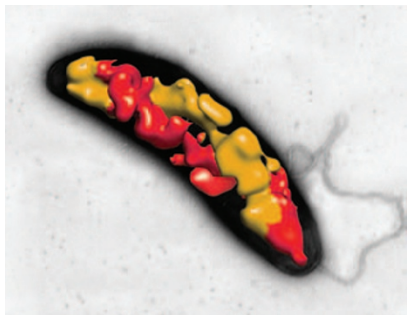
GENOMICS

Three-dimensional genetics

A combination of chromosome conformation capture carbon copy (5C), modeling and automated imaging renders an empirical three-dimensional model of a bacterial genome.



A 3D map of a bacterial chromosome. Chromosome arms are shown in orange and yellow. Image courtesy of J. Dekker.

In 2005 Job Dekker and his team at the University of Massachusetts, Worcester had just developed a technique that would become known as 5C when Dekker was approached by Mark Umbarger, a member of George Church's laboratory at Harvard Medical School, about investigating the organization of a bacterial chromosome.

Dekker found the proposal interesting on many levels. "Bacteria have a different solution to organizing their chromosome," he says. "It does not have nucleosomes, it is circular with an interesting topology, and genetic manipulation is easy." By first identifying and then changing structurally important elements, the researchers expected to be able to look at the interplay between genome structure and cellular functions.

The newly developed 5C method was ideally suited to obtain high-resolution information on the small genome of *Caulobacter crescentus*, about 4 megabases in size. After multiplexed ligation–mediated amplification with primers designed to restriction fragments of the chromosome, the researchers sequenced the resulting 5C library at high throughput, yielding information about the patterns of long-range chromatin interactions of all fragments.

But obtaining this information was only the beginning of a story seven years in the making.

To convert the 5C data into a three-dimensional (3D) model of the bacterial genome, the Dekker and Church groups asked Marc A. Marti-Renom from the Centro de Investigación Príncipe Felipe in Valencia, Spain to join the collaboration. To validate the model, they invited Lucy Shapiro from Stanford University. Her team provided automated fluorescence images that confirmed genomic

SEQUENCING

PEERING DEEPER INTO THE TRANSCRIPTOME

Targeting portions of the transcriptome for deep sequencing reveals very rare transcripts.

Deep sequencing of any transcriptome generates many spurious reads. There is spirited debate over the extent to which these scattered staccato signals, which are easy to dismiss as noise, are a sign of pervasive genomic transcription. Motivated by this question, John Rinn, John Mattick and colleagues working at Harvard University and the University of Queensland have come up with RNA CaptureSeq, a strategy that ramps up sensitivity by targeting only a specific portion of the transcriptome. With Jeffrey Jeddeloh of Roche NimbleGen, they designed tiling arrays to select genomic regions for sequencing at fantastic depths, gaining enough coverage to confidently detect the rarest of transcripts.

Like targeted genomic capture, hybridization of complementary DNA to overlapping probes on the arrays traps just a portion of the transcriptome for sequencing, allowing deeper sampling of the original RNA with the same amount of sequencing—what Mattick likens to a "transcriptomic microscope." To test their idea, the researchers designed an array for ~50 human genes, a few long noncoding RNAs (lncRNAs) and nearly 1,000 intergenic regions bearing epigenetic marks of active transcription. Applying the method to human fibroblast RNA, they achieved over 4,600-fold coverage, or a ~380-fold enrichment of reads that map to targeted regions.

What were the results of this closer scrutiny? "The transcriptional complexity of the genome is far greater than people thought," says Mattick. "It operates at a much finer scale." The researchers found over 200 new protein isoforms, including four from the heavily studied p53

interactions and enabled translation of contact probability into 3D distance. This made it possible to orient the model with respect to landmarks in the cell. "We put all this expertise together," says Dekker, "to build an empirical structure of a genome; then we went one step further to do structure-function studies."

Dekker describes it as a new way of looking at chromosomes, an approach he calls 3D genetics. In the *Caulobacter* genome the teams identified short sequences, known as *ParS* sites, to be essential for anchoring the chromosome to the cell pole. They showed that moving these sites around rotates the chromosome and causes genes to change position. Notably, this swapping of positions did not affect gene expression, but it did lead to segregation defects.

Having proven the concept of 3D genetics in a prokaryote, Dekker is ready to move to higher organisms. Of course, with increasing genome size the complexity of interactions and the stochastic variation between cells increases. For the current modeling approach to work, there needs to be a lot of communality between cells in a population, and it may not be easily applicable to the full yeast or even larger genomes. Dekker estimates that the size limit of genome sections that can be queried is in the several megabase range, anything larger would have too much variation for current modeling methods.

Another limitation to looking at interactions in larger regions is the resolution and scale of the chromosome conformation capture methods.

In 2009 Dekker and his team developed Hi-C, an unbiased approach, in contrast to 5C, that does not rely on the amplification of interacting fragments by specific primers but pulls down all interactions via a biotin link. It allows a much more comprehensive view of global interactions but requires much deeper sequencing. To achieve the same resolution for larger genomes as the 5C approach showed for *Caulobacter*, a sequencing depth would be needed that is currently prohibitive.

But often one is not interested in an entire genome but has more specific questions. Dekker likens 5C to other genome-enrichment methods in which only regions of interest, such as the exome, are captured, amplified and sequenced.

Dekker sees this work as a path forward to identify *cis* elements that define chromatin structure or tether sites to subnuclear regions. He sums it up: "3D genetics will allow one to test the hypothesis that genomes actually have a large number of elements defining its 3D structure."

Nicole Rusk

RESEARCH PAPERS

Umbarger, M.A. *et al.* The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol. Cell* **44**, 252–264 (2011).

tumor suppressor locus. They also detected 163 new lncRNAs near protein-coding regions and gathered robust evidence of lncRNA splicing.

Results from intergenic regions were no less dramatic. The researchers detected transcription from nearly every base, and almost half of the regions gave rise to polyadenylated messages with conventional structure including introns, exons and features typical of lncRNAs. Another key finding was that many of these transcripts were exceedingly rare: present at less than one copy per 1,000 cells. "Each cell is transcriptionally unique," says Mattick, pointing to the scale of cellular heterogeneity that these results imply for development. "It's not 200 cell types; it's a hundred trillion cells doing precise things in different places."

Beyond informing our general view of the transcriptome, the team sees great potential in the method. Tim Mercer, who led the work, notes that many applications can benefit from additional depth. The group is exploring the application to genome-wide association studies that link multiple complex diseases to genomic regions that lack annotated genes. RNA CaptureSeq offers a way to reanalyze these regions for functional changes in rare transcripts. Likewise, small RNAs can be comprehensively mapped using the technique, even if they are expressed from a minority of cells. The bandwidth savings also facilitates greater mutliplexing of sequencing experiments.

"It is almost impossible to provide both a comprehensive and deep view of the transcriptome at current sequencing capacity," says Mercer. Targeted capture can help solve this by focusing on regions of interest with high sensitivity. Many aspects of the transcriptome certainly bear taking such a deep look, with the potential to radically change how we see fundamental concepts such as cell identity and genomic activity.

Tal Nawy

RESEARCH PAPERS

Mercer, T.R. *et al.* Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* advance online publication (13 November 2011).