

nature genetics

volume 33 no. 3 march 2003

Bankable assets?

Despite concerns over how to ensure the privacy of genetic information and its fair use by researchers (see *Nat. Genet.* 3, 195–196 (2001)), population databases that combine genetic and other medical information are becoming more commonplace and increasingly ambitious in scope. The latest example is the UK's Biobank, a £45 (\$73)-million project that organizers hope will make it possible to explore the interplay of genes and environmental influences in the pathogenesis of common diseases.

Starting next year, up to half a million healthy Britons aged 45–69 will be asked to contribute a blood sample from which DNA will be isolated, subject themselves to a brief medical examination and fill out a questionnaire about lifestyle details and medical history. The information will end up in a huge database and will be tracked, over a period of at least ten years, against the participants' medical records through the national health care system. Since study participants will be middle-aged people, in a relatively short time many of them will have developed some of the more common (and serious) diseases. By 2014 over 11,000 of the study's participants are expected to have developed diabetes mellitus; 8,000 myocardial infarction; and 6,000, 5,500 and 3,000, respectively, breast, colorectal and prostate cancer.

The project sponsors—The Wellcome Trust, the Medical Research Council and the Department of Health—will make information in the database, coded in a way to protect the confidentiality of individual participants, freely available to researchers, including those in the pharmaceutical industry, thereby providing an unprecedented resource for identifying risk factors and mechanisms for disease. The ultimate promise is that physicians will one day be able to provide patients with personalized treatment and prevention plans based on this information.

The concept of a genetic database is not new. For the last 30 years, physicians and patient advocates have organized registers of patients with specific genetic diseases, which have been extremely useful for finding disease-associated genes. But large-scale genetic databases are a relatively new trend. The Icelandic Health Sector Database initiative, announced in 1998, was the first in this arena. It aims to link health records of most Icelanders (about 280,000 individuals) with genealogical information and information about genotype. The Icelandic government's decisions to allow a single company, deCODE Genetics, to create the database and receive potential profits from discoveries, and to use anonymous patient health records unless the patients specifically request otherwise (presumed consent), were greatly criticized by scientists and patient advocates alike. Since then, 7% of Icelanders have opted not to be included in the project. Perhaps taking a lesson from the controversy, Biobank organizers have taken care to obtain informed consent from all participants and to hold information

and samples in the public domain. They will also establish an independent oversight body to monitor responsible use of the samples and data collected.

The Icelandic project has not yet started uploading medical records into the database, but by compiling genetic data on a subset of individuals with specific diseases, deCODE has already reported the possible locations of genes associated with these diseases. The unique aspect of Biobank, aside from its unprecedented size, is that the database will contain detailed information about lifestyle and environmental exposures in addition to information about genotypes and medical records.

Other countries have plans to prepare large-scale genetic databases. For example, the Estonian Genome Project plans to collect genotype data for at least three quarters of the population (1.4 million) and combine them with clinical histories and genealogical information. In the United States a national database similar to Biobank or the Icelandic database is not feasible because the country lacks a national health care system; several genetic databases are, however, being planned by individual health-care providers. As one example, the Mayo Clinic plans to assemble a comprehensive database of medical records, demographic data, physician notes and other information on some 4 million individuals. Perhaps one of the more creative approaches to the genetic database idea surfaced about two-and-a-half years ago, when DNA Sciences, a start-up company in Mountain View, California, recruited people to donate their DNA and medical histories through the Internet.

Although many researchers expect that Biobank will yield useful information, opinion is split as to whether its cost will be justified by the returns and what the optimal design of the database should be. Perhaps the most serious concern regarding database design is that the initial examinations and follow-up of individuals in Biobank will be done by physicians scattered across the UK, resulting in inconsistent diagnoses. In contrast, investigators associated with the Framingham Heart Study, an ongoing epidemiological study that is now collecting DNA and other information from the grandchildren of the original cohort, track participants by taking quantitative measurements of cardiovascular disease and its risk factors every couple of years. Will a clinical diagnosis and a health record be sufficiently detailed?

Despite these concerns, large population databases will be necessary to understand the interaction between genes and environment for common diseases, where the contribution of individual genotypes and exposure might be very small and where some combinations of genotypes and exposures might have beneficial effects and others detrimental ones. How best to design these studies remains to be seen, but the initial results from the Biobank should at the very least prove useful in informing the design of future databases. It also remains to be seen how concerns about potential abuses of the information contained in genetic databases will play out. A recent report by the World Health Organization on genomics and world health (see *Nat. Genet.* 32, 213–214 (2002)) gave voice to the major concerns:

“It is not certain whether individuals who donate DNA samples for these databases are fully aware of the potential risks involved, and it is even less clear whether some of the arrangements that have been made with the private sector, which is becoming increasingly involved in these enterprises, are adequately controlled. It is also not apparent how information, particularly unexpected findings, will be handled in these large population studies and how these DNA samples will be used above and beyond the stated aims of those who are establishing the databases.”

Biobank organizers should address these concerns carefully and ensure the proper safeguards are in place. Ongoing public support will be critical to the success of Biobank and similar projects.

