

Research Commentary

The Estonian Genome Project

Andres Metspalu*

Estonian Genome Project Foundation, University of Tartu, Estonian Biocentre, Tartu, Estonia

Strategy, Management and Health Policy				
Enabling Technology, Genomics, Proteomics	Preclinical Research	Preclinical Development Toxicology, Formulation Drug Delivery, Pharmacokinetics	Clinical Development Phases I-III Regulatory, Quality, Manufacturing	Postmarketing Phase IV

ABSTRACT The Estonian Genome Project (EGP) is a large population-based databank that was established with health records and biological samples from a large portion of the population for use in biomedical and genetic research to improve the future of public health care in Estonia. A nonprofit foundation, the Estonian Genome Foundation presented the EGP to the Estonian government in June 2000 leading to a new legislative act, the "Human Genes Research Act," that provides a road map for future gene-related activities and guidelines for the oversight of the databank via a Supervisory Board, Ethics Committee, and Scientific Advisory Board. Unlike other gene discovery efforts, participants in population-based projects are not selected by specific disease type but rather via a random sampling process. The unbiased nature of the recruitment process provides a more accurate measure of the disease risk provided by particular genetic variants and is anticipated to collect 100,000 samples by the end of 2007. The EGP is part of the first international consortia, P3G, set up between Cart@gene from Canada and GenomEUtwin, a FP5-funded project coordinated by the University of Helsinki, Finland. With Estonia joining the European Union in 2004, the EGP will also function to build scientific excellence in Estonia. Drug Dev. Res. 62:97–101, 2004. © 2004 Wiley-Liss, Inc.

Key words: biobank; population genetics; health records

WHAT IS THE ESTONIAN GENOME PROJECT (EGP) AND HOW WAS IT BORN?

It was clear in 1999 that the International Human Genome Project, which was in competition with the Celera Genomics project, would be finished ahead of schedule, that a SNP database would be produced, and eventually a human haplotype map would be constructed using existing genotyping technologies. Yet it became evident that a more individualized approach would be needed for future health care research and that the genetic variability of individuals would be used in both drug development (e.g., better selection of the patients for drug trials) and drug therapy (responders vs. nonresponders and ADR). Additionally, deCODE Genetics reached a 5-year commercial agreement with Roche in 1998 demonstrating the interest from "Big Pharma" in using large-scale population-based genetic research as an additional tool for new drug discovery.

In view of these events, a large population-based databank was established in Estonia with health records and biological samples from a large portion of the population for use in biomedical and genetic research plus future public health care. Initiation of the process involved a small group of Estonian scientists, physicians, entrepreneurs, and politicians incorporated into a nonprofit foundation, "Estonian Genome Foundation" (EGF). This group prepared a first version of

Grant sponsor: Targeted Funding by EMRE; Grant number: 0182582s03.

*Correspondence to: Andres Metspalu, IMCB of University of Tartu, Molecular Diagnostic Centre of University of Tartu Clinicum and Estonian Biocentre, Tartu, 23 Riia Street, Estonia 51010. E-mail: andres@ebc.ee

Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/ddr.10371

the EGP documentation, and presented it to the Estonian government in June 2000. The government supported the idea and one of the first steps was to ensure a firm legal and ethical environment for the EGP.

A new legislative act, the “Human Genes Research Act” (HGRA), was passed by the Estonian Parliament (Riigikogu) in December 2000. This gave the road map for future activities and control of the proposed database. The next step was to incorporate the new nonprofit organization, the “Estonian Genome Project Foundation” (EGPF) in March 2000 with the sole task of making the EGP a reality.

The main objectives of the EGPF are:

- promoting the development of genetic research
- collecting health and genetic data of the Estonian population
- using the results of genetic research to develop new products and services in order to improve public health

Contrary to the usual gene discovery study design, the participants in population-based projects are not selected by having a specific disease but rather via a random sampling process. The unbiased nature of the recruitment process provides a more accurate measure of the disease risk provided by particular genetic variants. There are many ways to build biobanks, but in the Estonian case we have a longitudinal population-based cohort where clinical data can be added over certain intervals in the future. This size makes a difference, and according to current plans we intend to collect 100,000 samples by the end of 2007. The HGRA oversees the Supervisory Board, Ethics Committee and Scientific Advisory Board among many other issues of the EGP.

Recently, biobanking has become more popular and several new initiatives are surfacing [Kaiser, 2002] including the first international consortia, P3G, as one set up between the Estonian Genome Project, Cart@gene from Canada, and GenomEUtwin, a FP5-funded project coordinated by the University of Helsinki, Finland. England, Japan, Australia, Spain, and other countries are also developing biobanks.

Finally, while several approaches will be needed for discovering disease-susceptibility genes, there is an increasing understanding that the primary resources needed for addressing the key question, e.g., how genetic variation influences phenotypic variation, will be large collections of samples from

human populations that include extensive clinical and environmental information together with comprehensive data on clinical, educational, lifestyle, and pedigree aspects.

As Collins et al. [2003] noted: “...a large longitudinal population-based cohort study, with collection of extensive clinical information and ongoing follow-up, would be profoundly valuable to the study of all common diseases.” This is what the EGP is all about.

WHY ARE WE DOING THE EGP?

Estonia joined the European Union in May of 2004. Science is international as is the competition in biomapping, and only excellence counts in doing science. One has to be best in at least one area in order to compete and collaborate with the whole of Europe (and the world) to have a chance of making important discoveries in human molecular, clinical, and population genetics. Biomedical research has, thus, been selected as one of three priorities of Estonian basic science activities. The EGP is an initiative for Estonia to compete in the European research arena in addition to promoting the emerging, nascent biotechnology industry in the country. This will allow us in coming years to integrate genetic education, extensive new databases, genotyping technology, and individual genetic variations into everyday health care. In 2002, an agreement was reached between the EGPF and EGeen, Inc. (U.S.) to fund the EGP and to date about €5M has been allocated. From December 2004 this agreement is changed and EGP is now a fully public project without any exclusive contracts and is seeking public funding.

WHAT DO WE EXPECT FROM THE EGP?

Firstly, we expect to attract more funding for competitive research and development including investments into the private sector. With the EGP database and high throughput genotyping technology, we will be able to conduct large-scale association studies in order to associate gene and environmental factors to complex disease phenotypes (cardiovascular and neuropsychiatric diseases, cancer, etc.). This is a long-term goal of more than a decade. Such information will be useful in developing new diagnostic tests and therapeutics. A nearer term, 5-year goal is to conduct pharmacogenetic studies for new and existing drugs and basic research in clinical and human genetics. This will be done via collaborative studies.

Secondly, we hope to reverse the westward (and northbound) “brain-drain” trend of scientists’ migration out of Estonia. Already, we have had initial positive

signs of repatriation helped by both government programs and by the EU, The Wellcome Trust, HHMI, and EBC. By creating more Estonian jobs in academia and industry in internationally competitive projects, we hope to attract young talented investigators from everywhere.

SOME DETAILS ON ESTONIA

Estonia is the northern-most of the three Baltic countries, with a population of 1.4 MM, an almost single health care provider, the Estonian Health Insurance Fund, and annual GDP growth rate around 6%. The Estonian population consists approximately of 72% Estonians and 27% Russian speakers. Estonia has a single medical faculty at the University of Tartu founded in 1632. The Estonian human development index is considered high, being 81.2 (compared with 92.1 in Germany) and in business creativity and research Estonia ranks 26th globally, with a factor of 3.94, just between Greece and Slovenia (Germany is 13th with a factor of 4.98) (Economist "World in figures 2003").

ORGANIZATIONAL ASPECTS

General Principles of HGR Act

- Voluntary participation and informed consent
- Confidentiality of gene donor (GD)
- GD's right to opt out and apply for destruction of identification data
- GD's right to know or not to know data
- Non-discrimination by employers and insurers

The Ethics Committee

- Funded independently from the EGP
- Shall act pursuant to generally recognized ethical rules and international conventions
- Aims to ensure the protection of the health, human dignity, identity, privacy, and other fundamental rights and freedoms of GD's
- Draws the attention of the supervisory board and management of the EGPF to circumstances that might be in conflict with ethical norms

Scientific Research

- Free access for Estonian (and in collaboration with Estonian) academic research institutes
- Projects approved by Ethics Committee of EGPF and IRB of the applicant institution
- With only anonymous records (not less than 5)

PUBLIC ATTITUDE TOWARDS THE EGP

In March 2003, public opinion was analyzed by a professional polling company (1,000 randomly selected individuals, telephone interviews). Replies were as follows:

- 32%: I do support the idea of the EGP
- 31%: I have not decided yet
- 24%: I need more information about the EGP
- 6%: I am against the EGP
- 7%: I do not know

This public support is the result of the ongoing educational effort mostly by EGCF and EGPF. Main activities included over 25 special programs (some of it as a series) on Estonian TV, radio broadcasts, more than 300 articles in the Estonian mass media, plus lectures, EGP leaflets, annual scientific meetings, e.g., "Gene Forum" in Tartu, Estonian Society of Human Genetics, special educational programs to physicians participating in the EGP, and so on.

The EGP was also covered internationally, including the scientific journals *Nature*, *Nature Biotechnology*, and *Science*, and general media such as *FAZ*, *Die Zeit*, *Merian*, *BostonGlobe*, and the *NY Times*, and TV programs like Euronews. How many individuals finally enroll in the database will be determined mostly by available funding. From a population point of view, we anticipate consent from 2/3 of the people over time.

WHAT DO WE HAVE TODAY?

We have effectively a working organization, EGPF, a network of physicians (currently over 300 and growing), laboratory facilities, and public support. The EGP went through the pilot phase (September 2002 to April 2003) collecting data only from 3 regions of the country. During this phase, all steps and solutions were tested and necessary corrections were introduced. Mostly, these were improvements in extensive software packages developed for the project. We currently have nearly 10,000 samples collected.

The EGPF is included in an international network of 3 biobank projects under the name Public Population Project in Genomics, P3G. We are using a computer-assisted method for filling the questionnaires (taking about 1–1.5 h to complete at a physician's office). The blood samples (50 ml) together with the signed informed consent forms are transported by a special courier service to the central laboratory facility in Tartu. Informed consent allows commercial access to the anonymous database and samples and is restricted to biomedical research and

development and medical statistics. This research is not restricted to current scientific knowledge nor to future technologies. Blood samples are chilled but not frozen before processing into fractions (DNA, plasma, buffy coat) within 48 h [Steinberg et al., 2002]. All samples and health records are double-coded and a bar-code LIMS system has information and time on every step of sample processing. To date, all DNA samples were run on agarose gels, A_{260} was measured, and standard PCR performed. All data are stored in the database. DNA, plasma, and samples of WBC are kept in liquid nitrogen tanks in MAPI CBS straws. Completed questionnaires are monitored by a special team to ensure quality. The EGPF received the ISO9001: 2000 certification in September 2003. The HGR Act allows the EGPF to go back to individuals (provided that they did not forbid it in their informed consent form) and, after obtaining a new informed consent form, ask additional specific questions and take more measurements including new biological samples. In the future, when the "Digital Health Record" system is functional in Estonia, GD medical information can be collected directly from the primary databases. Physicians are using ICD-10 as the clinical nomenclature recommended by WHO (including the use of pharmaceutical drugs), asking for personal data (including education and occupation), genealogy (three ancestral generations), health behavior (smoking, alcohol, nutrition, physical activity), and objective phenotypic data including weight, height, blood pressure, and pulse. Family physicians who are the main data collectors have preexisting information on their patients and can use this for the EGP including information from special care hospitals on specific diagnoses, diagnostic procedures, and their results. The EGP can monitor other databases such as the Estonian Cancer Registry. In addition to the general EGP questionnaire, more specific phenotypes can be collected in defined (and agreed with partners beforehand) disease areas using hospital settings and their databases, special physicians, and laboratory or instrumental analysis. Several research projects have already been approved by the ethics committee and are in progress at the University of Tartu and the Estonian Biocenter (EBC and EGeen, Inc.).

In conclusion, high-quality phenotypes are difficult to collect and there are no easy solutions. We, and others [Jick et al., 1992], firmly believe that general practitioners do have structured information in their computers and are able to collect quality data useful for many research projects and clinical trials. Several reports are devoted to the current biobanks including the EGP [Austin et al., 2003a,b; Metspalu et al., 2004]

POTENTIAL PROBLEMS

The EGP is a scientific infrastructure for the next generation of research and as such is difficult to finance, since funding agencies and venture capital are more concerned with immediate results and capital return. In addition, academia and private capital have different expectations of the project and, therefore, in order to please both parties, compromises are needed. Finally, the management of the large, ethically sensitive and socially complex public-private projects is always difficult and includes risks to funding agencies as well as to scientists.

KEY TO SUCCESS

- Education of the population and physicians on genetics, biotechnology, and the potential of personalized medicine
- Careful preparatory work and IT solutions
- Transparency of the financing and management
- Quality of the collected data, phenotypes, pedigree information
- Include the best scientists on each phenotype being studied

ACKNOWLEDGMENTS

I thank all physicians who are collecting data for the EGP and, last but not least, close to 10,000 individuals who have taken time to go through the interview and donated samples. A.M. is supported by Targeted Funding by EMRE (0182582s03).

WEB RESOURCES

1. Estonian Genome Project Foundation: www.geenivar-amu.ee
2. Estonian Genome Foundation: www.genomics.ee
3. Dept. of Biotech., IMCB Univ. of Tartu: www.biotech.ebc.ee
4. Asper Biotech: www.asperbio.com
5. Estonian Health Insurance Fund: www.haigekassa.ee/HK/in_english.htm
6. Egeen, Inc.: www.egeeninc.com
7. "Gene Forum" in Tartu: www.geneforum.ee
8. Public Population Project in Genomics, P3G: www.p3gconsortium.org
9. Estonian Cancer Registry: www.ekmi.ee/EARST/evr.pdf
10. UK biobank: www.ukbiobank.ac.uk

REFERENCES

- Austin MA, Harding S, McElroy C. 2003a. Genebanks: a comparison of eight proposed international genetic databases. *Community Genet* 6:37–45.

- Austin MA, Harding SE, McElroy CE. 2003b. Monitoring ethical, legal, and social issues in developing population genetic databases. *Genet Med* 5:451–457.
- Collins FS, Green ED, Guttmacher AE, Guyer MS. 2003. A vision for the future of genomics research. *Nature* 422: 835–847.
- Jick H, Terris BZ, Derby LE, Jick SS. 1992. Further Validation of Information Recorded on a General Practitioner Based Computerized Data Resource in the United Kingdom. *Pharmacoepidemiol Drug Safety* 1:347–349.
- Kaiser J. 2002. BIOBANKS: Population databases boom from Iceland to the U.S. *Science* 298:1158–1161.
- Metspalu A, Kohler F, Laschinski G, Ganten D, Roots I. 2004. Das estnische Genomprojekt in kontext der europaischen Genomforschung. *Dtsch Med Wochenschr* 129:S25–S28.
- Steinberg K, Beck J, Nickerson D, Garcia-Closas M, Gallagher M, Caggana M, Reid Y, Cosentino M, Ji J, Johnson D, Hayes RB, Earley M, Lorey F, Hannon H, Khoury MJ, Sampson E. 2002. DNA banking for epidemiologic studies: a review of current practices. *Epidemiology* 13:246–254.