Title: Study Research Protocol for Phenome India-CSIR Health Cohort Knowledgebase (PI-CHeCK): A Prospective multi-modal follow-up study on a nationwide employee cohort

Running Title: ***Protocol paper for Phenome India project***

Authors: Phenome India Consortium

Correspondence: Dr Shantanu Sengupta, CSIR-Institute of Genomics and Integrative Biology

Abstract

The ability to predict the health trajectories of individuals based on their personalized risk scores can help formulate a preventive roadmap - of a disease or its complications. Currently, most of these risk prediction algorithms are based on epidemiological data from the Caucasian population and there is liberal evidence that they fail to work well for the Indian population due to ethnic diversity, varied dietary and lifestyle patterns, and altered risk profiles. In this multi-centric pan-India study, we aim to address these challenges and develop clinically relevant personalized risk prediction scores of cardio-metabolic diseases for the Indian population. This multi-centric program will involve the longitudinal collection and bio-banking of samples from ~10,000 CSIR employees, pensioners, and their spouses of which of the baseline sample collection is now completed. Multi-parametric data collected during baseline sampling includes a clinical questionnaire, lifestyle and dietary habits, anthropometric parameters, assessment for lung function, liver elastography, ECG, biochemical data, followed by molecular assays, including genomics, plasma proteomics, metabolomics, and fecal microbiome. In addition to mining the data for associations between the different parameters and their cardio-metabolic outcomes, we intend to develop models using artificial intelligence algorithms (AI) to predict phenotypic conditions. The study may be a step towards precision medicine for the Indian population, especially middle-income group strata, and aid in refining the normative values of healthy/disease parameters in the Indian population.

Keywords: Phenome, Cardio-metabolic, multi-omics, cohort, risk scores

Introduction

Non Communicable Diseases (NCDs) contribute significantly to human morbidity and mortality globally. Amongst NCDs, cardiovascular diseases (CVD) result in the highest number of deaths - about 17.9 million people each year, followed by cancer (9.3 million), respiratory disorders (4.1 million), and diabetes (2 million). NCDs, in total, account for 71% of the total 55 million deaths globally (1). Studies have shown that about 17 million NCD deaths occur in individuals younger than 70 years, of which 86% of early deaths are reported in low and middle-income countries (LMICs) (1). India accounts for 5.8 million of these, which is nearly 10 percent of the global mortality due to NCDs. Most importantly, the onset of CVD in this population is nearly a decade earlier (2, 3). In India, for all ages, the disability-adjusted life-years DALY rate per 100,000 population was 36,300 of which NCDs contributed 46·6% overall. This number increased to 55·0% in urban areas (4). Nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steato-hepatitis (NASH) add to increasing the risk of cardiovascular events and hence become important to be studied in the context of this framework. They have a global prevalence of 24% to 40% and are expected to rise to greater than 55% by 2040 (5-7). Liver and alcohol-related conditions now contribute to one of the top 15 NCDs for DALYs derived majorly from Years of Life Lost (YLL) (4). While NCDs result from genetic, physiological, environmental, and behavioral factors interplay, the risk factors contributing to NCDs could be an amalgam of all these. To reduce the exposure to risk factors, it is critical to prioritize specific preventive measures promoting overall health (8). There is a need for more extensive population-based studies of NCDs in the Indian population to understand the risk factors specific to the population in higher granularity than the guidelines from the Caucasian population that form the current basis for evidence-based medicine. India's wide variation of ethnicities, geographical heterogeneity, and age distribution contributes to nearly one-sixth of the world's population as it is the genetic melting pot for the world population (9). Moreover, Indian population forms a large part of the diaspora in other countries; hence, it is furthermore important to understand risk factors concerning the Indian population. Knowing risk factors specific to India for cardio-metabolic diseases may help formulate relevant

preventive, predictive, personalized, and participatory policies and guidelines for precision medicine.

Precision medicine allows the prediction of the health trajectories of individuals for early mediation and prevents the onset or complications of the disease. The development of risk scores has traditionally accomplished such prediction of health outcomes based on data from prospective cohort studies. The advent of multi-omics data and artificial intelligence-based big-data analytical tools has opened an unprecedented opportunity for developing novel personalized risk metrics to predict health outcomes (10, 11). A longitudinal population study, consisting of phenotypic measurements at baseline and long-term follow-up, can provide opportunities to understand the complex interrelationships between environmental, genetic, and other molecular factors with the risk of subsequent disease. It has the advantage of measuring parameters well before the disease onset, thereby eliminating selection bias and issues of reverse causality. Such epidemiological data is crucial for developing risk prediction tools and is vital in establishing precision health and medicine. Currently, most of these risk prediction algorithms are based on epidemiological data from the Caucasian population, and there is ample evidence that they are inaccurate for the Indian population due to ethnic diversity, varied dietary and lifestyle patterns, and altered risk profiles (12-14). More importantly, several studies currently available for the Indian populations are limited in the phenotypic information or diversity.

The Indian population is known to have a predisposition to a centrally obese phenotype that increases their risk of cardio-metabolic diseases (15, 16). Hence, a detailed phenotyping and understanding of these risks' genetic and metabolic underpinnings is essential. A longitudinal approach involving gathering phenotypic data from individuals across multiple time points ranging from physiological function, multi-omics data, assessment of cellular function, and microbiome diversity may aid in understanding the risks better. Integrating different OMIC data, including phenome, metabolome, proteome, genome, etc., is crucial to developing preventive protocols for various NCD risks. Phenomics can provide a strategy for discovering newer risk factors and

diagnostic biomarkers and unveiling novel therapeutic targets and precision therapies for realizing precision medicine (17).

With this objective of understanding risk factors of cardio-metabolic diseases, especially diabetes and NAFLD, a multi-centric pan-India study has been undertaken The current study, “Phenome India-CSIR Health Cohort Knowledgebase” (PI-CheCK) is initiated by the Council of Scientific and Industrial Research (CSIR) and spans its 37 constituent laboratories and associated centers across India, spanning 17 states and 2 union territories (UT) (18) (Figure 1).

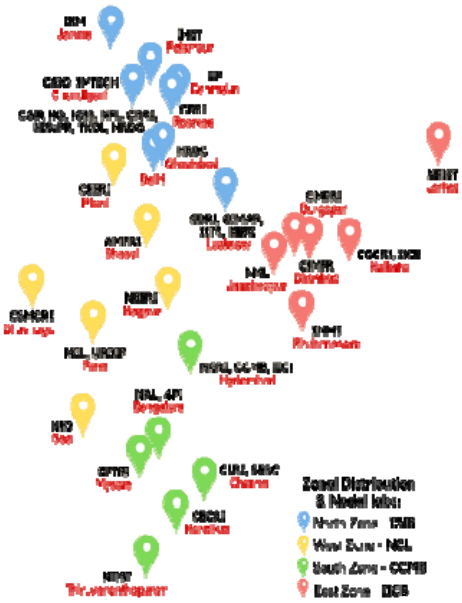

Figure 1: Sampling zones, sites in Phenome India Project, and coordinating labs. Zones are colored and defined for logistics purposes, not in any conventional manner.

On one side, CSIR encompasses various ethnicities, geosocial, habitats and occupational exposures with majorly urban and semi-urban distribution, thus bringing in variables that could play a role in the objectives being studied; on the other hand, CSIR is fairly homogenous from a perspective of socio-economic status, with most of its staff belonging to urban and semi-urban middle- and higher-middle-income groups. This planned cohort (~ N=10,000) comprises a captive population of well-literate working and

retired staff and their spouses. India has some significant cohort studies, albeit localized or limited by geographical entities, population heterogeneity, or the monitored parameters (18-22). Thus, to our knowledge, such a pan-country longitudinal cohort has not been previously reported. The planned study is a prospective collection (initailly over five years at three time points) of extensive health information, including but not limited to the physiological assessment of lung function through spirometry and oscillometry, cardiac function and heart rate variability (HRV) analyses by ECG, body composition analysis, liver function tests by transient elastography, dermatological assessment of skin moisture, sebum, hydration, anthropometry and molecular and biochemical assays of blood samples for analysis in addition to the creation of a sample and data repository for multiple omics and microbiome of an active Indian population.

The goal is to generate a first-of-its-kind Indian dataset on different health-related and biomedical parameters from a country-wide employee cohort, followed up for years together, which will be of significant public health importance with the possibility of developing diagnostic and prognostic biomarkers. An employee cohort favors a lower dropout rate, being educated and literate in realizing the study's potential and discussing clinically significant results with the study participants, and encouraging seeking appropriate medical attention also results in a healthy and productive scientific workforce for CSIR.

Study Objectives

The primary objectives of the PI-CHeCK cohort study are as follows:

a) To develop a prospective longitudinal cohort of consenting CSIR employees and their spouses, including pensioners/retired employees.

b) To develop clinically useful personalized risk prediction scores for cardio-metabolic disorders for pre-diabetes, diabetes, and fatty liver disease based on multi-omics data integration.

The long-term objectives for the listed disorders of the study are:

1. Develop risk prediction models for diabetes and liver fibrosis:

(i) Glycaemic control: Develop a binary prediction model for the normo-glycaemic population to develop to:

(a) Diabetes (prior HbA1c <5.7 and FBS < 100 to HbA1c >= 6.5 and Fasting blood sugar (FBS) >126 mg/dL on American Diabetes Association (ADA) criteria)

(b) Pre-Diabetes (with prior HbA1c <5.7 and FBS =<100 to HbA1c >=5.7 and <6.5 and FBS =>100 but less than <126 mg/dL on ADA criteria)

(ii) NAFLD/Liver Fibrosis: Develop a binary prediction model for F0/F1 (no fibrosis) to F2/F3/F4 (moderate to severe fibrosis/cirrhosis) on liver Fibroscan (Transient Elastography)

2. Create and maintain a database and linked bio-repository.

Methods

Design of the study

The study design is primarily longitudinal cohort-based with additional two-time points for the follow-up other than the present first phase of baseline collection which is now completed.

Sampling sites and sample size determination

All 37 CSIR labs and associated centers are categorized into four zones primarily to assist in better logistics management. Zones were defined based on the geographical locations of the labs; the current number of regular employees was taken as a baseline count from each CSIR lab. The baseline count was multiplied by four to accommodate spouses of current employees, an equal number of pensioners, and pensioners' spouses. Here, the assumptions made were a) the pensioners are of at least the same number as the employees of each lab, and b) employees and pensioners have spouses. Thus, an overall count for each lab was obtained and further weighted proportionately so that each zone had at least 2,500 samples. This was done to keep the follow-up studies' attrition rate below 20-30%. The information on the four zones and respective labs is given in Supplementary Table S1. Four zonal nodal Institutes (Indian Institute of Chemical Biology (CSIR-IICB), East zone; National Chemical Laboratory (CSIR-NCL), West zone; Institute of Genomics and Integrative Biology (CSIR-IGIB), North zone; Centre for Cellular and Molecular Biology (CSIR-CCMB), South zone) were designated to coordinate this activity. (Figure 1).

Inclusion/Exclusion criteria

All CSIR permanent staff, i.e., employees, pensioners, and their spouses, were eligible to be enrolled in this study. The exclusion criteria for this study were pregnancy and under 18. Further, the participants were excluded from a specific scanning modality if he/she gave a history of medical surgery or complication defined for that particular

modality, as mentioned in Supplementary Table S2. The participants would fast for a minimum of 10-12 hours before coming to the sampling site and phlebotomy.

Tests and scans planned

Three main categories of assays were planned*: blood-based, scanning-based*, and stool sample-based. A brief description is given in Table 1. A few modalities such as Spirometry, Oscillometry, TE- for Liver, and BCA may be done twice or more for a selective population at the same or different time points to understand multiple

Blood-based	
1. Biochemistry Analysis and Immunoassays	5. Genomics-Global Screening Array (GSA), Whole Genome Sequencing and Whole Exome sequencing
2. Proteomics and Metabolomics	6. Metal Content
3. Cytokine and Immuno-phenotyping	7. Telomere Length and Activity
4. RNA-Micro RNA profiling and RNA sequencing	
Scanning-based	
1. Transient Elastography (TE) for Liver-Ultrasound based	5. Anthropometry-Height, Skinfold Caliper, Dynamometer for grip strength, and anthropometry tape-based measurements of Chest, Waist, and Hip Circumference
2. Spirometry for Lung	6. ECG

physiological variable states or outcomes based on molecular assays. Consent was

3. Oscillometry for Lung	7. Skin Health Analysis
4. Body Composition Analysis (BCA)	
Stool Sample-based	
1. Gut microbiome analysis	

Table 1: Details of the different assays planned in Phenome India Project.

obtained initially before their enrollment.

Contraindications to the imaging/scanning/other tests being done

Specific contraindications for the tests are listed in Supplementary Table 2, along with specific query-based enrollment if the volunteering participants are doubtful of their medical history.

Cohort enrolment and sample collection

The first sample collection phase was planned for six months, from December 2023 to June 2024. In each lab, a week was spent sampling nearly 50 participants per day. The second and third sampling phases are planned to be carried out in 2025-26 and 2026-27

To ensure maximum participation, awareness about this drive of the program's importance and participatory nature was conducted at each of the constituent laboratories of CSIR through brochures, web-based and personal interactions. A signed physical consent form was obtained from the participants during the sample collection.

Registration & Survey Questionnaire, Slot Booking

Once a lab is made aware of the project, the participants were registered through a secure online portal-the CSIR Cohort portal (available at https://www.csircohort.org/) inspired by HIPAA guidelines complying with considerations such as two-factor authentication, backend data encryption, user timeouts, role-based access (administrator types), data integrity controls and logging privileges. The portal is equipped with various supporting documents, making the participants aware of the objectives, the scientific rationale of the protocols, and the interpretation of results. Using two-factor authentication, a participant may access his/her recorded health parameters and reports anytime from the portal. Various administrators were assigned to the defined roles based on the data modalities.

Participants must fill out a baseline questionnaire, including questions about their family and residence. This information was recorded for authentication purposes only. The backend data structure of the baseline questionnaire would be placed in a different table, which is not connected to any other data pipelines. The entire table will be encrypted and not accessible via the back end. The entries will be indexed with Registration IDs, which become the only identifying factor for a participant.

Once the respective coordinator approves, the participant can complete the questionnaire online. The participant's personal information was converted to the computer-generated registration ID once the coordinator approved the participant. Henceforth, the participant's identity is represented by this registration ID, which only the participant knows. The questionnaire is embedded in REDCap, which enables the participant to provide self-reported information in different fields. Details of the questionnaire are provided in Supplementary Table 3 and Table 4. The participant will only be eligible for slot booking when the questionnaire is submitted successfully. The slot booking procedure includes precautionary questions for the testing day. Participants can then book the slot date and time with eligible tests. The reports section is accessible via two-factor authentication, which fetches the reports with date and time.

Once the participant has filled out the questionnaire, the participant can book the slot for the sample collection. During this process, the participant must consent to blood collection, which is mandatory for further consideration when a participant agrees to participate. Once selected, the date and time of sample collection can be finalized by the participant. Registered participants from each laboratory will be informed of the dates for sample collection at their chosen lab, and they can then schedule the date and time of their visit through the online portal at their convenience. If a participant chooses to provide a stool sample, the kit can be obtained from the collection center 48 hours before the sampling date, and the participant can bring a stool sample on the sampling day. The reminders through email and mobile will be conveyed to the participants enabled through the portal in which the participant has registered.

Data Acquisition (Materials and Methods)

All the stations at all labs and centers had the same equipment except Oscillometry.

Stations for different scanning modalities are given in Supplementary Table 5 and briefly described as follows:

Anthropometry

Anthropometry will be carried out to measure height, weight, chest (CC), waist (WC), Abdominal (AC), and hip circumferences (HC).

Height (in cm) - As recommended by ICDS guidelines, a standard stadiometer was utilized at all centers.

Weight (in kg) - Weight was measured using Accuniq Body Composition Analyzer (Manufacturer- SELVAS Healthcare, Korea Model BC 380)

Body Circumferences (in cm) - CESCORF measuring tape (Manufacturer – Cescorf, Brazil) was utilized to measure CC, WC, AC, and HC.

The following definitions were adopted to measure circumferences thrice at each site (23, 24).

CC-"Measure the chest circumference at the most significant part of the chest, which is usually across the level of the nipple line in males and just above the breast tissue in females. Measurement should be taken at the end of a normal expiration".

WC- "A horizontal measure is taken at the midpoint between the lower margin of the last palpable rib and the top of the iliac crest".

AC- "The tape should be held behind the participant with one edge at the horizontal plane through the center of the umbilicus".

HC- "The participant is standing erect, and the feet close together. A horizontal measure is taken around the widest portion of the hips and buttocks".

Body Composition Analysis

Body Composition Analysis (BCA) was done on the Accuniq Body Composition Analyzer (manufacturer- SELVAS Healthcare, Korea Model BC 380) as per the standard protocol specified by the manufacturer.

Transient Elastography of Liver

The FS Mini 430 Plus model (Manufacture-Echosens, France) with M or XL probes was used to acquire the scan and data. Fibroscan is a simple and safe technique to assess liver stiffness and fat content and takes 5-10 minutes on average per examination. Preliminary preparation required the subject to remain fasting for 6 hours before the procedure, which was met through the fasting condition specified. The subject is then asked to lie supine with the right arm in full abduction. The examiner begins by placing the probe along the 5th to 7th intercostal space in the mid or anterior axillary line to obtain a view of the right lobe of the liver. Once the area is identified, measurements were obtained using the fibro scan probe.

Liver Stiffness Measurement (LSM) and Continuous Attenuation Parameter (CAP) measurements will be captured by trained personnel blinded to patients' clinical data. Although the literature points out the use of an XL probe for BMI greater than 30, here we utilized and worked with the recommender system of the device for probe selection. Failure of transient elastography is defined as the inability to obtain valid LSM or CAP value (25-27). Manufacturer guidelines were followed to acquire a minimum number of valid measurements, usually 10.

SPIROMETRY

EasyOne Air hand-held spirometer (Manufacturer-NDD, Switzerland), a widely utilized handheld spirometer based on ultrasound sensing, was used in the study to capture spirometry data. ATS/ERS guidelines and interpretation were followed to capture spirometry data (28-31). FEV1 and FVC are generally used to diagnose lung function abnormalities, with an FEV1/FVC ratio of <0.7 indicative of obstructive airway disease (OAD). If the ratio is normal, i.e.≥0.7, FEV1 and/or FVC of <80% predicted indicate low

lung function with restricted lung function measured through FVC and preserved ratio through FEV1 (32).

Oscillometry

Oscillometry was performed before spirometry using the Tremoflo C-100 Airwave Oscillometry System (Thorasys, Canada) or Resmon Pro V3 System. Oscillometry is used to measure perturbations to sound signals of different frequencies superimposed onto normal tidal breathing, which helps characterize the resistance of small airways and airway distensibility. The signal was recorded for at least three artifact-free breaths by Tremoflo or 45 seconds by Resmon Pro (automated in device itself), which sets the QC before capturing acceptable maneuvers. Resistance and Reactance at 5 Hz (R5 and X5), Resistance at 20 Hz or 19 Hz (R20/R19), Resistance of small airways, 5 to 20 Hz (R5-20 or R5-R19) and frequency response (Fres) would be monitored primarily

ECG

The study included ECG of around 10,000 samples at a 500 Hz sampling rate using Pisces 1012 model 12 lead ECG, (Manufacturer Allengers, India). The ECG uses a 12-lead setup with a high-pass frequency filter at 0.05 Hz, a low-pass filter at 35 Hz, and a 50 Hz notch filter. The sweep speed is 25 mm/s, and the amplitude is 1.00 cm/mV. This process ensures the accurate and efficient collection of ECG data. The ECG machine is connected to a laptop, and participant details are entered into the software. A two-minute recording was taken while the participant remained motionless. The data gets saved automatically, generating a PDF uploaded to the participant's dashboard. A hard copy of the ECG was provided to the participant. The station administrator uploads all data in PDF and Excel formats to the portal by the end of the day.

Skin analysis and phenotyping

To develop standards for objective assessment of skin barrier in Indian skin types, we measured skin barrier function in around 10,000 participants utilizing Courage Khazaka (Manufacturer Courage Khazaka Electronic GmBH, Germany) in terms of four

biophysical parameters, employing distinct probes for measurement: Skin sebum content (Sebumeter): The measurement process involves using grease spot photometry with the Sebumeter® SM 815. The matte tape contacts the skin or hair, becoming transparent based on surface sebum. A photocell measures transparency when the tape is inserted into the device aperture, indicating sebum content by light transmission. The Sebumeter measures sebum levels on the forehead and left cheek, with readings taken over 30 seconds at each site. Transepidermal Water Loss (Tewameter): The Tewameter® probe measures the density gradient of the water evaporation from the skin (Δc) proportional to the TEWL. The probe takes three 10-second measurements on the forehead, left cheek, and left volar forearm to assess TEWL. Stratum Corneum Hydration (Corneometer®): The measurement is based on capacitance measurement of a dielectric medium, here the stratum corneum, the uppermost layer of the skin. With increasing hydration, its dielectric properties change. The measurement is based on the fact that water has a higher dielectric constant than most other substances (mainly < 7) and is done on the forehead, left cheek, and left volar forearm. Skin Elasticity (Cutometer): The measuring principle of the Cutometer® is based on the suction method, where negative pressure deforms the skin mechanically. The device creates pressure, drawing the skin into the probe's aperture. After a defined time, it is released again.

Blood collection

Since several parameters were being investigated, we expected to collect about 40-50 samples/ day. The samples collected were transported to CSIR-IGIB in a controlled temperature environment. CSIR-IGIB will provide a sample code for each sample, and aliquots will be transported to the CSIR labs, which will be designated to perform the molecular analysis for that lab. The remaining aliquots will be stored (Bio-banking). A few of the routine biochemical analyses were outsourced to a private diagnostic laboratory that will use the same equipment for all the samples collected and meet the predefined quality control requirements of the consortium in all phases of sample collection. Other biochemical analyses will be performed at IGIB (Details in Supplement). The clinically relevant results will be uploaded to the central database by

CSIR-IGIB, and the participants will be provided passcode-enabled access to their results. The data will be stored in a data repository at IGIB with a backup at 4PI. The sample flow and processing work plan is shown in Figure 2.

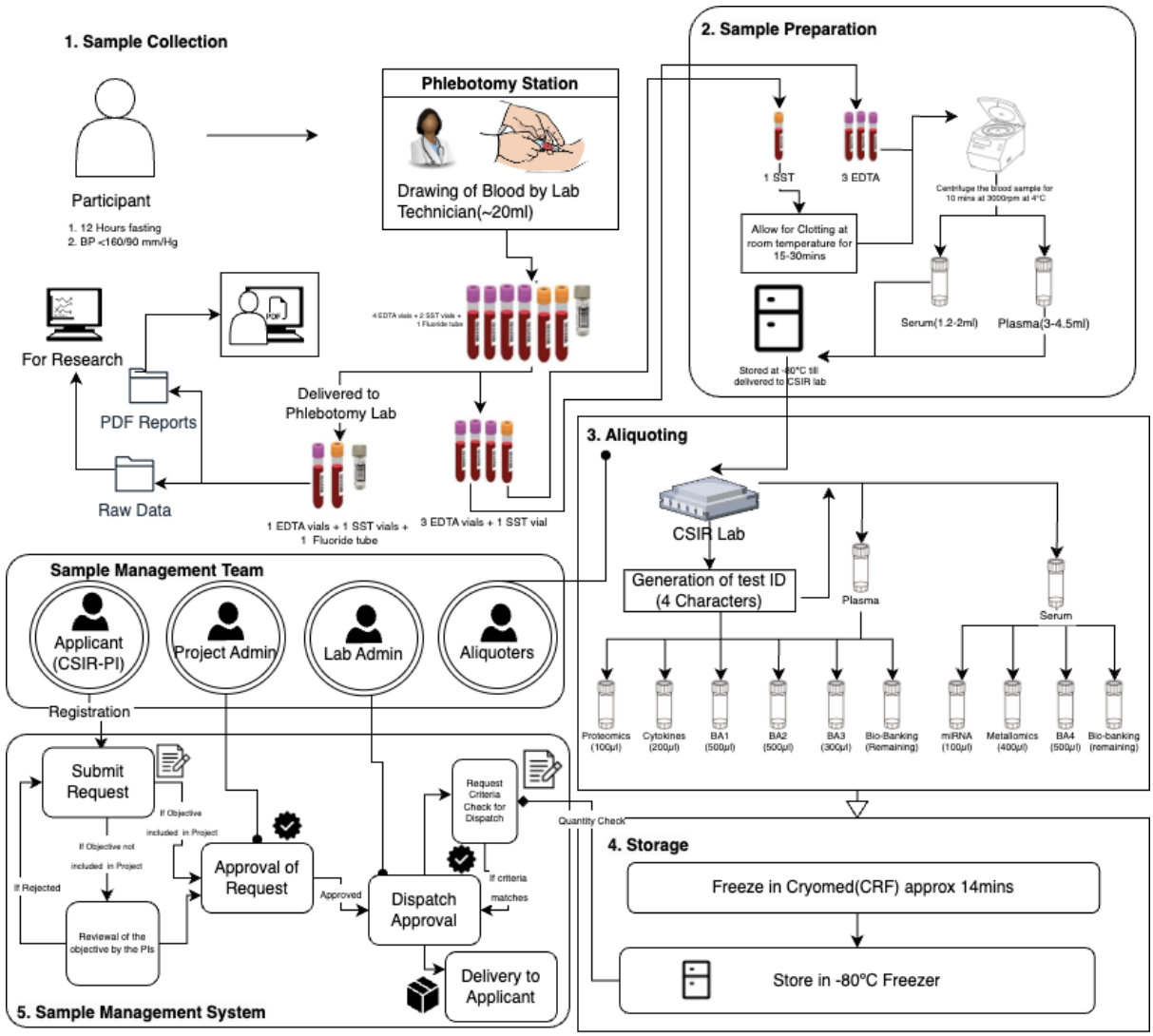

Figure 2: Sample flow and processing of Phenome India Project

Bio-Banking

Long-term Bio-banking of samples plays a crucial role in developing and maintaining longitudinal prospective cohorts like the CSIR-cohort, which involves the collection of more than 30,000 samples over five years, including baseline assessments and two follow-up evaluations. This will allow us to study common diseases over time, focusing on identifying risk factors and prevention strategies from the data acquired in large, diverse populations, understanding disease mechanisms, and developing targeted treatments. Not only does bio-banking hold immense potential to foster proactive public health strategies, but it is also crucial to address ethical, legal, and social implications to maximize the benefits of bio-banking while safeguarding participant rights and data integrity. Anonymization and de-identification are the key features that are employed in the maintenance of the biorepository to protect participant privacy, which is done through the assignment of a LAB-ID for each participant at the time of sample collection and is later converted to a TEST-ID for various assays based on the study and its objectives. This process ensures that the data remains linked to the correct samples while maintaining confidentiality. The unused samples will be stored in 0.5 ml 2D tubes, each marked with unique QR codes and tube numbers for precise traceability. This meticulous system supports robust and reliable research by accurately correlating samples with participants and health outcomes. Maintaining a standard bio-bank can expand global biobank networks and comprehensive data-sharing frameworks to enhance the scope and impact of research further, allowing for more representative studies across diverse populations.

Assays to be performed and planned

Biochemical and Molecular Analysis

The molecular studies, including Genomics, transcriptomics, proteomics, metabolomics, metallomics, immunological and microbiome studies, will be performed by different CSIR Labs. It is proposed that a Global Screening array (GSA 3.0, Illumina, USA) based SNP array will be used for genomic studies. Mass spectrometry-based MRM

assays will be performed for proteomic and metabolomics studies, and inductively coupled optical electron microscopy studies (ICP-OES) will be used for metallomics studies. This comprehensive profiling will be done in a subset of 50% of the participants who will be chosen based on randomization controlled for gender and geographic distribution (zonal). The list of biochemical assays that would be done in the samples is provided in Supplementary Table 6.

Targeted Quantitative Metabolomics and Proteomics

It will be done using MxP Quant 500 Kit (Biocrates), and analysis will be done using Web IDQ Software to calculate the concentration of plasma metabolites. MRM (Multiple Reaction Monitoring) based absolute quantification of a panel of important proteins associated with Cardio-metabolic diseases from human plasma (~10 ul) in a single LC-MS/MS run will be performed through Scheduled-MRM method by using heavy isotope-labeled proteotypic peptides. In brief, total proteins from plasma samples will be extracted, estimated, and digested per standard protocols (33-35). The MRM of the targeted proteotypic peptides will be run in a triple quadrupole mass spectrometer using standard and validated methods. The peak area of endogenous and heavy isotope-labeled peptides will be utilized to calculate the concentration of endogenous peptides. The concentration of targeted plasma proteins will then be calculated from the concentration of endogenous proteotypic peptides by the ‘Protein-peptide equivalent molarity concept’.

Metallomics in Phenome-India Population

Metals are crucial for the development and indispensable for the growth and development of organisms. The metal profile of the Indian population is required to establish the role of metal ions in disease biology in general and cardiovascular and diabetes in particular. Though small pockets of cohort studies in India have been found related to a particular disease or particular metal ions in a defined region, large-scale

studies across India taking multiple elements have not yet been reported (36). Few other countries have reported the metal profile in extensive cohort studies using multi-element standards or multiple calibrations to measure the ppb to ppm level of metal ions (37). Nevertheless, metals play a very vital role in physiology and biochemistry. Thus, their prime role in biochemical pathways and the proper functioning of cellular systems cannot be ruled out. Keeping this in view, the Phenome India project aims to analyze more than 30 essential, trace, and toxic serum metals of participants and correlate them to the objectives under study. Moreover, the metal profile of the Indian population will be established for the first time according to the geographical location and food habits from the filled questionnaire.

Phenome India will use the matrix-matched serum calibrators (Recipe) to measure more than 30 elements (Al, As, Au, Ag, Ba, Be, Bi, Cs, Cd, Cr, Co, Cu, Fe, Hg, I, Li, Mg, Mn, Mo, Ni, Pd, Pt, Pb, Sr, Se, Sb, Sn, Ti, Ti, V and Zn) through Inductively Coupled Plasma Mass Spectrometry (ICP-MS).

RNA-seq Assay protocol

We will be using Illumina TruSeq Stranded Total RNA Library Prep Gold (cat. no 20020599) to prepare the sequencing libraries from 250 ng of total blood RNA extracted of participants, as per the manufacturer's reference guide (1000000040499 v00). In summary, cytoplasmic and mitochondrial rRNA will be depleted using target-specific biotinylated oligos and Ribo-Zero rRNA removal beads, and pure RNA will be fragmented using a divalent cation at a high temperature. AMPure XP bead with a sample-to-bead ratio of 1:1 (A63881; Beckman Coulter) will be used for cleaning the final library. The library quality will be checked using an Agilent 2100 bioanalyzer. The libraries will be sequenced on the NextSeq 2000 platform P4 sequencing reagent kit, with 2 x 151 cycles of sequence by synthesis (SBS) and at a final loading concentration of 650 pM. The initial quality assessment of raw sequencing reads will be conducted using FastQC, and subsequent removal of adapter sequences will be performed using Trimmomatic. Before identifying editing sites, we will perform read filtering to ensure stringent read quality, removing reads with a phred score < 30 and poor quality bases,

which will be trimmed. Therefore, all the reads selected for the analysis have a minimum error rate of 0.001 per base. Trimmed sequencing reads were taken for differential gene expression analysis. The reads will be aligned to the human genome using the Salmon quasi-mapping tool to quantify read abundance or transcript expression levels. DESeq2 will be employed for the differential gene expression analysis, utilizing Wald's statistics test. Genes with a p-adjusted value of ≤0.05 and a Log2 fold change of ≥ ± 1.5 will be considered differentially expressed.

Genomic studies (GSA)

To discern genetic markers associated with cardio-metabolic diseases, we intend to genotype about 5000 participants in the study using Global Screening Array (GSA) 3.0 (Illumina, USA) on iScan (Illumina, USA). The GSA 3.0 includes a broad set of single nucleotide polymorphisms (SNPs) that are informative across global populations, covering common rare and low-frequency variants. In addition, several disease-related and pharmacogenetically significant SNPs in the array enable us to analyze their prevalence in the context of Indian populations. We expect the analysis will enable us to generate polygenic risk scores with improved accuracy for cardio-metabolic diseases in Indians. The data shall also be used to explore previously unexplored associations of these SNPs with fecal microbiome, serum metal content, skin hydration status, etc.

Serum miRNA profiling

The human serum contains an array of miRNAs, which may be promising biomarkers for various conditions, especially cardiovascular disease, cancer, and neurodegenerative conditions, where the lack of objective diagnostic criteria is a long-standing problem (38). However, no baseline data is available for serum miRNA profiles from apparently normal Indians. Here total RNA will be isolated from 100 microL of serum collected from subjects, using the RNA Isoblood reagent as per manufacturer's instruction. A polyA tail will be added using E. coli Poly(A) polymerase; NEB #M0276). Further, using the polyA tail to anchor a universal adaptor (5' CTCAATCGTACATAGAAACAGGGATCTTTTTTTTTTTTTTTTTTVN 3'), the cDNA

corresponding to the RNA will be prepared as per manufacturer's protocol (M-MuLV RT; NEB #M0253L). A DNA oligonucleotide corresponding to the miRNA sequence will be used as the forward primer, along with the universal reverse primer (CTCAATCGTACATAGAAACAGGGATC) and SyBR PCR mix (KAPA SYBR FAST qPCR Master Mix (2X) Universal; Roche #KK4618) to amplify selected miRNAs in the RT-PCR (Applied Biosystems QuantStudioTM 6-Flex). Lastly, the threshold cycle (CT) value of each miRNA will be compared to U6 RNA, to correct for differences in starting amount of total RNA. The expression level of miRNAs will be estimated using ΔCT where ΔCT = CT (U6)- CT (miRNA) (39).

Telomere

Telomeres present at the chromosome ends offer genomic stability. With each round of cell division, telomeres shorten, and in different physiological and pathological conditions, accelerated shortening of telomeres has been reported (40, 41). The length of telomeres is broadly understood as a molecular marker that discerns healthy and diseased states. Evaluating telomere length over a longitudinal cohort of healthy volunteers might provide interesting insights into whether and how telomeres relate to healthy people over the years. Moreover, if short or long telomeres in volunteers correlate to other biomarkers and/or physiological conditions. We use a reported quantitative PCR-based method to analyze telomeres length from about 0.5-1 microgram of isolated DNA (42). Optimized by us, the method uses amplification of telomeric repeats by PCR, which is then normalized to a single copy reference gene.

Gut Microbiome

Stool sample were collected by the participant at home or the sample collection center using the feces catcher (Zymo Research) and sterile hand gloves (Kimtech) provided in the stool sample collection kit in Sterile Multi-Purpose Clinical Sample Collector with spoon (Himedia). Samples were collected and stored at 0-4℃ immediately until DNA extraction. Participants who reported having taken antibiotics in the last two weeks before sample collection were excluded from the study. Metagenomic DNA extraction

were done on the same day with 300 mg of stool sample using QIAamp Powerfecal Pro DNA kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in Tris-EDTA buffer instead of the C6 solution provided in the kit. The V3-V4 region of the 16S rRNA gene were sequenced using the '16S Metagenomic Sequencing Library Preparation' protocol by Illumina. The 341F and 785R 16S rRNA gene primers with Illumina indexing overhangs were used. The libraries were sequenced on the MiSeq™ platform using the MiSeq™ Reagent Kit v3 (2 × 300 bp; Illumina) with a depth of 100,000 reads per sample. The taxonomic composition of metagenomes were profiled and analyzed by the QIIME2 pipeline.

Cytokine Analysis

In our study, we utilized multiplexing to analyze cytokine concentrations in human plasma using a bio-plex pro-human cytokine assay containing 48 cytokines (cat no – 12007283), which enables simultaneous detection and quantification of 48 different cytokines, facilitating a more efficient and comprehensive analysis of the cytokine profile. In this method, 40 µL of plasma isolated from blood samples is diluted with the sample diluent provided in the kit in a 1:2 dilution. Then the sample preparation for analysis is done by following the instructions provided by the kit manufacturer. After the preparation of the samples they were analyzed on the Bioplex 200 platform (Manufacturer-Biorad, USA)

Data acquisition and storage

Data Management

Technical Specifications: The backend API and frontend application were integrated and deployed primarily on a virtual machine provided by IGIB. Data was stored in a relational database management system-based architecture on the back end. The front end was developed using PHP-based architecture and framework with modular supports from JavaScript, CSS, and web server-related programming widgets. Robust data security and access management protocols were deployed, and development

using non-standard protocols will be avoided to prevent potential ‘SQL Injection attacks. Data backup, recovery, and mirrored deployment (disaster management protocols) were implemented (Figure 3).

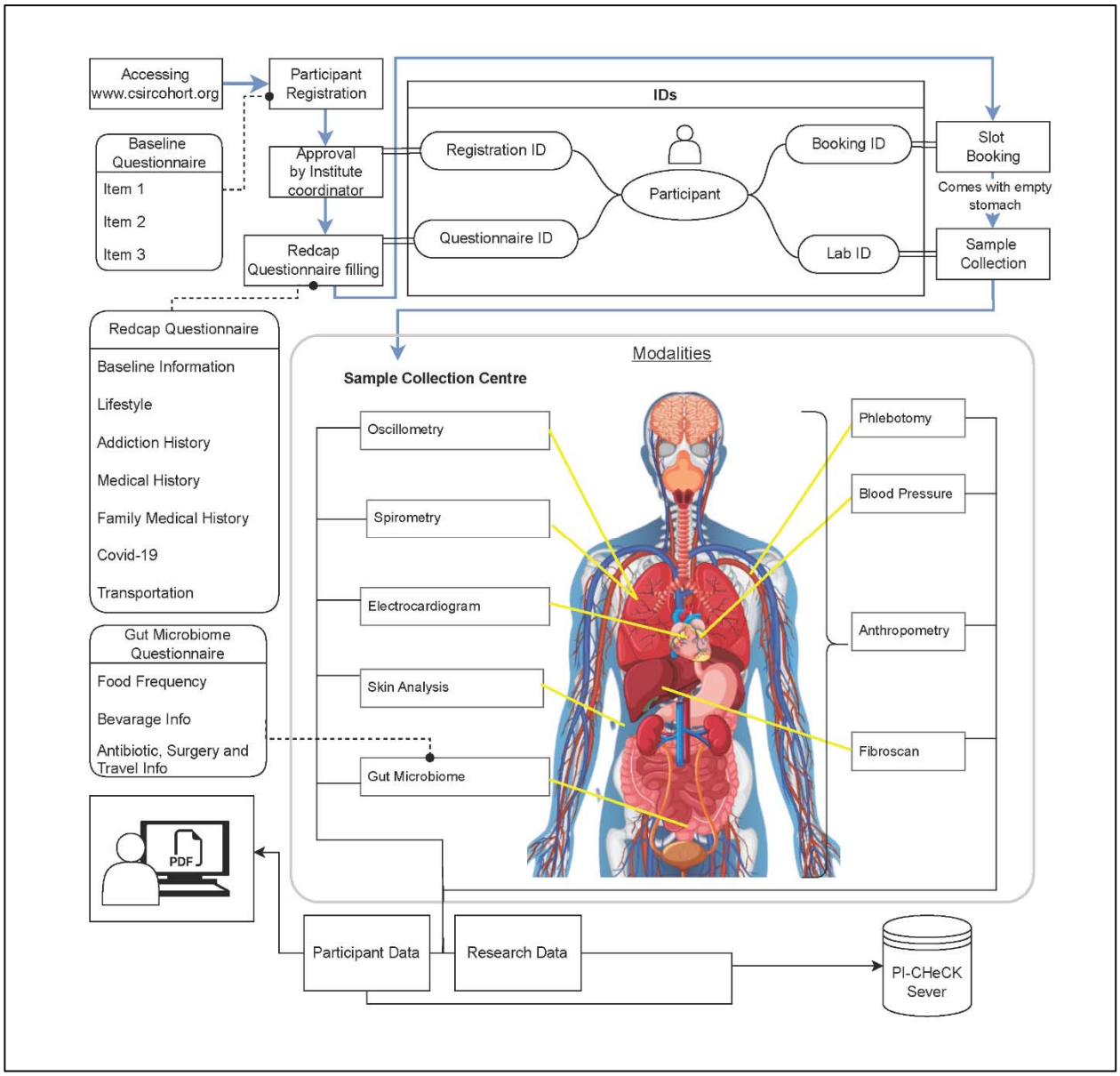

Figure 3: Overall flow of data in the Phenome India Project.

Data acquisition

The data acquisition process for the PI-CHeCK study, which focuses on examining the long-term health outcomes related to cardio-metabolic risks, involves collecting data from 10,000 participants of diverse backgrounds over a five-year period. The study will utilize various tools and methods, including surveys, laboratory tests, clinical tests, and imaging techniques. Medical instruments like ECGs, spirometers, oscillometers, Fibroscan, blood analyzers, dynamometers, body composition analyzers, stadiometers, and various skin probes. These instruments undergo regular calibration and maintenance to ensure data accuracy. Data is uploaded via secure methods, including direct uploads, manual entry, and data importation, with stringent quality assurance measures, such as automated validation checks, manual reviews, and proper personnel training. To protect data security and privacy, participant information is de-identified and Base64 encoded, with restricted access granted only to authorized personnel. The data is stored in the in-house server, established at CSIR-IGIB, where the PI-CHeCK portal is hosted. The data backup is managed by a cron job to store it on the same server and at a remote FTP (File Transfer Protocol) server at regular intervals. Researchers need access to participant's data stored in a centralized database for their studies. However, access to this data will be controlled to ensure patient privacy and compliance with regulations like HIPAA (Health Insurance Portability and Accountability Act).

Data security and privacy:

Utmost care has been taken to maintain data privacy, i.e., data is secured at rest, during transit, and editing using HIPAA guidelines. Collected multi-omics/clinical/Lifestyle data from different CSIR institutes is stored in a Cohort database. The database can be queried using a secured Graphical User Interface (GUI) application. Individuals (who provided samples) were treated as 'normal users' of the application and allowed only to view the submitted data. Power users with 'Lab' privileges can generate sample IDs (multiple samples for each 'normal user'), add/link protocols, and upload data files (raw and processed) against anonymized sample IDs only.

A centralized data platform of about 500 terabytes (TB) and appropriate data security, including data protection and database development, were set up at CSIR-IGIB. CSIR-4PI has been recognized as a data recovery site. To enable collaboration amongst the participating CSIR labs, curated de-identified data will be made available for in-depth analysis.

Each sample was provided with a unique identifier during the sample registration. The unique ID is then printed as a small sticker and pasted on the collection tube. Users with 'Administrator' privileges can Upload 'Announcements', 'News' etc. Through a content management system (CMS) included in the application. The GUI includes static pages for a broad project description with a gallery view (image grid with auto-slider) (https://csircohort.org/).

Return of the results

Results of scanning and phlebotomy were provided to the participants within 72 hours. Participants were able to access the portal only through the OTP-enabled login.

Data Sharing

Data Accessibility: Data accessibility refers to the ability of authorized users or systems to access specific data. It involves establishing protocols and permissions to ensure that only authorized users can access the necessary data.

Researchers need access to participants' data stored in a centralized location for the purpose of the research. However, access to this data must be controlled to ensure patient privacy and compliance with regulations. Only authorized researchers with approved projects and predefined objectives should be able to access the data. Technical controls, such as role-based access control (RBAC) and encryption, will be implemented to prevent unauthorized access, and audit logs should be maintained to track data access and modifications. Requesters must fill out the data access request form, which the data access committee will review. The following will assess the objective/purpose of the data access does not overlap with the existing research

objectives. Once approved, the requester will be granted data access for a defined period. The requester will not share data with anyone else who is not mentioned in the request form.

Secure data sharing using SSH File Transfer Protocol (SFTP) is essential for maintaining sensitive research data's confidentiality, integrity, and availability, ensuring compliance with ethical and regulatory standards. SFTP provides a secure alternative to traditional methods such as email and FTP by encrypting data during transmission and implementing robust authentication mechanisms. Our protocol includes configuring the SFTP server, establishing user groups with appropriate permissions, and setting up secure directories, with additional measures like VPN configuration and tools such as FileZilla or command-line interfaces to enhance security and efficiency. Data requests must be submitted via email, detailing the intended use, and upon committee approval, requestors receive login credentials and a user guide. An MD5 checksum file is provided for each data directory or file, requiring users to perform an MD5 checksum comparison to verify the completeness and integrity of the downloaded data, adding a crucial layer of verification to the process.

Data backup and mirroring

Data backup will be done regularly on different physical machines. Data mirroring efforts will be made at different locations, preferably other CSIR labs located at a location different from the primary data storage site.

Development of the CMD risk scores

The Framingham Risk Score (FRS) is a widely used tool for estimating the 10-year cardiovascular disease (CVD) risk based on traditional risk factors like treated or non-treated hypertension, smoking, high-density lipoprotein, total cholesterol, and fasting blood sugar. However, its application across diverse populations often necessitates recalibration for demographic, environmental, and genetic variations that may influence CVD risk in specific cohorts. For PI-CHeCK, recalibration of the FRS is essential to

enhance the accuracy of risk prediction models and ensure they reflect the unique characteristics of the population under study. This involves a systematic strategy that begins with validating the FRS on the in-house dataset, followed by adjusting the survival coefficients as per WHO's reported mortality rate for CVD in both genders and calculating the new reference score that captures the specific risk profile of the population using the mean of log values. Such recalibration improves the predictive power of the FRS and supports more personalized and effective preventive strategies tailored to the population's needs.

In addition to data visualization, we plan to analyze inter-parameter linear/nonlinear correlation keeping given age/gender covariates, use open source tools to identify patterns within the data, develop predictive models using ML/AI algorithms to predict certain phenotypic conditions, create an integrative network of various genomics, proteomics, and metabolomics parameters to understand the systemic regulation (Figure 4).

Sample and indexed patient similarity can be achieved through clustering and Euclidean/ Mahalanobis distance-based methods, which can be used to quantify these similarities. These similarity scores can then be fed and used to train personalized models for risk score assessment.

Using collaborative filtering techniques, individuals having similar phenotypes can be clubbed together. Employing cohort participants' similarities will help identify a subset proximal to an index patient, which will then be used to train a personalized model in a disease-specific manner. The approach for the personalized model will depend on the type of disease and ground truth label criteria. Recent studies have used various algorithms ranging from recurrent neural networks and ensemble techniques as personalized models for disease state prediction (10, 11). Multi-view datasets will allow us to explore the quantitative association of collected risk factors with the comorbidity- and disease-specific outcomes. The developed framework will be based on explainable AI, rendering the weight of each contributing factor to the predicted outcome.

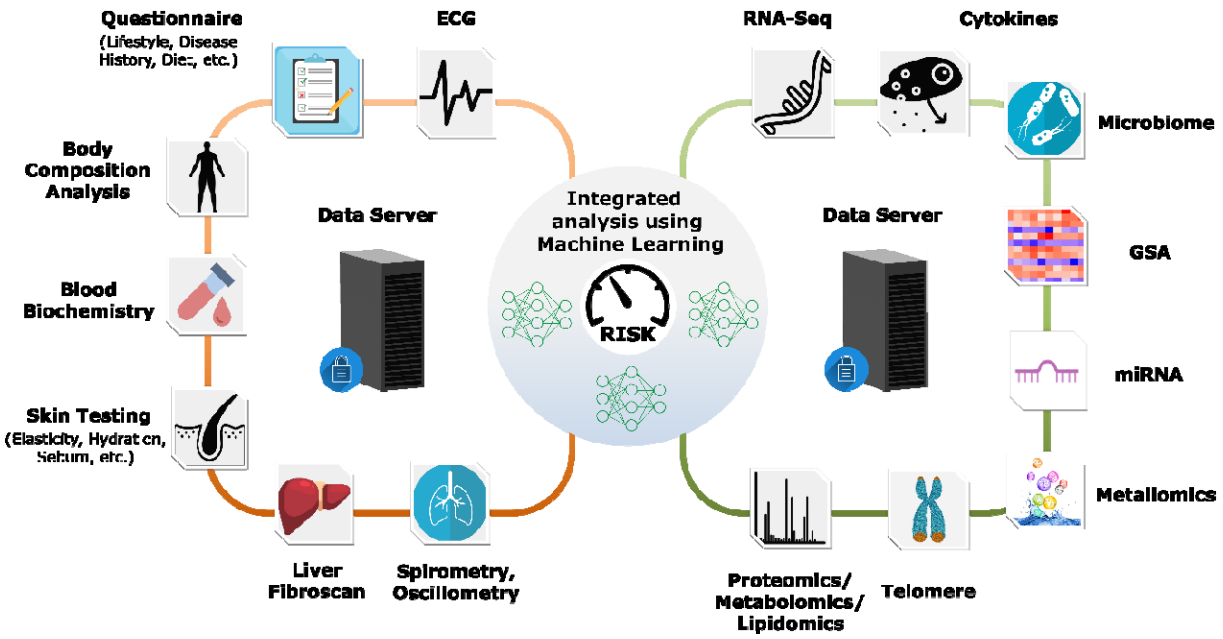

Figure 4: Schematics depicting multi-modal data integration for risk score development.

Discussion

The incidences of metabolic and other non-communicable complex diseases are increasing at an alarming rate in our country. India is now considered to be the capital of cardiovascular diseases and diabetes in the world. Thus, from the public health perspective, it is of prime importance to understand the mechanisms underlying the increased risk of these diseases in the Indian population and develop new strategies for risk stratification, prevention, and management of these major diseases.

Large-scale longitudinal cohort studies are critical for investigating the cause of disease and establishing links between risk factors and health outcomes. For instance, the Framingham heart study, an ongoing study that began in 1948, has provided the key information on most of the currently well-accepted risk factors for heart disease, including the role of lipids, diet, exercise, hypertension, cigarette smoking, etc. It has led to the development of a 10-year cardiovascular risk prediction score popularly called the Framingham risk score (FRS) (43-47). However, the FRS has low value for Indian patients due to varied lifestyles and risk factors, necessitating the development of our own risk matrix and prediction scores based on long-term prospective data. Therefore, an indigenized longitudinal population study, consisting of phenotypic measurements at baseline and long-term follow-up, is essential to understand the complex interrelationships between environmental, genetic, and other molecular factors with disease risk in our country.

Over the last few decades, several studies were initiated to identify diagnostic and prognostic markers in India and based on localized case-control studies where a small number of biochemical markers were considered. Such studies were powered only to look at the association of these markers with the disease. Some of these large-scale studies were cross-sectional, where critical time trajectories of diseases could not be analyzed (48-52). With these limitations, concluding these studies from the pan-Indian perspective is difficult. In India, CSIR, with its more than forty constituent laboratories and centers spread all over the country, represents a wide range of ethnicity subclasses, geo-social habitats, and occupational exposures and is ideally placed to overcome such shortcomings. This has been demonstrated during the recent pandemic of COVID-19 where a longitudinal cohort helped determine infection dynamics, antibody response dynamics, progression toward herd immunity, and the likelihood of significant outbreaks in a population. The study helped obtain sentinel data about the spread pattern and different characteristics of SARS-CoV-2 infection across India. The multi-centric Indian cohort showed that a large number of Indians, probably exceeding a hundred million, had contracted asymptomatic SARS-CoV-2 infection by September 2020 (53).

Further, advances in molecular techniques in genomics, transcriptomics, proteomics, metabolomics, and metagenomics in the last decade have allowed us to comprehensively understand several facets of human biology. CSIR, through the Phenome India project, intends to leverage its strength in these advanced techniques and combination with the prospective collection of biological samples from a diverse population amenable to long-term follow-up, fill in the current gaps in efforts for the establishment of causation and development of diagnostic and prognostic biomarkers for chronic non-communicable diseases in the country. Establishing such a cohort with longitudinal biological and physiological parameter sampling will align with the framework of UN-SDG and the National Health Mission. It will facilitate the development of national reference standards to aid clinical decision-making and national healthcare policy decisions.

As described above, the extensive phenotyping of individuals makes this cohort extremely valuable and unique. To mention a few, tests like fibroscan, oscillometry, gut microbiome, and skin barrier assessments are probably being performed in the country's diverse and large population of this extent for the first time. Prospective collection and analysis of fecal microbiomes adds another unexplored dimension to cardio-metabolic disease risk. Evaluation of metal content in serum and plasma samples of the participants and comparison of these with the information available regarding these metals in various geographies and environments is expected to illuminate their complex genetic, dietary, and environmental regulation in addition to their impact on metabolic diseases. This deep phenotyping is also expected to unravel several novel genotypic and phenotypic associations in the Indian population.

To conclude, this multicenter multi-omic prospective cohort study is an essential step towards precision medicine, especially in the context of the Indian population. This study is expected to generate normative data for several parameters that will impact health care in India shortly. Such comprehensive phenotyping is being attempted for the first time in the country. Coupled with data analytics, AI/ML approaches will hopefully lead to the development of reliable risk prediction tools for cardio-metabolic disorders.

Ethics

The study was approved by IHEC at IGIB (reference no.: CSIR-IGIB/IHEC/2023-24/16). CCMB (IEC95-R1/2023). The study was registered vide CTRI number-CTRI/2024/01/061807 .

Funding

The study is funded by CSIR through grant HCP47.

Acknowledgments

We acknowledge the support from CSIR, prospective participants, and volunteers to get this project started.

Author Contributions

Conflicts of Interest

All the authors declare no conflict of interest.

Data Availability

Anonymized Data for public use may be made available after 3 yrs from completion of baseline phase of study or as per advisory from Monitoring Committee of the project if any revisions thereof.

List of and DeidentifiedAbbreviations

AC	:	Abdominal Circumference
ADA	:	American Diabetes Association
AEC	:	Absolute Eosinophil Count
Ag	:	Silver
AI	:	Artificial Intelligence
Al	:	Aluminium
ALC	:	Absolute Lymphocyte Count
ALP	:	Alkaline Phosphatase
ALT/SGPT	:	Alanine Aminotransferase
ANC	:	Absolute Neutrophil Count
API	:	Application Programming Interface
As	:	Arsenic
AST/SGOT	:	Aspartate Aminotransferase
ATS	:	American Thoracic Society
Au	:	Gold
Ba	:	Barium
BCA	:	Body Composition Analysis
Be	:	Beryllium
Bi	:	Bismuth
BMI	:	Body Mass Index
BNP	:	brain natriuretic peptide
BUN	:	Blood Urea Nitrogen
CAP	:	Continuous Attenuation Parameter
CC	:	Chest Circumference
CCMB	:	Centre for Cellular and Molecular Biology
Cd	:	Cadmium
cDNA	:	Complementary deoxyribonucleic Acid
CMS	:	Content Management System
Co	:	Cobalt
Cr	:	Chromium
Cs	:	Caesium
CSIR	:	Council of Scientific and Industrial Research
CSS	:	Cascading style sheet
Cu	:	Copper
CVD	:	Cardio Vascular Disease
DNA	:	Deoxyribonucleic Acid
ECG	:	Electro Cardio Gram
ERS	:	European Respiratory Society
ESR	:	Erythrocyte Sedimentation Rate
FBS	:	Fasting Blood Sugar
Fe	:	Iron

FRS	:	Framingham Risk Score
FTP	:	File Transfer Protocol
GFR	:	Glomerular Filtration Rate
GGT	:	Gamma Glutamyl Transferase
GSA	:	Global Screening Array,
GUI	:	Graphical User Interface
HALE	:	Healthy Life Expectancy
Hba1c	:	Glycosylated Haemoglobin
HBC	:	High Blood Cholesterol
HBP	:	High Blood Pressure
HDL	:	High Density Lipoprotein
Hg	:	Mercury
HIPAA		Health Insurance Portability and Accountability Act
HP	:	Hip Circumference
HRV	:	Heart Rate Variability
I	:	Iodine
ICDS	:	Integrated Child Development Scheme*
ICP-MS	:	Inductively Coupled Plasma Mass Spectrometry
ICP-OES	:	Inductively Coupled Plasma Optical Emission Spectroscopy
IGIB	:	Institute of Genomics and Integrative Biology
IICB	:	Indian Institute of Chemical Biology
LC/MS	:	Liquid chromatography-mass spectrometry
LDL	:	Low Density Lipoprotein
Li	:	Lithium
LMIC	:	Low- and Middle-Income Countries
LSM	:	Liver Stiffness Measurement
MCH	:	Mean Corpuscular Haemoglobin
MCHC	:	Mean Corpuscular Haemoglobin Concentration
MCV	:	Mean Corpuscular Volume
MD5	:	Message-Digest Algorithm 5
Mg	:	Magnesium
miRNA	:	micro ribonucleic acid
ML	:	Machine Learning
Mn	:	Manganese
Mo	:	Molybdenum
MPV	:	Mean Platelet Volume
MRM	:	Multiple Reaction Monitoring
NAFLD	:	Non-Alcoholic Fatty Liver Disease
NASH	:	Non-Alcoholic SteatoHepatitis
NCDs	:	Non-Communicable Diseases
NCL	:	National Chemical Laboratory
NDD	:	new diagnostic design
Ni	:	Nickel
OTP	:	One Time Password
Pb	:	Lead
PCR	:	Polymerase Chain Reaction

PCV	:	packed cell volume
Pd	:	Palladium
PHP	:	Hypertext Preprocessor
PI-CHeCK	:	Phenome India-CSIR Health Cohort Knowledgebase
PLT	:	Platelet
PSA	:	Prostate-Specific Antigen
Pt	:	Platinum
RBAC	:	Role-Based Access Control
RBC	:	Red Blood Cell
RDW - CV	:	Red Cell Distribution Width - Coefficient of Variation
RDW - SD	:	Red Cell Distribution Width - Standard Deviation
RDWI	:	Red Cell Distribution Width Index
RNA	:	Ribonucleic Acid
rRNA	:	Ribosomal ribonucleic acid
SARS-COV	:	Severe Acute Respiratory Syndrome Coronavirus 2
Sb	:	Antimony
SBS	:	Sequence By Synthesis
Se	:	Selenium
SFTP	:	Secure Shell File Transfer Protocol
Sn	:	Tin
SNP	:	Single Nucleotide Polymorphisms
SQL	:	Structured Query Language
Sr	:	Strontium
SSH	:	Secure Shell
TB	:	Terabyte
TE	:	Transient Elastography
Ti	:	Tennessine
TIBC	:	Total Iron Binding Capacity
TLC	:	Total Leucocyte Count
TSH	:	Thyroid Stimulating Hormone
UIBC	:	Unsaturated Iron-Binding Capacity
UN-SDG	:	United Nations Sustainable Development Goals
UT	:	Union Territories
V	:	Vanadium
VLDL	:	Very Low-Density lipoprotein
VPN	:	Virtual Private Network
VRS	:	Voluntary Retirement Scheme
WC	:	Waist circumference
WHO	:	World Health Organization
Zn	:	Zinc

References

1. Noncommunicable diseases 2023 [Available from: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.
2. Joshi P, Islam S, Pais P, Reddy S, Dorairaj P, Kazmi K, et al. Risk Factors for Early Myocardial Infarction in South Asians Compared With Individuals in Other Countries. Jama. 2007;297(3):286-94.
3. Prabhakaran D, Jeemon P, Roy A. Cardiovascular Diseases in India. Circulation. 2016;133(16):1605-20.
4. Menon GR, Singh L, Sharma P, Yadav P, Sharma S, Kalaskar S, et al. National Burden Estimates of healthy life lost in India, 2017: an analysis using direct mortality data and indirect disability data. The Lancet Global Health. 2019;7(12):e1675-e84.
5. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nature Reviews Gastroenterology & Hepatology. 2018;15(1):11-20.
6. Li J, Wang Q, Ni W, Liu C, Li Z, Qi X. Global health burden of cirrhosis and other chronic liver diseases (CLDs) due to non-alcoholic fatty liver disease (NAFLD): A systematic analysis for the global burden of disease study 2019. Global Transitions. 2023;5:160-9.
7. Prabhakar T, Prasad M, Kumar G, Kaushal K, Shenoy PS, Dubey S, et al. High prevalence of MAFLD in general population: A large cross-sectional study calls for concerted public health action. Alimentary Pharmacology & Therapeutics. 2024;59(7):843-51.
8. Ramesh S, Kosalram K. The burden of non-communicable diseases: A scoping review focus on the context of India. Journal of education and health promotion. 2023;12:41.
9. Majumder PP, Basu A. A genomic view of the peopling and population structure of India. Cold Spring Harbor perspectives in biology. 2014;7(4):a008540.
10. Sharma A, Lysenko A, Jia S, Boroevich KA, Tsunoda T. Advances in AI and machine learning for predictive medicine. Journal of Human Genetics. 2024.
11. Ahmed Z, Wan S, Zhang F, Zhong W. Artificial intelligence for omics data analysis. BMC Methods. 2024;1(1):4.
12. Bien SA, Wojcik GL, Hodonsky CJ, Gignoux CR, Cheng I, Matise TC, et al. The Future of Genomic Studies Must Be Globally Representative: Perspectives from PAGE. Annual review of genomics and human genetics. 2019;20:181-200.
13. Mehta LS, Velarde GP, Lewey J, Sharma G, Bond RM, Navas-Acien A, et al. Cardiovascular Disease Risk Factors in Women: The Impact of Race and Ethnicity: A Scientific Statement From the American Heart Association. Circulation. 2023;147(19):1471-87.
14. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. Nature. 2022;611(7934):115-23.
15. Wells JC. Commentary: Why are South Asians susceptible to central obesity? — the El Niño hypothesis. International journal of epidemiology. 2007;36(1):226-7.
16. Chaudhary M, Sharma P. Abdominal obesity in India: analysis of the National Family Health Survey-5 (2019–2021) data. The Lancet Regional Health - Southeast Asia. 2023;14.
17. Ying W. Phenomic Studies on Diseases: Potential and Challenges. Phenomics (Cham, Switzerland). 2023;3(3):285-99.
18. Aggarwal M, Garg NM, Agrawal A, Sardana V. Lung Function Reference Equations for Indian Ethnic Groups Based on a Handheld Forced Oscillation Device for Age 9-19 Years. Indian journal of pediatrics. 2023;90(1):61-8.
19. Kondal D, Patel SA, Ali MK, Mohan D, Rautela G, Gujral UP, et al. Cohort Profile: The Center for cArdiometabolic Risk Reduction in South Asia (CARRS). International journal of epidemiology. 2022;51(6):e358-e71.

20. Anjana RM, Unnikrishnan R, Deepa M, Pradeepa R, Tandon N, Das AK, et al. Metabolic non-communicable disease health report of India: the ICMR-INDIAB national cross-sectional study (ICMR-INDIAB-17). The Lancet Diabetes & Endocrinology. 2023;11(7):474-89.
21. Huffman MD, Prabhakaran D, Osmond C, Fall CH, Tandon N, Lakshmy R, et al. Incidence of cardiovascular risk factors in an Indian urban cohort results from the New Delhi birth cohort. Journal of the American College of Cardiology. 2011;57(17):1765-74.
22. Rastogi A, Rath I, Varadarajan A, Ramakrishna G, Bihari C, Maiwall R. Non-alcoholic fatty liver disease (NAFLD) in lean individuals – Single centre large cohort clinicopathologic and immunophenotypic study. Pathology - Research and Practice. 2022;238:154112.
23. Norton K, & Eston, R. (Eds.). . Kinanthropometry and Exercise Physiology (4th ed.).2018.
24. Peterson DD. Periodic Fitness Testing: Not Just for Athletes Anymore. Strength & Conditioning Journal. 2018;40(5).
25. Rout G, Kedia S, Nayak B, Yadav R, Das P, Acharya SK, et al. Controlled Attenuation Parameter for Assessment of Hepatic Steatosis in Indian Patients. Journal of Clinical and Experimental Hepatology. 2019;9(1):13-21.
26. Shalimar, Kumar R, Rout G, Kumar R, Yadav R, Das P, et al. Body mass index-based controlled attenuation parameter cut-offs for assessment of hepatic steatosis in non-alcoholic fatty liver disease. Indian journal of gastroenterology : official journal of the Indian Society of Gastroenterology. 2020;39(1):32-41.
27. Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2019;156(6):1717-30.
28. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. European Respiratory Journal. 2005;26(2):319-38.
29. Moore VC. Spirometry: step by step. Breathe. 2012;8(3):232-40.
30. Graham BL, Steenbruggen I, Miller MR, Barjaktarevic IZ, Cooper BG, Hall GL, et al. Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. American Journal of Respiratory and Critical Care Medicine. 2019;200(8):e70-e88.
31. Wanger J, Clausen JL, Coates A, Pedersen OF, Brusasco V, Burgos F, et al. Standardisation of the measurement of lung volumes. European Respiratory Journal. 2005;26(3):511-22.
32. Miura S, Iwamoto H, Omori K, Yamaguchi K, Sakamoto S, Horimasu Y, et al. Preserved ratio impaired spirometry with or without restrictive spirometric abnormality. Scientific reports. 2023;13(1):2988.
33. Mohammed Y, Goodlett D, Borchers CH. Absolute Quantitative Targeted Proteomics Assays for Plasma Proteins. In: Greening DW, Simpson RJ, editors. Serum/Plasma Proteomics: Methods and Protocols. New York, NY: Springer US; 2023. p. 439-73.
34. von Zychlinski A, Williams M, McCormick S, Kleffmann T. Absolute quantification of apolipoproteins and associated proteins on human plasma lipoproteins. Journal of proteomics. 2014;106:181-90.
35. Uchida Y, Tachikawa M, Obuchi W, Hoshi Y, Tomioka Y, Ohtsuki S, et al. A study protocol for quantitative targeted absolute proteomics (QTAP) by LC-MS/MS: application for inter-strain differences in protein expression levels of transporters, receptors, claudin-5, and marker proteins at the blood-brain barrier in ddY, FVB, and C57BL/6J mice. Fluids and barriers of the CNS. 2013;10(1):21.
36. Sharma E, Vaidya N, Iyengar U, Zhang Y, Holla B, Purushottam M, et al. Consortium on Vulnerability to Externalizing Disorders and Addictions (cVEDA): A developmental cohort study protocol. BMC Psychiatry. 2020;20(1):2.

37. . !!! INVALID CITATION !!! (29-32).
38. Halushka PV, Goodwin AJ, Halushka MK. Opportunities for microRNAs in the Crowded Field of Cardiovascular Biomarkers. Annual review of pathology. 2019;14:211-38.
39. Shi R, Chiang VL. Facile means for quantifying microRNA expression by real-time PCR. BioTechniques. 2005;39(4):519-25.
40. Bär C, Blasco MA. Telomeres and telomerase as therapeutic targets to prevent and treat age-related diseases. F1000Research. 2016;5.
41. Shammas MA. Telomeres, lifestyle, cancer, and aging. Current opinion in clinical nutrition and metabolic care. 2011;14(1):28-34.
42. Cawthon RM. Telomere measurement by quantitative PCR. Nucleic acids research. 2002;30(10):e47.
43. Dawber TR, Meadors GF, Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. American journal of public health and the nation's health. 1951;41(3):279-81.
44. Gordon T, Kannel WB. Predisposition to atherosclerosis in the head, heart, and legs. The Framingham study. Jama. 1972;221(7):661-6.
45. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. The American journal of cardiology. 1976;38(1):46-51.
46. D'Agostino RB, Russell MW, Huse DM, Ellison RC, Silbershatz H, Wilson PW, et al. Primary and subsequent coronary risk appraisal: new results from the Framingham study. American heart journal. 2000;139(2 Pt 1):272-81.
47. Hemann BA, Bimson WF, Taylor AJ. The Framingham Risk Score: An Appraisal of Its Benefits and Limitations. American Heart Hospital Journal. 2007;5(2):91-6.
48. D K, Swaminathan K, Velmurugan G, Ramakrishnan A, Alexander T, Raghupathy AK, et al. Association between serum uric acid levels and cardiovascular risk factors among adults in India. Nutrition, metabolism, and cardiovascular diseases : NMCD. 2023;33(7):1330-8.
49. Shrivastava U, Misra A, Mohan V, Unnikrishnan R, Bachani D. Obesity, Diabetes and Cardiovascular Diseases in India: Public Health Challenges. Current diabetes reviews. 2017;13(1):65-80.
50. Unnikrishnan AG, Sahay RK, Phadke U, Sharma SK, Shah P, Shukla R, et al. Cardiovascular risk in newly diagnosed type 2 diabetes patients in India. PloS one. 2022;17(3):e0263619.
51. Geldsetzer P, Manne-Goehler J, Theilmann M, Davies JI, Awasthi A, Danaei G, et al. Geographic and sociodemographic variation of cardiovascular disease risk in India: A cross-sectional study of 797,540 adults. PLoS medicine. 2018;15(6):e1002581.
52. Mallinson PAC, Kulkarni B, Bhogadi S, Kinra S. Association between parents' socioeconomic conditions and nutritional status during childhood and the risk of cardiovascular disease in their adult offspring: an intergenerational study in south India. Journal of epidemiology and community health. 2021;75(11):1091-7.
53. Naushin S, Sardana V, Ujjainiya R, Bhatheja N, Kutum R, Bhaskar AK, et al. Insights from a Pan India Sero-Epidemiological survey (Phenome-India Cohort) for SARS-CoV2. eLife. 2021;10:e66537.

S. No.	Lab	Name	City	Zone

Supplementary Information

Supplementary Table S1.

1	CSIR-4PI	CSIR-Fourth Paradigm Institute	Bangalore	South
2	CSIR-AMPRI	CSIR-Advanced Materials and Processes Research Institute	Bhopal	West
3	CSIR-CBRI	CSIR-Central Building Research Institute	Roorkee	North
4	CSIR-CCMB	CSIR-Centre for Cellular Molecular Biology	Hyderabad	South
5	CSIR-CDRI	CSIR-Central Drug Research Institute	Lucknow	North
6	CSIR-CECRI	CSIR–Central Electrochemical Research Institute	Karaikudi	South
7	CSIR-CEERI	CSIR-Central Electronics Engineering Research Institute	Pilani	West
8	CSIR-CFTRI	CSIR-Central Food Technological Research Institute	Mysore	South
9	CSIR-CGCRI	CSIR-Central Glass and Ceramic Research Institute	Kolkata	West
10	CSIR-CIMAP	CSIR-Central Institute of Medicinal Aromatic Plants	Lucknow	North
11	CSIR-CIMFR	CSIR-Central Institute of Mining and Fuel Research	Dhanbad	East
12	CSIR-CLRI	CSIR-Central Leather Research Institute	Chennai	South
13	CSIR-CMERI	CSIR-Central Mechanical Engineering Research Institute	Durgapur	East
14	CSIR-CRRI	CSIR-Central Road Research Institute	New Delhi	North
15	CSIR-CSIO	CSIR-Central Scientific Instruments Organisation	Chandigarh	North
16	CSIR-SERC	CSIR-Structural Engineering Research Centre	Chennai	South
17	CSIR-CSMCRI	CSIR-Central Salt and Marine Chemicals Research Institute	Bhavnagar	West
18	CSIR-HQ	Council of Scientific and Industrial Research	New Delhi	North
19	CSIR-HRDC	CSIR-Human Resource Development Centre	Ghaziabad	North
20	CSIR-HRDG	CSIR-Human Resource Development Group	New Delhi	North
21	CSIR-IGIB	CSIR-Institute of Genomics and Integrative biology	New Delhi	North
22	CSIR-IHBT	CSIR-Institute of Himalayan Bioresource Technology	Palampur	North
23	CSIR-IICB	CSIR-Indian Institute of Chemical Biology	Kolkata	East
24	CSIR-IICT	CSIR-Indian Institute of Chemical Technology	Hyderabad	South
25	CSIR-IIIM	CSIR-Indian Institute of Integrative Medicine	Jammu & Srinagar	North
26	CSIR-IIP	CSIR-Indian Institute of Petroleum	Dehradun	North
27	CSIR-IITR	CSIR-Indian Institute of Toxicology Research	Lucknow	North
28	CSIR-IMMT	CSIR-Institute of Minerals & Materials Technology	Bhubaneswar	East
29	CSIR-IMTECH	CSIR- Institute of Microbial Technology	Chandigarh	North
30	CSIR-TKDL	CSIR-Traditional Knowledge Digital Library Unit	New Delhi	North
31	CSIR-NAL	CSIR-National Aerospace Laboratories	Bangalore	South
32	CSIR-NBRI	CSIR-National Botanical Research Institute	Lucknow	North
33	CSIR-NCL	CSIR-National Chemical Laboratory	Pune	West
34	CSIR-NEERI	CSIR-National Environmental Engineering Research Institute	Nagpur	West
35	CSIR-NEIST	CSIR-North East Institute of Science and Technology	Jorhat	East
36	CSIR-NGRI	CSIR-National Geophysical Research Institute	Hyderabad	South
37	CS-NIIST	CSIR-National Institute for Interdisciplinary Science and Technology	Trivandrum	South
38	CSIR-NIO	CSIR-National Institute of Oceanography	Goa	West
39	CSIR- NIScPR	CSIR-National Institute of Science Communication and Policy Research	New Delhi	North
40	CSIR-NML	CSIR-National Metallurgical Laboratory	Jamshedpur	East

41	CSIR-NPL	CSIR-National Physical Laboratory	New Delhi	North
42	CSIR-SERC	CSIR-Structural Engineering Research Centre	Chennai	South
43	CSIR-URDIP	CSIR-Unit for Research and Development of Information Products	Pune	West

Supplementary Table 2: Details of exclusion criteria for different scanning modalities.

	BCA	**Liver Elastography**	**Skin**	**Oscillometry**	**Spirometry**	**ECG**	**Dynamometer**
Pacemakers, intra-cardiac defibrillators/Loop recorders, or any implantable electronic device	E	E					
Contagious disease/skin disease	E	E	E	E	E	E	E
Suspicion or ongoing treatment of active pulmonary tuberculosis	E	E	E	E	E	E	E
Active injury or an active bleed in measurement area	E		E				
Metal implant, exclude dental implants (Specific Query)	E						
History of myocardial infarction/heart attack/ recent thoracic (chest) or abdominal surgery/Pneumothorax/Hemoptysis, recent eye or ear surgery/sinus surgery or any major surgery/ies (within the previous eight weeks)					E		E
Recent history of bleeding from the mouth or vomiting blood				E	E		
Loose teeth or have facial pain/TM joint issues				E	E		
Respiratory infection in the past month				E	E		
History of Thoracic (chest), abdominal, or cerebral aneurysm/stroke					E		E
Untreated/uncontrolled/ pulmonary hypertension					E		E

Supplementary Table 3: Description of self-reported Main questionnaire in Phenome India cohort.

Domain	Questions
Baseline information	Consent, Gender, Date of Birth
Lifestyle	Brushing habit, Nail biting habit, Delivery mode, Exercise, Stress, Sleep, Bowel, Diet
Female-specific	Menstruation, Menopause, Pregnancies, Live birth, Abortion/stillbirths, Gynecological issues
Addiction	Smoking, Alcohol, Chewing, Other addictions
Health issues	High blood pressure (HBP), Diabetes, High blood cholesterol (HBC), Thyroid, Neurological problems, Heart problems, Respiratory problems, Liver problems, Bone/joint problems, Kidney problems, Eye problems, Gastro-intestinal problems, Hearing problems, Cancer, Depression, Anxiety, Allergies, Recurrent infections, Surgery
Family medical history	Diabetes (Family), High blood cholesterol (Family), Thyroid problems (Family), Neurological Problems (Family), Heart problems (Family), Respiratory problems (Family), Kidney problems (Family), Gastro-intestinal problems (Family), Liver problems (Family), Bone/joint problems (Family), Eye problems (Family), Hearing problems (Family), Cancer (Family), Depression (Family), Anxiety (Family), Allergies (Family), Recurrent infections (Family)

COVID-19	COVID-19 Infection, COVID-19 vaccination, Long term COVID-19 complications
Transportation	Family own car, Family own two-wheeler, Family own cycle, family own no vehicle, Metro travel, Car travel, Two wheeler travel, Bus travel, Cycle Travel, Walking.

Supplementary Table 4: Description of the self-reported Gut questionnaire in the Phenome India cohort.

Domain	**Questions**
Food item consumption	Rice, Wheat (Paratha, chapati, roti, etc), Other cereals (Bajra, jowar, ragi etc), Bread(Yeast), Pulses (Arhar, tur, moong, urd daal etc), Sprouts, Eggs, Poultry (Chicken), Meat (Mutton, pork etc), Fish and other seafoods, Curd, Paneer, Ghee, Butter, Ice cream, Cake and pastries, Chips, namkeen, Pickles, Biscuits, Chocolate, Carbonated beverages (Coke, sprite etc), Fruit juice and shakes, Tubers and starchy vegetables (Potato, corn etc), Green leafy vegetables, Fruits (Banana, apple etc), Dry fruits (Resins, almonds, dates etc), Fermented foods (Idli, dosa, dhokla etc), Fermented drinks, Vitamin supplements, Protein supplements, Probiotics
Alcohol	Beer consumption within 2 days, Alcohol consumption within two days

Travel & eating	Types of green leafy vegetables consumed last week: Travel out of city or country in the last 15 days.
Antibiotic consumption	Last antibiotic consumption: Antibiotics consumption within two weeks
Hospitalization	Surgery in last one month, Hospitalized in last one month
Stool type	Bristol stool type

Supplementary Table 5: Summary of different sampling stations

Modality	Machine	Time (min)	Output Format	Information
Anthropometry	Dynamometer	~1-2	Strength reading entry in portal	Muscle strength
BCA	8-electrode BIA Analyzer	3- 5	PDF and Excel	Body Composition
Phlebotomy	(Details in Supplementary Table)	3- 5	PDF	Blood parameters
Transient Elastography	Fibroscan	6-10 after positioning	PDF and Excel	LSM and CAP score
ECG	Allenger Pisces 1200	6-8 after positioning	PDF report	Heart condition
Spirometry	NDD Easy One	3-10 for three readings	PDF and Excel	Lung Function and Volume
Oscillometry	Tremoflo C100,Resmon Pro V3	3-5 for three readings	PDF and excel	Small Airway Assessment
Skin	Courage Khazaka	8-10	PDF report	Skin health

Supplementary Table 6: Details of the phlebotomy parameters, including biochemistry assays.

Diagnostic Lab-Outsourced					
S. No.	**Units**	**Parameter**	**Sample Type**	**Platform**	**Sample Size**
1	10^3/uL	Absolute.Basophil.Count	Whole Blood	DxH-900 (Beckman Coulter)	All
2	10^3/uL	Absolute.Eosinophil.Count.(AEC)	Whole Blood	DxH-900 (Beckman Coulter)	All
3	10^3/uL	Absolute.Lymphocyte.Count.(ALC)	Whole Blood	DxH-900 (Beckman Coulter)	All
4	10^3/uL	Absolute.Monocyte.Count	Whole Blood	DxH-900 (Beckman Coulter)	All
5	10^3/uL	Absolute.Neutrophil.Count.(ANC)	Whole Blood	DxH-900 (Beckman Coulter)	All
6	U/L	Alanine.Aminotransferase.(ALT/SGPT)	Serum	AU5800 (Beckman Coulter)	All
7	Ratio	Albumin/Globulin.Ratio	Serum	AU5800 (Beckman Coulter)	All
8	U/L	Alkaline.Phosphatase.(ALP)	Serum	AU5800 (Beckman Coulter)	All
9	mg/dL	APO-A1	Serum	AU5800 (Beckman Coulter)	All
10	mg/dL	APO-B	Serum	AU5800 (Beckman Coulter)	All
11	U/L	Aspartate.Aminotransferase.(AST/SGOT)	Serum	AU5800 (Beckman Coulter)	All
12		Average.Estimated.Glucose.-.plasma	Calculated	DxH-900 (Beckman Coulter)	All
13	%	Basophils	Serum	DxH-900 (Beckman Coulter)	All
14	mg/dl	Blood.Urea	Serum	AU5800 (Beckman Coulter)	All
15	mg/dl	Blood.Urea.Nitrogen.(BUN)	Serum	AU5800 (Beckman Coulter)	All
16	Ratio	Bun/Creatinine.Ratio	Calculated	AU5800 (Beckman Coulter)	All
17	%	Eosinophils	Whole Blood	DxH-900 (Beckman Coulter)	All
18	mm/1st hour	ESR	Whole Blood	ALIFAX TEST-1	All
19	ng/mL	Ferritin	Serum	ATELLICA IM (SIEMENS)	All
20	U/L	Gamma.Glutamyl.Transferase.(GGT)	Serum	AU5800 (Beckman Coulter)	All

21	mL/min/1.73m2	GFR,.ESTIMATED	Calculated	AU5800 (Beckman Coulter)	All
22	mg/dl	Glucose,.Fasting	Fluoride	AU5800 (Beckman Coulter)	All
23	Ratio	Green.and.king.index	Whole Blood	DxH-900 (Beckman Coulter)	All
24	g/dL	Haemoglobin.(HB)	Whole Blood	DxH-900 (Beckman Coulter)	All
25	%	Hba1c.(Glycosylated.Hemoglobin)	Whole Blood	BIORAD D-100 (BIORAD)	All
26	Ratio	HDL./.LDL.Cholesterol.Ratio	Serum	AU5800 (Beckman Coulter)	All
27	%	Hematocrit.(PCV)	Whole Blood	DxH-900 (Beckman Coulter)	All
28	Ratio	LDL./.HDL.Cholesterol.Ratio	Serum	AU5800 (Beckman Coulter)	All
29	%	Lymphocytes	Whole Blood	DxH-900 (Beckman Coulter)	All
30	pg	Mean.Corp.Hb.(MCH)	Whole Blood	DxH-900 (Beckman Coulter)	All
31	g/dL	Mean.Corp.Hb.Conc.(MCHC)	Whole Blood	DxH-900 (Beckman Coulter)	All
32	fL	Mean.Corp.Volume.(MCV)	Whole Blood	DxH-900 (Beckman Coulter)	All
33	Ratio	Mentzer.Index	Whole Blood	DxH-900 (Beckman Coulter)	All
34	%	Monocytes	Whole Blood	DxH-900 (Beckman Coulter)	All
35	fL	MPV	Whole Blood	DxH-900 (Beckman Coulter)	All
36	%	Neutrophils	Whole Blood	DxH-900 (Beckman Coulter)	All
37	mg/dl	Non-HDL.Cholesterol	Serum	AU5800 (Beckman Coulter)	All
38	10^3/uL	Platelet.Count(PLT)	Whole Blood	DxH-900 (Beckman Coulter)	All
39	%	RDW.-.CV	Whole Blood	DxH-900 (Beckman Coulter)	All
40	fL	RDW.-.SD	Whole Blood	DxH-900 (Beckman Coulter)	All
41	Ratio	RDWI	Whole Blood	DxH-900 (Beckman Coulter)	All
42	10^6/uL	Red.Blood.Cell.Count.(RBC)	Whole Blood	DxH-900 (Beckman Coulter)	All
43	g/dL	Serum.Albumin	Serum	AU5800 (Beckman Coulter)	All
44	mg/dL	Serum.Bilirubin,.(Direct)	Serum	AU5800 (Beckman	All

				Coulter)	
45	mg/dl	Serum.Bilirubin,.(Indirect)	Serum	AU5800 (Beckman Coulter)	All
46	mg/dl	Serum.Bilirubin,.(Total)	Serum	AU5800 (Beckman Coulter)	All
47	mg/dl	Serum.Calcium	Serum	AU5800 (Beckman Coulter)	All
48	mmol/L	Serum.Chloride	Serum	AU5800 (Beckman Coulter)	All
49	mg/dl	Serum.Creatinine	Serum	AU5800 (Beckman Coulter)	All
50	gm/dl	Serum.Globulin	Serum	AU5800 (Beckman Coulter)	All
51	mg/dl	Serum.HDL.Cholesterol	Serum	AU5800 (Beckman Coulter)	All
52	ug/dl	Serum.Iron	Serum	AU5800 (Beckman Coulter)	All
53	mg/dl	Serum.LDL.Cholesterol	Serum	AU5800 (Beckman Coulter)	All
54	mg/dl	Serum.Phosphorus	Serum	AU5800 (Beckman Coulter)	All
55	mmol/L	Serum.Potassium	Serum	AU5800 (Beckman Coulter)	All
56	mmol/L	Serum.Sodium	Serum	AU5800 (Beckman Coulter)	All
57	µg/dl	Serum.Total.Iron.Binding.Capicity.(TIBC)	Serum	AU5800 (Beckman Coulter)	All
58	g/dL	Serum.Total.Protein	Serum	AU5800 (Beckman Coulter)	All
59	mg/dL	Serum.Triglycerides	Serum	AU5800 (Beckman Coulter)	All
60	mg/dL	Serum.Uric.Acid	Serum	AU5800 (Beckman Coulter)	All
61	mg/dl	Serum.VLDL.Cholesterol	Serum	AU5800 (Beckman Coulter)	All
62	Ratio	SGOT/SGPT.Ratio	Calculated	AU5800 (Beckman Coulter)	All
63	ulU/mL	Thyroid.Stimulating.Hormone.(TSH)-Ultrasensitive	Serum	ATELLICA IM (SIEMENS)	All
64	Ratio	Total.CHOL./.HDL.Cholesterol.Ratio	Calculated	AU5800 (Beckman Coulter)	All
65	mg/dL	Total.Cholesterol	Serum	AU5800 (Beckman Coulter)	All
66	10^3/uL	Total.Leucocyte.Count.(TLC)	Whole Blood	DxH-900 (Beckman Coulter)	All
67	%	Transferrin.Saturation.%	Serum	AU5800 (Beckman Coulter)	All

68	pg/mL	UIBC	Serum	AU5800 (Beckman Coulter)	All
69		Urea/Creatinine.Ratio	Calculated	AU5800 (Beckman Coulter)	All
70	pg/mL	VITAMIN.B12	Serum	ATELLICA IM (SIEMENS)	All
71	ng/mL	VITAMIN.D.(25.-.OH.VITAMIN.D)	Serum	ATELLICA IM (SIEMENS)	All
In -House					
S.No		**In House Parameter**	**Sample Type**	**Platform**	**Sample Size**
1	ng/mL	C-Peptide	Plasma	Abbott Architect	All
2	mU/mL	Insulin	Plasma	Abbott Architect	All
3	pg/mL	iPTH	Plasma	Abbott Architect	All
4	ng/mL	Folate	Plasma	Abbott Architect	All
5	mmol/dL	Magnesium	Plasma	Abbott Architect	All
6	ug/dL	Cortisol Morning	Plasma	Abbott Architect	All
7	umol/L	Homocysteine	Plasma	Abbott Architect	Restricted
8	g/L	Haptoglobin	Plasma	Abbott Architect	Restricted
9	pg/mL	BNP	Plasma	Abbott Architect	Restricted
10	AU/mL	SARS-Cov2 IgG Semi Quantitative	Plasma	Abbott Architect	Restricted
11	U/mL	Anti-TPO	Serum	Abbott Architect	Restricted
12	ng/mL	PSA	Serum	Abbott Architect	Restricted
13	ng/mL	Prolactin	Serum	Abbott Architect	Restricted
14	pmol/L	Active B12	Serum	Abbott Architect	Restricted