The Genetic Linkage Map of the Human X Chromosome

Dennis Drayna and Ray White

Summary. A database useful for mapping the human X chromosome has been established. The data consist of the genotypic characterizations obtained at more than 20 DNA marker loci from a set of 38 selected families. Multilocus linkage analysis has provided an initial genetic map completely spanning the distance from the distal short arm to the distal long arm of the chromosome, for a total genetic length of at least 185 recombination units. Analysis of the recombinational behavior of fully marked chromosomes suggests that the number of recombination events on the X chromosome may be nonrandom. Linkage studies of six families that carry the mutation which causes Duchenne muscular dystrophy were combined with linkage data from a large number of normal families. This permitted mapping of the locus for Duchenne muscular dystrophy with greater precision and statistical confidence than studies in which disease families alone provided the genotypic database. This observation suggests that the normal linkage map of this chromosome should be especially valuable in the mapping of rare X-linked diseases.

The X chromosome has been studied more intensively than any other human chromosome, because its unique inheritance allows identification of recessive X-linked traits. A linkage map provides an essential framework for understanding and predicting chromosome behavior at the molecular genetic level. Because of the importance of the X chromosome for the study of human genetics, we have devoted our efforts to the construction of a genetic linkage map for this chromosome. The map is based on meiotic recombination frequencies rather than on physical localization. A widely applicable genetic map of this chromosome will be useful in the diagnosis and prediction of many X-linked genetic diseases.

The primary genetic defect is unknown for many X-linked diseases. Some of these disorders, such as Duchenne muscular dystrophy and fragile X-linked mental retardation, are genetic diseases of major importance. A genetic trait can be mapped by linkage without a detailed knowledge of the biochemical defect causing the disease. Specifically, even in the absence of a direct test for a specific mutant gene, it is possible to diagnose or predict a disease if ubiquitous closely linked markers are available. For such markers, we have relied on cloned DNA segments that reveal polymorphism in the length of the fragments produced by cleavage of genomic DNA with certain bacterial endonucleases (restriction enzymes). These restriction fragment length polymorphisms (RFLP's) can be detected in human DNA by hybridization of a single-copy DNA probe to a Southern blot of total genomic DNA digested with restriction enzymes (*1*). RFLP's have several important advantages as genetic markers. They can be found in virtually all regions of the genome, and they act as codominant systems, providing complete genotypic information at their respective loci. Furthermore, a high degree of polymorphism can often be developed in RFLP's, and this makes them useful as genetic markers for most families.

To construct the map, we determined the genetic linkage relationships among a number of X chromosome RFLP markers in a series of normal families, because a normal population is the best source of families with pedigree structures optimal for efficient accumulation of linkage data. These markers serve to identify loci of genes associated with disease (OTC, HPRT, FVIII, and FIX) as well as arbitrary loci defined solely by DNA. The DNA marker loci used in our study together with their characteristics and sources are listed in Table 1.

The X chromosome map resulting from our linkage studies (Fig. 1) has allowed us to examine the meiotic behavior of X chromosomes as a population and to improve the resolution of linkages between DNA markers and X-linked diseases.

Building the Map

Two types of information contribute to the construction of a genetic map of the X chromosome: localization of the DNA markers by physical methods, and determination of the various genetic linkage relationships among them. The physical locations of these markers were approximated by in situ hybridization to metaphase chromosomes (*2*) or by hybridization to the DNA of a panel of hybrid cell lines, each containing a different portion of the human X chromosome (*3*). In some instances, both techniques were used.

We determined the genetic linkage relationships among 21 DNA markers (Table 1) by examining DNA from 38 normal families. These families are characterized by the presence of all four grandparents and by large sibships, averaging nine children each (*4*). We gathered complete genotypic information for each marker in each family in which the marker was informative—that is, the mother was heterozygous. Because the maternal grandfather in each family passes an exact copy of his only X chromosome to his daughter, the genotype of the maternal grandfather explicitly determines the disposition of alleles on the two maternal chromosomes (that is, their haplotype phase). When phase is known, the recombination frequency between two marker loci can be estimated by counting the recombinant and nonrecombinant chromosomes.

In the construction of the linkage map from our genotypic data we considered two features, genetic distance and genetic order. The distinction between these two parameters is important, for while gene distances are expected to change as the data set enlarges, it would be useful for the gene order to remain largely invariant (*5*).

Gene distances. We obtained a first estimate of genetic linkage distance between loci by examining the recombination frequency between pairs of marker loci. The number of meioses generating X chromosomes in which recombination could be scored between any two marker loci ranged from 0 to 88, with a mean of

Dennis Drayna was formerly at the Howard Hughes Medical Institute, University of Utah Medical Center, Salt Lake City 84132, and is now in the Department of Molecular Biology, Genentech, South San Francisco, California 94080. Ray White is associated with the Howard Hughes Medical Institute and the Department of Human Genetics, University of Utah Medical Center.

26.5. Listed in Table 2 are the recombination fractions observed between each marker locus and the four marker loci that emerged as the closest adjacent loci in the final map. Examining pairwise cross data alone, we can often estimate a consistent marker order. However, the standard errors associated with each linkage distance are frequently such that several different orders are possible.

Gene order. To determine gene order, we began by considering the physical evidence for localization of our 21 DNA markers. As shown in Table 1, the physical order of many markers along the chromosome is known with sufficient resolution that order is left uncertain for only a subset of markers in certain regions. On the long arm, for example, the physical order is cen-DXYS1-S21-HPRT-52A-DX13 (*2*); this establishes a framework that can be used to order other loci on the long arm.

When physical methods were insufficient for resolving marker order, we relied on genetic evidence derived from multifactor crosses. These can be scored in progeny chromosomes from mothers simultaneously informative for three or more linked markers. Multifactor crosses are particularly powerful for determining gene order because they provide an internal accounting of crossover events among the loci being examined. If, for example, the order of three loci is A-B-C, then a crossover between A and B will reveal recombination between A and B and between A and C, but not between B and C, neglecting double exchanges. Likewise, a single crossover between B and C will show recombination between B and C and between A and C, but not between A and B (Fig. 2). The likelihood of the order A-B-C over the alternative orders can then be quantified on the basis that chromosome classes requiring double exchanges will be the least frequently observed. This logic can be extended to more than three markers, and confidence in gene order can be quantified in the same fashion.

Likelihood Calculations

The support for a gene order can be quantified by the method of maximum likelihood; the relative likelihood of each order is calculated at the most likely recombination fractions between the markers (*6*). These overall likelihoods can then be compared to determine the relative support for one gene order over another. Computerized systems for multilocus likelihood calculations, recently developed by Lathrop *et al.* (*7*), greatly facilitate these analyses. We have made extensive use of the linkage analysis program ILINK for this purpose (*7*). The ILINK program iteratively calculates the most likely recombination fraction for several loci considered jointly, taking into account information available from both pairwise and multifactor crosses. We have analyzed each successive set of four loci with ILINK. The program was run serially, each time specifying a different gene order consistent with known physical locations. This analysis produced a series of likelihoods, one for each marker order. ILINK analysis also provided the most likely genetic distances among the four loci; in consequence, we were able to build the map with the most likely parameters for both distance and order, based on markers taken four at a time.

Examination of markers four at a time allows the map locations for one group of four loci to be conditional on the information from the previous, overlapping group of four loci. For example, in a given set of four loci, the most likely order and distances between loci 1, 2, 3, and 4 were determined. The resulting distance between loci 2 and 3 was then fixed and used to determine the most likely order and distances for loci 2, 3, 4, and 5. The map produced by this technique integrates the information on marker order and marker distance for all 21 loci. The most distal and highly polymorphic locus, St14, was chosen as the zero point of the map.

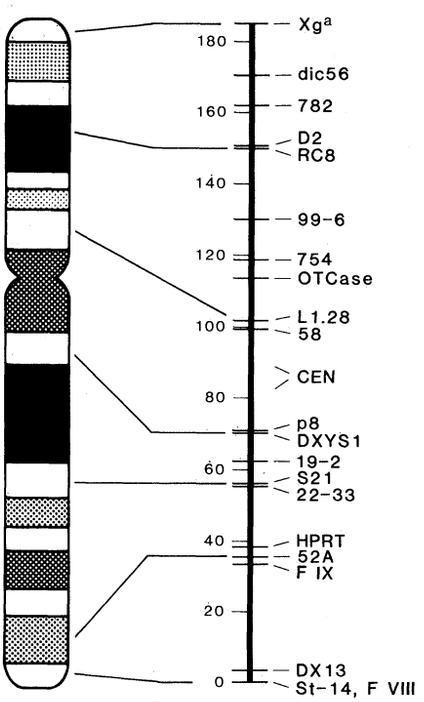

Fig. 1. Linkage map of the human X chromosome. Genetic distances are represented in recombination units, and physical locations of selected markers are indicated on the ideogram.

For each set of four marker loci, the results of multifactor analysis with ILINK are listed in Table 2, which gives the most likely gene order within each successive group of four loci. Table 2 also lists the likelihood of that order when compared to the second most likely order. The second most likely order was usually that of the middle two loci inverted. In some cases, pairwise cross information on one interval was limited; either there were no informative meioses (such as in the case of 58 × L1.28), or we observed no recombinants in a small number of meioses (as in D2 × RC8). In these cases, the genetic evidence for relative order had to be based solely on two-factor cross information with neighboring markers. This relatively weak method of determining gene order results in limited support for a given order, and the order shown in Fig. 1 is in some instances based solely on physical location, as noted in Table 2. In these cases, the confidence in the gene order is presented with respect to the second most likely order that is consistent with the physical locations of those markers. Where the $\log_{10}$ of the odds for a unique order is less than 3, we must consider the most likely order as only tentative. However, this method does provide evidence as to where the map is strong and where it is weak.

In regions of the map in which genetic evidence for order was limited, loci were analyzed five at a time with ILINK to extract the most information possible from the genotypic data available. For example, the position of marker 754 was determined by analyzing 754 with RC8, 99-6, OTCase, and L1.28, and again with 99-6, OTCase, L1.28, and 58. The final map reveals that the human X chromosome is at least 185 recombination units in length; the short arm displays roughly the same genetic length as the long arm.

The results obtained in 38 normal Utah kindreds are generally consistent within the set. However, one instance represented striking inconsistency between families in the interval DX13 to St14. In seven families, no recombinants were observed in 47 meioses. In the eighth family, these two markers displayed three recombinants in seven meioses. A recombination fraction of 0 obtained in 47 meioses would place these two markers within 5.6 recombination units of each other (95 percent confidence interval). The probability of observing three recombinants out of seven if the true recombination fraction was 0.056 is less

than 2.0×10^{-4}. It is difficult, however, to determine whether this inconsistency is the result of statistical fluctuation or of some fundamental heterogeneity.

Distribution of Crossover Events

The current data set contains 285 X chromosomes in the third generation—that is, chromosomes in which recombination can be scored. The average X chromosome in this set is informative for 7.3 markers, with a fraction of this set (105 chromosomes) informative for eight or more markers evenly distributed throughout the length of the chromosome. When this subset of well-marked chromosomes was examined for the number and distribution of recombination breakpoints, we found that chromosomes can undergo zero to four recombination events per meiosis, as illustrated in Fig. 3. At this level of resolution, the X chromosomes displayed a mean of 1.5 recombination events per chromosome, with 14 percent of the X chromosomes showing no recombination, 34 percent showing a single exchange, 35 percent showing two exchanges, 12 percent showing three exchanges, and 1 percent showing four exchanges. The distribution of crossover events shows deviation from the Poisson expectation which is significant at a level of 0.05, according to the χ^2 test. The deviation indicates that the number of recombination events on the X chromosome may be nonrandom.

Mapping Genetic Diseases

The current distribution of RFLP markers on the X chromosome is such that any X-linked disease locus will map within ten recombination units of at least two markers. Because of this proximity of markers, we should be able to locate disease loci whose ultimate genetic defect is unknown. Toward this end, we have studied linkage relationships among a number of DNA markers and several X-linked diseases.

For example, we sampled six families in Utah with Duchenne muscular dystrophy (DMD). Each family contained three or four generations. These six families provided information on the linkages between DMD and the markers 99-6, 754, OTCase, and 58. The most likely linkage distances and marker order, calculated with ILINK, indicate that the DMD locus lies between the markers 754 and 99-6, being 15 ± 6 (standard error of the mean) recombination units proximal to 99-6 and 21 ± 11 units distal to 754. Findings were similar in a recent collaborative study (*8*).

Disease families have also been studied for linkage between DNA markers and several other X-linked diseases, including X-linked retinitis pigmentosa (*9*), hemophilia A (*10*), X-linked retinoschisis (*11*), and choroideremia (*12*).

Using the Normal Map to Refine Disease Linkages

The linkage results obtained for the diseases Duchenne muscular dystrophy, hemophilia A, and X-linked retinitis pigmentosa indicate that a reasonable level of certainty of genetic location (and thus

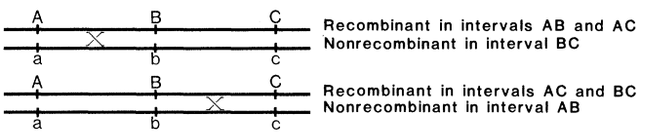

Fig. 2. Logic for definition of gene order when multifactor crosses are used. Gene orders other than A-B-C will require double exchanges to produce chromosomes that display this distribution of recombination and nonrecombination among the three markers.

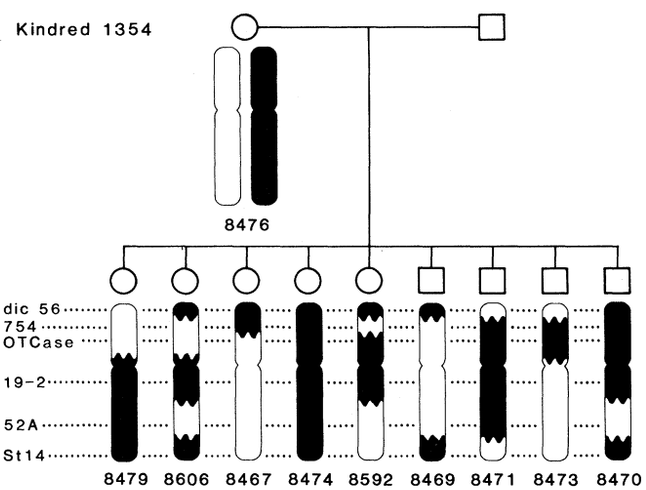

Fig. 3. Distribution of recombination breakpoints among X chromosomes in one family. The contribution of each maternal X chromosome toward the constitution of the X chromosomes in her progeny is shown by shading. Since the precise locations of recombination breakpoints between any two loci in the progeny X chromosomes are not known, the breakpoints are indicated by wavy lines.

Table 1. DNA markers on human chromosome X.

Marker	HGM8 workshop symbol	Physical location	Polymorphism	Allele frequency	Reference
Xg[a]		Xp22.3	Protein polymorphism	0.35/0.65	(*26*)
dic56	DXS143	Xp22.3	Bcl I	0.45/0.55	(*25*)
pD2	DXS43	Xp21-p22	Pvu II	0.29/0.71	(*19*)
RC8	DXS9	Xp21	Taq I	0.15/0.85	(*3*)
p99-6	DXS41	Xp21-p22	Pst I	0.29/0.71	(*19*)
754	DXS84	Xp21	Pst I	0.40/0.60	(*24*)
OTCase	OTC	Xp21	Msp I	0.30/0.25/0.25/0.20	
			Bam HI	0.35/0.65	(*23*)
L1.28	DXS7	Xp1	Taq I	0.35/0.65	(*22*)
p58	DXS14	Xcen-p21	Msp I	0.35/0.65	(*19*)
p8	DXS1	Xcen-q13	Taq I	0.12/0.88	(*19*)
DXYS1	DXYS1	Xq12	Taq I	0.48/0.52	(*21*)
p19-2	DXS3	Xq13-q22	Msp I	0.20/0.80	(*19*)
S21	DXS17	Xq22	Taq I	0.35/0.65	(*17*)
			Msp I	0.10/0.90	
22–23	DXS11	Xq24-Xqter	Taq I	0.16/0.84	(*19*)
HPRT	HPRT	Xq26	Bam HI	0.70/0.22/0.08	(*20*)
43–15	DXS42	Xq24-qter	Bgl II	0.19/0.81	(*19*)
52A	DXS51	Xq27	Taq I	0.50/0.50	(*17*)
F IX	F9	Xq27–28	Taq I	0.30/0.70	(*18*)
DX13	DXS15	Xq28	Bgl II	0.35/0.65	(*17*)
F VIII	F8	Xq28	Bcl I	0.30/0.70	(*10*)
St14	DXS52	Xq28	Taq I	0.36/0.20/0.15/0.12/0.11/0.05/0.01	(*16*)
			Msp I	0.35/0.35/0.19/0.11	

Table 2. Linkage data between DNA markers on the human X chromosome. N.I., not informative.

Marker		Two-factor crosses	Recombinants per informative meioses	Recombination fraction	Most likely gene order	Log_{10} of ratio of likelihood of most likely order to next most likely order
St14	×	F VIII	0/57	0.0 ± .048	No recombinants between St14 and FVIII; order not determined	
		DX13	3/54	.055 ± .021		
		F IX	17/57	.285 ± .059		
		52A	31/88	.352 ± .051		
		43-15	23/53	.434 ± .068		
FVIII	×	DX13	0/27	0.0 ± .074	(St14, FVIII) DX13 FIX	0.14
		52A	6/27	.222 ± .080		
		43-14	12/26	.461 ± .096		
		HPRT	6/20	.300 ± .102		
DX13	×	FIX	13/33	0.393 ± .085	52A FIX DX13 St14	32.14
		52A	20/69	.290 ± .055		
		43-15	6/19	.316 ± .107		
		HPRT	5/13	.385 ± .135		
FIX	×	52A	1/55	0.018 ± .017	43-15 52A FIX DX13	0.43
		43-15	0/7	0.0 ± .125		
		HPRT	1/13	.077 ± .074		
		22-33	1/7	.143 ± .132		
52A	×	43-15	3/35	0.086 ± .047	HPRT 43-15 52A FIX	3.49
		HPRT	1/20	.050 ± .049		
		22-23	4/22	.182 ± .082		
		S21	7/29	.241 ± .079		
43-15	×	HPRT	1/25	0.040 ± .039	22-33 HPRT 43-15 52A	0.26
		22-33	3/14	.214 ± .110		
		S21	2/16	.125 ± .083		
		19-2	3/8	.375 ± .171		
HPRT	×	22-33	4/14	0.285 ± .121	S21 22-23 HPRT 43-15	8.64
		S21	2/7	.285 ± .171		
		19-2	2/11	.182 ± .116		
		DXYS1	4/7	≧.50 ± .189		
22-33	×	S21	0/7	0.0 ± .125	19-2 S21 22-23 HPRT	0.01
		19-2	0/8	0.0 ± .112		
		DXYS1	5/15	.300 ± .118		
		p8	N.I.			
S21	×	19-2	1/17	0.059 ± .057	DXYS1 19-2 S21 22-23	0.66
		DXYS1	4/32	.125 ± .058		
		p8	N.I.			
		58	1/10	.100 ± .095		
19-2	×	DXYS1	1/23	0.044 ± .043	p8 DXYS1 19-2 S21	0.91
		p8	N.I.			
		58	1/19	.053 ± .051		
		L1.28	6/16	.375 ± .121		
DXYS1	×	p8	N.I.			
		58	0/16	0.0 ± .070	L1.28 58 p8 DXYS1	Order based on physical location
		L1.28	8/15	≧.50 ± .130		
		OTCase	19/73	.261 ± .051		
p8	×	58	N.I.			
		L1.28	4/8	0.50 ± .176	58 p8 DXYS1 19-2	Order based on physical location
		OTCase	6/16	.375 ± .121		
		754	3/8	.375 ± .171		
58	×	L1.28	N.I.			
		OTCase	1/9	0.111 ± .105	OTC L1.28 58 p8	Order based on physical location
		754	3/10	.300 ± .145		
		99-6	5/9	≧.50 ± .130		
L1.28	×	OTCase	4/22	0.182 ± .082	754 OTC L1.28 58	1.41
		754	4/32	.125 ± .058		
		99-6	3/6	≧.50 ± .204		
		RC8	7/20	.350 ± .107		
OTCase	×	754	4/88	0.045 ± .022	99-6 754 OTC L1.28	0.84
		99-6	12/67	.179 ± .047		
		RC8	9/27	.333 ± .098		
		D2	16/40	.400 ± .077		
754	×	99-6	3/37	0.081 ± .045	RC8 99-6 754 OTC	5.30
		RC8	5/19	.263 ± .101		
		D2	18/62	.290 ± .098		
		782	9/33	.273 ± .077		

Table 2 (continued).

Marker		Two-factor crosses	Recombinants per informative meioses	Recombination fraction	Most likely gene order	Log_{10} of ratio of likelihood of most likely order to next most likely order
99-6	×	RC8	5/34	0.147 ± .061	D2 RC8 99-6 754	2.86
		D2	4/17	.235 ± .103		
		782	9/30	.300 ± .084		
		dic56	8/23	.348 ± .065		
RC8	×	D2	0/7	0.0 ± .125	782 D2 RC8 99-6	0.00
		782	2/12	.167 ± .087		
		dic56	2/12	.167 ± .087		
		Xg^a	5/14	.357 ± .128		
D2	×	782	3/16	0.187 ± .074	dic56 782 D2 RC8	0.50
		dic56	2/17	.118 ± .075		
		Xg^a	N.I.			
782	×	dic56	2/18	0.111 ± .073	Xg^a dic56 782 D2	3.71
		Xg^a	3/8	.375 ± .092		
dic56	×	Xg^a	0/6	0.0 ± .140		

certainty of predictive value of the DNA markers) is possible if large numbers of families are available to build linkage data. Most X-linked diseases, however, are much less common than DMD, hemophilia A, or retinitis pigmentosa, and frequently occur in only one family or a few families (*13*). How can such diseases be accurately mapped and predicted?

We suggest that RFLP linkage data gathered in normal families can greatly improve the quality of linkage information in cases of rare genetic diseases. When only a few informative meioses are available, the linkage distances from RFLP markers to a disease locus will be characterized by a large variance. This variance can be constrained, however, by the relatively large amount of information available on the linkage distance between the RFLP markers themselves, obtained from normal families.

An example of the usefulness of the normal map is provided by our analysis of the limited amount of linkage information we obtained in the six Utah DMD kindreds. These families were genotyped at the loci RC8, D2, 99-6, 754, L1.28, and 58. The number of meioses informative for DMD and for each of the DNA markers ranged from none (for RC8 and L1.28) to 11 for 99-6; very little information on linkage between DNA markers was obtained. From these families alone, information on the pairwise linkage distances between each marker and DMD (calculated with LINKMAP) was very weak, as shown in Fig. 4A; the pairwise LOD scores for linkage to DMD did not exceed +0.40 for any DNA marker.

By contrast, the linkage database gathered on the same DNA markers in normal families is substantial. We took advantage of the information in the larger database by examining the genetic linkages between the DMD locus and the four closest loci—namely, RC8 and 99-6 distal, and 754 and OTCase proximal—pooling the DMD and normal family data. We analyzed the entire data set with ILINK to determine the most likely genetic distances between the loci taken five at a time. With normal and DMD data sets combined, the most likely position of the DMD locus was established between 99-6 and 754, with the distance from 99-6 to DMD now estimated at 9.1 ± 3.6 units, and the distance from DMD to 754 at 12.9 ± 3.5 units (Fig. 4B). This example demonstrates the potential usefulness of combining in a multifactor analysis the normal map with data from disease families in the effort to map rare genetic diseases accurately.

Discussion

A genetic map represents a description of an important part of the fundamental biology of an organism. In addition, it serves as a tool for the experimental study, and ultimately for experimental manipulation, of that organism. While technical and ethical considerations make experimental modification of the human genome problematic, the genetic map of the human X chromosome is of immediate value in the study and diagnosis of human X-linked genetic diseases.

The data set on which the X chromosome linkage map is based is sufficiently developed to allow efficient mapping of new markers. The average X chromosome in our set is currently informative for more than seven markers, and it will be possible to add new markers rapidly by observing their segregation in X chromosomes that contain defined crossover breakpoints. The genotypic data derived from this study, as well as DNA samples from family members, will be made available through the Centre d'Etude du Polymorphisme Humain in Paris, to assist in an international collaboration to construct a linkage map of the human genome.

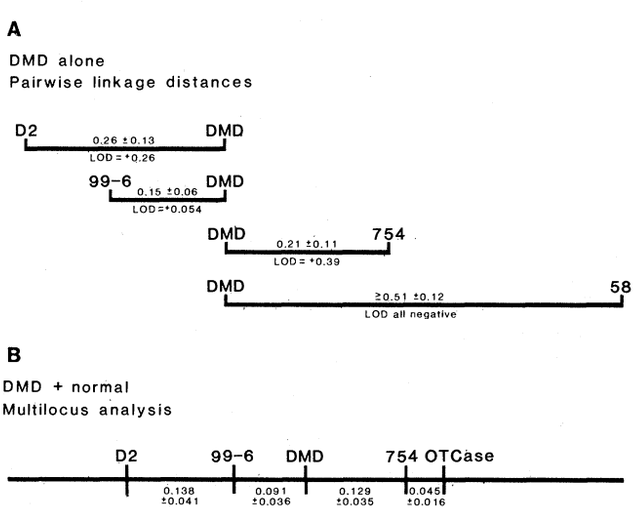

Fig. 4. Refinement of estimates of gene distance by combining the normal map with linkage information on DMD disease families. The numbers on the distance map represent the recombination fractions with standard errors. Placement of DNA markers proximal or distal to the DMD locus was based on physical mapping (*25*).

Finally, the genetic map will provide information important to the task of spanning large regions of the chromosome by physical methods now being developed (*14*, *15*). The defined loci will provide essential reference points for studies that seek to "walk" along the X chromosome.

References and Notes

1. D. Botstein, R. White, M. Skolnick, R. Davis, *Am. J. Hum. Genet.* **32**, 314 (1980).
2. D. Hartley, K. Davies, D. Drayna, R. White, R. Williamson, *Nucleic Acids Res.* **12**, 5277 (1984).
3. J. Murray *et al.*, *Nature (London)* **300**, 69 (1982).
4. R. White *et al.*, *ibid.* **313**, 101 (1985).
5. C. Bridges and T. Morgan, publication of the Carnegie Institution of Washington, Washington D.C. (1923).
6. A. Edwards, *Likelihood: An Account of the Statistical Concept of Likelihood and its Application to Scientific Inferences* (Cambridge Univ. Press, Cambridge, England, 1972).
7. G. Lathrop, J. Lalouel, C. Julier, J. Ott, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3443 (1984).
8. E. Bakker *et al.*, *Lancet* **1985-I**, 655 (1985).
9. S. Bhattacharya *et al.*, *Nature (London)* **309**, 253 (1984).
10. J. Gitschier, D. Drayna, E. Tuddenham, R. White, R. Lawn, *ibid.* **314**, 738 (1985).
11. P. Wieacker *et. al.*, *Hum. Genet.* **64**, 143 (1983).
12. R. Nussbaum, personal communication.
13. V. McKusick, *Mendelian Inheritance in Man* (University Park Press, Baltimore, Md., ed. 6, 1984).
14. D. Schwartz and C. Cantor, *Cell* **37**, 67 (1984).
15. F. Collins and S. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6812 (1984).
16. I. Oberle, D. Drayna, G. Camerino, R. White, J. Mandel, *ibid.* **82**, 2824 (1985).
17. D. Drayna *et al.*, *ibid.* **81**, 2836 (1984).
18. G. Camerino *et al.*, *ibid.*, p. 498.
19. J. Aldridge *et al.*, *Am. J. Hum. Genet.* **36**, 546 (1984).
20. R. Nussbaum, W. Crowder, W. Nyhan, C. T. Caskey, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4035 (1983).
21. D. Page *et al.*, *ibid.* **79**, 5352 (1982).
22. K. Davies *et al.*, *Nucleic Acids Res.* **11**, 2302 (1983).
23. R. Rozen, J. Fox W. Fenton, A. Horwich, L. Rosenberg, *Nature (London)* **313**, 815 (1985).
24. B. Bakker and P. Pearson, personal communication.
25. L. Kunkel, personal communication.
26. R. Race and R. Sanger, *Blood Groups in Man* (Blackwell, Oxford, ed. 6, 1975), p. 610.

27. Supported by a postdoctoral fellowship from the Muscular Dystrophy Association (to D.D.) and by a Howard Hughes Medical Institute investigator award (to R.W.). We thank investigators who generously agreed to share X chromosome DNA probes before publication, including L. Kunkel and J. Aldridge, P. Pearson and B. Bakker, and J.-L. Mandel and I. Oberle. We also thank F. Ziter for his role in gathering and sampling muscular dystrophy families and M. Hoff for excellent technical assistance. We thank P. Callahan and T. Elsner for help with linkage analysis and B. Ogden for help in sampling muscular dystrophy families.

Plasticity of the Differentiated State

Helen M. Blau, Grace K. Pavlath, Edna C. Hardeman
Choy-Pik Chiu, Laura Silberstein, Steven G. Webster
Steven C. Miller, Cecelia Webster

Tissue-specific phenotypes result from a sequence of developmental stages. Totipotent cells in the early embryo give rise to stem cells specific to three distinct layers—endoderm, ectoderm, and mesoderm. Although the lineage, or progression from stem cell to tissue-specific phenotype is not always fixed (*1*), once a cell is determined, it is generally destined for specialization along a specific pathway, such as erythropoiesis or myogenesis. The option to generate other phenotypes no longer exists for the determined vertebrate cell, and its progeny stably inherit its limited potential. The determined cell will give rise to other phenotypes only under unusual experimental conditions, such as at sites of regeneration in amphibian limbs where transdifferentiation has occurred or after treatment with a drug such as 5-azacytidine; even then, only derivatives of the same embryonic lineage are obtained (*2*). At some point in development, the determined cell expresses its phenotype, and the genes necessary for its role in the function of a particular tissue are transcribed.

To obtain tissue-specific phenotypes, a sequence of regulatory mechanisms must exist that determine when in a cell's history specific genes are transcribed. The genetic composition of eukaryotic cells is generally stable and heritable. Chromosomes are not lost in the course of cell specialization. This is evident since entire frogs can be generated from the transplantation of nuclei of specialized intestine cells into enucleated oocytes and since a diversity of normal tissue-specific cell types can be generated from malignant tumor cells introduced into early mouse embryos (*3*). The current model for the differential expression of genes characteristic of tissues at different points in development requires regulation by DNA sequences on the same chromosome (*cis*-acting) and on different chromosomes (*trans*-acting). *Cis*-acting DNA sequences that impart tissue-specific regulation have been identified from the study of the expression of cloned genes after transfection into cultured cells (*4*). The diffusible products of *trans*-acting genes are assumed to be negative or positive regulators of the *cis*-acting gene sequences. Although some general mediators of gene transcription and gene-specific binding proteins have been characterized in eukaryotes (*5*), with the exception of the factors that bind the *Drosophila* alcohol dehydrogenase gene (*6*), no tissue-specific *trans*-acting regulators have yet been isolated. An understanding of how the expression of tissue-specific genes is activated is not only of fundamental biological interest but also of practical importance in implementing genetic engineering and possibly gene therapy.

Muscle provides a model system for studies of the mechanisms controlling the appearance of tissue-specific functions. For a number of species, developmentally distinct stages are readily recognized by their morphological and biochemical properties, and conversion from one stage to another can be mimicked under the controlled conditions of tissue culture (*7*) (Fig. 1). First, a mesodermal stem cell gives rise to a myoblast, destined for myogenesis. The determined myoblast is capable of recognizing and spontaneously fusing with other myoblasts leading to the production of a differentiated myotube. The multinucleated myotube no longer divides or synthesizes DNA but produces muscle proteins in large quantity. These include constituents of the contractile apparatus and specialized cell-surface components essential to neuromuscular transmission. Eventually the differentiated muscle cell exhibits characteristic striations and rhythmic contractions. A further step in this pathway is maturation: the contractile apparatus in muscle at different stages of development contains distinct isoforms of muscle proteins such as myosin and actin, encoded by different members of multigene families (*8*, *9*).

The authors are in the Department of Pharmacology at Stanford University School of Medicine, Stanford, California 94305. The present addresses for C.-P.C. and G.K.P. are the Laboratory of Molecular Carcinogenesis, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, and the Department of Biological Sciences, Stanford University, respectively.