PUBLIC HEALTH

Biobanks in Developing Countries: Needs and Feasibility

Technological advances coupled with use of existing resources can be used to create biological repositories that may lead to better health in developing countries.

S. K. Sgaier,[1]* P. Jha,[1]† P. Mony,[1,2] A. Kurpad,[2] V. Lakshmi,[3] R. Kumar,[4] N. K. Ganguly[5]

Over 90% of the global burden of disease is in developing countries, yet only 10% of global research addresses many of these diseases. Between 1975 and 1999, ~1% of new marketed drugs were for tropical diseases and tuberculosis (*1*). Repositories of biological samples linked with medical data from individuals (biobanks) are infrastructures for sustained research on the biological determinants of disease and promise to accelerate the discovery of vaccines, drugs, and diagnostics. However, the distribution and focus of current biobanks suggests that their discoveries will not sufficiently benefit those living in developing countries. Innovative use of recent technological advances and existing infrastructure platforms make biobanks cost-effective and feasible in developing countries.

Biobanks
(with sample sizes of ≥200,000 or in a developing country)

Name of biobank	Size (age group)	Start-up cost* (cost per person)	Chronic disease	Infectious disease
U.K. Biobank	500,000+ (40–69 yr)	$120 million (~$240)	Yes	No
Estonian Genome Project	~ 1 million†	$2.5 million‡ (~$250)	Yes	No
Icelandic deCode Biobank§	~250,000	$212 milion (~$850)	Yes	No
Kadoorie Study of Chronic Diseases in China	500,000 (35–74 yr)	$22 million (~$50)	Yes	Some
The Mexico City Prospective Study	160,000 (35+ yr)	Not available	Yes	Some
The Gambian National DNA Bank	~57,000	$0.6 million‖ (~$15)	No	Yes¶
The Indian National Biobank§,#	~2–3 million (18+ yr)	$20–$30 million (~$10)	Yes	Yes

*Estimated, in U.S. dollars. †This is 75% of the Estonian population. ‡For a pilot project of 10,000 participants. §Familial linkage studies are possible for these two biobanks, and genetic association studies are possible for all biobanks listed. ‖For the first 5 years. ¶Powered to discover correlates of nonfatal TB and/or malaria and not HIV/AIDS, which has low prevalence in The Gambia. #Presently at the design stage.

Biobanks as a Platform

The Human Genome Project, annotation of millions of single-nucleotide polymorphisms (SNPs) within the genome, development of ultrahigh-throughput genotyping, small-molecule detection methods, and powerful software to analyze the mass of data that is generated, now make possible the discovery of the allelic and biological variants that underlie complex diseases (such as cardiovascular diseases, cancer, diabetes, tuberculosis, and AIDS). Common genetic variants likely involve moderate effects, such as a relative risk of 1.2. Reliable assessment of these variants in different populations (including documenting any interactions between genes and other risk factors) requires studies with thousands, or even tens of thousands, of cases and controls.

A few biobanks have already been established in developing countries such as the Chinese Kadoorie study (*2*), the Mexican biobank (*3*) and the Gambian national DNA bank (*4*) (see table, above). The Chinese and Mexican biobanks were designed primarily to discover correlates of noncommunicable diseases in adults over age 35 years, and the Gambian study is relatively small. These studies are not sufficiently representative of the major causes of death and disability in developing countries or of the age groups at which disease strikes. In particular, HIV/AIDS, tuberculosis, and malaria require larger studies in diverse populations.

Disease Burdens in Developing Countries

More than 75% of the 5 million deaths worldwide due to HIV/AIDS, tuberculosis, or malaria are in developing countries. Even infectious diseases have cofactors, which make their acquisition or conversion to clinical disease more likely, such as smoking and tuberculosis (*5*).

Genetic or undiscovered copathogens may help explain the unprecedented increases in HIV-1 in eastern and southern Africa. Natural resistance to HIV-1 appears to be evolving among select populations (*6*) that are constantly being challenged by the virus. Understanding how the immune systems have so evolved to fight these infections can enable new drugs and vaccines. Focused biological research on the genetic and other biological correlates of infectious diseases is under way, most notably by the Grand Challenges in Global Health (*7*). These need to be complemented by more open-ended platforms for unpredictable discoveries, as is possible with larger biological repositories linked to medical data.

Developing countries, unlike most developed countries, suffer from the dual burdens of chronic (chiefly noncommunicable) diseases and infectious diseases (*8*). Already, four out of five chronic disease deaths occur in developing countries. The genetic and environmental variations that contribute to complex chronic diseases are not necessarily the same in geographically segregated populations. Indeed, common chronic diseases can have surprising correlates. A 10-year prospective study in China, for example, found higher risks of vascular deaths among people with excessively low, as well as those with excessively high, body mass even after adjusting for smoking and blood pressure (*9*). Two small studies in India (*10*) and in Iraq (*11*) found that low body mass and diabetes were correlated with tuberculosis history and a positive tuberculin test, respectively.

[1]Centre for Global Health Research, St Michael's Hospital (LKSKI/KRC), University of Toronto, Toronto, ON, M5B1C5, Canada. [2]St. John's Research Institute, Bangalore 560034, India. [3]Nizam's Institute of Medical Sciences, Hyderabad, 500082, India. [4]School of Public Health, Post Graduate Institute of Medical Education and Research, Chandigarh, 160012, India. [5]Indian Council of Medical Research, New Delhi, 110029,India.

*Present address: Howard Hughes Medical Institute, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02115, USA.

†Author for correspondence. E-mail: Prabhat.jha@utoronto.ca

Published by AAAS

A Way Forward

Typical biobanks are expensive partially because the serum samples that they collect and keep must be kept cold continuously. Dried blood spots (DBSs), in contrast, do not require refrigeration during collection or transport (*12*). DBS samples can be easily collected and safely transported by regular mail. The higher acceptability by participants of DBS versus whole-blood collection, lower costs, and ease of handling also enable much larger sample sizes to be achieved within a given budget.

MDS health surveyor collecting a DBS sample in the field.

Despite their small volume, DBS samples have been increasingly used for molecular, enzymatic, immunological, biochemical, and hematopoietic analyses. SNPs within the whole genome can now be reliably assessed from DBSs (*13*). The continuous decline in the sample volume and the cost of SNPs and the development of high-throughput analyses, mean that biobanks based on DBSs are becoming economical, as well as scientifically practicable. The major roadblock is getting reliable epidemiological evidence about the relevance of variables measured to the development of disease. Appropriately large-scale epidemiological fieldwork in developing countries to acquire blood samples systematically linked to relevant measures of disability and future mortality is crucial. Biobanks in Western countries use their national health systems, with physicians collecting samples and medical data for their patients. In developing countries, however, fewer people have access to medical care, and linkage to routine health care is not yet possible.

Several developing countries have established disease and mortality surveillance systems, which could be cost-effective platforms for biobanks. A good example is the antenatal clinic HIV surveillance system recommended by the World Health Organization to track changes in HIV prevalence among pregnant women in 132 countries. This widely implemented system is not used adequately to understand the transmission and correlates of HIV infection, primarily because collected samples are usually discarded after HIV testing. Modest enhancement of this system with additional demographic and medical information, as well as reliable archiving of samples, would provide a widely practicable resource to investigate the biological correlates of HIV. Other examples of established surveillance systems include surveys of malaria parasites and the INDEPTH network of 37 demographic surveillance sites—26 of which are in Africa (*14*).

India: A Case Study

Over the past 30 years, the Indian government has built a population-monitoring framework called the Sample Registration System—a nationally representative sample of 7.6 million people in 1.3 million households across the country. The "Million Death Study" (MDS) (*15*) is under way within this system. The Indian Council of Medical Research, Registrar General of India, and the University of Toronto are collaborating to explore the logistics of building a national Indian biobank to assess the underlying risk factors and correlates of disease in India. The MDS is unique because it collects data that will allow both family-based genetic studies and case-control association studies. The design and scale should ensure that for each complex disease of interest, several thousand cases and controls (efficiently tested through "nested" case-control design) will be available for association studies with sufficient statistical power to detect modest but medically relevant associations.

We are exploring methodological and logistical issues of building a national biobank in India. Medical and family history, blood pressure, body mass, smoking and alcohol use, other variables, and DBSs were collected in household surveys in 2006 from 2700 adults aged 18 years or older in six districts in three states (see figure, left). Only 5% of the people interviewed refused to give DBSs, in contrast to refusal rates of nearly 40% in a similar study collecting a serum sample (*16*). The likely costs of a DBS-based Indian biobank would be at least 1/20th the cost per person of the ongoing serum-based U.K. Biobank (see table, p. 1074).

Toward a Global Consortium of Biobanks

National or regional biobanks are a first step. Others have called for a global consortium of biobanks to address common ethical issues, data ownership, and data sharing (*17*, *18*). We see these consortia arising for two reasons. First, joint analyses of important, but uncommon, gene variants will be needed to generate more definitive results than can be generated from individual (and likely underpowered) studies. Second, reasonable expectations from funders and beneficiaries will push toward collaboration, as has happened with the Human Genome Project (*19*) and the Global HIV Vaccine Enterprise (*20*). The promise is enormous: accessible and affordable studies in diverse populations to permit imaginative search for common and rare genetic and other biological correlates of global diseases.

References and Notes

1. P. Trouiller *et al.*, *Lancet* **359**, 2188 (2002).
2. Z. Chen *et al.*, *Int. J. Epidemiol.* **34**, 1243 (2005).
3. R. Tapia-Conyer *et al.*, *Int. J. Epidemiol.* **35**, 243 (2006).
4. G. Sirugo *et al.*, *Nat. Genet.* **36**, 785 (2004).
5. V. Gajalakshmi, R. Peto, T. S. Kanaka, P. Jha, *Lancet* **362**, 507 (2003).
6. J. L. Heeney, A. G. Dalgleish, R. A. Weiss, *Science* **313**, 462 (2006).
7. H. Varmus *et al.*, *Science* **302**, 398 (2003).
8. C. D. Mathers, D. Loncar, *PLoS Med.* **3**, e442 (2006).
9. Z. Chen *et al.*, *Int. J. Epidemiol.* **35**, 141 (2006).
10. N. Shetty, M. Shemko, M. Vaz, G. D'Souza, *Int. J. Tuberc. Lung. Dis.* **10**, 80 (2006).
11. W. Al Kubaisy, A. Al Dulayme, D. S. Hashim, *East Mediterr. Health J.* **9**, 675 (2003).
12. K. Steinberg *et al.*, *Epidemiology* **13**, 246 (2002).
13. R. A. Paynter *et al.*, *Cancer Epidemiol. Biomarkers Prev.* **15**, 2533 (2006).
14. INDEPTH Network, www.indepth-network.org/dss_site_profiles/dss_sites.htm
15. P. Jha *et al.*, *PLoS Med.* **3**, e18 (2006).
16. Prospective Urban and Rural Epidemiological Study, www.ccc.mcmaster.ca/pure/index.html.
17. H. E. Hagen, J. Carlstedt-Duke, *Nat. Med.* **10**, 665 (2004).
18. A. Cambon-Thomsen, *Nat. Rev. Genet.* **5**, 866 (2004).
19. Human Genome Project, www.ornl.gov/sci/techresources/Human_Genome/home.shtml.
20. Global HIV Vaccine Enterprise, www.hivvaccineenterprise.org/

10.1126/science.1149157