American Society of Human Genetics (ASHG) Virtual 2021 Conference

All of Us Research Program Invited Workshop Session

Friday, October 22, 2021

3:30-5:00 p.m. ET

Global Resources for Precision Medicine

Featured Speakers:

Joshua C. Denny, MD, MS
All of Us Research Program

Latrice Landry, PhD
Harvard Medical School and Harvard School of Public Health

Naomi Allen, MD, PhD
UK Biobank

Laura Lyman Rodriguez, PhD
Patient-Centered Outcomes Research Institute (PCORI)

Heidi Rehm, PhD
Massachusetts General Hospital and Broad Institute of MIT and Harvard

Global Resources for Precision Medicine

NIH *All of Us* Research Program: Diversity and Scale in Precision Medicine
Joshua C. Denny, **MD, MS**

UK Biobank: 2021 and Beyond
Naomi Allen, MD, PhD

The IHCC Experience Bringing Cohorts Data Together to Advance Precision Health Research Around the Globe
Laura Lyman Rodriguez, PhD

GA4GH Standards to Enable Global Access and Interoperability of Data to Inform Precision Health
Heidi Rehm, PhD

Q&A Segment
MODERATORS: Joshua Denny and Latrice Landry

Disclosure Slide

Financial Disclosure for
Latrice Landry, PhD

I have nothing to disclose

NIH's *All of Us* Research Program: Diversity and Scale in Precision Medicine Research

October 22, 2021

Joshua C. Denny, MD, MS
Chief Executive Officer
All of Us Research Program

@AllofUsCEO

National Institutes of Health

Disclosures

Vanderbilt University Medical Center, my former employer, licensed PheWAS running within Vanderbilt's DNA biobank to Nashville Biosciences. I receive a portion of those royalty payments.

All of Us Research Program Mission

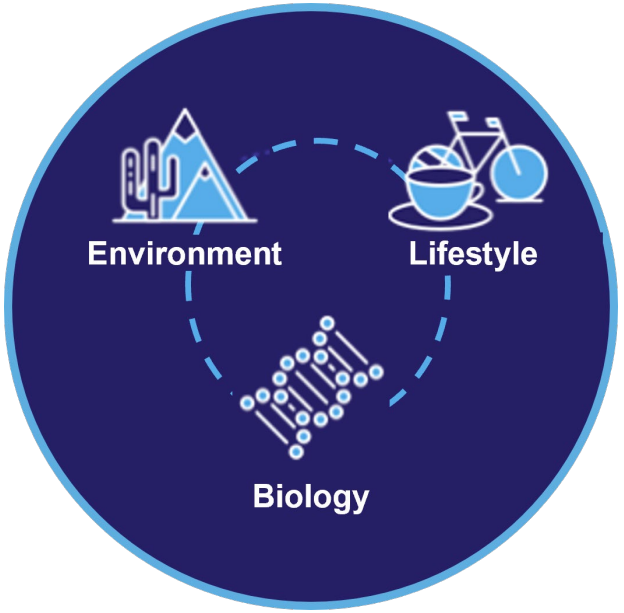
Nurture relationships
with **one million or more** participant partners, from all walks of life, for decades

Deliver one of the largest, richest biomedical dataset
that is secure and easy to access

Catalyze a robust ecosystem
of researchers and funders hungry to use and support it

The *All of Us* Research Program: An Innovative Research Effort

- **Diversity at the scale of one million people or more**
- **Focus on participants as partners,** with return of value as a priority
- **Longitudinal design** and ability to recontact participants
- **Multiple data types:** EHR, surveys, baseline physical measurements, biospecimens, genomics, and more
- **National, open resource for all:** broadly accessible to all researchers with open-source software and tools
- **Security and privacy safeguards** for all participants



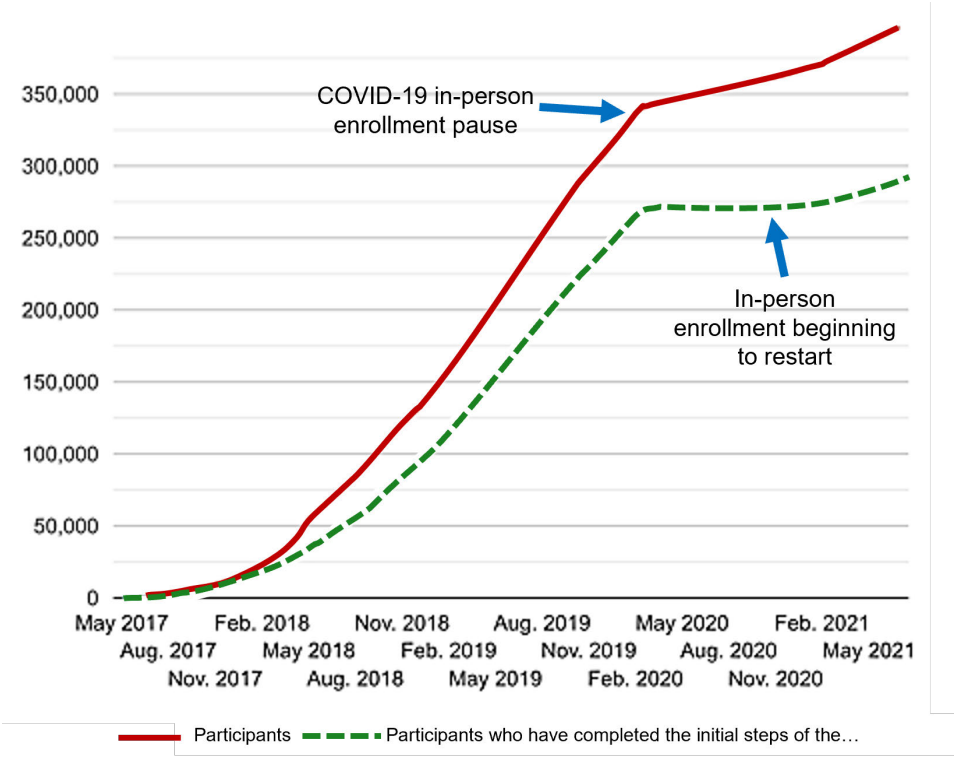
Status of the *All of Us* Research Program (as of September 21, 2021)

411,000+
Participants

253,000+
Electronic Health Records

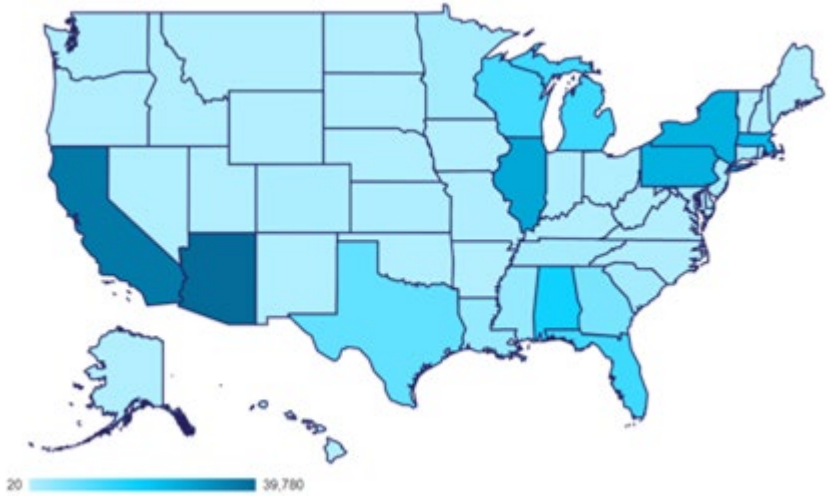
297,000+
Participants who have completed initial steps of the program

313,000+
Biosamples

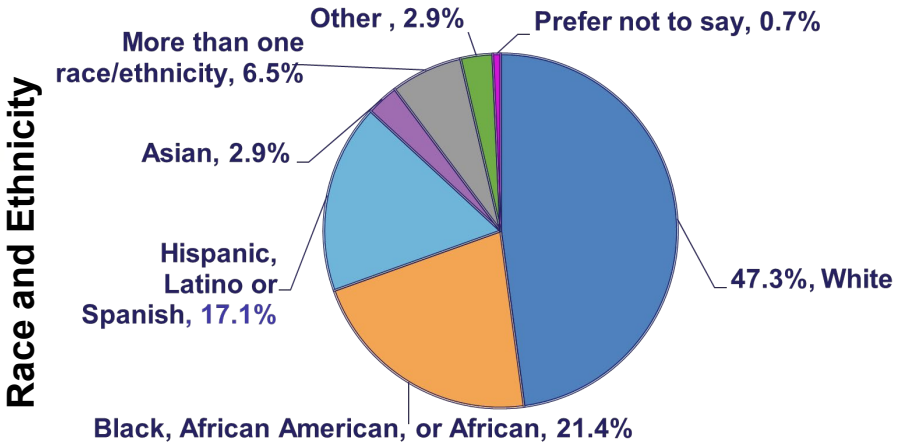


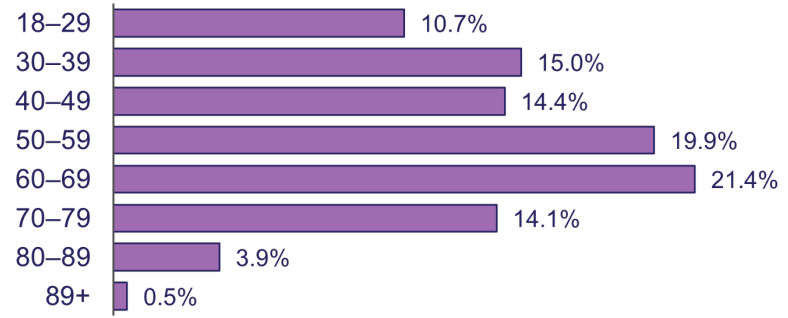
researchallofus.org

Status of the *All of Us* Research Program (as of September 21, 2021)



More than 80% of *All of Us* participants are underrepresented in biomedical research





researchallofus.org

All of Us Research Program Core Values
Return of Information

Participation is **open** to all.

Participants have **access** to their information.

Participants reflect the rich **diversity** of the U.S.

Data will be accessed **broadly** for research purposes.

Participants are **partners**.

Security and privacy will be of highest importance.

Trust will be earned through **transparency**.

The program will be a catalyst for **positive change** in research.

Returning Value for Participants: Genetic Information

Non-Health Genetic Traits

Traits

Genetic Ancestry

Currently Returning to Participants

>59,000 participants have viewed traits or ancestry results

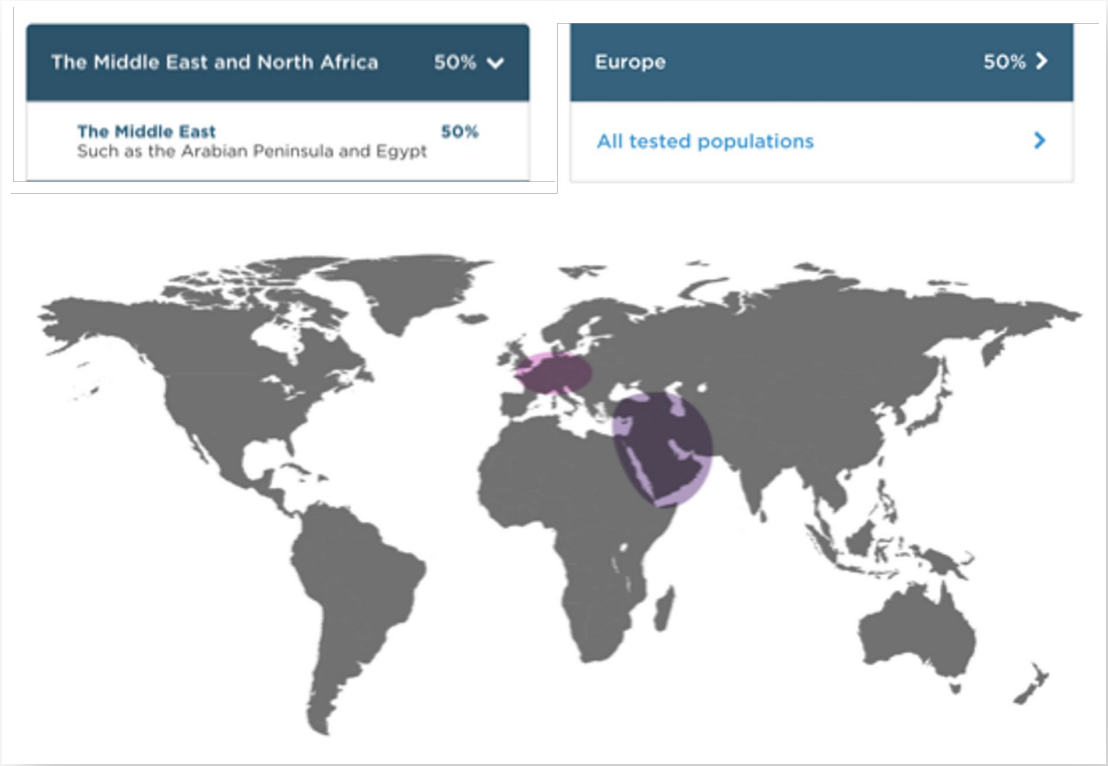
Health-Related Genetic Traits

Hereditary Disease Risk (ACMG59)

Medicine and Your Health (Pharmacogenomics)

Launching in Early 2022

Genetic Ancestry Results



The Middle East and North Africa

This genetic group represents people from these areas:

- The Middle East
- North Africa
- Western Africa
- The Caucasus

Connections near and far

People with recent ancestors from Asia, Europe, and sub-Saharan African may have patterns of DNA from this genetic ancestry group. This is likely because of significant trade and migration through the region that continues to this day. The Silk Road and Incense Route connected the Middle East and North Africa to Europe and Asia. Trans-Siberian trade routes connected North Africa to sub-Saharan Africa.

Trait Results

Ancestry

Genetic Ancestry
Genetic ancestry can be very interesting, but you may also learn information you didn't expect. Learn more

Get Started

Traits

Bitter taste perception
Learn what your genes can tell you about your ability to taste bitter things.

View

Cilantro preference
Smell and taste work together to influence your cilantro preference.

View

Earwax type
Flaky or sticky? Earwax type is encoded in your genes.

View

Lactose intolerance
Your genes code for lactase, which helps you digest milk.

View

Cilantro preference

Some people like the taste of cilantro and others think it tastes like soap.

What we looked at and why

We looked at a place in your DNA that influences if you have a slightly higher chance of liking or disliking cilantro.[1] The percent of people across the world who dislike cilantro ranges from 3-21%.[2]

- **People who have slightly higher chances of liking cilantro** may find it fragrant and citrusy.
- **People who have slightly higher chances of disliking cilantro** may find it soapy or moldy.

This place in your DNA only predicts a small amount of your chances of liking or disliking cilantro. Environmental and other genetic factors also play a role.

DNA Marker*	Gene	Your result*
rs72921001	*Near OR6A2*	C A

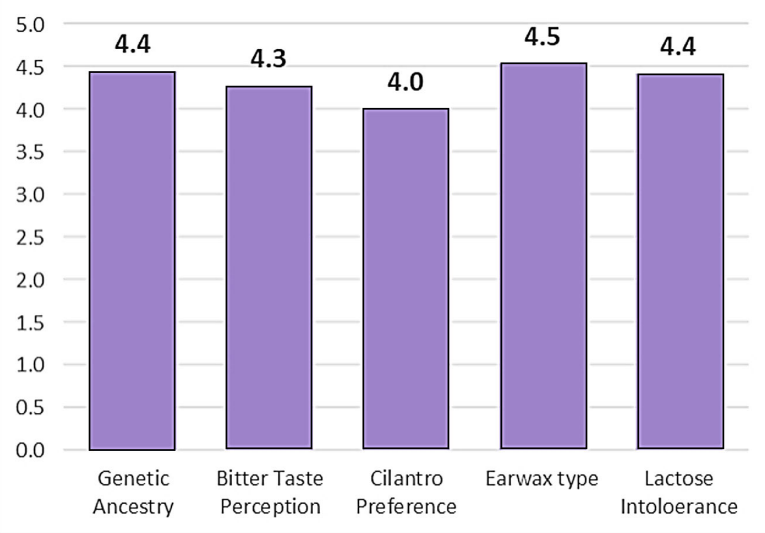
* Each of your parents provides you with a nucleotide at this position, but we don't know which parent gave you which nucleotide.

Scientific details

OR6A2 makes a sensor in the nose that helps us perceive smells. Changes near *OR6A2* may impact whether you find cilantro fragrant and citrusy, or soapy or moldy.[1]

Genetic Ancestry and Traits: Preliminary Participant Satisfaction Survey Results

How satisfied were you with your [genetic ancestry, bitter taste perception, cilantro preference, earwax type, lactose intolerance] results?

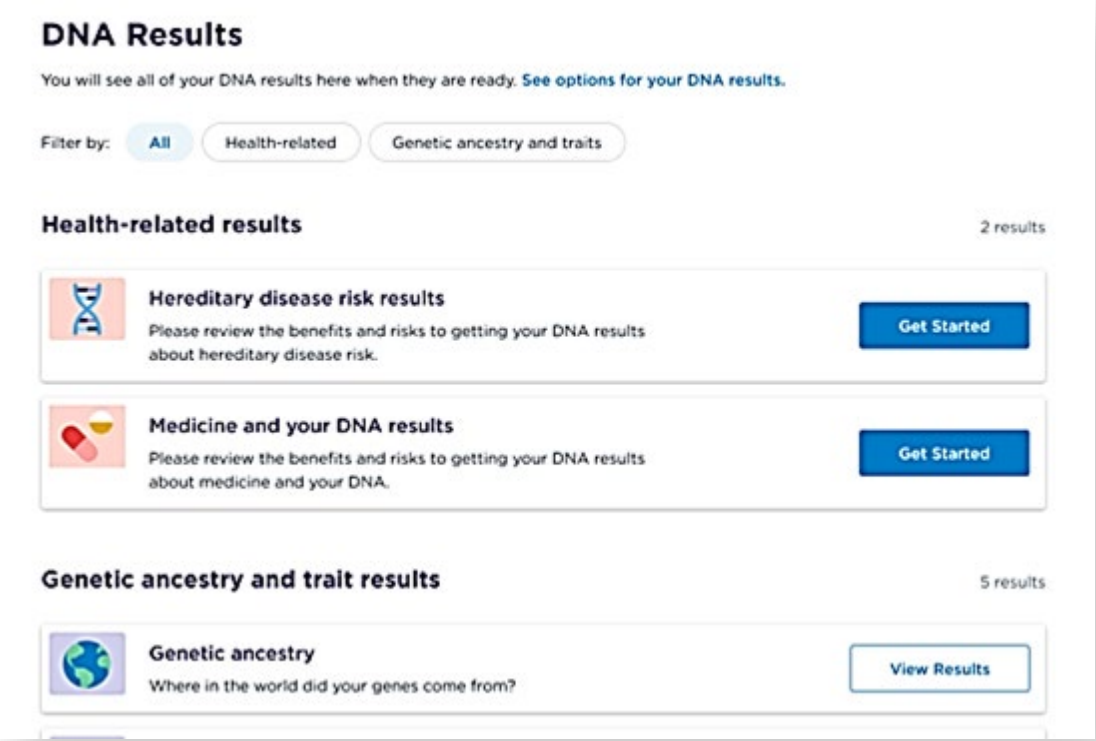


Average Response Scale of 1–5 (Very Satisfied = 5)

Was there **anything else you were hoping to learn about** [genetic ancestry, bitter taste perception, cilantro preference, earwax type, lactose intolerance]?

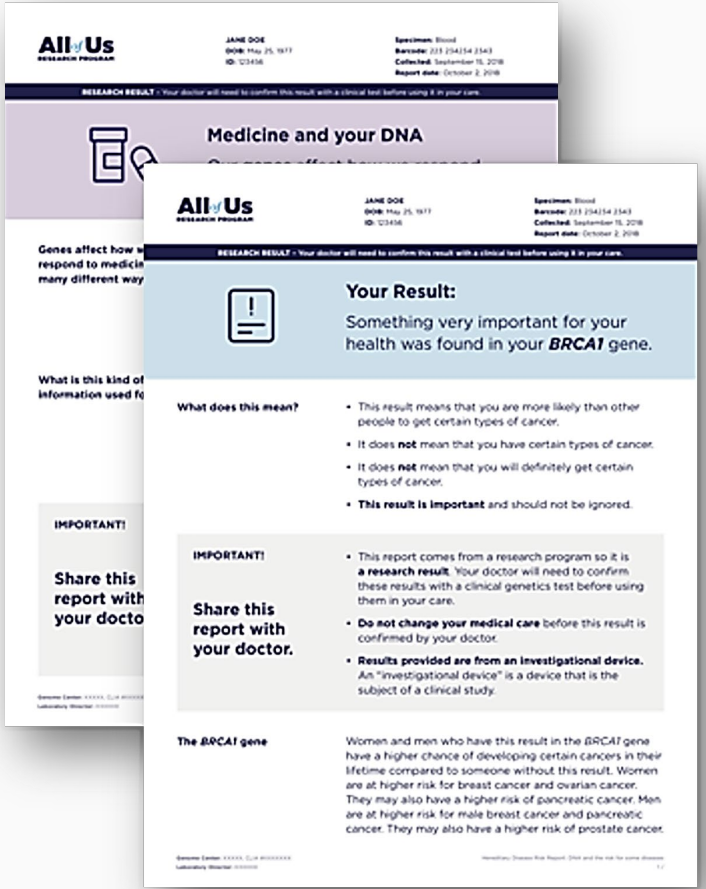
More DNA Information	• "Again, I know there is more so I'm anxiously waiting." • "Very interesting! I hope to learn more tendencies or traits." • "I welcome and look forward to any further information on my DNA."
Other	• "Just that I think this is totally amazing!!" • "This whole study is interesting, and not about stuff I would have ever have thought to be significant"

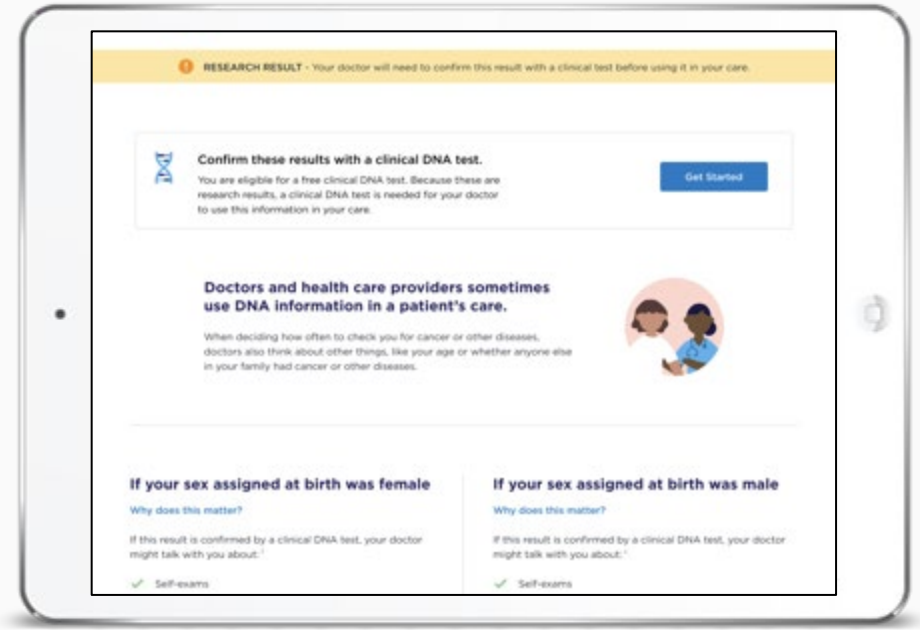
Coming in 2022: Health-Related Genetic Return of Results



- Hereditary disease risk (ACMG59) and medicine and your DNA (pharmacogenomics)
- Participants can choose results they want
- Hereditary disease risk and medicine and your DNA interpretation begin at Clinical Validation Laboratories

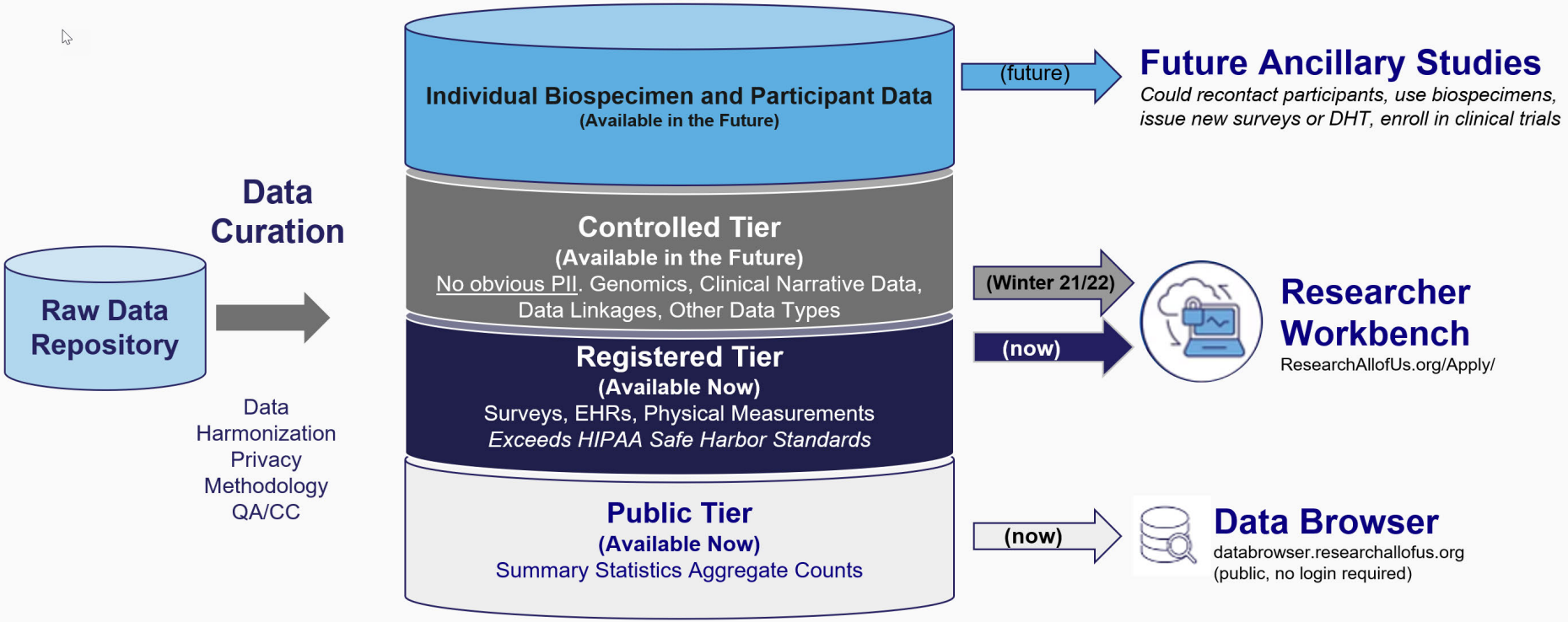
Deliver an End-to-End Genetics Experience



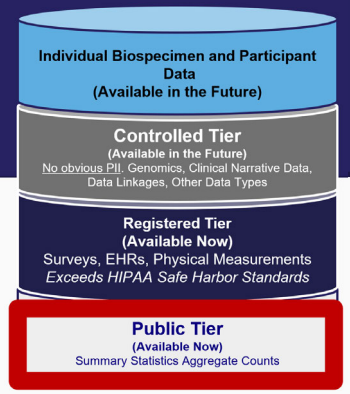


All Hereditary Disease Risk results will be returned through a genetic counselor, and we will offer confirmatory clinical genetic testing.

Researcher Data Access

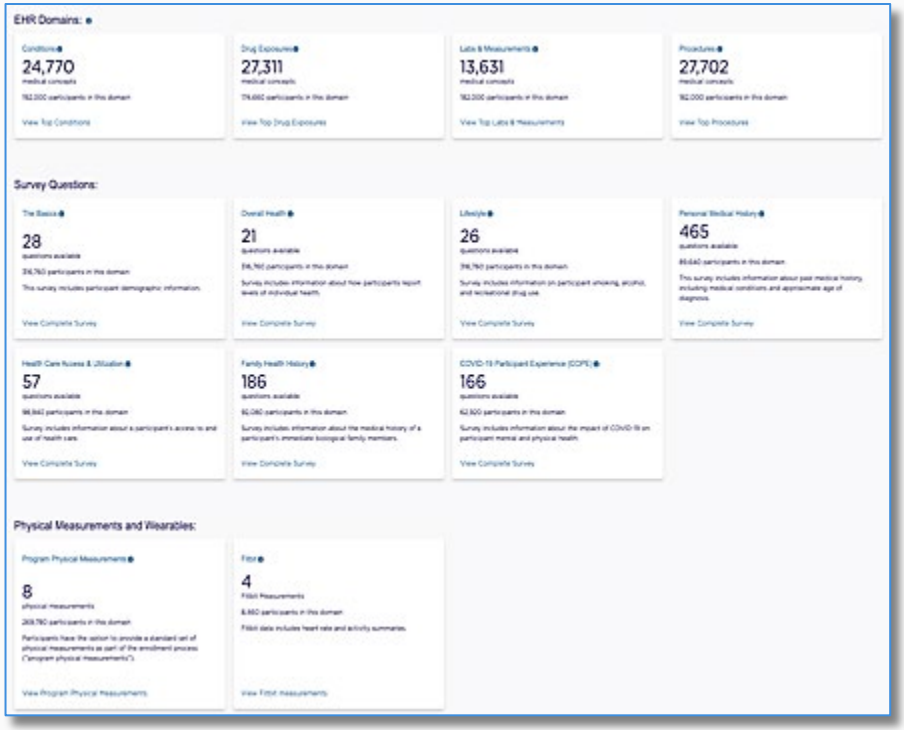


All of Us Research Hub: Public Data Browser



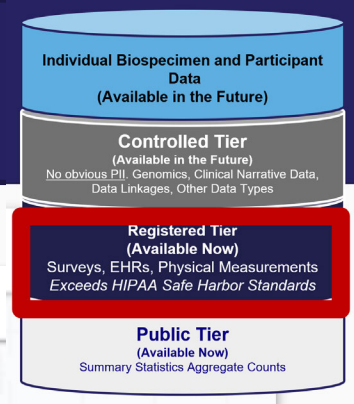
Summary statistics of participant data

- **EHR data** (conditions, drug exposures, lab and measurements, procedures)
- **Survey questions** (including COVID-19 surveys)
- **Physical measurements**
- **Open access** (no login required)



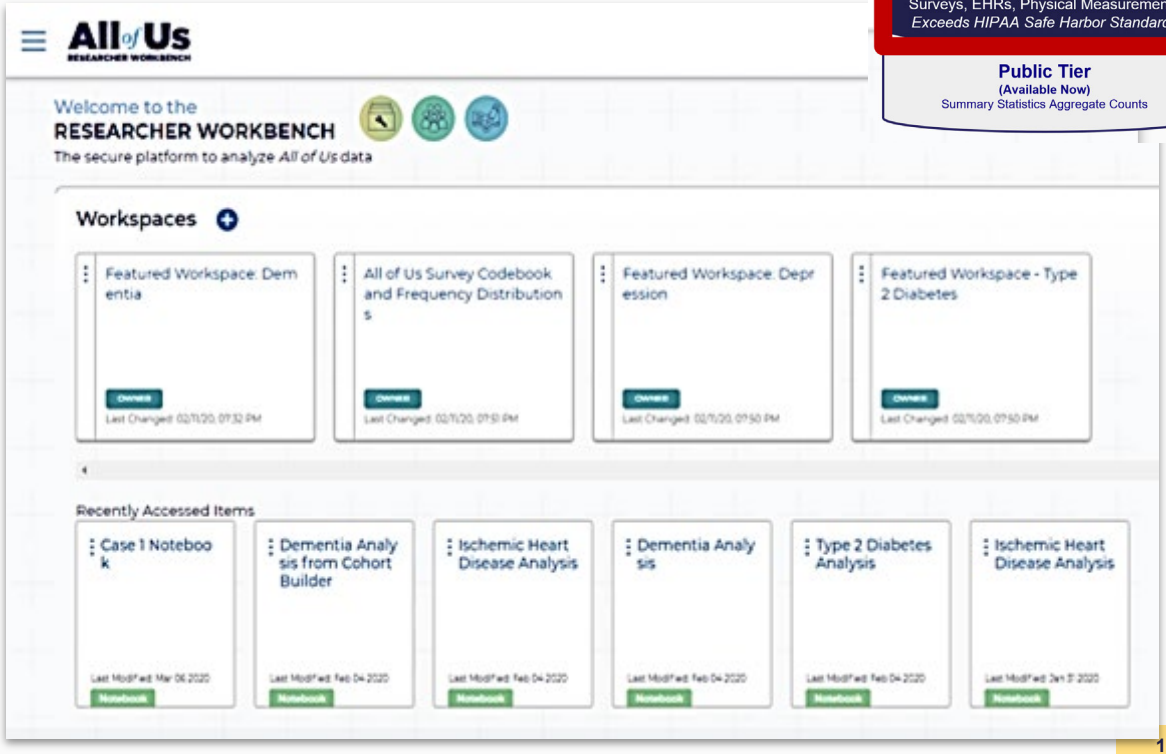
DataBrowser.ResearchAllofUs.org

All of Us Researcher Workbench: Access to Row-Level Data for Analysis



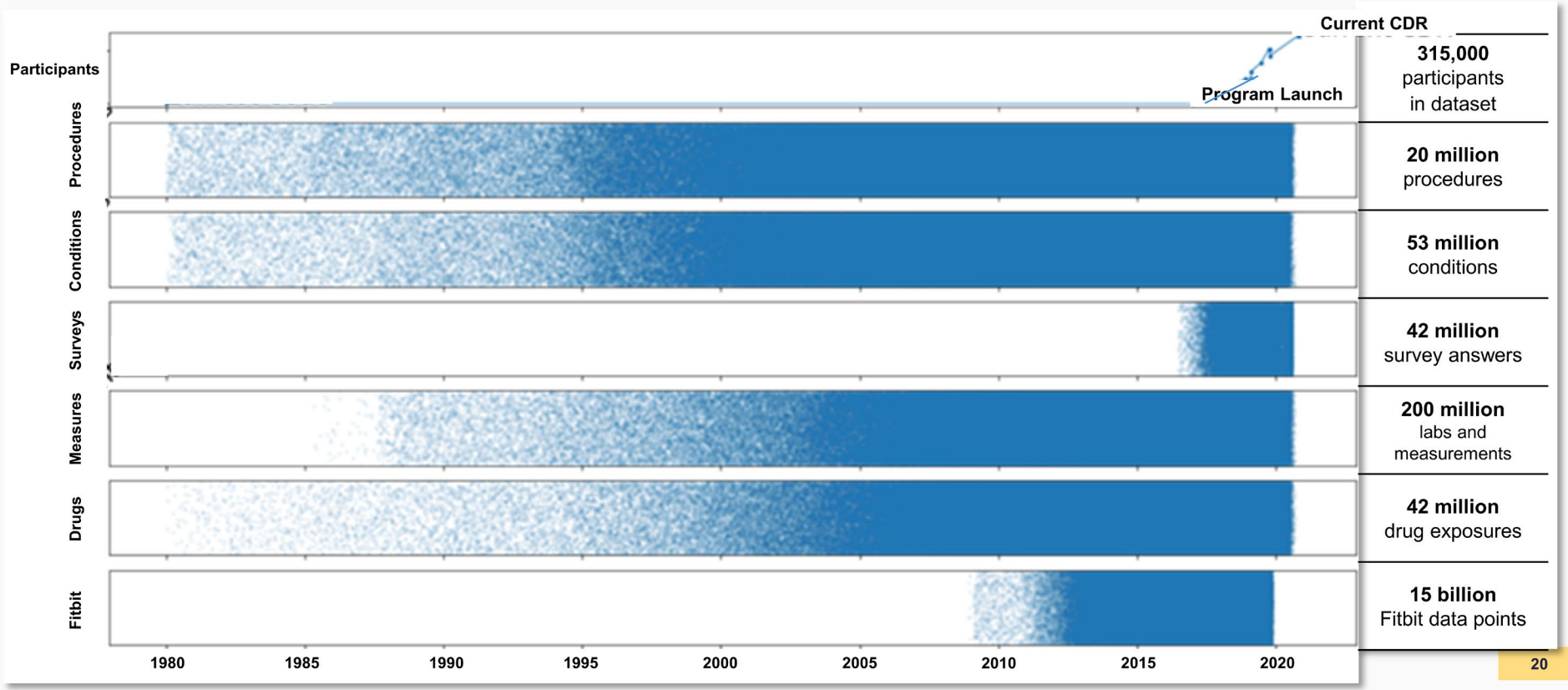
Researcher Workbench Beta
Launched on May 27, 2020

- Cloud-based central resource
- Personally identifiable information is removed
- **Passport access model**—just create, describe your workspace, and get to work! No separate IRB approval needed
- During beta phase, access requires eRA Commons ID and limited to U.S. nonprofits

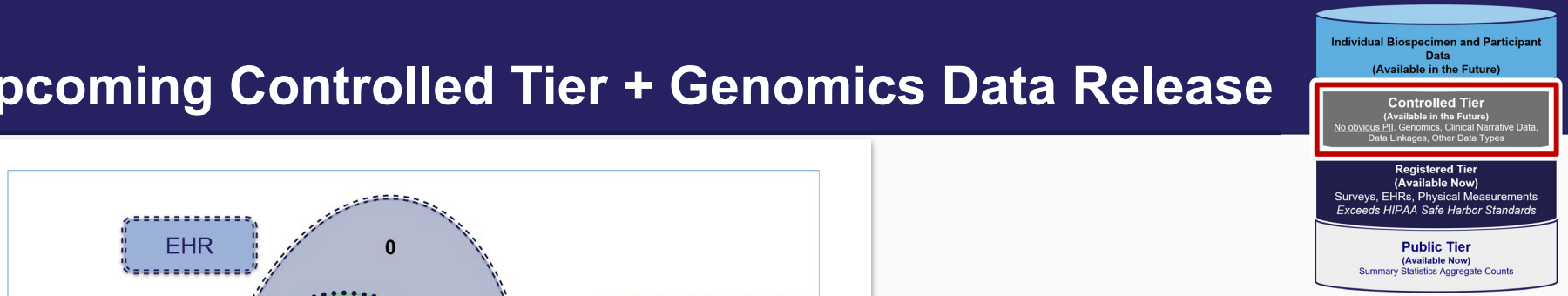


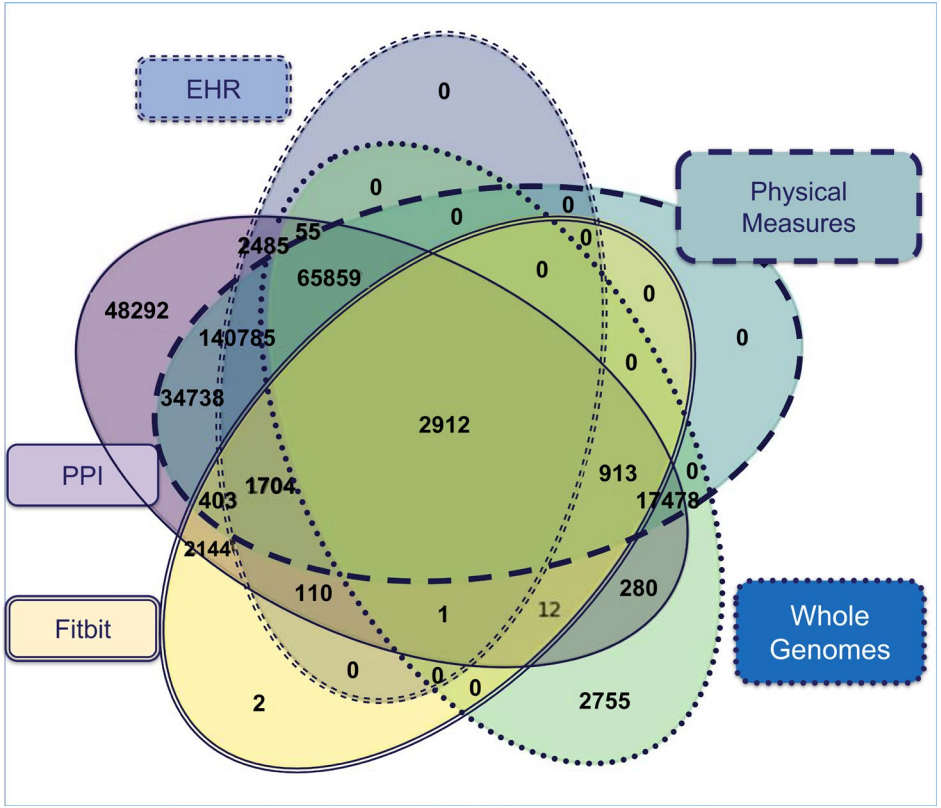
ResearchAllofUs.org

Participant EHRs and Fitbit Provide Longitudinal Data



Upcoming Controlled Tier + Genomics Data Release

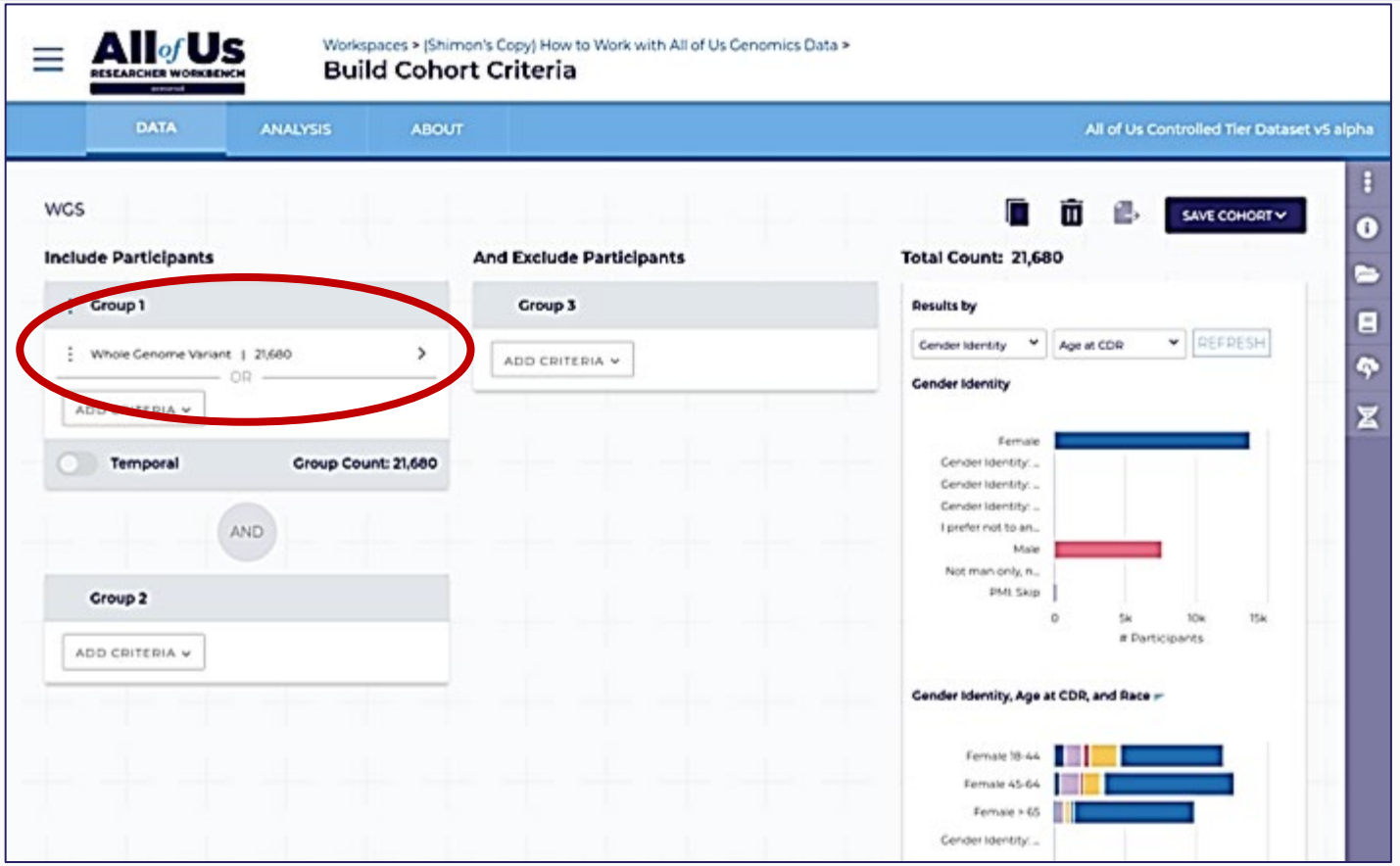




Coming Winter 2021/2022

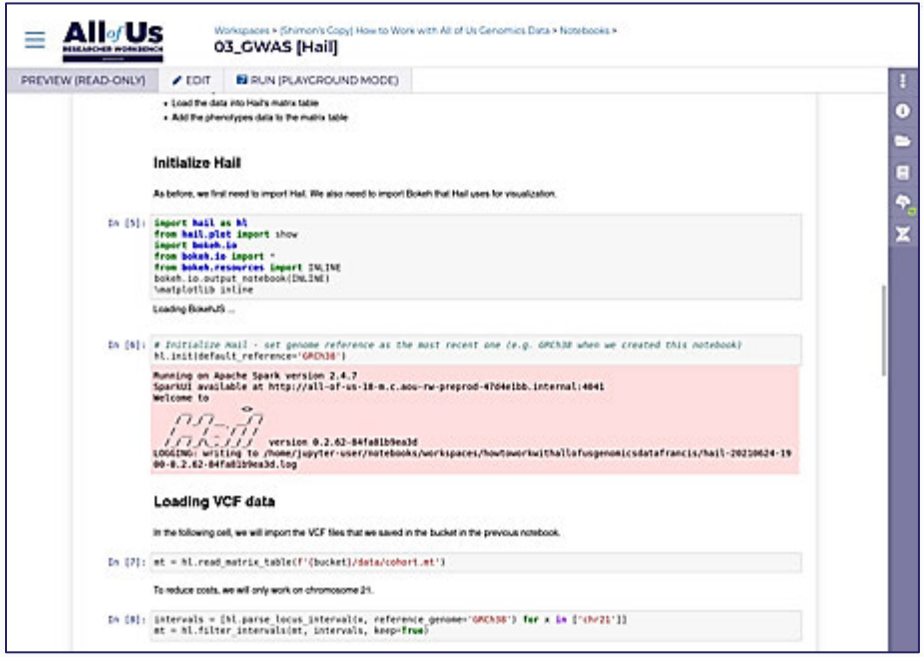
- Expected 90,000 WGS + 120,000 arrays
- More participants
- COVID diagnoses and surveys
- More detailed demographic data
- More Fitbit data

Genomics-Enabled Cohort Builder (in Alpha Testing)

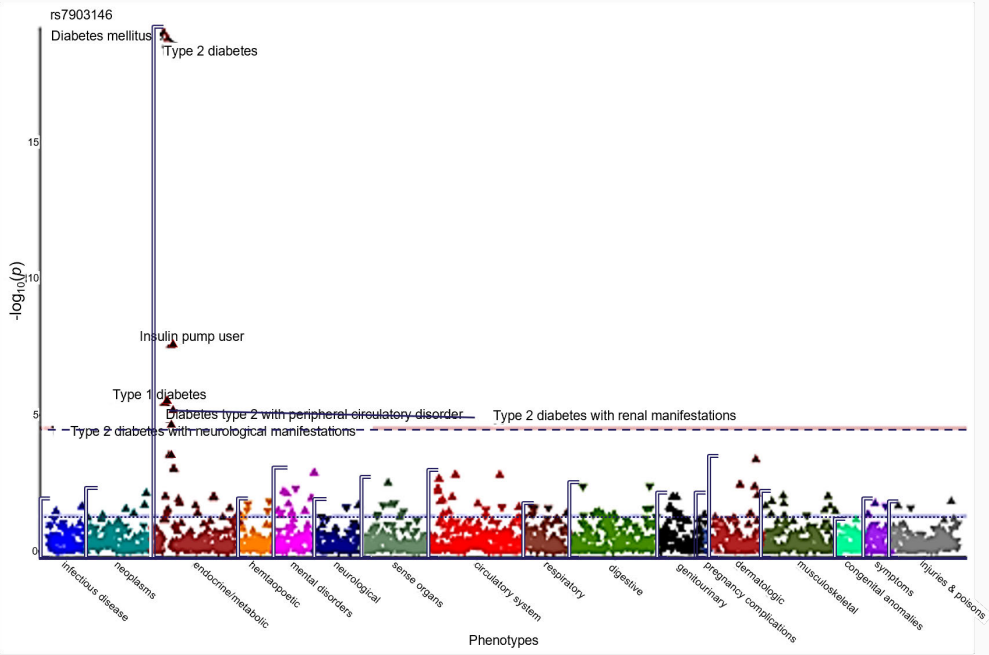


In Alpha Testing Now With 20K WGS Subset

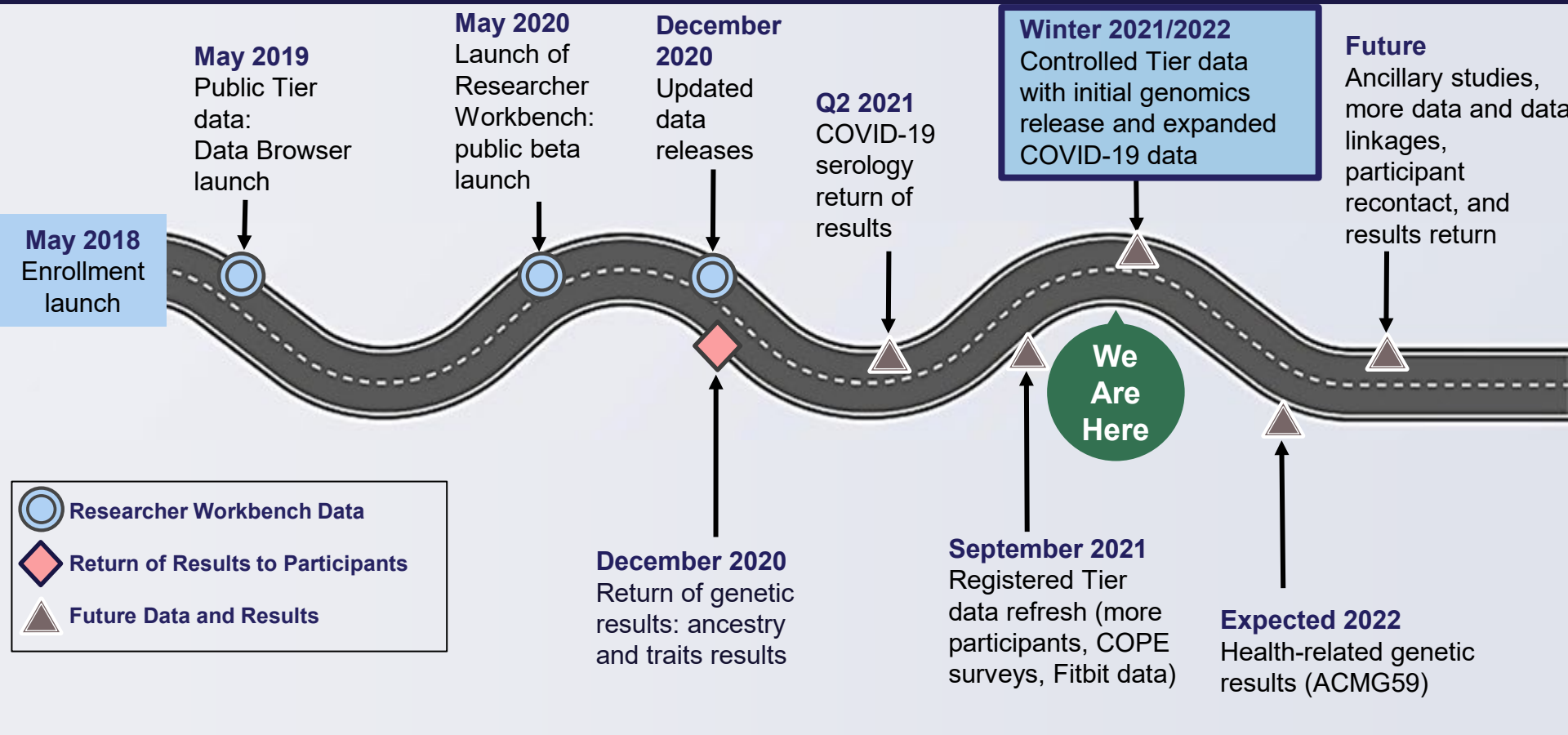
Will support standard tools like PLINK and Hail



PheWAS of rs7903146 (*TCF7L2*)



All of Us Roadmap



Thank You!

ResearchAllofUs.org

National Institutes of Health

AllofUs.nih.gov

@AllofUsResearch
@AllofUsCEO
#JoinAllofUs

biobank uk

Enabling scientific discoveries that improve human health

UK Biobank: 2021 and Beyond

Naomi Allen, MD, PhD
UK Biobank Chief Scientist
University of Oxford

@UK_Biobank

Disclosure Slide

Financial Disclosure for
Naomi Allen, MD, PhD

I have nothing to disclose

UK Biobank: 2021 and Beyond

- Enabling scientific discoveries that improve human health
- Further increasing the breadth and depth of participant characterisation
- Democratising access to the UK Biobank resource

Overview of UK Biobank Recruitment

UK Biobank has a **unique** combination of:

- Large-scale data
  - 500,000 participants
  - 40–69 years
  - Recruited 2006–2010
- Deep characterization (lifestyle, physical measures, etc.)
- Data highly standardized and curated
- Biological samples (blood, urine, saliva)
- Linkage to electronic health records over time (now has 10 years of follow-up)

- **Readily accessible to researchers worldwide**

The Importance of Genomic Data

- Custom-built genome-wide array (850,000 variants) with imputation to 90,000,000+ variants; released in 2017

- These data are transforming population-based genetic research
  - GWAS of thousands of traits now publicly available (e.g., Stanford, Broad, Edinburgh)
  - Made possible through:
    - Large size
    - Standardized measurement
    - Linkage to health records over time

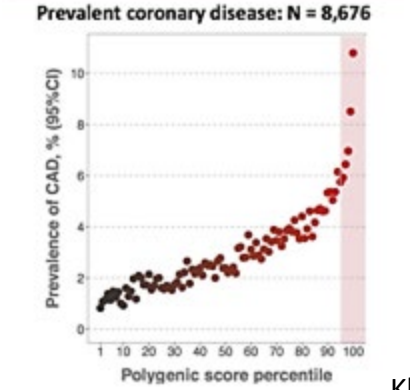
Development of Polygenic Risk Scores

Creation of **polygenic risk scores**

E.g., PRS for heart disease:

- Risk in top 5% equivalent to many monogenic disorders
- Independent of known risk factors
- Modifiable by lifestyle and/or medication
- Clinical utility to be established

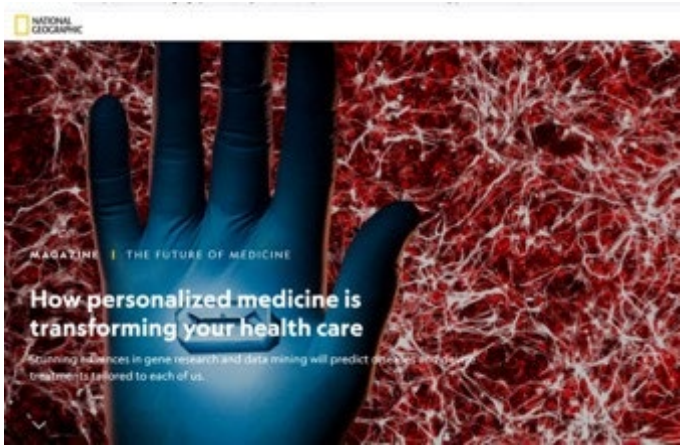
Results: polygenic score in the general population
UK Biobank: N=288,890



High polygenic score definition	Odds ratio
Top 5%	3.3
Top 2.5%	4.0
Top 1%	4.7
Top 025%	6.3

Khera et al., *Nat Genet*. 2018;50:1219–1224.

- Much we still don't understand about genetic variation (in coding and noncoding regions) on health



Whole Exome Sequencing

- Coding regions (~2% of genome)
- Regeneron-led commercial funding
- Timeline of data availability:
  - 50,000 in March 2019
  - 200,000 in Oct. 2020
  - 300,000 in Sept. 2021
  - **Full cohort end: Oct. 2021**

"This precompetitive collaboration has further strengthened the ties between academia and industry and provided teams an unprecedented opportunity to interact with and learn from the wider research community."

Advancing Human Genetics Research and Drug Discovery through Exome Sequencing of the UK Biobank

Joseph D. Szustakowski, Suganthi Balasubramanian, Ariella Sasson, Shareef Khalid, Paola G. Bronson, Erika Kvikstad, Emily Wong, Daren Liu, J. Wade Davis, Carolina Haefliger, A. Katrina Loomis, Rajesh Mikkilineni, Hyun Ji Noh, Samir Wadhawan, Xiaodong Bai, Alicia Hawes, Olga Krasheninina, Ricardo Ulloa, Alex Lopez, Erin N. Smith, Jeff Waring, Christopher D. Whelan, Ellen A. Tsai, John Overton, William Salerno, Howard Jacob, Sandor Szalma, Heiko Runz, Greg Hinkle, Paul Nioi, Slavé Petrovski, Melissa R. Miller, Aris Baras, Lyndon Mitnaul, Jeffrey G. Reid

Consortium:

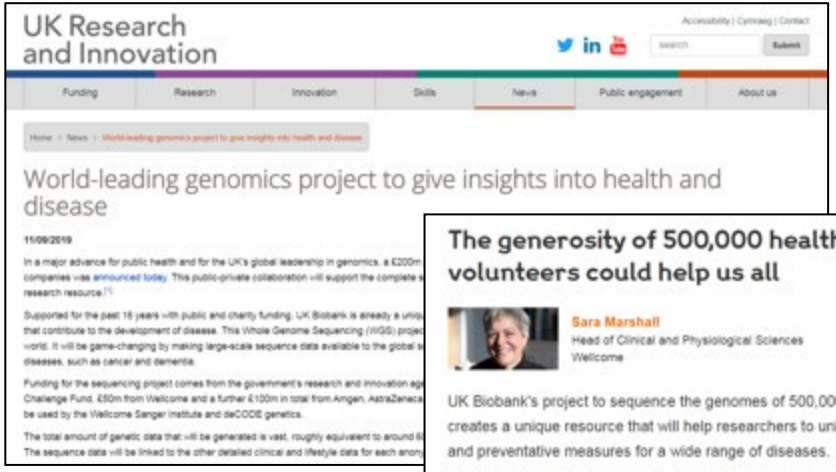
Regeneron	**Biogen**
GSK	**Alnylam**
Pfizer	**Takeda**
Bristol Myers Squibb	**AstraZeneca**

Whole Genome Sequencing

- Sequencing the entire genome
- Government, charity, and industry funding
- Sequencing by Sanger and deCODE
- Timeline of data availability:
  - 200,000 in Nov. 2021
  - Full cohort Q1 2023

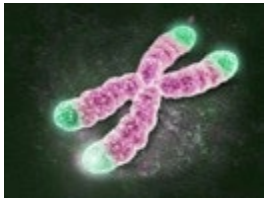
"…the most ambitious sequencing effort of whole human genomes ever undertaken…"

Consortium:
Wellcome Trust
UKRI
Amgen
AstraZeneca
Janssen
GSK



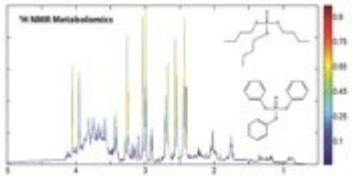
What Next? Moving Beyond Genomics

- Telomere length in all 500,000
  – Data available March 2021



University of Leicester

- NMR-metabolomics in all 500,000
  – Data available for 120,000 March 2021



Nightingale Health

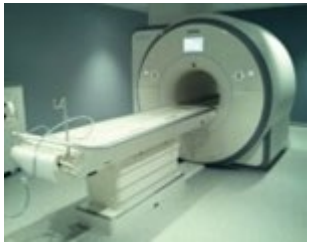
- Proteomics in 57,000 samples (initially)
  – Olink platform (3,000 proteins)
  – Data available 2022

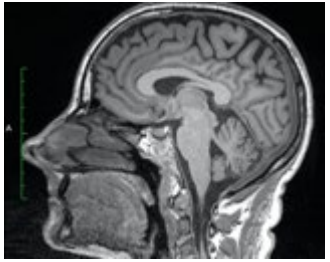
Consortium:
AstraZeneca
Biogen
Bristol Myers Squibb
Calico
Genentech
GSK
Janssen
Novo Nordisk
Pfizer
Regeneron
Takeda

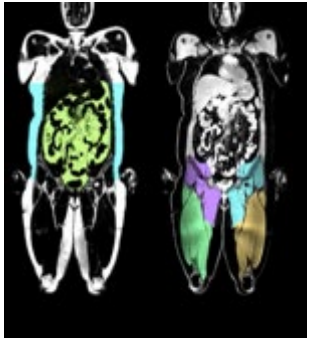
- Cohort-wide assays more flexible than case-control comparisons

Enhanced Participant Phenotyping: Imaging

- Largest study in the world (by far) to undertake population-based multimodal imaging
  - MRI—brain, heart, body
  - Carotid ultrasound—measures of large artery
  - DXA—low-dose x-ray of bones and joints
  - 12-lead ECG
  - Cardiac monitor (subset)
- 50,000 of 100,000 participants imaged
- Repeat imaging of up to 70,000 participants to enable assessment of longitudinal change in imaging phenotypes
- Enables research into disease pathways and early detection

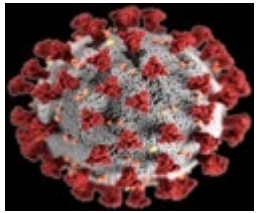


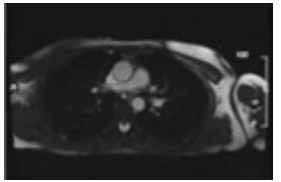




SARS-CoV-2 Imaging Study

- Background: Emerging data suggests that multi-organ injury with SARS-CoV-2 infection is common and associated with medium- to long-term consequences
- Aim: to generate unique data to assess effect of SARS-CoV-2 on changes in internal organs by performing imaging scans on individuals before and after infection
- 50,000 participants imaged before the pandemic received a home-based SARS-CoV-2 antibody lateral flow test
- Invited 2,000 individuals (half seropositive, half seronegative) for a second imaging assessment
- Only study in the world with both pre- and post-infection imaging data





Long-Term Follow-Up of Participants via Linkage to Health Records

Existing linkages

- Death registry
- Cancer registry
- Hospital admissions
- Primary care (latest data COVID-19 research only)
- SARS-CoV-2 PCR tests

More than 200 publications related to COVID-19 research

Potential future linkages

- Microbiology
- Clinical disease audits
- Mental health services
- Ophthalmic records

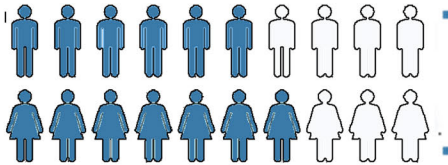
New approaches needed to combine data across diverse health records to characterize health outcomes that are:

- Valid
- Comprehensive
- Scalable

A Truly International Resource

MORE THAN 27,000 APPROVED RESEARCHERS IN 70+ COUNTRIES

2012



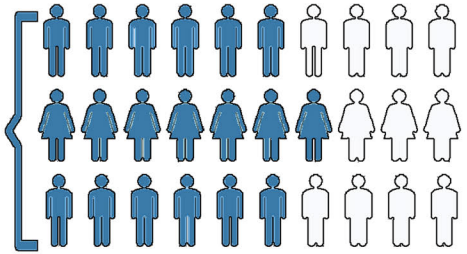
307

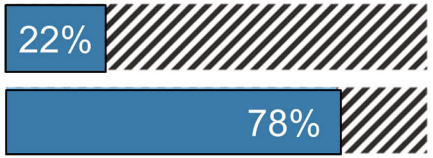
UK

International

2020

6,585





UK

International

Future Exploitation of the Resource

Today

- A lending library
- Researchers download data, perform research, publish results, return derived data, delete data

Data moves to the compute power

Tomorrow

DNA nexus
Amazon Web Services

- Also a reading library
- Researchers access the data in situ, perform analysis, publish results, generate derived data

Compute power moves to the data

- Democratization of access
- Flexible, scalable, secure
- Launched Sept. 28, 2021

Acknowledgements

UK Biobank: Executive Team and Coordinating Centre staff, Steering Committee, International Scientific Advisory Board, Scientific Working Groups, Oxford University team, Cardiff University Participant Resource Centre

Funders:

MRC
Wellcome Trust
British Heart Foundation
Cancer Research UK
Diabetes UK

And, of course, our 500,000 participants:

ihccglobal.org

The IHCC Experience Bringing Cohort Data Together to Advance Precision Health Around the Globe

Laura Lyman Rodriguez, PhD
IHCC Steering Committee

International 100K Cohort Consortium

Disclosure Slide

Financial Disclosure for
Laura Lyman Rodriguez, PhD

I have nothing to disclose

Disclaimer

All statements, opinions, or discussions are solely representative of my personal views and are not reflective of any positions of the Patient-Centered Outcomes Research Institute (PCORI)

Premise for Consortium

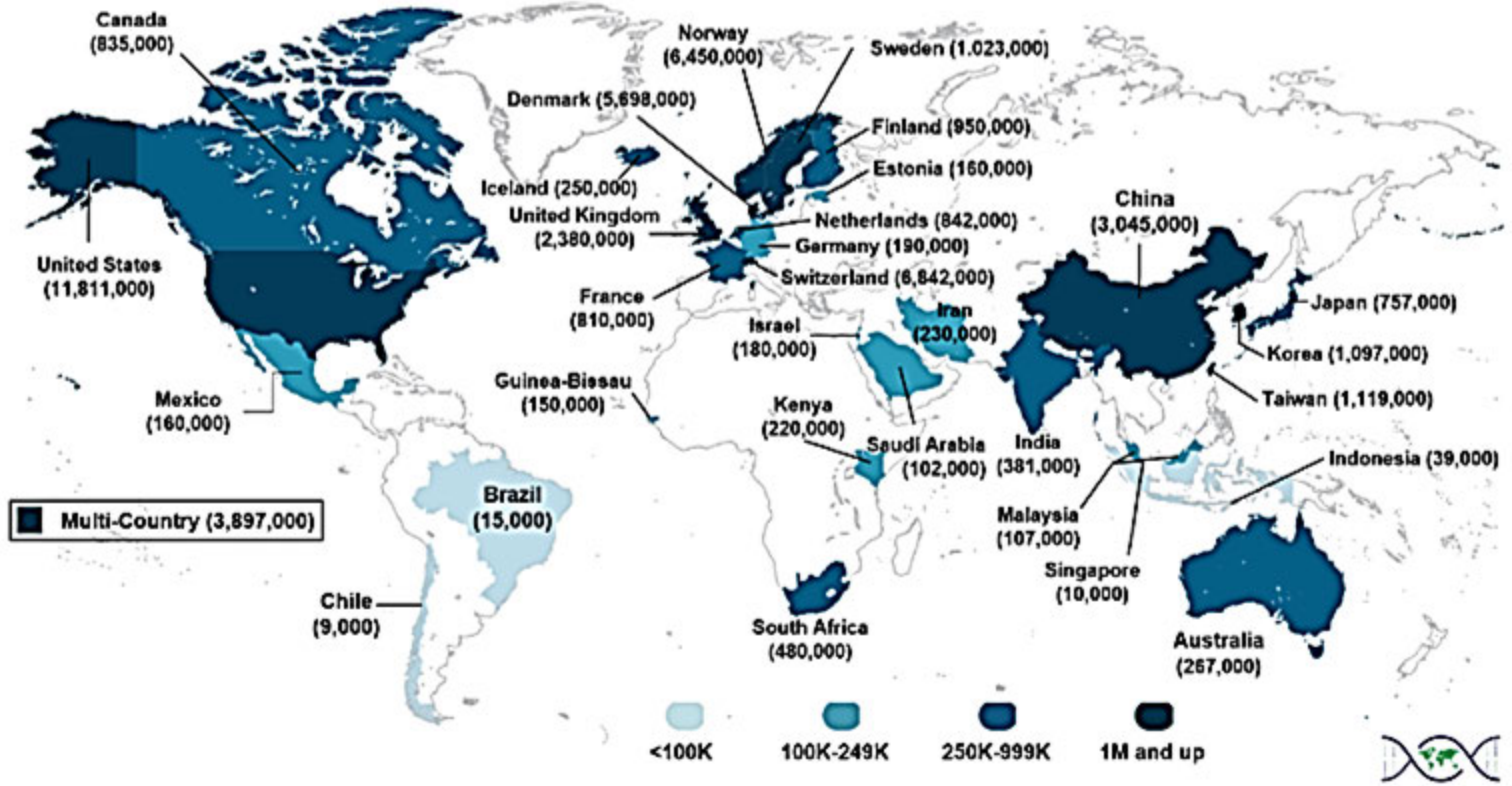
International 100K+ Cohorts Consortium (IHCC)

- Large cohort studies have been established worldwide (some for decades)
- Each constrained by size, ancestral origins, and geographic boundaries
- Constraints limit analyses (e.g., subgroup, exposures, interactions)
- Combining data from these cohorts enables addressing pressing global health questions none can answer alone
    - Enhance value of each
    - Leverage enormous investments in them

https://doi.org/10.1016/S2589-7500(20)30242-9

Power and Potential of IHCC for Research

IHCC Member Cohorts across the World



~110 cohorts, ~60M participants

What Does IHCC Intend to Add to the Community?

Vision

A global community of cohorts working together to advance science and improve health for all.

Mission

To forge cohort connections that revolutionize population health science by providing sustainable data infrastructure, cultivating a collaborative research environment, and promoting policies and best practices that foster connectivity, interoperability, and reciprocity.

Building the Framework

Guiding Principles:

Promotion of inclusivity of cohorts within IHCC activities (focus on LMIC)

Open and timely dissemination of research findings

Transparency about IHCC activities

- Scientific Project Proposal Process
- Publication Policy
- Guidance for Collaboration with Industry
- Core Data Sharing Principles

https://ihccglobal.org/resource-center/

IHCC: Core Values

Action-Oriented, Inclusive and Equity-Focused, Audacious Intentions

We are INTENTIONAL	We focus on building the systems, structures, policies, and practices that enable and sustain cohort collaboration. We do great science rather than just talking about it. We communicate and disseminate our work widely.
We have INTEGRITY	We value transparency, honesty, fairness, and respect. We approach our research with absolute integrity as the basis for the trust imperative to our collective and collaborative endeavors.
We embrace DIVERSITY	We strive to be truly global and inclusive. Any qualifying cohort that wants to be part of our community is welcome. We believe that all cohorts have something to gain and something to contribute and that reciprocity in learning is key to our success.
We strive for EQUITY	We create a supportive environment in which all cohorts and colleagues will excel. We enhance the capacity of each cohort and across cohorts. We see all cohorts as equal in terms of stature and opportunities to contribute.
We act with AUDACITY	We take moonshots and tackle difficult challenges to make progress. We encourage innovation in the discovery and translation of breakthroughs.

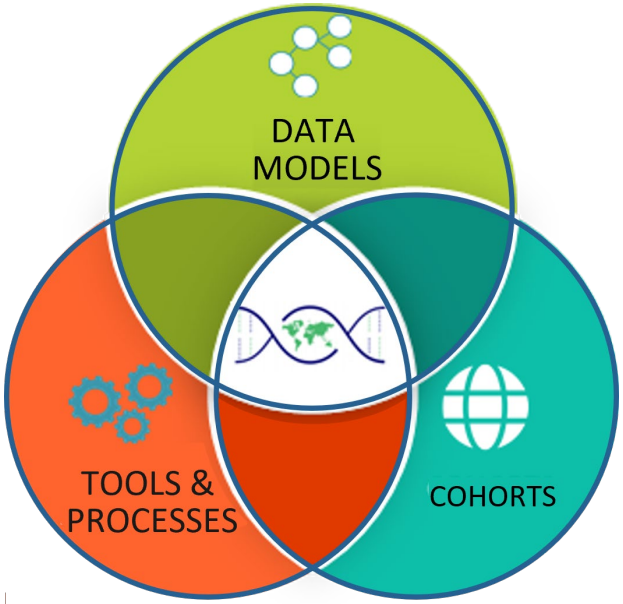
Thomas Keane (UK) & Philip Awadalla (Canada), Co-Leads

- IHCC brings together several axes of cohort data (e.g., disease status, data use, sample collection parameters, genotype, phenotype)
- IHCC consists of a highly diverse set of more than 100 cohort data dictionaries
- IHCC Cohort Atlas aims to:
  - Survey and collate cohort data dictionaries for all IHCC cohorts
  - Semantically harmonize the cohort metadata
  - Develop an online cohort atlas to enable discovery across IHCC cohorts

EMBL-EBI

OICR Ontario Institute for Cancer Research

Building a Common Framework



1. Data models to represent both access conditions and cohort data
2. Tools and processes for implementations
3. Deployment over clinical cohorts

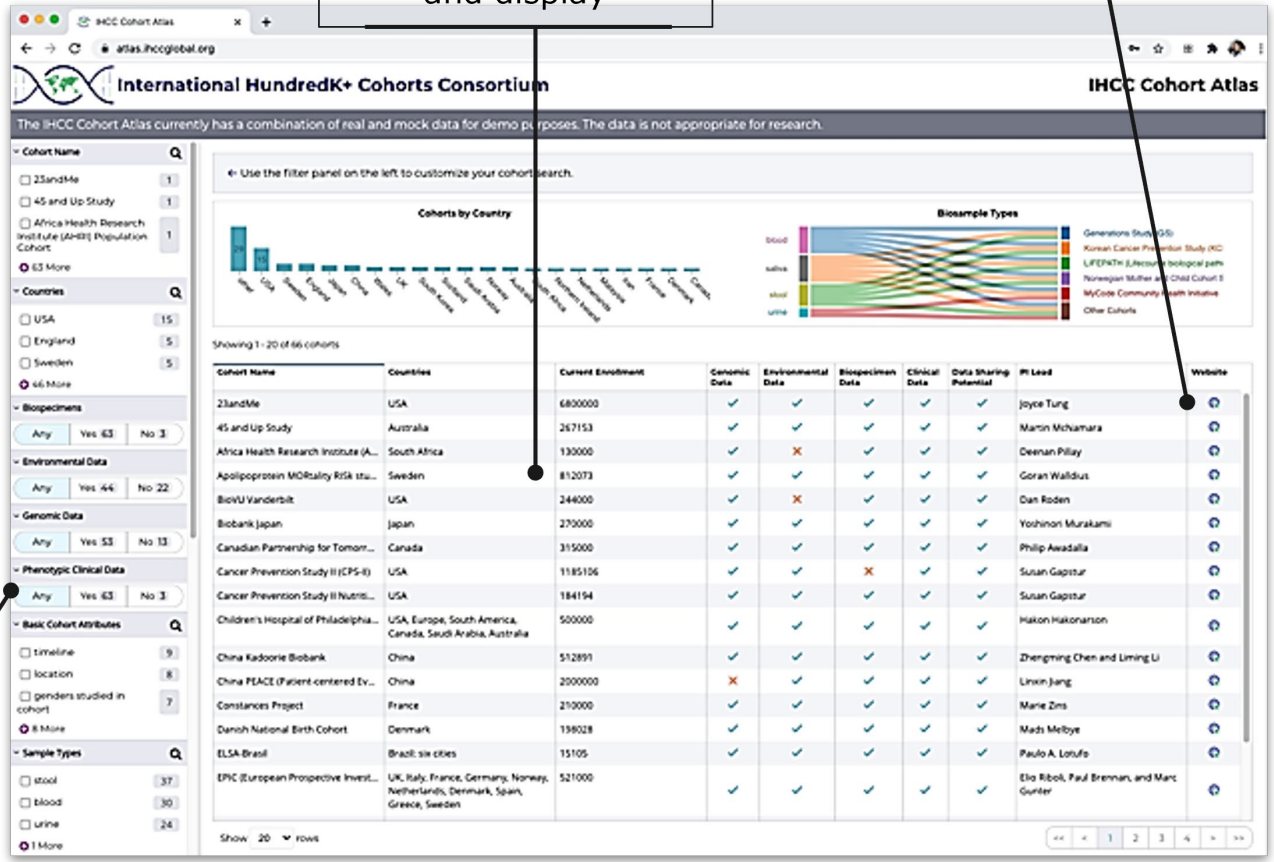
Slides adapted from C. Yung
https://ihccglobal.org/2021-virtual-summit/

IHCC Cohort Atlas Project

- 13 cohorts searchable; >100 variables
- Atlas pipeline applied to second IHCC Project
- Workshops to learn more about the atlas to be held this fall

Cohort presentation and display

Reference to external cohort sites



Intuitive filtering by cohort metadata and data dictionary attributes

atlas.ihccglobal.org/

ihccglobal.org/eventlist/

IHCC Funded Projects: August 2021

Project	PI/PM	Institution	Funders	Year	# cohorts	# LMIC cohorts
Polygenic Risk Scores (PRS)	Hákon Hákonarson (USA)	Children's Hospital of Philadelphia (CHOP)	NIH/WT	2020	6	2
Exploring the Role of Genetically Determined BMI in Infancy, Childhood, and Early Adulthood on Colorectal Cancer Development in Later Life	David J. Hughes (Ireland)	University College Dublin, International Agency for Research on Cancer (IARC), University of Texas	NIH/WT	2021	4	1
High-Throughput Metabolomic Biomarker Measures in Diverse Ancestries	Hákon Hákonarson (USA)	CHOP	NIH/WT	2021	4	2
Opioid Cohort Consortium (OPICO) to Investigate the Effects of Regular Opioid Use on Mortality and on Cancer Development	Paul Brennan (France)	IARC	NIH/WT	2021	10	4
Global Mental Health Impact of the COVID-19 Pandemic	Jordan Smoller (USA) Sarah Bauermeister (UK) & Andre Brunoni (Brazil)	Massachusetts General Hospital, Oxford University, University of Sao Paulo Medical School	NIH/WT	2021	12	3
Novel Coronavirus Host Susceptibility Study in South Africa (COVIGen-SA)	Michele Ramsay (South Africa)	Wits Health Consortium	NIH/WT	2021	3	3
Strengthening Biospecimen Collection for Global Longitudinal Population Studies in the COVID-19 Era	John Chambers (Singapore)	Nanyang Technological University	CZI	2021	4	3
Davos Alzheimer's Collaborative: Pilot PRS	Davos Alzheimer's Collaborative: Pilot	CHOP	DAC	2021	7	4

https://ihccglobal.org/ihcc-funded-projects/

Polygenic Risk Score Implementation Pilot

Hákon Hákonarson (USA), PI

- Late 2019: NIH funded IHCC to conduct a cross-network pilot
  - A proof-of-principle to demonstrate feasibility across a condensed timeline (4 months)
- All IHCC cohorts invited to participate
  - Federated model, whereby sites generate PRSs locally and share summary statistics for centralized analyses
- 2 common traits (BMI, BP) and 2 complex diseases (T2D, asthma)
  - Present across the global population
  - Requirement that each trait had recent large scale meta-analysis reported with **publicly available** genome-wide summary stats
- Scientific goal
  - Compare trans-ancestry and ancestry specific PRS scores in different populations

Slides adapted from P. Sleiman https://ihccglobal.org/2021-virtual-summit/

Polygenic Risk Score Implementation Pilot

Cohort Name	PI: Lead	Participating Countries	Current Enrollment
ELSA-Brasil	Paulo A. Lotufo	Brazil: six cities	15,105
Norwegian Mother and Child Cohort Study (MoBa)	Per Magnus	Norway	284,000
Children's Hospital of Philadelphia (CHOP) Biorepository	Hákon Hákonarson	USA, Europe, South America, Canada, Saudi Arabia, and Australia	500,000
NHS (Nurses' Health Study, NCI)	Meir Stampfer, Rulla Tamimi	USA	121,700
NHSII (Nurses' Health Study II, NCI)	Walter Willett, Heather Eliassen	USA	116,430
Shanghai Men and Women's Health Study (2 cohorts)	Wei Zheng	Shanghai, China	136,000
UK Blood Donor Cohorts	Emanuele Di Angelantonio, John Danesh	UK	100,000

Interested but could not participate

23andMe	Joyce Tung	USA	6,800,000
East London Genes and Health	David van Heel	UK	41,500
Estonian Genome Project	Andres Metspalu	Estonia	200,000

Polygenic Risk Score Implementation Pilot

Results/Conclusions

- Trans-ancestry scores outperform population specific scores in non-European cohorts with similar predictive values
- A modular approach of generating scores at one site and applying across the consortium, while it has issues, does appear to work well
- Additional population-specific LD files, such as South Asians, may be beneficial
- For some phenotypes, such as BP and asthma, summary stats are not yet sufficiently good to generate predictive PRS
- Data sharing agreements between the participating cohorts would greatly facilitate work; by sharing individual-level genotypes, we can generate common weights across the participant sites

Metabolomic Markers in Diverse Ancestries

A. Butterworth (UK), A. Brunoni (Brazil), A. Etemadi (US), H. Hákonarson (US)

- Chronic diseases impose a high burden on the health system.
- Health outcomes can be significantly improved through early diagnosis and intervention.
- Early diagnosis often unavailable particularly for individuals in low- and middle-income countries and minority populations in high-income countries.
- Metabolic profiling represents a highly-scalable model for risk prediction and prevention.

Slides adapted from A. Etemadi
https://ihccglobal.org/2021-virtual-summit/

Metabolomic Markers in Diverse Ancestries

Participating Cohorts

Cohort Name	Study Samples	Principal Investigator/Lead(s)
South Asian Cohorts (BELIEVE)	1,500 samples of South Asian ancestry from Dhaka, Bangladesh	Adam Butterworth
ELSA-Brasil	1,000 samples from Brazilian civil servants	Andre Brunoni
Golestan Cohort Study	1,000 samples from Northeast Iranian general population	Arash Etemadi
Children's Hospital of Philadelphia (CHOP)	1,500 samples of African American children	Hákon Hákonarson

- Project developed with intention of scaling up beyond initial pilot studies
- Each study has own cohort-specific outcomes
- Data from samples recently received; analysis underway

Overarching IHCC Goals: 2–3 Years

- **Demonstrate That IHCC Generates Impactful Science**
  - Provide proof of concept that IHCC generates impactful science through ambitious scientific projects that require scale and diversity and improve health for all
- **Enable Discovery and Connectivity of Cohorts for Collaboration**
  - Facilitate cohort discoverability and data access by enhancing IHCC's Atlas
- **Make It Possible for All Cohorts to Contribute to IHCC Scientific Challenges**
  - Promote development and/or adoption of policies and best practices and enhance cohort capabilities and competencies to improve the practice of collaboration
- **Build a Strong Governance and Operational Foundation**

IHCC Funding Organizations

NIH National Institutes of Health
Turning Discovery Into Health

wellcome

CHAN ZUCKERBERG INITIATIVE

Davos Alzheimer's Collaborative

For more information and updates:

ihccglobal.org

Comments or questions to llymanrodriguez@gmail.com

GA4GH Standards to Enable Global Access and Interoperability of Data to Inform Precision Health

Heidi Rehm, PhD

Massachusetts General Hospital and Broad Institute of MIT and Harvard
Boston, MA, USA

Disclosures for Heidi Rehm

I receive NIH funding to support GA4GH (U24HG011025)

Why do we need the

Global Alliance for Genomics and Health,

and what does it do?

Challenges

- Data is typically in silos:
  - By type, by disease, by country, by institution
- Analysis methods are non-standardized; few work at scale
- Different approaches to regulation, consent, and data sharing limit interoperability

Why data sharing is needed:

- Increases statistical significance of analyses
- Builds evidence for gene–disease causality by gathering rare disease patients
- Leads to “stronger” variant interpretations by aggregating evidence
- Supports more informed clinical decisions

These challenges and opportunities led to the launch of GA4GH in 2013

The GA4GH Mission

The Global Alliance for Genomics and Health aims to accelerate progress in genomic science and human health by developing standards and framing policies for responsible genomic and health-related data sharing.

About GA4GH

GA4GH aims to...

Enable **international** data sharing

Promote sharing across the **translational continuum** (research, clinical, industry)

Encourage technology-enabled **federated approaches** (bring analysis to the data) where needed to access data

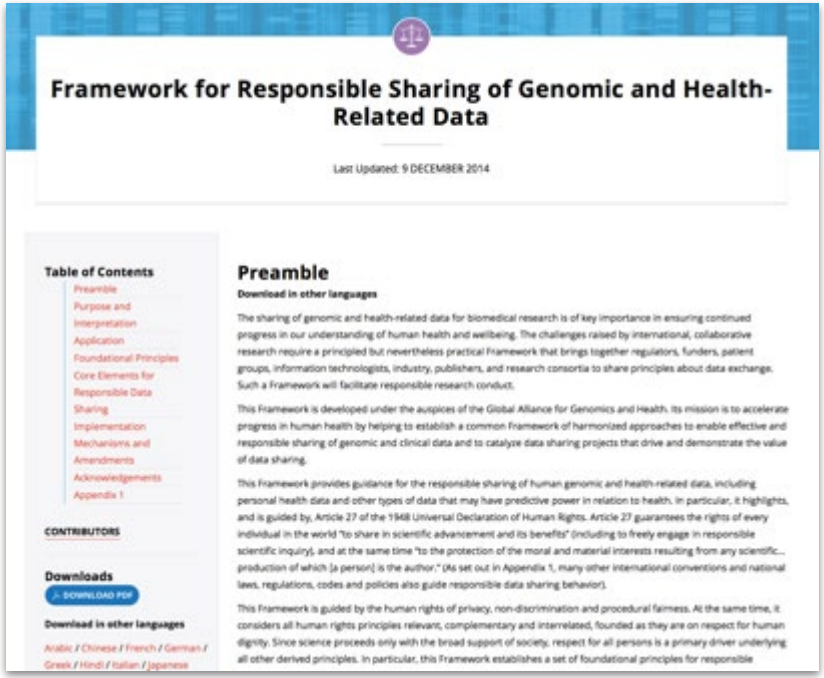
Promote **interoperability** (scientific, technical, ethical)

GA4GH achieves this by...

- **Convening** stakeholders
- **Catalyzing** sharing of data
- **Creating** harmonized approaches
- **Acting** as a clearinghouse
- **Fostering** innovation
- **Committing** to responsible data sharing
- But **not** generating data or performing research/care for patients

Core Framework for Responsible Sharing of Genomic and Health-Related Data

Built on **Universal Declaration of Human Rights (1948, 27(1): "The Right to Science"):** "**Everyone** has the right…**to share in scientific advancement and its benefits.**"



Framework for Responsible Sharing of Genomic and Health-Related Data

Last Updated: 9 DECEMBER 2014

Table of Contents

Preamble
Purpose and Interpretation
Application
Foundational Principles
Core Elements for Responsible Data Sharing
Implementation Mechanisms and Amendments
Acknowledgements
Appendix 1

CONTRIBUTORS

Downloads

DOWNLOAD PDF

Download in other languages

Arabic / Chinese / French / German / Greek / Hindi / Italian / Japanese

Preamble

Download in other languages

The sharing of genomic and health-related data for biomedical research is of key importance in ensuring continued progress in our understanding of human health and wellbeing. The challenges raised by international, collaborative research require a principled but nevertheless practical Framework that brings together regulators, funders, patient groups, information technologists, industry, publishers, and research consortia to share principles about data exchange. Such a Framework will facilitate responsible research conduct.

This Framework is developed under the auspices of the Global Alliance for Genomics and Health. Its mission is to accelerate progress in human health by helping to establish a common Framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data and to catalyze data sharing projects that drive and demonstrate the value of data sharing.

This Framework provides guidance for the responsible sharing of human genomic and health-related data, including personal health data and other types of data that may have predictive power in relation to health. In particular, it highlights, and is guided by, Article 27 of the 1948 Universal Declaration of Human Rights. Article 27 guarantees the rights of every individual in the world "to share in scientific advancement and its benefits" (including to freely engage in responsible scientific inquiry), and at the same time "to the protection of the moral and material interests resulting from any scientific... production of which [a person] is the author." (As set out in Appendix 1, many other international conventions and national laws, regulations, codes and policies also guide responsible data sharing behavior).

This Framework is guided by the human rights of privacy, non-discrimination and procedural fairness. At the same time, it considers all human rights principles relevant, complementary and interrelated, founded as they are on respect for human dignity. Since science proceeds only with the broad support of society, respect for all persons is a primary driver underlying all other derived principles. In particular, this Framework establishes a set of foundational principles for responsible

GA4GH Organization Structure

Global Alliance for Genomics & Health

STRATEGIC ADVISORY

- Funders Forum
- SAB

ORGANIZATIONAL GOVERNANCE

- Board of Directors
  - Finance, Audit, & Risk Committee
  - Executive Committee
  - Nominating Committee
- CEO
- Secretariat
  - Ops Team
  - Technical Standards Team
  - Work Stream Managers
  - Comms Team

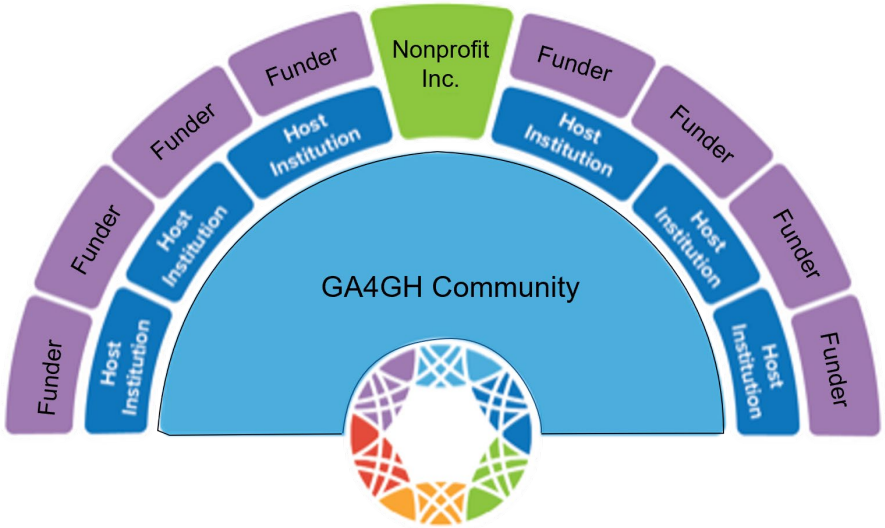
TECHNICAL GOVERNANCE

- Standards Steering Committee
- Technical Alignment Sub-Committee

Executive Leadership

Peter Goodhand, Chief Executive Officer
Ewan Birney, Chair
Kathryn North, Vice Chair
Heidi Rehm, Vice Chair
Angela Page, Director of Strategy and Engagement
Susan Fairley, Chief Standards Officer

GA4GH Inc.—Incorporated as a Nonprofit Organization



Host Institutions

OICR Ontario Institute for Cancer Research
wellcome sanger institute
BROAD INSTITUTE
EMBL-EBI

Current Funding Sources

NIH
BROAD INSTITUTE
wellcome
NIHR | National Institute for Health Research
MRC Medical Research Council
OICR Ontario Institute for Cancer Research
CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada
GenomeCanada

How GA4GH Works

Technical Work Streams

- Discovery
- Large-Scale Genomics
- Data Use & Researcher IDs
- Cloud
- Genomic Knowledge Standards
- Clinical & Phenotypic Data Capture

Foundational Work Streams

- Regulatory & Ethics
- Data Security

Input from Driver Projects as exemplars of community

Output of standards that are taken up by the community

Partner Engagement

RARE DISEASES

Matchmaker Exchange

COMPLEX TRAITS

AUTISM SPEAKS

NHLBI TOPMed

CANCER

ARGO

NIH National Cancer Institute

EU CAN.

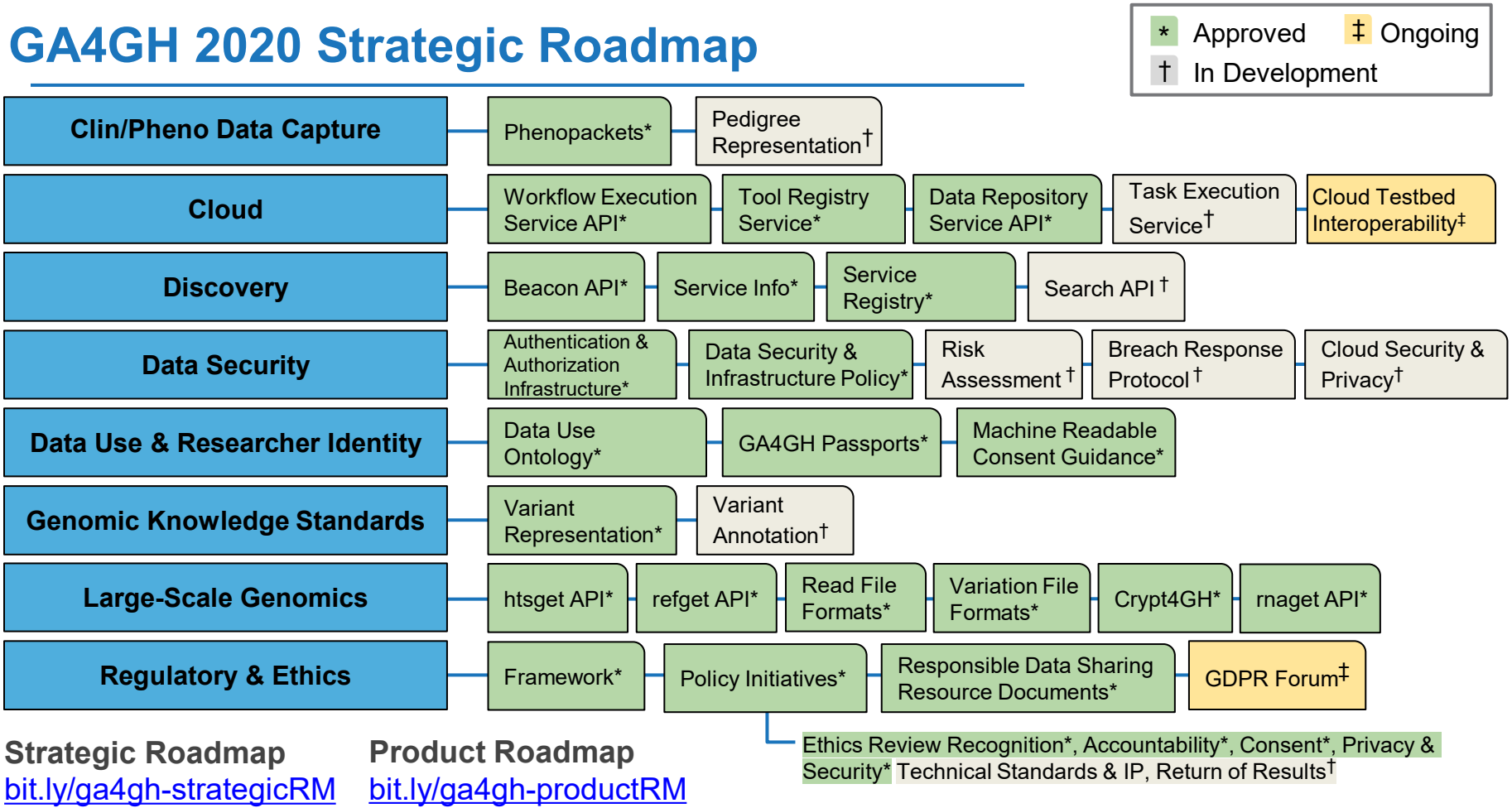
VICC Variant Interpretation for Cancer Consortium

BASIC BIOLOGY

24 Driver Projects

Real-World Driver Projects

GA4GH 2020 Strategic Roadmap



Strategic Roadmap
bit.ly/ga4gh-strategicRM

Product Roadmap
bit.ly/ga4gh-productRM

Strategic Roadmap Alignment to F.A.I.R. Principles

Findable

- Beacon API
- Data Use Ontology
- refget API
- Search API
- Service Registry Prototype
- Tool Registry Service (TRS)

Accessible

- Authentication & Authorization Infrastructure
- Data Repository Service
- Data Use Ontology
- Researcher ID & Bona Fide Status

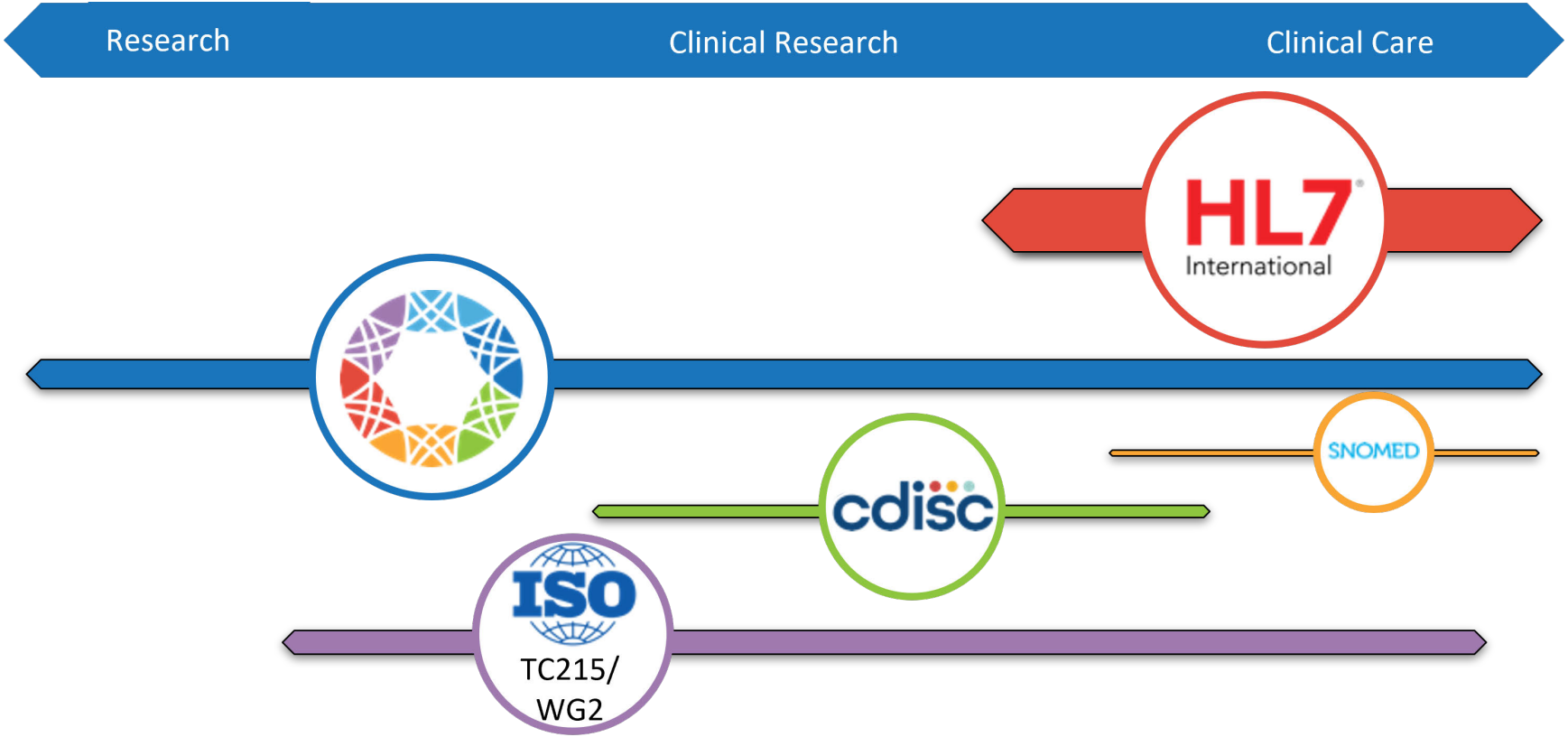
Interoperable

- Phenotype Representation
- Phenopackets/FIHR
- Pedigree Representation
- Genetic Variant File Formats
- Read File Formats
- RNASeq Expression Matrix
- RNASeq API
- Crypt4GH
- Variant Annotation
- Variant Representation
- Task Execution Service
- Testbed Interoperability Demonstration
- Tool Registry Service
- Workflow Execution Service

Reusable

- htsget Streaming API
- refget API
- Variant Annotation
- Workflow Execution Service
- Testbed Interoperability Demonstration

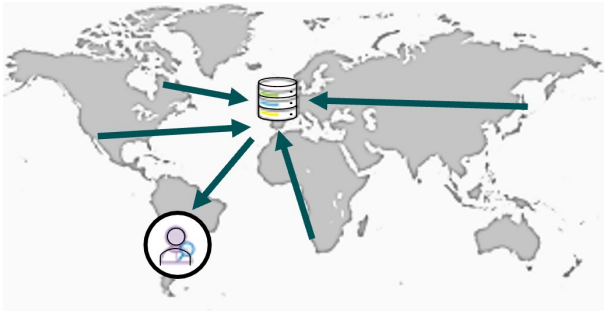
Alignment With Other Genomics Standards Organizations



Different Approaches to Data Sharing

Central Database

Genomic knowledge base

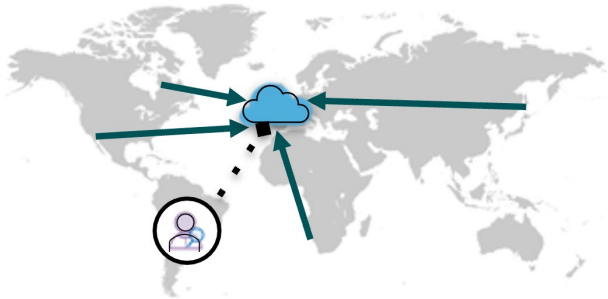


Aggregate data globally

Download and analyze locally

Secure Cloud

Large-scale research datasets

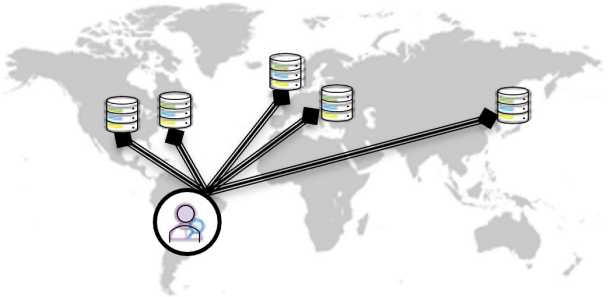


Aggregate data globally

Analyze centrally in secure cloud

Federation

Connecting national genomics initiatives



Host data locally

Visit data remotely and collate results

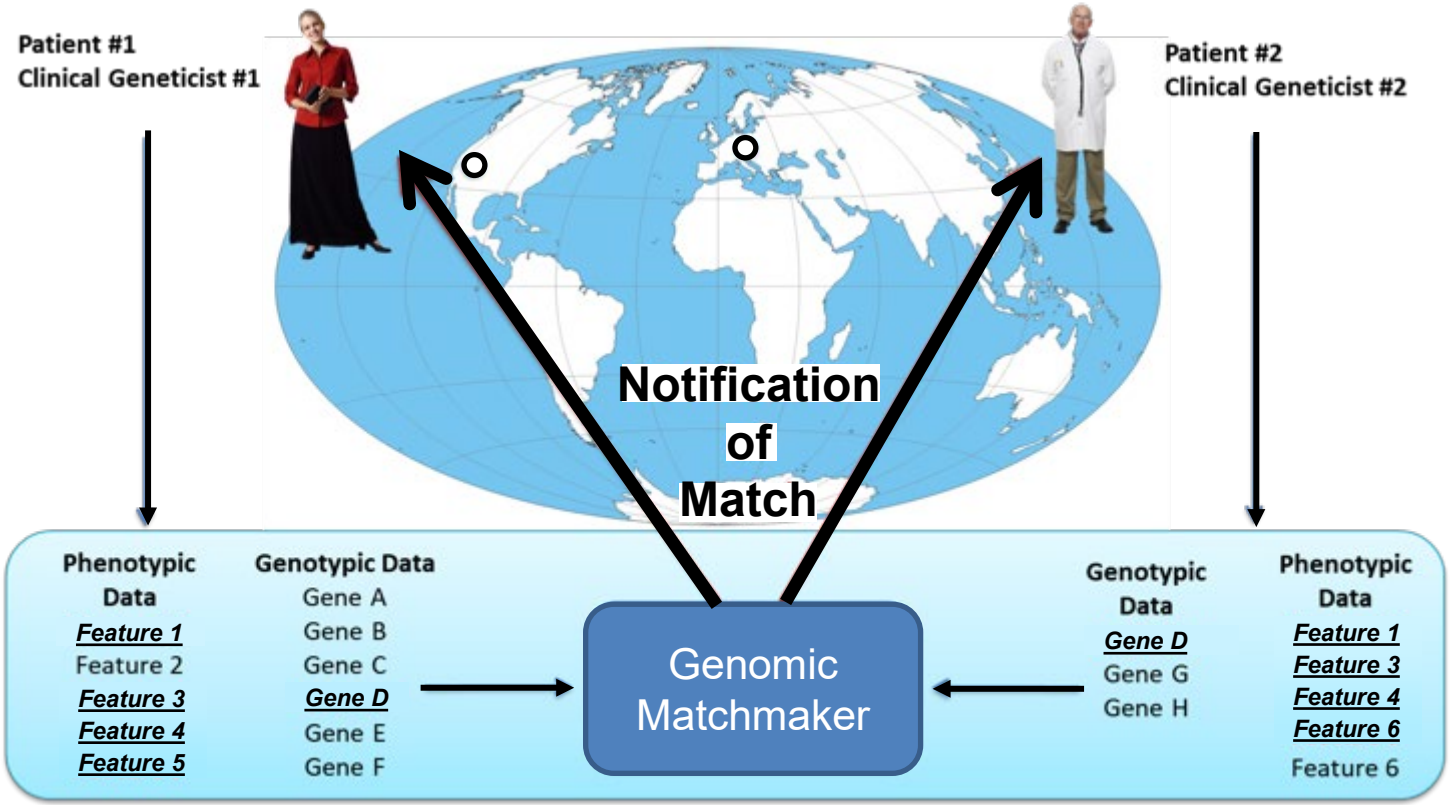
User

Data transmission

Secure access

Rare Disease Genomic Matchmaking

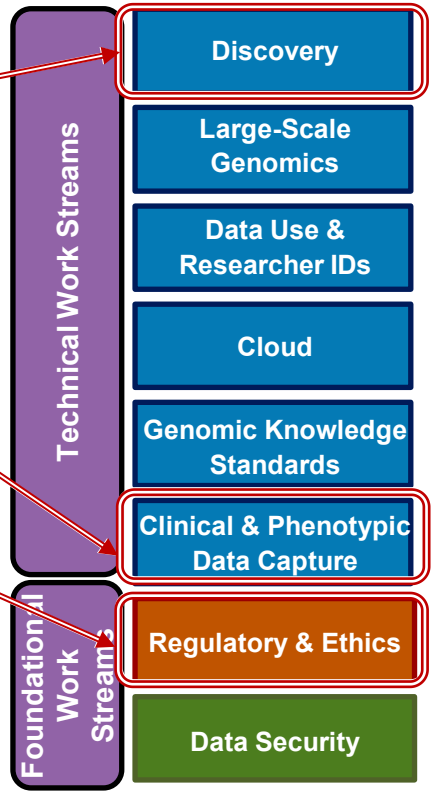
A Use Case for Federation

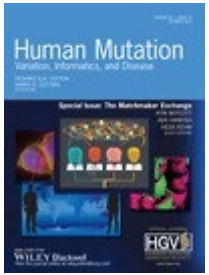


Developing the MME Federated Network Using GA4GH Standards

Use of GA4GH standards:

- API for data exchange
  - ID (Mandatory) +/- Label
  - Submitter (Mandatory)
  - Phenotypic Features and/or Gene Names (Mandatory)
  - Disorders (Optional)—OMIM or OrphaNet
  - Sex, Age of Onset, Inheritance (Optional)
- Clinical and phenotypic data capture standards
- Consent framework for data sharing





Philippakis et al. **The Matchmaker Exchange: a platform for rare disease gene discovery.** *Hum Mutat*. 2015;36(10):915–21.

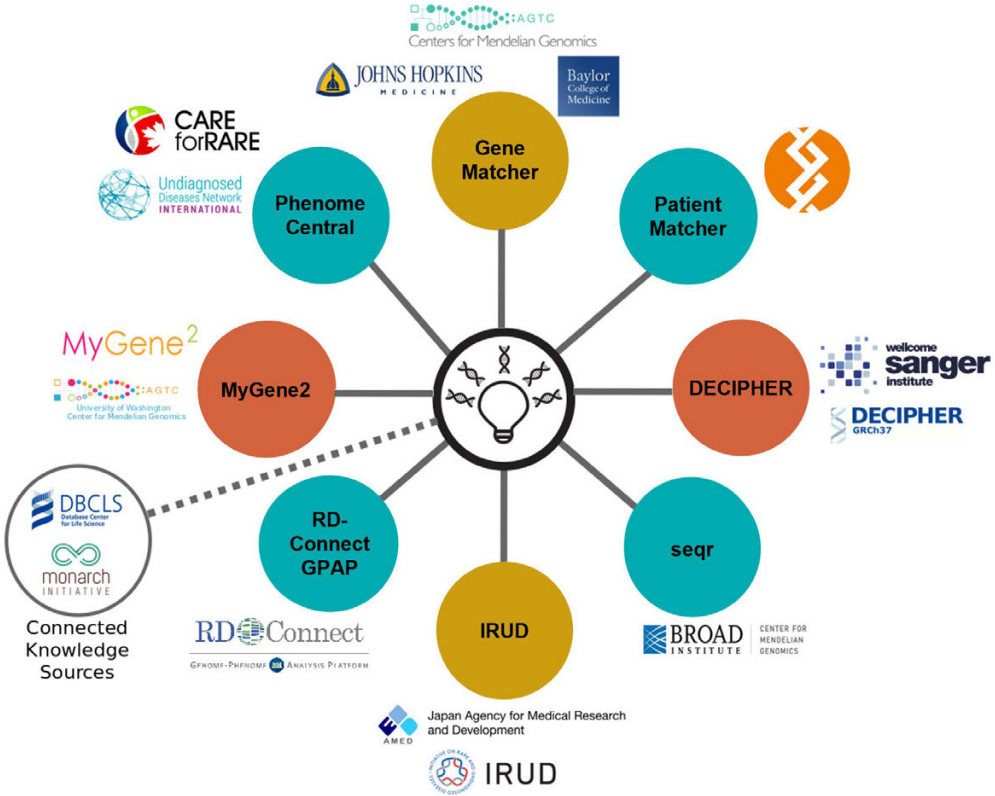
Buske et al. **The Matchmaker Exchange API: automating patient matching through the exchange of structured phenotypic and genotypic profiles.** *Hum Mutat*. 2015;36(10):922–7.

16 papers in a special issue of *Human Mutation* (Vol 36, Issue 10, Oct 2015)

The Matchmaker Exchange

Supporting the Discovery of Novel Causes of Rare Disease

A Successful Example of a Federated Platform

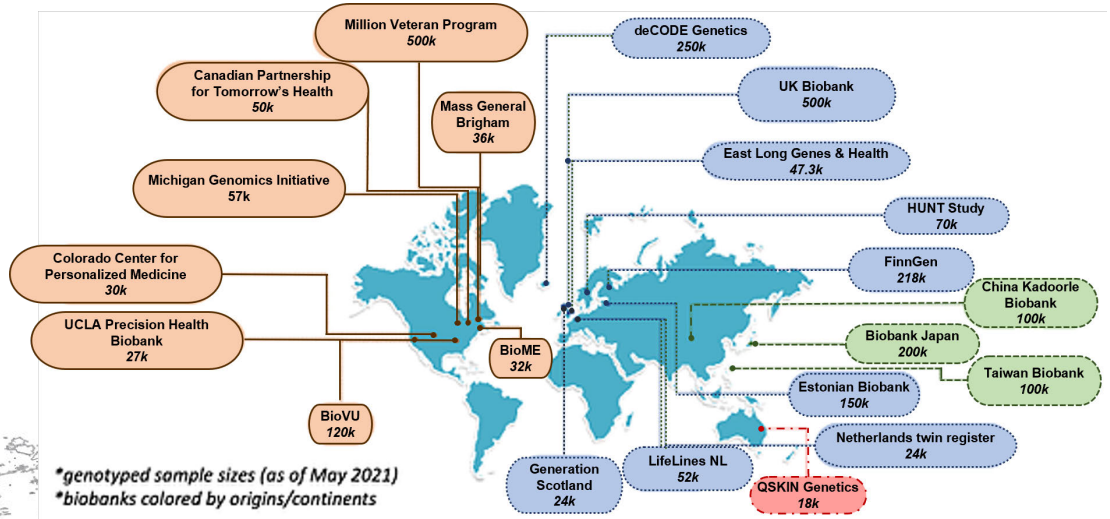


www.matchmakerexchange.org

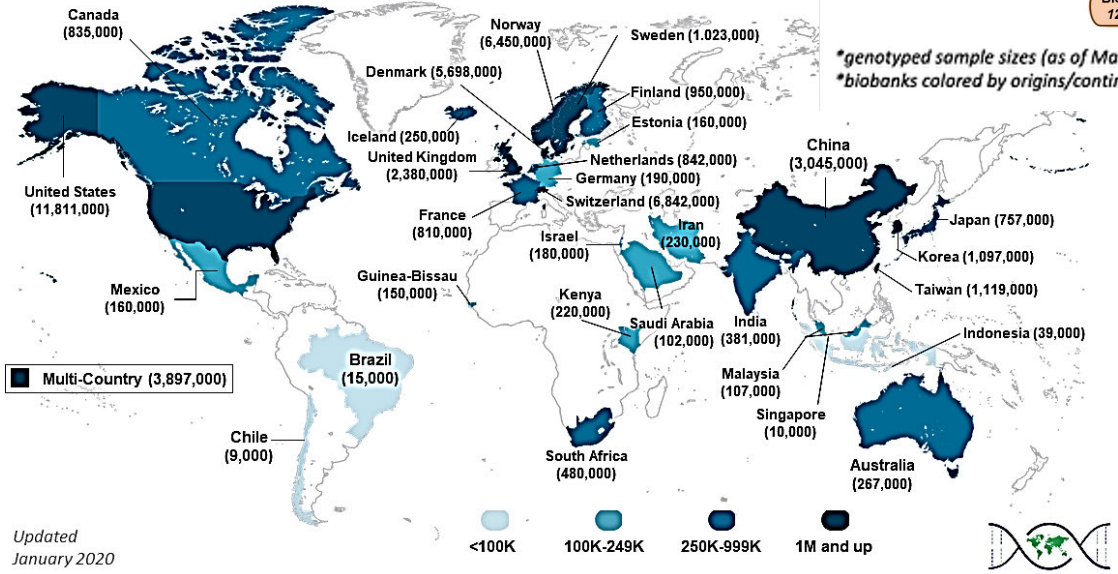
Datasets Around the Globe

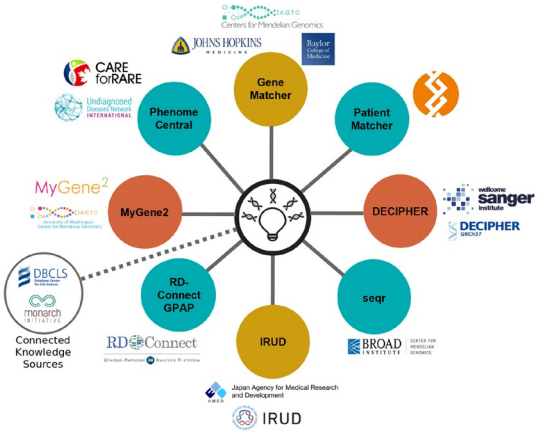
21 biobanks with different origins and ancestries have joined GBMI

>2.6 million genotyped samples

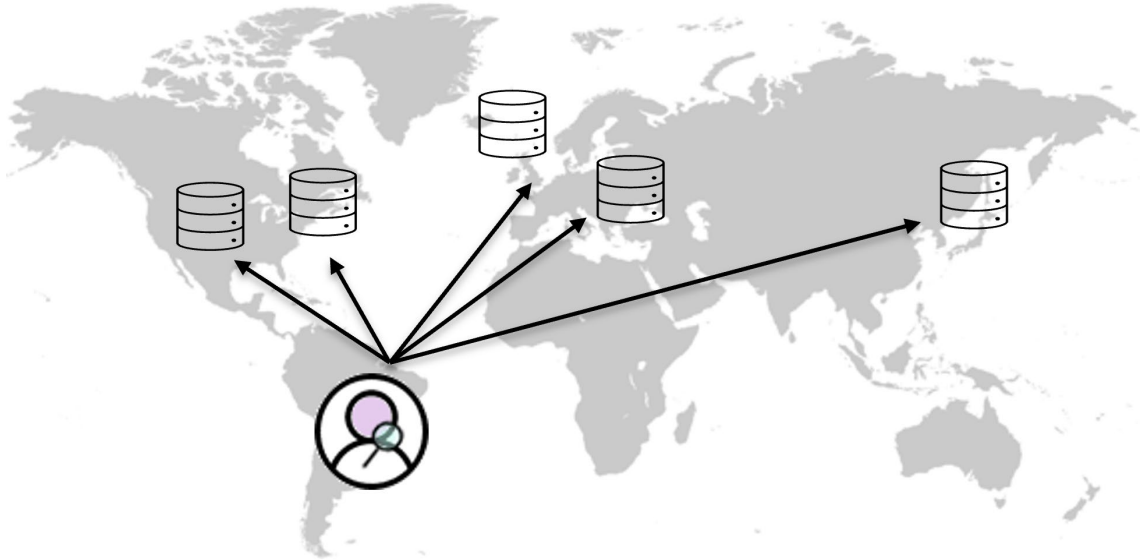


IHCC Member Cohorts across the World





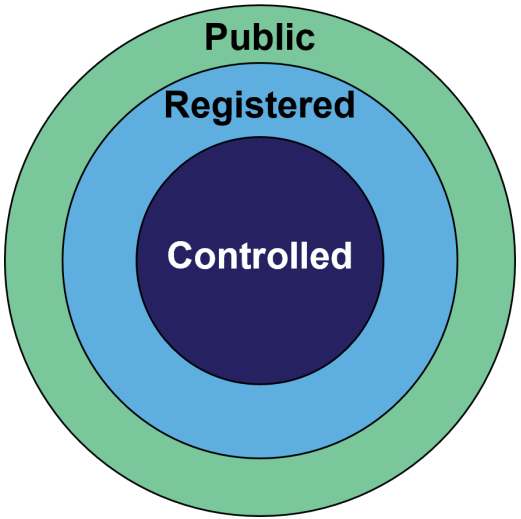
Federated Data Ecosystems



Requires more sophisticated models for data access and analysis

All of Us Research Program Model for Data Access

Data Tiers



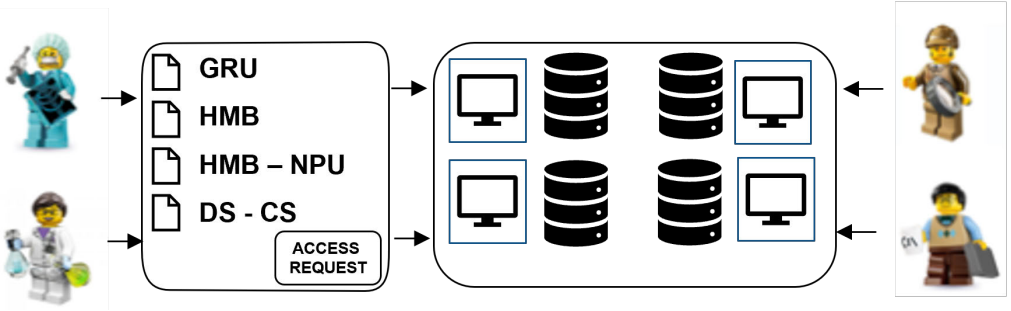
- Public—data with **minimal risk** to participant privacy
- Registered—data with **some privacy risk** to participants
- Controlled—data with most **significant privacy risks**—researchers must be approved by a DAC to access it

Access granted as **researcher-based**, rather than project-based, using a **data passport**

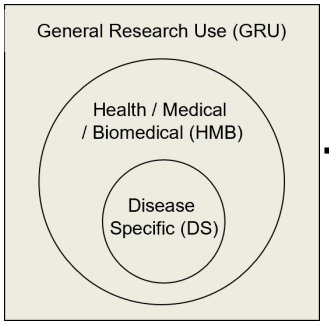
Data Use Ontology v1

The Data Use Ontology (DUO) allows users to semantically tag genomic datasets with usage restrictions, allowing these sets to become automatically discoverable based on a health, clinical, or biomedical researcher's authorization level or intended use.

Approved: January 28, 2019



Controlled Access



Modifiers

POA	Population origins or ancestry research
NMDS	No general methods research
GSO	Genetic studies only
NPU	Not-for-profit use only
PUB	Publication required
COL	Collaboration required
IRB	Ethics approval required
GS	Geographical restriction
MOR	Publication moratorium
RT	Return to database/resource

Example Users

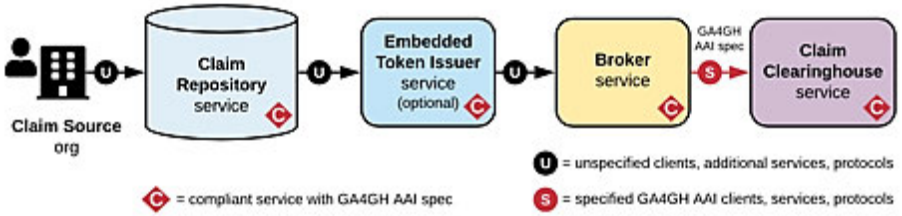
All of Us RESEARCH PROGRAM · BROAD INSTITUTE · NIH NATIONAL CANCER INSTITUTE · NHLBI TOPMed · EUROPEAN GENOME-PHENOME ARCHIVE

GA4GH Passports v1 and AAI

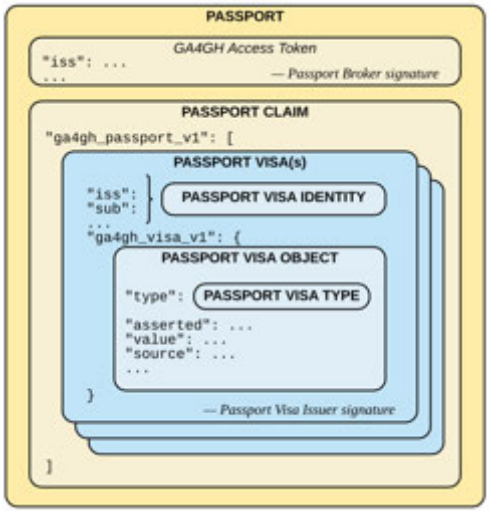
The GA4GH Passports and Authentication & Authorization Infrastructure standards work together to reliably authenticate a researcher's digital identity and automate their access to requested genomic datasets.

Approved: October 23, 2019

Flow of Claims:



GA4GH Passport Structure:



Example Users

elixir

Google Cloud Platform

Get Involved! Visit ga4gh.org

Join a Work Stream!

Contact secretariat@ga4gh.org

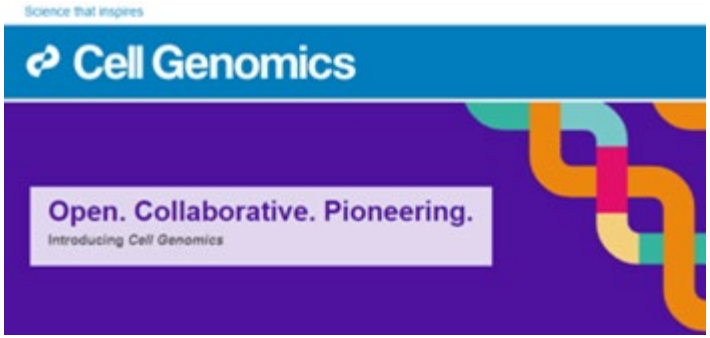
Become an Organizational Member

ga4gh.org/members

Subscribe to GA4GH Updates

ga4gh.org/subscribe

GA4GH Marker Paper and Other GA4GH Work Product Publications Coming in November Issue of *Cell Genomics*



GA4GH: international policies and standards for data sharing across genomic research and healthcare

Heidi L. Rehm[1,2], Angela J.H. Page[1,2], Lindsay Smith[4,5], Jeremy B. Adams[4,5], Gil Alterovitz[6], Lawrence J. Babb[7], Maximilian P. Barkley[8], Michael Baudis[9,10], Michael J.S. Beauvais[9,11], Tim Beck[12], Jacques S. Beckmann[13], Sergi Beltran[14,15,16], David Bernick[7], Alexander Bernier[8], James K. Bonfield[17], Tiffany F. Boughtwood[18,19], Guillaume Bourque[8,20], Sarion R. Bowers[21], Anthony J. Brookes[22], Michael Brudno[23,24,25,26], Matthew H. Brush[27], David Bujold[8,28,29], Tony Burdett[30], Orion J. Buske[31], Moran N. Cabili[1], Daniel L. Cameron[25,26], Robert J. Carroll[27], Esmeralda Casas-Silva[25], Debyani Chakravarty[32], Bimal P. Chaudhari[9,33], Shu Hui Chen[22], J. Michael Cherry[34], Justina Chung[4,5], Melissa Cline[35], Hayley L. Clissold[9], Robert M. Cook-Deegan[36], Mélanie Courtot[37], Fiona Cunningham[30], Miro Cupak[8], Robert M. Davies[38], Danielle Denisko[9], Megan J. Doerr[27], Lena I. Dolman[39], Edward S. Dove[40], L. Jonathan Dursi[20,26], Stephanie O.M. Dyke[8], James A. Eddy[27], Karen Eilbeck[41], Kyle P. Elliott[22], Susan Fairley[23,3], Khalid A. Fakhro[42,43], Helen V. Firth[43,15], Michael S. Fitzsimons[44], Marc Fiume[8], Paul Flicek[23], Ian M. Fore[29], Mallory A. Freeberg[23], Robert R. Freimuth[45], Lauren A. Fromont[32], Jonathan Fuerth[8], Clara L. Gaff[46,7], Weiniu Gan[33], Elena M. Ghanaim[46], David Glazer[47], Robert C. Green[48,49], Malachi Griffith[50], Obi L. Griffith[50], Robert L. Grossman[44], Tudor Groza[51], Jaime M. Guidry Auvil[29], Roderic Guigó[52,53], Dipayan Gupta[22], Melissa A. Haendel[33], Ada Hamosh[54], David P. Hansen[22,6], Reece K. Hart[24,1,55], Dean Mitchell Hartley[56], David Haussler[25], Rachele M. Hendricks-Sturrup[36], Calvin W.L. Ho[57], Ashley E. Hobb[8], Michael M. Hoffman[9,20,21], Oliver M. Hofmann[58], Petr Holub[59,60], Jacob Shujui Hsu[50], Jean-Pierre Hubaux[61], Sarah E. Hunt[23], Ammar Husami[62], Julius O. Jacobsen[63], Saumya S. Jamuar[64,65], Elizabeth L. Janes[66,7], Francis Jeanson[29], Aina Jené[52], Amber L. Johns[67,68], Yann Joly[8], Steven J.M. Jones[69], Alexander Kanitz[70,8], Kazuto Kato[71], Thomas M. Keane[23,72], Kristina Kekesi-Lafrance[8,3], Jerome Kelleher[73], Giselle Kerry[23], Seik-Soon Khor[74,75], Bartha M. Knoppers[8], Melissa A. Konopko[76], Kenjiro Kosaki[77], Martin Kuba[59], Jonathan Lawson[1], Rasko Leinonen[23], Stephanie Li[7,3], Michael F. Lin[78], Mikael Linden[79,80], Xianglin Liu[66], Isuru Udara Liyanage[23], Javier Lopez[101], Anneke M. Lucassen[81], Michael Lukowski[64], Alice L. Mann[9,3], John Marshall[82], Michele Mattioni[52], Alejandro Metke-Jimenez[83], Anna Middleton[84,85], Richard J. Milne[84,85], Fruzsina Molnar-Gabor[86], Nicola Mulder[87], Monica C. Munoz-Torres[32], Rishi Nag[23], Hidewaki Nakagawa[88,89], Jamal Nasir[90], Arcadi Navarro[91,92,52,93], Tristan H. Nelson[94], Ania Niewielska[23], Amy Nissele[95,77,24], Jeffrey Niu[20], Tommi H. Nyrönen[79,80], Brian D. O'Connor[1], Sabine Oesterle[8], Soichi Ogishima[96], Laura A.D. Paglione[97,98], Emilio Palumbo[52,3], Helen E. Parkinson[23], Anthony A. Philippakis[1], Angel D. Pizarro[99], Andreas Prlic[100], Jordi Rambla[52,3], Augusto Rendon[101], Renee A. Rider[46], Peter N. Robinson[102,103], Kurt W. Rodarmer[104], Laura Lyman Rodriguez[105], Alan F. Rubin[25,26], Manuel Rueda[52], Gregory A. Rushton[1], Rosalyn S. Ryan[106], Gary I. Saunders[78], Helen Schuilenburg[23], Torsten Schwede[70,8], Serena Scollen[78], Alexander Senf[107], Nathan C. Sheffield[108], Neerjah Skantharajah[4,5], Albert V. Smith[109], Heidi J. Sofia[46], Dylan Spalding[79,80], Amanda B. Spurdle[110], Zornitza Stark[18,19], Lincoln D. Stein[8,19], Makoto Suematsu[111], Patrick Tan[64,112,113], Jonathan A. Teixido[78], Alastair A. Thomson[22], Adrian Thorogood[8,114], Timothy L. Tickle[1], Katsushi Tokunaga[74,75], Juha Tornroos[79,80], David Torrents[91,92], Sean Upchurch[115], Alfonso Valencia[116,92], Roman Valls Guimera[36], Jessica Vamathevan[23], Susheel Varma[117,23], Danya F. Vears[85,17,26,118], Coby Viner[9,20], Craig Voisin[119], Alex H. Wagner[21,33], Susan E. Wallace[120], Brian P. Walsh[22], Vivian Ota Wang[29], Marc S. Williams[84], Eva C. Winkler[121], Barbara J. Wold[122], Grant M. Wood, J. Patrick Woolley[123], Chisato Yamasaki[71], Andrew D. Yates[23], Christina K. Yung[4,21], Lyndon J. Zass[87], Ksenia Zaytseva[9,22], Junjun Zhang[4], Peter Goodhand[4,3], Kathryn North[17,26], Ewan Birney[23,22],

Acknowledgments

GA4GH Executive Leadership:
Ewan Birney: European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK
Peter Goodhand: Ontario Institute for Cancer Research, Toronto, Canada
Kathryn North: Murdoch Children's Research Institute, Melbourne, Australia
Heidi L. Rehm: Broad Institute of MIT and Harvard, Cambridge, MA, USA
Angela Page: Broad Institute of MIT and Harvard, Cambridge, MA, USA
Susan Fairley: EMBL-EBI, Hinxton, UK

GA4GH Secretariat:
Alexa Frieberg: Broad Institute of MIT and Harvard, Cambridge, MA, USA
Alice Mann: Wellcome Sanger Institute, Hinxton, UK
Amreen Mohamed: Ontario Institute for Cancer Research, Toronto, Canada
Jeremy Adams: Ontario Institute for Cancer Research, Toronto, Canada
Justina Chung: Ontario Institute for Cancer Research, Toronto, Canada
Kristina Kékesi-Lafrance: McGill University, Montreal, Canada
Lindsay Smith: Ontario Institute for Cancer Research, Toronto, Canada
Maili Raven-Adams: Wellcome Sanger Institute, Hinxton, UK
Michael Beauvais: McGill University, Montreal, Canada
Neerjah Skantharajah: Ontario Institute for Cancer Research, Toronto, Canada
Stephanie Li: Broad Institute of MIT and Harvard, Cambridge, MA, USA

Current GA4GH WSLs:
Michael Baudis: Universität Zürich, Zürich, Switzerland
David Bernick: Broad Institute of MIT and Harvard, Cambridge, MA, USA
Edward Dove: University of Edinburgh, Edinburgh, UK
Marc Fiume: DNAstack, Toronto, Canada
Robert Freimuth: Mayo Clinic, Rochester, MN, USA
David Glazer: Verily Life Sciences, South San Francisco, CA, USA
Melissa Haendel: Oregon Health & Science University, Portland, OR, USA
David Hansen: Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia
Oliver Hofmann: University of Melbourne, Melbourne, Australia
Jean-Pierre Hubaux: École polytechnique fédérale de Lausanne, Lausanne, Switzerland
Yann Joly: Centre of Genomics and Policy, McGill University, Montreal, Canada
Thomas M. Keane: EMBL-EBI, Hinxton, UK
Tommi Nyrönen: CSC – IT Center for Science, Espoo, Finland
Brian O'Connor: University of California, Santa Cruz, CA, USA
Andrew Yates: EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge, UK

Past GA4GH WSLs:
Bartha Knoppers: Centre of Genomics and Policy, McGill University, Montreal, Canada
Madeleine Murtagh: Newcastle University, Newcastle upon Tyne, UK
Harindra Arachchi: Broad Institute of MIT and Harvard, Cambridge, MA, USA (currently Editas Medicine, Inc., USA)
Dixie Baker: Martin, Blanck & Associates, Redondo Beach, CA, USA
Moran Cabili: Broad Institute of MIT and Harvard, Cambridge, MA, USA (currently Foundation Medicine, Inc., USA)
Paul Flicek: EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge, UK
Ravi Pandya: Microsoft, Redmond, WA, USA
Anthony Philippakis: Broad Institute of MIT and Harvard, Cambridge, MA, USA

Global Alliance for Genomics & Health