

# **Modeling Methylation Patterns with Long Read Sequencing Data**

Michelle Lacey, PhD

Tulane University, Department of Mathematics, New Orleans-LA

## **Abstract**

Cytosine methylation is a fundamental epigenetic process that regulates gene expression. Variation in cytosine methylation at CpG dinucleotides is often observed in genomic regions, and analysis typically focuses on estimating the proportion of methylated sites observed in a given region and comparing these levels across samples to determine association with conditions of interest. While sites are tacitly treated as independent, when observed at the level of individual molecules methylation patterns exhibit strong evidence of local spatial dependence. We previously developed a neighboring sites model to account for correlation and clustering behavior observed in two tandem repeat regions in a collection of ovarian carcinomas. We now introduce extensions of the model that account for the effect of distance between sites as well as asymmetric correlation in de novo methylation and demethylation rates. We apply our models to published data from a whole genome bisulfite sequencing experiment using long reads, estimating model parameters for a selection of CpG-dense regions spanning between 21 and 67 sites. Our methods detect evidence of local spatial correlation as a function of site-to-site distance and demonstrate the added value of employing long read sequencing data in epigenetic research.