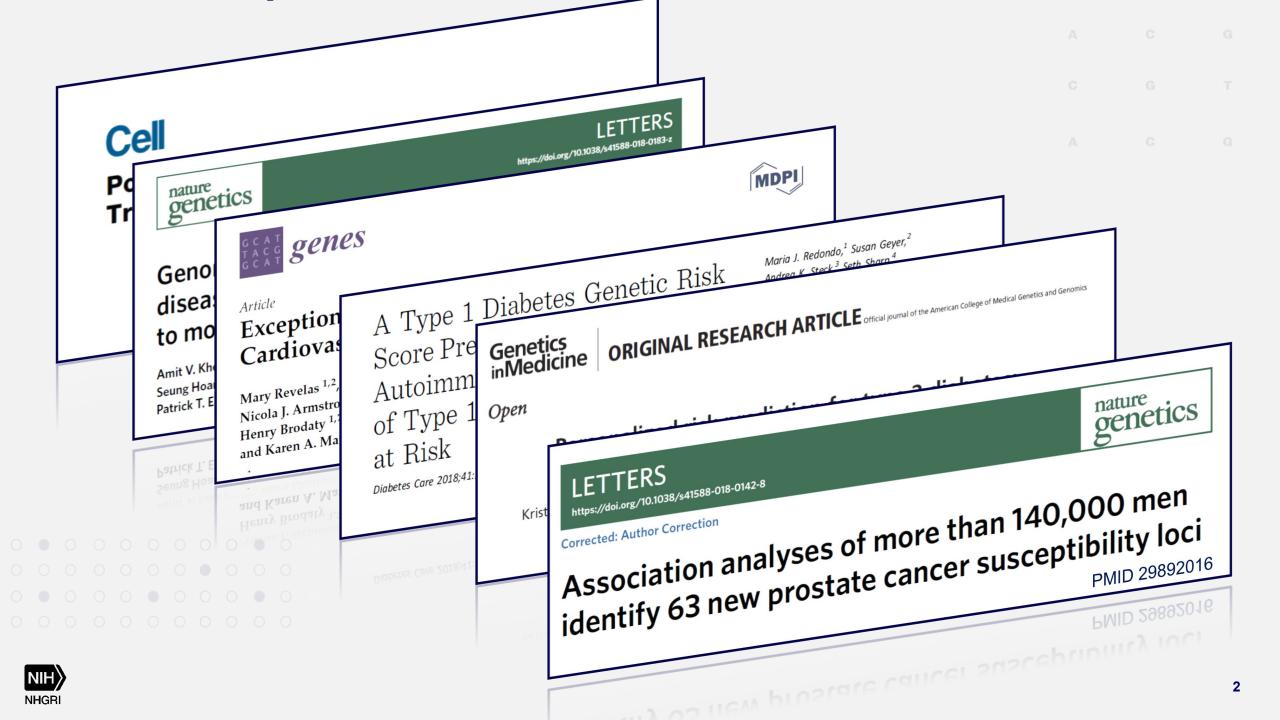
Concept for Sept. 2019 Council: Collaborative Data Integration and Analysis of Polygenic Risk Scores (PRS) from Diverse **Populations**

Lucia Hindorff

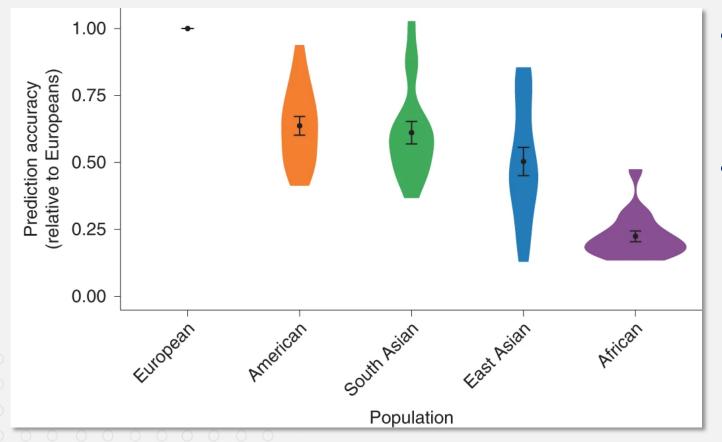
National Advisory Council on Human Genome Research September 16, 2019



The **Forefront** Genomics



Poorer PRS prediction in non-European populations



American = Hispanic/Latinx Martin, et al. *Nature Genet* 2019. PMID 30926966

- 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

Use the data we have

Summary	statistics	data
---------	------------	------

- dbGaP
- EGA
- UK Biobank
- GWAS Catalog

Large-scale genomic + phenotype data

- Centers for Common Disease Genomics
- Trans-Omics for Precision Medicine
- Population Architecture using Genomics and Epidemiology

"Hard-to-get" data

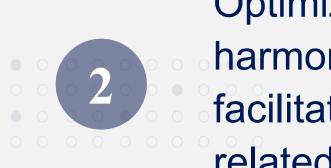
- Industry
- International

PRS concept: goals





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



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



Accelerating collaboration

Importance of sample size in improving performance of PRS

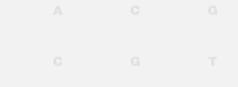
) Use of summary statistics in lieu of individual level data

) One approach \rightarrow multiple health/disease measures

Rapid translation of findings from epidemiological to clinical setting

Nascent global efforts requiring data sharing

PRS Centers (PRSCs)



PRSC <

Strongly encouraged*

High priority* ~

 At least 1 non-European ancestry (EA) group with >10,000 participants, OR

• One application representing one or more cohorts

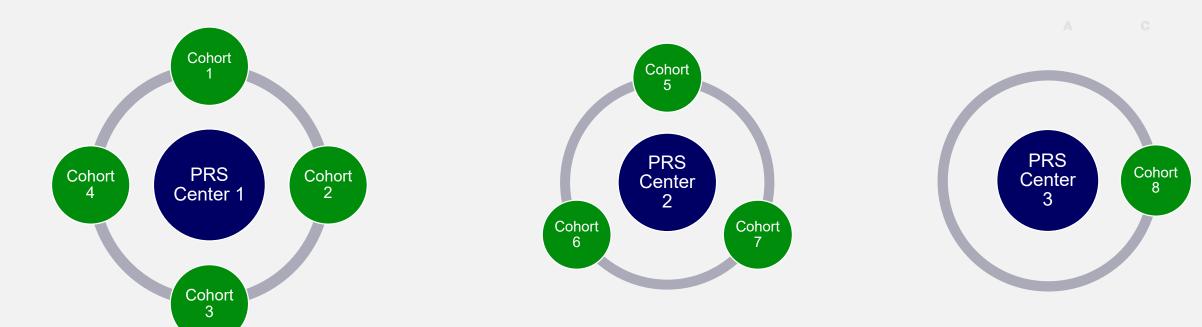
• At least 20,000 participants, with at least 50% of participants from non-EA populations

- 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 PRSC, across all participating cohorts



Examples of PRSCs



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
```



PRSC contributions to cross-consortium analysis

- Maximize sample size, genetic diversity available for crossconsortium 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



Coordinating Center

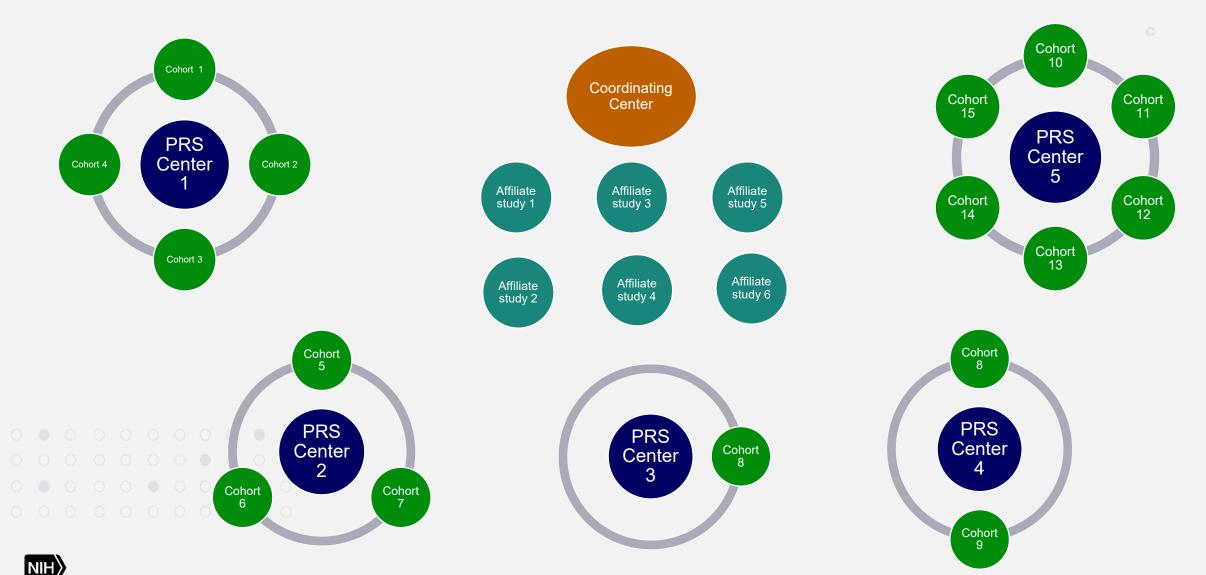
- Overall logistic and scientific coordination
- Lead data science aims
 - Propose FAIR approaches to data integration and analysis
 - Work with AnVIL and external standards groups
- Lead genotype imputation
- Disseminate PRS results
- Provide/convene ELSI expertise
- 0 0 0 0 0 0 0 0 0 0
- Provide limited support for affiliate studies
 - Provide limited genotyping

PRS Consortium

NHGRI



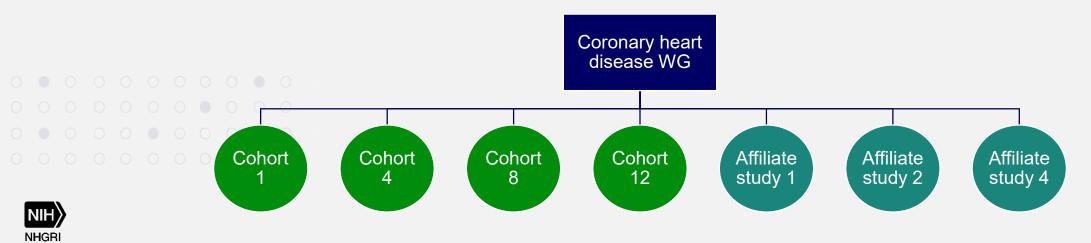
C G



11

PRS Working Groups

- Focal point for trait-specific, cross-consortium PRS analysis
- Leverage domain expertise
- Analysis efforts from all contributing cohorts

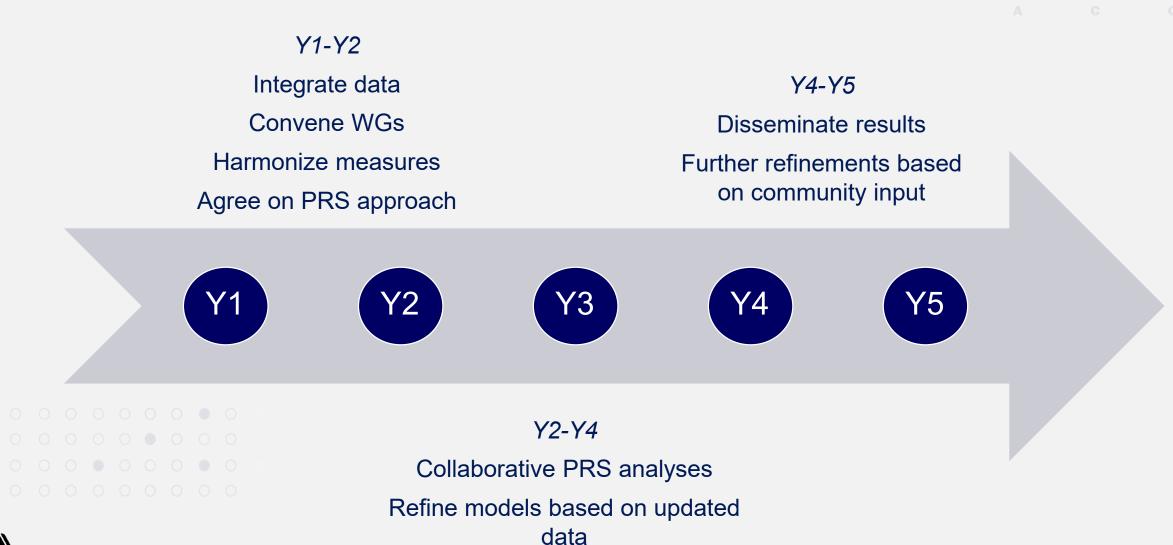


Deliverables

- Harmonized data (summary statistics, meta-data; controlled access where possible)
- PRS model: SNPs, weights, covariates
- Tools/resources developed by PRS investigators
- Policies and standards to enable data sharing, including ELSI
- $\bullet \circ \circ \circ \circ \circ \circ \circ \bullet \circ$
 - Data and approaches facilitating validation in clinical
 - setting



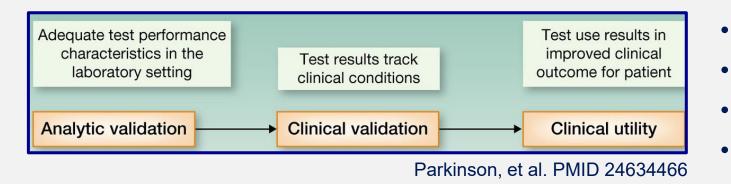
Timeline

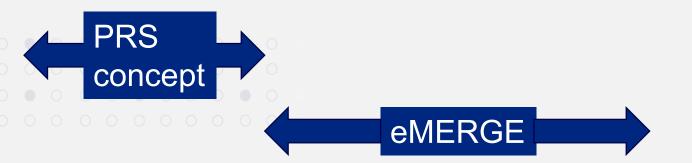




Relationship to other efforts

Other PRS efforts





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



Budget (in progress; in 1000's)

	Y1 (FY21)	Y2 (FY22)	Y3 (FY23)	Y4 (FY24)	Y5 (FY25)	5 year total
PRS Centers (5)	5,000	5,000	5,000	5,000	5,000	25,000
Coordinating Center	1,600	1,600	1,600	1,600	1,600	8,000
Limited genotyping	775	0	0	0	0	775
TOTAL	7,375	6,600	6,600	6,600	6,600	33,775



