



Chromatin untangled: New methods map genomic structure

With some 2 meters of DNA squished into the micron-sized nucleus of every mammalian cell, to call that space “cozy” is an understatement. Yet for cells, the genome poses issues not only of compaction but of functionality: The DNA must be packaged such that it remains accessible to gene regulatory factors. Indeed, chromatin architecture plays a key role in gene regulation, but until recently, researchers had few tools with which to study that relationship. **By Jeffrey M. Perkel**

For decades, researchers have recognized that cancer cells often carry a raft of chromosomal abnormalities. A translocation joining pieces of human chromosomes 9 and 22 produces the so-called “Philadelphia chromosome,” a structural abnormality associated with chronic myeloid leukemia. And certain forms of prostate cancer are associated with gene fusions that place ETS-family transcription factors under the control of androgen-responsive promoters. In both of those events, no genetic material is lost; it simply moves from one location to another. But genes are fused in the process, and proteins that normally are tightly regulated become constitutively active, leading inexorably to disease.

Some forms of acute myelogenous leukemia (AML) also harbor specific genetic aberrations—in this case, on the long arm of chromosome 3. That region, called “3q26,” contains *EVI1*, a gene that is normally expressed only in early hematopoietic progenitors but is turned off at differentiation. But in certain cases of AML, *EVI1* is turned on permanently.

To find out why, H. Ruud Delwel, professor of molecular leukemogenesis at **Erasmus University Medical Center** in Rotterdam, the Netherlands, and his team sequenced the translocation breakpoints for 3q26. It turns out that 3q26 could break either upstream or downstream of *EVI1*—it didn’t seem to

matter, says Delwel. But the other breakpoint was tightly restricted to another piece of chromosome 3 located some 40 megabases away: an 18,000-base-pair segment of 3q21, sandwiched between *GATA2* and *RPN1*. There’s no gene or promoter there, so no gene fusion per se occurs. But there is a regulatory element present, a 1,000-base pair sequence that normally drives the *GATA2* promoter in hematopoietic stem cells. The AML translocation forces that enhancer to switch genetic allegiances, activating *EVI1* while concurrently shutting down *GATA2*.

The net result is a pair of genetic anomalies, both of which have been independently associated with leukemia. Prognosis for these patients is exceptionally poor, Delwel says. But disease does not result simply from changes in linear distance on the chromosome. After all, it’s theoretically possible that an enhancer, even one located megabases away, could reach out and activate a distant but otherwise silent promoter. Yet, such interactions are largely precluded by the basic folding architecture of the genome. The translocation associated with AML fundamentally alters that architecture on chromosome 3, making such interactions far more likely.

Today, armed with an increasingly sophisticated, high-resolution toolset, researchers are decoding the architectural secrets of chromatin with unprecedented clarity. Some are using population-based sequencing approaches, while others are exploiting the power of microscopy to study chromatin architecture on a cell-by-cell basis. Still others are developing methods to probe the genome at the single-cell level and in living cells. There’s much left to do. But according to Job Dekker, a Howard Hughes Medical Institute investigator at the **University of Massachusetts Medical School**, a corner has been turned: Researchers have advanced from simply cataloging chromatin structure to actually manipulating it. “We are now at the stage where we can do classical structure–function studies,” he says.

Chromatin in the key of C

Researchers have long known that the eukaryotic nucleus must be ordered, thanks to “chromosome painting” studies suggesting each chromosome exists in its own domain. Yet the structure within those domains remained mysterious for several years. Then, in 2012, researchers proposed that eukaryotic chromosomes fold into relatively predictable structures called “topologically associating domains” (TADs), and that these structures prevent enhancers from activating genes all over the genome by effectively restricting their actions to within a few hundred thousand bases. “TADs gave us a physical understanding of where the boundaries of such elements could act,” explains Bing Ren, professor of cellular and molecular medicine at the **Ludwig Institute for Cancer Research** in La Jolla, California, whose lab was one of three to first describe these domains.

To define TADs, researchers use a variety of techniques, collectively called the “C methods.” In 2002, Dekker developed the first such approach, chromosome conformation capture (3C), as a postdoc in Nancy Kleckner’s laboratory—a strategy he dreamed up as a way to study chromatid pairing during meiosis.

Though they differ in the particulars, C methods all rely on proximity ligation, using DNA ligase to suture together two

Upcoming Features

Neurotechniques—November 4 ■ Cell Analysis—November 25 ■ Genomics—January 13



“Even in phenotypically identical cells, the way chromosomes fold is highly variable from cell to cell.”
—Peter Fraser, Babraham Institute, UK

linearly distant DNA fragments that happen to be close together in 3D space. In 3C, linked fragments are isolated and identified by PCR, using primers to both candidate regions; subsequent variants of 3C have multiplexed the method, improved its resolution, and broadened its reach to identify interactions over broad chromosomal regions or even genome-wide. Today, they represent the foundational technology of chromatin architecture analysis.

Alternative strategies include ChIA-PET (chromatin interaction analysis using paired-end tag sequencing), which effectively blends a genome-scale variant of 3C with chromatin immunoprecipitation (ChIP) to isolate those interactions associated with specific proteins; and DamID (DNA adenine methyltransferase identification), which identifies lamina-associating domains—chromatin regions associated with the inner face of the nuclear membrane, which tend to be transcriptionally silent—by fusing lamin proteins with a bacterial methyltransferase and monitoring the location of methylated sequences.

Epigenetics product vendor **Active Motif** offers a commercial tool that it says may also be used to map chromatin loops. The enChIP (engineered DNA-binding molecule-mediated ChIP) assay kit uses an antigen-tagged, enzymatically inactive (“dead”) Cas9 protein to map the specificity of the nuclease for genome editing applications via ChIP. But according to product manager Kyle Hondorp, the assay can also capture chromatin-looping events.

Perhaps the most widely used approach these days, however, is Hi-C. Erez Lieberman Aiden, assistant professor of genetics at **Baylor College of Medicine**, developed that approach in collaboration with Dekker while a graduate student with Eric Lander at the Broad Institute of MIT and Harvard. As he recalls, “There was a seminar or something, and Eric commented that the Illumina sequencers generated so much data that it was often in one’s interest to figure out how to translate a [genetic] problem into a sequencing problem, and then it would just be this game-changer.”

Aiden recognized that coupling nuclear ligation with an unbiased, high-throughput sequencing-based approach—essentially updating a gel-based method published in 1993 by Vanderbilt University researcher Katherine Cullen—could represent just such a game-changer. But his original protocol, published in 2009, was “a bit of a letdown,” says Aiden. Hi-C data generally take the form of heat maps of contact

frequencies. Contact domains stand out in such data as a sequence of dark squares along the diagonal. But with Hi-C resolutions in the tens of thousands of bases, loops between enhancers and promoters were simply too small to be detected.

Aiden’s team identified two problems with the original method, he says. One was sequencing throughput. A “back-of-the-envelope” calculation suggested he might need 10 billion reads in order to actually capture promoter–enhancer interactions—this in 2007, when 10 billion sequencing reads was more “than had ever been generated in the history of the world.” But the method was also laborious and clunky, and produced inherently blurry maps. So, his team, led by Suhas Rao and Miriam Huntley, spent five years honing the method and waiting for sequencing throughput to catch up. One of the most important changes, he says, was restoring the ligation step to the intact nucleus, as in Cullen’s original work (in 3C, the nucleus is disrupted first). Eventually, in 2014, they published an updated method called “in situ Hi-C” that featured kilobase resolution sharp enough to capture chromosomal looping events, which appear as dark, off-axis dots on Hi-C heat maps.

Going single-cell

Today, researchers are using C methods to address a number of vexing biological questions. Edith Heard, head of the Genetics and Developmental Biology Department at the Institut Curie in France, worked with Dekker and applied allele-specific Hi-C (a method in which each sequencing read can be assigned to either the paternal or maternal chromosome) to inactivation of one of the X chromosomes in female mammals. In 2014, Aiden’s team demonstrated that the inactive X assembles not into TADs (as the active X does) but into two massive “superdomains” separated by a “hinge” at the DXZ4 macrosatellite region. Heard and Dekker (and Aiden, working independently) showed in July 2016 that disruption of that region, or loss of the X-inactivation-associated noncoding RNA Xist (X-inactive specific transcript), alters this partitioning as well as the expression of the few genes that normally manage to “escape” inactivation. And Danny Reinberg, a Howard Hughes Medical Institute investigator at **New York University (NYU) Langone Medical Center**, has used C methods to probe and disrupt the exquisite temporal regulation of the *Hox* body-plan-regulating gene cluster.

Yet for all that researchers can learn from such methods, these techniques describe populations, not individual cells, each of which may display a different chromosomal conformation, or be at a different stage of the cell cycle, says Peter Fraser, head of the Nuclear Dynamics Programme at the **Babraham Institute** in Cambridge, United Kingdom. As a consequence, the resulting contact maps may not actually represent any cell in the population at all—they are the equivalent, he explains, of trying to capture the dynamics of soccer by photoaveraging hundreds of snapshots of a game. “It would just give you sort of a lump,” he says.

Fraser’s solution to that problem is single-cell Hi-C, a method he first described in 2013. The key to the method, Fraser says, is keeping the nuclei intact prior to ligation. “It makes it cleaner and less noisy,” he explains, than disrupting the nuclei first (as in Aiden’s original Hi-C protocol). But initially, his team could probe only about 2.5% of the contacts in any given cell, in perhaps a dozen cells at a time—today, improvements have boosted coverage about 10- to 20-fold, **cont.>**

Featured Participants

Active Motif www.activemotif.com	Institut Curie www.institut-curie.org
Albert Einstein College of Medicine www.einstein.yu.edu	Janelia Research Campus, Howard Hughes Medical Institute www.janelia.org
Babraham Institute www.babraham.ac.uk	Ludwig Institute for Cancer Research www.ludwigcancerresearch.org
Baylor College of Medicine www.bcm.edu	New York University Langone Medical Center nyulangone.org
Center for Integrative Biology cbi-toulouse.fr/eng	University of Illinois College of Medicine at Chicago www.chicago.medicine.uic.edu
Department of Genome Sciences, University of Washington www.gs.washington.edu	University of Massachusetts Medical School www.umassmed.edu
Erasmus University Medical Center www.erasmusmc.nl/?lang=en	
Harvard University www.harvard.edu	

and upped throughput to some 400 cells a week. “We can do experiments with thousands of single-cell Hi-C datasets, which allow us to see how a population behaves, but from the single-cell level,” he explains. The data suggest that TADs are less stable than researchers believed. “Even in phenotypically identical cells, the way chromosomes fold is highly variable from cell to cell,” he says.

More recently, Jay Shendure of the **University of Washington’s Department of Genome Sciences** and his colleagues reported in *Science* on a highly multiplexed approach that uses DNA barcoding (rather than physical cell isolation) to extend single-cell Hi-C to thousands of cells at once.

Still other researchers probe chromatin architecture at the single-cell level using fluorescence in situ hybridization (FISH), an inherently single-cell technique that uses fluorescent probes and statistical analysis to infer the relative positioning of chromosomal domains in the nucleus. Heard, for instance, used FISH probes positioned on either side of DXZ4 to visually demonstrate that chromosome folding differs on the active and inactive X chromosomes in mice, thus validating her Hi-C findings.

But unlike C methods, FISH is low throughput (in terms of the number of genomic loci that can be analyzed at once) and low resolution. Xiaowei Zhuang, a Howard Hughes Medical Institute investigator at **Harvard University**, and her colleagues recently provided a solution to the former problem, at least, with the development of a method to multiplex FISH by using Oligopaint probes and sequential hybridization to image tens of thousands of genomic loci and possibly more. Using this approach, the team mapped the position of each TAD on a chromosome one by one. By connecting the dots defined by the resulting signals,

the team effectively traced the 3D topology of a given chromosome in individual fixed cells—a strategy they applied to human chromosomes 20, 21, 22, and X.

According to Zhuang, the resulting data are consistent with population-based Hi-C datasets—for instance, they also catch a glimpse of the DXZ4 “hinge” in the inactive X chromosome—but also reveal spatial features that were not previously known. “The compartment assignments of TADs in our approach, which is an entirely different approach compared to Hi-C ... are essentially identical,” she says. “It’s really remarkable.”

And we’re live

Ultimately, of course, what matters is not chromatin structure in fixed cells, but how that structure changes as the cells grow, divide, and respond to stimuli. “That requires a live-cell technique, and there isn’t a good one actually developed yet,” says Robert Singer, chair of anatomy and structural biology at the **Albert Einstein College of Medicine** in New York.

Many groups are addressing that problem, however, including several funded by the National Institutes of Health Common Fund’s 4D Nucleome program, says Jane Skok, professor of pathology at NYU, who sits on the project’s advisory board. Skok, for instance, has used dead Cas9 nuclease and modified guide RNAs to visually tag specific (albeit repetitive) chromosomal regions, including the immunoglobulin heavy-chain gene, in live cells. Others have published similar approaches. For example, Singer and Wulan Deng, a project scientist at the **Howard Hughes Medical Institute’s Janelia Research Campus**, dubbed their approach “CASFISH,” and Hanhui Ma of the University of Massachusetts Medical School developed a six-color strategy called “CRISPRainbow.”

Another option is ANCHOR, developed by Kerstin Bystricky, head of the Chromatin and Gene Expression group at the **Center for Integrative Biology** and professor at the University of Toulouse, France. ANCHOR involves inserting a handful of binding sites for the bacterial ParB protein near the locus to be imaged, and expressing a ParB-fluorescent protein fusion in the same cells. Unlike Cas9-based approaches, which require one DNA target site for every binding event, ParB has the useful property of tending to accumulate, explains Bystricky. Thus, binding of a single protein dimer quickly nucleates more molecules. As each protein molecule is fluorescently tagged, the net effect is a highly amplified fluorescent signal.

By combining that approach with live-cell RNA visualization, Bystricky’s team has found that actively transcribing genes tend to be less mobile than silent genes—a counterintuitive observation, she says. “There was this belief that transcription activation would increase mobility,” she explains.

According to Dekker, the pieces are falling into place for researchers to begin transitioning from structural characterization to manipulation. But there remains the challenge of data integration, he says. Often, data produced by one method only partially matches that of another. “I’m an optimist,” he says. “I like to think [that FISH and Hi-C methods] are both true. But we don’t understand how they both can be true at the same time.” Given the rapid advancement in chromatin conformation methods, the explanation may not be long in coming.

Jeffrey M. Perkel is a freelance science writer based in Pocatello, Idaho.

DOI: 10.1126/science.opms.p1600108