

# Biobanking in India: Road to Translation Research and Precision Medicine Initiative

Suchi Vaidya<sup>1</sup>, Khushbu Shah<sup>2</sup>, Pooja Utekar<sup>3</sup>, Simran Keshwani<sup>4</sup>, Abhishek Prajapati<sup>5</sup>, Prashant Khadke<sup>6</sup>

<sup>1, 2, 3, 4, 5, 6</sup>Pensieve Health Pvt Ltd, Dr. Annie Besant Road, Worli, Mumbai, India

<sup>1</sup>Email: [suchivora\[at\]pensievehealth.com](mailto:suchivora[at]pensievehealth.com)

<sup>2</sup>Email: [khushbushah\[at\]pensievehealth.com](mailto:khushbushah[at]pensievehealth.com)

<sup>3</sup>Email: [poojautekar\[at\]pensievehealth.com](mailto:poojautekar[at]pensievehealth.com)

<sup>4</sup>Email: [simrankeshwani\[at\]pensievehealth.com](mailto:simrankeshwani[at]pensievehealth.com)

<sup>5</sup>Email: [abhishekprijapati\[at\]pensievehealth.com](mailto:abhishekprijapati[at]pensievehealth.com)

<sup>6</sup>Email: [prashantkhadke\[at\]pensievehealth.com](mailto:prashantkhadke[at]pensievehealth.com)

**Abstract:** *Non-Communicable diseases are on a rise and India reports one of the highest number of cases for diabetes, cardiovascular diseases, chronic kidney diseases, mental health disorders and cancer. These cases have increased and made the people even more vulnerable after the SARS COV - 19 pandemic. In the age of Precision Medicine, it is time to tap the diversity in the country in terms of culture, ethnicity and genetic variability. Tapping the diversity will not only help in the drug discoveries but also to help make the treatment regimens more streamlined and with lesser side effects. This can be achieved with the help of biobanking which is in its infancy stage as of now in India. There are multiple small biobanks in the country however, forming a consortium of these biobanks is the need of the hour.*

**Keywords:** Biobanks, Data, Electronic Health Records (EHR), Regulatory Guidelines

## 1. Introduction

A Biobank is an entity which stores the biological samples from human beings. It could be in the form of blood, tissue, DNA, RNA and clinical records. In the recent past, the concept of Biobank has been accepted in India, however, the establishment of biobanks is still in its struggling phase. At present, India hosts 19 registered biobanks of the 340 global biobanks with many others yet to be recognised globally. [1]

*“as legal entities or parts of a legal entity that perform biobanking and states that biobanking is the process of acquisition and storing, together with some or all of the activities related to collection, preparation and preservation and testing, analysing and distributing defined biological material as well as related information and data” - ISO20387: 2018*

### Need for Biobanking in India

Non communicable diseases (NCD's) are on a rise all over the world and becoming a major concern in India as well. It has been published by the Ministry of Health and Family welfare that the deaths due to NCD's have risen from 37.9% in 1990 to 61.8% in 2016. The non - communicable diseases include cardiovascular diseases, diabetes, chronic respiratory diseases, mental health and neurological disorders, musculoskeletal disorders, cancers, and chronic kidney disease. More than 6 million people died in India due to Non communicable diseases in 2016, with ischemic heart disease being the leading cause of death. Institute of Health metrics and evaluation highlights under nutrition, air pollution, and a group of risks causing cardiovascular disease and diabetes, as the major risk factors contributing to health loss in the country. The group includes an unhealthy diet, high blood pressure, high fasting plasma glucose, high cholesterol and being overweight. These non - communicable diseases are not

only lifestyle related diseases but also based on the genetic make - up of the individual.

Communicable diseases remain one of the major causes for the deaths and illness in the country having its burden at about 27.5% for 3 decades. [2]. Severe Acute Respiratory Syndrome, Coronavirus 2 (SARS CoV 2) which hit the world in 2019, has resulted in a number of changes in the patients in terms of mental health and other metabolic diseases. Data 4 Life, a decentralised cohort study is one such example which tries to tap the everlasting effects post Covid in 10, 000 Indian participants. [3]. The study of genetic make - up and its effects on the health and the treatment regimens becomes increasingly important today where the complex diseases get even more complex with the advent of SARS CoV 19 pandemic. Availability of a large number of samples in a biobank, together with their complete electronic clinical data and genealogical information will help us to boost our research towards preventive, diagnostic and treatment protocols; and translate the research into high - quality patient care.

At present the available genomic data from 3700 GWAS studies so far is based on the European, US and Icelandic studies. There is no scientometric study performed on these GWAS studies. The lack of ancestral, demographic diversities increases the need of data from Asian and African countries to contribute in the genetic data pool. As reported based on the genetic studies conducted by the Indian scientists, computational analysis led to the identification of 55, 898, 122 single nucleotide variants of which 18, 016, 257 (32%) variants were unique to India genome dataset.

### India: Diversity at Genomics and Phenomics level

India is the most populated country in the world with the estimated population of 1, 428, 627, 663. It contributes about 17.76% of the total world population. India has 4500 anthropologically different distinct populations, differentiated on the basis of caste, tribe, and religious groups that differ in

terms of cultural practices, geographical locations, climatic conditions, physical features, marriage practices, linguistics, as well as their genetic architecture. India has the second largest tribal community in the world consisting of nearly 100 million tribal people contributing to about 8.6% of the Indian population. These tribes mostly reside in the north eastern, central and southern parts of India. [4] The Gonds, Bhils, Santhals, Munda, Khasi Garo are the tribal communities residing in India to name a few.

India has witnessed migrants from Africa through southern coastal routes due to which the genetic diversity has been enriched even further. The Siddi Community, wherein the Bantu tribe from East Africa were brought to India as slaves approximately 200 - 300 years ago and ever since they reside in the states of Gujarat, Karnatak, Goa and Telangana. Indian genetic diversity has been further stratified into endogamous groups and is also characterised by consanguineous marriages. This has resulted in many recessive alleles in the Indian population.

India is home to more than 1 billion people and is a land of large human diversity. It is interesting to know that a DNA molecule of two unrelated individuals differs by a mere 0.1% and rest 99% is identical. The 0.1% of the DNA which is not identical harbors millions of base pairs and is a storehouse to a rich source of information. Endogamous marriages in India (marriage within population) has resulted into population specific diseases and some genetic variations have resulted into India specific diseases. Accumulating the genetic data from different regions of India in the form of biobanks will not only help develop a correlation matrix but also will build the scope to enrich the knowledge in terms of etiology, epidemics and epigenetics of the diseases.

### Types of biobanks

The biobanks are broadly classified based on a number of factors like:

#### 1) Approach Based:

- a) Population based
- b) Hospital or Academic Based
- c) Disease oriented based

#### 2) The sample type:

- a) Omics based bio bank (DNA/RNA/ Plasma)
- b) Tissue/ multiple specimen (biobanking for translational diabetes research in india)
- c) Image Biobanking (digital X ray, MRI, histopathology slides)
- d) Microbiome Biobank

### Requirements of Biobank

#### 1.1 Regulatory Requirement in the Biobank

##### 1.1.1 Ethical Consent and Institutional Review Board (IRB)

For a biobank to function smoothly, it must have an ethics committee who approve the projects of the Biobank and makes sure that the protocols and the processes are in compliance with the laid guidelines. An ethical committee reviews the project proposal submitted by the biobank independently without any bias and gives its approval by providing an ethical oversight to different aspects of the Biobank including legal and *guidelines for sustaining DBT* -

*Bio - banks and Cohorts*) bioethics, which consists of giving the participant an overview of the project and his readiness to participate in the research study program. [5]

Any amendments made to the SOP, medical history form or the consent form should be informed to the ethics committee and if required an approval should be taken from the committee members. Ethics Committee is an integral part of the biobanking project as it shoulders the major responsibility of approving the project which the biobank wishes to carry out.

##### 1.1.2 Data Monitoring Committee (DMC)

Data confidentiality and its security is of utmost importance to any biobank. Primary function of any biobank remains to de - identify and safeguard the medical history and the sequenced genomic data (for a biobank who keeps the genetic information) of the participants who supplied it. A privacy policy should be made available for the participants to view as and when they require. It is the biobank's responsibility to secure the personal information received from the participants and for this they can implement access control to retrieve any participant related information. DMC works in close relation with Principle investigator to oversee the integrity, assess the risk & benefits and make recommendations to maintain the security and the confidentiality of the data in the biobank. DMC can consist of a bio - statistician, lawyer, clinicians and an ethicist.

##### 1.1.3 Data Storage: Requirements and management

Data in a Biobank comprises of sample - associated data, metadata about the sample history, quality, and its proper usage. [6] Data management is one of the key factors in a Biobank to make the right data available at the right time. A good quality data is based on its Findability, Accessibility, Interoperability and Reusability (FAIR). [6]. The data management, confidentiality and its security is of prime importance for any Biobank.

Management of the data starts from the time the sample enters the Biobank. It consists of registering the sample in the Biobank by assigning its unique identity, uploading the medical history, consent, the supporting data (clinical reports) and its associated genetic data. As per the current ICMR rules, it is mandatory to store the hard copy and the soft copy of the consent forms in the Biobank. Owing to this mandate, it becomes necessary to have the consent form in a safe custody under strict vigilance and restricted access area.

Data management also includes one of the important factors which is the segregation of the data. India is yet not well equipped with maintaining systematic Electronic Health Records (EHR) which makes its even more difficult to obtain relevant medical history and other information from the patients. Obtaining a sample from the patient is relatively easier than to obtain a correct medical history, phenomic data and its supporting documents (in the form of clinical reports/ evidences of the pertaining disease). A sample in the Biobank is worthy only if it has its supporting documents and the respective consent form duly signed by the patient. This in total will make the Biobank rich not in only terms of the medical data but also when any kind of data is generated from the results obtained during the study. Lack of systematic EHR is also one of the major concerns, as the Biobank would not

have the longitudinal medical history and reports of the patients. This in total becomes a tedious process to achieve the quality data which could be very useful for the research purposes.

Developing a robust EHR is the need of the hour which will help concentrate the medical history and information about the patient which will help in (LHS) Learning healthcare systems. LHS are characterised by continuous loop of health information and then extracting information for its application in the clinical use. LHS helps in comprehending the patient data shadow and in turn is helpful in personalization of care. The government of India has initiated the ABHA card for a one step, paperless and secured digital health services. This initiative will help record all the medical history, diagnosis and the treatment plan under one number. This will make the EHR in India not only seamless but very convenient to access the medical information in one step.

General Data Protection Regulation (GDPR), a legal framework for the people who reside in the European Union, sets guidelines for the collection and usage of the personal information. GDPR includes Pseudonymization where in the personal data security is based on masking the information with a code which can be reversible. This code is available with a very narrow group of people who can decode to identifiable personal information. Pseudonymization of the samples is performed on those patients on whom long term research is going to be performed.

Data is considered anonymous only when it is not possible to identify the person to whom the data belongs to. Health Insurance Portability and Accountability Act (HIPAA) was passed in the USA in 1996 and based on its guidelines there are two methods described for anonymization namely, Safe Harbor and Expert Determination.

Anonymization and pseudonymization are not the same. Anonymization should be used when the data is not required for the specific purpose and no personal information is going to be required. Pseudonymization should be used when the data is going to be used for a specific purpose and re-identification of the personal details may be required.

#### **1.1.4 Certifications and accreditations: Global and Indian perspective**

Standardising processes in association with sampling, storing of samples and quality control along with good governance form an important pillar in the creation, reliability and sustainability of the Biobank [13] [14]. Quality assurance with good lab practices play a critical role to ensure that the processes followed in the laboratory are reliable and constant. Usually in any molecular or a pathology lab, there are external quality assurance programs or external bodies like College of American Pathologists (CAP) or National Accreditation for laboratories (NABL) to govern the functioning of the laboratories. Accreditations or certifications for a biobank so far are limited or are specific to tumor repositories. Gangadharan. C; et al have mentioned about the certification program by the Canadian Tissue Repository Network, ISO 20387: 2018, CAP. In India, Indian Council of Medical Research (ICMR) has laid guidelines for biobank in “

National Ethical Guidelines for Biomedical and Health Research involving Human Participants” but there are no biobank specific regulatory authorities or certification programs for the biorepositories.

Pensieve Health is one of the first certified biobanks in India to be compliant with the standard requirements of “ISO 20387: 2018 General requirements for Biobanking Standards”. This was achieved through the documentation of all the biobanking processes, demonstration of personnel competence, the stringent control of documents and records and ongoing evaluation of processes and the Quality Management System (QMS) [9]. An independent third party was appointed for the verification and aligning the documents for the ISO certification. The team consisted of the certified ISO auditor who conducted the internal Audit first followed by arranging the external (final) ISO certification audit. Post 1 year of the ISO 20387: 2018 certification, a surveillance audit was conducted and now Pensieve Health has successfully entered the second year with the ISO certification.

#### **1.1.5 Essentials of Biobanking Course**

International Society for Biobanking and Environment Repositories (ISBER) is the global forum that addresses scientific, technical legal and ethical issues for biobanks and biorepositories. ISBER have launched the course “Essentials of Biobanking” which is applicable to all the biobanks, is an online course and is in line with the ISO 20387: 2018 standard.

## **2. Current challenges in the Biobank: Indian Perspective**

Biobanking in India is a newly coined concept and is still in its infancy stage. A Biobank is the library of comprehensive medical information which when put together can pave a path for translational research. In India, there are several small scale biobanks and some are disease specific biobanks. However there are a number of challenges faced in India with respect to biobanks which needs to be addressed.

- **Awareness:** In India there is lack of awareness among the citizens due to which they are reluctant to share their samples and the medical history. There is lack of trust in the people to give their informed consent and share the health information.
- **Financial Support / Funding:** Biobanking is one of the fields where a strong funding and financial support is a must and the Return on Investments is long term.
- **Unavailability of Electronic Health Records:** Developing a robust Electronic Health Record is a must in India to have a one step access to the information. Currently in India, most of the Electronic Health records are limited only to the demographic information and the date of admission and discharge in the hospitals. This does not help in gathering a comprehensive health information of the patient.
- **Dedicated guidelines for Biobanking:** Indian Council for Medical Research (ICMR) currently has laid the basic guidelines for the biobanks in India. However, there is no dedicated body for the governing of the biobanks.

**Table 1:** Top ten biobanks in the world

Sr. No	Name of the Biobank	Type of Samples	Disease / Cohort	Public/ Govt/ Private
1	Biobank Graz - (Austria)	Tissue, DNA, RNA, Whole Blood, Peripheral Blood Cells, Biological Fluids	All Diseases	Public
2	Shanghai Zhangjiang Biobank - (China)	Tissue, DNA, RNA,, Whole Blood, Peripheral Blood Cells, Biological Fluids	All Diseases	Public
3	The International Agency for Research on Cancer (IARC) Biobank (IBB) - WHO	Plasma, Serum, Urine, DNA, RNA	Cancer	Inter Government
4	Helsinki Biobank/Finngen Biobank - (Finland)	Tissue, DNA, RNA,, Whole Blood, Peripheral Blood Cells, Biological Fluids	All disease	Public - Private
5	Medpace Labs Biorepository - (United States)	Tissue, DNA, RNA,, Whole Blood, Peripheral Blood Cells, Biological Fluids	All Disease	Private
6	“All of Us” Biobank (United States)	Whole Blood, DNA, RNA, Saliva, Urine	All Disease	Public
7	Canadian Biosample Repository (Canada)	Serum, Plasma, Buffy coat, Urine	All Disease	Public
8	China Kadoorie Biobank (China)	Whole blood	Non Communicable Diseases	Public
9	UK Biobank (UK)	Blood, Urine, Saliva	All Disease	Public
10	Tissue For Research (BiobankOnline) (UK)	Tissue, DNA, RNA,, Whole Blood, Peripheral Blood Cells,	All Disease	Public

### 3. Summary and Conclusion

Biobanks are the institutions created at the intersection of science, technology and translational research. A Biobank is not only the facility to store the biospecimen and its information but also it extracts the information from the data which helps accelerate finding the solutions for bio diversity, bio evaluation, drug targets for the specific diseases. A Biobank can be well established only with good public support. The willingness and awareness of the people to share their health information and eventually the genetic information can help shape a diagnostic marker for any disease or fasten the process to achieve a drug target. A recent example for this can be taken from the UK Biobank, where data from the bio bank helped the researchers in the US discover a novel gene that protects against obesity and

diabetes. At present they are trying to develop a drug which could mimic the effects of this gene.

At present, the genomic representation of Indians is very low the global level. It is not only important to have the global genomic representation of the Indians but also to address the important illnesses like TB, Cancer, Cardiac, Diabetes, IBD and Chronic Kidney Diseases at the molecular level. There are several disease specific biobanks in India but are at the smaller scale. If all the biobanks join hands and form a consortium, personalized medicine in India would not be a far - fetched target. Personalized medicine does not mean tailor made therapy approach for every individual but for those who fall in a particular category of the genetic pool. As we have increasing evidences from the UK Biobank, data from Indian biobanks could also bring a breakthrough the field of medicine.

**Table 2:** List of Biobanks in India

Sr. No.	Biobank/Institute	Disease	Government/ Private
1	NCTB (Chennai, Tamil Nadu)	Cancer	Government
2	NLDB (New Delhi)	Liver diseases	Government
3	Rajiv Gandhi Cancer Institute and Research Centre (New Delhi)	Cancer	Government
4	NIMHANS (Bangalore, Karnataka)	Brain disorders	Government
5	TiMBR (Kolkata, West Bengal)	Cancer	Government
6	National Cancer Institute - All India Institute of Medical Sciences (Jhajjar, Haryana)	Cancer	Government
7	Sapien Biosciences (Hyderabad, Telangana)	Commercial biobank	Private
8	National repositories for cell lines/hybridomas NCCS (Pune, Maharashtra)		Government
9	ADBS (Bangalore, Karnataka)	Brain disorders	Government
10	Pensieve Health (Mumbai, Maharashtra)	Non - communicable diseases,	Private
11	Global Gene Corp (Mumbai, Maharashtra)	Non - communicable diseases,	Private
12	ACTREC Biorepository	Cancer	Government

### References

- [1] S. A. Jaykumar, “India’s potential influence on the need for global governance of biobanks, ” *Observer Researcher Foundation*, p.5, 2023.
- [2] B. R. R. Thakur, “Epidemiology and Economic Burden of Continuing Challenge of Infectious Diseases in India: Analysis of Socio - Demographic Differentials, ” *Frontiers in Public Health*, 2022.
- [3] S. S. A. M. G. N. N. A. W. C. A. S. R. S. K. J. K. S. S. P. K. & P. J. Josh Schilling, “Development of a decentralized cohort for studying post - acute sequelae of COVID - 19 in India in the Data4life Study, ” *Nature*, 2023.
- [4] “History of TRIFED, ” Tribal Co - Operative Marketing Development Federation of India Limited Ministry of Tribal Affairs, Govt. of India.

- [5] R. R. W. M. M. A. H. G. M. L. a. W. H. Y. Mitra D. Harati, "An introduction to starting a biobank," Vols.10.1007/978 - 1 - 4939 - 8935 - 5\_2. PMID: 30539430; , no.1897: 7 - 16, 2019.
- [6] G. D. M. L. C. S. P. K. Z. Heimo Müller, "Biobanks for life sciences and personalized medicine: importance of standardization, biosafety, biosecurity, and data management," *Science Direct.*, pp.45 - 51, 2020.
- [7] S. W. R. K. V. a. A. S. Charitha Gangadharan, "Biobanking for Translational Diabetes Research in India," *Bio Research*, vol.9.1, no. biores.2019.0052, pp.183 - 189, 2020.
- [8] D. P. G. B. G. S. A. B. G. B. E. M. C. A. P. D. M. S. S. A. D. M. C. V. P. B. P. A. B. M. M. C. Annaratone L and A. C. i. C. (. P. a. B. W. Gro, "Basic principles of biobanking: from biological samples to precision medicine for patients," *Virchows Arch.*, vol.479 (2), no.10.1007/s00428 - 021 - 03151 - 0., pp.233 - 246, 2021.
- [9] S. J. G. D. L. A. M. G. C. Lara Mouttham, "A Biobank's Journey: Implementation of a Quality Management System and Accreditation to ISO 20387," Vols. Jun; 19 (3): 163 - 170, no.10.1089/bio.2020.0068, 2021.
- [10] D. R. Mathur, NATIONAL ETHICAL GUIDELINES FOR BIOMEDICAL AND HEALTH RESEARCH INVOLVING HUMAN PARTICIPANTS, 2017.
- [11] J., M. P. - d. - l. - P.,. A. C. L. L.,. D. v. E. I.,. M. R.,. A. J.,. C.,. S. F. A.,. V. B. - F.,. V. P. V. M.,. D. R. Yllka Kodra, "Recommendations for Improving the Quality of Rare," *International Journal of Environmental Research and Public Health*, p.22, 2018.
- [12] C. L. T. B. G. T. A. B. C. U. L. S. M. Meystrea, "Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress," *IMIA Yearbook of Medical Informatics*, pp.38 - 52, 2017.
- [13] J. Drepper, "Data protection in biobanks from a practical point of view: what must be taken into account during set - up and operation?," *Journal of Laboratory Medicine*, pp.301 - 309, 2019.
- [14] A. K.,. A. S.,. T. B. A. S. A. I. a. A. K. Kristina Malsagova, "Biobanks—A Platform for Scientific and Biomedical Research," *Diagnostics - Basel*, p.21, 2020.



**Simran Keshwani** received her MSc in Genetic Engineering from D Y Patil University. She has been working with Peansieve Health Pvt Ltd since 2019



**Abhishek Prajapati** received his BSc in Chemistry in 2017 and MSc in Nutraceuticals in 2019 from Kishinchand Chellaram College and K J Somaiya College of Science and Commerce respectively. He has been working with Pensieve Health Pvt Ltd since 2019.



**Prashant Khadke** is a PhD from BARC and Post Doctorate from USA. He is Geneticist having more than 15 years' of research and techno commercial expertise. He is expert in Next Generation DNA Sequencing. With more than 15 international publications, worked with major KOLs across country. With his networking skills as a strength Pensieve health started this initiative in India. In his previous experience he worked with leadership role in life sciences and health care multinational companies

## Author Profile



**Suchi Vaidya** received her MSc degree in Biological Sciences in 2013 from NMIMS University. She has been working with molecular biology laboratories and hospital like Agilus, Kokilaben Dhirubhai Ambani Hospital respectively. At present she is with Pensieve Health Pvt Ltd and is persuing her PhD in genetics from Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan, India



**Khushbu Shah** received her MSc degree in Nutraceuticals in 2019 from K J Somaiya College of Science and Commerce. She has been working with Pensieve Health Pvt Ltd since 2019. At present she is persuing her PhD in genetics from Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan, India



**Pooja Utekar** received her MTech Degree in Bioprocess Technology in 2019 from Institute of Chemical Technology. She has been working with Pensieve Health Pvt Ltd since 2019.