34

Interpretation of genetic data for medical and public health uses

by Paul A. Schulte

THE GROWTH OF genetic data presents various resources that may be of use to clinicians and public health practitioners. However, before there can be effective and safe application of genetic information various interpretive issues need to be addressed. These include: (1) the need for population thinking about genetic information; (2) the extension of the biomedical paradigm to include it; and (3) attention to issues in the application of genetic information to qualitative and quantitative risk assessments. These issues have scientific, ethical, and social aspects.

Genetic databases are derived from various types of clinical, laboratory, and field studies involving single or multiple genes, expression products, and covariate information. These are derived from a range of efforts including microarray experiments and examination of expression profiles, and assessments of case series or etiologic epidemiologic studies. Epidemiologists, worldwide, often establish genetic databases when conducting epidemiologic studies (Steinberg et al. 2002). Sometimes these databases are not the primary objective, but since the investment to conduct population-based field studies and obtain data and specimens from subjects is so large, it is cost effective to collect and bank DNA for the current and future research.

Research on genetic factors in people can be conceived on a continuum that has been described as leading from basic research through population research to medical and public health practice (Schulte 2004). Regardless of whether a genetic factor has been analytically and clinically validated, the mere collection of DNA and the subsequent assay and results are de facto genetic tests and ethical, legal, and social issues of genetic testing need to be considered. There is a rich literature on these topics (e.g., see Grody 2003; Laberge and Knoppers 1992; Davison, Macintyre, and Smith 1994).

For the most part, the discussions of the ethical and social issues in genetic testing have involved one or a few genes. The ability to assess hundreds or thousands of genes or gene products in one microarray amplifies issues that have already been identified for individual genes, and creates some situations with new implications (Grody 2003; Schulte 2004). These implications involve how scientists and nonscientists conceive of variability. To clearly define these terms, there is a need to apply the disciplines of population science—the understanding that individual research subjects or patients are part

of population groups that can be defined by genetics and other factors (Knoppers 2000). Finally, the applicability of genetic information to the socially desirable practice of risk assessment has various implications that should be considered.

Need for population thinking

Genetic analysis at the macro and micro levels has provided an opportunity to quantify variability in populations. "Variation" has been the topic that sparked creative thinking in biology and epidemiology. In the view of Mayr (1982), Western thinking for 2000 years after Plato was dominated by "essentialism," or the belief that there were a limited number of immutable essences. Much of epidemiologic (and for that matter, biomedical scientific) thinking to date has been essentialist. For example, the early statistics used in public health by Graunt (1620–1674) and Quetelet (1796–1874) attempted to calculate true values to overcome the confusing effects of variation. Quetelet, who attempted to use a mathematical function of height and weight as a biomarker, hoped to calculate characteristics of the average person. To him and like thinkers, variations were nothing but "errors" around mean values. In the nineteenth century, new ways to view nature began to spread; a concept called "population thinking" was developed that stressed the uniqueness of everything in the organic world. To population thinkers there is no "typical" individual, and mean values are considered abstractions (Mayr 1982).

Lloyd (1998) has elegantly set the stage for further discussions of variation by recognizing that there is a diversity of theories and models from the different scientific disciplines involved in the Human Genome Project that provide a unique challenge to producers and consumers of DNA-sequencing information. While it is generally thought that science provides guidance as to variation, there are actually social views and constructs intertwined with the level of scientific analysis that can lead to divergent interpretations of genetic data. Specifically, science deals with observation of events under various circumstances. However, depending on the scientific model, the interpretation may change. For example, for a mechanistic model such as the "biochemical causal-pathway model," the emphasis is usually on function and the model presents, in detail, a picture of normal or proper functioning. As Lloyd (1998) notes, there is no room for simple variation in the explanatory scheme of such a model. Variation is seen as nonfunctioning. For example, with sickle cell anemia, the goal of the causal model is to explain at least one causal chain from DNA in its initial state arranged on a chromosome to the ultimate state involving iron molecules arranged in hemoglobin carrying around oxygen in the body. There is no variation in function that is addressed. In contrast, very different types of descriptions are used when populations of people are considered. Population thinking would suggest tri-level analysis: individual, propagating family, and community. At each level, in population sciences, a great deal of information is needed to delineate the variation. These include: (1) the distribution and frequencies of the types of genes, phenotypes, and functions associated with these genes, and (2) the range of environmental variables and the related parameters of reaction (Lloyd 1998). This kind of information is not provided in biochemical causal models that are prominent in genetics. Therefore, depending on the level of scientific analysis, it is possible to get

interpretations of the role of genetic factors that are either deterministic or probabilistic. These interpretations have different implications for users and consumers of genetic information. Too often genetic reductionism leads to flawed thinking about the role of environmental factors in genetic diseases. Of the more than 7,000 genetic disorders known, only a few have been studied for environmental factors in their phenotypic expression (Samuels 2003).

The definition and range of variation that can occur in microarray studies or that will be stored in databases poses a rich, but problematic, challenge to medical and public health practitioners. Beyond the issue of standardization of approaches and scales are the issues of the extent to which the people in the database represent those who are not in the database. That is, are they representative in terms of genetic and ethnic factors as well as in terms of the various other host or environmental factors? Of concern is whether various ethnic groups have similar opportunities to be in a database. If not, characterization in a database can make one ethnic group appear more or less "susceptible" than another ethnic group without the same opportunity for characterization.

Utilization of genetic bases for public health requires that variation in the population be accurately categorized and that the concept of abnormal be thought of more in terms of susceptibility, then determined appropriately, and interpreted probabilistically. For public health purposes, there is a need to define concepts such as susceptibility on a population level. As Lloyd (1998) concluded "public and scientific misconceptions of susceptibility are probably one of the most prominent problems facing those interested in the development of genetic medicine."

Extension of the biomedical paradigm

Predictive genetic testing is a departure from the traditional paradigm of laboratory medicine (which involves laboratory testing or confirmation in a patient with signs or symptoms) in that disease or high risk can now be diagnosed years or decades before the clinical appearance of signs or symptoms (Grody 2003). Such predictive knowledge forces us to re-examine the definition of disease or pre-existing condition and the way society deals with people in those states of health. Increasingly, there will be a need for pre- and post-test counseling. Press and Burke (2000) have suggested that this type of predictive genetic testing and information be assessed in terms of how well it fits into the progression of biomedicine that has increasingly moved from bringing people into the medical system at the onset of symptoms to bringing them in as part of a prevention and early detection paradigm in which risk factors, including genetic information, trigger monitoring or intervention. The concept of "high risk management," however, needs to be based on epidemiologic research that links the genetic marker or markers to risk and, ultimately, to diseases (Samuels 2003). The mere presence of an allele in a test is not a sufficient indicator of risk or disease. Before a predictive genetic test can be useful for clinical or public health purposes, the analytical and clinical validity and clinical utility should be determined (Burke et al. 2002). At present, much of the data in databases derived from microarray studies have not been validated or linked to diseases or risks in populations. Additionally, such testing extends beyond the immediate per-

son being tested and may impact all with that germline. Microarray technology will make such testing too easy and may further bring a departure from traditional laboratory medicine (Grody 2003).

Finally, one of the vexing problems that limits use of genetic information is the difficulty in translating genetic information on groups to individuals. Epidemiologic research pertains to group risk assessment, not to individuals in the group. Information about the presence or absence of variants of genetic factors in groups may not be useful for individual classification without research specifically designed for such translational purposes (McCanlies et al. 2002).

Risk assessment

One area for application of genetic data is in risk assessment. The term "risk assessment" has different meanings in different countries. In some countries, it is an effort to provide society with estimates of the likelihood of illnesses and injury as a consequence of exposure to various hazards. Risk assessments are needed when social policy decisions are in dispute, when alternative policies in question are not subject to direct measurements, and when scientific analysis of a hazard is not complete (IPCS 2001). Risk assessment can be qualitative or quantitative at the individual or population level. The term quantitative risk assessment describes the response associated with a specific level of exposure. The promise is for a more refined assessment of risk through the identification of gene-gene and gene-environment interactions, and also for the focusing of prevention and control programs for high-risk individuals. The perils of using genetic information in risk assessment include ethical and social issues such as stigmatization, discrimination, and the misconception that removing a susceptible person from the exposure scenario without reducing exposure opportunities will reduce risk effectively, when it may not, on a comparative basis (Vineis and Schulte 1995).

Most of the research on genetic modifiers of exposure-disease associations has involved single genes which are polymorphic for a particular enzyme. These have been referred to as "metabolic polymorphisms." They have also been referred to as susceptibility markers. One example of the use of susceptibility markers examined how changing glutathione-*S*-tranferase theta (GSTT1) genotype frequencies would impact cancer risk estimates from dichloromethane by the application of Monte-Carlo simulation methods in combination with physiologically based pharmacokinetic (PBPK) models (El-Masri, Bell, and Portier 1999). The investigators reported that average and median risk estimates were 23% to 30% higher when GSTT1 polymorphism was not included in the models.

Although the scientific press has been replete with promises of the use of genetic information in risk assessment, the use of even a single metabolic polymorphism in quantitative risk assessment is rare. More recently, similar pronouncements have been made for toxicogenomics; however, there is a large amount of work necessary to devise effective risk assessments based on these technologies. The ability to interpret data sets from toxicogenomics and other microarray studies will be quite difficult due to the large number of potential combinations of gene expression states (Morgan et al. 2002). Before

more use is made of this type of information and extended to include multiple genes from microarray generated databases, there is need for new thinking on how to incorporate the information in risk assessments. Using the results of human studies and animal data, risk assessors define environmental exposures that may be linked to disease in a portion of the population (Waters et al. 2003). Often these determinations are based on the use of default assumptions to reflect limitations in knowledge. When used in risk assessment, genetic information can provide increased mechanistic insight and replace default assumption in species and other extrapolations. There is also need for consideration of how assessments that identify differential risks in populations will be incorporated into laws, regulations, or practices in ways that do not have untoward effects. If protective measures based on genetic risk assessments are to be provided to particular groups or individuals, criteria should be established for how those groups are identified and how their rights will be preserved (Hornig 1988).

References

Burke, W., D. Atkins, M. Gwinn, A. Guttmacher, J. Haddow, J. Lau, G. Palomaki, et al. 2002. Genetic test evaluation: Information needs of clinicians, policy makers, and the public. *American Journal of Epidemiology* 156:311–18.

Davison, D., S. Macintyre, and G. D. Smith. 1994. The potential social impact of predictive genetic testing for susceptibility to common chronic diseases: A review and proposed research agenda. *Sociology of Health and Illness* 16:340–371.

El-Masri, H. A., D. A. Bell, and C. J. Portier. 1999. Effects of glutathione transferase theta polymorphism on the risk estimates of dichloromethane to humans. *Toxicology and Applied Pharmacology* 158:221–30.

Grody, W. W. 2003. Ethical issues raised by genetic testing with oligonucleotide microarrays. *Molecular Biotechnology* 23:127–38.

Hornig, D. F. 2002. Conclusion. In *Variations in susceptibility to inhaled pollutants*, ed. J. D. Brain, B. D. Beck, A. J. Warren, and R. A. Shaikh. Baltimore, MD: The John Hopkins University Press.

International Programme on Chemical Safety (IPCS). 2001. Environmental Health Criteria 222: Biomarkers in risk assessment; Validity and validation. Geneva: World Health Organization.

Knoppers, B. M. 2000. From medical ethics to "genethics." *Lancet* 356, Suppl. 1:s38.

Laberge, C. M., and B. M. Knoppers. 1992. Rationale for an integrated approach to genetic epidemiology. *Bioethics* 6:317–30.

Lloyd, E. A. 1998. Normality and variation: The human genome project and the ideal human type. In *The philosophy of biology*, ed. D. L. Hull and M. Ruse. Oxford: Oxford University Press.

Mayr, E. 1982. The growth of biological thought: Diversity, evolution, and inheritance. Cambridge, MA: Harvard University Press.

McCanlies, E., D. P. Landsittel, B. Yucesoy, V. Vallyathan, M. L. Luster, and D. S.

Sharp. 2002. Significance of genetic information in risk assessment and individual classification using silicosis as a case model. *Annals of Occupational Hygiene* 46:375–81.

Morgan, K. T., H. R. Brown, G. Benavides, L. Crosby, D. Sprenger, L. Yoon, H. Ni, et al. 2002. Toxicogenomics and human risk assessment. *Human and Ecological Risk Assessment* 8:1339–1353.

Press, N., and Burke W. 2000. "Genetic exceptionalism" and the paradigm of risk in US biomedicine. Paper presented at "A Decade of ELSI Research" Conference, January 16, Washington, DC.

Samuels, S. W. 2003. Occupational medicine and its moral discontents. *Journal of Occupational and Environmental Medicine* 45:1226–33.

Schulte, P. A. 2004. Some implications of genetic biomarkers in occupational epidemiology and practice. *Scandinavian Journal of Work Environment and Health* 30:71–79.

Steinberg, K., J. Beck, D. Nickerson, M. Garcia-Closas, M. Gallagher, M. Caggana, Y. Reid, et al. 2002. DNA banking for epidemiologic studies: a review of current practices. *Epidemiology* 13:246–254.

Vineis, P., and P. A. Schulte. 1995. Scientific and ethical aspects of genetic screening of workers: The case of the N-acetyltransferase phenotype. *Journal of Clinical Epidemiology* 48:189–197.

Waters, M. D., J. K. Selkirk and K. Olden. 2003. The impact of new technologies on human population studies. *Mutation Research* 544:349–60.

Blood & Data

Ethical, Legal and Social Aspects of Human Genetic Databases

Published for the International ELSAGEN Conference,
University of Iceland, Reykjavík, 25–28 August 2004

Edited by

Gardar Árnason
Salvör Nordal
Vilhjálmur Árnason

University of Iceland Press
& Centre for Ethics
Reykjavík 2004

Cover design: Róbert Guillemette
Layout: Brynjólfur Ólason
Printing: Gutenberg

This book is a publication of the ELSAGEN project (Ethical, Legal and Social Aspects of Human Genetic Databases: A European Comparison), which is financed between 2002 and 2004 by the European Commission's 5th Framework Programme, Quality of Life (contract number QLG6-CT-2001-00062).

PRINTED IN ICELAND

ISBN 9979-54-593-3