Real-world data, real-world problems, and real-world opportunities for AD/ADRD

Jiang Bian, PhD

Chief Data Scientist & Chief Research Information Officer, University of Florida Health
Professor and Division Chief, Health Outcomes & Biomedical Informatics, College of Medicine
Director, Biomedical Informatics program, UF Clinical and Translational Science Institute
Director , Cancer Informatics Shared Resource (CISR), UF Health Cancer Center
Chief Data Scientist, OneFlorida+ Clinical Research Consortium

University of Florida
Gainesville 32610, Florida, USA
e-mail: bianjiang@ufl.edu

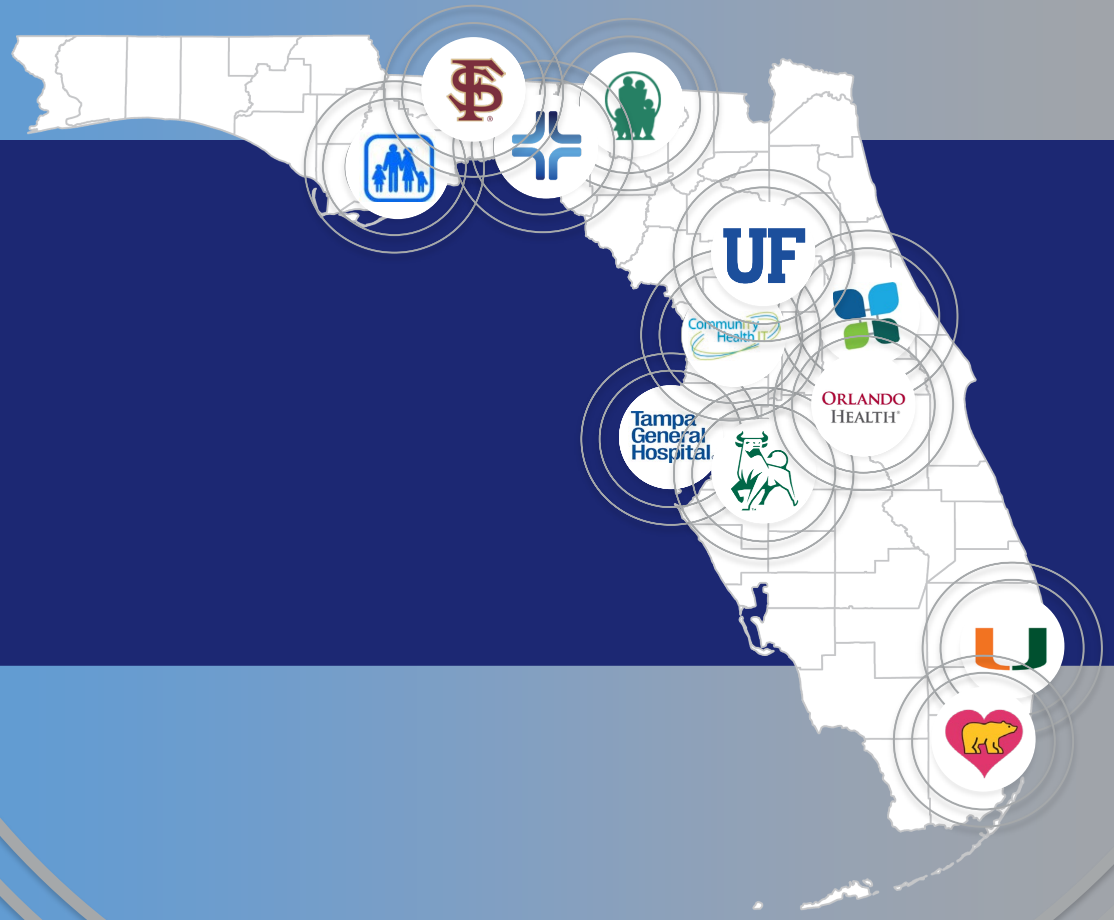
Disclosures

- **Sources of support**
  - US National Institutes of Health (NIH)
    - NCATS, NIA, NCI, NIDDK
  - US National Science Foundation (NSF)
  - Patient-Centered Outcomes Research Institute (PCORI)
  - Centers for Disease Control and Prevention (CDC)
  - Florida Department of Health (FDoH)
  - American Heart Association (AHA)
- **Current PI/MPI Grants**: NIA RF1AG077820, NIA R01 AG080624, NIA R01 AG080991; NCATS U01 TR003709; NIA R01 AG076234; CDC U18 DP006512, NCI R01 CA246418, PCORI ME-2018C3-14754, NIA R56 AG069880, NIA R21 AG068717, NCI R21 CA245858, NCI R21 CA253394, NIEHS R21 ES032762, NIA R21 AG061431, NSF SBE #1734134;

Agenda

- The OneFlorida+ Clinical Research Consortium
  - OneFlorida+ RWD
- Real-world data have real-world problems
- Select AD/ADRD studies using OneFlorida+ RWD

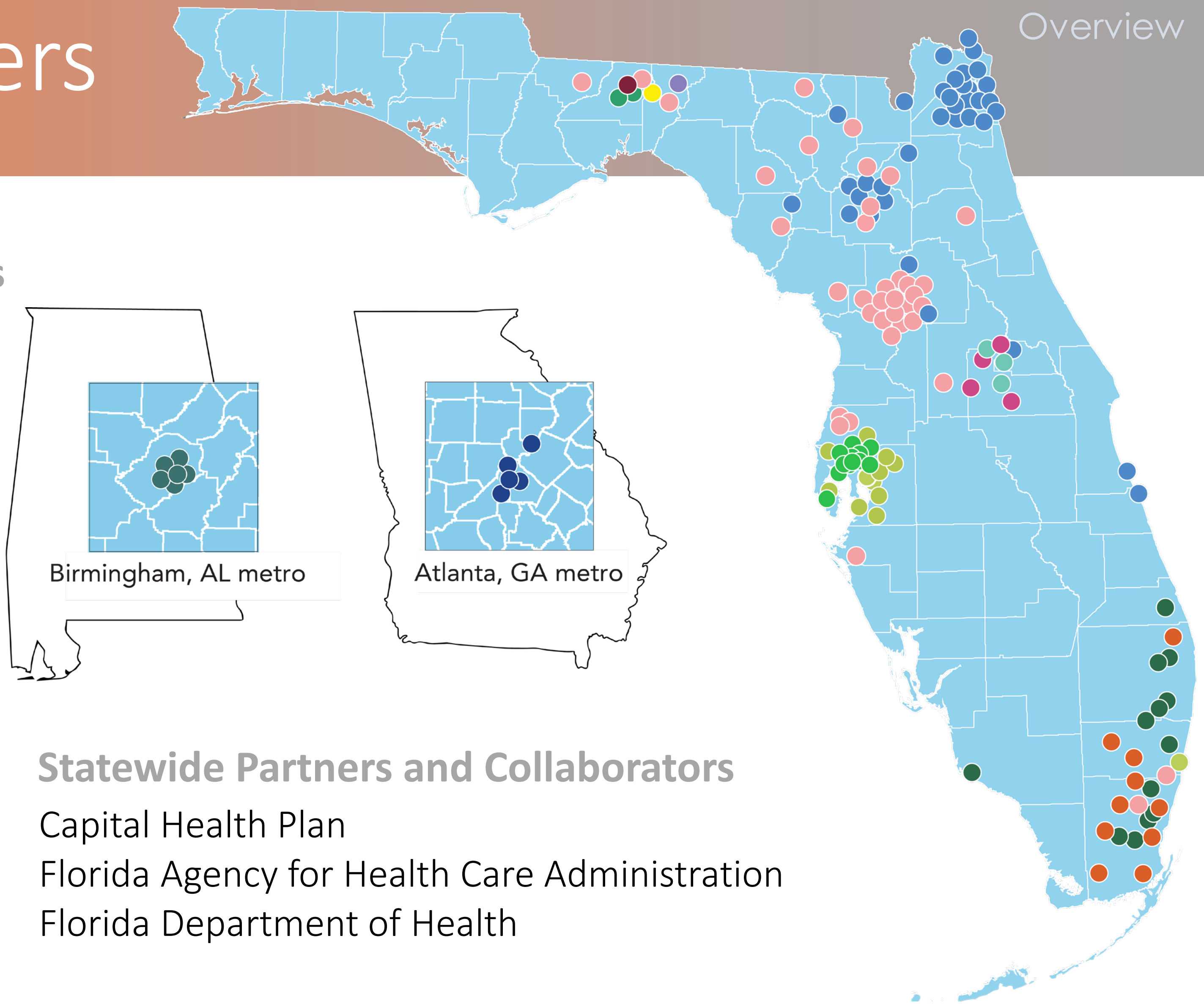
ONEFLORIDA+


UAB
UF
Community Health IT
ORLANDO HEALTH
Tampa General Hospital

OneFlorida+ Partners

Health Care Systems and Affiliated Practices

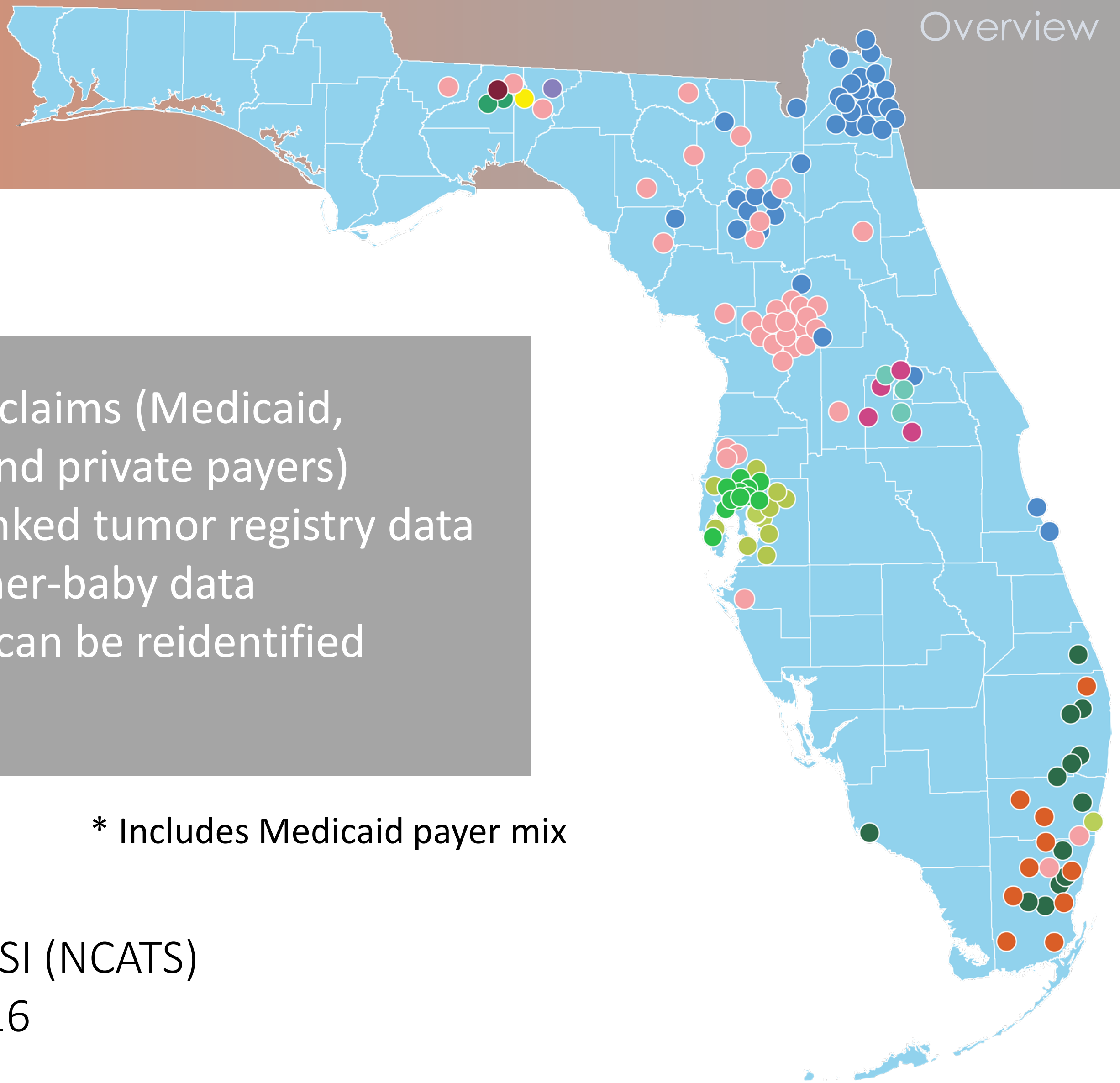
- University of Florida and UF Health
- Florida State University
- University of Miami and UHealth
- Orlando Health System
- AdventHealth
- Tallahassee Memorial HealthCare
- Tampa General Hospital
- Bond Community Health Center Inc.
- Nicklaus Children's Hospital
- CommunityHealth IT
- University of South Florida and USF Health
- University of Alabama at Birmingham
- Emory University



Statewide Partners and Collaborators

Capital Health Plan
Florida Agency for Health Care Administration
Florida Department of Health

OneFlorida+ Partners



Current Key Features:

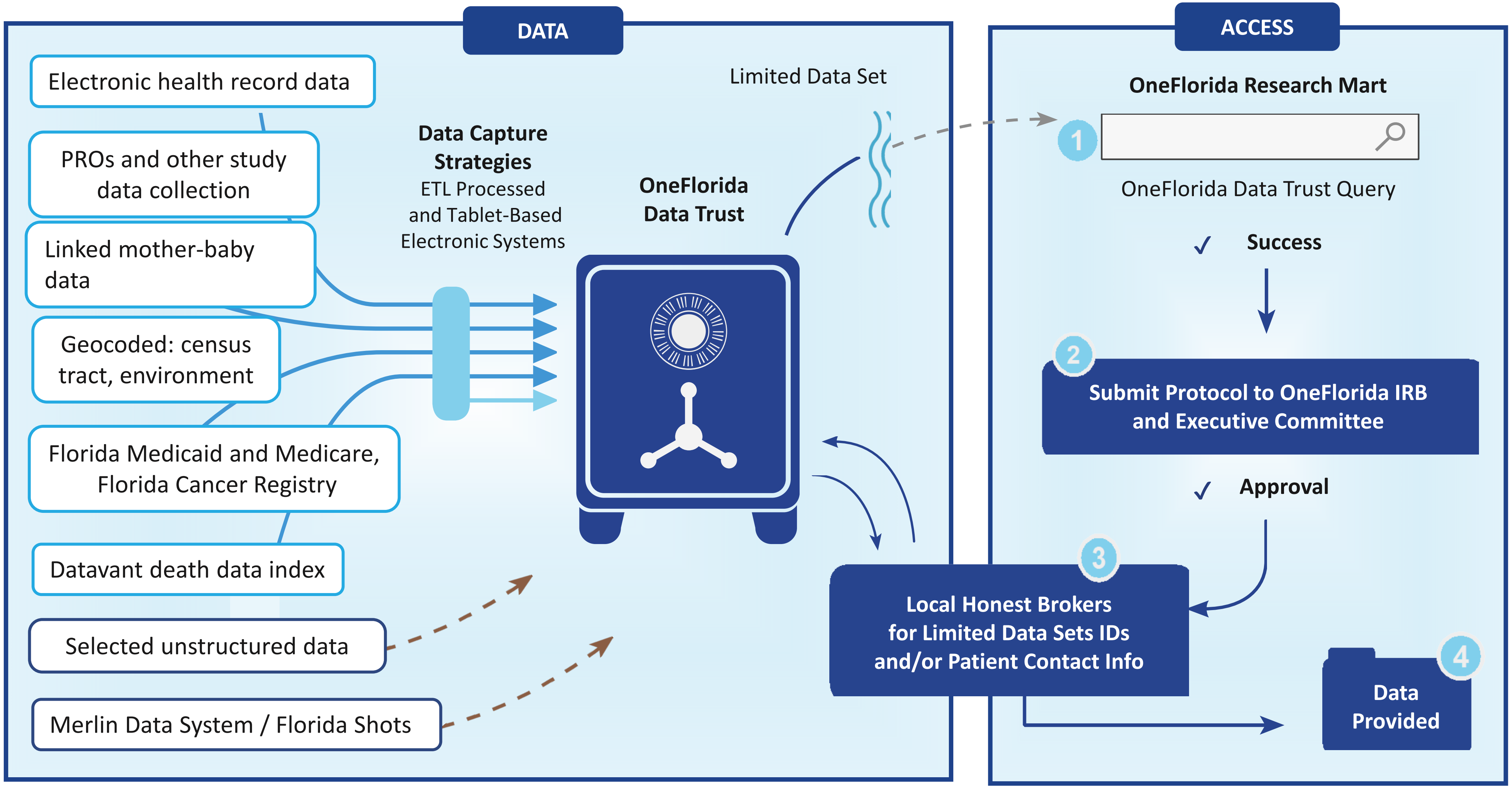
- 17 M Floridians*
- 2.1M Georgians and 1.1M Alabamans
- Centralized Data Trust
- Clinical data (electronic health records)
- Health care claims (Medicaid, Medicare, and private payers)
- Sites with linked tumor registry data
- Linked mother-baby data
- All patients can be reidentified

* Includes Medicaid payer mix

Infrastructure Funding:

UFHCC; PCORI CDRN-1501-26692 and 2020-005; CTSI (NCATS) UL1 TR001472 and UL1 TR000064; FL DOH JEK 4KB16

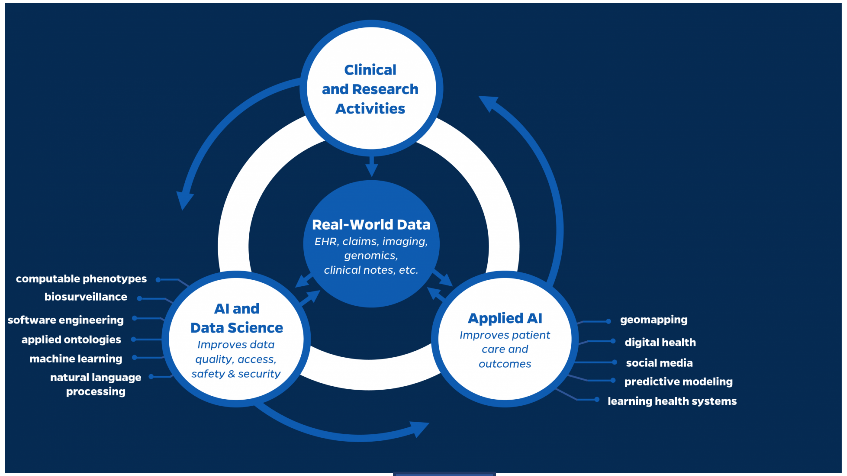
The OneFlorida+ Data Trust:
A Central Data Repository to Facilitate Research



Hogan WR, Shenkman EA, Robinson T, Carasquillo O, Robinson PS, Essner RZ, Bian J, Lipori G, Harle C, Magoc T, Manini L, Mendoza T, White S, Loiacono A, Hall J, Nelson D. The OneFlorida Data Trust: a centralized, translational research data infrastructure of statewide scope. ***J Am Med Inform Assoc***. 2022 Mar 15;29(4):686-693. doi: 10.1093/jamia/ocab221. PMID: 34664656; PMCID: PMC8922180.

Real-world data have real-world problems!

- RWD like EHRs are **M E S S Y ! e.g.,**
  - **Misclassification errors -> computable phenotypes (still has misclassification errors)**
    - Take Alzheimer's' disease as an example
      - Validated CP through manual chart reviews is only a "**silver standard**"
      - True "***gold standard***" of AD diagnosis needs biomarkers and imaging studies
  - **Missing information:** a multi-fold challenge!
    - "80% of clinical information is locked in free-text narratives"
      - **natural language processing**
    - A long list of ***other data types*** that are not readily accessible for research:
      - ***Imagining***, ***genomics***, ***microbiome***, etc.
        - ??? A better data infrastructure



It is not just about developing methods and tools, but also has a **real-world impact on improving clinical and community outcomes!**

"transition from infrastructure to impact"

Domains of Influence (Over the Lifecourse)		Levels of Influence*			
		Individual	Interpersonal	Community	Societal
	Biological	Biological Vulnerability and Mechanisms	Caregiver–Child Interaction Family Microbiome	Community Illness Exposure Herd Immunity	Sanitation Immunization Pathogen Exposure
	Behavioral	Health Behaviors Coping Strategies	Family Functioning School/Work Functioning	Community Functioning	Policies and Laws
	Physical/Built Environment	Personal Environment	Household Environment School/Work Environment	Community Environment Community Resources	Societal Structure
	Sociocultural Environment	Sociodemographics Limited English Cultural Identity Response to Discrimination	Social Networks Family/Peer Norms Interpersonal Discrimination	Community Norms Local Structural Discrimination	Social Norms Societal Structural Discrimination
	Health Care System	Insurance Coverage Health Literacy Treatment Preferences	Patient–Clinician Relationship Medical Decision-Making	Availability of Services Safety Net Services	Quality of Care Health Care Policies
Health Outcomes		Individual Health	Family/ Organizational Health	Community Health	Population Health


Computational Drug Repurposing

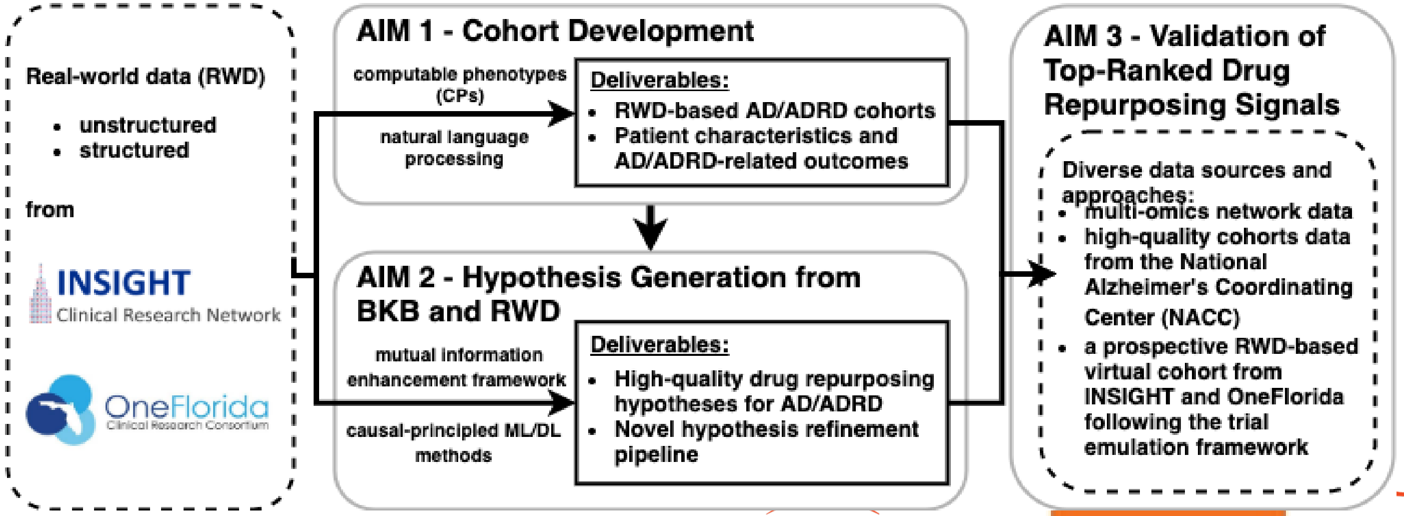
Jiang Bian, University of Florida/OneFlorida; Fei Wang, Weill Cornell Medicine/INSIGHT

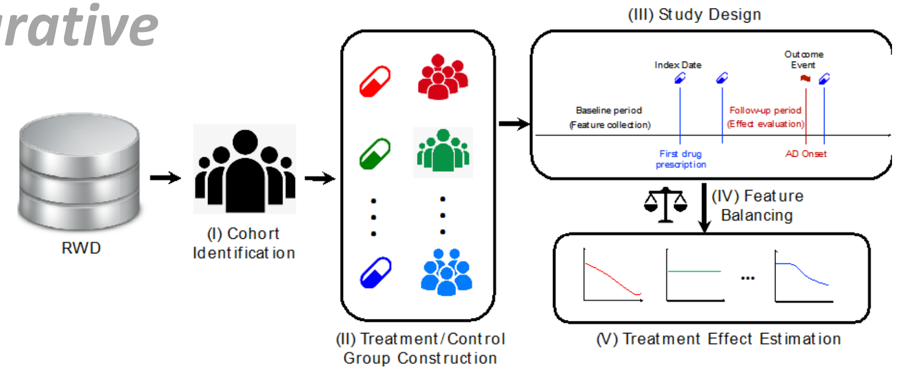
R01AG076234 (2022 – 2027)

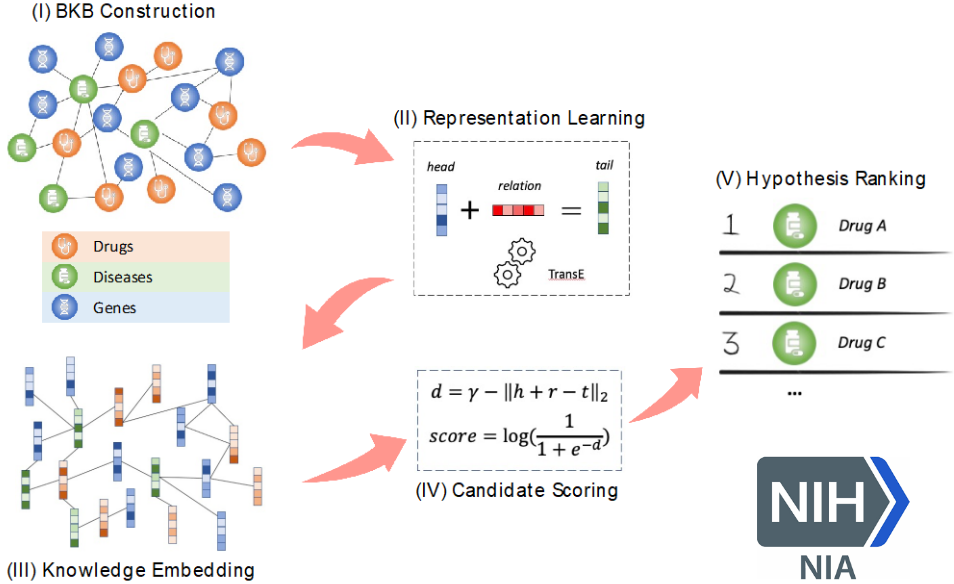
OneFlorida: Novel Methods

Computational Drug Repurposing for AD/ADRD with Integrative Analysis of Real-World Data and Biomedical Knowledge

- Combining machine learning (ML) and causal inference to estimate treatment effects from RWD
- Integrative learning from both biomedical knowledge base (BKB) and RWD to generate high-quality drug repurposing hypotheses







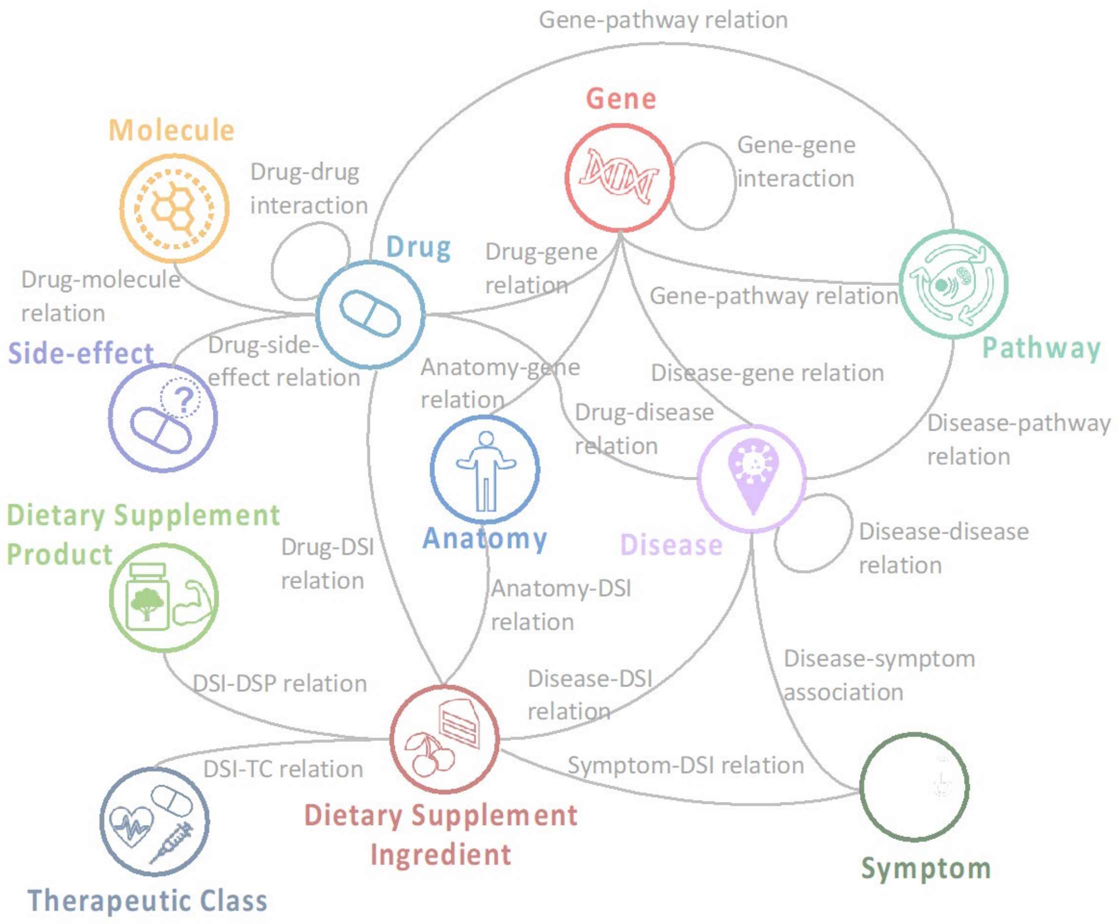
NIH NIA

UFHealth UNIVERSITY OF FLORIDA HEALTH

UF

UFHealth CANCER CENTER

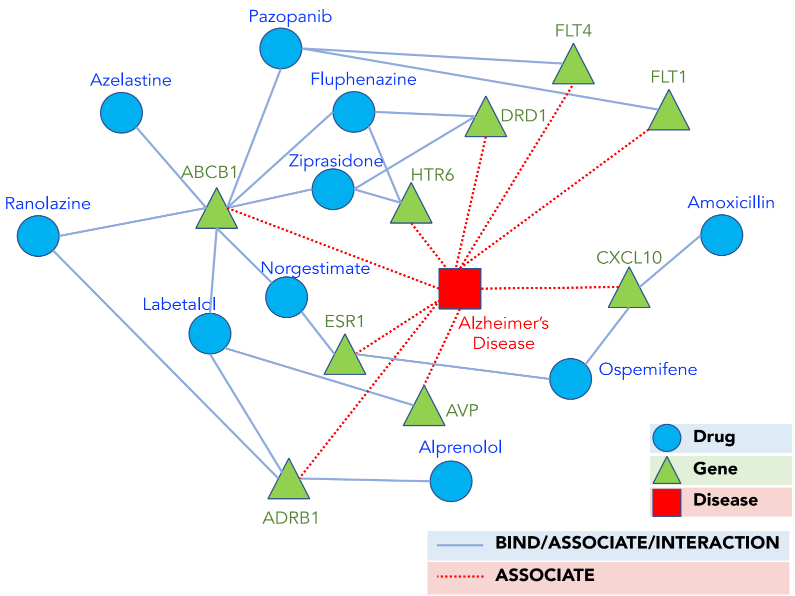
integrative Biomedical Knowledge Hub (iBKH)



https://github.com/wcm-wanglab/iBKH

Summary statistics of current iBKH.		
Existing BKBs	Entities	Relations
Bgee	60,072	11,731,369
BRENDA Tissue	6,478	-
Cell Ontology	2,200	-
ChEMBL	1,940,733	-
ChEBI	155,342	-
CTD	73,922	38,344,568
Disease Ontology	10,648	-
DrugBank	15,128	28,014
Repurposing KG	97,238	5,874,261
Hetionet	47,031	2,250,197
HUGO Gene	41,439	-
iDISK	144,536	705,075
KEGG	33,756,186	43,464
PharmGKB	43,112	61,616
Reactome	13,589	13,732
TISSUE	26,260	6,788,697
Uberon	14,944	-
Total	36,448,948	65,840,993

Repurposing for Alzheimer's Disease



Procedure 1: RCT Emulation

Input: Patient data: assigned treatment, outcome, censoring time, observations of hypothesized confounders

Output: Estimated effect (and P-value) of treatment on the outcome

1: Identify major confounders from the list of hypothesized confounders by their association with the outcome

2: **repeat**

3: Compute balancing weights for the treatment and control cohorts

4: Set the state *is_balanced* to **true** if no major imbalance exists between the reweighted treatment and control cohorts; Otherwise set it to **false**

5: **if not** *is_balanced* **then**

6: Increase positivity likelihood by patient exclusion

7: **until** *is_balanced* **or** number of patients in the treatment or control cohorts is too small

8: **if not** *is_balanced*

9: **return** "Cannot evaluate treatment effect"

10: **else**

11: Estimate the treatment effect (using causal inference method)

12: Estimate the P-value of the computed effect (using bootstrapping)

13: **return** the estimated effect and P-value

P(outcome | high feature values) – P(outcome | low feature values)
For event-based outcomes, we computed this difference with Kaplan-Meier estimators and used ~~bootstrapping~~ to assess its statistical significance

Inverse probability weighting

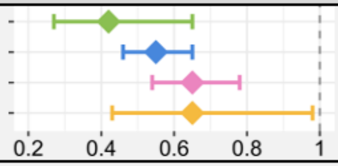
$$d = \frac{|\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}|}{\sqrt{(s^2_{\text{treatment}} + s^2_{\text{control}})/2}}$$

excluded patients whose propensity scores lay outside the overlap of the treatment and control cohorts

$$P^a(\text{outcome}) = \prod_{t=1}^{\text{follow-up length}} \left(1 - \frac{\sum_{\{i:\, A_i=a\}} w_{t,x_i} 1[T_i^{\text{outcome}} = t]}{\sum_{\{i:\, A_i=a\}} w_{t,x_i} 1[T_i \geq t]}\right)$$

$$P^a(\text{outcome}) = \frac{1}{n}\sum_i P[\text{outcome}|X = x_i, A_i = a]$$

Drug	Survival analysis result: Hazard ratio with 95% CI	P	5-year conversion rate (%): Treatment	Control	5-year conversion time (days): Treatment	Control
Ziprasidone	0.42 (0.27-0.65)	<0.005	5.60%	7.79%	683.34	600.25
Ibuprofen	0.55 (0.46-0.65)	<0.005	5.94%	7.98%	641.75	563.03
Amoxicillin	0.65 (0.54-0.78)	<0.005	5.22%	5.86%	612.37	559.63
Azelastine	0.65 (0.43-0.98)	0.04	4.21%	8.31%	718.23	626.66



Q & A ?