Benefits and Risks of Emerging Genetic Technologies: The Need for Regulation

Nell A. Holtzman

Efforts to localize human genes to specific chromosomes and then to identify these genes and their role in specific diseases are increasing. These developments will expand the repertoire of genetic tests that are claimed to be predictive of future disease in the person being tested (presymptomatic tests) or in future offspring of those being tested (tests for reproductive options). In single-gene (mendelian) disorders, such tests may fail to detect all of the alleles capable of causing disease. In multifactorial disorders, positive results of tests for specific alleles will not always predict future disease because the other necessary factors may be absent. Policies are already in place to provide safe and effective clinical laboratory tests and interventions for subjects with positive test results, and to assure the quality of the laboratories performing such tests, but are not widely known or enforced for genetic tests.

Indexing Terms: heritable disorders/screening/quality control

This paper is not about genetic technology for detecting DNA of specific infectious agents. It is not about tests for human somatic cell mutation, although in the future, such tests could be very important for early detection of cancer. It is not about the diagnosis of human genetic disease, that is, the use of genetic tests to confirm clinical symptoms and signs. What I will discuss is predictive genetic testing, testing to detect inherited genetic variants—alleles—that will either cause future disease or increase susceptibility to it. I will first briefly classify genetic tests and genetic diseases and then turn to assessing the validity of predictive genetic tests. I will conclude by considering policies to provide safe and effective genetic tests and interventions in individuals with positive test results, and policies to assure the quality of the laboratories performing genetic tests.

Classification of Genetic Tests

Predictive genetic tests are either for presymptomatic detection or for assisting in reproductive planning or management of pregnancy. Presymptomatic testing is appropriate when an intervention that can prevent or allay overt disease is efficacious only when started before symptoms appear. There are few such interventions. One is the low-phenylalanine diet for phenylketonuria (PKU).1 We screen virtually all newborns in the US for PKU because benefit occurs only when this special diet is introduced before the disease becomes apparent clinically (1).

When no efficacious interventions for a genetic condition are available, predictive genetic tests provide people who are at risk of having children with that condition options they would otherwise not have. Such testing before pregnancy provides women or couples at risk the option of avoiding conception of an affected child. Such testing early in pregnancy provides the options of termination of pregnancy or preparing for the birth of an affected child.

Uncertainties about whether a condition will appear, how severe it will be, and whether interventions will be efficacious complicate decisions about the use of predictive genetic tests. For instance, the survival of children with cystic fibrosis (CF) is highly variable, despite improved therapy over the past decade (2). If trials of inserting the normal CF transmembrane regulator gene into lung epithelial cells of affected children prove effective, interest will shift from predictive genetic testing for reproductive planning to presymptomatic testing for early treatment.

Sometimes a family might be interested in both presymptomatic testing and tests for reproductive options. Suppose that a presymptomatic genetic test indicates that a 25-year-old woman carries an allele that greatly increases her chance of breast cancer. Although monitoring or preventive mastectomy or early therapeutic intervention might improve her outcome, she does not want any future daughter to go through the same tribulations. On average, 50% of her female children would inherit her allele. If she becomes pregnant, she could use prenatal diagnosis and consider abortion of female fetuses who have inherited the susceptibility-conferring allele. Predictive genetic testing for Huntington disease (HD) is already in use, both to detect presymptomatic individuals and to give them the option of avoiding having affected children—although few families have demonstrated interest in the latter use (3).

Classification of Genetic Disease

Let me turn now from types of genetic tests to a classification of genetic diseases. First are the single-gene disorders such as PKU, CF, sickle cell anemia, and hemophilia. The list is very large and will grow larger as we identify genes through the Human Genome Project. For the most part, these are rare disorders, but their pattern of inheritance is very clear, generally following the rules discovered by Mendel. Once we identify someone who has a disease-causing allele for one of these genes, we can give very precise predictions of the risk to...
future offspring, although we cannot predict severity of the disease as precisely.

Second are multifactorial disorders, in which an allele at a single gene locus may contribute to the occurrence of disease but is insufficient, and often not necessary, to cause the disease because other factors must also be present. For instance, possession of a certain allele at the BRCA1 gene locus on chromosome 17 does not mean a woman’s risk of developing breast cancer is 100%. Rather, her risk is ~80% by age 65—obviously much higher than women who don’t have that allele in the general population, but by no means certain (4).

Finally, some diseases may be single gene in origin (mendelian) in a few families, but multifactorial in most people who get that disease. A gene on chromosome 21 accounts for early-onset Alzheimer disease in a few families (5); whether that gene plays a role in other people is not yet known.

Each class of genetic disease presents somewhat different problems in assessing the value of tests. Because many single-gene disorders can result from multiple mutations at a single gene locus, tests of the DNA may be unable to detect all of those mutations. This means there is a sensitivity problem with this kind of testing. Furthermore, DNA-based tests for multifactorial disorders will be able to detect specific mutations, but will not predict with certainty whether individuals who have those mutations are going to develop the disease. Thus there are also problems of specificity and low predictive value of positive test results.

Issues in Genetic Testing

Predictive genetic tests are conducted either in families in which disease has already occurred or in the general population. As new genes are localized to specific regions of specific chromosomes but not yet specifically identified, linkage tests must be used; for linkage studies to be informative, several family members must agree to be tested. That is the situation today for the BRCA1 gene for breast cancer: Once the gene is identified and the specific mutations that can lead to cancer are elucidated, we have the opportunity of developing “direct” tests for those mutations. Such tests employ various different technologies, and more are being developed. The use of the polymerase chain reaction greatly facilitates testing. The costs of testing, even when looking for multiple mutations at the same time, are dropping fairly quickly. Consequently, population-based screening with direct tests for mutations could be offered at reasonable cost. Whether screening should be undertaken or not depends on assuring that the issues discussed in the following sections are addressed.

Direct tests for mutations often are initially validated in families in which the disease resulting from the mutation has already appeared. Although some information may be extrapolated from the findings of such studies to population-based screening, pilot screening studies in the target population will be needed for a full assessment of the validity of genetic tests and the follow-up interventions. Findings in high-risk families may not be representative of those in the population.

Test Sensitivity

How can the sensitivity of a test be established before the test is made widely available? This is an important question because of the genetic heterogeneity of most mendelian disorders. Genetic heterogeneity, as I use the term, takes two different forms: In one, alleles at more than one gene locus are capable of causing the same clinical entity; in the other, also referred to as allelic diversity, several different mutations at one gene locus are each capable of causing the same clinical symptomatology. With current technology, DNA-based tests may not detect all of the alleles accounting for either form of heterogeneity.

Until very recently, HD was detectable in families only through linkage studies; now the mutation has been identified (6). One concern of Gusella and his colleagues after they localized the HD gene by linkage studies to chromosome 4 was whether HD could be caused by genes at other loci (7). Consequently, they contacted several other investigators who had access to HD families and, through a collaboration, finally established the extremely high sensitivity of testing for linkage to their markers on chromosome 4. In the meantime, some clinicians not in their collaborative study were eager to use this test and asked Gusella for the probes. Gusella refused. The clinicians objected in a public letter (8):

If Wassermann had published his test for syphilis, which was far from reliable, in a way which delayed its application, neurosyphilis might now be more common than Huntington’s chorea. This infectious disease involves far more difficult problems in handling patients and their families; many individuals must have been distressed by investigations based on error, and there were probably a few suicides. But the disease is now rare in Northern Europe. In no field of effective medicine can techniques be applied without casualties.

This is a chilling statement. If one is unwilling to risk a few suicides, and suicide has been a problem in HD, the approach used by Gusella to establish sensitivity of his linkage tests was necessary.

A second example relating to genetic heterogeneity is CF. In the first series of papers reporting the discovery of the gene, one mutation was found to account for 68% of CF carriers (9). Do we consider this level of sensitivity sufficient to begin to test for carriers in the general population? Although the survival of those with CF is improving, and gene therapy holds hope for a cure, it is a debilitating disease for many. Carrier screening before or early in pregnancy provides the options of prenatal diagnosis and abortion. What are the implications of carrier testing of couples when the sensitivity of the test is 70%? If both parents are found to carry one of the detectable mutations they have a one in four chance of having a child with CF with each pregnancy. Because both parents’ mutations can be detected, so can the CF alleles that the fetus inherits. Thus prenatal diagnosis will be able to predict with certainty whether the fetus

CLINICAL CHEMISTRY, Vol. 40, No. 8, 1994 1653
has CF, is a carrier like the parents, or has inherited normal alleles at the CF locus from each parent. However, what if the carrier test identifies one parent as being a carrier? The negative test result in the other parent does not preclude him or her from carrying a CF mutation that cannot be detected by the test. Nor can prenatal diagnosis, should the woman become pregnant, establish any more than whether the fetus is a carrier of the detectable mutation. Given the carrier frequency in the white population of 4%, ~1 in 80 people in whom the test result is negative would in reality be a carrier of an undetectable CF mutation. The chance that a person detected by the assay as a carrier would mate with such an undetected person and have a child with CF is almost 1 in 300.

The American Society of Human Genetics (10) and the National Institutes of Health (11) have issued policy statements saying that, with such low sensitivity, CF screening in the general population should not be a standard of care. They urge pilot studies to examine the effects of imperfect sensitivity. Studies in which people are informed of the imperfect sensitivity will tell us the extent to which test uncertainty determines their interest in having screening (12).

Since the discovery of the CF gene, >300 mutations of it have been discovered. About 20 of those 300 account for ~90–95% of all the known cases. The remaining mutations occur in only one or a few families. Thus we can now offer screening of ~20 mutations, which will have a combined test sensitivity of ~90%.

When a disorder is associated with numerous mutations, tests for structural or functional alterations in gene products (enzymes or other proteins) may be more sensitive than tests for specific mutations in DNA. This is because the many genotypes capable of causing a mendelian disease usually operate through a final common pathway: changing the protein for which the gene encodes. Unfortunately, many of the affected proteins are not present in readily accessible tissues. New test technologies are in development to overcome this problem (13).

Sensitivity is also important in considering screening for genetic factors in disorders that are usually multifactorial in origin; the absence of that factor by no means rules out the possibility that the person will develop the disease. Even the alleles at a single gene locus that have a major effect (that is, their presence predicts future disease with high probability but not certainty) in common multifactorial disorders seldom contribute to more than 5–10% of all patients with the disease, e.g., inherited forms of breast and colon cancer. A small proportion of some diseases that are usually multifactorial in origin may be mendelian, e.g., an early-onset form of Alzheimer disease (5). If the risk attributable to the presence of a testable genetic factor is small, screening may not be an efficient approach to dealing with a multifactorial disorder. Moreover, telling people that their genetic test result is negative may give them a false sense of security. Before routine screening, studies should be undertaken to determine the role played by the genetic factor in a representative sample of those with the disease in question, and not to be limited to families in which the disease has occurred repeatedly. Studies of peoples' reactions to such screening are also critical, to ensure that more good than harm is accomplished (14).

Specificity and Positive Predictive Value

These parameters are less important than sensitivity for mendelian disorders. For such disorders, once a test detects a disease-causing mutation, the probability is high that, in appropriate dosage (e.g., double dose for a recessive disease), the disease will manifest itself clinically. For most mendelian disorders, however, positive test results cannot predict how severe that disease is going to be in any affected individual. People who inherit the same mutations for CF vary in the severity of their disease (15).

The problem of predictive value is much more serious for multifactorial disorders. Because other factors are also involved in the causation, one must be very conservative in making estimates of how frequently a positive test result for the genetic predisposing factor will predict future disease.

A recent report described an association between a specific allele for human leukocyte antigen (HLA) and chronic berylliosis (16). Companies in which workers handle beryllium might want to screen job applicants for the HLA allele and refuse to hire those who have it. However, not all of the workers who had chronic berylliosis had this particular HLA allele. Moreover, many workers exposed to beryllium and who were free of the disease also had the allele.

Another example involves the association of Alzheimer disease with a specific apoE4 allele at the apolipoprotein (apo) E gene locus on chromosome 19. The data in Table 1 are recalculated from another study (17), which was limited to ~40 families in which Alzheimer disease occurred after age 60. Ages of the family members with Alzheimer were 60 to 91 years; those without it were 66 to 94. Thus, many of the people in the unaffected group would already have shown signs of the

<p>| Table 1. Association of apolipoprotein E4 allele with late-onset Alzheimer disease in families with affected members. |
|---|---|---|---|---|---|</p>
<table>
<thead>
<tr>
<th>No. of apoE4 alleles</th>
<th>Affected (age 60–91)</th>
<th>Unaffected (age 65–94)</th>
<th>Total</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>21</td>
<td>2</td>
<td>23</td>
<td>8.07</td>
</tr>
<tr>
<td>1</td>
<td>55</td>
<td>63</td>
<td>118</td>
<td>2.84</td>
</tr>
<tr>
<td>1 or 2</td>
<td>76</td>
<td>65</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>74</td>
<td>93</td>
<td>1.00</td>
</tr>
<tr>
<td>95</td>
<td>139</td>
<td>234</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Sensitivity | = 76/95 = 80% |
| Specificity | = 74/139 = 53% |
| Predictive value of positive result | = 76/141 = 54% |

Calculated from data of Corder et al. (17).
disease if they were going to get it. Within these families, those who inherited two apoE4 alleles had an eightfold greater chance of developing Alzheimer than those who had inherited no apoE4 allele. Those who inherited one allele had almost triple the chance of disease, which also is highly significant statistically. Converting these data into sensitivity and specificity statements, however, gives a somewhat different picture, which cautions us about genetic testing. The sensitivity is 80%, the specificity 53%, and the positive predictive value only 54%. That is, of all the individuals with positive test results within these families, only about half had Alzheimer disease; most of the others were probably not going to get it. Flipping a coin would be almost as informative as doing the test. Within the general population, moreover, screening for apoE4 alleles might yield still lower results for sensitivity and positive predictive value. Some factors that may have made the likelihood of Alzheimer disease higher in these families may not be operating in many people in the general population.

Before we screen, therefore, it is critical that we establish sensitivity and specificity in populations comparable with the one on which we plan to target the test.

In considering the problems of false positives, we can learn from experience with the Wassermann test for syphilis mentioned earlier. Walsh McDermott, late professor of medicine at Cornell Medical School, discussed the consequences of not validating the test before its widespread use (18):

Of all those people yielding positive reactions, only about one half were actually syphilitic, but this ‘validation’, so to speak, of the test, this characterization of its inadequacies, was only performed decades after large-scale public health campaigns (including laws governing premarital examinations) had brought thousands of people under treatment .... These four or five decades, during which thousands of patients who did not have syphilis were subjected to the shame and dangers of antisyphilitic therapy, are not from the era of bleedings and leechings, but from the modern era of interventionist technology.

Safety and Efficacy

Genetic screening is not an end in itself. The intervention that follows can range from prenatal diagnosis after identification of a high-risk pregnancy by screening, and abortion of those fetuses found to be affected, to the institution of a drug or special diet to prevent the appearance of symptoms after symptomatic screening. Here I’ll limit this discussion to symptomatic screening. Because, by definition, the person identified is free of symptoms, the question arises: Will symptoms appear even if the person is not treated? If the positive predictive value of a test is low, then many people who are treated as a result of screening would never have developed symptoms in the absence of treatment. If, however, all those with positive test results are given an unproven therapy, we may mistakenly attribute the absence of symptoms to the efficacy of the intervention rather than to a false-positive test result.

Once again, syphilis proves instructive. In a Scandinavian study, people with primary syphilis defined clinically who were not treated with arsenicals were traced over their lifetime. Only about 20% of them went on to develop tertiary syphilis (19). So, even when the diagnosis has been established, the natural history of the disorder cannot be predicted clearly. I will return to this point when we consider the efficacy of interventions.

Assume that scientists have identified the BRCA1 gene and a few mutations at that locus that are strongly associated with breast cancer in high-risk families. It then would become possible to develop a test to screen women in the general population for those mutations. Although 80% of women in high-risk families who have those alleles are likely to develop breast cancer over their lifetimes, it is by no means certain that the same mutations—or other mutations at the same locus—will indicate such a poor prognosis in the general population. Given the likelihood that somatic mutations are needed before breast cancer occurs in those who inherit a BRCA1 mutation (4), it is quite possible that the presence of other genetic or environmental factors explains the high likelihood in high-risk families. Some or all of these factors may be lacking in a woman found by population screening to have a BRCA1 mutation. If screening identifies women with these mutations but without a family history of breast cancer, what should these women be told? Without following them over a life-time, we can’t collect data on what their risk of developing cancer would be. If we offered them treatment, such as bilateral total mastectomy (although the efficacy of mastectomy in preventing cancer in residual breast tissue has not been established), we wouldn’t be able to tell how many would go on to develop cancer. We might think that the number of women so treated who then did not develop cancer was evidence that surgery was a “cure,” when, in fact, many of the women undertaking it might not have developed cancer anyway.

One way around this dilemma is to enroll women into combined pilot screening studies and randomized controlled trials in which it is explained to them that we don’t know the risk of cancer when the screening test result is positive and that we don’t know the efficacy of bilateral mastectomy (or whatever intervention is on trial) in preventing cancer. Ideally, women with positive screening test results would be randomized to prophylactic surgery or to a control group. The rate of breast cancer in the positive control group would tell us the positive predictive value of the screening test. A comparison of the rates in the two groups would indicate the efficacy of mastectomy.

This kind of pilot program/randomized trial might take a long time to yield conclusive results. The question arises, should testing be made available routinely and interventions given to all women who want them, once a cohort of sufficient size is randomized but before results are available? I’ll discuss that question shortly.

Policy and Regulatory Issues

Does the regulatory atmosphere today facilitate or require studies to assess validity of tests and efficacy of
interventions before they become widespread, and is the quality of the laboratories providing tests assumed?

Validity and Efficacy

The US Food and Drug Administration (FDA) considers tests to be medical devices and subject to scrutiny by the FDA if they are to be sold as kits or reagents for clinical purposes. Today, however, many commercial laboratories do not sell genetic test kits but are offering genetic tests as services, performed in their own laboratories. Such services are not directly covered by FDA regulations, but if the reagents or probes used have not been approved by the FDA for clinical use, they fall under FDA regulations for investigational use (20). This requirement is not widely known and not widely enforced. University research laboratory directors often are innocently ignorant of the requirements for investigational use as they begin to use tests developed for research in clinical settings. Under FDA regulations, a laboratory using an “investigational use” reagent cannot make a profit on tests that involve the reagent and must submit a protocol to a human research Institutional Review Board to obtain permission for use as well as present results to the Board periodically. If the laboratory is not testing anonymously, it must obtain informed consent from the patient.

Given the amount of interest in developing new genetic tests, it is not clear, either in commercial laboratories or in some academic laboratories, that rigorous premarket or preroutine-use studies will be conducted. In some situations, collecting data on test validity and efficacy will, as we have discussed, take a long time. The difficulties of such testing may be ameliorated by a proposed federal policy that permits provisional release of promising tests or interventions before the final results of validation or efficacy studies are in. Under this policy, once a test’s sponsor enrolls a cohort of sufficient size, e.g., in a randomized trial to determine both the positive predictive value of a test and the efficacy of an intervention, the manufacturer might be allowed to market the test or intervention and be able to get a fair price for either of them. However, the manufacturer would have very stringent postmarketing reporting responsibilities so the FDA could keep track of test validity and efficacy of interventions. Depending on the data, the FDA would have the authority to terminate use of the test or give full premarket approval.

Laboratory Quality

Any test that is used in the clinical management of a patient falls under the Clinical Laboratory Improvement Amendments of 1988 (CLIA88). CLIA88 is directed at the quality of laboratories, not the validity of tests (except as they are performed in an individual laboratory). If a laboratory performs a test whose validity has never been established in the first place, there is not much CLIA88 can do about it except to make sure that the test is being done in accord with the manufacturer’s protocol. Under the CLIA88 definitions of complexity levels of tests, most of the few genetic tests that have been classified are either of high or moderate complexity; that is, they require proficiency testing. At the moment, however, relatively few organizations are providing proficiency tests for DNA-based tests (20). A new test not yet classified under CLIA88 must be treated as a high-complexity test, and laboratories performing it must develop their own proficiency programs.

Because of the consequence and novelty of genetic tests, interpretation of results is critical for making sure a patient—or a referring physician—understands their implications. As already discussed, genetic tests have the chance of giving false-positive and false-negative results, and they often cannot predict how severe a disease is going to be. With relatively little training in genetics or in interpreting test results in general, some physicians providing tests may not understand the implications of results or may not communicate information accurately to consumers. I use the term consumers because we are talking about screening healthy people, not patients who are usually considered sick. Therefore, if is consumers who need to understand what they are getting into when they decide to have a test or not. In a study of primary care physicians’ knowledge of genetics and genetic tests, results were not as bad as we anticipated, particularly in specialties in which genetic tests are increasingly used, e.g., obstetrics (21). In our pilot CF study, we have also begun to look at consumers’ understanding of genetic tests, ways to improve their understanding (Bernhardt et al., paper submitted for publication). Because of limited consumer understanding, the process for obtaining informed consent becomes an important vehicle for explaining tests and making sure consumers understand their implications. Unlike many other laboratory tests, predictive genetic tests are offered to asymptomatic people who may not have thought about having them; it will often be the physician or health provider who introduces them. They are not foremost in the thoughts or on the tips of the tongues of most people obtaining healthcare today. Unfortunately, few physicians are willing or able to spent time to inform people and obtain truly informed consent.

In conclusion, the localization and identification of genes is, today, one of the most rapidly growing areas of research, as evidenced by increasing funding for the Human Genome Project. As associations between newly discovered genes and specific diseases are made, the temptation to screen people for the presence of disease-related alleles will be great. The issues discussed here should sound a cautionary note about the use of genetic tests for predictive purposes. If the public is to be protected, stringent application of existing clinical laboratory regulations to genetic testing will be needed as well as new policies to assure long-term validation of test results and efficacy of interventions.

This work was made possible by grants R01-HG00026 and R01-HG00481 from the National Center for Human Genome Research, NIH.
References