Selection, optimization, and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse populations

Niall J Lennon*, Leah C Kottyan*, Christopher Kachulis, Noura Abul-Husn, Josh Arias, Gillian Belbin, Jennifer E Below, Sonja Berndt, Wendy Chung, James J. Cimino, Ellen Wright Clayton, John J. Connolly, David Crosslin, Ozan Dikilitas, Digna R. Velez Edwards, QiPing Feng, Marissa Fisher, Robert Freimuth, Tian Ge, The GIANT Consortium, The All of Us Research Program, Joseph T. Glessner, Adam Gordon, Candace Guiducci, Hakon Hakonarson, Maegan Harden, Margaret Harr, Joel Hirschhorn, Clive Hoggart, Li Hsu, Ryan Irvin, Gail P. Jarvik, Elizabeth W. Karlson, Atlas Khan, Amit Khera, Krzysztof Kiryluk, Iftikhar Kullo, Katie Larkin, Nita Limdi, Jodell E. Linder, Ruth Loos, Yuan Luo, Edyta Malolepsza, Teri Manolio, Lisa J. Martin, Li McCarthy, James B Meigs, Tesfaye B. Mersha, Jonathan Mosley, Bahram Namjou, Nihal Pai, Lorenzo L. Pesce, Ulrike Peters, Josh Peterson, Cynthia A. Prows, Megan J. Puckelwartz, Heidi Rehm, Dan Roden, Elisabeth A. Rosenthal, Robb Rowley, Konrad Teodor Sawicki, Dan Schaid, Tara Schmidlen, Roelof Smit, Johanna Smith, Jordan W. Smoller, Minta Thomas, Hemant Tiwari, Diana Toledo, Nataraja Sarma Vaitinadin, David Veenstra, Theresa Walunas, Zhe Wang, Wei-Qi Wei, Chunhua Weng, Georgia Wiesner, Yin Xianyong, Eimear Kenny.

* These authors contributed equally.
Author Affiliations at end of manuscript.

Corresponding editor: Niall Lennon (nlennon@broadinstitute.org)

Abstract

Polygenic risk scores (PRS) have improved in predictive performance supporting their use in clinical practice. Reduced predictive performance of PRS in diverse populations can exacerbate existing health disparities. The NHGRI-funded eMERGE Network is returning a PRS-based genome-informed risk assessment to 25,000 diverse adults and children. We assessed PRS performance, medical actionability, and potential clinical utility for 23 conditions. Standardized metrics were considered in the selection process with additional consideration given to strength of evidence in African and Hispanic populations. Ten conditions were selected with a range of high-risk thresholds: atrial fibrillation, breast cancer, chronic kidney disease, coronary heart disease, hypercholesterolemia, prostate cancer, asthma, type 1 diabetes, obesity, and type 2 diabetes. We developed a pipeline for clinical PRS implementation, used genetic ancestry to calibrate PRS mean and variance, created a framework for regulatory compliance, and developed a PRS clinical report. eMERGE's experience informs the infrastructure needed to implement PRS-based implementation in diverse clinical settings.

Introduction

Polygenic risk scores (PRS) are being calculated and disseminated at a prodigious rate[1], but their development and application to clinical care, particularly among ancestrally diverse individuals, present substantial challenges[2–4]. Incorporation of genomic risk information has the potential to improve risk estimation and management [3,5], particularly at younger ages [6]. Clinical use of PRS may ultimately prevent disease or enable its detection at earlier, more treatable stages [6–8]. Improved estimation of risk may also enable targeting of preventive or therapeutic interventions to those most likely to benefit from them while avoiding unnecessary testing or over-treatment [9,10].

PRS for individual conditions are typically generated from summary statistics derived from genome-wide association studies (GWAS), which are themselves derived from populations that are heavily over-represented by individuals of European ancestry [11]. Such scores have been shown to have limited prediction accuracy with increasing genetic distance from European populations [11,12]. PRS can be improved if developed and validated using multi-ancestry cohorts [13]. Clinical and environmental data combined with genomically-derived risk measurements can improve risk prediction [14]. Approaches for combining genomic and non-genomic information, optimizing models for genomically diverse populations and across age groups, and conveying this information to clinicians and patients have yet to be developed and applied in clinical care.

The Electronic Medical Records and Genomics (eMERGE) Network is a multicenter consortium established in 2007 to conduct genomic research in biobanks with electronic medical records [15,16]. In 2020, eMERGE embarked on a study of genomic risk assessment and management in 5,000 children and 20,000 adults of diverse ancestry, beginning with efforts to identify and validate published PRS

across multiple race-ethnic groups (and inferred genetic ancestries) in 10 common diseases with complex genetic etiologies. This paper describes identification, selection, and optimization of these PRS; calibration of ancestry for PRS estimation using a novel method developed for eMERGE; and development and launch of clinical reporting tools.

Results

PRS Auditing and Evaluation

To select the PRS for clinical implementation, the network conducted a multi-stage process to evaluate proposed scores (Figure 1). An initial set of 23 conditions was selected based on considerations including relevance to population health (condition prevalence and heritability), strength of evidence for PRS performance, clinical expertise in the eMERGE Network, and data availability that would facilitate validation of the PRS in diverse populations. Network sites completed comprehensive literature reviews on 23 proposed conditions and the corresponding PRS. A summary of the features of the PRS for each of the final conditions chosen is shown in Supplemental Table 1. The collated information included *analytic viability* - a description of covariates, the age, and ancestry effects of the original PRS model; *feasibility* - access to sufficiently diverse validation data sets (race/ethnicity and age) as well as condition prevalence and relevance to preventative care; *potential clinical actionability* - existing screening or treatment strategies, and magnitude (odds ratio) of risk in the high risk group; and *translatability* - expected public health impact across diverse populations. Candidate PRS were restricted to those that were either previously validated and published (journal or pre-print) or for which there was sufficient access to information to develop and/or optimize new PRS, which could then be validated.

In auditing and evaluating evidence of PRS performance, the eMERGE Steering Committee (SC) considered PRS for conditions that could be implemented in pediatric and/or adult populations, and for diseases with a range of age of onset (0 to >65 years of age). We considered published SNP-based heritability estimates available for 10 of the 23 conditions, ranging from 3% to 58%. The majority of PRS under consideration aimed to identify individuals at high risk for disease; however, PRS to predict disease severity and drug response were also considered. Two of the conditions, breast cancer and prostate cancer, were only considered for implementation in individuals whose biological sex was female or male, respectively. As the eMERGE network plans to enroll >50% participants from underrepresented groups (including racial and ethnic minority groups; people with lower socioeconomic status (SES); underserved rural communities; sexual and gender minority (SGM) groups) [17] emphasis was placed on the PRS that were already available for, or could be developed and validated in, diverse population groups.

To define population groups, study-level population descriptors were first extracted from published literature, pre-prints or information shared directly by collaborators on data used to develop and/or optimize and/or validate PRS. Methods for using population labels across studies ranged from self-reporting, extraction from health system data, and/or analysis of genetic ancestry. We designated four population groups; European ancestry (EA) (i.e. study population descriptors included European, European-American, or other European descent diaspora groups), African (African, African-American (AA), or other African descent diaspora groups), Hispanic (HL) (i.e. Hispanic, Latina/o/x, or those who have origins in countries in the Caribbean and Latin America), and Asian (Asn) (i.e. South Asian, East Asian, South-East Asian, Asian-American or other diaspora Asian groups).

Of the 23 conditions initially selected, six were excluded at the outset (August 2020) due to lack of diversity in the PRS training or optimization data, lack of access to diverse datasets for validation, or lack of available clinical expertise in the network (Figure 1). A further four conditions were dropped in March 2021 due to insufficient confidence in PRS performance or lack of validation datasets. Conditions not prioritized for implementation continued on a 'developmental' pathway for further refinement. Each of the 12 conditions that were selected to move forward from the March 2021 review were assigned a 'lead' and 'co-lead' site which worked together to develop, validate, and transfer the score to the clinical laboratory for instantiation and CLIA validation. Assignment of leads was based on site preference, expertise, and distribution of workload.

Selection, optimization, and validation

A systematic framework was developed to evaluate the performance for the remaining 12 PRS, in accordance with best practices outlined in Wand *et al* [18]. An in depth evaluation matrix of the 12 chosen conditions can be found in Supplemental Table 2. Clinical use of Eurocentric PRS in diverse patient samples risks exacerbating existing health disparities[11][19,20]. The Network carefully considered a variety of strategies to optimize PRS generalizability and portability. The Network prioritized validation across four ancestries with an emphasis on African and Hispanic ancestry due to their underrepresentation in genetic research and projected representation within the study cohort. We determined that a PRS was validated if the genomic predictor was significantly discriminative and the odds ratios were statistically significant in a minimum of two and up to four ancestral populations: African/African-American, Asian, European Ancestry; and Hispanic/Latino. The PRS Working Group members conducted an extensive scoping exercise to identify suitable datasets of multiple ancestries for disease-specific PRS validation. These included datasets from early phases of eMERGE (2007-2019) as well as external datasets such as the UK Biobank and Million Veteran Program (MVP). A standardized set of questions were addressed by the disease leads that included the source of discovery and validation datasets, the availability of multi-ancestry validation datasets, availability of cross-ancestry PRS, proposed percentile thresholds for identifying high risk status, model

discrimination (AUC), and effect sizes (odds ratios) associated with high risk vs. not-high risk status (Supplemental Table 2). For 7 out of the 12 candidate scores, no further optimization of the original model was performed. For 5 scores, an additional optimization effort was undertaken to further refine the score performance in multiple ancestries. Details of the optimization can be found in Supplemental Table 3. A specific score optimization was applied for CKD. This optimization consisted of adding the effect of APOL1 risk genotypes to a polygenic component, which has been found to improve risk predictions in African ancestry cohorts[21].

For final selection, the steering committee considered the score performance summaries (presented by condition leads) in addition to the actionable and measurable recommendations relevant for return, for each condition, in the prospective cohort. Two conditions (colorectal cancer (CRC) and abdominal aortic aneurysm (AAA)) were moved to the developmental pathway (Figure 1). While the PRS for CRC was not included in the prospective cohort, as several Mendelian genes for Lynch syndrome were included in the custom panel generated by Invitae, the network decided to include CRC in the list of conditions returned with the overall risk report (GIRA) and only report on monogenic and family history risk.

Population-based z-score calibration

In this study, the focus is on integration and implementation of validated PRS in clinical practice rather than novel PRS development. Ultimately, the Network opted to balance generalizability and feasibility by validating and returning cross-ancestry PRS. However, even with cross-ancestry scores, differences remain in the distribution of z-scores across genetic ancestries that can result in inconsistent categorization of individuals into ‘high’ or ‘not high’ polygenic risk categories for a given condition [22]. To that end, the Network chose to develop methods to determine each participant's ancestry and calibrate the distribution of resulting z-scores through a population-based calibration model[22][23] (see below). An alternative would have been to apply existing PRS in available samples of different ancestries and derive ancestry-specific effect estimates. However, returning ancestry-specific risk estimates is challenging in real world implementations as it would require self-reporting of ancestry by patients (who may not be able to provide this with accuracy) and developing multiple ancestry-specific reports for each health condition. In addition, such PRS would be problematic to return to patients of mixed ancestry.

Polygenic risk scores often have different mean and standard deviation for individuals from different genetic ancestries. While some of these differences could be due to true biological differences in risk, they also result from allele frequency and linkage disequilibrium (LD) structure differences between populations [24]. This problem is more acute when a PRS is calculated for an individual whose ancestry does not match the ancestries used to develop the PRS. A clinically implemented PRS test to return disease risk estimates, therefore, must be adjusted to account for these differences due to

ancestral background. A calibration method based on principal component analysis (PCA) which was initially described by Khera *et al*. [22] was modified to model both the variance and means of scores as ancestry dependent, as compared to the previous method, which modeled only the means as dependent on ancestry. This modification was found to be necessary because some conditions were found to exhibit highly ancestry-dependent variance (see for example, Figure 3 in Online Methods), which would have led to many more or fewer participants of certain ancestries receiving a 'high PRS risk' determination than was intended. The model was fit to a portion of the All of Us Research Program (https://www.researchallofus.org/) cohort genotyping data, which allowed for continuous return of results to participants without needing to wait for the entire study dataset to be available. More details can be found in Online Methods.

Transfer and Implementation

Once the final 10 conditions had been selected, condition-leads worked with computational scientists at the clinical laboratory (Clinical Research Sequencing Platform, LLC at the Broad Institute) to transfer the PRS models. Condition-specific models were run with outputs from the lab's genotyping (Illumina Global Diversity Array), Phasing (Eagle2 [25] https://github.com/poruloh/Eagle), and imputation (Minimac4 [26] https://genome.sph.umich.edu/wiki/Minimac4) pipelines to assess genomic site representation (see Online Methods for more information on the architecture and components of the pipeline). Several rounds of iteration between the clinical laboratory and condition-leads followed in which any issues with the pipeline were resolved and the effect of genomic site missingness was assessed (Table 1). The final version of the implemented models was returned to the condition leads to recalculate effect sizes in the validation cohorts.

Finally, as part of the implementation of the PRS pipelines as a clinical test in a CLIA laboratory, a validation study was performed (See Online Methods for a detailed description, Table 1 summarizes some of the results). Briefly, this study leveraged 70 reference cell lines from diverse ancestry groups (Coriell) where 30X whole genome sequencing data was generated to form a variant truth set from which the technical accuracy and reproducibility of imputation and PRS calling was assessed. A second sample set of 20 matched donor blood and saliva specimens was procured to assess the performance of the pipeline with different input materials. A set of three samples, each with 6 replicates, was run end-to-end through the wet lab and analytical pipelines as an assessment of reproducibility. As a verification of the clinical validity of the scores, cohorts of cases for 8 of the 10 conditions were created using the eMERGE phase III imputed dataset (available on https://anvil.terra.bio/#workspaces/anvil-datastorage/AnVIL_eMERGE_GWAS/data (registration required)). PRS performance measures were calculated to confirm associations between scores and conditions. Due to limitations in the eMERGE phase III imputation (no chromosome X, different imputation pipeline) the ORs from this analysis were not included in the final reports, rather the ORs

calculated in the condition-specific validation cohorts (using the final clinical lab pipeline) were used (Figure 3 and Table 1). A validation report was created for each condition. This report was reviewed and approved by the Laboratory Director in compliance with CLIA regulations for the development of a laboratory developed test. Personnel were trained on laboratory and analytical procedures, and standard operating procedures were implemented. Data review metrics were established, sample pass/fail criteria were defined, and order and report data transfer pipelines were built as described in Linder *et al.* [27]

Creation of report and pipeline for report creation, review, sign-out, release

A software pipeline was built to facilitate data review and clinical report generation in both document (pdf) and structured data formats (sample report included in Supplementary Material). Logic was built into the PRS and reporting pipeline to account for differences in return based on age and sex at birth for certain conditions. For instance, the PRS for breast cancer is only calculated for participants who report sex at birth as female; similarly prostate cancer scores are only generated for participants who report sex at birth as male. Age-related restrictions were similarly coded into the pipeline to account for study policies on return. Data review by an appropriately qualified, trained individual is required for high complexity clinical testing. In the PRS clinical pipeline this review takes the form of a set of metrics that are exposed by the pipeline to the reviewer. These include a z-score range for each condition (passing samples will have a score $-5 < z < +5$), a PCA plot per batch against a reference sample set (visual representation of outlier samples), monitoring the z-score range for each control per condition (one control on each plate; NA12878), and flagging any samples with multiple 'High Risk' results for further review.

Each participant's sample is also run on an orthogonal fingerprinting assay (Fluidigm biomark) that creates a genotype-based fingerprint for that DNA aliquot. Infinium genotyping data is compared to this fingerprint as a primary check of sample chain-of-custody fidelity and to preclude sample or plate swaps during lab processing.

Reviewed and approved data for a participant is processed into a clinical report. The text and format of this report were created during an iterative review process by consortium work groups. For this pragmatic clinical implementation study, two results are returned to participants: "High Risk" or "Not High Risk" based on the PRS [27]. In the clinical report a qualitative framework has been developed to indicate for which condition(s) a participant has been determined to have a high PRS (if any). Quantitative values (Z-scores) are not included for any condition in the main results panel. For breast cancer and CHD, the z-score is presented in another section of the report for inclusion in integrated score models for those conditions. For breast cancer specifically, the provided z-score is used with the BOADICEA[28] model to generate an integrated risk that is included in the genome-informed risk assessment (GIRA) as described in Linder *et al*.[27]

Overview of first 2500 clinical samples processed.

Between launch in July 2022 and May 2023, 2500 participants have been processed through the clinical PRS pipeline (representing ~10% of the proposed cohort). Of the first 2500 participants processed, 64.5% (1612) indicated sex at birth as female, while 35.5% (886) indicated male. Median age at sample collection was 51 years (range 3 years to 75 years). Participants self-reported race/ancestry, with 32.8% (820) identifying as “White (e.g. English, European, French, German, Irish, Italian, Polish, etc)”; 32.8% (820) identified as “Black, African American, or African (e.g. African American, Ethiopian, Haitian, Jamaican, Nigerian, Somali, etc.)”; 25.4% (636) identified as “Hispanic, Latino, or Spanish (e.g. Colombian, Cuban, Dominican, Mexican or Mexican American, Puerto Rican, Salvadoran, etc.)”; 5% (124) identified as “Asian (e.g. Asian, Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, etc.)”; 1.5% (38) identified as American Indian or Alaska Native (e.g. Aztec, Blackfeet Tribe, Mayan, Navajo Nation, Native Village of Barrow (Utqiagvik) Inupiat Traditional Government, Nome Eskimo Community, etc.); 0.9% (22) identified as Middle Eastern or North African (e.g. Algerian, Egyptian, Iranian, Lebanese, Moroccan, Syrian, etc.); 0.8% (21) selected “None of these fully describe [me_or_my_child]”; 0.7% (17) selected “Prefer not to answer”; 0.1% (2) participants had incomplete data. A summary of the performance of the first 2500 samples and resulting high PRS metrics are shown in Figure 4. In the first 2500 participants, we identified 515 participants (20.6%) with a high PRS risk for one of the 10 conditions, 64 participants (2.6%) had high PRS risk for two conditions, and two participants (0.08%) had a high risk for three conditions. The remaining 1919 participants had no high PRS found. High PRS participants spanned the spectrum of genetic ancestry when projected onto principal component space (Figure 4).

Discussion

While the predictive performance of PRS has improved significantly in recent years, challenges remain in ensuring that PRS are applicable and effective in diverse populations. In particular, the vast majority of GWAS have focused on individuals of European ancestry, and the predictive accuracy of PRS declines with increasing genetic distance from the discovery population[29][24][4]. This risks exacerbating existing health disparities, as clinical use of Eurocentric PRS in diverse patient samples may not accurately reflect disease risk in non-European populations. To address these challenges, the eMERGE Network has conducted a multi-stage process to evaluate and optimize PRS selection, development, and validation. The network has prioritized conditions with high prevalence and heritability, existing literature, clinical actionability, and the potential for health disparities, and has developed strategies to optimize PRS generalizability and portability across diverse populations. In particular, the network has emphasized performance across four major ancestry groups (African, Asian,

European, Hispanic, as reflected by self-identified race/ethnicity) and has developed a pipeline for clinical PRS implementation, a framework for regulatory compliance, and a PRS clinical report. The potential impact of PRS-based risk assessment in clinical practice is significant. By enabling targeted interventions and preventative measures, PRS-based risk assessment has the potential to reduce the burden of a range of conditions [27]. Moreover, the development of PRS-based risk assessment in diverse populations has the potential to reduce health disparities by ensuring that clinical use of PRS accurately reflects disease risk in diverse populations.

However, challenges remain in the successful implementation of PRS-based risk assessment in clinical practice. These include concerns about genetic determinism, the potential for stigmatization, and the need for robust regulatory frameworks to ensure that PRS-based risk assessment is deployed safely and effectively. Additionally, one of the biggest challenges is the implementation of effective disease prevention strategies after the return of the results. Return of the results won't result in a benefit without effective disease prevention or early detection strategies. The eMERGE Network's work provides a promising blueprint for addressing these challenges, but ongoing research and evaluation will be necessary to ensure that PRS-based risk assessment is implemented in a responsible and effective manner.

In conclusion, the eMERGE Network's work in PRS development represents a significant step forward in the implementation of PRS-based risk assessment (in combination with other risk estimates from monogenic testing and family history) in clinical practice. By leveraging the power of genetics to predict disease risk and enable targeted interventions, genetically-informed risk assessment has the potential to revolutionize personalized medicine and usher in a new era of precision health. While challenges remain in ensuring that PRS are applicable and effective in diverse populations, the eMERGE Network's work provides a promising foundation for the continued development and evaluation of PRS-based risk assessment in clinical practice.

Funding

This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG011172 (Cincinnati Children’s Hospital Medical Center); U01HG011175 (Children’s Hospital of Philadelphia); U01HG008680 (Columbia University); U01HG011176 (Icahn School of Medicine at Mount Sinai); U01HG008685 (Mass General Brigham); U01HG006379 (Mayo Clinic); U01HG011169 (Northwestern University); U01HG011167 (University of Alabama at Birmingham); U01HG008657 (University of Washington); U01HG011181 (Vanderbilt University Medical Center); U01HG011166 (Vanderbilt University Medical Center serving as the Coordinating Center). The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA#: AOD

16037; Federally Qualified Health Centers: 75N98019F01202.; Data and Research Center: 1 OT2 OD35404; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 OT2 OD030043; Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276.

Acknowledgements

The authors would like to thank the past and future participants of the eMERGE network projects. The authors would like to thank Mary O'Reilly for help with figure creation. We thank our All of Us Research Program colleagues, Andrea Ramirez, Sokny Lim, Brandy Mapes, Ashley Green and Anji Musick for providing their support and input throughout the ancestry calibration demonstration project lifecycle. We also thank the All of Us Science Committee and All of Us Steering Committee for their efforts evaluating and finalizing the approved demonstration projects. The All of Us Research Program would not be possible without the partnership of contributions made by its participants. To learn more about the All of Us Research Program's research data repository, please visit https://www.researchallofus.org/.

Author Information

These Authors contributed equally Niall Lennon, Leah C.Kottyan

Authors and Affiliations

Broad Institute of MIT and Harvard
Marissa Fisher, Candace Guiducci, Maegan Harden, Chris Kachulis, Amit Khera,Katie Larkin, Niall J. Lennon, Edyta Malolepsza, Li McCarthy, Nihal Pai, Heidi Rehm, Diana Toledo

Children's Hospital Boston
Joel Hirschhorn

Children's Hospital of Philadelphia
John J. Connolly, Joseph T. Glessner, Hakon Hakonarson, Margaret Harr

Cincinnati Children's Hospital Medical Center
Leah C. Kottyan, Lisa J. Martin, Tesfaye B. Mersha, Bahram Namjou, Cynthia A. Prows

Columbia University
Wendy Chung, Atlas Khan, Krzysztof Kiryluk, Chunhua Weng

Fred Hutchinson Cancer Center and University of Washington
Li Hsu, Ulrike Peters

Icahn School Mount Sinai
Gillian Belbin, Clive Hoggart, Roelof Smit, Zhe Wang

Invitae
Tara Schmidlen

Mass General Brigham
Tian Ge, Elizabeth W. Karlson, Elizabeth W. Karlson, James B Meigs, Jordan W. Smoller

Mayo Clinic
Ozan Dikilitas, Robert Freimuth, Iftikhar Kullo, Dan Schaid, Johanna Smith

National Institutes of Health
Josh Arias, Sonja Berndt, Teri Manolio, Robb Rowley

Northwestern University
Adam Gordon, Yuan Luo, Lorenzo L. Pesce, Megan J. Puckelwartz, Konrad Teodor Sawicki, Theresa Walunas

University of Alabama at Birmingham
James J. Cimino, Ryan Irvin, Nita Limdi, Hemant Tiwari

University of Copenhagen
Ruth Loos

University of Michigan
Yin Xianyong

University of Washington
David Crosslin, Minta Thomas, David Veenstra

University of Washington Medical Center
Gail P. Jarvik, Elisabeth A. Rosenthal

Vanderbilt University Medical Center
Jennifer E Below, Ellen Wright Clayton, Digna R. Velez Edwards, QiPing Feng, Jodell E. Linder, Jonathan Mosley, Josh Peterson, Dan Roden, Nataraja Sarma Vaitinadin, Wei-Qi Wei, Georgia Wiesner

The GIANT Consortium

The All of Us Research Program

Conflict of Interest

The authors have no conflicts of interest to declare.

Tables and Figures

Figure 1

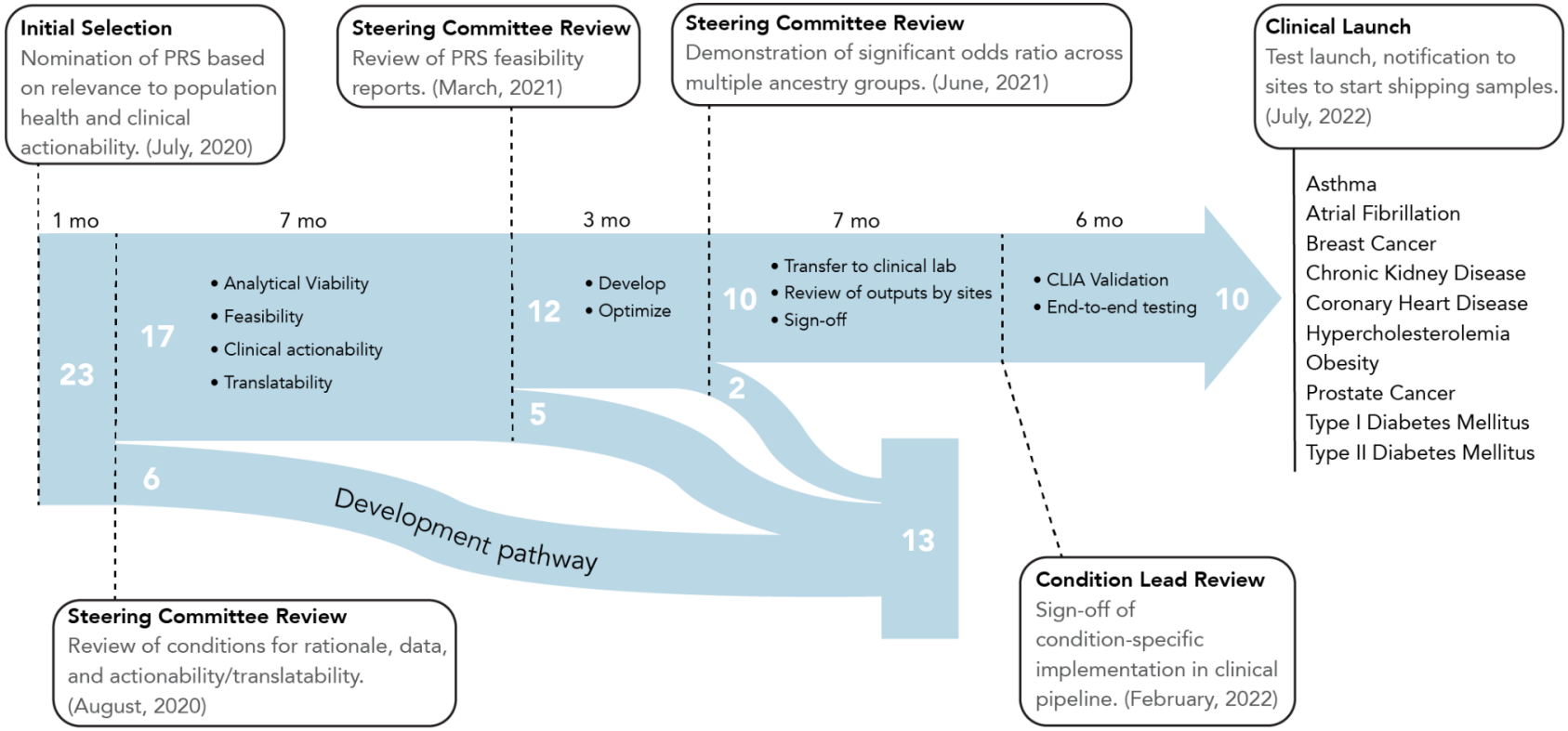

Figure 1. Timeline and process for selection, evaluation, optimization, transfer, validation, and implementation of the clinical PRS test pipeline. Dotted lines represent pivotal moments in the progression of the project.

Figure 2

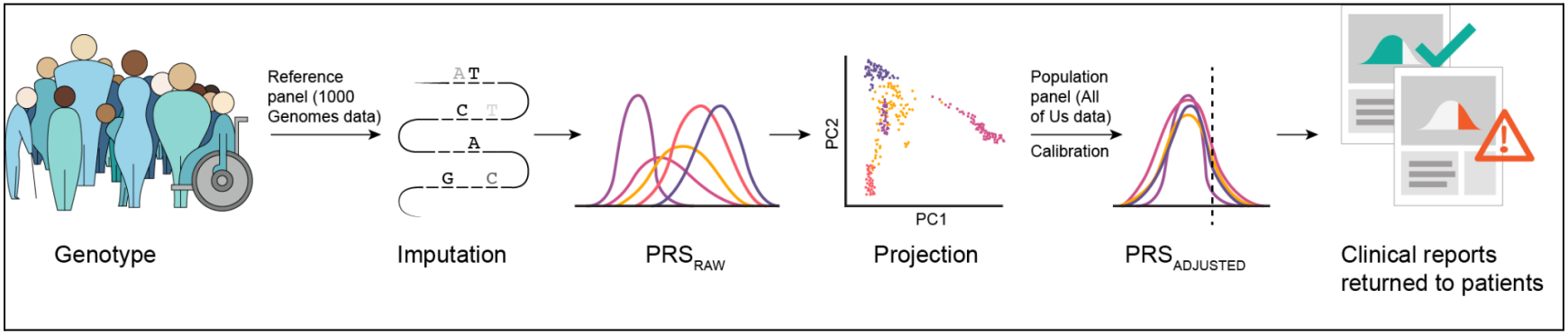

Figure 2. Overview of the eMERGE PRS process. Participant DNA is genotyped using the Illumina Global Diversity Array which assesses 1.8M sites. Genotyping data is phased and imputed with a reference panel derived from the 1000 Genomes Project. Raw PRS scores are calculated for each condition. For each condition an ancestry calibration model is applied based on model parameters derived from the All of Us Research Program. Participants whose adjusted scores cross the pre-defined threshold for high PRS are identified and a pdf report is generated. The report is electronically signed after data review by a clinical laboratory director and delivered to the study portal for return to the clinical sites.

Figure 3

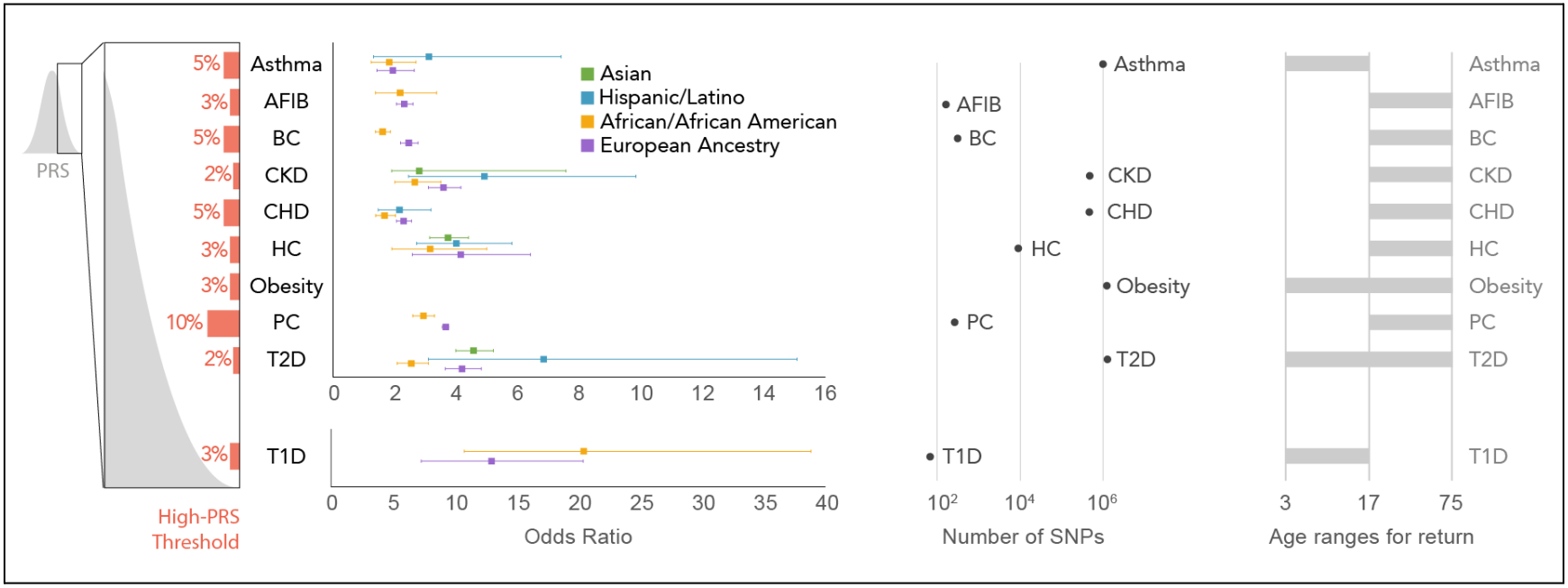

Figure 3. Summary of the ten conditions that were implemented. “High-PRS Threshold” represents the percentile that is deemed to be the cut-off for a specific condition above which a high PRS result is reported for that condition. The Odds Ratios are the OR of the implemented scores, 95% confidence interval shown in the whiskers (with the exception of Obesity for which the OR will be published by the GIANT consortium). “Number of SNPs” represents the range of numbers or sites included in each score. “Age ranges for return” indicates the participant ages at which a PRS is calculated for a given condition. AFIB= Atrial fibrillation; BC = Breast Cancer; CKD = Chronic Kidney Disease; CHD = Coronary Heart Disease; HC = Hypercholesterolemia; PC = Prostate Cancer; T2D = Type 2 Diabetes; T1D = Type 1 Diabetes.

Table 1. Performance measures from the PRS pipeline validation study at the clinical laboratory.

		Asthma	Atrial Fibrillation	Breast Cancer	Chronic Kidney Disease	Coronoary Heart Disease	Hypercholesterolemia	Obesity/BMI	Prostate Cancer	Type 1 Diabetes	Type 2 Diabetes
PRS Accuracy: Pearson Correlation between PRS from array and WGS (%)		99.3	98.6	93.0	98.3	98.2	95.9	99.5	96.4	99.5	98.8
PRS Precision: PRS pipeline repeatability (%)		100	100	100	100	100	100	100	100	100	100
PRS Precision: PRS pipeline reproducibility (z-score std dev)		0.0020	0.0010	0.0040	0.0001	0.0010	0.0050	0.0020	0.0006	0.0001	0.0010
PRS site missingness (%)		0.69	1.20	0.32	0.69	0.46	1.20	0.70	2.97	2.97	0.70
Odds Ratio (95% Confidence Interval)	European	1.95 (1.43-2.65)	2.32 (2.07-2.61)	2.47 (2.20-2.77)	3.6 (3.11-4.17)	2.3 (2.07-2.56)	4.16 (2.59-6.44)		3.67 (3.57-3.76)	12.97 (7.29-20.40)	4.21 (3.66-4.84)
	African American	1.83 (1.24-2.70)	2.19 (1.38-3.38)	1.61 (1.38-1.87)	2.66 (2.01-3.51)	1.68 (1.39-2.03)	3.16 (1.92-5.01)		2.95 (2.60-3.30)	20.45 (10.77-38.830	2.55 (2.09-3.11)
	Hispanic	3.12 (1.32-7.44)	2.27 (1.09-4.50)	2.05 (1.10-3.83)	4.93 (2.46-9.89)	2.16 (1.47-3.19)	4.02 (2.72-5.83)		n.d.	n.d.	6.87 (3.11-15.15)
	Asian	n.d.	n.d.	2.22 (1.99-2.470	3.81 (1.91-7.59)	n.d.	3.75 (3.15-4.42)		n.d.	n.d.	4.58 (4.00-5.23)

Table 1 Legend. PRS pipeline accuracy is assessed as the Pearson correlation between scores derived from PCR-free 30X WGS and those derived from imputed genotyping data (GDA) in the same 70 specimens. Pearson correlation shown in the mean correlation across all ancestry groups tested. PRS pipeline precision (repeatability) is the measure of concordance in PRS scores calculated from the same 70 specimens, run through the pipeline 10 times over the course of two weeks. PRS pipeline precision (reproducibility) is assessed using three samples, each run 6 times end-to-end and then compared in a pairwise manner. The z-score standard deviation is used as a measure of variability. PRS site missingness is the percentage of genomics sites in the original score that are missing from the final imputed dataset. Odds Ratios for high PRS vs Not high pRS are derived from the condition-specific cohorts and calculated by each condition lead group across the ancestries available. Odds ratio information for Obesity/BMI is in preparation for publication by the GIANT consortium.

Figure 4 - Summary of first 2500 clinical samples

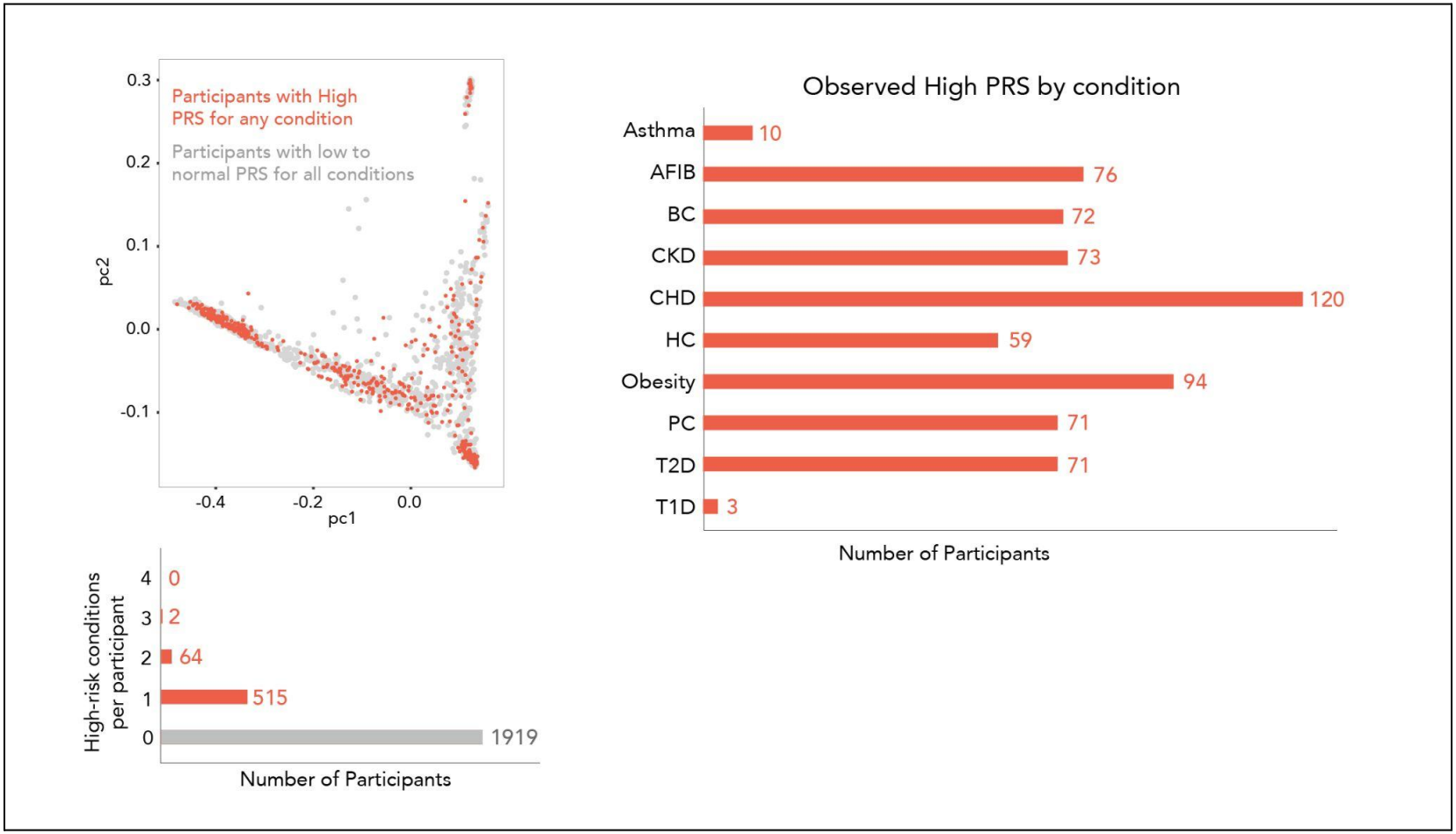

Figure 4. Upper left - Principal component of ancestry indicating participants with a result of 'high PRS' for any condition (red dots) compared to participants who did not have a high PRS identified (gray dots). Lower Left - summary of number of high risk conditions found per participant. Left - Observed numbers of high risk PRS called per condition. Note not all participants get scored for every condition based on age and sex at birth filters. AFIB= Atrial fibrillation; BC = Breast Cancer; CKD = Chronic Kidney Disease; CHD = Coronary Heart Disease; HC = Hypercholesterolemia; PC = Prostate Cancer; T2D = Type 2 Diabetes; T1D = Type 1 Diabetes.

References

1. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat Genet. 2021;53: 420–425.

2. Polygenic Risk Score Task Force of the International Common Disease Alliance. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nat Med. 2021;27: 1876–1884.

3. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19: 581–590.

4. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. 2019;10: 3328.

5. Hurson AN, Pal Choudhury P, Gao C, Hüsing A, Eriksson M, Shi M, et al. Prospective evaluation of a breast-cancer risk model integrating classical risk factors and polygenic risk in 15 cohorts from six countries. Int J Epidemiol. 2022;50: 1897–1911.

6. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol. 2018;72: 1883–1893.

7. Guo F, Edelmann D, Cardoso R, Chen X, Carr PR, Chang-Claude J, et al. Polygenic Risk Score for Defining Personalized Surveillance Intervals After Adenoma Detection and Removal at Colonoscopy. Clin Gastroenterol Hepatol. 2023;21: 210–219.e11.

8. Fantus RJ, Helfand BT. Germline Genetics of Prostate Cancer: Time to Incorporate Genetics into Early Detection Tools. Clin Chem. 2019;65: 74–79.

9. Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med. 2008;358: 2796–2803.

10. Willoughby A, Andreassen PR, Toland AE. Genetic Testing to Guide Risk-Stratified Screens for Breast Cancer. J Pers Med. 2019;9. doi:10.3390/jpm9010015

11. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51: 584–591.

12. Ruan Y, Lin Y-F, Feng Y-CA, Chen C-Y, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. Nat Genet. 2022;54: 573–580.

13. Márquez-Luna C, Loh P-R, South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol. 2017;41: 811–823.

14. Elliott J, Bodinier B, Bond TA, Chadeau-Hyam M, Evangelou E, Moons KGM, et al. Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. JAMA. 2020;323: 636–645.

15. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics. 2011;4: 13.

16. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med. 2013;15:

761–771.

17. Overview. In: NIMHD [Internet]. [cited 20 Apr 2023]. Available: https://www.nimhd.nih.gov/about/overview/

18. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al. Improving reporting standards for polygenic scores in risk prediction studies. Nature. 2021;591: 211–219.

19. Mars N, Lindbohm JV, Della Briotta Parolo P, Widén E, Kaprio J, Palotie A, et al. Systematic comparison of family history and polygenic risk across 24 common diseases. Am J Hum Genet. 2022;109: 2152–2162.

20. Wang Z, Choi SW, Chami N, Boerwinkle E, Fornage M, Redline S, et al. The Value of Rare Genetic Variation in the Prediction of Common Obesity in European Ancestry Populations. Front Endocrinol . 2022;13: 863893.

21. Khan A, Turchin MC, Patki A, Srinivasasainagendra V, Shang N, Nadukuru R, et al. Genome-wide polygenic score to predict chronic kidney disease across ancestries. Nat Med. 2022;28: 1412–1420.

22. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50: 1219–1224.

23. Ge T, Irvin MR, Patki A, Srinivasasainagendra V, Lin Y-F, Tiwari HK, et al. Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. Genome Med. 2022;14: 70.

24. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 2017;100: 635–649.

25. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016;48: 1443–1448.

26. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44: 955–959.

27. Linder JE, Allworth A, Bland ST, Caraballo PJ, Chisholm RL, Clayton EW, et al. Returning integrated genomic risk and clinical recommendations: The eMERGE study. Genet Med. 2023;25: 100006.

28. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med. 2019;21: 1708–1718.

29. Ding Y, Hou K, Burch KS, Lapinska S, Privé F, Vilhjálmsson B, et al. Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk stratification. Nat Genet. 2022;54: 30–39.

Online Methods/Supplemental Information Clinical PRS Paper

A. Analytical and Technical Validation Studies

Broad Imputation Pipeline Overview

An imputation pipeline that takes as an input a variant call format (VCF) file generated from a genotyping microarray and imputes the genotypes at additional sites across the genome was developed. The pipeline architecture and composition was based on the widely used University of Michigan Imputation Server which uses a software called *Eagle* (https://github.com/poruloh/Eagle) for phasing and *Minimac4* (https://genome.sph.umich.edu/wiki/Minimac4) for the imputation. The pipeline uses a curated version of the 1000 Genomes Project (1KG, www.internationalgenome.org) as the reference panel. Additional details on the imputation pipeline can be found at https://broadinstitute.github.io/warp/docs/Pipelines/Imputation_Pipeline/README.

Broad Curated 1KG Reference Panel

During the validation process we determined that some sites in the 1KG reference panel were incorrectly genotyped compared to the sites in matching whole-genome sequencing data. In order to increase accuracy of the imputation and PRS scoring, we curated the original panel by removing sites that were likely incorrectly genotyped based on comparing allele frequencies to those reported in gnomAD v2 (https://gnomad.broadinstitute.org/). Documentation of this curation can be found at: https://broadinstitute.github.io/warp/docs/Pipelines/Imputation_Pipeline/references_overview and a publicly available version of the panel at: gs://broad-gotc-test-storage/imputation/1000G_reference_panel/

Selection of a reference panel for imputation as an input to PRS is an important consideration. Some reference panels (e.g. TOPMed) have more samples than the default used in our pipeline (i.e. 1KG). This leads to more variants being imputed. The question is whether this would materially change the PRS calculated from samples imputed with the TOPMED panel. Access to this panel computationally is restricted (and local download prohibited) so it was deemed

infeasible to implement in our clinical production environment. The performance of a non-eMERGE PRS (for CHD, Khera et al.) using the two different reference panels was determined for 20 GDA saliva specimens and for 42 AoU v1 specimens. The cohort was imputed both by the Broad imputation pipeline with curated 1KG as the reference panel as well as on the TOPMed Imputation Server with TOPMed as the reference panel. Imputed arrays were scored by the PRS pipeline.

The PRS percentiles computed with each method are highly concordant for both cohorts. The Pearson correlation coefficient is 0.996 for both cohorts, the p-value of the Welch two sample t-test is equal to 0.93 and 0.85 (indicating no statistical difference between the methods) for GDA and AoU v1 cohorts, respectively.

Performance verification of the Imputation Pipeline

Imputation accuracy was determined for 42 specimens that were processed through a genotyping microarray (AoU v1 array - the precursor to the commercial Global Diversity Array) and imputed with curated 1KG as the reference panel where corresponding deep-coverage (>30X) PCR-free whole genome sequencing data were used as a truth call set to calculate sensitivity and specificity. The arrays were also imputed on the Michigan Imputation Server with 1KG as the reference panel.

Within the cohort, four different ancestries were represented - Non-Finnish Europeans (NFE), East Asians (EAS), South Asian (SAS), African (AFR), together with the results for three arrays of undetermined ethnicity (NA).

Ethnicity	SNP				INDEL			
	Mean Sensitivity (95% CI)		Mean Specificity (95% CI)		Mean Sensitivity (95% CI)		Mean Specificity (95% CI)	
	Broad	Michigan	Broad	Michigan	Broad	Michigan	Broad	Michigan
NFE (non Finninsh European)	98.5% (98.4 - 98.6)	98.5% (98.4 - 98.6)	99.8% (99.8 - 99.9)	99.8% (99.8 - 99.9)	97.15% (97.0 - 97.3)	97.1% (97.0 - 97.3)	99.4% (99.4 - 99.4)	99.4% (99.4 - 99.4)
EAS (East Asian)	98.0% (97.1 - 98.8)	98.0% (97.1 - 98.8)	99.8% (99.8 - 99.8)	99.8% (99.8 - 99.8)	96.4% (95.5 - 97.3)	96.3% (95.5 - 97.0)	99.2% (99.1 - 99.4)	99.3% (99.0 - 99.5)
SAS (South Asian)	97.9% (97.9 - 98.0)	97.9% (97.9 - 97.9)	99.8% (99.8 - 99.8)	99.8% (99.8 - 99.8)	96.4% (96.3 - 96.5)	96.3% (96.2 - 96.3)	99.2% (99.2 - 99.2)	99.3% (99.2 - 99.3)
AFR (African)	97.3% (97.3 - 97.4)	96.8% (96.7 - 96.8)	99.6% (99.5 - 99.7)	99.5% (99.4 - 99.5)	95.7% (95.7 - 95.8)	95.3% (95.2 - 95.3)	98.7% (98.6 - 98.9)	98.7% (98.6 - 98.9)

Table 1. Sensitivity and specificity of SNP and INDEL imputation from the Broad Imputation Pipeline (Broad) and Michigan Imputation Pipeline (Michigan) with the curated 1000Genomes reference panel when compared to matched whole genome data for 42 AoU v1 samples.

Broad imputation pipeline sensitivity for SNPs is >97% and INDELs >95% for all ethnicities. Similarly, specificity for SNPs from the Broad imputation pipeline are above 99% and the specificity for INDELS is >98%. Results were highly concordant with those returned by the remote server at Michigan.

Performance evaluation of different input material types.

To assess the performance of specimens derived from both saliva and whole blood a set of 20 matched blood saliva pairs were run through the GDA genotyping process and the resulting VCFs were imputed using the Broad pipeline to be compared against results for matched blood derived whole genome data. The Pearson correlation between sensitivity and specificity of blood and saliva derived samples are equal to 100% and 100%, respectively. For the same pairs, the Welch two-sample t-test statistic is 0.997 and 0.987, respectively. There is no significant difference between the different input sample types.

Imputation repeatability and reproducibility

Imputation pipeline repeatability was assessed by repeating imputation of a cohort of 1000 GSA arrays ten times over the course of two weeks and was found to be 100% concordant. Imputation pipeline precision (reproducibility) was also tested on technical replicates. Three individual samples derived from saliva were each genotyped six times, followed by an imputation in a cohort of all saliva derived samples. In each set of technical replicates all pairs and variants in each pair were compared (making a total of 45 pairs for which genotypes were compared). Reproducibility is measured using Jaccard scores. "Reproducibility over variants" was calculated only over sites where at least one of the two replicates in a pair calls a non hom-ref genotype and was found to be 99.91% (95CI: 99.89-99.93) for SNPs and 99.87%

(95CI: 99.85-99.90) for InDels. "Reproducibility over all sites" was calculated over all genotyped sites, including sites genotyped as hom-ref in both replicates and was found to be 100% (95CI: 100-100) for both SNPs and InDels.

Imputation performance as a function of variant frequency

Because we expect accuracy to be impacted by the frequency of a variant in the population (rare variants are less likely to be in the reference panel and therefore less accurately imputed) we further subdivided the performance assessment by allele frequencies on two cohorts: 42 AoU v1 arrays and 20 blood-saliva pairs of GDA arrays. Accuracy of imputation of variants as a function of population allele frequency performed as expected with rare variants in the population not being as accurately represented. Imputation is more accurate for variants that are more frequently observed in the population (≥0.1 allele frequency). This is predicted to have a low impact on the accuracy of PRS calculations from imputed variants as PRS scores are also typically derived from common variants.

Impact of genotyping array call rate on imputation performance.

The impact of call rate on the imputation was assessed by generating a downsampled series of 42 arrays, each with call rates of 90%, 95%, 97% and 98%. Pearson correlation values for SNPs and INDELs were calculated across bins of allele frequencies, assessed against gnomAD common variants (AF >0.1), for the cohorts with downsampled call rates. Call rates below 95% were found to produce suboptimal results. At this rate the mean R^2 dosage score for sites with AF ≥0.1 was found to be 0.98% (95CI: 0.98-0.98) for both SNPs and InDels compared to 0.99% for call rates of 97% and 98%.

Impact of imputation batch size on performance.

Batch size effect of the imputation pipeline was assessed by imputing and analyzing arrays in a cohort of size 1000 (randomly chosen), ten cohorts of size 100 (non-overlapping subsets of the 1000 cohort), and ten cohorts of size 10 (non-overlapping subsets of one of the 100 cohorts). Pearson correlations of dosage scores were calculated across bins for allele frequencies (assessed against gnomAD) for smaller cohorts versus larger cohorts. The data show that imputation is highly correlated across batch sizes with batches down to as few as 10 samples producing acceptable performance. The mean R^2 correlation of dosage scores for sites with allele frequency greater or equal to 0.1 is above 0.97 in all cases both for SNPs and INDELs and increases to 0.98 for the larger studied cohorts. Increasing batch sizes produces very slight improvements in imputation but these are not significant and the choice of imputation batch size (above or equal to 10 samples) can be made on practical and operational grounds

Broad PRS Pipeline Overview

The polygenic risk score (PRS) pipeline begins by calculating a raw score using plink2 (https://www.cog-genomics.org/plink/2.0/). For each condition, effect alleles and weights are defined for a set of genomic sites stored in a weights file. At each site, the effect allele dosage observed in the imputed vcf is multiplied by the effect weight in the weights file. The raw score is the sum of these products over all the specified sites.

Validation of technical and analytical performance of the PRS pipeline.
For each of the 10 conditions chosen by the consortium for clinical return, a validation study was performed to assess the technical and analytical performance as well as to verify the association between score and disease risk.

PRS Pipeline Accuracy. Accuracy of the pipeline was determined by calculating the Pearson correlation between PRS scores calculated from 70 specimens imputed from GDA array data and PRS scores of corresponding deep-coverage PCR-free whole genome sequencing data (used as a truth call set).

Input Material Performance. Accuracy of PRS scoring when different sample types (blood or saliva) are used as inputs was determined by comparing the PRS scores from matched blood and saliva pairs collected from 20 individuals.

PRS pipeline repeatability. PRS pipeline repeatability was assessed by running the pipeline on the same dataset of 70 imputed GDA arrays ten times over the course of two weeks (without call caching). Scores generated from the different processing runs were compared to determine if there are any differences observed for a given PRS when the pipeline is run at different times.

PRS pipeline reproducibility. PRS pipeline precision (reproducibility) was assessed using three samples each run 6 times end-to-end and then compared in a pairwise manner. The z-score standard deviation is used as a measure of variability.

PRS site representation. The SNP weight sites that are not called during genotyping or imputation were determined. These are sites not present in the intersection of an imputed GDA array and the reference panel. Ideally, all sites required for PRS calculation are present either as genotyped or imputed sites; however, in practice a small number of sites are not present due to differences in the data used to create the score and the specific array and imputation reference panel used in this study.

	PRS Accuracy: Pearson Correlation between PRS from array and WGS*	PRS Pipeline Performance concordance between blood and saliva input types.	PRS Precision: PRS pipeline repeatability	PRS Precision: PRS pipeline reproducibility	PRS Limit of Detection: Score site missingness
Asthma	99.3%	100.0%	100%	0.002	0.69%
Atrial Fibrillation	98.6%	100.0%	100%	0.001	1.20%
Breast Cancer	93.0%	100.0%	100%	0.004	0.32%
Chronic Kidney Disease	98.3%	100.0%	100%	0.0001	0.69%
Coronoary Heart Disease	98.2%	100.0%	100%	0.001	0.46%
Hypercholesterolemia	95.9%	100.0%	100%	0.005	1.20%
Obesity/BMI	99.5%	100.0%	100%	0.002	0.70%
Prostate Cancer	96.4%	100.0%	100%	0.0006	2.97%
Type 1 Diabetes	99.5%	100.0%	100%	0.0001	2.97%
Type 2 Diabetes	98.8%	100.0%	100%	0.001	0.70%

PRS pipeline accuracy is assessed as the pearson correlation between scores derived from PCR-free WGS and those derived from imputed genotyping data (GDA) in 70 specimens. Pearson correlation shown is the mean correlation across all ancestry groups tested. PRS pipeline performnance as a function of sample input was assessed by comparing the scores from 20 matched blood and saliva pairs. PRS pipeline precision (repeatability) is the measure of concordance in PRS scores calculated using the same data from 70 specimens, run through the pipeline 10 times over the course of two weeks. PRS pipeline precision (reproducibility) was assessed using three samples each run 6 times end-to-end and then compared in a pairwise manner. The z-score standard deviation is used as a measure of variability. PRS site representation is a measure of the percentage of the original score sites are missing from the final imputed dataset.

Table 2. Validation measures summary.

Performance verification using eMERGE I-III cohort. A cohort of samples with known phenotypic information was used to verify the relationship between polygenic risk score as determined by our pipeline and disease risk. For conditions where cases and controls could be identified in the eMERGE I-III cohort we determined performance using metrics outlined in the ClinGen working group recommendations (Wand et al.). Specifically, we determined the PRS distributions for cases and controls, we examined the impact of ancestry adjustment on the distributions (Fig), and we examined the relationship between observed and predicted risk. There are some limitations to this analysis: i) The eMERGE I-III dataset being used for this analysis was generated from different array platforms and was imputed with a different pipeline including a different version of 1KG reference panel than the one currently implemented; ii) The eMERGE I-II-III imputed dataset does not include variants from Chromosomes X or Y. For these reasons, the PRS disease association analysis represents a verification of the clinical validation performed by eMERGE IV condition leads rather than the quantitative measure of the impact of the score on risk. The clinical associations (odd ratios) that are reported on the clinical report for each condition were independently determined by eMERGE IV disease-specific expert teams.

Validation of pipeline and ancestry adjustment in original case control cohorts. The final pipeline was made available to computational scientists at each of the eMERGE IV disease-specific expert teams who had access to appropriate case control cohorts. These groups confirmed the performance of the final pipeline on their cohorts. The odds ratios for each condition that are reported on the clinical reports come from these cohorts rather than the eMERGE cohort for the reasons described above.

B. PRS Ancestry Calibration Overview

PCA method description

For a polygenic risk score which is a sum of SNP effects (linear weights), the central limit theorem states that the distribution of scores in a homogenous population will tend towards a normal distribution as the number of SNPs becomes large. When two different homogenous populations are randomly mixed, the additive property of prs leads the resulting distribution to be similarly normally distributed, with mean and variance depending on the mean and variance of the original homogenous populations. We can therefore model the distribution of prs scores as being normally distributed, with mean and variance being functions of genetic ancestry. Practically, we implement this as

$$PRS_{raw} = N(\mu, \sigma^2)$$

$$\mu = \alpha_0 + \sum \alpha_i PC_i$$

$$\sigma^2 = exp(\beta_0 + \sum \beta_i PC_i),$$

with genetic ancestry being represented by projection into principal component space. The $\boldsymbol{\alpha}$ and **β** parameters are found by jointly fitting them to a cohort of training data. This fit is performed by minimizing the negative log likelihood:

$$-\log L = \sum_i \log \sigma_i + 1/2\left(\frac{prs_i - \mu_i}{\sigma_i}\right)^2$$

where ι runs over the individuals in the training cohort, prs_i is the i'th individuals raw prs score, and μ_i and σ_i are calculated using Eq X by projecting the i'th individual into PC space. Note that, due to the simplicity of the model, overfitting is unlikely to be a problem, and so no regularization or other overfitting avoidance technique is implemented. An individual's PRS z-score can then be calculated as

$$z - score = \left(\frac{prs - \mu}{\sigma}\right),$$

where μ and σ have again been calculated based on the specific individuals projection into PC space. In this way, once the model has been trained, the z-score calculation is fully defined by the fitted model parameters, and z-scores can be calculated without needing additional access to the original training cohort.

Generating trained models from All of Us data

Generating the trained models consisted of three steps: 1. Selecting the training cohort. 2. Imputation of the training cohort. 3. Training the models on the training cohort. A test cohort was also generated in order to test the performance of the training.

Ancestry balanced training and test cohorts were generated by subsampling from an initial cohort of around 100,000 All of Us samples. For the purposes of balancing the cohort,

each sample was assigned to one of the five 1KG Super Populations. Principal component analysis was first performed on a random subset of 20,000 samples. 1KG samples were projected onto these principal components, and a support vector machine (SVM) was trained on 1KG to predict ancestry. The SVM was then used to assign 108,000 AoU samples to one of the five 1KG Super Populations. A balanced training cohort was selected based on these predicted ancestries, and principal components were recalculated using this balanced training cohort. A similarly balanced test cohort was selected based on ancestries estimated from projection on the training set PCs. The resulting breakdown of the cohorts by estimated ancestry is shown in Table 3.

	Training Cohort	**Test Cohort**
AMR	1817	1500
AFR	1664	1500
EAS	1137	1436
EUR	1823	1500
SAS	444	654
TOTAL	**6885**	**6590**

Table 3

Both the training and testing cohorts include a number of individuals with highly admixed ancestry. Admixture was quantified using the tool Admixture (Alexander *et al. PMID: 19648217)* with 5 ancestral populations. The resulting admixtures of each cohort are shown in Figure 1, and the most common admixed ancestries in each cohort are summarized in Table 4.

Each cohort was imputed using the imputation pipeline described above, with 1KG as the reference panel. By keeping the imputation pipeline identical to the pipeline used for the eMERGE dataset, and because the AoU dataset uses the same GDA array as the eMERGE dataset, any potential biases resulting from differing data production and processing methods were removed. The training cohort was scored for each of the ten conditions, and model parameters were fit by minimizing the negative log likelihood as described. The test cohort was then used to evaluate the generalizability of these model parameters.

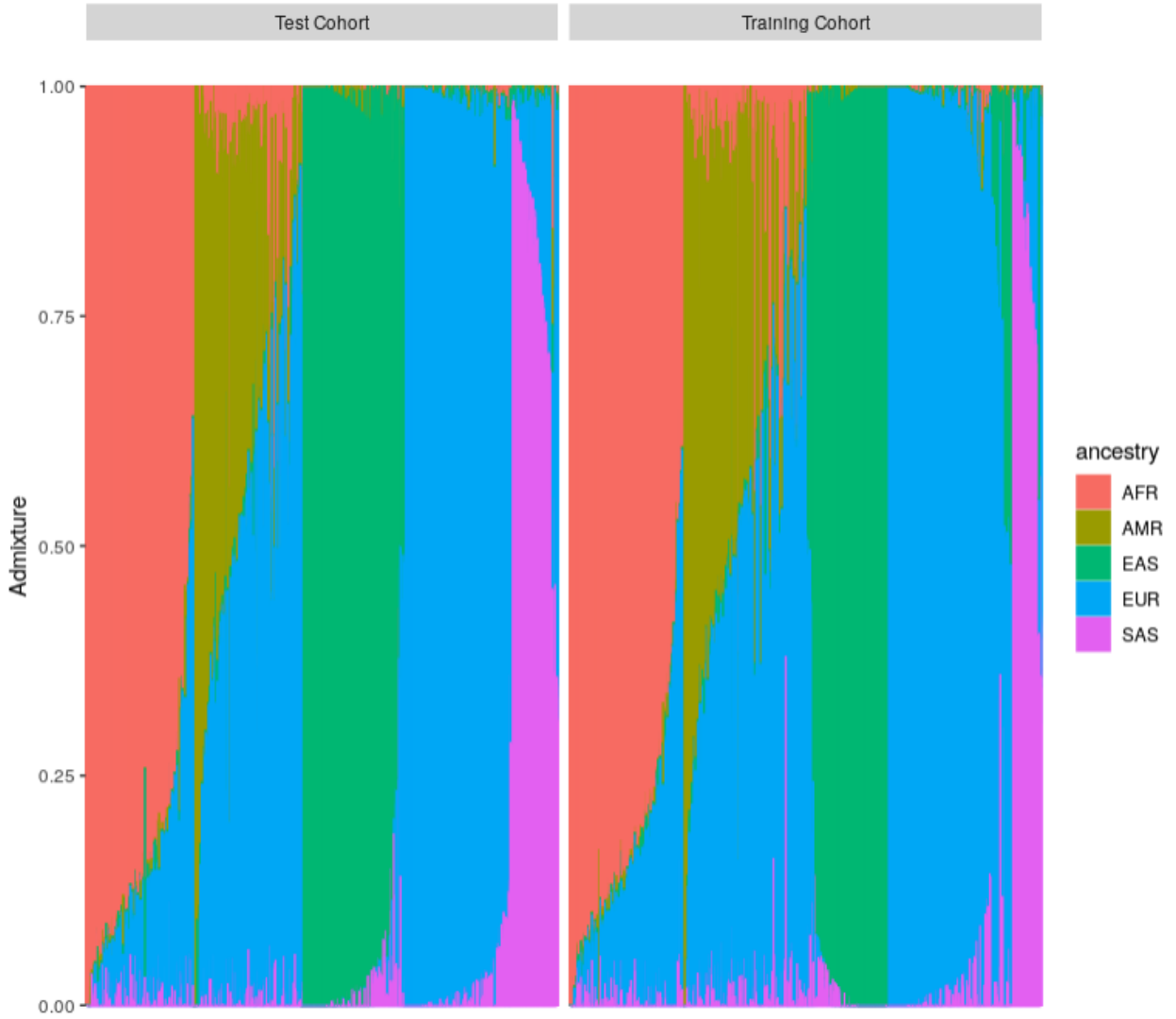

Figure 1

Admixed Ancestry	Training Cohort	Test Cohort
AFR-EUR	590	556
AMR-EUR	1238	883
EAS-EUR	236	102
EUR-SAS	191	229

Table 4. Admixed ancestries are defined as ancestries for which an individual's admixture fraction is greater than 20%. For example, an individual who is indicated by admixture to be

45% AFR, 37% EUR, 12 % AMR, 5% EAS, 1% SAS would be included in the AFR-EUR row of this table.

Performance on Test Cohort

Figure 2 the distribution of calibrated z-scores in the test cohort using the parameters fit in the training cohort. As can be seen, all ancestries show the intended standard normal distribution of calibrated scores.

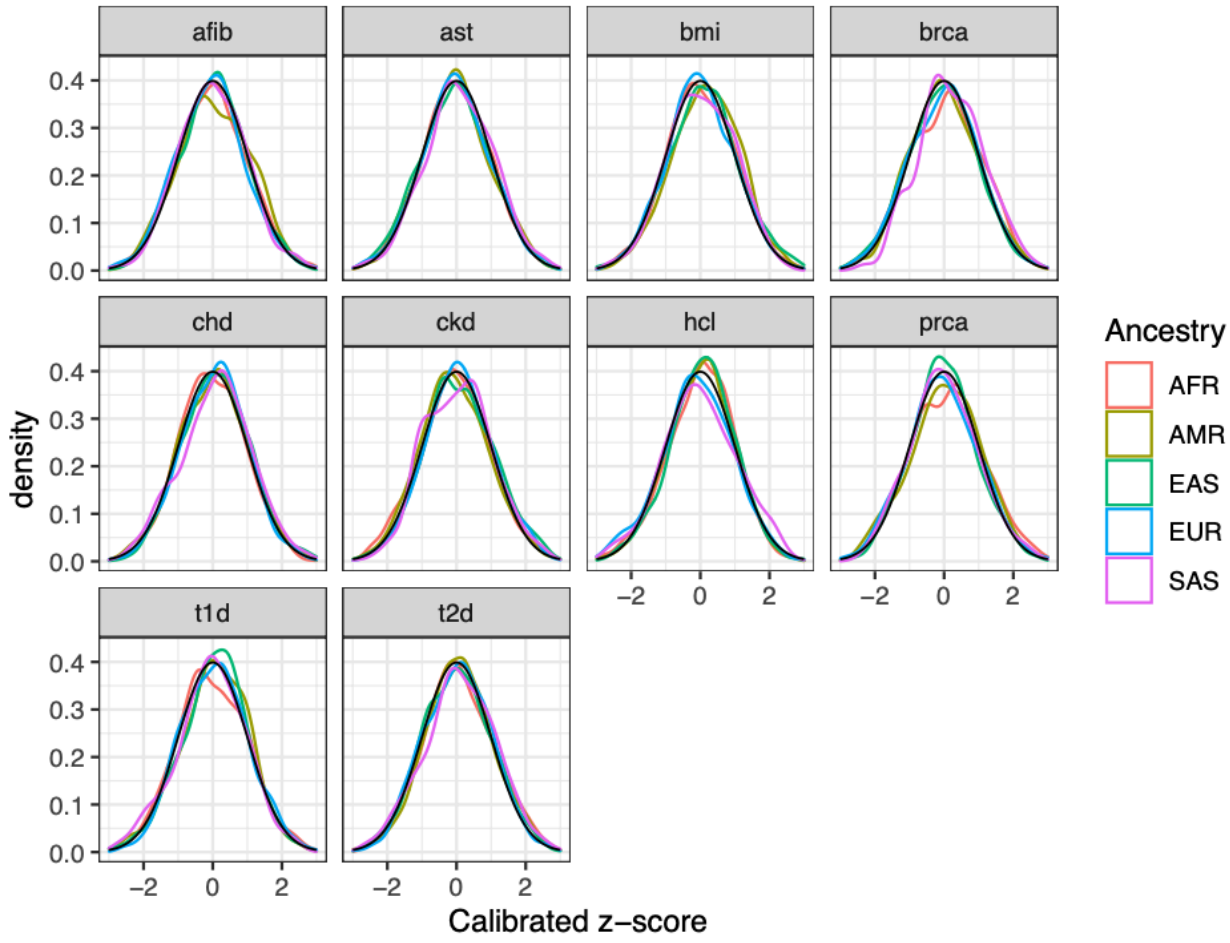

Figure 2

One of the main improvements of this method over previous methods is the inclusion of an ancestry dependent variance in addition to the ancestry dependent mean. The importance of this is shown for the Hypercholesterolemia PRS in Figure X below. As can be seen, the variance of this score differs significantly across ancestries, so that a method which only fits the mean of the distribution as ancestry dependent can result in z-score distributions which have been attenuated towards zero or expanded away from zero for some ancesties. By also treating

variance as ancestry dependent, this method results in z-score distributions which are more standardized across ancestries.

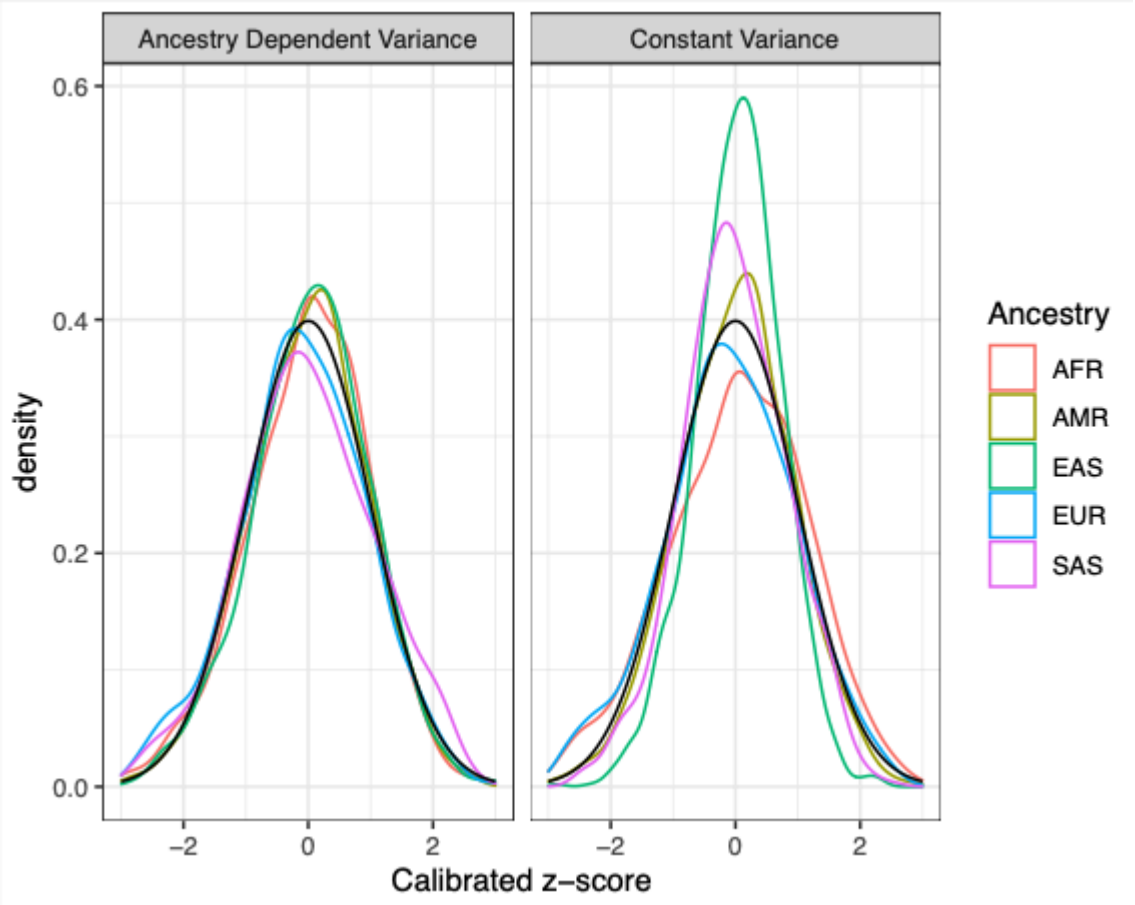

Figure 3. Hypercholesterolemia PRS calibrated z-scores of training cohort. Note the improvement when an ancestry dependent variance is used over a constant variance method.

In addition to improving calibration across ancestries, this method can improve calibration within ancestries, particularly for highly admixed individuals. An example of this can be seen in Figure 4. As can be seen, because no ancestry group is homogenous, when individuals are compared directly to other individuals in their assigned population group, a dependence between admixture fraction and PRS score can result. This dependence is removed by the described PCA calibration method, and the resulting calibrated PRS scores are independent of admixture fraction.

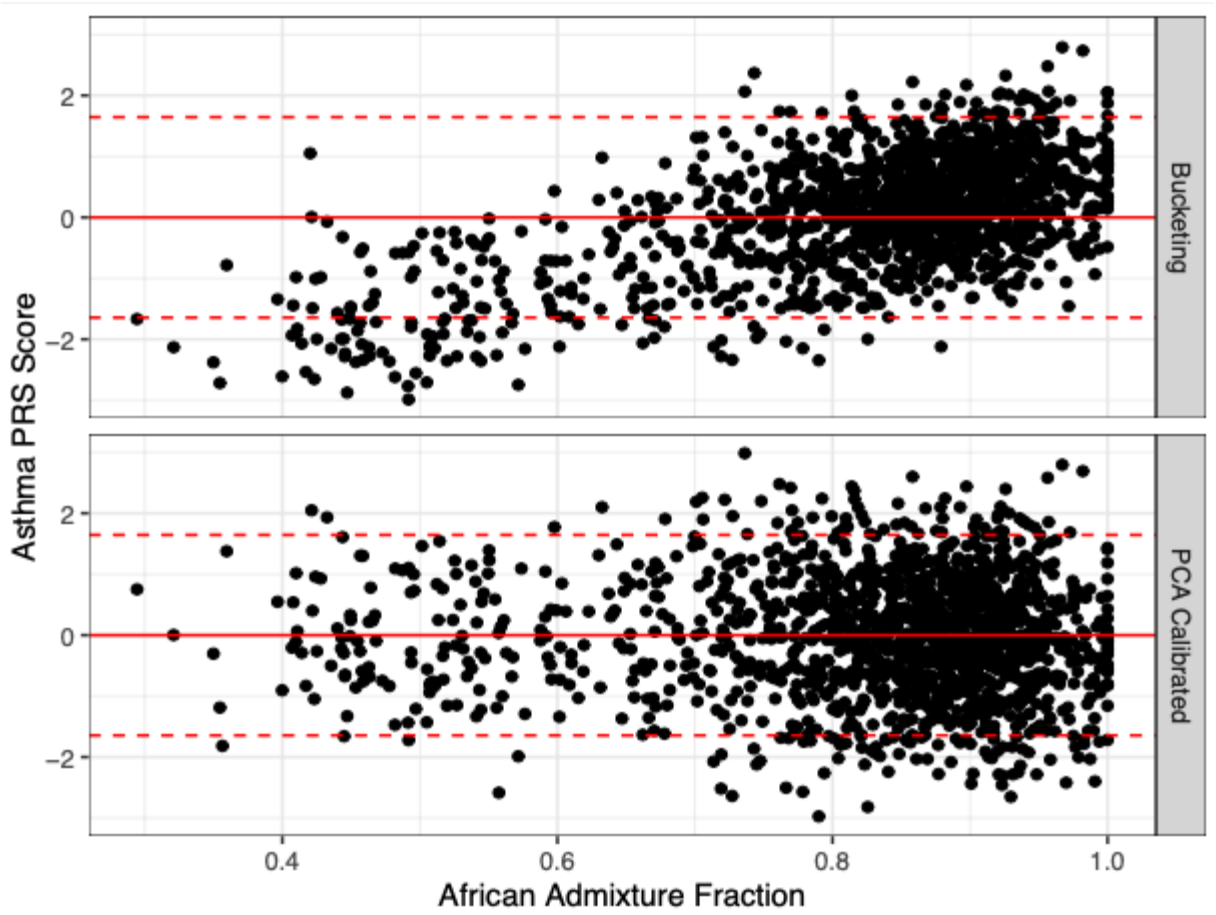

Figure 4. PRS z-score as a function of African Admixture Fraction, for individuals of African ancestry. In the “Bucketing” method, a z-score is calculated by comparing to the mean and variance of all individuals of African ancestry in the cohort. The “PCA Calibrated” method is the method described above. Note the dependence on admixture fraction in the “Bucketing” method, which has been removed in the “PCA Calibrated” method.

Supplemental Table 1

		Abdominal Aortic Aneurysm	Age-Related Macular Degeneration	Asthma	Atopic Dermatitis	Atrial Fibrillation	Bone Mineral Density	Breast Cancer	Crohn's Disease	Chronic Kidney Disease	Colorectal cancer	Coronary Heart Disease	Depression	Hypercholesterolemia	Hypertension	Ischemic Stroke	Lupus	Nonalcoholic Fatty Liver Disease	Obesity	Primary open angle glaucoma	Prostate Cancer	Rheumatoid Arthritis	Type I Diabetes	Type II Diabetes
Analytical Viability - existing PRS	What does PRS predict?	Case status	Disease severity and progression	Case status	Not developed	Case status	Case status	Case status	Case status; drug response	Not developed	Case status	Case status	Case status	Case status	Not developed	Case status	Not developed	Not developed	Case status	Case status	Case status	Case status	Not developed	Case status
	Validated PRS available?	Yes	Yes	No	Not developed	Yes	No	Yes	Yes	Not developed	Yes	Yes	No	Yes	Not developed	Yes	Not developed	No	Yes	Yes	Yes	Yes	Not developed	Yes
	Validation populations	EU, AA	EU	EU	Not developed	EU	NP	EU, AA, HL, Asn	EU	Not developed	EU, AA, HL, Asn	EU, AA, HL, Asn	NP	NP	Not developed	EU, Asn, HL	Not developed	NA	EU, AA, HL, Asn	EU, AA, HL, Asn	EA	EU	Not developed	EU, AA, HL, Asn
	Number of SNPs	29 - 3699	52	15	Not developed	30-6 million	NP	34-290,000	10,799-909,763	Not developed	1.2 million	12 - 6.6 million	NP	223	Not developed	NP	Not developed	NA	NP	1250	25-110	NP	Not developed	>100,000
	Age range	adults	adults	9-28 years	Not developed	NP	NP	adults	NP	Not developed	NP	adults	NP	adults	Not developed	adults	Not developed	all ages	pediatrics	adults	adults	adults	Not developed	adults
Analytical Viability - future PRS development	AUC or R^2, not asked for in 2 pager	Not published (NP)	0.72-0.82	NP	Not developed	NP	NP	0.51-0.69	0.69	Not developed	0.654	0.81	NP	NP	Not developed	NP	Not developed	NA	0.55-0.69	0.77	NP	NP	Not developed	0.66
	GWAS availability for PRS development and optimization	Yes	PRS already developed	Numerous	11 GWAS; 1 with African ancestry	Numerous	Yes	Yes	Yes	Yes	Numerous	Numerous	Numerous	Numerous	Numerous	Numerous	Numerous	Numerous	Numerous	numerous	PRS already developed	Numerous	Numerous	NP
Feasibility	Phenotype definition	Yes	Yes	NP	NP	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Family-based heritability	NP	NP	NP	0.75	NP	0.5-0.8	NP	NP	0.32-0.43	NP	0.4-0.6	NP	0.22-0.61	NP	NP	NP	0.4-1.0	NP	NP	34% with 28.4% from common variants, 6% from rare	NP	NP	NP
	SNP-based heritability	NP	0.3-0.58	0.82	NP	As high as 0.22	NP	0.18-0.41	0.48	> 0.07 .	0.07	0.22	0.09 - 0.11	NP	NP	NP	NP	NP	NP	0.03	NP	NP	NP	0.15
	Datasets for independent validation	eMERGE III	Three GWAS - AS and EU	eMERGE III + institutional dataset	eMERGE III + institutional dataset	eMERGE III + additional GWAS	BioVU	Numerous	Numerous	UK Biobank, eMERGE III, institutional datasets	Numerous	Numerous	Numerous	Numerous	Numerous	Numerous, with limited data in AA	Numerous	NP	Numerous	Numerous	eMERGE III	Numerous in EU, with limited data sets for other ancestries	Numerous	Numerous
	Age of disease onset	> 65	> 50	< 18	< 18	NP	> 60	> 18	NP	NP	NP	> 20	NP	< 18	NP	NP	>20	0-99	0-99	>50 years	> 55	NP	< 18	varies; > 50
	Population based prevalence	1.3% in males 45-54 y to 12.5% in males 75-84 y. For females, the prevalence ranges from 0% in the youngest to 5.2% in the oldest age groups	2.6%-15.0% based on economic status of country	5-12% in children; disproportionately affects AA	10-20% of children. AD especially is very common in those with asthma in which up to 50% may be affected	In those over 65 yrs, 15% prevalence in Blacks	In the US, over 10 million people are afflicted and another 43.4 million are considered at risk with low bone mass	Breast cancer is the most common cancer and the second leading cause of cancer-related death among women in the U.S.	CD has a prevalence of 100 to 300 per 100,000 people in Western Europe and North America	In the United States, an estimated 13.6% of adults have CKD	Lifetime risk in general USA population is ~5%.	Based on NHANES 2013-2016 data, 18.2 million Americans ≥20 y have CHD (prevalence 6.7%).	10.4% past-year and 20.6% lifetime prevalence estimated in a nationally representative survey of US adults	Approximately 7% of US adults have severe hypercholesterolemia (untreated LDL-C >= 190 mg/dL), the majority of whom do not harbor a monogenic FH-associated variant	Decades of scientific evidence implicate elevated blood pressure (BP) in the etiology of cardiovascular disease, including coronary artery disease, peripheral arterial disease, and stroke, as well as renal and ocular damage. Elevated BP accounts for at least 13% of annual deaths worldwide.	Stroke affects an estimated 7 million people (2.7% EA, 4.1% AA) each year	NP	Global prevalence is approximately 25%	NP	5-13% at age 80	2nd most common cancer in men – 96.7 diagnoses per 100k EA men, 163.8 for Black/AA men	1.3-1.5 million people affected	NP	NP
Actionability	Timing of intervention	NP	NP	Pediatrics	Pediatrics	Adults	At ages	Adults	NP	NP	40 years old	Pediatrics	NP	Pediatrics	Adults	NP	NP	All participants	Pediatrics	NP	Adults	NP	Pediatrics	Adults
	Actionability/Intervention	Abdominal ultrasonography (USG)	Supplements to slow progression (vitamin C, vitamin E, beta carotene, zinc); VEGF inhibitors for late disease.	Parental education and guidelines such as avoiding newborn babies from asthma triggers, and making sure child maintain a healthy weight are part of our goals.	Emollients in the neonatal period appear to reduce the incidence of atopic dermatitis.	Anticoagulation to reduce stroke risk, medication to control heart rhythm & rate, ablation	Calcium / vitamin D supplementation, regular exercise	Enhanced screening (breast MRI alternating with mammogram); risk reducing mastectomy, reproductive, breast feeding decisions, avoidance of HRT; lifestyle factors	Colonoscopy	Serum Cr and urine microalbumin checks, dietary intervention (low salt), smoking cessation, BP control	Colonoscopy; Removal of polyps during screening can reduce risk for CRC.	Screening tests such as exercise stress testing and coronary calcium scan. Blood to assess need for statin therapy.	Cognitive-behavioral treatments, psychopharmacological options, mood monitoring, lifestyle changes	Measurement of lipid levels; initiation or optimization of lipid-lowering therapy as appropriate	Earlier or more intense anti-hypertensive treatment, lifestyle interventions, home and clinical BP monitoring	Aspirin recommended in (1) patients with 10-year ASCVD risk >10% (2) women, including those with diabetes, in which benefit outweighs risk (3) considered in patients with CKD (NOT stage 4 or 5); cilostazol recommended in patients with peripheral artery disease.24	Treating early with milder therapies (e.g., hydroxychloroquine) that have been established to prevent the progression in disease severity in untreated SLE patients. Avoiding sun-bathing and attending to unexplained rashes, will help prevent disease and help diagnose the onset of SLE at its earliest stages.	to be determined	Alerting individuals to the risk may provide them with sufficient time to modify their behavior / lifestyle to avoid excessive weight gain.	If detected by pressure measurement (tonometry), increased IOP can be treated with eye drops, laser surgery, or microsurgery, preventing visual loss	GRS guides entry into existing, common guidelines for PrCa surveillance via PSA testing. Identifying those at low genetic risk could improve PSA over-testing.	Dependent upon age and other clinical findings	Measure autoantibodies, A1C, fasting blood sugar	Lifestyle change/weight loss
Translatability and Development potential	Other known predictors of risk	Age (especially ≥65), male sex, cigarette smoking, atherosclerosis, hypertension, and hyperlipidemia	Family history, age, smoking, hypertension, night vision	family history, maternal smoking history, living location, coexisting chronic respiratory infection, presence of allergy, atopy and obesity	Asthma, family history, gestational diabetes	BMI, hypertension, tobacco use, diabetes mellitus, history of myocardial infarction and heart failure	Age, sex, BMI, metabolic health, smoking, alcohol use, race/ethnicity	BMI, hormone replacement therapy (HRT), alcohol consumption, physical activity, diet, breast density, atypical hyperplasia, breast inflammatory disease, and parity	Environmental factors	Age, sex, diabetes, hypertension, smoking, family history of kidney disease	BMI, sex, diet, smoking, age, family history, alcohol, diabetes, hormone replacement therapy, exercise, education, NSAID use	Age, male sex, hyperlipidemia, obesity, hypertension, diabetes, cigarette smoking, and family history of CHD.	Stressful life events; race; socioeconomic factors; sex/gender; family history of psychiatric illness	NP	Race, age, sex, and BMI are well-known factors that are addressed in our default models. Additionally, smoking, alcohol, diet, stress, diabetes, kidney disease, sleep apnea, and family history are known to be associated with increased BP.	Non-modifiable: age >55, gender, race, low birth weight (≤5 lbs) Modifiable: hypertension, cardiovascular disease, diabetes, dyslipidemia, lifestyle (alcohol, smoking, sedentary), medications (hormones, oral contraceptives) Other potentially modifiable: sleep apnea, sickle cell disease, drug & alcohol abuse, hyperhomocysteinemia, inflammatory and infectious disease, migraine.	Sex, ancestry	ALT levels, fatty liver index uses BMI, waist circumference, serum triglycerides and GGT	NP	Cardiovascular disease, diabetes, hypertension, smoking, alcohol, sex, age, African ethnicity, family history, myopia	NP	NP	NP	Ancestry
	Public Health and medical impact	Ruptured AAAs have a mortality estimated at 81%.	Leading cause of irreversible blindness in adults over 50 years old with 11 million individuals affected in the United States	Asthma is the most common chronic disease among children, which affects more than 330 million people worldwide. Asthma is the third leading cause of hospitalization in United States.	Atopic dermatitis (AD) is a chronically relapsing skin disorder with an immunologic basis that occurs in approximately 10-20% of children.	Most common cardiac arrhythmia, affects over 3 million Americans and 33 million people worldwide. AF causes increased risk of stroke, the leading cause of serious, long-term disability in the US; heart failure, dementia and death	Osteoporosis is the most prevalent metabolic bone disease in the United States with over 10 million people afflicted and another 43.4 million considered at risk with low bone mass.	Breast cancer is the most common cancer and the second leading cause of cancer-related death among women in the U.S.	NP	CKD is associated with a high burden of comorbidities and increased mortality. CKD is frequently under-recognized and under-treated; patient awareness is less than 10% for those with stages 1-3 CKD, and 52% in patients with severe CKD (stage 4+). In addition to early CKD detection, genetic risk may play an important role in selection of living kidney donors.	CRC is the 3rd most deadly cancer in the world.	The annual incidence of myocardial infarction in the United States is 580,000.	Depression is the leading cause of disability in the US and globally, and associated with poor outcomes including suicide and a broad range of downstream medical comorbidities	ASCVD remains the leading cause of morbidity and mortality in the U.S., and improvements in cardiovascular outcomes have not been shared equally across populations	A substantial proportion of deaths and morbidity worldwide are due to effects of hypertension.	Stroke accounts for an estimated $34 billion in costs each year.	Mortality is high with accounting for one of the highest death rates of any disease in the USA in the 25 to 35-year old female population	Given its high correlation with obesity, NAFLD is skyrocketing in prevalence worldwide, and a recent estimate of the global prevalence is approximately 25%	NP	Glaucoma is the second highest cause of blindness, after cataract;	Prostate Cancer is the 2nd most common cancer in the US. PSA surveillance guided by genetic risk could have a large impact on potentially harmful PSA over-testing in the broader population	RA accounts for accounting for an estimated > $30 billion in health care costs.	NP	T2D is one of this centuries greatest chronic, non-communicable health scourges.
	Health disparities	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	NP	NP	Yes
Reason for phenotype was moved to Development pathway before June 2021		PRS only validated in EU and AA. Inability to pull major risk factor (smoking) from EHR. Low disease prevalence in AS and HA populations.	Less common disease prevalence in EU ancestry. Genetic tests are being used to assess progression not case/control status.	Clinical Pathway	Limited multiancestry GWAS and lack of immediate access to validation datasets.	Clinical Pathway	No validated phenotype EHR algorithum; phenotype requires DEXA scan. No genetic association found to date in HA and AA populations.	Clinical Pathway	Lack of multiancestral validation and reliance of availble PRS on ImmunoChIP.	Clinical Pathway	The PRS development and validation was not complete for all ancestral groups in June 2021.	Clinical Pathway	The PRS development and validation was not on track for clinical implementaton by June 2021.	Clinical Pathway	PRS had a relatively low predictive value and the value of return was questioned relative to other phenotypes.	The PRS development and validation was not on track for clinical implementaton by June 2021.	Low population prevalence; ethical considerations of returning results for rare disease with substantial false positives. Lack of timely access to validation data sets. Lack of clinically validated preventative strategies.	Performance of PRS in March 2021 was not predictive across ancestral groupus.	Clinical Pathway	Deprioritized based on avalible expertise at site that proposed the phenotype.	Clinical Pathway	Lack of data in subjects of AA ancestry to support Development and Validation of PRS.	Clinical Pathway	Clinical Pathway
Moved to Developmental Pathway		**June 2021**	**Aug 2020**		**Aug 2020**		**Aug 2020**		**Aug 2020**		**June 2021**		**March 2021**		**March 2021**	**March 2021**	**March 2021**	**March 2021**		**Aug 2020**		**Aug 2020**		

Supplemental Table 2

Datasource: Data generated by eMERGE sites, stadardized metrics defined and reported on PRS Grid and PRS grid instructions

		Performance metrics of PRS validation model											Feasibility in pipeline	
Condition	**# case:control in validation cohort(s)**	**Is the score cross-ancestry, if so, for which populations? (European (EA), African (AA), Hispanic (HL), Asian (Asn))**	**Was validation analysis restricted to specific age range? If yes, list**	**Odds ratio per SD**	**Estimate of model discrimination (AUC) with CI of genomic predictor only**	**Estimate of model discrimination (AUC) with CI for non-genetic covariates only**	**Estimate of model discrimination (AUC) with CI of the full model (i.e. with genomic predictor and non-genetic covariates)**	**Does the model include covariates beyond Age & Sex? If yes, list**	**Proposed high risk cut off**	**Odds ratio & CI of proposed high risk cut off compared to rest**	**P value of OR**	**Sensitivity and specificity at cut off (SN:SP)**	**Positive predictive value and negative predictive value at cut off (PPV:NPV)**	**Are all sites genotyped or imputed on Global Diversity array? If not, #/total**	**Does the model use a standard sites and weights only format? If no, what are the additional components/variables that will need to be evaluated for implementation feasibility?**
Abdominal Aortic Aneurysm	European: 4,165:42,843 African: 42:4,492 Hispanic: N/A Asian: N/A	No	18+	Overall: 1.63 (1.51, 1.74) European: 1.49 (1.38, 1.62) African: 1.62 (1.25, 1.90) Hispanic: N/A Asian: N/A	European: 0.60 African: 0.59 Hispanic: N/A Asian: N/A	European: 0.74 African: 0.80 Hispanic: N/A Asian: N/A	European: 0.75 African: 0.82 Hispanic: N/A Asian: N/A	Age, BMI, 10 PCs	5%	European: 2.21 (2.04-2.37) African: 3.34 (2.53-4.14) Hispanic: N/A Asian: N/A	European: p<2.00E-16 African:p=0.0033 Hispanic: N/A Asian: N/A	European: 0.08:0.94 African: 0.17:0.94 Youden cutoff: European: 0.78:0.62 African: 0.85:0.68	European: 0.057:0.96 African: 0.028:0.99 Youden cutoff: European: 0.079:0.98 African: 0.026:0.822	Yes	PRS is only sites and weights; the genome informed risk assesment (GIRA) includes age, sex, bmi and smoking status
Asthma	European: 3,835/43,531; ped 1,048/7,416 African: 1,373/ 6,013; ped 1,404/2152 Hispanic: 476/3,049; ped 106/535 Asian: 132/2,637; ped 70/542	No	Pediatric only - less than 18 years old	European: 1.26 (1.18, 1.35) African: 1.27 (1.17, 1.39) Hispanic: 1.60 (1.16, 2.20) Asian: 1.48 (1.15, 1.91)	European: 0.55 (0.53, 0.57) African: 0.58 (0.56, 0.59) Hispanic: 0.67 (0.61, 0.72) Asian: 0.61 (0.54, 0.68)	European: 0.59 (0.57-0.61) African: 0.57 (0.55-0.59) Hispanic: 0.67 (0.62-0.72) Asian: 0.63 (0.57-0.70)	European:0.61 (0.60-0.62) African: 0.59 (0.57-0.61) Hispanic: 0.70 (0.65-0.76) Asian: 0.65 (0.58-0.71)	Yes + first 3 PCs, based on scree plot	5%	European: 1.97 (1.54, 2.51) African: 1.45 (1.07, 1.95) Hispanic: 2.43 (1.12, 5.29) Asian: 2.02 (0.79, 5.13)	European: p<0.0001 African: p=0.01 Hispanic: p=0.02 Asian/: p=0.13	European: 0.08:0.96 African: 0.06:0.96 Hispanic: 0.09:0.96 Asian: 0.08:0.96	European: 0.17:0.90 African: 0.19:0.86 Hispanic:0.29:0.86 Asian: 0.18:0.90	Only variants overlap with 1000G and discovery cohort will be used.	Yes
Breast Cancer	European: 3,939:28,860 African: #1: 274:3,527, #2: 9241:10,193, #3: 246:4,376 Hispanic: 147:2,049 Asian: #1: 45:431, #2: 15755: 16,483	No	age ≥ 18 years old and females only	European: 1.37 (1.33, 1.42) African: #1: 1.15 (1.01, 1.3), #2: (1.27), #3: (1.25) Hispanic: 1.27 (1.11, 1.40) Asian: #1: 1.45 (1.04, 2.01), #2: 1.52 (1.49,1.56)	European: 0.59 (0.58, 0.60) African: #1: 0.53 (0.50, 0.57), #2: 0.57 (0.56, 0.58), #3 0.56 (0.52, 0.6) Hispanic 0.53 (0.48, 0.58) Asian: #1: 0.61 (0.52, 0.69), #2: (0.61)	European: 0.66 (0.64,0.65) African: #1: 0.70 (0.67, 0.73), #2: NA, #3: 0.69 (0.67,0.72) Hispanic: 0.70 (0.67, 0.74) Asian: #1: 0.69 (0.62,0.75)), #2: NA	European: 0.67 (0.66,0.68) African: #1: 0.70 (0.67, 0.73)), #2 N/A, #3 .71 (0.68, 0.74) Hispanic: 0.70 (0.67, 0.74) Asian: #1: 0.70 (0.63, 0.77)), #2: N/A	yes, top 3 PCs and eMERGE site (females only - sex not included)	5%	European: #1: 2.12 (1.87, 2.4)), #2: (2.47 (2.20, 2.77)) African: #1:1.3 (0.77, 2.22), #2: 1.61 (1.38, 1.87). #3: 1.84 (1.1, 2.94) Hispanic: 2.05 (1.1, 3.83) Asian: #1: 2.75 (0.85, 8.89), #2: 2.22 (1.99-2.47)	European: p=8.90E-32 African: #1: p=0.33, #2 p=1.32E-09, #3: p=1.42e-2 Hispanic: p=0.02 Asian: p=0.091 #2: N/A	European: 0.09:0.96 African: #1: 0.06:0.95, #2: N/A, #3: 0.08:0.95 Hispanic:0.09:0.95 Asian: #1: 0.09:0.95, #2: N/A	European: 0.23:0.88 African: #1: 0.15:0.88, #2: N/A, #3: 0.18:0.88 Hispanic: 0.17:0.90 Asian: 0.18:0.90, #2: N/A	305/313	Yes
Atrial fibrillation	European: 8,613:26,900 African: 338:2,392 ; VU: 502:3,408 Hispanic: 132:933 Asian: 45:400	No	≥40 years old	European: 1.44 (1.4, 1.47) African: #1: 1.29 (1.15, 1.45), #2: 1.25 (1.14, 1.38) Hispanic: 1.39 (1.14, 1.69) Asian: 1.64 (1.1, 2.46)	European: 0.56 (0.59, 0.60) African: #1: 0.57 (0.54, 0.61), #2: 0.56 (0.54, 0.59) Hispanic: 0.568 (0.51, 0.62) Asian: 0.62 (0.53, 0.70)	European: 0.62 (0.62, 0.63) African: #1: 0.64 (0.61, 0.68), #2: 0.65 (0.63, 0.68) Hispanic: 0.67 (0.62, 0.71) Asian: 0.64 (0.56, 0.72)	European: 0.660 (0.65, 0.66) African: #10.65 (0.62-0.68), #2: 0.67 (0.64, 0.69) Hispanic: 0.67 (0.62, 0.72) Asian: 0.680 (0.60, 0.76)	PCs 1-4	3%	European: 2.46 (2.18, 2.78) African: #1: 1.75 (1.02, 3.01), #2: 2.19 (1.38, 3.38) Hispanic: 2.88 (1.16, 7.17) Asian: 4.61 (1.28, 16.67)	European: p=1.33E-47 African: p=0.042 Hispanic: p=0.023 Asian: p=0.020	European: 0.057:0.975 African: 0.053:0.970 ; VU: 0.06:0.97 Hispanic: 0.053:0.975 Asian: 0.089:0.98	European: 0.051:0.978 African: #1: 0.027:0.985, #2: 0.03:0.99 Hispanic:0.021:0.990 Asian:0.030:0.993	162/166 (98%)	Yes
Chronic Kidney Disease	European: 23,364:117,883 African: 5,235:16,467 Hispanic: 1,492:2,984 Asian: 1,030:9,896	Yes - EA, AA, HL, Asn	≥40 years old	European: 1.50 (1.49, 1.52) African (meta-validation): 1.29 (1.25, 1.32) Hispanic (meta-validation): 1.42 (1.33, 1.51) Asian (meta-validation): 1.60 (1.52, 1.67)	European: #1: 0.61 (0.61, 0.62), #2: 0.60 (0.59, 0.60), #3: 0.65 African: #1: 0.54 (0.52, 0.56); #2: 0.57 (0.55, 0.59); #3: 0.57 (0.51, 0.63) ;#4: 0.562 (0.54, 0.58), #5 0.61 (0.59, 0.63), #6: 0.55 (0.49, 0.61) #7: 0.57 Hispanic: #1: 0.57 (0.54, 0.60), #2: 0.63 Asian: #1: 0.59 (0.52, 0.66), #2: 0.61 (0.59, 0.63); #3: 0.56 (0.48, 0.64); #4: 0.57	European: #1: 0.69 (0.69,0.70), #2: 0.76 (0.75, 0.76); #3: NA African: #1: 0.69 (0.67, 0.70), #2: 0.78 (0.76, 0.79); #3: 0.76 (0.71, 0.81); #4: 0.77 (0.76, 0.79); #5: 0.73 (0.71,0.75), #6: 0.70 (0.65, 0.75); #7: NA Hispanic: E-III 0.86 (0.84-0.88); BioMe: NA Asian: UKBB-E,Asian: 0.86 (0.82-89), UKBB-SW Asian: 0.73 (0.70-0.74); E-III: 0.91 (0.87-0.95); BioMe: NA	European: #1: 0.72 (0.72, 0.73); #2: 0.77 (0.76, 0.78); #3: 0.91 African: #1: 0.69 (0.67, 0.71); #2: 0.79 (0.77, 0.80); #3: 0.77 (0.72, 0.82); #4: 0.78 (0.76, 0.79), #5: 0.75 (0.73, 0.77); #6: 0.71 (0.66, 0.76); #7: 0.81 Hispanic: #1: 0.87 (0.85, 0.89); #2: 0.89 Asian: #1: 0.86 (0.83, 0.90), #2: 0.75 (0.73, 0.77); #3: 0.92 (0.88, 0.96); #4: 0.92	Yes, diabetes and PCs	2%	European: 3.41 (3.31, 3.52) African: 2.33 (2.12, 2.54) Hispanic: 4.44 (3.78, 5.10) Asian: 3.59 (3.31, 3.88)	European: pi=3.1E-117 (meta 3 cohorts) African: p=6.6E-16 (meta 7 cohorts) Hispanic: p=9.1E-06 (meta 2 cohorts) Asian: p=1.7E-09 (meta 4 cohorts)	European (3 cohort meta-analysis): 0.54:0.96 African (7 cohort meta-analysis): 0.16:0.97 Hispanic (2 cohort meta-analysis): 0.74:0.87 Asian (4 cohort meta-analysis): 0.48:0.99	European (3 cohort meta-analysis): 0.66:0.93 African (7 cohort meta-analysis): 0.51:0.85 Hispanic (2 cohort meta-analysis): 0.45:0.95 Asian (4 cohort meta-analysis): 0.88:0.93	Yes	Yes
Colorectal cancer	European: #1: 1,311/53,722 ; #2: 573/37,641 African: #1: 56/2,409 ; #2: 42/4,067 Hispanic: #1: 70/5,221 ; #2: 8/1,042 Asian: #1: 96/5,758 ; #2: 3/375	Yes - EA, Asn	> 18	European: #1: 1.7(1.6, 1.8) ; #2: 1.6 (1.6, 1.7) African: #1: 1.1(0.9, 1.5) ; #2: 1.4 (1.1, 1.8) Hispanic: #1: 1.6(1.6, 1.8) ; #2: 1.9 (1.3, 2.6) Asian: #1: 1.43(1.2, 1.8) ; #2: 2.7 (1.5, 4.0)	European: #1: 0.67 (0.66, 0.69); #2: 0.64 (0.6, 0.7) African: #1: 0.57 (0.54, 0.65); #2: 0.59 (0.5, 0.7) Hispanic: #1: 0.63 (0.56, 0.70); #2: 0.7 (0.4, 0.96) Asian: #1: 0.65 (0.60, 0.71); #2: 0.78 (N/A)	Does not include non-genetic covariates	Does not include non-genetic covariates	No. We do not include age, sex or other non-genetic covariates.	7%	European: #1 2.9(2.5, 3.1) ;#2: 2.4 (2.1, 2.6) African: #1: 1.2(0.4-3.3), #2: 1.9 (0.9,2.97) Hispanic: #1: 1.7(0.7-3.9), #2: 4.2 (2.6,5.8) Asian: #1: 2.32 (1.3-4.2), #2: N/A	European: #1: p=8.9e-72; #2p=5e-13 African: #1 p=0.28; #2 p=0.2 Hispanic: #1 p=0.0003; #2 p=0.08 Asian: #1 p=0.00038; #2: N/A	European: #1: 0.17:0.93, #2: 0.03:0.99 African: #1: 0.1:0.93, #2: 0.02:0.99 Hispanic: #1: 0.08:0.93), #2: 0.03:0.99 Asian: #1: 0.16:0.93; #2: 0.5:0.99	European: #1: 0.02:0.99; #2: 0.02:0.99 African: #1: 0.01:0.99; #2: 0.02:0.99 Hispanic: #1: 0.01:0.99; #2: 0.04:0.99 Asian: #1: 0.02:0.99; #2: 0.37:0.99	1017562/1020292 (99.7%)	Yes
Cotonary heart disease	European: 5,970:53,171 African: 1,427:16,290 Hispanic: 337:6,809 Asian: 32:690	Yes - EA, AA, HL, Asn	Yes, >= 18 years old	European: 1.51 (1.46, 1.55) African: 1.23 (1.17, 1.30) Hispanic: 1.31 (1.17, 1.47) East Asian: 1.91 (1.32, 2.77)	European: 0.600 (0.59, 0.61) African: 0.556 (0.54, 0.57) Hispanic: 0.587 (0.57, 0.62) East Asian: 0.655 (0.55, 0.76)	European: 0.689 (0.68, 0.70) African: 0.640 (0.62, 0.66) Hispanic: 0.71 (0.68, 0.73) East Asian: 0.72 (0.65, 0.80)	European: 0.73 (0.71, 0.72) African: 0.65 (0.64, 0.67) Hispanic: 0.72 (0.70, 0.75) East Asian: 0.76 (0.68, 0.84)	No, just age and sex	5%	European: 2.36 (2.12, 2.62) African: 1.76 (1.41, 2.19) Hispanic: 2.23 (1.51, 3.29) East Asian: 2.67 (0.74, 9.61)	European: p=5.63E-57 African: p=3.75E-07 Hispanic: p=5.69E-05 East Asian: p=0.13	European: 0.085 : 0.956 African: 0.080 : 0.955 Hispanic: 0.092 : 0.952 East Asian: 0.094 : 0.952	European 0.20:0.89 African 0.16:0.91 Hispanic 0.13:0.93 East Asian 0.11:0.94	539,986/542,218 (99.6%)	Yes
Hypercholesterolemia	European: 323:4,810 African: 422:3,741 Hispanic: 539:5,780 Asian: 25:618	Yes: EA, AA, HL, Asn	Yes, >= 18 years old	European: 1.96 (1.73, 2.23) African: 2:1 (1.86, 2.37) Hispanic: 2.03 (1.83, 2.27) Asian: 1.76 (1.15, 2.77)	European: 0.67 (0.64, 0.70) African: 0.68 (0.66, 0.71) Hispanic: 0.65 (0.62, 0.67) Asian: 0.66 (0.56, 0.76)	European: 0.56 (0.53, 0.59) European: 0.59 (0.56, 0.61) Hispanic: 0.65 (0.62, 0.67) Asian: 0.63 (0.54, 0.72)	European: 0.68 (0.65, 0.71) African: 0.7 (0.68, 0.73) Hispanic: 0.72 (0.70, 0.74) Asian: 0.71 (0.62, 0.80)	Yes, + top 10 PCs	3%	European: 4.43 (2.89, 6.60) African: 2.98 (1.88, 4.60) Hispanic: 3.89 (2.68, 5.55) Asian: N/A	European: p=1.31E-12 African: p=1.73E-06 Hispanic: p=2.02E-13 Asian: N/A	European: 0.10:0.97 African: 0.07:0.97 Hispanic: 0.08:0.98 Asian: N/A	European: 0.20:0.93 African: 0.15:0.93 Hispanic: 0.23:0.93 Asian: N/A	8996/9009 (99.9%)	Yes
Obesity/BMI	European: 1,226:6,134 African: 1,396:3,215 Hispanic: 1821:4,710 East/South-East Asian: 123:724 South Asian: 159:575	Yes - EA, AA, HL, Asn	Yes, >= 18 years old	info redacted for publication by GIANT consortium	info redacted for publication by GIANT consortium	info redacted for publication by GIANT consortium	info redacted for publication by GIANT consortium	Yes, + top 4 PCs (across collumns P-R only not included in P)	3%	info redacted for publication by GIANT consortium	info redacted for publication by GIANT consortium	info redacted for publication by GIANT consortium	info redacted for publication by GIANT consortium	Yes	Yes
Prostate cancer	European: 3,529:27,729African: 260:3,035Hispanic: 62:1,528Asian: 10:337	Yes - EA, AA, HL, Asn	yes, age 35+	European: 1.94 (1.90, 1.98) African: 1.79 (1.64, 1.93) Hispanic: 1.68 (1.39,1.98) Asian: N/A	European: 0.663 (0.65, 0.67) African: 0.646 (0.61, 0.68) Hispanic: 0.67 (0.60, 0.73) Asian: 0.640 (0.50, 0.78)	European: 0.71 (0.70, 0.71) African: 0.83 (0.80, 0.85) Hispanic: 0.90 (0.86, 0.93) Asian: 0.91 (0.84 - 0.98)	European: 0.77 (0.76, 0.77) African: 0.85 (0.83, 0.87) Hispanic: 0.91(0.88, 0.94) Asian: 0.84 (0.74, 0.93)	yes, top 10 PCS and eMERGE site (males only - sex not included)	10%	European: 3.67 (3.57, 3.76) African: 2.952 (2.60, 3.30) Hispanic: 2.367 (1.70, 3.04) Asian: N/A	European: p=1.961e-162 African: p=0.77e-09 Hispanic: p=0.012 Asian: N/A	European: 0.26:0.90 African: 0.17:0.93 Hispanic: 0.09:0.97 Asian: N/A	European: 0.06:0.99 African: 0.063:0.98 Hispanic: 0.034:0.99 Asian: N/A	264/269 (98.1%)	Yes
Type 1 Diabetes	European: 361:1,944 African: 168:1,371 Hispanic: N/A Asian: N/A	No	Yes. =< 21 years old	European: 3.79 (3.27, 4.38) African: 3.48 (3.15, 3.85) Hispanic: N/A Asian: N/A	European: 0.84 (0.82, 0.85) African: 0.82 (0.80, 0.85) Hispanic: N/A Asian: N/A	European: 0.55 (0.52, 0.58) African: 0.52 (0.48, 0.57) Hispanic: N/A Asian: N/A	European: 0.84 (0.82, 86) African: 0.82 (0.79,0.85) Hispanic: N/A Asian: N/A	Yes + top 3 PCs	3%	European: 12.19 (7.29, 20.4) African: 20.45 (10.77, 38.82) Hispanic: N/A Asian: N/A	European: p=2.54E-32 African: p=1.08E-35 Hispanic: N/A Asian: N/A	European: 0.13:0.98 African: 0.18:0.98 Hispanic: N/A Asian: N/A	European: 0.09:0.99 African: 0.15:0.99 Hispanic: N/A Asian: N/A	Yes	SNPs & Weight + SNP2HLA imputation
Type 2 Diabetes	European: 2,460:22,818 African: #1\:1,659:5,086; #2 2,776:2,722; #3 401:1,494; #4 352/422 Hispanic:N/A Asian: N/A	Yes - EA, AA, HL, Asn	No	European: 1.98 (1.89, 2.07) African: #1: 1.69 (1.55, 1.83); #2: 1.78 (1.63, 1.95); #3: 1.53 (1.32, 1.76); #4: 1.34 (1.10, 1.64) Hispanic: N/A Asian: N/A	European: 0.67 (0.66, 0.68) African: #1: 0.59 (0.57, 0.60); #2: 0.59 (0.57, 0.60); #3: 0.57 (0.54, 0.60); #4: 0.56 (0.52, 0.61) Hispanic: N/A Asian: N/A	European: 0.70 (0.69, 0.71) African: #1 0.57 (0.56, 0.59); #2: 0.58 (0.55, 0.59); #3: 0.72 (0.70, 0.75), #4: 0.60 (0.56, 0.64) Hispanic: N/A Asian: N/A	European: 0.76 (0.75, 0.77) African: #1 0.62 (0.61,0.64); #2 0.62 (0.61,0.64); #3 0.74 (0.71,0.76); #4 0.63 (0.58,0.67) Hispanic: N/A Asian: N/A	eMERGE: Yes + top 20- PCs + site; MGB Biobank Yes + top 20 PCs; UAB cohorts: Yes + top 10 PCs	2%	European: 4.44 (3.60, 5.49) African: #1: 1.88 (1.32,2.69); #2: 2.35 (1.54,3.60); #3: 1.76 (0.84,3.68); #4: 1.88 (0.60,5.97) Hispanic: N/A Asian: N/A	European: p=1.94E-43 African: #1: p=0.0001; #2: p<0.0001; #3: p=0.0001; #4: p=3.7E-03 Hispanic: N/A Asian: N/A	European: 0.06:0.98 African: #1: 0.03:0.98; #2: 0.03:0.99; #3: 0.03:0.98; #4: 0.02:0.98 Hispanic: N/A Asian: N/A	European: 0.23:0.93 African: #1 0.13:0.91; #2: 0.24:0.91; #3: 0.13:0.91; #4: 0.09:0.91 Hispanic: N/A Asian: N/A	99%	Yes

Supplemental Table 3

Condition	Starting point			Development and Optimization						Multiancestral Validation				
				Demographics						Demographics				
	de novo PRS generation	GWAS	Existing Validated PRS	% female	age (median; range)	Ancestry definitions (self-reported, genetic ancestry, or both)	Genotype array	SNP Imputation strategy (method, populations (1000 genomes etc...)	Cross-ancestry approach (fine-mapping/algorithmic/none)	% female	Ancestry definitions (self-reported, genetic ancestry, or both)	Statistical method	Cross-ancestry approach (fine-mapping/algorithmic/none)	Covariants included in model
Atrial Fibrillation	No		PMID:30061737							~54%	genetic	P+T	Meta-analyses effects from multiple ancestries	PCs, age, and sex
Breast Cancer	No		PMID: 30554720							100%	both	PMID: 30554720	Meta-analyses effects from multiple ancestries	PCs, age
Coronary Heart Disease	No		PMID: 35915156							52-60%	genetic	PRS-CSx	Meta-analyses effects from multiple ancestries	PCs, age, and sex
Hypercholesterolemia	No		PMID: 34887591							~63%	both	P+T	Trans-ancestral meta-analysis and reference panel	PCs, age, and sex
Obesity/BMI *	No		To be described in upcoming (currently embargoed) paper							54-65%	self-reported	PRS-CS	Trans-ancestral meta-analysis and reference panel	PCs, age, and sex
Prostate Cancer	No		PMID: 33398198							0%	genetic	PMID: 33398198	Meta-analyses effects from multiple ancestries	PCs, age, substudy
Type II Diabetes *	No	PMIDs: 30297969, 34594039, 25102180	N/A	PRS-CSx was used without further model development and optimization					N/A	49-63%	both	PRS-CSx PMID: 35513724	Yes- PRS-CSx jointly models GWAS summary statistics from multiple populations	PCs, age, sex, and study site
Abdominal Aortic Aneurysm	Yes	PMID: 32981348 + FinnGen	N/A	1%	76.2 (+/- 8.4)	Both	MEGA	HRC	N/A	23%	both	PRS-CS + P-value thresholding	none	PCs, age, sex, smoking, and bmi
Asthma *	Yes	PMID=29273806	N/A	52%	12.82 (1-19)	Both	Illumina_Affymetrix	Minimac3 and IMPUTE4 for the two cohorts	N/A	52%	both	Bayesian	none	PCs, age, and sex
Chronic Kidney Disease	Yes	PMID:31152163	N/A	54%	56.65 (37-73)	Genetic	Illumina_Affymetrix	IMPUTE4	Meta-analyses effects from multiple ancestries (PMID=31152163)	46-60%	both	P+T	none	PCs, age, sex, and diabetes
Colorectal Cancer	Yes	PMID:30510241	PMID: 32758450 was further optimized	56-61%	64 (19-95)	Genetic	Affymetrix Axiom	1000 Genome	Meta-analyses effects from multiple ancestries	52-72%	genetic	LDPred rho=0.03	none	PCs, age, and sex are included as confounders (not as covariates)
Type I Diabetes *	Yes	PMIDs: 19956093, 30305743, 30655379	PMID: 30655379 was further optimized	~50%	13 (8-17)	Both	Illumina 550k SNP array	Hidden Markov Model, Minimac4, TOPMed Imputation Server, TOPMed (Version R2 on GRC38) Reference Panel	AA and EA cross-ancestry used local ancestry specific LD to prioritize SNPs	48%	genetic	Beta shrinkage, LD approach, BOLT-LMM, Plink, https://github.com/huiqi-qu/GRS2	AA and EA cross-ancestry used local ancestry specific LD to prioritize SNPs	PCs and sex

* pediatric conditions