Cohort Overview

Mia M. Gaudet, Ph.D.

Connect Chief Scientist

Trans Division Research Program
Division of Cancer Epidemiology and Genetics
National Cancer Institute

August 2, 2021

Study Objectives

Comprehensive research resource using new technologies and methods for the scientific community to study:

Cancer Etiology
   Precursor to Tumor Progression
Cancer Risk Prediction
Early Detection of Cancer
Survivorship

connect
for cancer prevention study

Connect Designed to Address Key Research Priorities

Emerging exposures

Novel biomarkers, genomics

Cutting-edge methodology
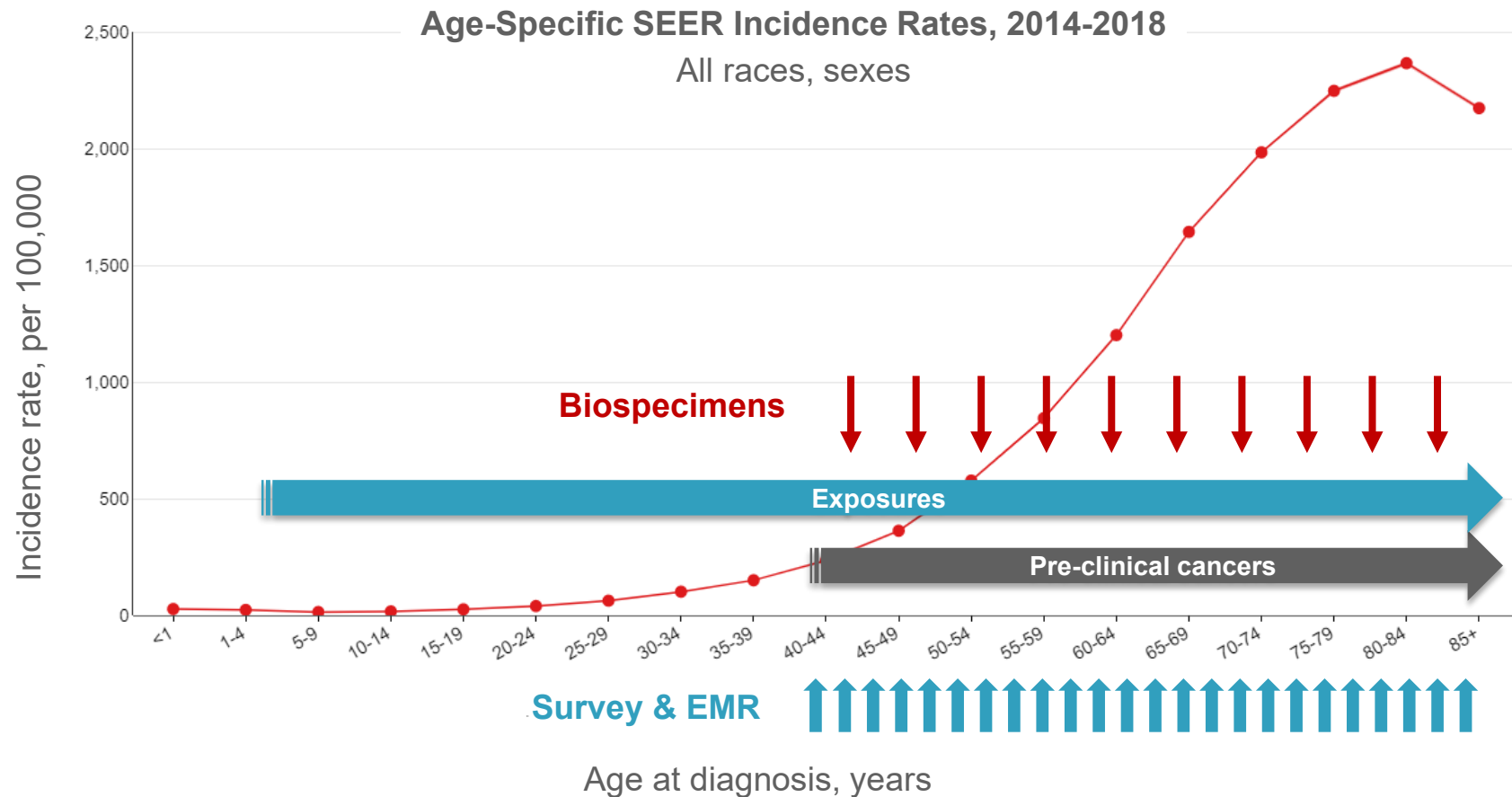
Diverse and special populations

Cohort Study Design Overview

Key Features of Connect as a Trans-NCI Resource
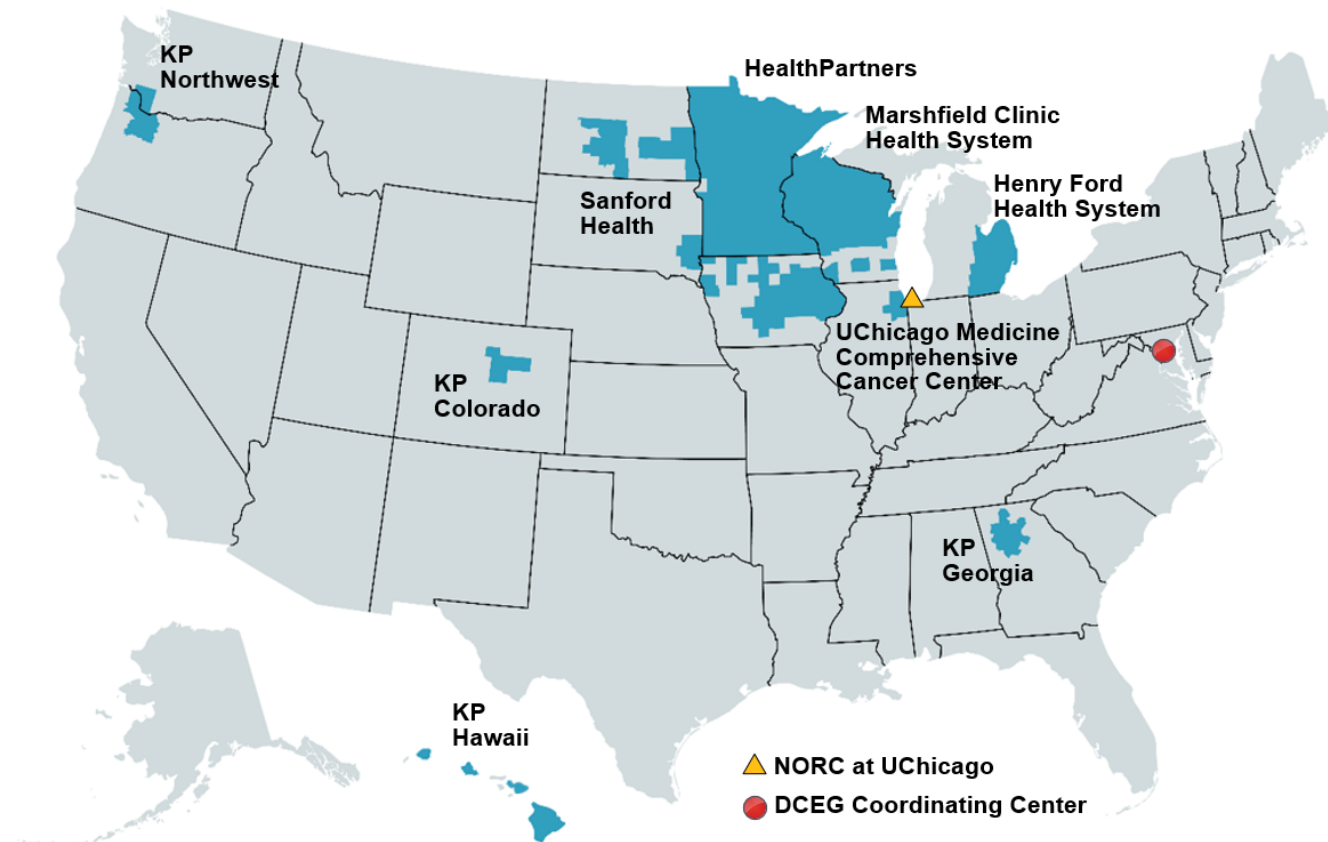
▶ 200,000 adults across the US

- ✓ Aged 40-65 years
- ✓ No history of cancer
- ✓ Patients or members of partner health care systems

▶ Long-term follow-up

▶ Serial assessments

▶ Comprehensive cancer outcomes

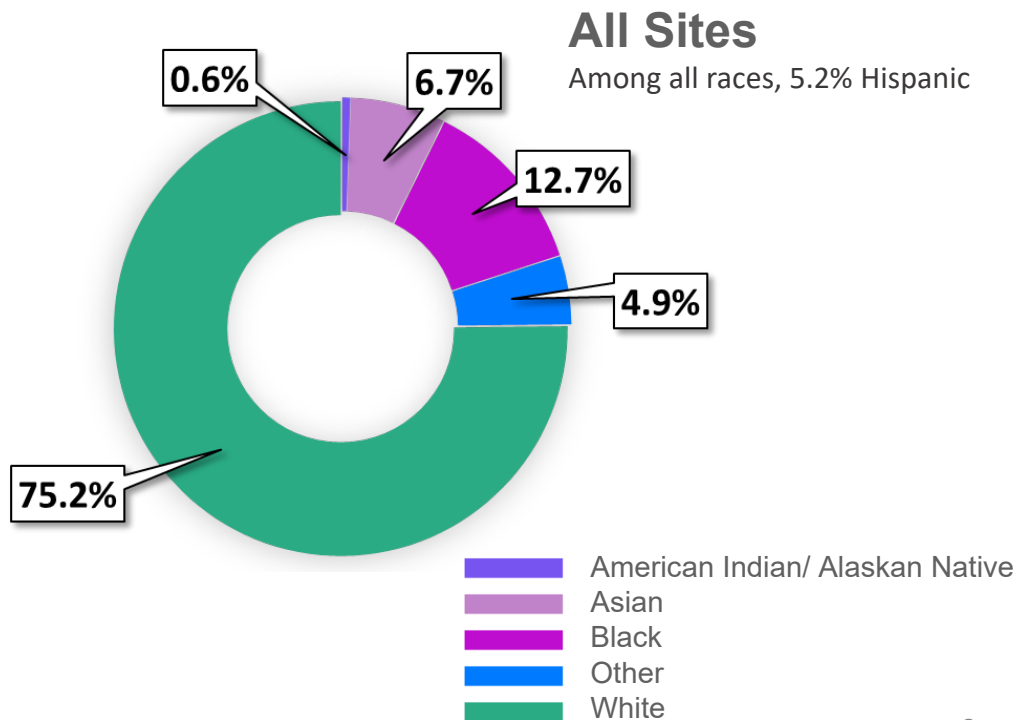▶ Flexible infrastructure for enhancement studies

connect
for cancer prevention study
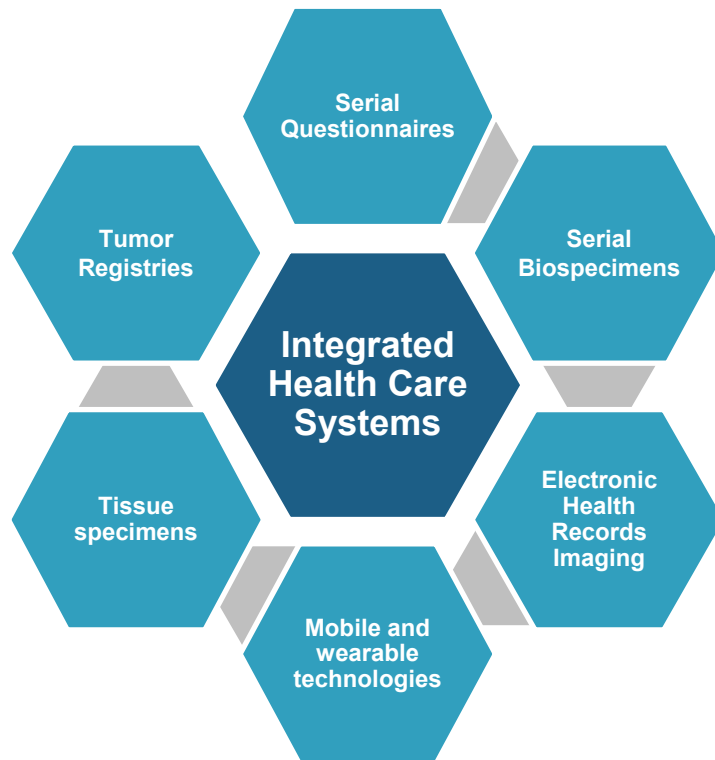
Serial Assessments to Observe Longitudinal Changes

Age-Specific SEER Incidence Rates, 2014-2018
All races, sexes

Recruitment at Partner Health Care Systems

Diversity of Catchment Population: Race and Ethnicity

Kaiser Hawaii

Health Partners (Minnesota)

Kaiser Georgia

Marshfield (Wisconsin)

Henry Ford (Detroit)

Sanford (Dakotas)

University of Chicago

Kaiser Colorado

Kaiser Northwest (Portland, OR)

All Sites
Among all races, 5.2% Hispanic

0.6%

6.7%

12.7%

4.9%

75.2%

- American Indian/ Alaskan Native
- Asian
- Black
- Other
- White

connect
for cancer prevention study

8
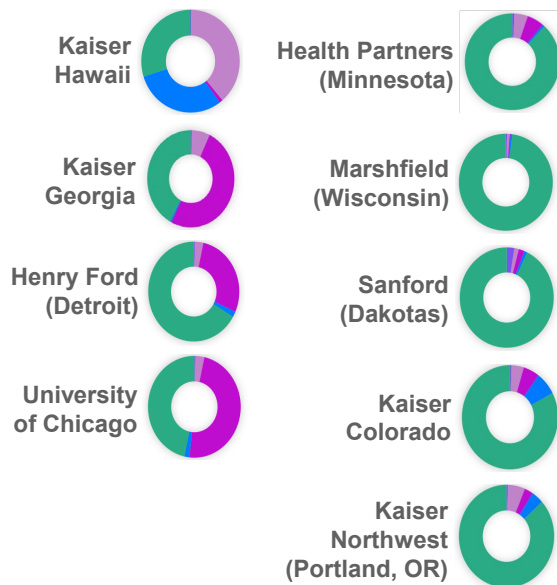
Participant Engagement

Recruitment Materials & Strategy

- Value proposition to anchor communications

- Prevention-focused messaging

- Materials designed to engender trust, enthusiasm, and inclusivity

- Incentives

- Active and passive recruitment

Connect today. Prevent cancer tomorrow.

connect
for cancer prevention study

Retention Strategies

- Candid messaging about participant burden for the long-term

- Responsive communication using multiple channels

- Incentives for retention, such as return of results

connect
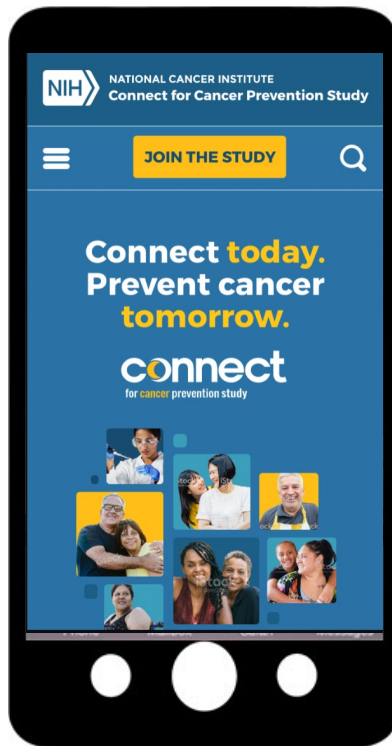for cancer prevention study

Ongoing Engagement Strategies

- Digital engagement

- Regular contact every 6 months beyond those for study activities

- "Study anniversary" mailing

- Monitor real-time metrics to pivot engagement strategy

connect
for cancer prevention study

MyConnect Participant App: Communication and Study Activity Hub

Reminders & Communication

Healthcare Provider App

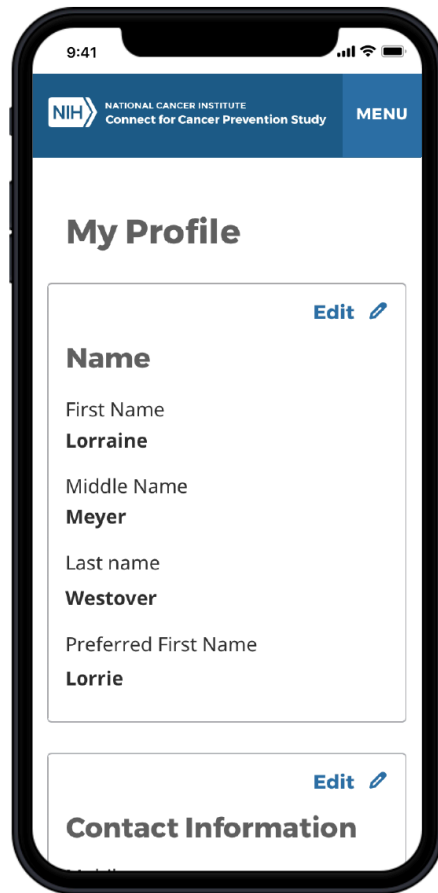Health App

Mobile/ Wearable Technologies

MyConnect Participant App

Messages and Data Input

Messages

New 2   Read 1

Complete Your Profile
6 days ago

Please complete your profile section so you can move on to surveys!

Start

Sign Your Consent Forms
8 days ago

If you would like to participate in this study, please sign the consent forms.

Start

9:41

NIH NATIONAL CANCER INSTITUTE
Connect for Cancer Prevention Study

MENU

My Profile

Edit ✎

Name

First Name
Lorraine

Middle Name
Meyer

Last name
Westover

Preferred First Name
Lorrie

Edit ✎

Contact Information

Educational Module before Long Form Consent

Overview of Study Activities and Data Collection

Study Activities

Timing	Type of Study Activity	
	Participant actively completing…	**Occurring in the background…**
Recruitment	Surveys	EHR data transfer
	Specimen donation	Benign breast biopsies, colon adenomas, other tissues
Follow-up	Surveys	EHR data transfer
	Specimen donation	Clinical archived/ discard specimens
	Physical measurements	Linkages
	Wearable devices, downloaded apps	Pathology and radiology imaging

Biospecimen Collections

Biospecimen Collections at Recruitment
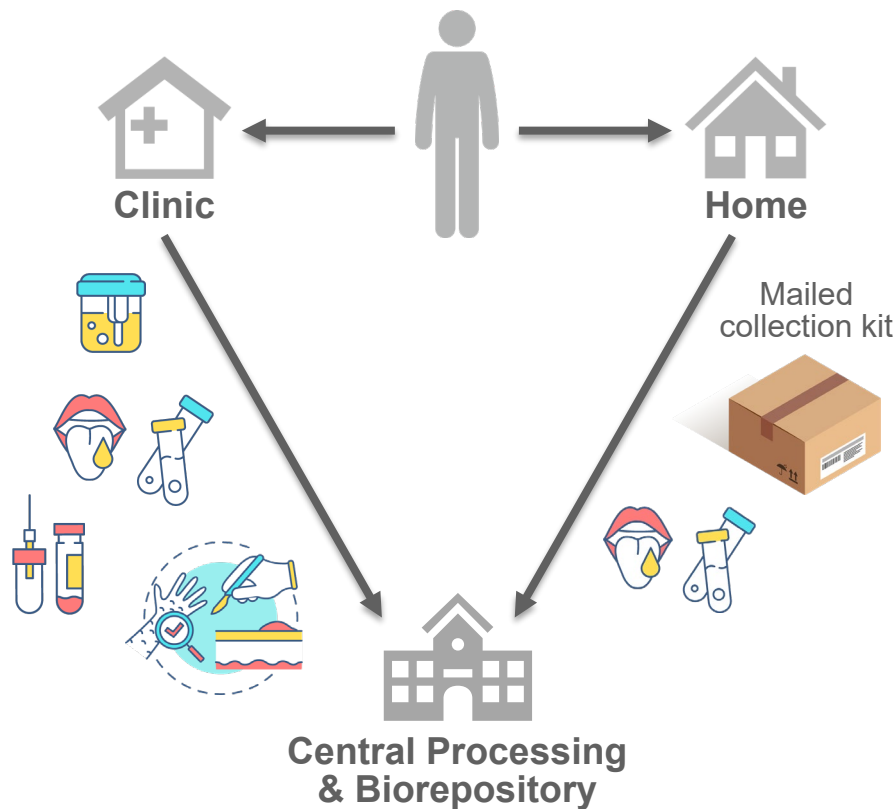
	No. Tubes	Fraction	Tube Type
Buccal	1		Saliva-mouthwash kit
Blood	2	Serum	10ml Red top
	1	Plasma/ buffy coat	10ml Lavender top (EDTA)
	1	Plasma/ buffy coat	10ml Green top (Heparin)
	1	Whole blood	6ml ACD
Urine	1		10ml Urine kit
Tissue			Precursor tissue blocks, if available

connect for cancer prevention study

Biospecimen Collections at Follow-up

- Timed on average two to three years

  - More frequent for participants at higher risk of cancer

- Series of same biospecimens and new types of biospecimens

- Located at healthcare system, external clinic, home

- Paired with questionnaires and other relevant data collection

connect
for cancer prevention study

Surveys

Baseline Questionnaire Module Topics & Display in MyConnect

Background and Overall Health
- Background information
- Medical History
- Family History of Cancer
- General Health
- Education and Occupation

Medications, Reproductive Health, Exercise, and Sleep
- Medications
- Women's Reproductive Health
- Physical Activity
- Sleep

Smoking, Alcohol, and Sun Exposure
- Tobacco
- Marijuana
- Alcohol
- Sun Exposure

Where You Live and Work
- Residential History
- Commuting

NIH | NATIONAL CANCER INSTITUTE
Connect for Cancer Prevention Study

Sign Out

Surveys   Samples   Agreements   About the Study   What to Expect   Privacy   My Profile   Messages

1 — Introduction
2 — Background Information
3 — Medical History
4 — General Health
5 — Family History
6 — Education and Occupation

How often do you have children in the household from birth to 5 years of age in your household?

- ⦿ Every day, at least 2 hours each day
- ○ Every day, less than 2 hours each day
- ○ Most days of the week
- ○ One Day per week
- ○ A few days per year
- ○ Never

Clear

connect
for cancer prevention study

Biospecimen Questionnaires

Blood	Urine	Mouthwash
• Covid-19 exposure and vaccination • Menstrual cycle dates (last, next) • Recent exposures to food, drink, pharmaceuticals	• Menstrual cycle dates (last, next) • Recent exposures to food, drink, pharmaceuticals	• Oral health and care • Recent exposures to food, drink, pharmaceuticals • Collection parameters, if home collection

connect
for cancer prevention study

Planned Follow-up Surveys

6 Months	Within 24 Months	12 Months	2 Years
Diet food frequency questionnaire (DHQ)	Repeated 24-hour recalls for diet (ASA24) & physical activity (ACT24)	Screening Quality of life Psychosocial	Update on baseline exposures, additional questions of covered and other topics

Survey at cancer diagnosis

Data Linkages

Data Linkages

Cancer registries, virtual tumor registry

Mortality (NDI)

Geospatial data (such as outdoor air quality data, other EPA MyEnvironment, WATERS data, US Census)

External health care records (such as Medicare, other EHR for participants that leave IHCS)

Health outcomes data (such as HIV registry)

connect
for cancer prevention study
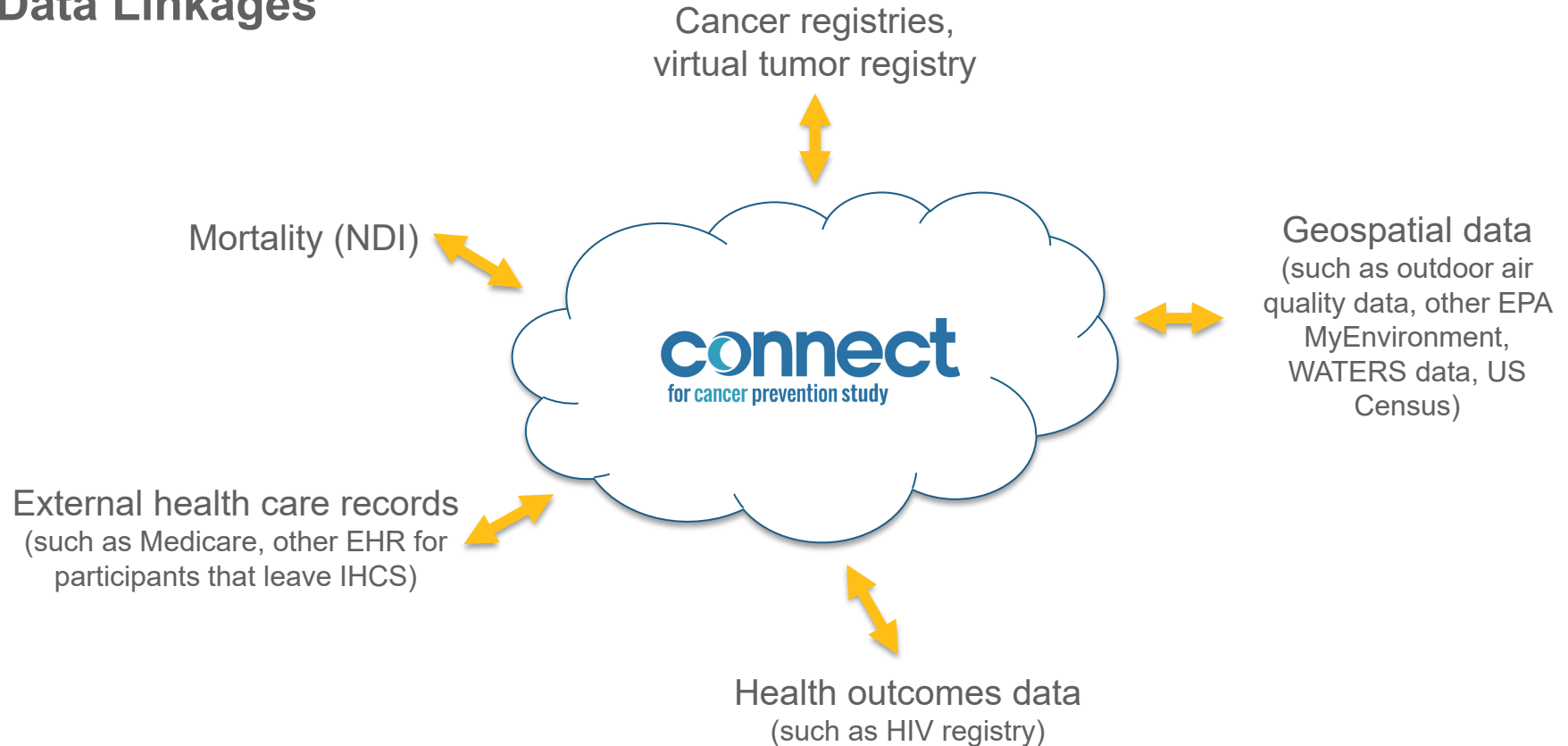
Geographic Information: Future Activities

- Geocode 20-year address histories (collected on baseline survey)

- Enable linkages to datasets for numerous external risk factor and exposure information

- Useful for planning ancillary field studies

- Add daily movement and travel to exposure information via MyConnect

Fore et al., 2020
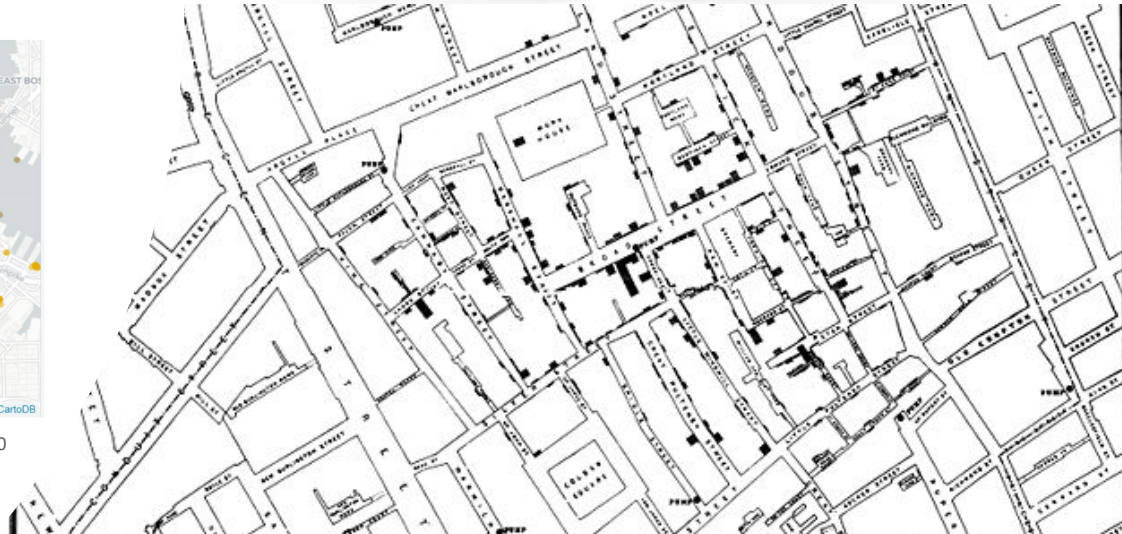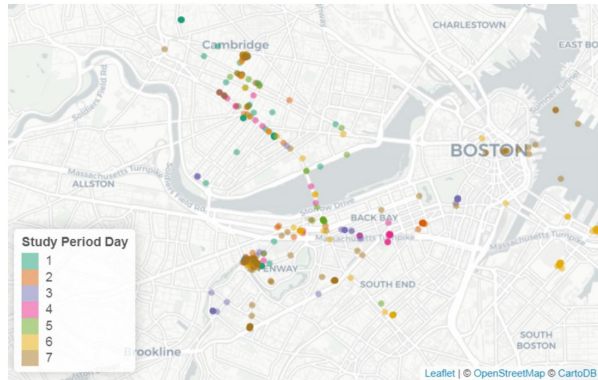
Wearable Devices and Apps

- Connect will have research-grade and consumer-oriented devices available

- Key domains where tech can improve assessment of exposures and risk factors:

    - sleep
    - light at night
    - air pollution
    - physical activity

Estimated Outcomes

Estimated total number of incident invasive cancers

Cancer site	Total Number of Expected Incident Cancers	
	10 years of follow up	15 years of follow up
Prostate (males only)	1,951	3,595
Breast (females only)	1,566	2,553
Lung & Bronchus	1,151	2,165
Colorectal	877	1,566
Melanoma of the Skin	410	681
Non-Hodgkin Lymphoma	369	649
Uterine Corpus	345	579
Urinary Bladder	327	622
Thyroid	232	356
Pancreas	221	410
Leukemia	203	365
Ovary	146	246

Estimated Total Number of Precursor Lesions

Lesions	Total Number of Expected Precursor Lesions	
	Retrospective at Baseline	Prospective at 10 Years of Follow Up
Colorectal Adenomas	5,778	18,311
Benign Breast Biopsies (females only)	2,339	4,942
Dysplastic Nevus	2,311	3,779
Cervical Biopsies (females only)	2,137	2,822
High-Grade Prostatic Intraepithelial Neoplasia (males only)	238	384
Lung Nodules	165	349

Timeline

Launching July 2021

Timeline

Retention

Recruitment

Evaluation based on metric analysis

Initial phase of recruitment

| 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 |

- 12-month duration
- Goal to recruit 10% of study population
- Evaluation of 12-month study metrics
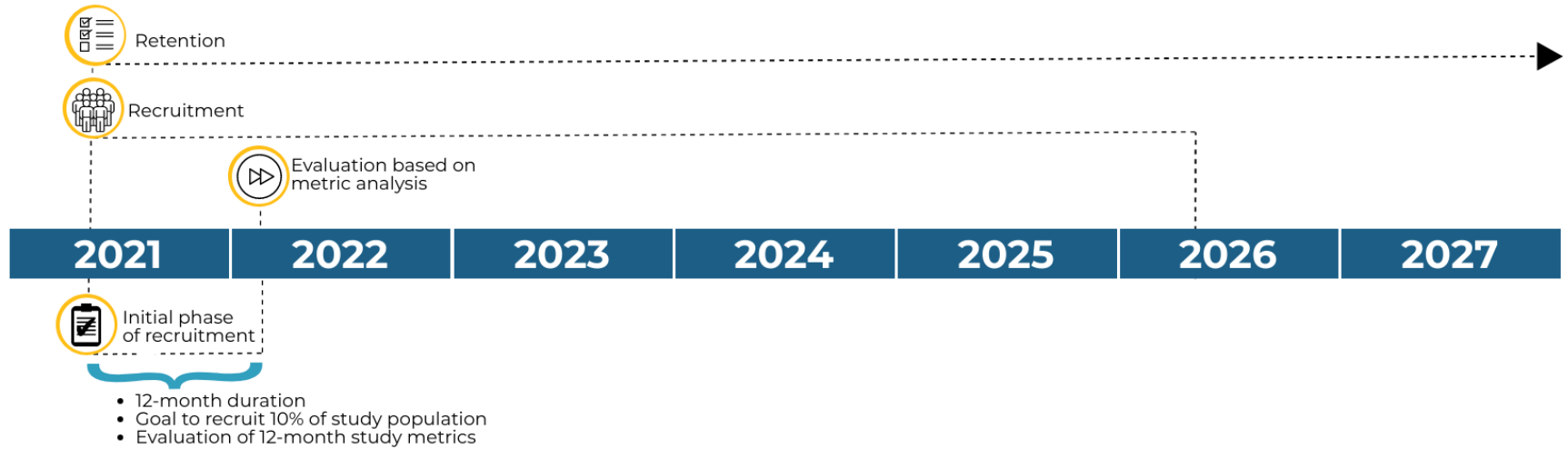
Resource Access Principles

Connect Resource Access Principles

Research resource for scientific community

Broad data sharing policies

Participant privacy and confidentiality

F.A.I.R. data infrastructure

Resource Access for All*

Serial Questionnaires

Tumor Registries

Serial Biospecimens

Connect Resource & Infrastructure

Tissue specimens

Electronic Health Records Imaging

Mobile and wearable technologies

Intramural NCI Investigators

International Investigators

Industrial Partners

Extramural NCI/ NIH Grantees

Other NIH Intramural Investigators

Integrated Health Care System Partners

connect
for cancer prevention study

*Governance procedures forthcoming

Levels of Data Access

Public Access
Univariate distributions and aggregated counts

Controlled Access
Individual-level survey, electronic health record, and derived data

Sensitive Variables
Variables with group or individual stigma or risk for reidentification

PHI Linkages
No shared access to 18 HIPAA identifiers; in-house linkages available to all investigators

connect
for cancer prevention study

Data System Infrastructure & Digital Assets

Data Ecosystem Provides Flexible Infrastructure

Front-End

User Interface
Participant Apps, Dashboards for Study Management/Tracking

Middle Layer

API Ecosystem (via Web)

Back-End

Cloud-based Data Storage

https://episphere.github.io/
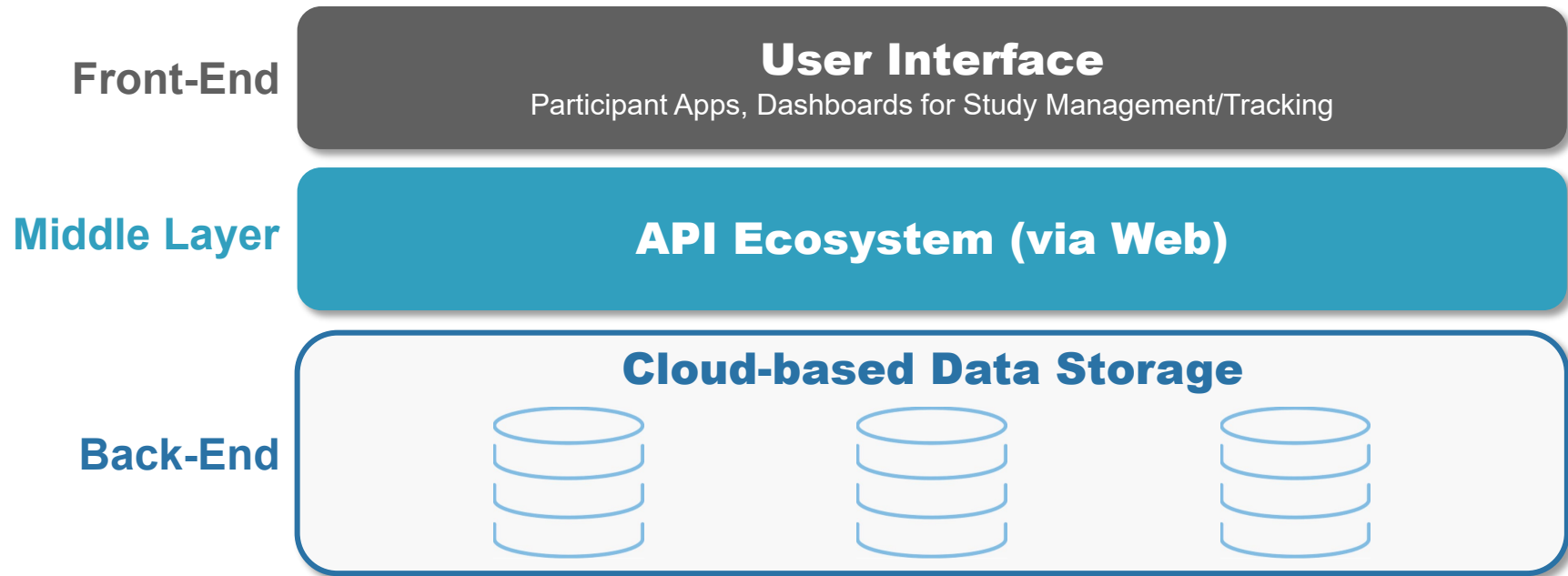
Data Moving at the API Level

Baseline Recruitment

- Deidentified data on eligible members/patients
- Verification of consented individuals in MyConnect
- Study activity flags

Data Linkages

- Health outcomes: cancer registries, virtual tumor registry, NDI mortality, HIV registry
- Geospatial data: outdoor air quality data, other EPA data, US Census

connect
for cancer prevention study

Study Services

- Communications, Twilio
- Incentives

Participant-Provided Access

- Wearable devices: FitBit, air monitors, etc.
- External health care records: Medicare, other EHR for participants that leave partner healthcare systems, etc.

Front-End User Interfaces

Integration with other data ecosystems (e.g., wearable devices) or services (e.g., Twilio)

Apps

Dashboards

Analytics

Authentication
Authorization

Middle Layer "API Phone Lines"

API

NCI/DCEG

Back-End Data Storage

Use of Cloud functions

NIH STRIDES

Cloud Storage

https://datascience.nih.gov/strides

42

Governance & Organizational Structure

Connect Governance

Connect Leadership and Coordinating Center

External Advisory and Review Groups

Governing Board
Stephen Chanock (Chair)

Executive Leadership

Steering Committee
Chief Scientist (Chair), DCEG Connect PIs, Coordinating Center Leadership, DCEG Associate Directors

Executive Committee
Chief Scientist (co-Chair), Rotating Site PI (co-Chair), Steering Committee Members, Site PIs, Participant Advisory Board Lead, Support Service Contractor Leads

International Scientific Advisory Board
Naomi Allen, Eric Boerwinkle, Julie Buring, Chris Haiman, David Hunter, Roger Milne, Alpa Patel

National Cancer Institute (NCI) Board of Scientific Counselors (BSC) – Clinical Sciences and Epidemiology

Participant Advisory Board (PAB)

Connect Executive Committee

DCEG Senior Investigators

Montse Garcia-Closas, PI of Record

Mia Gaudet, Chief Scientist

Paul Albert

Amy Berrington

Nico Wentzensen

Christian Abnet

Jonas Almeida, Chief Data Scientist

DCEG Staff

Mandy Black, Contract Rep

Hannah Yang, Contract Rep

Michelle Brotzman, Operations Lead

Stephanie Weinstein, Resource Access Lead

MJ Horner, Scientific Development Lead

Site PIs

Ben Rybicki, Henry Ford

Pam Pawloski, Health Partners

Robert Greenlee, Marshfield

Heather Feigelson, KP Colorado

Stacey Honda, KP Hawaii

Katrina Goddard, KP NW

Jenn Gander, KP Georgia

Habib Ahsan, U Chicago

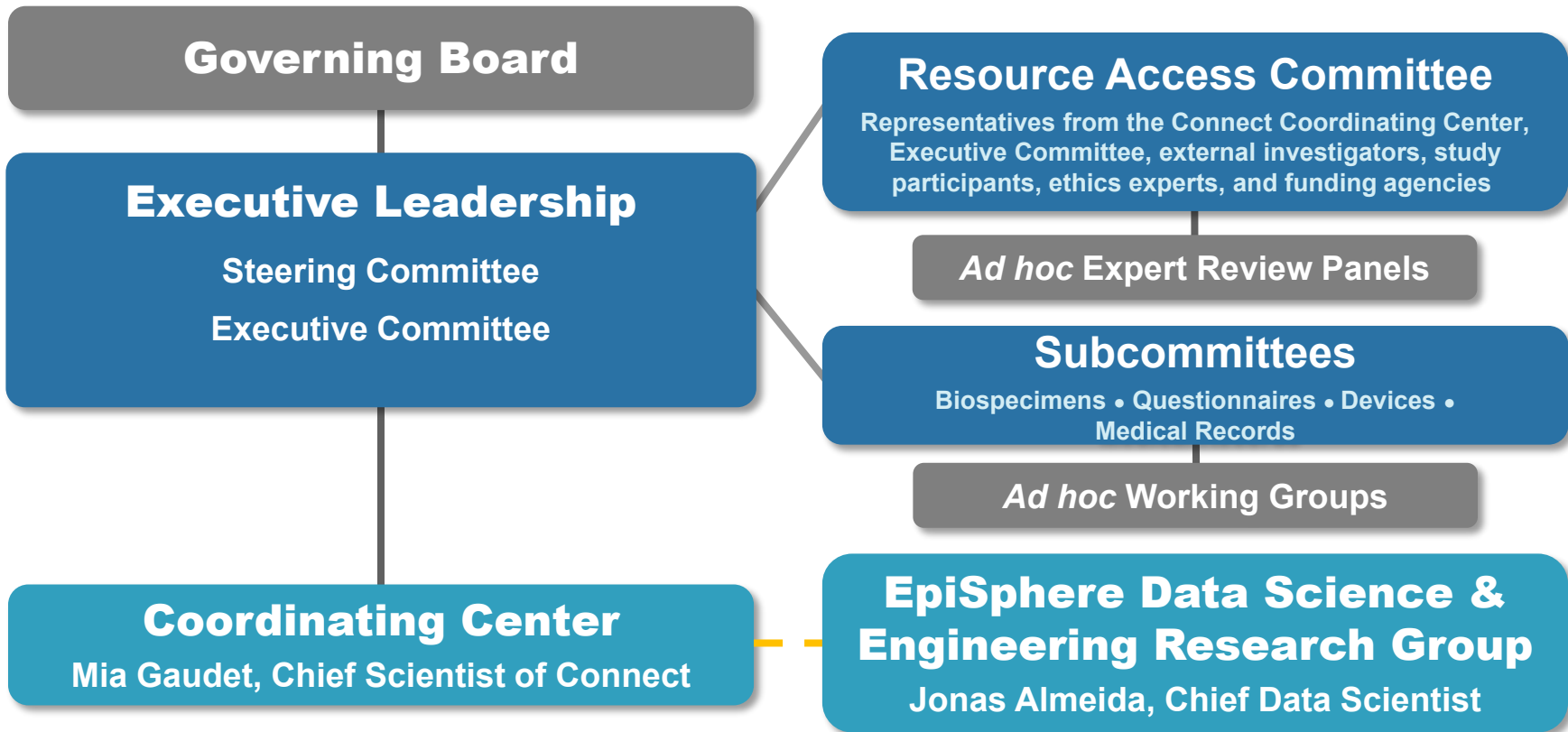Chun-Hung Chan, Sanford Health

**Other Contract Leads**

Carmit McMullen, PAB Lead

Stephen Smith, NORC Lead

Connect Leadership and Organization

Governing Board

Executive Leadership

Steering Committee

Executive Committee

Resource Access Committee

Representatives from the Connect Coordinating Center, Executive Committee, external investigators, study participants, ethics experts, and funding agencies

Ad hoc Expert Review Panels

Subcommittees

Biospecimens ● Questionnaires ● Devices ● Medical Records

Ad hoc Working Groups

Coordinating Center

Mia Gaudet, Chief Scientist of Connect

EpiSphere Data Science & Engineering Research Group

Jonas Almeida, Chief Data Scientist
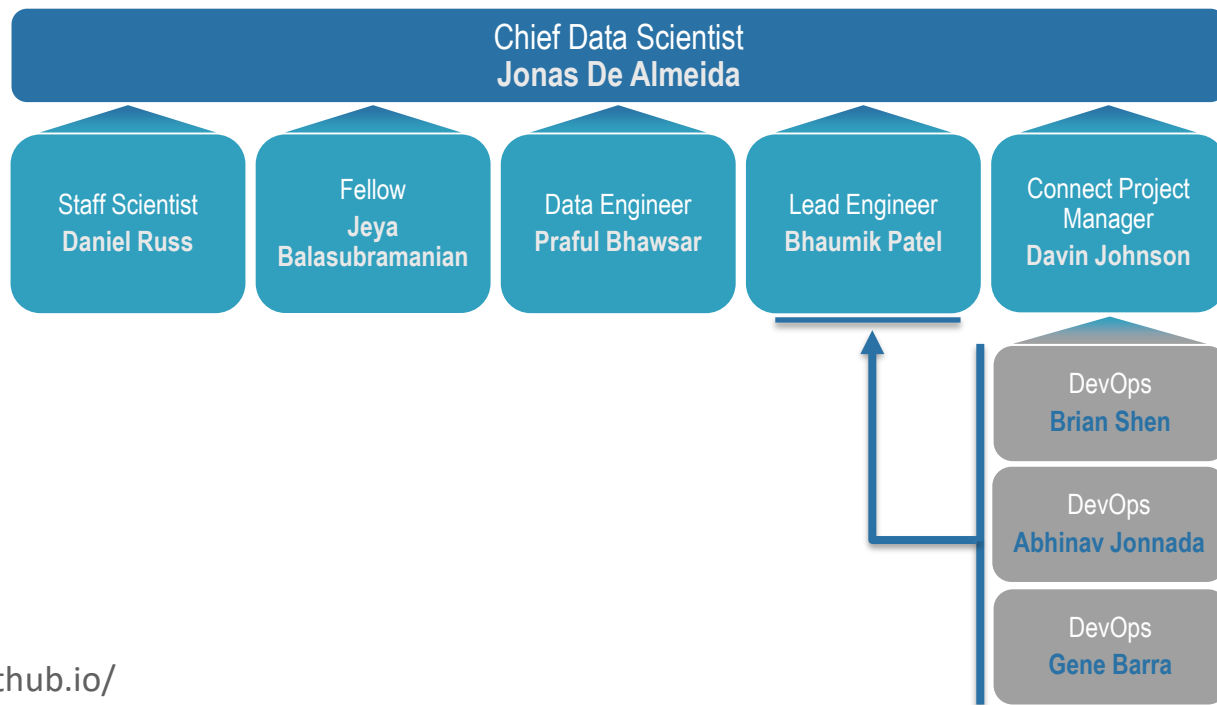
connect
for cancer prevention study

Connect Coordinating Center

Organizes and develops scientific proposals and protocols, operations, and resource utilization; monitors study progress and budget; manages contracts to IHCS sites, support services, and biorepository; obtains NIH IRB study approvals
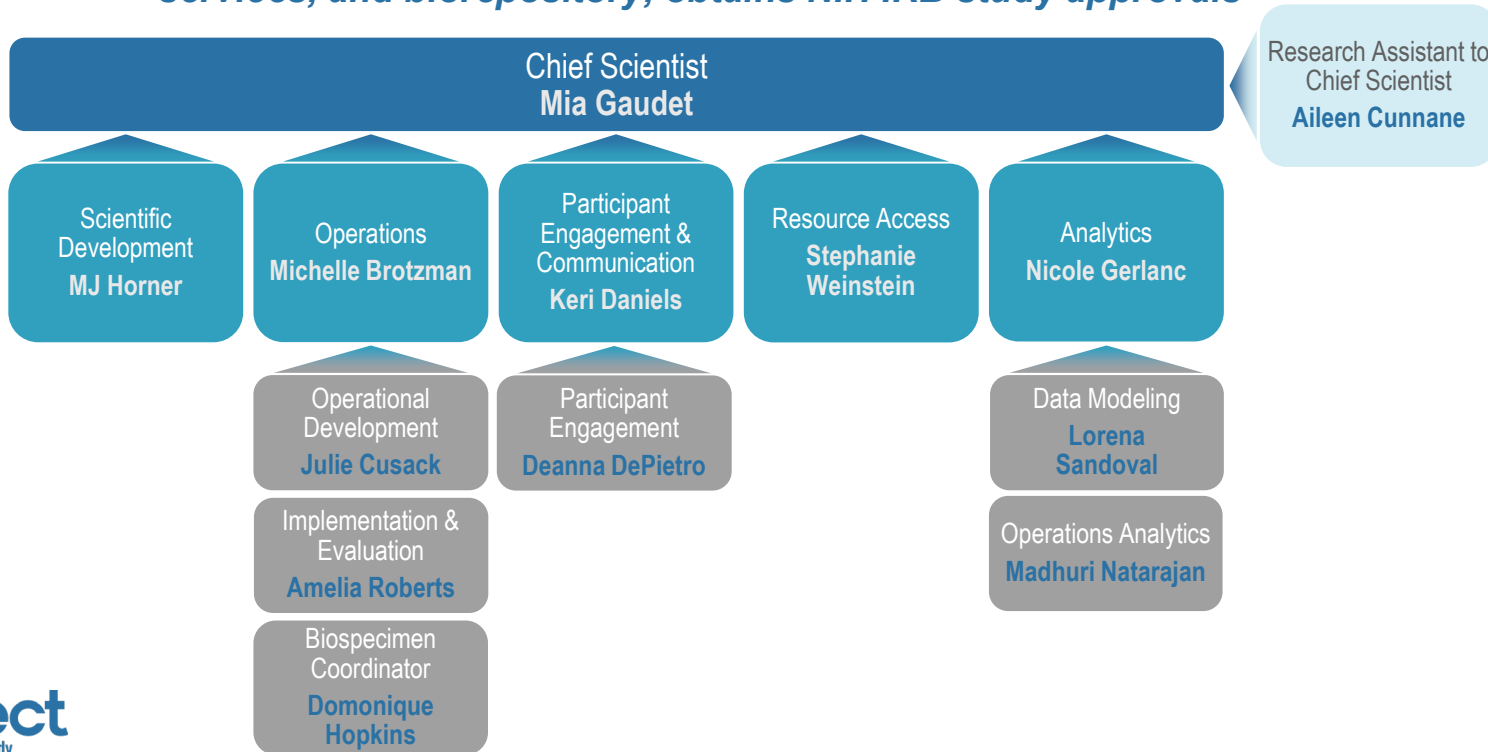
EpiSphere Data Science & Engineering Group

Advances research and infrastructure for data-intensive precision prevention studies

Chief Data Scientist
Jonas De Almeida

| Staff Scientist **Daniel Russ** | Fellow **Jeya Balasubramanian** | Data Engineer **Praful Bhawsar** | Lead Engineer **Bhaumik Patel** | Connect Project Manager **Davin Johnson** |

DevOps
Brian Shen

DevOps
Abhinav Jonnada

DevOps
Gene Barra

https://episphere.github.io/