Algorithms for inferring multi-contact interactions from chromosome conformation capture data

The objective of the project is to develop new algorithms to interpret three-dimensional organization of the genome, as measured by chromosome conformation capture experiments. The organization of the genome is typically studied by quantifying pairwise interactions between genomic loci, either in the form of adjacency matrix (inferred from Hi-C experiments), or linear adjacency profiles obtained for a number of 'viewpoints' (using 4C or Capture-C techniques). Many such interactions are observed when functional regulatory elements, located at a large linear distance along the genome, are brought into spatial proximity to their target genes.

Here, I propose to develop algorithms to fully exploit the potential of chromosome conformation capture data by inferring concurrent interactions between multiple loci. The identification of such multi-contact interactions would allow for better understanding of the underlying grammar of regulatory elements. In particular, we will determine whether individual genes are typically contacted by multiple regulatory elements concurrently, or whether these contacts are mutually exclusive.

The proposed research is organized into three specific aims. First, we will develop computational methods to infer multi-contact interactions from existing types of data, and apply them to published datasets to infer patterns which were previously missed. We will also generalize our approach to handle long sequencing reads originating from recently developed genome sequencing technologies. Secondly, we will propose new methods to interpret multi-contact interactions in allele-specific manner. This will allow us to analyse contacts between homologous chromosomes and identify the cases of transvection. Finally, we will scrutinize co-occurrence of transcription factor binding sites in regulatory elements at multi-contact interaction loci.