The Promising Future of Biobanks: Building a Global Perspective

Jordi Camí
Institut Municipal d'Investigació Mèdica (IMIM)
and Universitat Pompeu Fabra
Barcelona, Spain

Jaume Bertranpetit
Universitat Pompeu Fabra
Barcelona, Spain

BIOBANKS ARE LARGE collections of biological specimens (tissue and blood samples) and individual information intended for increasing knowledge and research. Recent advances in basic genomic science and technology have been crucial for opening the possibility of obtaining genetic information in a faster, cheaper and easier manner. These advances make feasible the analysis of many parts of the genome in large collections of DNA from individuals for which there were careful descriptions of their health history and life events.

Biobanks have recently been developed under the auspices of public or private initiatives as suitable tools for biomedical research purposes, but should be used for in-depth investigation of biological biodiversity of human beings. As shown below, biobanking is a scientific and a social endeavour.

1. BACKGROUND: THE EXPERIENCE FROM STUDIES BASED ON LARGE COLLECTION OF SAMPLES

We know that many lives are saved thanks to studies based on large collections of samples. The classical Framingham study in the 1960s showed that people with elevated cholesterol were more likely to develop heart disease and later work clarified which lipoprotein fractions were impor-

tant. Now we have drugs that, in some groups, reduce cholesterol and the risk for ischemic heart disease by 30%. Using archived autopsied tissue from 198 soldiers, it was possible to sequence almost 10% of the genome of the influenza virus that, in 1918, killed over 20 million people (Murray, 2004; UK Biobank, 2004). These are only a few examples. Most of such findings in the literature came from stored materials preserved without prior knowledge of their interest in future analysis. It is now time to establish storage and information frameworks for collected samples, because a better understanding of our evolutionary past can shed light on patterns of disease today.

2. THE VALUE OF BIOLOGICAL SPECIMENS AND MEDICAL INFORMATION FOR MEDICAL RESEARCH

There is a large prospective literature in relation to the opportunities offered by genomics for biomedical research purposes (Collins and McKusick, 2001). A common denominator is that our future medicine based on the genomic science will be more predictive and will have the opportunity to be used more individualised.

In fact, main examples of specific research studies requiring the collection of large DNA samples with detailed medical information are as follows (GeneWatch UK Biobanks, 2004). *Linkage studies* are researches on families directed to identify the genes or genome regions associated with specific disorders usually as causative in the form of DNA sequence anomalies. *Association studies of diseases* are based on the differences in allele frequency between affected individuals and controls, usually using Single Nucleotide Polymorphisms (SNPs) that have been previously identified. This research looks for the complementary or partial genetic contribution in common diseases. Linkage and association studies serve to unravel the genetic bases of diseases. *Genetic epidemiology studies* are aimed to determine the interactions between genes and the environment and their role in disease, as would be the case of complex diseases in families or specific groups who have been exposed to toxic chemicals as part of their work. *Pharmacogenetics* is a knowledge based on the genetic basis of human drug metabolism. Adverse drug reactions (ADR) are very common and constitute a health challenge as well a deterrent in drug commercialization. The availability of appropriate individual genetic information will make pos-

sible to identify the people susceptible to an ADR, so it is envisaged that in the future pharmacogenetic studies will lead to a new era of "personalised" drug therapeutics.

3. "REVERSE GENETICS" STUDIES AS THE NEXT PROMISING APPROACH

Reverse genetics represented a main step forward in biomedical research as nothing had to be known to unravel the genetic and molecular defect for a given genetic disease: from the disease to the molecular pathology base in one step thanks to the genetic analysis. Intervention could then be designed with a clear target. This has been a powerful tool for Mendelian conditions that fails with complex and common diseases. A naïve view created expectations for a rapid dissection of complex diseases, but the success has become more and more rare and even if we hear about some of the main achievements, they only correspond to small fractions of patients having very specific features in which there is a differential susceptibility.

There is a consensus regarding the need to change the approach of genomic research for medical purposes. The direct bridge "genome for health" that was established some years ago with the genetic revolution (mainly through reverse genetics) is much less exploited than expected and promised. The output of genomics in health cannot wait the wisdom path of going first from genomics to biology and then applying it into health. Besides the scientific interest of having first a detailed understanding of the molecular basis of life to be able to fully understand disease and intervention (e.g., drugs, vaccines), a more direct path is needed.

4. THE NEED OF LARGEN COLLECTIONS OF SAMPLES FROM HUMAN POPULATIONS THAT INCLUDE EXTENSIVE CLINICAL INFORMATION

This contemporary approach means that biobanks with medical and associated personal information are extremely valuable, turning into a key tool for the progress and advance of the biomedical knowledge. In summary, biobanks should be able to provide the materials on which genetic and epidemiological approaches would be successful in the study of health determinants. In deep, biobanks should help to determine the biological

bases of the participant's health status (phenotypes), understanding the complex interactions between genes and environment (including nutrition, and environmental exposures).

In recent years, a number of national and multinational efforts have led to the launching of population-based biobanks and longitudinal cohort studies to be used for large-scale genetic research. However, consensus in the scientific and medical world holds that there are major issues to be addressed prior to transferring new genetic discoveries to the health care system. These issues range from the essential step of validating the initial finding in other populations, to the dissection of the genetic effect in relation to disease manifestations as well as other cofactors, such as environment. Last but not least, the challenge is in what extent the materials collected today will be useful and interesting for the questions that will be asked in the future.

At this point, it is time to present a brief description of the current large initiatives in biobanking as well as other related projects.

The UK Biobank
http://www.ukbiobank.ac.uk/

The UK Biobank aims to study the separate and combined effects of genetic and non genetic risk factors in the development of multifactorial diseases of adult life. The pilot study will start along 2004 and try to involve half a million people of 45–69 years of age. The project is developed under a tripartite collaboration between the Wellcome Trust, the Medical Research Council, and the Department of Health. It has an allocated budget of £61 million. The selected infrastructure for the biobank is a charitable company limited under guarantee, UK Biobank Ltd, with a Board of Directors. It is centrally coordinated by a Host Coordinating Centre based in Manchester. This Centre will also coordinate the activities of the six scientific Regional Collaborating Centres, each a consortium of academic and research institutions responsible locally for recruitment and collection of data and samples. UK Biobank is provided of a Scientific Committee and an independent Ethics and Governance Council.

In addition, the "GeneWatch UK Biobanks Database" project is aimed to give descriptions of some of the biobanks that have been or are being compiled in the UK. The database mainly includes only DNA collections

where genetic information is linked to personal medical and/or lifestyle data, and also includes the DNA collection of the Sanger Centre—which is part of the Human Genome Project (GeneWatch UK Biobanks Database, 2004).

GenomeEUTwin
http://www.genomeutwin.org/

GenomEUtwin is a project aiming to analyze twin and general cohorts in order to determine the influence of genetic and non genetic factors on five traits: obesity, stature, coronary heart disease, stroke, and longevity, and to create synergies in genetic epidemiology. The implementation is coordinated by the Finland National Public Health Institute and University of Helsinki and it builds on existing twin cohorts from seven European countries (Finland, Sweden, Denmark, the Netherlands, Italy, and UK) and Australia. The project has received funding from the European Union for four years under its 5th Framework program "Quality of Life and Management of Living Resources." To this day 600 000 samples have been collected, accumulated by the various twin cohort studies undertaken in Europe. The first stage of the project consists of genotyping 10 000 samples.

One branch of this project is the MORGAM population cohort, a multinational collaborative study to explore the relationship between the development of cardiovascular diseases and risk factors, including genetic determinants.

The Estonian Genome Project
http://www.geenivaramu.ee/

The Estonian Genome Project aims not only to enable research on the genetic and non genetic components of common diseases, but also to create biological descriptions of a large and representative sample of the Estonian population. It aims to collect the data of up to one million people.

Following the adoption of the *Human Genes Research Act* in 2000, the Estonian Genome Project Foundation (EGPF), responsible for the database coordination, was established in March 2001 with the aim of coordinating the creation of a central Gene Bank database of health

and genetic data of the Estonian population. The project is based on a public and private partnership between the EGPF, owner of the database and privacy shelter, and Egeen (a public limited company), the exclusive commercial licensee of the database responsible for the financing of the project.

A pilot phase of the project (October 2002–February 2003), funded by foreign venture capital funds and private investors, has been completed in three counties and the main project has been launched in twelve counties. On January 2004, the Estonian Government has allocated funds for the development of the database and to set up a special working group to ensure the projects' compliance with the *Human Genes Research Act*. As the end of December 2003, the bank contained samples from approximately 10 000 donors from the 30-50 age groups.

The Cartagene Project
http://www.cartagene.qc.ca/

The CARTaGENE project aims to study genetic variation in Quebec's population in order to discover new susceptibility genes to tailor Quebec health care system to the needs of its population. It takes advantage of the founding effect of Quebec population as well as their detailed demographic and genealogical data.

It has been developed in Quebec since 1999 by the multidisciplinary team of the Quebec Network of Applied Genetic Medicine. The initial project has subsequently been modified with the team's decision, in November 2003, to opt for a semi-longitudinal study and the creation of a database of double-coded information. The project will be governed by the Institute for Populations and Genetics (IPEG), a not-for-profit organization, incorporated in December 2003, and ethically overseen by an independent Ethics Review Board. The IPEG was recently created to control and manage, in the public interest, major population projects in genetics, genomics and proteomics, including the CARTaGENE project.

The team is currently seeking financial support. An ongoing public engagement phase will be followed by the collection of 60 000 individuals aged 24 to 75 randomly selected from the Health Insurance (RAMQ) registry over a period of 4 years.

The Swedish National Biobanking Program
http://www.biobanks.se

The 2002 Swedish Act on Biobanks is a recent healthcare law that regulates how human biological samples can be stored and used, and also determines the quality and security demands of biobanks. The act defines the concept ”Biobank” as ”biological material from one or several human beings collected and stored indefinitely or for a specified time and whose origin can be traced to the human or humans from whom it originates”. This definition do not difference between samples collected for research or during routine medical care provided that the biological material is identifiable, coded or uncoded. According to this definition, the total number of samples in the biobanks of the Swedish Health Care system is estimated to about 50-100 million human samples, increasing with about 3–4 million samples per year.

Based on the opportunities of large biobanks existing in Sweden from routine medical health care, and further complementary comprehensive registries with health care information, mortality and genealogical data a “National Biobank Program” has been launched to establish a system for quality assurance, compile information, and increase knowledge on the usability of Swedish biobank samples, among other objectives. The program is the result of joining two Swedish investment initiatives on functional genomics, Swegene and the Wallenberg Consortium North.

deCODE genetics in Iceland
http://www.decode.com

“deCODE genetics” is a private company headquartered in Reykjavik. After a huge discussion in the parliament and the whole population of Iceland between 1998 and 2000, Decode received a licence to build and run the Icelandic Health Sector Database. At present Decode has created a biobank of genetic samples from 100 000 Icelandic volunteers. Decode is linking these data with information from the Icelandic Health Sector Database and from public genealogical records and takes advantage of the founding effect of the initial Icelandic population. Already, it has mapped genes involved in more than 50 common diseases and identified several specific disease-causing genes, including ones behind schizophrenia, stroke, and osteoporosis.

Decode company tries to apply its discoveries to develop new drugs and DNA-based diagnostics that target the underlying biology of disease. The Decode initiative has raised questions about informed consent and the ability to protect confidentiality and patient rights.

5. THE PUBLIC POPULATION PROJECT IN GENOMICS (P3G)

In 2003, four different but complementary genomics research projects involving whole populations, i.e., Quebec's CARTaGENE, GenomEUtwin project (involving eight countries), Estonia's genome project and the UK Biobank, came together to create an international consortium called P3G: *Public Population Project in Genomics* (http://www.p3gconsortium.org/index.cfm). A parallel effort to P3G occurred in Europe, under the auspices of COGENE coordinated from Norway on behalf of the Forum of Genomes Program Managers with representatives from 25 European countries. Given the similar natures and overlapping memberships of P3G and the European initiative COGENE, leaders of both projects chose to join forces at a P3G meeting in Manchester in December 2003.

P3G is envisaged as a resource, service and infrastructure for existing and future partners involved in large-scale population genomics projects. It is recognized that specific projects will have differences in terms of study design, governance, and outcomes and is expected that there will be many common elements (such as phenotypes, samples storage, genetic tests, governance, and consent issues) that can be shared in order to increase the knowledge created by these projects. Ultimately, P3G will foster international collaboration and public access to population genomics data, according to prevailing ethical and legal norms. At the current time (January 2004) (Knoppers, unpublished data), legal advice is being sought on the merits and process in regards to incorporation as a not-for-profit organization.

The overall objectives of the consortium are the following:

- To provide necessary coordination, harmonization and standardization for data collection, production and storage to foster international collaboration, advance science and maximize public health benefits.
- To develop common understanding of the socioethical and legal issues.

- To foster a deeper understanding of the relative contribution of genetic and non genetic determinants to health and disease.
- To transfer this knowledge to the international community so as to optimize benefits for public health care worldwide, with the aim to provide a resource for data sharing between public population genomics projects.

P3G initiative should be seriously considered as one of the best positioned platforms to achieve the new challenges of biobanking as a scientific and a social endeavour. In forthcoming initiatives P3G may play a key role providing standards and technology transfer.

6. CURRENT CHALLENGES OF BIOBANKING: BIOETHICAL AND PROPERTY ISSUES

The key legal and property issues, privacy and bioethical aspects as well as methodological issues in research with human biological materials are a matter of huge discussion.[1] In particular, the debate focuses on questions related to the consent modalities and contents for archived tissue specimens, including the specific problem of sharing data with private biobanks and the problems of accountability for participants and conflicts of interests for the institutions (Winickoff and Winickoff, 2003). How should we balance the concerns for confidentiality, privacy and the possibility of discrimination with the desire to do the best possible science? How anonymous samples are "anonymised"? How to manage the consequences and legitimacy of commercialization of information? When human tissue, blood, and medical data are already being collected and sold, it is indispensable to balance the property rights and the benefits to the community. In this respect, benefit to the population has become one of the critical issues in determining the ethical justification for the study itself, and sharing benefits with the population is critical in preventing exploitation (National Research Council, 1997).

Bioethical and legal aspects of biobanking should be a matter of intercultural and interregional discussion. The present proposal for a worldwide network of "reference biobanks" will be useless in the absence of an

[1]See the last report of the Nuffield Council on Bioethics related to Genetic Screening: Ethical Issues. Sept 2003. http://www.nuffieldbioethics.org/home/

international standardization of the ethical requirements for the control and access of DNA samples and personal data, an harmonization pursued since 1998 by the International Ethics Committee of the Human Genome Organization (HUGO).[2]

7. ADDED VALUE OF BIOBANKS FOR HEALTH: AN EVOLUTIONARY PERSPECTIVE

Beyond the fundamental role of population genomics for its implication for health, there are other aspects that may bridge the existing gap between genome and society. These include the social implications of genetic diversity, exploring how individuals and communities can take advantage of the biological diversity knowledge and assessing the differences in health status in communities with different genomes and environments. Evolutionary biology studying the processes of diversification and adaptation of individuals, populations, species, or ecological systems may provide the clue for a better understanding of our humanistic meaning, the structure of its biological basis and the process that gave raise to humanity.

The human genome sequence, for example, is a clear example of an evolutionary process. The coding elements of the genome for the synthesis of proteins account for only 1.5%. More than 50% is formed by elements that are repeated once or multiple times, frequently inactivated by random mutations, persisting as inert remnants in our genome, but as valuable information on our origin, our ancestors, as well as our personal and familiar lineage.

For the advance of knowledge from an evolutionary perspective, future scientific efforts will allow both understanding the dynamics of the genome and the meaning of its different parts with its functional and historical constraints. In this context, biobanking providing information on human diversity takes a special singularity. If we would like to have an integral and integrated knowledge of our species, the development of biobanks in different parts of the world seems outstanding to get overall data from populations around the world. At any rate, to develop *a global project* with universal ambitions it will be necessary to

[2]Statement on DNA sampling: control and access. HUGO Ethics Committee. http://www.gene.ucl.ac.uk/hugo/sampling.html

add to the present challenges, the peculiarities derived from the perception of cultural diversity in the collection of biological samples for genetic studies. Therefore, a cross-cultural consensus to promote donor participation and to stimulate research that will benefit the humanity appears as a matter of thinking and debate. So we need to move forward a global perspective in the requirements and obligations for good research governance in biobanking.

8. FUTURE CHALLENGES

It is time to encourage a global discussion to promote a future culture where each human being considers his/her genetic information as a public good of interest for future generations. Every genome is a variation on "the" human genome that is ready to generate the humans of the future. The desired participation of large populations will have to take care of personal aspects of the participation, but also to have a pedagogical dimension to make individual's participation active and engaged.

According to the OECD definition, biobanks are a type of "Biological resource centres," an essential part of the infrastructure underpinning life sciences and biotechnology (Biological Resources Centres, 2001). But in agreement with the Global Forum on Health Research, the existing 10/90 gap related to the worldwide amount spent in health research versus the world's health problems should not be worse by excluding developing countries from the potential benefits of the revolutionary advances in genomics (Pang, 2002) . Any proposal for a worldwide network of "reference biobanks" should not be built if it does not involve new large collections of other world regions beyond the most developed ones. A smart management of these challenges will enforce the possibilities of merging samples, methods, and results as well as the desired global harmonization of bioethical and property issues.

Specific granting initiatives and collaborative efforts should be provided to facilitate the development of new biobanks under the sovereignty of the countries and the cultural complicity of its inhabitants. An international coalition should be promoted to prevent less developed countries or less funded laboratories to lay behind the mainstream of biomedical research and to suffer from abuses, such as "take the blood and run."

CONCLUSIONS

Biobanks are novel and valuable tools to advance in the health research determinants and the complex interactions between genes and environment.

Biobanks should also be envisaged as novel tools for research in biological diversity. So biobanking is a scientific and a social endeavour.

Building biobanks is a balancing act between enabling research in a novel perspective, protecting information, and winning public support.

Efforts should continue for a cross-cultural consensus in biobank development. International standardization of methodological, ethical and property management of collected DNA samples and personal data is mandatory.

Looking for a global benefit, specific efforts should be provided to involve less developed countries in biobanking endeavour.

REFERENCES

Biological Resource Centres. OECD 2001.

Collins, F. S., and McKusick, V. A. (2001). Implications of the Human Genome Project for Medical Science. JAMA *285,* 540–544.

GeneWatch UK Biobanks Database. http://www.genewatch.org/HumanGen/default.htm

Murray, T. H. (2004). Ethical Challenges in Research with Human Biological Materials. The Online Ethics Center for Engineering and Science at Case Western Reserve University (Onlineethics.org) http://onlineethics.org/reseth/mod/biores.html

National Research Council. Committee on Genome Diversity (1997). Evaluating human genetic diversity. Washington, D.C.: National Academy Press.

Pang, T. (2002). Genomics and Health Research. *In* 10/90 Report on Health Research 2001–2002. Global Forum for Health Research. http://www.globalforumhealth.org/pages/index.asp

Winickoff, D. E., and Winickoff, R. N. (2003). The Charitable Trust as a Model for Genomic Biobanks. New Engl. J. Med. *349,* 1180–1184.