*Etiology of Human Disease at the DNA Level,*
edited by Jan Lindsten and Ulf Pettersson.


# 1

# Linkage Genetics in Humans: Origins and Prospects

David Botstein

*Department of Genetics, Stanford University School of Medicine, Stanford, California 94305*

The rediscovery of Mendel in the first years of this century was immediately followed by the recognition that some human traits, indeed some human diseases, are inherited according to Mendel's laws. It soon became unmistakably clear that the mechanisms of heredity in humans are entirely typical of those in all higher eukaryotes. Among the earliest post-Mendelian discoveries was the principle of genetic linkage and the idea of a linkage map (Sturtevant, 1913). Yet linkage mapping in humans was not practiced on a large scale until after 1980, even though the applicability of the principle and, indeed, a number of essential statistical methods for detecting linkage had been in place since the 1930s (see Ott, 1985, for a review). It is the purpose of this short chapter to sketch briefly how linkage mapping was finally applied to inherited human diseases, to place the ideas behind human genetic linkage mapping in their proper historical context, and to give some inkling of the future and limits to linkage mapping in humans.

## LINKAGE MAPPING WITH DNA POLYMORPHISMS

The limiting factor that made linkage mapping difficult between the 1930s and the 1980s, even for diseases obviously inherited in a simple Mendelian way, was the supply of adequately polymorphic genes that could serve as markers. Figure 1 illustrates that Mendelian inheritance and, indeed genetic linkage, can be observed readily, given only loci (marked A, B and C) such that the four alleles at each locus in the parents (A1, A2 in the father and A3, A4 in the mother; B1, B2 in the father and B3, B4 in the mother; and so on) can all be distinguished. As shown in Fig. 1, when all the alleles can be distinguished, it is easy to see segregation of the parental alleles (each child gets either A1 or A2 from the father and A3 or A4 from the mother, etc.).

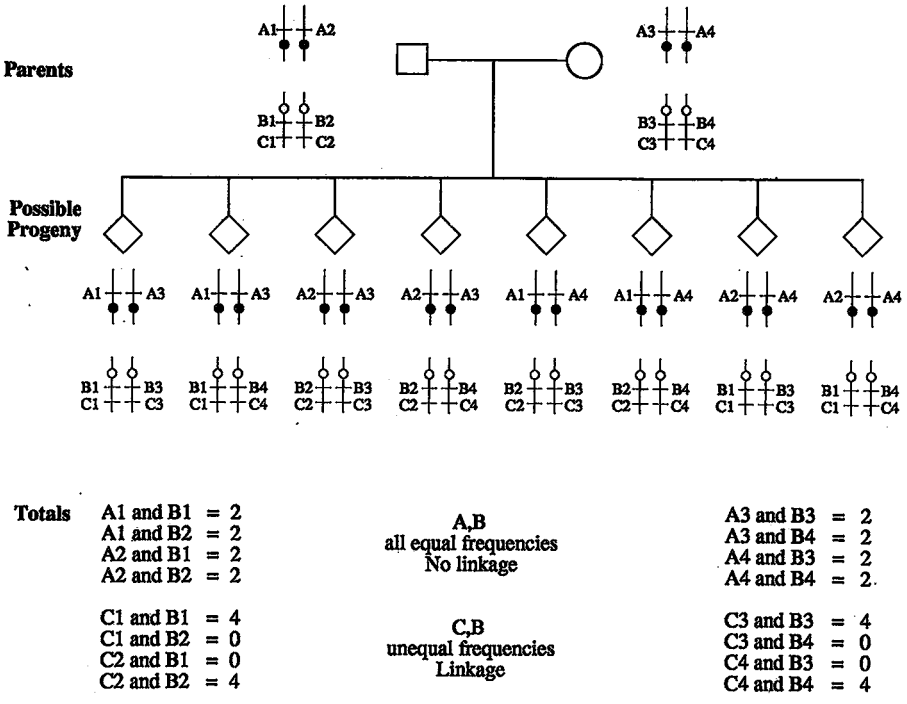


**FIG. 1.** Mendelian inheritance, and genetic linkage, can be observed given loci (A, B, C) such that the four alleles at each locus in the parents can be observed (A1, A2 in the father and A3, A4 in the mother, etc.). See text for details.

Independence of inheritance of alleles residing on different chromosomes is equally easily seen (e.g., A1 from the father is accompanied by B2 as frequently as B1; the combinations A3,B3; B3,B4; A4,B3; A4,B4 are equally frequently inherited from the mother; etc.). Linkage is easily observed as well, as the inheritance of C1 is highly correlated with the inheritance of B1; C2 is correlated with B2 (both from the father); on the mother's side we see correlation of C3 with B3 and C4 with B4 in Fig. 1. Note also that we see, from these correlations, not only that loci B and C are linked, but also the so-called linkage phase (i.e., the fact that C3 and B3 lie on the same chromosome in the mother).

Recombinant DNA technology provided a source of polymorphic markers in the form of restriction fragment length polymorphisms (RFLPs). In 1980 my colleagues and I (Botstein et al., 1980) noted that if there is enough variation among the DNA sequences of humans, differences in pattern of digestion by sequence-specific endonucleases (the *restriction enzymes*) would be found from individual to individual. These differences in restriction fragment length could be used as codominant genetic markers, just as we used the

hypothetical alleles A1, etc. in Fig. 1. The method of detection proposed in 1980 (and still the most common) is to use single-copy DNA probes derived from human genomic clones as hybridization probes in gel-transfer experiments by the method of Southern (1985). Genomic DNA is extracted from white blood cells, cut with restriction enzymes, separated by size by gel electrophoresis, "blotted" onto filter paper, hybridized with labeled (radioactive or fluorescent) DNA probe, and analyzed in comparison with other samples from the same family. It should be emphasized that the DNA probe used to elicit the RFLP has a dual nature: it is a *genetic marker* that can be placed on a *genetic map* by *linkage* via polymorphism it reveals; it is also a *physical marker* that can be placed on a *physical map* because it is a single-copy DNA sequence. RFLPs thus bind together the genetic and physical maps of the human genome.

Of course, real RFLP markers are never completely polymorphic, and thus not every mating is "informative," in the sense of allowing the distinction between parental alleles. For a given locus, the inheritance of the marker could be followed in only a fraction of the families under study; in the remainder the marker is homozygous in key individuals, and thus yields no information. For this reason, we emphasized the importance of the "informativeness" of markers for mapping, proposing a measure of usefulness we called the *polymorphism information content* (PIC). The problem of informativeness of markers made us also propose that a map consisting mainly of polymorphic markers be constructed, so that even if a marker near a particular region turned out to be uninformative, the next marker on the map could be used in its stead, albeit with some loss of resolution. We proposed in 1980 that standard likelihood measures ("LOD scores") would serve admirably for disease mapping as well as for the construction of RFLP linkage maps.

## DISEASE MAPPING WITH RFLP MARKERS

In 1983, the first autosomal disease gene was mapped using RFLPs: Gusella et al. (1983) found that the gene causing Huntington's disease is linked to a RFLP marker located near the end of the short arm of chromosome IV. Since then, a large number of other disease genes, including notably the gene for the recessive disease cystic fibrosis, have been mapped (Table 1). Because the RFLP markers are useful in physical as well as genetic mapping, they are crucial in the isolation of actual DNA sequences corresponding to the disease gene; the particular recent success in isolating the gene that causes cystic fibrosis was an example of this use of RFLP markers (Rommens et al., 1989; Kerem et al., 1989).

Despite the success in mapping some of the major inherited disease genes by linkage mapping using individual markers, theoretical analysis shows that

**TABLE 1.** *Diseases mapped using RLFPs*

| Disease | Chromosome | Reference |
|---|---|---|
| Duchenne's muscular dystrophy | X | Davies et al., 1983 |
| Huntington's disease | 4 | Gusella et al., 1983 |
| Retinoblastoma | 13 | Cavenee et al., 1983 |
| Cystic fibrosis | 7 | Tsui et al., 1985; Knowlton et al., 1985; White et al., 1985 |
| Adult polycystic kidney disease | 16 | Reeders et al., 1985 |
| Familial colon cancer | 5 | Bodmer et al., 1987 |
| von Recklinghausen's neurofibromatosis | 17 | Barker et al., 1987; Seizinger et al., 1987 |
| Bilateral acoustic neurofibromatosis | 22 | Rouleau et al., 1987 |
| Multiple endocrine neoplasia type 2A | 10 | Simpson et al., 1987 |

use of a set of mapped markers whose linkage relationships are known and that span all of the genome will be more efficient and powerful (Lander and Botstein, 1986a). In large part, this is because of the increased power of linkage tests with markers flanking a disease gene as compared with single markers lying to one side. With single markers, single crossover events can completely reverse the relationship between the marker and the disease gene, but with flanking markers, only a double-crossover (an exceedingly rare event) will suffice to switch the marker-disease gene relationship completely.

Several techniques that take advantage of a complete linkage map have been suggested that allow either disease gene mapping without family study (*homozygosity mapping*; Lander and Botstein, 1987), mapping of diseases showing heterogeneity of cause (*simultaneous search*; Lander and Botstein, 1986b) or even mapping of genes contributing to quantitative traits (Lander and Botstein, 1988). This last method, which is applicable only to model systems, has been used to map several traits specified by as many as five contributing quantitative trait loci in tomato (Paterson et al., 1988).

In 1987, the first reasonably complete genetic linkage map of the human genome was published by Donis-Keller et al. This map has been preceded by maps of individual chromosomes and chromosome arms, notably the X chromosome (Drayna and White, 1985). The Donis-Keller et al. map covers, in the sense of showing continuous linkage, about 95% of the genome. However, not all the markers in it are highly informative and routine use as suggested above is not quite yet a reality. Improvement of the map, both with respect to informativeness of markers and density of markers, is a short-term goal of the Human Genome Initiative. As the maps become better,

with more informative markers spaced at even intervals, the techniques involving the entire map can be applied to diseases with complex etiologies (using simultaneous search) and very rare recessive diseases (homozygosity mapping).

## INTELLECTUAL ORIGINS OF THE RFLP MAPPING STRATEGY

As indicated above, the idea of a linkage map based on frequency of recombination is very old, having been published by Sturtevant in his historic paper of 1913. In the same paper, Sturtevant put forward the principle that double-crossovers will be rare, and he used this principle to order markers. Thus there is nothing in the idea of linkage mapping per se that is new in RFLP mapping. Quite to the contrary, it is very much in the tradition of genetic mapping as understood by Sturtevant and generations of geneticists who followed him.

Physical markers that can be scored in genetic crosses, a central attribute of RFLP markers, are also not new. The first use of a physical marker in a genetic cross was by Creighton and McClintock in 1931; this paper was the first to correlate cytological crossing over and genetic recombination. Correlations between genetic and physical (usually cytogenetic) markers became a major preoccupation of geneticists of all kinds since then.

Molecular polymorphisms were also old antecedents of the RFLP strategy. The first such marker in humans was hemoglobin-S, the protein with an altered beta-chain that is the cause of sickle-cell anemia (Neel, 1949; Pauling et al., 1949). Many protein polymorphisms were subsequently discovered in humans and it was these that were the main markers in pre-RFLP linkage studies.

The first correlations between physical and genetic markers at the DNA level were carried out with bacteriophage λ (Davis and Parkinson, 1971; Parkinson and Davis, 1971). These papers describe the generation and physical mapping of deletion and substitution mutations by formation of DNA heteroduplexes followed by measurements of the molecules under the electron microscope. Some of these very deletions and substitutions led to, and even are still present in, today's recombinant DNA vectors.

The same year saw the publication of the first map of a genome with restriction enzymes (Danna and Nathans, 1971). It was quickly understood, since this was very much in the tradition of bacteriophage and animal virus genetics, that maps of the recognition sites of these DNA-sequence-specific endonucleases constituted an important new tool for physical mapping with a technical simplicity and theoretical resolving power quite beyond that of DNA electron microscopy. The method was limited, however, to small genomes in which all the fragments produced after a digest could be separated from each other on a gel.

The first use of a difference in restriction fragment recognition sites as a passive genetic marker in genetic crosses was done by Sambrook's group at Cold Spring Harbor (Grodzicker et al., 1974). Using simple restriction mapping as introduced by Danna and Nathans (1971), they located temperature-sensitive (*ts*) mutations on the adenovirus genome by crosses between strains (Ad2 and Ad5) that differed in their restriction maps. By making crosses between *ts* strains that also differed in restriction pattern, they could select temperature-independence and then score the presence or absence of particular restriction enzyme recognition sites. The design of this experiment goes back to Creighton and McClintock (1931), i.e., the first correlation between genetic and physical maps.

The method for extracting, by hybridization, restriction enzyme cleavage pattern information from complex genomes with thousands of fragments was invented by Southern (1975). His method, which involves transfer ("blotting") of DNA from the medium of size separation of fragments (usually agarose gel) to filter paper on which hybridization is performed, was one of the central elements of the revolutionary methods now lumped under the rubric *recombinant DNA technology*. It was first applied to the case of a single-copy DNA sequence in a mammalian genome by Botchan et al. in 1976. They followed the integration of SV40 into cellular DNA by gel-transfer hybridization using viral DNA as probe, and found many obviously single-copy integrants. This paper made clear the possibility of following single genes by gel-transfer, and was the basis for our expectation that polymorphism in restriction fragment length would be routinely detectable, since the different integration sites of SV40 were readily detected.

Having surveyed the origins of the elements of RFLP analysis, namely, linkage mapping with physical markers, restriction fragment length differences as molecular markers, and gel-transfer to visualize the restriction fragment length difference, we come to the first use of RFLPs as genetic markers in complex genomes. Two groups discovered and applied the technique to genes of yeast (*Saccharomyces cerevisiae*) independently of each other; both applications were published in 1977. Petes and Botstein (1977) found a polymorphism in the restriction pattern of the ribosomal DNA of yeast in a diploid strain that was heterozygous with respect to this property. By sporulating that strain and performing tetrad analysis, it was possible to show that all the 100-rDNA copies comprise a single tandem array at a single locus (subsequently mapped to chromosome XII). Olson et al. (1977) sought ways to distinguish the eight different genes in yeast that encode tyrosine tRNAs that can mutate to become ochre suppressors. Taking advantage of the different restriction fragment lengths of the different genes, and performing crosses between strains that were polymorphic at one or more of the loci of these genes, Olson et al. were able to map these genes to 8 loci on six different chromosomes.

Application to human genetics began with Kan and Dozy (1978), who used a restriction fragment length difference revealed by probing with the β-globin gene itself to carry out antenatal diagnosis of sickle-cell anemia. The principle of linkage was not directly invoked, and no linkage mapping was proposed. This was followed by the proposal of Botstein et al. (1980) to find RFLPs deliberately, construct a linkage map, and use it to find disease genes by linkage. As mentioned above, the first such mapping of a gene by RFLP analysis was Huntington's disease (Gusella et al., 1983).

Thus we see that RFLP linkage mapping grew naturally out of classical and molecular genetics. Its intellectual antecedents are the same basic papers that are the antecedents of much of genetic analysis, and much of what we call recombinant DNA technology.

In this context, it must be remarked that the commonly used concept of "reverse genetics," as used by some in the human genetics community, makes no historical sense. "Reverse" genetics is meant by human geneticists to mean finding of a gene by its effect (i.e., phenotype) followed by its mapping (by linkage) and only subsequently by molecular isolation of the gene's DNA or protein product. The paradigm case is cystic fibrosis. Yet Sturtevant mapped genes by their phenotypic effects, and the DNA corresponding to one of his genes (white eyes) was only recently isolated. Surely we do not want to say Sturtevant practiced "reverse" genetics! He was, after all, nothing less than the *inventor* of linkage mapping!

If the above is not enough reason to abandon the term "reverse genetics," consider the problem posed by Charles Weissman, who in 1978 (2 years before the publication of the RFLP mapping idea; Weissman, 1978) proposed "reversed genetics" as a term for an approach in which " . . . a mutation is first generated in a predetermined area of the genome by site-directed mutagenesis and the effect of the lesion is then studied either in vivo or in vitro" (Weissman et al., 1979). This use of the term "reversed genetics" at least makes historical sense (Sturtevant is now facing forward again). Nevertheless, in view of the confusion, and lest we never know whether we are coming or going, in forward or reverse direction, let us agree to abandon entirely the idea of "reverse" genetics.

In conclusion, the idea of mapping genes using polymorphic DNA markers has allowed the mapping of many disease genes and promises to allow the mapping of even more. The idea grew naturally out of the history of genetics and molecular biology. As the human DNA marker map becomes better, in the sense of more polymorphic markers at shorter intervals that are easy to use, we can envision finding genes that contribute to the inheritance of more complex diseases than the simple Mendelian. We can also look forward to the use of RFLP or other DNA markers in aligning the physical and genetic maps of the human genome. This may turn out to be of major importance to the Human Genome Initiatives that are underway in the

United States and around the world. The DNA markers have, in reality, made it possible to have a real genetics of humans, in the sense of Sturtevant and McClintock, in the sense of *Drosophila* and yeast—a genetics based on linkage and phenotype as well as molecules.

## REFERENCES


Barker D, Wright E, Nguyen K, et al. Gene for von Recklinghausen neurofibromatosis is in the pericentromeric region of chromosome 17. *Science* 1987;236:1100–1102.

Bodmer WF, Bailey CJ, Bodmer J, et al. Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* 1987;328:614–616.

Botchan M, Topp W, and Sambrook J. The arrangement of simian virus 40 sequences in the DNA of transformed cells. *Cell* 1976;9:269.

Botstein D, White RL, Skolnick M, and Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980;32:314–331.

Cavenee WR, Dryja TP, Phillips RA, et al. Expression of recessive alleles by chromosome mechanisms in retinoblastoma. *Nature* 1983;305:779.

Creighton HB and McClintock B. A correlation of cytological and genetical crossing-over in *Zea mays*. *Proc Natl Acad Sci USA* 1931;17:492–497.

Danna KJ and Nathans D. Specific cleavage of SV40 DNA by restriction endonucleases of *Hemophilus influenzae*. *Proc Natl Acad Sci USA* 1971;68:2913.

Davis RW and Parkinson JS. Deletion mutants of bacteriophage lambda III. The structure of att-p. *J Mol Biol* 1971;56:403.

Davies KE, Pearson PL, Harper PS, et al. Linkage analysis of two cloned DNA sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. *Nucleic Acids Res* 1983;11:2303.

Donis-Keller H, Green P, Helms C, et al. (1987) A genetic linkage map of the human genome. *Cell* 51:319–337.

Drayna D and White R. The genetic linkage map of the human X chromosome. *Science* 1985;230:753–758.

Grodzicker T, Williams J, Sharp P, and Sambrook J. Physical mapping of temperature-sensitive mutations of adenoviruses. *Cold Spring Harbor Symp Quant Biol* 1974;39:439.

Gusella JF, Wexler NS, Conneally PM, et al. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983;306:234.

Kan YW and Dozy AM. Antenatal diagnosis of sickle-cell anemia by DNA analysis of amniotic-fluid cells. *Lancet* 1978;2:910.

Kerem B, Rommens JM, Buchanan JA, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;245:1073–1080.

Knowlton RG, Cohen-Haguenauer O, Van Cong N, et al. A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* 1985;318:380.

Lander E and Botstein D. Mapping complex genetic traits in humans: new methods using a complete RFLP linkage map. *Cold Spring Harbor Symp Quant Biol* 1986a;51:49–62.

Lander ES and Botstein D. Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. *Proc. Natl. Acad. Sci. USA* 1986b;83:7353–7357.

Lander ES and Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 1987;236:1567–1570.

Lander ES, and Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 1988;121:185–194.

Morgan TH. Sex limited inheritance in *Drosophila*. *Science* 1910;32:120.

Neel JV. The inheritance of sickle cell anemia. *Science* 1949;110:56–66.

Olson MV, Montgomery DL, Hopper AK, et al. Molecular characteristics of the tyrosine tRNA genes of yeast. *Nature* 1977;267:639.

Ott J. *Analysis of Human Genetic Linkage*. Johns Hopkins University Press, Baltimore, 1985.

Parkinson JS and Davis RW. A physical map of the left arm of the lambda chromosome. *J Mol Biol* 1971;56:425.

Paterson AH, Lander ES, Hewitt JD, et al. Resolution of quantitative traits into Mendelian factors by using a complete RFLP linkage map. *Nature* 1988;335:721.

Pauling L, Itano HA, Singer SJ, and Wells IC. Sickle cell anemia, a molecular disease. *Science* 1949;110:534–548.

Petes TD and Botstein D. Simple Mendelian inheritance of the reiterated ribosomal DNA of yeast. *Proc. Natl. Acad. Sci. USA* 1977;74:5091–5095.

Reeders ST, Breuning MH, Davies KE, et al. A highly polymorphic DNA marker linked to adult polycystic kidney diseases on chromosome 16. *Nature* 1985;317:542.

Rommens JM, Iannuzzi MC, Kerem B, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 1989;245:1059–1065.

Rouleau GA, Wertelecki W, Haines JL, et al. Genetic linkage of bilateral acoustic neurofibromatosis to a DNA marker on chromosome 22. *Nature* 1987;329:246–248.

Seizinger BR, Rouleau GA, Ozelius LJ, et al. Genetic linkage of von Recklinghausen neurofibromatosis to the nerve growth factor receptor gene. *Cell* 1987;49:589–594.

Simpson NE, Kidd KK, Goodfellow PJ, et al. Assignment of multiple endocrine neoplasia type 2A to chromosome 10 by linkage. *Nature* 1987;328:528–530.

Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975;98:503–517.

Sturtevant AH. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool* 1913;14:43–49.

Tsui LC, Buchwald M, Barker D, et al. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 1985;230:1054.

Weissmann C. Reversed genetics. *Trends Biochem. Sci.* May 1978:109–111.

Weissman C, Nagata S, Taniguchi T, Weber H, and Meyer F. The use of site directed mutagenesis in reversed genetics. In: *Genetic Engineering: Principles and Methods*. Volume 1. Eds. Jane K. Setlow and Alexander Hollaender. Plenum Press, New York, pp. 133–150, 1979.

White R, Woodward S, Leppert M, et al. A closely linked genetic marker for cystic fibrosis. *Nature* 1985;318:382.