Gemini Digital Twins Identified Both Common and Disease-specific Drivers of Cognitive Progression in Huntington's and Alzheimer's Diseases

L.Sun[1], S.Sathe[2], X.Shen[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

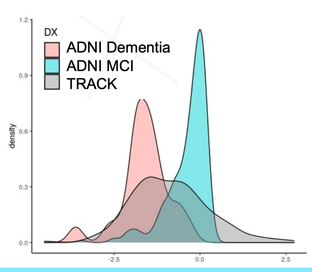
Objective

- To investigate molecular or clinical biomarkers with common or disease-specific effects on disease progression measures, using Gemini Digital Twins, generated from causal network models integrating multi-modal data across Huntington's Disease (HD) and Alzheimer's Disease (AD) studies in combination with *in silico* counterfactual experiments

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 Considered *(Feature Size of the Train Data)*	• Demographics *(8)* • Clinical biomarkers *(8)* • Genomic variants *(14,914)* • Gene expressions in blood *(14,746)* • MRI imaging variables *(3)*	
Clinical Outcomes Considered	• Rate of change of blood NfL • Rate of change of TMT-B • Rate of change of Harmonized Cognitive Score*	
Sample Size of the Train Data	N=73	N=275
Age	Mean=45.4 (std=9.9)	Mean=71.5 (std=7.4)
Sex	Female 57.5 % Male 42.5 %	Female 44.4 % Male 55.6 %

*For the harmonization of the SDMT and MMSE, available only in HD and AD cohort data, respectively, we investigated the distributions of the progression rates and confirmed their comparability.



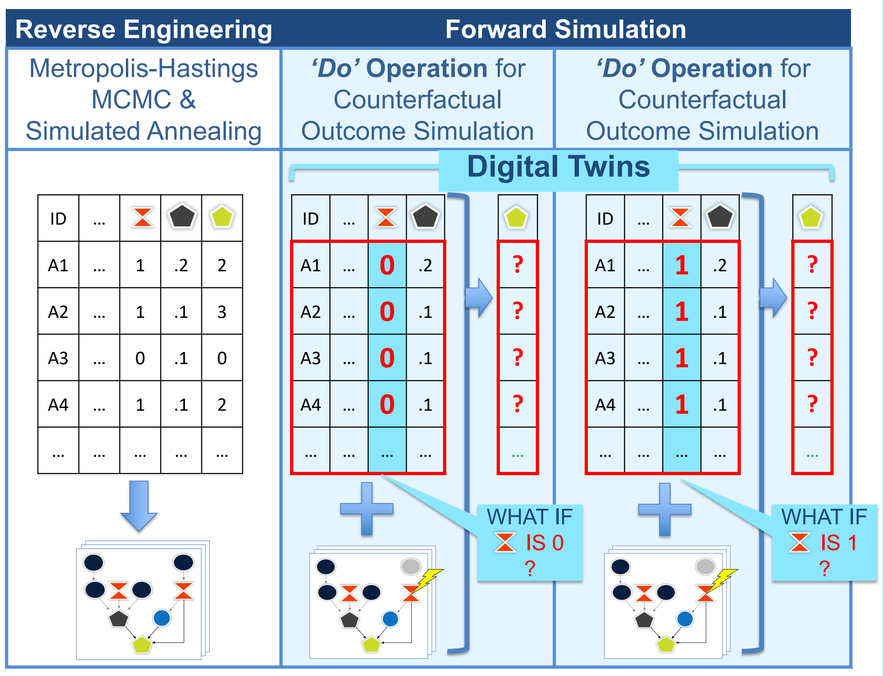
Methods

Reverse Engineering

- In REFS™ AI platform[2], an ensemble of Bayesian network models are learned and optimized using the Metropolis-Hastings Markov Chain Monte Carlo algorithm and simulated annealing techniques.

Forward Simulation

- The REFS™ causal models can estimate *in silico* counterfactual outcome values effectively and appropriately, by 'Do' operations.



Digital Twins

- 'Digital Twins' are virtual patients simulated using forward simulation based on the same ensemble of Bayesian network models with only differences in the variables of our interest (e.g. potential outcome drivers). The causal effects of these variables could be estimated by comparing simulated outcome values from these 'Digital Twins'.
- These estimations are done fully adjusting for any confounding effects identified in the models, which is necessary in causal inference as emphasized in randomized experiments.

Results

Figure 1. Study Workflow and Summary of Results

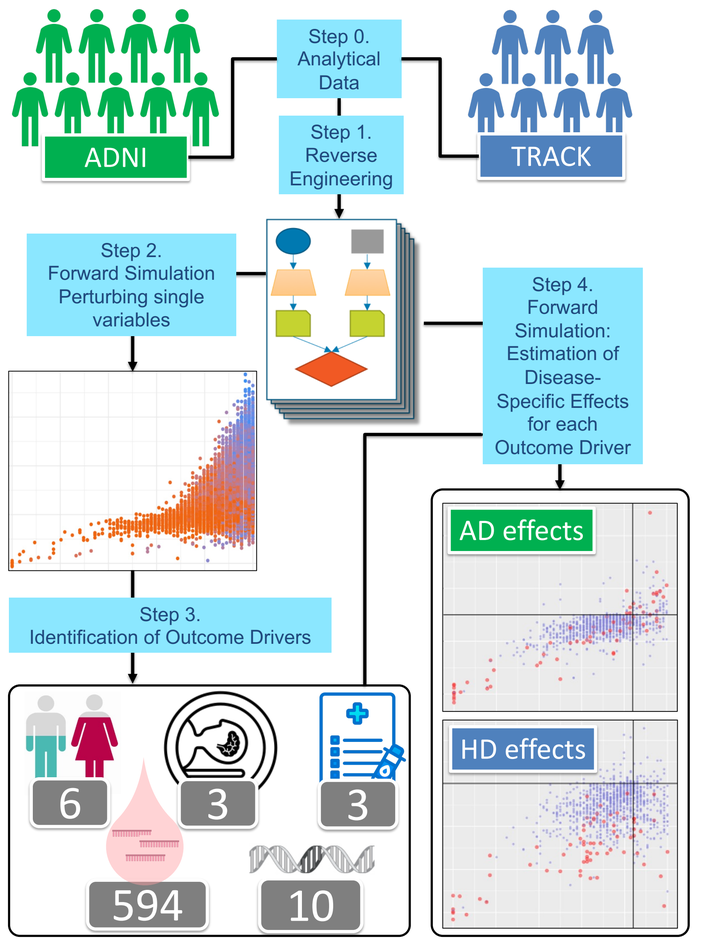


Figure 2. Outcome Drivers Identified from Demographic, Imaging, and Clinical Variables, with Their Effects Estimated in the Combined HD and AD Patients

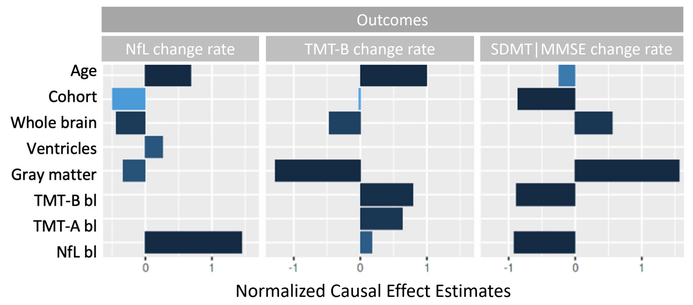
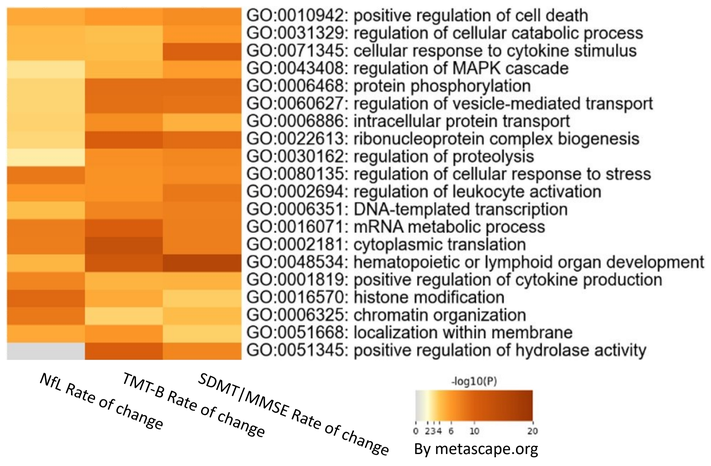


Figure 3. GO Terms Enriched with Genes Driving Rate of Change of blood NfL, Genes Driving Rate of Change of TMT-B, and Genes Driving Rate of Change of Harmonized SDMT & MMSE Scores

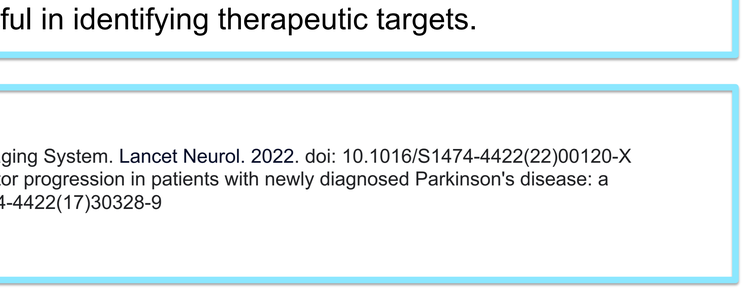


- Using the Gemini Digital Twins, the causal effects of 29,680 variables for three clinical outcomes were estimated. Outcome drivers with the strongest causal effects were identified first, and their disease-specific effects were further evaluated **[Figure 1].**
- Our study recapitulated known and expected drivers of disease progression measures (including age and corresponding baseline values) **[Figure 2]** and identified many potential novel genes (based on the blood RNA expression) among which the genes driving blood NfL change rate were enriched in the histone modification pathway, and the genes driving the two cognitive decline rates were enriched in GO terms for protein phosphorylation and regulation of hydrolase activity among many **[Figure 3]**. It was also found that drivers of the two cognitive decline rates largely overlapped (297 shared drivers out of 428 and 362), while drivers of blood NfL change rate were more distinct.
- Additional forward simulation for disease-specific effects enabled us to label the outcome drivers into four different types: 572 drivers with similar disease-specific effects, 37 drivers with distinctive disease-specific effects, 19 HD-specific drivers and 8 AD-specific drivers **[Table 1]**. Interestingly, only 28% of the drivers of blood NfL change rate showed similar effects between HD and AD, while most of the drivers of the two cognitive decline outcomes were HD-AD-common **[Table 1]**. The genes driving blood NfL change rate with distinctive disease-specific effects were enriched in GO terms for chromatic organization and histone modification **[Figures 3 and 4]**.

Table 1. Common and Disease-Specific Outcome Drivers

Type of Drivers	Number of Drivers for Rate of Change of		
	blood NfL	TMT-B	Harmonized SDMT & MMSE Scores
Drivers with Common Disease-Specific Effects	18	305	249
Drivers with Distinctive Disease-Specific Effects	22	9	6
HD-Specific Drivers (with HD-Only Effects)	18	0	1
AD-Specific Drivers (with AD-Only Effects)	6	1	1

Figure 4. GO Term Enriched with Outcome Driver Genes with Common Effects, Outcome Driver Genes with Distinctive Disease-Specific Effects or AD-Only Effects, and Outcome Driver Genes with HD-Only Effects



Conclusions

- The data-driven, hypothesis-free Gemini Digital Twins successfully replicated several known drivers for blood NfL change rate and the cognitive decline rates across HD and AD, and provided some interesting and potentially novel insights as well.
- Increase in the blood NfL at baseline appears to increase its rate of change in AD patients, however, decreases the rate of change in HD patients.
- On the other hand, increase in the blood NfL at baseline consistently worsens the cognitive decline rate in both AD and HD patients. The blood NfL change rate (estimated from the baseline up to 3 years) does not seem to be related with the cognitive decline rate (estimated from the baseline up to 3+ years).
- Common drivers of the two cognitive decline rates, measured by TMT-B and harmonized SDMT & MMSE, suggest a shared mechanism for cognitive progression in HD and AD despite the distinct inciting pathogenic triggers.
- Further work on the converging pathways may be useful in identifying therapeutic targets.

References

1. Tabrizi *et al.* A biological classification of Huntington's disease: the Integrated Staging System. Lancet Neurol. 2022. doi: 10.1016/S1474-4422(22)00120-X
2. Latourelle *et al.* Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. Lancet Neurol. 2017. doi: 10.1016/S1474-4422(17)30328-9