Post-Seminar Journal Entry for Yu Zhang Paper

Conclusions: Neural Networks are artificial models of neuronal networks found within the nervous system of biological Humans. Artificial networks consist of a series of connected nodes, which transmit information or 'signals' between each other. Signals or outputs are transmitted to another node, if the sum of the weights of the inputs or signals received by the original node reaches a certain threshold. The complexity and nonlinear function through which the signals are processed forms the reasoning behind their use in predicting age using DNA methylation levels. If the model becomes too complex, however, the network can result in the use of duplicated outputs to find a data pattern. A high variance model results in this case leading to incorrect age predictions when new DNA methylations are introduced to the constructed model.

To correct for this disadvantage found in basic neural networks, regularization techniques and further approaches were utilized to reduce the complexity of the system. The techniques: LASSO, elastic net, drop out, and pre-filtering neural networks, basic neural networks, and two statistical models were assessed on their average number of errors accumulated for their accuracy in the prediction of age by DNA methylation levels. The results of the assessment found neural networks, regardless of the use of a regulator, to predict age more accurately than the commonly used two statistical approaches. With the use of regulators, all except one approach coupled with neural networks were shown to increase the accuracy of age prediction by the model. LASSO was the only model found to decrease the accuracy of the model due to the increase in prediction errors. While more can still be achieved in the discovery of the LASSO approach and its interaction with neural networks, it is assumed the extraction of random connections that produce the same output to create a sparse model leads to difficulty in finding a pattern between the DNA methylation levels and the predicted biological age. In contrast, the correlation pre-filtered neural network approach (CPFNN) was found to be the only model leading to a statistically significant reduction in the average number of errors produced by the model. The approach decreased the amount of inputs used in the creation of the model but maintained the importance of high associations between certain CpG sites and age leading to a distinct and accurate pattern.

CPFNN was applied with the two statistical approaches to evaluate the age acceleration of individuals with and without schizophrenia and down syndrome. The results demonstrated CPFNN to be a more accurate age predictor even in applied situations, for even in diseases that presented a weak association with age, CPFNN was able to recognize these patterns or associations unlike the two statistical models. Overall, the results of this study point to a new, more accurate approach to predicting age by an individual's DNA methylation levels, and imply a medicinal application through the ability to determine the difference between biological age and chronological age leading to improved knowledge about a patient's condition and whether they have a disease. Moreover, it can be used to discover which diseases have an association

with age, and can allow physicians to begin the process of scanning for age-related illnesses within a patient before the onset of symptoms.

Importance: The results of the paper provide new informative information detailing the discovery of an improved method to assess age discrepancies between your chronological and biological ages. Implications of this knowledge, once made readily available to healthcare professionals, can lead to not only an improved assessment of the patient's health, but also early scanning and diagnosis of age-related diseases prior to the onset of symptoms. Furthermore, aside from patient care, this can lead to advanced knowledge between the association of age and various diseases. Utilizing this new knowledge, health care professionals can further test for the presence of age-related diseases from an improved and current list. Results for this study were consistently controlled throughout the study with the prevention of splitting bias, unequal age distributions, and through added control by hyperparameters for the learning/training process. Even though the model could have a greater effect if it pointed to a specific age-related disease or range of age-related diseases, it still presents valuable, reliable, previously unknown information that can be utilized by health care professionals in the improved treatment and diagnosis of patients, but also in the advancement of current knowledge regarding various diseases and age. Thus, the results are important.