Genes regulating cytoskeleton organization identified as neuro-common drivers of blood NfL change rate and disease-specific clinical progression in AI driven digital twins

*X. SHEN[1], D. SHOKEEN[1], O. ISACSON[2], R. HARRISON[1], S.-Y. SHIN[1], J. LATOURELLE[1]
[1]Aitia, Somerville, MA; [2]Neuroregeneration Res. Inst., McLean Hosp. / Harvard Med. Sch., Belmont, MA

aitiabio.com

NEUROSCIENCE 2023

#10057

Background

- Elevated Neurofilament Light Chain (NfL) is commonly observed in neurodegenerative disorders (NDs), including AD and PD, and is widely recognized as a potential biomarker (Khalil et al. 2018). A recent study suggested that the rate of change in blood NfL may be a better predictor of AD progression than the absolute NfL at baseline (Preishe et al. 2019).
- Here, we use causal AI-based digital twins to identify genes and pathways driving the rate of change in blood NfL robustly across AD and PD.

Data

Training Data used for AD Digital Twins and PD Digital Twins

	Alzheimer's Disease	Parkinson's Disease
Cohort	Alzheimer's Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu)	Parkinson's Progression Markers Initiative (PPMI) & Genetic Cohort (www.ppmi-info.org)
Study Population	MCI (191) + Dementia (29) + Control (97)	De novo Idiopathic PD (277) + LRRK2 mutation carrier (49) + Unaffected LRRK2 mutation carrier (38) + Control (15)
Data Modality *(Feature Size)*	• Demo (4); Clinical Biomarker (5); MRI Imaging (15) • Genomic Variants (18,733) • Gene Expression in Blood (18,105)	• Demo (4); Clinical Biomarker (5); MRI Imaging (2) • Genomic Variants (21,008) • Gene Expression in Blood (Version **IR2**) (16,792)
Sample Size	N=317	N=514
Age	Mean = 72.4 (SD = 7.3)	Mean = 61.3 (SD = 9.7)
Sex	Female: 46%; Male: 54%	Female: 38%; Male: 62%

Methods

Gemini Digital Twins

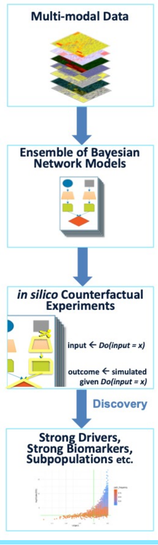
- 'Gemini Digital Twins' are virtual patients modeled and simulated using the REFS™ AI platform[2] as follows.

Reverse Engineering

- Each Gemini Digital Twins is comprised of a total of 128 Bayesian network models (called an ensemble) built from the training data.
- A Bayesian network model is a directed graphical representation of relationships between variables where each node denotes a variable, and each arrow denotes a conditional dependency.

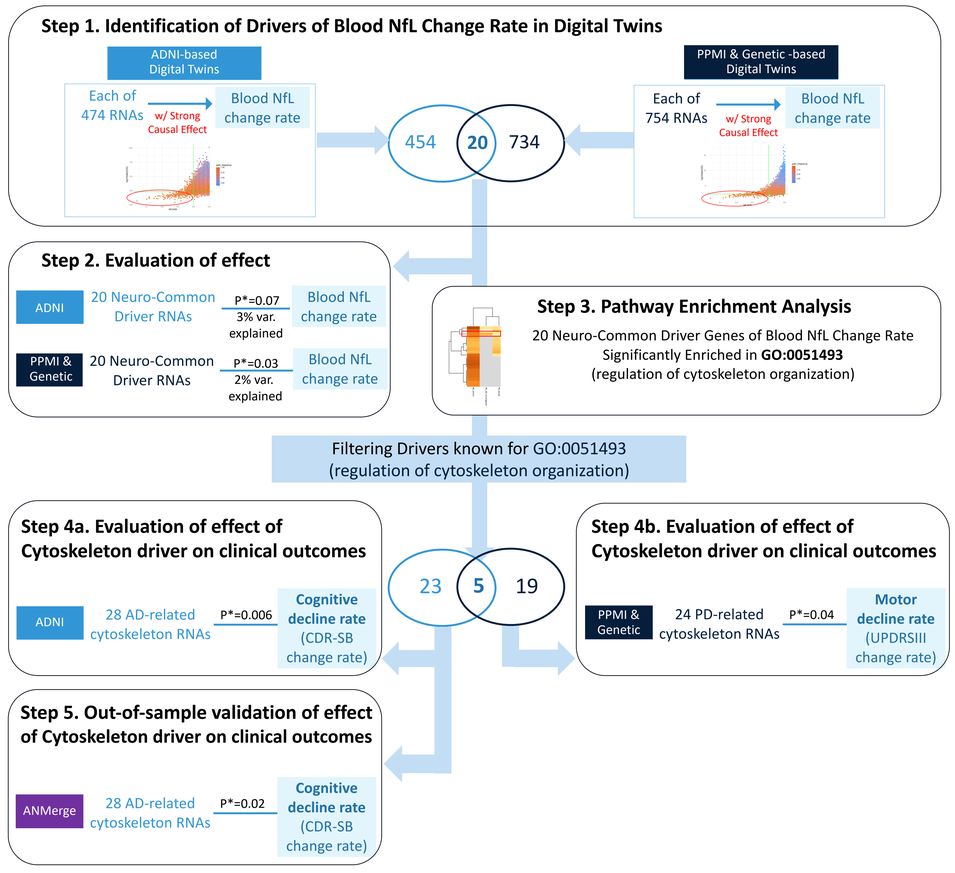
Forward Simulation

- Patient-level outcome values can be simulated in the Gemini Digital Twins, by *in silico* counterfactual experiments which computationally estimate the outcome values through model interventions, known in causal inference as 'Do' operations.
- These estimations are done fully adjusting for any confounding effects identified in the causal models, which is necessary in causal inference as emphasized in randomized experiments.



Results

Figure. Workflow and Results



*Multivariate linear regression model adjusting for Age, Gender, Genetic risk factor (Presence of pathogenic mutation for PD, and APOE4 genotype for AD)

Step 1. AD and PD Digital Twins identified 20 genes driving the rate of change of blood NfL in both diseases (neuro-common genes).

Step 2. These 20 genes, quantified at the blood transcript level, explain 3% of the variance of the NfL change rate in the AD train dataset, adjusting for age, sex and APOE4 genotype (p=0.07); and 2% in the PD train dataset adjusting for age, sex, and pathogenic variants (p=0.03).

Step 3. Pathway enrichment analyses found the 20 overlapping genes significantly overrepresented in various gene sets including GO:0051493, known for the regulation of cytoskeleton organization ($p < 0.0002$).

All additional gene drivers of NfL change rate in either PD or AD that are represented GO:0051493 were used to create disease-specific cytoskeleton gene signatures and evaluate their effect on clinical disease progression.

Step 4. The 24 total PD-related cytoskeleton genes were shown to drive motor decline in the PD cohort, measured as rate of change in UPDRS III score (p=0.04), while the 28 AD-related cytoskeleton genes strongly affected cognitive decline, measured by the rate of change in CDR-SB score (p=0.006) in the AD train dataset.

Step 5. The AD related signature was validated in an independent out-of-sample AD dataset (ANMerge, p=0.02).

Conclusions

- AI-based Digital Twins provided evidence that AD and PD may share common genes and pathways causally driving a neuro-common biomarker (NfL) as well as disease-specific clinical progression rates.
- Our finding that genes driving the NfL change rate are involved in the mechanism underlying the regulation of cytoskeleton organization is consistent with reports that cytoskeleton disorganization in neurons is a known early pathogenic event in multiple NDs and observed along with the excessive extra neuronal abundance of NfL.
- Further investigation of cytoskeleton genes may benefit potential drug targets for multiple NDs.

Reference

- Preische, Oliver, et al. "Serum Neurofilament Dynamics Predicts Neurodegeneration and Clinical Progression in Presymptomatic Alzheimer's Disease." Nature Medicine, 21 Jan. 2019
- Khalil, Michael. "Neurofilaments as Biomarkers in Neurological Disorders." *Nature Reviews Neurology*, vol. 14, no. 10, 31 Aug. 2018, pp. 577–589