# Team Uses PacBio Data to Detect and Phase Bacterial DNA Methylation at Single Molecule Level

Jun 15, 2015 | a GenomeWeb staff reporter

NEW YORK (GenomeWeb) – A team led by scientists at the Icahn School of Medicine at Mount Sinai has developed a method for detecting and phasing DNA methylation at the single molecule level using Pacific Biosciences' sequencing technology.

In a paper published online in Nature Communications today, the researchers, led by Eric Schadt and Gang Fang of the Department of Genetics and Genomic Sciences and the Icahn Institute for Genomics and Multiscale Biology, described their approach, called single-molecule modification analysis of long reads (SMALR), and tested it on seven bacterial strains.

"We found that a typical clonal bacterial population that would otherwise be considered homogeneous using conventional techniques has epigenetically distinct subpopulations with different gene expression patterns," said Fang, an assistant professor of genetics and genomics, in a statement.

While their study focused on cultured bacterial strains, the methodology could also be applied to mixed populations of bacteria, they wrote, such as clinical isolates or microbiome samples.

Also, in addition to bacteria, it could be used to analyze DNA from viruses and human mitochondrial DNA.

In their study, the researchers used PacBio's single-molecule real-time (SMRT) sequencing technology, but their approach could be modified, they wrote, to analyze data from other real-time sequencing techniques, such as nanopore sequencing.

PacBio sequencing has already been utilized in other studies to detect almost 20 different types of DNA modifications, including DNA methylation. The technology detects modified DNA by measuring the time between the incorporation of nucleotides by a DNA polymerase, called the inter-pulse duration (IPD). Changes in IPD, or kinetic variation, are correlated with a specific type of DNA modification in the template.

However, studies so far have analyzed IPD data from several cells in aggregate rather than from individual cells, which "fundamentally limits the ability to resolve epigenetic heterogeneity within the sample," the researchers wrote.

To study heterogeneity of DNA methylation in greater detail, they developed SMALR, which relies on two complementary methods that both use IPD data from single molecules to infer their methylation states. SMALR is freely available to other research groups through GitHub.

The first method relies on circular consensus sequencing of short inserts of about 250 base pairs, where the polymerase goes around in a circle and reads each base several times. Comparing IPD values from native DNA with IPD values from whole-genome amplified DNA, which has lost its methylation, the researchers calculated a score that allowed them to detect methylation at specific sequence motifs within a single DNA molecule.

The second method uses long-insert libraries of about 3,000 to 7,000 base pairs, so each PacBio read represents a long, contiguous DNA molecule. Those reads are used to phase methylation at single-molecule resolution, again using IPD data from the single reads and IPD data from the same sequence motifs in whole-genome amplified DNA. This method can reveal small fractions of cells in a population that contain active or inactive methyl transferase enzymes.

To start, the researchers tested the sensitivity and specificity of their methylation detection method by analyzing one sequence motif in an E. coli strain and a matched whole-genome-amplified sample for N6-methyladenine. Sensitivity increased with coverage per molecule and reached 98.5 percent, while specificity was 99.5 percent. They were also able to detect 4-methylcytosine, with slightly lower sensitivity and specificity, but did not try to find other modifications.

In addition, they found that even with a low percentage of native DNA compared to whole-genome amplified DNA, they obtained good estimates of methylation. This, they noted, "could have implications for the characterization of in vivo isolates, for which low sequencing coverage due to limited DNA input is often a challenge."

They also applied their approach to six other bacterial species, including C. salexigens, H. pylori, C. crescentus, G. metallireducens, and C. jejuni, and found that it improved resolution and revealed distinct types of epigenetic heterogeneity.

Their current approach looks at the kinetic signature of a single base, but, in principle, it could integrate signatures from several surrounding bases, they wrote, allowing them to interrogate other types of DNA methylation, as well as different kinds of DNA damage.

"The application of SMALR and its integration with other single molecule- or single cell-level data, such as RNA and protein expression, will enable a more detailed understanding of the functions of DNA methylation in bacterial physiology," they wrote.