Alzheimer's Disease Mechanistic Pathways were Discovered through *in silico* Experiments in Causal AI based Digital Twins

Apoorva Bharthur Sanjay, Deepanshi Shokeen, Xinyu Shen, So-Youn Shin and Jeanne Latourelle; Aitia, Somerville, MA, USA

Introduction

- We created Digital Twins of Alzheimer's disease (AD) to identify gene pathways with direct causal relationships to AD-related outcomes such as cognitive decline rates.
- Beyond a single-gene-based approach, pathway-based approaches have become important in drug development.

Methods

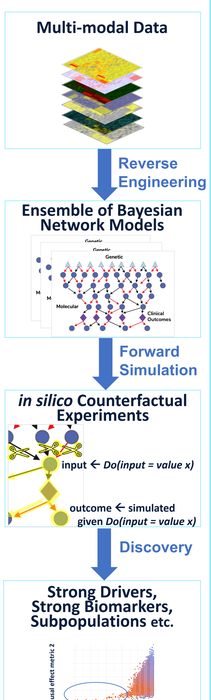
- Two independent and complementary AD Digital Twins were built using ADNI and ANMerge consortium data available at https://adni.loni.usc.edu and https://doi.org/10.7303/syn222528881, respectively.

Reverse Engineering

- Each *in silico* patient of the AD Digital Twins is comprised of an ensemble of Bayesian network models built from the training data using REFS™ causal AI platform [https://aitiabio.com].
- A Bayesian network model is a directed graphical representation of relationships between variables where each node is a variable, and each arrow is a conditional dependency.

Forward Simulation

- Patient-level outcome values can be estimated in the AD Digital Twins, by *in silico* counterfactual experiments which computationally simulate patient outcome values through model interventions, known in causal inference as 'Do' operations.
  - Briefly, in *in silico* counterfactual experiments, we 1) intervene the input variable with a specific value, 2) which won't be affected by its parent variable values, and 3) propagate its effect onto the downstream outcome variable through all the paths across the networks in an ensemble.
- 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.



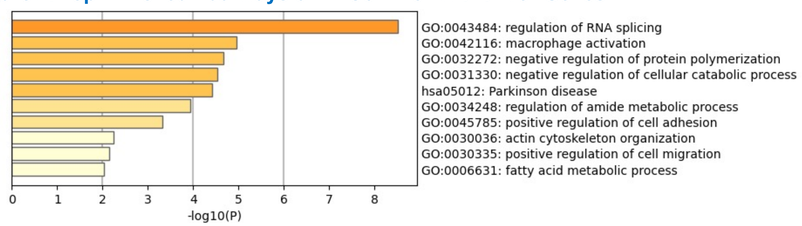
Results

- Using two AD Digital Twins, the average causal effect of each gene (out of 18k genes at the transcript expression level in blood) on each outcome was estimated through *in silico* counterfactual experiments.
- We identified a total of 573 and 300 AD-driver-genes from ADNI- and ANMerge-based Digital Twins, of which 41 overlapped (overlap p-value = 6e-15) **[Figure 1]**.

Figure 1. Study Workflow and Summary of Results

- These 41 robustly validated genes were related to several gene ontology (GO) biological processes and KEGG pathways by the gene set enrichment analysis [https://metascape.org] **[Figure 2]**.
- **AD Digital Twins recapitulate well-known biology directly from the data:** As expected, well-known processes related to AD were described by the models, including 'regulation of RNA splicing' [Koch L., 2018] and 'macrophage activation' [Costarelli *et al.,* 2017; Mammana *et al.,* 2018; Munawara *et al.* 2021]. For example, RNA binding and splicing proteins are shown to be linked to the biochemical phase of AD associated with aberrant amyloid precursor protein (APP) processing, Aβ accumulation, and tau hyperphosphorylation [Chen *et al.,* 2022]. Also, the overlap of the AD gene set with the gene set known for Parkinson's disease is consistent with previous studies [Morgan *et al.* 2022].
- **AD Digital Twins provide new evidence for less well-known biological mechanisms:** Other less well-known processes such as 'actin cytoskeleton organization' and many others were revealed through causal simulations.

Figure 2. Top Enriched Pathways of 41 Common AD-Driver-Genes



- Furthermore, the AD-driver-genes from ADNI-based Digital Twins representing a sample population with a higher proportion of MCI patients, were related to the processes such as 'cell activation, regulation of cell activation and positive regulation of response to external stimulus' associated with immune/inflammatory responses that occurs with aging that may be pathological during the prodromal stage of AD [Song *et al.,* 2022]; and 'ribosomal sub-complexes pathways', previously known to be involved specifically in early-stage AD [Ding *et al.,* 2005] **[Table 1]**.
- On the other hand, the genes selected from ANMerge-based digital twins, were enriched in the processes related to those identified from common genes (i.e. 'mRNA metabolic process' and 'ribose phosphate biosynthetic process' corresponding to 'regulation of RNA splicing'); as well as many others including 'oxidative stress' and 'mitochondrial respiration' **[Table 1]**.

Table 1. Top 10 Enriched Pathways of AD-driver-genes in Both Digital Twins

ADNI-based AD-Driver-Genes		ANMerge-based AD-Driver-Genes	
Geneset	-log10(P)	Geneset	-log10(P)
cell activation	-28.60	mRNA metabolic process	-12.64
regulation of cell activation	-18.59	ribose phosphate biosynthetic process	-10.73
positive regulation of response to external stimulus	-15.86	hemopoiesis	-10.66
Prion disease	-15.68	positive regulation of organelle organization	-9.78
Nop56p-associated pre-rRNA complex	-15.62	positive regulation of cytokine production	-8.14
cellular response to cytokine stimulus	-14.84	regulation of cytoskeleton organization	-7.81
regulation of cellular response to stress	-14.70	cellular macromolecule catabolic process	-7.77
cellular response to lipid	-14.36	platelet formation	-7.58
negative regulation of catabolic process	-13.55	response to antibiotic	-7.14
positive regulation of cell migration	-13.41	vacuole organization	-6.78

Conclusions

- We created two independent Digital Twins to **identify and validate causal driver genes** directly from data in a hypothesis-free way, and to **prioritize AD-related pathways** for drug discovery.
- Our findings **recapitulated some well-known AD pathways** (e.g. RNA splicing regulation, cell activation) and further **illuminated several less well-known pathways** (e.g. cytoskeleton regulation).
- Investigating causal genes from two complementary consortium studies with a different proportion of early- vs. late-stage AD patients may allow discovery of disease-stage-specific and -common signatures in the pathway-based drug discovery efforts.