A Genome-Wide Analysis Tool to Identify Functional Regulatory Single Nucleotide Polymorphisms (rSNPs) Impacting Disease

Ng, Natalie

My research develops a tool to enable the first ever genome-wide analysis of genetic variants, specifically single nucleotide polymorphisms (SNPs), in transcription factor binding sites. The vast majority (88%) of disease associated SNPs lie in non-coding regions. If a SNP lies in a transcription factor binding site, characterizing this variant could pinpoint its mechanism of action and lead to improved diagnostic and treatment strategies. My project has three phases: (1) development of an intragenomic analysis tool, (2) development of a methylation analysis module, and (3) applying the tool to identify new relationships between transcription factors and disease. In Phase 1, I proposed and developed a novel intragenomic analysis tool, which seamlessly integrates three types of genomic data: sequence (PWM), transcription factor binding (chIP-seq), and open chromatin (DNase-seq). SNP Effect Matrix (SEM) scores, generated through the developed tool, predict downstream gene expression changes better than the current standard. In Phase 2, I developed the methylation analysis module, which enables the first ever genome wide analysis of DNA methylation in transcription factor binding sites. In Phase 3, I successfully applied the tool to analyze a list of 55,000+ disease associated SNPs. The tool has been prepared as an open-source software package using GIT source code management and will be released as an addendum to an upcoming publication. Major accomplishments of my research are (1) developing a comprehensive framework to characterize genetic variants in regulatory regions, (2) enabling the first ever genome-wide analysis of the regulatory role of DNA methylation, and (3) identifying statistically significant relationships between transcription factors and disease.

Awards Won:

Second Award of $2,000