Alina Grigorovitch
Epigenetics and Gene Expression
November 2016

The Importance of Epigenetics in Bioinformatic Analysis

As more information emerges about genetic and epigenetic regulation, the field of bioinformatics can no longer afford to seek for answers strictly from DNA sequences alone. A wealth of information is available in epigenetic sources, such as the methylome (pattern of DNA methylation) and epigenome (the record of chemical changes to DNA and associated histones throughout an organism's life). Due to their mutability and higher sensitivity to environmental stressors, the methylome and epigenome are better markers, and at times causes, of disease than the genome itself.

Understanding epigenetic mechanisms and signals is crucial to any area of study and disease treatment directly involving DNA. These mechanisms – DNA methylation, histone modifications, RNAi – both organize DNA and are the language by which DNA communicates with transcriptional machinery and other regulatory factors. Greater portions of DNA once thought to be "junk" because they did not code for gene products carry epigenetic significance, and this suggests that a majority of our DNA might have a use in mechanisms that have not yet been uncovered. Regions of DNA outside protein sequences function as promoters, enhancers, and inhibitors, and may be *cis* or *trans*, creating an important marker in a seemingly random location. The orientation of these regions is also related to the packaging and folding of DNA, as well as conformational changes that occur in response to stimuli. One motif sees a segment of DNA either becoming exposed or changing conformation in response to machinery recruited by an enhancer element. This change in conformation may now expose a promoter, inhibitor, or other marker. Another feature of found in "junk" DNA is CpG islands, a relatively rare "CG"-rich sequence preceding transcription start sites that are prone to methylation, another regulatory feature. In short, much of what has been thought of as useless DNA serves to regulate DNA transcription and packaging. Cellular mechanisms in higher eukaryotes are redundant to allow for the fine tuning which all of the interconnected processes in complex systems need.

Epigenetic signals also regulate cell differentiation by permanently repressing genes from transcription based on cell type, and in temporarily repressed genes, histone modifications are also responsible for converting those segments between euchromatin and heterochromatin. Therefore, epigenetic factors play a key role in phenotype, and epigenetic dysregulations may account for differences between genotype and phenotype when there is no error in the sequence. This is also true of certain diseases, including neurodevelopmental, cardiovascular, and type 2 diabetes, in which epigenetics are important for disease progression. Finally, changes in epigenetic markers account for the "nurture" aspect of the nature vs nurture debate, as monozygotic twins who spent their lives in different environments carried different epigenetic markers. Thus, epigenetic markers are also thought to be one of the causes of aging.

Similarly, epigenetic markers are complicit in conferring immortality to stem and cancer cells by hypermethylating the regulatory inhibitors of key growth factors and allowing for unchecked cell proliferation and reduced apoptosis. Because epigenetic changes are prevalent in virtually every cancer, the current direction of cancer research is in investigating the epigenetic basis of disease, looking for epigenetic markers, and identifying mutations or dysregulation in histone deacetylases and other histone modifiers. A better understanding of he tepigenetic mechanisms involved in reprogramming, such as *de novo* methylation and the methylation markers that are still carried from somatic cells to stem cells to induced pluripotent stem cells will allow for the development of safer and more effective

therapies with these cell lines. Surveys of the epigenome could be useful in providing real-time information and may be a method for assessing the effectiveness of stem cell transplants as they provide biomarkers of cellular events.

In summary, epigenetics and gene regulation account for DNA's dynamic and mutable qualities, and are essentially another layer of code atop the genetic code.