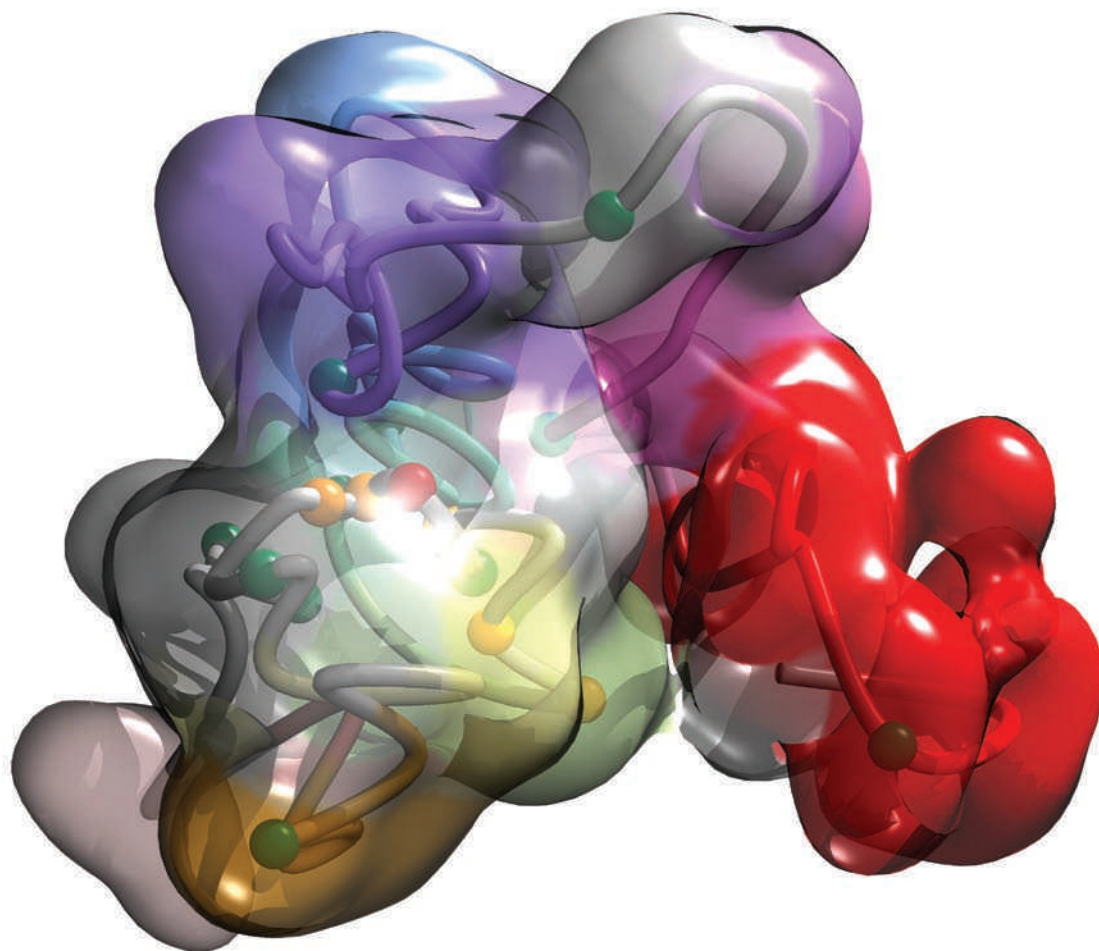# GENOMES IN THREE DIMENSIONS

A DNA sequence isn't enough; to understand the workings of the genome, we must study chromosome structure.

Globular conformation of a 500-kilobase gene-rich domain on human chromosome 16.

BY MONYA BAKER

The next frontier of genomics is space: the three-dimensional structures of chromosomes coiled in the nucleus. Far from being the random result of packing 2 metres of DNA into a sphere perhaps 10 micrometres across, the structures vary across cell types and exert an as-yet-mysterious influence on gene expression. Efforts to decipher the effects of structure face many

difficulties, not least that researchers are still trying to find out how chromosomes shift as cells change, says Thomas Cremer, a geneticist at the Ludwig Maximilian University of Munich in Germany, who has studied the spatial organization of the genome since the 1970s. "The nucleus is still an uncharted landscape and it is embarrassing how little undoubtedly proven knowledge we have about its dynamic topography," he says.

The basics have been known for decades:

DNA double helices coil around proteins called histones, forming 'chromatin' strands that in turn are bundled into chromosomes. But when it came to the twisting and turning of chromosomes themselves, "it wasn't clear what role genome organization was playing or even if there was that much organization", says Peter Fraser, a genome biologist at the Babraham Institute in Cambridge, UK. Long-range interactions seemed implausible. "People assumed that sequences 50 kilobases away couldn't find

each other in the nucleus," he says.

These days, scientists know that such interactions happen all the time. In 2002, Fraser's laboratory was among the first to detect 'long-range looping interactions' that bring gene sequences into physical contact with far-off regulatory elements[1].

More-global changes also occur. For example, inactive chromatin is generally shunted to the nuclear periphery, but that arrangement is inverted in mouse retinal cells, allowing more light to reach photoreceptors[2]. That the spatial organization of the genome is important is also demonstrated by the havoc that alterations can wreak. A cancer of the lymphatic system called Burkitt's lymphoma occurs after a chunk of chromosome 8 ends up on chromosome 14 and vice versa. This happens because of the way that chromosomes arrange themselves in white blood cells[3] — translocations occur more often between genes that physically come together during transcription[4]. Various types of cancer have been found to be connected with mutations in proteins that affect chromatin structure, and researchers have speculated that long-range interactions can be altered by disease-associated mutations in stretches of DNA that do not code for genes.

ANSWERS IN THE STRUCTURE

Researchers have long known that DNA sequences and histones are tagged with chemical modifications that turn genes on and off; the cataloguing of such 'epigenetic' modifications is well under way. It is now becoming clear that the three-dimensional organization of chromatin reflects a higher order of epigenetic regulation, says Yijun Ruan, a biologist at the Genome Institute of Singapore, who has developed techniques to find long-range interactions mediated by specific proteins[5]. Instead of assuming that gene activity is determined entirely by chemical attachments along a linear DNA sequence, researchers are looking for answers in the ways that chromatin folds, moves and communicates.

"You start out with a difficult problem, and you convert it through a series of molecular steps to a simple problem."
Job Dekker

Discussions are beginning to include phrases such as 'chromatin network', 'chromosome interactome' and 'spatial epigenetics'.

A suite of technological innovations is starting to reveal the significance of such concepts. New microscopes are letting researchers look more closely at more nuclei, for example, and experiments are allowing researchers to
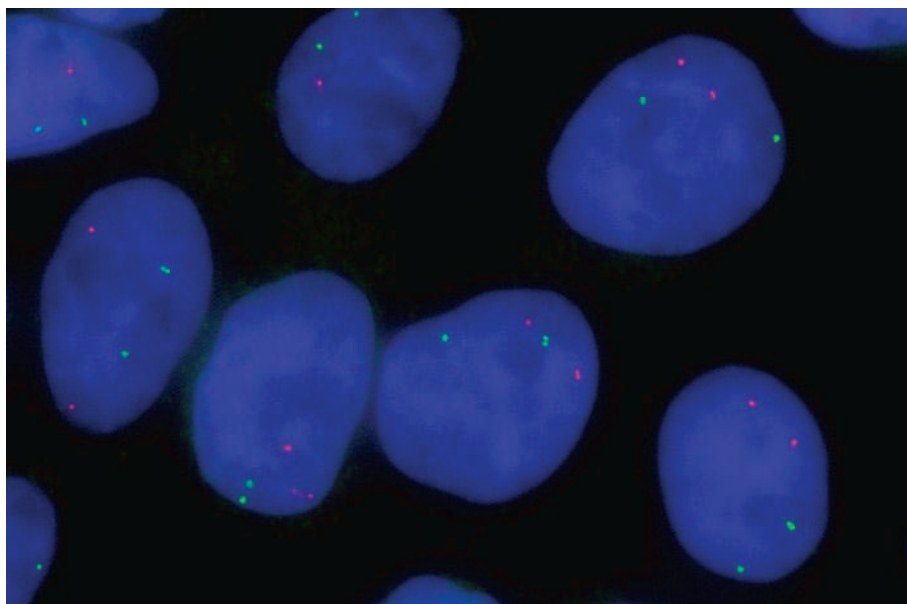
Fluorescence *in situ* hybridization can illustrate the positions of genes within nuclei, such as the *MYC* (red) and *MMP1/3/12* (green) genes, shown here in breast tissue.

identify interacting sequences or to locate sequences within the nucleus. But challenges remain: chromosomal movements are dynamic and non-deterministic, so detecting what is where, and when, is difficult. Even more difficult is figuring out when and how genome architecture affects gene activity.

Until the beginning of this century, nearly all techniques that were used to study chromosome arrangements relied on microscopy. Researchers could label certain DNA sequences or DNA-associated molecules, and see where the labelled areas were inside the nucleus. But a strand of chromatin is only about 10 nanometres thick, and conventional fluorescence microscopy has a resolution at best of 200 nanometres. Thus, microscopy can reveal that two loci are close to each other, but not whether they come into contact. Moreover, if an interaction is fragile or short-lived, microscopy can miss it altogether.

When Job Dekker was a postdoctoral researcher studying the mechanics of cell division at Harvard University in Cambridge, Massachusetts, he wanted to map the DNA sequences that mediated interactions between chromosomes. One day, while commuting to his lab, he hit on the idea of capturing an interaction by chemically snagging two strands of chromatin that approached one another, then fusing the DNA from both into a single molecule. "You start out with a difficult problem — where are two loci in three dimensions — and you convert it through a series of molecular steps to a simple problem, just sequencing a piece of DNA," says Dekker, now a genome biologist at the University of Massachusetts Medical School in Worcester.

Dekker's idea became a technique, described in the literature in 2002, known as chromosome conformation capture (3C; ref. 6). It has since

spawned many variations (see 'Investigating the architecture'), but the basic principles are the same. Protocols begin with 'cross-linking': dousing cells with formaldehyde to glue the DNA to its associated proteins, and those proteins to each other. Then the DNA is cut up with restriction enzymes or sheared by sonication, leaving behind 'hairballs' of tangled DNA and protein.

The next steps vary between protocols, but all combine free strands of DNA to create hybrid molecules: ligation products of DNA strands that had been close together on the same hairball. Researchers interested in genes that are associated with a particular transcription factor or other DNA-associated protein use specially designed antibodies to capture the relevant hairballs. In some techniques, chemically modified nucleotides are incorporated into hybrid molecules to ease purification, whereas in others, judicious application of PCR amplifies DNA sequences near loci of interest.

THE MEDIUM MATTERS

No matter which technique is used, researchers need to be careful when choosing their restriction enzymes. For example, those that cut at sites made up of 6-base-pair sequences produce large fragments that may not capture important interactions, whereas enzymes that recognize sequences of 4 base pairs may produce more and smaller fragments, perhaps generating so much background information that real interactions cannot be detected.

Researchers also need to keep in mind that most of the hybrid DNA molecules produced by this technique are the result of random interactions, particularly between loci that are just a few kilobases apart on the same chromosome; separating the signal from the background noise requires involved

INVESTIGATING THE ARCHITECTURE

Technique	Detects	Protocol after sample is cross-linked and digested	Detection method
3C (chromosome conformation capture)[4]	Interactions between two loci	Free DNA ligated. PCR is performed with one primer for each locus, and products are amplified	Quantitative PCR (qPCR)
4C (circular 3C; also known as 3C on a chip)[16,17]	Interactions between one locus and the rest of the genome	DNA molecules self-ligate into circles. Inverse PCR amplifies molecules containing loci of interest	Sequencing or microarrays
5C (3C carbon copy)[18]	Interactions between multiple selected loci (for example, several within one chromosome)	Extension primers are designed for each interaction to be analysed. These allow selected DNA to ligate and be amplified by PCR	Sequencing or microarrays
ChIP-loop (chromatin immunoprecipitation loop) assay[19]	Interactions between two loci bound by a particular protein	DNA bound to protein is purified and then ligated. PCR is performed with one primer for each locus, and products are amplified	qPCR
ChIA-PET (chromatin interaction analysis by paired-end tag sequencing)[5]	Genome-wide interactions for loci bound by a particular protein	DNA bound to protein is purified and then ligated. Ligation incorporates biotin-containing primers that allow ligation junctions to be purified	Deep sequencing
DamID (DNA adenine methyltransferase identification)[20]	Sequences that occur near nuclear landmarks or other proteins	Genetically modified cell lines are produced so that a DNA-tagging enzyme is fused to a protein of interest	Sequencing that detects methylated adenine
e4C (enhanced ChIP-4C)[7]	A more-sensitive version of 4C that does not require inverse PCR	Primer extension with biotinylated primer that allows ligated fragments to be purified	Sequencing or microarrays
Hi-C (high-throughput 3C)[12]	Genome-wide interactions at a resolution of about 1 megabase	Ligation step incorporates biotin-containing nucleotides that allow ligation junctions to be purified	Deep sequencing

bioinformatics and replicated experiments. "It used to be, even two years ago, that getting the data would be an endpoint of the project. Now it's the start," says Dekker.

On the plus side, preparing libraries of ligation products requires only very general reagents: formaldehyde, a variety of buffers and the enzymes that cut DNA and join it back together. Moreover, all the necessary reagents can be purchased from established companies: Life Technologies of Carlsbad, California; New England Biolabs of Ipswich, Massachusetts; QIAGEN of Hilden, Germany; Sigma-Aldrich of St Louis, Missouri; and Thermo Fisher Scientific of Waltham, Massachusetts. Researchers can also order specially synthesized primers for DNA amplification or ligation from a large range of (generally smaller) providers.

Different techniques generate different information. A million sequenced molecules (or 'reads') for Hi-C (high-throughput 3C) provides a low-resolution map of the whole human genome, whereas a million reads for 4C (circular 3C) produces a detailed interaction map for a gene of interest, and in ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) the same amount of data indicates which transcription-factor binding sites interact with which gene promoters.

This summer, Life Technologies plans to launch a kit that bundles together reagents for 3C experiments. The kit would allow researchers to monitor and optimize digestion, use less of the sample for ligation and produce a library of ligation products in 1.5 days, says Shoulian Dong, a technology developer at Life Technologies. But perhaps the most important factor for throughput is the increasing availability of next-generation sequencers from companies such as Applied Biosystems of Carlsbad, California, and Illumina of San Diego, California, which can quickly sequence the hundreds of thousands of short hybrid DNA molecules produced in these experiments.

FROM SEQUENCES TO IDEAS

The ability to detect specific interacting loci is already revealing previously unknown biology. Last September, researchers led by Richard Young, a molecular biologist at the Massachusetts Institute of Technology in Cambridge, described evidence for a biological system that juxtaposes separate stretches of DNA. Together, these stretches control gene expression. The team found that a 'mediator' protein complex was often bound to enhancer sequences and core promoters of genes transcribed in embryonic stem cells[7]. Another protein, cohesin, which can connect two DNA segments, was bound along with mediator, and purified with it. Follow-up 3C studies on four genes showed increased interactions between promoter and enhancer sequences in stem cells, but not in another type of cell in which the genes were inactive[7].

For Wouter de Laat, a genome biologist at the Hubrecht Institute in Utrecht, the Netherlands, who showed how 3C can be used to match a gene with its regulatory elements[8], the most exciting applications of chromosome capture technology are global: working out which sites interact with which genes in different tissues. "There are many more sites with regulatory potential than we have genes, and the only way to know which site is acting on which gene is to get three-dimensional," he says. "That's the next level of what we need in functional genomics."

Current techniques are not powerful enough to match regulatory elements and genes across the genome, but de Laat and other labs are working on more far-reaching methods, which they hope to describe in the literature this year. It is useful to ask genome-wide questions because, otherwise, researchers tend to interpret their results only in the context of the gene they happen to be studying, says de Laat. But because every gene is part of a chromosome, those observations could have less to do with the gene under study than with its neighbours.

Adding to the challenge is another signal-to-noise problem: all the current techniques have to be carried out on between 10 million and 20 million cells at once, which means that the observed interactions represent an averaged
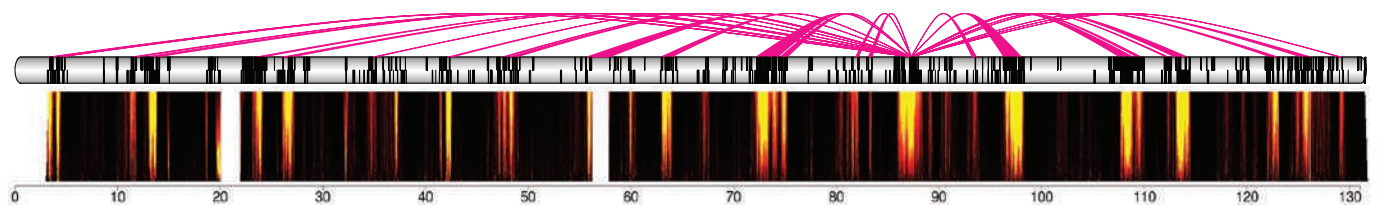
Regions on mouse chromosome 8 that interact with the *Rad23a* gene. The interactions were uncovered using conformation capture techniques.

reading. No one believes that all the interactions identified by sequencing technologies occur in any one cell, says Tom Misteli, who studies the cell biology of genomes at the US National Cancer Institute in Bethesda, Maryland. "Any interaction that happens will appear as a signal, but it doesn't tell you how often it happens in cells," he adds. "That makes the interpretation of the sequencing data a little bit complicated."

SEEING IS BELIEVING

To find out how often interactions occur, researchers have to count labelled cells under a microscope. For live-cell imaging, they can insert genes for fluorescent proteins that bind to desired DNA sites into the cell, but the technique is labour-intensive and tedious. A fixed-cell technique, fluorescence *in situ* hybridization (FISH), is more common. Nuclei are treated with formaldehyde, then denatured just enough to allow the entry of DNA probes that fluorescently label certain sequences.

In general, interactions identified by chromosome conformation studies are observed in only about one in ten cells under the microscope, says Misteli. That doesn't mean that the interaction isn't real; randomly selected loci are seen near each other even less often. Instead, such rates show just how dynamic and varied chromosome arrangements are, and how difficult they can be to study.

Last year, Fraser and his colleagues combined chromosome capture technology with microscopy to show that a single transcription factor, Klf1, helps to bring target genes from distant loci into a cluster in a common space[9]. Such studies of 'transcription interactomics' could reveal secrets of cell differentiation and stability, but mastering the necessary technologies is a formidable task. To separate relevant hybrid molecules from background signals, the researchers made significant tweaks to the 4C technique. And to show that multiple loci came together at the same time, lead author Stefan Schoenfelder looked at some 50,000 cells under a microscope: the equivalent, Fraser says, of spending half a year in a dark room.

That situation is familiar to Misteli, who in 2009 used FISH to show how genes reposition themselves in cancer[10]; such knowledge could aid diagnosis. Genes generally move from the periphery of the nucleus towards the centre when they become active, but individual genes move in unpredictable ways. No one has yet been able to look at gene positioning comprehensively, to discover how it might vary across different cell types, says Misteli. "It's all based on small sample numbers and people's favourite genes. So you want to look at more genes and that's simply not possible."

Technologies are improving, letting

researchers look at more cells; Fraser says that currently available microscopes with faster autofocus and more-agile robotic stages would now let Schoenfelder perform the same number of experiments in a month or less. Platforms are available: PerkinElmer in Waltham, Massachusetts, sells the Opera high-content screening system, which keeps the objective lens immersed in water. This allows it to work at the high resolutions required to determine where sequences are in the nucleus. The instrument automatically moves along wells on a plate to collect the necessary data, and its four different-coloured lasers can light up several probes in each cell.

The Opera instrument can examine loci in hundreds of cells a minute — considerably faster than stand-alone microscopes — and can make difficult techniques more accessible to non-experts, says Achim von Leoprechting, vice-president of imaging at PerkinElmer. "We're seeing FISH moving out of specialized

Loci can interact more (red) or less (blue) than would be expected given their distance in the genome.

labs," he says, "so from an imaging standpoint we need to make sure they can use these platforms and get high-quality data without being trained as microscopists." Researchers who are already studying the position of genes in the nucleus are particularly keen to examine more cell types under different conditions, says Aaron Risinger, a specialist in high-content screening at PerkinElmer. "For individuals who were doing one-off experiments, the natural progression is to move to high-throughput," he says. In fact, Misteli is doing just that by incorporating the platform into a new US National Cancer Institute facility aimed at ultra-high-throughput cell biological imaging.

Lower-throughput techniques also have their advocates. Ana Pombo, a cell biologist at Imperial College London, has developed the cryoFISH technique: rather than fixing and denaturing intact cells, researchers embed cells

in a sugar solution, carefully freeze them, cut them into thin slices, then add DNA probes[11]. The process is technically demanding but produces fewer artefacts and better resolution than standard FISH because the probes don't need to move through an entire nucleus. Pombo has used cryoFISH to show that chromosomes keep largely to their own 'territories' but intermingle extensively[9].
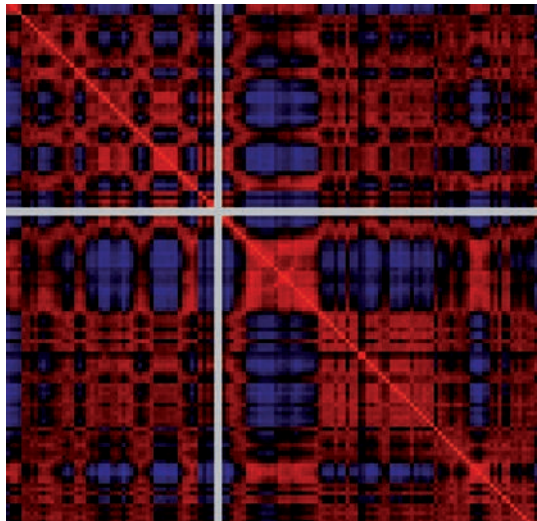
Electron microscopy has very high resolution, but the staining and imaging of cells can take days. In the past three years, researchers have turned to super-resolution optical microscopy, which uses techniques such as synchronized laser pulses to focus on structures as small as 15–20 nanometres — well below the 200-nanometre resolution limit of conventional optical microscopy — even in living cells. Companies selling these new microscopes include Applied Precision of Issaquah, Washington; Leica of Wetzlar, Germany; Nikon of Shinjuku, Japan; and Zeiss of Oberkochen, Germany, but the instruments have not yet reached most laboratories.

A THIRD WAY

Ultimately, all microscopy is a coarse detection technique, says Rolf Ohlsson, an epigeneticist at the Karolinska Institute in Stockholm. Standard fluorescence microscopy cannot distinguish between loci that are near each other and those that are in contact; even super-resolution microscopy cannot do so definitively. On the other hand, sequencing techniques cannot show which interactions occur together, says Ohlsson. "Somewhere between DNA FISH and chromosome conformation capture is the truth," he adds. But even accurate representations will not be enough: ascertaining that an interaction occurs is far easier than showing that it affects function. "Is what you see an interaction?" asks Ohlsson. "Or just a collision?"

Several groups are attempting to use conformation capture to build computational models that show the positions of chromosomes in different cell types and at different stages of the cell cycle. To construct these models, researchers do not actually measure distances between two loci; instead, they use algorithms to process captured DNA sequences. The programs produce 'proximity profiles' from sequencing data by measuring how frequently regions of the genome are observed to interact with one another, and comparing that with what would be predicted from chance.
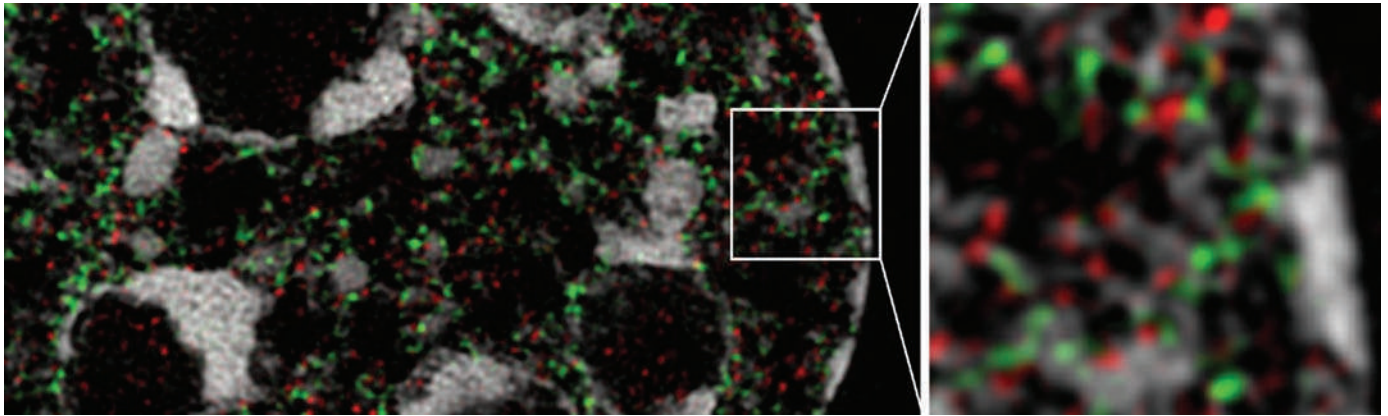
In 2009, Dekker and his colleagues constructed a model of human cells that breaks the 3-billion-base-pair genome into 3,000 pieces and maps long-range interactions[12]. That resolution is too poor to show individual genes, let alone predict which binding sites might help to generate a particular conformation, but

Super-resolution image of part of a mouse-cell nucleus, showing dense regions of chromatin separated by DNA-free channels. RNA production (red) and DNA replication (green) occur in a layer of decondensed chromatin on these domains. Strands of chromatin occasionally loop long distances between domains.

creating a more detailed picture is difficult. Constructing the interaction map required some 30 million reads of fused DNA molecules; improving resolution by a factor of 10 (to 100-kilobase pieces) would require some 3 billion reads, because the number of reads required increases exponentially as the resolution improves linearly. Even so, Dekker and his colleagues' maps agreed with established ideas about chromosome territories, indicating that gene-rich areas lie close together.

WHOLE-GENOME MODELS

This year, researchers led by Dekker and Marc Marti-Renom, a bioinformatician at the Prince Felipe Research Centre in Valencia, Spain, published the results of 3C carbon copy (5C) performed on two different types of cell. They used the data to build a three-dimensional model of a 500-kilobase region of human chromosome 16 (ref. 13). This region contains a cluster of housekeeping genes active in most cell types, and another set of genes active in only some cells. Using interaction-frequency maps, the researchers generated chromatin models for both cell types. These predicted the existence of compact chromatin structures in which active genes were clustered. In the cells in which both sets of genes were active, the chromatin in the model folded into two 'globules'. In cells in which only the housekeeping genes were active, only one globule formed. FISH experiments confirmed the overall size and shape of this region of chromatin in individual cells.

It is possible to construct genome-wide models at higher resolution, by starting with smaller genomes. Last year, Ken-ichi Noma, who studies gene expression at the Wistar Institute in Philadelphia, Pennsylvania, and his colleagues took this approach, generating a very high-resolution genome-wide model of the fission yeast *Schizosaccharomyces pombe*, which has only three chromosomes, containing a total of about 14 million base pairs and 5,000 genes[14]. The researchers calculated how close different pieces of chromatin were to each other

by dividing the genome into sections of just 20,000 base pairs, and confirmed several results with microscopy. Earlier that year, a multilaboratory team had built a kilobase-resolution model of the genome of the budding yeast *Saccharomyces cerevisiae*, which has 16 chromosomes[15].

The challenge starts with gathering reliable data: picking out real interactions from background reads. "The hardest step was going from sequence data to a set of interactions we could trust and interpret functionally. We had the data in hand for a year before the paper was published," says William Noble, a genome biologist at the University of Washington, Seattle, who leads one of four labs that produced the budding yeast model. The structure provides a visual interpretation that the human brain can understand, says Noble, but that interpretation can be taken only so far. "The structure isn't introduced until the very end because we didn't want to base any of our conclusions on the structure itself," he says.

"Any interaction will appear as a signal, but it doesn't tell you how often it happens. That makes it complicated."
Tom Misteli

Other researchers acknowledge that such models could be useful, but worry that they could be misleading. "When you say that two points are folded together, what's in between? We don't have the physical parameters to predict what's really happening there," says Ruan. The distance estimates from high-throughput data represent an "unrealistic average" that does not take into account that chromatin is in constant, often non-directed, motion, says Pombo. "You make protein structures when you crystallize a protein," she says. "Nuclei are not like that."

Model builders reply that in future,

representations will reflect the dynamic, semi-random movements of chromosomes, and that current versions can still be valuable, by showing overall tendencies. "By imaging you highlight the variability. By chromosome capture you highlight the commonalities," says Dekker.

But Cremer suggests that researchers should spend at least as much time with their microscopes as with their computers. Before people can really understand what high-throughput sequencing data tell us about higher-order chromosome arrangements, he says, the field needs many more descriptive studies. "One has to be very careful about making generalizations at this moment, and we need a lot more data." ∎

Monya Baker *is technology editor for* Nature *and* Nature Methods.

1. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y.-F. & Fraser, P. *Nature Genet.* **32**, 623–626 (2002).
2. Solovei, I. *et al. Cell* **137**, 356–368 (2009).
3. Osbourne, C. S. *et al. PLoS Biol.* **5**, e192 (2007).
4. Roix, J. J., McQueen, P. G., Munson, P. J., Parada, L. A. & Misteli, T. *Nature Genet.* **34**, 287–291 (2003).
5. Fullwood, M. J. *et al. Nature* **462**, 58–64 (2009).
6. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. *Science* **295**, 1306–1311 (2002).
7. Kagey, M. H. *et al. Nature* **467**, 430–435 (2010).
8. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & de Laat, W. *Mol. Cell* **10**, 1453–1465 (2002).
9. Schoenfelder, S. *et al. Nature Genet.* **42**, 53–61 (2010).
10. Meaburn, K. J., Gudla, P. R., Khan, S., Lockett, S. J. & Misteli, T. *J. Cell Biol.* **187**, 801–812 (2009).
11. Branco, M. R. & Pombo, A. *PLoS Biol.* **4**, e138 (2006).
12. Lieberman-Aiden, E. *et al. Science* **326**, 289–293 (2009).
13. Bau, D. *et al. Nature Struct. Mol. Biol.* **18**, 107–114 (2011).
14. Tanizawa, H. *et al. Nucleic Acids Res.* **38**, 8164–8177 (2010).
15. Duan, Z. *et al. Nature* **465**, 363–367 (2010).
16. Simonis, M. *et al. Nature Genet.* **38**, 1348–1354 (2006).
17. Zhao, Z. *et al. Nature Genet.* **38**, 1341–1347 (2006).
18. Dostie, J. *et al. Genome Res.* **16**, 1299–1309 (2006).
19. Horike, S.-I., Cai, S., Miyano, M., Cheng, J.-F. & Kohwi-Shigematsu, T. *Nature Genet.* **37**, 31–40 (2004).
20. Greil, F., Moorman, C. & van Steensel, B. *Methods Enzymol.* **410**, 342–359 (2006).