Am. J. Hum. Genet. 44:409–421, 1989

Polymorphic DNA Haplotypes at the LDL Receptor Locus

Eran Leitersdorf,* Aravinda Chakravarti,† and Helen H. Hobbs*

*Department of Molecular Genetics, University of Texas Southwestern Medical Center, Dallas; and †Human Genetics Division, University of Pittsburgh, Pittsburgh

Summary

Mutations in the low-density lipoprotein (LDL) receptor gene result in the autosomal dominant disorder familial hypercholesterolemia (FH). Many different LDL receptor mutations have been identified and characterized, demonstrating a high degree of allelic heterogeneity at this locus. The ability to identify mutant LDL receptor genes for prenatal diagnosis of homozygous FH or to study the role of the LDL receptor gene in polygenic hypercholesterolemia requires the use of closely linked RFLPs. In the present study we used 10 different RFLPs, including three newly described polymorphisms, to construct 123 independent haplotypes from 20 Caucasian American pedigrees. Our sample contained 31 different haplotypes varying in frequency from 0.8% to 29.3%; the five most common haplotypes account for 67.5% of the sample. The heterozygosity and PIC of each site were determined, and these values disclosed that eight of the RFLPs were substantially polymorphic. Linkage-disequilibrium analysis of the haplotype data revealed strong nonrandom associations among all 10 RFLPs, especially among those sites clustered in the 3′ region of the gene. Evolutionary analysis suggests the occurrence of both mutational and recombinational events in the generation of the observed haplotypes. A strategy for haplotype analysis of the LDL receptor gene in individuals of Caucasian American descent is presented.

Introduction

The low-density lipoprotein (LDL) receptor is a transmembrane cellular protein that is pivotal in cholesterol homeostasis (Brown and Goldstein 1986). Most plasma cholesterol is contained within the LDL particle. At the cell surface, LDL binds to its receptor and is transported into the cell via receptor-mediated endocytosis. The intracellular LDL-derived cholesterol down-regulates both the LDL receptor gene and genes encoding enzymes in the cholesterol biosynthetic pathway. Any interruption of LDL trafficking into cells results in both an elevation of plasma LDL-cholesterol (LDL-C) and an increase in endogenous cellular production of cholesterol (Brown and Goldstein 1986).

Individuals with a mutation in one of their two LDL receptor alleles have heterozygous familial hypercholesterolemia (FH) (Goldstein and Brown 1983). These individuals produce half the normal number of LDL receptors, resulting in a twofold elevation in plasma LDL-C. The elevated plasma LDL-C is associated with an increased risk of ischemic heart disease. Heterozygous FH has a population frequency of 0.2% (Goldstein and Brown 1983) but accounts for 5% of patients with myocardial infarction who are less than 60 years of age (Goldstein et al. 1973). Homozygous FH occurs rarely and can be due to the same mutation (a "true" homozygote) or different mutations (a compound heterozygote) in the two LDL receptor genes. Patients with homozygous FH have marked hypercholesterolemia and often die in childhood of premature coronary atherosclerosis.

Prior to cloning the LDL receptor gene it was clear that FH was a genetically heterogeneous disease. Biosynthetic and binding characteristics of the LDL receptor in cultured fibroblasts from FH patients demonstrated four different functional classes of mutations (Tolleshaung et al. 1982, 1983). Cloning of the LDL receptor gene allowed the molecular characterization of specific mutations in each of these four classes. A myriad of different mutations including deletions, in-

Received August 8, 1988; revision received November 2, 1988.

Address for correspondence and reprints: Helen H. Hobbs, Department of Molecular Genetics, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75235.

0002-9297/89/4403-0015$02.00

sertions, missense, and nonsense have thus far been described (Lehrman et al. 1985*a*, 1985*b*, 1986, 1987*a*, 1987*b*, 1987*c*; Davis et al. 1986; Hobbs et al. 1986, 1987*a*; Horsthemke et al. 1987; Russell et al., in press). With two exceptions (Hobbs et al. 1987*a*, Lehrman et al. 1987*c*), each mutation analyzed at the LDL receptor locus has been observed in only a single pedigree. This impressive allelic heterogeneity mandates the use of genetic markers at the LDL receptor locus for prenatal diagnosis, except in the unusual situation where the mutation is known and can be assayed directly.

Only in a minority of hypercholesterolemic patients,—those with the monogenic disorder FH—is the etiology of the disease known (Goldstein and Brown 1983). In the vast majority of patients, the etiology of the hypercholesterolemia is at best poorly defined. Hypercholesterolemia clusters in families (Goldstein et al. 1973), which suggests an underlying genetic mechanism, but, with few exceptions, the gene(s) responsible has not been identified (Breslow 1988). Alterations in the LDL receptor gene, alone or together with other genes, may have etiological significance in polygenic hypercholesterolemia. In an attempt to define genetic factors contributing to hypercholesterolemia, molecular markers consisting of RFLPs (Kan 1978) have been identified for the LDL receptor gene (Hobbs et al. 1985, 1987*b*; Humphries et al. 1985; Funke et al. 1986; Kotze et al. 1986, 1987; Leitersdorf and Hobbs 1987) as well as for other genes in the lipoprotein metabolic pathway (Breslow 1988).

RFLPs are differences in restriction-fragment lengths which result either from mutations that destroy or create new restriction sites or from insertions or deletions between invariant restriction sites (Kan 1987). RFLPs can be used to follow the segregation of a closely linked gene within families if all the parental alleles can be distinguished. The informativeness of RFLP analysis can be markedly improved by constructing haplotypes consisting of a defined pattern of RFLPs linked together on a chromosome. In addition to its use in linkage studies in pedigrees, haplotype analysis has been applied to population studies and can be useful in determining the origin and migration of particular alleles (Wainscoat et al. 1983; Chakravarti et al. 1984*b*; Orkin and Kazazian 1984). Haplotype analyses at the β-globin locus and at the phenylalanine hydroxylase gene locu have proved useful in the molecular analysis and diagnosis of β-thalassemia (Orkin et al. 1982; Orkin and Kazazian 1984) and phenylketonuria (DiLella et al. 1986, 1987; Chakraborty et al. 1987), respectively.

In the current study, we have used 10 RFLPs to construct parental haplotypes from individuals of 20 Caucasian American families. Seven of the 10 RFLPs have been reported elsewhere (Hobbs et al. 1985, 1987*b*; Humphries et al. 1985; Funke et al. 1986; Kotze et al. 1986, 1987; Leitersdorf and Hobbs 1987), and three new RFLPs are described here. The polymerase chain reaction (PCR) (Saiki et al. 1985) has been adapted to assay RFLPs located within the coding region of the gene (Leitersdorf and Hobbs 1988). Statistical analysis of the haplotypes indicates marked linkage disequilibrium between the 10 sites and throughout the region containing the LDL receptor gene. Owing to the high degree of linkage disequilibrium over the entire locus, not all the RFLPs were informative. Each RFLP was ranked according to its informativeness, and a strategy is presented for the optimal selection of RFLPs for pedigree analysis.

Material and Methods

Materials

Restriction endonucleases and DNA polymerase I (Klenow fragment) were obtained from New England Biolabs (Beverly, MA) and Boehringer Mannheim (Indianapolis). *Thermus aquaticus* DNA polymerase I was obtained from Perkin Elmer Cetus Corp. (Norwalk, CT). [α-^{32}P] dCTP (3,000 Ci/mmol) and [γ-^{32}P] ATP (7,000 Ci/mmol) were purchased from New England Nuclear (Boston) and from ICN Radiochemicals (Irvine, CA), respectively. Biotrans™ nylon membranes were from ICN Biomedical (Irvine, CA), and Hoeschst dye 33258 was obtained from Polyscience, Inc. (Warrington, PA). Oligonucleotides were synthesized on a DNA synthesizer model 380A, and DNA was extracted by a nucleic acid extractor model 340A; both machines were purchased from Applied Biosystems (Foster City, CA). DNA was amplified using the DNA Thermal Cycler from Perkin Elmer Cetus (Norwalk, CT). Plasma lipids were determined on a Technicon Analyzer model 2 (Technicon Instruments Corp., Tarrytown, NY).

Subjects

One hundred thirty-two Caucasian American individuals from 20 unrelated families living in the Dallas area were studied. Fasting plasma cholesterol and triglyceride levels were determined as described elsewhere (Block et al. 1965; Kessler and Lederer 1965). Analysis of the cholesterol content of plasma lipoproteins was performed according to the Lipid Research Clinic Program's *Manual of Laboratory Operations* (1974).

Genomic Blot Analysis

High-molecular-weight genomic DNA was prepared from peripheral leukocytes either as described elsewhere (Hobbs et al. 1986) or by using an automatic nucleic acid extractor. The DNA was quantified by fluorescence assay (Labarea and Paigen 1980) and diluted to a concentration of 90–100 ng/μl. Genomic DNA (8 μg) was digested to completion with each restriction endonuclease, in the buffer suggested by the manufacturer. Subsequently, the DNA was subjected to electrophoresis on a 0.8% (w/v) agarose gel and transferred to a nylon membrane. Single-stranded [α-^{32}P] dCTP–labeled probes derived from the LDL receptor cDNA were prepared by the method of Church and Gilbert (1984) and were hybridized (5×10^6 cpm/ml) to filters at 42°C for 16 h in 50% (v/v) formamide , 0.5% (w/v) each of BSA, Ficoll 400, and polyvinyl pyrrolidone 360, 5 × SSPE (0.9 M NaCl, 50 mM sodium phosphate, 5 mM EDTA, pH 7.4), 1% (v/v) SDS, and 100 μg denatured and sonicated salmon sperm DNA/ml. Following hybridization, filters were washed for 4 h. at 68°C in 0.5 × SSC (75 mM NaCl, 7.5 mM sodium citrate) and 1% (v/v) SDS and subjected to autoradiography on Kodak XAR-5 film with a Quanta III (Dupont, Wilmington, DE) intensifying screen for 24–48 h.

Primer-directed DNA Amplification and Restriction Analysis

Two 25-base PCR-oligonucleotides, S-1 and S-2, complementary to intron sequences flanking exon 8 of the LDL receptor gene, were synthesized. The 5′ PCR-oligonucleotide, S-1, was end-labeled with [γ-^{32}P] ATP (Maniatis et al. 1982, p. 122) One microgram of diluted genomic DNA was subjected to gene amplification using PCR (Saiki et al. 1988) with the following modifications: (1) the annealing and extension reactions were performed at 68°C for a total of 6 min, and (2) the denaturation was performed at 95°C for 1 min. The amplified DNA was size-fractionated on a 6% polyacrylamide gel in buffer A (0.09 M Tris-borate, 0.09 M boric acid, 2 mM EDTA) at 200 V for 1 h. The gel was exposed to Kodak XAR-5 film for 1 min. The amplified radiolabeled DNA band was excised from the gel, eluted, extracted with equal volumes of phenol and chloroform, and precipitated in 0.3 M sodium acetate (pH 5) and 2.5 vol ethanol. Approximately 100 ng (3×10^5 Cerenkov cpm) of the amplified DNA was subjected twice to either *Stu*I or *Bg*II digestion with 8 units of a given enzyme. The digested DNA was size-fractionated on a 6% polyacrylamide gel in buffer A and exposed to Kodak XAR-5 film for 5 min.

Haplotype Analysis

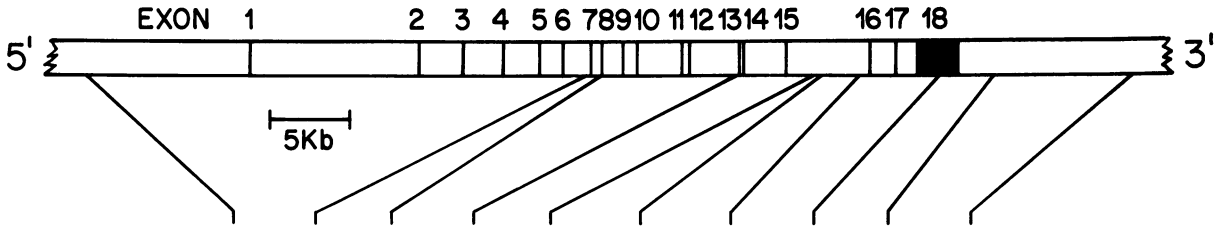
Genotypes at 10 RFLP sites were determined for individual members of the 20 families. Haplotypes for each individual were constructed by analysis of the joint segregation of the RFLPs in all members of each pedigree. For each pedigree, only the independent haplotypes were included in the analysis. Unambiguous haplotypes could be determined for 123 chromosomes. Construction of the haplotypes was based on the assumption that there had been no recombination within the LDL receptor locus in these families.

Statistical Methods

The allele frequencies, standard errors, and heterozygosities of each RFLP were calculated from the haplotype frequency data by standard methods for codominant markers (see, e.g., Chakravarti et al. 1984*a*); PIC values, a measure of the proportion of informative families, were calculated according to the formula of Botstein et al. (1980). Standardized linkage disequilibrium statistics (Δ) (Hill and Robertson 1968) and the relative amount of disequilibrium (D′) (Lewontin 1964) were calculated. The global test of linkage disequilibrium was performed according to the simulation method of Blanton and Chakravarti (1987) as described by Antonarakis et al. (1988). Evolutionary trees were constructed by using the principle of maximum parsimony and the computer program PAUP (Swofford 1985).

Results

A total of 264 LDL receptor alleles from 132 members of 20 different pedigrees were analyzed with 10 RFLPs. Seven of the 10 RFLPs enzymes were previously reported: *Pvu*II (Hobbs et al. 1985; Humphries et al. 1985), *Stu*I (Kotze et al. 1986), *Pst*I (Funke et al. 1986), *Nco*I (Kotze et al. 1987), *Ava*II (Hobbs et al. 1987*b*), and *Apa*LI (two separate polymorphic sites) (Leitersdorf and Hobbs 1987). Three additional RFLPs (*Bsm*I, *Sph*I, and *Spe*I) are included in the present analysis. Figure 1 shows a map of the LDL receptor gene, including the location of each RFLP. The RFLPs are distributed over the entire gene, extending from 10 kb 5′ to exon 1 to 13 kb 3′ to exon 18 and span a total distance of 62.8 kb. Two of the 10 RFLPs, *Ava*II (data not shown) and *Stu*I (fig. 2), are located within the coding region and are due to point mutations. The polymorphic *Nco*I site is located in the 3′ noncoding region of exon 18. The additional seven sites map to introns or flanking regions of the LDL receptor gene (fig. 1).



Haplotype	BsmI	SphI	StuI	AvaII	SpeI	ApaLI	PvuII	NcoI	PstI	ApaLI	n	%
1	+	+	+	+	−	+	−	+	+	+	36	29.3
2	+	−	+	−	−	−	+	+	−	+	15	12.7
3	+	−	+	−	−	−	−	−	−	−	20	16.3
4	+	+	−	−	−	−	+	+	−	+	1	0.8
5	+	+	+	+	−	−	−	−	−	−	3	2.4
6	+	+	+	+	−	+	−	+	+	−	3	2.4
7	+	−	+	−	−	−	−	−	−	+	1	0.8
8	−	+	+	−	+	+	−	+	−	+	2	1.6
9	−	+	+	+	−	+	−	+	+	+	2	1.6
10	+	+	+	+	−	+	−	+	−	+	1	0.8
11	−	−	+	−	−	−	−	−	−	−	3	2.4
12	−	−	+	−	−	−	+	+	−	+	1	0.8
13	+	−	+	−	−	−	+	+	+	+	1	0.8
14	+	−	+	+	−	−	+	+	−	+	1	0.8
15	−	+	+	−	−	−	−	−	−	−	3	2.4
16	−	+	−	−	−	−	+	+	−	+	6	4.9
17	+	−	+	+	−	−	+	+	+	+	1	0.8
18	+	−	+	+	−	+	+	+	+	+	1	0.8
19	+	+	+	−	+	+	−	+	+	+	1	0.8
20	+	−	+	−	−	+	+	+	−	+	1	0.8
21	+	−	+	−	−	+	−	+	+	+	2	1.6
22	+	−	+	+	−	+	−	+	+	+	2	1.6
23	−	+	+	+	−	+	−	−	+	+	2	1.6
24	+	+	+	−	−	+	−	+	+	+	2	1.6
25	+	−	+	−	+	+	−	+	−	+	1	0.8
26	+	+	+	−	−	−	−	+	−	−	1	0.8
27	−	−	+	−	−	−	−	+	−	+	1	0.8
28	−	+	+	−	−	−	+	−	+	+	1	0.8
29	+	+	+	−	−	−	+	+	−	+	1	0.8
30	+	+	+	+	−	+	−	−	+	+	1	0.8
31	+	+	+	−	+	+	−	+	−	+	6	4.9

TOTAL NUMBER OF CHROMOSOMES: 123

Figure 1 Physical map of the 10 RFLP sites used to construct the LDL receptor gene haplotypes. A total of 123 independent haplotypes were identified from analysis of 20 pedigrees that included 132 individuals. The numbering system reflects the order in which the haplotypes were encountered. The plus sign (+) and minus sign (−) denotes the presence or absence of the restriction site, respectively. The frequency of each given haplotype and its percentage (%) in the population studied are shown.

Each RFLP is diallelic and segregates in a codominant Mendelian fashion. Southern blot analysis using other restriction enzymes flanking each RFLP showed no evidence of insertions or deletions (data not shown). Thus, each polymorphism is probably due to a point mutation or to a small insertion or deletion that cannot be detected by genomic blotting.

*Bgl*I and *Stu*I can be used to assay an identical poly-

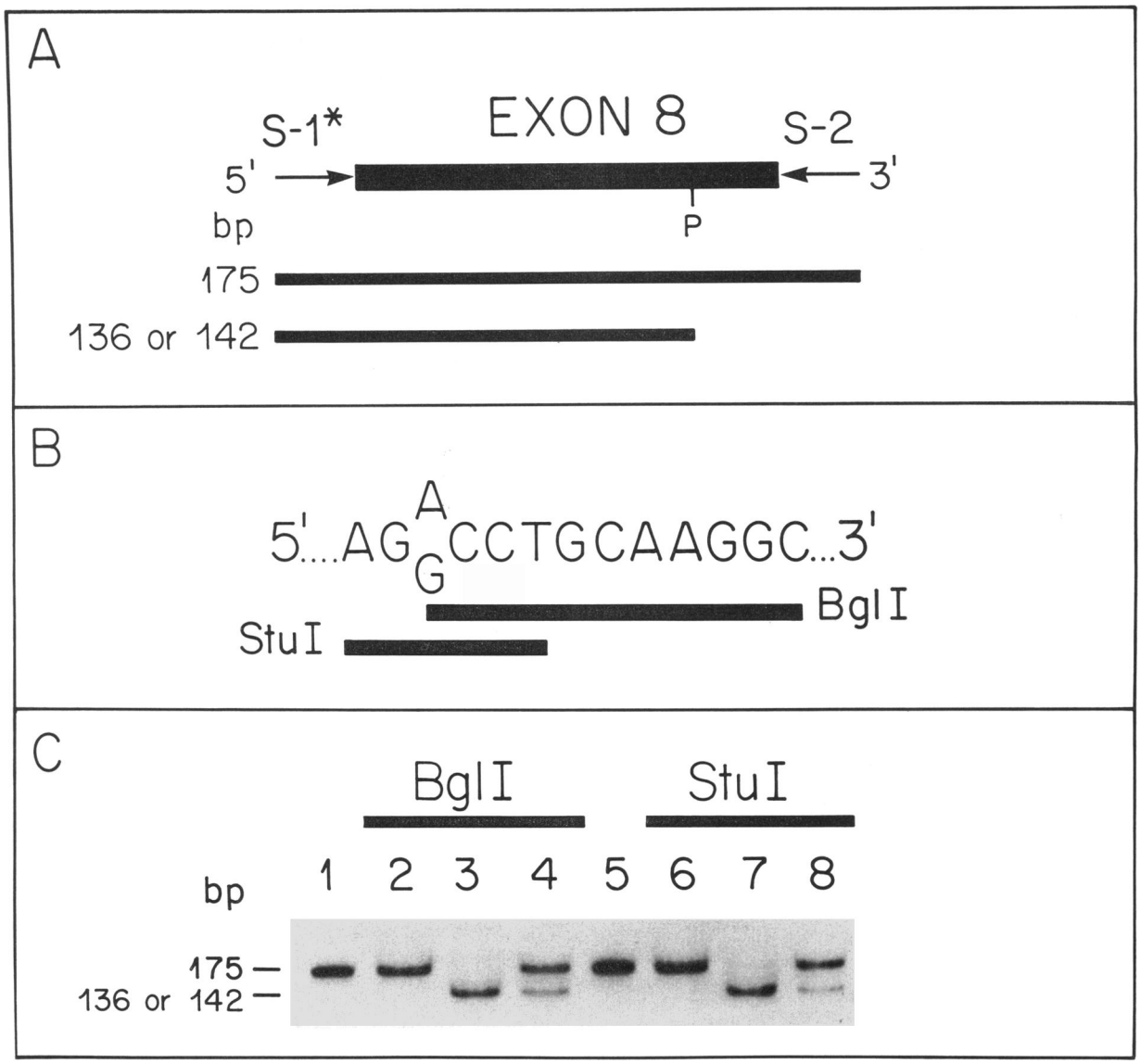


Figure 2 Amplification of exon 8 for analysis of the *StuI* and *BglI* polymorphic site. Panel *A* shows the location of the two PCR oligonucleotides used to amplify exon 8, S-1 (5′-CCAAGCCTCTTTCTCTCTCTCTTCCA-3′) and S-2 (5′-CCACCCGCCGCCTTCCCGTGCTGA-3′). S-1 was end-labeled with [γ-^{32}P] ATP (42) prior to amplification (43). The amplified product was size-fractionated on a nondenaturing 6% polyacrylamide gel in buffer A, and the 175-bp fragment was purified. Panel *B* shows the sequence of exon 8 containing the polymorphic site. Genomic DNA from an individual who was homozygous for the absence of the *StuI* site was selectively amplified and sequenced (Maxam and Gilbert 1980). The sequence was compared with the sequence of an allele in which the *StuI* site was preserved. A guanine-to-adenosine substitution was found within the consensus sequence for both *StuI* and *BglI*, destroying both restriction sites. Approximately 100 ng (300,000 Cerenkov cpm) of PCR-amplified DNA from individuals known to have different genotypes for *StuI* was subjected to either *BglI* (lanes 2–4) or *StuI* (lanes 6–8) digestion. The DNA was size-fractionated on a 6% polyacrylamide nondenaturing gel (panel *C*). The 175-bp undigested amplified DNA fragment is shown in lanes 1 and 5. Amplified DNA from an individual homozygous for the absence of the restriction site also produces the 175-bp fragment shown in lanes 2 and 6. Amplified DNA encoding exon 8 from an individual homozygous for the presence of the restriction site results in a 136-hp fragment when *StuI* is used and a 142-bp fragment when *BglI* is used (lanes 3 and 7). Lanes 4 and 8 show the results obtained for a heterozygous individual.

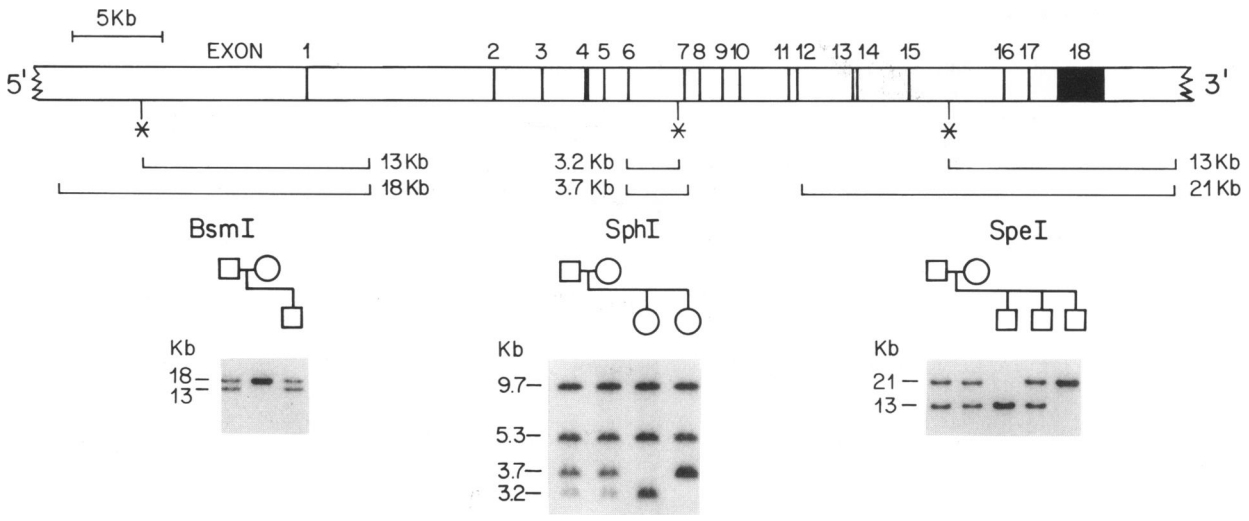


Figure 3 Map of three LDL receptor gene RFLPs. A schematic representation of each polymorphic restriction fragment is illustrated below the physical map of the LDL receptor gene. A Southern blot of a pedigree informative for each RFLP site is shown below the restriction enzyme used for analysis. Eight micrograms of DNA from each family member was digested with the indicated enzyme, subjected to electrophoresis in a 0.8% agarose gel, transferred to a nylon membrane, and then probed with a [α-^{32}P] dCTP–labeled exon-specific probe as described in Methods. The filters were washed and subjected to autoradiography. λ-*Hind*III-digested DNA was used for size markers.

morphic site in exon 8. Exon 8 of the LDL receptor gene was amplified from genomic DNA of an individual homozygous for the absence of the *Stu*I site by using [γ^{32}P] ATP end-labeled PCR oligonucleotides S-1 and unlabeled S-2 (fig. 2A). The amplified fragment was purified and sequenced (Maxam and Gilbert 1980), and a point mutation (guanine to adenosine) was found at nucleotide position 111 in exon 8. This base change is located in the consensus sequence for both *Stu*I and *Bgl*I (fig. 2*B*). These enzymes have overlapping consensus sequences although they are not isoschizomers. Thus, this particular RFLP can be assayed by either *Stu*I or *Bgl*I. Figure 2*C* shows the results of the PCR-amplified fragment following digestion with *Stu*I or *Bgl*I. If the restriction site is present, the fragment sizes expected are 136 and 142 bp for *Stu*I and *Bgl*I, respectively.

Three previously unpublished RFLPs were included in the analysis: *Bsm*I, *Sph*I, and *Spe*I (fig. 3). The *Bsm*I site is located approximately 10 kb 5′ to exon 1. When an exon 1–specific probe is used, the presence and absence of this restriction site results in either a 13-kb or 18-kb restriction fragment. The *Sph*I polymorphic site is located in intron 6 and results in two fragments of 3.7 and 3.2 kb which can be detected by a cDNA-derived probe spanning exons 5–9. *Spe*I reveals a diallelic polymorphism in intron 15 resulting in the formation of a 21- or 12.5-kb fragment detected by a probe derived from exon 18. The frequency of each of these RFLPs in the Caucasian American population was determined by analyzing 123 parental chromosomes. The least frequent alleles for *Bsm*I, *Sph*I, and *Spe*I RFLPs have frequencies of 17%, 41%, and 8%, respectively.

Haplotypes were constructed for 132 individuals of 20 pedigrees by using the 10 RFLPs. A total of 31 unique haplotypes were obtained, and the number and frequency of each are shown in figure 1. The frequency distribution of individual haplotypes ranged from 0.8% to 29.3%. Five of the most common haplotypes (1, 2, 3, 16, and 31) accounted for 67.5% of all chromosomes. Sixteen (≈52%) of the 31 haplotypes appeared only once in the sample. Table 1 presents the characteristics of each RFLP, including its frequency (with standard error), heterozygosity, and PIC value, as estimated from the haplotype data. Individually, the sites are fairly polymorphic, with heterozygosities ranging from .283 to .500, except for the *Stu*I and *Spe*I sites, which display low levels of polymorphism. Taken together, these 10 RFLPs provide a highly polymorphic locus with a cumulative haplotype heterozygosity of 86.3% (PIC value = 85.3%). These values are close to 90% of the maximum variability that is possible for a locus with 31 alleles.

Linkage disequilibrium was studied for each pair of RFLPs by using the Δ statistic (Hill and Robertson 1968). In this analysis, $\Delta = D/[(g_1+g_2)(g_1+g_3)(g_3+g_4)(g_2+g_4)]$, where $D = g_1g_4 - g_2g_3$ and $g_i = n_i/n$ for

Table 1

Frequencies and Polymorphism of 10 RFLPs at the LDL Receptor Gene Locus among 123 Chromosomes

RFLP	Frequency of + Site (Mean ± Standard Error)	Heterozygosity	PIC Value
*Bsm*I	.829 ± .034	.283	.243
*Sph*I	.585 ± .044	.485	.368
*Stu*I	.943 ± .021	.107	.102
*Ava*II	.431 ± .045	.490	.370
*Spe*I	.081 ± .025	.149	.138
*Apa*LI (5′)	.512 ± .045	.500	.375
*Pvu*II	.244 ± .039	.369	.301
*Nco*I	.724 ± .040	.400	.320
*Pst*I	.447 ± .045	.494	.372
*Apa*LI (3′)	.732 ± .040	.393	.316
Overall		.863	.853

$i = 1, \ldots, 4$, where $n_1, \ldots, n_4$ are the numbers of ++, +−, −+, and − haplotypes among n chromosomes (Chakravarti et al. 1984*a*, 1984*b*, 1986). $\Delta = 0$ under random association and can take extreme values of ± 1; for a particular pair of RFLPs, the range of Δ values depends on the observed gene frequencies (Weir 1979). The value of Δ may be positive or negative depending on the relative frequencies of the double heterozygotes ++−− and +−/−+. The significance of Δ (i.e., $\Delta = 0$ or not) can be tested by using $n\Delta^2$, which is distributed as a χ^2 with 1 df. We also computed the relative linkage disequilibrium $D' = D/D_{max}$, where D_{max} is the maximum possible value for D given the observed frequencies at the two RFLPs (Lewontin 1964).

Pairwise Δ and D' values among all RFLPs are presented in table 2. Forty-five pairwise comparisons were made, and the significance of each Δ value was taken at the stringent level of .0011 so that the overall significance level for the 45 tests was 5%. To achieve significance, $\chi^2 = n\Delta^2$ (1 df) has to exceed the value 10.5; since $n = 123$, all $|\Delta|$ values greater than .292 are significant. Of the 45 pairs, 23 are significant. Note that for the two RFLPs with the lowest levels of polymorphism, *Stu*I and *Spe*I, the linkage disequilibrium value is not significant ($\Delta = .073$) but that, since $D' = 1$, the observed Δ value is the maximum value attainable. Furthermore, when *Spe*I is compared with other RFLPs, no significant disequilibria are observed even though $D' = 1$ in five of eight comparisons; for

Table 2

Pairwise Disequilibrium Values (Δ, above the Diagonal) and Relative Degree of Disequilibrium (D′, below the Diagonal)

	Bsm	*Sph*I	*Stu*I	*Ava*II	*Spe*I	*Apa*LI	*Pvu*II	*Nco*I	*Pst*I	*Apa*LI
*Bsm*I		−.163	.448[a]	.220	−.023	.206	−.145	.154	.191	.018
*Sph*I	.426		−.207	.566[a]	.190	.631[a]	−.329[a]	.365[a]	.525[a]	.347[a]
*Stu*I	.828	1.000		.214	.073	.252	−.433[a]	−.152	.221	−.149
*Ava*II	.558	.772	1.000		−.259	.685[a]	−.380[a]	.318[a]	.802[a]	.305[a]
*Spe*I	.035	.759	1.000	1.000		.290	−.169	.184	−.208	.180
*Apa*LI	.442	.732	1.000	.807	1.000		−.506[a]	.524[a]	.780[a]	.510[a]
*Pvu*II	.181	.487	1.000	.768	1.000	.870		.309[a]	−.358[a]	.344[a]
*Nco*I	.210	.498	1.000	.590	1.000	.828	.879		.410[a]	.816[a]
*Pst*I	.468	.693	1.000	.829	.776	.888	.702	.737		.434[a]
*Apa*LI	.024	.482	1.000	.578	1.000	.823	1.000	.832	.797	

[a] Significant at $\alpha \leq .0011$.

*Stu*I, two of nine Δ values are significantly different from zero, while eight of nine of these comparisons have $D' = 1.0$. The moderately polymorphic *Bsm*I site also fails to demonstrate significant linkage disequilibrium when Δ or D' is used, except with the *Stu*I site. With *Bsm*I, D' values range between .04 and .56 with all sites except the *Stu*I site, where $D' = .83$. If the low to moderately polymorphic sites (*Bsm*I, *Stu*I, and *Spe*I) are considered as a group A, and the remainder as a group B, then, within groups A and B, one of three comparisons and 21 of 21 comparisons, respectively, are significant; also, only 1 of 21 comparisons between groups A and B are significant. Thus, in our data, loci that have low polymorphism have disequilibrium values (D) that are low relative to the absolute value (Δ) but high relative to the maximum possible value (D'). Thus, the lack of linkage disequilibrium between some RFLPs as measured by Δ correlates with their low polymorphism content and, as judged by the D' values, suggests high linkage disequilibrium throughout the locus.

An alternative global test of linkage disequilibrium for the locus was performed. By simulation methods (Blanton and Chakravarti 1987; Antonarakis et al. 1988), we estimated the distribution of the number of different haplotypes expected under linkage equilibrium if 123 chromosomes were sampled and the individual RFLP frequencies were as observed (table 1). By simulation, a total of 250 samples were generated, and these contained 77–93 haplotypes, with an average of 85.6 ± 3.0 haplotypes. Since only 31 haplotypes were actually observed in our sample, the simulation results imply that there is marked linkage disequilibrium at this locus.

Since extensive linkage disequilibrium exists, not all 10 RFLPs are useful for genetic analysis. In table 3 we present the order in which RFLPs should be chosen in order to maximize the informativeness. These calculations were made with the algorithm of Chakravarti and Buetow (1985). The 5′ *Apa*LI site has a heterozygosity value of .4997 and should be the first site analyzed. *Pst*I is the next most polymorphic site; however, linkage disequilibrium between *Apa*LI 5′ and *Pst*I ($\Delta = .870$) is greater than that between *Apa*LI 5′ and *Sph*I ($\Delta = .631$), and thus *Sph*I is the next most useful site. Addition of the 3′ *Apa*LI, *Ava*II, and *Bsm*I sites sequentially increases the heterozygosity index to a cumulative value of .8442. These data indicate that analysis of these five RFLPs provides the highest information content and that the additional sites listed in figure 1 are of limited value, except for selected pedigrees where these RFLPs may be useful.

Table 3

LDL Receptor Gene RFLP Site Selection

RFLP Site	Location of Site	Cumulative Heterozygosity Index
*Apa*LI (5′)	Intron 15	.4997
*Sph*I	Intron 6	.6446
*Apa*LI (3′)	3′ Flanking	.7378
*Ava*II	Exon 13	.8071
*Bsm*I	5′ Flanking	.8442
*Pst*I	Exon 18	.8550
*Nco*I	Exon 18	.8623
*Pvu*II	Intron 15	.8628
*Spe*I	Intron 15	.8631
*Stu*I	Exon 8	.8632

Our sample contains 31 haplotypes for 10 segregating sites. This would be unexpected if recombination were rare within this region, since in the absence of recombination each RFLP can delineate at most one unique haplotype. Of course, multiple mutations at each RFLP could produce further unique haplotypes. The problem of multiple mutations at polymorphic restriction sites has been addressed by R. R. Hudson, (personal communication; see also Antonarakis et al. 1988). Hudson shows that for a Wright-Fisher selectively neutral model with fixed population size N at genetic equilibrium, identical mutation rates (u) for all nucleotide sites, and no recombination, this probability (P) depends on three quantities: n, the sample size of haplotypes; r, the number of base pairs in the recognition site for a restriction endonuclease; and $\theta = 4Nu$. P is an increasing function of all three quantities, r, n, and θ. If $r = 6$, $n = 100$, and $\theta = .005$, then $P = .113$. The θ value has been estimated to be a maximum of .005, from analysis of several human gene loci (e.g. see Chakravarti et al. 1986). Thus, if all such multiple mutations created new haplotypes, we would expect approximately 12 haplotypes in our samples; the observed number of 31 is clearly in excess.

An alternative way to visualize the evolution of the haplotypes is to construct a phylogenetic tree or network (unrooted tree) of haplotypes. In figure 4 we present such a tree, constructed in accordance with the principle of maximum parsimony, and with the assumption that each RFLP and each haplotype evolves in an independent manner. The computer program PAUP (Swofford 1985) was used to construct a tree requiring the fewest number of mutations. In figure 4 each link between the indicated haplotypes requires at least one

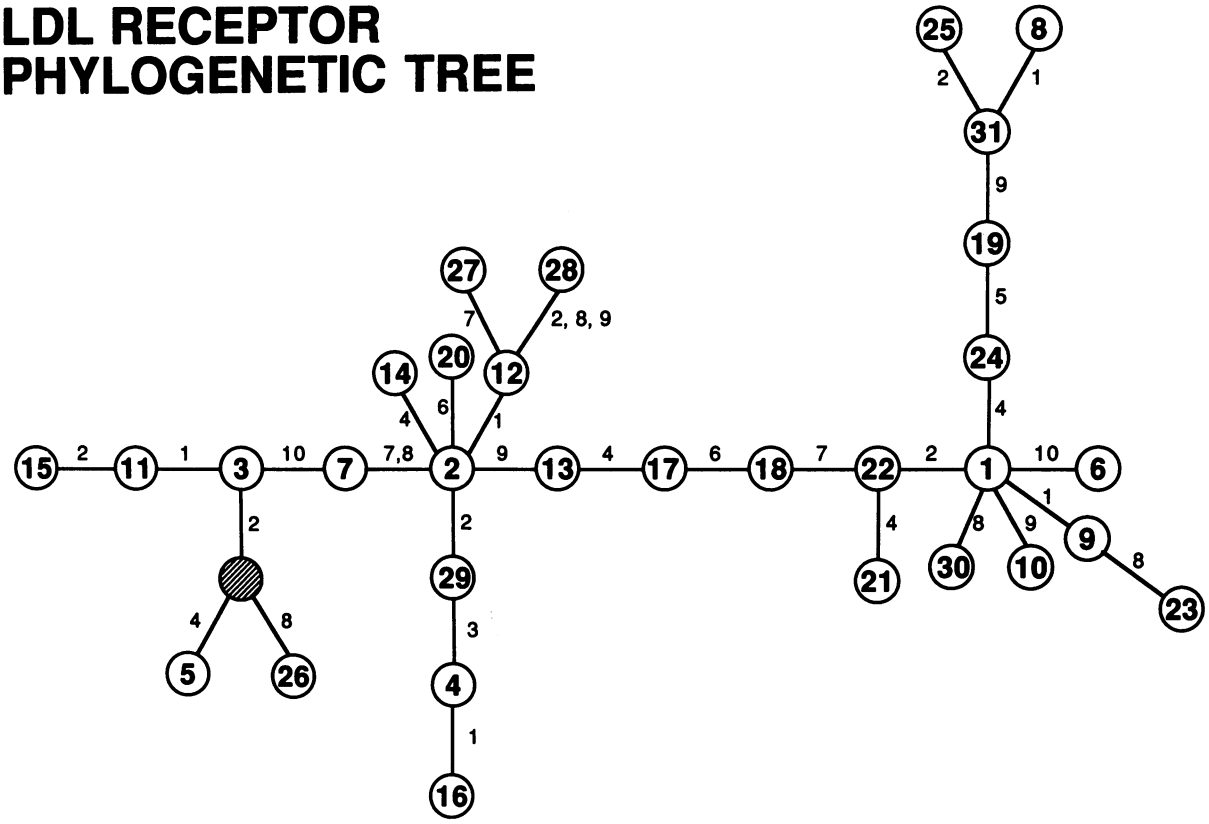


Figure 4 A possible phylogeny for LDL receptor haplotypes. The phylogenetic tree was constructed, using maximum parsimony and the computer program PAUP (Swofford 1985). Each circle represents an individual haplotype with the number corresponding to the haplotype number in fig. 1. The shaded circle is an intermediate haplotype presumed to occur but not observed. The numbers associated with each link in the network represent the restriction sites that have mutated. The numbers correspond to the following restriction sites: 1 = *Bsm*I; 2 = *Sph*I; 3 = *Stu*I; 4 = *Ava*II; 5 = *Spe*I; 6 = *Apa*LI(5′); 7 = *Pvu*II; 8 = *Nco*I; 9 = *Pst*I; 10 = *Apa*LI(3′).

mutation. According to our data, this tree requires 34 mutations; that is, on average, each RFLP has mutated 3.4 times. Specifically, the number of mutations at each RFLP in a 5′-to-3′ direction (i.e., from *Bsm*I to the 3′ *Apa*LI site) were 5, 6, 1, 5, 1, 2, 3, 5, 4, and 2. The large number of multiple changes seen at each RFLP would be consistent with each of these RFLPs being a mutational hot spot. On the other hand, a more plausible explanation would be the occurrence of recombination events, since this would put an RFLP on different haplotypes and simulate the effect of multiple mutations.

In all probability both multiple mutations and recombination events are necessary to explain the history of the haplotypes. Once a polymorphism is established, either event would cause us to observe discordant haplotypes. For a pair of RFLPs, if ++ were ancestral and polymorphism were created, the haplotypes +− and −+, or +− and −−, or −+ and −− would appear; the fourth remaining haplotype could only be created by recombination or further mutation. Candidates for ancestral haplotypes may be identified as those that are polymorphic for all 10 sites and contain no discordancies. Under the assumption of selective neutrality, the age of an allele is, on average, directly proportional to its frequency (Watlerson and Guess 1977). Thus, ancestral haplotypes may be more common than derived haplotypes. Starting with the most common haplotypes, we searched for candidate ancestral types. If a selected set of haplotypes did not satisfy the properties of being ancestral, each haplotype was replaced by one from the remainder until the properties of being ancestral were met. This analysis shows that haplotypes 1, 3, 16, 24, and 31 were ancestral; these haplotypes are polymorphic for all 10 sites and contain no discordancies. The remaining 26 haplotypes can be explained by multiple

mutation and recombination events. Surprisingly, the frequent haplotype 2 appears to be a recombinant between the ancestral haplotypes 3 and 16, whereas an infrequent haplotype (24) appears to be ancestral.

Discussion

In this study we have determined the LDL receptor gene haplotypes in 132 Caucasian American individuals from 20 different families. A total of 31 different haplotypes ranging in frequency from 0.8% to 29.3% were identified. Association studies between the RFLP sites demonstrate a high degree of linkage disequilibrium encompassing the entire locus. No region of linkage equilibrium was found to suggest a recombinational hot spot, as has been observed in both the human insulin (Chakravarti et al. 1986) and β-globin loci (Chakravarti et al. 1984*a*). Within these two genes there are localized regions where clusters of linked RFLPs on either side of the hot spot demonstrate linkage equilibrium with each other. It has been hypothesized that *Alu* sequences (Schmid and Jelinek 1982) are hot spots for recombinational events at the LDL receptor locus (Lehrman et al. 1987*a;* Russell et al, in press). *Alu* sequences are highly redundant 300-bp sequences composed of two tandem repeats (Schmid and Jelinek 1982) which are estimated to compose 9% of the human genome. The LDL receptor gene contains many *Alu* sequences in its introns and flanking sequences as well as in the 3′ noncoding region (Yamamoto et al. 1984; Lehrman et al. 1987*a*), and every deletion or insertion at the LDL receptor gene that has been characterized at the molecular level to date has involved at least one *Alu* sequence. Linkage analysis of 10 RFLPs spanning the locus disclosed no localized regions of linkage equilibrium flanking any of the known *Alu* sequences as might be anticipated.

The Δ values correlate well with the physical distance between RFLP sites, with one exception: the *Bsm*I site located 5′ to the gene was not in linkage disequilibrium with its next closest site, *Sph*I, though it was in linkage disequilibrium with the more distant site, *Stu*I. Precedence for this exists at the apolipoprotein AI-CIII-AIV and D11S12 loci, where it has been observed that two RFLPs are in linkage disequilibrium with each other but not with an intervening site (Barker et al. 1984; Hiroko et al. 1987; Thompson et al. 1988). The likely explanation for all of these observations is that the sample sizes were too small and the polymorphisms too infrequent to accurately determine linkage between these particular sites.

The high degree of linkage disequilibrium between the RFLPs at the LDL receptor locus reduces their informativeness in pedigree analysis. However, the appropriate analysis of selected sites can be highly informative. In the population studied, analysis of five RFLPs was almost as informative as an analysis of all 10 sites. Efficient evaluation of a pedigree can be performed by sequential analysis of each RFLP site according to table 1 until the haplotypes of the four parental alleles are differentiated.

As has been shown for β-[thalassemia and phenylketonuria (Orkin et al. 1982; Orkin and Kazazian 1984; DiLella et al. 1986, 1987; Chakraborty et al. 1987), haplotype analysis can facilitate the molecular definition of specific mutant alleles in diseases where there is allelic heterogeneity. Functional and biosynthetic studies of the LDL receptor protein in cultured fibroblasts from FH homozygotes have demonstrated four distinct classes of receptor mutants (Tolleshaug et al. 1982, 1983). LDL receptor haplotype analysis has assisted in differentiating and identifying particular mutant alleles for molecular characterization. For example, LDL receptor haplotype analysis of the subset of FH homozygotes who have no immunoprecipitable LDL receptor protein enabled differentiation of alleles that were indistinguishable by biochemical analysis (Hobbs et al. 1988). As has been observed at the β-globin and phenylalanine hydroxylase loci, alleles with different haplotypes have been found to be associated with different mutations, particularly if the haplotype is of low frequency in the general population. As well, haplotype analysis can be used to distinguish compound heterozygotes from "true" homozygotes. In the Caucasian American population, if all 10 sites are analyzed, the heterozygosity is 86.3%. Therefore, if an FH homozygote is homozygous for a rare haplotype, it is likely that he or she is homozygous for the same mutation. Identification of the mutation by sequence analysis of either cloned or PCR-amplified DNA (Saiki et al. 1988) is greatly simplified if both alleles are identical.

We have shown that LDL receptor haplotype analysis using 10 RFLP sites dramatically improves the informativeness of this locus. However, one haplotype (no. 1) accounts for 29.3% of LDL receptor alleles and it can be expected that up to 8.6% of individuals will be homozygous for this haplotype. Any additional RFLPs found within the LDL receptor gene will likely be in linkage disequilibrium with the currently known sites and thus will not increase the informativeness of the analysis. Therefore, efforts are underway to locate RFLPs in the distant flanking regions of the LDL recep-

tor gene. It is to be hoped that identification of a few additional RFLPs will allow differentiation of alleles that are currently indistinguishable.

Acknowledgments

We thank Drs. M. S. Brown, J. L. Goldstein, and D. W. Russell for helpful discussions. Shellie Craig, Irena Pankiewicz, and Pam St-Jean are thanked for their valuable technical assistance. This study was supported by research grants HL20948 and GM33771 from the National Institutes of Health. H.H.H. is supported by the Syntex Scholar Program and E.L. is the recipient of fellowship IF-05-TW03742 from the Fogarty International Center of the National Institutes of Health Research. A.C. is the recipient of NIH Research Career Development Award HD00774.

References

Antonarakis, S. E., P. Oettgen, A. Chakravarti, S. L. Halloran, R. R. Hudson, L. Feisee, and S. K. Karathanasis. 1988. DNA polymorphism haplotypes of the human apolipoprotein APOA1-APOC3-APOA4 gene cluster. Hum. Genet. **80**:265–273.

Barker, D., T. Holm, and R. White. 1984. A locus on chromosome 11p with multiple restriction site polymorphisms. Am. J. Hum. Genet. **36**:1159–1171.

Blanton, S. H., and A. Chakravarti. 1987. A global test of linkage disequilibrium. Am. J. Hum. Genet. [Suppl.] **41**: A250.

Block, W. D., K. J. Jarret, and J. B. Levine. 1965. Use of a single color reagent to improve the automated determination of serum total cholesterol. Pp. 345–247 *in* Technicon Symposia; Automation in Analytical Chemistry. Technicon, Tarrytown, NY.

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

Breslow, J. L. 1988. Apolipoprotein genetic variation and human disease. Physiol. Rev. **68**:85–132.

Brown, M. S., and J. L. Goldstein. 1986. A receptor-mediated pathway for cholesterol homeostasis. Science **232**:34–47.

Chakraborty, R., A. S. Lidsky, S. P. Daiger, F. Guttler, S. Sullivan, A. G. DiLella, and S. L. C. Woo. 1987. Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationship with phenylketonuria. Human Genetics **76**:40–46.

Chakravarti, A., K. H. Buetow, S. W. Antonarakis, P. G. Waber, C. D. Boehm, and H. H. Kazazian. 1984*a*. Non-uniform recombination within the human β-globin gene cluster. Human Genet. **36**:1239–1258.

Chakravarti, A., J. A. Phillips III, K. H. Mellits, K. H. Buetow, and P. H. Seeburg. 1984*b*. Patterns of polymorphism and linkage disequilibrium suggest independent origins of the human growth hormone gene cluster. Proc. Natl. Acad. Sci. USA **81**:6085–6089.

Chakravarti, A., and K. H. Buetow. 1985. A strategy for using multiple linked markers for genetic counseling. Am. J. Hum. Genet. **37**:984–997.

Chakravarti, A., S. C. Elbein, and M. A. Permutt. 1986. Evidence for increased recombination near the human insulin gene: implication for disease association studies. Proc. Natl. Acad. Sci. USA **83**:1045–1049.

Church, G. M., and W. Gilbert. 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.

Davis, C. G., M. A. Lehrman, D. W. Russell, R. G. W. Anderson, M. S. Brown, and J. L. Goldstein. 1986. The J. D. mutation in familial hypercholesterolemia: substitution of cysteine for tyrosine in cytoplasmic domain impedes internalization of LDL receptors. Cell **45**:15–24.

DiLella, A. G., J. Marvit, A. S. Lidsky, F. Guttler, and S. L. C. Woo. 1986. Tight linkage between a splicing mutation and a specific DNA haplotype in phenylketonuria. Nature **322**:799–805.

DiLella, A. G., J. Marvit, K. Brayton, and S. L. C. Woo. 1987. An amino-acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype 2. Nature **327**:333–336.

Funke, H., J. Klug, P. Frossard, R. Coleman, and G. Assmann. 1986. *Pst*I RFLP close to the LDL receptor gene. Nucleic Acids Res. **14**:7820.

Goldstein, J. L., and M. S. Brown. 1983. Familial hypercholesterolemia. Pp. 672–712 *in* J. B. Stanbury, J. B. Wyngaarden, D. S. Fredrickson, J. L. Goldstein, and M. S. Brown, eds. The metabolic basis of inherited disease. 5th ed. McGraw-Hill, New York.

Goldstein, J. L., H. G. Schrott, W. R. Hazzard, E. L. Bierman, and A. G. Motulsky. 1973. Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. J. Clin Invest. **52**:1544–1568.

Hill, W. G., and Robertson, A. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38**:226–231.

Hiroko, P., D. Galton, and J. Stocks. 1987. DNA polymorphic patterns and haplotype arrangements of apo A-1, apo C-III, apo A-IV gene cluster in different ethnic groups. Hum. Genet. **75**:264–268.

Hobbs, H. H., M. S. Brown, J. L. Goldstein, and D. W. Russell. 1986. Deletion of exon encoding cysteine-rich repeat of LDL receptor alters its binding specificity in a subject with familial hypercholesterolemia. J. Biol. Chem. **261**: 13114–13120.

Hobbs, H. H., M. S. Brown, D. W. Russell, J. Davignon, and J. L. Goldstein. 1987*a*. Deletion in the gene for the LDL receptor in majority of French Canadians with familial hypercholesterolemia. N. Engl. J. Med. **317**:734–737.

Hobbs, H. H., V. Esser, and D. W. Russell. 1987*b*. *Ava*II polymorphism in the human LDL receptor gene. Nucleic. Acids. Res. **15**:379.

Hobbs, H. H., M. A. Lehrman, T. Yamamoto, and D. W. Russell. 1985. Polymorphism and evolution of *Alu* sequences in the human low density lipoprotein receptor gene. Proc. Natl. Acad. Sci. USA **82**:7651–7655.

Hobbs, H. H., E. Leitersdorf, J. L. Goldstein, M. S. Brown, and D. W. Russell. 1988. Multiple crm- mutations in familial hypercholesterolemia: evidence for 13 alleles, including four deletions. J. Clin. Invest. **81**: 909–917.

Horsthemke, B., U. Beisiegel, A. Dunning, J. R. Havinga, R. Williamson, and S. Humphries. 1987. Unequal crossing-over between two *Alu*-repetitive DNA sequences in the low density lipoprotein receptor gene: a possible mechanism for the defect in a patient with familial hypercholesterolemia. Eur. J. Biochem. **164**:77–81.

Humphries, S. E., A. M. Kessling, B. Horsthemke, J. A. Donald, M. Seed, N. Jowett, M. Holm, D. J. Galton, V. Wynn, and R. Williamson. 1985. A common DNA polymorphism of the low-density lipoprotein (LDL) receptor gene and its use in diagnosis. Lancet **1**:1003–1005.

Kan, Y. W. 1987. New application for DNA polymorphism. N. Engl. J. Med. **316**:478–480.

Kan, Y. W., and A. M. Dozy. 1978. Polymorphism of DNA sequence adjacent to human β-globin structural gene: relationship to sickle mutation. Proc. Natl. Acad. Sci. USA **75**:5631–5635.

Kessler, G., and H. Lederer. 1965. Fluorometric measurement of triglycerides. Pp. 341–344 *in* Technicon Symposia: Automation in Analytical Chemistry. Technicon, Tarrytown, NY.

Kotze, M. J., E. Langenhoven, E. Dietzsch, and A. E. Retief. 1987. A RFLP associated with the low-density lipoprotein receptor gene (LDLR). Nucleic Acid. Res. **15**:376.

Kotze, M. J., A. E. Retief, P. A. Brink, and H. F. H. Welch. 1986. A DNA polymorphism in the human low-density lipoprotein receptor gene. S. Afr. Med. J. **70**:77–79.

Labarea, C., and K. Paigen. 1980. A simple, rapid and sensitive DNA assay procedure. Anal. Biochem. **102**:344–352.

Lehrman, M. A., J. L. Goldstein, M. S. Brown, D. W. Russell, and W. J. Schneider. 1985*a*. Internationalization-defective LDL receptors produced by genes with nonsense and frameshift mutations that truncate the cytoplasmic domain. Cell **41**:735–743.

Lehrman, M. A., J. L. Goldstein, D. W. Russell, and M. S. Brown. 1987*a*. Duplication of seven exons in LDL receptor gene caused by *Alu-Alu* recombination in a subject with familial hypercholesterolemia. Cell **48**:827–835.

Lehrman, M. A., D. W. Russell, J. L. Goldstein, and M. S. Brown. 1986. Exon-*Alu* recombination deletes 5 kilobases from low density lipoprotein receptor gene, producing a null phenotype in familial hypercholesterolemia. Proc. Natl. Acad. Sci. USA **83**:3679–3683.

———. 1987*b*. *Alu-Alu* recombination deletes splice acceptor sites and produces secreted LDL receptor in a subject with familial hypercholesterolemia. J. Biol. Chem. **262**: 3354–3361.

Lehrman, M. A., W. J. Schneider, M. S. Brown, C. G. Davis, A. Elhammer, D. W. Russell, and J. L. Goldstein. 1987*c*. The Lebanese allele at the LDL receptor locus: nonsense mutation produces truncated receptor that is retained in endoplasmic reticulum. J. Biol. Chem. **262**:401–410.

Lehrman, M. A., W. J. Schneider, T. C. Südhof, M. S. Brown, J. L. Goldstein, and D. W. Russell. 1985*b*. Mutation in LDL receptor: *Alu-Alu* recombination deletes exons encoding transmembrane and cytoplasmic domains. Science **227**:140–146.

Leitersdorf, E., and H. H. Hobbs. 1987. Human LDL receptor gene: two *Apa*LI RFLPs. Nucleic Acids Res. **15**:2782.

———. 1988. *Hin*cIII polymorphism detected by gene amplification. Nucleic Acids Res. **15**:2782.

Lewontin, R. C. 1964. The interaction of selection and linkage. I. General considerations of heterotic models. Genetics **49**:49–64.

Lipid Research Clinic Program. 1982. Lipid and lipoprotein analysis. Manual of laboratory operations. Department of Health, Education, and Welfare publ. NIH/75–628. Government Printing Office, Washington, DC.

Maniatis, T., E. F. Fritsch, and J. Sambrook, eds. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with basic-specific chemical cleavages. Methods. Enzymol. **65**:499–560.

Orkin, S. H., and H. H. Kazazian, Jr. 1984. The mutation and polymorphism of the human β-globin gene and its surrounding DNA. Annu. Rev. Genet. **18**:131–171.

Orkin, S H., H. H. Kazazian, Jr., S. E. Antonarakis, S. C. Goff, C. D. Boehm, J. P. Sexton, P. G. Waber, and P. J. V. Giardina. 1982. Linkage of β-thalassaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in the human β-globin gene cluster. Nature **296**:627–631.

Russell, D. W., V. Esser, and H. H. Hobbs. The molecular basis of familial hypercholesterolemia. Atherosclerosis (in press).

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **230**:487–491.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**: 1350–1354.

Schmid, C. W., and W. R. Jelinek. 1982. The *Alu* family of dispersed repetitive sequences. Science **216**:1065–1070.

Swofford, D. L. 1985. Phylogenetic analysis using parsimony (PAUP), version 2.3. Illinois Natural History Survey. Champaign.

Thompson, E. A., S. Deeb, D. Walker, and A. G. Motulsky. 1988. The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am. J. Hum. Genet. **42**:113–124.

Tolleshaug, H., J. L. Goldstein, W. J. Schneider, and M. S.

Brown. 1982. Posttranslational processing of the LDL receptor and its genetic disruption in familial hypercholesterolemia. Cell **30**:715–724.

Tolleshaug, H., K. K. Hobgood, M. S. Brown, and J. L. Goldstein. 1983. The LDL receptor locus in familial hypercholesterolemia: multiple mutations disrupting the transport and processing of a membrane receptor. Cell **32**: 941–951.

Wainscoat, J. S., J. I. Bell, S. L. Thein, D. R. Higgs, G. R. Sarjeant, T. E. A. Peto, and D. J. Weatherall. 1983. Multiple origins of the sickle mutation: evidence from the β^S globin gene cluster polymorphisms. Mol. Biol. Med. **1**:191–197.

Watlarson, G. A., and H. A. Guess. 1977. Is the most frequent allele the oldest? Theor. Popul. Biol. **11**:141–160.

Weir, B. S. 1979. Inferences about linkage disequilibrium. Biometrics **35**:235–254.

Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell. 1984. The human LDL receptor: a cysteine-rich protein with multiple *Alu* sequences in its mRNA. Cell **39**:27–38.