

```
Drupal.behaviors.print = function(context) {window.print();window.close();}>
```



# To Map a Methylome

June 01, 2011

## To Map a Methylome

By Tracy Vence

Nature? Nurture? How about a bit of both?

Investigations into the complexities that govern the dynamic relationship between genotype and phenotype have shown that the two are by no means mutually exclusive. But despite seeing an exponential growth in genomic data, researchers remain puzzled by an intangible, but substantive, source of variation — the epigenome.

"Many researchers in the broader area of human disease studies are trying to identify the missing variable that cannot be explained by genetic variation alone," says Kun Zhang at the University of California, San Diego. "Epigenetic modifications and variation is clearly one of the areas that people are looking at now."

Indeed, in recent years, researchers have shown an increasing interest in deciphering how histone modifications and differential DNA methylation affect gene expression, particularly as both epigenetic marks have been shown to play demonstrable roles in development and disease. To place the missing pieces within the phenotypic puzzle, many researchers have opted to map histone modifications and methylation status genome-wide. For the latter, the Broad Institute's Alexander Meissner says researchers are only now beginning to get a handle on how high-resolution methylation mapping efforts could enhance human development and disease research.

[Click here to subscribe to Genome Technology](#)

"There are only about a dozen whole-methylome maps at this time, so we're still sort of in the infancy of learning what they could mean and how much power they could have," Meissner says. "There are a few studies that are less comprehensive that definitely show [methylation has] a lot of power. The key is to now learn how much additional power we can gain from looking at the entire genome."

Current approaches to mapping methylation status across the genome run the gamut from methylation-specific PCR to restriction enzyme-based assays and, more recently, to sequencing the entire bisulfite-converted genome — though this method remains prohibitively expensive for most labs. While there are a handful of reduced representation and targeted capture methods that, when coupled to arrays or sequencing, are commonly and reliably used to map

methylation on a modest budget, there is no clear front-runner among them. Each approach has its limitations.

On the technology development front, some investigators are working to refine targeted capture protocols to reduce sample input requirements and enhance multiplexing capabilities, while others intend to refine detection to single-cell and even single-molecule resolution. Taken together, advances in approaches to map methylomes are poised to advance researchers' efforts to understand the epigenome.

## Options and limitations

The Broad's Meissner is mapping a multitude of methylomes as co-PI on a grant that has provided more than \$3 million annually since 2008 as part of the National Institutes of Health's Roadmap Epigenomics Project. In their efforts to produce reference epigenomes, he and his colleagues are working to generate comprehensive, high-resolution maps of chromatin state and methylation status for 100 human cell types. So far, they have generated reference epigenomes for both human embryonic and induced pluripotent stem cells — which were published in *Cell* in February — though Meissner says the Broad has several dozen additional cell types in its mapping pipeline. Priority is assigned to cell types based on their relevance to disease and ease of isolation, he says.

[pagebreak]

"Liver is fairly homogeneous tissue, so that's doable. For the brain we can actually isolate from specific regions, and for the blood system it's very easy to isolate based on the surface markers of certain cell types," Meissner says. "But beyond that, it actually takes a lot of additional resources and work to actually get the sample set."

Current sample input recommendations for methylome-mapping studies are on the order of a million cells, he adds, and certain cell types are more difficult to assess than others. "Even if you think about hematopoietic stem cells — you can't get an unlimited supply of those," he says.

Working to reduce sample requirements was one of the reasons that in 2005, Meissner — then at Whitehead Institute — and his colleagues at the Broad and the University of Edinburgh developed the reduced representation bisulfite sequencing approach for analyzing genome-wide methylation patterns. In contrast to genomic tiling microarray-based methods — most of which require more than 500 micrograms of input DNA — researchers can now perform reduced representation bisulfite sequencing with as little as 30 nanograms of starting material.

While, theoretically, whole-genome bisulfite sequencing offers researchers a comprehensive view of the methylome, it does so at both a technical and financial cost. Sequencing coverage is generally biased against regions of very high and low GC content, and, even at 20- to 40-fold coverage, whole-genome bisulfite sequencing still fails to detect certain methylation differences at regions of low CpG island density. For this, the Broad's Christoph Bock says that while "sequencing deeper is the solution, cost is the main problem." Until whole-genome bisulfite sequencing becomes more affordable, reduced representation bisulfite sequencing is one of several approaches researchers can take to selectively scan the methylome.

When coupled to sequencing, both methylated DNA immunoprecipitation — MeDIP — and methylated DNA capture by affinity purification — MethylCap — allow researchers to specifically interrogate methylated regions of interest genome-wide. In October, Bock, Meissner, and their colleagues published a quantitative comparison of reduced representation bisulfite sequencing against MeDIP-seq, MethylCap-seq, and an array-based assay in *Nature Biotechnology*. While they primarily aimed to measure the performance of the Broad's in-house reduced

representation bisulfite sequencing approach pitted against other, more commonly used methods, Bock says he and his colleagues also hoped to be able to "make fact-based recommendations for new groups entering epigenome mapping."

In its comparative analysis, the team found that each approach it evaluated had its pitfalls. "In our hands, MethylCap-seq worked better than MeDIP-seq, although the difference was not massive, and it may be possible to optimize MeDIP-seq to a degree that it performs as well as MethylCap-seq," Bock says. Between reduced representation bisulfite sequencing and MethylCap-seq, Bock says that the latter method "performed better in terms of genomic coverage," but showed an increased susceptibility to experimental bias.

[pagebreak]

Choosing which approach to take largely "depends on the samples and the biological question," Bock says. For example, "when conducting DNA methylation profiling on a small to moderate number of samples in a single lab with sufficient amounts of DNA ... that is of similar quality between samples, MethylCap is likely to provide best genomic coverage and best value for money, especially if the expected DNA methylation differences are relatively large and broadly scattered over the genome," Bock says. However, if any of these conditions are not met, he says reduced representation bisulfite sequencing is often the better choice because it "decouples sequencing coverage from DNA methylation read-out. It is much less susceptible to error sources such as batch effects, between-laboratory variation, or variation induced by low and different input DNA quality."

Overall, the methods his team tested "are useful, but none is comprehensive in the sense that it would identify all DNA methylation differences that are present in a sample," Bock says, adding that every approach performed "better at identifying differentially methylated regions in CpG-rich regions of the genome than in CpG-poor regions."

For investigators at the Broad and elsewhere who are working to produce reference epigenomes, Bock says whole-genome bisulfite sequencing has become the standard. Even so, its practical pertinence for medically relevant epigenomics research is currently limited. "It is usually a lot more valuable to sequence 50 cases and 50 controls using RRBS or MethylCap-seq than to complete whole-methylome data for one case and one control," which echoes the situation in which genomic researchers have found themselves, he says. "Although whole-genome sequencing is feasible — and, to a limited extent, affordable — many current studies continue to focus on whole-exome sequencing in many samples rather than whole-genome sequencing in few."

In May 2010, researchers at Pacific Biosciences reported in *Nature Methods* that they are able to detect DNA methylation, without bisulfite conversion, using the company's single-molecule, real-time sequencer. In harnessing the polymerase kinetic signals intrinsic to PacBio's platform, combined with circular consensus sequencing, the researchers found they could detect epigenetic modifications on single-molecules at base-pair resolution. Catching epigenetic data along with sequence is likely to increasingly be a focus for vendors going forward.

### **A smaller scale**

With an eye toward analyzing methylation within complex tissues, Rob Mitra at Washington University in St. Louis is developing MethylMap, a technology that combines multiplexed amplification, sample-specific barcodes, and single-molecule bisulfite sequencing on laser capture microdissected cells. "By using laser capture microdissection to isolate different cell types from a complex tissue, one can confidently assign methylation patterns," Mitra says, adding that his team's goal is "to be able to analyze DNA from 10 to 100 groups of 100 micro-

dissected cells."

Using MethylMap, Mitra's group plans to characterize tumor samples and surrounding tissues from patients with colorectal cancer, endometrial cancer, and uveal melanomas. While he says his group began its technology development project primarily to "understand how patterns of methylation are specified in developing tissues and to understand the role of methylation in tumor evolution," Mitra believes MethylMap shows potential for clinical application as well. "The ability to analyze genome-wide methylation in a small number of cells may also ... identify biomarkers of disease state," he adds.

[pagebreak]

Like Mitra, Cornell University's Paul Soloway aims to assess epigenomic alterations on a smaller scale. But rather than characterizing methylation in single cells, Soloway is co-developing a nanofluidics-based technology that's able to detect differential methylation at the level of single molecules. Working in collaboration with his colleague Harold Craighead, Soloway is working to optimize a technology the team developed called SCAN — single-chromatin analysis at the nanoscale.

"In principle, it's similar to flow cytometry," Soloway says. "Except instead of working with whole cells, we're working with fragments of chromatin. And instead of using antibodies against cell-surface markers, we're using antibodies against histone proteins and histone protein variants as well as [proteins against] methylated DNA," such as NBD1, which recognizes methylation in the context of double-stranded DNA, he adds.

Using SCAN, researchers can flow single, tagged molecules in aqueous solution through nanoscale channels by way of a voltage gradient, imaging each individual molecule as it passes through an inspection volume using laser-induced fluorescence confocal microscopy. Since he and his colleagues are "looking at single molecules, not single cells, [and] working with nanoscale channels, not micron-scale channels," Soloway says they "need much higher sensitivities of photon detection than are routinely used for cell sorting."

While not yet ready for routine use, Soloway says that, once optimized, the SCAN approach could have several advantages over other, more commonly used techniques, like those based on immunoprecipitation. "The first is that you can query a sample for multiple epigenetic marks simultaneously," he says. Rather than performing systematic re-immunoprecipitation experiments to validate two or more independent epigenetic marks, by querying multiple marks simultaneously, "we can unambiguously identify molecules that do in fact contain two or more different epigenetic marks [at once]," he says. Because it requires extremely low sample input, Soloway says another advantage to SCAN is that researchers "can potentially do this analysis with vanishingly small inputs of cells." SCAN could also prove to be a useful tool for those who seek "a quick assessment of the density of a given epigenetic mark in the genome," he adds.

Soloway says he and Craighead are currently working to optimize throughput and to modify the SCAN platform so that they could eventually run multiple parallel nanofluidic channels. The team also hopes to incorporate a sorting capacity to the SCAN platform, thus mirroring the dual functions of flow cytometers in their analytical and preparative modes.

Overall, Soloway says he is most hopeful for how SCAN could accelerate his own basic epigenomics research. "One of the questions I've been pursuing in my own lab in the last few years is trying to understand how it is that particular regions of a genome acquire epigenetic marks. ... How certain combinations of epigenetic marks either collaborate or antagonize one another to dictate local epigenetic states. I'm interested in trying to define some of those rules," he says, adding that to do so "really requires being able to query multiple epigenetic marks simultaneously ... [so] a tool like this could be very useful."

[pagebreak]

## On the horizon

At UCSD, Kun Zhang is working to refine the epigenomics field's "equivalent to exome sequencing to analyze the human genome." In an extension of the padlock capture method he developed in collaboration with Harvard University's George Church and others, Zhang and his team are designing padlock probes — circularized oligonucleotides consisting of two end segments that are complementary to a target sequence and connected by a linker sequence — for the targeted capture of bisulfite-converted DNA.

As bisulfite conversion significantly reduces genomic complexity — by converting all non-methylated cytosines into uracils — Zhang says that cooperation between the two ends of a padlock probe gives a significant advantage in terms of specificity over other capture agents. In a 2009 *Nature Biotechnology* paper, Zhang says he and his colleagues reported their creation of about 30,000 padlock probes to "capture every single CpG island on two human chromosomes."

Zhang says his group has since expanded its genomic targets by "taking advantage of the large number of publications in the past two years on global methylation profiling of a variety of different cell types." At present, his group's probe set has a target size of 30 megabases, roughly that of the human exome. The team is now working to simplify the padlock probe protocol and has a keen interest in achieving what he calls two-dimensional multiplexing — "we want to multiplex both on CpG sites and on many different samples," he says.

Using this bisulfite padlock capture approach, Zhang says his team can efficiently process loaded 96-well plates to prepare sequencing libraries. "I think we're getting to the point where we can process a large panel of samples and deliver highly consistent results," he says. "Our technology is just getting to prime time."

Zhang is hopeful that his team's targeted capture technology will enable researchers to efficiently perform epigenome-wide association studies. "To enable that kind of study, we need some kind of assay that can be cheap, accurate, and provides an unbiased survey across the whole genome," he says. "This community really needs a ... tool that allows them to look at a large number of samples — a cohort of maybe thousands — in order to detect weak signals."

## More questions

During the course of his targeted capture technology development effort, Zhang collected what he calls "very interesting data" that suggest that long-range co-methylation and neighborhood dynamics appear to affect the probability of methylation among adjacent CpG sites. "With methylation reads, we can characterize not only the methylation status, but also genetic variants," Zhang says. This realization "suddenly allowed us to directly interrogate both genetic and epigenetic information on the same sample, with the same assay," he adds. Zhang and his colleagues then developed an algorithm that enables them to do just that, and using that information, they found that there is "a lot of interaction between the genome and the methylome, and a lot of interaction seems to be mediated by genetic polymorphisms, particularly SNPs that directly hit on CpG dinucleotides," he says.

The bisulfite sequencing data Zhang and his colleagues have generated have opened up a variety of questions they seek to address going forward. "How exactly are those CpG sites on the same chromosome molecule co-regulated? And what are the regulators? How does that relate to chromosome folding and high-dimensional structure? How does this relate to other

chromatin marks [and] epigenetic modifications?" he asks.

To begin to chip away at these and other questions, Zhang's team proposes to build a human haplotype so that they can distinguish paternal chromosomes from maternal ones, map CpG sites to each, and assess differential methylation between them. Having this information in hand could help researchers to "separate allelic differences in methylation — due to genomic imprinting — from [those] due to cis-regulatory polymorphisms," he says.

## Related Stories

- [Don't Know Much About Phylogeny](#)  
June 1, 2011 / [Genome Technology](#)
- [The Clinical Push](#)  
May 1, 2011 / [Genome Technology](#)
- [Variations on a Theme](#)  
April 1, 2011 / [Genome Technology](#)
- [Exome En Vogue](#)  
March 1, 2011 / [Genome Technology](#)
- [The Undiscovered Variants](#)  
March 1, 2011 / [Genome Technology](#)

