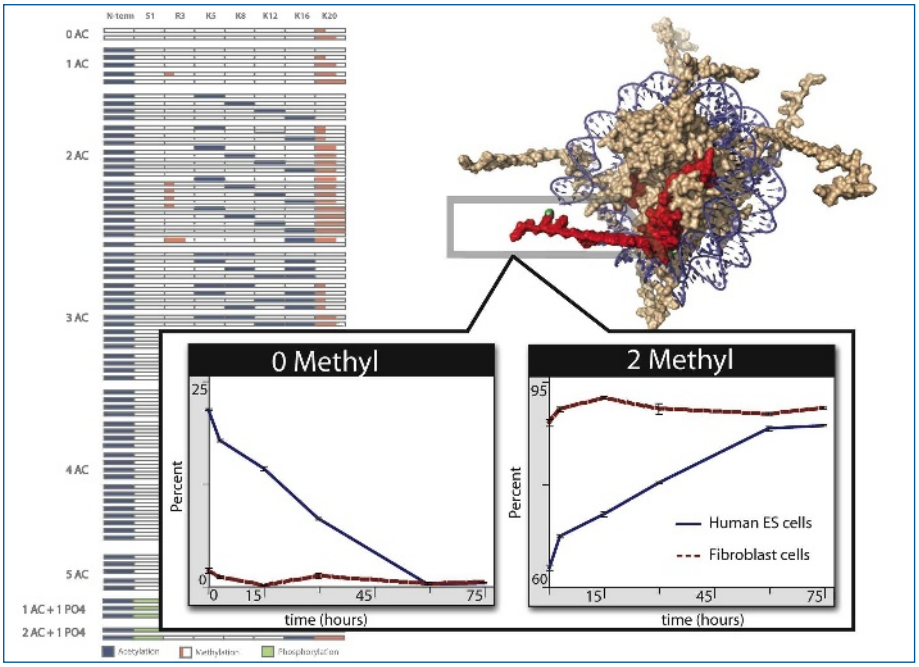
Detailed analysis

Researchers now have access to an burgeoning collection of tools for unravelling the epigenome, which could lead to new drug targets and ways to track disease. Laura Bonetta reports.

Just as DNA mutations can cause disease, so too can changes in gene activity that don't depend on DNA sequence. For example, the addition of methyl groups to cytosines within the promoter regions of a gene, a process known as DNA methylation, typically represses the gene's activity. Similarly, the methylation, acetylation or ubiquitination of certain amino acids on histones — the main protein components of chromatin — can change the way DNA is wrapped around the histones and can affect which nearby genes are available for activation by regulatory proteins such as transcription factors.

Such 'epigenetic' processes can have dire consequences. Many cancers and other diseases have aberrant DNA methylation at particular sites in the genome. And certain patterns of histone modification are key to biological events such as embryonic development, ageing and cell-cycle regulation.

"We have found a gene that is highly methylated in more than 90% of colorectal cancers, regardless of stage and histology," says Achim Plum, senior vice-president for corporate development at Epigenomics in Berlin,



J. COON

Mass spectrometry analysis of histone H4 isoforms in human stem cells. When differentiation was induced in the cells, levels of unmethylated isoforms fell and dimethylated isoforms rose.

TACKLING THE EPIGENOME

So far, the approach to the epigenome has been somewhat piecemeal, with disparate groups cataloguing just the modifications in their cells or organisms of interest. Technological advances that allow faster and cheaper mapping of epigenetic modifications are now helping to unite these efforts.

The Human Epigenome Project, for example, was established in 1999, when researchers in Europe teamed up to identify, catalogue and interpret genome-wide DNA methylation patterns in human genes. The project is a collaboration between the Wellcome Trust Sanger Institute in Hinxton, UK, the National Centre for Genotyping near Paris, France, and the German company Epigenomics in Berlin. The consortium completed a pilot study of methylation patterns within the major histocompatibility complex followed by a more thorough analysis of the methylation patterns on three chromosomes. After showing the feasibility of such an approach "the consortium has decided to leave it to others to continue the endeavour", says Achim Plum, senior vice-president for corporate development at Epigenomics.

ENCODE, or the Encyclopedia of DNA Elements, was launched in September 2003 by the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. The project aims to identify all the functional elements in the human genome sequence. Following initial success, the NHGRI last October pledged $80 million over four years to scale up the project. Many of the functional elements so far identified by ENCODE are being studied for changes in methylation under a range of conditions.

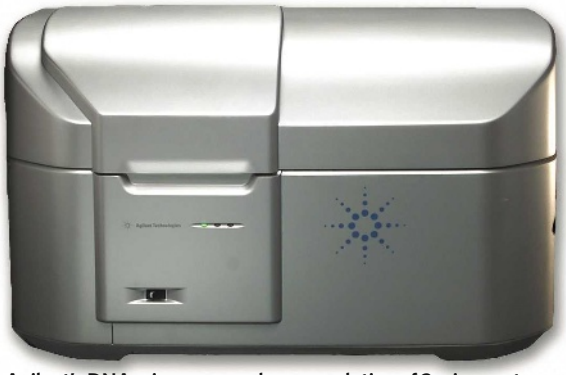
The Epigenome Network of Excellence, headed by Thomas Jenuwein of the Research Institute of Molecular Pathology in Vienna, Austria, brings together 81 laboratories from 10 European countries under a €12.5-million (US$19.7-million) grant from the European Union to advance epigenetic research over the next five years. The network, which officially began in June 2004, supports members and non-members so that they can attend conferences, workshops and training visits, as well as offering them shared resources.

The US American Association for Cancer Research is championing the formation of the Alliance for the Human Epigenome and Disease (AHEAD) project. Under the leadership of Peter Jones, director of the University of Southern California's Norris Comprehensive Cancer Center in Los Angeles, AHEAD would coordinate an international interdisciplinary project to map a defined subset of epigenetic markers in a limited number of human tissues at different stages of development (see page 711). This would provide a reference epigenome to which samples from various diseases could be compared. The alliance would also try to develop a bioinformatics infrastructure to support the collection of epigenomic data.

AHEAD's planned efforts will be helped by January's announcement from the National Institutes of Health that it will invest more than $190 million over the next five years to accelerate epigenomics research. Grant applications are now being accepted for epigenome mapping centres, epigenomics data analysis and coordination, and technology development. L.B.

Germany. "You will have a hard time finding a single somatic mutation that is as prevalent in this particular cancer."

Because these epigenetic changes can potentially be reversed by drugs, they are good targets for the prevention and treatment of disease. To assess the effect of epigenetic regulation on health and disease, researchers are now cataloguing epigenetic variations across the genome — or epigenome — in different tissues and at various stages of development.



AGILENT TECHNOLOGIES

Agilent's DNA microscanner has a resolution of 2 micrometres.

Chips with everything

Cytosine is typically methylated when it is next to guanine in what is known as a CpG dinucleotide. Although most CpGs are methylated in mammals, some are not methylated and these are usually grouped in clusters called CpG islands. These tend to be located in the 5′ regulatory regions of genes. In many cancers, these CpG islands become hypermethylated, resulting in the heritable silencing of transcription of downstream genes.

DNA methylation can be detected in several ways. One method compares what happens when genomic DNA is digested by enzymes that are either sensitive or insensitive to methylation. Another approach uses chromatin immunoprecipitation, or ChIP. This involves crosslinking DNA with its associated proteins and then shearing the DNA. The fragments that contain methylated cytosine are extracted by immunoprecipitation with antibodies specific for 5-methylcytosine or fragments associated with other proteins such as transcription factors or histones. The immunoprecipitated DNA is purified, amplified and labelled with a fluorescent tag. This is then applied to the surface of a DNA microarray containing a set of probes — a procedure commonly referred to as ChIP-on-chip.

"The genome is methylated at a low level everywhere, but it is more of a sprinkling. What we are looking for are regions with higher densities of methylation," says Mary Harper, chief scientific officer at Genpathway in San Diego, California. "These can be regulatory regions and there are tens of thousands of these regions across the genome."

ChIP-on-chip has grown in popularity with the availability of high-density arrays of contiguously tiled oligonucleotide probes. Affymetrix of Santa Clara, California, offers whole-genome tiling yeast arrays with a resolution of 5 base pairs (a total of 3.2 million probes). The company also makes human and mouse whole-genome tiling arrays, each as a set of 14 arrays containing around 45 million probes at a spacing of 35 base pairs.

Along with other companies such as NimbleGen in Madison, Wisconsin, and Agilent Technologies in Santa Clara, California, Affymetrix provides several array formats for ChIP-on-chip experiments. In addition to whole genome arrays, these companies sell CpG islands arrays, promoter arrays, ENCODE arrays (see 'Tackling the epigenome') or custom-made arrays. The choice of the array depends mainly on the type of experiment being done, the resolution needed and cost.

"Different arrays are not that different,"

TOOLS OF THE TRADE

For researchers tackling the epigenome, there are two key tools: bisulphite conversion kits and ChIP-grade antibodies.

Getting complete conversion of unmethylated cytosine to uracil has been one of the main challenges for the bisulphite treatment of DNA, not least because the reaction conditions can cause the DNA to degrade. A number of companies have tackled the problem.

"There are incredibly good kits for bisulphite conversion," says Achim Plum, senior vice-president for corporate development at Epigenomics in Berlin. "If you use home brews you can make many mistakes, but with kits this is no longer an issue." QIAGEN of Venlo, the Netherlands, sells EpiTect, a kit based on Epigenomics' technology that Plum says offers nearly 100% conversion without DNA degradation. In June, QIAGEN expanded the EpiTect product line to include a standardized workflow for methylation analysis from sample collection, stabilization and purification to bisulphite conversion, and real-time endpoint PCR methylation analysis or sequencing.

Epigentek in Brooklyn, New York, also sells epigenetic products and kits for DNA methylation and histone modification. Its Methylamp DNA modification kit takes two hours to convert cytosine to uracil with more than 99% accuracy and requires only 50 picograms of starting DNA, according to Adam Li, the company's chief scientific officer. The sample can then be applied for methylation-specific PCR or a methylation array.

In June last year, Sigma-Aldrich of St Louis, Missouri, launched a number of products for epigenetic research based on Epigentek's technologies. Other providers of bisulphite conversion kits and related epigenetics products include Human Genetic Signatures in Sydney, Australia; Invitrogen of Carlsbad, California; Promega of Madison, Wisconsin; and Zymo Research of Orange, California.

EPIGENTEK

Epigentek's Methylamp kit takes two hours to convert cytosines to uracil.

When it comes to antibodies that can be used in ChIP experiments, a fresh set of problems presents itself. "For most transcription factors there are not that many antibodies," says Kevin Struhl, a ChIP pioneer at Harvard Medical School in Boston , Massachusetts. "And a lot of antibodies don't work well on ChIP."

ChIP-on-chip requires highly specific antibodies that recognize a protein's epitope in free solution and under fixed conditions. "Because an antibody that works well in ChIP must recognize the target while it is bound to DNA, the epitope that the antibody recognizes must be available in a spatial conformation that is not obscured by the shape the protein takes while bound to DNA," explains Dana Meents, a product manager at Active Motif in Carlsbad, California. Active Motif is one of several companies now selling products that have been ChIP validated.

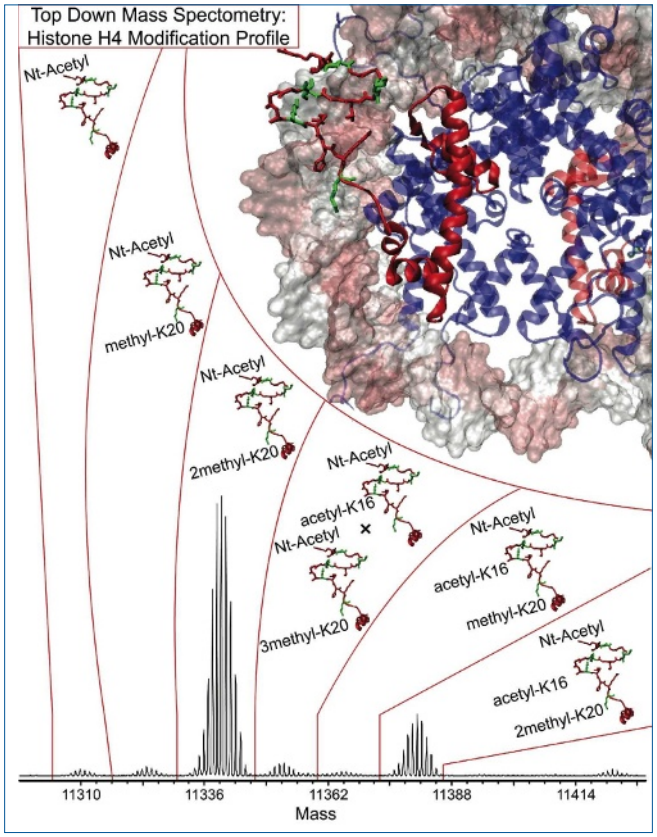
Abcam in Cambridge, UK, sells just over 200 ChIP-grade antibodies, many of them specific to histone post-translational modifications.

As an alternative to antibodies, the HaloCHIP system developed by Promega allows researchers to perform ChIP-on-chip experiments by cross-linking DNA to a protein of interest fused to a modified haloalkane dehalogenase tag (HaloTag). The protein–DNA complex can then be captured on a resin that recognizes the HaloTag for further microarray analysis. L.B.

N. KELLEHER

says ChIP pioneer Kevin Struhl at Harvard Medical School in Boston, Massachusetts. He notes that the greatest source of variation among ChIP-on-chip experiments comes from the sample preparation and immunoprecipitation steps. He and his colleagues have found that although most commercially available arrays perform similarly, those carrying longer oligonucleotides, such as the ones from NimbleGen (50–75-mer) or Agilent (60-mer), are slightly more sensitive than shorter ones for lower levels of enrichment[1].

Agilent make its arrays by inkjet printing, which allows it to respond quickly to design changes, says Kevin Meldrum, manager of the company's genomic collaborations. "Once an array design has been completed, the sequences are simply sent to the printer and the chip is produced." This allows Agilent to focus on custom arrays. "Right now many scientists are conducting broad, whole genome scans," Meldrum says. "But once they identify regions of interest, it will be more cost effective to do more targeted studies."



The mass range profile of possible histone H4 modifications.

All kinds of arrays

DNA microarrays have many applications in epigenomics. SwitchGear Genomics in Menlo Park, California, uses custom-designed oligonucleotide arrays from Affymetrix to profile DNA methylation patterns in genomic DNA treated with methylation-sensitive enzymes. The array design covers nearly all of the annotated CpG islands in the genome and about 20,000 additional CpG-rich regions the company has identified as potential regulatory regions. "We can profile 50,000 distinct 200–800-base-pair regions in the genome and quickly determine whether they are methylated or not," says Nathan Trinklein, co-founder and chief executive of the company.

EPIGENOMICS

Epigenomics uses a slightly different microarray design to offer a similar service, which it calls differential methylation hybridization (DMH). "The data we produce with DMH are highly reproducible and comparable across different projects," says Plum. "We are building a database of profiles for healthy and disease tissues."

Achim Plum is working to identify epigenetic changes associated with various cancers.

One of the advantages of using arrays for methylation studies, rather than for monitoring the expression of thousands of mRNAs at once (expression profiling), is that the signal is much easier to score. "Compared with expression profiling there is better signal-to-noise," says Meldrum. "Generally what you are measuring is 'have I or have I not had any binding?'."

In sequence

As well as using microarrays to analyse ChIP results, the method can also be subjected to sequencing. This technique uses sequencing to localize the methylation sites or interaction points of proteins or histones.

ChIP sequencing offers a couple of advantages over ChIP-on-chip. It is better suited for regions of DNA close to repetitive sequences, which can create 'noise' in a microarray experiment, and it can cover a larger portion of the genome. "Tiling arrays require masking of the repetitive regions of the genome," says Harper. "With sequencing, genome coverage is greater. We are evaluating the next-generation sequencing in combination with our methylated DNA assays to determine the potential for attaining the same high level of information with sequencing that we do with arrays."

The main drawback of ChIP sequencing is that the throughput is much lower than is obtained with microarray screens. Although this will probably increase, sequencing studies are likely to remain a more expensive proposition compared with microarrays.

One of the first companies to perform faster 'next generation' sequencing was 454 Life Sciences of Branford, Connecticut, but its technology is not well suited to ChIP sequencing. "For ChIP sequencing you don't need long read runs and super accuracy like with the 454 technology," says Struhl. "Right now Illumina [of San Diego, California] has a better technology for ChIP sequencing and others are being developed."

Other sequencing technologies that can be applied to ChIP sequencing include the SOLiD system from Applied Biosystems in Foster City, California, and the single-molecule sequencing process from Helicos BioSciences in Cambridge, Massachusetts.

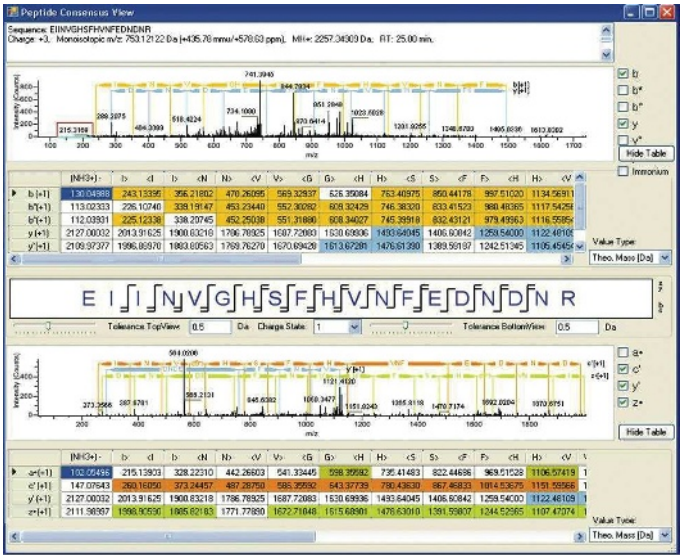
Bisulphite treatment

Another way to detect methylation is to modify DNA using sodium bisulphite, which converts unmethylated cytosine into uracil. This technique used to be fraught with difficulties, but that is no longer the case, says Plum. In fact, the technique is growing in popularity thanks to the number of effective kits available (see 'Tools of the trade').

Bisulphite-treated DNA is analysed in different ways depending on the resolution and throughput required. Methylation-sensitive PCR uses primers that anneal only those sequences that contain 5-methylcytosines, which don't get converted by bisulphite. For higher-throughput screens, bisulphite-treated DNA is hybridized to microarrays containing two sets of oligonucleotides, one of which is complementary to the unaltered methylated sequence and the other to the converted unmethylated sequence. For a more detailed look at the cytosines that are methylated and their precise location, bisulphite-treated DNA can be sequenced directly. Earlier this year, this 'bisulphite sequencing' was used to decipher DNA methylation patterns in the *Arabidopsis* genome at nucleotide resolution[2].

Epigenomics combines these analytical techniques to identify biomarkers for cancer. The company has so far identified several genes that show changes in methylation in breast, colon, prostate and lung cancers. "We believe that there are some technical advantages to using epigenetic markers," says Plum. "For one thing, we are looking at a signal on the DNA. We can analyse the DNA in paraffin sections. We can look for fragments shed by tumours in the blood stream."

Although methylation can change in different tissues or as a result of ageing and environmental factors, it is not as dynamic as gene expression. As a result, a single methylation



THERMO FISHER SCIENTIFIC

Thermo Fisher Scientific's Proteome Discoverer software (left) along with its LTQ Orbitrap XI ETD system are being used to analyse histone modifications.

marker can often be used to detect a disease, says Plum, whereas several markers would be needed for expression profiling.

Sequenom, based in San Diego, California, provides genomic services and has taken a different approach to analysing bisulphite-treated DNA. The company's EpiTYPER platform identifies and quantifies methylated sequences by gene-specific amplification of bisulphite-treated DNA followed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. "Although it does not provide genome-wide analysis, the value of our technique is fine mapping of sample cohorts," says Mathias Ehrich, Sequenom's group leader for epigenetics.

The two main advantages of the technique are its quantitative capabilities — allowing a researcher to determine whether, for example, methylation at a specific gene increased from 20% to 40% in a tumour sample — and its long read lengths. "If you have a 2,000-base-pair CpG island, you want to look at the entire island for changes in methylation. If you only look at single CpG sites using PCR or a microarray you will not have enough information," says Ehrich.

Critical mass

When it comes to mapping the combinations of histone modifications, mass spectrometry offers a way to get at important details. "Individual modifications are like words in a sentence, but you have to know the context," says Joshua Coon, a biomolecular chemist at the University of Wisconsin, Madison. "For example, in human embryonic stem cells, methylation at arginine 3 of histone H4 is found only in the presence of dimethylation at lysine 20. That is the context that mass spectrometry can deliver. It is challenging to determine such context with antibodies alone."

Unlike the use of antibodies, in which researchers select their antibodies based on the particular modification they are seeking, mass spectrometry is an unbiased approach. It also avoids the problem caused by one bound antibody inhibiting another from binding a nearby modification.

To determine the modification patterns, intact histones — or at least entire N-termini peptides — have to be analysed. Such proteins are much larger than those usually examined in conventional 'bottom-up' mass spectrometry.

It also requires highly sensitive and precise mass measurements: the difference in mass between a trimethylation and acetylation is only 36 millidaltons, for example. This issue is being successfully addressed by hybrid mass spectrometers, which bring two types of analyser into one instrument.

For example, a combination of a very-high-resolution Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer with a relatively low-resolution ion trap mass spectrometer has produced some spectacular results, says Coon. "The hybridization with ion traps resulted in significant gains in routine mass accuracy and resolution, because the ion trap regulates the ion population going to the FT-ICR," he explains.

"You can get much more detailed information from hybrid mass specs," says Andreas Hühmer, proteomics marketing director at Thermo Fisher Scientific in Waltham, Massachusetts. These instruments mean that it should be possible to characterize every protein modification in different types of cells, he adds.

Another component to analysis involves breaking all the bonds between the amino-acid residues so that the mass spectrometer can read their sequence. This used to be done by colliding ions in a neutral gas, but recent developments have introduced more effective techniques. These include electron capture dissociation (ECD) and electron transfer dissociation (ETD), which induce fragmentation by transferring electrons to positively charged peptides.

Coon's group has used an ETD-enabled linear ion trap mass spectrometer to map all the modifications that occur in the first 23 residues of the N-terminal tail of the histone H4 in differentiating human embryonic stem cells[3]. Similarly, Neil Kelleher, a chemist at the University of Illinois at Urbana-Champaign, has used ECD to identify combinations of modifications on the first 50 amino acids of the histone H3 protein, finding more than 150 forms of the protein[4].

The next step for mass spectrometry is the development of methods to determine levels of modifications and the ability to track patterns of change under different conditions. Waters, a liquid chromatography and mass spectrometry firm in Milford, Massachusetts, is working with researchers at the University of Southern Denmark in Odense to determine histone modifications in normal cells and cells undergoing senescence. "We have combined ion mobility and mass spectrometry," says James Langridge, director of the proteomics mass spectrometry business at Waters. "We are also pushing the boundaries to understand where the limits are and how to improve the technology."

The next five years will see a boom in epigenomic research thanks to advances in mass spectrometers, array design and sequencing technologies. Several groups and consortia are taking advantage of these developments to characterize the entire epigenome in both healthy and disease states. The knowledge obtained will increase our understanding of gene regulation and should yield new biomarkers for disease. ■

Laura Bonetta is a science writer based in Garrett Park, Maryland.

1. Johnson, D. S. *et al. Genome Res.* **18,** 393–403 (2008).
2. Cokus, S. J. *et al. Nature* **452,** 215–219 (2008).
3. Phanstiel, D. *et al. Proc. Natl Acad. Sci. USA* **105,** 4093–4098 (2008).
4. Garcia, B. A., Pesavento, J. J., Mizzen, C. A. & Kelleher, N. L. *Nature Methods* **4,** 487–489 (2007).