mary database, in addition to making it available on their own FTP servers, it would facilitate one-stop shopping and eliminate the need to query each data repository individually.

Primary databases should establish working relations with each of the data repository centers to ensure that newly produced data is transmitted to them quickly. In some cases, this will require working with the informatics group at the centers to output the data in a standard format. However, genome data is often generated in laboratories that do not have the luxury of dedicated individuals to assist in the data collection and cataloging. It becomes largely the responsibility of the primary database to provide the tools necessary to allow such groups to submit their data efficiently [for example, Authorin, a program developed to transmit data to GenBank (*2*)].

Genome-wide research is now in vogue in a few major centers, and this is resulting in the generation of large amounts of low-resolution mapping data. The challenge to the database is to integrate this information with that garnered by all the smaller laboratories who produce high-resolution maps of their favorite chromosomal region. For example, all the detailed map information derived from a particular yeast artificial chromosome (YAC) must be correctly placed on the low-resolution map of YACs. The primary mapping database must both store and allow integration of all information relating to that one YAC, irrespective of the original source of the data. Even if a virtual database was a practical proposition, the essential integration of information would not take place without a single authority taking responsibility for carrying it out. It is logical that the primary mapping database, in consultation with appropriate members of the genome mapping community, provide such integration.

Although it is in the interest of the community that all relevant mapping information be made available in a central fashion, it cannot be the function of the primary database to act as a "data police" for ensuring timely entry. Foremost responsibility must rest on the research community itself to promptly submit data so that the primary database increases in utility for the entire genome community. The requirements of scientific journals to have the data submitted to a primary database as a condition for publication of an article is a significant motivating factor for timely data submission. Furthermore, this relieves the journals from having to publish large amounts of sequence information that is largely unintelligible on a printed page but can be studied and analyzed extensively once in electronic form. A similar requirement for mapping information must also be established so that detailed information on, for example, probe descriptions, polymorphisms, and chromosome breakpoints are consigned to the primary mapping database and can simply be referred to in journal articles by their accession numbers. This mechanism would also allow retrieval of all information relating to one publication from the database in a single action, such as retrieving all the PCR primers and their associated polymorphic information from a particular linkage study.

Although it is the major directive of each primary database to collect data, maintain its integrity, and make data available in as many different formats as possible, the limited resources do not permit an unlimited number of access methods. Many tools for accessing and analyzing the data contained with the DNA sequence and protein databases have already been developed by both the research community and commercial enterprises. As increasing amounts of mapping information are accumulated in the primary mapping database, tools for accessing such data are already being developed by several groups throughout the world using sophisticated graphics for map displays. Although it will still be the responsibility of the primary mapping database to provide some tools to the public for accessing the data, it is also their task to provide information to the research community to assist in the development of new access methods.

References

1. M. J. Cinkosky, J. W. Fickett, P. Gilna, C. Burks, *Science* **252**, 1273 (1991).
2. T. B. Shows *et al.*, *Cytogenet. Cell Genet.* **46**, 11 (1987).

Presymptomatic Diagnosis: A First Step Toward Genetic Health Care

C. Thomas Caskey

Rapid progress is being made by the international human genome initiative in the discovery of genes responsible for human disease. The establishment of a genetic map of the human, a goal of the initiative, will provide medicine with the most rapid expansion of new knowledge in recent history. The application of this knowledge will begin a new era of molecular medicine, in which the risk of disease can be accurately assessed by DNA-based diagnostic procedures. Furthermore, disease pathogenesis and progression can be logically and efficiently studied, once the genes associated with a particular disease are known. The ability to detect individuals at risk for a disease prior to any pathologic evidence of the disease theoretically offers to medicine a new strategy—anticipation of disease and preemptive therapy. This vision will not be realized by the mapping and sequencing efforts of the Human Genome Project alone, but also requires study of gene function by disease specialists.

DNA-based diagnostics, coupled with the discoveries of new genes, can benefit health care in the short term by reducing the incidence of severe or presently untreatable diseases. Screening programs for Tay-Sachs and β-thalassemia initiated in the 1970s have reduced the incidence of these diseases by 20-fold (*1*). This remarkable success can be attributed to recognition of the requisite features of accurate diagnosis, education (public and professional), and freedom for individuals to make reproductive decisions in avoiding the disease in their families. The discovery in recent years of common disease genes, such as those causing cystic fibrosis (CF) (*2*) (1 in 3700) and fragile X syndrome (*3*) (1 in 1200), and their ancestral mutations or premutations, makes possible the expansion of genetic screening. The Ethical, Legal, and Social Implications (ELSI) component of the Human Genome Project within the National Institutes of Health is studying the feasibility and utility of population-based screening for CF in the United States. The first reports of this study will be available in the fall of 1993. The experiences from Tay-Sachs, β-thalassemia, sickling hemoglobinopathies, and CF screening should point the way for effective and acceptable broader usage of genetic screening in other common heritable diseases, such as Gaucher disease, α_1-antitrypsin deficiency, myotonic dystrophy, and spinal muscular atrophy (*4*). Couples at risk for disease in their offspring have had increased options since the 1970s for reproductive planning, including prenatal diagnosis, preimplantation diagnosis of embryos, artificial insemi-

The author is an investigator in the Howard Hughes Medical Institute and Henry and Emma Meyer Professor in the Institute for Molecular Genetics, One Baylor Plaza, Baylor College of Medicine, Houston, TX 77030.

nation by donor, in vitro fertilization, and embryo donation. In the future, genetic screening programs will operate more cost-effectively by testing multiple disease loci, perhaps through the use of DNA "chips," which have high allele capacities and can be interpreted automatically (*5*). Discoveries of genes for particular diseases through the Human Genome Project provide the knowledge for this new direction. Although the public and medical attitude toward population-based screening is often negative, this could be dramatically reversed by the development of new, effective therapy options.

Screening of newborns is of great utility in diseases such as phenylketonuria (PKU) (*6*) and hypothyroidism (*7*). Early intervention with diet or thyroid hormone administration has significantly reduced the mental retardation associated with these diseases. Screening tests for newborns currently detect metabolites or hormones with accurate, automated, and economical laboratory tests. A shift to DNA-based screening of newborns would broaden the number of diseases that can be screened and has already proved feasible and less costly. Currently, tests are only used for childhood diseases that have effective therapy options, and therefore yield considerable overall health care advantages in terms of morbidity and cost. It is likely that the age range for disease screening and therapeutic intervention will be expanded from newborns to adults. Two examples where this might be useful are in α_1-antitrypsin deficiency and heritable causes of atherosclerosis (*8*); in these cases, smoking abstinence and diet and medication can reduce the severity of disease. Here, the knowledge of disease risk offers the opportunity to preempt the pathology. In the future, enthusiasm for newborn screening could be increased by the development of successful gene therapy protocols for diseases that currently have little or no effective treatment. Considerable research effort is being put into gene therapy for diseases that have pathology that progresses from birth, such as CF, Duchenne muscular dystrophy, immune deficiencies, PKU, and urea cycle defects (*7*). The earlier the genetic correction of these diseases, the higher the likelihood of long-term success. Thus, expansion of newborn screening for heritable disorders may well be coupled to the success of gene therapy.

There is every reason to examine carefully the potential for reducing the morbidity and mortality of adult-onset heritable diseases, for which current care is minimal and the medical cost high. Examples include myotonic dystrophy, spinocerebellar ataxia, Huntington disease, and spinal and bulbar muscular atrophy. In these diseases, the mutations are unstable trinucleotide repeats that cause advancing disease in adults and can readily be detected presymptomatically. DNA diagnostic methods are available for some cancers, including retinoblastoma (*9*), adenoma polyposis coli (*10*), renal cancer (von Hippel-Lindau disease) (*11*), and also for adult polycystic kidney disease, which is the most common cause of heritable renal failure and whose risk can be determined genetically in advance of disease detection (*12*). Dedicated efforts to isolate the gene accounting for a form of breast cancer should broaden the presymptomatic diagnosis of cancer (*13*). The success of early identification of retinoblastoma (leading to the sparing of the eyesight of children) is an encouraging example of the presymptomatic detection of cancer risk. We must determine the utility of early surgical or medical intervention for common cancers, such as colon, breast, and prostate, after identification of individuals at high risk by DNA diagnosis. Current therapy awaits the patient's diagnosis of cancer, that is, after initiation and progression of the disease. The preemption of these events, on the basis of knowledge of gene function and pathogenic progression, would provide an entirely new approach to these common disorders. There are a host of other common diseases, such as diabetes mellitus, rheumatoid arthritis, ulcerative colitis, amyotrophic lateral sclerosis, Alzheimer's disease, and multiple sclerosis, in which genetic factors or genes are clearly identified. Since many of these have an immune response associated with their pathogenesis, methods to preempt the initial or progressive immune response in individuals identified to be at risk have the potential to reduce disease morbidity and incidence. The identification of the involved genes and determination of their functions could open entirely new opportunities for medical intervention.

There are many unresolved issues that will profoundly influence this new direction of molecular medicine. These include the understanding of the function of various genes, drug and surgical intervention, education of the public, acceptance of new approaches to intervention, and patient compliance with preemptive therapy. The Human Genome Project is providing the techniques, materials, and methods to greatly accelerate our efforts in the gene discovery. Nevertheless, it is a different knowledge that must be gathered to determine the utility of this information in medicine. Presently, we can assign functions to less than 40% of sequenced genes on the basis of sequence analysis (*14*). The ability to alter the cellular expression of genes in a specific manner is critical to disease intervention. Such technology will include gene therapy, inhibition of gene expression, and drug development. These new scientific directions are derivatives of the Human Genome Project and have great potential for health benefits.

Only an educated public and medical profession can take full advantage of this era of molecular medicine. The ELSI program of the National Institutes of Health and the Department of Energy have responded to this need with an increasing number of sponsored activities to study specific diseases, medical insurance coverage, and ethical considerations for new approaches to disease intervention. As the pace of medical and scientific discoveries increases, it will be necessary to focus the attention of ELSI and other organizations on education and the appropriateness of new health care proposals.

A change in the health care insurance system is critical to the success of preemptive therapy following presymptomatic diagnosis. Appropriate use of preemptive care will benefit the patient and lower the cost of disease. If a presymptomatic diagnosis is equated to identification of a "prior existing condition," the patient will lose privacy, actuarial rates become complex, presymptomatic diagnosis neglected, and a health care opportunity lost. We should consider underwriting our nation's heritable disease care by a national pool of insured citizens to ensure privacy, health care access, and opportunity for disease intervention.

References

1. M. G. Blitzer and G. A. McDowell, *Clin. Lab. Med.* **12**, 463 (1992); A. Cao, M. C. Rosatelli, R. Galanello, *Curr. Opin. Genet. Dev.* **1**, 48 (1991).
2. J. M. Rommens *et al.*, *Science* **245**, 1059 (1989); J. R. Riordan *et al.*, *ibid.*, p. 1066; B.-S. Kerem *et al.*, *ibid.*, p. 1073.
3. A. J. M. H. Verkerk *et al.*, *Cell* **65**, 905 (1991).
4. C. T. Caskey, *J. Am. Med. Assoc.* **269**, 1986 (1993).
5. S. P. A. Fodor *et al.*, *Nature* **364**, 555 (1993).
6. M. E. O'Flynn, *Curr. Probl. Pediatr.* **22**, 159 (1992).
7. S. M. Willi and T. Mosnang Jr., *Pediatr. Clin. North Am.* **38**, 555 (1991).
8. K. Mitani and C. T. Caskey, *Trends Biotechnol.* **11**, 162 (1993).
9. J. Wiggs *et al.*, *N. Engl. J. Med.* **318**, 151 (1988).
10. S. M. Powell *et al.*, *Nature* **359**, 235 (1992).
11. F. Latif *et al.*, *Science* **260**, 1317 (1993).
12. W. M. Bennett, *Semin. Nephrol.* **10**, 552 (1990).
13. B. B. Biesecker *et al.*, *J. Am. Med. Assoc.* **269**, 1970 (1993); M. C. King, S. Rowell, S. M. Love, *ibid.*, p. 1975.
14. M. D. Adams, M. B. Soares, A. R. Kerlavage, C. Fields, J. C. Venter, *Nat. Genet.* **4**, 373 (1993).