nature genetics

Perspective

https://doi.org/10.1038/s41588-022-01274-x

The Singapore National Precision Medicine Strategy

Received: 30 March 2022

Accepted: 30 November 2022

Published online: 19 January 2023

Check for updates

Eleanor Wong[1], Nicolas Bertin[1], Maxime Hebrard[1], Roberto Tirado-Magallanes[1], Claire Bellis[1,2], Weng Khong Lim[1,3,4,5], Chee Yong Chua[6], Philomena Mei Lin Tong[7], Raymond Chua[7], Kenneth Mak[7], Tit Meng Lim[8], Wei Yang Cheong[7], Kwee Eng Thien[9], Khean Teik Goh[7], Jin-Fang Chai[10], Jimmy Lee[11,12], Joseph Jao-Yiu Sung[12], Tien Yin Wong[13], Calvin Woon Loong Chin[14], Peter D. Gluckman[15,16], Liuh Ling Goh[17], Kenneth Hon Kim Ban[18,19], Tin Wee Tan[18,19], SG10K_Health Consortium*, Xueling Sim[10], Ching-Yu Cheng[13,20,21,22], Sonia Davila[3,4,23], Neerja Karnani[15,19,24], Khai Pang Leong[17], Jianjun Liu[1,25], Shyam Prabhakar[1], Sebastian Maurer-Stroh[24,25,26], Chandra Shekhar Verma[24], Pavitra Krishnaswamy[27], Rick Siow Mong Goh[28], Irenaeus Chia[29], Clarissa Ho[29], Doreen Low[29], Suchin Virabhak[29], Jacklyn Yong[29], Weiling Zheng[29], Shih Wee Seow[29], Yee Kwang Seck[30], Mingshi Koh[31], John C. Chambers[12,29,32] ✉, E. Shyong Tai[10,25,29,33] ✉ & Patrick Tan[1,3,5,29,34] ✉

Precision medicine promises to transform healthcare for groups and individuals through early disease detection, refining diagnoses and tailoring treatments. Analysis of large-scale genomic–phenotypic databases is a critical enabler of precision medicine. Although Asia is home to 60% of the world's population, many Asian ancestries are under-represented in existing databases, leading to missed opportunities for new discoveries, particularly for diseases most relevant for these populations. The Singapore National Precision Medicine initiative is a whole-of-government 10-year initiative aiming to generate precision medicine data of up to one million individuals, integrating genomic, lifestyle, health, social and environmental data. Beyond technologies, routine adoption of precision medicine in clinical practice requires social, ethical, legal and regulatory barriers to be addressed. Identifying driver use cases in which precision medicine results in standardized changes to clinical workflows or improvements in population health, coupled with health economic analysis to demonstrate value-based healthcare, is a vital prerequisite for responsible health system adoption.

Advances in molecular biology and genetics have yielded remarkable achievements in medicine, including biological therapies targeting key proteins[1] and harnessing RNA molecules for pandemic vaccines[2]. Despite these advancements, many national health systems remain under pressure, driven by aging populations, increasing prevalence of chronic conditions and rising numbers of patients presenting with advanced stage disease. It is estimated that most individuals will spend 13.2% of their lives in ill health, requiring a minimal global healthcare spending of 6.6% gross domestic product[3,4]. This is projected to increase over coming decades, ultimately leading to unsustainable healthcare costs.

Faced with these challenges, healthcare systems and funders are exploring transformative approaches to deliver high-quality healthcare while maintaining equitable and fair access. Precision medicine involves the integration of biological, medical, lifestyle and environmental information to inform treatments and intervention strategies

A full list of affiliations appears at the end of the paper. ✉e-mail: john.chambers@ntu.edu.sg; mdctes@nus.edu.sg; patrick.tan@precise.cris.sg

for specific groups and individuals. Individual genomic profiling is a key enabler of precision medicine, as genetics has been proposed to explain up to 30% of healthcare outcomes[5]. Genomic signals have shown success in pharmacogenomics[6] and patient stratification for cancer therapy[7] (here, the term 'genomics' is used to cover DNA-based information such as genetics and epigenetics, RNA–based transcriptomics, and also protein–metabolite information). Precision medicine can also be harnessed for precision public health (PPH), where population health data is combined with genetics to identify multimodal risk factors for targeted intervention and interception before the presentation of overt disease[8].

Precision medicine carries further potential for new insights into disease mechanisms, facilitating target identification, drug repurposing and asset development as therapies with supporting genetic evidence have higher clinical success[1,9–12]. Together with PPH, precision medicine brings opportunities for public–private collaborations involving academic, public healthcare and private entities to develop advanced technological and analytical capabilities leveraging data to provide a comprehensive picture of an individual's disease risk and strategies for promoting healthspan.

Driven by these promises, large-scale efforts in precision medicine have been initiated worldwide[13], including the All of US program (USA), the 100,000 Genomes Project (UK) and Australian Genomics[14–16]. While each program is tailored to the needs of funding agencies and stakeholders, a common foundational feature of these programs, driven by the falling cost of genome sequencing, is a focus on large-scale genomic databases in which other data modalities such as proteomics and clinical and research phenotypes are layered. The increasing proportion of older populations in many countries has also raised substantial interest in applying PPH to research and policies related to healthy aging. These efforts require the convergence of technology, ethics, education, industry and social sciences to optimize national benefits and minimize potential harms.

Global disparities in Asian precision medicine

Despite its promise, concerns are growing that the benefits of precision medicine will be restricted to a 'privileged few' populations sufficiently equipped to carry out such efforts, resulting in growing disparities in this fast-emerging field[17,18]. Potential inequities either globally or within countries can pervade multiple areas, including the transferability of findings across populations, access to innovative therapies, limited healthcare financing to introduce new clinical workflows and having a sufficiently skilled workforce to exploit precision medicine findings. These disparities are reflected by current biases in our understanding of global genomic variation and, specifically for this Perspective, the under-representation of Asia. Currently, 78% of genome-wide association studies predominantly contain participants from North America and Europe[19]. While recent efforts by gnomAD and TOPMed have attempted to increase diversity, the percentage of individuals with Asian ancestry in these two databases is below 10% (refs. [20,21]). The lack of genomic diversity in precision medicine data comes with risks and missed opportunities. For instance, applying the wrong ancestral genetic background in managing patients with genetic conditions can lead to incorrect diagnoses[22] and genetic risk models such as polygenic risk scores constructed from one population may exhibit lower prediction accuracies when applied to other populations[23–25]. Relying only on one population for disease and genetic association studies may also limit potential discoveries, as evidenced by recent 'multi-ancestry' analyses[26–28]. Beyond genetics, many clinical conditions are known to manifest differently in Asian populations. For example, South Asians and East Asians develop type II diabetes at a lower body mass index than Europeans[29,30], likely attributable to higher abdominal fat deposition in the former leading to increased insulin resistance.

While established studies such as BioBank Japan, the China Kadoorie Biobank, the South Asia Biobank and the UK Biobank have begun to address the need for precision health research involving Asian populations, these resources have important limitations. Biobank Japan focuses on disease cases, limiting the identification of prospective disease relationships and biomarkers. While China Kadoorie, the South Asia Biobank and the UK Biobank have large sample sizes and a prospective design, none of these studies focus on Southeast Asia, a region of considerable genomic and cultural diversity.

Singapore National Precision Medicine program

Singapore, with a land area of 724 km^2, has a total population of 5.5 million. This includes Chinese (74.2%), Malay (13.7%) and Indian (8.9%) of East Asian, Southeast Asian and South Asian ancestry, respectively, along with Eurasian and Peranakan communities representing admixtures of European and Asian descent and Chinese with Malay–Indonesian heritage[31,32]. This diverse composition allows the conduct of translational and clinical programs across multiple populations, particularly for ancestries under-represented in existing population genomic research. Singapore's stable sociopolitical environment, built on strong public trust in the government and robust collaborations between public agencies and local research communities, has allowed the country to embark on a large-scale precision medicine initiative. The nation's advanced research infrastructure enables comprehensive participant phenotyping and extensive biological sample collection, permitting generation of contemporary data types such as microbiomes and time-series behavioral profiles inferred through wearables. Singapore's data science capabilities facilitate accurate interpretation of precision medicine data with safe data governance and sharing, seeding opportunities for deep gene and environment interaction analyses distinctive to Southeast Asian populations. In parallel, the country's restricted geographical footprint facilitates the longitudinal analysis of consented patient trajectories in a 'birth-to-death' fashion, and the availability of a unique national identifier system enables automatic data capture of electronic health records (EHRs) and linkage with genomic and other national-level datasets relevant to health and disease, including disease registries, health-promotion programs and educational, financial, geospatial and environmental records.

In healthcare, Singapore is known for its impressive clinical outcomes with one of the world's longest life expectancies, while spending 2.2% of gross domestic product on government health expenditure in 2019. However, Singapore is also one of Asia's most rapidly aging societies, with 22.5% of the nation's population consisting of people >65 years by 2030 (refs. [4,33]). Unfortunately, many of these individuals will live with poor health, estimated at one in three older Singaporeans suffering from hypertension, one in seven with diabetes and two in five with high cholesterol[34]. Driven by these concerns, the Singapore Ministry of Health (MOH) is exploring transformative approaches for healthy aging and sustainable healthcare, emphasizing pan-life-course health promotion, disease prevention and affordable, high-quality medical services. Precision medicine has been identified as one promising approach.

The Singapore National Precision Medicine (NPM) program is a whole-of-government 10-year strategy to establish precision medicine as a peak of research excellence for Singapore, ultimately improving the nation's health by identifying clinical applications that are cost-effective, sustainable and relevant to Singapore communities. Launched in 2017 and building on initiatives abroad and local pilot studies[35–37], NPM adopted a three-phase approach to provide tangible short-term deliverables while having clear sight on long-term objectives (Fig. 1). NPM pursued a collaborative philosophy leveraging on existing infrastructure and national partners, including public healthcare systems, research agencies and government agencies including the MOH, Chief Health Scientist Office, Government Technology Agency, and Integrated Health Information

https://doi.org/10.1038/s41588-022-01274-x

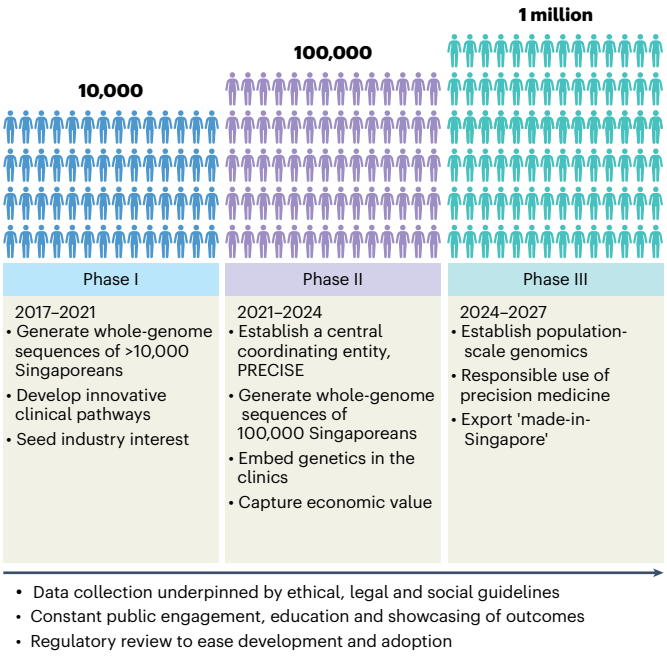

Fig. 1 | Overview of Singapore's NPM program. Singapore's NPM program is a whole-of-government 10-year roadmap. The program adopted a three-phased approach designed to have tangible short-term deliverables while having a clear vision of long-term objectives to prepare Singapore for precision medicine.

Systems for linkage of EHRs to genomic data for consented participants. Crucially, early engagement with these partners enabled the formulation of pathways to mitigate future barriers and concerns, thereby increasing the likelihood of adopting strategies that add value to local healthcare systems and are aligned to national population health goals.

Phase I (2017–2021) was a 'proof-of-concept' phase demonstrating the feasibility of large-scale genomic data generation with EHR linkage. Pathways were also developed to return incidental genetic findings to individuals participating in genomic research studies[38] and resource-utilization studies (RUS) conducted to facilitate the clinical adoption of precision medicine. NPM Phase II (ongoing) is a 'proof-of-value' phase, in which NPM is expanding sequencing efforts to a large-scale population cohort incorporating other phenotypic and molecular data to enrich the genomic data, and clinical adoption of precision medicine pursued through clinical implementation pilots (CIPs). In Phase III, NPM will expand up to one million Singaporeans, providing further depth and coverage of the Asian genotype, incorporating social and environmental information. While the specific Phase III design is still under evaluation, it will certainly consider the most contemporary findings and lessons from other programs, and the target of one million genomes is consistent with the guidance of key opinion leaders that realizing the value of population-scale sequencing will require sequencing tens of millions of individuals to enable the unbiased, comprehensive study of our genome and the variation therein[39]. At the end of NPM's three phases, it is envisioned that validated precision medicine applications will be routinely used in Singapore, with a prepared healthcare system and workforce. NPM will also galvanize the Singapore research community to investigate deep multidisciplinary questions focused on Southeast Asian populations and cultivate public–private partnerships, strengthening not only the nation's genomic and biotechnology ecosystem but also adjacent sectors in information technology, cybersecurity and artificial intelligence in line with national strategies for homegrown deep technology companies.

NPM workgroups and early insights into the Singapore genome

Due to its multifaceted nature covering ethics, law, society, technology, healthcare and industry, the success of NPM Phase I required coordination and integration across multiple domains. Six NPM workgroups were established to engage, interact and work with public agencies and stakeholders whose support and participation were deemed critical to ensure active representation. These workgroups included Regulation and Ethics, Public and Community Trust, Enabling Platforms, Clinical Adoption, Industry Development and Workforce Development (Table 1). Each workgroup was tasked to identify critical barriers impeding precision medicine in Singapore and strategies to overcome these barriers. Some key achievements of the NPM Phase I workgroups included issuing a national moratorium[40] on the use of genetic information arising from biomedical research for insurance underwriting or employment, rolling out regulatory sandboxes for clinical genetic services, and public consultations assessing community attitudes towards sharing precision medicine data with industry. Notably, public-engagement exercises by the Public and Community Trust workgroup revealed that the Singaporean community was comfortable with academic and healthcare researchers using precision medicine data for research but were more cautious about sharing the information with specific industry segments, such as internet and social media companies, and also insurers unless there was clear evidence of public benefit[41]. While initial studies suggest that such risk perceptions are common across ancestry groups[41,42], the workgroup also noted differences between community groups on the interpretation of specific words and language, which required cultural sensitivity to accurately communicate the concepts of precision medicine[43]. Future investigations are planned to deepen our understanding of risk communication, perception and management among different communities.

A headline deliverable of the Enabling Platforms workgroup in Phase I was generating a genomic reference database of 10,000 healthy Singaporeans (SG10K_Health). While underpowered to detect rare pathogenic genetic variants, SG10K_Health provided a starting catalog of common genetic variants in Chinese, Malay and Indian (allele frequency >1%) individuals, representing a baseline reference for population-based filtering based on allele-frequency thresholds to better interpret related diseases. To establish SG10K_Health, NPM collaborated with six local prospective cohorts[44–47] to sequence 10,323 healthy consented individuals. Key challenges included setting up a laboratory workflow with sufficient flexibility to accommodate DNA samples of varying quality from different cohorts, establishing a whole-genome sequencing (WGS) analysis workflow for population-scale single-sample-level variant calling and performing joint calling of variants across the SG10K_Health cohort using cloud computing. After stringent quality control following GATK best practices[48,49], the SG10K_Health joint called dataset was created for the community (n = 9,770 genomes; 58.4% Chinese, 21.8% Indian and 19.5% Malay). To provide better coverage of the particularly under-represented South Asian and Southeast Asian populations, SG10K_Health contained a higher proportion of Indian and Malay genomes than Singapore's resident population distribution (Methods). Analysis of SG10K_Health using the human genome GRCh38 reference identified 179,418,917 small variants (158,331,366 single-nucleotide variants and 21,087,551 indels), of which 92.8% (166,559,124) had minor allele frequencies of less than 1% (Table 2 and Methods). Using Variant Effort Predictor (VEP)[50] and Loss-of-Function Transcript Effect Estimator (LOFTEE)[21] to predict functional consequences, 1,147,135 protein-coding missense, 595,783 protein-coding synonymous and 104,569 (including 63,224 high-confidence) protein-coding loss-of-function (LoF) variants were identified. In total, 25% (278,715 of 1.1 million) of human allelic variations previously associated with clinically relevant phenotypes[51] were also observed in SG10K_Health, of which nearly half (131,720 variants) were reported

Table 1 | NPM Phase I workgroups

Workgroup	Challenges and objectives	Major NPM Phase I achievements
Regulation and Ethics	• Lack of guidelines for the provision of clinical genomic services • Lack of legal frameworks for preventing discrimination due to genetics	• Issued code of practice for delivery of clinical genetic services[52] • Passed moratorium on the use of genetic information for insurance underwriting[40]
Public and Community Trust	• Few data on Singapore public's understanding and acceptance of precision medicine and sharing of healthcare data • Conduct public and community consultation, education and outreach	• Population surveys on public acceptance of precision medicine[41,42] • Citizen's jury on sharing of precision medicine data with industry
Enabling Platforms	• Lack of scalable infrastructure for large-scale genomic sequence data production • Lack of information technology software pipelines for at-scale whole-genome analysis • Formulate data-sharing policies and platforms for accessing NPM data	• Generation of 10,323 whole-genome sequences • Established genomic analytic pipeline compatible with international gold standards • Successful proof of concept integrating genomic data with national-scale EHRs
Clinical Adoption	• Absence of defined processes for converting research-grade findings into clinical management • Lack of accepted economic models for assessing the cost-effectiveness of precision medicine • Lack of a Singapore genome database as a reference for patients with genetic disease	• Development of clinical pathways for return of genetic incidental findings • Commissioning of RUS for precision medicine use cases • Accelerate clinical workflows through provision of Asian reference databases of genomic normality
Industry Development	• Overall unfamiliarity of overseas companies with the Singapore precision medicine and research ecosystem • Promotion of Singapore as a gateway to Asia	• Establish public–private partnerships for genomic data production • Consortium framework for companies interested in engaging NPM
Workforce Development	• Overall lack of critical mass in workforce skill sets required for precision medicine • Map overall numbers of staff required for different areas of precision medicine (for example, genetic counselors, bioinformatics staff) • Define pathways to send qualified staff for external training or repurposing staff from adjacent industry sectors	• Workforce projections to estimate numbers of qualified staff required for national-scale precision medicine • Provision of in-service scholarships for clinical genetic training • Developed workforce development roadmap for tiered genetic counseling services

as benign (Table 3). The SG10K_Health dataset can be accessed via various NPM data-access platforms (https://npm.a-star.edu.sg/).

Developing clinical pathways for precision medicine

Knowledge and insights derived from population-scale genomic data are necessary but insufficient to transform healthcare. Therefore, to drive the clinical adoption of precision medicine, the Clinical Adoption workgroup proactively engaged stakeholders, including clinicians, patients, funders and healthcare systems, to comprehensively map the complete end-to-end journey for embedding genomic technologies in the clinic (Fig. 2). Such engagements were essential to identify barriers to implementation and potential strategies for overcoming them, highlighting not only the essentiality of identifying appropriate clinical indications in which outcomes would be improved by genomic testing but also clinician champions for these indications.

One hurdle to implementation was the heterogeneity of laboratory-developed assays with no clear regulations around their use. Thus, the Regulation and Ethics workgroup developed a Code of Practice[52] to outline standards for providing clinical laboratory genetic testing and services, covering categories of tests needed to be prescribed by a qualified medical geneticist and scenarios in which pre- and post-test genetic counseling were required. In parallel, the Workforce Development workgroup identified a shortfall in genetic counselors and initiated processes to clearly define competencies and licensing for counselors in a way that would offer 'title protection' and career progression paths[53]. Refinement of the roles and functions of counselors also enabled assessment of the extent to which other allied health professionals could be re-trained to facilitate the expansion of clinical capacity while avoiding counselor overloading. For example, the workgroup found that competencies required to construct a three-generation pedigree could be readily delivered by allied health professionals after specific training, while applying knowledge of genetics in the risk assessment of individual patients would require deeper skill sets.

Another challenge addressed by the Clinical Adoption workgroup, in preparation for large population studies, was the establishment of clinical pathways to return incidental findings, in keeping with growing expectations that such findings should be proactively managed[54,55]. Similar to other studies[56–59], 1–2% of participants in NPM harbored previously undetected pathogenic variants in clinically actionable genes. The workgroup collaborated with local healthcare institutions to establish clinical frameworks in which participants consented for recall were recontacted and guided through a systematic workflow conducted by qualified medical geneticists and counselors, involving a full health assessment, redrawing of blood samples, validation of sequencing results in a clinically accredited laboratory and collecting detailed family medical histories[38]. After evaluation by a multidisciplinary team, the results were returned to participants by a genetic counselor with subsequent referral to appropriate specialty clinics. This return-of-findings platform will be expanded in future NPM phases to cope with increasing levels of population-scale sequencing and as additional variants of medical relevance are identified.

A third important challenge was assessing the true impact of precision medicine on healthcare costs as most healthcare interventions are cost-effective rather than cost saving[60]. Given finite resources, NPM's stakeholders posited that adopting new technologies such as genome sequencing would likely involve important tradeoffs with other medical services or treatments. While health technology assessment is often used to guide the evaluation of these tradeoffs, deeper conversations revealed uncertainties whether current health economic models, frequently used to evaluate new therapeutics, were relevant to precision medicine. Stakeholders also raised concerns about the specific indications in which genomic tests would be warranted, how downstream therapeutic decisions would be influenced based on the test results, as such information is required for economic assessment.

Table 2 | Variants observed in all individuals in NPM Phase I SG10K_Health WGS

	Variants in WGS		Median variants per genome			
	Number of variants	Number of variants with MAF <1%	Number of variants	s.d.	Number of variants with MAF <1%	s.d.
Total variants	179,418,917	166,559,124	4,106,905	207,880.84	393,367	37,222.15
Variant type						
SNVs	158,331,366	148,665,318	3,501,477	129,196.15	346,317	32,594.75
Indels	21,087,551	17,893,806	602,448	84,126.34	47,139	4,794.15
Functional prediction						
Synonymous	595,783	566,740	10,940	405.29	1,084	123.49
Missense	1,147,135	1,114,839	11,147	447.75	1,469	155.03
LoF	104,569	101,917	722	44.33	82	11.64
LoF (high confidence)	63,224	62,324	207	19.61	35	6.23

Median counts and s.d. counts per genome are also presented. See Methods for details on genome mapping, joint variant calling and quality-control filters. MAF, minor allele frequency.

Table 3 | Observed number of clinically relevant variants categorized based on ClinVar clinical relevance values: benign, likely benign, likely pathogenic, pathogenic and variants of unknown relevance

Clinical relevance	Total variants		
	Private	Rare	Common
Benign	10,888	37,397	83,435
Likely benign	19,547	46,116	7,214
Likely pathogenic	1,029	756	3
Pathogenic	1,315	1,072	20
VUS	25,150	41,735	1,352
Other	328	704	654

Numbers of private (variant found in only one individual), rare (allele frequency ≤1% and more than one individual) and common (allele frequency >1%) variants are shown. See Methods for details on variant classification. VUS, variants of unknown significance.

Other concerns included the potential for technologies such as WGS to generate large numbers of incidental findings ironically resulting in more, rather than less, healthcare demand. These concerns are not unique to Singapore, as policy makers, administrators and regulators require a compelling value proposition to approve more therapies incorporating precision medicine[61,62].

In an initial attempt to generate models that would inform implementation, NPM commissioned RUSs in pediatric genetic diseases, familial hypercholesterolemia and hereditary cancer (Table 4). The RUSs focused on establishing national consensus on clinical pathways in which genomic tests could guide diagnosis and treatment locally. To limit uncertainties around incidental findings, for familial hypercholesterolemia and hereditary cancer, the technology considered was a targeted panel rather than WGS and whole-exome sequencing. In each RUS, clinicians worked with health economists to construct economic models comparing the proposed pathways (with genomic testing) to existing pathways to evaluate potential impacts on resource utilization. Notably, sensitivity analyses identified certain parameters likely to substantially impact the cost and the cost-effectiveness of the identified clinical pathway, highlighting important but sometimes under-appreciated considerations for improving the economics of genomic testing. For example, the familial hypercholesterolemia and hereditary cancer RUSs underscored the need to improve processes to increase the proportion of first-degree relatives coming forward for testing after identifying an index case. The familial hypercholesterolemia study also identified a need to optimize treatment with low-cost cholesterol-lowering medications (statins) to reduce the number of patients who require high-cost medications (proprotein convertase subtilisin–kexin type 9 (PCSK9) inhibitors). These parameters thus represent targets for focused efforts to improve the economics of precision medicine by modulating these factors.

Ambitions and goals of NPM Phase II

NPM Phase II (2021–2024) is a key initiative under Singapore's Research, Innovation and Enterprise 2025 (RIE2025) master plan. Reflecting Phase II's expanded scope and ambition, 'Precision Health Research, Singapore' (PRECISE, https://www.npm.sg) was established under the MOH to coordinate and administer the program. Key activities to be undertaken in NPM Phase II include establishing the PRECISE-SG100K longitudinal population cohort and CIPs to continue the clinical adoption of precision medicine.

The PRECISE-SG100K population cohort

Identifying mechanisms underlying disease risk in the Singapore population and developing affordable, effective and scalable strategies for disease-risk stratification and prevention are urgent national priorities. Longitudinal population studies, comprising phenotypic measurements and biological sample collection at baseline, coupled with long-term follow-up for clinical outcomes, can provide unique insights into complex behavioral, environmental and genetic interactions causing subsequent disease or health. A key strength of prospective cohorts is that behavioral and environmental exposures are measured before the onset of disease, thus avoiding selection bias and problems of reverse causality. SG100K is a longitudinal population study of >100,000 Singaporeans comprising comprehensive phenotypic characterization and high-quality biological samples. The cohort was established by harmonizing existing prospective cohorts (n = 50,000) with additional prospective recruitment of ~50,000 new participants. These new participants have expanded phenotypic characterization in key cardiovascular and metabolic domains through deep phenotyping (including nutrition, diet and advanced imaging tests) and enhanced biological sample collection to enable genomic, transcriptomic and microbiome studies. To ensure substantial numbers of participants and incident events from different genetic ancestries, SG100K is recruiting men and women aged over 21 years, with sociodemographic features representative of the broader Singapore population and purposive sampling of South Asian and Southeast Asian individuals.

https://doi.org/10.1038/s41588-022-01274-x

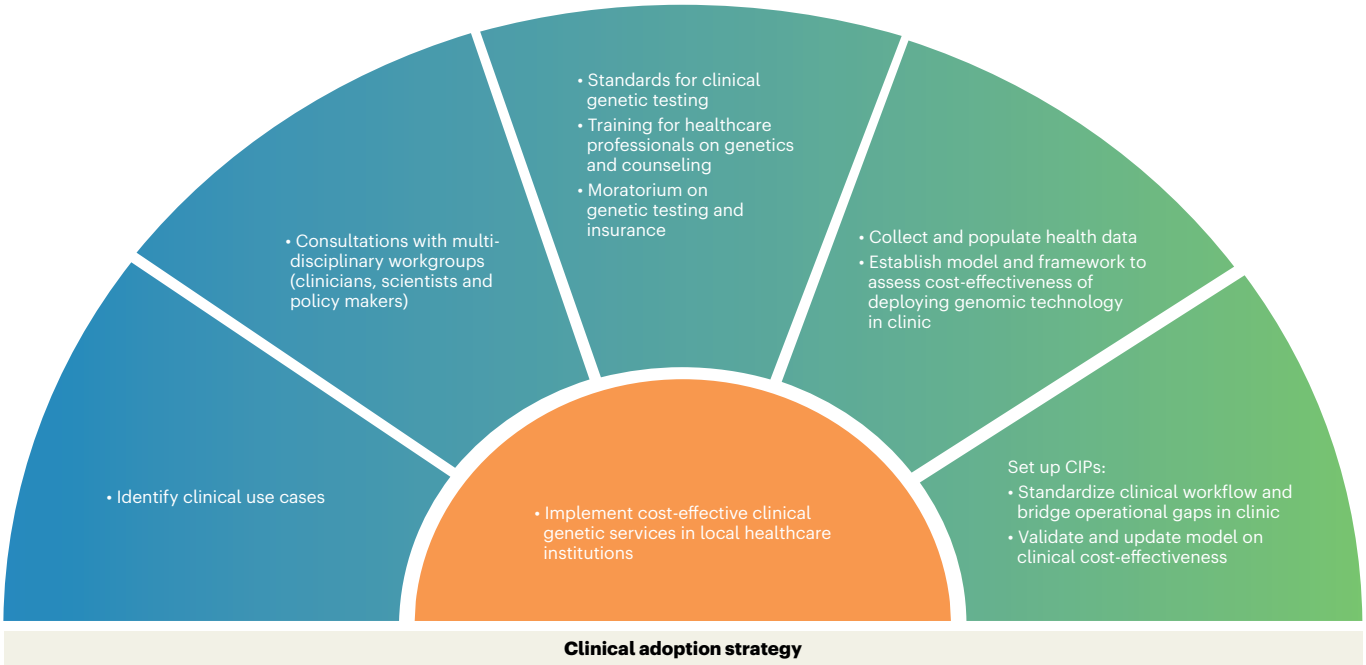

Fig. 2 | Singapore NPM clinical adoption strategy. A systematic landscape survey of barriers and potential strategies for clinical adoption of precision medicine in Singapore.

Table 4 | NPM Phase I RUS

RUS to assess feasibility of precision medicine
Aims I. Assess resource utilization with or without precision medicine intervention II. Determine minimum datasets required for health economic analyses III. Identify and address barriers to precision medicine implementation
Clinical use cases I. Pediatric genetic disease II. Familial hypercholesterolemia III. Hereditary breast and ovarian cancers
Outcomes • All three studies reported that using genomic testing in their identified clinical pathways is likely to be beneficial • The pilot studies also showed that it is feasible to have clinicians agree on a clinical pathway and generate an economic model that could be used to inform healthcare decisions
Modifiable factors that have a large impact on the cost-effectiveness of the pathway • A more efficient process that increases the proportion of first-degree relatives coming forward for testing after identifying an index case • Improve adoption and adherence to therapeutic and screening guidelines in patients with positive tests

In Phase II, PRECISE is partnering with SG100K to sequence the SG100K participants, jointly creating a PRECISE-SG100K resource of comprehensive genomic and research phenotypes. In parallel, PRECISE is also collaborating with the MOH Chief Health Scientist Office to establish national-scale infrastructure linking consented participants to their longitudinal EHRs through the Trusted Research/Real World-Data Utilisation and Sharing Tech (TRUST) platform (https://trustplatform.sg/). Integration of PRECISE-SG100K baseline data with EHRs will allow identification of individuals who develop disease, alongside their laboratory measurements, diagnosis codes and drug prescriptions. PRECISE-SG100K thus represents a unique platform for state-of-the-art epidemiological, translational and precision medicine research in the genetically distinct and culturally diverse Singapore population that may also contribute to chronic disease management at a regional and international scale.

Clinical adoption via CIPs

In Phase II, informed by design choices and learning lessons from Phase I RUS, a competitive grant call for prospective CIPs was performed. Clinician teams, with health economists as co-investigators, submitted proposals suggesting suitable clinical use cases, genomic tests to conduct, and treatments to prescribe based on test results. CIP applicants were also asked to propose economic models assessing the adoption of the proposed new clinical pathways whereby studies could be conducted to improve their health economics. This exercise demonstrated that, through a focused request for proposals, clinicians from diverse healthcare institutions could reach consensus, agree on downstream clinical pathways and generate economic models to inform healthcare decisions. To date, PRECISE has commissioned five CIPs, evaluating breast cancer polygenic risk scores, familial hypercholesterolemia, hereditary cancer, pharmacogenomics and primary glomerular kidney disease.

In a 'competition–collaboration' model, clinicians and health economists of separate shortlisted CIP applications converged to identify commonalities across the selected CIPs in which joint development of infrastructure or shared services would enable further economies of scale and reduce variation in implementing precision medicine across use cases. Economic models were further refined through discussions with stakeholders. For influential model parameters where data are either unavailable or deemed unsuitable for the Singapore context, this information will be collected either through the CIPs or from additional data sources facilitated by NPM. Clinical teams in the CIPs will also work closely with PRECISE-SG100K investigators to actively reclassify variants of unknown significance in Asian populations, thereby integrating NPM's research and clinical activities. We foresee that the CIPs will be pivotal to underpin informed healthcare decisions about which clinical pathways to adopt into Singapore's mainstream clinical practice.

NPM as part of the global precision medicine community

As a small country, Singapore's NPM will not operate in isolation but will consciously advocate responsible sharing of genomic and health-related data through a three-pillar strategy. First, NPM will

contribute to the global precision medicine community through international collaborations and participation in major international consortiums such as the Global Alliance for Genomics and Health (GA4GH), the Global Genomic Medicine Collaborative (G2MC), the International HundredK+ Cohorts Consortium (https://www.ihccglobal.org) and the International Common Disease Alliance. Second, NPM will seek to strategically partner with regional genomic initiatives, such as the Tohoku Medical Megabank Project (https://www.megabank.tohoku.ac.jp/english) and the Genomics Thailand Initiative (https://www.genomicsthailand.com). Notably, NPM is already working with the Thailand Health Intervention and Technology Assessment Program to generate frameworks for health economic analysis. Third, NPM recognizes the importance of data reusability to further add value to the data. The current lack of interoperable healthcare data presents substantial challenges with a plethora of diverse medical standards widening interoperability gaps. To counter this, NPM will adopt international standards such as OMOP (Observational Medical Outcomes Partnership) Common Data Models FAIR (findability, accessibility, interoperability, reusability) principles, and participate in GA4GH workstreams involving genomic and phenotypic data.

Conclusion

Translating an individual's genetic profile, in conjunction with traditional forms of health information, will enable clinicians to deliver more accurate diagnoses, target appropriate interventions and reduce potential incompatibilities of medication. However, the current lack of diversity in genetic databases, particularly for Asian populations, poses challenges for realizing precision medicine for the 4.6 billion people in the Asia-Pacific region. Hence, defining Asian genomic diversity through initiatives such as Singapore's NPM and other Asian-focused initiatives is vital for advancing healthcare globally. Beyond research, implementing precision medicine in routine clinical workflows will require investment and infrastructure development, with ongoing efforts to maintain public trust. Partnerships will be vital in creating new opportunities, whether in translating research, driving industry innovation or creating high-value jobs in the healthcare and biotechnology sectors. Finally, forging a common ethical and legal framework with dedicated cultural discussions with community, religious and social leaders in a sensitive and respectful manner is necessary to enable the potential of precision medicine in healthcare transformation.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01274-x.

References

1. Shapiro, M. D., Tavori, H. & Fazio, S. PCSK9: from basic science discoveries to clinical trials. *Circ. Res.* **122**, 1420–1438 (2018).
2. Sahin, U., Karikó, K. & Türeci, Ö. mRNA-based therapeutics—developing a new class of drugs. *Nat. Rev. Drug Discov.* **13**, 759–780 (2014).
3. *Global Spending on Health: a World in Transition* (WHO, 2019).
4. *Global Health Estimates 2019: Life Expectancy, 2000–2019* (WHO, 2020).
5. Schroeder, S. A. Shattuck Lecture. We can do better—improving the health of the American people. *N. Engl. J. Med.* **357**, 1221–1228 (2007).
6. Roden, D. M. et al. Pharmacogenomics. *Lancet* **394**, 521–532 (2019).
7. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
8. Olstad, D. L. & McIntyre, L. Reconceptualising precision public health. *BMJ Open* **9**, e030279 (2019).
9. Middleton, P. G. et al. Elexacaftor–tezacaftor–ivacaftor for cystic fibrosis with a single Phe508del allele. *N. Engl. J. Med.* **381**, 1809–1819 (2019).
10. Cully, M. Target validation: genetic information adds supporting weight. *Nat. Rev. Drug Discov.* **14**, 525 (2015).
11. King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* **15**, e1008489 (2019).
12. Ochoa D. et al. Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. *Nat. Rev. Drug Discov.* **21**, 551 (2022).
13. Stark, Z. et al. Integrating genomics into healthcare: a global responsibility. *Am. J. Hum. Genet.* **104**, 13–20 (2019).
14. The All of Us Research Program Investigators et al. The 'All of Us' Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).
15. Turnbull, C. et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* **361**, k1687 (2018).
16. Stark, Z. et al. Australian genomics: a federated model for integrating genomics into healthcare. *Am. J. Hum. Genet.* **105**, 7–14 (2019).
17. Cohn, E. G., Henderson, G. E. & Appelbaum, P. S. Distributive justice, diversity, and inclusion in precision medicine: what will success look like? *Genet. Med.* **19**, 157–159 (2017).
18. Caffrey, M. Disparities in cancer care: has precision medicine widened the gap? *AJMC* (22 April 2021).
19. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).
20. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
21. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
22. Manrai, A. K. et al. Genetic misdiagnoses and the potential for health disparities. *N. Engl. J. Med.* **375**, 655–665 (2016).
23. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
24. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
25. Privé, F. et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am. J. Hum. Genet.* **109**, 12–23 (2022).
26. Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
27. Atkinson, E. G. et al. Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* **53**, 195–204 (2021).
28. Mahajan, A. et al. Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet.* **54**, 560–572 (2022).
29. Chambers, J. C. et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case–control study. *Lancet Diabetes Endocrinol.* **3**, 526–534 (2015).
30. Chandalia, M. et al. Insulin resistance and body fat distribution in South Asian men compared to Caucasian men. *PLoS ONE* **2**, e812 (2007).
31. Wu, D. et al. Genetic admixture in the culturally unique Peranakan Chinese population in Southeast Asia. *Mol. Biol. Evol.* **38**, 4463–4474 (2021).
32. *Population Trends, 2021* (Singapore Department of Statistics, 2021).

33. *Government Health Expenditure and Healthcare Financing* (Singapore Ministry of Health, 2018).
34. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1789–1858 (2018).
35. Huang, K. K. et al. Genomic and epigenomic profiling of high-risk intestinal metaplasia reveals molecular determinants of progression to gastric cancer. *Cancer Cell* **33**, 137–150 (2018).
36. Metspalu, A., Köhler, F., Laschinski, G., Ganten, D. & Roots, I. The Estonian Genome Project in the context of European genome research. *Dtsch. Med. Wochenschr.* **129**, S25–S28 (2004).
37. Chen, J. et al. Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* **52**, 177–186 (2020).
38. Bylstra, Y. et al. Implementation of genomics in medical practice to deliver precision medicine for an Asian population. *NPJ Genom. Med.* **4**, 12 (2019).
39. McGuire, A. L. et al. The road ahead in genetics and genomics. *Nat. Rev. Genet.* **21**, 581–596 (2020).
40. *Life Insurance Association (LIA) Moratorium on Genetic Testing and Insurance* (Singapore Ministry of Health, 2021).
41. Lysaght, T. et al. Trust and trade-offs in sharing data for precision medicine: a national survey of Singapore. *J. Pers. Med.* **11**, 921 (2021).
42. Lysaght, T. et al. 'Who is watching the watchdog?': ethical perspectives of sharing health-related data for precision medicine in Singapore. *BMC Med. Ethics* **21**, 118 (2020).
43. Ong, S., Ling, J., Ballantyne, A., Lysaght, T. & Xafis, V. Perceptions of 'precision' and 'personalised' medicine in Singapore and associated ethical issues. *Asian Bioeth. Rev.* **13**, 179–194 (2021).
44. Majithia, S. et al. Cohort profile: the Singapore Epidemiology of Eye Diseases study (SEED). *Int. J. Epidemiol.* **50**, 41–52 (2021).
45. Tan, K. et al. Cohort profile: the Singapore Multi-Ethnic Cohort (MEC) study. *Int. J. Epidemiol.* **47**, 699–699j (2018).
46. Soh, S. et al. Cohort profile: Growing Up in Singapore Towards healthy Outcomes (GUSTO) birth cohort study. *Int. J. Epidemiol.* **43**, 1401–1409 (2014).
47. Yap, J. et al. Harnessing technology and molecular analysis to understand the development of cardiovascular diseases in Asia: a prospective cohort study (SingHEART). *BMC Cardiovasc. Disord.* **19**, 259 (2019).
48. Van der Auwera G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* 1st edn (O'Reilly Media, 2020).
49. Poplin, R. et al. Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at *bioRxiv* https://doi.org/10.1101/201178 (2018).
50. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
51. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
52. *Code of Practice for Key Office Holders Under the Healthcare Services Act* (Singapore Ministry of Health, 2019).
53. Doyle, D. L. et al. 2013 review and update of the genetic counseling practice based competencies by a task force of the accreditation council for genetic counseling. *J. Genet. Couns.* **25**, 868–879 (2016).
54. Christenhusz, G. M., Devriendt, K. & Dierickx, K. To tell or not to tell? A systematic review of ethical reflections on incidental findings arising in genetics contexts. *Eur. J. Hum. Genet.* **21**, 248–255 (2013).
55. Green, R. C. et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* **15**, 565–574 (2013).
56. eMERGE Clinical Annotation Working Group. Frequency of genomic secondary findings among 21,915 eMERGE network participants. *Genet. Med.* **22**, 1470–1477 (2020).
57. Kuo, C. W. et al. Frequency and spectrum of actionable pathogenic secondary findings in Taiwanese exomes. *Mol. Genet. Genomic Med.* **8**, e1455 (2020).
58. Chen, W. et al. Secondary findings in 421 whole exome-sequenced Chinese children. *Hum. Genomics* **12**, 42 (2018).
59. Van Hout, C. V. et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).
60. Cohen, J. T., Goodell, S. & Neumann, P. J. The cost savings and cost-effectiveness of clinical preventive care. In *Synthesis Project* (Robert Wood Johnson Foundation, 2009).
61. Hatz, M. H., Schremser, K. & Rogowski, W. H. Is individualized medicine more cost-effective? A systematic review. *Pharmacoeconomics* **32**, 443–455 (2014).
62. Kasztura, M., Richard, A., Bempong, N. E., Loncar, D. & Flahault, A. Cost-effectiveness of precision medicine: a scoping review. *Int. J. Public Health* **64**, 1261–1271 (2019).

[1]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [2]Centre for Genomics and Personalised Health, Genomics Research Centre, School of Biomedical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia. [3]SingHealth Duke-NUS Institute of Precision Medicine, Singapore, Singapore. [4]SingHealth Duke-NUS Genomic Medicine Centre, Singapore, Singapore. [5]Cancer & Stem Cell Biology Program, Duke-NUS Medical School, Singapore, Singapore. [6]Integrated Health Information Systems, Singapore, Singapore. [7]Ministry of Health, Singapore, Singapore. [8]Science Centre Singapore, Singapore, Singapore. [9]Sentosa Development Corporation, Singapore, Singapore. [10]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. [11]Department of Psychosis, Institute of Mental Health, Singapore, Singapore. [12]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. [13]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. [14]Department of Cardiology, National Heart Centre Singapore, Singapore, Singapore. [15]Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research, Singapore, Singapore. [16]Liggins Institute, University of Auckland, Auckland, New Zealand. [17]Personalised Medicine Service, Tan Tock Seng Hospital, Singapore, Singapore. [18]National Supercomputing Centre, Singapore, Singapore. [19]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [20]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore. [21]Centre for Innovation & Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [22]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [23]Cardiovascular and Metabolic Disorders Program, Duke-NUS Medical School, Singapore, Singapore. [24]Bioinformatics Institute, Agency for Science, Technology and Research, Singapore, Singapore. [25]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [26]Department of Biological Sciences, National University of Singapore, Singapore, Singapore. [27]Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore, Singapore. [28]Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore, Singapore. [29]Precision Health Research, Singapore, Singapore. [30]Biomedical Sciences Industry Partnership Office, Singapore, Singapore. [31]Chief Health Scientist's Office, Ministry of Health, Singapore, Singapore. [32]Department of Epidemiology and Biostatistics, School of Public Health Faculty of Medicine, Imperial College London, London, UK. [33]Duke-NUS Medical School, Singapore, Singapore. [34]Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: john.chambers@ntu.edu.sg; mdctes@nus.edu.sg; patrick.tan@precise.cris.sg

SG10K_Health Consortium

Rob M. Van Dam[10,35,36], Yik Ying Teo[10], Marie Loh[12,32,37], Paul Eillot[32], Eng Sing Lee[12,38], Joanne Ngeow[12,39], Elio Riboli[32], Rinkoo Dalan[40], Irfahan Kassam[12], Lakshmi Narayanan Lakshmanan[12], Tock Han Lim[41], Hong Kiat Ng[12], Theresia Mina[12], Darwin Tay[12], Charumathi Sabanayagam[13,42], Yih Chung Tham[13,42], Tyler Rim[13,42], Tin Aung[13,42], Miao Ling Chee[13], Hengtong Li[13], Miao Li Chee[13], Khung Keong Yeo[14], Stuart Alexander Cook[14], Chee Jian Pua[14], Chengxi Yang[14], Yap Seng Chong[15,43], Johan Gunnar Eriksson[15,43], Kok Hian Tan[33,44], Fabian Yap[44], Chia Wei Lim[17], Pi Kuang Tsai[17], Wen Jie Chew[17], Wey Ching Sim[17], Li-xian Grace Toh[17], Clarabelle Bitong Lin[1], Yee Yen Sia[1], Tat Hung Koh[1], Wee Yang Meah[1], Joanna Hui Juan Tan[1], Justin Jeyakani[1], Jack Ow[1], Shimin Ang[1], Ashar J. Malik[24] & Dimitar Kenanov[24]

[35]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [36]Exercise and Nutrition Sciences, Milken Institute School of Public Health, the George Washington University, Washington, USA. [37]National Skin Centre, Singapore, Singapore. [38]National Healthcare Group, Singapore, Singapore. [39]National Cancer Centre Singapore, Singapore, Singapore. [40]Endocrinology, Tan Tock Seng Hospital, Singapore, Singapore. [41]Ophthalmology, Tan Tock Seng Hospital, Singapore, Singapore. [42]Ophthalmology & Visual Sciences Academic Clinical Program, Duke-NUS Medical School, Singapore, Singapore. [43]Department of Obstetrics & Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [44]KK Women's and Children's Hospital, Singapore, Singapore.

Methods

SG10K_Health project participant details

The SG10K_Health project comprises 10,323 healthy individuals of three major ethnic groups: Chinese (C), Indian (I), Malay (M) and others (O), from six prospective cohorts, including (1) the Singapore Multi-Ethnic Cohort (MEC, $n = 3,082$; C, 1,148; I, 950; M, 981; O, 3), (2) the Health for Life in Singapore Study (HELIOS, $n = 2,400$; C, 1,907; I, 210; M, 178; O, 105), (3) the Singapore Epidemiology of Eye Diseases cohort (SEED, $n = 1,534$; C, 602; I, 654; M, 278), (4) the Growing Up in Singapore Towards healthy Outcomes study (GUSTO, $n = 1,000$; C, 581; I,175; M, 243; O, 1), (5) the SingHealth Duke-NUS Institute of Precision Medicine (PRISM, $n = 1,350$; C, 1,181; I, 69; M, 71; O, 29) and (6) Tan Tock Seng Hospital (TTSH, $n = 957$; C, 517; I, 244; M, 196). Genomic DNA was extracted from whole blood or buffy coat except for the GUSTO cohort, in which DNA was extracted from cord blood. All participants have provided informed consent for research. IRB approvals from each participating cohort were obtained from their respective contributing organizations: the MEC cohort (National University of Singapore IRB, B-16-158), HELIOS (Nanyang Technological University IRB, 2016-11-030 and 2017-11-006-01), SEED (SingHealth Centralised Institutional Review Board, 2012/487/A, 2010/392/A and 2015/2279), GUSTO (SingHealth Centralised Institutional Review Board, 2018/2767 and National Health Group Domain Specific Review Board, D/2009/00021 and B/2014/00406), PRISM (SingHealth Centralised Institutional Review Board, 2013/605/C) and TTSH (National Health Group R&D Office, TTSH/2014-00040).

Whole-genome sequencing

Paired-end 151-bp WGS was performed on the Illumina HiSeq X with 15× and 30× sequencing depth. The paired-end reads were adaptor trimmed using Trimadap and mapped using BWA-MEM to the GRCh38 reference. Duplicated reads were discarded using Picard MarkDuplicates with GATK 'germline short variant per-sample calling' reference implementation defined parameters. Base quality scores were recalibrated using GATK BaseRecalibrator and/or ApplyBQSR on a per-library basis.

Joint variant calling, quality-control filters and annotation

Single-sample-level CRAM files were used to call individual SNPs and indels with GATK HaplotypeCaller in gVCF mode, using GATK 'germline short variant per-sample calling' reference implementation defined parameters and companion files (GATK resource bundle GRCh38). Single-sample gVCF files were indexed for cohort-wide and collection-wide joint variant-calling steps. Variants failing the VQSR filter were removed, sex was imputed based on the mean depth ratio of chrX/chr20 and chrY/chr20 of each sample, and samples with abnormal ploidy were also excluded. Subsequently, samples with call rate <95%, contamination rate >2% and error rate >1.5% were also excluded. Next, median average deviation was computed on the autosome only for the ratios insertion/deletion, transition/transversion and heterozygote/homozygote alternative, where samples with a deviation of more than 6× median average deviation were excluded to derive the final SG10K_Health dataset of 9,770 genomes. Finally, genotypes on chrX and chrY were corrected according to the imputed sex. Next, genotype with allele balance >0.8 or allele balance <0.2, read depth <5 or genotype quality <20 were excluded. The remaining variants were annotated using VEP95 in merged mode (GENCODE and RefSeq reference).

Characterization of ClinVar clinical relevance values

The ClinVar public archive reporting evidence-supported relationships between human variations and phenotypes (clinvar_20220328.vcf.gz) was downloaded and used to annotate the SG10K_Health variant catalog based on ClinVar interpretations. We grouped ClinVar's 'Uncertain_significance' and 'Conflicting_interpretations_of_pathogenicity' interpretations as 'VUS'. We further grouped ClinVar's 'Pathogenic/Likely pathogenic' and 'Likely Pathogenic' as 'Likely Pathogenic'. We retained ClinVar's interpretation of 'Likely benign', 'Pathogenic' and 'Benign', and all other ClinVar reported interpretations were grouped as 'Other'.

Full data descriptions, methods and results will be described in subsequent manuscripts.

The codes to perform all analyses in this study are available at https://github.com/c-BIG/sg10k-health.

Acknowledgements

We thank all investigators, staff members and study participants of the contributing cohorts and studies: (1) the HELIOS study at the Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore; (2) the GUSTO study jointly hosted by the National University Hospital, KK Women's and Children's Hospital, the National University of Singapore and the Singapore Institute for Clinical Sciences, the Agency for Science Technology and Research (A*STAR); (3) the SEED cohort at the Singapore Eye Research Institute; (4) the MEC, National University of Singapore; (5) the PRISM cohort; and (6) the TTSH Personalised Medicine Normal Controls cohort. We also thank the National Supercomputing Centre, Singapore (https://www.ncss.sg) for computation resources. The SG10K_Health project is funded by the Industry Alignment Fund (Pre-Positioning) (IAF-PP, H17/01/a0/007); the project made use of participating study cohorts supported by the following funding sources: (1) the HELIOS study by grants from a Strategic Initiative at Lee Kong Chian School of Medicine, the Singapore MOH under its Singapore Translational Research Investigator Award (NMRC/STaR/0028/2017) and the IAF-PP (H18/01/a0/016); (2) the GUSTO study by the Singapore National Research Foundation under its Translational and Clinical Research Flagship Program and administered by the Singapore MOH's National Medical Research Council Singapore (NMRC/TCR/004-NUS/2008, NMRC/TCR/012-NUHS/2014) with additional funding support available through the A*STAR and the IAF-PP (H17/01/a0/005); (3) the SEED study by NMRC/CIRG/1417/2015, NMRC/CIRG/1488/2018 and NMRC/OFLCG/004/2018; (4) the MEC by individual research and clinical scientist award schemes from the Singapore National Medical Research Council (including MOH-000271-00) and the Singapore Biomedical Research Council, the Singapore MOH, the National University of Singapore and the Singapore National University Health System; (5) the PRISM cohort study by NMRC/CG/M006/2017_NHCS, NMRC/STaR/0011/2012, NMRC/STaR/0026/2015, the Lee Foundation and the Tanoto Foundation; and (6) the TTSH cohort study by NMRC/CG12AUG2017 and CGAug16M012. This research is also supported by the National Research Foundation Singapore under its NPM program Phase II funding (MOH-000588) and administered by the Singapore MOH's National Medical Research Council.

Author contributions

Conceived and led the NPM program: P.T., E.S.T. and J.C.C. Cohort recruitment and sample collection: J.Lee, J.J.Y.S., T.Y.W., C.W.L.C., P.D.G., L.L.G., X.S., C.Y.Cheng, S.D., N.K., K.P.L., E.S.T. and J.C.C. Sample processing and data analysis: N.B., M.H., R.T.M., C.B., W.K.L., J.F.C., J.Liu, S.P., S.M.S., C.S.V., P.K. and R.S.M.G. Enabling Platform workgroup: P.T., C.Y.Chua, K.H.K.B. and T.W.T. Regulation and Ethics workgroup: P.M.L.T. and R.C. Clinical Adoption workgroup: K.M., I.C., D.L., S.V. and M.K. Public and Community Trust workgroup: T.M.L., C.H. and S.W.S. Industry Development workgroup: W.Y.C., K.E.T., J.Y., W.Z. and Y.K.S. Workforce Development workgroup: K.T.G. The SG10K_Health Consortium was involved in sample collection and processing and data analysis. The manuscript was co-written by E.W., P.T., E.S.T. and J.C.C.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence should be addressed to John C. Chambers, E. Shyong Tai or Patrick Tan.

Peer review information *Nature Genetics* thanks Mark Caulfield and Philip Wilcox for their contribution to the peer review of this work.