Federated Analyses to Accelerate Precision Medicine

Precision Medicine
Federated Learning

Lucila Ohno-Machado, MD, PhD

Biomedical Informatics & Data Science

Yale School of Medicine

10/09/23 – MIRACUM-DIFUTURE MII Symposium

Section of Biomedical Informatics & Data Science (BIDS, est. Jan 1, 2023)

The hub for biomedical collaboration:

Innovate new approaches to the analysis of big data across the biomedical research spectrum from basic genetic, proteomic, cellular, and systems biology to medicine and the understanding of the social determinants of health

Bring informatics to the clinic and the bedside

Work in concert with colleagues in data science

Three pillars

Research: informatics and clinical research collaboration

Education: MS, PhD, Postdoc training programs, Certificate Program

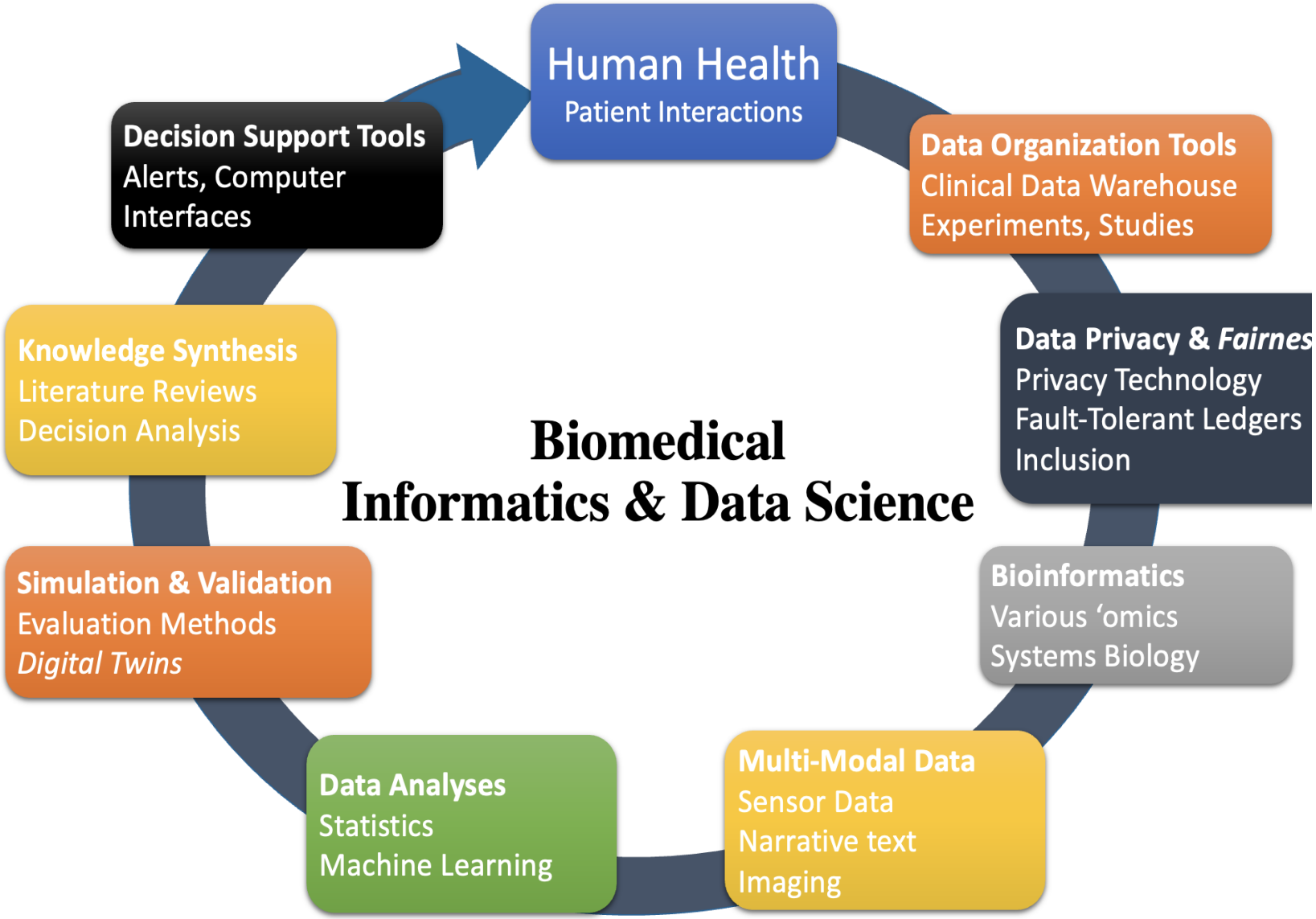
Service: Research Information Office, collaborations with the library, etc.

BIDS Team

Lucila Ohno-Machado, MD, MBA, PhD

Chair of BIDS

- Ladder faculty - 11
- Research faculty – 12
- Affiliated faculty – 31
- Trainees – 10
- Staff - 14




covid19questions.org
Covid-19 Clinical Data Consult
Home
Questions & Answers
Answers in Progress
Ask Questions
About Us
COVID-19 Data Discovery from Clinical Records
Press here for an important note

Covid-19 Clinical Data Consult

In 2020

- 928,255 patients tested
- 59,074 diagnosed with COVID-19
- 19,022 hospitalized
- 2,591 deceased

- Data from >45M patients for comparisons

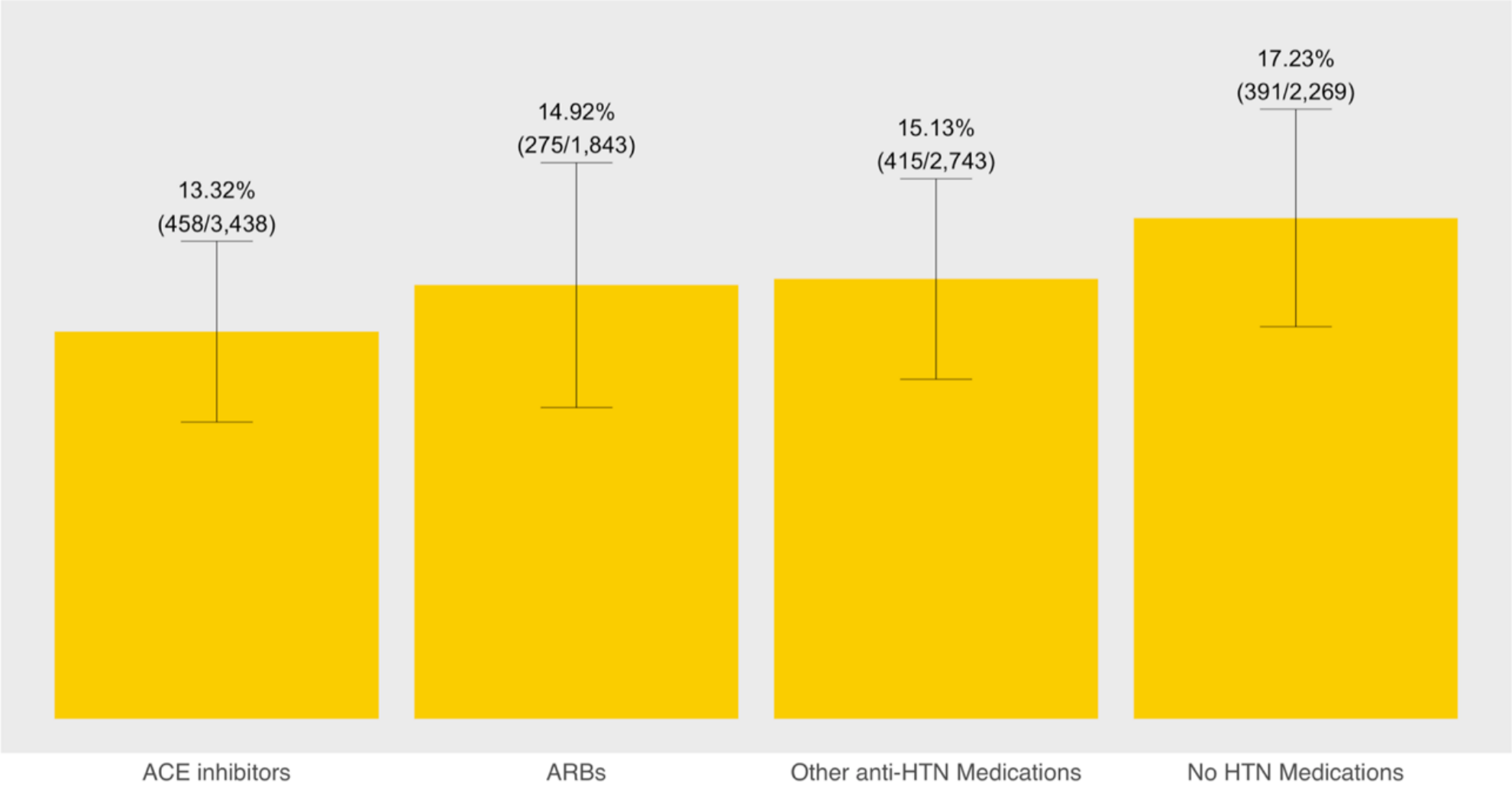
Moore Foundation award

In-hospital mortality of hypertensive patients per medication use group*

* Patients with medication use within one year prior to hospital encounter

Numbers in parentheses = in-hospital deaths / total patients per group

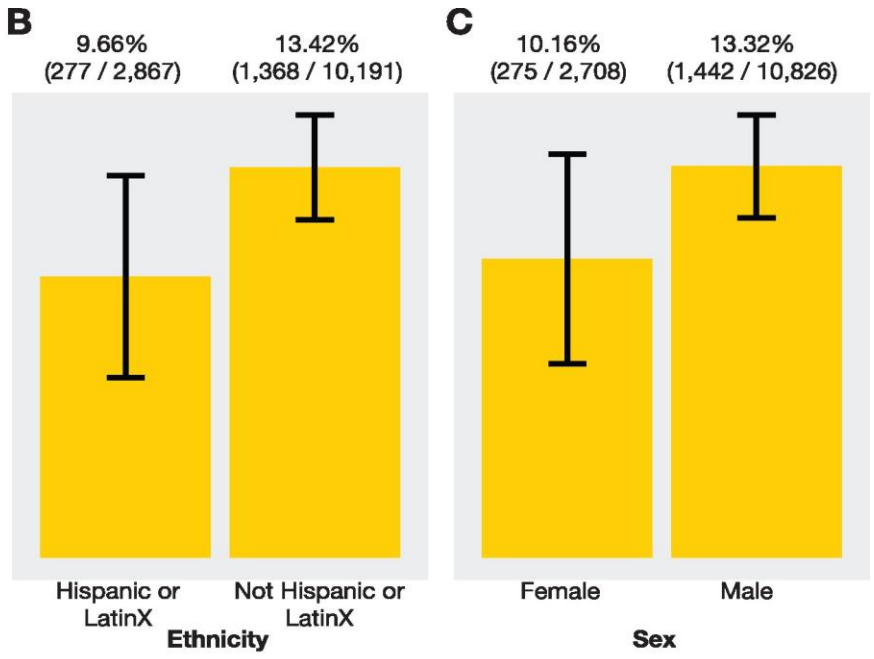
ACE = Angiotensin-converting enzyme, ARBs = Angiotensin II receptor blockers, HTN = Hypertension

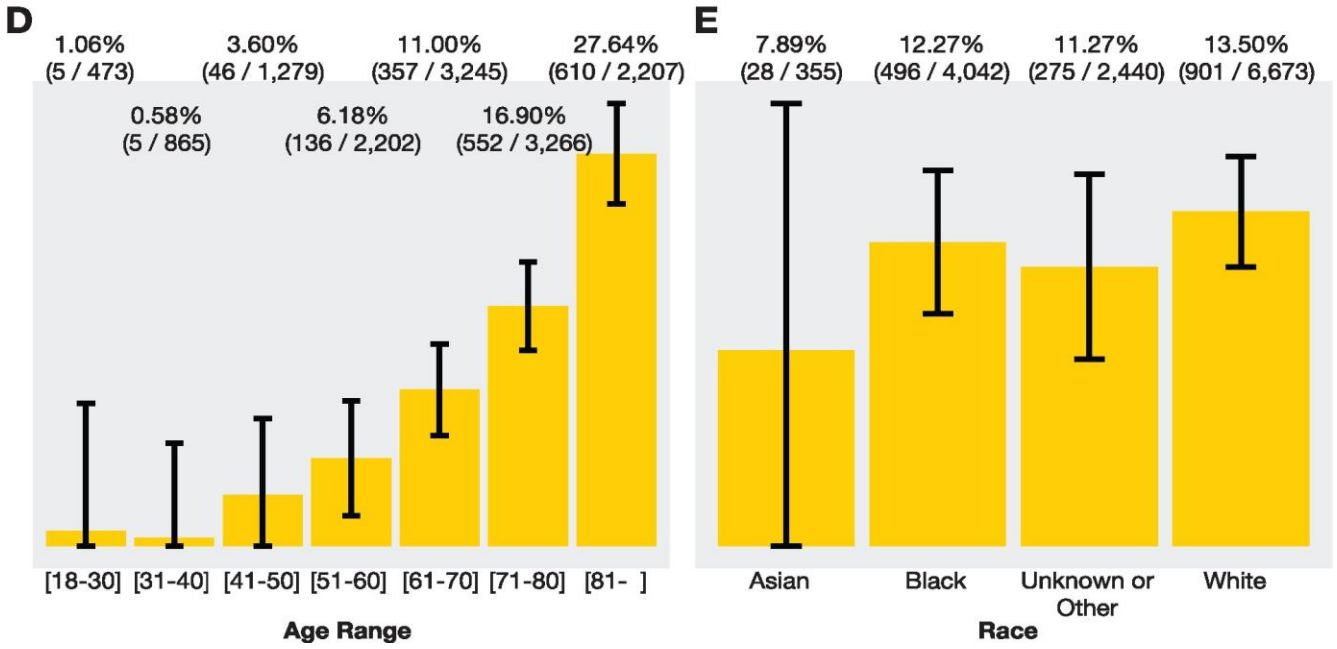


Source: 10,293 adult patients from 10 institutions
Data retrieved October 15, 2020 - November 03, 2020

Univariate Mortality

- 928,255 tested for SARS-CoV-2
- 59,074 diagnosed with COVID-19
- 19,022 hospitalized
- 2,591 deceased





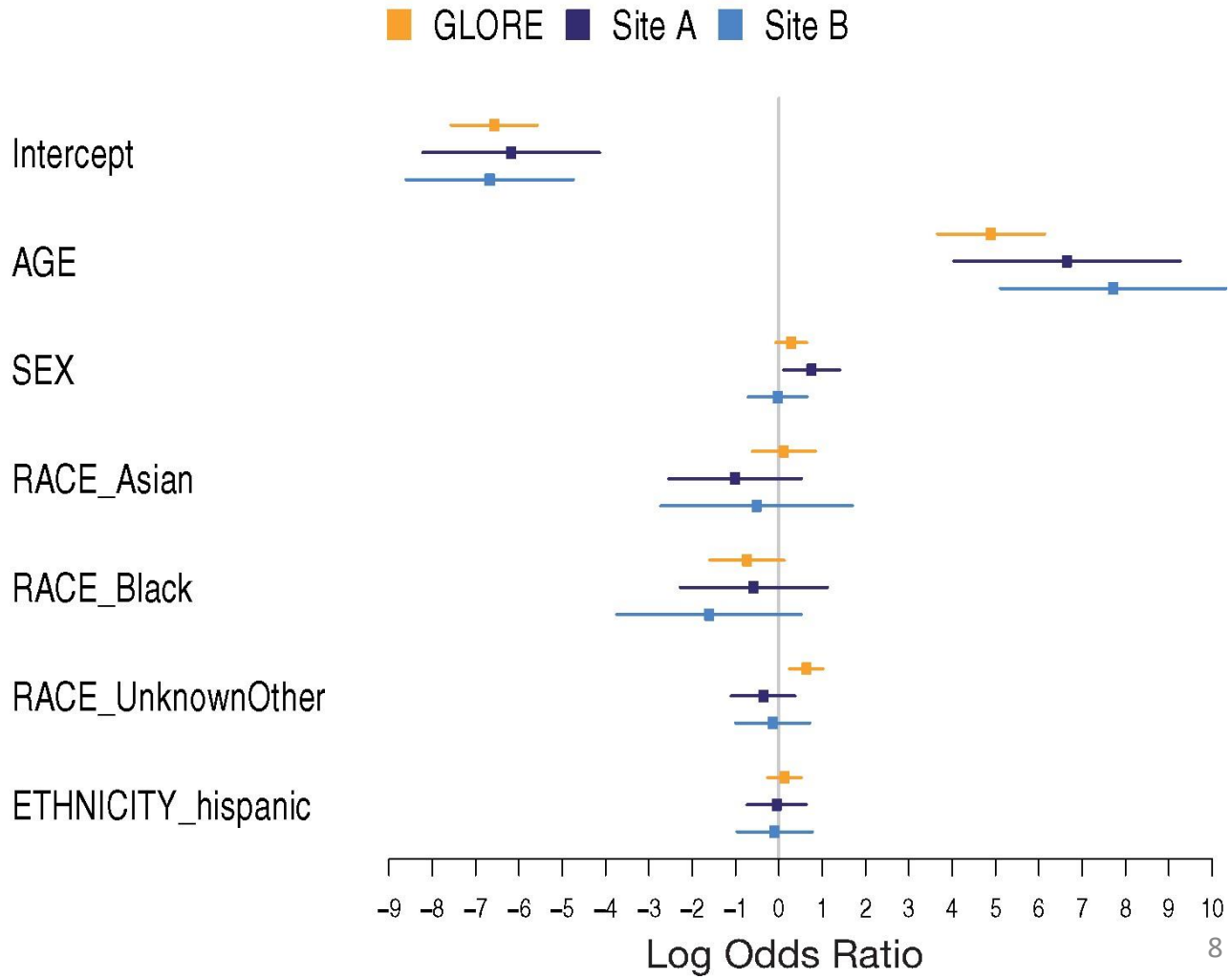
Multivariate Analyses

Kim et al. Privacy-protecting, reliable response data discovery using COVID-19 patient observations. J Am Med Inform Assoc. 2021 Jul 30;28(8):1765-1776.

A

Variable	Coefficient	Standard Error	Z–statistic	P–value	Lower 95% CI	Upper 95% CI
Intercept	–6.572	0.508	–12.942	0.000	–7.567	–5.576
AGE	4.898	0.632	7.744	0.000	3.659	6.138
SEX_Male	0.290	0.182	1.591	0.112	–0.067	0.647
RACE_Asian	0.118	0.373	0.316	0.752	–0.613	0.849
RACE_Black	–0.739	0.437	–1.689	0.091	–1.596	0.119
RACE_UnknownOther	0.633	0.196	3.228	0.001	0.249	1.017
ETHNICITY_hispanic	0.130	0.198	0.654	0.513	–0.259	0.518

B



CAST

Genomics for Everyone

Center for Admixture Science and Technology

An NHGRI-funded Center of Excellence in Genome Science

Disease risk is influenced by many factors

Trait
e.g. LDL, cancer diagnosis

Genetic determinants

Variant effects
SNPs, HLA, TRs,…

Ancestry
SNPs, PCs,…

Social determinants
income, education, diet,…

$$E[y_i] \sim X_i\beta + Q_i\gamma + C_i\delta$$

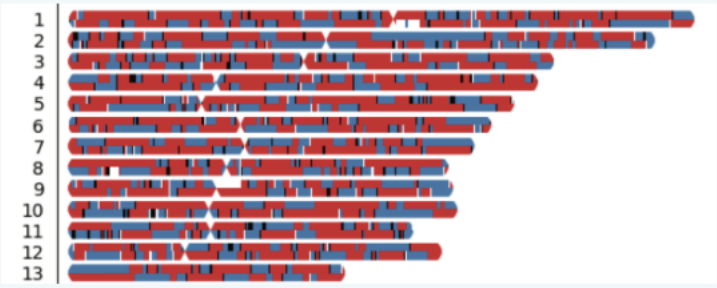
Algorithms and tools to analyze data that can stay in their enclaves

Adapted from a slide by Melissa Gymrek

Polygenic Risk Scores (PRS)

- Can we improve the methods?
- Can the scores be continuously evaluated, and models updated as needed?
- Can we do all this without introducing more biases that lead to more discrimination? Can we protect privacy?
- What to do with individuals from mixed ancestries?

ANCESTRY-AWARE POLYGENIC RISK SCORES



We are developing models that consider each individual's unique patchwork of ancestry to accurately predict disease risk.

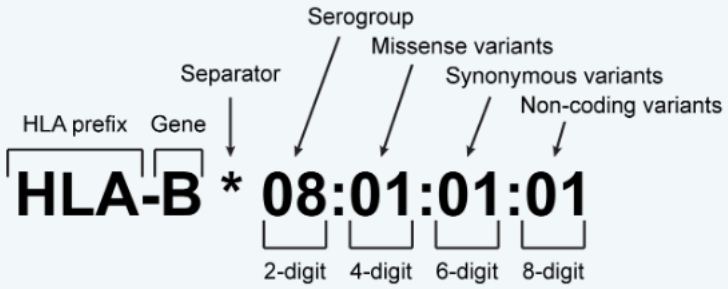
TANDEM REPEATS

```
CGATCGAGCAG---------ACTACAACTAGG
CGATCGAGCAGCAG------ACTACATCTACG

CGAACGAGCAGCAGCAG---ACTACAACTAGG
CGATCGAGCAGCAG------ACTACATCTACG
```

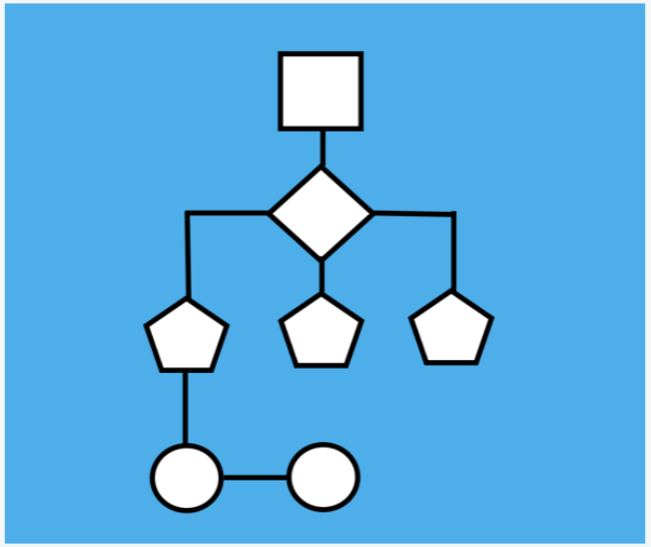
CAST is leveraging methods to analyze highly polymorphic tandem repeats in diverse populations, and understand their role in medically relevant complex traits.

HLA ANALYSIS



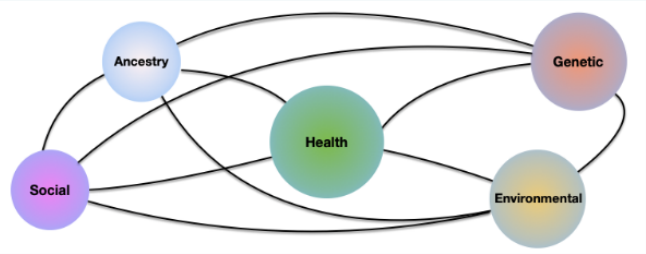
We are developing methods to perform HLA typing in diverse populations, and characterize their associations with complex traits and disease.

DISTRIBUTED ANALYTICS



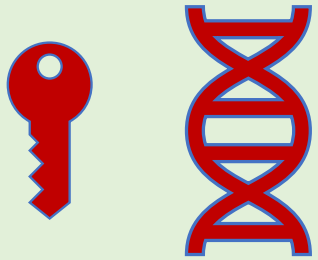
We are developing algorithms that allows us to compute with data that are not located at a central database, but instead are distributed across different databases.

SOCIOECONOMIC DETERMINANTS OF HEALTH



We are developing methods that incorporate factors like genetics, ancestry, socioeconomic factors, and environmental factors for health evaluation in diverse populations.

GENOME PRIVACY



We are studying best ways to protect the privacy of individuals while allowing authorized researchers to compute with sensitive data.

Distributed Analysis for Precision Medicine

All of Us (AOU) Research Program

409k participants (Sept '23)
Health and genomic data (250k WGS)

Secure federated algorithms*

Million Veteran Program (MVP)

>950k participants (Sept '23)
Health and genomic data

Access limited to AoU analysis platform

Access limited to VA analysis platform

*Other algorithms funded by NIH R01GM118609

Our Vision

- No one will be left out
- Replace “race” and ”ethnicity” with genetic, environmental, and social determinants of health
- Develop new methods and tools that allow genetic findings to be applicable to all

Working together, we can enable a LHS where everyone has equal opportunity

- Data science research & applications, linking across disciplines, delivering in practical settings
- Promoting excellence in transdisciplinary training
- Enabling biomedical researchers
- Will make Precision Health a reality <u>for all</u>

Thank you

Lucila.Ohno-Machado@yale.edu

NIH R01HG011066
iAGREE: A Multi-Center, Networked Patient Consent Study

NIH R01GM118609
Decentralized differentially-private methods for dynamic data release and analysis

NIH R01HL136835
Protecting Privacy and Facilitating Shared Access of Clinical and Genetic Data of Special Populations

NIH U24LM013755
RADx-rad Data Coordinating Center

NIH OT2OD026552
California All of Us

Moore Foundation
Rapid Response Data Discovery

NIH RM1HG011558
Center for Admixture Science & Technology

NIH U54HG012510
Bridge2AI Center

NIH T15LM007056
Biomedical Informatics Research Training at Yale