measured. In the ChIP-loop assay, immunoprecipitation enriches the sample for fragments bound by a specific protein, and restriction fragments are ligated to each other on the beads. In ChIP-loop and 3C, ligation frequencies are measured by quantitative PCR, using a unique primer set for each ligation junction analyzed. In 5C, ligation events are amplified by ligation-mediated amplification with T7 and T3 primers, and then analyzed by large-scale sequencing or microarray. In 4C, ligation junctions are first trimmed by a frequently cutting secondary restriction enzyme, then subjected to ligation to form circles and inverse PCR to amplify captured fragments. If a frequently cutting enzyme is used in the first digestion, the second digestion can be omitted (see **Fig. 3**). The 4C PCR product is analyzed by large-scale sequencing or microarray analysis.

nology is a quantitative assay, and a meaningful analysis critically relies on an accurate comparison of interaction frequencies between multiple DNA segments.

3C and 3C-based technologies provide information about the frequency, but not the functionality, of DNA interactions. Thus, additional, often genetic, experiments are required to address whether an interaction identified by 3C-based technologies is functionally meaningful. For example, many of the interactions identified by 4C technology[14] between genomic regions far apart on the same chromosome or on other chromosomes may well be nonfunctional and merely the consequence of general folding patterns of chromosomes[15].

During most of the cell cycle, one mammalian cell provides maximally two events for 3C analysis, in that it contains only two copies of a given restriction fragment, each end of which can be ligated to maximally one other restriction site during the 3C procedure. This implies that a meaningful (that is, quantitative) 3C PCR analysis must be performed on a DNA template that represents many genome equivalents. It also implies that DNA interactions can be quantified accurately only if they occur in a substantial proportion of the cells. Sites separated by large genomic distances (hundreds of kilobases or more) or present on different chromosomes often do not form enough ligation products for accurate quantification, even if microscopy studies suggest that they come together in a substantial proportion of cells. To study such long-range interactions, we recommend using high-throughput 4C technology.

Below, a more detailed outline of the experimental steps involved in all 3C-based technologies will be presented to allow a better appreciation of the potentials and limitations of these methods.

Common experimental steps in 3C-based technologies

Step 1: formaldehyde cross-linking. Formaldehyde is used to cross-link protein-protein and protein-DNA interactions by means of their amino and imino groups. Advantages of this cross-linking agent are that it works over a relatively short distance (2 Å) and that cross-links can be reversed at higher temperatures[16–18]. Although cross-linking is sometimes performed on isolated nuclei, it is preferentially done on living cells, because this better guarantees that a faithful snapshot of the chromatin conformation will be obtained. Routinely, cells are cross-linked at room temperature (18–22 °C) for 10 min, using a formaldehyde concentration of 1–2%, but optimal fixation conditions depend on the frequency and stability of the interactions analyzed and have to be redefined for every new 3C experiment. Many 3C experiments demonstrate preferential interactions between transcription regulatory DNA elements. These sites are known to carry transcription factors and often contain fewer histone proteins, hence their hypersensitivity to nuclease digestion. A concern often raised is that the 3C assay may be biased because of better cross-linking ability of these sites. However, evidence that the contrary may be true comes from recently developed formaldehyde-assisted isolation of regulatory elements (FAIRE)[19,20]. FAIRE involves phenol-chloroform extraction of formaldehyde–

cross-linked and sonicated chromatin and isolates regulatory DNA sequences based on their tendency to end up in the aqueous phase more than other genomic regions (as they contain fewer histones, they are less cross-linkable to proteins).

Formaldehyde is also used under similar experimental conditions in chromatin immunoprecipitation (ChIP) experiments as the cross-linking agent that captures protein-DNA interactions. It is conceivable that formaldehyde often produces complex aggregates containing more than two DNA fragments. In support of this, it has been found that a single restriction fragment frequently captures two or more other restriction fragments together in a 4C experiment[21]. This notion would imply that both ChIP and 3C-like technologies also pick up indirect interactions.

Step 2: restriction enzyme digestion. After cross-linking, nuclei are isolated and digested with a restriction enzyme. The choice of restriction enzyme will mainly depend on the locus to be analyzed. The restriction enzyme should dissect the locus such that it allows for the separate analysis of the relevant regulatory elements (gene bodies, promoters, enhancers, insulators and so on). Analyzing the topology of small loci (<10–20 kb) requires the use of frequently cutting restriction enzymes such as *Dpn*II or *Nla*III (four-base cutters; hereafter called four-cutters). In analyzing larger loci, six-base cutters (six-cutters) can also be used. Not all enzymes digest cross-linked DNA equally well, and we prefer to use *Eco*RI, *Bgl*II or *Hin*dIII[3]. When we digest overnight with a large excess of one of these restriction enzymes, we do not observe notable preferential digestion for specific regions of the genome such as open chromatin. This may be different with different enzymes and conditions though, and we recommend that for each new 3C experiment one rule out the possibility of a bias in the assay owing to preferential digestion of some sites over others. Digestion efficiency also decreases with increasing cross-linking stringency[3]. We recommend that at least 60–70% of the DNA, but preferably 80% or more, be digested before continuing with the ligation step.

Step 3: ligation. A critical selective step in the procedure is the ligation step carried out under conditions that favor intramolecular ligation events between cross-linked DNA fragments. This step creates the actual 3C library that is enriched for ligated junctions between DNA fragments that originally were close together in the nuclear space. It is relevant to know how frequently a given ligation occurs. We have carefully quantified the abundance of the most frequently formed ligation products (**Fig. 2**). Independently of the restriction site analyzed, two types of junction are always over-represented. The first most abundant junction is with the neighboring DNA sequence. This junction is the result of incomplete restriction enzyme digestion and can constitute up to 20–30% of all the junctions; this number drops when less stringent fixation conditions are used. The second most abundant junction is with the other end of the same restriction fragment, as a consequence of restriction fragment circularization. This product can be formed independently of the cross-linking step and can account for up to 5–10% of all the junctions formed. Interestingly, this percentage goes up when less stringent cross-linking conditions are used (data not shown), suggesting that under such conditions fewer restriction fragments are cross-linked together. The formation of other junctions is much less efficient. For example, ligation to ends of directly neighboring restriction fragments (which will always be close together in the nuclear space and therefore should also ligate relatively efficiently) already occurs only 0.2–0.5% of the time. This percentage quickly drops to <0.1% with increasing genomic site separation, unless two sites are engaged in a specific interaction. However, even sites thought to frequently interact with each other, such as sites in the β-globin locus control region and the active β-globin genes 30–50 kb away, account for only 0.2–0.5% of the junctions formed with each of them. It is therefore clear that to accurately quantify such rare events that often occur in less than 1 in 1,000 cells, it is necessary to include many genome equivalents in a PCR reaction.

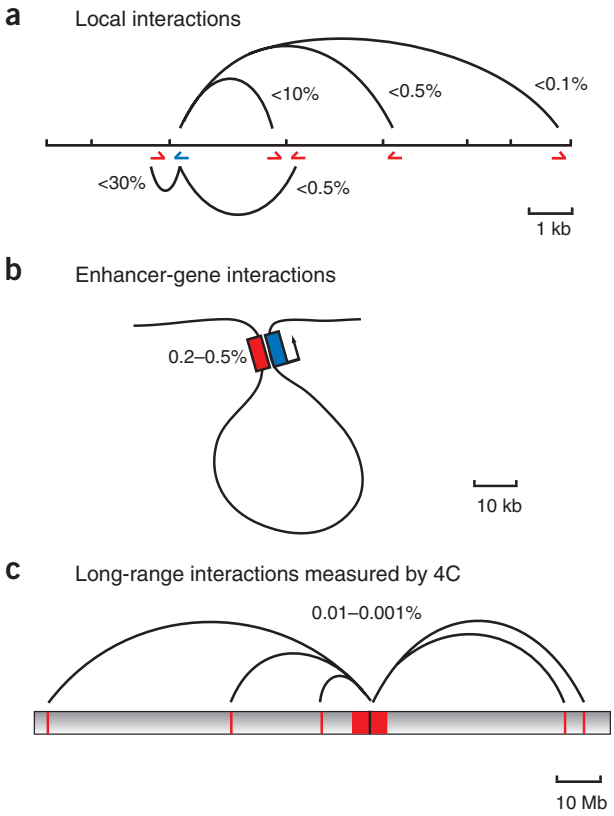

Figure 2 | Ligation events measured at the β-globin locus. (**a**–**c**) The different types of interactions at the mouse β-globin locus are local interactions between neighboring restriction fragments (**a**), enhancer-gene interactions over 30–100 kb (**b**) or long-range interactions in *cis* and *trans* (>1 Mb; **c**). Typical values for ligation frequencies (in % alleles) of a 'bait' restriction fragment end with a given other restriction fragment end are indicated. Arrows below the restriction fragments indicate the location and direction of 3C primers (bait primer indicated in blue; **a**). Ligation frequencies measured using 3C-qPCR (**a**,**b**). Ligation frequencies estimated from 4C data (**c**).

Final steps: PCR in 3C. After reversal of the cross-links, ligation frequencies of restriction fragments are analyzed by PCR, using primers specific for the restriction fragments of interest. We routinely use 50–200 ng of 3C template, or ~8×10^4–3×10^5 genome equivalents, per PCR reaction. A meaningful 3C analysis critically relies on the accurate quantification of the different ligation products, and measurements therefore have to be taken when each DNA amplification reaction is in the linear range. The standard 3C PCR protocol uses a standard number of PCR cycles and a standard amount of DNA template for the analysis of all different ligation

products. This approach is only semiquantitative and prone to inaccuracies, in that measurements may be taken outside the linear range of the amplification reaction. To overcome this limitation, a real-time PCR approach using TaqMan probes, called 3C-qPCR, was developed[22,23]. A single probe and fixed PCR primer are used that hybridize to opposite strands of the restriction fragment of interest and work in combination with a series of test PCR primers hybridizing to other restriction fragments. This configuration ensures that the fluorescent signal provided by the probe is strictly specific to the amplification of the ligation product selected for analysis[5].

Different primer pairs will have different amplification efficiencies; to account for this, these efficiencies must be assessed. This is done on a control template containing all ligation products in equimolar amounts[1–3,7,8], mixed with the same amount of genomic DNA as is present in the 3C PCR reaction. To account for possible differences in quality and quantity between 3C templates, interaction frequencies are analyzed between segments in a control locus that is expected to adopt a similar conformation in the different cell types of interest[2,3,7].

The ChIP-loop assay

It is often found by 3C technology that, in the population of cells analyzed, a single DNA site interacts with multiple other sites. In many cases this is likely to reflect cell-to-cell differences in chromatin conformation, and it is quite possible that in different subpopulations of cells distinct proteins bind to such a given site and mediate the different DNA interactions. The ChIP–combined loop (ChIP-loop) assay was developed to investigate this[24–26]. The method involves formaldehyde cross-linking of cells, restriction enzyme digestion and urea gradient purification of cross-linked chromatin, immunoprecipitation using an antibody against the protein of interest, ligation of precipitated DNA fragments (still coupled to the beads) and PCR analysis of the junctions (**Fig. 1**).

In our opinion, several technical aspects complicate the analysis of results obtained by ChIP-loop. First, current protocols ligate the fragments when they are bound and concentrated to the beads. Concentrating the DNA on the beads before ligation is expected to facilitate the formation of junctions between bead-associated, but not necessarily formaldehyde–cross-linked, DNA fragments, hence also producing results that reflect loops formed on the beads rather than in nuclear space. Unless the user can demonstrate that such undesired events do not take place, we would argue that it is better to carry out the precipitation after the ligation step, which has to be performed under conditions described in the 3C procedure.

Second, accurate quantification of ligation products, already very challenging in standard 3C, is even more complicated in ChIP-loop assays because it must take into account the relative enrichment of each site on the beads. For example, we would argue that ChIP-loop assays should only be applied to the analysis of fragments that are both enriched by ChIP. Indeed, we tend to question the relevance of analyzing, via ChIP-loop, interactions between DNA segments that are not bound by the protein of interest, or between DNA segments of which only one is enriched by the antibody. If a sequence is coprecipitated because it is cross-linked to a target sequence of the protein of interest, it should also be found enriched in the ChIP assay.

It may be possible to obtain unique information, not obtainable from ChIP or 3C only, when studying loops formed between sites that are both precipitated because of their association to a protein of interest. As in 3C, however, the mere detection of a ligation product may reflect a random collision rather than a specific interaction. Therefore, interaction frequencies have to be quantified accurately and compared to other interactions, which requires taking into account the genomic site separation between each pair of segments and the relative enrichment of each site on the beads. ChIP-loop assays can be useful to identify proteins participating in long-range interactions in *cis* (that is, over hundreds of kilobases or more) or in *trans*, because interpretation of these results will not be complicated by frequent random collisions.

5C technology

Large-scale mapping of, for example, several hundred chromatin interactions using standard 3C is time-consuming and difficult. The introduction of the 3C–carbon copy (5C) method generates the possibility of such large-scale locus-wide analysis[6,27]. The method uses a multiplex ligation-mediated amplification step to amplify selected ligation junctions, thereby generating a quantitative carbon copy of a part of the initial 3C library, which is subsequently analyzed by microarray detection or high-throughput sequencing (**Fig. 1**). Ligation-mediated amplification involves using a combination of test and fixed 5C primers that hybridize to the sense and antisense strand, respectively, of the restriction fragment ends analyzed. Fixed and test primers will be directly juxtaposed when a ligation junction is formed between the corresponding restriction sites, allowing subsequent primer-primer ligation. Universal tails protruding from the test and fixed primers, such as T7 and complementary T3 promoter sequences, subsequently permit massive parallel quantitative amplification of all investigated ligation products. Interestingly, 5C technology provides the opportunity to analyze a locus from a single or multiple fixed points. It can generate a complex matrix of interaction frequencies for a given genomic region, which can be used to reconstruct the intricate topology of this region. However, the size of the genomic region that can be studied is limited by the number of 5C primers that can be used simultaneously. Scanning hundreds of megabases of the genome will require using tens of thousands of 5C primers, which makes the technology less suitable for genome-wide scans[6].

4C technology

3C and 5C technologies have been developed to identify interacting elements between selected parts of the genome, and both techniques require the design of primers for all restriction fragments analyzed. Recently, new related strategies, collectively referred to as 4C technology, have been developed that allow screening the entire genome in an unbiased manner for DNA segments that physically interact with a DNA fragment of choice[14,21,23,28].

An outline of 4C technology is provided in **Figure 3**. Like 3C, 4C technology depends on the selective ligation of cross-linked DNA fragments to a restriction fragment of choice (the 'bait'). In 4C technology, all the DNA fragments captured by the bait are simultaneously amplified via inverse PCR, using two bait-specific primers that amplify from circularized ligation products. Essentially two strategies can be pursued to obtain these DNA circles (**Fig. 3**). One strategy relies on the formation of circles during the standard 3C ligation step, that is, while the DNA is still cross-linked[21,28]. Here circle formation requires both ends of the bait fragment to be ligated to both ends of a captured restriction fragment. After de-cross-linking, captured DNA fragments are directly amplified by inverse PCR, using

bait-specific primers facing outward. Four-cutters are preferred in this method[21], because they produce smaller restriction fragments (average size 256 bp, versus ~4 kb for six-cutters) and linear PCR amplification of the captured DNA fragments requires that the average product size be small.

The second strategy relies on the formation of DNA circles after the chromatin has been de-cross-linked. Here the standard 3C procedure is followed, using a six-cutter as the restriction enzyme and yielding a de-cross-linked 3C template. The ligation junctions are then trimmed using a frequently cutting secondary restriction enzyme and re-ligated under conditions that favor the formation of self-ligated circles. Inverse PCR primers hybridizing to the bait are used to linearly amplify (the small outer ends of) captured DNA fragments[14,23]. Because the two strategies have not yet been worked out in similar detail, it is currently difficult to compare them. Theoretically, though, each strategy will have its own advantages and disadvantages.

The first approach presents the advantage of requiring fewer processing steps, and the use of four-cutters provides a higher resolution (256 bp for a four-cutter versus 4 kb for a six-cutter). This should allow for a better definition of the site of interaction, which is expected to be particularly useful for identifying *cis*-regulatory DNA elements that locate away from a gene of interest. A potential concern exists if formaldehyde cross-links multiple DNA fragments together. As a consequence of this, circles formed between cross-linked DNA fragments may contain more than two captured fragments, which will often be too large to be amplified in a linear fashion. This, in turn, may affect the reproducibility of the approach. It is also not clear how efficient circle formation is between DNA fragments that reside in cross-linked chromatin aggregates. Analysis of fragments captured by this approach has so far been limited to the sequencing of relatively small numbers of clones, 114 (ref. 21) and 320 (ref. 28), respectively. Although these studies identified interesting DNA fragments, it is not clear whether such small numbers of clones provide a fair representation of the complex library of ligation junctions.

The advantage of the second approach is that it depends on the ligation of only one end of the bait to one end of a cross-linked DNA fragment, which will be more efficient than forming a circle between cross-linked DNA fragments. Circle formation takes place when the DNA is naked, which will also be more efficient than when the DNA is cross-linked. Products to be amplified will generally be smaller, because the circles will not contain more than one captured fragment and will therefore be easier to amplify in a linear fashion. Because the strategy selectively amplifies the ligated outer ends of the restriction fragments created by the six-cutter, the complexity of the genomic library to be analyzed is markedly reduced. One can take advantage of this by designing tailored microarrays containing only probes located directly adjacent (within 100 bp) to each recognition site of a given six-cutter (for example, *Hin*dIII) in the genome to analyze the captured DNA fragments[14]. This design allows a large representation of the genome to be spotted on a single array. In fact, current designs cover the complete human or mouse genome on a single Nimblegen microarray (400,000 probes), allowing the identification of interactions at a resolution of ~7 kb (unpublished data).

Tailored microarrays were used to simultaneously analyze hundreds of thousands of fragments captured by the second approach. Replicate experiments performed on biologically independent samples demonstrated that this strategy is highly reproducible. Whatever the bait chosen for analysis—and we have now analyzed interactions with more than 15 different baits—it is always found that sequences physically close on the linear chromosome template are largely overrepresented (**Fig. 4a**). In fact, restriction fragments within 5–10 Mb from the bait are always captured so efficiently that they saturate every corresponding probe present on the array, precluding a quan-

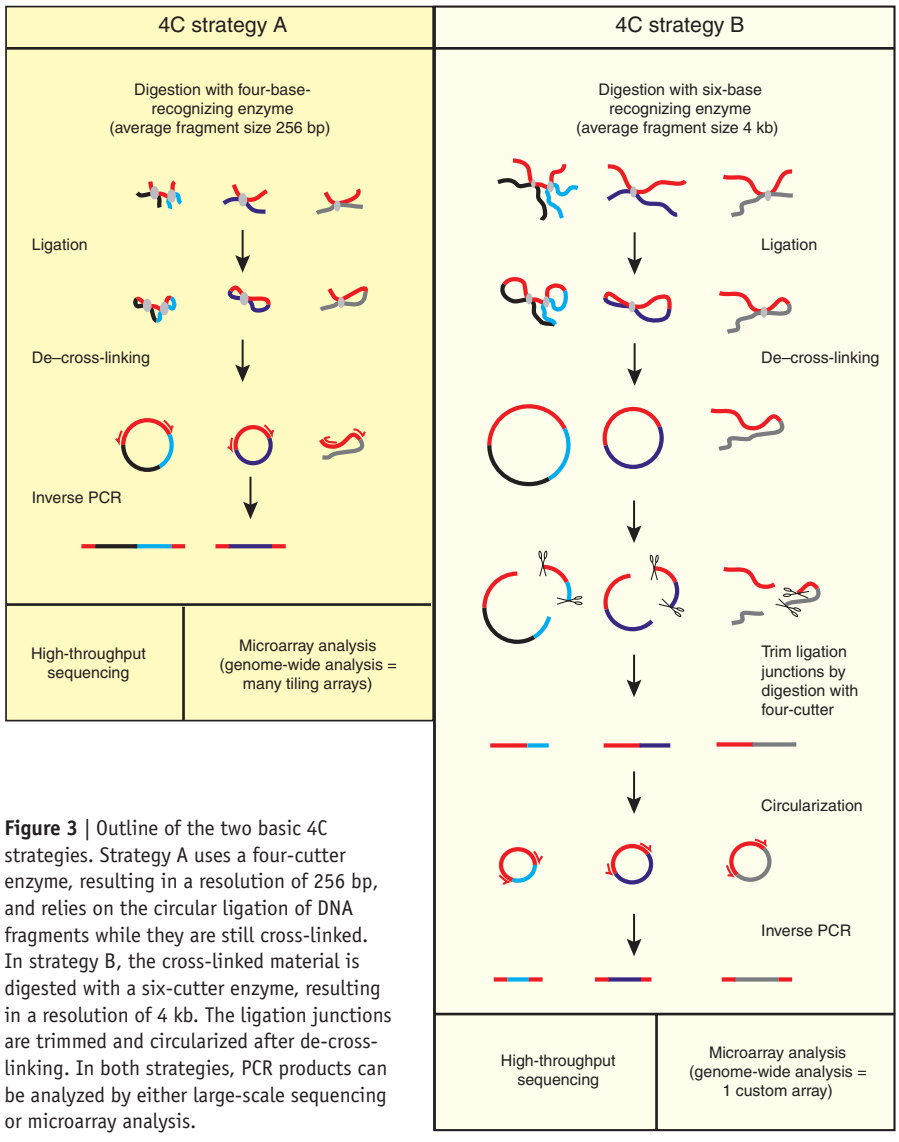

Figure 3 | Outline of the two basic 4C strategies. Strategy A uses a four-cutter enzyme, resulting in a resolution of 256 bp, and relies on the circular ligation of DNA fragments while they are still cross-linked. In strategy B, the cross-linked material is digested with a six-cutter enzyme, resulting in a resolution of 4 kb. The ligation junctions are trimmed and circularized after de-cross-linking. In both strategies, PCR products can be analyzed by either large-scale sequencing or microarray analysis.

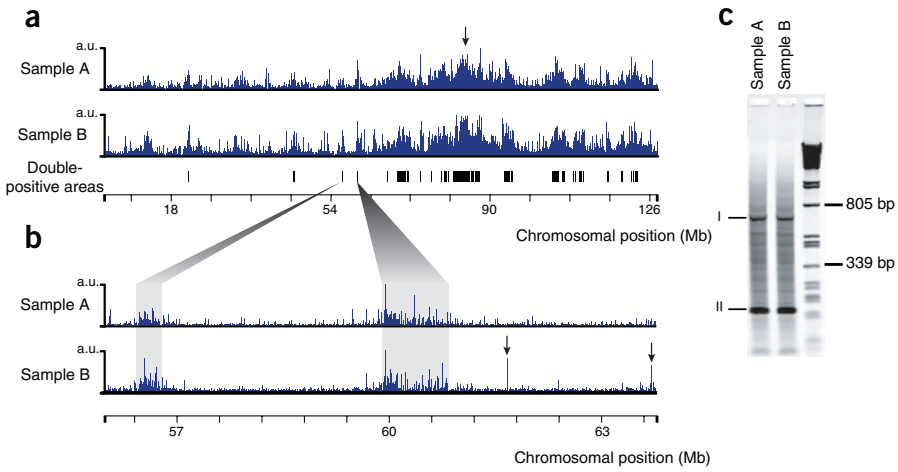

Figure 4 | Results obtained by 4C technology are highly reproducible. Example of 4C data analyzing mouse *Rad23a*. (**a**) Unprocessed data on the *cis* chromosome of two independent experiments on biological replicates. The arrow indicates the position of the 4C primers. (**b**) Unprocessed data from two independent experiments showing reproducible clustering of high signals. Arrows indicate irreproducible, isolated high signals, representing random ligation events. (**c**) Two 4C PCR products of biological replicates analyzed by gel electrophoresis. The appearance is highly reproducible. I indicates the self-circularized fragment; II indicates the undigested product.

titative analysis of local signal intensities. Farther away from the bait and on other chromosomes, clusters of 20–50 neighboring restriction fragments can be identified that all show increased hybridization signals (**Fig. 4b**). Because each probe analyzes an independent ligation event and only two fragments can be captured per cell, such clustering of interacting DNA fragments strongly indicates that this genomic region contacts the bait in multiple cells. Notably, high-resolution cryo-fluorescence *in situ* hybridization (cryo-FISH) confirmed in an independent manner for more than 20 of these regions that they truly represent interacting regions in *cis* and in *trans*[14]. These experiments also showed that 4C technology identifies *trans*- and *cis*-interacting regions even if they are together in only 4% and 6% of the cells, respectively (cryo-FISH background: <2% in *trans* and <4% in *cis*).

Potential pitfalls of 4C technology

Number of cells. Whichever strategy is followed, several critical steps have to be considered. First, the analysis must be performed on a relatively large population of cells. Even frequent interactions between fragments close together on the linear chromosome template often are captured in less than 1 in 500 cells, and we think that the *trans* and long-range *cis* interactions that we identify are captured in only 1 in 10,000 or even 1 in 100,000 cells. We routinely process 10 million cells and perform 16 inverse PCR reactions on 200 ng of template, which we subsequently pool and label for microarray hybridization. Hence, we analyze an equivalent of approximately 1 million interactions on a single microarray.

PCR. The advantage of 4C (and 5C) over 3C is that only two primers are required to amplify all products, circumventing the problem of differences in primer pair efficiencies. All PCR-based methods have the limitation that different amplicons amplify with a different efficiency. By performing the same PCR on a control template containing all ligation products in equimolar amounts, one can correct for these differences in 3C and 5C, but not in 4C. It is absolutely critical to optimize the 4C PCR step, because this step will select the DNA fragments for analysis, which need to correctly represent the fragments captured by the bait.

Typically, 80% of the DNA fragments are smaller than 600 bp when samples are processed first with a six-cutter and then with a four-cutter, but one also wants larger fragments to be amplified in a linear fashion. Different polymerases will perform this task with different levels of success (data not shown). One can use 3C primers and real-time PCR to test whether the abundance of different-sized products is similar before and after the inverse PCR step in 4C. We have used this strategy to define conditions that allow fragments up to 1.2 kb to be amplified at very similar efficiencies (less than twofold bias after 30 cycles of PCR; see **Supplementary Protocol** online). When separated by gel electrophoresis, biological replicates should give a similar smear of PCR products and several more prominent bands that are reproducible between the samples (**Fig. 4c**). One should also check if the theoretically most abundant products that originate from the undigested template and from the self-ligated circle are prominently present, which also confirms that the inverse PCR works (**Fig. 4c**).

High-throughput analysis. Although sequencing of even hundreds of clones may reveal potentially interesting DNA fragments, we strongly recommend high-throughput analysis of captured DNA fragments, using either microarrays or large-scale sequencing, to exclude the possibility that the analysis is focused on a misrepresentation of the actual library of captured fragments. Indeed, whatever the bait chosen for analysis and whatever the 4C strategy used, the great majority of captured fragments will always be located close to the bait on the linear chromosome template[14,23].

Analyzing 4C data. High-throughput microarray analysis shows that probes with high signals are found across the chromosome and to a lesser extent also on other chromosomes. Many of these captures are random, though, as they are not reproducible between independent duplicate experiments (**Fig. 4b**). Thus, highly specific long-range intra- and interchromosomal interactions with single restriction fragments may exist, but it is very difficult to discriminate them from random captures. The presence of genomic clusters of restriction fragments that show increased hybridization signals in biologically replicate experiments reveals interacting regions, as explained above (**Fig. 4b**). These regions can be identified by the application of a sliding-window approach that provides a measure for the relative abundance of ligated fragments per genomic area[14].

Verification of 4C data. 3C technology may be used as a first verification of data obtained by 4C technology. However, they are not independent technologies, and long-range interactions identified by 4C technology should therefore always be verified by completely

independent methods such as FISH. Preferably this should be done by high-resolution FISH studies, such as 3D-FISH or cryo-FISH[29], that use fixation conditions which preserve the nuclear ultrastructure well. It has to be demonstrated that two regions identified by 4C technology as interacting indeed come together more frequently in the population of cells than two randomly chosen loci.

Concluding remarks and perspectives

The development of 3C technology has contributed enormously to our understanding of the intricate folding of gene loci and has revealed, for example, that transcriptional regulatory DNA elements loop toward their target genes to regulate the expression. On the basis of 3C technology, several new approaches have recently been developed. The ChIP-loop assay may direct structure analysis to specific protein-bound DNA sequences, but correct interpretation is currently still complicated, in that it requires a quantitative comparison between ChIP-loop, ChIP and 3C data. 5C technology is expected to provide unprecedented insight into the conformational fine structure of selected regions in the genome. Like 4C, it may help researchers in screening a genomic region for DNA elements that interact with a DNA segment of choice, such as a gene (promoter), an insulator sequence, an enhancer, an origin of replication and the like. 4C technology is expected to contribute substantially to a comprehensive understanding of nuclear architecture[15], picking up interactions not previously anticipated and putting the relative frequency of interactions in perspective. Current 4C microarray studies allow identification of long-range interactions in *cis* over tens of megabases and in *trans* between chromosomes. The large over-representation of fragments closer to the bait precludes a quantitative analysis of local interactions, but it is to be expected that 4C can be modified to also identify loops formed in smaller genomic regions. In the near future more new 3C-based methods may be expected. Their potential should be evaluated not so much on the exciting nature of the interactions identified but on the independent evidence, obtained for example by FISH, that is provided to demonstrate that interactions are real.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank F. Grosveld and the members of our laboratory for discussion. This work was supported by grants from the Dutch Scientific Organization (NWO) (016-006-026) and (912-04-082) to W.d.L.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
2. 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* **10**, 1453–1465 (2002).
3. Splinter, E., Grosveld, F. & de Laat, W. 3C technology: analyzing the spatial organization of genomic loci *in vivo*. *Methods Enzymol.* **375**, 493–507 (2004).
4. Miele, A., Gheldof, N., Tabuchi, T.M., Dostie, J. & Dekker, J. Mapping chromatin interactions by chromosome conformation capture (3C). in *Current Protocols in Molecular Biology* (eds. Ausubel, F.M. *et al.*) 21.11.1–21.11.20 (John Wiley & Sons, Hoboken, New Jersey, USA, 2006).
5. Hagege, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722–1733 (2007).
6. Dostie, J. & Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nat. Protoc.* **2**, 988–1002 (2007).
7. Dekker, J. The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat. Methods* **3**, 17–21 (2006).
8. Palstra, R.J. *et al.* The β-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**, 190–194 (2003).
9. Murrell, A., Heeson, S. & Reik, W. Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat. Genet.* **36**, 889–893 (2004).
10. Spilianakis, C.G. & Flavell, R.A. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* **5**, 1017–1027 (2004).
11. Liu, Z. & Garrard, W.T. Long-range interactions between three transcriptional enhancers, active Vκ gene promoters, and a 3′ boundary sequence spanning 46 kilobases. *Mol. Cell. Biol.* **25**, 3220–3231 (2005).
12. O'Sullivan, J.M. *et al.* Gene loops juxtapose promoters and terminators in yeast. *Nat. Genet.* **36**, 1014–1018 (2004).
13. Skok, J.A. *et al.* Reversible contraction by looping of the *Tcra* and *Tcrb* loci in rearranging thymocytes. *Nat. Immunol.* **8**, 378–387 (2007).
14. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348–1354 (2006).
15. de Laat, W. & Grosveld, F. Inter-chromosomal gene regulation in the mammalian cell nucleus. *Curr. Opin. Genet. Dev.*, published online 18 September 2007 (doi:10.1016/j.gde.2007.07.009).
16. Solomon, M.J. & Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. *Proc. Natl. Acad. Sci. USA* **82**, 6470–6474 (1985).
17. Orlando, V., Strutt, H. & Paro, R. Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods* **11**, 205–214 (1997).
18. Jackson, V. Formaldehyde cross-linking for studying nucleosomal dynamics. *Methods* **17**, 125–139 (1999).
19. Hogan, G.J., Lee, C.K. & Lieb, J.D. Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet.* **2**, e158 (2006).
20. Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R. & Lieb, J.D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877–885 (2007).
21. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341–1347 (2006).
22. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
23. Wurtele, H. & Chartrand, P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res.* **14**, 477–495 (2006).
24. Horike, S., Cai, S., Miyano, M., Cheng, J.F. & Kohwi-Shigematsu, T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.* **37**, 31–40 (2005).
25. Cai, S., Lee, C.C. & Kohwi-Shigematsu, T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* **38**, 1278–1288 (2006).
26. Kumar, P.P. *et al.* Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat. Cell Biol.* **9**, 45–56 (2007).
27. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
28. Lomvardas, S. *et al.* Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006).
29. Branco, M.R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, e138 (2006).