Abstract Title: Genetic and brain imaging phenotype joint prediction of longitudinal Parkinson's Disease subtypes
Control Number: 646
Topic: 18. Bioinformatics, Machine Learning and Statistical Methods
Presentation Preference: Oral Presentation
Applied for Early Career Award and/or Fellowship:

Authors:
Linda Polfus*[1, 2], Hamed Masnadi Shirazi[1, 2], Alexandra Reardon[3], Matthew Varga[1, 2], Alexander Tchourbanov[1], Taylor Gosselin[3], Anant Dadu[4, 5], Faraz Faghri[4, 5], Mike Nalls[4, 5], Jacob Hesterman[3], Roger Gunn[2, 6], Adam Chamberlin[1, 2], Brian Avants[2, 3]

[1]Ambry Genetics, Advanced Analytics, Aliso Viejo, CA, United States, [2]REALM IDx, Inc., Aliso Viejo, CA, United States, [3]Invicro, Needham, MA, United States, [4]Data Tecnica International, Washington DC, United States, [5]National Institute On Aging, Center for Alzheimer's and Related Dementias, Bethesda, MD, United States, [6]Invicro, A Konica Minolta Company, London, United Kingdom

Consortium name:

Background/Objective: The characterization of Parkinson's Disease (PD) progression involves several heterogeneous clinical attributes, such as motor function or cognitive decline, to define disease stage. PD subtyping to identify fast or slow disease progressors remains especially challenging. We utilize -OMICs data (genomic and brain imaging) along with HPO (Human Phenotype Ontology)/GO (gene ontology) term embedding (node2vec), CSF biomarkers, and longitudinal clinical measurements to predict PD progression subtypes.

Methods: A Neural Net based Long Short-Term Memory (LSTM) model was employed on participants of the Parkinson's Progression Marker Initiative (PPMI, N=600) with clinical features at 2 time points (baseline/2 years) capturing motor changes, cognitive impairment and memory scores. Whole genome sequenced data filtered on 90 known PD associated genetic variants along with HPO/GO terms and imaging data (T1-weighted neuroimaging for volume and thickness) were used to train the LSTM model into a 2-D embedded vector.

Results: We identified specific genes (*CTNNB1, SHH,* and *GLI2*), HPO embeddings/biologic pathways via GO terms combined with brain regions characterizing fast/slow PD progressors independent of Hoehn and Yahr disease stage. Utilizing embedding algorithms along with HPO/GO terms increased the Area Under the Curve (AUC) by ~5% compared to using genetic variants directly. Clinical+imaging inputs provided the highest AUC for two (92.2%) or three (77.4%) subtype models.

Conclusions: Our findings provide mapping between genetic risk variants, symptom progression along with brain network health useful for clinical trial inclusion. Future personalized medicine applications include specific drug targets by brain region and PD-related biologic mechanism.