

## TECHNOLOGY FEATURE

# READING THE SECOND GENOMIC CODE

*Transient changes to the genome make its code more complex to interpret but they still put a gleam in the eye of drug and technology developers.*

ALENGO/ISTOCKPHOTO



BY VIVIEN MARX

**D**NA is famous as the instruction manual of life — the multi-billion-base-pair data tape that directs how a fertilized egg turns into the specific cells, tissues and organs of, say, a sharp-eyed soccer pro who is musically inclined but who also battles depression.

But DNA works with many partners, including ‘epigenetic’ factors, which influence gene expression in ways that don’t involve changes to the underlying sequence (see ‘Polygamous DNA’). An important example is methylation, in which methyl groups are tacked on to various locations along the double helix to control the activity of particular genes. Methylation

also affects histones, the spool-like proteins around which DNA is tightly wound inside the nucleus: the chemical modifications help to control when this protein–DNA complex, called chromatin, opens up so that the genetic instructions can be read.

Figuring out when and how such epigenetic changes get made — or damaged — has become a crucial part of scientists’ efforts to understand both the normal development of cells and their progression into cancer and other diseases. It can be painstaking work. Sometimes, says Andrew Feinberg, an epigeneticist at Johns Hopkins University in Baltimore, Maryland, the available techniques often pick up only “little biochemical shadows” of events going on at a particular location, while

the complete set of players and their mechanisms remain mysterious. And even when you can identify an epigenetic molecule, says Tony Kouzarides, a molecular biologist at the University of Cambridge, UK, “you have to work out why it is there, and what it is doing there”.

Nonetheless, epigeneticists have made remarkable progress over the past two decades. Their tool kit now includes advanced sequencing techniques, targeted antibodies and even laser cell sorting — and it should soon encompass ultrasensitive nanofluidic and nanopore sequencing methods. The community is also turning to advanced bioinformatics to cope with the sheer volume of data — especially the wealth of epigenetic information from the Encyclopedia of DNA ►

► Elements (ENCODE) project, which this year released more than 1,600 genome-wide data sets covering more than 100 cell types<sup>1</sup>.

Technology development is now kicking into high gear as epigenetics researchers push to decipher the genome's many partners, and to deepen understanding of health and disease.

## BEYOND A PRECIPITATING HEADACHE

The standard method used to study epigenetic histone modifications is called chromatin immunoprecipitation (ChIP), coupled with sequencing<sup>2</sup>. The basic idea is to shear DNA while it is still wrapped around the histones, use antibodies to capture specific protein–DNA complexes from the fragments, and then study which DNA sequences are attached to which proteins. The approach helps to unpick how the interactions are tuning genes — activating some, silencing others.

The technique has its drawbacks, however. Sriharsa Pradhan, an RNA biologist at New England Biolabs in Ipswich, Massachusetts, says that he is often unable to reproduce work from published epigenetics studies. “Most of the failures happen if the antibody is not good,” Pradhan says. It might pick up too many DNA–protein complexes — “every Tom, Dick and Harry” in a sample — and so does not offer the resolution that scientists seek.

Kouzarides agrees that the quality of the antibody matters greatly for ChIP and many other lab procedures. That's what led him to co-found Abcam, an antibody supplier with headquarters now in Cambridge, UK. The goal is exceptionally high quality, says Kouzarides, who is on Abcam's board of directors — but it is a constant struggle. “You are at the mercy of the rabbits,” he says, referring to the animals used to generate the antibodies. “Some generate good antibodies, some generate bad antibodies” — and there is no predicting which is which.

Monoclonal antibodies could offer more reliability, says Kouzarides, because they avoid the problem of batch-to-batch variability. But for reasons that still aren't clear, he says, some of them do not work well for ChIP. For now, the field has to use animals to generate the antibody mixes useful for ChIP. “You have to put up with the unreliable nature of antibodies because it's the only way to do such experiments at the moment,” he says.

Another drawback with standard ChIP is its bias, says Alan Tackett of the University of Arkansas for Medical Sciences in Little Rock. Although the technique lets scientists localize

a specific protein acting on a genomic site, “you have to know what protein or histone modification you are targeting”. And scientists need to have on hand an antibody that matches the protein of interest. So ChIP is not easily multiplexed to profile multiple areas of the genome at the same time.

In response to this shortfall, Tackett and his Arkansas colleagues, along with scientists at the Johns Hopkins School of Medicine, have developed chromatin affinity purification with mass spectrometry (ChAP-MS)<sup>3</sup>. The approach involves cutting out a 1,000-base-pair region of a chromosome, purifying it and determining all the epigenetic changes that are present. The team has used the approach in yeast to detect different chromatin states, silenced genes and other regions in which genes are still active. And Tackett says that around ten other labs have begun exploring it, too.

He is now readying the technique for use in human cell lines and tissues. “We are working on the mammalian version and anticipate having that complete within the year,” he says. One challenge for ChAP-MS is that the analysis requires  $10^7$  to  $10^{10}$  cells, so Tackett and his colleagues are trying to lower that number. And Tackett is confident about the technology's promise. “We see this ultimately taking the place of ChIP in epigenetics labs,” he says, with mass spectrometry being available through proteomics core facilities on campuses, he says.

Other scientists are proposing different alternatives to ChIP, which “is not a very efficient process”, says Paul Soloway at Cornell University in Ithaca, New York. In addition to the challenges involved in sample processing, ChIP usually queries just one epigenetic mark at a time in a population of cells. That means that the results of multiple ChIP-seq experiments have to be aligned to determine if some cells have one mark and others have another, or if, perhaps, all cells have both.

Soloway wants to offer scientists greater resolution for ChIP analysis. He also wants the approach to be scalable, delivering detail and screening for multiple epigenetic marks in a single experiment. His answer is a nanofluidic device based on a silica wafer that is in the prototype stage and which comes in two formats<sup>4</sup>. One of them quantifies the molecules with at least one epigenetic mark. The other, a

branched nanofluidic device, sorts and quantifies the molecules. Using fluorescent labels and optics-based sorting, the molecules are shunted to one chamber or another for later analysis, such as DNA sequencing. “Because silica is clear and non-fluorescent, we can make measurements of individual molecules using highly sensitive optics,” he says.

Ultimately, Soloway would like to be able to go through whole genomes in a rapid, multiplexed way. He says that standard ChIP is still ahead of his technique because it can generate materials in the amounts needed for sequencing, whereas he still needs to get from single molecules to the pico- and nanograms needed.

Soloway believes that his technology will find a home in drug development, helping researchers to quickly and quantitatively characterize how drug candidates affect epigenetic marks. Clinical applications could include helping to monitor how patients fare when treated with epigenomic drugs, and identify how epigenetic marks vary during the course of a disease such as cancer, he says. In January, together with the Cornell engineers Harold Craighead and Stephen Levy who worked on the technology, he founded Odyssey Molecular in Ithaca, to commercialize the device.

## FINDING OTHER MARKS

DNA methylation has important roles in cells, including the regulation of genes during development and disease. One of several methods used to find these sections of the genome is methylated-DNA immunoprecipitation, which uses an antibody that locates 5-methylcytosine, a methylated form of the DNA base cytosine.

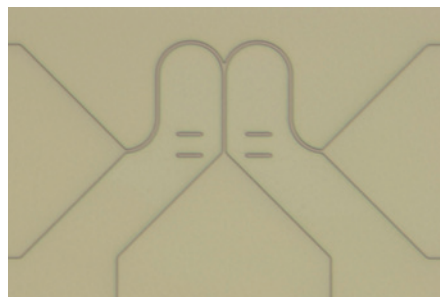
A different approach targets methylated parts of the genome in ‘CpG islands’, which are characterized by a specific chemical bond between the DNA bases cytosine and guanine. In an analysis of methylation levels for 240,000 of the several million CpG islands in the ENCODE data, John Stamatoyannopoulos at the University of Washington in Seattle and his colleagues found a strong association between methylation and accessibility for genes to be read<sup>5</sup>. As Wendy Bickmore from the Medical Research Council Human Genetics Unit at the University of Edinburgh, UK, notes, the results support the idea that DNA methylation is blocked where the transcription factors that read DNA bind. This mechanism, she says, is relevant to the interpretation of disease-associated sites that show altered DNA methylation<sup>6</sup>.

One widely used technique to determine DNA methylation patterns across a genome is bisulphite sequencing. The addition of bisulphite to DNA converts cytosine to uracil, but skips methylated cytosines, thereby allowing the methylation status of DNA segments to be determined through high-throughput sequencing. Many companies offer bisulphite conversion kits. “It's cheap enough now and there are statistical tools for understanding it, so there's no reason to use another method,” says



**“Whatever you see in one moment will change in the next.”**

Tony Kouzarides



**A nanofluidic device can sort through DNA molecules to find those with epigenetic marks.**



Feinberg.

Yet detecting methylation is time-consuming, so scientists in academia and industry have been exploring ways to improve the approach. Some teams, including one at Osaka University in Japan and one at the University of Oxford, UK, are exploring the use of nanopores, tiny gates through which to run a DNA strand. And Pacific Biosciences, a sequencing firm in Menlo Park, California, is using tags to prepare single strands of DNA for high-throughput sequencing.

At Washington University in St Louis, meanwhile, Rob Mitra is leading an effort to be more precise in capturing methylation data, because this information can, for example, be an early sign of tumour development. Mitra and his team, including graduate student Maximiliaan Schillebeeckx, have developed a technique that uses lasers to separate out the cells of interest. He calls the technique laser capture microdissection–reduced representation bisulphite sequencing. Among the advantages, says Mitra, is that the technique covers “the majority of the CpG islands and it’s relatively inexpensive”.

Reduced representation bisulphite sequencing is similar to whole-genome bisulphite sequencing, but sequences only the parts of the genome that include CpG-dense regions. The technique uses enzymes to cut up purified genomic DNA into fragments that contain CpG islands. The fragments are then processed, and those of a certain size are subjected to bisulphite conversion, amplified and then sequenced.

The approach is geared to work on small amounts of DNA — perhaps even less than a nanogram — and in formalin-fixed, paraffin-embedded tissue, which is “typically not in as good shape as good fresh frozen DNA”, Mitra says. This type of tissue fixation is typically used in biobank samples.

His technique could be a tool for researchers who work with specific cell types or with complex tissues, such as neurological samples, in which it is hard to isolate the cell type of interest, he says. The method also avoids the need for multiple labour-intensive purifications. And, he says, “at each point in space, you get a genome-wide profile of methylation, so now you can start to correlate methylation profiles spatially”, Mitra says. A researcher can see, for example, if similar regions of complex tissue are methylated similarly. By coupling genome-wide methylation analysis with laser capture to isolate targeted cell populations, the tool can help researchers to address questions in these challenging tissues, he says.

#### EXPANDED REACH

Along with the flood of data that ENCODE brought to epigenetics came data standards, quality metrics, software tools and ways to convey how experiments are done, allowing comparisons between labs. This development has heightened awareness about the “good technologies” needed to study how the genetic

code is put into action, says Adam Pettersson, a senior scientist at Zymo Research in Irvine, California, which is one of many companies offering epigenetics services to academics as well as drug-discovery companies.

Such awareness is going to become ever more important as epigenetics grows to encompass not just multiple cell types, but multiple species. The modENCODE project ([www.modencode.org](http://www.modencode.org)) is mapping regulatory patterns in two frequently used model organisms, the fruitfly *Drosophila melanogaster* and the nematode worm *Caenorhabditis elegans*, and the Mouse ENCODE consortium is focusing on epigenomic mapping of the mouse. “A huge way to understand function is by comparative epigenomics,” says Feinberg, who would like to see efforts across many more species.

These developments will inevitably require increased reliance on massive computation, says Kouzarides, who sees bioinformatics as a rate-limiting step in epigenetics. Researchers need ways to integrate and do global analyses



**“Now you can start to correlate methylation profiles spatially.”**

Rob Mitra

of the emerging maps of epigenomic marks and their effects, as well as ways to do high-resolution analyses, preferably at the single-cell level (see page 27). Without such computational tools, Kouzarides says, “it’s almost impossible to appreciate the complexity of the information”.

For scientists who would rather not dig into the data themselves, Michael Snyder and his team

at Stanford University in California have developed Regulome-DB ([regulome.stanford.edu](http://regulome.stanford.edu)), an automated tool to explore non-coding regions of the human genome. Manolis Kellis at the Massachusetts Institute of Technology in Cambridge and his group have set up Haplo-Reg ([www.broadinstitute.org/mammals/haploreg](http://www.broadinstitute.org/mammals/haploreg)), a tool that helps to link non-coding variant patterns to possible clinical conditions.

#### TRANSIENT DRUGS

The potential for clinical applications is an important motivator for epigenetics research. The transient nature of epigenetic changes gives drug developers and biomedical researchers reasons to dream about how their efforts might reverse changes that contribute to disease. “Those sorts of things that are more malleable are likely the things that we can target,” Feinberg says.

Four drugs that act on epigenetic pathways have been approved by the US Food and Drug Administration (FDA), and the next wave of candidates is being readied in biotech and

pharmaceutical companies. Kouzarides, for example, is looking at chromatin modifications and develops drug candidates that could right the wrongs in cancers in which, for example, epigenetic influences lead to the misregulation of cell pathways<sup>7</sup>.

Targeting an aggressive form of leukaemia for which treatments are lacking, Kouzarides and his team have explored how to inhibit bromodomain and extraterminal (BET) proteins and remove them from chromatin. BET proteins belong to a class of epigenetic reader that targets histones, recruits multi-protein complexes to the spot where they attach and instructs cellular processes involved in reading genetic information.

The journey from the lab to the clinic is not usually quick, Kouzarides says. In this case, however, a candidate under development for inflammation was found to be applicable for the leukaemia. Now, the small-molecule inhibitor of the BET protein is in clinical development at GlaxoSmithKline, headquartered in London.

Kouzarides believes that chromatin-modification pathways are promising drug targets because they involve proteins interacting with other proteins. In the past, drugs have tended to target enzymes, and it has not been considered feasible to target protein–protein interactions with small molecules. But his work<sup>8</sup>, along with that of others, has shown that it is possible to develop specific small molecules against the BET proteins that recognize a small epigenetic modification present on chromatin.

Constellation Pharmaceuticals in Cambridge, Massachusetts, is also exploring the BET family, as well as other enzymes that modify chromatin<sup>9</sup>. These therapies are going to be part of the second-generation epigenetic drugs that target specific modifications with a role in disease, explains Keith Dionne, the company’s president and chief executive. The past, more coarse scientific understanding of chromatin has shifted to an appreciation of the “subtle distinctions” between chromatin states, explains James Audia, the company’s chief scientific officer.

Earlier this year, Constellation and Genentech began collaborating on the development of inhibitors of epigenetic modifiers. Constellation also has its own programmes targeting inhibitors of BET proteins and another class of epigenetic modifier, the EZH2 chromatin-writers. These proteins seem to be part of a complex that represses gene expression; mutated versions have been linked to some cancers.

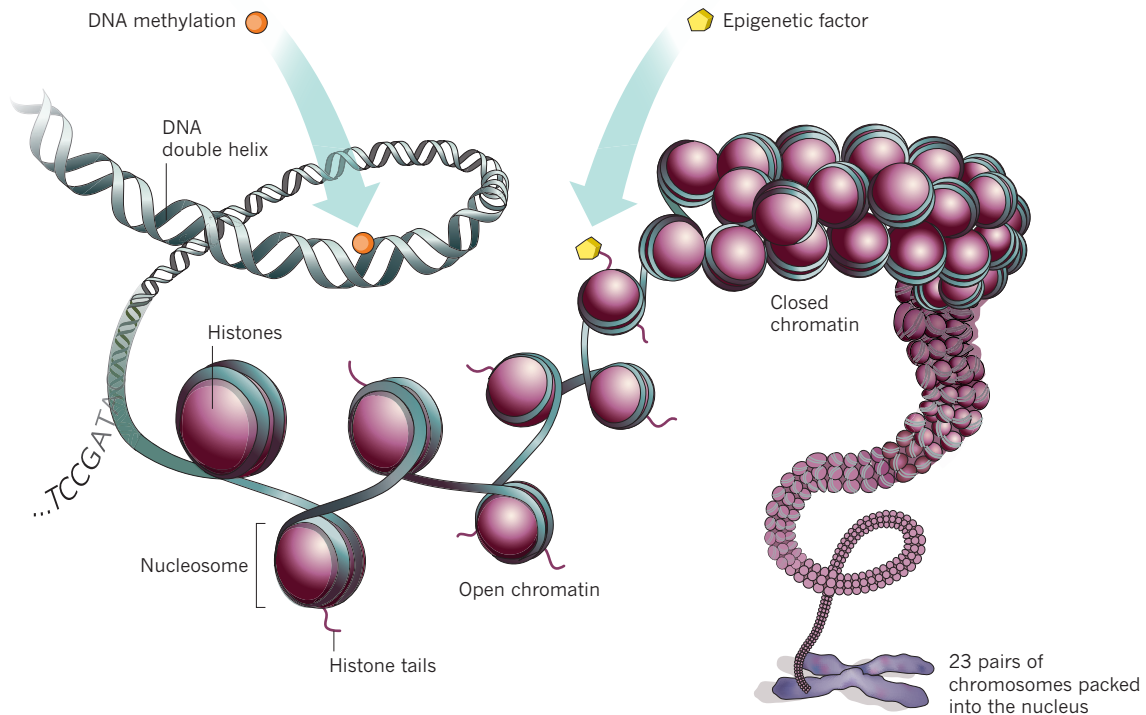
As Patrick Trojer, director of biology at Constellation Pharmaceuticals explains, cancers use chromatin modification to gain an advantage, for example to inactivate a pathway that creates room for unhindered tumour growth. As part of the company’s drug-discovery programme, he and his colleagues develop techniques to study the details of chromatin changes. The understanding of chromatin biology is one of the company’s strong suits, he says.

To support this application-based research,

B. BOSTON

## POLYGAMOUS DNA

DNA works with many partners. DNA methylation, for example, influences the way that genes are expressed without changing the underlying DNA sequence, and other epigenetic factors bind to histones to control when chromatin complexes open up and allow their DNA to be read.



Trojer and his colleagues use a number of epigenetic techniques. ChIP-seq is a lab standard in which antibodies are “the key” to the technique, he says. But the company has also made histone mass spectrometry a priority, because it allows the scientists to query the chromatin changes without using antibodies and to query a number of modifications at once. The company set up an in-house high-throughput facility to screen for potential compounds.

Although other companies tend to outsource these tasks, the company wants to integrate findings about chromatin biology into drug discovery with an in-house suite of tools that includes mass spectrometry and biophysics analyses, Dionne explains.

Another Cambridge-based epigenetics company, Epizyme, focuses on a family of proteins called histone methyltransferases. These epigenetic modifiers act on histones, by catalysing the transfer of methyl groups onto specific positions in the protein. The company has partnerships with the pharmaceutical companies Glaxo-SmithKline; Celgene Corporation in Summit, New Jersey; and Eisai in Woodcliff Lake, New Jersey, as well as the Leukemia and Lymphoma Society in White Plains, New York, and the Multiple Myeloma Research Foundation in Norwalk, Connecticut. So far, 96 histone methyltransferases have been identified in humans, says Robert Copeland, Epizyme’s chief scientific officer. “We believe there are at least 20 of those enzymes that are high-value targets for human cancers.”

The company’s goal is to find a molecule that

blocks an enzyme active in an epigenetic pathway but not its nearest neighbours, he says. It is a selectivity that has been difficult to come by in the development of biotherapeutic drugs.

Copeland believes that epigenetics drugs fit into a trend of defining a cancer not by its anatomical location but by its molecular profile, which includes epigenetic signatures. Like many companies in this field, he and his colleagues mine the publicly available databases, noting that many genetic alterations in epigenetic pathways are found in human cancers.

Kouzarides believes that many cancer cells will be very vulnerable to epigenetic drugs because they rely on only one or two epigenetic pathways, whereas normal cells draw on several pathways for their functions. At the same time, he believes that epigenetics researchers and technology developers will still want to develop and refine experimental methods, for example to explore the three-dimensional structure of epigenomic events, to see how chromatin is changing throughout the genome. “It’s very difficult to look at chromatin itself,” he says. “Technology still has to evolve to look at *in vivo* chromatin effects.” The available epigenetic data are “extensive, but still a very small snapshot” of epigenetic changes, he says. They represent a situation at a specific time in a specific cell.

Epigenetics might find its way into preventive medicine, too. Scanning the epigenome could be a way to detect disease well before symptoms arise. The blood pricked from the heels of newborn babies is one way to begin. In many countries, the blood spots are placed on Guthrie

cards and stored indefinitely by hospitals and health-care systems. Scientists at Queen Mary, University of London are exploring how DNA methylation patterns change between newborns and in cells from the same children when they are three years old. Differences in epigenetic marks could be clues to health.

If the sequencing companies are betting right, then genome sequencing could become commonplace for many patients, perhaps even part of an annual physical examination. An epigenetic read-out, updated at regular intervals, might be an important companion file to that genome sequence.

But this type of progress depends on deeper understanding of epigenetic mechanisms and technology that has yet to evolve, Kouzarides says. Because epigenetic events change constantly in the cell, “whatever you see in one moment will change in the next”, he says. ■

**Vivien Marx** is technology editor at *Nature* and *Nature Methods*.

1. The ENCODE Project Consortium *Nature* **489**, 57–74 (2012).
2. Park, P. J. *Nature Rev. Genet.* **10**, 669–680 (2009).
3. Byrum, S. D., Raman, A., Taverna, S. D. & Tackett, A. J. *Cell Rep.* **2**, 198–205 (2012).
4. Cipriany, B. R. *et al. Proc. Natl Acad. Sci. USA* **109**, 8477–8482 (2012).
5. Thurman, R. E. *et al. Nature* **489**, 75–82 (2012).
6. Ecker, J. R. *et al. Nature* **489**, 52–55 (2012).
7. Dawson, M. A. & Kouzarides, T. *Cell* **150**, 12–27 (2012).
8. Dawson, M. A. *et al. Nature* **478**, 529–533 (2011).
9. Mertz, J. A. *et al. Proc. Natl Acad. Sci. USA* **108**, 16669–16674 (2011).