

# UNFINISHED SYMPHONY

To correctly 'play' the DNA score in our genome, cells must read another notation that overlays it — the epigenetic code. A global effort to decode it is now in the making, reports **Jane Qiu**.

**M**anuel Esteller's phone did not stop ringing for weeks. It was summer 2005, and he and his team at the Spanish National Cancer Centre in Madrid had just published a study comparing the activity of DNA in identical twins. The anxious callers were invariably twins whose sibling had developed a serious disease such as cancer or diabetes. Could the study help predict whether they too would succumb, they asked. Did the identical DNA sequence they shared with their afflicted twin mean they had the same genetic predisposition to illness?

Surprisingly, the answer to the second question is 'not necessarily'. Researchers have known for years that, despite their common genes, identical twins can have very different physical constitutions and develop different diseases. The traditional explanation for this is that our environment somehow interacts with our genes to produce our physical attributes, or phenotype, but no one knew exactly how.

The study by Esteller and his team<sup>1</sup> showed

that the missing link between nature and nurture could lie in a phenomenon known as epigenetics: a cryptic chemical and physical code written over our genome's DNA sequence. The term 'epigenetics' was first coined in the 1940s by British embryologist and geneticist Conrad Waddington, to describe "the interactions of genes with their environment, which bring the phenotype into being"<sup>2</sup>. The term now refers to the extra layers of instructions that influence gene activity without altering the DNA sequence.

**"Deciphering the epigenetic code will illuminate some of the most profound questions in biology."**

— Stephan Beck

on gene activity: the number of genes that differ in activity between 50-year-old twins was more than three times that in pairs aged 3. "So we are more than our genes," says Esteller. "Not only is the DNA sequence important but

also how gene activity is regulated in response to environment. This might explain why many identical twins have different susceptibility to disease."

## Spot the difference

As well as offering answers to identical twins, deciphering this epigenetic code promises to dramatically alter our understanding of disease in the wider population (see 'Tagged for disease', overleaf). Many cancers might be triggered by epigenetic faults, for example. It should also fill some big gaps in our grasp of how the environment affects a creature's constitution — epigenetic changes explain how simply altering the diet of a pregnant mouse, for example, can completely change the coat colour of her pups<sup>3</sup>, or even alter their response to stress<sup>4</sup>.

"It will illuminate some of the most profound questions in biology," says Stephan Beck, an immunologist at the Wellcome Trust Sanger Institute, Cambridge, UK, who worked on the Human Genome Project. How a given cell executes its unique genomic programme in time and space could shed fresh light not only on development and disease but also

on what makes us human, he says.

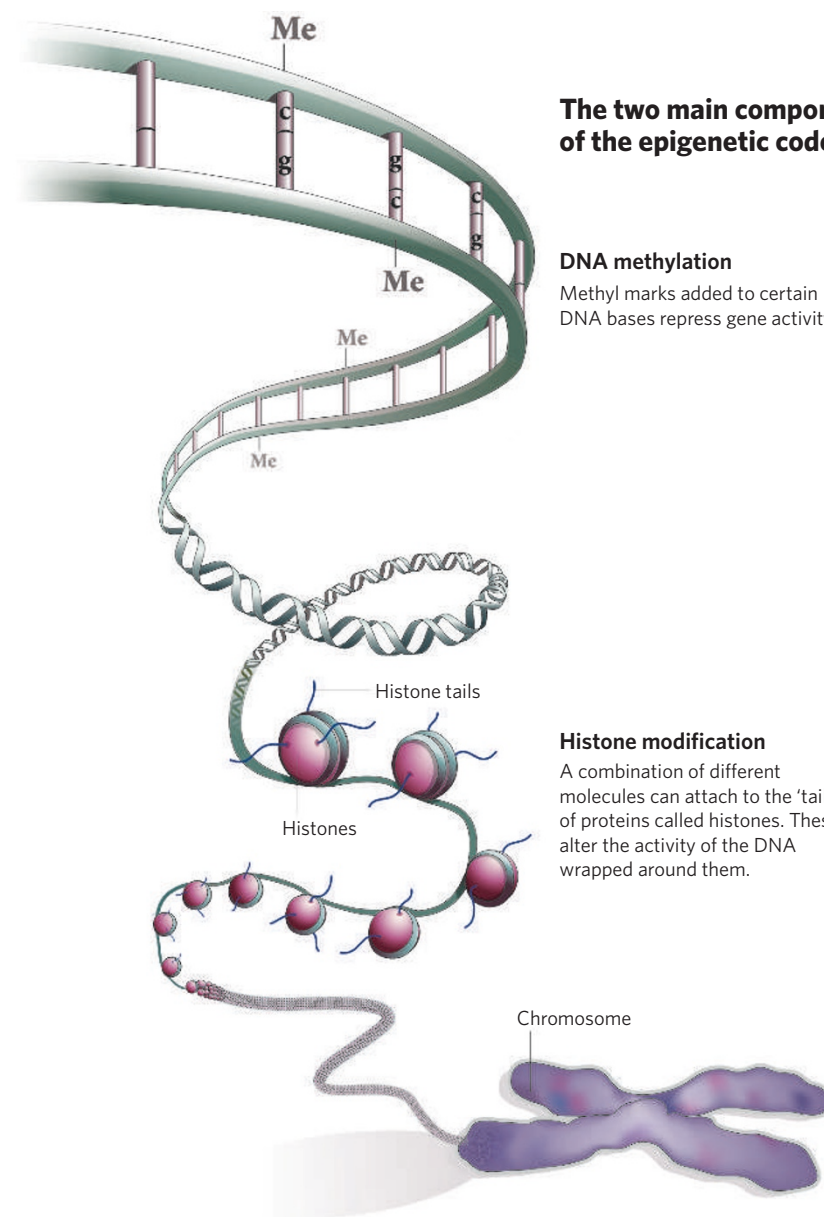
The complete epigenetic code of our genome, its 'epigenome' has increasingly been the focus of research over the past decade, and scientists are now embarking on an ambitious attempt to crack it. The International Human Epigenome Project, or IHEP, first suggested by Beck and colleagues in 1999, is the logical next step after the Human Genome Project, which published the draft sequence of the human genome's 3 billion DNA letters in 2001. But the IHEP faces daunting challenges. The sequence of the human genome is the same in all our cells, whereas the epigenome differs from tissue to tissue, and changes in response to the cell's environment. Can researchers really hope to pin down this vast, complex and ever-changing code in a meaningful way?

### Clever packaging

If the DNA sequence of the genome is like the musical score of a symphony, then the epigenome is like the key signatures, phrasing and dynamics that show how the notes of the melody should be played. Epigenetic control of gene expression occurs in two main ways: either the DNA itself is chemically altered, or the proteins that package DNA into chromatin (the main component of chromosomes), are modified. These proteins, called histones, determine whether the chromatin is tightly packed, in which case gene expression is shut down (or silenced), or relaxed, in which case gene expression is active.

The first kind of alteration takes the form of methyl groups added to the DNA — frequently to the base cytosine when it is immediately followed by guanine — by a process known as DNA methylation (see graphic). The methyl group can be sensed by proteins that turn gene expression on or off through regulating chromatin structure. The second, more complex kind of alteration involves changes to the histones around which chromosomal DNA is wrapped. Each histone has a protruding 'tail' to which more than 20 chemical tags can attach, like charms on a bracelet. Some of these tags, or certain combinations of them, dubbed the histone code, give rise to relaxed chromatin; others have the opposite effect.

Epigenetic codes are much more subject to environmental influences than the DNA sequence. "This could explain how lifestyle and toxic chemicals affect susceptibility to diseases," says Vardman Rakyen, a researcher at the Sanger Institute. "Up to 70% of the contribution to a particular disease can be non-genetic." Indeed, one key finding of Esteller and his team's study was that epigenetic profiles of twins who had been raised apart or had noticeably distinct lifestyles differed more than



### The two main components of the epigenetic code

#### DNA methylation

Methyl marks added to certain DNA bases repress gene activity.

#### Histone modification

A combination of different molecules can attach to the 'tails' of proteins called histones. These alter the activity of the DNA wrapped around them.

those who had lived together for a while or shared similar environments and experiences<sup>1</sup>. Rakyen himself is studying a cohort of identical twins, where one twin has type 1 diabetes and the other does not, with the aim of finding epigenetic changes associated with the disease.

Although different labs around the world, such as Rakyen's, are already working on their own individual studies, several researchers argue that it is time for a coordinated effort. A series of international workshops, and expert and government reports have emerged in recent months that address the value and scope of an international human epigenome project. The ultimate goal of such a project would be to identify all the chemical modifications of DNA and histone proteins for all chromosomes in all types of normal human tissue.

As for the HGP, an international consortium

would set priorities, coordinate research efforts, centralize materials and resources, create the necessary technologies and monitor research progress.

### Piece by piece

"Epigenomics is at a stage where genomics was 30 years ago, when everyone was working on their part of the puzzle," remarks cancer biologist Peter Jones at the University of Southern California, Los Angeles. Jones was formerly president of the American Association for Cancer Research (AACR), which is based in Philadelphia. "We need to see the bigger picture. It takes concerted efforts on an international scale. And this is how the IHEP would make a difference."

Although a number of funding bodies — such as the Wellcome Trust, the AACR, the US



## TAGGED FOR DISEASE

Thanks to its ability to instruct cells how to 'play' the genetic instructions spelled out in their DNA, the epigenetic code is proving to be central to processes such as development, ageing, cancer, mental health and infertility. Cancer researchers, in particular, are studying it to develop diagnostic and prognostic tools and drugs.

Epigenetics may contribute as much to cancer development as mutations in the DNA itself. For example, *Wnk2*, a gene whose activity is thought to suppress tumour development, is more often shut down by epigenetic

changes in certain brain tumours than it is lost by genetic deletions.

Epigenetic marks can also help predict clinical outcomes. Researchers have identified a pattern of epigenetic modifications within the oestrogen receptor gene that correlates with a cancer patient's chances of survival in response to treatment with the drug tamoxifen<sup>9</sup>.

Epigenetic changes are easier to reverse than genetic mutations, by adding or removing the chemical tags involved. Drugs that do this are now at the forefront of cancer treatments. For example,

Vidaza, a drug made by Pharmion Corporation of Boulder, Colorado, has been approved by the US Food and Drug Administration for the treatment of myelodysplastic syndromes, also known as 'preleukaemia'.

In theory, this drug, which blocks DNA methylation, should be active in all cells. In practice, it seems to be active only in cancerous cells, stimulating the expression of many genes, including those that suppress cancer development. A number of other epigenetic drugs are at various stages of clinical trials.

J.Q.

the lesser of two evils, but exactly how different the epigenomes of cell lines are compared with normal tissues remains to be seen. The inclusion of cell lines in some pilot studies in the proposed IHEP should be able to resolve this issue.

### Final frontier

Perhaps the greatest challenges facing the IHEP are technological: mass-production-style tools must be developed to decode the epigenome, and the morass of data will have to be stored and analysed. At the moment, the main method used to determine DNA methylation sites is reliable, but extremely expensive, and the technology used to study histone marks is prone to problems with accuracy and reproducibility.

Scientists hope to tackle these problems by linking the IHEP to projects on the epigenomes of lab workhorses, such as the yeast, fruitfly and mouse, for which techniques are more advanced. Computational scientists are also developing the sophisticated bioinformatics tools needed to store and analyse multi-dimensional epigenome data.

Given these technological challenges, it is only natural to question whether the research community is ready for such an enormous undertaking. Drawing on the experience of the early days of planning for the HGP, researchers working on epigenetics are unanimous in thinking they can do it. They have drawn up a plan of how to manage the international assets available for the

IHEP<sup>5,7-9</sup> and say that, like the HGP, the IHEP will catalyse its own development. "One can never be 100% ready. We have 60% of the technology to go for the real thing," says Thomas Jenuwein, a molecular biologist at the Research Institute of Molecular Pathology at the Vienna Biocentre, Austria. "The rest will happen once the momentum is built up. We should have that vision to go in big."

Jane Qiu is an editor at *Nature Reviews Neuroscience*.

1. Fraga, M. F. *et al. Proc. Natl Acad. Sci. USA* **102**, 10604-10609 (2005).
2. Waddington, C. *Endeavour* **1**, 18-20 (1942).
3. Waterland, R. A. & Jirtle, R. L. *Mol. Cell. Biol.* **23**, 5293-5300 (2003).
4. Weaver, I. C. G. *et al. J. Neurosci.* **25**, 11045-11054 (2005).
5. Lieb, J. D. *et al. Cytogenet. Genome Res.* **114**, doi:10.1159/000091922 (2006).
6. Raky, V. K. *et al. PLoS Biol.* **2**, e405 (2004).
7. Feinberg, A. P. & Jones, P. A. Workshop on Defining the Epigenome <http://dcb.nci.nih.gov/Workshoprpt.cfm> (2006).
8. Rauscher, F. J. III *Cancer Res.* **65**, 11229 (2005).
9. Jones, P. A. & Martienssen, R. *Cancer Res.* **65**, 11241-11246 (2005).

National Cancer Institute (NCI), and the US National Human Genome Research Institute (NHGRI) — have shown interest by taking part in the discussion, funding agencies have yet to commit to financing and leading the project.

### All aboard

Since the completion of the Human Genome Project, there have been many multi-centre schemes, each of which costs millions or even billions of dollars. Some of these initiatives, such as the US National Institutes of Health Human Cancer Genome Atlas (which aims to identify and catalogue genetic mutations in human cancers), have prompted arguments over scale and cost-effectiveness. A key question for funding bodies is whether the IHEP would be yet another multi-million-dollar project. Proponents say no. "The goal of the IHEP is not to create another big enterprise, but to make things as cost effective as possible, to interface with wonderful projects that are under way and to fund important pilot projects," says Andrew Feinberg, director of the Centre for Epigenetics of Common Human Disease at the Johns Hopkins University in Baltimore, Maryland.

A number of smaller scale multi-centre epigenome projects are already under way or under discussion in Europe, the United States, India and Japan. Most prominent is that set up by the European Human Epigenome Project (HEP) Consortium in 2000. Following the publication of a pilot project in 2004 (ref. 6), the European HEP Consortium will soon make its data on the epigenetics of the entire

chromosomes 6, 20 and 22 publicly available.

Although few people doubt the importance of an international human epigenome project, how to go about it remains a subject of debate.

A key challenge is defining what the epigenome entails and what cell types to study. Some researchers argue that the project should first tackle blood cells, because they are easy to collect and work with, and are our main 'window' into the epigenome of both healthy and diseased individuals. Once a high-resolution blood epigenome is determined, it will serve as a reference with which other epigenomes, including those of diseased or ageing tissues, could be compared.

But the diversity of epigenomes in different cell types means that it may not make sense to restrict pilot projects to one single tissue, or to a particular time in a tissue's development. After intense discussion in three recent international workshops<sup>5,7-9</sup>, researchers in the epigenetic community now agree that initially eight to ten tissues, including the blood, should be studied simultaneously. Ultimately, the epigenome of all tissues, including embryonic stem cells, will be mapped out.

Another question is whether to study cells grown in the lab or biopsies of tissues taken from people. Biopsies contain different cell types, which would muddy the picture, but lab-grown cells might contain abnormal epigenetic tags. At the moment, some biologists are leaning towards lab-grown cells as being

**"Epigenomics is where genomics was 30 years ago, when everyone was working on part of the puzzle."**

— Peter Jones