didate genes that require extensive subsequent validation.

We are entering a phase in which we shall see more functional genomics data and hear less hype. Despite the unprecedented volume of data being generated from individual experiments, rigorous, reproducable design must be the watchword, so that emerging technologies can be fairly evaluated. The traditional format of scientific publication cannot reflect the scope and depth of data being produced. Summaries of results and conclusions in publications are certainly of interest, but are not very useful for subsequent analysis or utilization of the data by others and may not even be adequate for effective peer review. A key issue regarding the access to data from publicly funded, genome-scale, functional analyses must be addressed. A great legacy of the structural genomics era is the philosophy and practice of the public release of data that we hope will carry over to the functional genomics age. The timely submission of expression data, for example, in some standard format independent of specific technique, would lead to the most effective analysis and utilization of the results by the scientific community.

REFERENCES AND NOTES

1. V. McKusick, *Genomics*, in press.
2. A. Goffeau *et al.*, *Science* **274**, 546 (1996).
3. S. Oliver, *Trends Genet.* **12**, 241 (1996).
4. V. Smith, D. Botstein, P. O. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6479 (1995); D. D. Shoemaker *et al.*, *Nature Genet.* **14**, 450 (1996).
5. S. Fodor *et al.*, *Nature* **364**, 555 (1993).
6. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995).
7. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *ibid.*, p. 467.
8. V. E. Velculescu *et al.*, *Cell* **88**, 243 (1997).
9. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
10. L. Zhang *et al.*, *ibid.* **276**, 1268 (1997).
11. S. Fields and O. Song, *Nature* **340**, 245 (1989).
12. J. Yates, J. Eng, A. McCormack, D. Schieltz, *Anal. Chem.* **67**, 1426 (1995).
13. M. Wilm *et al.*, *Nature* **379**, 466 (1996).
14. P. Goodfellow, *Nature Genet.* **16**, 209 (1997).
15. Symposium on "Genomics and Gene Therapy: Meaning for the Future of Science and Medicine," Harvard Institute of Human Genetics, Cambridge, MA, 26 March 1997.
16. http://www.ncbi.n1m.nih.gov/ncicgap/
17. G. Miklos, L. Gabor, G. M. Rubin, *Cell* **86**, 521 (1996).
18. S. Fields, *Nature Genet.* **5**, 325 (1997).

19. We thank J. Roskams, S. Fields, and D. Bassett for critical reading of the manuscript and helpful input.

Predictive Genetic Testing: From Basic Research to Clinical Practice

Neil A. Holtzman,* Patricia D. Murphy,† Michael S. Watson, Patricia A. Barr‡

Announcements of discoveries of disease-related genes often suggest that tests to predict people at risk of future disease will soon be available (*1*). Few regulatory barriers stand in the way (Table 1A). If a commercial or academic clinical laboratory wants to offer a genetic test service (whereby it receives specimens, analyzes them, and reports results), it must register under the Clinical Laboratory Improvement Amendments (CLIA) and receive certification from the Health Care Financing Administration, which is the federal agency primarily responsible for the administration of CLIA (*2*). The process is not expensive and causes no delays in offering tests. For an organization that wants to market kits that independent laboratories, health care providers, or consumers can use to perform the test, the process is longer and more complex (*3*). In this case, the organization must first notify the Food and Drug Administration (FDA). If the test kit is not substantially equivalent to others already on the market, the FDA will require the organization to go through a premarket approval process during which it must collect data under an Institutional Review Board (IRB)–approved protocol to demonstrate that the test is clinically valid for the use intended by the manufacturer. Clinical validity includes determining test sensitivity and the predictive value of a positive test result (PVP) (*4*). Few genetic tests emerging from genome discoveries are being marketed as kits today, primarily because of the complexities of the assays and the interpretations. Under CLIA, there are no requirements for demonstrating clinical validity, but in the certification process, each laboratory may be called on to provide data on the analytic validity of its tests (*5*).

If the discovery of a disease-related gene provided sufficient information on a test's validity and other aspects of its effectiveness, this regulatory environment might be adequate. Seldom is this the case. First, the data collected on research subjects may not be representative of the findings in others at risk of the disease. Second, additional questions regarding the benefits and risks of testing, which are unlikely to be part of the original research, need to be considered. Under current regulations, the acquisition of sufficient data to warrant the transition of predictive genetic testing into health care cannot be ensured. More data must be collected in an investigative stage, during which results may be given to subjects (through their providers) if they have consented to participate and receive results. Before consenting, subjects must be informed of the questions the study is designed to answer and the potential risks and benefits.

The Task Force on Genetic Testing, was convened by the National Institutes of Health–U.S. Department of Energy (NIH-DOE) Working Group on Ethical, Legal, and Social Implications of Human Genome Research to review the state-of-the-art of genetic testing in the United States and to make recommendations when necessary to ensure (i) development of safe and effective genetic tests, (ii) their performance in laboratories of assured quality, (iii) their appropriate use by health care providers and consumers, and (iv) the continued delivery of tests for rare diseases. The Task Force, representing a wide array of stakeholders, has just issued its final report, concluding that, for the most part, genetic testing for Mendelian disorders in the United States has developed successfully, providing options for avoiding, preventing, and treating inherited disorders (*6*). However, problems arise in attaining each of the goals. Below we will consider the steps needed to establish the safety and effectiveness of a genetic test before it is incorporated into health care and the relevant Task Force recommendation (Table 1B). In its report, the Task Force makes recommendations on the last three goals as well. The focus of the Task Force on potential problems in no way is intended to detract from the benefits of genetic testing. Its overriding goal is to recommend policies that will reduce the likelihood of damaging effects so the benefits of testing can be fully realized.

N. A. Holtzman is with the Genetics and Public Policy Studies, Johns Hopkins Medical Institutions, 550 North Broadway, Suite 511, Baltimore, MD 21205, USA. P. D. Murphy is in the Department of Obstetrics and Gynecology, Albany Medical Center, Albany, NY 12208, USA. M. S. Watson is in the Departments of Pediatrics and Genetics, Washington University School of Medicine, St. Louis Children's Hospital, One Children's Place, St. Louis, MO 63110, USA. P. A. Barr is with Barr, Sternberg, and Moss, PC, 507 Main Street, Bennington, VT 05201, USA. N. A. Holtzman was the chair, M. S. Watson was the co-chair, and P. A. Barr and P. D. Murphy were members of the Task Force on Genetic Testing.

*To whom correspondence should be addressed.
†Consultant to Oncormed, a company that provides genetic tests.
‡Member of the board of the National Breast Cancer Coalition.

Table 1. Steps in the development of predictive genetic tests. (**A**) Current policy. (**B**) Task Force recommendations.

Research* →	Investigative† ↔	Regulatory‡ ↔			Health care
	Type of test§				
A Current policy					
Gene identified	Service	Establish analytic validity	CLIA certification		Market
	Kit	Establish analytic, clinical validity	IRB-approved protocol‖	FDA approval	Market¶
B Task Force recommendations					
Gene identified	Service	Establish analytic, clinical validity	IRB-approved protocol‖	CLIA certification	Market¶
		Initiate post-test intervention studies	IRB-approved protocol#	External review	
	Kit	Establish analytic, clinical validity	IRB-approved protocol‖	CLIA certification	Market¶
		Initiate post-test intervention studies	IRB-approved protocol#	FDA approval / External review	

*In the research step, results are not given to subjects. †In the investigative step, results may be given to subjects. ‡The regulatory component may relate to both the investigative and health care steps. §As technology improves for tests provided as either services or kits, components in the investigative step may need to be repeated. Test development is an iterative process. ‖Protocols establish sensitivity, and positive predictive value of each intended use in each intended group or population. ¶Marketing may be contingent on postmarket surveillance recommended by external review (service or kits) or FDA (kits). #Protocols assess safety and effectiveness of proposed post-test interventions.

Extrapolating from Research Findings

The original discovery of a disease-related gene is usually made in families with multiple affected relatives (*7*). Subsequent study in other affected individuals often reveals additional inherited mutations capable of causing, or predisposing to, the disease (genetic heterogeneity) (*8*). Consequently, tests capable of detecting only the original mutations will have low sensitivity when applied to other families or the general population. For common, complex disorders, other genetic as well as environmental factors may exacerbate the deleterious effects of an inherited susceptibility mutation, thereby affecting the penetrance of the allele. Consequently, the PVP in the families originally studied, in which these other factors may be aggregated (9), can be higher than in others with a less marked, or absent, family history of the disease (*10*).

To ensure that adequate data are available to demonstrate a test's validity among a wider range of people than were involved in the early discovery studies, the Task Force recommends that "[a]nalytical sensitivity and specificity of a genetic test must be determined before it is made available in clinical practice" and that "[d]ata to establish the clinical validity of genetic tests . . . must be collected under investigative protocols." The Task Force stipulates that "[i]n clinical validation, the study sample must be drawn from a group of subjects representative of the population for whom the test is intended" and that "[f]ormal validation for each intended use of a genetic test is needed" (6, p. 28).

Benefits and Risks

The benefits of some predictive genetic tests can be substantial. Screening of newborns permits the administration of prophylactic antibiotics in healthy appearing infants with sickle cell anemia and of low-phenylalanine diets in asymptomatic infants with phenylketonuria. In the first case, infant mortality is significantly reduced, and in the second, mental retardation can be prevented (*11*). Testing for specific inherited mutations in the *ret* proto-oncogene in children at risk of familial medullary thyroid carcinoma (MTC) or multiple endocrine neoplasia type 2a can spare those found not to have inherited the mutation from frequent monitoring for early signs of MTC. In those who are found by testing to have inherited the mutation, prophylactic thyroidectomy eliminates the risk of thyroid cancer, which can be fatal (*12*).

Unfortunately, relatively few interventions have yet been devised to improve the outcome of most Mendelian disorders (*13*). When no effective treatments have been developed, carrier screening and prenatal diagnosis in couples at risk offer people options of avoiding the birth of children with very severe Mendelian diseases. For some common complex disorders identified by predictive testing, such as breast cancer, optimal modes of preventing or ameliorating disease have not yet proven to be risk-free or totally effective (*14*).

Collecting evidence on the safety and effectiveness of potential interventions in people who have positive test results is more easily undertaken when tests are still in the investigative stage in which clinical validity is being assessed. Clinical trials and other methods studying effectiveness can be initiated simultaneously. Once the test is made available in medical practice, it will be far more difficult to organize studies, to involve providers (who can use the test without participating in a study), or to recruit subjects (*15*). The Task Force believes that "[b]efore a genetic test can be generally accepted in clinical practice, data must be collected to demonstrate the benefits and risks that accrue from both positive and negative results" (6, p. 29).

The Task Force recognizes that for some disorders, collaborative studies will be needed to obtain answers on validity and utility efficiently. It calls on NIH and the Centers for Disease Control and Prevention to support consortia and other collaborative efforts to facilitate collection of data on the safety and effectiveness of new genetic tests (6, p. 35). It also calls for the creation and maintenance of a central repository of cell lines and DNA (without patient identifiers) that can be used by test developers in the validation of new tests, as well as in quality control after tests are developed (6, p. 53).

The Task Force also recognizes that the collection of data may take a long time and may deter the development of genetic tests.

Once protocols are in place to address questions of clinical validity, safety, and effectiveness, and preliminary evidence justifies expanding testing, the test could be made widely available to providers and consumers on condition that they will adhere to the protocols—for example, by agreeing to be recontacted periodically or enrolling in clinical studies to assess post-test interventions. If the FDA has jurisdiction (for example, in the development of kits), it could grant conditional premarket approval and allow the developer to include a profit in its price, which it does not allow in a strictly investigational stage. The developer essentially agrees to continue to collect data proactively after the test is marketed (postmarket surveillance).

Genetic tests also have risks. Some genetic tests are imperfect predictors of future disease. The uncertainty is sufficiently troublesome to lead some people to forego testing (*16*). Others, however, will want to be tested despite the uncertainly. Some who are tested may not foresee the impact on them of getting a positive test result for an untreatable condition (*17*). On the other hand, in getting a negative result for a common complex disease, some people may not appreciate that they still are at risk of the disease. Problems of test uncertainty are not unique to predictive genetic testing. Providers and consumers may not be sufficiently aware of the imperfect validity of many tests.

The public is also concerned about other risks of genetic testing, notably genetic discrimination and invasion of privacy. These latter concerns have been considered recently (*18*) and will not be discussed further, except to say that before testing, people must be advised of them, as well as of the uncertainty of test results, so they can make an informed decision about being tested.

External Review

Researchers sometimes take it on themselves to collect some of the requisite information on validity, risks, and benefits before making a test available in health care or, when benefits are unclear, to introduce other safeguards. This was the case, for example, after the gene for Huntington disease was localized (*19*). On the other hand, tests have been introduced into health care prematurely. Carrier screening for cystic fibrosis was made routinely available despite professional statements that clinical sensitivity was inadequate (*20*). A commercial developer of genetic testing for inherited susceptibility to breast cancer made testing broadly available (*21*) without publishing its own preliminary data on test validity and despite statements from consumer and professional organizations that testing should remain investigative (*22*). One company briefly made testing for apolipoprotein E4 (apoE4) available as a predictor of Alzheimer's disease (AD), in part prompting several professional societies to state that with poor PVP and no preventive modalities, such testing was inappropriate (*23*).

Investigative studies are needed to ensure the collection of adequate data. IRBs are the most appropriate organizations to consider whether the scientific merit of protocols for the development of genetic tests warrant the risk to subjects who are invited to participate in investigative studies. The Task Force states that "[p]rotocols for the development of genetic tests that can be used predictively must receive the approval of an institutional review board (IRB) when subject identifiers are retained and when the intention is to make the test readily available for clinical use, i.e., to market the test. IRB review should consider the adequacy of the protocol for: (a) the protection of human subjects involved in the study, and (b) the collection of data on analytic and clinical validity, and data on the test's utility for individuals who are tested . . . Tests under development must be conducted in CLIA-certified laboratories if the results will be reported to patients or their providers" (6, p. 30). This applies to both academic and commercial test developers. The Task Force recognizes that the number of genetic tests being developed could overwhelm IRBs. Therefore, it recommends that tests be prioritized for review on the basis of characteristics such as their ability to predict future disease in healthy or apparently healthy people, their likely use for predictive purposes, and the absence of independent confirmatory tests (*24*).

Decisions about using new genetic or other tests are often left to individual physicians who seldom have access to data on a test's validity and utility. They and their patients would benefit from having an appraisal of such data, and indications for a test's use, by an organization independent of the test developer. Consequently, the Task Force recommends that test developers submit their validation and clinical utility data to independent external review. Such review (akin to technology assessment) can be undertaken by professional organizations, health care insurers (who will reimburse for the test), government agencies, and consensus panels (*25*). It is obviously an impossible task to subject all new predictive tests, even all predictive genetic tests, to external review. A scheme similar to the one for prioritizing IRB review is needed here as well.

To develop policies for the implementation of these and other of its recommendations, the Task Force calls on the Secretary of Health and Human Services to establish an advisory committee on genetic testing in the Office of the Secretary. Like the Task Force, its members should represent the various stakeholders in genetic testing and have strong liaison with government agencies.

Few scientists engaged in basic gene discovery are likely to divert their efforts to clinical applications, such as test development. Through patents they may file or hold, however, they have a stake in test development. Moreover, in promoting their work, the claims they make about practical applications, either directly or in response to journalists' questions, may lead to public expectations that cannot be satisfied until additional studies are performed. Scientists can use their stature in our society to ensure that genetic testing will be safe and effective.

REFERENCES AND NOTES

1. N. A. Holtzman, *Account. Res.* **5**, 95 (1996). Headlines and news stories sometimes raise expectations of the clinical applications of new discoveries that cannot be fulfilled without additional study. See, for example, "Rare breast cancer link is tied to European-Jewish ancestry; discovery said to promise community screening test," *Los Angeles Times* (news service), *Baltimore Sun*, 29 September 1995, p. 3A; "Gene causing colon cancer found; discovery at Hopkins expected to save thousands of lives," *Baltimore Sun*, 6 May 1993, p. 1.
2. Public Law 100-578, Clinical Laboratory Improvement Amendments of 1988, 42 USC 263a *et seq.*
3. Although the FDA has stated that genetic tests fall under its regulatory authority as "medical devices," the agency has chosen only to regulate clinical laboratory tests marketed as kits or other tangible products, not those marketed as services.
4. Sensitivity is the probability that people who have or will develop a disease can be detected by the test. PVP is the probability that a person with a positive test result will develop the disease the test is designed to detect. Often, a test can be used for other purposes (off-label use). For example, in addition to being used predictively to detect those at risk of the Li-Fraumeni syndrome, a test for p53 mutations can be used to diagnose the syndrome in individuals with cancer and a suggestive family history, or to stage a tumor by detecting somatic p53 mutations. The requirements for sensitivity and PVP could differ depending on the intended use.
5. Analytical validity is the ability of the test to correctly identify the analyte it is designed to detect. In the CLIA certification process, a clinical laboratory must demonstrate the tests' analytical validity to outside surveyors who inspect laboratories approximately every 2 years.
6. Task Force on Genetic Testing, *Promoting Safe and Effective Genetic Testing in the United States, Final Report*, N. A. Holtzman and M. S. Watson, Eds. (National Institutes of Health, Bethesda, MD, 1997). Also available at www.nhgri.nih.gov.
7. E. S. Lander and N. J. Schork, *Science* **265**, 2037 (1994).
8. Over 150 BRCA1 ISMs have been identified, but they do not account for all patients with linkage to chromosome 17q21 [F. J. Couch, B. L. Weber, the Breast Cancer Information Core, *Hum. Mutat.* **8**, 8 (1996)]. Locus heterogeneity also occurs, further reducing sensitivity. At least two genes, BRCA1 and BRCA2, increase susceptibility to breast cancer [B. Healy, *N. Engl. J. Med.* **336**, 1448 (1997)], and five to colon cancer [S. J. Laken *et al.*, *Nature Genet.* **17**,

79 (1997); N. W. Toribara and M. H. Sleisenger, *N. Engl. J. Med.* **332**, 861 (1995)].
9. The first report of linkage of familial breast cancer to chromosome 17q21 [J. M. Hall *et al.*, *Science* **250**, 1684 (1990)] involved 23 extended families with an average of 6.3 cases per family. The study describing the risk of breast and ovarian cancer in BRCA1 mutation carriers was based on families in which a combination of at least four cases of ovarian cancer at any age or breast cancer before age 60 were present [D. F. Easton, D. Ford, D. T. Bishop, the Breast Cancer Linkage Consortium, *Am. J. Hum. Genet.* **56**, 265 (1995)].
10. A. Schatzkin, A. Goldstein, L. S. Freedman, *J. Natl. Cancer Inst.* **87**, 1126 (1995). Struewing *et al.* recently found that the risk of developing breast cancer among women drawn from the general population of Ashkenazi Jews before age 70 years was 56% [J. P. Struewing *et al.*, *N. Engl. J. Med.* **336**, 1401 (1997)], which is lower than the 85% risk of breast cancer in families with multiple affected members used in the research studies [D. Easton, *Nature Genet.* **16**, 210 (1997)]. The association of late-onset AD in people with the apoE4 polymorphism is greater in families with multiple affected members than in sporadic cases [S. Seshadri, D. A. Drachman, C. F. Lippa, *Arch. Neurol.* **52**,.1074 (1995); National Institute on Aging, *Lancet* **347**, 1091 (1996)].
11. M. H. Gaston *et al.*, *N. Engl. J. Med.* **314**, 1593 (1986); E. Vichinsky, D. Hurst, A. Earles, K. Kleman, B. Lubin, *Pediatrics* **81**, 749 (1988); N. A. Holtzman, R. A. Kronmal, W. van Doorninck, C. Azen, R. Koch, *N. Engl. J. Med.* **314**, 593 (1986).
12. B. Pasini, I. Ceccherini, G. Romeo, *Trends Genet.* **12**, 138 (1996)
13. E. Treacy, B. Childs, C. R. Scriver, *Am. J. Hum. Genet.* **56**, 359 (1995).
14. W. Burke *et al.*, *J. Am. Med. Assoc.* **277**, 915 (1997); W. Burke *et al.*, *ibid.*, p. 997.
15. N. A. Holtzman, *Adv. Oncol.* **13** (no. 1), 9 (1997).
16. E. S. Tambor *et al.*, *Am. J. Hum. Genet.* **55**, 626 (1994); R. T. Croyle, D. S. Dutson, V. T. Tran, Y. Sun, *Women's Health Res. Gender Behav. Pol.* **1**, 329 (1995).
17. N. S. Wexler, *FASEB J.* **6**, 2820, (1992).
18. K. L. Hudson, K. H. Rothenberg, L. B. Andrews, M. J. E. Kahn, F. S. Collins, *Science* **270**, 391 (1995); K. H. Rothenberg *et al.*, *ibid.* **275**, 1755 (1997); N. A. Holtzman and L. B. Andrews, *Epidemiol. Rev.* **19**, 163 (1997).
19. Until they were confident that locus heterogeneity did not exist, the Huntington researchers refused to make their probes available outside of investigative protocols [J. F. Gusella, *Nature* **320**, 329 (1986); *ibid.* **323**, 118 (1986)]. In addition, concern about the psychological impact of identifying people at risk of a fatal, untreatable disease led to pilot programs with extensive guidelines for counseling and support for those considering predictive testing, many of which have been retained as testing has become available in health care [C. De Somviele *et al.*, *J. Med. Genet.* **27**, 34 (1990)].
20. National Institutes of Health, *N. Engl. J. Med.* **323**, 70 (1990); American Society of Human Genetics, *Am. J. Hum. Genet.* **46**, 393 (1990).
21. O. Smith, *Nature Med.* **2**, 613 (1996).
22. American Society of Clinical Oncology, *J. Clin. Oncol.* **14**, 1730 (1996); American Society of Human Genetics Ad Hoc Committee, *Am. J. Hum. Genet.* **55**, 1 (1994).
23. American College of Medical Genetics and American Society of Human Genetics Working Group on ApoE and Alzheimer Disease, *J. Am. Med. Assoc.* **274**, 1627 (1995); National Institute on Aging, *Lancet* **347**, 1091 (1996).
24. Additional characteristics include the likelihood that a test will have low sensitivity (because of genetic heterogeneity) and low PVP (because of incomplete penetrance), and the absence of an intervention proven to be effective in those with positive test results. Tests for disorders of high prevalence, for screening populations, or for use selectively in ethnic groups with higher incidence or prevalence of the disorder are other characteristics that might increase the priority for review.
25. See, for example, R. R. Howell *et al.*, *NIH Consensus Development Conference Statement: Genetic Testing for Cystic Fibrosis*, in press.
26. Preparation of this paper was supported in part by the National Human Genome Research Institute. The views expressed in this paper are entirely those of the authors and do not necessarily represent the organizations with which they are affiliated.

Sequencing the Human Genome

Lee Rowen,* Gregory Mahairas, Leroy Hood

At the end of 1997, we are halfway through the time allotted for completing the Human Genome Project. The Human Genome Project aims to sequence the genomes of the human and selected model organisms, identify all of the genes, and develop the technologies required to accomplish these objectives. Significant progress has been made, particularly in identifying and mapping genes, developing a stable DNA-sequencing technology, and in building the computational tools required for the analysis of sequence data. Yet, the large-scale sequencing of the 3 billion base pairs of the human genome has barely begun (Table 1). Approximately 60 million base pairs have been analyzed to date. Of these, the longest contiguous stretch of human DNA sequence in a public database is less than 1.5 million base pairs (*1*). Here we discuss today's challenges for sequencing the human genome.

What Has Been Done So Far?

Gene identification. The expressed genes from hundreds of different human tissues have been partially sequenced after copying the messenger RNAs into complementary DNA libraries. About 800,000 of these so-called expressed sequence tags (ESTs) are available in public databases and at various Web sites (*2*). These represent perhaps 40,000 to 50,000 genes of the estimated total of 70,000 to 100,000 human genes. ESTs from a variety of model organisms are also available.

Mapping. Mapping requires the identification of unique genome markers [for example, ESTs or sequence-tagged sites (STSs)] and their localization to specific chromosomal sites. STSs are unique addresses generated by polymerase chain reaction primers that amplify just a single chromosomal site. Three techniques have been used for marker localization: genetic mapping (generally 1- to 10-Mb resolution), fluorescence in situ hybridization (~1-Mb resolution), and radiation hybrid mapping (down to 50-kb resolution). By means of these techniques, markers have been placed on average every 200 kb across the genome (*3*). Using STS landmarks to identify and order clones, researchers have constructed a framework physical map for most of the human genome from large inserts of human DNA cloned into yeast artificial chromosomes (YACs) in 1993 (*4*). A genetic map with more than 5000 highly polymorphic simple sequence repeats is also available (*5*).

Clone library construction. Human chromosomes cannot be sequenced directly. Rather, human DNA must be isolated, randomly fragmented, and cloned into vectors capable of stable propagation in a suitable host such as the bacterium *Escherichia coli* or yeast. Before sequencing, clones must be selected from libraries with chromosomal markers as probes, verified for their fidelity to the genome, and ordered in a minimal-overlapping tiling path spanning a portion of a chromosome. Several cloning systems with insert sizes varying from hundreds of base pairs to megabases have been successfully developed. The ideal clone library for genomic sequencing has the following features: (i) the clones are highly redundant, covering the entire human genome many times; (ii) the clone coverage is random and not biased toward or against specific regions of the genome; and (iii) the clones are stable, not subject to deletion or rearrangement during the propagation process. A signifi-

Table 1. Current state of genome sequence, as of September 1997.

Organism	Size (Mb)	Sequenced	Percent finished
Microbial genomes (~11)	0.6–4.2	0.6–4.2	100
E. coli	4.6	4.6	100
Yeast	13	13	100
Nematode	100	71	71
Drosophila	130	8	6
Mouse	3000	6	0.2
Human	3000	60	2

The authors are in the Department of Molecular Biotechnology, University of Washington, Post Office Box 357730, Seattle, WA 98195–7730, USA.

*To whom correspondence should be addressed.