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.

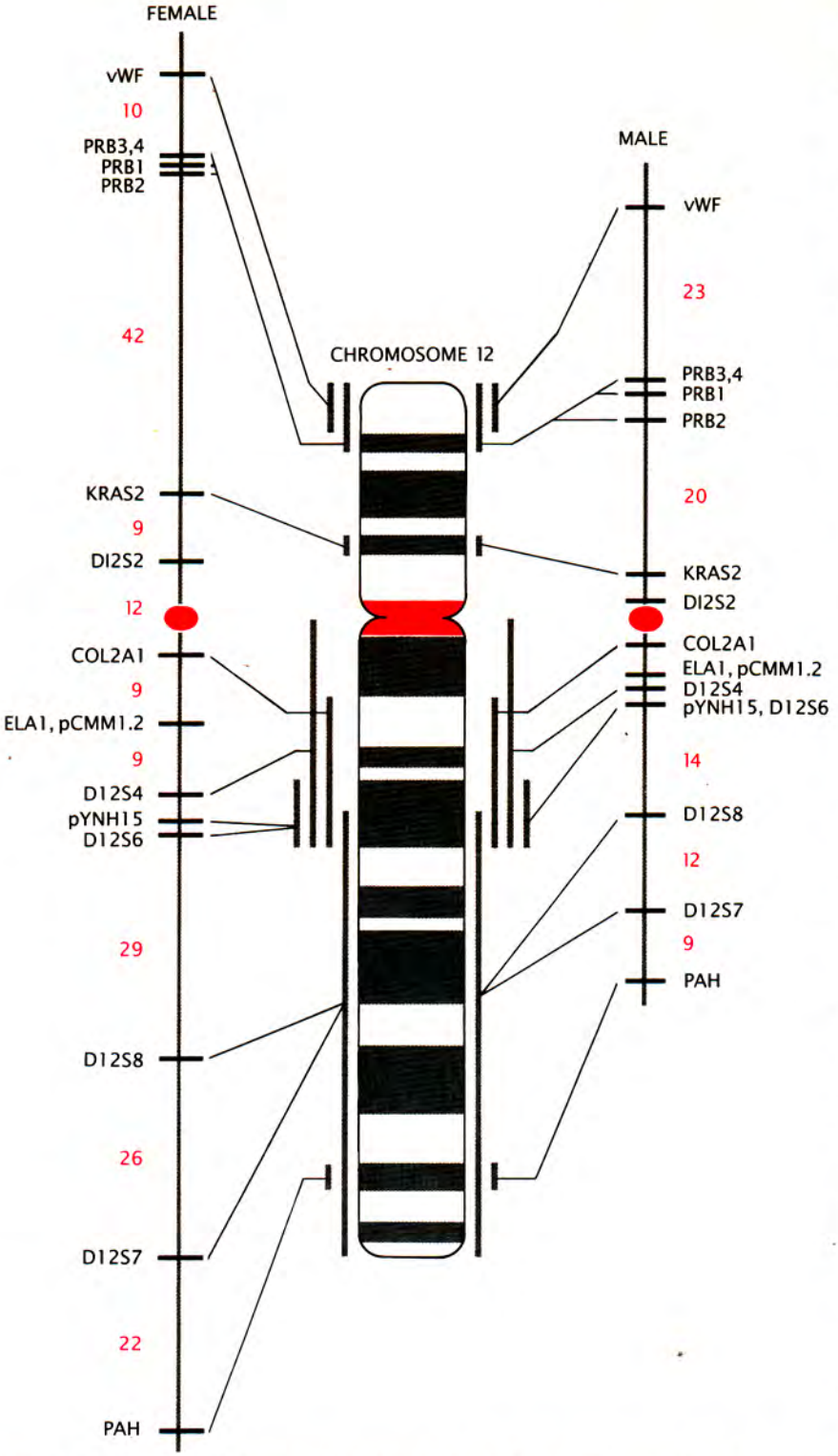
Other markers lying farther from the disease gene will recombine with the gene more frequently, so that the disease will be less likely to be inherited together with any given marker allele. In the extreme case, for a marker and a disease lying well apart on a chromosome, the recombination frequency 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

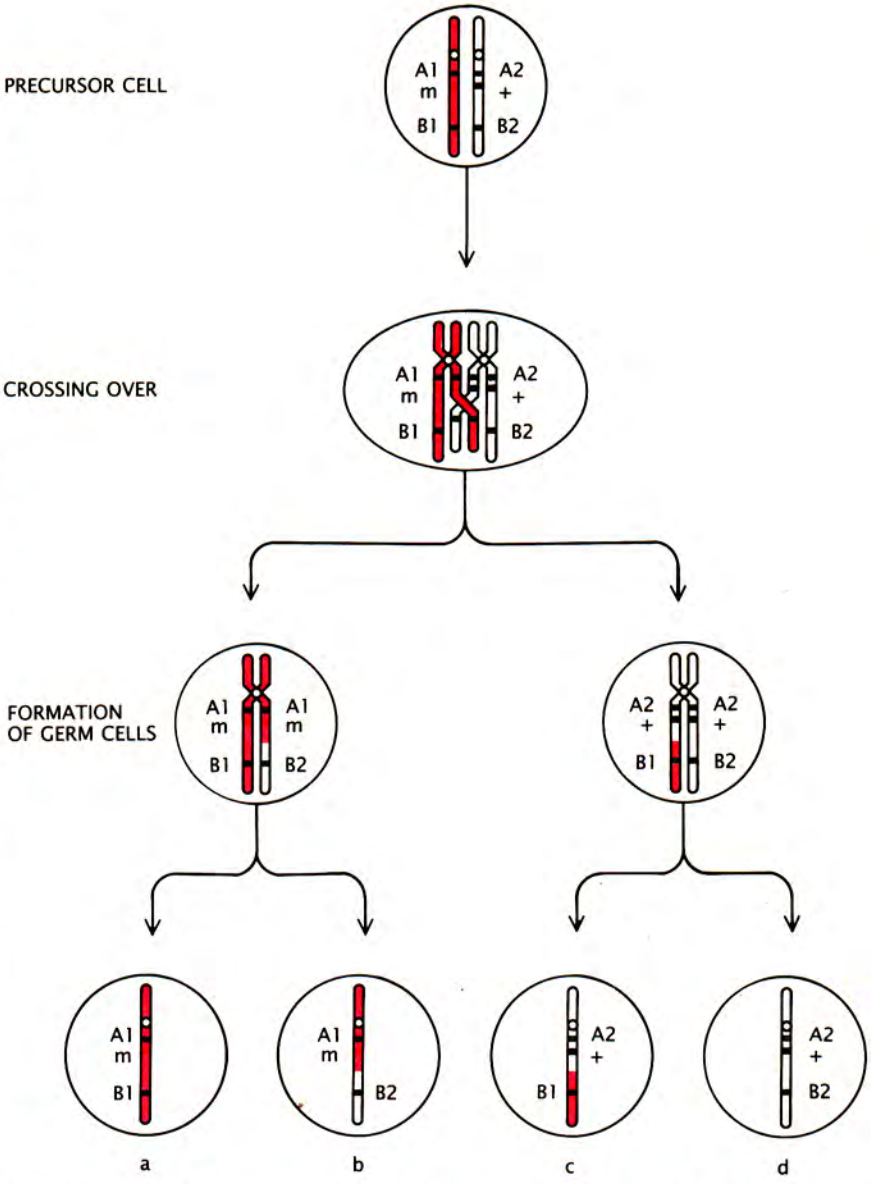
radioactive label makes it possible to detect the position of the fragments, which reveals their size.

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.

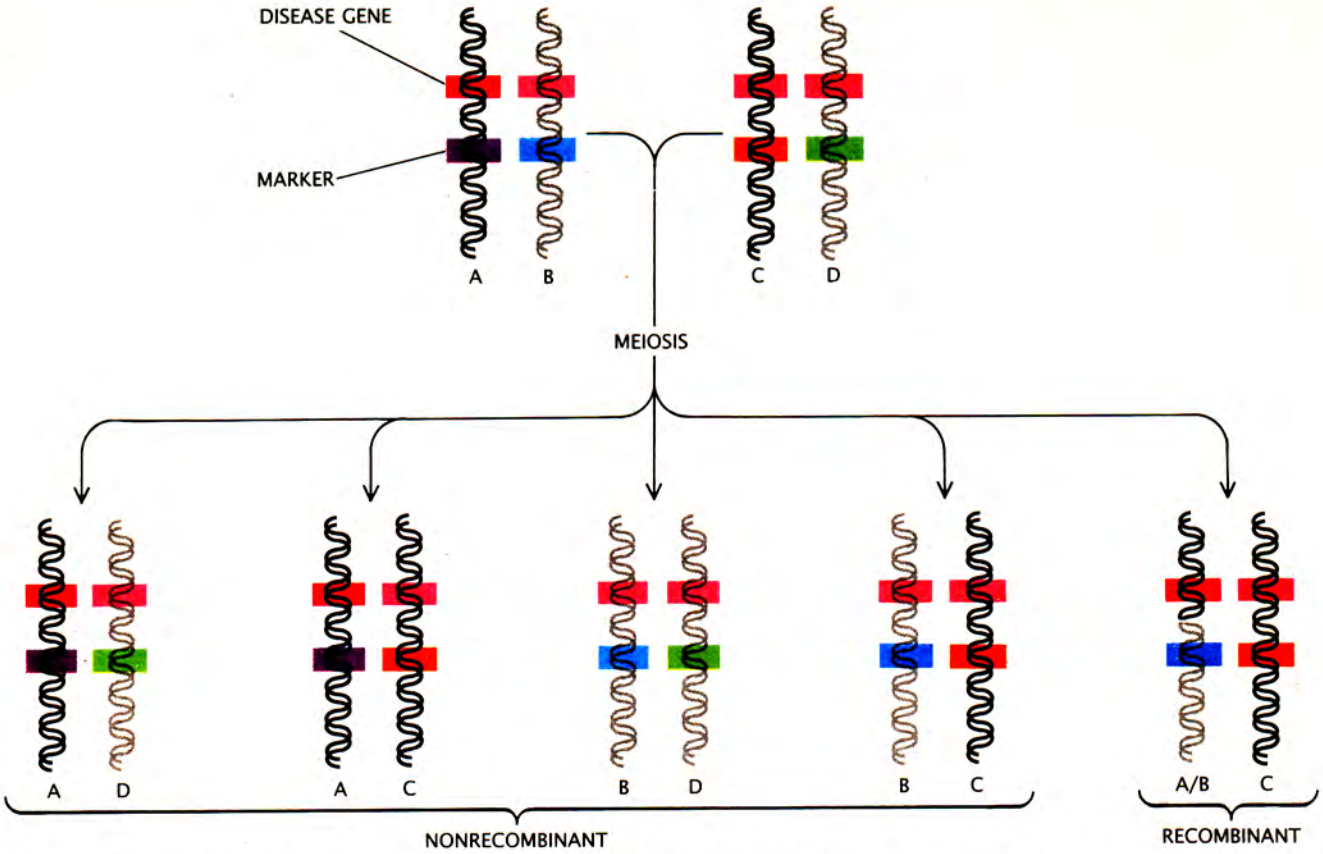
This DNA marker, defined by the RFLP, is found in one form or another in every individual, healthy or diseased. But if a genetic disease is passed down a pedigree together with a particular allele of the RFLP, the mutant gene can be assumed to lie in the same chromosomal region as the marker. In a second 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.)

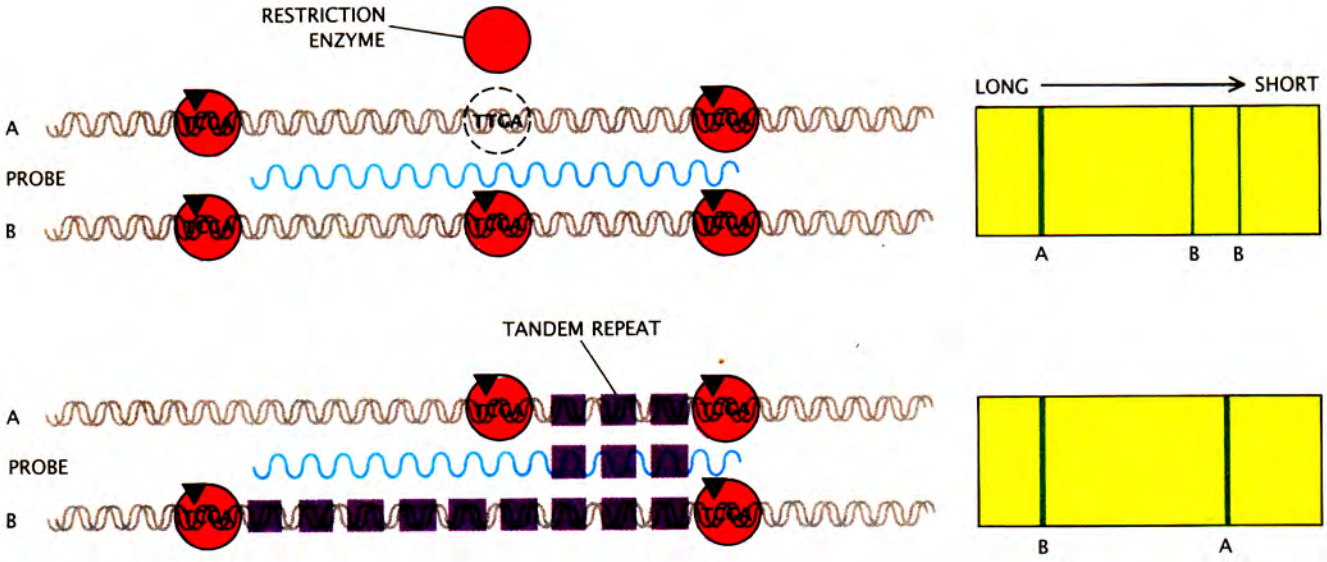
Another kind of DNA polymorphism creates many different versions of an RFLP. At many sites on the human DNA a single sequence that does not code for any protein is repeated many times. The origin and significance of these "tandem repeats" is a mystery, but for linkage mapping they offer a practical advantage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

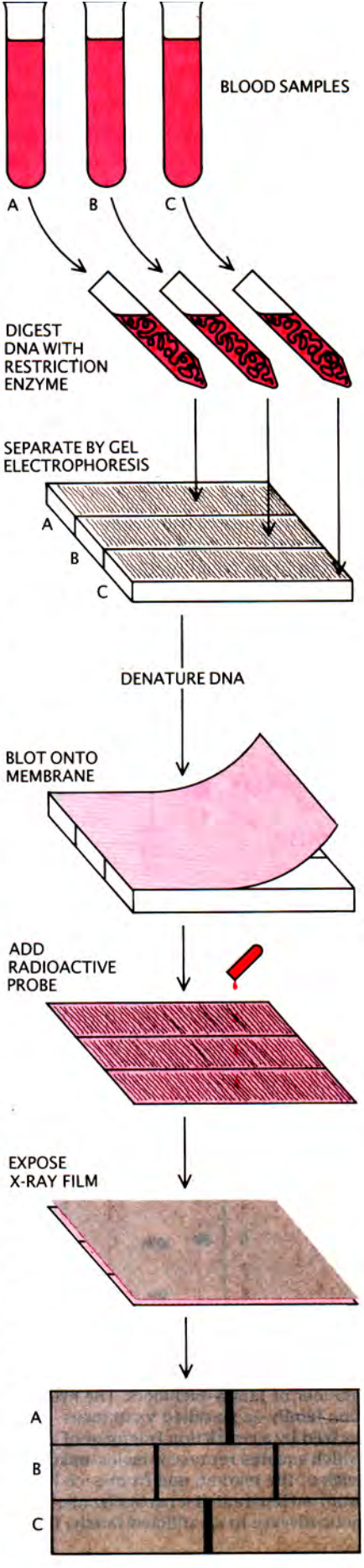
Probes for markers based on variations in the number of tandem repeats (VNTR's) can be developed more systematically than probes for ordinary markers. Alec J. Jeffreys of the University of Leicester recently recognized that the repeated sequences at many VNTR loci in different parts of the genome show 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.

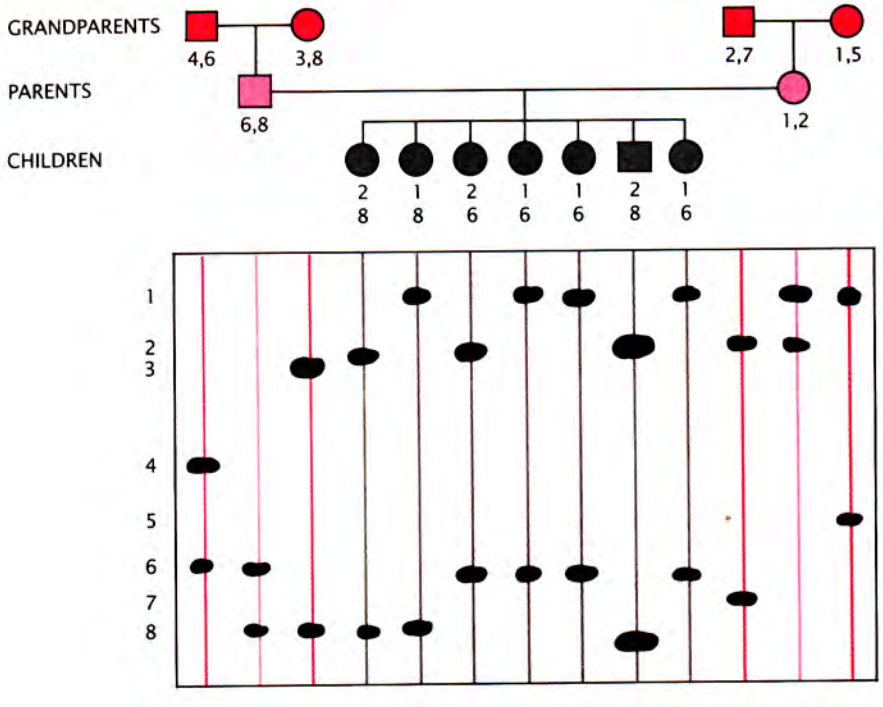
Tightly linked, flanking markers for cystic fibrosis, peripheral neurofibromatosis and familial polyposis are in hand, and a new, tightly linked marker has been identified for Huntington's disease. The search for the causative gene of each disease is in high gear. The approaches vary, but a common tactic 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,



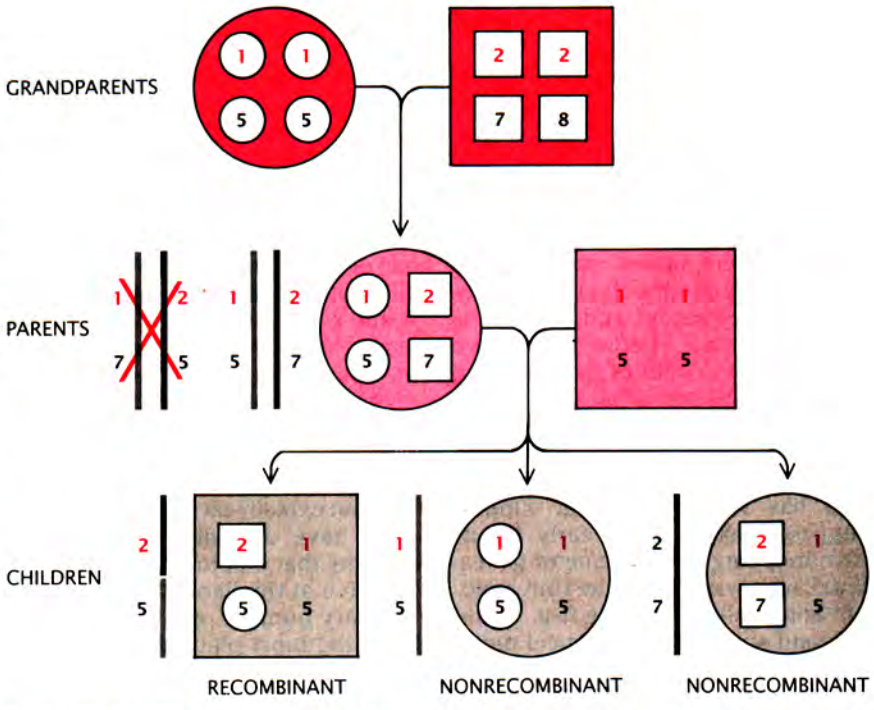
INHERITANCE OF AN RFLP can be traced by comparing restriction fragments from a number of family members. The RFLP marker that was analyzed in this three-generation family—a so-called VNTR locus—has many different alleles, each of them characterized by a restriction fragment of a specific size. Each individual 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.

MARK KINSINGER (*left*) and his younger sister, Ruth (*right, not their real names*), were born with the same genetic disorder. Mark, not diagnosed until age four, suffered irreversible brain damage. Since then, a pilot screening program and early intervention have fully prevented disability in other children with the condition, including Ruth.

MEDICINE

GENOMICS FOR THE PEOPLE

A children's clinic raised and supported by Amish and Mennonites proves that high-tech genetics research can be harnessed right now to prevent disease

By Kevin A. Strauss

Levi and Emma Kinsinger operate a small greenhouse in southern Pennsylvania. On November 6, 2002, they traveled 450 miles round-trip by taxi, at a rate of a dollar per mile, to bring their eldest son—Mark—to the Clinic for Special Children in Strasburg, Pa. At age four, Mark was frail and socially detached. He lay on the floor in constant, restless motion. His eyes roamed but did not fix, and he was unmoved by sound. From time to time, a guttural noise escaped his throat as he shook violently. The Kinsingers' question, one I've heard countless times in my work as a pediatrician, gave voice to their quiet desperation:

"What can we do to help our child?"

Photographs by Grant Delin

Kevin A. Strauss, who earned his M.D. from Harvard Medical School, is medical director of the Clinic for Special Children in Strasburg, Pa.

Our clinic is a medical home for children like Mark. (For privacy, I have changed the names of all patients and their families.) Its sturdy timber frame, erected by Amish and Mennonite hands, encloses a modern pediatric office equipped with an arsenal of high-tech gene-sequencing tools. We serve the North American "Plain" communities descended from European Anabaptists who fled to the New World in the 1600s to 1800s seeking religious asylum. Today's Plain people live in small, isolated Christian settlements throughout North America and eschew modern ways. Electricity and telephones are commonly forbidden in the home, codes of dress and conduct emphasize group cohesion, private and government insurance are rejected, and members distrust technologies that erode social interdependence.

The Plain people choose to live differently in the modern world, but every parent knows what it means to fear for a sick child: "Will my daughter ever walk?" "Can you stop the seizures?" "Is it autism?" Such are the questions that move us to translate the complex language of modern biochemistry and genetics into meaningful answers for children and families. To date, our laboratory has identified more than 170 different disease-causing gene mutations disproportionately represented among the Plain people. Nearly half endanger the developing brain and, left untreated, cause death or disability in children. Rapid, affordable, on-site molecular testing opens a precious window; it allows us to expose future health threats, craft more precise therapies and preempt disease before it strikes.

Our collaborative relationship with the Plain people also provides a glimpse into how genomics research will transform understanding of more common diseases. With the cooperation of a few dedicated Amish families, we recently discovered a specific genetic variation that appears to be linked to bipolar (manic-depressive) disorder, which affects between 2 and 4 percent of people worldwide and remains woefully underdiagnosed and undertreated. Linking a genetic variation to bipolar disorder moves genomics one step closer to the medical mainstream; it challenges the medical research community to close the gap between what we know about the causes of human suffering and what we can do for the people who need our help.

PROGRESS, ONE CHILD AT A TIME

WHAT THE KINSINGERS NEEDED was clarity. Within a few days we detected a constellation of chemical abnormalities in Mark's blood that implicated deficiency of an enzyme—5,10-methylenetetrahydrofolate reductase (MTHFR)—as the cause of his disabilities. Our lab director, Erik Puffenberger, worked quickly to discover an error in both of Mark's MTHFR-coding genes. This knowledge allowed us to diagnose three more affected children from the Kinsingers' home settlement.

I searched the medical literature and found the first description of MTHFR deficiency, published 30 years earlier by S. Harvey Mudd and his colleagues. Mudd was a legend in the small world of research devoted to intermediary metabolism, the collective processes that convert food into the energy and building blocks of cells. He elucidated what came to be known as the transsulfuration pathway—a complex network of chemical reactions that recycles an essential amino acid, methionine, while simultaneously supplying methyl groups (CH_3) to molecules throughout the body. Methionine is indispensable for the growth of the brain and other tissues, and methyl tags profoundly affect how these tissues function. MTHFR is a vital link in the chemical supply chain; lacking this enzyme, Mark had suffered the devastating neurological consequences of cerebral methionine and CH_3 deprivation.

I called Mudd, then age 75 and emeritus researcher at the National Institute of Mental Health. He generously guided me through the complexities of transsulfuration and suggested a treatment: an over-the-counter compound called betaine, which supplies the brain with methionine and CH_3 via an alternative metabolic pathway and can be administered as a dietary powder for just 60 cents a day. In the months that followed, I frequently made the four-hour trip to the Kinsingers' settlement with clinic nurse Christine Hendrickson. We traveled from farm to farm, carefully assessing the effects of betaine on our young patients. Armed with a cooler of dry ice, a portable centrifuge and a power inverter in my car's cigarette lighter, we spun and froze blood samples in the field. We shipped them to Mudd, who called on his network of colleagues to analyze methionine, betaine and a host of other chemicals in the transsulfuration pathway. This partnership allowed us to correlate the dose of betaine to its specific therapeutic actions and thereby establish a treatment protocol that we published together in 2007.

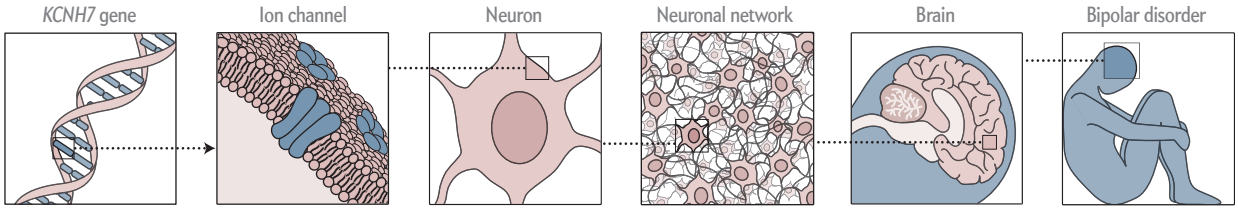
Weeks after starting betaine, Mark took his first steps and awoke to light and sound. Other patients also made quick and decisive progress, but we learned a poignant lesson about the arrow of biological time. Mark and other children who started betaine later in life were left with permanent disabilities traced to stagnant brain growth during infancy. The dense matrix of neural connections that form within this narrow window becomes an enduring substrate for our mental life. Once that window closes, the damage is done. Mark's case brought the tragedy of an entire community into sharp relief. During the three decades after Mudd's

IN BRIEF

The Clinic for Special Children in Strasburg, Pa., in collaboration with the Amish and Mennonite families it serves, has closed the gap between growing scientific knowledge of human genetics and its translation into effective medical care. **Genetic information**—gathered with high-tech, low-cost approaches—enables the nonprofit clinic to efficiently diagnose and treat dozens of potentially crippling or fatal genetic conditions. **The clinic's practice** is a model for improving medical care in underserved communities throughout the world. **A recent study** spearheaded by the clinic links a gene mutation to bipolar disorder and shows how research in isolated communities might enrich understanding and treatment of common human afflictions.

How Genetic Mutations Lead to Disease

Gene mutations can disrupt biology at multiple levels (molecules, cells, tissues and organs) to cause disease. Certain mutations are particularly prevalent in Amish and Mennonite populations. For each patient the clinic sees, it applies advanced technologies to identify the individual's genetic variants, understand their causal links to disease, and devise ways to alleviate or prevent the mutations' harmful effects. In related work, the clinic and its collaborators recently identified a gene mutation linked to bipolar disorder among the Amish, and they are now constructing a picture of how it might impair emotional regulation (*below*). This knowledge could lead to a deeper understanding of bipolar disorder in the general population and to new strategies for prevention and treatment.



Gene
A gene consists of a sequence of DNA "letters" that spell out the amino acids needed to make a protein. Proteins are the main workhorses of cells. A mutation in a gene can alter the functioning of the encoded protein. The bipolar study pinpointed a mutation in a gene called *KCNH7*.

Protein
To function properly, proteins must have the right structure, location and abundance in each cell. *KCNH7* encodes a protein that spans the cell membrane, forming a channel that regulates the flow of potassium ions. The mutant is altered at just a single amino acid, but this subtle change affects potassium movement across the membrane.

Cell
All cells contain the same genes, but many genes are expressed (that is, give rise to proteins) only in select cell types. The ion channel encoded by *KCNH7* is used by neurons throughout the brain. Potassium currents critically shape each neuron's electrical behavior, and the mutant alters the cells' firing patterns.

Tissue
Tissues can contain a mixture of cell types. Brain tissue, for instance, includes neurons and supporting cells called glia. The mutant *KCNH7* gene would be expected to disrupt the operation, not only of individual nerve cells, but of whole neuronal circuits, such as those regulating emotions and behavior.

Organ
Nerve cells throughout the brain make the ion channel encoded by the *KCNH7* gene, but the channel is most abundant in brain regions underlying emotions and cognition. Consistent with that finding, mutation of the gene has been tied to mania observed in laboratory animals.

Behavior
Bipolar disorder is marked by a spectrum of behaviors that can include depression, mania and psychosis. New insight into how the *KCNH7* mutation affects each level of biology—from misspelled protein to perturbed brain function—could lead to fresh ideas for interrupting the chain of events underlying the disorder.

publication about MTHFR deficiency, children like Mark lived and died in obscurity, shrouded in confusion and sorrow.

While working out the details of therapy, we developed a test to screen young couples for the genetic defect and were alarmed to find that 30 percent of healthy Amish from the Kinsingers' settlement carried one mutant copy of MTHFR. From this figure, we could infer that one in 50 of their babies would be born with the disease. In 2003, recognizing the critical role of preemptive therapy, we reached out to biochemist Edwin Naylor at his pioneering newborn screening lab in Pittsburgh. Together we were able to develop and implement a method for detecting the MTHFR mutation from the dried filter-paper blood spots collected on every newborn as part of mandatory state screening for various hereditary disorders.

Remarkably, the first child diagnosed by this novel filter-paper method was Mark's sister Ruth, born September 2003, just 10 months after the Kinsingers first brought Mark to our clinic. Ruth started betaine therapy during her second week of life and has flourished during 12 years of follow-up. Today she is an accomplished student, affectionate daughter and formidable stickball player.

In 2009 Mudd and his wife had the opportunity to meet the Kinsingers at a Clinic for Special Children 20th anniversary celebration. As the Mudds spoke with Ruth's parents, Ruth quietly climbed into Mudd's lap. He told me later that it was the finest moment of his scientific career.

Mudd died in January 2014 at age 86. Several weeks later his widow received a handmade card that read: "Dear Mrs. Mudd, Greetings of love are being sent your way. How are you today? I'm fine. It is a foggy morning, and looks as if it would be sunny. I am looking forward to going barefoot. Love, Ruth."

GRASSROOTS GENOMIC MEDICINE

THE UNUSUALLY HIGH INCIDENCE of MTHFR deficiency and other genetic disorders among the Plain people is rooted in their unique social and cultural history. Small bands of Anabaptists who survived the trans-Atlantic migration composed a meager gene pool. Like all of us, these individuals unknowingly harbored damaging sequence variants (more commonly called mutations) in their genetic code. In isolated populations, such gene variants can propagate silently through carriers over generations, randomly drifting up or down in prevalence, until a child inherits two copies of the damaging genetic change from parents who share a common ancestry. This recessive pattern of inheritance is an important mechanism of genetic disease in isolated communities across the world. Among modern Anabaptists, the ancestral constellation of gene variants causes much individual and communal suffering, a problem compounded by limited science

Illustration by Amanda Montañez

MENNONITE BOY on the left has maple syrup urine disease [*see box on opposite page*] and lives 23 miles from the clinic. The boy on the right has glutaric aciduria type 1. His family relocated so that he could be cared for at the clinic.

education and poor access to the U.S. health care system.

In the early 1960s the late Victor McKusick, a pioneer of modern medical genetics, first recognized the potential for studying hereditary disease patterns among the Amish and launched a comprehensive field study to this end. Though wary of technology's power to undermine social relationships, Plain people opened their homes to McKusick and his collaborators in the hope that future generations might benefit. This work culminated in the 1978 publication of *Medical Genetic Studies of the Amish,* which catalogued 18 previously recognized and 16 newly diagnosed genetic disorders among the Amish of North America. These early research efforts established many key principles about human genetic disorders but did little to help the population under study. Many Amish grew weary of doctors who were interested in investigating their patterns of disease but unable or unwilling to care for them.

A decade later a young physician by the name of D. Holmes Morton would take a different approach. In 1988, while Morton was a fellow in biochemical genetics at Children's Hospital of Philadelphia, a colleague asked him to analyze a urine sample from a six-year-old Amish boy named Danny (his real name) who had suffered an abrupt and unexplained regression of motor skills at 14 months of age. Local doctors called it cerebral palsy, but Morton, using a technique called gas chromatography/mass spectrometry, detected a substance called 3-hydroxyglutaric acid in the boy's urine. This distinctive chemical footprint implicated a rare genetic disorder called glutaric aciduria type 1 (GA1)—not cerebral palsy—as the cause of Danny's brain injury.

Morton visited Danny at his home in Lancaster County, where he learned of the many families who communicated in letters about their children with so-called Amish cerebral palsy. In 1991 he and his colleagues published a report of 10 definite cases of GA1 among the Amish, doubling the number of published cases worldwide. He listened to harrowing stories from parents who had fallen into a kind of learned helplessness; generation after generation, they watched their children struck down by a mysterious brain disease only then to be vexed by a medical system too remote, too fragmented and too expensive to help them. This wheel of anguish convinced Holmes and his wife, Caroline, of the need for a local clinic—a medical home—where uninsured Plain families could bring their special children for affordable and compassionate care.

Thus began a health care experiment fundamentally different from the profit-driven U.S. health system: a grassroots collaboration between the Mortons and a handful of committed parents who knew firsthand the pains inflicted by genetic disease. An Amish farmer who had two grandchildren with GA1 offered two and a half acres of his field to site the clinic. Other Plain community members provided timber and labor to raise its mortise-and-tenon frame. From then until now, the Plain communities have continued to support the project as a valuable investment in their children. Nearly 75 percent of the current annual operating budget of $2.6 million comes from charitable sources, including more than $850,000 raised by Plain people at benefit auctions that offer quilts, furniture, plants, ponies, barbecued chicken, handmade pretzels, whoopie pies, and more. This support limits out-of-pocket clinical and lab fees to between $50 and $150 per visit, 70 to 90 percent below the cost of comparable services at academic health centers.

The Mortons recognized from the outset that the most effective approach for treating GA1 and other disorders was to start with healthy newborns, detect genetic risks before disease onset and provide informed, local services across the arc of youth. Yet preemptive strategies are easier to conceptualize than to implement. And the details matter: an accurate genetic diagnosis is meaningless if it comes too late, and a clever molecular therapy is useless if it costs too much. The clinic is a place where science is harnessed to practical ends, empowering communities to better care for their own while shielded from medical bankruptcy.

Our ground floor is equipped with an array of advanced gene-sequencing tools. The on-site lab team, led by Puffenberg-

The Economics of Prevention

The Clinic for Special Children's progress in managing a disorder called maple syrup urine disease (MSUD) illustrates the practical economic benefits of integrating biochemical and genetic science with the everyday practice of medicine. MSUD is rare worldwide but common among Mennonite settlements of Pennsylvania, where it affects about one in 380 newborns. It is a dangerous disorder; before the clinic opened its doors in 1989, 39 percent of MSUD victims died during childhood, and most survivors were left with severe mental and physical disabilities.

Children with MSUD lack an enzyme necessary for degrading three dietary amino acids. Consequently, certain chemicals reach concentrations that poison the brain. In excess, these chemicals spill into the urine, giving it the characteristic odor of maple syrup. Affected children appear normal at birth, but within three to five days become inconsolable and then develop forceful, involuntary muscle spasms. Left untreated, accumulating toxins cause brain swelling, coma and death.

Before the clinic's inception, health services for children who had rare and complex genetic disorders were woefully inadequate in rural communities. Those with conditions such as MSUD had medical care that was fragmented, costly and ineffective. During each medical crisis, families were forced to travel 100 miles or more to reach the nearest academic medical center, where they remained for weeks and paid cash rates of $50,000 to $100,000 for emergency services. This reactionary cycle encumbered the Mennonite people with medical debt but did not alleviate the burden of disease.

Since 1989 our clinic has managed 80 Mennonite MSUD patients from the time they were newborns. Half were diagnosed on-site between 12 and 24 hours of life and transitioned safely at home. The remainder were diagnosed by mandatory state newborn screening and discharged safely to home after an average five-day hospital stay. Over 25 years we have made incremental improvements in the monitoring and treatment of MSUD that include inexpensive filter paper–based chemical testing from home, sophisticated intravenous nutritional mixtures used to reduce toxin levels, and new dietary formulas designed to optimize the chemical environment of the brain. These innovations have decreased hospitalizations from 7.0 to 0.1 days per patient per year. A 98 percent decrease in hospital costs applied to all the MSUD patients under our care saves the community at least $4.8 million annually—nearly twice the clinic's operating budget.

Advanced technologies have the reputation of being prohibitively expensive, but the cost depends in large part on how they are deployed. Investing resources in preemptive diagnosis and disease prevention can be instrumental in reducing unnecessary and wasteful medical spending. —*K.A.S.*

er, has worked closely with clinic physicians to discover between five and 15 population-specific damaging gene variants each year since 1998. Focused molecular strategies allow the team to accurately diagnose most genetic disorders within fewer than 24 hours for a cost of $50. Precise genetic knowledge allows us to look into the future, understand when and how disease is likely to unfold, and take actions to keep children safe.

In the instance of GA1, Morton worked closely with Naylor to implement elective statewide newborn screening in 1994. A few years later Stephen I. Goodman of the University of Colorado School of Medicine deciphered the specific genetic change underlying "Amish" GA1, which enabled Puffenberger to perform rapid, inexpensive molecular testing. By identifying affected children before disease onset and intensifying their medical care, we were able to reduce the risk of disability from 94 to 36 percent, but we still agonized each time an affected infant succumbed to brain injury.

Then, in 2006, I collaborated with Richard Finkel, a founder of the nutrition-supplement company Applied Nutrition, to design what we call a "medical formula"—a prescription diet—for infants and children with GA1. We knew that the mutation responsible for GA1 caused glutaric acid and other toxins—produced from the amino acid lysine—to accumulate in the brain and that the presence of a different amino acid, arginine, could limit lysine's entry into the brain. By judiciously manipulating the relative dietary proportions of these two amino acids (with the help of computer modeling), we thought we could reduce cerebral lysine uptake and thereby limit neurotoxin production by the brain.

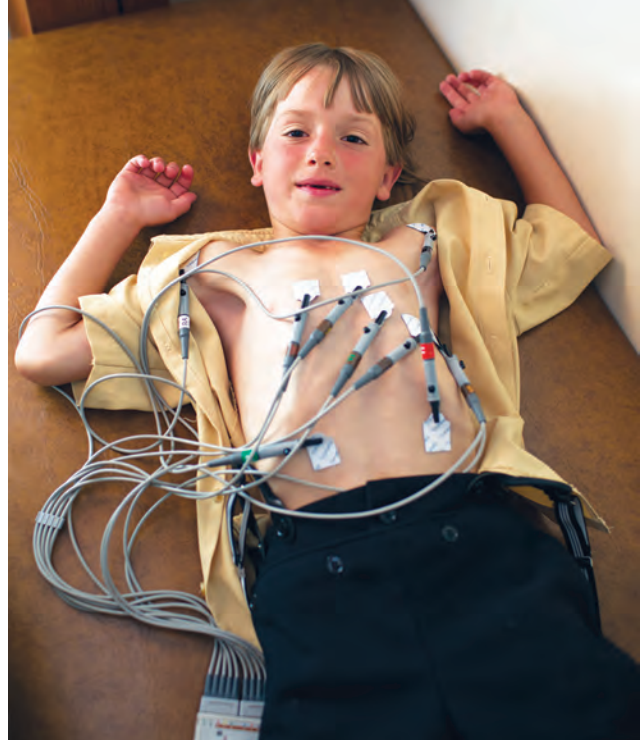
I tested the new approach on 12 affected infants in a clinical trial conducted between 2006 and 2011. Treated infants had half the toxin excretion, a third of the hospitalizations and complete protection from brain injury. We published our findings in 2011 and to date have treated a total of 25 consecutive newborns with the customized medical formula. The results have been durable; the brain injury rate remains less than 5 percent, and nearly all children born today with GA1 grow up healthy. For many other genetic disorders we treat, similar stepwise innovations in diagnosis and treatment have enabled us to reduce rates of disability, hospitalization and death by between 50 and 95 percent—a powerful testament to the idea that science guided by conscience can do much to prevent human misery.

MANY POPULATIONS, ONE BIOLOGY

THE STUDY OF RARE GENETIC DISORDERS has a special role to play in the growth of biological science. Only by carefully observing the medical consequences of a gene mutation can we fully appreciate how the normal gene contributes to human biology. William Harvey foresaw this in 1657, when he suggested that investigation of rare disorders is the best way to reveal nature's "secret mysteries" and thereby advance mainstream medical practice. Three and a half centuries later we understand Harvey's axiom in modern terms. By attending closely to the dynamic interplay between a rare gene variant and mental health, we recently gained key insight into one of the most common human afflictions.

It was a crisp autumn morning when I first met Katie, a woman of about 40 who had agreed to participate in our research study of bipolar disorder among the Pennsylvania Amish. She preferred we meet in the barn where her husband, David, repaired small engines. Machine parts were strewn about carelessly in a manner unexpected for an Amish shop. Most days David did the work of two—Katie's bipolar disease had dominated their shared life for more than a decade, and David often struggled in isolation to raise their five children.

Bipolar disorder first exacted its toll on Katie after the birth



CHILDREN treated at the clinic have many serious health issues, such as heart arrhythmias (*top panels*), brain malformations (*bottom left*), and hereditary attention deficit disorder (*all three, bottom right*).

of her second child. She began to speak fast, sometimes very fast, and often followed random trains of thought into obliquity. At intervals she stayed up nights on end, cleaning and recleaning the house. "These floors are filthy. Filthy. Filthy." During the dark periods that followed, Katie lay in bed ruminating, hopeless and racked with guilt. Familiar voices—her husband, children, and parents—incessantly whispered to her from behind: "You're worthless." But her biggest concern, conveyed repeatedly when we first met, was a mass that filled her abdomen and tormented her relentlessly, a chronic perceptual hallucination that she called her "miserableness."

Mental disorders—including bipolar disorder—are common worldwide, affecting 12 to 47 percent of different populations. In the U.S., psychiatric disease accounts for 40 percent of medical disability among young adults, and suicides outnumber homicides two to one. Isolated groups such as the Amish provide distinct advantages for investigating the heritability of psychiatric illness and other common medical conditions. One such effort, the Amish Study of Major Affective Disorders, began in 1976 and tracks several large, multigeneration Amish pedigrees with a high prevalence of bipolar disorder. Over three decades the cohort swelled to include more than 400 subjects and remains one of the most intensively studied in the history of medical genetics.

On October 31, 2011, Puffenberger and I attended a Family Meeting hosted by Alan Shuldiner and the University of Maryland's Amish Research Clinic. Leading psychiatric investigators addressed a gathering of Plain people concerned about mental illness within their families and communities. As the meeting drew to a close, researchers summarized 35 years of Amish bipolar research with a dispiriting message: "There is not a lot new to tell you." On the way to the parking lot, I was stopped by three Amish sisters whose family had participated in familial bipolar research for more than two decades. Nine of 11 siblings from their generation had spent much of adult life debilitated

SCIENTIFIC AMERICAN ONLINE Read more about the clinic's work at ScientificAmerican.com/dec2015/clinic

by mania or depression. They wondered if our clinic, which had a reputation for taking on intractable problems, might help them better understand if "some gene was involved."

The timing was right. We had recently begun collaborating with the Broad Institute in Cambridge, Mass., to explore the utility of whole exome sequencing for investigating rare genetic disorders in children. The exome consists of all coding letters, or nucleotides, that are "read" to construct the body's 19,000 proteins.

Although the exome represents only about 1 percent of the human genome, it contains the vast majority of genetic changes that can cause disease, and whole exome sequencing is at present the most efficient and lowest-cost method for disease-gene discovery.

Although our clinic has historically focused on pediatric health, psychiatric disorders pervade every aspect of family and community life, and our collaborators in Cambridge allowed us to reserve seven exome samples for Amish adults with bipolar disorder. Remarkably, all seven people shared an exceedingly rare variation in a gene that encodes the KCNH7 protein. This single-letter substitution, called a missense change, alters the structure of KCNH7 at an amino acid conserved across the evolution of many different animal species; changes in such conserved regions often critically alter the way a protein functions.

Over the next two years, Sander Markx and Michael First of Columbia University's department of psychiatry helped us expand the study to more individuals and implement a method to rigorously categorize their symptoms. Ultimately we were privileged to collaborate with investigators at Weill Cornell Medical College, the University of Pennsylvania, Franklin & Marshall College and the McKusick-Nathans Institute of Genetic Medicine. This team approach also allowed us to track the movement of KCNH7 protein in cells, demonstrate how its mutant form alters electrical firing in neurons and establish a statistical foundation for our discovery. For the first time, we identified a specific genetic change that signals a strong predisposition to bipolar disease among the Amish. We published our findings in 2014; they now allow investigators worldwide to explore the connection between KCNH7 and mental illness in other populations.

KCNH7 is especially abundant in brain regions that affect mood and cognition, where it forms channels that mediate potassium movement across cell membranes. These ephemeral waves of ions, moving in and out of membranes too thin to see, are directly linked to what we think and feel. Our everyday experience belies this fact; it is difficult to imagine electrochemical signals at the root of violence, addiction, psychosis and suicide. But our research suggests that a subtle change in the threshold and timing of ion currents can cast a person into repeated cycles of madness and despair.

To realize at the genetic level that the mind is embodied in this way allows us to understand mental suffering in concrete terms. Discovery of the KCNH7 variant is important because it provides a foothold for rational discussions among scientists and patients and helps to strip away the layers of guilt and shame that surround mental illness. In the near term, knowledge that connects genetic variation to bipolar disease can lead to more timely and effective medical care for people such as Katie. On a longer time scale, it might be possible to design drugs that modulate the KCNH7 ion channel—a form of precision medicine that could open up a whole new class of therapeutics for the treatment of bipolar disorder in all populations.

TIME AND OPPORTUNITY

THE STUDY OF BIPOLAR GENETICS in the Amish is a parable about the future of medicine—about how genetic information might be used to predict *your* future. At the clinic, we now have a simple and inexpensive blood test that can be collected from umbilical cord blood at birth to inform us about a child's risk for bipolar disorder 30 years hence. Because adult-onset psychiatric disorders are often preceded by erratic thought and behavior in youth, early detection of a genetic risk factor could enable more timely and effective mental health care across a lifetime. But should we begin screening Amish newborns for the damaging KCNH7 variant?

Such questions are accruing quickly and pertain to all of us. Peer into your exome, and you will find between 20,000 and 40,000 deviations from the so-called normal human sequence of nucleotide letters. Twenty percent of the variants in your DNA code have the potential to alter protein function, and about 1,000 are exceedingly rare, possibly even unique to you. How many of these changes can predict your future? And if so, what can or should be done about it? The answer depends, in part, on what knowledge we deem actionable for any particular person at a specific time. That is perhaps one key to our clinic's success: cumulative population knowledge—painstakingly acquired over the past 25 years—works like a Rosetta Stone. It allows us to decipher the meaning of genomic data within a specific social context and thereby tailor medical care to the individual: the right treatment for the right person at the right time.

In all populations, this kind of deep knowledge about gene action will allow scientists to visualize cellular machinery in exquisite detail and understand how the various molecular parts interact in health and disease. But it is people—not their component parts—who suffer. Clinicians and molecular biologists who work side by side at an appropriate scale, one patient at a time, can weave genomics into the physician's craft, yielding strategies that are preemptive rather than reactive.

Pediatric practice is a good place to put this idea to the test. At our clinic, knowledge and treatment grow in lockstep as we explore the complex interactions between gene variation and environment that play out over the formative stage of life. Caring for children challenges us to leverage the predictive power of genetic knowledge, focusing on outcomes that matter most. One child at a time, we continue to close the gap between genomic research and the day-to-day practice of medicine, which at our clinic is a practical endeavor, driven by what this child needs, right now. SA

MORE TO EXPLORE

Prevention of Brain Disease from Severe 5,10-methylenetetrahydrofolate Reductase Deficiency. Kevin A. Strauss et al. in *Molecular Genetics and Metabolism*, Vol. 91, No. 2, pages 165–175; June 2007.

Genetics, Medicine, and the Plain People. Kevin A. Strauss and Erik G. Puffenberger in *Annual Review of Genomics and Human Genetics*, Vol. 10, pages 513–536; September 2009.

One Community's Effort to Control Genetic Disease. Kevin A. Strauss, Erik G. Puffenberger and D. Holmes Morton in *American Journal of Public Health*, Vol. 102, No. 7, pages 1300–1306; July 2012.

A Population-Based Study of KCNH7 p.Arg394His and Bipolar Spectrum Disorder. Kevin A. Strauss et al. in *Human Molecular Genetics*, Vol. 23, No. 23, pages 6395–6406; December 1, 2014.

FROM OUR ARCHIVES

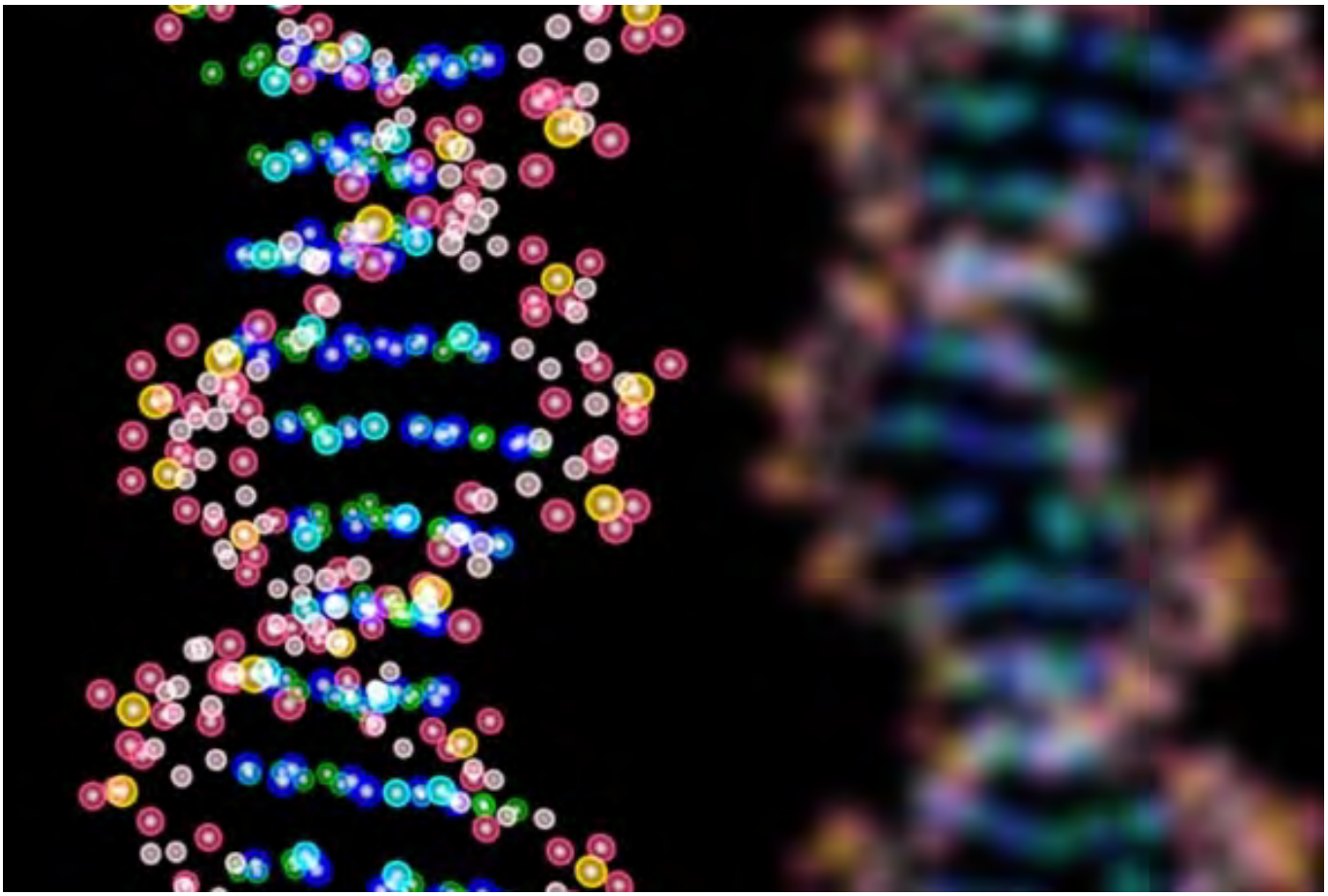
When DNA Means "Do Not Ask." Dina Fine Maron; Science of Health, January 2015.

HEALTH

Genomics Can Improve Health Care--Right Now

A unique clinic in Strasburg, Pa., shows the way

By Kevin A. Strauss on December 1, 2015



Shunyu Fan ©iStock.com

Shutterstock.com

A tale of three sisters offers a glimpse into how, with the right systems in place, genomics research can already be applied to alleviate human suffering.

In January 2007 Jesse and Anna, members of an Old Order Amish settlement in Pennsylvania, brought their daughter Esther to the Clinic for Special Children in Strasburg, Pa. (For privacy, pseudonyms are used for all children mentioned in this article.) Esther, just a few hours old, was born with inflamed skin, patchy hair loss and a swollen liver—telltale signs of Omenn syndrome, a rare and lethal form of severe combined immune deficiency. Our clinic is a shelter for children like Esther, who enter the world facing dangerous genetic risks.

Esther's sister Mary was born 14 years earlier, in 1993, and lived out a short life of misery marked by infections, needles, ventilators and feeding tubes. She died of pneumonia before a specific diagnosis could be confirmed and left behind a shattered family with more than $400,000 of hospital bills.

Esther's life would be different. From her first hours, she had a medical home where her care was informed by a deep but pragmatic understanding of biology, culture and the places where they meet. Using advanced molecular techniques, our laboratory director, Erik Puffenberger, rapidly homed in on Esther's genetic diagnosis (*recombination-activating gene 1, RAG1*) and proved that her sister Mary succumbed to the same condition. Using these same molecular data, Puffenberger identified an ideal bone marrow transplant donor among Esther's siblings that enabled her to have a lifesaving transplant at 65 days of age. The entire process—from clinical presentation to genetic diagnosis to donor identification—took less than two weeks, saved the family about $80,000 and cleared the path to a cure.

Severe combined immune deficiency (SCID), first recognized in 1950, encompasses a variety of genetic disorders that render the immune system powerless. Without radical

variety of genetic disorders that render the immune system powerless. Without radical treatment in the form of bone marrow–cell transplant, children with SCID inevitably die from infections by two years of age. In 1972 a paper in *The Lancet* described the first specific protein deficiency linked to SCID, an enzyme called adenosine deaminase. In the years that followed immunogenetics research advanced at an explosive pace. Doctors today know of 18 different molecular causes of SCID, four of which are found in high prevalence among the Amish and Mennonites of North America (collectively called "Plain" people). And yet as this knowledge unfolded, 26 of 41 (63 percent) Plain children born with SCID died by two years of age because of late diagnosis, geographic and cultural isolation, and unaffordable access to care. It was not the research community but anguished parents—people like Jesse and Anna—who brought this problem to our front door.

Esther is now eight years old and healthy. In May 2013 her baby sister Annie arrived. Using blood from the umbilical cord, we diagnosed Annie with Omenn syndrome at four hours of age for a cost of less than $50. Annie's father peered anxiously over the shoulders of our laboratory team as they inspected molecular markers to identify a bone marrow donor for this daughter. Annie was successfully transplanted in the first month of life and has thrived ever since.

This tale of three sisters, rising from a community's tragic history, is a paradigm of how genomics can shape the everyday delivery of medical care. These three Amish sisters give us a peek into the future of genomics, which has the power to make the healing arts preemptive rather than reactive. Progress along this track must be measured one patient at a time, and in terms people understand. For this family, help did not come in the form of new knowledge but from choices made about how to use what we know. Applying molecular tools to the task was not so much an innovative as a *sensible* thing to do.

Kevin A. Strauss, who earned his MD degree from Harvard Medical School, is the medical director of the Clinic for Special Children, in Strasburg, Pa.