Genomics in 3D and 4D

Vivien Marx

DNA folding shapes gene expression. Emerging techniques promise to reveal the intricacies of this architectural language of chromosomes.

Two meters is the length of a mammalian cell's DNA if the chromosomes are placed end to end. In the cell, these two meters are spooled around proteins and folded into the nucleus, which is an organelle around 0.00001 meters, or 10 micrometers, in diameter. That's like folding a 100-kilometer-long sleeping bag into a camping stuff sack.

Chromosomes are not static occupants of the nucleus; they constantly move and sample different positions, says Mitchell Guttman, a researcher at the California Institute of Technology. All loci wiggle, says Job Dekker, a researcher at the University of Massachusetts Medical School. Not every movement matters, but a structural change can have small to dramatic effects: it can determine cell type or developmental stage or create the folding that is characteristic of cancer cells. Chromosomes do not have one overarching 3D structure: an embryonic stem cell's chromosomes are structurally unlike those in a mature neuron. "3D structure is both a cause and consequence of gene regulation," says Guttman.

In a folded chromosome, DNA regions that are distant on the linear molecule become interacting neighbors. Some interactions lead parts of the genome to be transcribed into messenger RNA, whereas other genomic regions interact with lamins, the network of proteins on the inside of the nuclear envelope, and remain quiet. "So in a way the

In addition to the C-technologies, new methods will help expand the view of chromosome topology, says Ana Pombo.

DNA is compacted 200,000-fold and packed into the nucleus in ways that labs are eager to characterize.

genome-folding state could be diagnostic for the cell state," says Dekker.

Labs have found that there are compartments in the nucleus in which genes tend to be silenced and others in which genes are activated, says Ana Pombo, a researcher at the Max Delbrück Center for Molecular Medicine in Berlin. This organizational principle is like packing some books in the attic and keeping others at eye level on a shelf in the living room. Each cell type displays clear choices related to gene expression control, gene activation and DNA packing, she says.

Tools now and soon

With existing techniques to study chromosomal conformation, nuclei are isolated and fixed to hold interacting chromatin areas in place, and the chromatin is cross-linked, digested, ligated and then analyzed. The original chromosome conformation capture (3C) protocol includes ligation followed by PCR to find two defined interacting loci, and 3C variations apply high-throughput sequencing techniques. Some methods find interactions between one locus and many or all others; other techniques probe

interactions of many sites with many or all others. The methods vary in terms of how much sequencing they involve, how time-consuming they are and the signal-to-noise ratio in their data[1–4].

By applying current methods, researchers have gained a more complex view of gene regulation and structure, says Guttman. Now, his lab and others seek genome-wide, unbiased, quick but comprehensive assessments of chromatin conformation, new types of imaging techniques and data integration and visualization, all of which are the remit of the 4D Nucleome program at the US National Institutes of Health (NIH) Common Fund. A first round of awards last year sent $25 million to 29 research groups characterizing nuclear organization and its effect on the cell and genome, in part for developing and validating tools with which to measure chromatin interactions and decipher function.

Labs have moved from studying a single or isolated gene to looking at whole genomic transactions in an intact nucleus under native conditions and taking into account that a genome is constantly multitasking, says Ananda Roy, who directs the NIH program. To help labs interpret and predict chromosomal folding, one goal is to create reference maps of chromatin interactions in space and time in many cell types and across developmental phases. With that goal in mind, labs are increasing the sensitivity and throughput of assays and sharing new tools to get them ready for community-wide use.
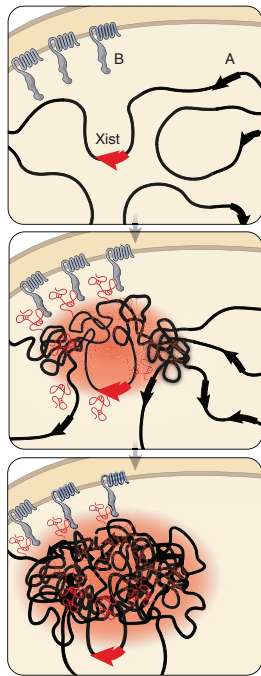
The 4D Nucleome program plans to offer common cell lines so different methods can work reproducibly in these cells and cell lines, says Roy. The cells can eventually become a community resource. Other research initiatives devoted to the 4D nucleome are in the making, for example, in Japan and in Europe.

Xist, a long noncoding RNA, silences one of the two X chromosomes (A) in female mammals. Xist interacts with proteins (B) on the nuclear membrane, then the chromosome folds up.

S. Knemeyer, Guttman lab, Caltech; Erin Dewalt/Nature Publishing Group

The traditional C-technologies have been and are still going to be transformative, says Pombo, but alternative methods will help expand a wider view of chromosome topology. C-based chromatin contacts are often validated by fluorescence *in situ* hybridization (FISH). Cryo-FISH, or FISH performed on thin cryosections of fixed cells embedded in sucrose, is also used and allows higher-resolution imaging. Although both FISH and cryo-FISH are informative, she says, new approaches will help address the challenges of automating probe production and of collecting and analyzing images. Automated image collection is accessible to only a few labs and is expensive to set up.

Pombo is working on a 4D Nucleome project led by University of California at San Diego researchers Bing Ren and Cornelis Murre to optimize protocols for contact-mapping technologies and validate them with imaging and other approaches. "I find it especially important to develop novel technologies that do not require the chemical locking of conformation through DNA ligation, which has been a main focus of my lab's research," she says, but it's too soon to talk about details. Guttman is also working on ligation-free ways to capture chromatin conformation.

All 3C methods use cross-linking, and they don't work well without it, says Dekker. Detergents help to loosen up the cross-linked chromatin without risking the loss of chromatin interactions and allow the chromatin to be more evenly digested with enzymes. Without this step, he says, digestion is biased toward open and active chromatin sites, and interactions in closed chromatin can escape detection. Even so, cross-linking can lead to artifacts and blind spots, which is why he looks forward to live-cell studies and assays in development that do not use cross-linking. His 4D Nucleome project focuses on new ways to perturb chromosomes, such as by editing genes to explore how chromosome structure turns some genes off and others on.

Loops in the family

Over time, the C-family of methods has grown. The methods take a similar strategy for chromosome proximity ligation and differ mainly in the post-processing of the ligation products, says Amos Tanay, a researcher at the Weizmann Institute. One can get many contacts for fewer regions or fewer contacts for more or all regions.

Tanay would like more "knobs" to play with to, for example, make an experiment more sensitive and less specific or more specific and less sensitive. A few other methods are emerging with this goal, such as next-generation Capture-C[5]. To enhance the sensitivity of 3C data and allow a quantitative interpretation, he, his team and colleagues at other institutions in Israel developed an approach and software called UMI-4C that uses unique molecular identifiers (UMIs) to label every interaction[6].

Labs can use UMI-4C to zoom in on, for example, 30–60 loci and obtain accurate and high-resolution chromatin contact profiles about them without needing to do too much sequencing. Looking at many more loci with capture Hi-C, in which specific targets are enriched, or looking genome-wide with plain, unenriched Hi-C involves much more sequencing to deliver the numbers of interaction sites per locus. As the view of the genome broadens, the amount of specificity information about the contacts per DNA element decreases unless the sequencing depth is dramatically increased.

The desire to scan chromatin contacts more quickly has been driving methods development. Dekker and his colleagues developed 5C (carbon-copy chromosome conformation capture) to achieve higher throughput and to detect matrices of inter-

actions, "as only matrices of interactions can reveal domains and loops," he says.

Proteins have folding motifs that affect protein function in a variety of ways. Chromosomes, too, may be shown to have different classes of folds, says Dekker. Proteins have evolved semi-deterministic folding, and although proteins are dynamic, the same peptide will usually fold into the same structure, says Tanay. Rather than folding deterministically, chromosome loops evolved to fold so as to preserve essential but general structural features that allow dynamic function.

"We do not know the nature of these folds at this moment," says Dekker, nor do scientists know how to detect them in the contact maps made using 3C or Hi-C approaches, or even how to recognize them in other types of data such as from FISH or live-cell imaging. Chromatin contact maps describe loops and domains, and the research community has to fit these data into models and characterize dynamic chromatin folding.

The folding pattern can relate to disease, says Roy; for example, mutations might cause defective looping, or a protein that tethers a loop in place can become defective. Reference maps will help scientists compare, contrast and predict cell states, and capture snapshots of genomic configurations as structure changes, he says. Building a map or atlas of
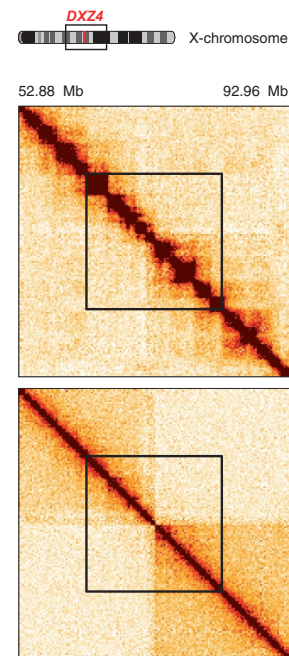
Chromatin interactions captured in Hi-C maps are plentiful in the active X chromosome (top map) but largely absent in the silenced X chromosome (bottom map).

B. Lajoie, J. Gibcus, Dekker lab, UMassMed; Erin Dewalt/Nature Publishing Group

the nucleus with three-dimensional coordinate positions and chromatin views at differing levels of resolution involves integrating results from genomic, imaging and data visualization techniques, says Guttman. Mapping all aspects of DNA and RNA as they change genome-wide in all three dimensions, across time and at all resolutions—"that's probably a little more science fiction than reality today," he says, although he believes it will come together.

The genome-folding state could indicate the cell state, says Job Dekker.

UMassMed

X marks the spot

Guttman seeks to integrate RNA into the emerging picture of 3D and 4D nuclear structure. Because RNA is the readout of transcription, integrating RNA into these maps can broaden the view of how structure shapes transcription and will help researchers study the role of noncoding RNAs. He draws on Hi-C to map out chromatin structure, using FISH to image RNA and DNA simultaneously in the cell. He is also manipulating RNA expression with genetic modifications to track what happens over time when that RNA contacts the 3D structure of DNA.

The long noncoding RNA Xist got Guttman into studying nuclear structure. Xist inactivates one of the two X chromosomes in female mammals. The silenced chromosome interacts minimally with others and is visible as a compacted structure at the periphery of the nucleus. But genome-wide maps of Xist's target sites weren't offering enough information about how Xist finds its regulatory destinations on the X chromosome. Hi-C has offered more clues, says Guttman, indicating that Xist reshapes the X chromosome and that this folding pattern determines where Xist binds[7,8].

When a cell divides and sends DNA to a daughter cell, the chromatin is unpacked and then repacked. During this process, says Tanay, not only do chromosomes maintain their dynamic function of transcription and replication, but some parts of the chromosome stay active while others are repressed, and the different functions do not interfere with one another. It's like a crowded airline flight, says Tanay: one passenger wants to work, eat or listen to music, and her neighbor wants to sleep. Changes in chromosome configuration help to define gene regulation, and the configuration is affected by other elements such as transcription factors. Exploring 3D chromosome folding makes it easier to understand how different components in the gene regulation machinery work, says Tanay.

The genome contains compartments of large stretches of open and closed chromatin that interact preferentially with each other. It is not yet clear, says Dekker, to what degree the interactions between certain compartments differ in frequency or in kind. C-based techniques have also revealed much smaller domains, between hundreds of kilobases and several megabases long, called topologically associating domains (TADs). Experiments that disrupt the boundaries between TADs show that gene expression in one TAD interferes with expression in the adjacent TAD, says Edith Heard, a researcher at the Institut Curie in Paris. Active TADs interact with one another, and learning about their role matters, says Roy. In cancer, a normally inactive locus can be turned active.

TADs can interact with one another over long distances. For example, Pombo and colleagues in several countries analyzed data in matrices from Hi-C experiments to explore chromatin changes during the development of mouse neurons; they also studied data from several human cell types. They found that long-range interaction between different TADs correlates with, for example, gene expression and epigenomic changes. These TAD–TAD interactions possibly facilitate how chromatin is compacted or genes are activated.

Finding TADs in Hi-C matrices is computationally straightforward, says Tanay, because of their prominent signal. At the same time, TADs do not have a canonical definition because they appear in the data at multiple levels and scales. "So you can decide to set the threshold high and get fewer TADs or zoom in and find more," he says.

Using current methods, labs can

Eventually it will be possible to map all aspects of DNA and RNA, genome-wide in all three dimensions, across time and at all resolutions, says Mitchell Guttman.

L. Rubinstein

agree on their TAD results, says Dekker, but there is plenty of discussion about how to interpret data about these domains, how they form, and what functions they fulfill. Higher-resolution data will let researchers zoom in or out, revealing perhaps smaller domains embedded within larger ones.

Single-cell view

With current approaches, plenty is still unexplained. "What does it mean for two pieces of DNA to lie within 700 nanometers when the chromosome diameter is 2 micrometers?" says Pombo. The extent and strength of local interactions are shaped by many factors such as the binding of transcription factors, which in turn shape the large-scale arrangement of subchromosomal regions and possibly entire chromosomes, she says. The nucleus and the genome make up a system in which local changes can propagate effects all the way to the behavior of a whole chromosome.

Current methods capture the ensemble of chromatin interactions in millions of cells, says Guttman. He is working on ways to allow simultaneous measurement of all interacting DNA and RNA molecules. Some of the interactions are mutually exclusive, he says, which is where single-cell analysis will help. In a 4D Nucleome project, Guttman and colleagues at Caltech are mapping RNA targets on chromatin in many thousands of single cells and at single-molecule resolution and building microfluidic devices with which to measure RNA–DNA nuclear compartments in hundreds of thousands of nuclei.

Single-cell Hi-C techniques now allow labs to visualize the structure of nuclei, says Tanay. Time and again researchers see that nuclei are dynamic and stochastic; each chromosome looks different. "The challenge is to understand what key properties of chromosomes are recurrent and how they relate to chromosome functions," he says.

Ensemble analysis indeed does not always translate to single-cell analysis, says Roy. For a given question, labs will need to decide how much single-cell analysis they need. The human body has nearly 100 trillion cells, but it might, for example, be sufficient to characterize the major cell types. "This is where modeling and mathematical predictions will come in handy before we delve into 100 trillion combinations," he says.

Genomic methods to probe chromosomal configuration have helped the field progress, says Dekker, as have imaging advances such as super-resolution techniques and new types of dyes. Next up is data integra-

npg
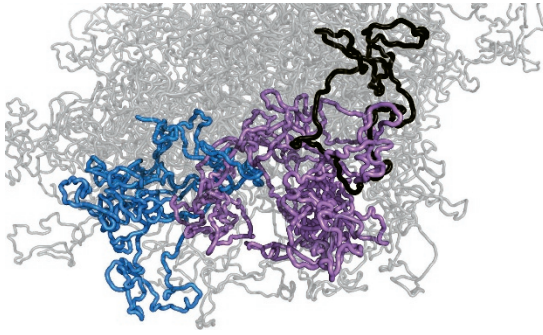
C-based data and a polymer model show how topologically associated domains (TADs; blue and purple) interact. Molecular motors at the base of some loops (black) likely move TADs into place.

tion to derive comprehensive models of the spatial and dynamic organization of chromosomes, drawing on physics principles and modeling. Labs often have FISH and Hi-C data, "and when we compare imaging and genomic data we often think they are in conflict with each other," he says. "I am an optimist and think both data types are actually correct; we just don't understand how they both can be correct." New ways to integrate the data obtained with different methods will help scientists build models of chromosome configuration.

At the current time, global genome-scale analysis is more advanced than imaging scales, says Roy, leading to two different "languages" about chromosomal architecture. The plan is to learn how to match imaging scales with genomic scales to create what he calls an "illustrated book of life." The 4D Nucleome Data Coordination and Integration Center at Harvard University Medical School, led by Peter Park, is addressing this challenge.

Physics and models
In Pombo's view, chromatin behaves like a functionalized polymer in which each part has specific properties that determine its behavior. She and her colleagues use polymer physics calculations to explore different folding mechanisms; they compare these results with experimental data and then return to the model. "The polymer modeling helps us focus our efforts on most likely and interesting mechanisms of folding," she says.

To date, questions about chromatin fiber structure at the nucleosome level and about 3D folding of this fiber in chromosomes are

separate. "Bridging those two questions in a unifying model is currently the 'holy grail' many groups are chasing," says Julien Mozziconacci, a physicist at Pierre and Marie Curie University in Paris. Experiments that draw on genomics and microscopy help with this bridge-building, but his "gut feeling" is that the only way this model can be achieved is by considering the function of this multi-scale architecture.

Labs use molecular dynamics and empirical force fields to model the deformation and interactions of macromolecules such as DNA, proteins, RNA and lipids, says Mozziconacci. Statistical physics helps to describe how a large number of entities behave when the interaction between them is known. Polymer physics is a way to model chromosomes using C-data. "Depending on the way the polymer folds on itself, you can predict the exponent of the decrease of the contact frequency when the genomic distance between two loci increases," he says. "Unfortunately, different polymer models can lead to the same exponent so that more experiments are needed at this point."

Many labs are bridging the gap between questions about chromatin fiber structure at the nucleosome level and 3D folding of this fiber in chromosomes, says Julien Mozziconacci.

Cross-training
Building bridges across disciplines is easier when labs collaborate in close-knit, exchange-oriented networks, but doing science in this connected way appears to have fallen out of favor of late with European funders, says the Institut Curie's Heard.

Complementing the NIH 4D Nucleome project, she and a number of other scientists have proposed the 4D Nucleome Initiative in Europe (http://www.4DNucleome.eu) for labs to characterize chromatin architecture in many model organisms and advance techniques such as live-cell imaging and single-molecule tracking. The group hopes to hear soon about whether this effort can get off the ground, says Heard.

Interdisciplinary science is a vocation, says Mozziconacci. As a physicist, he did a postdoctoral fellowship in a molecular biology lab and then worked in a computationally oriented one. It has been a struggle to learn language, tools, the differing concepts and knowledge bases, he says, but the challenging road has become satisfying and mind-opening because he can talk with colleagues in different fields and tackle transversal problems. "Having said that, I think that there is a pitfall that biologists should not fall into: complex interdisciplinary work should not only be done by one postdoc, for instance, a computer scientist or a physicist, analyzing data from the wet lab," he says. He or she will have ideas that need to be challenged and to grow within a collaborative, multidisciplinary group.

It has taken a while to convince the scientific community that 3D chromatin folding matters, says Pombo. Beyond needing new and multidisciplinary approaches, this research also takes courage, time and mutual commitment, she says. Researchers have to step out of a comfort zone, reach across disciplines to others and build effective ways to work together. And, she says, "for many it's only when the knowledge in one's comfort zone fails to explain a given observation that one is forced to reach out."

1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Science **295**, 1306–1311 (2002).
2. Pombo, A. & Dillon, N. Nat. Rev. Mol. Cell. Biol. **16**, 245–257 (2015).
3. Dekker, J., Marti-Renom, M.A. & Mirny, L.A. Nat. Rev. Genet. **14**, 390–403 (2013).
4. Kolovos, P. et al. Epigenetics Chromatin **7**, 10 (2014).
5. Davies, J.O.J. et al. Nat. Methods **13**, 74–80 (2016).
6. Schwartzman, O. et al. Nat. Methods **13**, 685–691 (2016).
7. Chen, C.-K. et al. Science http://dx.doi.org/10.1126/science.aae0047 (2016).
8. Giorgetti, L. et al. Nature **535**, 575–579 (2016).

Vivien Marx is technology editor for Nature Methods (v.marx@us.nature.com).