

FEBRUARY 17, 2008

## A Widescreen View of Methylation

If the wrong gene is turned on at the wrong time, it can wreak havoc in the cell. To prevent this from happening, organisms rely on DNA methylation to keep unneeded genes turned off. Despite its importance, however, methylation has remained enigmatic — largely because researchers have been forced to examine individual genes one at a time, a myopic and laborious process.

Now, novel molecular biology techniques and software developed by Howard Hughes Medical Institute (HHMI) investigator Steven Jacobsen and colleagues will give scientists the tools they need to identify large-scale methylation patterns in complete genomes. The team published their findings February 17, 2008, in an advanced online publication of the journal *Nature*.

---

"There's lots of evidence that when genes' methylation patterns are not properly maintained, that is a major cause of cancer. So if we can learn enough about these methylation mechanisms, we may some day learn to manipulate them and treat cancer."

— Steven E. Jacobsen

---

During DNA methylation, small molecules called methyl groups are added to specific sites on DNA. The methyl groups are attached only to specific cytosine (C) bases — one of the four building blocks of DNA. The methyl group serves as a beacon, saying that 'that stretch of DNA should be silenced,' explained Jacobsen, whose lab is at the University of California, Los Angeles (UCLA).

Jacobsen has been working to understand how gene silencing happens in the plant *Arabidopsis thaliana*, a common model organism. His lab has created mutant strains of *Arabidopsis* that have been useful in observing just how cells attach the chemical signals in exactly the right places.

Methylation is vital to most organisms - even the small *Arabidopsis* genome contains 13 million methylated cytosines. But Jacobsen says that researchers have struggled to see its exact footprint on the genome. To find methylated

genes, researchers usually turn to a technique called bisulphite sequencing, which chemically changes normal cytosine to thymidine (T; another DNA base), while leaving methylated cytosines unchanged. Once this step is complete, the modified DNA can be sequenced.

In the past we have had to use very labor-intensive, painstaking techniques and look at one gene at a time, Jacobsen said. This restricted researchers to looking at only a few genes in a genome—stifling their ability to uncover large-scale patterns of methylation.

In the experiments reported in *Nature*, Jacobsen's group collaborated with computational biologists Matteo Pellegrini and Shawn Cokus, as well as researchers from the biotechnology companies Illumina and New England BioLabs to create new molecular biology methods, algorithms, and software to analyze bisulphite sequences. The techniques allowed the team to correctly and rapidly identify more than 90 percent of the methylated cytosines across the entire genome of *Arabidopsis*.

Their analysis confirmed methylation patterns that other researchers' experiments had hinted at, and offered clues into how the enzymes that methylate DNA do their work. For example, they found that among pairs of cytosines, those that were 10 bases apart had the best chance of both being methylated. The reason? Ten bases is exactly the length of the turn of the double helix of DNA, Jacobsen said. Their observation suggests that methylating enzymes may travel along one face of the DNA molecule, placing multiple methyl groups at once.

The team also found that the chance of methylation increased at intervals of 167 bases. This matches the spacing between the histone proteins that package DNA, Jacobsen noted. We think it's because the enzymes are being pulled in by the histones, and methylating the DNA right nearby.

According to Jacobsen, his team's findings improve understanding of how cells control gene expression, and might one day find use in medicine. There's lots of evidence that when genes' methylation patterns are not properly maintained, that is a major cause of cancer, he said. So if we can learn enough about these [methylation] mechanisms, we may some day learn to manipulate them and treat cancer.

Jacobsen said that the *Arabidopsis* study's success validates their new tools. Now we're looking at other organisms to see how widely conserved these [patterns] are, he said. The team is looking for help to search as many genomes as possible; they've posted the source code and additional information about their experiments on their website, <http://epigenomics.mcdb.ucla.edu/BS-Seq/>. They don't know quite what they'll find, Jacobsen said. We don't have this kind of data for any other organism yet.