npj | digital medicine

Perspective

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-02335-1

The OurHealth Study: A digital genomic cohort for cardiometabolic risk mechanisms in US South Asians

Check for updates

Shriienidhie Ganesh[1,2,20], Romit Bhattacharya[1,3,20] ✉, Aarushi Bhatnagar[1,4], Rishi Madnani[1,5], Christine Russo[1,4], Sara Haidermota[1,4], Bhaavana Oruganty[4], Harshil Bhavsar[6], Priyansh Shah[6,7], Sarah Pitafi[5], Nishant Uppal[3,5], Namrata Sengupta[8], Kenneth Rice[9], Matthew P. Conomos[9], Ravi Dave[10], Abha Khandelwal[11], Aniruddh P. Patel[1,3], Kaavya Paruchuri[1,3], Yamini Levitsky[12,13], Sanchita Singal Parulkar[3], Rohan Khera[14,15], Martha Gulati[16,17], Amit V. Khera[18,19], Whitney E. Hornsby[1,4], Latha Palaniappan[11] & Pradeep Natarajan[1,3] ✉

South Asians experience disproportionately elevated cardiometabolic disease risk yet remain underrepresented in genomic research. The OurHealth Study builds a digital biobank of US South Asian adults, integrating remote surveys, mailed biospecimens for sequencing, and electronic health record sharing to identify genetic and non-genetic drivers of cardiometabolic disease. By pairing remote participation with culturally tailored outreach, OurHealth enhances accessibility, supports granular phenotyping, and addresses logistical barriers to genomic research inclusion.

Diasporic South Asians have greater atherosclerotic cardiovascular disease (ASCVD) morbidity and mortality compared to other resident ethnic groups, documented consistently across studies done in the US, UK, and Canada[1–8]. In recognition of this disproportionate risk, the American Heart Association and the American College of Cardiology designated South Asian ancestry, defined as lineage from Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, or Sri Lanka, as a 'risk-enhancing factor' for ASCVD in 2018[9,10]. Cohort studies have yielded insights into the adverse metabolic effects of certain South Asian diets, lower levels of physical activity, high rates of visceral fat with or without obesity, and higher levels of emotional and physiological stress related to cultural factors[2,11–15]. Research suggests metabolic disease as the primary driver of cardiovascular disease (CVD) risk in this population, but the mechanisms underlying proposed pathways of early insulin resistance, beta-cell dysfunction, and a pro-inflammatory milieu are largely unclear[9].

Despite comprising 23% of the world population, South Asian individuals remain starkly underrepresented in genetic research, limiting the discovery of ancestry-enriched CVD risk variants[16–18]. As of 2021, there were 5.7 million South Asians living in the US, a 48% relative increase in size from 2010, with the largest ethnicities reported being Indian, Pakistani, and Bangladeshi[19]. Major initiatives with genomic data, such as the UK Biobank and the US-based All of Us Research Program, while transformative, continue to show underrepresentation of South Asian individuals relative to estimated country prevalences[20]. In contrast, targeted efforts within other countries, such as the UK-based Genes & Health Study and the Pakistan Risk of Myocardial Infarction Study (PROMIS), have rapidly yielded actionable insights[21,22]. For example, one largely genetic biomarker, serum lipoprotein(a) (Lp(a)), has been found to account for a greater proportion of CVD risk in South Asians compared to other ethnicities[23–25]. Additionally, at the single-nucleotide polymorphism level, impaired metabolizer alleles of *CYP2C19* are enriched and associated with poor response to clopidogrel in South Asians (Fig. 1)[22].

Larger cohorts with relevant feature-ascertainment provide greater granularity in participant characteristics, a key advantage given the cultural, linguistic, and religious diversity of South Asian populations. Distinct patterns of admixture, migration, and dietary practices, along with partially isolated genetic pools shaped by cultural and linguistic boundaries, complicate interpretation when individuals are grouped broadly as "South Asian" or "Asian – Other"[26]. Disaggregated analyses powered by large, diverse cohorts can enable genetic and clinical insights into CVD risk that would otherwise remain obscured.

The impact of building a large-scale South Asian cohort extends beyond this population. Endogamy among ethno-religious-linguistic groups and resulting patterns of consanguinity throughout South Asian history have led to an enrichment of autozygous (i.e., homozygous by descent) genotypes[27–30]. Autozygosity enriches for the possibility of identifying 'human knockouts' (i.e., individuals possessing two protein-truncating variants), and identification in South Asian individuals has already aided in drug development[29].

Digitized trials improve inclusion of underrepresented communities by reducing participant burden to expand reach beyond traditional research infrastructure and enabling integration of health tools and ancillary

A full list of affiliations appears at the end of the paper. ✉e-mail: rbhattacharya@mgh.harvard.edu; pnatarajan@mgh.harvard.edu

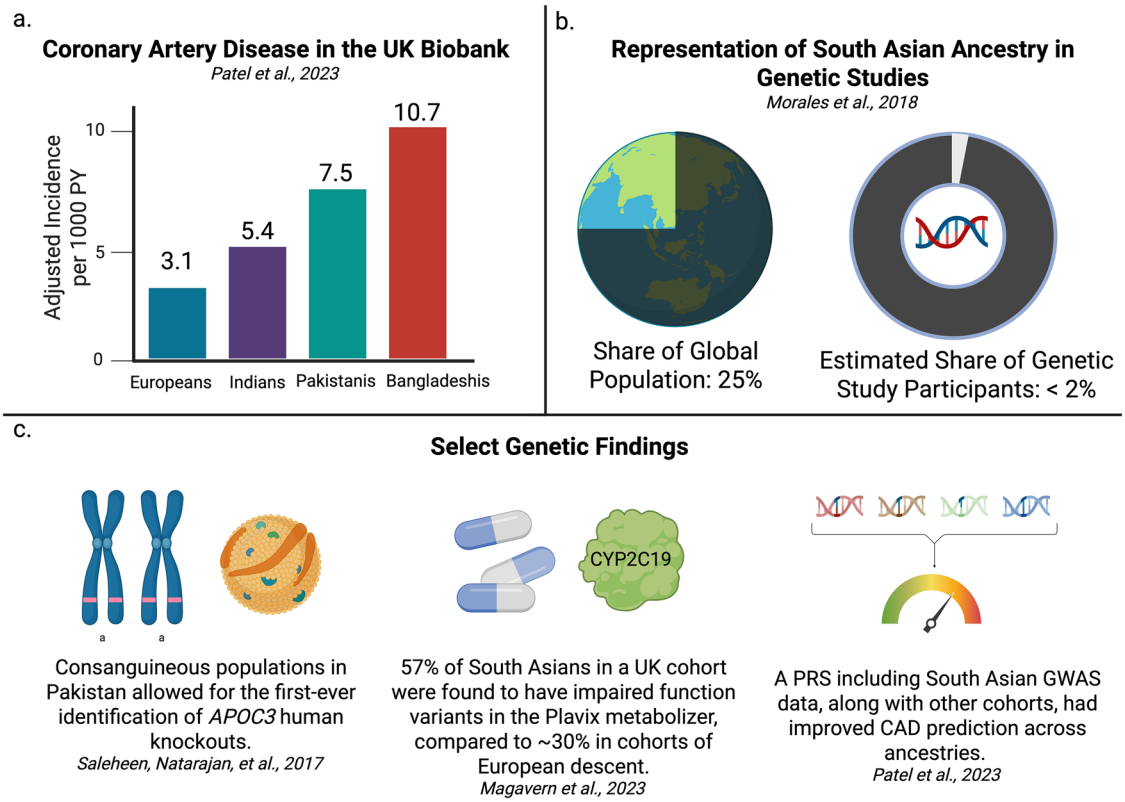

Fig. 1 | South Asian cardiovascular risk and the importance of expanding genomic data collection in this population. a An analysis of coronary artery disease incidence among South Asians and White European individuals (n = 457,473) in the UK Biobank showed that South Asians carried a hazard ratio of 2.03 relative to White European individuals, with further differences by nation of origin. **b** A 2018 analysis of the GWAS catalog examined combined cohort ancestry, finding that 2% of enrolled participants were categorized as "Other Asian." **c** Clinically relevant findings from South Asian genomic data demonstrate the utility of targeted outreach and the construction of South Asian ancestry-specific biobanks. *Created in BioRender. Madnani, R. (2025)* https://BioRender.com/3z0m6uy.

studies[31]. As digital literacy increases, decentralized models can scale recruitment across geographically dispersed and underrepresented populations, democratizing access to clinical research[32,33]. Digital studies present an opportunity to facilitate genetic discovery by improving access and enabling larger sample sizes[34–36].

We introduce the OurHealth Study, a digital nationwide biobank that investigates the elevated cardiometabolic risk of South Asians in the US. OurHealth deploys a novel digital platform used for study recruitment, study coordination, and collection of health outcomes, genomic samples, and electronic health record (EHR) information to identify genetic and non-genetic drivers of CVD risk. Its digital design also supports bidirectional communication with participants and implementation of nested studies, including OurHealth-PRS. OurHealth-PRS is a sub-study returning polygenic risk scores (PRS) for coronary artery disease (CAD) to participants to evaluate PRS acceptability and understanding. The digital infrastructure enables direct-to-participant return of genomic information and provides a foundation for evaluating the utility and acceptability of PRS-based risk stratification in an underrepresented population.

Methods

OurHealth is conducted remotely through the study's platform website (https://ourhealthstudy.org), optimized for desktop and mobile devices, and has a unified, broad consent for general research use. Institutional Certification was obtained, enabling the submission of large-scale human genomic data generated from OurHealth to an NIH-designated data repository consistent with the NIH Genomic Data Sharing Policy, study participants' informed consent, and research use limitations[37]. OurHealth aims to recruit a diverse cohort representing the full spectrum of South Asian ancestry living in the US by self report, inclusive of individuals regardless of migration history, generational status, or degree of admixture, as cardiometabolic risk profiles may vary across groups. Inclusion criteria include (a) self-identification with South Asian ethnicity, (b) age ≥18 years, and (c) residence in the US. Potential participants answer eligibility questions to verify inclusion criteria, after which they create an account and proceed with the self-paced electronic informed consent module. Participants complete health surveys, connect their EHR, and donate saliva biospecimens for sequencing.

Participants complete questionnaires using the data donation platform interface, which includes seven questionnaires: Basics, Cardiometabolic Medical History, Other Medical History, Medications, Lifestyle, Mental Health, and Family History (Table 1)[38,39]. OurHealth survey instruments include detailed ascertainment of country or territory of origin in South Asia, language identification, citizenship status, migration patterns, and family structure to enable granular phenotyping (Supplementary Data 1). Basic demographics are aligned and harmonized with other major biobanks, such as All of Us, to ease data harmonization for comparison of South Asian health data to other ancestral backgrounds[39]. After completion of the Basics and Cardiometabolic Medical History surveys, participants receive saliva collection kits in the mail, which are completed and returned to the Broad Institute's Genomics Platform for sequencing.

Juniper data platform

The Broad Institute's Genomics Platform hosts Juniper, a secure registry platform enabling direct-to-participant engagement for consent, data collection, and recontact via intuitive web and mobile interfaces[40–42]. The Juniper interface allows the study team to design, edit, track participants, and manage the data. Juniper has the functionality to send automated messaging to participants including study outcome reminders, new survey notifications, educational information, and opportunities to participate in community webinars and ancillary studies.

Table 1 | OurHealth surveys and brief descriptors

Survey	Description
Basics*†	South Asian identity, demographic information, and education/employment details
Cardiometabolic Medical History*	Heart, blood, metabolic health history, and women's health history, if applicable
Other Medical History†	Medical conditions, including endocrine, gastrointestinal, renal, and other conditions
Medications	All medications that participants are currently taking
Lifestyle†	Physical activity, smoking, and drinking habits
Mental Health†	Patient Health Questionnaire (PHQ-8) and Generalized Anxiety Disorder (GAD-7) formatted into a web-friendly module
Family History†	Participants' family health history, specifically that of biological mother, father, sibling(s), children, and grandparent(s)

*Surveys required to receive a biospecimen donation kit. †Surveys aligned with All of Us Research Program instruments.

Genomics platform

The Genomics Platform coordinates saliva sample kit shipment, biospecimen receipt, DNA isolation, sequencing, and data storage for centralized analysis. Participants receive Genotek's OGR-600 DNA collection kit via FedEx, along with an instruction sheet for completion and return of de-identified biospecimens.

OurHealth's samples undergo DNA isolation and sequencing using the blended genome exome (BGE) method, which uses the NovaSeqX 10B Flowcell followed by Dynamic Read Analysis for GENomics (DRAGEN) analysis for alignment, mapping, and variant calling. BGE uses low-coverage whole-genome sequencing (2–3x) and deep-coverage exome sequencing (30–40x), improving on SNP arrays for common variant imputation while also capturing rare variants in non-European populations[43–45]. External South Asian cohorts with whole-genome sequencing will be used for BGE imputation.

EHR integration

OurHealth uses Hugo Connect (Arboretum LifeSciences, Inc.), to offer participants the opportunity to link EHR and pharmacy data, subject to additional consent[46–48]. This functionality also provides the research team with cross-sectional and longitudinal access to clinical data. Encrypted EHR data is normalized and harmonized with de-duplication, automated ontology mapping, and multi-site integration before upload to a Secure File Transfer Protocol (SFTP) server accessible to the OurHealth research team. Hugo Connect is an approved member of the Creating Access to Real-Time Information Now Through Consumer-Directed Exchange (CARIN) alliance and adheres to strict protocols governing the sharing of any data.

Data models, quality control, and data privacy

Participants can contribute data through three sources: (1) online surveys, (2) mailed saliva kits, and (3) optional EHR sharing. Survey responses are collected and stored on Juniper, de-identified, and securely transferred to Terra, a cloud-based research environment enabling scalable storage, integration, and analysis of large-scale biomedical data[49]. Genomic data is also stored initially on Terra.

In parallel, participants consent to EHR sharing through Hugo Connect's engagement platform, which aggregates records across health systems and curates the data before transmitting de-identified EHR data via the Broad Institute's SFTP server. De-identified survey and genomic data from Terra can also be transmitted by the study team to the SFTP server, allowing for harmonization of data streams and centralized analysis within a unified secure environment (Fig. 2).

Only data and/or specimens necessary for the conduct of the study are collected, and all data are governed by Institutional Review Board (IRB)-approved research protocols, which include provisions for data security and confidentiality that persist regardless of institutional transactions. Electronic data are maintained in a secure location with appropriate protections such as password protection, encryption, and physical security measures, including locked files or restricted access areas. Similarly, all collected specimens are stored in secure, access-controlled locations such as locked laboratory spaces. Data and specimens are only shared with individuals who are part of the IRB-approved research team or are otherwise approved under the current IRB protocol. When data or specimens must be transported, either physically or electronically, secure methods are used, including encrypted files, password protection, and chain-of-custody procedures where applicable. All electronic communications with participants comply with Mass General Brigham's secure communication policies and procedures. Identifiers are removed or coded as soon as feasible, and access to the linkage between identifiers and coded data or specimens is restricted to the minimum number of research team members necessary to conduct the study. If a merger or acquisition were ever to occur, any transfer of research data would remain subject to all applicable consent terms, confidentiality protections, and legal/regulatory requirements governing human subjects research. All research staff are trained in and will adhere to Mass General Brigham's confidentiality policies and procedures for handling research data and specimens.

Data sharing

With support from the Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium [PMID: 39561770], de-identified data are made available to the broader scientific community. Data access requests are made to a dbGaP [PMID: 24297256] study accession (phs003821), and data are accessible on the NHGRI AnVIL cloud platform [PMID: 35199087]. Genomic and phenotypic data files in AnVIL are organized according to the PRIMED data model[50]. This open-science model enables broad and secure data sharing to advance genomic discovery.

Engagement ecosystem

Prior population studies have identified the need for an integrative recruitment strategy employing digital and personal approaches, including community-based organization partnerships, media outreach, and health professional advocacy[51–55]. In alignment, OurHealth uses four intersecting approaches in a multi-tier engagement strategy to optimally engage with South Asian communities (Fig. 3). (1) The clinical arm of OurHealth implements hospital inpatient and clinic outpatient workflows, South Asian cardiovascular clinic partnerships, physician referrals, and EHR-integrated research recruitment messages to reach potential participants. (2) The community arm engages participants through local and national South Asian community partnership, aligning study advertisements with community events. (3) The academic arm has implemented the Community Leader Program, engaging South Asian student groups nationally to recruit participants to OurHealth. (4) OurHealth reaches national audiences via social media platforms (Twitter/X, Instagram, Facebook), the website's education portal, and webinars with question and answer sessions focused on South Asian cardiovascular health and study updates. OurHealth's engagement strategy aims to recruit participants and build awareness about cardiovascular health and disease.

Education

OurHealth aims to be a reliable, evidence-based resource for those seeking to learn about optimizing cardiometabolic risk in South Asians, including lifestyle modifications and medication considerations where appropriate.

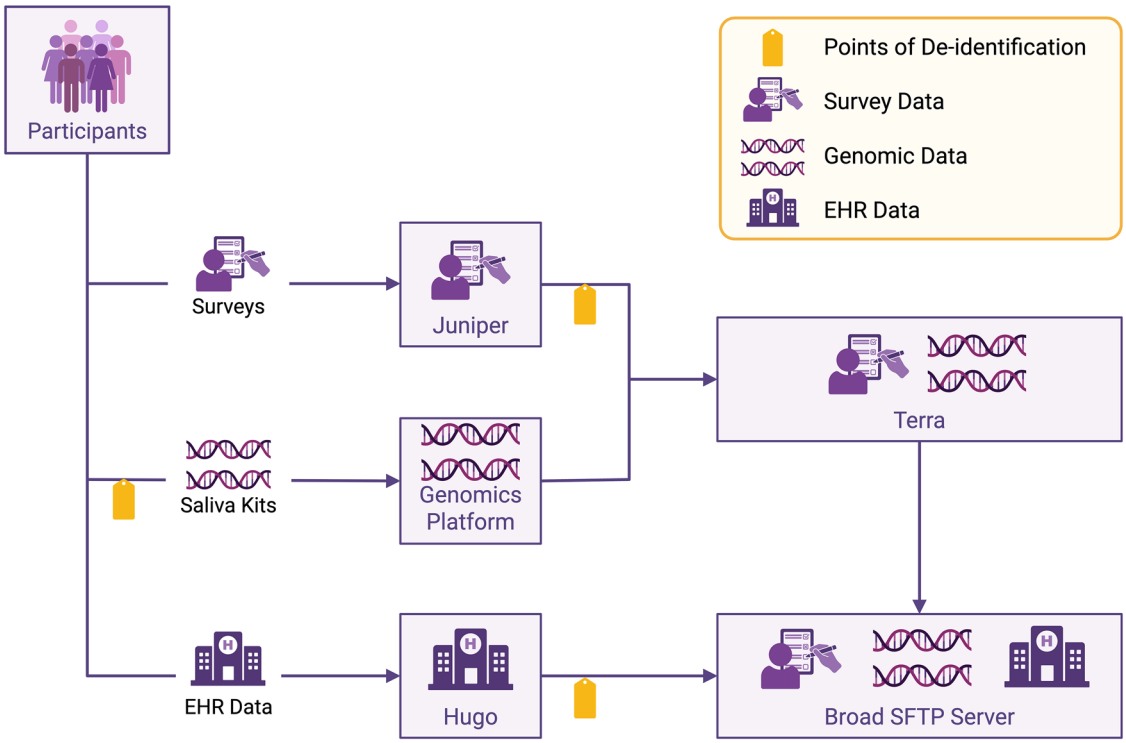

Fig. 2 | OurHealth data flow across survey, genomic, and electronic health record sources. Participants contribute survey data, mail back saliva kits, and optionally consent to EHR sharing. Survey data is tabulated and stored on the Juniper platform, after which it can be de-identified and securely transferred to the Terra platform. Saliva kits are mailed in a de-identified manner, tagged with participant ID only. Saliva kits are sequenced by the Genomics Platform, and genetic data is stored in Terra. EHR data is parsed by Hugo Connect and transferred securely to the Broad SFTP server. Genomic and survey data are securely uploaded to the Broad SFTP server, allowing for centralized analysis. EHR Electronic health record, SFTP Secure file transfer protocol. *Created in BioRender. Ganesh, S. (2025)* https://BioRender.com/aav0q4s.

OurHealth has integrated educational material from peer-reviewed manuscripts and national organizations, such as the American Heart Association and National Lipid Association, and created partnerships with groups such as the South Asian Heart Center and the Nourish Initiative at Stanford University. Educational information is available on the website's education portal (www.ourhealthstudy.org/education). OurHealth seeks to longitudinally engage with participants and the larger South Asian community through additional webinars, video, and print content to build a larger community and generate discourse.

Data analytics

OurHealth's analytical framework includes (a) genome-wide association studies (GWAS) to identify common variant associations with cardiometabolic traits and diseases, (b) rare variant burden analyses to assess the cumulative effect of low-frequency coding variants, (c) gene-based association testing, and (d) development and validation of PRS tailored to South Asian population with performance benchmarked against existing large cohorts. To address within-group heterogeneity, stratified analyses by country or territory of origin, as determined from participant survey responses, will be conducted. Additionally, to disentangle environmental from genetic contributions to disease risk, analyses will incorporate covariate adjustment for social determinants of health and other cultural factors. BGE sequencing enables detection of common regulatory variants genome-wide while providing high-confidence calls for protein-coding variants that may be population-specific. Improved imputation upon the BGE sequences relative to array-derived genotypes will allow for subsequent GWAS and PRS-based analyses.

OurHealth-polygenic risk score ancillary study

To enhance the value of the program for the study participants and raise awareness for prevention, an OurHealth ancillary single-arm observational study entitled OurHealth-PRS returns CAD PRS to consenting participants. This study received ethical and regulatory approval from the Mass General Brigham IRB. Eligible participants with sequenced genomic data use the online platform to provide consent, complete pre-disclosure genetic and health literacy surveys, receive their CAD PRS report, and complete interpretation surveys after report disclosure. The aim of this study is to assess participants' genetic literacy and acceptability of CAD genetic risk information.

The digital nature of OurHealth enables the scalable, participant-directed return of PRS, allowing integration of consent, education, report delivery, and follow-up in a centralized and accessible manner. PRS for CAD will be generated from the genetic data and traits to generate a final CAD-PRS for consenting participants. GPSMult and other polygenic scores for CAD from the Polygenic Score Catalog are computed as the weighted sum of risk alleles for each participant and standardized against the 1000 Genomes reference population[26,56,57]. An integrative score optimized in All of Us and validated in the Mass General Brigham Biobank will be utilized[58–60].

PRS return requires clear communication of the probabilistic nature of the score and limitations of interpretation to minimize psychological risk and ensure responsible genetic risk communication[61,62]. The CAD-PRS will be returned to participants via their OurHealth portal by virtual report containing general information about polygenic risk and resources to help participants understand their risk[63,64]. The first iteration of the polygenic risk report will only return PRS for CAD. Additional disease-specific PRS will be incorporated into the report once they are validated for the South Asian population.

Discussion

OurHealth demonstrates the potential of remote platforms to advance precision medicine efforts among historically underrepresented populations in research, such as South Asians. It facilitates the collection of

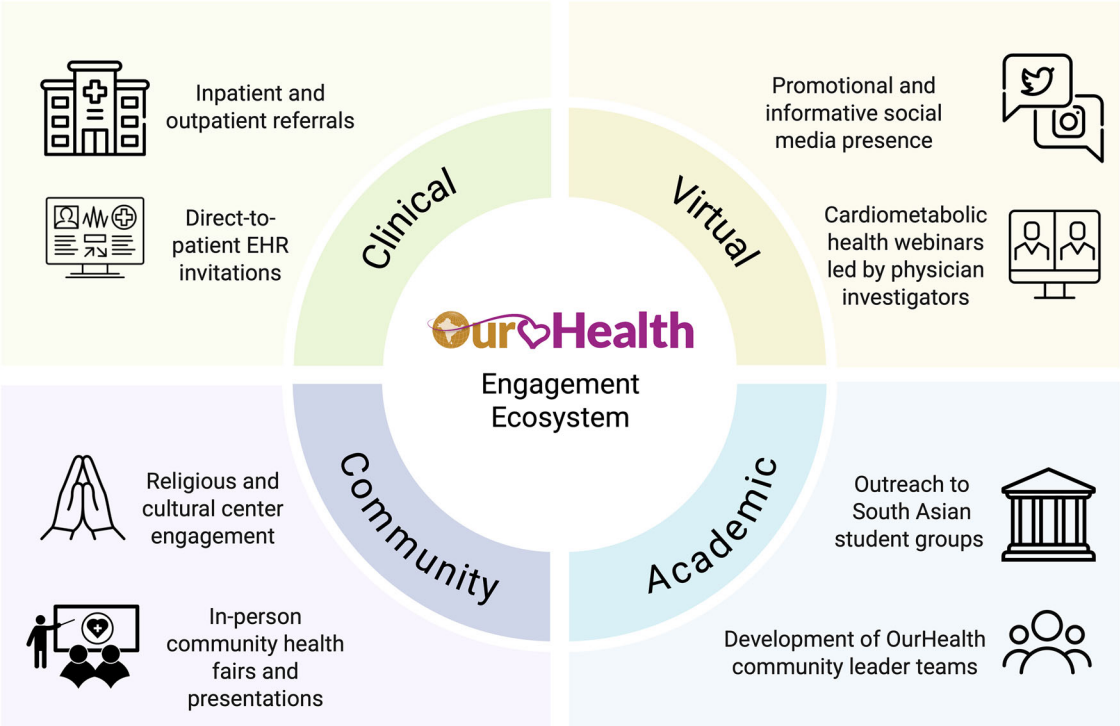

Fig. 3 | OurHealth engagement ecosystem with four main pillars. The four pillars include clinical outreach via physician relayed messages and EHR invitations, digital dissemination of study information using social media and webinars, community-based connections, and nationwide student group-led outreach initiatives. *Created in BioRender. Madnani, R. (2025)* https://BioRender.com/uivgclq.

information on regional ancestry, linguistic and cultural identities, and endogamy-driven genetic structures, all uniquely relevant to South Asian disease risk. This granularity can enhance precision medicine approaches, improve cardiometabolic disease risk stratification, address health disparities unique to South Asian populations, and open avenues for discovery in other sub-populations.

Remote recruitment strategies, such as those used by OurHealth, are increasingly supported by large-scale initiatives, which have demonstrated that digital enrollment can improve cohort diversity[65,66]. The digital nature of OurHealth offers several advantages. Online survey administration and mail-in biospecimen collection eliminate the need for in-person visits, expanding geographic reach, reducing participant burden, and increasing data collection efficiency. The digital infrastructure also enables the integration of ancillary studies, expanding the scope of research beyond baseline data collection. One such study is OurHealth-PRS, which returns PRS for CAD to South Asian participants through the secure online portal. Future ancillary studies may incorporate additional polygenic scores, behavioral interventions, and mobile health tools, demonstrating the adaptability of the digital platform for precision medicine research.

However, digital biobanks can introduce new challenges. Studies have shown that remote recruitment may exclude individuals with limited digital literacy, unreliable internet access, language barriers, or data privacy concerns[67,68]. These barriers often intersect with sociodemographic factors including age, income, education, and immigration status, necessitating the supplementation of digital strategies with community-based outreach[68,69]. To address this, OurHealth has partnered with local and national community-based organizations, leveraging trusted relationships to tailor outreach and build credibility. Additionally, a future direction will be translation of study materials into major South Asian languages to improve accessibility. In-person engagement through community events, local health fairs, and cultural programming remains a core strategy for improving cohort diversity and inclusion, particularly among individuals less likely to engage digitally. Still, OurHealth faces limitations with recruitment. As a self-enrolled, online cohort, the study is susceptible to volunteer bias, with participants differing systematically from the general South Asian population[70]. Ongoing evaluation and necessary steps will be needed to ensure equitable representation across gender, region, religion, and sociodemographic strata within the South Asian diaspora. Additionally, OurHealth does not currently include a comprehensive dietary assessment instrument, as no validated tool exists that adequately captures both traditional South Asian and Western dietary patterns. The development and validation of such an instrument represents an important near-term priority for future work, as dietary acculturation is a key factor in understanding cardiometabolic risk in diasporic populations.

The compilation of genomic information with lifestyle and family history survey data, prevalent disease survey data, and EHR data forms a powerful discovery cohort. OurHealth's BGE sequencing approach enables detection of common variants genome-wide and rare coding variants, though with reduced power for rare noncoding variants compared to high-coverage whole-genome sequencing. This design prioritizes deep coverage of protein-coding regions where population-specific rare variants are more readily interpretable for clinical translation. While rates of obesity, diabetes, and CVD rise globally, South Asians have long been found to have higher rates of disease as well as unique cultural histories leading to a higher concentration of rare variants in homozygous genotypes. The OurHealth Study has been designed to allow efficient identification of polygenic disease risk that interacts with lifestyle and modifiable risk factors. Discovery of unique pathways and mechanisms of disease in the South Asian population may inform prevention and treatment strategies applicable beyond this cohort.

Data availability

No datasets were generated or analyzed during the current study.

Received: 25 July 2025; Accepted: 31 December 2025;
Published online: 16 January 2026

References

1. Talegawkar, S. A., Jin, Y., Kandula, N. R. & Kanaya, A. M. Cardiovascular health metrics among South Asian adults in the United States: prevalence and associations with subclinical atherosclerosis. *Prev. Med.* **96**, 79–84 (2017).
2. Joshi, P. et al. Risk factors for early myocardial infarction in South Asians compared with individuals in other countries. *JAMA* **297**, 286–294 (2007).
3. Rana, A., de Souza, R. J., Kandasamy, S., Lear, S. A. & Anand, S. S. Cardiovascular risk among South Asians living in Canada: a systematic review and meta-analysis. *CMAJ Open* **2**, E183–E191 (2014).
4. Patel, A. P., Wang, M., Kartoun, U., Ng, K. & Khera, A. V. Quantifying and understanding the higher risk of atherosclerotic cardiovascular disease among South Asians—results from the UK Biobank prospective cohort study. *Circulation* **144**, 410–422 (2021).
5. Kuppuswamy, V. C. & Gupta, S. Excess coronary heart disease in South Asians in the United Kingdom. *BMJ* **330**, 1223–1224 (2005).
6. Beckles, G. L. A. et al. High total and cardiovascular disease mortality in adults of Indian descent in Trinidad, unexplained by major coronary risk factors. *Lancet* **327**, 1298–1301 (1986).
7. Wainwright, J. Cardiovascular disease in the Asiatic (Indian) population of Durban. *SA Med. J.* **43**, 136–138 (1969).
8. Walker, A. R. P. The epidemiology of ischaemic heart disease in the different ethnic populations in Johannesburg. *SA Med. J.* **57**, 748–752 (1980).
9. Volgman, A. S. et al. Atherosclerotic cardiovascular disease in South Asians in the United States: epidemiology, risk factors, and treatments: a scientific statement from the American Heart Association. *Circulation* **138**, e1–e34 (2018).
10. Arnett, D. K. et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* **140**, e596–e646 (2019).
11. Daniel, M., Wilbur, J., Fogg, L. F. & Miller, A. M. Correlates of lifestyle physical activity among South Asian Indian immigrants. *J. Community Health Nurs.* **30**, 185–200 (2013).
12. Shah, A. D., Vittinghoff, E., Kandula, N. R., Srivastava, S. & Kanaya, A. M. Correlates of pre-diabetes and type 2 diabetes in US South Asians: findings from the mediators of atherosclerosis in South Asians Living in America (MASALA) study. *Ann. Epidemiol.* **25**, 77–83 (2015).
13. Lauderdale, D. S. & Rathouz, P. J. Body mass index in a US national sample of Asian Americans: effects of nativity, years since immigration and socioeconomic status. *Int. J. Obes.* **24**, 1188–1194 (2000).
14. Chow, C. K. et al. Association of diet, exercise, and smoking modification with risk of early cardiovascular events after acute coronary syndromes. *Circulation* **121**, 750–758 (2010).
15. Lear, S. A., Chockalingam, A., Kohli, S., Richardson, C. G. & Humphries, K. H. Elevation in cardiovascular disease risk in South Asians is mediated by differences in visceral adipose tissue. *Obesity* **20**, 1293–1300 (2012).
16. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
17. Tcheandjieu, C. et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat. Med.* **28**, 1679–1692 (2022).
18. Aragam, K. G. et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).
19. Bureau, U. C. American Community Survey Data. *Census.gov.* https://www.census.gov/programs-surveys/acs/data.html.
20. Kathiresan, N. et al. Representation of race and ethnicity in the contemporary US Health Cohort All of Us Research Program. *JAMA Cardiol.* **8**, 859–864 (2023).
21. Saleheen, D. et al. The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other determinants of myocardial infarction in South Asia. *Eur. J. Epidemiol.* **24**, 329–338 (2009).
22. Magavern, E. F. et al. CYP2C19 genotype prevalence and association with recurrent myocardial infarction in British–South Asians treated with clopidogrel. *JACC Adv.* **2**, 100573 (2023).
23. Bilen, O., Kamal, A. & Virani, S. S. Lipoprotein abnormalities in South Asians and its association with cardiovascular disease: current state and future directions. *World J. Cardiol.* **8**, 247–257 (2016).
24. Paré, G. et al. Lipoprotein(a) levels and the risk of myocardial infarction among 7 ethnic groups. *Circulation* **139**, 1472–1482 (2019).
25. Patel, D. et al. Role of lipoprotein(a) in atherosclerotic cardiovascular disease in South Asian individuals. *J. Am. Heart Assoc.* **14**, eJAHA/2024/040361–T (2025).
26. Patel, A. P. et al. A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat. Med.* **29**, 1793–1803 (2023).
27. Rausell, A. et al. Common homozygosity for predicted loss-of-function variants reveals both redundant and advantageous effects of dispensable human genes. *Proc. Natl. Acad. Sci. USA* **117**, 13626–13636 (2020).
28. Narasimhan, V. M. et al. Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474–477 (2016).
29. Saleheen, D. et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).
30. Wall, J. D. et al. South Asian medical cohorts reveal strong founder effects and high rates of homozygosity. *Nat. Commun.* **14**, 3377 (2023).
31. Inan, O. T. et al. Digitizing clinical trials. *NPJ Digit. Med.* **3**, 1–7 (2020).
32. Jean-Louis, G. & Seixas, A. A. The value of decentralized clinical trials: inclusion, accessibility, and innovation. *Science* **385**, eadq4994 (2024).
33. Natarajan, P. Exceptional genetics, generalizable therapeutics, and coronary artery disease. *N. Engl. J. Med.* **391**, 957–959 (2024).
34. Flores, L. E. et al. Assessment of the inclusion of racial/ethnic minority, female, and older individuals in vaccine clinical trials. *JAMA Netw. Open* **4**, e2037640 (2021).
35. Warren, R. C., Forrow, L., Hodge, D. A. & Truog, R. D. Trustworthiness before trust - COVID-19 vaccine trials and the black community. *N. Engl. J. Med.* **383**, e121 (2020).
36. Kasahara, A. et al. Digital technologies used in clinical trial recruitment and enrollment including application to trial diversity and inclusion: a systematic review. *Digit. Health* **10**, 20552076241242390 (2024).
37. Smith, J. L. et al. Data sharing in the PRIMED Consortium: design, implementation, and recommendations for future policymaking. *Am. J. Hum. Genet*. **112**, 754–1768 (2025).
38. The All of Us Research Program Investigators The "All of Us" Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).
39. Survey Explorer – All of Us Research Hub. https://www.researchallofus.org/data-tools/survey-explorer/.
40. Broad Clinical Laboratories, The Broad Institute. Juniper [Hosted Computer Software]. (2023).
41. Bhakhri, P. et al. Count Me In: patient-partnered research to address disparities for rare cancer patients. *Ther. Adv. Rare Dis.* **5**, 26330040241304440 (2024).
42. The Heart Hive. *HeartHive*. https://thehearthive.org/.
43. DeFelice, M. et al. Blended Genome Exome (BGE) as a cost efficient alternative to deep whole genomes or arrays. Preprint at https://doi.org/10.1101/2024.04.03.587209 (2024).
44. Martin, A. R. et al. Low-coverage sequencing cost-effectively detects known and novel variation in underrepresented populations. *Am. J. Hum. Genet.* **108**, 656–668 (2021).

45. Boltz, T. A. et al. A blended genome and exome sequencing method captures genetic variation in an unbiased, high-quality, and cost-effective manner. Preprint at https://doi.org/10.1101/2024.09.06.611689 (2024).
46. Khera, R. et al. Assessment of health conditions from patient electronic health record portals vs self-reported questionnaires: an analysis of the INSPIRE study. *J. Am. Med. Inform. Assoc.* **32**, 784–794 (2025).
47. Hugo Health. *Hugo Health* https://hugo.health.
48. Hugo Health, Inc. *Hugo Connect*.
49. Broad Data Sciences Platform, The Broad Institute. Terra [Hosted Computer Software] (2023).
50. UW Genetic Analysis Center. *Primed Data Models.* (2025).
51. Chaudhary, N., Vyas, A. & Parrish, E. B. Community based organizations addressing South Asian American Health. *J. Community Health* **35**, 384–391 (2010).
52. Satagopan, J. M. et al. Experiences and lessons learned from community-engaged recruitment for the South Asian breast cancer study in New Jersey during the COVID-19 pandemic. *PLoS ONE* **18**, e0294170 (2023).
53. Kanaya, A. M. et al. Recruitment and retention of US South Asians for an epidemiologic cohort: Experience from the MASALA study. *J. Clin. Transl. Sci.* **3**, 97–104 (2019).
54. Islam, N. S. et al. Evaluation of a community health worker pilot intervention to improve diabetes management in Bangladeshi immigrants with type 2 diabetes in New York City. *Diab. Educ.* **39**, 478–493 (2013).
55. Mukherjea, A., Ivey, S. L., Shariff-Marco, S., Kapoor, N. & Allen, L. Overcoming challenges in recruitment of South Asians for health disparities research in the United States. *J. Racial Ethn. Health Disparities* **5**, 195–208 (2018).
56. Lambert, S. A. et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* **53**, 420–425 (2021).
57. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
58. Truong, B. et al. Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genomics* **4**, 100523 (2024).
59. Misra, A. et al. Instability of high polygenic risk classification and mitigation by integrative scoring. *Nat. Commun.* **16**, 1584 (2025).
60. Koyama, S. et al. Genetics and context for precision health in Greater Boston. *Nat. Commun.* **16**, 11661 (2025).
61. Abu-El-Haija, A. et al. The clinical application of polygenic risk scores: a points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* https://www.gimjournal.org/article/S1098-3600(23)00816-X/fulltext (2023).
62. Wand, H. et al. Clinical genetic counseling and translation considerations for polygenic scores in personalized risk assessments: A Practice Resource from the National Society of Genetic Counselors. *J. Genet. Couns.* **32**, 558–575 (2023).
63. National Human Genome Research Institute. *Polygenic Risk Scores*. https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores (2020).
64. Broad Institute. *Polygenic Scores Explained.* http://polygenicscores.org/explained/ (2025).
65. Klein, D. et al. Building a digital health research platform to enable recruitment, enrollment, data collection, and follow-up for a highly diverse longitudinal US cohort of 1 million people in the All of Us Research Program: design and implementation study. *J. Med. Internet Res.* **27**, e60189 (2025).
66. Naz-McLean, S. et al. Feasibility and lessons learned on remote trial implementation from TestBoston, a fully remote, longitudinal, large-scale COVID-19 surveillance study. *PLoS ONE* **17**, e0269127 (2022).
67. Tomiwa, T. et al. Leveraging digital tools to enhance diversity and inclusion in clinical trial recruitment. *Front. Public Health* **12**, 1483367 (2024).
68. Goodson, N. et al. Opportunities and counterintuitive challenges for decentralized clinical trials to broaden participant inclusion. *NPJ Digit. Med.* **5**, 58 (2022).
69. Rebbeck, T. R. et al. A framework for promoting diversity, equity, and inclusion in genetics and genomics research. *JAMA Health Forum* **3**, e220603 (2022).
70. Guo, X., Vittinghoff, E., Olgin, J. E., Marcus, G. M. & Pletcher, M. J. Volunteer participation in the Health eHeart study: a comparison with the US population. *Sci. Rep.* **7**, 1956 (2017).

Acknowledgements

We gratefully acknowledge the participants of the OurHealth study, without whom this research would not be possible. Research reported in this publication was supported by the National Institutes of Health for the project "Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium", with grant funding for Study Site FFAIR-PRS (U01HG011719) to P.N., and the Coordinating Center (U01HG011697) to P.N., M.P.C., and K.R. R.B. is supported by the Harvard Catalyst K12/CMeRIT Award (1K12TR004381-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

P.N. conceived and supervised the study. S.G., R.B., W.E.H., and R.M. drafted the manuscript. S.G. and R.M. contributed graphical illustrations. A.B., C.R., S.H., B.O., H.B., P.S., S.P., N.U., N.S., K.R., M.P.C., R.D., A.K., A.P.P., K.P., Y.L., S.S.P., R.K., M.G., A.V.K., and L.P. reviewed the manuscript and provided comments. All authors read and approved the final version of the manuscript.

Competing interests

P.N. reports research grants from Allelica, Amgen, Apple, Boston Scientific, Cleerly, Genentech / Roche, Ionis, Novartis, and Silence Therapeutics, personal fees from Allelica, Apple, AstraZeneca, Bain Capital, Blackstone Life Sciences, Bristol Myers Squibb, Creative Education Concepts, CRISPR Therapeutics, Eli Lilly & Co, Esperion Therapeutics, Foresite Capital, Foresite Labs, Genentech / Roche, GV, HeartFlow, Magnet Biomedicine, Merck, Novartis, Novo Nordisk, TenSixteen Bio, and Tourmaline Bio, equity in Bolt, Candela, Mercury, MyOme, Parameter Health, Preciseli, and TenSixteen Bio, royalties from Recora for intensive cardiac rehabilitation, and spousal employment at Vertex Pharmaceuticals, all unrelated to the present work. A.V.K. is an employee and holds equity in Verve Therapeutics and has received consulting fees from Arboretum Therapeutics. R.B. received consulting fees from Casana Care, Inc, and Novartis unrelated to the present work. M.G. received consulting fees from Medtronic, Bayer and New Amsterdam, and serves on a DSMB for Merck; all unrelated to this present work. N.U. has worked at the American Cancer Society unrelated to the submitted work.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-025-02335-1.

Correspondence and requests for materials should be addressed to Romit Bhattacharya or Pradeep Natarajan.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

[1]Cardiovascular Research Center & Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [2]Case Western Reserve University School of Medicine, Cleveland, OH, USA. [3]Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [4]Program in Medical & Population Genetics and Cardiovascular Disease Initiative, Broad Institute of MIT & Harvard, Cambridge, MA, USA. [5]Harvard Medical School, Boston, MA, USA. [6]Albert Einstein College of Medicine, The Bronx, NY, USA. [7]World Youth Heart Federation, New York, NY, USA. [8]Broad Institute of MIT & Harvard, Cambridge, MA, USA. [9]Department of Biostatistics, University of Washington, Seattle, WA, USA. [10]Division of Cardiology, David Geffen School of Medicine, Los Angeles, LA, USA. [11]Department of Cardiovascular Medicine, Stanford School of Medicine, Stanford, CA, USA. [12]Heart and Vascular Institute, Newton Wellesley Hospital, Newton, MA, USA. [13]Department of Medicine, Tufts University School of Medicine, Boston, MA, USA. [14]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. [15]Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, CT, USA. [16]Barbra Streisand Women's Heart Center, Cedars-Sinai Smidt Heart Institute, Los Angeles, CA, USA. [17]Baim Institute for Clinical Research, Boston, MA, USA. [18]Verve Therapeutics, Boston, MA, USA. [19]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA, USA. [20]These authors contributed equally: Shriienidhie Ganesh, Romit Bhattacharya. ✉e-mail: rbhattacharya@mgh.harvard.edu; pnatarajan@mgh.harvard.edu