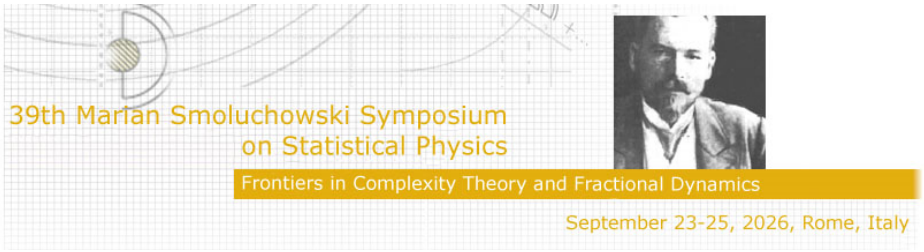

Contribution ID: **2** Type: **Invited talk**

Estimation of Methylation Entropy Distribution in Genomics

Friday, 25 September 2026 12:30 (30 minutes)

DNA methylation is an epigenetic modification that adds a methyl group to cytosine within CpG dinucleotides. CpG-rich regions often occur in gene promoters and regulatory elements. Methylation at these sites is a relatively stable, heritable mark that modulates transcription, chromatin state and genome stability [1]. Dysregulation is linked to several diseases including cancer [2].

Each CpG on a single DNA molecule is either methylated or unmethylated. While mean methylation levels capture the overall tendency of a locus to be methylated, they fail to differentiate between fundamentally different underlying stochastic configurations.

We use Shannon entropy to quantify randomness and develop a simulation framework to explore the entropy landscape [3]. We sample probability vectors from a symmetric Dirichlet prior on the simplex and condition them to a specified mean methylation. Accepted distributions are used to compute entropy and site-level summaries. Simulating ~10 million reads per setting and comparing to whole-genome nanopore calls from human blood, we found that entropy is low at extreme means and peaks near 0.5. Increasing CpG count expands the space of possible methylation patterns and so increases entropy, while the Dirichlet concentration controls dispersion.

Shannon entropy is thus a sensitive metric of epigenetic heterogeneity, while simple Dirichlet-based models are able to reproduce the characteristics of distributions seen in empirical data.

[1] Bird A. Genes & Development. 2002;16(1):6–21.
[2] Baylin SB, & Jones PA (2016). Cold Spring Harbor perspectives in biology, 8(9), a019505.
[3] Asante E. Estimation and Applications of Some Models for Dependent Data in DNA Methylation and Genomic Mutation Data. (Doctoral Dissertation) Rice University, 2026.

Primary author: ASANTE, Emmanuel (Department of Statistics, Rice University, Houston, TX, USA)

Co-authors: KURPAS, Monika (Department of Systems Biology and Engineering, Silesian University of Technology, Gliwice, Poland); LI, Meng (Department of Statistics and Ken Kennedy Institute,, Rice University, Houston, TX, USA); KIMMEL, Marek (Departments of Statistics and Bioengineering; Ken Kennedy Institute, Rice University, Houston, TX, USA; Department of Systems Biology and Engineering, Silesian University of Technology, Gliwice, Poland); VAN LOO, Peter (Department of Genetics, Division of Discovery Science and Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA); OGILVIE, Huw (The University of Texas MD Anderson Cancer Center, Houston, TX, USA)

Presenter: KURPAS, Monika (Department of Systems Biology and Engineering, Silesian University of Technology, Gliwice, Poland)

Session Classification: Session 8