New method for reading DNA sheds light on basis of cell identity

July 1 2007

As a fertilized egg develops into a full grown adult, mammalian cells make many crucial decisions — closing doors of opportunity as they adopt careers as liver cells, skin cells, or neurons. One of the most fundamental mysteries in biomedicine is how cells make such different career decisions despite having exactly the same DNA.

By using a new kind of genomic technology, a new study unveils a special code — not within DNA, but within the so-called "chromatin" proteins surrounding it — that could unlock these mysterious choices underlying cell identity.

A research team led by scientists at the Broad Institute of Harvard and MIT and the Massachusetts General Hospital has created genome-wide chromatin maps for embryonic stem (ES) cells and two cell types derived from them, by applying a powerful new technology for sequencing DNA.

The work, published in the July 1st advance online edition of *Nature*, provides a framework for mapping the complete chromatin landscape of almost any kind of cell. One of the most surprising findings suggests that cells contain an explicit chromatin-based code that reveals the developmental choices they have already made as well as those decisions that lie ahead.

"Unraveling the mysteries of chromatin holds great promise for understanding how cells in the body — with nearly identical DNA —

assume such different forms and functions," said co-senior author Bradley Bernstein, an associate member at the Broad Institute and an assistant professor at Massachusetts General Hospital and Harvard Medical School. "By applying a new technology for sequencing DNA, we have been able to look across the genome at chromatin, with greater resolution and efficiency than ever before."

Chromatin proteins are more than just packing material for the genome. By virtue of different chemical groups fastened to them, these proteins influence which parts of the double helix are open — or not — to the cellular machinery, thus controlling which genes get turned on or off.

To decipher this "epigenetic" code requires ways of determining precisely which chromatin proteins sit at which locations along a cell's DNA. In principle, scientists could infer the locations by using specialized DNA chips. In practice, though, the technique has proven slow and expensive to construct genome-wide maps of mammalian chromatin. But now, a new method of massively parallel DNA sequencing has given rise to a powerful approach for readily churning out whole-genome maps of chromatin structure. The technology — based on single-molecule sequencing — makes it possible to read billions of DNA letters simultaneously. "Single molecule-based methods for decoding DNA are now throwing open the doors to a plethora of unexplored questions in chromatin, epigenetics and many other areas of biology," said Bernstein.

Empowered by this new technology, the researchers set out to study chromatin in cells with drastically different behaviors. They analyzed an assortment of chromatin proteins, each with a distinct chemical tag that switches genes either on or off. The scientists examined these proteins in mouse ES cells — known for their unusual ability to form nearly any tissue — as well as two other types of descendant cells that are more limited in the developmental paths they can choose.

One of the most remarkable findings involves a way of using chromatin to look into a cell's past to determine the developmental decisions it has already made, and to peer into the future to read its potential choices. The fortuneteller lies in a unique form of modified chromatin known as a "bivalent domain", which marks the control regions of important genes. Such domains merge both activating and repressive chemical tags, keeping genes quiet yet poised for later activity.

Bivalent domains had been noted for their role in ES cells, helping keep these cells' developmental options wide open. But with the new genome-wide chromatin data, the scientists discovered that these domains also function in more specialized kinds of stem cells. In neural stem cells, for example, bivalent domains sit near genes important to various types of brain cells, but are notably absent from genes that would be active only in, say, skin cells or blood cells.

"Looking at a cell through a microscope often cannot tell you what kind of cell it is, or more importantly, what it has the potential to become," said first author Tarjei Mikkelsen, a Broad Institute researcher and a Harvard-MIT Health Sciences and Technology graduate student. "But by decoding its chromatin on a genomic scale, we can now begin to systematically address such questions."

"Our understanding of the basis of cell identity — the way that a liver cell knows that it is different from a skin cell — has been rather vague, much like our understanding of heredity was prior to our knowledge of DNA," said Broad Institute director Eric Lander, a co-senior author of the study. "The chromatin maps suggest that it may be possible to directly read out a complete description of all of a cell's past commitments and its future potential. If true, this would have enormous implications for our understanding of developmental biology and for guiding regenerative medicine."

In addition to shedding light on key developmental decisions, chromatin maps also contain other sorts of new biological information. One type of chromatin modification marks not the control regions of genes, but their "bodies" — from where genes first begin to where they end. The scientists found that these "body" marks identify not only typical genes — that is, the ones that encode proteins — but also so-called "non-coding" genes that only produce RNAs. These marks could provide a practical handle for precisely mapping all of the genes in the genome, a task that has proven quite challenging by other methods.

Source: Broad Institute of MIT and Harvard