

Title: **Prediction of *cis*-regulatory variations causal for Parkinson's disease**

¹Patrick Tan, ²Virginie Bernard, ²Carles Vilarino-Guell,
²Matthew Farrer and ^{2,3}Wyeth W. Wasserman

¹Bioinformatics Graduate Program, University of British Columbia, Vancouver BC, Canada

²Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute

³Department of Medical Genetics, University of British Columbia, Vancouver BC, Canada

Introduction

Most of the studies that have been conducted on the identification of causal variations of human-inherited diseases so far have only focused on mutations in protein-encoding exons. Nevertheless, many disorders result from genetic variabilities that are found in non-coding regions. Parkinson's disease (PD) is a neurodegenerative disorder that affects more than 1% of the aging population who are older than 65 years. It is characterized by impaired balance and walking, slowed movements, muscle rigidity, and resting tremors. It is associated with dopaminergic neuronal cell loss in the substantia nigra in the midbrain. While the molecular mechanisms of this complex disease are still poorly understood, it has been demonstrated that mutations in the coding regions of the leucine-rich repeat kinase 2 (LRRK2) and synuclein alpha (SNCA) genes are linked to pathogenesis. Researchers have recently reported a relationship between mutations in the non-coding regions of the DJ-1 gene and Parkinsonism. We hypothesize that variations in elements that regulate LRRK2 and SNCA may also contribute to PD.

Goal

To further understand the etiology behind PD, we propose a bioinformatics pipeline that will predict and score deleterious effects of *cis*-regulatory sequence variations on non-coding regions which will be validated in the lab.

Approach

Initially, we will map short-reads from high-throughput sequencing experiments to the reference genome and call for sequence variations. Those variations that meet quality criteria and are located outside of protein-encoding exons will be analyzed. Core promoter regions will be predicted based on sequence conservation and proximity to CpG islands. Next, we will use transcription factor binding site (TFBS) profiles from open-source databases such as JASPAR and also use PAZAR to build new profiles of TFBS from publicly available ChIP-seq data that are known to be involved in transcription. These TFBS profiles will be used to score the effects of the variations with respect to transcription factor binding. Variations that have different TFBS prediction scores between variant and reference sequences are predicted to have a significant impact on transcription factor binding which may have contributed to PD. These variations may also provide insights into the evolutionary relationships among similar protein family members.