https://doi.org/10.1038/s41588-025-02153-x

Mapping genetic diversity with the GenomeIndia project

Chandrika Bhattacharyya, Krithika Subramanian, Bharathram Uppili, Nidhan K. Biswas, Shweta Ramdas, Karthik Bharadwaj Tallapaka, Prathima Arvind, Khader Valli Rupanagudi, Arindam Maitra, Tulasi Nagabandi, Tiyasha De, Kuldeep Singh, Praveen Sharma, Nanaocha Sharma, Sunil K. Raghav, Punit Prasad, E. V. Soniya, Abdul Jaleel, Shijulal Nelson Sathi, Madhvi Joshi, Chaitanya Joshi, Mayurika Lahiri, Santosh Dixit, L. S. Shashidhara, Nachimuthu Senthil Kumar, H. Lalhruaitluanga, Lal Nundanga, Venkataram Shivakumar, Ganesan Venkatasubramanian, Naren P. Rao, Mohd Ashraf Ganie, Imtiyaz Ahmad Wani, Ganganath Jha, Ashwin Dalal, Murali Dharan Bashyam, Pritish Kumar Varadwaj, Sanjeev BS, Yogesh Simmhan, Chirag Jain, Durai Sundar, Ishaan Gupta, Pankaj Yadav, Himanshu Sinha, Manikandan Narayanan, Karthik Raman, Raghu Padinjat, Radhakrishnan Sabarinathan, GenomeIndia Consortium, Yadati Narahari, Vijayalakshmi Ravindranath, Thangaraj Kumarasamy, Divya Tej Sowpati, Mohammed Faruq, Analabha Basu & Bratati Kahali

The rich ethnolinguistic and sociocultural differences that exist in India offers a unique opportunity to study human diversity. With the whole genomes of 10,000 healthy and unrelated Indians from 83 populations, the GenomeIndia project captures the genetic diversity of one of the highly underrepresented populations in the global genomics landscape.

India, the most populous nation in the world with 1.44 billion people from over 4,600 distinct endogamous groups, represents an invaluable source of ethnolinguistic, sociocultural and genetic diversity. Deciphering the genetic makeup of human populations helps in understanding their ancestry, evolutionary history, admixture patterns, disease susceptibility and drug response. To this end, world over, genomics consortia have been established. Nevertheless, the global genomic landscape is predominantly Euro-centric[1]; although some projects have documented worldwide genetic diversity by discovering DNA sequence variation in several human populations, for example the 1000 Genomes project[2], followed by others across the world, recent inclusions being at biobank scale[3–7]. Unfortunately, India has been severely underrepresented in these global studies. To address this lacuna, a few studies have been initiated to understand the human genetic diversity of Asia (GenomeAsia)[8] and India (IndiGen and LASI-DAD)[9,10].

India experienced several waves of migration[11,12] with evidence of population expansion during the Paleolithic and Neolithic periods. The demographic history of India is also unique, with exposure to different environments and widespread admixture, followed by formation of numerous endogamous groups[13]. Many of the contemporary Indian populations have originated from a few founding groups and have maintained distinct identities through centuries of endogamy. These endogamous groups[13,14], apart from sharing genetic similarities, possess unique variations, including distinct disease-causing mutations with amplified frequencies within specific groups.

Overview of the project

In light of the above, the GenomeIndia national consortium was constituted in late 2017, comprising 20 institutions, with the Centre for Brain Research (CBR) at the Indian Institute of Science in Bangalore as the coordinating center. The GenomeIndia project was launched in January 2020, with funding from the Department of Biotechnology, Ministry of Science and Technology, Government of India.

The GenomeIndia Project is the largest, as well as the most comprehensive and well-designed study of India's genetic diversity. Compared with previous studies, the sampling strategy of GenomeIndia is extensive, nuanced and balanced, with respect to ethnic, socio-cultural, geographic, biogeographic and linguistic diversity of India. An emphasis on isolated populations from all corners of the country provided a robust representation of India's genetic landscape. Blood samples and associated phenotype data were collected from over 20,000 individuals, representing 83 population groups. These comprised 30 tribal and 53 non-tribal populations spread across India (Fig. 1). The tribal populations tend to be smaller groups (<1 million), some of which are isolated and restricted to certain geographical regions. Our sampling strategy was engineered at balancing between sampling from many populations and comprehensive sampling from fewer populations, yet with appropriate representation of population diversity. With approximately 100 samples per group, we aimed to estimate relatively rare alleles (at least 1% minor allele frequencies, MAFs), which are important to understand complex diseases. We focused on genetic diversity by sampling unrelated individuals, limiting the inclusion of first- and second-degree relatives, and emphasized trios (parent–child pairs), thereby ensuring accurate estimation of allele frequencies across groups. In addition, samples are available in the established biobank along with blood biochemistry, and anthropological, and socio-economic data for each participant, setting it apart from earlier studies. The objective was to

create a compendium of genomic variations, with regard to common and rare variants to gain insights into the genetic diversity, draw pharmacogenetic inferences, and construct a reference haplotype panel to facilitate imputation and therefore genotype–phenotype association studies for the Indian population.

Enrolment of subjects and sample collection

The GenomeIndia consortium followed the Helsinki Declaration for research protocols, sample collection, and ethical conduct, which was adapted across all the sampling centers (Fig. 1) and approved by their respective institutional human ethics committees. The spread of the language families in India overlaps with geography, and language is also an established proxy for genetic diversity in the Indian population[12,13]. We thus ensured that the four large major language families of Indo-European, Dravidian, Austro-Asiatic and Tibeto-Burman are appropriately represented in our sampling (Fig. 1), while accounting for the size of these ethnolinguistic groups. Moreover, within a broad geographic region, we sampled populations that belong to the distinct bio-geographies. To ensure that we do not miss out on the large array of rare variants in each population group, we sequenced a median of 159 unrelated individuals from each non-tribal group and 75 from each tribal group. We also included 3–6 trios in each group, for haplotype phasing and imputation, inference of de novo variations, and to assess the quality of called variants.

A total of 20,459 self-declared healthy adults (≥18 years) without any diagnosed monogenic diseases or chromosomal abnormalities consented to participate in the study (Fig. 2a). Informed written consent, in English and/or their native language, was obtained from all participants. The consenting individuals underwent detailed anthropometric assessments, such as height, weight, hip circumference, waist circumference and blood pressure, after assigning 12-digit IDs as part of the de-identification process. Blood samples (10 ml each) were collected from these individuals belonging to 83 populations, which inhabit over 100 distinct geographical locations (Fig. 1). Aliquots of samples were subjected to complete blood counts, as well as biochemical investigations that encompass glucose measures, lipid profiles, and liver and kidney function tests (Fig. 2b). This repository of samples with associated biomedical information distinguishes GenomeIndia from other population-scale projects. An important feature of this project is thus the biobank of over 20,000 samples that are available for future research.

Genotyping and WGS

Initially, 13,242 samples were genotyped on genome-wide SNP array to identify the subset used for whole-genome sequencing (WGS), primarily by checking for relatedness within a specific population. To maximize diversity, in addition to the selected trios, individuals in a population group who are unrelated to each other beyond first cousins were chosen for WGS (Fig. 2a). In total, DNA samples from 10,074 individuals were subjected to WGS. The four primary sequencing centers (CBR, CSIR-CCMB, CSIR-IGIB and BRIC-NIBMG) and one satellite sequencing center (GBRC) performed WGS, following uniform protocols in benchmarking and analysis, to ensure consistency and reliability. Subsequent quality checks were performed at the four primary sequencing centers (Fig. 2c). Variant quality was benchmarked with Genome in a Bottle (GIAB) samples, showing recall above 0.97, precision above 0.99, and an F1 score above 0.98 for single nucleotide variants (SNVs) and insertions and deletions (indels) for all sequencing centers. Five DNA samples, comprising one trio and two unrelated samples from the same population group, were sequenced and analyzed in a pairwise manner between two institutes. This process was repeated cyclically and cross-validated across four primary sequencing centers and achieved precision and recall of 0.99 and 0.97, respectively. The high concordance of genotype calls examined across sequencing centers suggests negligible batch effects. As the sequencing was conducted in five different centers, to eliminate any remaining possible batch effects, despite the above precautionary measures, we used additional tests on individual gVCF files before they were considered for joint genotyping (Fig. 2c). Our current genetic variant call-set is derived from 9,772 individuals − 4,696 male participants and 5,076 female participants.

Data handling and processing

To fasttrack a project of this magnitude, we distributed uniform protocols for raw data quality control, alignment and variant calling across the four primary sequencing centers (Fig. 1). In a first for the nation, each of these centers processed approximately 2,500 samples (totaling over 1 petabyte of raw data per center) from fastq to single-sample gVCF. Overall, this process consumed upwards of 0.7 million CPU hours and 4.5 petabytes of storage at the centers. These data are housed at a dedicated server at the Indian Biological Data Centre (IBDC), Regional Centre for Biotechnology (RCB), Faridabad, India.

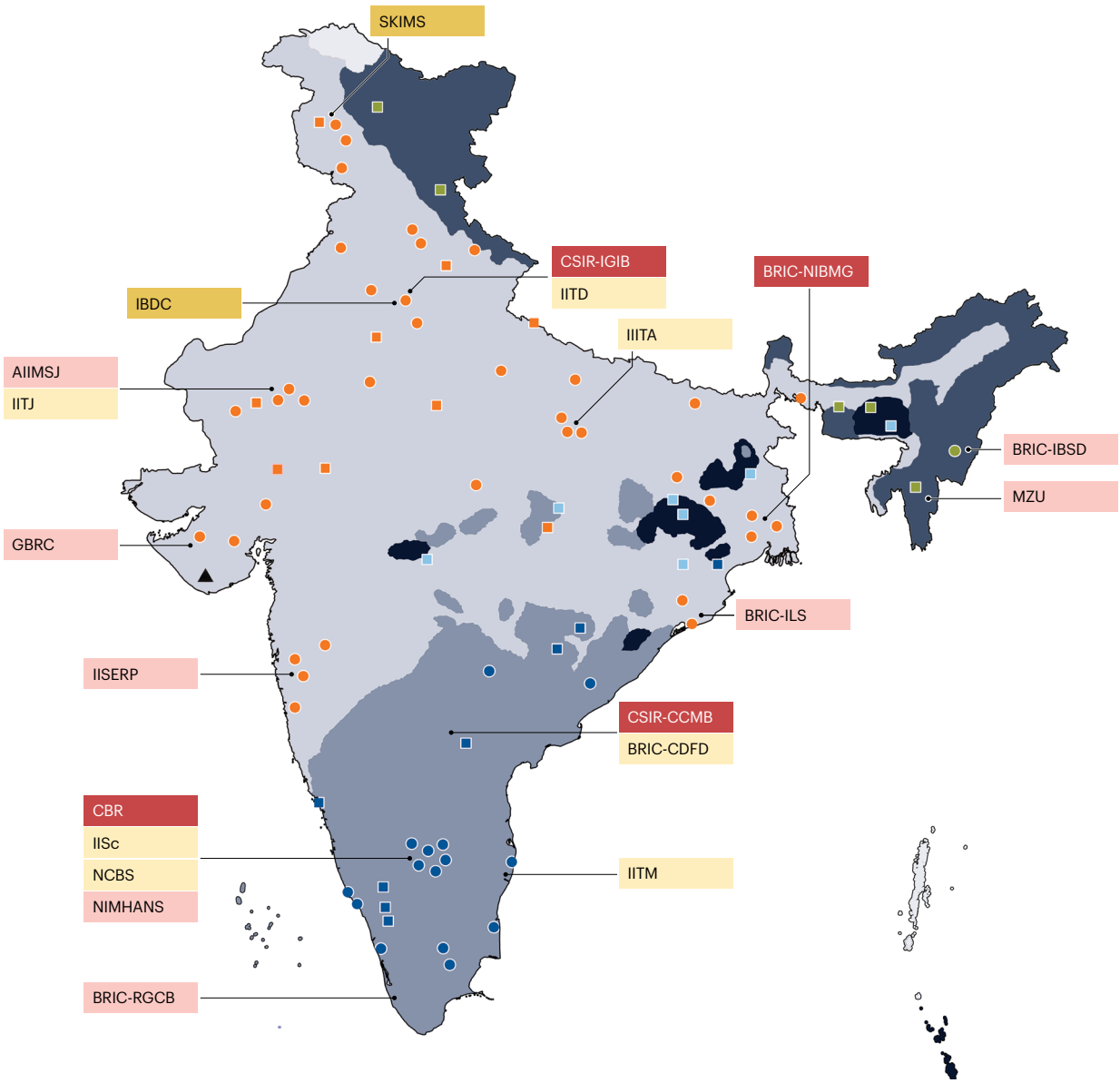
Preliminary findings after joint genotyping

The total number of raw genetic variants obtained were approximately 180 million, comprising biallelic SNVs and short indels, constituting biallelic singletons, doubletons and a small proportion of multi-allelic variants over autosomes, sex chromosomes and mitochondrial genome. After another round of quality checks for variants and genotypes at the sample level, after joint genotyping, 9,772 samples were retained for further consideration. This resulted in around 130 million autosomal variants. Excluding singletons and doubletons, as expected, most variants (65%) are ultra-rare, with a MAF of less than 0.1% in the overall population. Together, the identified genetic variants decode extensive genetic diversity that has been hitherto uncaptured in the Indian population.

Fig. 1 | Participating centers and populations sampled in the GenomeIndia project. The background colors of the map represent the distribution of the four major language families. Locations, language families and social groups associated with the samples collected are depicted on the map. Map source: https://surveyofindia.gov.in/pages/downloads, downloaded in December 2024. Twenty participating institutes across the nation, along with their responsibilities highlighted by a color code, are as follows. AIIMS-J, All India Institute of Medical Sciences, Jodhpur; BRIC-CDFD, Centre for DNA Fingerprinting and Diagnostics; BRIC-IBSD, Institute of Bioresources and Sustainable Development; BRIC-ILS, Institute of Life Sciences; BRIC-NIBMG, National Institute of Biomedical Genomics; BRIC-RGCB, Rajiv Gandhi Centre for Biotechnology; CBR, Centre for Brain Research (coordinating center); CSIR-CCMB, Centre for Cellular and Molecular Biology; CSIR-IGIB, Institute of Genomics and Integrative Biology; GBRC, Gujarat Biotechnology Research Centre (A Satellite Sequencing Centre); IISER-Pune, Indian Institute of Science Education and Research, Pune; IIITA, Indian Institute of Information Technology, Allahabad; IISc, Indian Institute of Science; IITD, Indian Institute of Technology, Delhi; IITJ, Indian Institute of Technology, Jodhpur; IITM, Indian Institute of Technology, Madras; MZU, Mizoram University; NCBS, National Centre for Biological Sciences; NIMHANS, National Institute of Mental Health and Neurosciences; SKIMS, Sher-i-Kashmir Institute of Medical Science.

The future ahead

Capacity-building and sharing digital public data. The successful execution of the GenomeIndia project has helped us to develop a robust ethical and legal framework to collect and govern the use of genomic data. A highlight of this project is the creation of digital public data, which will be available at IBDC and accessible for approved research in the foreseeable future. Similarly, the biobank containing over 20,000 samples will be a rich resource for multidisciplinary research. The project witnessed a synergy among 20 institutions with rigorous and uniform practices for consent, sampling, large-scale data generation and analysis. The capacity built in informatics infrastructure has provided a sustainable strategy that can form the basis for future genomics consortia in India, and contribute to South Asian specific genetic variations in global databases.



Spread of South-Asian language families

- Unclassified/language isolate
- Indo–European
- Dravidian
- Tibeto–Burman
- Austro–Asiatic

Description of the populations

- Tibeto–Burman tribe
- Indo–European tribe
- Dravidian tribe
- Austro–Asiatic tribe
- Continentially admixed outgroup
- Tibeto–Burman non-tribe
- Indo–European non-tribe
- Dravidian non-tribe

Centers and responsibilities

- Sample collection, sequencing and analysis
- Sample collection
- Method development
- Data storage

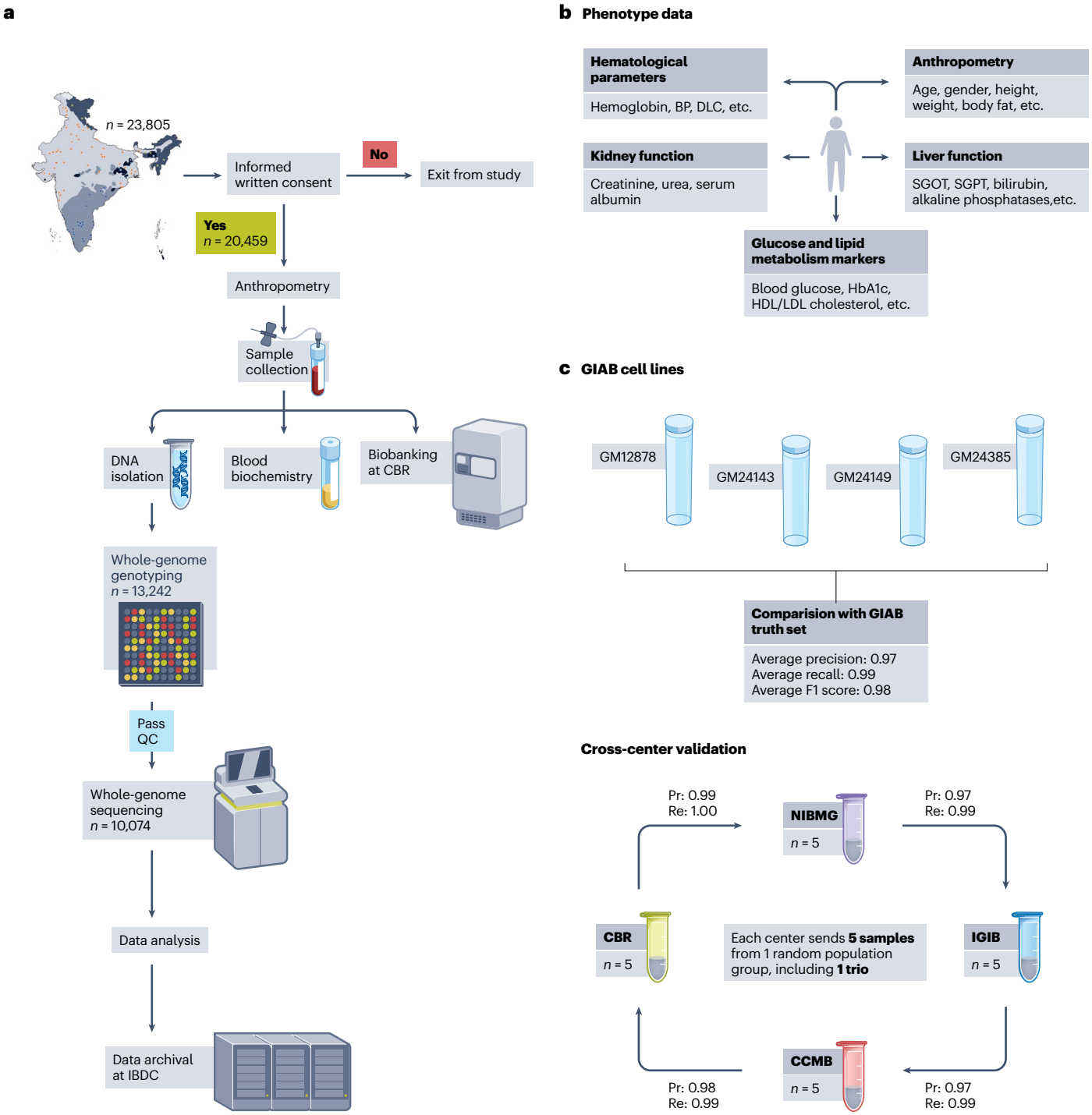


Fig. 2 | Sample processing workflow and quality checks. a, Volunteer recruitment, sample collection and processing workflow in the GenomeIndia project. **b**, Phenotype data collected from participants in the GenomeIndia project. Details included anthropometry, and various blood tests to study their hematological parameters, glucose and lipid metabolism markers, kidney and liver function indices. BP, blood pressure; DLC, differential leukocyte count; SGOT, serum glutamic oxaloacetic transaminase; SGPT, serum glutamic pyruvic transaminase. **c**, Quality checks (QCs) undertaken to ensure uniformity and reliability across the sequencing centers. Top, four (GIAB) cell lines were sequenced at each center and compared with the Genome in a Bottle (GIAB) truth set provided by National Institute of Standards and Technology (NIST). Bottom, five DNA samples (1 trio, 2 unrelated) from a randomly chosen population group were selected and sent to another sequencing center in a cyclical manner. The variants for each of these five samples were compared for concordance. Pr, precision; re, recall.

Identifying genetic diversity and uniqueness in Indian context. A direct outcome of the GenomeIndia data is decoding extensive genetic diversity from Indian populations that has been uncaptured until now. However, these data will also propel several ongoing and future compelling research questions and hypotheses. These questions include understanding the contribution of genetic variations to drug metabolism and modulation of drug responses, delving into patterns of natural selection that act on the genes and regulatory regions of the genome, as well as deciphering the unique regions of tandem repeats and transposable elements in the Indian population. The identified variants will enable us to create a larger set of benign innocuous variants that can be eliminated from candidate variant lists in clinical cases. Models that assess polygenic score (PGS) — an estimate of individual level genetic risk for susceptibility to a disease based on their genetic makeup — currently rely on allele frequencies of Euro-centric data and hence suffer from accuracy and portability issues in Indian population. The population-specific linkage disequilibrium structure unraveled by GenomeIndia data will enable training of better PGS models applicable to Indians, and perhaps South Asians in general. Overall, such research, which is being addressed by several working groups, will help to deepen our understanding of the genetic basis of different diseases. This will help to address health disparities that arise owing to a combination of sociocultural and environmental variations in various Indian population groups. Analysis towards these goals is ongoing, and we are actively preparing a detailed manuscript based on the results obtained so far and further insights anticipated in the near future.

Genome-wide arrays for South Asian ancestry. Our preliminary data indicate that imputing genotypes from Indian or South Asian ancestry, leveraging the reference haplotype panel constructed with rare and common variants obtained from GenomeIndia, will yield higher imputation power and accuracy, as well as allelic concordance compared with other widely used worldwide datasets, such as the HRC (Haplotype Reference Consortium) and TOPMed (Trans-Omics for Precision Medicine). This is owing to the benefits of ancestry-matched genomic linkage disequilibrium patterns, especially for rare and low-frequency variants in the Indian population. We are embarking on designing a genotype array for the Indian population, which we expect will notably improve on the recently developed SARGAM (South Asian Research Genotyping Array for Medicine)[15] array. Our focus is on the array content being tailored for individuals of Indian ancestry and will be centered around the discovery of common and rare genetic variants that is possible owing to the nuanced sampling in the GenomeIndia project. We foresee that this effort will successfully capture genetic variation across diverse population groups and facilitate future large-scale genetic association studies in the country. Importantly, many of these common variants are rare or non-existent in global variant databases, thereby highlighting the utility of this variant resource for clinical diagnostics and related applications in Indian and South Asian populations.

Concluding remarks

The GenomeIndia project is poised to provide valuable resources to advance medical genetics, enabling further clinical research. In-depth analysis of 9,772 diverse genomes along with the blood biochemistry and anthropometry data will improve disease diagnostics, predict the genetic basis of drug responses, and kickstart precision medicine efforts in India. This reference dataset could aid in the diagnostic process of rare diseases, potentially saving time and money, helping patients to circumvent an arduous road to healthcare. The findings from the project are also expected to inspire future functional studies that aim to determine the possible roles of population-specific genetic variants in various diseases and aiding in interventional and translational research in India.

Chandrika Bhattacharyya[1,24], Krithika Subramanian[2,3,24], Bharathram Uppili[4], Nidhan K. Biswas[1], Shweta Ramdas[2], Karthik Bharadwaj Tallapaka[5], Prathima Arvind[2], Khader Valli Rupanagudi[2], Arindam Maitra[1], Tulasi Nagabandi[5], Tiyasha De[4], Kuldeep Singh[6], Praveen Sharma[6], Nanaocha Sharma[7], Sunil K. Raghav[8], Punit Prasad[8], E. V. Soniya[9], Abdul Jaleel[9], Shijulal Nelson Sathi[9], Madhvi Joshi[10], Chaitanya Joshi[10], Mayurika Lahiri[11], Santosh Dixit[11], L. S. Shashidhara[11], Nachimuthu Senthil Kumar[12], H. Lalhruaitluanga[12], Lal Nundanga[12], Venkataram Shivakumar[13], Ganesan Venkatasubramanian[13], Naren P. Rao[13], Mohd Ashraf Ganie[14], Imtiyaz Ahmad Wani[14], Ganganath Jha[15], Ashwin Dalal[16], Murali Dharan Bashyam[16], Pritish Kumar Varadwaj[17], Sanjeev BS[17], Yogesh Simmhan[18], Chirag Jain[18], Durai Sundar[19], Ishaan Gupta[19], Pankaj Yadav[20], Himanshu Sinha[21], Manikandan Narayanan[21], Karthik Raman[21], Raghu Padinjat[22], Radhakrishnan Sabarinathan[22], GenomeIndia Consortium*, Yadati Narahari[2,18,25], Vijayalakshmi Ravindranath[2,25], Thangaraj Kumarasamy[5,25] ✉, Divya Tej Sowpati[5,25] ✉, Mohammed Faruq[4,25] ✉, Analabha Basu[1,25] ✉ & Bratati Kahali[2,25] ✉

[1]BRIC - National Institute of Biomedical Genomics (BRIC-NIBMG), Kolkata, India. [2]Centre for Brain Research (CBR), IISc Campus, Bengaluru, India. [3]Manipal Academy of Higher Education, Karnataka, India. [4]CSIR - Institute of Genomics & Integrative Biology (CSIR-IGIB), New Delhi, India. [5]CSIR - Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, India. [6]All India Institute of Medical Sciences (AIIMSJ), Jodhpur, India. [7]BRIC - Institute of Bioresources & Sustainable Development (BRIC-IBSD), Imphal, India. [8]BRIC - Institute of Life Sciences (BRIC-ILS), Bhubaneswar, India. [9]BRIC - Rajiv Gandhi Centre for Biotechnology (BRIC-RGCB), Thiruvananthapuram, India. [10]Gujarat Biotechnology Research Centre (GBRC), Gandhinagar, India. [11]Indian Institute of Science Education and Research (IISER), Pune, India. [12]Mizoram University (MZU), Aizawl, India. [13]National Institute of Mental Health & Neurosciences (NIMHANS), Bengaluru, India. [14]Sher-i-Kashmir Institute of Medical Sciences (SKIMS), Srinagar, India. [15]Vinoba Bhave University, Hazaribagh, India. [16]BRIC - Centre for DNA Fingerprinting and Diagnostics (BRIC-CDFD), Hyderabad, India. [17]Indian Institute of Information Technology (IIITA), Allahabad, India. [18]Indian Institute of Science (IISc), Bengaluru, India. [19]Indian Institute of Technology Delhi (IITD), New Delhi, India. [20]Indian Institute of Technology Jodhpur (IITJ), Jodhpur, India. [21]Indian Institute of Technology Madras (IITM), Chennai, India. [22]National Centre for Biological Sciences (NCBS), Bengaluru, India. [24]These authors contributed equally: Chandrika Bhattacharyya, Krithika Subramanian. [25]These authors jointly supervised this work: Yadati Narahari, Vijayalakshmi Ravindranath, Kumarasamy Thangaraj, Divya Tej Sowpati, Mohammed Faruq, Analabha Basu, Bratati Kahali.
*A list of authors and their affiliations appears at the end of the paper.
✉e-mail: thangs@ccmb.res.in; tej@ccmb.res.in; faruq.mohd@igib.in; ab1@nibmg.ac.in; bratati@cbr-iisc.ac.in

Published online: 08 April 2025

References

1. Popejoy, A. B. & Fullerton, S. M. *Nature* **538**, 161–164 (2016).
2. The 1000 Genomes Project Consortium. *Nature* **467**, 1061–1073 (2010).
3. Okada, Y. et al. *Nat. Commun.* **9**, 1631 (2018).
4. Choudhury, A. et al. *Nature* **586**, 741–748 (2020).
5. Halldorsson, B. V. et al. *Nature* **607**, 732–740 (2022).
6. Walters, R. G. et al. *Cell Genom.* **3**, 100361 (2023).
7. Bianchi, D. W. et al. *Nat. Med.* **30**, 330–333 (2024).
8. GenomeAsia100K Consortium. *Nature* **576**, 106–111 (2019).
9. Jain, A. et al. *Nucleic Acids Res.* **49**, D1225–D1232 (2021).
10. Kerdoncuff, E. et al. Preprint at *bioRxiv* https://doi.org/10.1101/2024.02.15.580575 (2024).
11. Thangaraj, K. et al. *Science* **308**, 996 (2005).
12. Kumar, S. et al. *BMC Evol. Biol.* **8**, 230 (2008).
13. Basu, A. et al. *Proc. Natl Acad. Sci. USA* **113**, 1594–1599 (2016).
14. Nakatsuka, N. et al. *Nat. Genet.* **49**, 1403–1407 (2017).
15. Wall, J. D. et al. *Nat. Commun.* **14**, 3377 (2023).

Acknowledgements

The GenomeIndia project was funded by the Department of Biotechnology, Ministry of Science and Technology, Government of India (BT/GenomeIndia/2018).

The consortium thanks all the participants who volunteered for this study. We extend our sincere gratitude to the dedicated social workers and government agencies at local and regional levels whose unwavering support has been instrumental in the successful collection of samples for the GenomeIndia project. We also thank the Illumina team of joint genotyping for help with the joint genotyper program. We thank the reviewers for their helpful suggestions for revising the article.

Competing interest

The authors have no conflict of interest to declare.

Additional information

Peer review information *Nature Genetics* thanks Mark Caulfield and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

GenomeIndia Consortium

Conceptualization and project writing

Analabha Basu[1,25], Vijayalakshmi Ravindranath[2,25], Bratati Kahali[2,25], Kumarasamy Thangaraj[5], Mitali Mukerji[4], Sridhar Sivasubbu[4], Vinod Scaria[4], Sunil K. Raghav[8] & Himanshu Sinha[21]

Project coordination and administration group

Suman K. Paine[1], Chandrika Bhattacharyya[1,24], Arindam Maitra[1], Nidhan K. Biswas[1], Analabha Basu[1,25], Vijayalakshmi Ravindranath[2,25], Bratati Kahali[2,25], Prathima Arvind[2], Yadati Narahari[2], Karthik Bharadwaj Tallapaka[5], Divya Tej Sowpati[5,25], Govindarajan Umapathy[5], Vinay K. Nandicoori[5], Rakesh Mishra[5], Kumarasamy Thangaraj[5], Mohammed Faruq[4,25], Sridhar Sivasubbu[4], Vinod Scaria[4], Kuldeep Singh[6], Praveen Sharma[6], Nanaocha Sharma[7], Dinabandhu Sahoo[7], Sunil K. Raghav[8], Ajay Parida[8], E. V. Soniya[9], Abdul Jaleel[9], M. Radhakrishna Pillai[9], Abitha Thomas[9], Madhvi Joshi[10], Chaitanya Joshi[10], L. S. Shashishara[11], Mayurika Lahiri[11], Nachimuthu Senthil Kumar[12], H. Lalhruaitluanga[12], Naren P. Rao[13], Ganesan Venkatasubramanian[13], Venkataram Shivakumar[13], Imtiyaz Ahmad Wani[14], Mohd Ashraf Ganie[14], Murali Dharan Bashyam[16], Ashwin Dalal[16], Pritish Kumar Varadwaj[17], Sanjeev BS[17], Y. Narahari[2,18], Yogesh Simmhan[18], Arun Kumar[18], Durai Sundar[19], B. Jayaram[19], Ishaan Gupta[19], Pankaj Yadav[20], Himanshu Sinha[21], Karthik Raman[21], Manikandan Narayanan[21], Padinjat Raghu[22] & Radhakrishnan Sabarinathan[22]

Sample collection group

Azad Ali[1], Mahabub Alam[1], Parveena Choudhury[1], Poulomi Ghosh[1], Sukanya Dhar[1], Saurav Roy[1], Nasrin Parvin[1], Rahul Modak[1], Sayan Bhowmick[1], Sourav Gangopadhyay[1], Devashish Tripathi[1], Analabha Basu[1,25], Chandrika Bhattacharyya[1,24], Suman K. Paine[1], Prathima Arvind[2], K. S. H. Shafeeq[2], G. Rajesh[2], C. Mohana[2], A. Divakar[2], Reddy P. Kommaddi[2], Neha Singh[5], Priya Pandey[5], Devavrat Desai[5], Mahfuj Hassan[5], Deepak Kumar Kashyap[5], Vasantha Kumar[5], Aman Kumar Suryan[5], Hema Sindhuja Rachiraju[5], A. Mahesh[5], Sushmita Nitta[5], Vijaya Mohan[5], Karthikeyan Meenakshisundaram[5], Jagamohan Chhatai[5], G. Mala[5], Karthik Bharadwaj Tallapaka[5], Kumarasamy Thangaraj[5], Sandeep Kumar Pal[4], Simmy Kaur[4], Mahino Fatima[4], Mohammed Akbar[4], Rahul C. Bhoyar[4], Pooja Sharma[4], Shreya Bari[4], Tiyasha De[4], Pratima Pandey[4], Anushree Mishra[4], Nishat Ashrafi[4], Syed Ahmad[4], Deepak Mudila[4], Sridhar Sivasubbu[4], Vinod Scaria[4], Mohammed Faruq[4,25], Arun Sree Parameswaran[6], Dolat Singh Shekhawat[6], Kuldeep Singh[6], Nayan Tada[6], Praveen Sharma[6], Tanuja Rajial[6], Varuna Vyas[6], Arvinda Thoudam[7], H. Moushmi Sharma[7], Khuraijam Dolly Devi[7], Nanaocha Sharma[7], Teresa Tangpua[7], Adyasha Mishra[8], Arup Ghosh[8], Deepak Jena[8], Punit Prasad[8], Soumendu Mahapatra[8], Sudarshana Jena[8], Sudeshna Datta[8], Sunil K. Raghav[8], E. V. Soniya[9], Abdul Jaleel[9], Shijulal Nelson Sathi[9], M. Radhakrishna Pillai[9], Abhitha Thomas[9], Udaya Lekshmi[9], R. A. Aswanth[9], Anjana S. Nair[9], Vasudev Paveri[9], T. S. Amal[9], Aman Tripathi[10], Bhagirath Dave[10], Bhumika Prajapati[10], Chaitanya Joshi[10], Madhvi Joshi[10], Ramesh Pandit[10], Sanman Samova[10], Ajay Malik[11], Kajal Gaikwad[11], L. S. Shashishara[11], Mayurika Lahiri[11], Santosh Dixit[11], Siddharth Gahlaut[11], Andrew Vanlallawma[12], H. Lalhruaitluanga[12], John Zohmingthanga[12], Lalawmpuii Pachuau[12], Lalchhandama Chhakchhuak[12], Lalnundanga[12], Nachimuthu Senthil Kumar[12], Ranjan Jyoti Sarma[12], Daddaladka Krishnayya Samartha[13], Ganesan Venkatasubramanian[13], Naren P. Rao[13], Paranthaman V. Kavya[13],

S. G. Tejaswini[13], Venkataram Shivakumar[13], Bashir Ahmad Charoo[14], Imtiyaz Ahmad Wani[14], Mahrukh Hameed Zargar[14], Mohd Ashraf Ganie[14] & Ganganath Jha[15]

Biobanking group

C. Mohana[2], G. Rajesh[2], A. Divakar[2], K. H. Rakesh[2] & Shobha Anilkumar[2]

Sequencing group

Mahabub Alam[1], Parveena Choudhury[1], Azad Ali[1], Poulomi Ghosh[1], Rahul Modak[1], Sukanya Dhar[1], Nasrin Parvin[1], Sayan Bhowmick[1], Sourav Roy[1], Shouvanik Sengupta[1], Chandrika Bhattacharyya[1,24], Indranil Bagchi[1], Subrata Patra[1], Suman K. Paine[1], Arindam Maitra[1], Khader Valli Rupanagudi[2], M. H. K. Mujawar[2], Vinayak Hosawad[2], Tulasi Nagabandi[5], Valli Undamatla[5], Neha Singh[5], Priya Pandey[5], Mahfuj Hassan[5], Vasantha Kumar[5], Devavrat Desai[5], Pratheusa Maccha[5], Sushmita Nitta[5], Aman Kumar Suryan[5], Hema Sindhuja Rachiraju[5], Karthik Bharadwaj Tallapaka[5], Pooja Sharma[4], Tiyasha De[4], Shreya Bari[4], Shahrumi Reza[4], Divya Goel[4], Rahul C. Bhoyar[4], Sandeep Kumar Pal[4], Bharathram Uppilli[4], Arushi Batra[4], Ashvarya Shankar[4], Gayatri Singh[4], Suman Mudila[4], Mahino Fatima[4], Divya Goel[4], Saima Iram[4], Mohamed Imran[4], Mohit Divakar[4], Vigneshwar Senthivel[4], Sridhar Sivasubbu[4], Vinod Scaria[4] & Mohammed Faruq[4,25]

Analysis working group

Chandrika Bhattacharyya[1,24], Devashish Tripathi[1], Vinay More[1], Arghya Dey[1], Shouvanik Sengupta[1], Haya Afreen[1], Mahabub Alam[1], Parveena Choudhury[1], Saurav Roy[1], Animesh Kumar Singh[1], Arnab Ghosh[1], Chitrarpita Das[1], Debashree Tagore[1], Subrata Das[1], Suman K. Paine[1], Nidhan K. Biswas[1], Analabha Basu[1,25], Krithika Subramanian[2], Shreya Chakraborty[2], Raghvendra Agrawal[2], Sauma Suvra Majumdar[2], Siddhi Jani[2], Akkshaya Rajesh[2], Debasrija Mondal[2], Anand Kumar[2], Debdutta Chatterjee[2], Priyanka Singh[2], A. Sohan Angelo[2], Tanmay Panigrahi[2], Eric Macwan[2], Rupanwita Majumder[2], S. Sagar[2], Samarpita Saha[2], Shweta Ramdas[2], Bratati Kahali[2,25], Payel Mukherjee[5], Pratheusa Maccha[5], Sreelekshmi MS[5], Jayesh Jain[5], Sofia Banu[5], Malini Nemalikanti[5], Sriram Sudarsanam[5], Divya Tej Sowpati[5,25], Shreya Bari[4], Bharathram Uppilli[4], Ankit Mukerji[4], Bani Jolly[4], Jupita Handique[4], Tiyasha De[4], Pooja Sharma[4], Sridhar Sivasubbu[4], Vinod Scaria[4] & Mohammed Faruq[4,25]

Joint genotyping group

Chandrika Bhattacharyya[1,24], Saurav Roy[1], Arnab Ghosh[1], Analabha Basu[1,25], Nidhan K. Biswas[1], Krithika Subramanian[2], Raghvendra Agrawal[2], Shreya Chakraborty[2], Siddhi Jani[2], Anand Kumar[2] & Bratati Kahali[2,25]

Data management group (data organization, data quality control, data storage, data archival, data security, data curation)

Chandrika Bhattacharyya[1,24], Saurav Roy[1], Shouvanik Sengupta[1], Animesh Kumar Singh[1], Arghya Dey[1], Arnab Ghosh[1], Azad Ali[1], Devashish Tripathi[1], Haya Afreen[1], Mahabub Alam[1], Parveena Choudhury[1], Vinay More[1], Suman K. Paine[1], Analabha Basu[1,25], Nidhan K. Biswas[1], Krithika Subramanian[2], Anand Kumar[2], Siddhi Jani[2], Shreya Chakraborty[2], Debasrija Mondal[2], V. Jothibasu[2], S. Karthik[2], Shweta Ramdas[2], Bratati Kahali[2,25], Sreelekshmi MS[5], Payel Mukherjee[5], Sriram Sudarsanam[5], Pratheusa Maccha[5], Jayesh Jain[5], Divya Tej Sowpati[5], Bharathram Uppilli[4], Shreya Bari[4], Pooja Sharma[4], Tiyasha De[4], Shahrumi Reza[4], Mohammed Faruq[4,25], Sanjay Deshpande[23], Deepak T. Nair[23] & Saurabh Raghuvanshi[23]

Method development group

Murali Dharan Bashyam[16], Ashwin Dalal[16], Asmita Gupta[16], Sumedha Avadhanula[16], Imlimaong Aier[17], Pritish Kumar Varadwaj[17], Rahul Semwal[17], B. S. Sanjeev[17], Ajeya Bhat[18], Arun Kumar[18], Chirag Jain[18], Nagakishore Jammula[18], Sai Manasa Chadalavada[18], Yogesh Simmhan[18], Nirmal Singh Mahar[19], Ishaan Gupta[19], Durai Sundar[19], Jyoti Sharma[20], Pankaj Yadav[20], Rajveer Singh Shekhawat[20], Soham Biswas[20], Ayam Gupta[21], Harshita Agarwal[21], Himanshu Sinha[21], Karthik Raman[21], Manikandan Narayanan[21], Venkatesh Kamaraj[21], Agastya Singh[22], Raghu Padinjat[22] & Radhakrishnan Sabarinathan[22]

[23]Indian Biological Data Centre, Haryana, India.