Programmability, complexity and computability of large-scale cellular epigenetic networks.

Konstantinos Krampis[1,2], Claudia Wultsch[1,3]

[1] Bioinformatics and Computational Genomics Laboratory, Department of Biological Sciences,, City University of New York, NY, USA
[2] Institute of Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA
[3] Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA

Correspondence: 413 E 69th St, Belfer Research Building, New York, NY 10021, USA. E-mail: agbiotec@gmail.com

Here, we provide a research summary that leverages Sequential Dynamical Systems (SDS), Cellular Automata (CA) and Algorithmic Information Theory (AIT) frameworks, for deciphering the complexity and computability of epigenetic networks. Specifically, we focus on cellular networks formed through long-range chromatin interactions, as these are captured chromatin conformation (Hi-C) genomic sequencing experiments. Using the theory of SDS, CA, AIT and modeling the networks as Boolean interactions between the elements, we analyze the spatio-temporal changes in the structure of such epigenetics interactions, using publicly available Next-Generation Sequencing (NGS) datasets from the International Human Epigenome Consortium, the ENCODE project or the Cancer Genome Atlas. These databases contain terabytes of epigenetic interaction data, mapped in great detail across different types of cells in healthy or cancer tissues, developmental stages or experimental treatments. Overall, the main objective of this research project is to discover the mechanistic processes and biological algorithms, that have evolved to operate on epigenetic networks adjusting their structure in response to internal or external perturbations. We also aim to identify the information processing capacity of these networks, formalize their capacity within a mathematical framework, while also quantify programmability from an evolutionary perspective as this is captured, for example using algorithmic complexity.

Currently, while complexity metrics such as clustering, centrality and other techniques have been applied to epigenetic interaction networks using the publicly available NGS datasets, no extensive study has been performed involving analysis of epigenetic networks from an algorithmic, computability or information theory perspective. The core computational hypothesis driving this research is whether there is a base set of interactions within the genome, uncovered for example when calculating the algorithmic complexity, and whether these base interactions catalyze the formation of the complete interaction network. Consecutively, the smallest programs found that catalyze the formation of the network, can be defined as the biological algorithms or causal factors defining the mechanistic processes that have evolved to code the epigenomic function through the chromatin structures.

By using the toolset of AIT, Kolmogorov and algorithmic complexity, we will also test the hypothesis whether there are minimal computable structures or programs that stay constant in size or vary in the instances of the epigenetic networks, and how this variance relates to the different cellular conditions where the chromatin interactions were captured. In order to find the smallest algorithm or programs encoded in the epigenome, we will model the epigenetic networks as Boolean networks based on matrices containing interacting genome loci, within the multiple public datasets for chromatin conformation (appropriate discretization will be applied). With the mathematical tools of SDS and CA we will encode the networks within these frameworks, and present results regarding the programmable structures and computability of the epigenetic networks, and whether there are minimal programs p that are responsible for mechanistically generating the various network instances found in the different biological conditions of either health or disease.

In conclusion, our research focuses on utilizing the toolset of algorithmic information, dynamical systems and computability theory, towards finding the innate cellular programs that control the structure, function and programmability of epigenomic networks. We believe that this research will have a broad impact by providing new insights when applied to large-scale, publicly available NGS datasets. With this novel approach, we can uncover the structure of the minimal programs p or otherwise epigenetic network modules acting across different cellular conditions, developmental stages, disease or perturbations. Furthermore, by comparing the minimal programs p in the different conditions, we can uncover whether there are foundational biological algorithms which generate the different network instances in both health and disease, though network programming and control within the cells. This presents an unprecedented opportunity for discovery of the foundations of molecular biology, and for identifying the causes of diseases such as cancer, which in great part result from malfunction of epigenomic agents of cellular control.