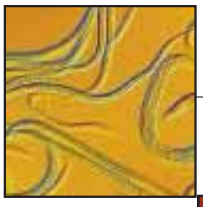
THE HUMAN GENOME BUSINESS TODAY



It's been a wild ride for the corporate and government parties who have deciphered the human genetic code. The fun has just begun

by Kathryn Brown

By the time this magazine hits your mailbox, you'll be able to read the entire genetic code of a human being over the Internet. It's not exactly light reading—start to finish, it's nothing but the letters A, T, C and G, repeated over and over in varying order, long enough to fill more than 200 telephone books. For biologists, though, this code is a runaway best-seller. The letters stand for the DNA chemicals that make up all your genes, influencing the way you walk, talk, think and sleep. "We're talking about reading your own instruction book," marvels Francis S. Collins, director of the National Human Genome Research Institute in Bethesda, Md. "What could be more compelling than that?"

Collins heads the Human Genome Project (HGP), so far a $250-million effort to write out the map of all our genes. The HGP is a publicly funded consortium that includes four large sequencing centers in the U.S., as well as the Sanger Center near Cambridge, England, and labs in Japan, France, Germany and China. Working together for more than a decade, over 1,100 scientists have crafted a map of the three billion DNA base pairs, or units, that make up the human genome. And they are not alone. In April a brash young company called Celera Genomics in Rockville, Md., beat the public consortium to the punch, announcing its own rough draft of the human genome. The rivalry has cast a spotlight on the human genetic code—and what, exactly, researchers now plan to do with it.

"For a long time, there was a big misconception that when the DNA sequencing was done, we'd have total enlightenment about who we are, why we get sick and why we get old," remarks geneticist Richard K. Wilson of Washington University, one partner in the public consortium. "Well, total enlightenment is decades away."

But scientists can now imagine what that day looks like. Drug companies, for instance, are collecting the genetic know-how to make medicines tailored to specific genes—an effort called pharmacogenomics. In the years to come, your pharmacist may hand you one version of a blood pressure drug, based on your unique genetic profile, while the guy in line behind you gets a different version of the same medicine. Other companies are already cranking out blood tests that reveal telltale disease-gene mutations—and forecast your chances of coming down with conditions such as Huntington's disease. And some scientists still hold out hope for gene therapy: directly adding healthy genes to a patient's body. "Knowing the genome will change the way drug trials are done and kick off a whole new era of individualized medicine," predicts J. Craig Venter, president of Celera.

Even with the human code in hand, however, the genomics industry faces challenges. Some are technical: it's one thing to know a gene's chemical structure, for instance, but quite another to understand its actual function. Other challenges are legal: How much must you know about a gene in order to patent it? And finally, many dilemmas are social: Do you really want to be diagnosed with a disease that can't be treated—and won't affect you for another 20 years? As scientists begin unraveling the genome, the endeavor may come to seem increasingly, well, human.

The "Race"

This spring all eyes were on the first finish line in the genome: a rough-draft sequence of the 100,000 or so genes inside us all. The HGP's approach has been described as painstaking and precise. Beginning with blood and sperm cells, the team separated out the 23 pairs of chromosomes that hold human genes. Scientists then clipped bits of DNA from every chromosome, identified the sequence of DNA bases in each bit, and, finally, matched each snippet up to the DNA on either side of it in the chromosome. And on they went, gradually crafting the sequences for individual gene segments, complete genes, whole chromosomes and, eventually, the entire genome. Wilson compares this approach to taking out one

SINCLAIR STAMMERS *SPL/Photo Researchers, Inc.* (*nematode*); ANDREW SYRED *SPL/Photo Researchers, Inc.* (*yeast*)

CELERA GENOMICS's gene-sequencing factory in Rockville, Md., has 300 automated DNA sequencers—as well as a nifty blue DNA helix on the ceiling.

page of an encyclopedia at a time, ripping it up and putting it together again.

In contrast, Celera took a shorter route: shredding the encyclopedia all at once. Celera's so-called shotgun sequencing strategy tears all the genes into fragments simultaneously and then relies on computers to build the fragments into a whole genome. "The emphasis is on computational power, using algorithms to sequence the data," says J. Paul Gilman, Celera's director of policy planning. "The advantage is efficiency and speed."

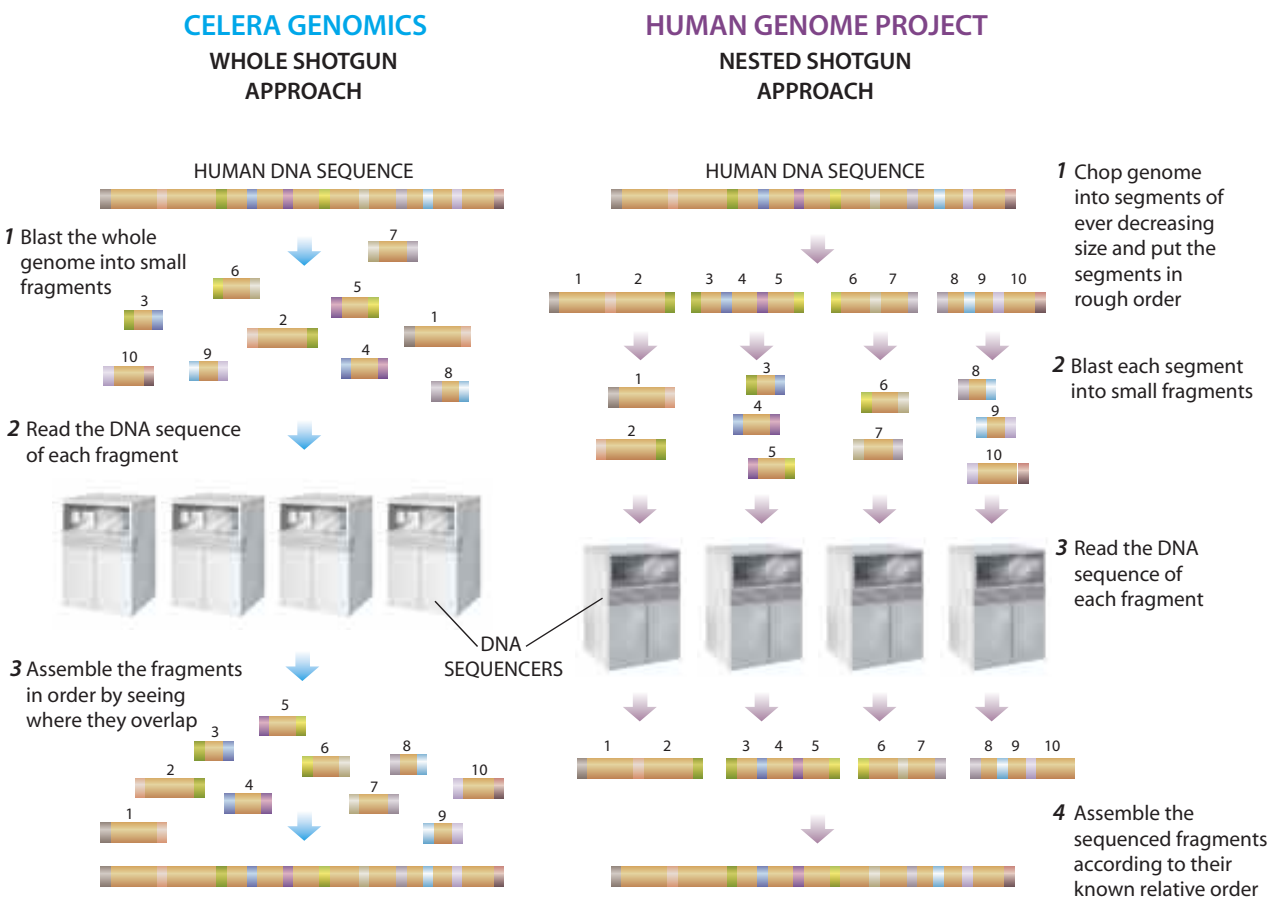
The HGP and Celera teams disagree over what makes a "finished genome." This spring Celera announced that it had finished sequencing the rough-draft genome of one anonymous person and that it would sort the data into a map in just six weeks. But the public team immediately cried foul, as Collins noted that Celera fell far short of its original genome-sequencing goals. In 1998, when the company began, Celera scientists planned to sequence the full genomes of several people, checking its "consensus" genome 10 times over. In its April announcement, however, Celera declared that its rough genome sequencing was complete with just one person's genome, sequenced only three times.

Although many news accounts have characterized the HGP and Celera as competing in a race, the company has had a decided advantage. Because the HGP is a public project, the team routinely dumps all its genome data into GenBank, a public database available through the Internet (at www.ncbi.nlm.nih.gov/). Like everyone else, Celera has used that data—in its case, to help check and fill the gaps in the company's rough-draft genome. Essentially Celera used the public genome data to stay one step ahead in the sequencing effort. "It does stick in one's craw a bit," Wilson remarks. But Gilman asserts that Celera's revised plan simply makes good business sense. "The point is not just to sit around and sequence for the rest of our lives," Gilman adds. "So, yes, we'll use our [threefold] coverage to order the public data, and that will give us what we believe to be a very accurate picture of the human genome." In early May the HGP announced it had completed its own working draft as well as a finished sequence for chromosome 21, which is involved in Down's syndrome and many other diseases. (For a full account of the chromosome 21 story, go to www.sciam.com/explorations/2000/051500chrom21 on the World Wide Web.)

Until now, the genome generators have focused on the similarities among us all. Scientists think that 99.9 percent of your genes perfectly match those of the person sitting beside you. But the remaining 0.1 percent of your genes vary—and it is these variations that most interest drug companies. Even a simple single-nucleotide polymorphism (SNP)—a T, say, in one of your gene sequences, where your neighbor has a C—can spell trouble.

KAY CHERNUSH

The Two Genome-Sequencing Strategies



Because of these tiny genetic variations, Venter claims, many drugs work only on 30 to 50 percent of the human population. In extreme cases, a drug that saves one person may poison another. Venter points to the type II diabetes drug Rezulin, which has been linked to more than 60 deaths from liver toxicity worldwide. "In the future, a simple genetic test may determine whether you're likely to be treated effectively by a given drug or whether you face the risk of being killed by that same drug," Venter predicts. While fleshing out its rough genome, Celera has also been comparing some of the genes with those from other individuals, building up a database of SNPs (pronounced "snips").

Other companies, too, hope to cash in on pharmacogenomics. Drug giants are partnering with smaller genomics-savvy companies to fulfill their gene dreams: Pfizer in New York City has paired with Incyte Genomics in Palo Alto, Calif.; SmithKline Beecham in Philadelphia has ties to Human Genome Sciences in Rockville; and Eli Lilly in Indianapolis has links to Millennium Pharmaceuticals in Cambridge, Mass. At this point, personalized medicine is still on the lab bench, but some business analysts say it could become an $800-million market by 2005. As Venter puts it: "This is where we're headed."

But the road is sure to be bumpy. One sticking point is the use of patents. No one blinks when Volvo patents a car design or Microsoft patents a software program, according to John J. Doll, director of the U.S. Patent and Trademark Office's biotechnology division. But many people are offended that biotechnology companies are claiming rights to human DNA—the very stuff that makes us unique. Still, without such patents, a company like Myriad Genetics in Salt Lake City couldn't afford the time and money required to craft tests for mutations in the genes *BRCA1* and *BRCA2*, which have been linked to breast and ovarian cancer. "You simply must have gene patents," Doll states.

Most scientists agree, although some contend that companies are abusing the public genome data that have been so exactingly sequenced—much of them with federal dollars. Dutifully reporting their findings in GenBank, HGP scientists have offered the world an unparalleled glimpse at what makes a human. And Celera's scientists aren't the only ones peering in—in April, GenBank logged roughly 35,000 visitors a day. Some work at companies like Incyte, which mines the public data to help build its own burgeoning catalogue of genes—and patents the potential uses of those genes. Incyte has already won at least 500 patents on full-length genes—more than any other genomics company—and has applied for roughly another 7,000 more. Some researchers complain that such companies are patenting genes they barely understand and, by doing so, restricting future research on

LAURIE GRACE, WITH ASSISTANCE FROM ERIC LANDER *Whitehead Institute*

The "Other" Genomes

Comparatively simple organisms are being harnessed to find new drugs for humans

by Julia Karow

What do we have in common with flies, worms, yeast and mice? Not much, it seems at first sight. Yet corporate and academic researchers are using the genomes of these so-called model organisms to study a variety of human diseases, including cancer and diabetes.

The genes of model organisms are so attractive to drug hunters because in many cases the proteins they encode closely resemble those of humans—and model organisms are much easier to keep in the laboratory. "Somewhere between 50 and 80 percent of the time, a random human gene will have a sufficiently similar counterpart in nematode worms or fruit flies, such that you can study the function of that gene," explains Carl D. Johnson, vice president of research at Axys Pharmaceuticals in South San Francisco.

Here's a rundown on the status of the genome projects of the major model organisms today:

The Fruit Fly

The genome sequence for the fruit fly *Drosophila melanogaster* was completed this past March by a collaborative of academic investigators and scientists at Celera Genomics in Rockville, Md.

The researchers found that 60 percent of the 289 known human disease genes have equivalents in flies and that about 7,000 (50 percent) of all fly proteins show similarities to known mammalian proteins.

One of the fly genes with a human counterpart is *p53*, a so-called tumor suppressor gene that when mutated allows cells to become cancerous. The *p53* gene is part of a molecular pathway that causes cells that have suffered irreparable genetic damage to commit suicide. In March a group of scientists, including those at Exelixis in South San Francisco, identified the fly version of *p53* and found that—just as in human cells—fly cells in which the P53 protein is rendered inactive lose the ability to self-destruct after they sustain genetic damage and instead grow uncontrollably. Similarities such as this make flies "a good trade-off" for studying the molecular events that underlie human cancer, according to one of the leaders of the fly genome project, Gerald M. Rubin of the Howard Hughes Medical Institute at the University of California at Berkeley: "You can do very sophisticated genetic manipulations [in flies] that you cannot do in mice because they are too expensive and too big."

The Worm

When researchers deciphered the full genome sequence of the nematode *Caenorhabditis elegans* in 1998, they found that roughly one third of the worm's proteins—more than 6,000—are similar to those of mammals. Now several companies are taking advantage of the tiny size of nematodes—roughly one millimeter—by using them in automated screening tests to search for new drugs.

To conduct the tests, scientists place between one and 10 of the microscopic worms into the pill-size wells of a plastic microtiter plate the size of a dollar bill. In a version of the test used to screen for diabetes drugs, the researchers use worms that have a mutation in the gene for the insulin receptor that causes them to arrest their growth. By adding various chemicals to the wells, the scientists can determine which ones restore the growth of the worms, an indication that the compounds are bypassing the faulty receptor. Because the cells of many diabetics no longer respond to insulin, such compounds might serve as the basis for new diabetes treatments.

The Yeast

The humble baker's yeast *Saccharomyces cerevisiae* was the first organism with a nucleus to have its genetic secrets read, in 1996.

Approximately 2,300 (38 percent) of all yeast proteins are similar to all known mammalian proteins, which makes yeast a particularly good model organism for studying cancer: scientists first discovered the fundamental mechanisms cells use to control how and when they divide using the tiny fungus.

"We have come to understand a lot about cell division and DNA repair—processes that are important in cancer—from simple systems like yeast," explains Leland H. Hartwell, president and director of the Fred Hutchinson Cancer Research Center in Seattle and co-founder of the Seattle Project, a collaboration between academia and industry. So far Seattle Project scientists have used yeast to elucidate how some of the existing cancer drugs exert their function. One of their findings is that the common chemotherapeutic drug cisplatin is particularly effective in killing cancer cells that have a specific defect in their ability to repair their DNA.

The Mouse

As valuable as the other model organisms are, all new drugs must ultimately be tested in mammals—and that often means mice. Mice are very close to humans in terms of their genome: more than 90 percent of the mouse proteins identified so far show similarities to known human proteins.

Ten laboratories across the U.S., called the Mouse Genome Sequencing Network, collectively received $21 million from the National Institutes of Health last year to lead an effort to sequence the mouse genome. They have completed approximately 3 percent of it, and their goal is to have a rough draft ready by 2003. But that timeline might be sped up: Celera announced in April that it is turning its considerable sequencing power to the task.

JULIA KAROW is an intern at Scientific American.

ALFRED PASIEKA *Peter Arnold, Inc. (fly)*; THE JACKSON LABORATORY *(mouse)*

THE MAJOR PLAYERS

Celera Genomics

A division of PE Corp.

www.celera.com
Stock Symbol: CRA
Headquarters: Rockville, Md.
Lead Executive: J. Craig Venter, president
Major Clients/Partners: Pfizer, Pharmacia, Novartis, Amgen and Takeda Chemical Industries
Strategy: Sell subscriptions to various annotated genomes on-line.
Financing This Year: $900 million
Key Challenge: Building a business around genome databases.
Competitive Advantages: Extensive DNA-sequencing infrastructure and a large amount of capital.

Human Genome Sciences

www.hgsi.com
Stock Symbol: HGSI
Headquarters: Rockville, Md.
Lead Executive: William A. Haseltine, chairman and CEO
Major Clients/Partners: SmithKline Beecham, Takeda Chemical Industries, Schering-Plough, Sanofi-Synthelabo and Merck
Strategies: Develop and market genomics-based drugs; provide drug targets to partners.
Financing This Year: $525 million
Key Challenge: Bringing genome-based drugs to market.
Competitive Advantages: Patents filed on more than 7,500 human genes; three genomic drugs in human clinical trials.

Incyte Genomics

www.incyte.com
Stock Symbol: INCY
Headquarters: Palo Alto, Calif.
Lead Executive: Roy A. Whitfield, CEO
Major Clients/Partners: 18 of the top 20 pharmaceutical companies
Strategy: Provide nonexclusive commercial access to genomic databases and sell access to DNA clones represented in the databases.
Financing This Year: $622 million
Key Challenge: Turning genomic information into sustainable business.
Competitive Advantage: A broad data set that includes gene sequences, patterns of gene and protein expression, and genetic variations among individuals.

Millennium Pharmaceuticals

www.mlnm.com
Stock Symbol: MLNM
Headquarters: Cambridge, Mass.
Lead Executive: Mark J. Levin, CEO
Major Clients/Partners: Bayer, Pharmacia, Pfizer and Eli Lilly
Strategies: Develop personalized therapeutics and medical tests; partner with biotech and drug firms in the field of pharmacogenomics.
Financing This Year: $700 million
Key Challenge: Translating genomic information into proprietary products, including drugs and tests.
Competitive Advantages: Existing alliances with drug developers; recently acquired LeukoSite.

The Human Genome Project

www.nhgri.nih.gov/HGP/
Headquarters: National Human Genome Research Institute (NHGRI), Bethesda, Md.
Joint Collaborators: NHGRI, Department of Energy (DOE) and Wellcome Trust
Lead Executives: Francis S. Collins, NHGRI; Ari Patrinos, DOE; and Michael Morgan, Wellcome Trust
Major Sequencing Centers: Washington University School of Medicine, St. Louis; Baylor College of Medicine, Houston; Sanger Center, Cambridge, England; Whitehead Institute, Cambridge, Mass.; DOE Joint Genome Institute, Walnut Creek, Calif.
Strategy: Map, sequence and annotate the human genome.
Grants Funded This Year: $112.5 million in 260 grants
Key Challenges: Understanding gene function; encouraging laws to ban genetic discrimination; teaching physicians to use genome information.
Competitive Advantages: Data available within 24 hours of sequencing, at no cost and with no restrictions, via GenBank. Also funding studies of the ethical, legal and social implications of genomics.

SLIM FILMS; SOURCES: THE COMPANIES LISTED HERE, NHGRI AND STELIOS PAPADOPOULOS *SG Cowen Securities*

those genes. "If data are locked up in a private database and only a privileged few can access it by subscription, that will slow discovery in many diseases," warns Washington University's Wilson.

Incyte president Randal W. Scott, however, sees things differently: "The real purpose of the Human Genome Project is to speed up research discoveries, and our work is a natural culmination of that. Frankly, we're just progressing at a scale that's beyond what most people dreamed of." In March, Incyte launched an e-commerce genomics program—like an amazon.com for genes—that allows researchers to order sequence data or physical copies of more than 100,000 genes on-line. Subscribers to the company's genomics database include drug giants such as Pfizer, Bayer and Eli Lilly. Human Genome Sciences has won more than 100 gene patents—and filed applications for roughly another 7,000—while building its own whopping collection of genes to be tapped by its pharmaceutical partners, which include SmithKline Beecham and Schering-Plough.

The federal government has added confusion to the patent debate. In March, President Bill Clinton and British prime minister Tony Blair released an ambiguous statement lauding open access to raw gene data—a comment some news analysts interpreted as a hit to Celera and other genomics companies that have guarded their genome sequences carefully. Celera and the HGP consortium have sparred over the release of data, chucking early talks of collaboration when the company refused to release its gene sequences immediately and fully into the public domain. The afternoon Clinton and Blair issued their announcement, biotech stocks slid, with some dropping 20 percent by day's end. A handful of genomics companies scrambled to set up press conferences or issue statements that they, indeed, did make available their raw genome data for free. In the following weeks, Clinton administration officials clarified that they still favor patents on "new gene-based health care products."

The sticky part for most patent seekers will be proving the utility of their DNA sequences. At the moment, many patent applications rely on computerized prediction techniques that are often referred to as "in silico biology" [see "The Bioinformatics Gold Rush," on page 58]. Armed with a full or partial gene sequence, scientists enter the data into a computer program that predicts the amino acid sequence of the resulting protein. By comparing this hypothetical protein with known proteins, the researchers take a guess at what the underlying gene sequence does and how it might be useful in developing a drug, say, or a diagnostic test. That may seem like a wild stab at biology, but it's often enough to win a gene patent. "We accept that as showing substantial utility," Doll says. Even recent revisions to federal gene-patent standards—which have generally raised the bar a bit on claims of usefulness—ask only that researchers take a reasonable guess at what their newfound gene might do.

Testing, Testing

Patents have already led to more than 740 genetic tests that are on the market or being developed, according to the National Institutes of Health. These tests, however, show how far genetics has to go. Several years after the debut of tests for *BRCA1* and *BRCA2*, for instance, scientists are still trying to determine exactly to what degree those genes contribute to a woman's cancer risk. And even the most informative genetic tests leave plenty of questions, suggests Wendy R. Uhlmann, president of the National Society of Genetic Counselors. "In the case of Huntington's, we've got a terrific test," Uhlmann avers. "We know precisely how the gene changes. But we can't tell you the age when your symptoms will start, the severity of your disease, or how it will progress."

Social issues can get in the way, too. After Kelly Westfall's mother tested positive for the Huntington's gene, Westfall, age 30, immediately knew she would take the test as well. "I had made up my mind that if I had Huntington's, I didn't want to have kids," declares Westfall, who lives in Ann Arbor, Mich. But one fear made her hesitate: genetic discrimination. Westfall felt confident enough to approach her boss, who reassured her that her job was safe. Still, she worried about her insurance. Finally, rather than inform her insurer about the test, Westfall paid for it—some $450, including counseling—out of pocket. (To her relief, she tested negative.)

The HGP's Collins is among those calling for legislation to protect people like Westfall. A patchwork of federal and state laws are already in place to ban genetic discrimination by insurers or employers, but privacy advocates are lobbying Congress to pass a more comprehensive law. Last February, President Clinton signed an executive order prohibiting all federal employers from hiring, promoting or firing employees on the basis of genetic information. It remains to be seen whether private companies will follow suit.

In the meantime, Celera is now ready to hawk its human genome, complete with crib notes on all the genes, to on-line subscribers worldwide. "It's not owning the data—it's what you do with it," Venter remarks. He envisions a Celera database akin to Bloomberg's financial database or Lexis-Nexis's news archives, only for the genetics set. Which 300 genes are associated with hypertension? What, exactly, does each gene do? These are the kinds of queries Celera's subscribers might pose—for a price. As of press time, Celera planned to offer a free peek at the raw genome data on-line, but tapping into the company's on-line toolkit and full gene notes will cost corporate subscribers an estimated $5 million to $15 million a year, according to Gilman. Academic labs will pay a discounted rate: $2,000 to $15,000 a year.

Internet surfers can now visit GenBank for free. With all this information available, will scientists really pay Celera? Venter thinks so. "We just have to have better tools," he says. For genomics, that is becoming a familiar refrain. SA

KATHRYN BROWN is a freelance writer based in Alexandria, Va.

Further Information

ARE SEQUENCERS READY TO ANNOTATE THE HUMAN GENOME? E. Pennisi in *Science*, Vol. 287, No. 5461, page 2183; March 24, 2000.

THE HUMAN GENOME PROJECT AND ITS IMPACT ON THE STUDY OF HUMAN DISEASE. E. Green in *Metabolic and Molecular Bases of Inherited Disease*. Edited by Charles R. Scriver. Eighth edition. McGraw-Hill, 2000.

For a primer on genetic testing and a directory of genetic tests, visit GeneTests at www.genetests.org

For more on the ethical, legal and social implications of human genome research, visit the National Human Genome Research Institute's Web site at www.nhgri.nih.gov/ELSI