Title: Association study on ENCODE data for the function of epigenetic marks on gene expression

Authors: Xulong Wang and Greg Carter

Abstract: Histone methylation and acetylation are critical in defining the epigenetic landscape of cells and tissues, and are closely associated with genetic features including promoters, enhancers, and repressors. To delineate the functions of epigenetic modifications, the ENCODE mouse project generated a significant amount of data on histone modification, transcription factor binding, and gene expression in 80 cell and tissue types. In this study, we focus on how different histone marks interact with each other and how multiple histone marks act jointly to effect gene expression by integrating histone modification, transcription factor binding, and gene expression data on 11 samples from the ENCODE mouse project. We found significant enrichment of histone modification activity in terms of the number of separate peaks in the genetic coding region compared to those in the rest of the genome. The pattern of H3k27ac marks showed significantly high levels of tissue specificity. To study the associations between epigenetic marks and gene expression, we defined segments of genomic regions with enhanced epigenetic activity as “epigenetic hotspots”, and assigned “epi-genotypes” for individual marks and samples in the hotspots. Multiple linear regression models were used to study the interactions between different epigenetic marks and the association between epi-genotypes and gene expression. Both trans- and cis- epi-genotypes were discovered as being strongly linked with gene expressions. Our study presents a unique venue for integrating large-scale genetic and epigenetic data. Our results suggest patterns of epigenetic interactions and the effects of epigenetic features on gene transcription.