Am. J. Hum. Genet. 47:887–891, 1990

1989 Allen Award Address: The American Society of Human Genetics Annual Meeting, Baltimore

David Botstein

Department of Genetics, Stanford University School of Medicine, Stanford

You have done Ray White and me a great honor in selecting us for the Allen Award this year, and I want first of all to express my appreciation of it. Even though I received my Ph.D. from the Human Genetics Department at the University of Michigan, I never imagined that I would contribute directly to the field of human genetics, let alone receive the Allen Award. The chairman of the department at Michigan in those days was Jim Neel, who believed that the discipline of "human genetics" made sense only in the context of the larger discipline, "genetics", that covered the study of inheritance in all living things. In this talk, I will tell you how the idea of making a genetic map by using restriction-fragment polymorphisms arose out of a fusion of classical and molecular genetics in other "model" organisms. I will try to show you that the basic ideas underlying RFLP mapping have a distinguished history dating back to the beginning of this century. My conclusion is that through the advent of DNA-based technology the study of human genetics, from the point of view of experimental practice and not just basic principle, has become reunified with general genetics, just as Jim Neel always imagined it must.

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 (for a review, see Ott 1985).

Received August 9, 1990.

Address for correspondence and reprints: David Botstein, Department of Genetics, Room S-337, Stanford University School of Medicine, Stanford, CA 94305.

0002-9297/90/4706-0001$02.00

DNA Polymorphisms and Linkage Mapping

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. Mendelian inheritance—and, indeed, genetic linkage—can be observed readily, given only loci polymorphic enough so that the parental alleles at each locus can be distinguished one from the other. In the ideal case each locus would be so polymorphic so that in virtually any mating the four parental alleles at each locus are different.

Recombinant DNA technology provided a source of polymorphic markers in the form of 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. 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 (1975). 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* be-

cause 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, then, 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.

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 (1931) 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). These papers describe the generation and physical mapping of deletion and substitution mutations by formation of DNA heteroduplexes followed 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, Topp, and Sambrook in 1976 (Botchan et al. 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. Their 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 constitute a single tandem array at a single locus (subsequently mapped to chromosome 12). 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 eight 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 deliberately find RFLPs, 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 for Huntington 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 of 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 usage of "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." (The quotation is from 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.

Application of Linkage Mapping to Human Diseases

The first autosomal disease gene was mapped, using RFLPs, by Gusella et al. (1983), who found the gene causing Huntington disease linked to an RFLP marker located near the end of the short arm of chromosome 4. Since then, a considerable 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 (Kerem et al. 1989; Rommens 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 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 1986*a*; 1986*b*). 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.

Table 1

Diseases Mapped by Using RFLPs

Disease	Chromosome	Reference(s)
Duchenne muscular dystrophy	X	Davies et al. 1983
Huntington disease	4	Gusella et al. 1983
Retinoblastoma	13	Cavenee et al. 1983
Cystic fibrosis	7	Knowlton et al. 1985; Tsui 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 Recklinhausen 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

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 disease gene mapping without family study ("homozygosity mapping"; Lander and Botstein 1987), mapping of diseases showing heterogeneity of cause ("simultaneous search"; Lander and Botstein 1986*b*), 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. (1987). This map had 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, with respect to both 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 (by using simultaneous search) and to very rare recessive diseases (by homozygosity mapping).

In conclusion, the idea of mapping genes by 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 diseases more complex 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 U.S. 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 on molecules.

References

Barker D, Wright E, Nguyen K, Cannon L, Fain P, Goldgar D, Bishop DT, et al (1987) Gene for von Recklinghausen neurofibromatosis is in the pericentromeric region of chromosome 17. Science 236:1100–1102

Bodmer WF, Bailey CJ, Bodmer J, Bussey HJ, Ellis A, Gorman P, Lucibello FC, et al (1987) Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature 328:614–616

Botchan M, Topp W, Sambrook J (1976) The arrangement of simian virus 40 sequences in the DNA of transformed cells. Cell 9:269–287

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

Cavenee WR, Dryja TP, Phillips RA, Benedict WF, GodBout R, Gallie BL, Murphree AL, et al (1983) Expression of

recessive alleles by chromosome mechanisms in retinoblastoma. Nature 305:779–784

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

Danna KJ, Nathans D (1971) Specific cleavage of SV40 DNA by restriction endonucleases of Hemophilus influenzae. Proc Natl Acad Sci USA 68:2913–2917

Davies KE, Pearson PL, Harper PS, Murray JM, O'Brien T, Sarfrazi M, Williamson R (1983) 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 2:2303–2312

Davis RW, Parkinson JS (1971) Deletion mutants of bacteriophage lambda. III. The structure of att-p. J Mol Biol 56:403–423

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, et al (1987) A genetic linkage map of the human genome. Cell 51:319–337

Drayna D, White R (1985) The genetic linkage map of the human X chromosome. Science 230:753–758

Grodzicker T, Williams J, Sharp P, Sambrook J (1974) Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harbor Symp Quant Biol 39:439–446

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234–238

Kan YW, Dozy AM (1978) Antenatal diagnosis of sickle-cell anemia by DNA analysis of amniotic-fluid cells. Lancet 2:910–912

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, et al (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080

Knowlton RG, Cohen-Haguenauer O, Van Cong N, Frezal J, Brown VA, Braman JC, Schumm JW, et al (1985) A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. Nature 318:380–382

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

——— (1986*b*) Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. Proc Natl Acad Sci USA 83:7353–7357

——— (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567–1570

——— (1988) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185–194

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

Olson MV, Montgomery DL, Hopper AK, Page GS, Horodyski F, Hall BD (1977) Molecular characteristics of the tyrosine tRNA genes of yeast. Nature 267:639–641

Ott J (1985) Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete RFLP linkage map. Nature 335:721–726

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

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

Reeders ST, Breuning MH, Davies KE, Nicholls RD, Jarman AP, Higgs DR, Pearson PL, et al (1985) A highly polymorphic DNA marker linked to adult polycystic kidney diseases on chromosome 16. Nature 317:542–544

Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, et al (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245:1059–1065

Rouleau GA, Wertelecki W, Haines JL, Hobb WJ, Trofatter JA, Seizinger BR, Martuza RL, et al (1987) Genetic linkage of bilateral acoustic neurofibromatosis to a DNA marker on chromosome 22. Nature 329:246–248

Seizinger BR, Rouleau GA, Ozelius LJ, Lane AH, Faryniarz AG, Chao MV, Huson S, et al (1987) Genetic linkage of von Recklinghausen neurofibromatosis to the nerve growth factor receptor gene. Cell 49:589–594

Simpson NE, Kidd KK, Goodfellow PJ, McDermid H, Myers S, Kidd JR, Jackson CE, et al (1987) Assignment of multiple endocrine neoplasia type 2A to chromosome 10 by linkage. Nature 328:528–530

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

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

Tsui LC, Buchwald M, Barker D, Braman JC, Knowlton RG, Schumm JW, Eiberg H, et al (1985) Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. Science 230:1054–1057

Weissmann C (1978) Reversed genetics. Trends Biochem Sci 109–111

Weissman C, Nagata S, Taniguchi T, Weber H, Meyer F (1979) The use of site directed mutagenesis in reversed genetics. In: Setlow JK, Hollaender A (eds) Genetic engineering: principles and methods, vol 1. Plenum, New York, pp 133–150

White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel JM, et al (1985) A closely linked genetic marker for cystic fibrosis. Nature 318:382–384