Big DataHighlight

BY KATHARINE MILLER

111 Reference Human Epigenomes

Each cell in a person's body contains the same three billion letter DNA encyclopedia. But the many different cell types throughout the body (brain, bone, heart, skin, etc.) represent different readers of that encyclopedia who have each highlighted their favorite parts, dog-eared certain pages, annotated interesting paragraphs, and crossed out things they find dull or uninteresting. These markings constitute the epigenome. "The epigenome tells us what are the important parts to read," says **Manolis Kellis, PhD**, professor of computer science at the Massachusetts Institute of Technology (MIT).

And now the marked up encyclopedias are available to all. A February 2015 *Nature* paper by the Roadmap Epigenomics Consortium, of which Kellis is a part, reports that they have mapped the epigenomes for 111 different human cell types. In addition, the researchers compared the different cell types' epigenomic signatures and even uncovered some possible ways the epigenome may play a role in disease.

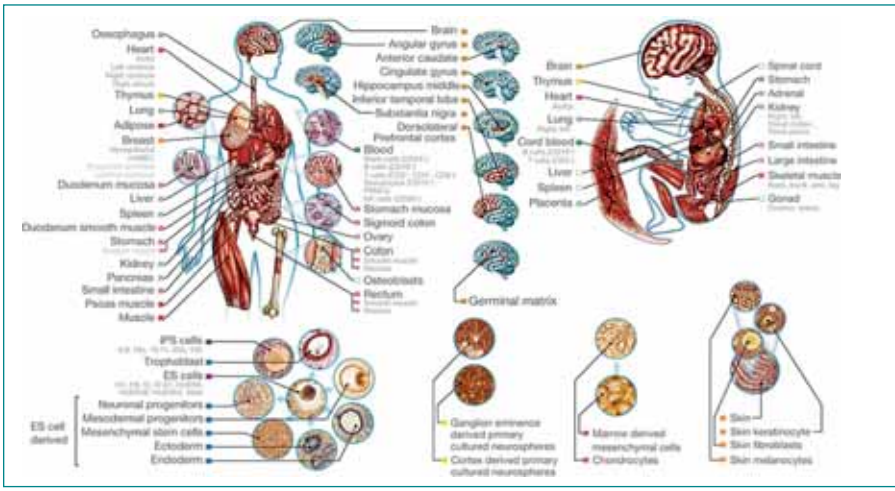
Epigenetic changes in different kinds of cells include such things as DNA methylation and modifications to the histones around which DNA is wrapped (like beads on a string) in the cell nucleus. These kinds of changes affect which genes in the cell are expressed at any particular time.

To map all such changes for 111 different human cell types—an enormous task in itself—the Consortium used a variety of assays to pull out the modified parts of the genome and then sequence the attached DNA to determine its location in the genome. This produced an epigenomic map showing each kind of modification for each type of cell.

The next step, Kellis says, was to figure out how the marks have meaning. For that, he and his colleagues turned to hidden Markov models capable of identifying underlying patterns in the epigenome maps. "Some patterns happen at the starts of genes; others in the gene; others in places that are repressed," Kellis says. These patterns revealed regulatory modules of coordinated activity as well as their likely activators and repressors.

Once that was understood, Kellis says, "Then we could start studying the differences." And some interesting ones were noted. For example, many cells derived from embryonic stem cells clustered closely with other embryonic stem cells rather than their corresponding tissues, suggesting their stem cell nature was still predominant. For one particular enhancer signal, fetal brain and germinal matrix cells cluster with neural stem cells rather than adult brain cells.

Going a step further, the researchers looked at how the epigenomic data sets squared with disease-associated variants identified in various genome-wide association studies (GWAS). In many cases, disease variants were enriched in epigenomic modifications for trait-specific tissues. For example, the team looked across the genome at all the genetic variants associated with blood pressure and found that they tend to be active in cells in the left ventricle of the heart. "These are genetic differences that affect the circuitry that turns the genes on and off," Kellis says. Thus, the way cells are controlled in the left ventricle has something to do with how much pressure builds up.



Researchers have sequenced the epigenomes of 111 different human cell types. Reprinted by permission from Macmillan Publishers Ltd: Roadmaps Epigenomic Consortium et al., Integrative analysis of 111 reference human epigenomes,* Nature *518:318 (2015).

A separate *Nature* paper by the Consortium also reported that the epigenomic signature for a cancerous tumor could help identify the originating site of the cancer—a piece of information that is sometimes unknown and confounds appropriate treatment.

This work is important, Kellis says, because it provides access not just to the protein coding part of the genome, but also to the hidden control circuitry. "We can read the circuitry of the genome and understand which genes are active and use that to understand where the genetic predispositions to disease lie within the genome," he says. □

This work is important, Kellis says, because it provides access not just to the protein coding part of the genome, but also to the hidden control circuitry.