A Novel Computational Model for Discovery of DNA Spatial Interactions: New Insights into Genomic Organization

Galasso, Joseph (School: Galasso Homeschool)

Recent research has determined that genes are regulated by a complex, changing network of DNA spatial interactions. However, current wet-lab technologies for characterizing the 3D genome are often inaccurate and generally find static, rather than dynamic, structural elements. This study created an efficient computational framework to derive the location of 4D interactions utilizing epigenetic and expression data from embryonic and immortalized cell lines. For this purpose, Random Forest machine learning classifiers were constructed to learn patterns in markers associated with DNA regulatory structures. These models were optimized with intelligent generation of training datasets and feature tuning, resulting in accuracies and sensitivities > 0.9. Predictions were validated using high-quality wet-lab structural data, and many show statistically significant ($p<0.001$) evidence of being dynamic. This enabled construction of a 4D mapping tool, which determines how interactions change as cells progress through their life cycle. 4D patterns were used to create a new model of genomic organization that fully incorporates proteomics findings from previous literature. This model was used to gain new insight into how structure is implicated in gene control and how structural changes may be associated with genomic phenomena in cancer, such as the Philadelphia Chromosome. Thus, this study's tools can inexpensively and accurately accelerate our current attempts to understand the genome as a dynamical system via elucidation of how DNA structure influences function.

Awards Won: