Day 2
November 4, 2021

2021 IHCC LINK

International 100K+ Cohort Consortium

Day 2 Welcome

NIH DS-I Africa Grant

Race, Ethnicity, and Ancestry Presentations

Geoffrey Ginsburg, MD, PhD

Director
Duke University, Center for Applied Genomics

Co-Chair
International HundredK+ Cohorts Consortium

USA

Day 2 Outline

- NIH DS-I Africa Grant Presentation
- Race, Ethnicity, and Ancestry Presentations
  - Live panel discussion
- Environmental and Climate Data Capture Presentations
  - Live panel discussion
- 15 minute break
- Funding Opportunities/ Resource Presentations and Workshop Summary
  - Live panel discussion

Nicky Mulder, PhD

Professor
University of Cape Town

Principal Investigator
H3ABioNet

South Africa

DS-I Africa Open Data Science Platform

Nicky Mulder

University of Cape Town

NIH DSI-Africa

- Harnessing Data Science for Health Discovery and Innovation in Africa
  - **Research Hubs:** Advance and demonstrate feasibility of data science research and innovation to improve health in Africa
  - **Training:** Increase capacity for data science research in Africa
  - **ELSI Research:** Explore Ethical, Legal, and Social Implications of data science research from an African perspective and contribute to policy discussion on the continent
  - **Open Data Science Platform & Coordination Center:** Facilitate the development of a trans-African network of data scientists

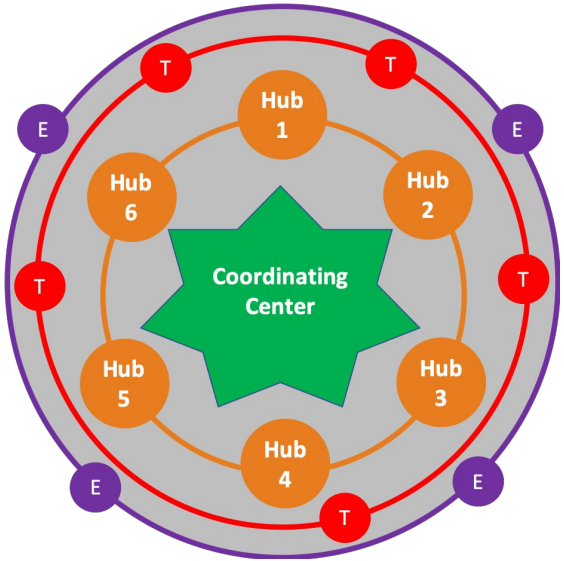


Image and text extracted from: https://commonfund.nih.gov/africadata

Funded project titles (1)

Research hubs

- Role of Data Streams In Informing **Infection Dynamics** in Africa- INFORM Africa
- UZIMA-DS: **UtiliZing health Information** for Meaningful impact in East Africa through Data Science
- Developing data science solutions to mitigate the **health impacts of climate change** in Africa: the HE2AT Center
- Harnessing Data Science to Promote Equity in **Injury and Surgery** for Africa
- Combatting **AntiMicrobial Resistance** in Africa Using Data Science (CAMRA)
- MADIVA (**Multimorbidity in Africa**: Digital innovation, visualisation and application)
- MUST Data Science Research Hub (MUDSReH)

Funded project titles (2)

ELSI:

- Research for Ethical Data Science in Southern Africa (REDSSA)
- Bridging Gaps in the ELSI of Data Science Health Research in Nigeria (BridgELSI)
- DS-I Africa - LAW
- Public Understanding of Big data in Genomics Medicine in Africa (PUBGEM-Africa)

7 Training grants on different aspects of data science & health

Open Data Science Platform & Coordinating Centre

Open Data Science Platform (ODSP) aim

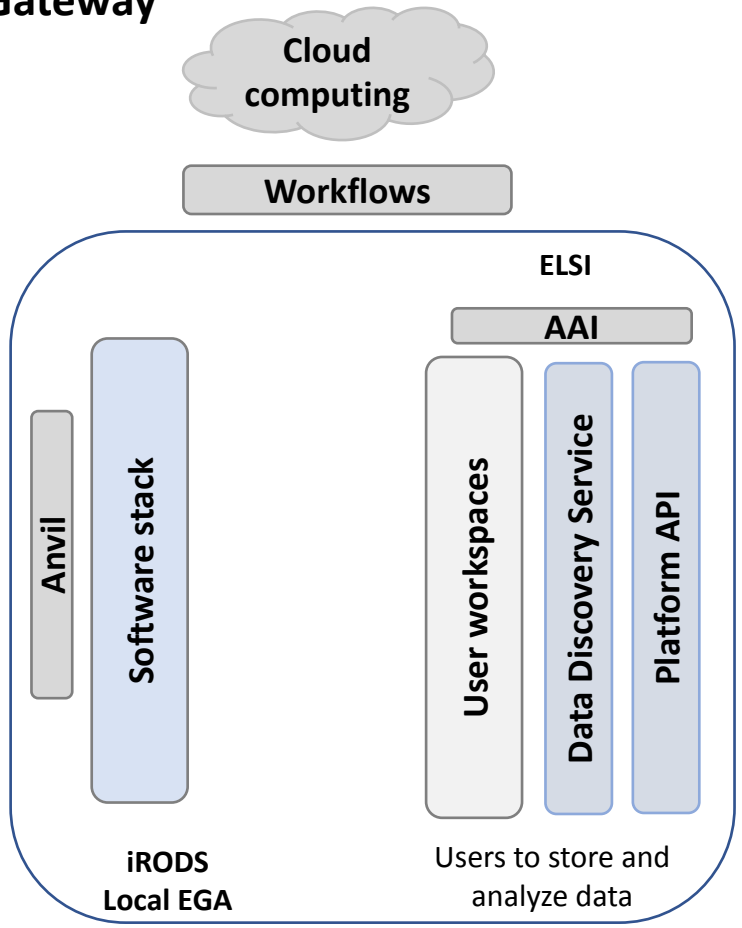
- To develop an African Open Data Science Gateway, including the eLwazi platform and associated resources, to support the Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa) consortium and beyond. eLwazi, will be a flexible, scalable Open Data Science Platform enabling the implementation of data science for health, that is relevant to the African context.

Ulwazi is the Xhosa word meaning “knowledge” or “information”,
and Olwazi means big rock in Luganda

eLwazi

eLwazi Data Science Gateway

AIM 1: Build Infrastructure



Cloud and local computing hosts

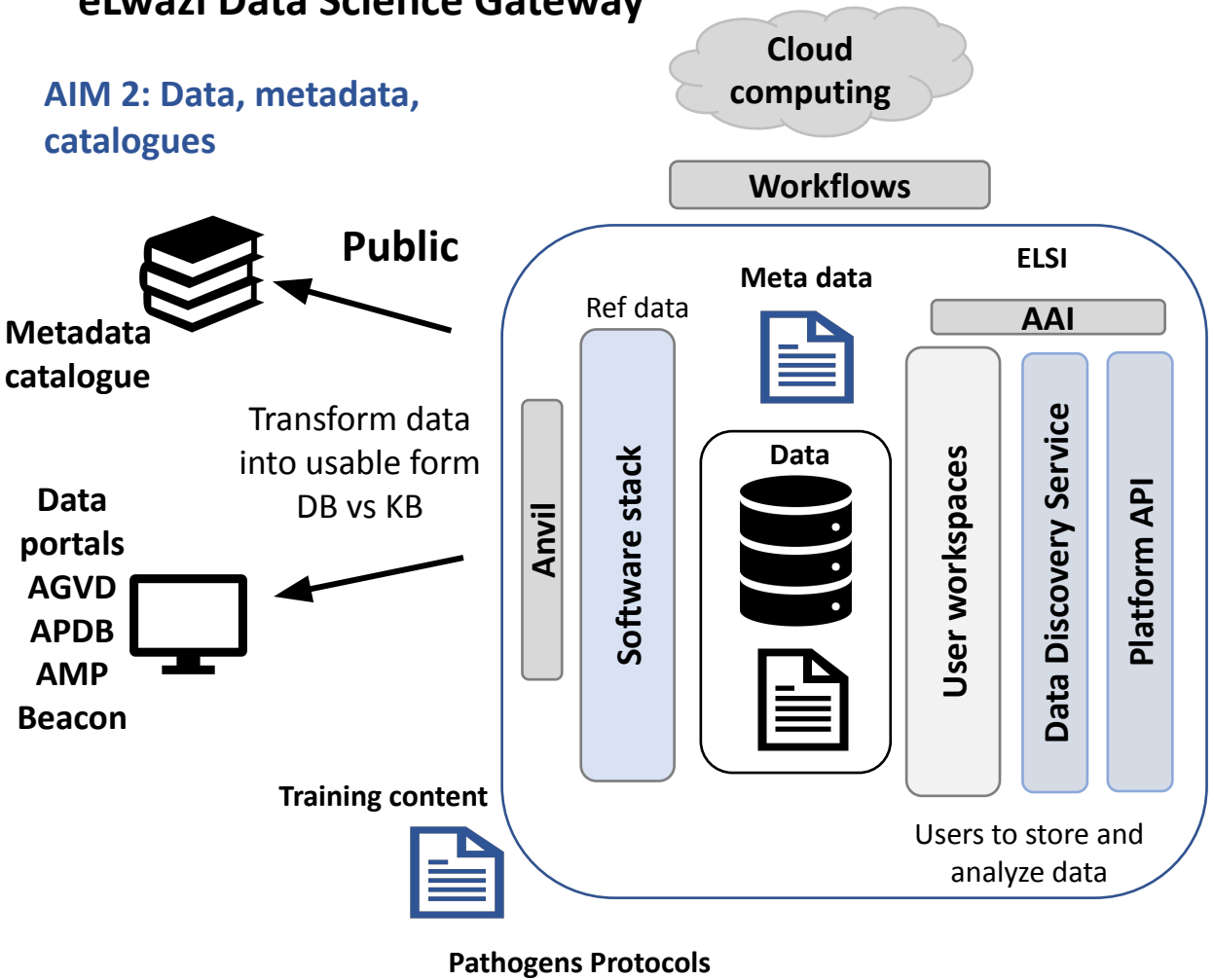
Security and access (AAI)

Aim 2. Access to data and resources

- Collect data types
- Reference datasets
- Metadata stds and curation, FAIR
- Data harmonization
- Cohorts atlas, catalogue
- Linking datasets
- Data Portals
- Other resources: SOPs, protocols, training materials

eLwazi Data Science Gateway

AIM 2: Data, metadata, catalogues

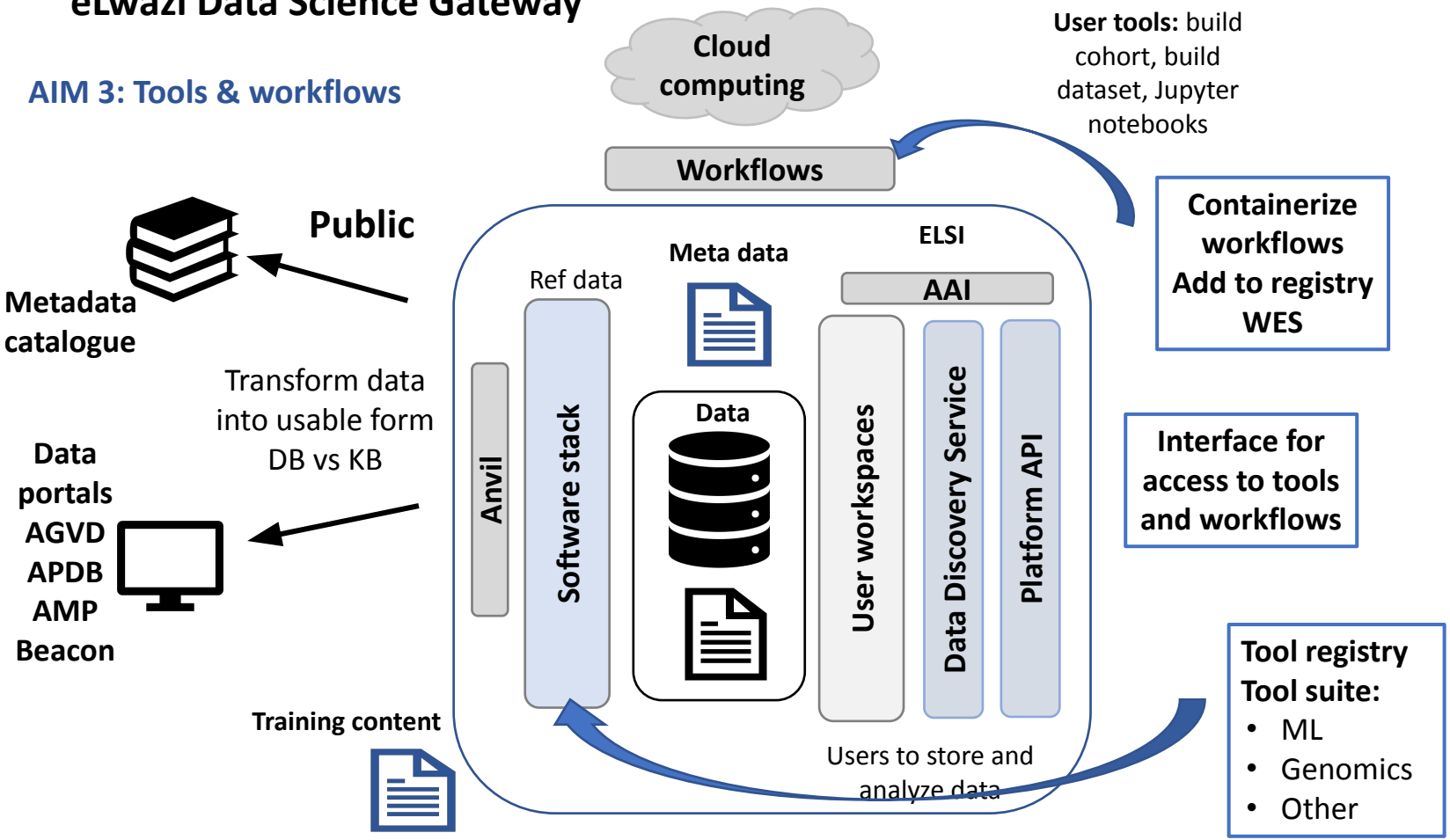


Aim 3. Tools and workflows

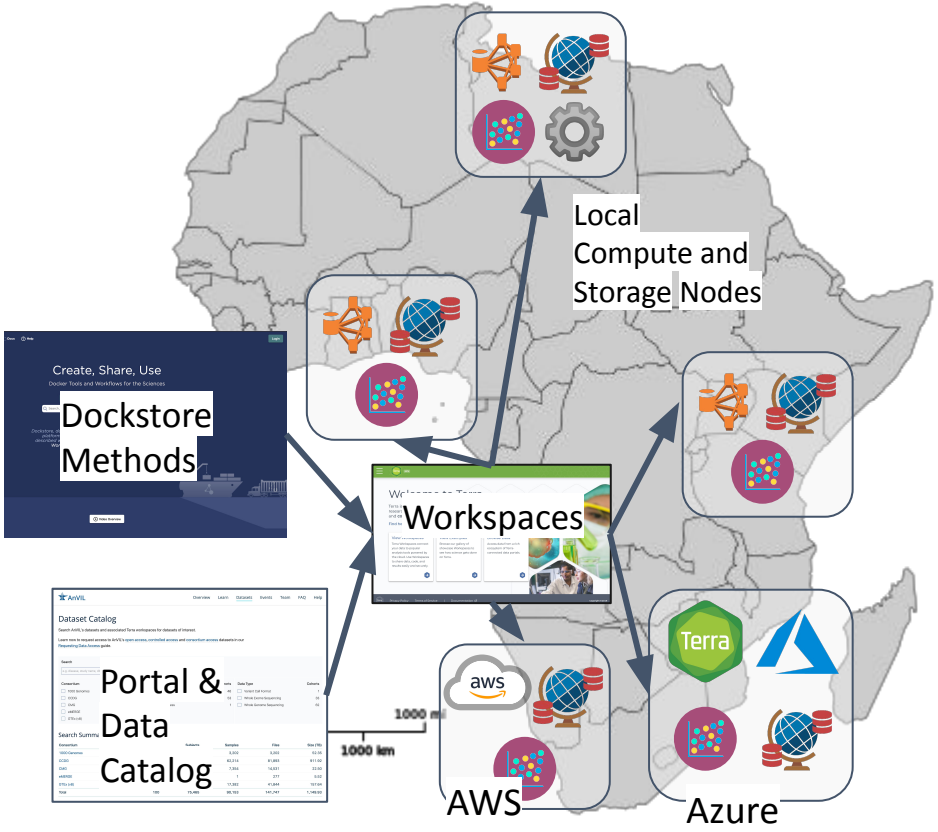
- Identify tools
- Installation on infrastructure
- Workflows
- Interface with Cloud providers
- Tool and workflow registries
- Galaxy
- Beginner and advanced access to tools

eLwazi Data Science Gateway

AIM 3: Tools & workflows



DS-I Africa ODSP hosting sites



Cloud & HPC Environments

- AWS, Azure, and Local

DRS for data access

WES for compute

TRS for workflow sharing

Workspaces for:

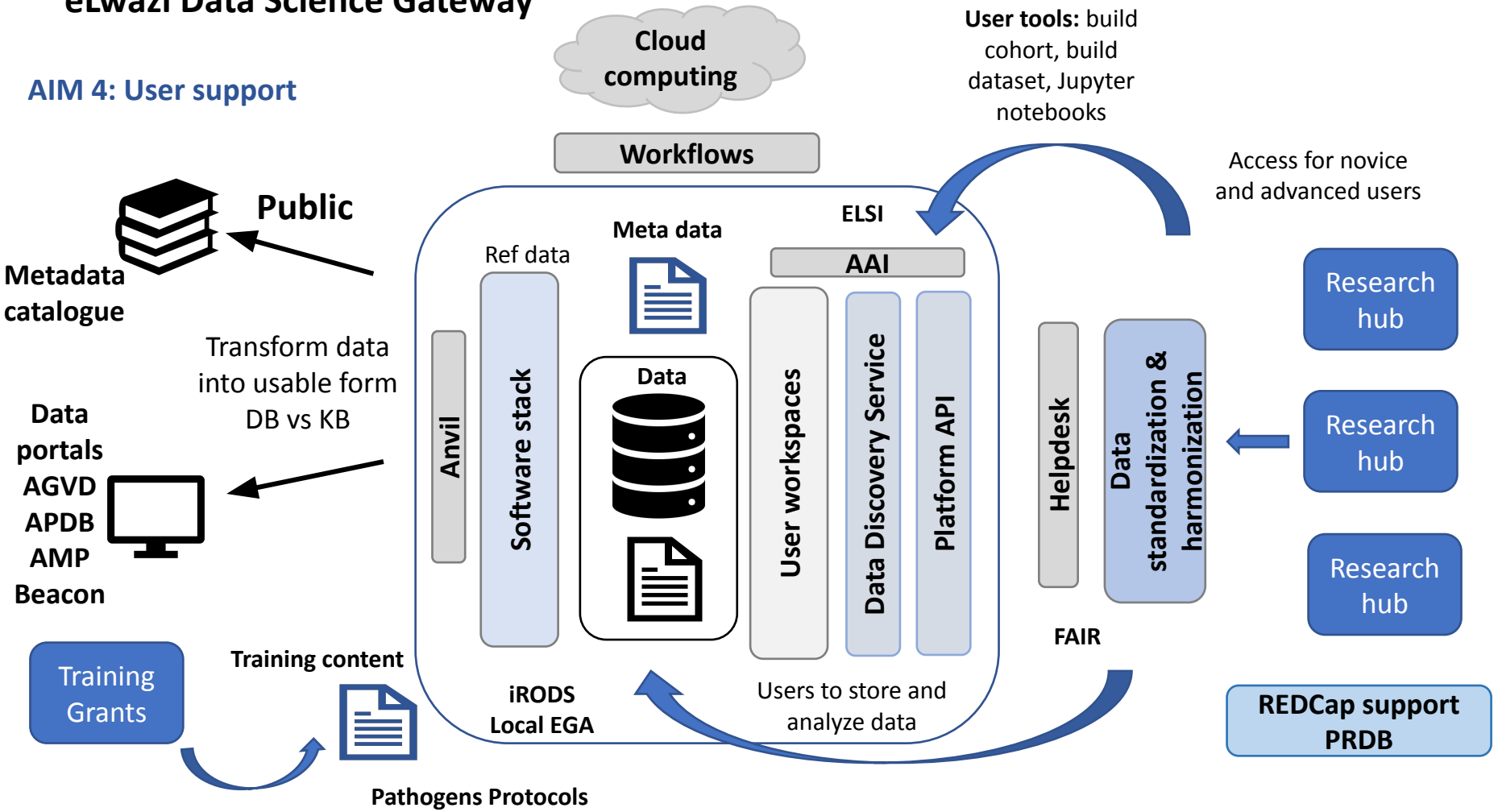
- Data modeling and access
- Compute on data in various locations
- Interactively analyze via Notebooks
- Sharing and Collaborating

Aim 4. User support

- User support type
  - User group meetings
  - Use case support
  - Clinical data support
- Helpdesk
- Training
  - Curriculum standards
  - Professional training
  - Coordinated with training programs

eLwazi Data Science Gateway

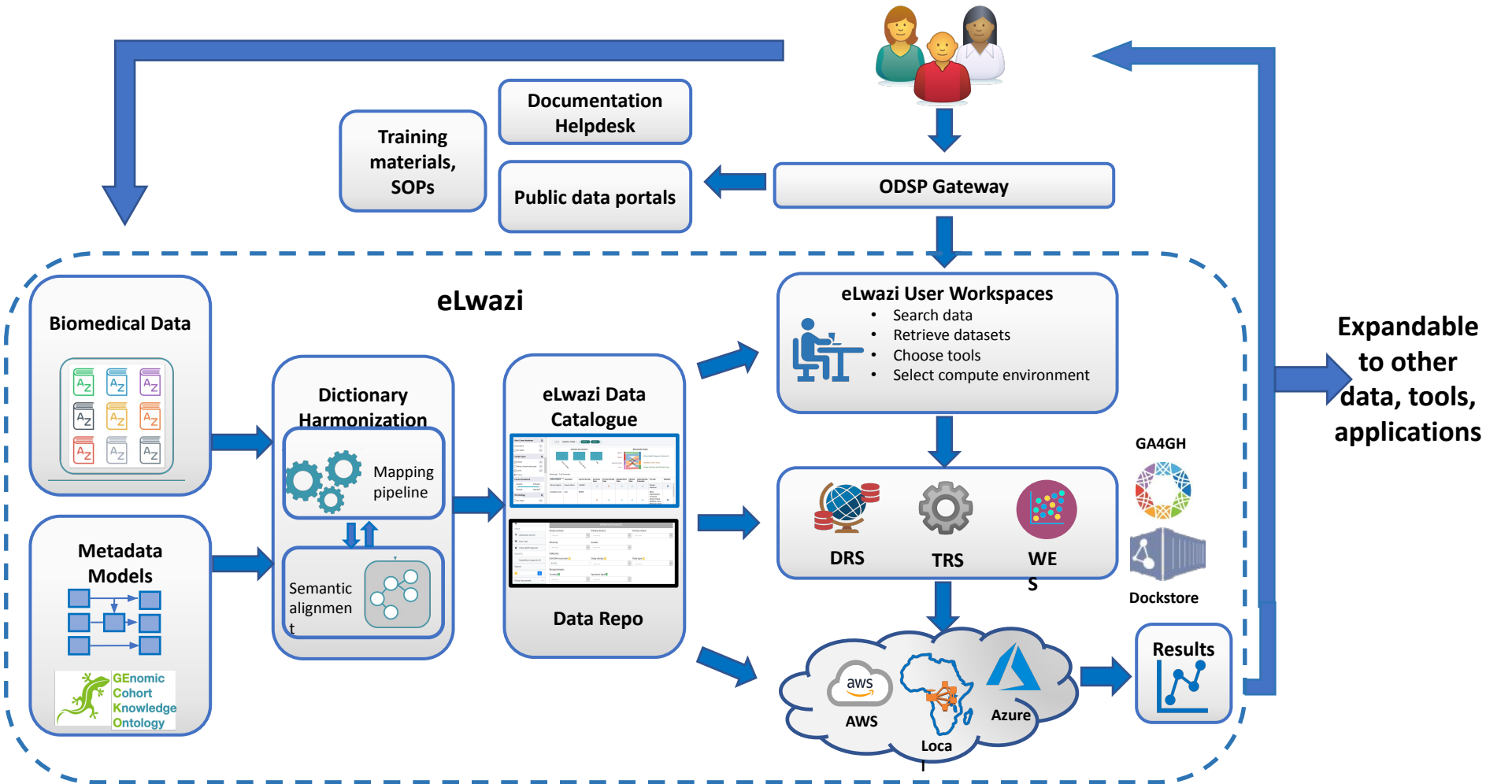
AIM 4: User support



Aim 5. Monitoring, outreach, sustainability

- Monitor platform usage and impact
- Dissemination
- Develop cost model and sustainability plan
- Engagement with external stakeholders
  - NRENS
  - Cloud providers
  - Industry

Overview of the ODSP



Acknowledgements

Coordinating Centre PI: Dr Michelle Skelton

ODSP Partners

- University of Cape Town
- University of the Witwatersrand
- University of the Western Cape (Ilifu)
- University of Kwazulu-Natal -> Stellenbosch
- University of Mauritius
- Uganda Virus Research Institute
- University of Khartoum
- USTTB, University of Bamako
- Broad Institute
- EMBL European Bioinformatics Institute
- University of California, Santa Cruz
- University of Chicago

Funding: DS-I Africa program through the NIH Common Fund

Questions?

Please put your questions in the chat

Race, Ethnicity, and Ancestry Presentations

Philip Awadalla, PhD

Director of Computational Biology

Ontario Institute for Cancer Research

Canada

Session Overview

Session Topics:

- Population Descriptors in Genomics Research: What classifications are used?
- Data Harmonisation

Session Speakers:

- Shawneequa Callier, MA, JD
- Mélanie Courtot, PhD

Shawneequa Callier, MA, JD

Associate Professor, Department of Clinical Research and Leadership
The George Washington University

USA

Population Descriptors in Genomics Research: What classifications are used?

Shawneequa L. Callier, JD, MA

Associate Professor
Clinical Research and Leadership
School of Medicine and Health Sciences
George Washington University

Special Volunteer
Center for Research on Genomics & Global Health
National Human Genome Research Institute
National Institutes of Health

International HundredK+ Cohorts Consortium
Virtual Conference
November 4, 2021

Agenda

1. The goal to harmonize cohorts in global genomics research
2. An IHCC cohort study to understand how researchers categorize populations
3. The challenges of using population descriptors in research
4. Opportunities for IHCC

Genomics Research Landscape

Goals

1. Global diversity and inclusion
2. Harmonization or standardization of terms
3. Determination of what descriptors we should use

IHCC Survey

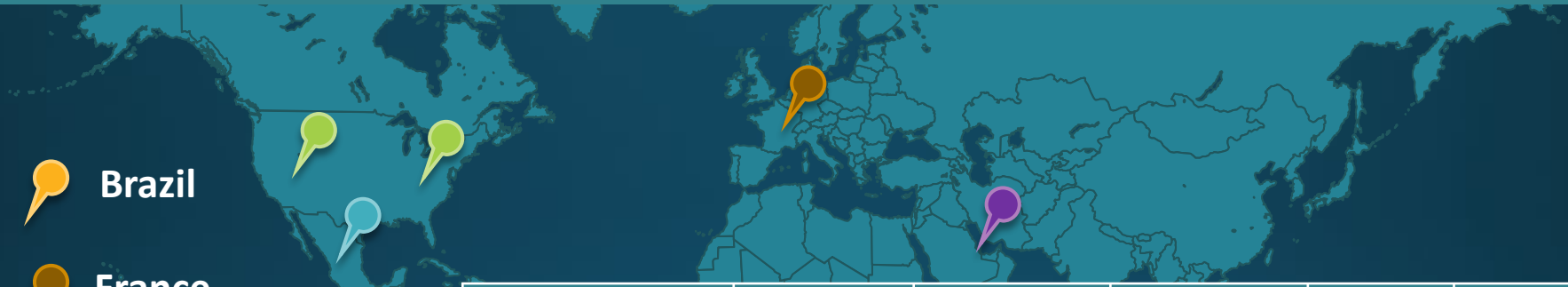
- Objectives
  - Understand how IHCC cohorts record population descriptors
  - Assess IHCC cohort opportunities and challenges related to collecting, defining, and reporting data
  - Use findings to inform discussion about harmonization and standardization of global data

- Status
  - IRB-approved study
  - Data-collection phase

IHCC Survey & Research Project

Harmonization of race, ethnicity, and ancestry measures across the globe

Collaboration with International HundredK+ Cohorts Consortium (IHCC)



- Brazil
- France
- Malaysia
- Mexico
- Qatar
- United States

	Race	Ethnicity	Ancestry	Tribe	Other
Brazil	Yes	Yes	Yes	No	No
France	No	No	Yes	No	Yes
Malaysia	Yes	Yes	Yes	No	No
Mexico	No	No	No	No	No
Qatar	No	No	No	No	No
US Cohort 1	Yes	Yes	Yes	No	No
US Cohort 2	Yes	Yes	No	No	No

Race / Ethnicity

Race:
1 White
2 Black/African-American
3 Asian
4 American Indian/Alaska Native
5 Native Hawaiian/Pacific Islander
6 Other

Ethnicity:
Spanish/Hispanic/Latino yes/no

What is your race?

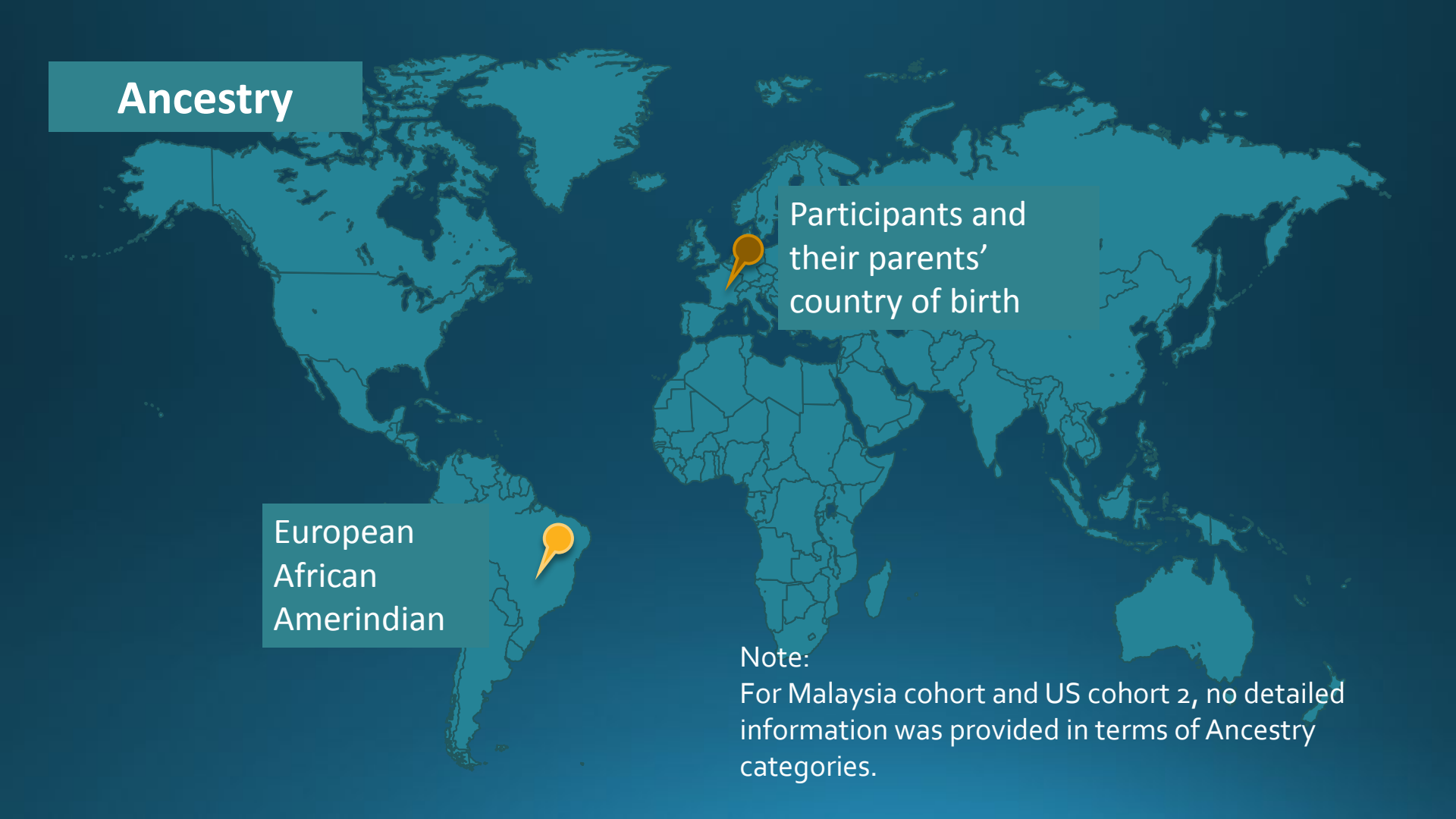
1 White
2 Black, African-American, or Negro
3 American Indian or Alaska Native
4 Asian Indian
5 Chinese
6 Filipino
7 Japanese
8 Korean
9 Vietname White
10 Other A Black
11 Native H Mixed
12 Guamar Asian ro
13 Samoan Indigenous
14 Other Pacific Islander

Are you Spanish/Hispanic/Latino?

0 No, not Spanish/Hispanic/Latino
1 Yes, Puerto Rican
2 Yes, Mexican, Mexican American, or Chicano
3 Yes, Cuban
4 Yes, other Spanish/Hispanic/Latina

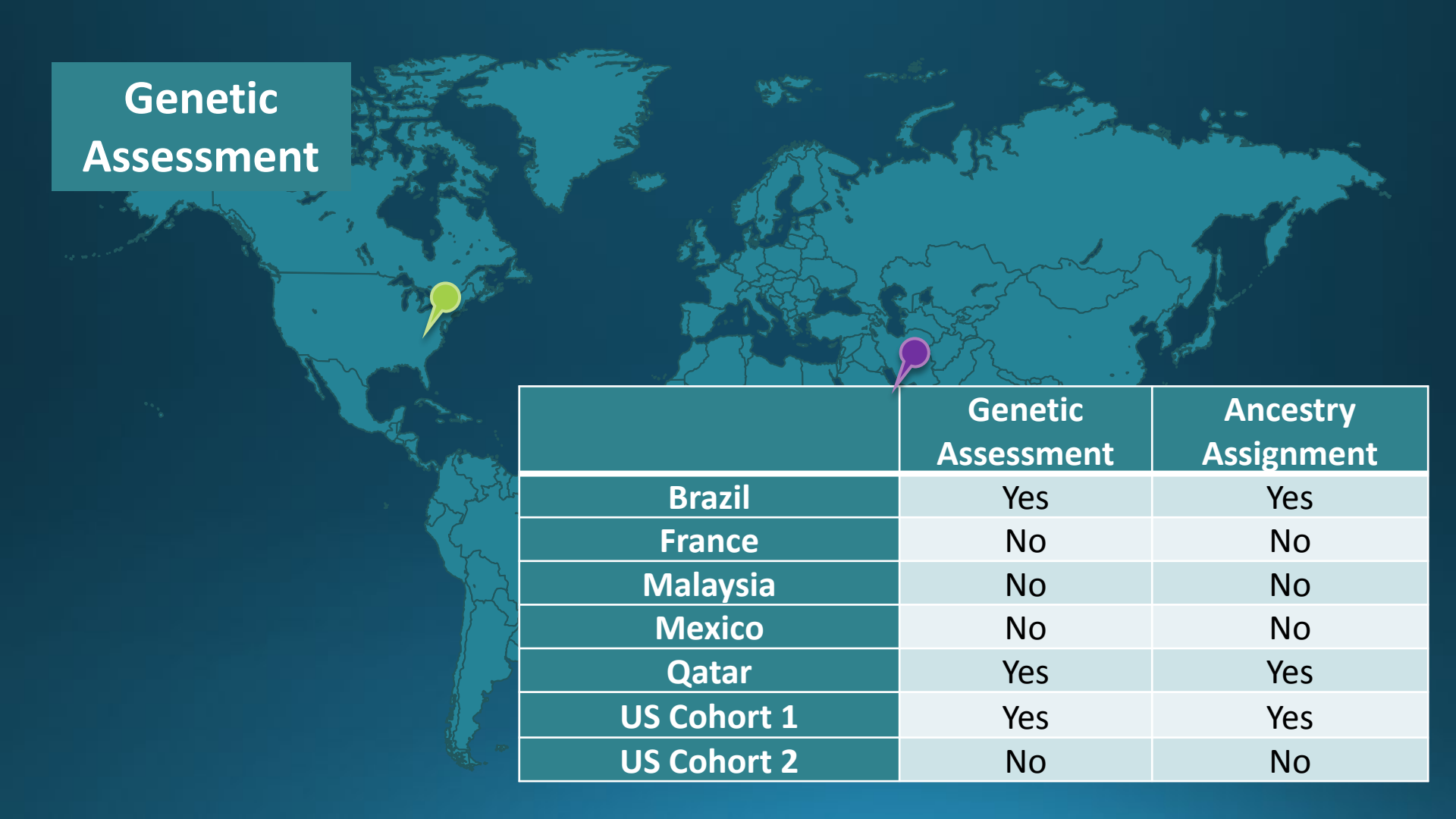
National Statistics Department (no detailed information provided)

Ancestry



Note:
For Malaysia cohort and US cohort 2, no detailed information was provided in terms of Ancestry categories.

Genetic Assessment



	Genetic Assessment	Ancestry Assignment
Brazil	Yes	Yes
France	No	No
Malaysia	No	No
Mexico	No	No
Qatar	Yes	Yes
US Cohort 1	Yes	Yes
US Cohort 2	No	No

The challenges of using population descriptors in research

Human Ancestry Ontology

Human ancestry ontology for the NHGRI GWAS Catalog

- ancestry category
  - Aboriginal Australian
  - African
    - Sub-Saharan African
  - African American or Afro-Caribbean
    - African American
    - Afro-Caribbean
    - Anguillan
    - Antiguan or Barbudan
    - Barbadian
    - Haitian
    - Jamaican

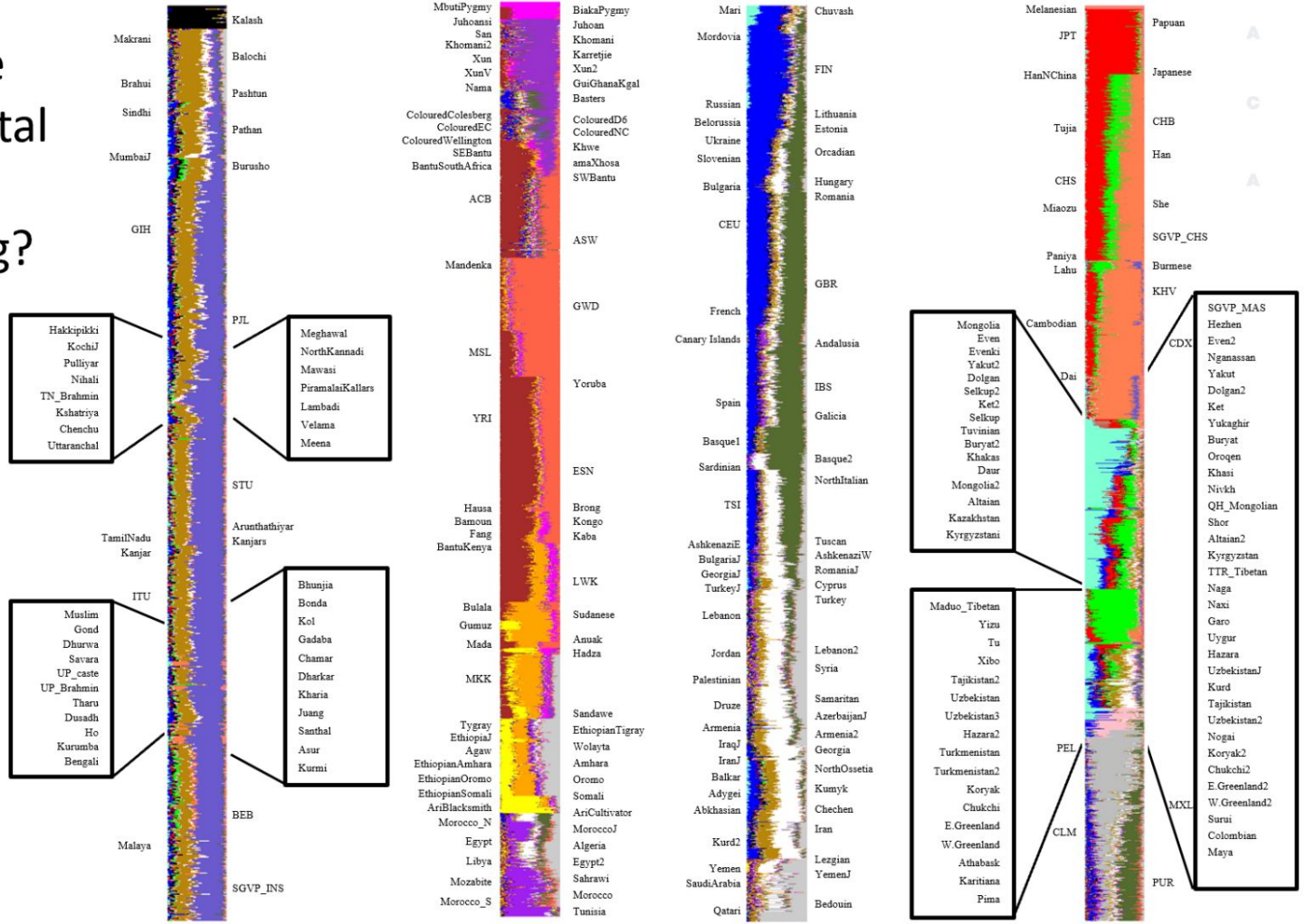
- ancestry status
  - admixed ancestry
  - genetically isolated ancestry
- continent
  - Africa
  - Asia
  - Europe
  - Latin America and the Caribbean
  - Northern America
  - Oceania
- country
- region

- Asian
  - Central Asian
  - East Asian
  - Filipino
  - Nepalese
  - South Asian
  - South East Asian

- Hispanic or Latin American
  - Brazilian
  - Costa Rican
    - Costa Rican founder
  - Cuban
  - Dominican
  - Mexican
  - Puerto Rican
- Native American
  - Pima Indian
  - Plains American Indian
- Oceanian
  - Kosraen
  - Micronesian
  - Native Hawaiian
  - Pacific Islander
  - Papua New Guinean
  - Solomon Islander

- European
  - Amish
  - British
  - Canadian
    - French Canadian founder
  - Dutch
    - Dutch founder
  - Erasmus Rucphen
  - European American
  - Finnish
  - Hutterite
  - Icelandic
  - Irish
    - genetically homogenous Irish
  - Italian
    - Carlantino
    - Cilento
    - Friuli Venezia Giulia
    - Italian founder
    - Sardinian
    - Talana
    - Val Borbera
  - Korculan
  - Old Order Amish
  - Orcadian
  - Romanian founder
  - Sorbian
  - South Tyrolean
  - Spanish
  - Split founder
  - Vis founder

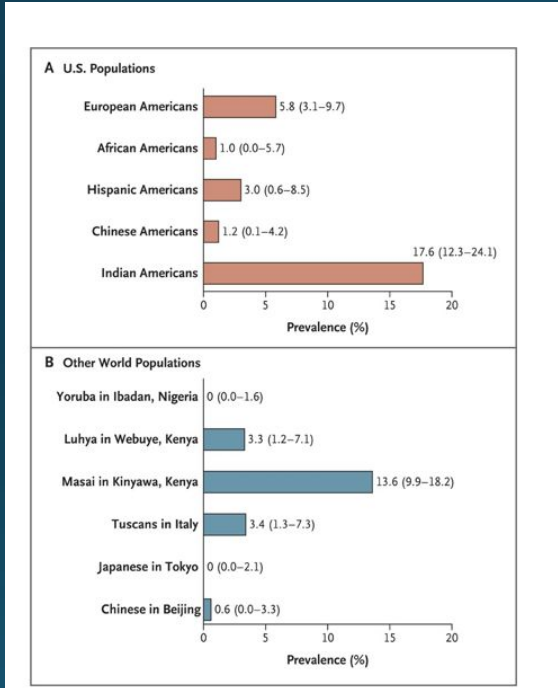
What are continental labels capturing?



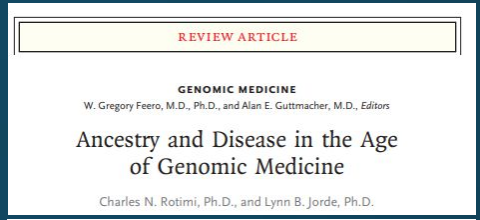
Baker, J.L., Rotimi, C.N. & Shriner, D. Human ancestry correlates with language and reveals that race is not an objective genomic classifier. *Sci Rep* **7,** 1572 (2017).

NIH NHGRI

What are continental labels capturing?



Rotimi CN, Jorde LB. N Engl J Med 2010;363:1551-1558.



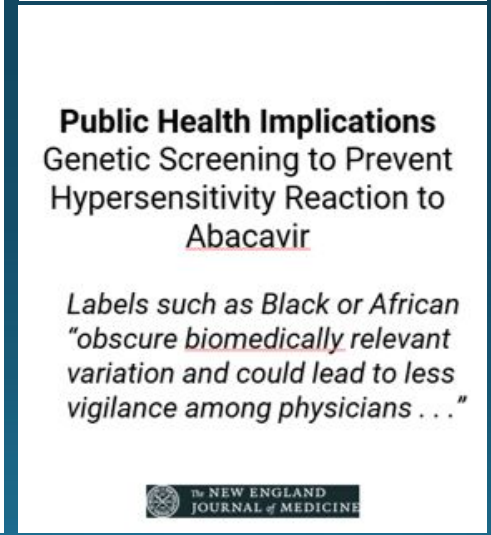


Figure 1. Variation in the HLA-B*5701 Locus in 11 HapMap Samples.

What are racial and ethnic labels capturing?

Nationalgeographic.com/magazine/2018/04/race-twins-black-white-biggs

Race in a Genetic World

MAY-JUNE 2008

HARVARD MAGAZINE

"I AM AN African American," says Duana Fullwiley, "but in parts of Africa, I am white." To do fieldwork as a medical anthropologist in Senegal, she says, "I take a plane to France, a seven- to eight-hour ride. My race changes as I cross the Atlantic. There, I say, '*Je suis noire*,' and they say, 'Oh, okay—*métisse*—you are mixed.' Then I fly another six to seven hours to Senegal, and I am white. In the space of a day, I can change from African American, to *métisse*, to *tubaab* [Wolof for "white/European"]. This is not a joke, or something to laugh at, or to take lightly. It is the kind of social recognition that even two-year-olds who can barely speak understand. '*Tubaab*,' they say when they greet me."

Dr. Duana Fullwiley
Anthropologist
Stanford University

ARTICLE

Chad Genetic Diversity Reveals an African History Marked by Multiple Holocene Eurasian Migrations

Marc Haber,[1,*] Massimo Mezzavilla,[1,2] Anders Bergström,[1] Javier Prado-Martinez,[1] Pille Hallast,[1,3] Riyadh Saif-Ali,[4] Molham Al-Habori,[4] George Dedoussis,[5] Eleftheria Zeggini,[1] Jason Blue-Smith,[6,10] R. Spencer Wells,[7] Yali Xue,[1] Pierre A. Zalloua,[8,9] and Chris Tyler-Smith[1,*]

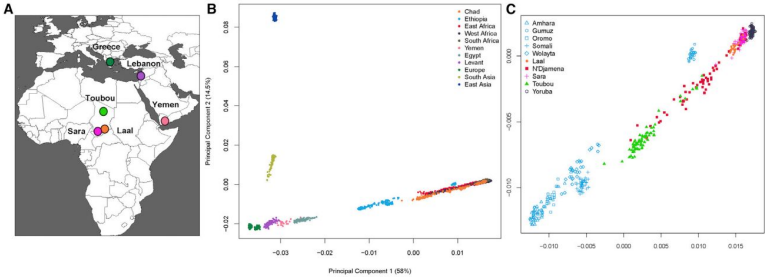


Figure 1. Population Locations and Genetic Structure
(A) The map shows the location of newly genotyped or sequenced populations.
(B) PCA of worldwide populations shows that Near Easterners and East Africans are intermediate to Eurasians and sub-Saharan Africans on PC1. Chad populations are close to sub-Saharan Africans and have some samples drawn toward Ethiopians.
(C) Magnification of the African PCA shows different affinities of the Chad populations to other Africans: the Toubou cluster close to Ethiopians, whereas the Sara and Laal speakers are close to the Yoruba. The mixed samples from N'Djamena, the capital, are intermediate to the Toubou, Sara, and Laal speakers.

The American Journal of Human Genetics 99, 1316–1324, December 1, 2016

Opportunities for IHCC

- Understand practices regarding population descriptors within a global consortium
- Assess ambiguities and inconsistencies in how cohorts collect, record, and report population data
- Contribute to the discourse on global harmonization in ethical and scientifically appropriate ways

Mélanie Courtot, PhD

Metadata Standards Coordinator, Archival Infrastructure and Technology Team
European Bioinformatics Institute (EMBL-EBI)

UK

Data harmonisation

Mélanie Courtot
EMBL-EBI
mcourtot@ebi.ac.uk

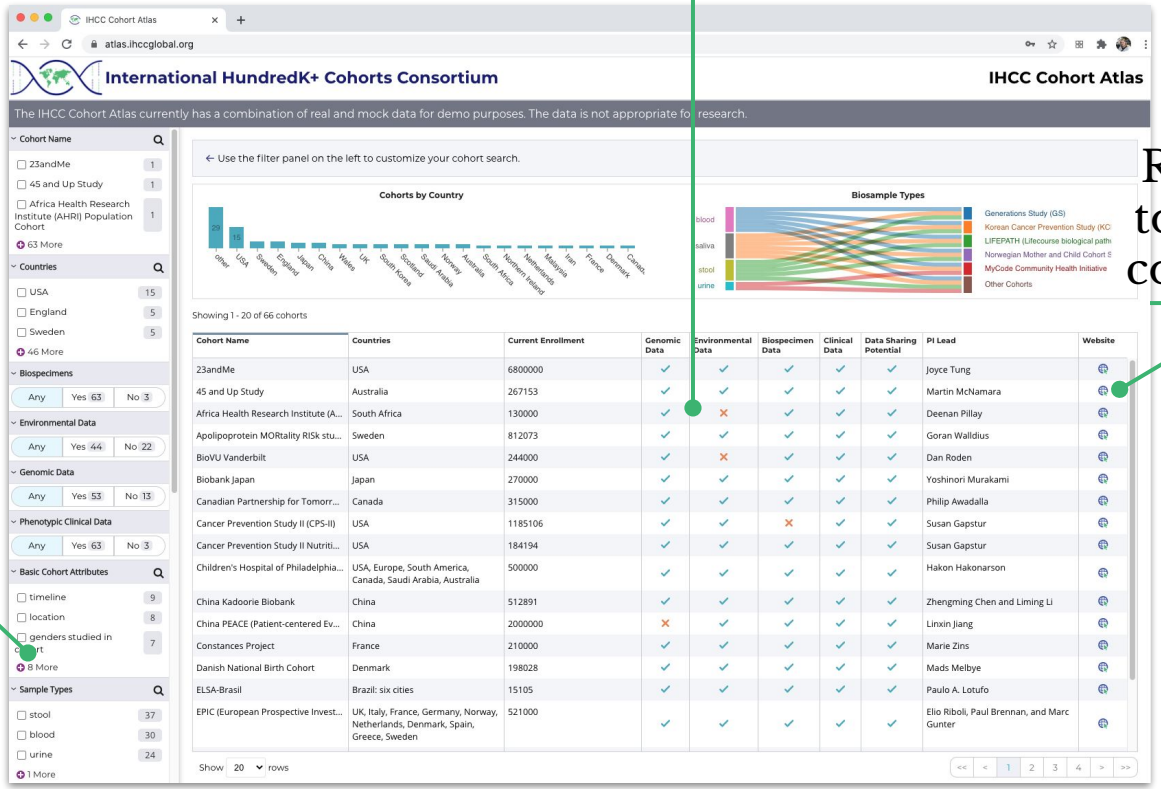
Picture: Alina Grubnyak @alinnnaaaa
https://unsplash.com/photos/ZiQkhI7417A

IHCC cohort atlas

Cohort presentation and display

Intuitive filtering by cohort metadata & data dictionary attributes

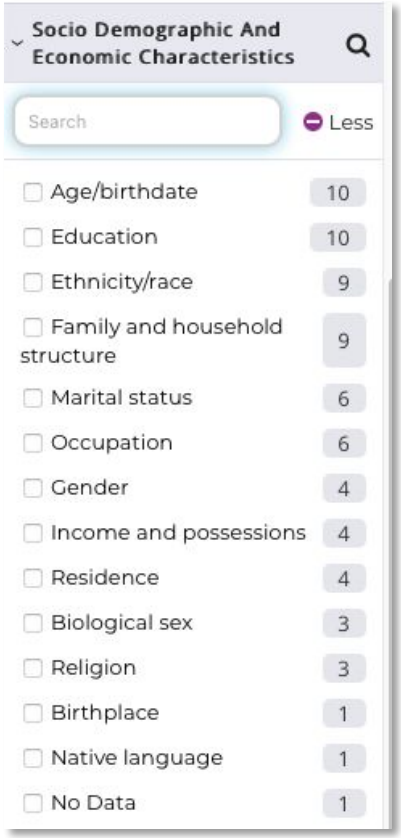
Reference to external cohort sites



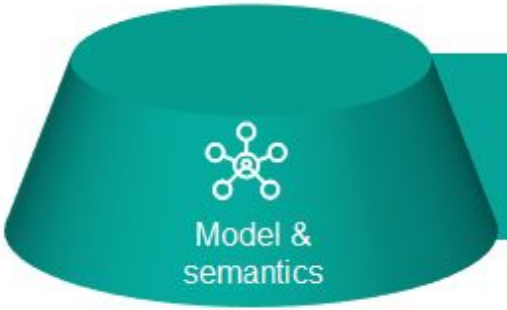
https://atlas.ihccglobal.org

Current status

- GECKO model
- Filters do not distinguish between race and ethnicity
- Cohorts collect different elements; some used as proxy, e.g., country, native language



Next steps

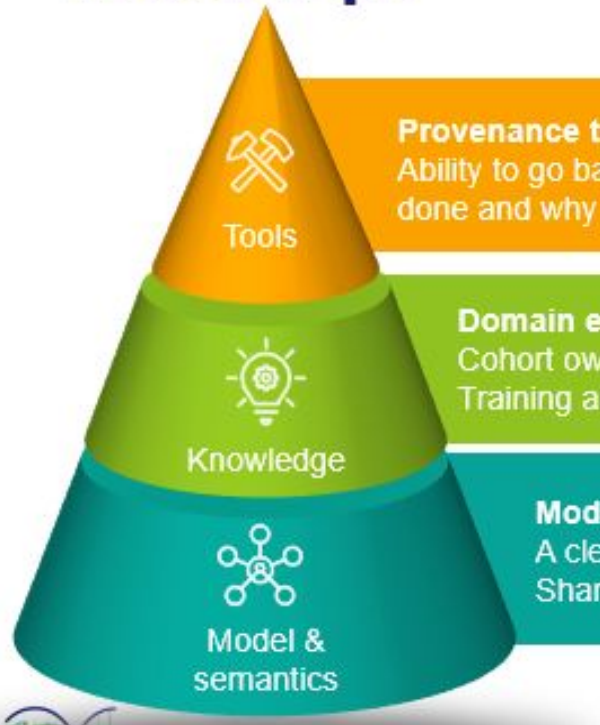


Model building
A clear definition of categories to be represented – label and textual definition
Shared community understanding and adoption

Next steps


Knowledge
Model &
semantics
Domain expertise
Cohort owner expertise to map dictionaries towards shared standard
Training and documentation
Model building
A clear definition of categories to be represented – label and textual definition
Shared community understanding and adoption

Next steps



Provenance tracking
Ability to go back to the original information and understand which mappings were done and why

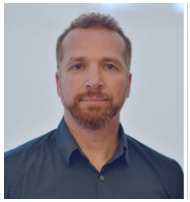
Domain expertise
Cohort owner expertise to map dictionaries towards shared standard
Training and documentation

Model building
A clear definition of categories to be represented – label and textual definition
Shared community understanding and adoption

Contact us! ihcc-browser@googlegroups.com

Acknowledgements

Thomas Keane



Philip Awadalla

Christina Yung

Rosi Bajari

Giselle Kerry

Melanie Courtot

Eric Plummer

KNOCEAN

James Overton

Rebecca Jackson

Nicolas Matentzoglu

Minh Ha

Brandon Chan

Carles Garcia

Isuru Liyanage

Dan Brake

Chris Lunt

elixir

CINECA

OICR Ontario Institute for Cancer Research

EMBL-EBI

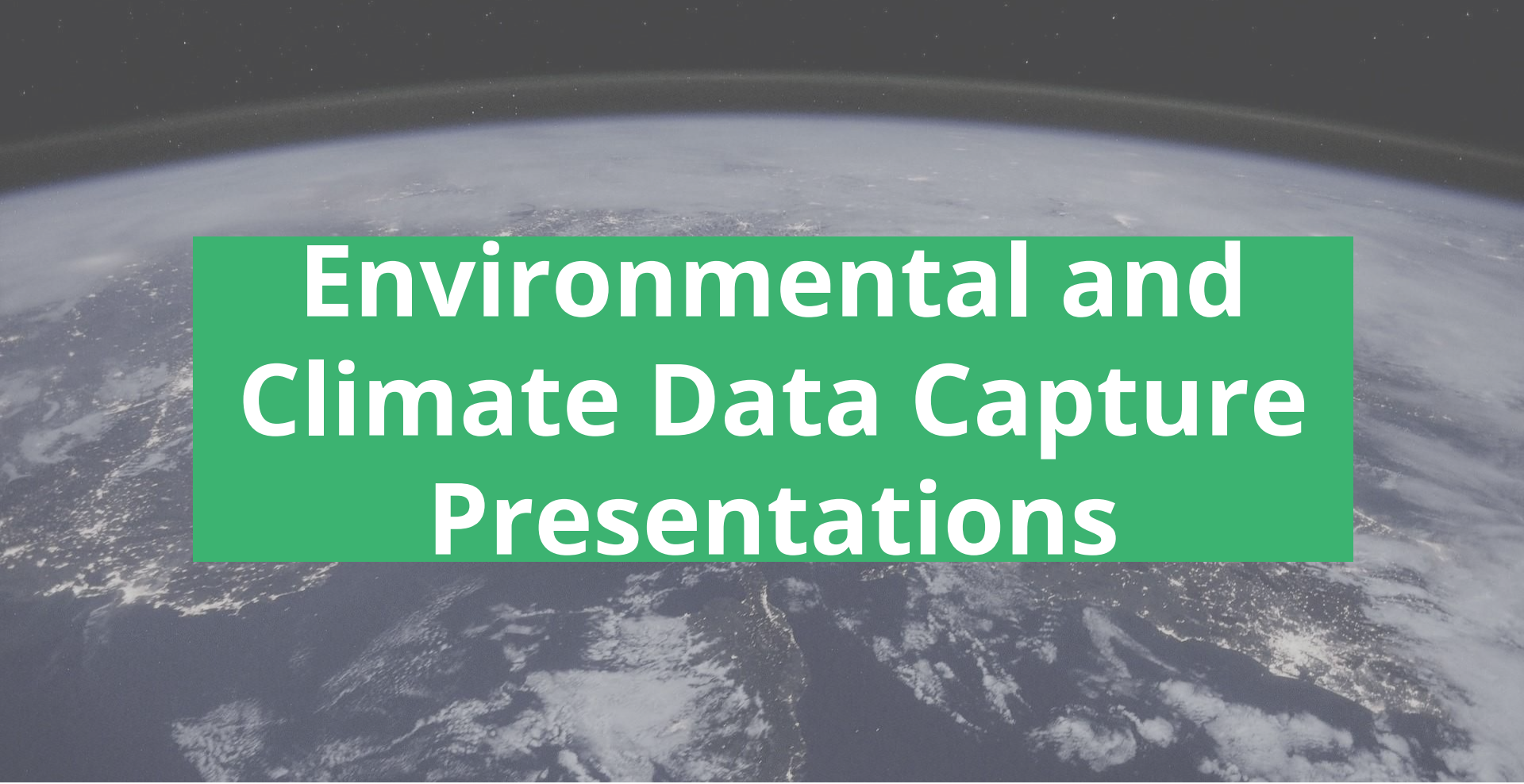
NIH National Institutes of Health Turning Discovery Into Health

Global Genomic Medicine Collaborative

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825775.

Live Panel Discussion

Please put your questions in the chat



Environmental and Climate Data Capture Presentations

Kyle Messier, PhD

Stadtman Investigator
Spatiotemporal Health Analytics Group

USA

Session Overview

Session Topics:

- Geospatial Exposures and Tools for Human Cohorts
- Design and Analysis of Longitudinal Population Studies for Climate-Health Research

Session Speakers:

- Alison Motsinger-Reif, PhD
- Peter Diggle, PhD

Alison Motsinger-Reif, PhD

Branch Chief

National Institute of Environmental Health Sciences

USA

Geospatial Exposures and Tools for Human Cohorts

Overview

Complex traits are the results of genes, environment, and their interactions

Exposome data tries to collect the totality of a person's lifetime exposures (an admittedly impossible goal)

Ongoing NIEHS efforts: PEGS cohort as a motivating example

Personalized Environment and Genes Study

Genotype

Phenotype

Environment

- **SNPs**
  (>11,000 participants)
- **WGS (Broad Institute)**
  (>4,700 subset)

- **Disease status (Qx)**
  (>9,000 participants)

- **Exposures** (>9,000)
  (eg diet, smoking)
- **Exposome Survey**
  (>3,000 participants)
- **Geospatial exp. data**

PEGS Laboratory

- DNA, RNA isolation
- Candidate SNP genotyping/analysis

Clinical Research Unit

- Translational projects
- Surveys, blood, plasma, serum, urine

Statistical Genetics

- WGS, exposome, eHR
- GWAS, vQTL, GxE PheWAS, PRS

Health and Environment Surveys

Health and Exposure (H&E)

- Health, family history of disease, occupational exposures, socioeconomic status, and lifestyle data

Exposome

- Data on internal and external exposures, including medications, physical activity, stress, sleep, diet, chemical and environmental exposures at home and work

Core medical history

- Provides medical history data on diseases and medical conditions, smoking, alcohol, sleep, and surgeries

Exposome Survey

Part A: 'External Exposome'
approximately 200 questions

- Characteristics of current and past homes
- Workplace characteristics
- Chemical and metal exposures at work
- Hobby exposures
- Ultraviolet light exposures

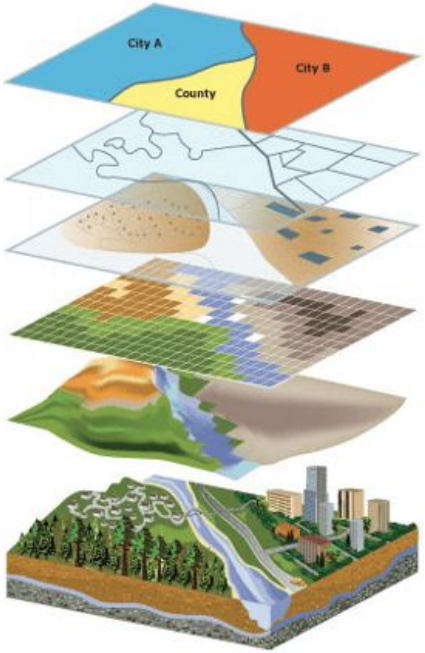
Part B: 'Internal Exposome'
approximately 240 questions

- Medications
- Vitamins, minerals, dietary supplements
- Chemotherapy/radiation therapy
- Physical activity
- Stress
- Infection
- Sleep
- Dietary behavior
- Dietary intake
- Siblings/twins/birth order
- Genetic history

GIS Exposure Estimates

Growing Number of Data Layers

- Airports
- CAFOs
- Cellular towers
- Drinking water
- Dry cleaners
- Hazardous waste
- Highways
- Nuclear sites
- Wastewater
- Population info
- Power lines
- PR landfills
- Railroads
- Spills
- Sanitary landfills
- Superfund sites
- Toxic release sites,
- Etc.



Address at the time of survey completion and longest lived childhood address

Immune Mediated Diseases

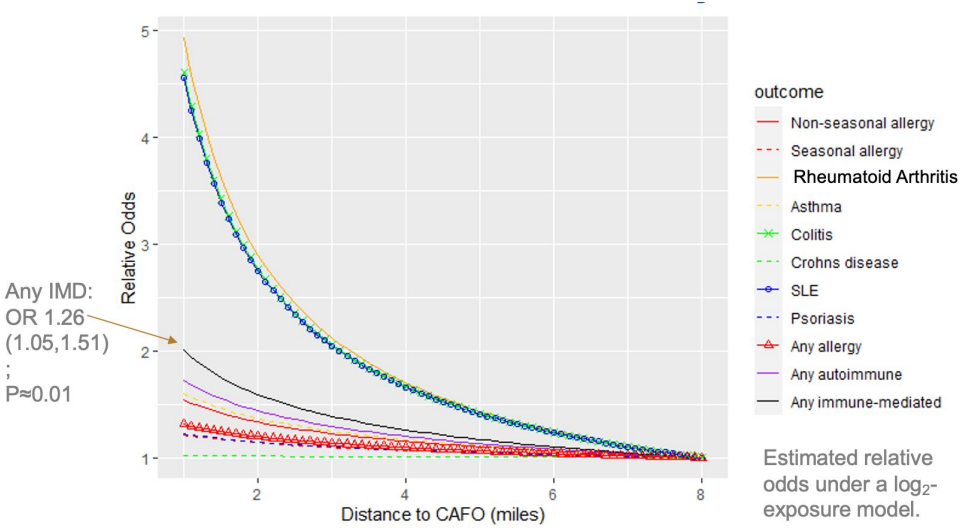
We hypothesize that proximity to swine CAFOs and IMD risk SNPs are associated with increased occurrence of immune-mediated diseases (IMD).

For array of IMD outcomes and mechanistic outcome groups we assessed:

1. The association between CAFO proximity and IMD
2. The association between IMD SNPs and IMD
3. Gene-environment interaction in IMD (GxE)[1]

[1]Interaction modeled according to Aliev et al (2014)

Association between CAFO Proximity and IMD



- Proximity to CAFOs was consistently associated with increased risk across IMD phenotypes.
- Significant associations were observed with rheumatoid arthritis, colitis, and any IMD (c=269).

Genetic associations with IMD

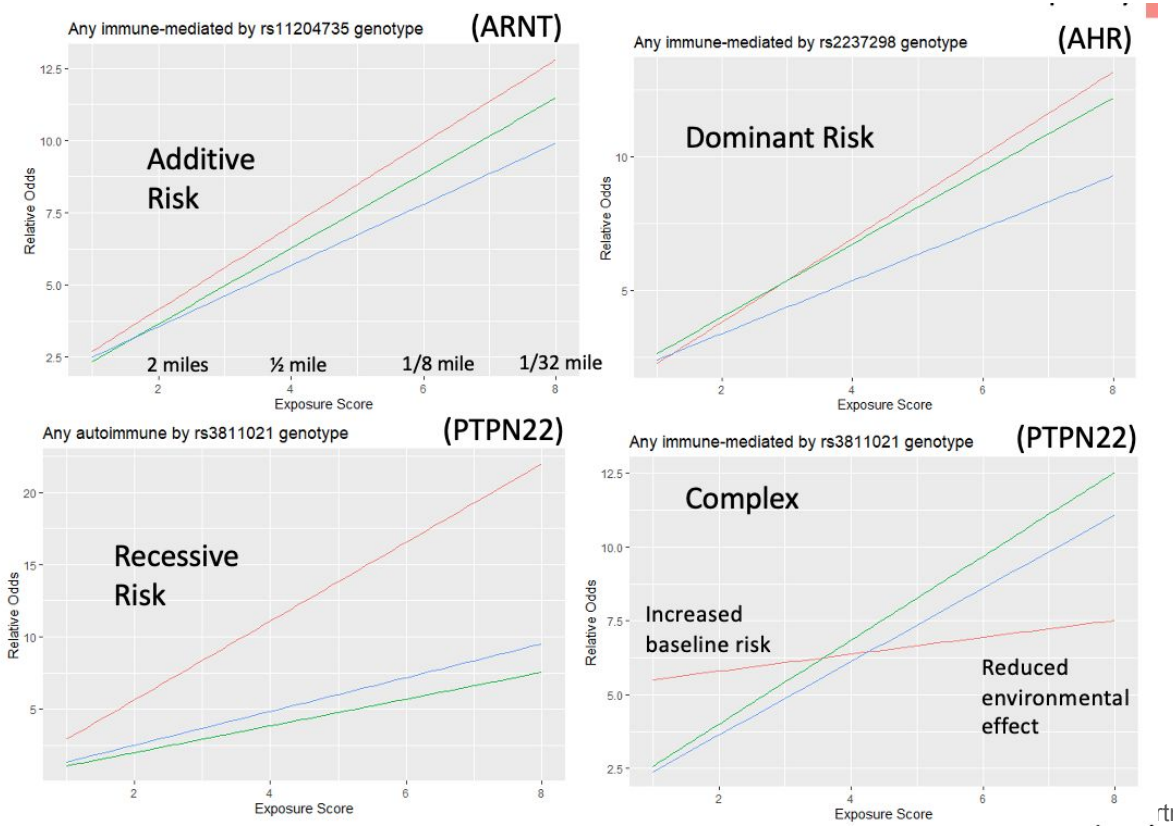
- Test for associations between SNPs previously associated with an IMD
  - AHRR, ARNT, PTPN22 and RSBN1

- rs2066853 (AHR) significantly associated with decreased odds of IMD, any allergy, and seasonal allergy

- Minor alleles increase disease, consistent across multiple IMD phenotypes
  - Suggests common mechanisms

Evidence for GXE Interactions

$ES = -\log_2(d)+3$



Ongoing Analysis

Many projects to understand the etiology of disease, and find gene-environment interactions

- Questionnaire based Exposome-wide Associations Studies
- Polygenic risk score development and GXE
- Methods development for prioritizing SNPs for GXE analysis
  - Genelist.niehs.nih.gov
- Genome-wide association analysis
- Association of geospatial exposures with diseases
- Linkage to health records from Duke and UNC Chapel Hill

Potential Sources of Exposome Data

Biomarker measurements in biospecimens

- Blood
- Urine
- Plasma
- Saliva
- Feces
- Placenta
- Umbilical cord blood
- Breast milk
- Hair, nails

Potential Sources of Exposome Data

Questionnaire data

- Residential exposures
- Occupational exposures
- Socioecomic status
- Educational
- Diet
- Lifestyle
- Medications

Potential Sources of Exposome Data

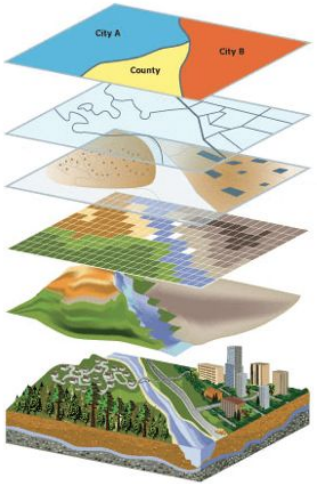
Geospatial Exposures – linked to addresses

- Residential address

- Address history

- Occupational address(es)

- Childhood address(es)



Potential Sources of Exposome Data

Data resources with exposure models

- Residential address
- Address history
- Occupational address(es)
- Childhood address(es)

Wearables and smart phones

Numerous exposure models and databases

- Weather and atmosphere
- Air pollution
- Groundwater contaminatio
- Distance to hazards
- Health disparities information
- Built environment

Challenges in data “munging”

- Different geographic extents
- Different data sources
- Different special resolutions
- Different temporal resolutions
- Not consistently structured

Challenges in data “munging”

- Different geographic extents
- Different data sources
- Different special resolutions
- Different temporal resolutions
- Not consistently structured

Data efforts at NIEHS: building software resources to empower assembling data resources

Pollutant	Geographic Extent	Type	Data Source	Spatial Resolution	Temporal Resolution	Link to data documentation
Aerosol Index	Worldwide	Orbital spectroscopy	Aura	2 km	Daily	https://earthdata.nasa.gov/earth-observation-data/near-real-time/hazards-and-disasters/air-quality
UV Aerosol Index	Worldwide	Orbital spectroscopy	Suomi	2 km	Daily	https://earthdata.nasa.gov/earth-observation-data/near-real-time/hazards-and-disasters/air-quality
Aerosol Optical Depth	Worldwide	Orbital spectroscopy	MODIS (Aqua/Terra)	1 km	Daily	https://earthdata.nasa.gov/earth-observation-data/near-real-time/hazards-and-disasters/air-quality
Carbon Monoxide	Worldwide	Orbital spectroscopy	Aqua/Aura/Terra	2 km	Day/Night	https://earthdata.nasa.gov/earth-observation-data/near-real-time/hazards-and-disasters/air-quality
Fires	Worldwide	Orbital spectroscopy	Aqua/Tetta/Suomi	1 km	Daily	https://earthdata.nasa.gov/earth-observation-data/near-real-time/hazards-and-disasters/air-quality
Nitrous Oxide	Worldwide	Orbital spectroscopy	Aura	5 km	Day/Night	https://earthdata.nasa.gov/earth-observation-data/near-real-time/hazards-and-disasters/air-quality
Ozone	Worldwide	Orbital spectroscopy	Aura / Suomi	2 km	Day/Night	https://earthdata.nasa.gov/earth-observation-data/near-real-time/hazards-and-disasters/air-quality
Sulfur Dioxide	Worldwide	Orbital spectroscopy	Aqua/Aura/Suomi	2km	Day/Night	https://earthdata.nasa.gov/earth-observation-data/near-real-time/hazards-and-disasters/air-quality
Land use	Worldwide	Orbital spectroscopy	MODIS (Aqua/Terra)	500m	6-month	https://modis.gsfc.nasa.gov/data/dataprod/mod12.php
Roadways	Worldwide	Survey/Administrative	Compilation	Long/Lat	~Decade	https://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1
Temperature	Worldwide	Meterology	Weather station network	5 degree grid	Monthly	https://www.ncdc.noaa.gov/data-access/marineocean-data/noaa-global-surface-temperature-noaaglobaltemp
Sulfur Dioxide	United States	Point emissions estimates	National Emissions Inventory	Long/Lat	3 Years	https://www.epa.gov/sites/production/files/2020-04/documents/nei2017_tsd_full_30apr2020.pdf
Volatile organic componds	United States	Point emissions estimates	National Emissions Inventory	Long/Lat	3 Years	https://www.epa.gov/sites/production/files/2020-04/documents/nei2017_tsd_full_30apr2020.pdf
Nitrogen oxides	United States	Point emissions estimates	National Emissions Inventory	Long/Lat	3 Years	https://www.epa.gov/sites/production/files/2020-04/documents/nei2017_tsd_full_30apr2020.pdf
Carbon dioxide	United States	Point emissions estimates	National Emissions Inventory	Long/Lat	3 Years	https://www.epa.gov/sites/production/files/2020-04/documents/nei2017_tsd_full_30apr2020.pdf
Lead	United States	Point emissions estimates	National Emissions Inventory	Long/Lat	3 Years	https://www.epa.gov/sites/production/files/2020-04/documents/nei2017_tsd_full_30apr2020.pdf
$PM_{2.5}$	United States	Point emissions estimates	National Emissions Inventory	Long/Lat	3 Years	https://www.epa.gov/sites/production/files/2020-04/documents/nei2017_tsd_full_30apr2020.pdf
PM_{10}	United States	Point emissions estimates	National Emissions Inventory	Long/Lat	3 Years	https://www.epa.gov/sites/production/files/2020-04/documents/nei2017_tsd_full_30apr2020.pdf
Ammonia	United States	Point emissions estimates	National Emissions Inventory	Long/Lat	3 Years	https://www.epa.gov/sites/production/files/2020-04/documents/nei2017_tsd_full_30apr2020.pdf
Carcinogens	United States	Point emissions estimates	Toxics Release Inventory	Long/Lat	Yearly	https://www.epa.gov/toxics-release-inventory-tri-program/tri-listed-chemicals
Acute respiratory toxicants	United States	Point emissions estimates	Toxics Release Inventory	Long/Lat	Yearly	https://www.epa.gov/toxics-release-inventory-tri-program/tri-listed-chemicals
Persistent organic pollutants	United States	Point emissions estimates	Toxics Release Inventory	Long/Lat	Yearly	https://www.epa.gov/toxics-release-inventory-tri-program/tri-listed-chemicals
Criteria air pollutants	United States	Fused Emissions Model	Synthesis of multiple datasets	Census Tract / BG	Yearly	https://www.caces.us/data
$PM_{2.5}$	United States	Downscaler model	EPA Air Quality System	Metro area	Daily	https://healthdata.gov/dataset/daily-census-tract-level-pm25-concentrations-2011-2014
PM_{10}	United States	Air monitor data	EPA Air Quality System	Metro area	Daily	https://www.epa.gov/aqs
Pesticides	United States	Survey/Land Use	USGS	County	Yearly	https://water.usgs.gov/nawqa/pnsp/usage/maps/
Roadway Density	United States	Survey/Land Use	USDOT	Census Block	~Decade	https://www.transportation.gov/mission/health/proximity-major-roadways#aboutthedata
Race/ethnicity	United States	Population survey	American Community Survey	Census Tract or ZIP	Yearly (rolling sample)	https://www.census.gov/programs-surveys/acs
Age/Sex	United States	Population survey	American Community Survey	Census Tract or ZIP	Yearly (rolling sample)	https://www.census.gov/programs-surveys/acs
Education/Income/Employment	United States	Population survey	American Community Survey	Census Tract or ZIP	Yearly (rolling sample)	https://www.census.gov/programs-surveys/acs
Poverry/public assistance	United States	Population survey	American Community Survey	Census Tract or ZIP	Yearly (rolling sample)	https://www.census.gov/programs-surveys/acs
Housing Indicators	United States	Population survey	American Community Survey	Census Tract or ZIP	Yearly (rolling sample)	https://www.census.gov/programs-surveys/acs
Social Vulnerability Index	United States	Derived Rank Measure	Derived from ACS Data	Census Tract	2 Years	https://svi.cdc.gov/data-and-tools-download.html
Quality of care	United States	Process indicator	Medicare/Medicaid Data	County	Yearly	https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures
Airports	United States	Point emissions sources	FAA	Long/lat	~Decade	https://catalog.data.gov/dataset/airports
Cellular Broadcast Towers	United States	Point emissions sources	FCC	Long/lat	~Decade	https://hifld-geoplatform.opendata.arcgis.com/datasets/0835ba2ed38f494196c14af8407454fb_0
Hazardous waste disposal sites	United States	Point emissions sources	EPA	Long/lat	~Decade	https://www.epa.gov/enviro/envirofacts-data-service-api
Nuclear	United States	Point emissions sources	EIA	Long/lat	~Decade	https://www.eia.gov/maps/layer_info-m.php
Power lines	United States	Point emissions sources	NGA	Long/lat	~Decade	https://hifld-geoplatform.opendata.arcgis.com/datasets/electric-power-transmission-lines
Railroads	United States	Point emissions sources	USDOT	Long/lat	~Decade	https://railroads.dot.gov/maps-and-data/maps-geographic-information-system/maps-geographic-information-system
Subtitle D landfills	United States	Point emissions sources	NGA	Long/lat	~Decade	https://hifld-geoplatform.opendata.arcgis.com/datasets/solid-waste-landfill-facilities
Superfund sites	United States	Point emissions sources	EPA	Long/lat	~Decade	https://catalog.data.gov/dataset/superfund-sites1e8f4
Wastewater treatment plants	United States	Point emissions sources	EPA	Long/lat	~Decade	https://hifld-geoplatform.opendata.arcgis.com/datasets/4b9bac25263047c19e617d7bd7b30701_0
Animal Feeding Operations	North Carolina	Point emissions sources	NCDEQ	Long/lat	Yearly	https://deq.nc.gov/about/divisions/water-resources/water-quality-regional-operations/afo
Dry cleaning solvent	North Carolina	Point emissions sources	NCDEQ	Long/lat	~Decade	https://deq.nc.gov/about/divisions/waste-management/dry-cleaning-solvent-cleanup-act-program
Pre-regulatory landfills	North Carolina	Point emissions sources	NCDEQ	Long/lat	~Decade	https://deq.nc.gov/about/divisions/waste-management/superfund-section/pre-regulatory-landfill-program
Hazardous waste spills	North Carolina	Point emissions sources	NCDEQ	Long/lat	~Decade	

What types of exposure would be the most relevant?

- Lifestyle
- Built environment
- Weather/atmosphere
- Stress
- Toxic or chemical exposures
- Diet and medications
- Social determinants of health
- Others?

NIEHS Acknowledgments

- PEGS Leadership
  - Janet Hall (co-PI)
  - David Fargo
  - Charles Schmitt
- GIS Modeling
  - Kyle Messier
  - Melissa Lowe
  - Kristin Eccles
- CAFOs Study
  - Fred Miller
  - Nat McNell
  - Shepherd Sherman
- Statistical Genetics
  - Farida Akhtari
  - John House
  - Eunice Lee
  - Xiaron Tong
  - Dillon Lloyd
  - Jasmine Mack
- Bioinformatics
  - Adam Burkholder
  - Jianying Li
- Data Scholar
  - Lara Clark
- Common Disease Working Group
  - Kelly Chandler
  - Kim McAllister
  - Aubrey Miller
  - Alex Merrick
  - Alison Harrill
  - David Balshaw
  - Yuxia Cui
  - Michelle Bennett
  - Cindy Lawler

Questions?

alison.motsinger-reif@nih.gov

National Institute of
Environmental Health Sciences

Peter Diggle, PhD

University Professor of Statistics

Faculty of Health and Medicine, Lancaster University

UK

Design and Analysis of Longitudinal Population Studies for Climate-Health Research

Challenges & Opportunities

Peter J Diggle

Nissan, Diggle and Fronterre, (2021). *Climate Science and Longitudinal Population Health Studies: bridging two research communities*. London: Wellcome Trust (to appear)

The Constituent Data…typical scenarios

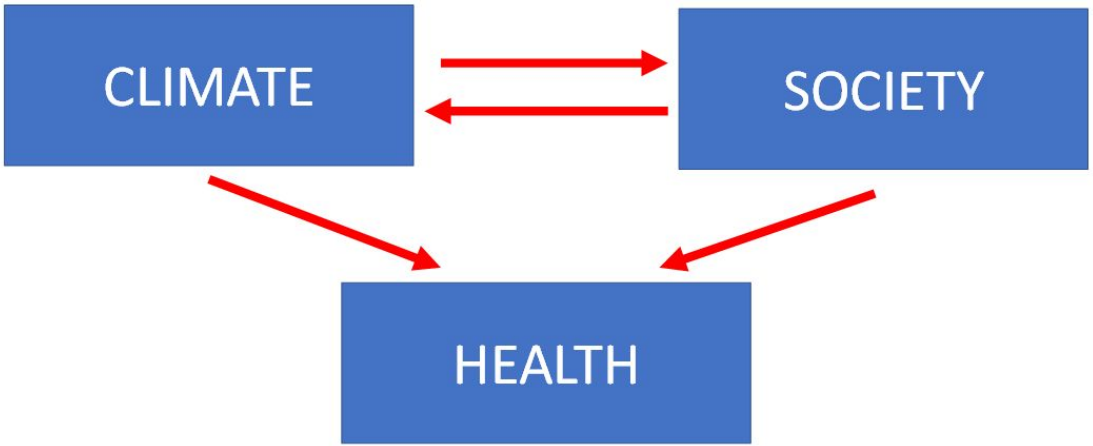
Longitudinal population studies

- Large number of individuals
- Small number of follow-up times
- Yearly (or longer) follow-up intervals
- Many variables
- Tight geographical span

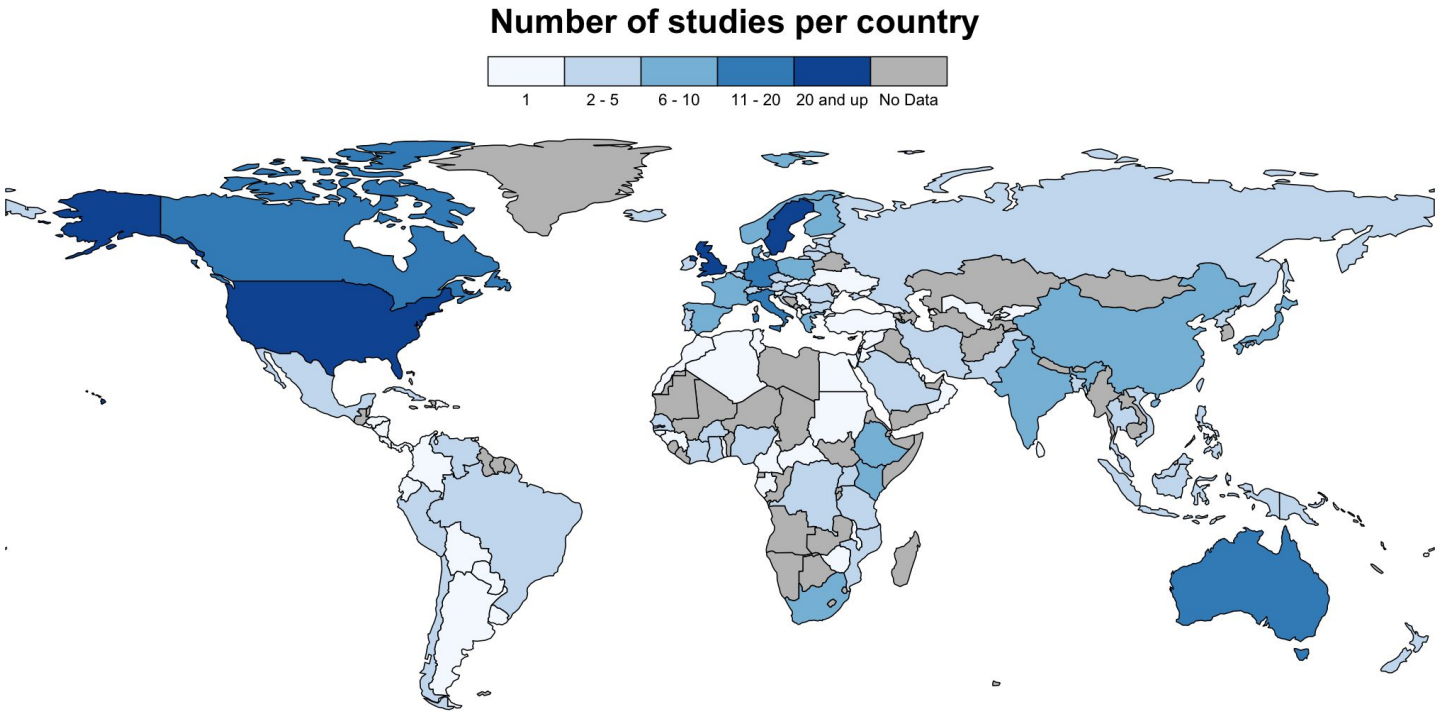
Climate data

- Network of monitoring sites
- Long time series at each site
- High-frequency data: daily, hourly,…
- Small number of variables
- Wide geographical span, but locally sparse

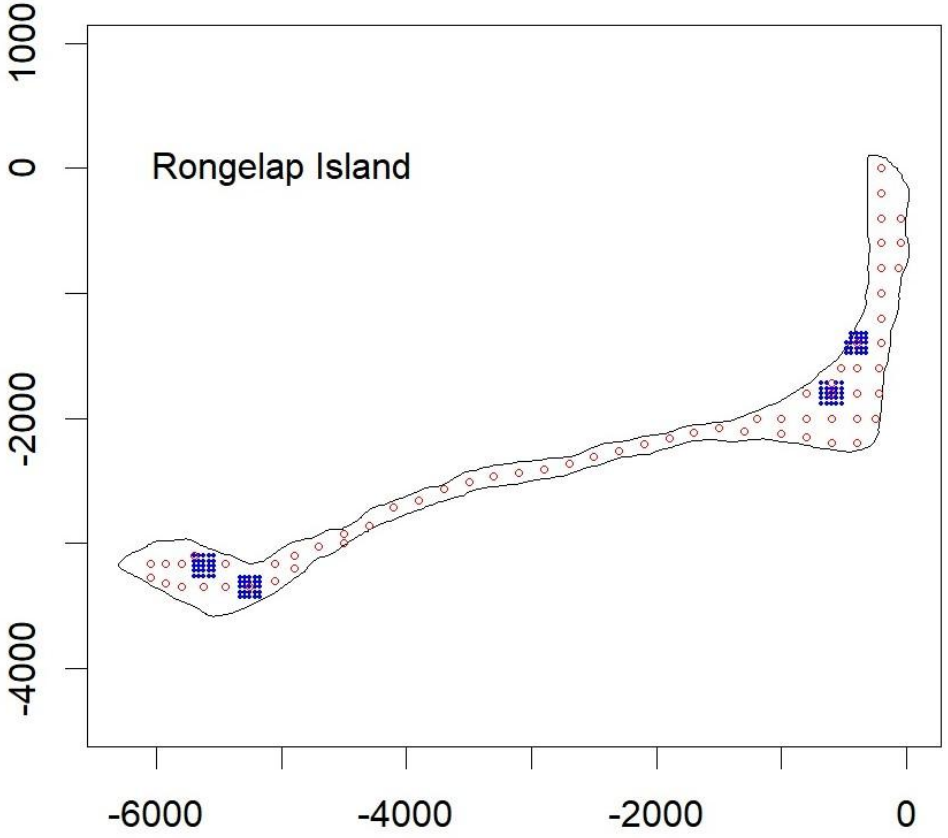
The Inter-disciplinary Challenge



The Geographical Challenge



The design challenge: multi-scale effects in time and space

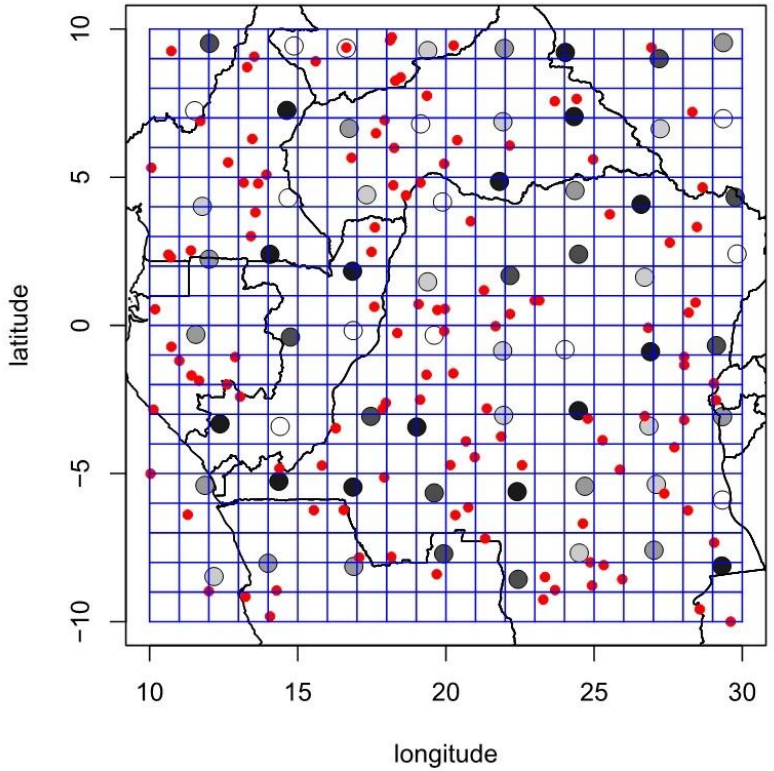


- Contaminated by nuclear testing in 1954
- Evacuated by Greenpeace in 1985
- Estimating residual contamination in 1997: what are the important spatial scales?

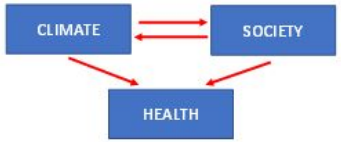
The analysis challenge: data-synthesis

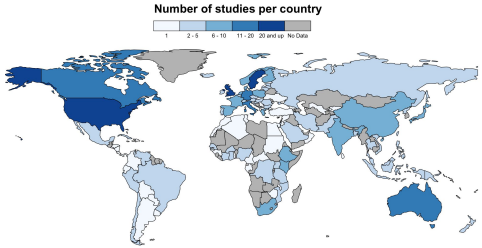
- Combining LPS data and electronic health records
- Example: Brazilian 100M study
- Exploiting spatial and temporal correlation for greater efficiency
- Combining point, area and image **data**, $D_1,\ldots,D_m$, to learn about a scientific **process**, P
- Hierarchical model structure:

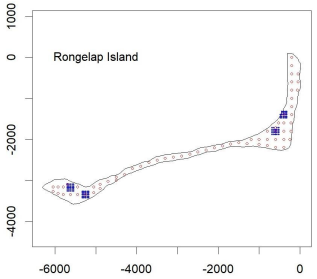
$$[P] \times [D_1|P] \times \ldots \times [D_m|P] \rightarrow [P|D_1,\ldots,D_m]$$

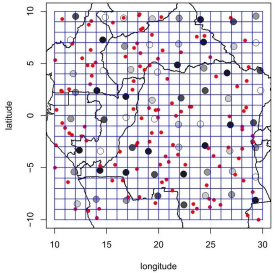


Challenges are also opportunities









Inter-disciplinarity **Geography** **Design** **Analysis**

How should Wellcome respond?

- Work with existing LPS and climate data?
- Modify existing LPS better to incorporate climate science?
- Fund new LPS with explicit health-climate focus?
- Single studies or consortia?
- A global-scale centre of excellence for climate-health research?

Our recommendations to Wellcome

- **Use existing grant and fellowship schemes** to fund:
  - secondary analyses of existing LPS and climate data
  - development of novel statistical and computational methods
  - projects to support better understanding of the indirect drivers in climate-health pathways

- **Construct a web-based platform** to visualise relevant metadata from different LPS, climate and other relevant datasets

- **Engage in discussion with Brazil 100M** and INPE with a view to developing an exemplar country-wide, real-time climate and health surveillance system.

- **Commission selected LPS consortia** to consider how they could re-orient some of their work towards climate-health research

Live Panel Discussion

Please put your questions in the chat

Returning at 13:30 UTC

Funding Opportunities/Resource Presentations

Patrick Tan, MD, PhD

Executive Director
Precision Health Research

Singapore

Session Overview

Session Topics:

- WT Discovery Research: new open funding schemes
- Introduction to Chan Zuckerberg Initiative
- NIH Funding History & Future Opportunities

Session Speakers:

- Bruna Galobardes, MBBCh, PhD, MPH
- Jonah Cool, PhD
- Rongling Li, MD, PhD, MPH

Bruna Galobardes, MD, PhD, MPH

Senior Research Manager in Discovery Research

Wellcome

UK

Discovery Research: new open funding schemes

Bruna Galobardes, MD, PhD, MPH

Senior Research Manager, Discovery Research

IHCC – 4 November 2021

Mission

We improve health for everyone by funding research, leading policy and advocacy campaigns, and building global partnerships.

We fund discovery research

We fund research across a wide range of disciplines that have the potential to make important discoveries about life, health and wellbeing.

We take on worldwide health challenges

We want to find solutions to three of the most urgent health challenges: mental health, infectious disease and climate.

New structure

Research Programmes

- **Discovery Research**
- **Health Challenges**
  - mental health
  - infectious disease
  - climate
- **Data for Science and Health**
- **Research Environment**: will address research culture, ethics and researcher-led engagement

Research Funding: manage funding operations

Discovery Research

Discovery Research: remit

- fundamental processes that underpin biology, to understand more about how life works
- complexities of human health and disease, including clinical and population-based approaches
- development of methodologies, conceptual frameworks, technologies, tools or techniques that could benefit health-related research

This includes research into the:

- needs, values and priorities of the people and communities affected by disease and health disparities
- social, cultural, political and historical contexts of human health and disease.

New funding schemes*

* Where your host organisation is based: UK, Republic of Ireland, Low- or middle-income countries (apart from India and mainland China)

	Eligibility	At a glance
Early Career Awards	This scheme provides funding for **early-career researchers** from any discipline who are ready to develop their **research identity**. By the end of the award, they will be ready to lead their own independent research programme.	Level of funding: **Salary** and **up to £400,000** for research expenses Duration of funding: **Usually 5 years**, but may be less for some disciplines, and longer if held on a part-time basis
Career Development Awards	This scheme provides funding for **mid-career researchers** from any discipline who have the potential to be **international research leaders**.	Level of funding: You **should ask for your salary (if required)** and the resources you need for your research programme Duration of funding: **Usually 8 years**, but may be less for some disciplines and longer if held on a part-time basis
Discovery Awards	This scheme provides funding for **established researchers and teams** for from any discipline	Level of funding: You should ask for the resources you need for your research programme Duration of funding: **Usually 8 years**, but may be less for some disciplines, and longer if held on a part-time basis.

Assessment

Two-stage process

1. written full application - shortlist
2. interview for shortlisted applicants

Peer assessment:

- shortlisted applications
- Technical: research proposal only

We will review

1. your research proposal
2. your skills and experience
3. your research environment

To be competitive – Research Proposal

Bold

- aims to deliver a significant shift in understanding
- provides a significant advance over existing methodologies or conceptual frameworks
- has the potential to stimulate new and innovative research

Creative

- novel approach
- it develops and tests new concepts, methods or technologies, or
- combines existing ideas and approaches in a new way

High quality

- well-designed & feasible
- clear, supported by evidence
- team has necessary expertise & skills

Questions

You can also email
grantenquiries@wellcome.org

Jonah Cool, PhD

Science Program Officer
Chan Zuckerberg Initiative

USA

Nov. 4, 2021

Introduction to Chan Zuckerberg Initiative

Jonah Cool, Ph.D
Science Program Officer
Lead, Single-Cell Biology

Chan Zuckerberg Initiative

Our Mission

Supporting the science and technology that will make it possible to cure, prevent, or manage all disease by the end of the century.

10 Year Plan

Accelerating biomedical science by developing **new tools and technologies** and supporting **open, collaborative** models of research.

Our Values

People

Technology

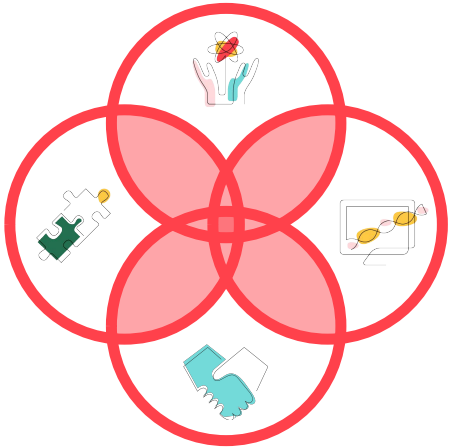
Collaboration

Open Science

Accelerating Biomedicine

We Fund
Grants and RFAs

We Collaborate



We Build
Science Tech Team

We Engage
Science in Society

Programs

Experiments in accelerating science

Building tools and resources, for and with scientists

Changing the culture of science

CZ Biohub

Imaging

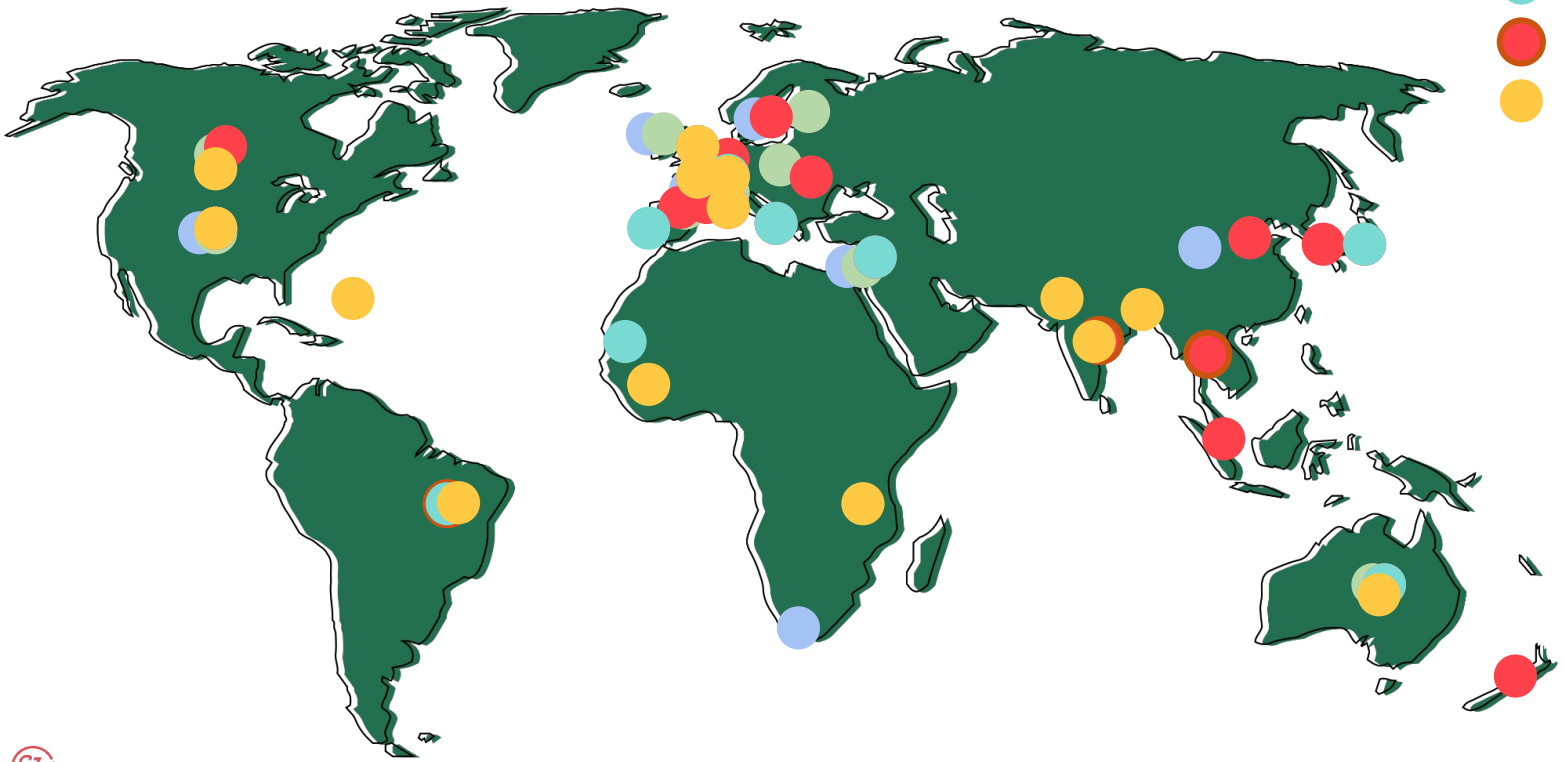
Neurodegeneration Challenge Network

Open Science

Science in Society

Single-Cell Biology

CZI Single-Cell Grantees



Thank you!

CZI-wide

https://twitter.com/ChanZuckerberg
https://www.facebook.com/chanzuckerberginitiative/
https://www.instagram.com/chanzuckerberginitiative
www.linkedin.com/company/chan-zuckerberg-initiative
https://www.youtube.com/channel/UCZioJ6fb9SuRdLIO7DlE09w
https://medium.com/czi-technology

CZI Science

https://twitter.com/cziscience
https://medium.com/@cziscience

jcool@chanzuckerberg.com
@jcoolscience

Rongling Li, MD, PhD, MPH

Program Director
Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health

USA

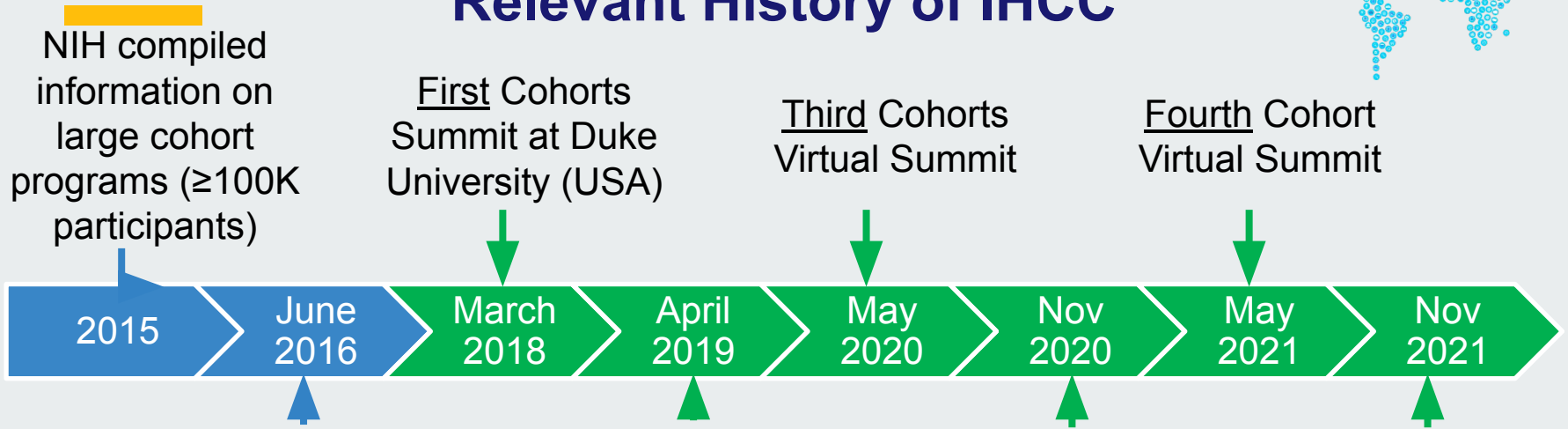
2021.11.4

NIH Funding History & Future Opportunities

Rongling Li, NHGRI, NIH

International 100K Cohort Consortium

Relevant History of IHCC



NIH National Institutes of Health
Turning Discovery Into Health

wellcome

Davos Alzheimer's Collaborative

Chan Zuckerberg Initiative

What has been supported?

- Secretariat – Coordinating the consortium activities
- Annual Summit – Bringing cohorts members together to establish collaborations, update consortium progress, propose future direction, and develop strategic plans
- Working Groups Activities
  - Data Interoperability and Connectivity – IHCC Atlas
  - Policies and Systems – Developing a policy agenda, addressing challenges and identifying common needs
  - Scientific Strategies – Stimulating research that builds upon rich and diverse datasets
  - Training, Sharing, and Capacity Development – Collaboration enhancement, training, education, and knowledge sharing

IHCC Pilot Projects

Project	PI/PM	Funders	Year
Polygenic risk scores (PRS)	H Hákonarson (USA)	NIH & Wellcome	2020
Exploring the role of genetically determined BMI in early life on colorectal cancer in later life	DJ Hughes (Ireland)	NIH & Wellcome	2021
High-Throughput Metabolomic Biomarker Measures in Diverse Ancestries	H Hákonarson (USA)	NIH & Wellcome	2021
Effects of regular opioid use on mortality and on cancer development (Opioid cohort consortium)	P Brennan (France)	NIH & Wellcome	2021
Global Mental Health Impact of the COVID-19 Pandemic	J Smoller (USA), S Bauermeister (UK) & A Brunoni (Brazil)	NIH & Wellcome	2021
Novel coronavirus host susceptibility study in South Africa (COVIGen-SA)	M Ramsay (S. Africa)	NIH & Wellcome	2021
Biospecimen collection for Global Longitudinal Population Studies in the COVID-19 era	J Chambers (Singapore)	Chan Zuckerberg Initiative (CZI)	2021
Davos Alzheimer's Collaborative (DAC) – Pilot PRS	Davos Alzheimer's Collaborative – Pilot	DAC	2021

NIH Policy for Foreign Applications

- NIH is part of the United States Department of Health & Human Services, which is the primary U.S. Federal agency that conducts and supports medical research
- The 27 Institutes and Centers (ICs) of NIH provide leadership and financial support to researchers both inside and outside the United States
- Detailed information for submitting foreign grants can be found at https://grants.nih.gov/grants/foreign/index.htm
- All applications must undergo peer review

Additional criteria for reviewing foreign applications

- Whether the project presents special opportunities for furthering research programs through the use of unusual talent, resources, populations, or environmental conditions in other countries that are not readily available in the United States or that augment existing U.S. resources.
- Whether the proposed project has specific relevance to the mission and objectives of the IC and has the potential for significantly advancing the health sciences in the United States
- Foreign applicants can learn more at the Information for Foreign Applicants and Grantees page.

How to Write an NIH Application

- **Write Your Application**
  https://grants.nih.gov/grants/how-to-apply-application-guide/format-and-write/write-your-application.htm
- **Important Writing Tips:**
  1. Make Your Project's Goals Realistic
  2. Be Organized and Logical
  3. Write in Clear Concise Language
  4. Sell Your Idea on Paper
  5. Edit Yourself, but also Enlist Help
  6. Share for Comments
- **Sample Applications**
  https://grants.nih.gov/grants/how-to-apply-application-guide/resources/sample-applications.htm
- **YouTube Videos (example** Best Practices for Junior Investigators When Writing an NIH ... **)**

Thank You!

email: lir2@mail.nih.gov
https://www.genome.gov/staff/Rongling-Li-MD-PhD-MPH

Live Panel Discussion

Please put your questions in the chat

Workshop Summary

Geoffrey Ginsburg, MD, PhD

Director
Duke University, Center for Applied Genomics

Co-Chair
International HundredK+ Cohorts Consortium

USA

Thank you!


• A global community of cohorts working together to advance science and improve health for all •
INTERNATIONAL HUNDREDK+ COHORTS CONSORTIUM
INTENTIONAL
DIVERSITY
AUDACITY
INTEGRITY
EQUITY