Leveraging 3D chromatin architecture for the study of non-coding regulatory regions variants.

Elisa Salviato¹, Vera Djordjilović², Judith Mary Hariprakash¹, Ilario Tagliaferri¹, Koustav Pal¹, and Francesco Ferrari³,*

¹ IFOM, the FIRC Institute of Molecular Oncology, Milan, 20139, Italy  
² Department of Economics, Ca’ Foscari University of Venice, Venice, 30100, Italy  
³ Institute of Molecular Genetics "Luigi Luca Cavalli-Sforza", National Research Council, Pavia, 27100, Italy

* Contact: francesco.ferrari@cnr.it

A growing amount of evidence in literature suggests that germline sequence variants and somatic mutations in non-coding distal regulatory elements may be crucial for defining disease risk and prognostic stratification of patients, in genetic disorders as well as in cancer. Their functional interpretation is challenging because genome-wide enhancer-target gene (ETG) pairing is an open problem in genomics. The solutions proposed so far do not account for the most updated knowledge on chromatin 3D architecture, which is organized in a complex hierarchy of structural domains.

In our work we introduce a paradigm shift based on multi-scale structural chromatin domains definitions integrated in a statistical framework to define ETG pairs. In this work 1) we develop a computational and statistical framework to reconstruct a comprehensive ETG regulatory network leveraging functional genomics data; 2) we demonstrate that the incorporation of chromatin 3D architecture information improves the accuracy in ETG pairing; and 3) we extensively benchmark our method against previous solutions for the genome-wide reconstruction of ETG pairs.

This solution will facilitate the annotation and interpretation of sequence variants in distal non-coding regulatory elements. We expect this to be especially helpful in clinically oriented applications of whole genome sequencing in cancer and undiagnosed genetic diseases research.

Key words: Computational genomics; Epigenetics; 3D chromatin architecture

Recent publications

   Global chromatin conformation differences in the Drosophila dosage compensated chromosome X.

2. Pal K, Tagliaferri I, Livi CM, Ferrari F.*
   HiCBricks: building blocks for efficient handling of large Hi-C datasets.

   Comparison of computational methods for the analysis of Hi-C data.
   Nature Methods, 2017 Jul; 14(7):679-685. (* co-corresponding/co-last authors)

   WoPPER: Webserver for Position Related data analysis of gene Expression in Prokaryotes.