Applicant webinar: PRS Diversity Consortium RFAs

Lucia Hindorff, Teri Manolio, Rudy Pozzatti, Catherine Sillari, Ken Wiley, NHGRI

Damali Martin, NCI

May 15, 2020

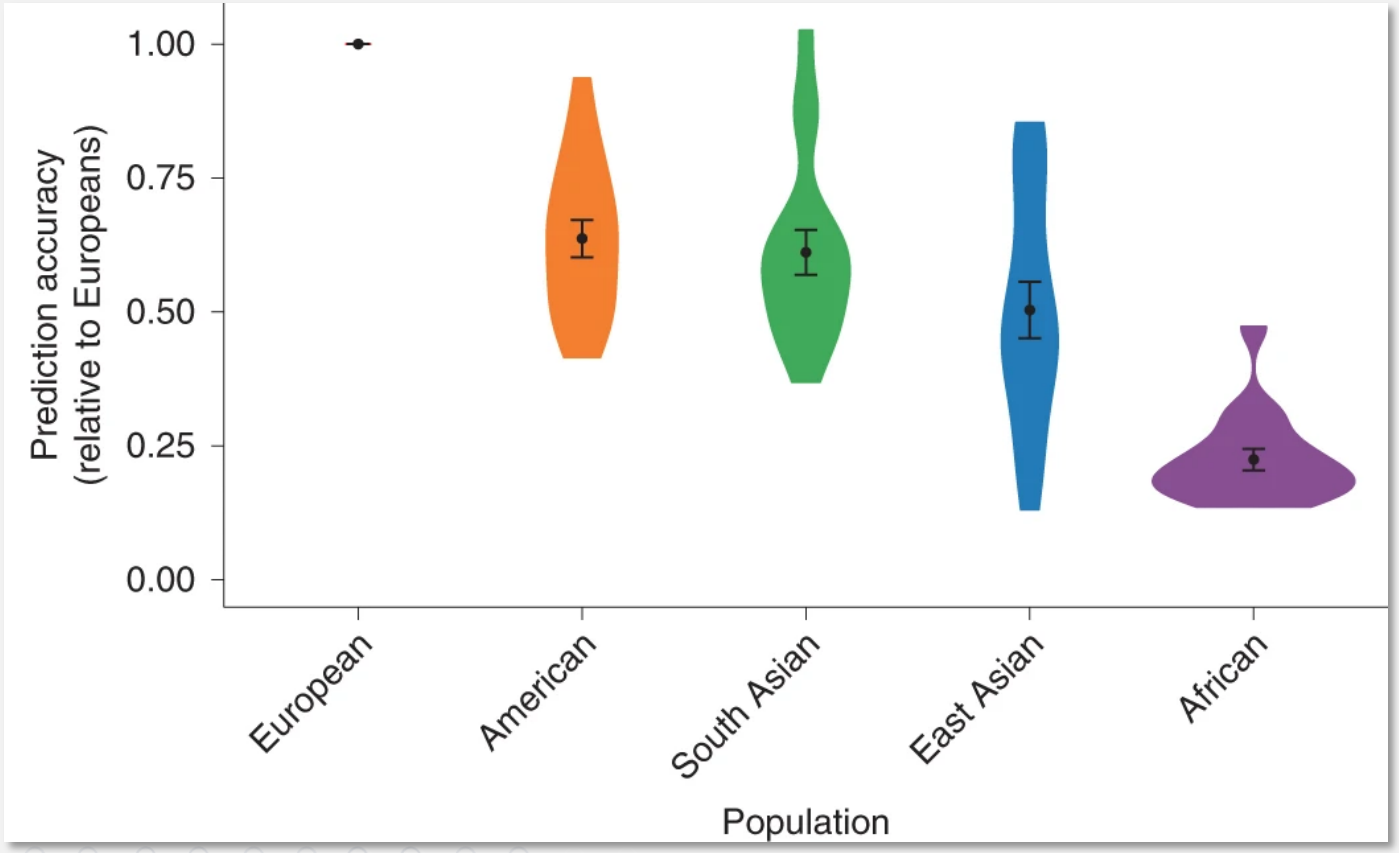
This webinar is being recorded.
WebEx questions or issues: email Catherine Sillari
Catherine.Sillari@nih.gov

National Human Genome Research Institute

NIH NATIONAL CANCER INSTITUTE

The Forefront of Genomics®

Poorer PRS prediction in non-European populations



- Prediction accuracy across 17 anthropometric and blood panel traits
- Lower prediction accuracy relative to Europeans:
  - American and South Asian: ~0.6-fold
  - East Asian: ~0.5-fold
  - African: ~0.25-fold

American = Hispanic/Latinx

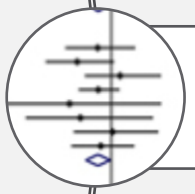
Martin, et al. *Nature Genet* 2019. PMID 30926966

NIH NHGRI

Accelerating scientific progress through collaboration

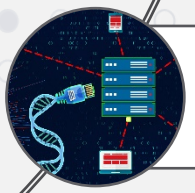
Methods development to improve applicability of PRS to diverse populations

Integration of large-scale cohort data dispersed across different resources



Importance of sample size in improving performance of PRS

One approach → multiple health/disease measures



Synergy with other data sharing, PRS efforts

NIH NHGRI

PRS concept: goals

1. Leverage genetic diversity to develop methods and improve the applicability of PRS across diverse populations and for a broad range of health and disease measures

2. Optimize the integration of large-scale, harmonized genomic and phenotype data to facilitate collaborative analysis, dissemination of PRS-related data, and development of related resources.

Common consortium objectives

Identify and integrate data for relevant cohorts

Standardize genomic and phenotype data, and map to existing ontologies

Develop and apply methods to generate and refine PRS for diverse populations

Establish external collaborations for PRS validation and implementation research

Identify secondary uses related to health and disease research

NIH NHGRI

“Diversity first”

Emphasize the use of non-EA data until maximum value has been extracted from them before exploring data from EA participants, even if the EA datasets are much larger and more frequently utilized.

Describe the scientific purpose and potential pitfalls of using data from potentially larger numbers of EA participants

Justify resulting biases (beyond simple convenience or expediency)

NIH NHGRI

Study site contributions

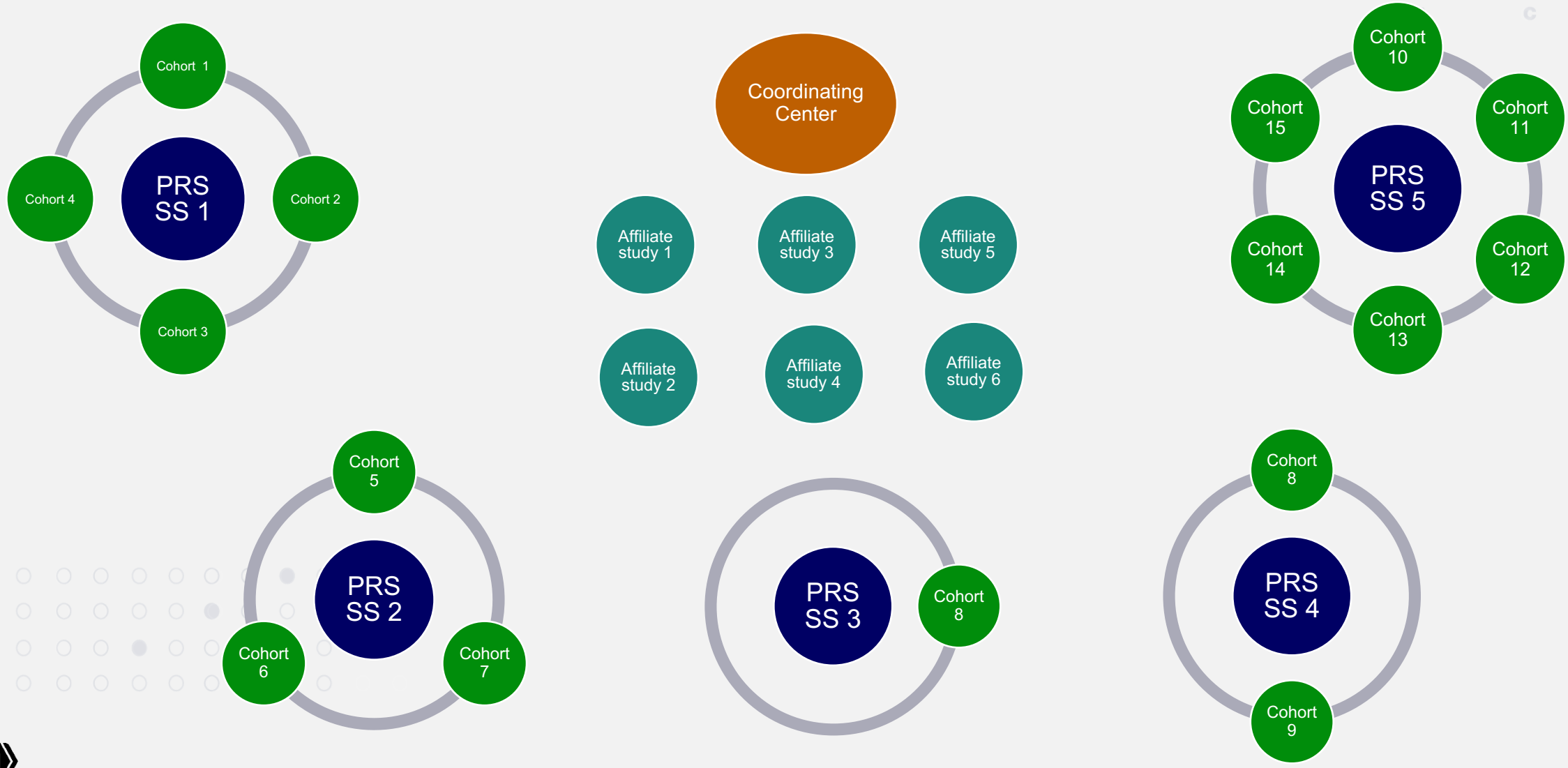
- Bring existing cohorts to maximize sample size, genetic diversity for cross-consortium analysis
- Address challenges related to differing availability of clinical data, data use limitations, availability of summary stats
- Identify and harmonize health/disease measures for analysis
- Integrate ancestry into analysis
- Identify metrics for improving PRS prediction
- Refine PRS based on updated data
- Participate in consensus approaches to developing and applying PRS
- Contribute to cross-consortium Working Groups
- See “Research Examples” section of RFA

Coordinating Center contributions

- Provide overall logistic and scientific coordination
- Lead data science aims
  - Propose FAIR approaches to data integration and analysis
  - Work with AnVIL and external standards groups
- Lead cross-consortium genotype imputation
- Lead cross-consortium outreach and dissemination efforts
- Provide/convene ELSI expertise
- Provide limited support for affiliate studies
- Provide limited genotyping

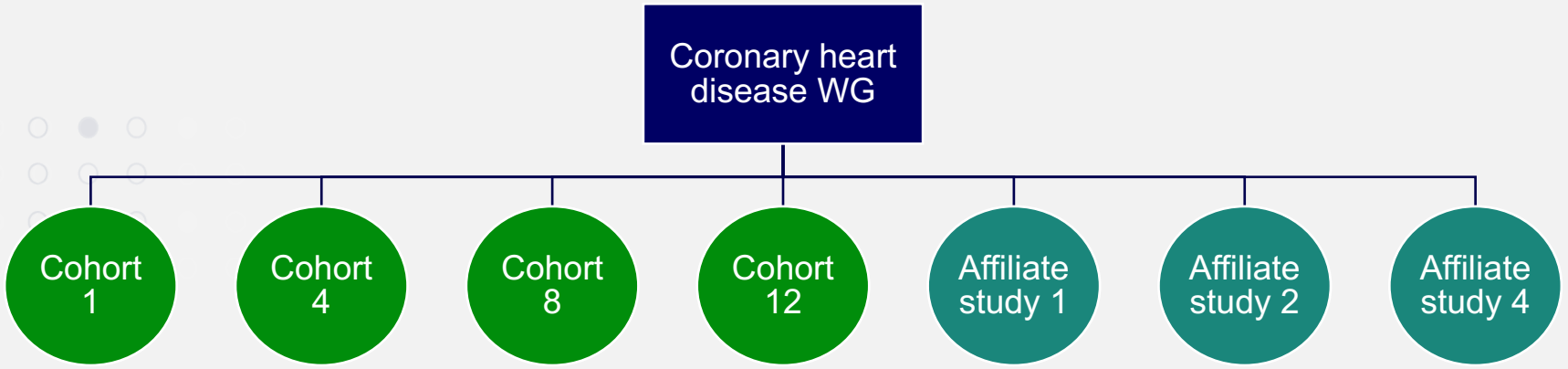
NIH NHGRI

PRS Consortium: hypothetical schematic



Example of cross-consortium focus: Working Groups

- Focal point for trait-specific, cross-consortium PRS analysis
- SS - contribute domain and analysis expertise
- CC - facilitate WG research



NIH NHGRI

Consortium deliverables

- Project datasets with harmonized data (summary statistics, meta-data; individual-level where possible)
- Consensus PRS models: SNPs, weights, covariates
- Tools/resources developed by PRS investigators
- Policies and standards to enable data sharing, including ELSI
- Data and approaches facilitating validation in clinical setting

AnVIL's Capabilities and Services

Cloud-based infrastructure and software platform

Shared analysis and computing environment

Data access and data security

Genomic datasets, phenotypes and metadata

Cost Control for Cloud services

User training and outreach

Participation in a federated genomic data commons ecosystem

Incorporation of scientific and technology advances

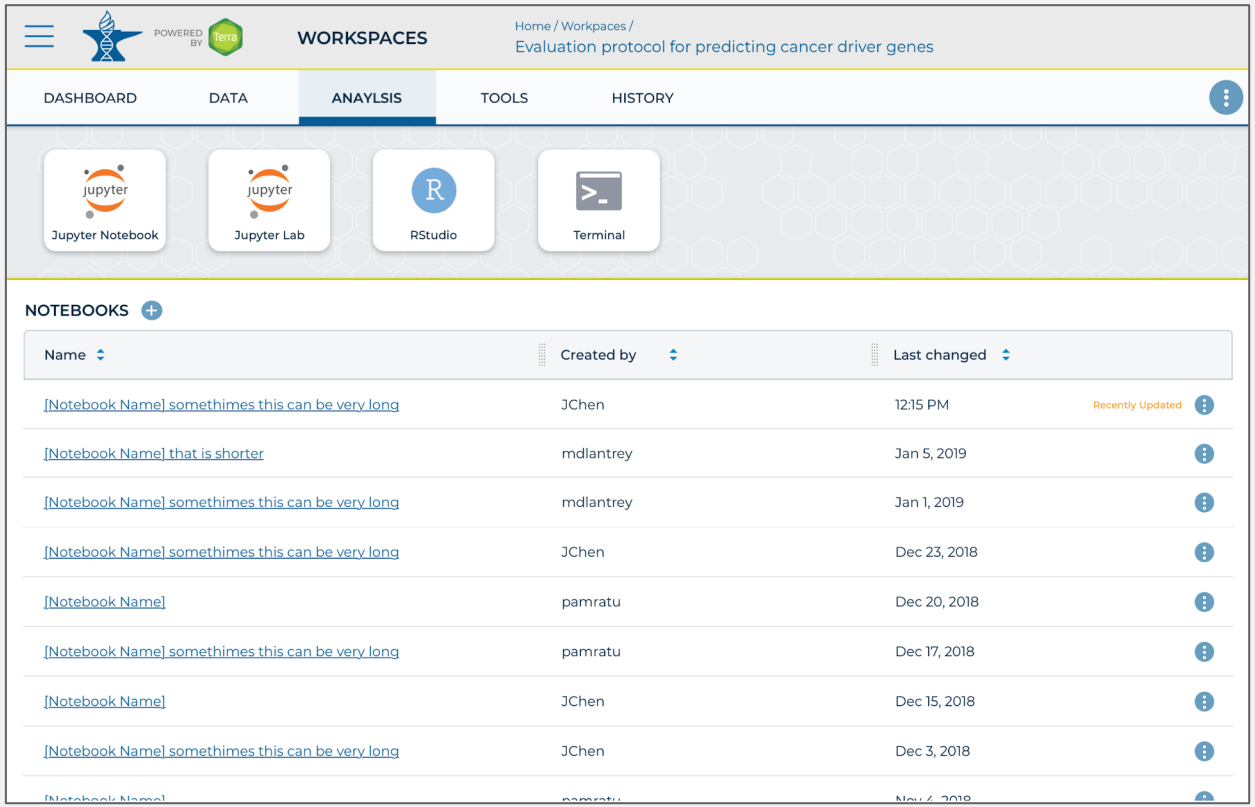
AnVIL / Terra: analysis workspaces and batch workflows

Faceted search

Established pipelines

Exploratory Analysis

Integrated development environment



Dockstore
Create, Share, Use

AnVIL / Dockstore: sharing containerized tools and workflows

RStudio

Galaxy

jupyter

{wdl}

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

UCSC
Genome Browser

NIH
NHGRI

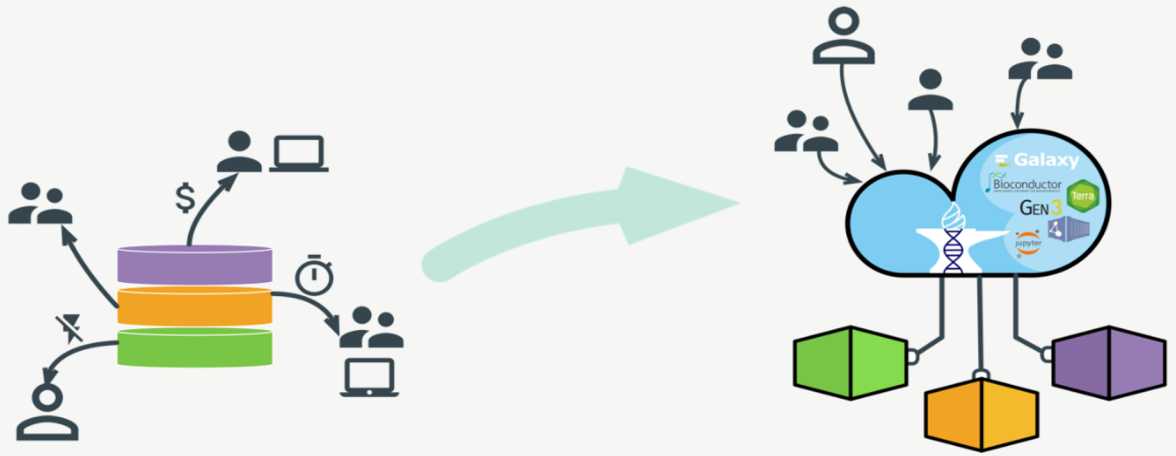
AnVIL

About | Data | Tools | Training | News | Events | Contact

Cloud-based Genomic Data Science

AnVIL – an Analysis, Visualization, and Informatics Lab-space for democratizing genomic data access, sharing and computing across large genomic-related data sets.

Learn More >



https://anvilproject.org

National Cancer Institute

- Cutting-edge research on cancer causes, treatment, and prevention
- Training the next generation of cancer researchers
- Funding and supporting the nation's vast network of scientists and cancer research institutions
- Informing and educating the American public and the world about cancer

National Cancer Institute

Goals of RFA are consistent with the NCI's priorities

- Addressing cancer disparities among minority populations as well as the cross-cutting theme of the Cancer Moonshot^{SM} to address health disparities.
- Underscores the urgency to ensure appropriate representation of minority populations to address translational gaps in genomic medicine
- Utilization of genetic information in prevention and treatment of cancer

Other RFA sections of note

- Data sharing in this Initiative
- Program Formation and Governance
- PHS398 Research Plan
- Application Review Information
  - Criteria
  - Review and Selection Process

Questions?

- WebEx: unmute yourself or type in chat box
- Email:
  - NHGRI: Lucia Hindorff hindorffl@mail.nih.gov
  - NCI: Damali Martin martinda@mail.nih.gov


The Forefront
of Genomics®

Extra slides

Applicant inquiries

- Relationship between site-specific and consortium priorities
- Can we “count” participants or datasets that
  - Are publicly available
  - Can’t share individual-level data
  - Don’t yet have genotype data
- What is the role of methods development
- Can one cohort participate in more than one application
- Any specific phenotypes on which to focus

Timeline

Y1-Y2

Integrate data

Convene WGs

Harmonize measures

Agree on PRS approach

Y4-Y5

Disseminate results

Further refinements based on community input

Y2-Y4

Collaborative PRS analyses

Refine models based on updated data

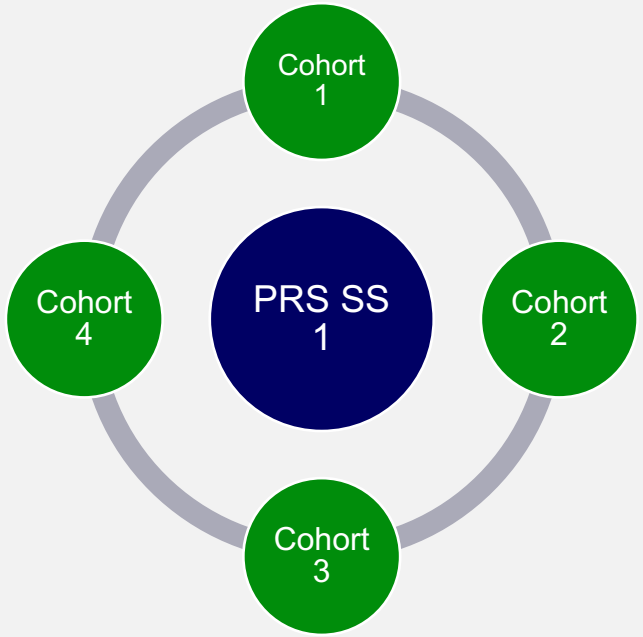
NIH NHGRI

PRS Study Sites (SS)

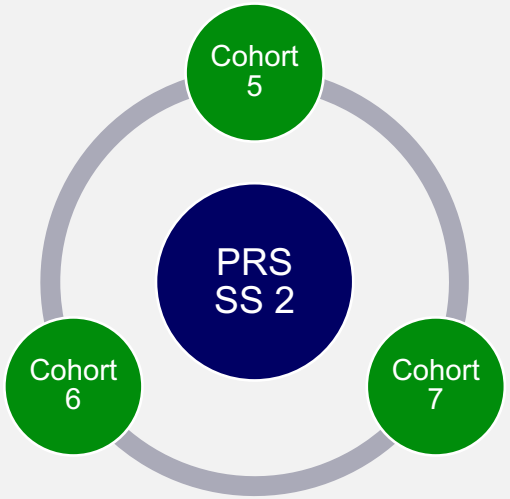
Study site	• One application representing one or more cohorts
Strongly encouraged*	• At least 1 non-EA group with ≥10,000 participants, OR • At least 20,000 participants, with at least 50% of participants from non-EA ancestry group
High priority*	• At least >50,000 participants • Large numbers (≥10,000) of non-EA participants • Broad phenotype information (multiple health and disease measures available) • Commitment to data sharing

* Within each SS, across all participating cohorts

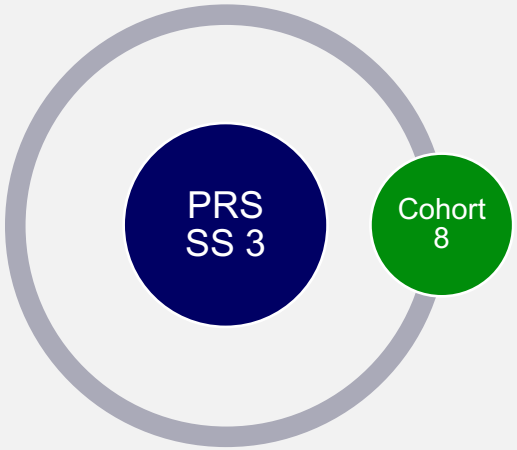
Examples of PRS SS



Cohort 1: 5,000 AA
Cohort 2: 6,000 AA, 2,000 H/L, 5,000 EA
Cohort 3: 1,000 Asian
Cohort 4: 8,000 EA



Cohort 5: 20,000 H/L
Cohort 6: 10,000 AA
Cohort 7: 25,000 EA



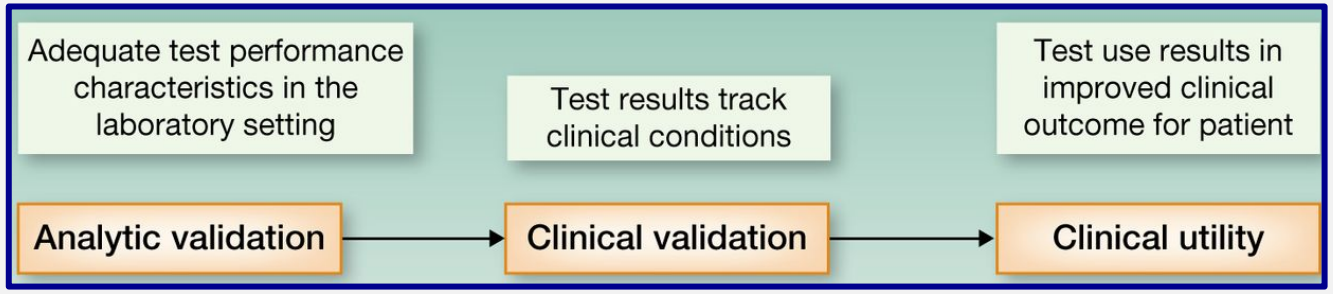
Cohort 8: 35,000 EA, 15,000 AA, 25,000 H/L

AA = African American
EA = European American
H/L = Hispanic/Latinx

NIH NHGRI

Relationship to other efforts

Other PRS efforts



Parkinson, et al. PMID 24634466

PRS concept

eMERGE

Examples of other cohorts eligible to apply

- All of Us
- Centers for Common Disease Genomics
- Electronic Medical Records in Genomics
- H3Africa
- International Common Disease Alliance
- International 100K Cohort Consortium
- Population Architecture using Genomics and Epidemiology
- Trans-Omics for Precision Medicine

NIH NHGRI