Functional data analysis of high-resolution "Omics" data

Marzia A. Cremona[1,2]; Francesca Chiaromonte[3,4]; Kateryna D. Makova[5]

[1]   Department of Operations and Decision Systems, Université Laval, Canada
[2]   CHU de Québec – Université Laval Research Center, Canada
[3]   Department of Statistics, The Pennsylvania State University, USA
[4]   EMbeDS, Sant'Anna School of Advanced Studies, Italy
[5]   Department of Biology, The Pennsylvania State University, USA

Abstract:
Recent progress in sequencing technology has revolutionized the study of genomic and epigenomic processes. These assays generate massive, high-resolution "Omics" data often suitable to be represented as curves over the genome, whose analysis poses great challenges to standard statistical methods. Indeed, these data pose several problems such as large amounts of noise and correlations among neighboring measurements. Functional data analysis (FDA) can overcome these problems, reducing data dimension and filtering noise through smoothing, while taking advantage of the correlations among neighboring genomic positions and leveraging shape information within curves.
I will present three main research directions in which FDA recently proved to be effective: (1) studying and contrasting the genomic landscape of regions/loci of interests; (2) incorporating shape in the analysis of high-resolution signals of the epigenome; (3) analyzing complex quantitative phenotypes that can be represented as curves or surfaces.

Keywords:
Functional data analysis; Permutation tests; Linear models; Clustering; Genomics

References:
1. Cremona, Xu, Makova, Reimherr, Chiaromonte, Madrigal (2019) Functional data analysis for computational biology. *Bioinformatics*, 35(17):3211–3213.
2. Chen, Cremona, Qi, Mitra, Chiaromonte, Makova (2020) Human L1 transposition dynamics unrevealed with functional data analysis. *Molecular Biology and Evolution*, 37(12): 3576–3600.
3. Cremona, Chiaromonte (2020) Probabilistic K-mean with local alignment for clustering and motif discovery in functional data. *arXiv* 1808.04773.
4. Guiblet, Cremona, Harris, Chen, Eckert, Chiaromonte, Huang, Makova (2021) Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nucleic Acids Research*, 49(3): 1497–1516.