Gemini Digital Twins Identified Neuro-Common and Disease-Specific Drivers of Blood NfL Change Rate

X.Shen[1], S.Sathe[2], L.Sun[1], P.Ashrap[1], K.Johnson[2], S.Sukhram[2], S.Reddy[1], S-Y.Shin[1], J.Latourelle[1], C.Sampaio[2]

[1] Aitia, Somerville, MA, USA, [2] CHDI Foundation, Princeton, NJ, USA

International Parkinson and Movement Disorder Society

Abstract #900

Objective

- To identify neuro-common or disease-specific drivers of NfL across neurodegenerative diseases, in two groups of Gemini Digital Twins: one built for combined Huntington's Disease (HD) and Alzheimer's Disease (AD) patients, and the other separately built for combined HD and Parkinson's Disease (PD) patients

Data

- Train Data Combining HD and AD data

	Huntington's Disease	Alzheimer's Disease
Cohort	TRACK-HD/TRACK-ON	ADNI
Study Population	Pre-manifest and Manifest (with HD-ISS[1] score of at least 2)	MCI and Dementia
Data Modality *(Feature Size)*	• Demo *(8)*, Clinical biomarkers *(8)*, MRI imaging *(3)* • Genomic variants *(14,914)* • Gene expressions in blood *(14,746)*	
Sample Size	N=73	N=275
Age	Mean=45.4 (std=9.9)	Mean=71.5 (std=7.4)
Sex	Female 57.5%, Male 42.5%	F 44.4%, M 55.6%

- Train Data Combining HD and PD data

	Huntington's Disease	Parkinson's Disease
Cohort	TRACK-HD/TRACK-ON	PPMI
Study Population	Pre-manifest and Manifest (with HD-ISS[1] score of at least 2)	*de novo* Idiopathic PD
Data Modality *(Feature Size)*	• Demographics *(5)*, Clinical biomarkers *(5)* • Genomic variants *(15,151)* • Gene expressions in blood *(14,340)*	
Sample Size	N=74	N=311
Age	Mean=45.6 (std=10.0)	Mean=61.8 (std=9.6)
Sex	Female 56.8%, Male 43.2%	F 35.4%, M 64.6%

Methods

Gemini Digital Twins

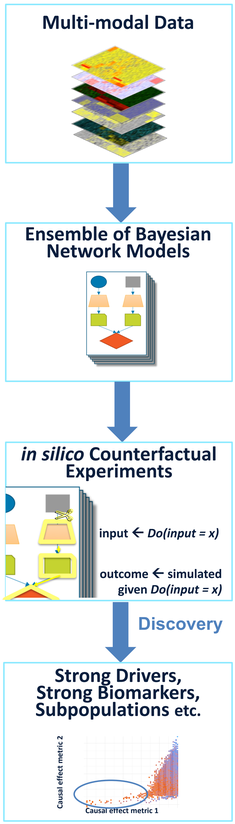
- 'Gemini Digital Twins' are virtual patients modeled and simulated using the REFS™ AI platform[2] as follows.

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 1. Study Workflow and Summary of Results

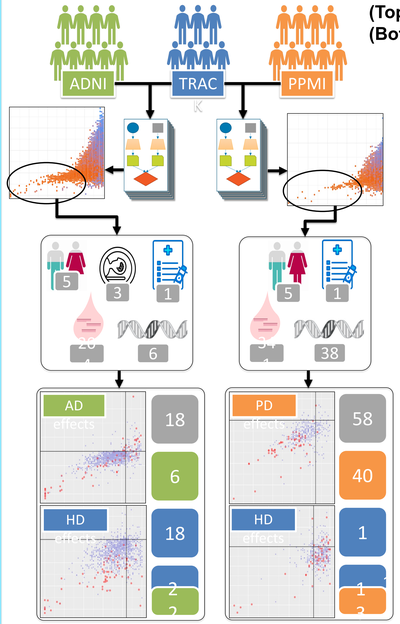
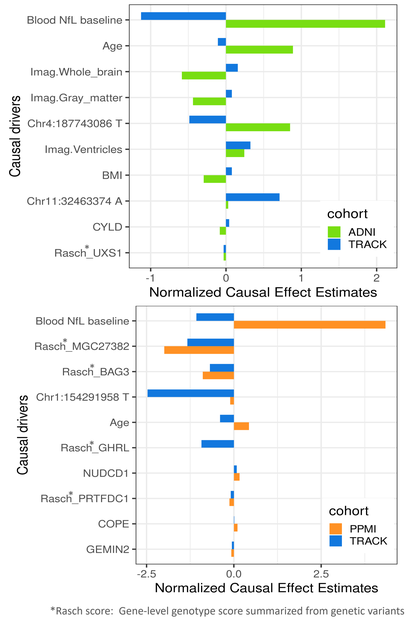


Figure 2. Disease-Specific Effects of Top Drivers of Blood NfL Change Rate, Defined by the Absolute Normalized Causal Effect: (Top) from the Model of Combined HD and AD, (Bottom) from the Model of Combined HD and PD

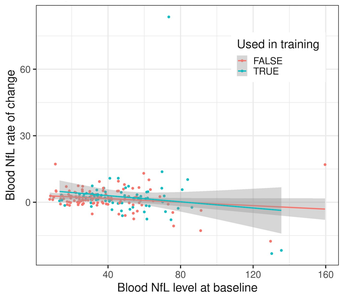


*Rasch score: Gene-level genotype score summarized from genetic variants

Table 1. GO Terms Enriched with Disease-Specific Genes Driving Blood NfL Change Rate in the Model of Combined HD and PD

Term	Description	Log(P)
GO:0016071	mRNA metabolic process	-6.7
GO:0016570	histone modification	-5.9
GO:0080135	regulation of cellular response to stress	-4.2
GO:0006974	cellular response to DNA damage stimulus	-4.1
GO:0006325	chromatin organization	-3.6
GO:1903706	regulation of hemopoiesis	-3.5
GO:0016055	Wnt signaling pathway	-3.2
GO:0045727	positive regulation of translation	-3.0
GO:0031647	regulation of protein stability	-2.9
GO:0051090	regulation of DNA-binding transcription factor activity	-2.4
GO:0048511	rhythmic process	-2.2

Figure 3. Correlation between the Baseline Blood NfL Level and the Blood NfL Change Rate Measured in the Train Data (HD only) and the Left-Out Independent HD Data



- Two separate sets of models were reverse-engineered from combined HD and AD patient populations, and combined HD and PD patient populations, separately.
- First, from the model from combined HD and AD, a total of 219 top causal drivers of blood NfL change rate were selected, which were further grouped as 18 HD-AD-common drivers, 18 HD-specific drivers, 6 AD-specific drivers, and 22 shared drivers with distinct disease-specific effects, through the Gemini Digital Twins simulation approach **[Figure 1]**. Notably, many top disease-specific drivers showed an opposite effect on NfL slope change (i.e. fast vs. slow accumulation) in HD and AD populations **[Figure 2]**. Disease specific drivers included age, BMI, blood NfL level at baseline, brain volume measures at baseline, a few genetic variants previously reported in the GWAS for AD and about forty genes (measured at the blood gene expression level) enriched in "mRNA metabolic process" and "histone modification" most significantly **[Table 1].**
- The second model based on HD and PD patient populations selected 385 causal drivers of blood NfL change rate, including 58 HD-PD-common drivers as well as 54 disease-specific drivers including age, BMI, education, blood NfL level at baseline, a few genetic variants or genetic variant summary scores, and some blood-based gene expressions **[Figures 1 and 2].**
- In particular, while the blood NfL level measured at baseline was a strong driver of its future change rate in all three diseases, the nature of that relationships was markedly different between diseases.
- The inverse relationship between baseline blood NfL and blood NfL change rate, was also observed in an independent HD patient dataset, supporting our finding **[Figure 3]**.

Conclusions

- The Gemini Digital Twins showed that higher baseline NfL in blood would likely lead to accelerated NfL accumulation in blood in both the AD and PD, but to slower accumulation in HD.
- This causal relationship was identified in combined multiple neurodegenerative patient populations using a data-driven, hypothesis-free REFS™ platform, and later independently confirmed in the out-of-sample validation data.
- This study is limited by the small samples size especially in the TRACK-HD/TRACK-ON, making the HD-specific effect estimation less powered than the other two disease specific effect or common effect.
- Although it may be difficult to validate the causal relationships discovered in this study, given that there is no comparable analytical tool available for in silico causal inference to our knowledge, these findings could provide a new insight on the relevance of biomarkers across neurodegenerative diseases.