Discovery of Novel Subtypes in Parkinson’s Disease

Xiang GU1, Hongxu Ding1, Jinghao Sun1

1. contribute equally, names not listed in order

Abstract

Introduction

Parkinson’s disease (PD), a long-term neurodegenerative disease, mainly affects the motor neural system. Patients of PD bear with disabilities in obvious shaking, rigidity and slow movement. Unfortunately, there is no available therapeutics against the underlying neurodegenerative process currently. Therefore, it is of great importance to understand how the disease develops, how many subtypes within the disease and how to perform health care for different patients. This study aims to classify the patients based on their genetic data to identify novel subtyping mechanism for PD.

Genetic information has been successfully applied in subtyping and classifying other diseases including complex disease as cancer. Genetic mutations in the protein coding region imply potential alteration of protein structure and function which contribute to the disease development and progression. Therefore, patients can be classified the mutations they harbor which is in correlation with the cause of their disease. This theory has been proved successful in many cancer studies, including breast cancer[ref] and glioblastoma[ref]. In this study, we analyzed the mutation information of 465 patients obtained via Exome-seq which sequenced only the protein coding regions in current cohort, to identify novel subtyping scheme for PD.

Subtyping and identifying potential subtypes of a disease falls in the regime of unsupervised clustering problem. Numerous methods for unsupervised clustering have been developed with each’s own advantages and disadvantages depending on the format and features of unlabeled data. These methods have been widely used in other fields to understand both real-life problems such as pattern mining in customers and biomedical problems such as subtyping of breast cancer[ref]. Under the concept of social networks, analysis on bipartite graph formulated by the patients and their mutations proposes a novel dimension in studying subtyping of a disease [ref]. Therefore, in this study, we apply both unsupervised machine learning algorithms and social networks to identify the subtypes of PD.

Method

Preprocessing of Exome-seq Data

Exome-seq of 465 patients was downloaded from Parkinson's Progression Markers Initiative (http://www.ppmi-info.org/). Variants were annotated using hg19 with ANNOVAR and package VariantAnnotation in R. Only the unique single nucleotide polymorphism (SNP) with amino acid sequence changed were selected. The overlapping filtered SNPs from both packages were selected as final candidate features used for subtyping.

Dimension Reduction

Unsupervised Clustering

Bipartite Graph Analysis

Result

Discussion

Conclusion

Reference