SCIENTIFIC AMERICAN February 1988 Volume 258 Number 2

Chromosome Mapping with DNA Markers

Variable sequences in the DNA of human chromosomes act as genetic landmarks. Individual markers serve for tracing defective genes; collectively the markers provide the elements of a chromosome map

by Ray White and Jean-Marc Lalouel

Say that a disease is known to run in families, following a classic Mendelian pattern of inheritance. Somewhere among the 100,000 genes on the 23 pairs of human chromosomes a single gene is defective. The symptoms and progress of the disease have been described in meticulous detail, but its biochemistry is an enigma, and even predicting who will actually get the disease is guesswork. Such has been the case not just for a handful of rare afflictions but for most of the 3,000 known genetic diseases, including such familiar scourges as Huntington's disease and cystic fibrosis. Where does one begin the search for a causative mechanism, a diagnostic test and, ultimately, a treatment?

It is now possible to start by closing in on the defective gene itself. The territory to be surveyed is vast: the human chromosomes consist of linear molecules of double-strand DNA with a total length of about three billion base pairs (the chemical subunits that encode information along DNA). A typical gene, a complete unit of genetic information, is minuscule by contrast, encompassing perhaps 10,000 base pairs. And yet by correlating the inheritance of a distinctive segment of DNA—a "marker"—with the inheritance of a disease, one can now localize the mutant gene to within one or two million base pairs, or less than a thousandth of the human genome (the total complement of DNA). That kind of precision puts the gene within reach of molecular tools for cloning DNA and testing its activity. The identification of a genetic marker that is closely linked with a disease also means the gene's inheritance can be followed. It opens the way to simple tests for diagnosing carriers and future disease victims.

The basic strategy, known as linkage analysis, is a venerable tool of classical genetics. In our laboratory at the University of Utah and in many others, however, it has gained new power from the techniques of molecular biology, which make available a greatly expanded set of markers: molecular variations known as RFLP (for restriction-fragment length polymorphism) markers. Linkage analysis has now revealed RFLP markers for a number of disease genes, and many more diseases will soon yield to the strategy. It is also serving a more general purpose. By following the inheritance of many RFLP markers simultaneously in healthy families, we and other workers have begun to plot their positions in relation to one another and map them onto the physical framework of the chromosomes. The goal is a complete map of markers: an array of reference points that spans the genome and makes it possible to pinpoint disease genes far more efficiently than can be done with isolated markers.

The linkage strategy exploits the way genes are inherited. An ordinary human cell contains 23 pairs of homologous, or matching, chromosomes, one chromosome per pair inherited from the mother and the other from the father. In meiosis, the series of cell divisions that gives rise to germ cells (sperm or eggs), the homologous chromosomes in a progenitor cell are duplicated and then distributed among four germ cells, each of which receives 23 single chromosomes. The parental chromosomes are not transmitted intact, however. In the course of meiosis homologous chromosomes repeatedly recombine: they "cross over" and exchange segments of equal length [*see illustration on page 43*]. As a result each chromosome that is transmitted in a germ cell is generally a patchwork of segments from the two parental chromosomes. Recombination is the phenomenon that enables one to find linkage between a marker and a disease.

What makes it possible to detect recombination and employ it in linkage analysis are the many differences between homologous chromosomes. They often carry two different alleles, or versions, of many of their matching genes and also of many apparently meaningless DNA sequences within and between genes. The recombinant chromosomes that are parceled out to the germ cells at meiosis represent new combinations of these features. An allele from a locus on one chromosome and an allele from a different locus on the other, homologous

chromosome can be combined and passed on together; at the same time the alleles at two loci on a single chromosome can be separated, so that only one of them is inherited.

The closer together two loci lie on the same parental chromosome, the less often their alleles are separated as DNA is exchanged between homologous chromosomes during meiosis. Hence one can gain a measure of the distance between a gene of particular interest—one that has a disease-causing mutant allele, for example—and a marker by correlating the inheritance pattern of their alleles. If the individuals in an afflicted family who develop the disease almost always inherit the same version of the marker, the mutant gene and the marker must lie very close together on the same chromosome. The marker and the disease gene are said to be linked.

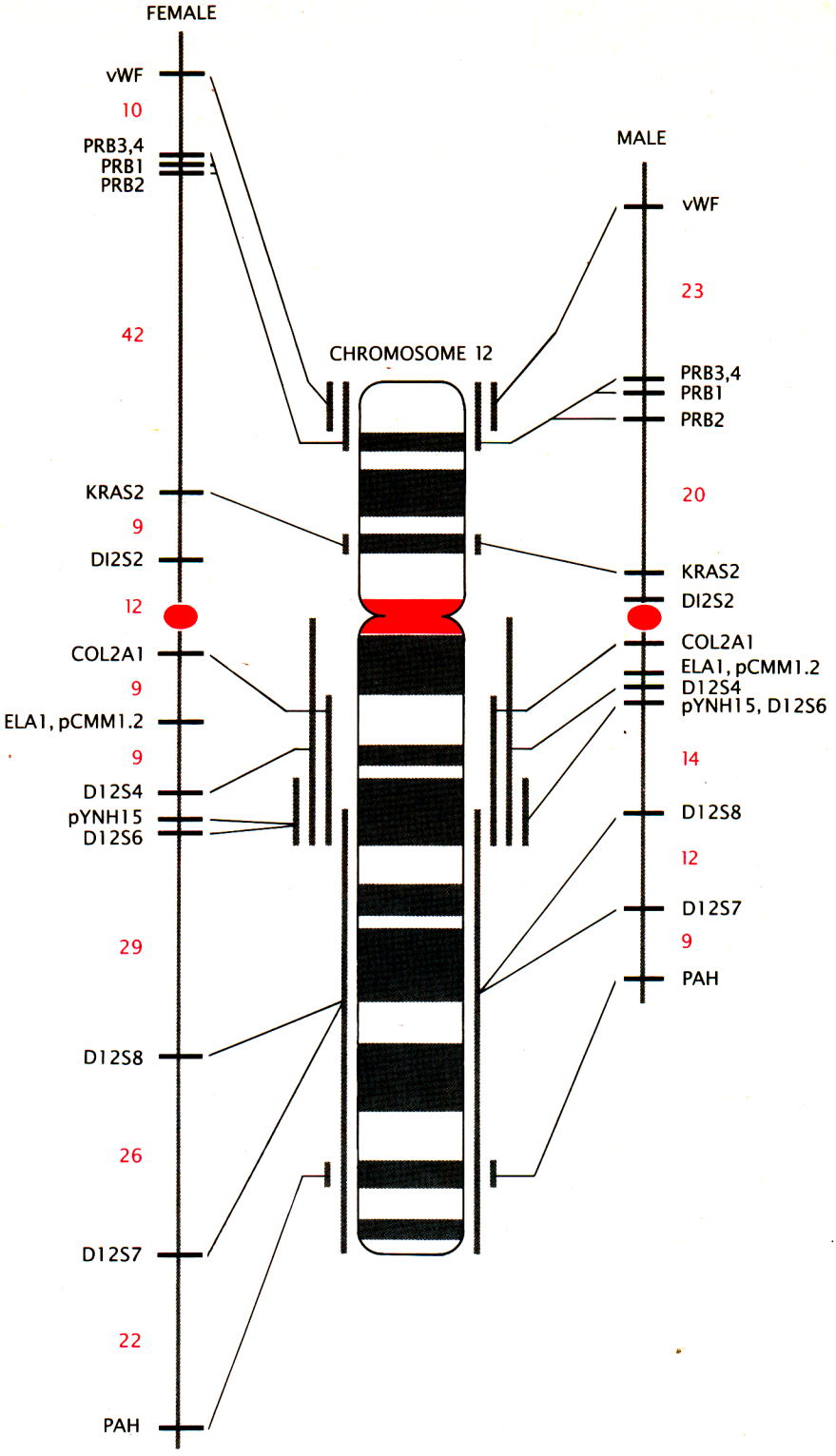
Other markers lying farther from the disease gene will recombine with the gene more frequently, so that the disease will be less likely to be inherited together with any given marker allele. In the extreme case, for a marker and a disease lying well apart on a chromosome, the recombination frequency reaches 50 percent. The marker and the gene are then unlinked: a given marker allele has only an even chance of being passed on with the disease. The same pattern of 50 percent coinheritance emerges when a marker and a mutant allele are borne on entirely different chromosomes.

Correlating the inheritance of a marker and a disease requires two things. The marker must be readily detectable, and it must be found in a number of distinguishable variants throughout the population. Linkage can be detected only if a person carrying mutant and normal alleles of a disease gene also carries two different versions of the marker; if the two marker alleles are indistinguishable, crossovers between the disease and the marker will be undetectable in the offspring. There will be no way to tell a linked marker from an unlinked one.

Until a few years ago only a limited set of markers met both criteria. The genes coding for certain enzymes, blood-group antigens (which determine blood type) and other proteins have multiple alleles, which manifest themselves by giving rise to protein polymorphisms: detectably different versions of the protein each gene codes for. Only 25 to 30 such marker systems of any value were known, however, covering only small sections of a few chromosomes. For want of markers most of the human genome remained inaccessible to the linkage approach.

With the advent of recombinant-DNA technology in the mid-1970's linkage mapping could be transformed into a practical and powerful tool for human genetics. The transformation can be dated to a genetics retreat sponsored by the University of Utah in April, 1978. There David Botstein of the Massachusetts Institute of Technology, Ronald W. Davis of Stanford University and Mark H. Skolnick of Utah proposed that the DNA sequence itself might yield numerous and readily detectable markers. Recognizing the potential power of the new approach, one of us (White) soon decided to test the hypothesis by committing his laboratory to the development of a set of DNA-based markers that would make it possible to detect linkage anywhere in the human genome. Botstein, White, Skolnick and Davis published the first paper detailing the approach in 1980. In the meantime

EXTENSIVE FAMILIES with living grandparents—modern counterparts to this turn-of-the-century family—are the ideal setting for studies of genetic linkage. In linkage studies the relative positions of sites in the chromosomes are inferred from the frequency with which genetic variations at those sites are passed on together from parents to children. By examining the inheritance of a genetic disease and arbitrary genetic markers in afflicted families one can assign a chromosomal location to the disease gene; by correlating inheritance of many markers in large, healthy families one can make maps of chromosomes.



MAP of chromosome 12 was made by tracing the inheritance of DNA markers: sites where the two copies of a chromosome often carry detectably different DNA sequences. The markers are arrayed in their statistically likeliest order and are separated by distances reflecting their recombination frequency, or the percent of the time marker versions carried on the same parental chromosome are separated by a recombination event during the formation of sperm or eggs. The recombination frequency between two markers rises with increasing physical separation, but the precise relation between recombination frequency and distance can vary depending on several factors, including sex. On chromosome 12, for example, the overall rate of recombination seen when the chromosome is passed on by a woman is higher than when it is passed on by a man, and so its genetic map is represented as being longer in women. An approximate chromosomal position has been determined for some of the DNA markers (*center*).

many other workers were beginning to find markers in human DNA and to speculate about their uses, and it was clear that this approach was an idea whose time had come.

The new linkage strategy gains its power from the very high level of normal polymorphism that can be found in the sequence of base pairs making up DNA. Between homologous chromosomes there is a difference in sequence, on the average, every 200 to 500 base pairs. Identifying these allelic variants would provide a practically limitless supply of markers scattered throughout the human chromosomes.

Molecular tools known as restriction enzymes provide a means of detection. Each restriction enzyme, made by a particular species of bacteria, binds to DNA wherever it finds a specific short sequence of base pairs and cleaves the molecule at a specific site within that sequence. A variation in DNA sequence that creates or eliminates a restriction site will alter the length of the resulting DNA fragment or fragments. The variation creates a restriction-fragment length polymorphism—an RFLP.

The RFLP defines a potential marker. A single restriction enzyme finds millions of cutting sites in the total human DNA, however. How can one or two variant fragments be detected among millions? The fragments are first sorted by electrophoresis: an electric field draws them through a gel, in which their mobility is inversely proportional to their length. A powerful and sensitive technique called Southern blotting after Edward M. Southern, who developed it at the University of Edinburgh, serves for picking out the fragments of interest.

Southern blotting relies on the unique character of the DNA molecule. The bases along two strands of DNA can pair only according to set rules, and so the sequence on one strand constitutes a unique match for the sequence on the other. A length of single-strand DNA can therefore act as a probe, detecting and binding to the complementary sequence in a sample of ordinary DNA that has been "denatured": heated or exposed to high *p*H in order to separate its strands. In Southern blotting the DNA fragments on an electrophoresis gel are denatured and blotted onto a membrane, where they are exposed to probe DNA labeled with a radioactive isotope. The probe hybridizes, or binds, only to the fragment or fragments that bear the complementary sequence of bases. The

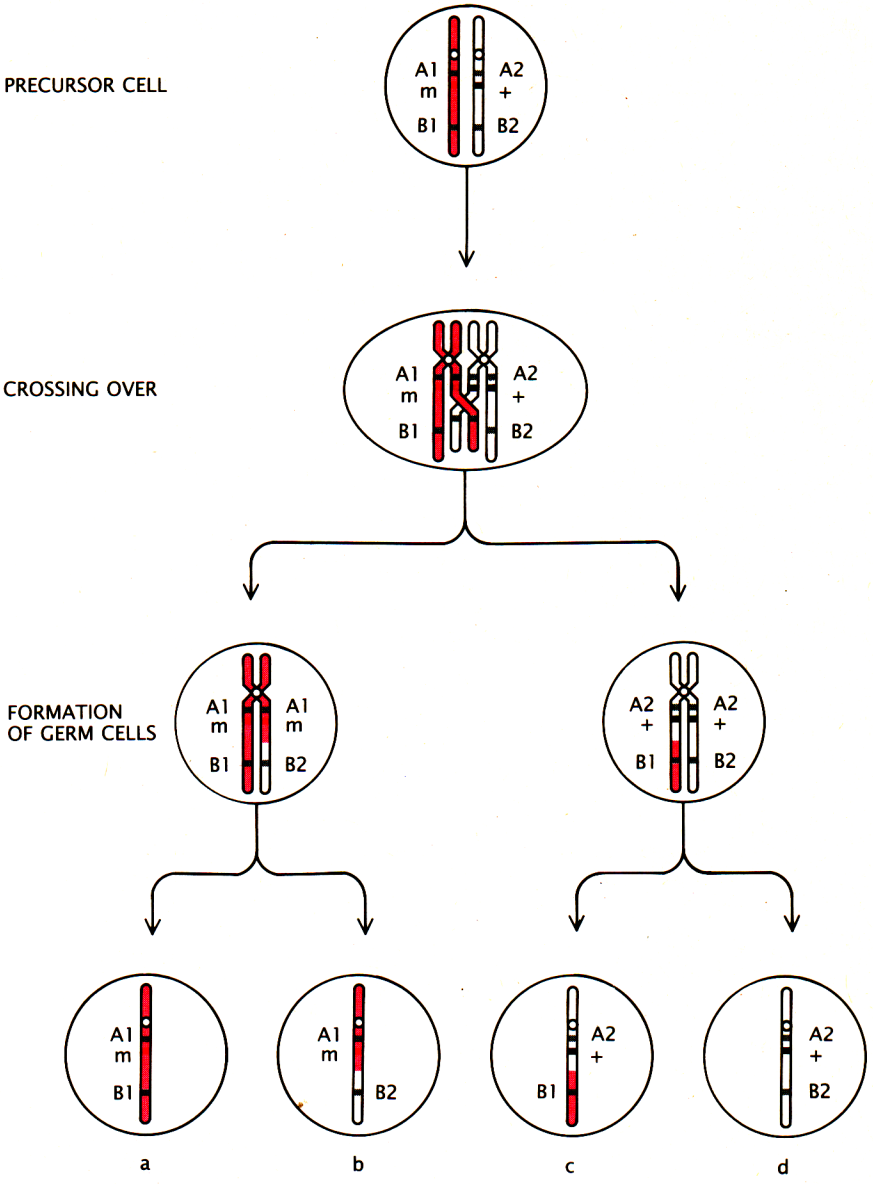
radioactive label makes it possible to detect the position of the fragments, which reveals their size.

To detect an RFLP, then, one needs to find a probe that is complementary to DNA near the restriction-enzyme cutting site. A segment of DNA is chosen, often at random, from a collection (a "library") of cloned DNA fragments representing the full human genome. It is denatured, made radioactive and applied to Southern blots of DNA samples that have been digested with a restriction enzyme. If the radioactive bands appear at different places on blots of DNA from different individuals, the cloned DNA has detected the variable cutting pattern that results from a DNA polymorphism. The probe and the RFLP it detects constitute a unique genetic marker system. With it one gains a point of reference in the genome: the short stretch of polymorphic DNA, whose inheritance pattern can now be traced.

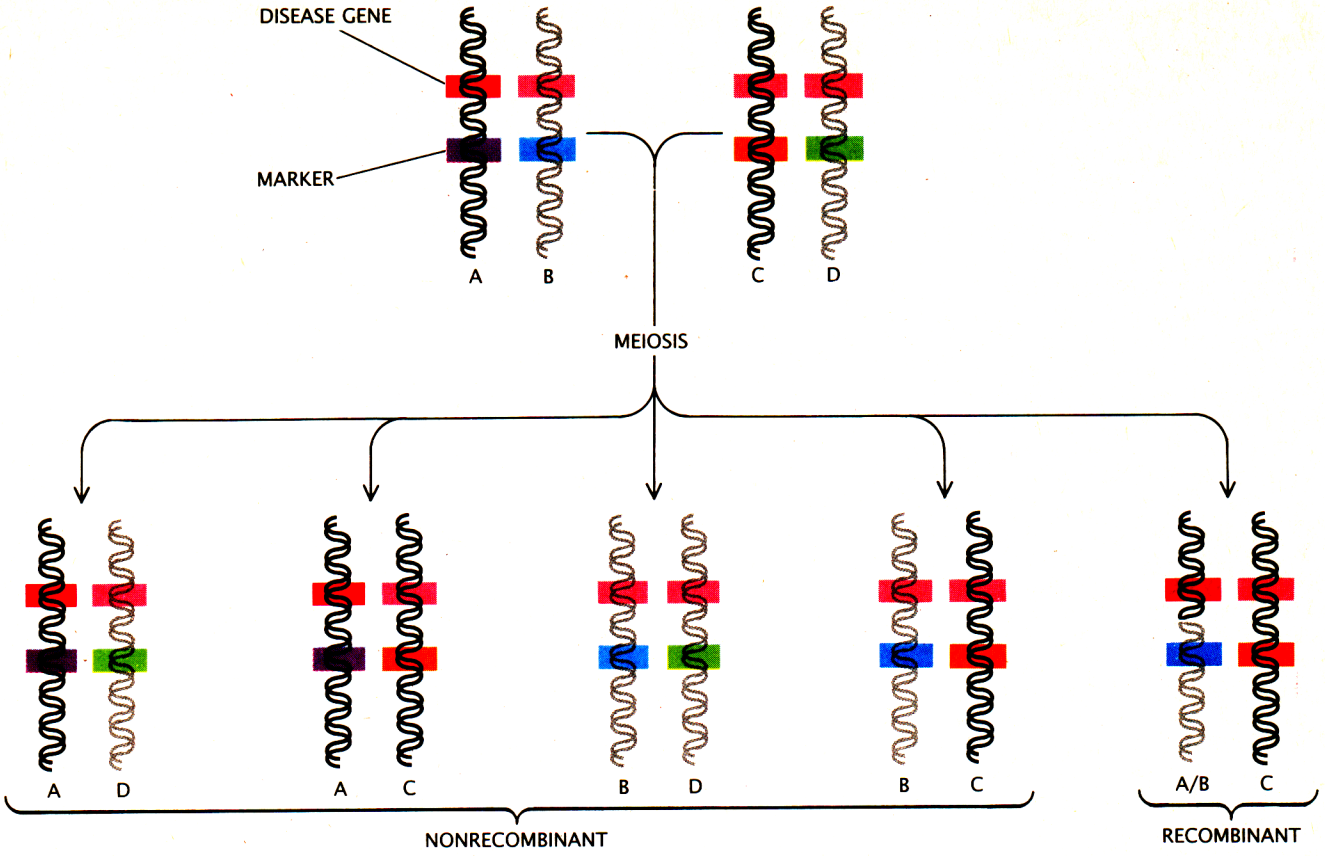
This DNA marker, defined by the RFLP, is found in one form or another in every individual, healthy or diseased. But if a genetic disease is passed down a pedigree together with a particular allele of the RFLP, the mutant gene can be assumed to lie in the same chromosomal region as the marker. In a second afflicted family the same marker will also show linkage, although the specific form of the marker that accompanies the disease may differ. Linkage to an arbitrary DNA marker reveals nothing about the physical position of the gene itself, of course, and for many purposes (such as diagnostic tests) physical location is immaterial. Nevertheless, the probe can also pick out the chromosome carrying the marker and the disease gene. If the probe is applied to a full set of human chromosomes, for example, it will hybridize to the chromosome bearing the marker site.

The value of any marker depends in large part on how many variants it displays throughout the population: the more versions of the marker there are, the more likely it is that an individual harboring a disease gene will carry two different alleles at the marker locus, making it feasible to detect recombination between the disease and the marker in offspring. Many RFLP's result from a change in a single base pair or the addition or deletion of a few base pairs at the restriction-enzyme cutting site. Such variation has a simple effect: the restriction site is either present or absent. The RFLP exists in only two forms, and so at least half of all individuals will probably be homozygous at the marker locus: they will carry the same variant on both homologous chromosomes. (Occasionally two restriction sites occur sufficiently close together to be detected by a single probe, yielding in effect a single marker with four alleles.)

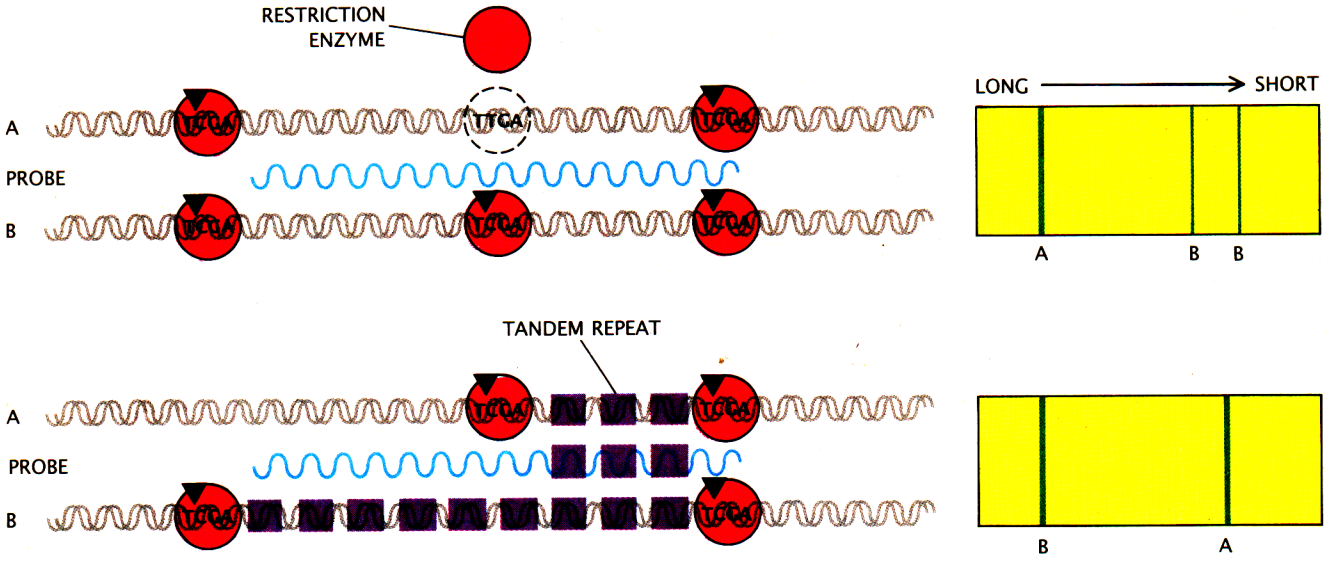
Another kind of DNA polymorphism creates many different versions of an RFLP. At many sites on the human DNA a single sequence that does not code for any protein is repeated many times. The origin and significance of these "tandem repeats" is a mystery, but for linkage mapping they offer a practical advantage in that the number of repeats at a given



RECOMBINATION makes it possible to detect genetic linkage. The diagram follows one idealized pair of homologous, or matching, chromosomes through meiosis, the process of cell division that produces germ cells (sperm or eggs). The chromosomes carry different alleles of two markers (*A, B*); one chromosome also bears a mutant, disease-causing allele of a gene (*m*) and the other chromosome bears the normal allele (+). In the precursor cell the disease is associated with allele *1* of both marker *A* and marker *B*. In the first phase of meiosis the chromosomes are replicated. The homologous chromosomes then "cross over," exchanging segments of equal length. Here crossing over takes place between loci *A* and *B*. The result is two germ cells (*a, d*) that carry the parental combinations of alleles and two (*b, c*) that contain recombinant chromosomes. In cell *b* the mutant gene is still found with allele *1* at locus *A* but is now joined by allele *2* at locus *B*. A low frequency of crossovers between the disease gene and marker *A* in many meioses would indicate that the disease and the genetic marker are closely linked.



LINKAGE between a disease gene and a marker is evident in the family history of the disease. Genetic features of a hypothetical couple and their children are shown. One parent suffers from a genetic disease caused by a single mutant allele (*red*); the other is healthy and hence carries only normal versions of the gene (*pink*). Children who inherit the disease usually also inherit a particular marker allele (*purple*) from the diseased parent, since the disease gene and the linked marker tend not to recombine.



DNA MARKERS—sites at which homologous chromosomes often differ in DNA sequence—are detected as RFLP's (restriction-fragment length polymorphisms). The DNA is digested with a restriction enzyme, which cuts wherever it finds a specific short sequence of nucleotides (in this case the base sequence *TCGA*). In one kind of marker (*top left*) a sequence difference causes a restriction site that is present on one homologous chromosome to be absent from the other. As a result the restriction fragments from each chromosome will differ in length. A DNA probe whose base sequence is complementary to that of DNA at the marker site reveals the fragments after they are sorted by electrophoresis (*top right*). Another kind of marker (*bottom left*) is characterized by a VNTR—a variation in the number of tandem repeats (short, repeated DNA sequences). The span between cutting sites differs between matching chromosomes, again resulting in distinctive fragments detected after electrophoresis (*bottom right*).

locus can vary from a few to hundreds of copies. Restriction fragments generated by cutting near these tandem repeats vary in length correspondingly [*see bottom illustration on opposite page*]; hence the RFLP comes in not just two forms but many. Given this variability in the population as a whole, the odds are good that a given individual will carry different versions of the RFLP on homologous chromosomes. A Southern blot will reveal two distinct fragments of different lengths, one from each homologous chromosome.

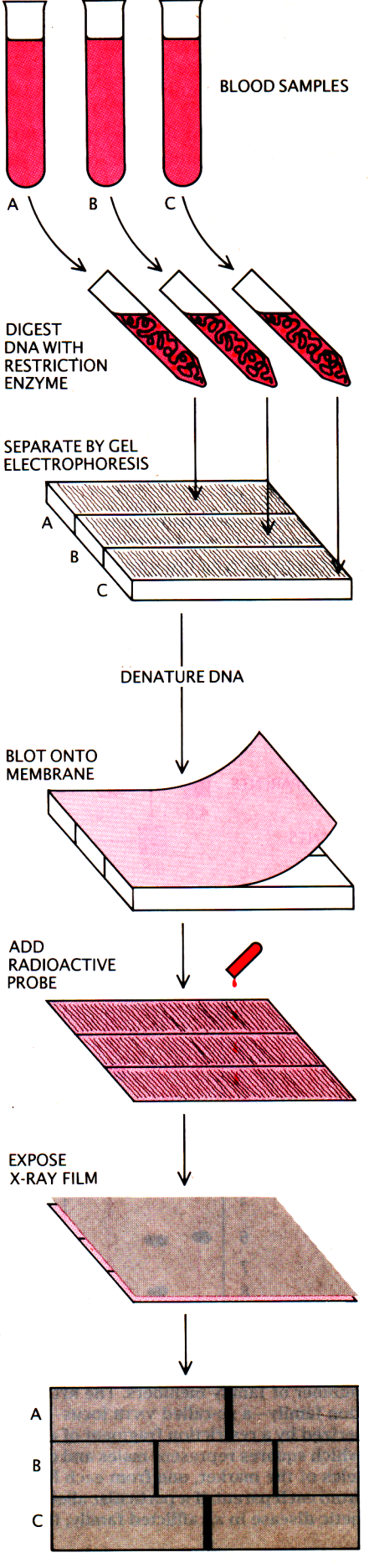
Probes for markers based on variations in the number of tandem repeats (VNTR's) can be developed more systematically than probes for ordinary markers. Alec J. Jeffreys of the University of Leicester recently recognized that the repeated sequences at many VNTR loci in different parts of the genome show similarities. The evolutionary explanation is again obscure, but the partial sequence homology means that under certain conditions a probe complementary to one VNTR locus can serve to pick out probes specific for other loci from a library of cloned DNA. Of the nearly 600 DNA markers developed so far in our laboratory, about 300 are VNTR's.

Such markers can serve as elements in an overall linkage map of the genome, or they can be developed for the more immediate purpose of tracking down a specific disease gene. Finding a marker whose inheritance is correlated with the appearance of a disease can be a staggering task on unmapped chromosomes. Since one often begins by knowing nothing about the chromosomal location of the disease gene or of any marker whose inheritance pattern is traced in an afflicted family, one can unwittingly search for linkage to tens of markers lying in a chromosomal region that is actually remote from the disease gene while leaving another, linked region unexamined. Nevertheless, the linkage strategy has already scored some remarkable successes.

If one knows which chromosome to search, the number of markers that must be tested can be reduced from several hundred, on the average, to as few as half a dozen. A genetic disease that almost always affects males but is inherited through the mother, for example, can be assumed to result from a recessive gene on the sex-determining *X* chromosome. (A mother carrying the disease has a second *X* chromosome bearing a normal copy of the gene, which masks the recessive disease gene; a son who inherits the mutation has only one *X* chromosome and therefore develops symptoms.) To find the gene, one need only test markers known to be carried on the *X* chromosome.

Genes for *X*-linked diseases were among the earliest to be traced through RFLP analysis; the first was the gene that causes Duchenne muscular dystrophy and probably also Becker muscular dystrophy (mapped by Kay Davies of the University of Oxford and Robert Williamson of St. Mary's Hospital in London). Yet an increasing number of diseases stemming from defects on the autosomes (the 22 pairs of nonsex chromosomes) have also yielded to the linkage strategy.

Huntington's disease became the first autosomal disease to be linked with a DNA marker when James F. Gusella of the Harvard Medical School and his colleagues studied afflicted families living in this country and near Lake Maracaibo in Venezuela. The group was fortunate in having to trace only eight markers through the families before finding one that was linked to the disease. Since then our laboratory and others have discovered markers for the genes causing disorders including cystic fibrosis, peripheral neurofibromatosis, or von Recklinghausen's disease (a disorder characterized by "café au lait" spots on the skin and a tendency to develop tumors and other disorders of the bone and nervous system), and familial polyposis coli (whose victims develop many colon polyps and run a very high risk of colon cancer).



RFLP ANALYSIS begins with a blood sample. DNA is extracted from the nuclei of white blood cells and digested with a restriction enzyme. The resulting DNA fragments are separated by gel electrophoresis, which sorts them in order of size. The RFLP is then detected by Southern blotting. First the DNA in the gel is heated to denature it, or separate its two strands, and is blotted onto a nylon membrane. A probe—a radioactively labeled segment of single-strand DNA that is complementary to the RFLP locus—is applied to the membrane. The probe hybridizes with the fragments from the locus; a sheet of X-ray film placed over the membrane detects the radioactively tagged fragments and thereby reveals which versions of the RFLP are present. In RFLP analysis of families, DNA samples from several individuals are often analyzed at the same time.

Tantalizingly, evidence of linkage has even been seen for forms of Alzheimer's disease and manic depression that run in families.

A "hit" can open the way to identifying the gene itself, which in turn provides a starting point for investigating the molecular mechanisms of the disease. By cloning the gene and determining its base-pair sequence one can deduce the composition of the protein it codes for and perhaps identify a specific defect. The protein can be synthesized and antibodies to the protein can be generated in experimental animals. Properly labeled, the antibodies can reveal the distribution of the protein in tissues affected by the disease. Such knowledge might hold the key to a treatment.

In many cases, however, the initial localization is too imprecise for a direct approach to the gene by current DNA technologies. The Huntington's disease gene, for example, recombines with its first identified marker at a frequency of about 5 percent, which implies that the marker lies as many as five million base pairs away from the gene. For identifying and cloning a gene the suspect stretch of DNA must be reduced to about a million base pairs, which means finding markers that recombine with the gene at a frequency of only about 1 percent. Ideally the markers will also flank the gene, bracketing the stretch of DNA to be tested.

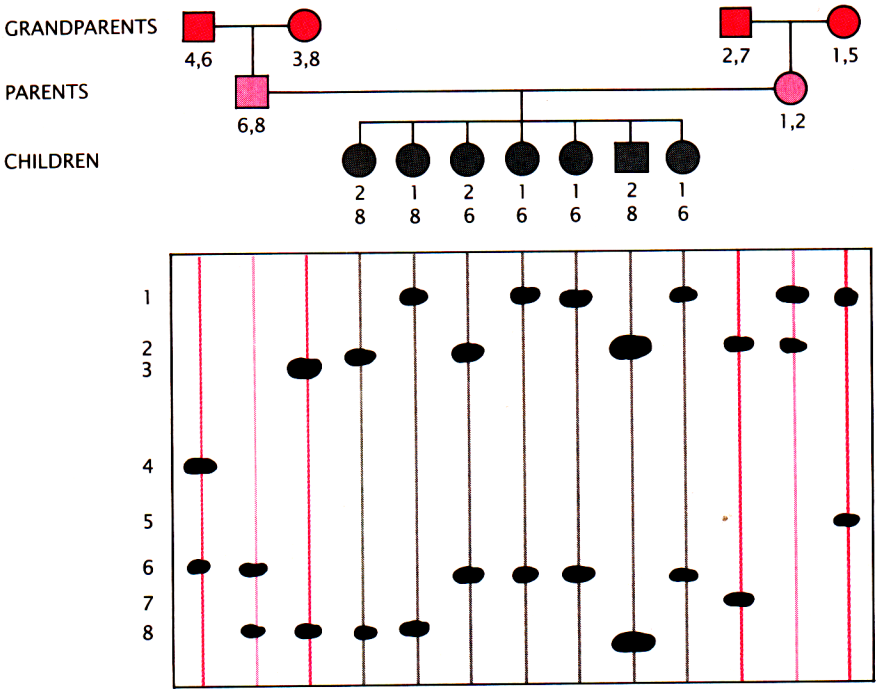
Tightly linked, flanking markers for cystic fibrosis, peripheral neurofibromatosis and familial polyposis are in hand, and a new, tightly linked marker has been identified for Huntington's disease. The search for the causative gene of each disease is in high gear. The approaches vary, but a common tactic is to comb a library of cloned chromosomal segments for one that is recognized by probes for both flanking markers. The segment—which presumably includes both markers and the gene they flank—can then be broken down further and each of the fragments cloned and tested for biological activity. Typically, a fragment can serve as a probe for messenger RNA (the sign that a gene is being expressed) in tissue affected by the disease. If it detects a messenger RNA that is unique to affected tissue, the probe itself may include all or part of the disease gene.

A different strategy has already culminated in the identification of the genetic defect in Duchenne muscular dystrophy. The region of the *X* chromosome that Davies and Williamson had linked with the disease shows missing segments in many patients; hence the disease may sometimes result from the absence of part or all of a normal gene. By identifying a region that is deleted in common among disease victims, Louis M. Kunkel of the Harvard Medical School and his colleagues were able to isolate and clone the gene.

Even before a disease gene is identified, linkage can sometimes point to possible causative mechanisms. The linked marker may fall near a gene of known function, which may then become a candidate for causing the disease. To take one instance, the marker for peripheral neurofibromatosis occurs on chromosome 17, which also carries the gene encoding the cellular receptor for nerve-growth factor (a substance that is vital for the survival and growth of nerve cells). That gene came under suspicion as a possible site of the mutation responsible for neurofibromatosis, but it was later found to lie some distance from the locus of the disease. Other genes on chromosome 17 may now become candidates for involvement in the disorder.

Reasonably tight linkage between a marker and a disease also makes it possible to devise tests for carriers and unborn victims—tests that are urgently needed given the frequency and insidious character of many genetic diseases. In populations of northern European descent, for example, one individual in 20 carries the cystic fibrosis gene. Because the gene is recessive, the carrier shows no symptoms, but if two carriers marry, their children stand a one-in-four chance of inheriting two defective genes and developing the disease. Huntington's disease is caused by a dominant gene (manifested even if the matching gene is normal), but its symptoms generally do not appear until middle age—after the unwitting victim has transmitted the disease to half of his or her children.

Before the presence of a disease gene can be established in an individual at risk, DNA from several other family members, both diseased and healthy, must be analyzed to determine which marker allele (or alleles,



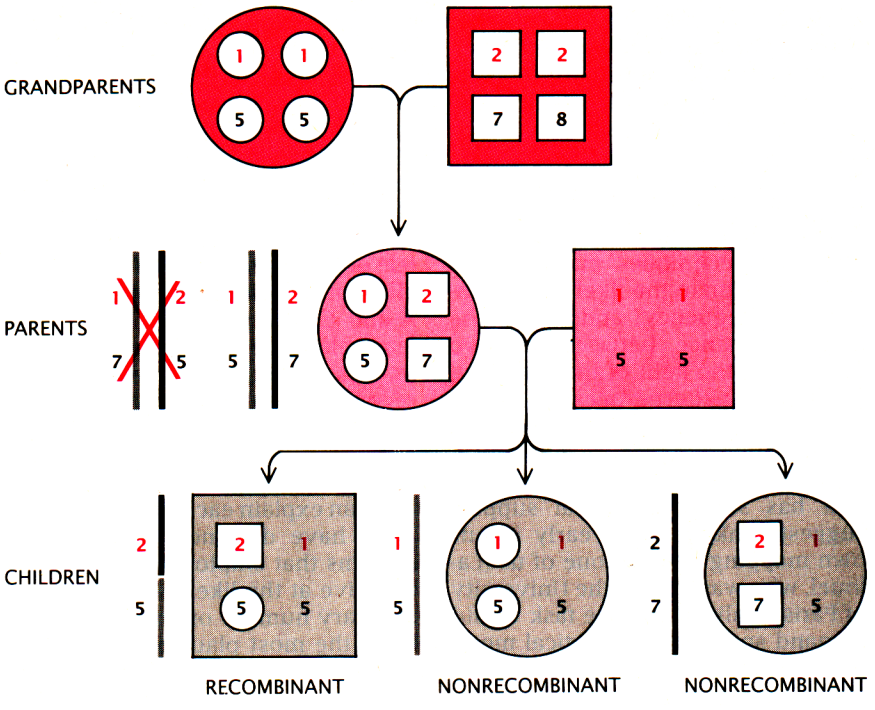
INHERITANCE OF AN RFLP can be traced by comparing restriction fragments from a number of family members. The RFLP marker that was analyzed in this three-generation family—a so-called VNTR locus—has many different alleles, each of them characterized by a restriction fragment of a specific size. Each individual in this pedigree (in which squares represent males and circles represent females) carries two different alleles of the marker, one from each homologous chromosome; children get one allele from each parent. If a particular allele of the RFLP is consistently associated with a genetic disease in an afflicted family, the marker and the defective gene may be linked.

in the case of a recessive disease that takes two copies of a gene to show itself) is inherited with the disease in this particular family. Finding a telltale allele in DNA from a prospective parent then indicates that he or she risks passing on the disease. Because DNA samples can be taken from a fetus soon after conception, the disease can also be diagnosed prenatally, enabling parents to make an informed decision about continuing the pregnancy. It is worth noting that in families at risk for some genetic diseases, fetal testing is actually increasing the number of births, simply because many couples would not conceive at all if they could not be sure the child was healthy before bringing it to term.

The construction of linkage maps, showing both arbitrary linkage markers and characterized genes arrayed along the chromosomes, has gone forward in parallel with the search for specific disease linkages. Linkage mapping represents a more deliberate and systematic approach to tracing mutant genes. From a complete linkage map workers trying to locate a disease gene will be able to choose and test a set of markers spaced at equal, large intervals along the chromosomes. Then, having discovered a linkage that restricts the gene's location to a specific chromosomal segment, they might test markers from a fine-scale map of the region in search of the tight linkage needed for further molecular studies.

The capacity to scan the genome for linkage not only will allow single-gene defects to be pinpointed more efficiently but also will hasten the search for the genetic bases of diseases caused by multiple aberrant genes. In addition, linkage maps will ultimately make it possible to check many points along the chromosomes simultaneously for a pattern of inheritance matching the family history of a disease, such as diabetes, coronary heart disease and certain cancers, to which susceptibility seems to be inherited. It might then be possible to close in on genes that confer predisposition to these illnesses.

Producing such a map extends the linkage strategy: now one is searching for linkage not between a DNA marker and a disease but between arbitrary DNA markers. Finding that alleles of different markers tend to be passed on together suggests the markers reside on the same chromosome, and the particular frequency



DATA FROM THREE GENERATIONS can solve genetic mapping's "phase" problem, posed by two markers on the same chromosome. Unless one knows the phase of two markers (*color and black*) in a parent—how their alleles (*numbers*) are distributed between the homologous chromosomes—one cannot unambiguously detect recombination in the children. Analyzing DNA from grandparents (the mother's parents in this case) can reveal which two alleles each grandparent contributed. Since the mother must have received alleles *1* and *5* on the chromosome she inherited from her mother, alleles *2* and *7* can be assigned to the matching chromosome from her father. The other configuration of alleles is thus ruled out, and a recombination event that has taken place in the mother's chromosome can be identified unambiguously in the first child.

with which the markers recombine reflects their "genetic distance."

Although linkage mapping is simple in concept, it presents an enormous bookkeeping and analytical challenge. A large-scale linkage map of the genome, sufficient to locate any disease gene within a span of between 10 and 20 million base pairs, would include between 100 and 200 evenly spaced markers. To have markers at even intervals, however, one must assemble a much larger set of random markers on the map. DNA must be collected from hundreds of individuals in dozens of large families and tested for the RFLP's characterizing each marker locus.

The analysis of these vast data sets is complicated by the fact that perhaps two-thirds of the markers in any individual are uninformative. They carry two identical alleles, with the result that linkage between the marker and any other locus cannot be detected in offspring. For two markers that may be linked, moreover, there is often no way to determine their "phase": how their alleles are distributed between the two homologous chromosomes. Without knowledge of which alleles are on the same chromosome in a parent, one cannot unambiguously detect recombination between the markers in the child.

These limitations are minimized when the data are gathered from extensive pedigrees. We have been fortunate in being able to draw on excellent family resources for our own mapping effort. For one thing, more than 50 Utah families with eight children or more have generously volunteered to give blood samples, from which we take DNA for analysis and establish permanent cell lines. The presence of many children means that the parents' chromosomes can be followed through a large number of meioses, giving more accurate estimates of recombination frequencies than could be had from families with few children. In addition almost all the Utah families we sampled have four living grandparents, whose DNA can often indicate the phase of markers in the parents. If one knows, for example, that allele 1

of marker *A* and allele 3 of marker *B* both originated in a grandfather, then his child—one parent—must carry both alleles on the same chromosome if the markers are linked.

Even so, the inevitable limitations in the data mean that the map must be founded on probabilities. Statistical techniques make it possible to estimate the likeliest recombination frequency, and hence the genetic distance, between any two markers in the light of the observed inheritance pattern. An estimated recombination frequency of 50 percent suggests two markers are unlinked; a smaller frequency—say 10 percent—that has strong statistical support suggests linkage. Very early in our own mapping venture one of us (Lalouel, who was then at the University of Paris) realized that the task would demand specialized statistical methodology and computer programs. He and his colleague Mark Lathrop designed algorithms and programs capable of both maintaining the huge data base and performing joint analysis of the inheritance patterns of many markers.

Having identified a set of linked markers, one still needs to determine their order along the chromosome. In principle one could calculate the probability of each possible order's giving rise to the observed inheritance pattern and choose the likeliest arrangement. As few as 15 linked marker loci, however, can be arranged in 6.5 times 10^{11} different orders, an impossibly large number. In practice one can quickly eliminate entire families of improbable orders by considering loci in subsets of, say, three at a time.

For a flavor of the reasoning, suppose that in a large family specific alleles of linked markers *A*, *B* and *C* are usually passed on as a group: a child inherits all or none of them. In one child, however, the original alleles of *A* and *C* are inherited with another allele of *B*; in a second child the original allele of *B* is joined by other alleles of *A* and *C*. The sequence *A-B-C* is the least plausible sequence because it implies that double recombination—recombination both between *A* and *B* and between *B* and *C*—took place in both cases. (Under either alternative order, *A-C-B* or *B-A-C*, one recombination can explain each observation.)

We have designed computerized systems that employ such strategies to arrive at the likeliest order for an arbitrary number of linked markers. Once the most plausible order for a cluster of linked markers has been established, they can be assigned to a specific chromosome, for example by hybridizing one of the marker probes to a set of intact chromosomes. Linkage clusters are thereby assembled into a chromosome map.

The genetic distances on a chromosome's linkage map are related to physical distances (numbers of base pairs), but the relation is by no means direct. We have found, for instance, that the recombination frequency of a given pair of markers often differs significantly between the sexes. That is, the probability that two markers on a chromosome inherited from the mother will have recombined during her meiosis may be quite different from the probability of recombination between the markers on the same chromosome inherited from the father. On chromosome 13, for example, recombination frequencies are several times higher in females. On chromosome 11 the opposite is true in one interval, and in an adjacent interval the two sexes show similar recombination frequencies. The molecular basis for these intriguing variations is mysterious, but as a practical matter we have been preparing two maps of each chromosome, one map for each sex, showing identical marker orders but different genetic distances.

We have completed preliminary maps for most of the human chromosomes. Another group has recently published a similar collection of preliminary maps, based on a smaller number of reference families. Yet linkage mapping is an inherently collaborative enterprise: every group is looking for landmarks on the same terrain. Markers developed and studied in one laboratory often complement markers from another laboratory, in some cases bridging gaps between linked clusters.

A framework for cooperation has been set up by Jean Dausset at the Center for the Study of Human Polymorphism (CEPH) in Paris. The CEPH has undertaken to collect, store and distribute DNA from 40 families. The collection draws mostly on our Utah families but also includes DNA from families studied by other workers. Investigators from around the world (including our own group) get complete sets of DNA from the collection; in return workers report their markers and inheritance patterns to the CEPH, which makes the data available to all investigators and so lays the groundwork for a single genetic map.

The completion, probably within the next few years, of a high-resolution map will consummate the transformation of the human genome from uncharted territory to well-surveyed ground. Such a map can be expected to yield precise locations for most of the remaining well-characterized genetic diseases. A complete linkage map will also prove invaluable for guiding another large-scale investigation of the genome, which is still in the planning stage: an effort to determine the complete base-pair sequence of human DNA. Small islands of DNA along the chromosomes will most likely be sequenced first. The linkage markers within each island will serve to locate it in the larger landscape of the genome.

DISEASE	CHROMOSOME	DATE
HUNTINGTON'S DISEASE	4	1983
DUCHENNE MUSCULAR DYSTROPHY	*X* (GENE CLONED)	1983
POLYCYSTIC KIDNEY DISEASE	16	1985
CYSTIC FIBROSIS	7	1985
CHRONIC GRANULOMATOUS DISEASE	*X* (GENE CLONED)	1985
PERIPHERAL NEUROFIBROMATOSIS	17	1987
CENTRAL NEUROFIBROMATOSIS	22	1987
FAMILIAL POLYPOSIS COLI	5	1987
MULTIPLE ENDOCRINE NEOPLASIA IIa	10	1987

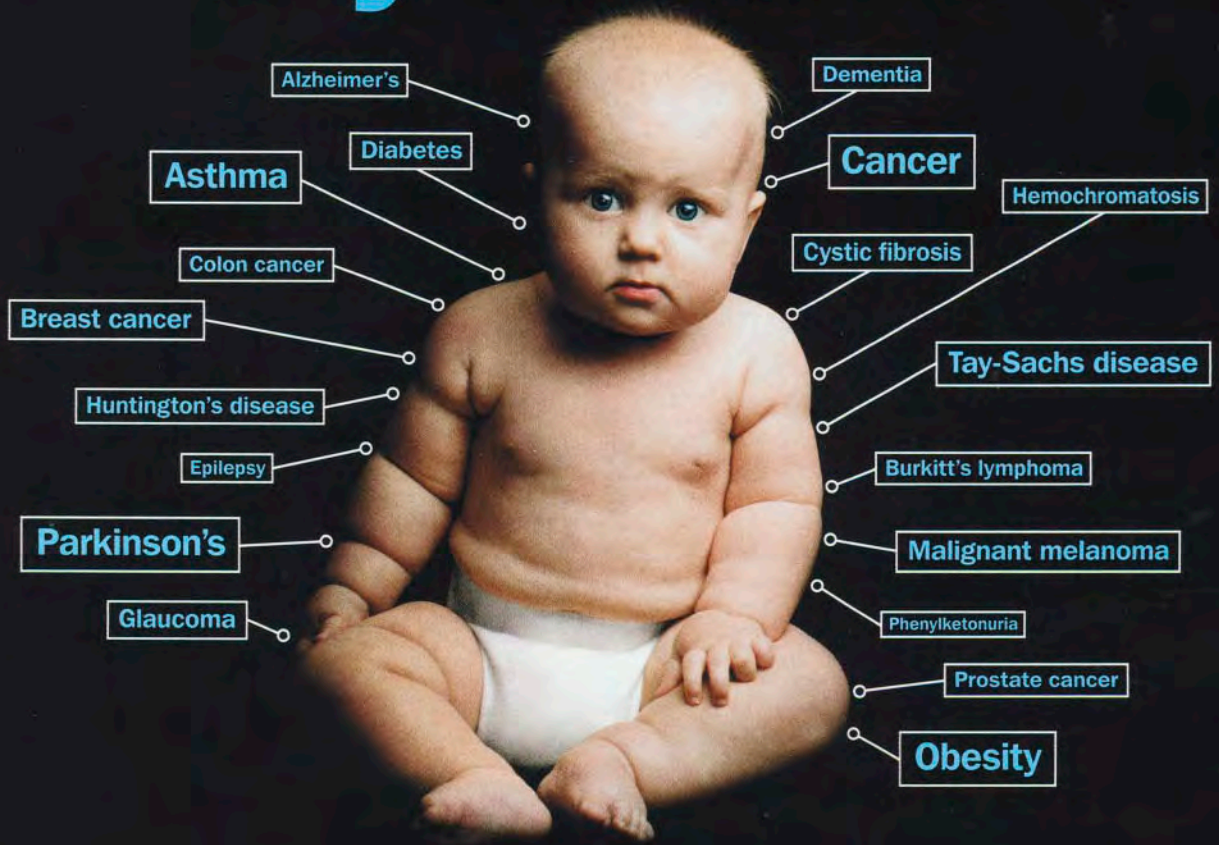
TABLE OF DISORDERS gives a small sample of the genetic diseases for which the chromosomal location of the defective gene has been determined with the help of linkage studies. The table also indicates which chromosome carries the gene and the linked marker and gives the year linkage was first reported. Reasonably tight linkage can make the marker useful for diagnosing the disease in members of an afflicted family; very tight linkage can open the way to identification and cloning of the disease gene.

DECEMBER 24, 2012

Egypt Divided / Pot's Big Moment / Best of 2012 Movies, Music, Books & More

TIME

Want to Know My Future?



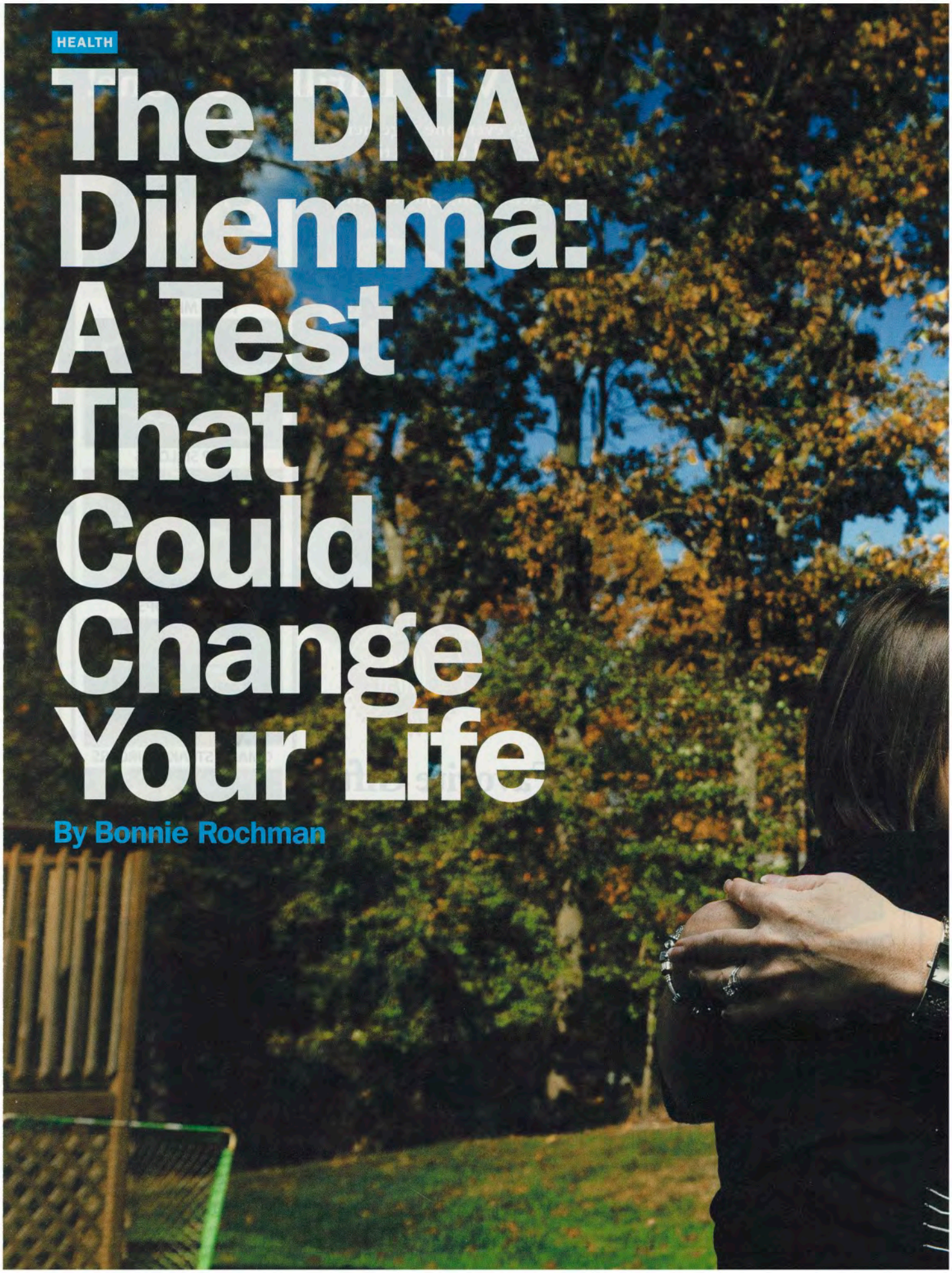
New genetic tests can point to risks—but not always a cure

BY BONNIE ROCHMAN

HEALTH

The DNA Dilemma: A Test That Could Change Your Life

By Bonnie Rochman



Know your enemy, we tell ourselves;

knowledge is power. Laurie Hunter wanted to know what disease was attacking her daughter Amanda, who by the age of 2 months was not developing normally. Her muscle tone was low. She wasn't lifting her head. She was slow to talk, and she didn't walk until she was 2.

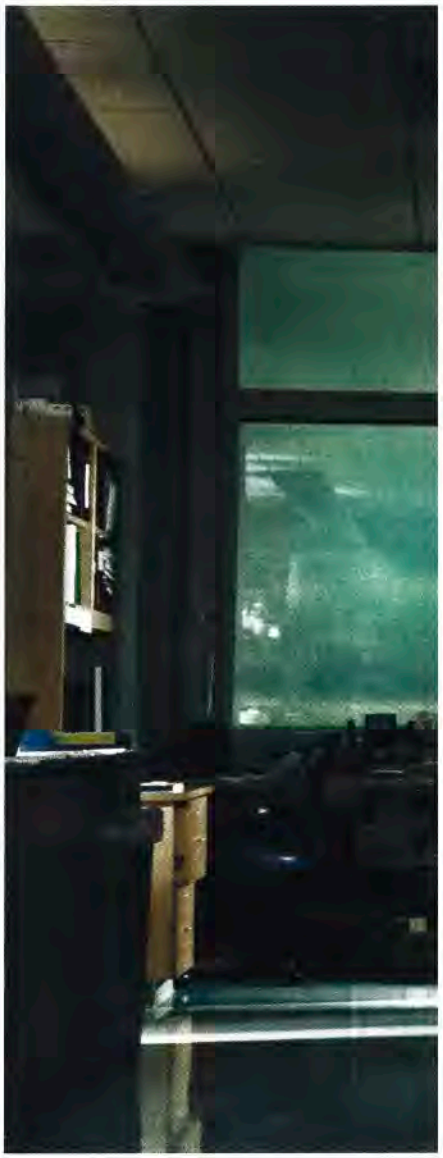
"As a mother, you know that everything that happens to your child is not your fault, yet you still feel responsible," says Hunter, 42, a high school English teacher who lives in Jackson, N.J. "We turned to genetic testing because I wanted answers." The first tests, done at the Children's Hospital of Philadelphia (CHOP) when Amanda was 4, came back normal. So did another round when she was 9. Doctors could not figure out what was making Amanda weak—even as she got weaker and slower and stopped being able even to blow her nose. "It's like her muscles are getting tighter and not moving in the way they should," Hunter said. But the doctors held out hope. Genetic testing grows more sophisticated every day, they said, allowing researchers to explore a child's health down to every last typo on a chromosome.

In March, a third round of tests found seven genes missing from Amanda's first chromosome. At last, Hunter thought, when the genetic counselor called and asked to see her. "It felt like finally I might have an answer." But it was not the answer she was looking for. The small deletion, the counselor said, did not explain Amanda's condition. That was still a mystery. And now a whole new threat appeared.

One of the seven deletions has been linked to very rare tumors. The geneticists wanted Amanda, who is 14, to be screened by an oncologist. "It was like, Oh, my God, now we are adding cancer to the mix," Hunter says. "Never in a million years did I think this would be an issue."

She was even more surprised when a counselor called after her own tests came back. "I know you're going to be upset," the counselor said, "but we found that you have the same deletion." And so might her other two children.

This is the world we are heading into: one with powerful new weapons against age-old diseases and a host of questions about how to use them wisely and not turn them on ourselves. Imperfect knowledge can make us crazy—or bankrupt—chasing down threats that may never materialize. The human genome is an exquisitely complex blueprint. Geneticists hunting for answers to mysterious symptoms invariably trip over incidental findings, genetic twists they were not even looking for that might signal a risk of cancer or Alzheimer's or Parkinson's in the near or distant future. But do doctors have to tell patients everything they learn, even about the risk of diseases for which there are not yet cures? Do parents have to tell their children what might await them as adults? And who will pay for all this? "Everyone at this point is flying by the seat of their pants," says Dr. James Evans, a medical geneticist at the University of North Carolina School of Medicine. "The technology is outpacing us."

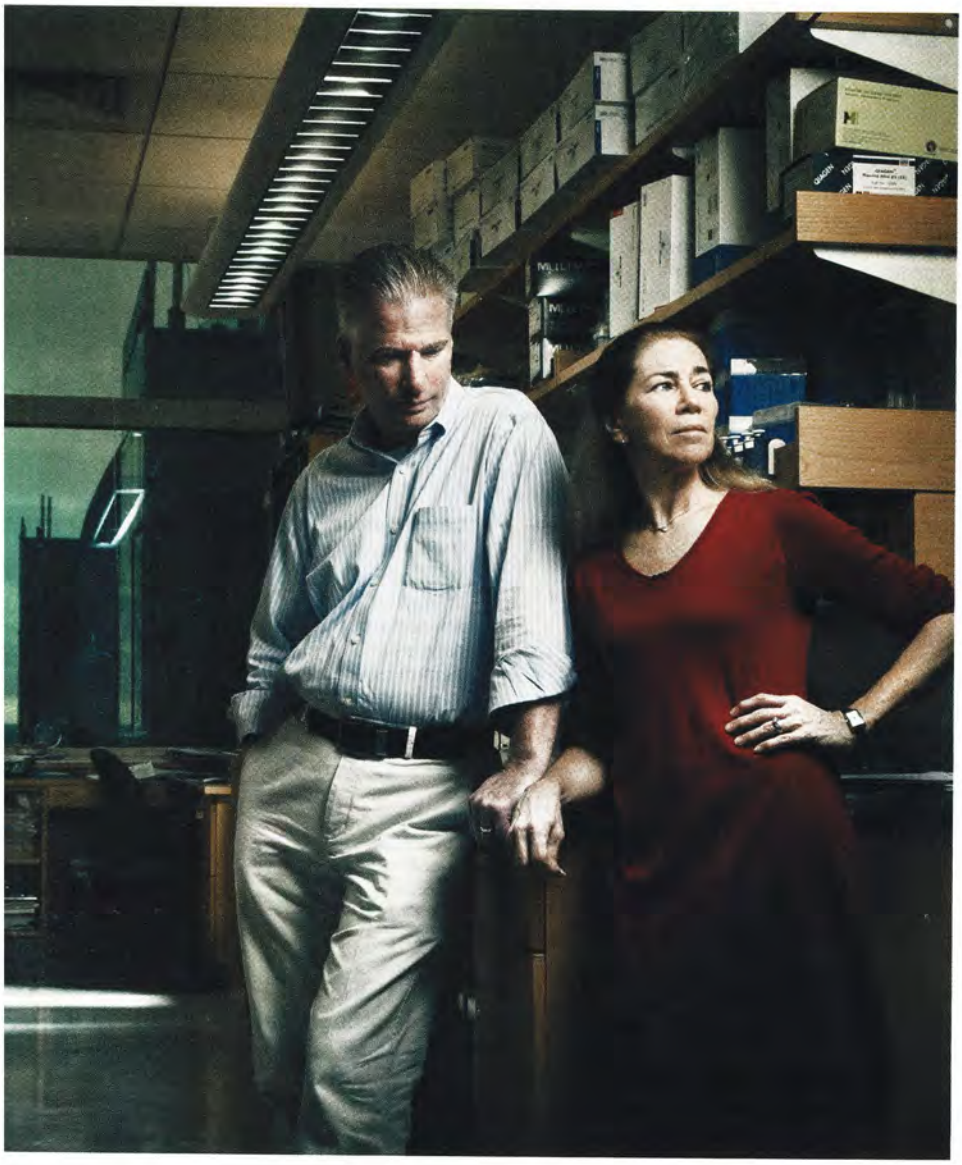


MILLER MOBLEY FOR TIME—GETTY FOR TIME

From Labs to Living Rooms

THE MAPPING OF THE HUMAN GENOME, completed in 2003, cost $2.7 billion. Now the cost for an individual's whole-genome sequencing (WGS) is $7,500 and falling fast. One day WGS could be as easy to get as a pregnancy test at the drugstore. To do the testing, lab technicians need less than a teaspoon of blood, which is chemically treated to burst open the cells so the DNA inside them can be collected. Those microscopic strands are then fed into sophisticated machines that read each of the 3 billion bits of information, called base pairs, that make up a person's genetic alphabet. Computers scan the data for the equivalent of spelling mistakes. Some mistakes cause disease; others don't. And in between is a vast gray area where scientists just don't know what the changes mean.

In an ideal world, genetic analysis could save money by catching diseases early, offering targeted treatments and identifying the most effective preventive measures. Dr. Katrina Armstrong, a professor at the University of Pennsylvania School of Medicine, notes that testing 21 genes could reveal which breast-cancer patients are unlikely to benefit from a particular chemotherapy—knowledge that could



Tough call *Dr. Ian Krantz and Nancy Spinner at the Children's Hospital of Philadelphia decided not to tell parents their baby will likely develop early-onset dementia*

spare women the treatment and save $400 million each year. "If genomics can help us understand who will get the most benefit and who will get little or no benefit from an intervention," Armstrong says, "it will take us a long way toward improving patient outcomes and saving money."

But a majority of doctors in a recent survey predicted that more testing will trigger higher costs, as patients with ambiguous results begin to seek frequent screenings—and potentially unnecessary procedures—for diseases they might never develop. "If we open the door to a test that has no clear, well-defined purpose, that is a recipe for unnecessary medical care," says Dr. Wylie Burke, a geneticist who chairs the department of bioethics and humanities at the University of Washington. "Instead, we could say, Here are the 1,000 mutations we should check in everyone." The American College of Medical Genetics and Genomics is already working on that, painstakingly assembling a list of a few dozen conditions that it says should be routinely looked for during genome sequencing. The hope is that focusing on certain hot spots—contenders include several syndromes that increase the risk of various cancers—will lead to improved analysis and, with it, better patient outcomes.

Nearly all the parents said they would want to know about every disease risk, even if there's no treatment available

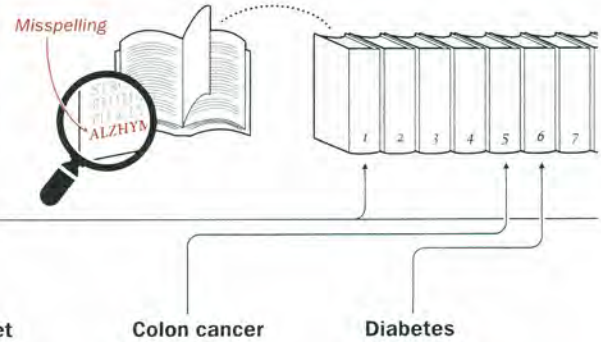
Some genetic testing has already moved out of the lab and into the living room. Companies like 23andMe offer DNA analysis directly to consumers—no doctor required. Since 23andMe's founding in 2006, more than 180,000 people have been tested as the price has fallen from $999 for information on 14 specific traits and health risks to $99 for more than 200. The promise boils down to "forewarned is forearmed." If parents learn that their child carries a gene called ApoE4, indicating a higher risk of Alzheimer's, they might discourage the child from playing youth hockey or football, since research has linked traumatic brain injuries with a greater likelihood of brain disease in people who test positive for ApoE4.

"I do believe at some point in time everyone will be genotyped at birth," says 23andMe co-founder and CEO Anne Wojcicki. Her husband, Google co-founder Sergey Brin, has a genetic mutation that increases the risk of Parkinson's disease up to 80%; she has already tested their two children. Wojcicki's grandmother had macular degeneration; when testing revealed that some of Wojcicki's nieces and nephews are at increased risk for it, she bought them high-quality sunglasses. If her kids were predisposed to developing diabetes, she says, she'd encourage healthier eating. "I want to do everything I can to potentially enable my children to be disease-free."

But having more-detailed genetic information does not always point to a clear path. Dr. Ian Krantz and Nancy Spinner, a husband-and-wife team at CHOP, are working with an $8.8 million federal grant to understand what genomic

Decoding Disease

INDIVIDUAL GENES
Think of human DNA as an encyclopedia. Testing a **specific gene** involves pulling out the right volume (chromosome) and looking for spelling errors on a particular page



CONNECTING THE DOTS
Some diseases are caused by a single mutation, while others involve a complex interplay among many genes and environmental factors

Early-onset Alzheimer's disease
Chromosomes 1, 14 and 21
Someone who inherits one of several mutations on chromosomes 1, 14 or 21 is almost certain to develop a rare form of Alzheimer's (accounting for less than 5% of cases) between the ages of 30 and 60

Colon cancer
Chromosome 5
Most cases of familial adenomatous polyposis, a rare form of colon cancer in which polyps have been detected in kids as young as 7, are caused by mutations in a tumor-suppressor gene on chromosome 5

Diabetes
Chromosome 6
Mutations on chromosome 6 play a role in Type 1 diabetes, but so do other factors, including early diet. If an identical twin has the condition, formerly called juvenile diabetes, the other twin has at most a 50% chance of developing it

information patients and parents want to know. Most parents go in looking for the cause of a mystery illness. "If you tell parents their child also has an increased risk for colon cancer or breast cancer," says Krantz, a pediatrician who oversees medical-genetics training at CHOP, "that's a whole different level of stress."

If you want to start an argument, ask doctors and patients what they think doctors should do when they discover genetic results they weren't looking for. It can be an emotional blow—and a lifelong burden—if a mom learns that her baby girl carries a mutation that increases her risk of ovarian cancer or a dad finds out that his aspiring linebacker is genetically predisposed to developing Alzheimer's. In focus groups that are part of Krantz and Spinner's study, nearly all the parents said they would want to know about every disease risk, even if there's no treatment available. But in groups of bioethicists, lab directors, geneticists, pediatricians and genetic counselors, the majority said only results that could be immediately acted on should be shared with families.

This year, the lab Spinner runs tested a baby with a mysterious illness and found a completely unrelated mutation that indicated that dementia would likely set in at around age 40. Endless discussions followed: Should they tell the baby's parents that their child would probably develop a progressive neurologic disease marked by incontinence, blurred vision and confusion? There is no current treatment or cure. Telling them would all but guarantee that their child would never be able to get disability or long-term-care insurance. "We came around to the realization that we could not divulge that information," says Spinner, who is a genetics professor at Penn's medical school. "One of the basic principles of medicine is to do no harm."

At about the same time, her lab discovered that a 2-year-old with kidney disease carried a genetic risk for a kind of colon cancer. In some cases, polyps have been known to develop as early as age 7. With this patient, withholding the information would have seemed unethical. "We feel good about that one," says Spinner. "Proper screening can make a huge difference."

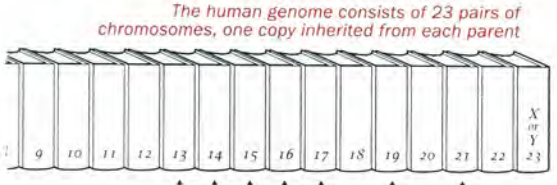
Genome sequencing isn't the first medical development that has forced doctors to grapple with the question of how much to tell patients. There have been cases of physicians' choosing to keep quiet when a test revealed a child's father was not his or her biological father. In years past, doctors have agreed not to share news of a terminal illness with an elderly patient if the consensus was that the knowledge would cause too much anxiety.

But genomes are vastly more complicated. "If you fall off your bike and get an X-ray looking for a fractured rib, the radiologist scans the entire X-ray and automatically reports back to your doctor if something else is going on," says Dr. Robert Green, a geneticist at Harvard Medical School. "More than a few cancers have been picked up this way. The problem with genomics is that everyone could have incidental findings."

Perhaps nowhere is the risk of overreacting to murky results greater than in the field of prenatal testing. This year two groups of researchers announced that they had each sequenced a fetus' DNA from cells gathered from the mother's blood, leading to concerns that in the not-too-distant future, women might abort a pregnancy if they learn their unborn baby has an increased risk for cancer. "Great, we can sequence the genome of a fetus. What the hell does it tell us?" says bioethicist Tom Murray, a visiting scholar at Yale. "Much less than most people probably believe. Probabilities are not the same as guarantees."

Faced with a growing need for protocols, the medical community is trying to hammer out some guidelines. This spring, the American College of Obstetricians and Gynecologists stated that though personalized gene profiles may be promising, they are "not ready for prime time" and should be discouraged. The American Academy of Pediatrics advises against genetic testing for children unless there is clear evidence of beneficial treatment or effective prevention strategies.

The challenge doctors face in determining how much to tell patients—or their parents—is complicated by a steady stream of new discoveries. Test results that are indecipherable today could be lifesaving in 2025. But waiting years to share sequencing



GENOME SEQUENCING
Scanning a person's **entire genetic code** can help diagnose a mysterious illness, but murky results can lead to a lot of anxiety

Breast cancer
Chromosomes 13 and 17
A woman who inherits mutations in either of two genes (BRCA 1 on chromosome 17 or BRCA 2 on chromosome 13) is about five times as likely to develop breast cancer as a woman who does not have such a mutation

Autism
Chromosomes 15 and 16
About 20% of autism cases can be traced to genetic abnormalities, including deletions or duplications on chromosomes 15 and 16. A new experimental blood test looking at 55 genes might help diagnose the condition earlier

Obesity
Chromosome 16
It's a mistake to attribute the obesity epidemic to DNA alone, but dozens of genes, including the fat mass and obesity-associated (FTO) gene on chromosome 16, appear to play a role in weight variation in adults

Alzheimer's disease
Chromosome 19
A common variant of the ApoE gene on chromosome 19 increases a person's risk of getting late-onset Alzheimer's, which develops after age 60. Mutations in several other genes have also been linked to the disease

information is a logistical nightmare, particularly considering that patients may not remain under that geneticist's care and may change addresses many times over. Genomic transcripts are also so massive—labs typically FedEx a hard drive because there's too much data to transmit digitally—that the information is often relegated to a hospital's archives, if it's saved at all.

One possible solution to the problem of what to do with the deluge of data is a new Web-based venture called My46. Named for the number of chromosomes in human DNA, the nonprofit will allow people to store their sequencing results online and choose what they want to know and when. For example, parents of a baby who gets sequenced could opt to learn right away any findings about childhood diseases and put everything else—from unclear results to increased risks of adult-onset diseases—in the digital equivalent of a locked drawer, where it can be stored forever and accessed whenever they want to open it.

"Right now, it's not unusual for researchers to say that they're not returning results because there's no good way to do it," says Dr. Michael Bamshad, chief of pediatric genetics at the University of Washington, who works with Burke and is helping develop My46. Eventually, he predicts, "everyone will have their genome stored in a cloud."

Living with the Results

FOR LAURIE HUNTER, THE NEWS OF HER own cancer risk was not actually a shock. The disease runs in her family. Her mother and aunt had breast cancer, and her brother died of testicular cancer when he was 27. "I'd resigned myself that it was part of my reality, but I didn't think about it being part of my kids' reality—not this young, anyway," she says. One of the genes she's missing increases her risk of extra-adrenal tumors, which can pop up in the head, neck, chest and abdomen. The average age of onset is 30. Hunter is 42. So she scheduled blood tests and a full-body MRI to see if any tumors had started growing. She was thinking not just of herself and Amanda but also of her son Ryan, 4, who has always been healthy, and of her youngest child Kailyn, who was born with a rare genetic disorder unrelated to Amanda's, called Wolf-Hirschhorn syndrome. At 2½, she cannot talk and can barely sit up. "I have two girls, one of whom will never speak, and they need to be cared for by somebody," she says. "I worry about, if something happens to me, who will take care of them." And then there is Ryan. What if she had passed the cancer risk on to him?

"I have shed more than a few tears since I learned about this gene deletion," Hunter says. "I love all my children equally, but I have reconciled myself that neither daughter will ever drive, go to college, get married or live on her own. The hardest part is thinking about my son. I have this one child in whom all my hopes and dreams lie, and now he may have this deletion too."

She considered not testing him. Maybe ignorance would be better than knowing the worst. "But I thought, God forbid, what if he was one of the ones who develops tumors at 10 years old and I didn't know. I'd be consumed with guilt."

Ryan was tested in the last week of September. The waiting was a kind of torment. "We got the results back the other day," Hunter says. "He does not have the deletion. I feel like I can breathe again."

But because of Amanda's increased risk, she is being closely monitored. An MRI found a spot on her neck that turned out to be an enlarged lymph node. The doctors still don't know what is causing her other health problems.

"If all three of my children were healthy and had no issues, I don't know if I'd want to know about those seven missing genes," says Hunter, whose own MRI detected a lesion above her diaphragm. She's waiting to learn whether it's a tumor. "Sometimes what you don't know is easier. I feel completely overwhelmed with information. Now it just feels like a waiting game."

This is often how medicine works. Our powers outpace our principles and protocols, so that we wake up one day to headlines that a sheep has been successfully cloned and have to figure out what that means for the future of reproduction. In the case of genetic testing, there is little doubt that greater knowledge will bring many blessings, but it comes with costs, literal and emotional, and patients entering this territory with imperfect maps need to reckon with the odds of getting lost. ■

latimes.com/news/science/la-sci-fetal-genome-sequence-20120607,0,7625263.story

latimes.com

Entire DNA of fetus revealed through risk-free testing

Researchers use blood from the mother and saliva from the father to determine a fetus' entire DNA sequence. If refined, the technique could provide a risk-free way to screen for genetic disorders.

By Rosie Mestel, Los Angeles Times

5:44 PM PDT, June 6, 2012

Scientists have pieced together the entire DNA sequence of an 18-week-old fetus without having to use any invasive tests that could result in a miscarriage — an advance that offers a glimpse of the future of prenatal testing.

Using blood drawn from the mother and a sample of saliva from the father, the researchers were able to scan the fetus' genome and determine whether it contained any of the myriad single-letter changes in the DNA code that can cause a genetic disorder. They could even pinpoint which mutations were inherited from Mom, which came from Dad, and which were brand-new.

If the technique is refined and the technology becomes inexpensive — as many experts anticipate — this type of prenatal testing could provide prospective parents with a simple, risk-free way to screen for a broad array of simple genetic disorders, according to the authors of a report in Thursday's edition of Science Translational Medicine.

The work is based on the fact that small fragments of fetal DNA circulate in the blood of pregnant women.

Several biotech companies are developing tests that capture those DNA fragments and screen them for signs of Down syndrome and other disorders that result from having an extra copy of an entire chromosome.

But that type of screening is far easier than searching for single-letter variations in individual genes, said senior author Jay Shendure, a geneticist at the University of Washington in Seattle.

An additional chromosome is "the equivalent of an extra chapter in a book," he said. "What we're trying to do is pick up a typo in a word."

To set about their task, Shendure's team started by sequencing the genome of an anonymous pregnant woman, using a complete sample of her DNA obtained from her blood cells. They also sequenced free-floating DNA fragments extracted from her blood plasma, repeating their work until they had decoded every part of the human genome 80 times.

That plasma contained a mix of 10% fetal DNA and 90% maternal DNA, all in tiny fragments. The scientists needed to be able to tell which pieces were from the mother and which belonged to the fetus.

To solve that problem, the scientists relied on the fact that genetic material is inherited in long strands of DNA, called chromosomes — and that tiny genetic variations on the same chromosome are usually inherited together, in blocks known as haplotypes. If a given haplotype was present in the fetus as well as in the mother, it would be detected in the plasma in extra amounts.

The scientists also sequenced the father's DNA, which was extracted from saliva. This allowed the team to figure out whether genetic variations in the fetus that didn't match the mother were inherited from the father or were new mutations. On average, about 50 new mutations show up in a fetus.

The scientists checked their results against a blood sample taken from the baby's umbilical cord after birth. Their calculations were more than 98% correct, they found, and they had detected 39 out of the 44 new mutations. None of those mutations had known medical consequences, the researchers said.

This approach could be used to devise a single test to screen for the 3,000 known disorders that are caused by mistakes in single genes. Individually, they are rare, but together they affect about 1% of births.

Technology like this could lead to more widespread screening of fetuses for genetic disorders that could benefit from early treatment, said Dr. Joe Leigh Simpson, senior vice president for research and global programs for the March of Dimes in White Plains, N.Y. It might even help doctors identify women at heightened risk for problems such as pre-term birth, he said.

rosie.mestel@latimes.com

The New York Times

June 6, 2012

DNA Blueprint for Fetus Built Using Tests of Parents

By ANDREW POLLACK

For the first time, researchers have determined virtually the entire genome of a fetus using only a blood sample from the pregnant woman and a saliva specimen from the father.

The accomplishment heralds an era in which parents might find it easier to know the complete DNA blueprint of a child months before it is born.

That would allow thousands of genetic diseases to be detected prenatally. But the ability to know so much about an unborn child is likely to raise serious ethical considerations as well. It could increase abortions for reasons that have little to do with medical issues and more to do with parental preferences for traits in children.

“It’s an extraordinary piece of technology, really quite remarkable,” said Peter Benn, professor of genetics and developmental biology at the University of Connecticut, who was not involved in the work. “What I see in this paper is a glance into the future.”

The paper, published Wednesday in the journal Science Translational Medicine, was written by genome scientists at the University of Washington. They took advantage of new high-speed DNA sequencing and some statistical and computational acrobatics to deduce the DNA sequence of the fetus with about 98 percent accuracy.

The process is not practical, affordable or accurate enough for use now, experts said. The University of Washington researchers estimated that it would cost $20,000 to $50,000 to do one fetal genome today.

But the cost of DNA sequencing is falling at a blistering pace, and accuracy is improving as well. The researchers estimated that the procedure could be widely available in three to five years. Others said it would take somewhat longer.

It is already possible to determine the DNA sequence of a fetus by acquiring fetal cells through

amniocentesis or chorionic villus sampling, which involves testing the placental tissue. But these procedures are invasive and carry a slight risk of inducing a miscarriage.

For couples worried about passing on a genetic disease, it is also possible to use in vitro fertilization and have an embryo genetically tested before implantation into the womb.

But the technique described in the paper would not require complete cells from the fetus and would make such DNA testing easier and less risky.

“If this sort of thing is ever to be used on a widespread basis, I think it necessarily has to be noninvasive,” said Jay Shendure, associate professor of genome sciences at the University of Washington, who supervised the research team.

The genome was determined from blood samples taken 18.5 weeks into the pregnancy, although the researchers said the technique could probably be applied in the first trimester, as early as or even earlier than some invasive techniques.

The technique takes advantage of the discovery in the 1990s that fragments of DNA from the fetus can be found in a pregnant woman’s blood plasma, probably the result of fetal cells dying and breaking apart.

These fragments can be genetically analyzed, providing that the fetal DNA fragments can be distinguished from the far more numerous fragments that come from the mother herself.

The analysis of fetal DNA fragments found in a pregnant woman’s blood is already used in new commercially available tests of the fetus’s gender, its paternity and whether it has Down syndrome. But reconstructing an entire genome from DNA fragments is much more difficult.

Such information would allow detection of so-called Mendelian disorders, like cystic fibrosis, Tay-Sachs disease and Marfan syndrome, which are caused by mutations in a single gene.

More than 3,000 such diseases collectively occur in about 1 percent of births. The mutations can be inherited from the parents or they can arise spontaneously in the fetus.

Researchers led by Dennis Lo at the Chinese University of Hong Kong first showed in 2010 that reconstructing a fetal genome would be possible. Other work toward this goal has been done by Stephen Quake and colleagues at Stanford University.

But Dr. Lo’s team used a maternal sample obtained invasively. And it could determine only the

inherited mutations, not the spontaneous ones.

The University of Washington researchers, using an approach partly developed by a graduate student, Jacob O. Kitzman, did not need an invasive test. And they were able to detect 39 of 44 such spontaneous mutations, though with a huge number of false positives.

“This will be a step toward having a better and better prenatal diagnosis that detects more and more at a reliable cost,” said Dr. Arthur L. Beaudet, chairman of molecular and human genetics at Baylor College of Medicine in Houston.

Dr. Beaudet, who was not involved in the work, said that spontaneous mutations account for about 10 percent of cases of mental retardation and other learning disabilities.

The ability to sequence an entire fetal genome is likely to raise numerous issues. “There are some scenarios that are extremely troubling,” said Marcy Darnovsky, associate executive director of the Center for Genetics and Society, a public interest group in Berkeley, Calif. The tests will spur questions on “who deserves to be born,” she said.

Use of the approach could lead to an increase in abortions because some parents might terminate the pregnancy if the fetus was found to have a genetic disease. But it is also possible that parents may be tempted to terminate if the fetus lacked a favorable trait like athletic prowess.

“You could start doing things more toward the direction of positive selection,” said Dr. Stephen A. Brown, associate professor of obstetrics and gynecology at the University of Vermont.

Moreover, a full fetal genome sequence would turn up numerous mutations for which information is lacking as to whether they cause disease, posing a dilemma for expectant parents and their doctors.

“Our capacity to generate data is outstripping our ability to interpret it in ways that are useful to physicians and patients,” the University of Washington researchers wrote their paper. “That is, although the noninvasive prediction of a fetal genome may be technically feasible, its interpretation — even for known Mendelian disorders — will remain a major challenge.”

The researchers sequenced the genomes of the mother and father. They then sequenced nearly three billion DNA fragments from the mother’s blood. The samples, obtained from a tissue bank, were from unknown donors.

Since people have two copies of each chromosome, they have two versions of each gene. Only one

version is passed to the baby.

Determining which version at any given spot in the father's genome was passed to the fetus was fairly straightforward, since any fragments of DNA in the mother's blood containing a sequence unique to the father had to have come from the fetus.

Determining which of two variants at a given location — call them A and B — the fetus inherited from the mother was more difficult. If the fetus inherited version A, then fragments containing A (which could come from either the fetus or the mother) would outnumber fragments containing B (which could come only from the mother). But since there are relatively few fetal fragments, the difference would be small and hard to detect.

The researchers used an approach they developed to figure out which variations in the mother's genome were likely to be passed to the baby together. That made the problem more tractable than trying to make a call individually at three million locations in the genome.

After it was determined what the fetus inherited from the mother and father, what was left in the fetus's DNA was considered a possible spontaneous mutation. There were initially 25 million such candidates, though statistical approaches narrowed that to 3,800. That still vastly exceeded the 44 such spontaneous mutations found after the baby was born and its cord blood sequenced. Having so many false positive findings of spontaneous mutations could worry parents and doctors.

"There's definitely plenty of room for improvement," Professor Shendure said. But, he added, "This is not science fiction anymore."

The New York Times

June 19, 2012

Before Birth, Dad's ID

By ANDREW POLLACK

It is an uncomfortable question that, in today's world, is often asked by expectant mothers who had more than one male partner at the time they became pregnant. Who is the father?

With more than half of births to women under 30 now out of wedlock, it is a question that may arise more often.

Now blood tests are becoming available that can determine paternity as early as the eighth or ninth week of pregnancy, without an invasive procedure that could cause a miscarriage.

Besides relieving anxiety, the test results might allow women to terminate a pregnancy if the preferred man is not the father — or to continue it if he is.

Men who clearly know they are the father might be more willing to support the woman financially and emotionally during the pregnancy, which some studies suggest might lead to healthier babies.

And if the tests gain legal acceptance, some lawyers say, women and state governments might one day pursue child support payments without having to wait until the birth. Under current law, "until and unless the pregnancy produces a child, any costs associated with it are regarded as the woman's personal problem," said Shari Motro, a law professor at the University of Richmond.

The testing itself, however, can be awkward because it requires a blood sample from at least one of the possible fathers.

Courtney Herndon, after breaking up with her boyfriend, had a brief relationship with a man she regarded more as a friend. She found herself pregnant at age 19, without knowing which man was the father.

The friend also wanted to know, so he agreed to the testing. He turned out to be the father, and the two agreed on child support even before the baby was born.

"I got the test done and was able to go on with my life," said Ms. Herndon, who lives in Fort Polk,

La.

Estimates of the extent of paternal uncertainty vary.

Studies have found a discrepancy rate — when the presumed father is not the biological father — of anywhere from 0.8 percent to 30 percent, with the median being 3.7 percent, according to one review of such studies. Another study found that about 9 percent of birth certificates in Florida, even excluding births to teenage mothers, did not list the full names of the father, though it was not clear how much of this was related to uncertainty. Infant mortality was higher in those cases than if the father's name was on the birth certificate.

It has already been possible to determine paternity during pregnancy using amniocentesis or chorionic villus sampling, the same medical procedures used to test a fetus for Down syndrome. But those procedures are invasive and carry a small risk of inducing a miscarriage, so they are rarely used for paternity testing.

By contrast, the new tests require only blood samples from the pregnant woman and the potential father. And doctors generally do not have to be involved.

That could vastly expand testing, said Sara Katsanis of Duke University's Institute for Genome Sciences and Policy. She is planning a study with one of the testing companies to see if prenatal paternity testing can reduce a pregnant woman's stress.

Some noninvasive paternity tests have been offered over the Internet for about a decade, and there have been various complaints about inaccurate or even fraudulent results.

But experts say the technology has advanced to the point that such testing can now be done reliably. A brief paper describing one such test, developed by a company called Ravgen, was published recently in the prestigious New England Journal of Medicine.

"I have no doubt that these tests will work clinically," said Dr. Mark I. Evans, a professor at the Mount Sinai School of Medicine and director of Comprehensive Genetics, a medical practice in New York that specializes in prenatal testing.

The tests analyze fragments of DNA from the fetus that are present in the mother's blood in tiny amounts. The same approach is now also being used to noninvasively determine the gender of the fetus or whether it has Down syndrome. And researchers recently demonstrated that they could even determine a fetus's entire genome this way.

Ravgen, a small company in Columbia, Md., has been offering its test on a limited basis and charges $950 to $1,650, depending on the circumstances, said Dr. Ravinder Dhallan, the chief executive.

Another test was developed by a company in Silicon Valley called Natera, and is marketed by DNA Diagnostics Center, a leading provider of conventional paternity tests. Thousands of the prenatal tests have been ordered since going on sale last August, executives say. The price is $1,775, compared with around $500 for a conventional postbirth paternity test.

Neither test has received a certification for accuracy that is necessary for use in child custody cases, though Natera has applied. The certifying organization, the AABB, is seriously considering whether it should certify prenatal tests, said Eduardo Nunes, senior director for policy, standards and global development at the organization, formerly known as the American Association of Blood Banks.

Still, some experts urge caution. Natera has not yet published any data about its test in peer-reviewed journals. Ravgen's paper in The New England Journal of Medicine discussed just 30 samples. (The test correctly distinguished the father from a randomly chosen man in all 30 cases.)

The tests could generate controversy if they led to more abortions. However, Matthew Rabinowitz, chief executive of Natera, said that if a woman were intent on terminating a pregnancy based on paternity, she could still get an invasive test. And Dr. Dhallan of Ravgen said the test could persuade women who learned they were pregnant after a rape to keep the baby if they learned the rapist was not the father.

Ravgen's test has been used in a murder case. In 2008, Michael Roseboro, a funeral home director in Lancaster County, Pa., was accused of killing his wife, Jan, whose body was found in the family swimming pool.

To establish a motive, prosecutors wanted to prove that Mr. Roseboro was having an affair with another woman, who was pregnant. But they did not want to wait until the baby was born.

"We became concerned that she might have an abortion, or something would happen and we'd never be able to determine whose child it was," said Craig Stedman, the district attorney in Lancaster County.

The evidence from the prenatal test was not introduced at trial, however, because Mr. Roseboro eventually conceded that he was the father. Mr. Roseboro, who still proclaims his innocence in his wife's death, was sentenced to life in prison.

It is possible that early testing could mean more paternal support for a pregnant woman.

One Seattle-area woman said that when she was pregnant, with two possible fathers, “Neither one really wanted to be involved and then find out the baby wasn’t theirs later.”

When the prenatal test showed that the father was her former boyfriend, he attended the delivery and supported the child. The woman spoke on the condition of anonymity, explaining, “I’m not proud of not knowing who my son’s father was.”

In some cases DNA is not destiny. Ms. Herndon’s test showed that the baby was not her ex-boyfriend’s. But they got back together and married, and he accepted the child, who is now 16 months old.

“We view our daughter as ours, mine and my husband’s,” Ms. Herndon said. The biological father sends gifts and pays child support.

Ariel Bleicher is a science writer and an associate editor at *IEEE Spectrum*. She has written for *Discover*, *Popular Mechanics* and *Scientific American Mind*.

Perils of Newborn Screening

Doctors may be testing infants for too many diseases

The first symptoms often appear a month or two after birth. The babies' muscles stiffen. They lose their hearing and vision, stop sleeping and scream in pain. Some develop seizures. By the time many parents learn that their children have Krabbe disease—a rare genetic disorder that degrades nerve cells—it is too late for the only viable treatment, a transfusion of umbilical cord blood stem cells from healthy donors. Children with full-blown Krabbe who do not receive medical treatment, as well as many who do get treated, usually die by age two.

In some cases, doctors can prevent this grim outcome by screening infants at birth for genetic harbingers of disease. Right now such tests are mandatory in only a few states—something that many parents want to change. "If we don't screen for this disease at birth, those children will never have a chance at life," says Jacque Waggoner, CEO of Hunter's Hope Foundation, one of several advocacy groups lobbying state politicians to add mandatory tests for Krabbe and other rare diseases. The politicians are starting to listen. In the past year four states have passed legislation that requires hospitals to check newborns for abnormal enzyme levels linked to as many as seven new diseases.

Within the medical community, however, doctors are debating the rapid expansion of screening programs. As a whole, the programs have saved many lives. But some experts worry that states may be aggressively demanding tests for diseases that do not always develop in those who show signs of risk or cannot be safely or effectively treated even when they are caught. Doctors who have recently started screening for Krabbe and similar rare diseases are swiftly realizing that, in many cases, the results of such mandatory tests unnecessarily frighten parents and fail to help the children the tests were designed to save.

HARD TO HEAL: Screening infants' blood for signs of disease may not make sense if effective treatment does not exist.

J. SCOTT APPLEWHITE AP Photo

THE BIRTH OF NEWBORN SCREENING

THE CURRENT DEBATE has origins in the earliest forms of newborn screening. By the early 1960s microbiologist Robert Guthrie had perfected a test for phenylketonuria (PKU) that simply required a drop of blood from a baby's heel. Children with PKU suffer brain damage and seizures because they cannot break down the amino acid phenylalanine, which is found in high-protein foods.

Although most states adopted the procedure, a few doctors worried that some babies who did not have PKU would test positive and suffer malnourishment as a consequence of a low-protein diet. Ultimately the doctors' fears proved unfounded. (In a 2006 review of the medical literature on PKU, Jeffrey Brosco and his colleagues at the University of Miami found "no published cases of children who suffered permanent harm after an erroneous [newborn screening] test and treatment for a condition they did not have.") States soon began using similar tests to screen for the likelihood of developing other easily treatable diseases, including congenital hypothyroidism and sickle cell disease.

Today all states require newborn screening for between 28 and 57 medical disorders. Overall, these mandatory programs mark "one of the most significant advances ever in public health," says Stuart Shapira, a medical geneticist at the Centers for Disease Control and Prevention. Of the four million babies born in the U.S. every year, newborn screening identifies 12,500 with medical disorders. Catching and treating many of these disorders early, Shapira says, can prevent intellectual and developmental disabilities, organ damage and death.

Recently, however, doctors have raised new concerns, this time about the repercussions of widespread newborn screening. By the 1990s a tool known as tandem mass spectrometry had drastically expanded the number of disorders laboratory technicians could detect with a single drop of blood—from one to as

many as 20. A mass spectrometer sorts and counts variously sized molecules in the blood, somewhat like a change machine sorts coins. Unusually high levels of certain molecules indicate the enzymes that normally break down these molecules are missing or deficient, which in turn suggests a genetic disorder.

Before 1995 no U.S. state had screened babies for more than eight disorders. A decade later some states were screening for anywhere from seven to 52. States lacked clear consensus on which disorders warranted mandatory screening, says Michael Watson, executive director of the American College of Medical Genetics and Genomics. To remedy the situation, the Health Resources and Services Administration commissioned Watson to review the scientific literature on 84 disorders and to determine which of the screens clearly benefited newborns.

In a report made in 2005 Watson recommended that all states screen for 29 disorders that doctors could clearly predict and treat. He further advised against screening for Krabbe and other diseases because there was not enough evidence that early intervention did more good than harm. Most states currently screen for all 29 recommended disorders, but some, like New York, also test for Krabbe or other conditions outside the uniform panel—including Pompe (a muscle-weakening disease) and Fabry (a metabolic disease causing severe pain). The outcomes of New York's decision to screen for Krabbe underscore why some doctors believe that enthusiasm for screening has gone too far.

When Should Doctors Screen?

According to guidelines proposed by the World Health Organization in 1968, doctors should screen for a medical condition only if:

- ✓ **The condition is an important health problem.**
- ✓ **Doctors can effectively treat the condition.**
- ✓ **Patients have access to diagnostic services and treatment.**
- ✓ **Doctors can recognize a latent stage and early symptoms.**
- ✓ **Doctors have devised an accurate test for the condition.**
- ✓ **The general population understands the rationale behind such tests.**
- ✓ **Doctors understand how the disease develops.**
- ✓ **Doctors agree on which patients should be treated.**
- ✓ **Screening is affordable.**
- ✓ **Doctors plan to continue screening new generations of children.**

PREMATURE ENTHUSIASM

SINCE ITS INCEPTION in 2006 New York's program has tested one million babies and identified more than 200 infants with unusually low levels of some enzymes, indicating risk for Krabbe. Lab technicians verify these results with both enzyme and genetic tests. What investigators have found has been surprising.

Of the 228 infants who tested positive for Krabbe, 24 were found to have genetic markers associated with the disease. So far, however, only four of those children have developed Krabbe symptoms, whereas the other 20 continue to appear healthy. In the vast majority of cases, symptoms of Krabbe appear in early infancy and quickly worsen. A few reports in patient registries describe infants who developed symptoms—albeit mild ones—later in life. The 20 New York infants who screened positive for genetic markers of Krabbe but have not yet shown symptoms may have this late-onset form of Krabbe.

But researchers do not understand late-onset Krabbe well enough to know when, if ever, any of these children will develop symptoms. Only when clinicians detect nerve damage in a battery of invasive neurological exams, including brain imaging and a spinal tap, can they be sure that a child has Krabbe. And only then are they certain that treatment justifies its inherent risks. Studies have shown that early stem cell transplants sometimes stop the disease from progressing, although around 30 percent of children do not survive the procedure and all who do still have trouble speaking and moving their limbs.

Many of the 20 children whose tests suggest late-onset Krabbe but who are not yet sick continue to get neurological exams about every four to six months. Some researchers call these children "patients in waiting." As Jennifer Kwon, a neurologist at University of Rochester Medical Center, puts it, "There's this whole group of children nobody expected to find." The problem, Kwon says, is that parents of patients in waiting do not know what to do with the information they receive from doctors or even what to expect. Parents begin to worry excessively, become overprotective, pursue risky tests and procedures, and avoid routine ones. "It's a huge burden for parents to carry around this knowledge that many of them didn't ask for," agrees Melissa Wasserstein, a pediatrician at Mount Sinai Hospital. "Every time their child so much as trips and falls, they're thinking, 'Oh, my God, does this mean the start?'"

Patricia K. Duffner, who directs the research arm of Hunter's Hope at the University of Buffalo, counters that many parents prefer to know about their child's risk because, if symptoms appear, they will not lose time searching for a diagnosis.

Other experts argue that forcing parents to participate in a public health program when the benefits of screening may not outweigh the emotional trauma is unfair. "So far what's come out of the Krabbe program is we've done a lot of screening, we've scared a lot of parents and we haven't truly helped a kid," says Lainie Friedman Ross, an ethicist and pediatrician at the University of Chicago. According to the doctors who cared for the four New York infants with early-onset Krabbe, one family refused a transplant and the baby died; a second baby died from complications of a transplant; and a third child's affliction continues to progress despite a successful transplant. Only one baby has clearly benefited from screening. At three years, though, he is the size of a one-year-old and recently lost his ability to walk.

Ross fears that newborn screening is destined for another rapid, premature expansion as genome-sequencing technologies become inexpensive enough to use routinely. "With these new test platforms, there is the potential to test for hundreds of conditions we don't fully understand," she says. "If adults can refuse these tests, why should we force them on children?"

Jeff Botkin, a medical ethicist at the University of Utah School of Medicine, has similar concerns. "I think people sometimes forget that we're talking about the state mandating these tests. That's a big deal. If we're going to say to parents, 'You don't have a choice,' there ought to be clear justification for doing a test. We shouldn't just add these things because we can." SA

SCIENTIFIC AMERICAN ONLINE
Comment on this article at ScientificAmerican.com/jul2012

KEEPING YOUR GENES PRIVATE

In spite of recent legislation, tougher laws are needed to prevent insurers and employers from discriminating on the basis of genetic tests

By Mark A. Rothstein

KEY CONCEPTS

- Genetic testing will expand quickly and soon, adding highly targeted data to people's medical records. As those records go electronic, outsiders will find it increasingly easy to peruse an individual's health information.
- Able to uncover private details, health and life insurers could deny coverage to someone with a complex medical condition, and employers could fire or refuse to hire the person to avoid burdening the company health plan.
- Existing laws offer weak protection at best; legislation is needed to give individuals more control over their own data, to limit unauthorized disclosures by others and to penalize wrongdoers.

—The Editors

In years gone by, if colon cancer ran in your family all you could do was wait and worry about whether you might get it, too. Today a genetic test can determine whether you have inherited a greater-than-average risk of the disease and so could benefit from preventive care. The more doctors know about your genes, the better able they are to prevent, treat or cure illnesses.

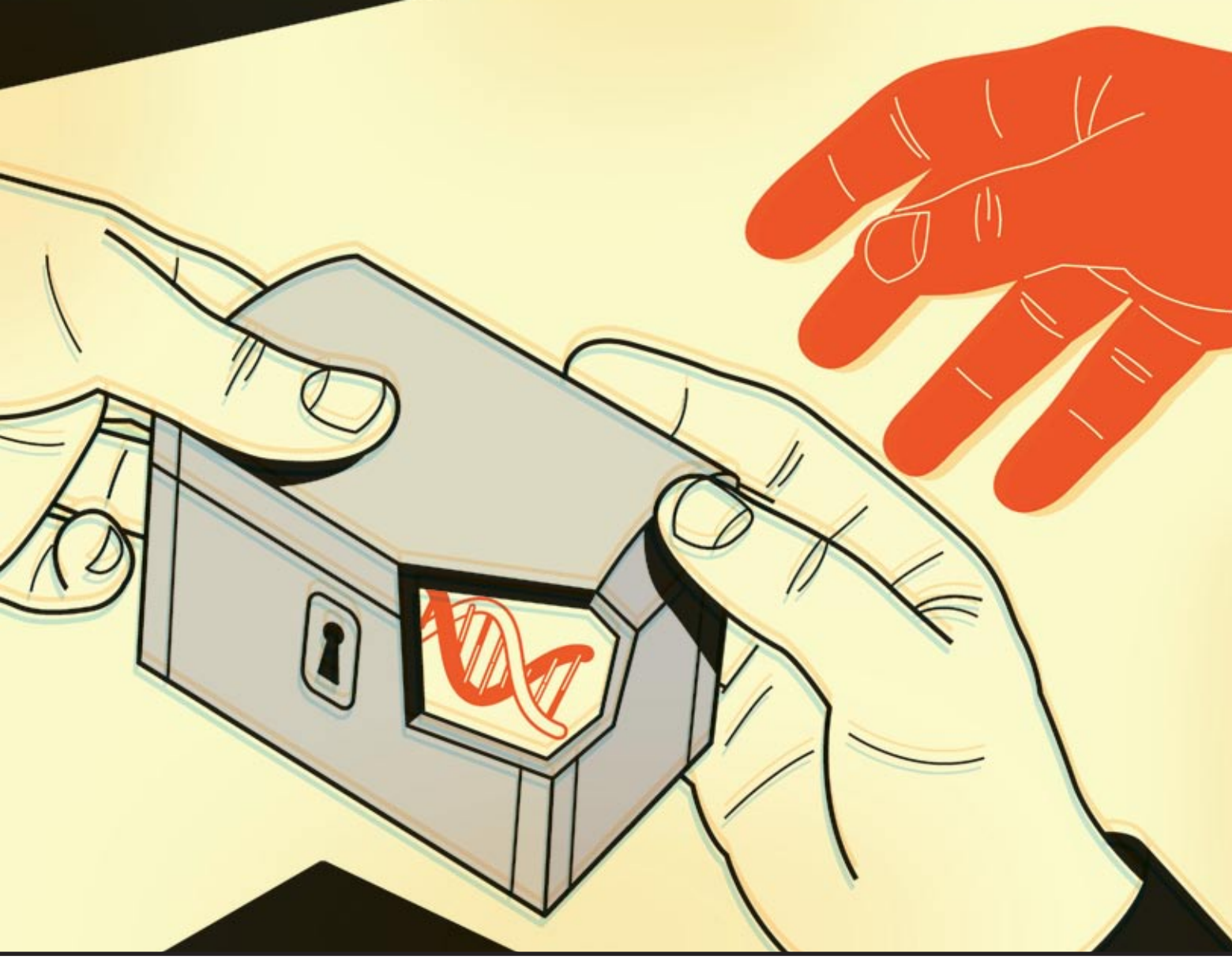
Excitement about such prospects surrounded the start of the Human Genome Project in 1990. But the enthusiasm was soon tempered by widespread concern about the need to protect the privacy of a person's genetic information. Simple tests that could readily reveal an individual's genetic endowment could also readily cause embarrassment or stigma. Furthermore, insurers could deny people health coverage or raise the premiums they have to pay. And employers seeing the results could deny people jobs or fire them. At the same time, scientists and public health officials recognized that the potential to improve health care based on genetic studies across large populations could never be achieved if legions of people refused to participate out of fear that the results could be misused.

Worries about discrimination have not come true—yet. Even though the Human Genome Project was completed in 2003, genetic testing has not become widespread, so there is little in the average person's health record to divulge. And genome-wide analyses remain costly—as much as several thousand dollars each. What is more, scientists still lack standard techniques for making whole-genome scans useful for health risk assessment.

Nevertheless, in many societies—particularly the wealthy ones—genetic testing for multiple disorders will soon become routine. New technologies and scientific discoveries are making the tests more useful and affordable. The health care sector's sweeping transition from paper to electronic records will also make genetic information more readily accessible. Safeguarding genetic privacy is more complicated than many people realize, and recently enacted laws such as the 2008 Genetic Information Nondiscrimination Act offer little protection. Better regulations must be developed soon, before testing spreads and abuses grow.

More Information Everywhere

Figuring out how best to secure genetic privacy would be simpler if "genetic information" and "genetic conditions" were easy concepts to define. But they are not. Medical investigators are finding that almost all illnesses have a genetic component. Distinguishing between genetic and nongenetic health information is becoming increasingly meaningless. Yet policymakers have been inclined to give special protection to genetic information. For legal purposes, the



most common definitions include the results of an individual's genetic tests, those of his or her family members, and the health histories of all these people (because disorders that run in families typically have a genetic link).

The data that fit into these categories are expanding noticeably. In the past decade genetic research and its clinical applications have shifted from disorders linked to a single gene, such as cystic fibrosis and muscular dystrophy, to more common and complex ills characterized by the interactions of multiple genes and environmental factors, including asthma, cancer, cardiovascular disease and diabetes. More than 1,500 genetic tests are now in use, and hundreds more are being developed. As these tools become part of standard medical practice, including primary care, most, if not all, health records will contain substantial genetic information.

Genome-wide analyses could vastly expand those contents. These tests can look for single changes in hundreds of thousands of nucleotide bases—the famous A, T, C and G "letters" of DNA code—associated with particular illnesses and conditions. Although most scientists think that it is premature to apply this technology routinely, some companies such as 23andMe in Mountain View, Calif., and deCODE Genetics in Reykjavik, Iceland, have started aggressively marketing genome-wide scans, even if they do not have a license to operate as a medical laboratory. Within a decade, whole-genome sequencing that reads all three billion bases in human DNA might well be available for less than $1,000.

At least two other factors will add to the amount of information in health records. The great desire for personalized medicine—drug therapies tailored to each person's body to improve effectiveness and reduce side effects—depends on genome-wide analytical tools. This "pharmacogenomic" testing is already becoming standard practice in selecting drugs and doses for treatment of certain cancers, and the trend

HARRY CAMPBELL

will continue. Likewise, "toxicogenomics"—the use of genome-wide tools to study how individuals respond to toxins—is becoming more important in assessing a person's health risks in the workplace and in the general environment.

Networks Amplify Risk

The challenge of protecting health information is compounded by an increasing reliance on digital data. Medical records of all kinds are shifting from largely paper-based systems to electronic health records (EHRs), which should improve the quality of care and reduce its cost. The transition is under way in many developed countries. In the U.S., a Nationwide Health Information Network (NHIN) is being developed as a "network of networks." Its key goal is establishing electronic formats that will make records of all kinds compatible and thus easy to transport across networks and across the country. Ultimately, a person's EHR will include all his or her medical information from "cradle to grave." The Office of the National Coordinator for Health Information Technology in the Department of Health and Human Services is leading the NHIN's development, but state governments and the private sector are engaged in research, development and trial implementation.

The NHIN raises contentious issues. In a paper-based system, privacy is mainly protected by chaos. Precisely because the system is fragmented, people find it impossible to compile, or even to locate, an individual's records from a multitude of providers in different locations over extended periods. But comprehensive, longitudinal records will inevitably contain sensitive information. Individuals will no longer have the option of "selective recall" in giving facts to health care providers or of obtaining care from one provider without the knowledge of another. Unlike today, an old diagnosis of depression made at a college mental health clinic or the results of a genetic test taken because of family history will become a permanent part of one's EHR. Many people with conditions that might stigmatize them, such as a history of substance abuse, might delay or forgo treatment. Such a result could be disastrous for individuals and for public health.

People fear they might not get a job if they could be a burden to the company medical plan.

INTRIGUED BUT WARY

According to a May 2008 Knowledge Networks survey:

Forty-seven percent of Americans are interested in using online personal health record services such as Google Health or Microsoft HealthVault. The services allow consumers to control their own medical records online.

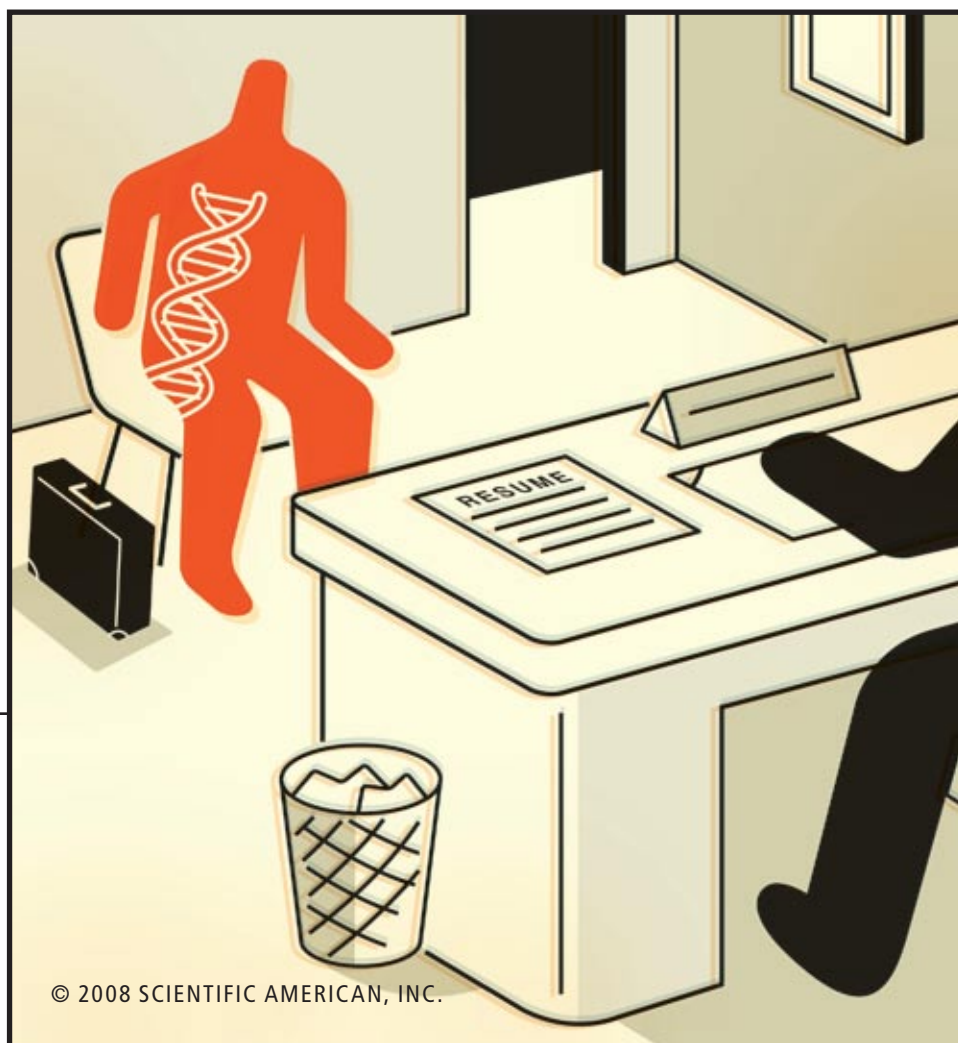
Ninety percent of the respondents, however, indicated they would be wary about the services' ability to keep records private.

In response, the Markle Foundation has recommended ways to make such systems as private as possible. Provisions would allow consumers to audit who is accessing their medical data and to dispute information provided by health care providers.

A full report is not needed to render effective care, however. A physician treating a sprained ankle does not need to know if a patient has a predisposition to breast cancer. A dentist filling a cavity does not need to find out about a family history of Huntington's disease.

To protect patients from unnecessary disclosures of sensitive information, countries such as Canada, the Netherlands and the U.K. are considering ways to restrict which information is revealed to which health care providers. These measures include giving patients complete control of their health records, permitting individuals to remove certain old information, limiting disclosures only to details needed for a given diagnosis or type of provider, applying special rules to sequester especially sensitive information, creating a subset of basic health data that would be available to all providers and establishing independent health record banks to disclose files according to a patient's direction. In Denmark's EHR network—one of the most advanced—people can "block" any information in their records. Although this option is rarely exercised, it is greatly valued.

The U.S. has no such measures in place. This



HARRY CAMPBELL

past February the National Committee on Vital and Health Statistics (which advises the secretary of health and human services) recommended that individuals be able to prevent the routine disclosure of sensitive health information in predefined categories, such as domestic violence, substance abuse, mental health, sexually transmitted diseases and genetic information. But methods for doing that have yet to be created. And how to strike the right balance between broad and narrow disclosure remains unclear. If patients have too much control, physicians will not have confidence in the accuracy or completeness of the records. In response, they will likely feel compelled to retake histories and order new tests, undermining the efficiencies of networks and adding cost to care. On the other hand, if patients have too little control, many may engage in defensive steps such as opting out of networks, paying cash for off-record services or declining certain care altogether.

Other issues must also be resolved. For example, should privacy rules be set for systems that scan electronic records and advise clinicians on possible drug interactions, so the systems do not divulge actual drugs taken? Should health care providers see an electronic notation in a patient's file indicating that certain health information has been made unavailable at the patient's request? And in such cases, will doctors have a way to lift those restrictions if the person needs emergency care?

Weak Laws

With more genetic information and far-reaching electronic networks on the horizon, legislation protecting health privacy is essential. Unfortunately, comprehensive laws do not exist in the U.S. The closest thing to a national safeguard is the 1996 Health Insurance Portability and Accountability Act (HIPAA) and the 2003 Privacy Rule attached to it. The Privacy Rule spells out the permissible uses and disclosures of individual health information by providers, plans and record clearinghouses.

There is a big loophole, however: the Privacy Rule applies only to entities that handle health claims data electronically. Hundreds of thousands of providers still do not, including doctors who take cash payments exclusively, fitness clubs that ask for medical information when putting members on workout plans and health care providers who work under contract to third parties, such as personnel in on-site employer clinics. A related problem is the lack of enforcement. About 36,000 complaints related to the Privacy Rule were filed with the Department of Health and Human Services's Office for Civil Rights between April 2003 and May of this year. Although corrections were made, only one civil monetary penalty has been assessed to date. Wrongdoers face few deterrents.

In addition, HIPAA only applies to entities involved in health care. The public, however, is most worried about stigma or discrimination from others. People fear complications when applying for a job, obtaining a life insurance policy or filing for workers' compensation benefits. Yet it is common for administrators involved in these and other everyday situations to require people to sign an authorization directing their provid-

DAVID ELLIS

Should Family Members Be Warned?

Sarah, a 40-year-old mother of three, has found out from various tests that she has an elevated risk of Alzheimer's disease, as well as of breast cancer. Does she have a legal or moral obligation to tell her children or close relatives that they, too, might be at high risk of getting these illnesses in the future?

The legal issue is straightforward: no court has held an individual liable for failing to warn a relative about genetic test results. The moral issue depends on many factors, including the severity of a genetic condition, the number of years before it is likely to produce symptoms, and whether the condition is treatable. The nature of relationships (parent and child) and their emotional closeness matter, too, as do relatives' ages, their interest in knowing about the chance of future ills, and the individual's own concern about not divulging his or her personal problems.

The nature of the danger often plays a strong role. In rare cases, genetic conditions can be lethal if combined with environmental stressors. For example, individuals with the genetic mutation for malignant hyperthermia can die during surgery if certain anesthesia is used. People with hypertrophic cardiomyopathy can suffer sudden death from strenuous exercise. The potential for these types of harm warrant notifying at-risk relatives.

Yet sharing one's genetic information with family members can be perilous. Testing may reveal, for instance, that the man everyone thought was a child's father actually is not, sending a family into turmoil. Genetic counselors can help people decide whether to undergo genetic testing and how to respond to possible results, but currently only 2,500 counselors practice in the U.S. The most common mistake is getting tested and waiting for results before considering what to do. Anyone contemplating testing should determine in advance whether to share the results with close relatives. There is no simple answer. The best advice is to consult with professionals and think ahead about the possible consequences. ***—M.A.R.***

ers to release their health information. According to one estimate, at least 25 million such authorizations occur every year in the U.S.

The parties requiring the disclosures are usually acting lawfully. And one's health can have legitimate bearing on decisions. An electric power company, for example, would not want to hire someone who is prone to seizures to fix wires at the tops of utility poles. The problem is the amount of information disclosed. The electric company has no need to know whether a job applicant has a genetic mutation that may increase susceptibility to heart disease decades from now. Judging a worker's compensation claim for a broken leg does not require reproductive health information. An automobile insurance adjuster handling a claim for a chipped tooth sustained in an accident does not need any genetic test result. But most of the laws authorizing disclosure of health information are written so broadly that no limits are placed on the scope of the requests.

Ironically, EHR networks could solve this problem. Software programs could scan electronic records and select only the data related to a specific inquiry. Yet this capability requires the use of "contextual access criteria"—software algorithms specifying that, for an inquiry of type X, only data A, B and C are needed. For example, contextual access criteria would disclose only information bearing on mortality risk to a life insurer. This technology is feasible but not yet available. And because commercial demand alone probably will not provide adequate incentives to develop the technology, laws may be needed to require it.

Legislation of Little Help

Given the general weakness of federal regulations, various state legislatures have enacted their own protection laws. In so doing, the states have adopted the notion of "genetic exceptionalism"—that genetic information is treated differently from other forms of sensitive health information. Whether this approach is desirable is an open question, but it parallels how some mental health, substance abuse and HIV information is handled.

Although the laws vary, 12 states require people to give written, informed consent for a genetic test, and 27 states require express consent to disclose test results. Nevertheless, these laws, like the federal regulations, continue to allow insurers and employers to legally require individuals to sign an authorization for the release of their medical information. As a result, 47 states have laws that prohibit insurers from denying or restricting coverage or charging different rates, based on an individual's genetic information. HIPAA already covers these cases for people in employer-sponsored group health plans, however, so the state laws in effect only extend protection to people who buy individual insurance.

Other laws in 35 states prohibit employers from requiring a genetic test as a condition of employment and from using predictive genetic information to deny an individual a job. Yet after a conditional offer of employment, the laws allow an employer to require prospective employees to authorize the release of their health records as a condition of being hired. The states differ on whether genetic information may be disclosed at this time, but that provision is largely immaterial: it is impracticable for anyone to excise genetic information from paper records and equally infeasible to exclude it from electronic records until the contextual access algorithms are devised.

Given such shortcomings, Congress has been under increasing pressure to improve privacy. In May members finally passed the Genetic Information Nondiscrimination Act (GINA), which had been pending since the mid-1990s. The act prohibits health insurance companies from discriminating in providing coverage, and in setting rates, on the basis of genetic predispositions. Unfortunately, the legislation is not much better than or even different from many state laws, and it doesn't cover life, disability or long-term care insurance.

Universal Solutions

The flaws in GINA, HIPAA and state regulations are not loopholes or oversights. They are the natural result of a health care system in which individual coverage is medically underwritten [see "Reflections on Privacy 2.0," by Esther Dyson, on page 50]. People in the U.S. can obtain insurance in one of three ways: a group health plan such as that offered by most employers, individual insurance, or federal programs such as Medicare and Medicaid. For group and individual plans, underwriters calculate the individual or collective health risks of those covered and impose premiums based on the relative risk they represent. Of course, one prime purpose is to protect the financial interests of the insurer. Insurers want to know about each person's past ailments and the possibility of future illnesses (genetic and otherwise) so they can bet-

GENE DETAILS COMING SOON

The 1000 Genome Project, an international research consortium started this year, intends to create a map of the human genome that is five times more detailed than the one created by the International HapMap Project.

HapMap discoveries spawned the recent explosion of genome-wide studies that have identified more than 130 genetic variants linked to a range of diseases, including type 2 diabetes, coronary artery disease, prostate and breast cancers, rheumatoid arthritis and mental illnesses.

In the next three years the 1000 Genome Project hopes to sequence the genomes of at least 1,000 people drawn from populations around the world. For more see www.1000genomes.org

[THE AUTHOR]

Mark A. Rothstein is chair of law and medicine and director of the Institute for Bioethics, Health Policy and Law at the University of Louisville School of Medicine. From 2001 to 2008 Rothstein chaired the Subcommittee on Privacy and Confidentiality of the National Committee on Vital and Health Statistics, which advises the U.S. secretary of health and human services.

LESLIE FRIESEN *University of Louisville*

HARRY CAMPBELL

ter determine price and ward off those who might make huge claims.

None of the privacy laws mentioned apply to Medicare or Medicaid, because technically these programs are entitlements, not insurance. Different laws attempt to protect information within these programs, but the government has no real incentive to look at anyone's genetic information because there are no rates to adjust.

Indeed, concerns about keeping information private are best addressed by a national system of universal health care, as in Canada. In universal plans, risk is spread across the entire population, and the plan is funded by the entire population. Whether any given person has a high risk for any disease has no bearing on the equation, so there is no incentive for others to seek protected information. The situation eliminates people's two greatest worries: that they will have trouble obtaining or will be dropped from health insurance, and that they will be denied a job because their medical conditions could impose a burden on the company's health plan.

Complications in obtaining life insurance must still be addressed, however. And health information still has to be made secure so records are not stolen or improperly disclosed. But the big incentives to discriminate largely disappear.

MORE TO EXPLORE

Genetic Privacy: A Challenge to Medico-Legal Norms. Graeme Laurie. Cambridge University Press, 2002.

Genetic Privacy. Pamela Sankar in *Annual Review of Medicine,* Vol. 54, pages 393–407; 2003.

Genetic Exceptionalism and Legislative Pragmatism. Mark A. Rothstein in *Hastings Center Report,* Vol. 35, No. 4, pages 27–33; July/August 2005.

Ensuring the Privacy and Confidentiality of Electronic Health Records. Nicolas P. Terry and Leslie P. Francis in *University of Illinois Law Review,* pages 681–735; 2007.

Canada and the Netherlands may give patients complete control of their health records.

The U.S., though, is unlikely to adopt universal health care anytime soon, even though it is front and center in the 2008 presidential campaign. Thus, better privacy laws must be enacted, even though some observers say new genetic technologies add little threat to privacy. Although very few legal cases have been brought over discrimination in employment or health insurance, almost all medical geneticists and genetic counselors know of numerous patients who have declined to undergo genetic testing because they feared possible discrimination or stigma. (According to Francis S. Collins, former director of the National Human Genome Research Institute, one third of eligible people decline to participate in genetic research because they fear discrimination.) Furthermore, the number of genetic tests and the number of people taking them, along with the tests' usefulness, will increase significantly in the next decade. And EHR networks will make it easy to disclose the information widely with the click of a mouse.

As the U.S. and other countries contemplate better ways to deal with genetic information, policymakers are seeing that protecting privacy is neither cheap nor easy. Improved security measures can keep information from being disclosed without authorization, but restricting the scope of authorized disclosures is equally important. It is essential, and challenging, to decide which individuals and entities have a right to which information and for what purposes.

Effective legislation should, at minimum, include four elements. First, it should address the underlying difficulties in gaining access to health insurance and carefully balance the rights of employers and employees. Second, legislation should limit nonmedical uses of predictive health information, including for life insurance, disability insurance and long-term care insurance. Third, any legislation should limit the scope of disclosures, penalize wrongdoers and provide remedies for people harmed by wrongful disclosures. And fourth, EHRs and EHR networks should be designed so that they can limit disclosures to relevant health information. Tackling these matters will provide an effective first step toward shaping the future of medical privacy. ■

THE WALL STREET JOURNAL.
WSJ.com

U.S. NEWS | Updated January 17, 2013, 3:18 p.m. ET

Researchers Identify Anonymous DNA Donors

By AMY DOCKSER MARCUS

Genetic information stored anonymously in databases doesn't always stay that way, a new study revealed, prompting a debate on how much privacy participants in scientific research can expect in the Internet era.

Tension has long existed between the need to share data to drive medical discoveries and the fact many people don't want personal health information disclosed. The growing use of genetic sequencing makes this even more challenging because genetic data reveals information not only about an individual, but also about his or her relatives.

Previously

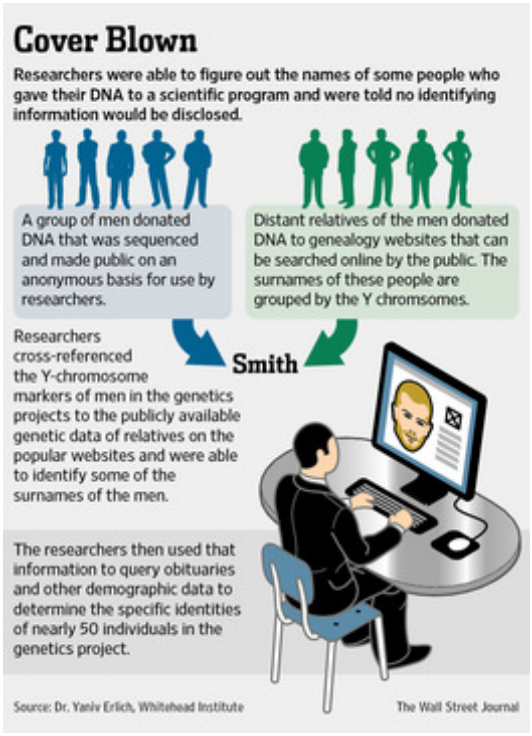
- Doubling Up on Research Using a Database of Twins 11/5/12
- Personalized Medicine Moves Closer 10/31/12
- Citizen Scientists Take Control of Their Health Data 12/3/11
- What You Should Ask Before You Give Up DNA 8/30/10

In a paper published Thursday in the journal Science, researchers were able to determine the identities of nearly 50 people who had submitted genetic information as part of scientific studies. The people were told that no identifying information would be included in the studies but warned of the remote possibility that at some point in the future, their identities might become known.

"We have been pretending that by removing enough information from databases that we can make people anonymous. We have been promising privacy, and this paper demonstrates that for a certain percent of a population, those promises are empty," said John Wilbanks, chief commons officer at Sage Bionetworks, a nonprofit organization that promotes data sharing, who wasn't involved in the study.

The public and scientific community are especially concerned about DNA privacy since they worry that genetic information—which can show susceptibility to certain diseases and other ailments—might be used by insurers, employers or others to discriminate against people.

In the new study, the researchers, led by the Whitehead Institute for Biomedical Research in Cambridge, Mass.,



used genetic information of people whose genomes had been anonymously published as part of the 1000 Genomes Project, an international collaboration to create a public catalog of data from at least 1000 people of different ethnic and population groups.

Using a computer algorithm, the researchers focused on identifying unique genetic markers on the Y chromosome of men in the project. They then searched publicly accessible genealogy databases that contain both Y chromosome information and men's surnames.

Such genealogy sites, which people join in hopes of compiling their family tree, sometimes include Y chromosome data because it is passed from father to son and can be traced back generations. Some sites group the genetic information with surnames.

When they got a match to a surname, the researchers ran numerous Internet searches to collect data on each individual's family tree, including obituaries, which often list the names of a deceased's family members. They also searched for demographic data on the public website of the Coriell Institute for Medical Research, a nonprofit in Camden, N.J., that houses collections of genetic material.

With the family-tree data, they were able to identify nearly 50 men and women who participated in genetic studies. "It only takes one male," said Yaniv Erlich, a Whitehead fellow, who led the research team. "With one male, we can find even distant relatives."

Dr. Erlich said the technique works best for people who have the highest participation in genetic genealogy services, upper- and middle-class Caucasian Americans. Based on their findings, they estimated their technique would have a success rate in identifying the last names of 12% of U.S. Caucasian males in similar studies.

The specific people in the Science study who names were discovered weren't disclosed in the paper.

Hank Greely, director of the Center for Law and the Biosciences at Stanford University, said the study raises important questions about expectations of privacy. In an age when genetic information is being collected as part of medical care and can be correlated with the personal information that people freely post online, Mr. Greely said the medical and scientific communities needs to be clear that "we cannot promise people confidentiality."

The issue of protecting privacy in genetic studies isn't new. In 2008, geneticist David Craig showed that he was able to trace pooled genetic data that was available online back to an individual who had participated. As a result, the National Institutes of Health and the Wellcome Trust tightened access to these collections of DNA. The scientific community protested, but officials felt they had no choice because participants had been explicitly promised anonymity, said

Eric Green, director of the National Human Genome Research Institute at NIH.

Dr. Green, who co-authored a perspective piece that accompanied Thursday's paper, said steps have already been taken to make it more difficult for others to duplicate the results of Dr. Erlich's team. For instance, the ages of the people in the study no longer are publicly available on the Coriell website, said Courtney Kronenthal, director of communications and development at Coriell.

David Altshuler, co-chair of the steering committee for the 1000 Genomes Project, said the people who enrolled in the study were told that all steps would be taken to keep them anonymous, but that technological advances one day might make it possible to identify them.

Dr. Altshuler said he favors offering research participants a variety of options about how much to share.

"If they choose to share that's a very admirable thing because by sharing freely, progress for everyone is accelerated, and if someone is not comfortable we should respect that too and find ways for them to still participate in research," he said.

Write to Amy Dockser Marcus at amy.marcus@wsj.com