GENOMICS

Capturing the human methylome

Pairing bisulfite conversion of the human genome with targeted enrichment and high-throughput sequencing allows a quantitative assessment of DNA methylation at base-pair resolution.

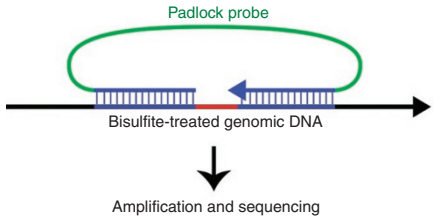
A DNA molecule is not just defined by its genetic components, the primary nucleotide sequence, but also by epigenetic modifications, such as the methyl groups added to cytosines at CpG dinucleotides. These methyl-cytosines, often referred to as the fifth base, are crucial for normal development, whereas aberrant cytosine methylation is the hallmark of many diseases.

To quantitatively measure methylation in the human genome, two independent research teams, led by George Church and Jin Billy Li from Harvard University (Ball *et al.*, 2009) and Kun Zhang from the University of California in San Diego together with Yuan Gao from the Virginia Commonwealth University (Deng *et al.*, 2009), combined traditional epigenetic tools such as bisulfite conversion of DNA with a recently developed targeted genome capture technique and high-throughput sequencing.

Both groups used padlock probes to enrich for selected parts of the genome. These linear oligonucleotide probes are designed so that each end hybridizes on either side of the targeted genomic region, then a DNA polymerase extends one end across the capture region, and after a final ligation step, the now circular DNA is amplified and sequenced.

Previously, Church and his colleague Jay Shendure had used padlock probes for exon capture; although the probes proved to be very specific for the targeted regions, they showed high allelic bias and poor reproducibility. Shendure's team has now addressed these problems for exon capture (Correspondence on p. 315), but using padlock probes for bisulfite-treated DNA poses an extra challenge: during bisulfite conversion, all methyl-cytosines are converted to uracil and subsequently to thymidine, thus reducing the complexity of the sequence and making specific probe design more difficult. Zhang's group, in collaboration with the Church team, optimized the capture protocol, including probe design.

To get a global profile of the methylome, Church and colleagues targeted their probes



Schematic of targeted genomic capture using padlock probes. Reprinted from *Nature Biotechnology*.

to selected genomic regions irrespectively of whether these regions are enriched in CpG dinucleotides, so-called CpG islands, and they complemented this targeted enrichment with a genome-wide counting of cuts after digest with a methylation-sensitive enzyme. After analyzing several cell lines including fibroblasts and induced pluripotent stem cells (iPSCs), they found that gene expression correlated negatively with methylation at the promoter region and positively with methylation in the body of the gene. Church is convinced that this is biologically very relevant, and he adds: "There are cases where a single methyl will matter to a biological function, and we

BIOINFORMATICS

SEQUENCE IS NOT EVERYTHING

A new algorithm for identifying evolutionary constraint incorporates information on local DNA topology, and leads to the finding that this topology is conserved across species.

Think of the structure of DNA, and what most probably comes to mind is the iconic double helix. Or, if you are a bioinformatician, like Elliott Margulies at the National Human Genome Research Institute, you may be prone to thinking of DNA as a string of characters on a computer screen.

But, of course, DNA is much more than this, and in a recent paper, Margulies and his collaborator Thomas Tullius at Boston University, and their colleagues, show that structural properties of DNA are important for how function is encoded in the genome.

In previous work, Tullius and colleagues had developed an algorithm to predict structural profiles of DNA sequences based on their hydroxyl-radical cleavage patterns. This reports on relatively subtle structural properties of DNA, which "have to do with the width of the major and minor grooves of the DNA helix and the bendability of the helix, and not with the chromatin packing that DNase hypersensitivity looks at," as Margulies explains. "These properties are important when a protein is scanning DNA and looking for a place to bind," he says. Notably, DNA molecules of different primary sequences can have similar structures.

This got Margulies thinking about incorporating such structural information into comparative genomics studies. Methods used to identify evolutionarily constrained regions of the genome typically only take primary sequence into account, but a number of functional sequences, in particular in noncoding regions of the genome, are often not identified as constrained by these approaches. "We thought that by incorporating structural profiles, maybe we'd see more subtle similarities in the DNA across species," Margulies says.

The researchers developed a new algorithm, Chai, to identify evolutionary constraint. Chai takes into consideration the effect a base change has on the DNA structural profile. To do this, the scientists first computed the structural profiles of all possible 11-mers and quantified the effect on structure of single-nucleotide substitutions at the central position.

Applying Chai to high-quality comparative sequence data on 36 mammalian species from the Encyclopedia of DNA Elements (ENCODE) pilot project, Margulies and colleagues found that, indeed, compared to a more traditional sequence-based algorithm, their structure-based approach identified about twice as many genomic regions as evolutionarily constrained (within the 1% of the genome included in the ENCODE pilot project). In addition, more of the Chai-identified regions overlap with functional

need to be open to the possibility that methyl groups could be just about anywhere." To follow up on this hypothesis, the Church team is working on high-throughput methods for allele-specific methylation.

The Zhang and Gao groups, in contrast, focused mostly on CpG islands, partly because those are the regions with higher methylation, and partly because they are clearly defined and thus present a stable set of targets. They compared the methylation patterns in all CpG islands on two chromosomes in iPSCs and human embryonic stem cells (hESCs). To their surprise, the researchers noted that only 10% of the regions show a difference in methylation between the cell lines. For Zhang, this underscores the advantages of a targeted strategy over genome-wide sequencing. "Full methylome sequencing is not cost-effective," he concludes, "because 90% of your data will not give you too much information."

As Church's and Li's teams, Zhang and his colleagues saw decreased promoter methylation and increased gene body methylation in highly expressed genes. In addition, they observed that the methylation patterns of iPSCs and hESCs differ. Zhang describes their findings: "iPSCs tend to be more methylated… and this could be causing an extra effort to do the re-differentiation." To assess this difference in more detail, Zhang plans to look at the methylation state in 'clean' iPSCs, that is, cells free of inducing factors, and their intermediate and fully differentiated descendents.

With these techniques, the role of the fifth base is becoming a lot more prominent.

Nicole Rusk

RESEARCH PAPERS

Ball, M.P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27**, 361–368 (2009).

Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **27**, 353–360 (2009)

sequences, like DNase I–hypersensitive sites and putative enhancers. Coding regions, in contrast, were significantly underrepresented. "When we first saw this, we were upset," says Margulies, "but then we realized that it underscores that the algorithm is picking up things that are not encoded in the primary sequence, since we know that in coding regions the primary sequence is important."

Might the effect that a sequence change has on structural profiles be a way to identify functional changes in noncoding regions of the genome? Possibly, says Margulies, although the experiments are still in progress. Notably, when the researchers examined single-nucleotide polymorphisms from the PhenCode project that are known to have associated phenotypes, and compared them with a set of neutral single-nucleotide polymorphisms, they found that the phenotype-causing variants are significantly more likely to produce large structural changes in the DNA.

Whether this approach will have predictive value for identifying functional noncoding variants remains to be seen, but the prospect is certainly exciting. And, as Margulies emphasizes, there may be still other informative ways of looking at DNA, "a living molecule, not just a sequence of letters," after all.

Natalie de Souza

RESEARCH PAPERS

Parker, S.C.J. *et al.* Local DNA topography correlates with functional noncoding regions of the human genome. *Science* advance online publication (12 March 2009).

NEWS IN BRIEF

STEM CELLS

Human induced pluripotent stem cells with a clean genome

Reprogramming of differentiated cells to induced pluripotent stem cells was initially achieved using integrating viruses to deliver the reprogramming factors. This can cause abnormalities in the cells and is undesirable particularly for therapeutic applications. Yu *et al.* now reprogram human fibroblasts using the oriP/EBNA1 vector to deliver the reprogramming factors. This vector forms a stable episome and is lost from cells in the absence of drug selection.

Yu, J. *et al. Science* advance online publication (26 March 2009).

PROTEOMICS

Finding inhibitors with ABPP

Activity-based protein profiling (ABPP), which uses reactive chemical probes to target enzyme active sites, can be used to elucidate the function of uncharacterized enzymes. Bachovchin *et al.* now adapt ABPP to a competitive inhibitor screen using fluorescence polarization, which monitors the apparent size of a fluorophore-tagged activity-based probe (whether it binds to the enzyme or not). This allows them to screen very large compound libraries, unlike the previous gel-based readout.

Bachovchin, D.A. *et al. Nat. Biotechnol.* **27**, 387–394 (2009).

GENOMICS

Inherited transcriptional errors

Mutations that are passed on to daughter cells are usually caused by changes in the DNA sequence. Errors in transcription, leading to mutation in the RNA, are as short lived as the RNA itself and thought to have no impact on the heritable phenotype. Gordon *et al.* now show with a bistable *lac* operon that the molecular noise caused by transcriptional errors can trigger a positive feedback loop that results in a heritable phenotypic difference in genetically identical cells in the same environment.

Gordon, A.J.E. *et al. PLoS Biol.* **7**, e1000044 (2009).

PROTEIN BIOCHEMISTRY

Reverse micelles for nuclear magnetic resonance

Membrane proteins are notoriously difficult to study with nuclear magnetic resonance spectroscopy owing to their need to be solubilized in large detergent micelles, which prevents rapid molecular tumbling needed for optimal data collection via triple-resonance experiments. Kielec *et al.* now use reverse micelles, a micelle that has flipped its orientation in a low-viscosity organic solvent, and thus facilitates fast molecular tumbling, to investigate the structure of a potassium channel.

Kielec, J.M. *et al. Structure* **17**, 345–351 (2009).

MICROSCOPY

Ultrastable AFM

Instrumental drift in atomic force microscopy (AFM) is a critical problem that limits imaging resolution. Sharper tips and high-sensitivity detection methods can improve resolution, but rapid scanning is still required to minimize instrumental drift. King *et al.* now describe a different solution to stabilize the AFM stage: they scatter laser light off the apex of the AFM tip to create a local frame of reference. This permits them to control the position of the AFM tip with high precision, allowing them to scan slowly, improving imaging resolution.

King, G.M. *et al. Nano Lett.* **9**, 1451–1456 (2009).