Title: A non-parametric method for joint association analysis of sequencing and Imaging data

Authors: Tong, X., Qing, L.

The rapid development of whole genome sequence (WGS) technology coupled with magnetic resonance image (MRI) data mandates the development of analytical methods that are capable of utilizing both WGS and MRI data to identify predictive biomarkers associated with neurodegenerative diseases, such as Alzheimer’s disease. The rich WGS/MRI data, however, brings the issue of "the curse of dimensionality" due to the vast number of sequencing variants and brain surface vertexes. In this work, we tackled the dimensionality issue of MRI data through a stacked denoising autoencoder (SDA) constructed using the deep learning algorithm, which reduces the dimensionality and maintains the majority of the information. For the WGS data, we use a weighted identity-by-state (IBS) kernel to aggregate information over multiple sequencing variants in a genetic region. A weighted U statistic is then used to evaluate the joint association of both imaging and sequencing data with the phenotype of interest. We show that our method maintains the correct type 1 error rate, while achieving high statistical power in comparison to methods using either sequencing or image data alone. To illustrate our approach, we apply the proposed method to the sequencing and image data from the Alzheimer's disease Neuroimaging Initiative.