Linkage Analysis for Monogenic Traits

6

Arno G. Motulsky and Michael Dean

Revised by M.R. Speicher

Abstract Linkage analysis, that is the observation of cosegregation of adjacent genetic markers or traits, is the principal means of constructing genetic maps, and locating genes that cause disease, or genetic traits. Although these methods have been partially supplanted by newer methods such as whole genome sequencing, linkage analysis still has considerable utility. The history of the method and the mathematical basis of linkage analysis are presented as well as specific applications to human genetics. Gene clusters consist of groups of adjacent genes that exist largely though mechanisms of gene duplication. In addition there exist clusters of genes that have related function, the best studied of which is the Major Histocompatibility Complex. The structure and evolutionary history of these clusters provides insight into the history of mammalian genomes.

Contents

> If "to take a possible example, an equally close linkage" (as between the genes for hemophilia and color blindness) "were found between the genes for blood group" and that "determining Huntington's chorea, we should be able, in many cases, to predict which children of an affected person would develop this disease and to advise on the desirability or otherwise of their marriage."
>
> J.B.S. Haldane and J. Bell (1937) The linkage between the genes for colour-blindness and haemophilia in man. Proc. Roy. Soc. B 123, 119.

6.1 Linkage: Localization of Genes on Chromosomes

Genes are located in a linear fashion on the chromosomes. This has the logical consequence that genes located on the same chromosome are transmitted

M. Dean (✉)
National Cancer Institute, Frederick, MD 21702, USA
e-mail: deanm@mail.nih.gov

M.R. Speicher et al. (eds.), *Vogel and Motulsky's Human Genetics: Problems and Approaches*,
DOI 10.1007/978-3-540-37654-5_6, © Springer-Verlag Berlin Heidelberg 2010

together, i.e., that their segregation is not independent. On the other hand, it is known from cytogenetics that chiasmata are formed during the first meiotic division, and that certain chromosome segments are exchanged between homologous chromosomes (crossing over; see Sect. 3.5). Hence, even genes located on the same chromosome are not always transmitted together; the probability of transmission of two linked genes depends on the distance between them and on the frequency with which they are separated by crossing over. If located on a fairly long chromosome, and if the distance is large enough that numerous crossing over events occur between them, genes located on the same chromosome may even seem to segregate independently. Such genes are syntenic but not linked. It was the great achievement of Morgan and his school in the first two decades of the twentieth century to exploit linkage for localizing genes relative to each other on chromosomes and developing gene maps in the fruit fly *Drosophila melanogaster*.

Studies on linkage and gene mapping in humans lagged behind this development for decades. Sophisticated statistical techniques were developed to get around the difficulty that directed breeding experiments are impossible in humans, and information from naturally occurring families must be used. The application of such techniques, however, was only sparsely rewarded by detection of linkage. A breakthrough occurred only when the new techniques of somatic cell genetics and especially cell fusion were introduced. These techniques permitted the assignment of genes to specific chromosomes and even chromosome segments. Later, methods taken from molecular biology, especially the development of restriction fragment length polymorphisms (RFLPs) revolutionized this progress [9]. Radiation hybrid methods and YAC and BAC clones provided the first high-resolution physical maps and paved the way to the complete sequence of the human genome [50, 104]. Annotation of the human genome sequence has allowed the location of all the known genes to be determined.

In the following we describe first the principle of the classic approach to gene localization as introduced by Morgan and his followers. This provides an opportunity to introduce some general concepts. We then discuss statistical methods for detecting and measuring linkage in humans. The various groups of DNA markers are described next, followed by the principle of cell fusion and its use in localizing genes on chromosomes, as well as the application of radioactive and nonisotopic in situ hybridization for this purpose. Genetic maps are compared to physical maps, and the use of linkage studies as analytical tools in genetic analysis of common diseases with complex etiology and pathogenesis is assessed.

6.1.1 Classic Approaches in Experimental Genetics: Breeding Experiments and Giant Chromosomes

According to Mendel's third law, segregation of two different pairs of alleles is independent; all possible zygotes of two pairs of alleles are formed by free recombination. Mating between the double heterozygote AaBb and the double homozygote aabb leads to:

Paternal gametes	AB	Ab	aB	ab
Maternal gametes ab	$^1/_4$ AaBb	$^1/_4$ Aabb	$^1/_4$ aaBb	$^1/_4$ aabb

The four genotypes are formed in equal proportions.

Soon after Mendel's laws were rediscovered Bateson et al. [4] found an exception from this rule in the vetch, *Lathyrus odoratus*. Certain combinations were observed more frequently and others less frequently than expected. In some cases, the two parental combinations – in our example AaBb (father) and aabb (mother) – were increased among the progeny; in other cases the two other types Aabb or aaBb were more frequent.

Paternal gametes	AB	Ab	aB	ab
Maternal gametes ab	AaBb	Aabb	aaBb	aabb
First case (coupling)	$\frac{1}{2} - \Theta$	Θ	Θ	$\frac{1}{2} - \Theta$
Second case (repulsion)	Θ	$\frac{1}{2} - \Theta$	$\frac{1}{2} - \Theta$	Θ

Θ = Recombination fraction Θ.

The alleles of the parental combination seemed either to attract one another or to repel one another. Bateson et al. [4] coined the terms "coupling" for the former phase and "repulsion" for the latter. Morgan in [65] recognized that coupling and repulsion are two aspects of the same phenomenon (i.e., location of two genes on the same or homologous chromosomes). He coined the term "linkage." Coupling occurs when the genes A and B are localized in the doubly heterozygous parent on the same chromosome $\frac{AB}{ab}$, and repulsion occurs when they are localized on homologous chromosomes $\frac{Ab}{aB}$. The terms *cis* and *trans* are more frequently used to refer to genes in coupling or repulsion, respectively. If linkage is complete, only two types of progeny can occur. More frequently, however, all four types are found, albeit two types in smaller numbers. Morgan explained this finding by exchange of chromosome pieces between homologous chromosomes during meiotic crossing over. He also recognized that the frequency of crossing over depends on the distance between two gene loci in one chromosome. Using recombination analysis as an analytic tool, he and his coworkers succeeded in locating a great number of gene loci in *Drosophila* and in establishing chromosome maps. Their results were confirmed in the early 1930s when Heitz, Bauer, and Painter discovered the giant chromosomes of some *Dipterae.* With this experimental tool many gene localizations known from indirect evidence could be confirmed by direct inspection when they were accompanied by small structural chromosomal variation. In the meantime linkage analyses have been carried out in a great number of species.

6.1.1.1 Linkage and Association

It is sometimes assumed that genes which are linked should always show a certain association in the population, i.e., that the chromosomal combinations AB or ab (coupling) occur more frequently than the combinations Ab or aB (repulsion). However, this is not the case in a randomly mating population. Even if the linkage is fairly close, repeated crossing over in many generations causes all four combinations, AB, ab, Ab, and aB, to be randomly distributed in the long run. As a rule, association of genetic traits does not point to linkage. This rule, however, has exceptions. Some combinations of closely linked genes do indeed occur more often than expected with random distribution. Such "linkage disequilibrium" was first postulated in humans for the rhesus blood types (Sect. 6.2.4) and has also been proven for the major histocompatibility complex (MHC), especially the HLA system (Sect. 6.2.5) and for many DNA polymorphisms. It has now been shown that there are blocks of linkage disequilibrium throughout the genome and these blocks vary by location and by population [25] (Sect. 16.3). Linkage disequilibrium may occur for three reasons:

1. The population under examination originated from a mixture of two populations with different frequencies of the alleles A,a and B,b, and the time elapsed since the mixing of the populations was not sufficient for complete randomization (admixture linkage disequilibrium).
2. Two mutants, for example, DNA markers, are located so closely together that an insufficient number of generations has elapsed to separate them by recombination since the two mutations occurred in one chromosome.
3. Certain combinations of alleles at linked gene loci are maintained in high frequency by natural selection.

These problems are discussed in greater detail in connection with the MHC system (Sect. 6.2.5) and in the discussion on association between HLA and disease (Sect. 6.2.5.4).

6.1.2 Linkage Analysis in Humans

6.1.2.1 Direct Observation of Pedigrees

Linkage analysis by classic methods in humans is difficult since no directed breeding occurs. However, in some cases pedigree inspection can provide information. Linkage is excluded, for example, if one of the genes under scrutiny can be localized to the X chromosome while the other is on an autosome. By the same token, there is a high probability of demonstrating formal linkage if both genes are X-linked. Even in this case, however, formal linkage may not be demonstrable since the loci may be so far from each other that crossing over separates them. Similar considerations hold for genes located on a given autosomal chromosome. The term synteny refers to two or more genes

being situated on the same chromosome, regardless of whether formal linkage can be demonstrated. Either a large pedigree or a number of smaller pedigrees must be screened to assess the extent of crossing over. Figure 6.1a shows a pedigree with red–green color blindness (303800, 303900) and hemophilia (306700). The males in the sibships at risk either have both conditions or are normal. The genes are in the coupling (or *cis*) state. The pedigree in Fig. 6.1b shows the opposite; here these genes are in the repulsion (or *trans*) phase.

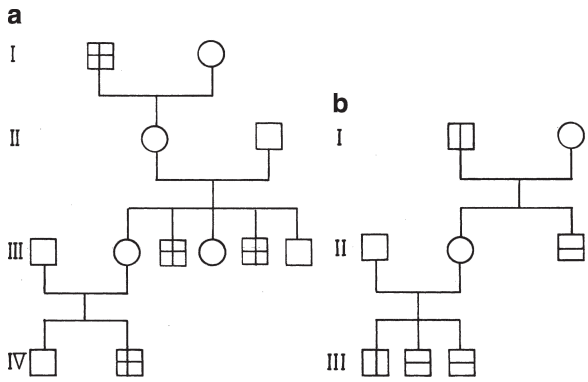

Fig. 6.1 (**a**, **b**) Pedigrees with red-green color blindness (⊟), hemophilia (◫), or both conditions (⊞). (**a**) Both abnormal genes in coupling. (From Madlener 1928) (**b**) In repulsion. From Stern [98]

In some exceptional cases linkage between gene loci localized on an autosome can also be established by simple inspection of an extensive pedigree. Figure 6.2 shows a large pedigree in which Huntington disease segregates together with a *Hin*dIII DNA polymorphism detected by a DNA marker, which was named "G 8" [29]. Four allelic variants of this probe are observed in this pedigree, A, B, C, and D. The Huntington gene invariably segregates together with allele C. One individual, VI, 5 (arrow), so far has been unaffected by Huntington disease, but she will be affected later, provided that her father, (who has not been tested), does not happen to have transmitted another chromosome that carries a C allele not linked to the Huntington gene. The pedigree points to close linkage between the locus for this DNA polymorphism and the Huntington gene. Some cross-overs in other such pedigrees have been detected, and the recombination fraction is 4% or less.

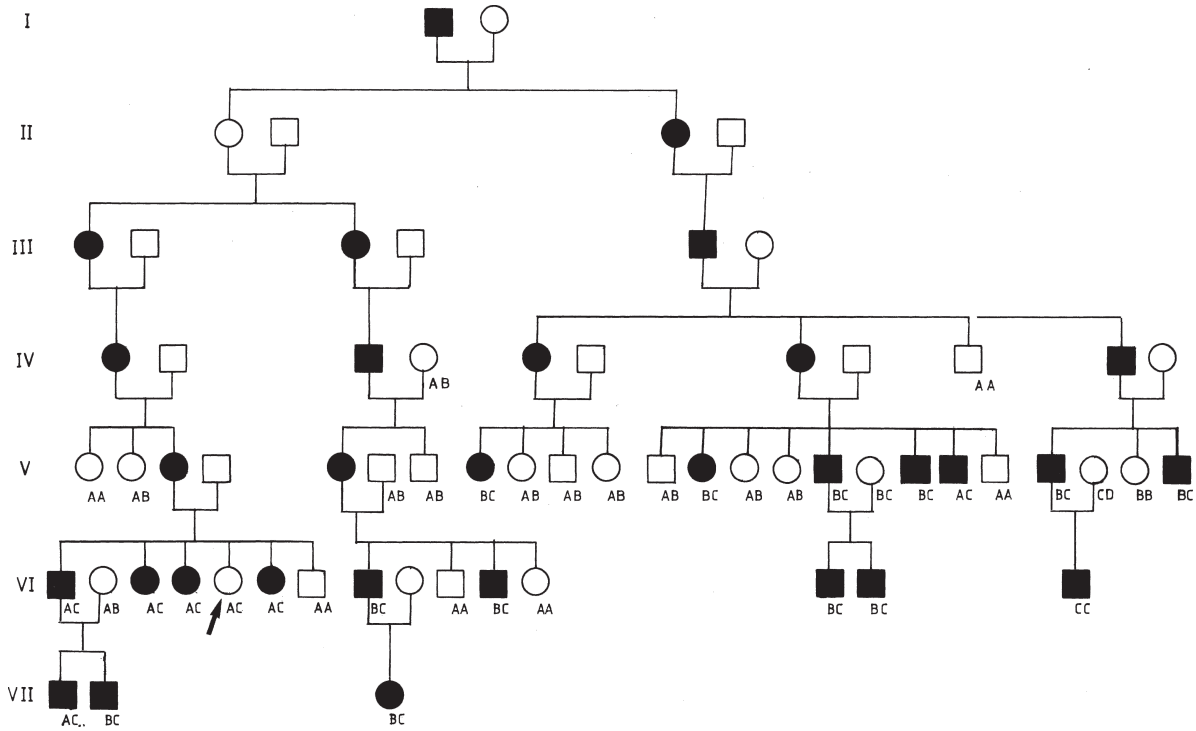

Fig. 6.2 Large pedigree from Venezuela with Huntington disease. *A, B, C, D*, four different "alleles" of a DNA polymorphism. The Huntington gene is transmitted together with "allele" C. One individual, VI,5 (*arrow*) has so far been unaffected. She will most likely be affected later (See text). From Gusella et al. [28]

The example of this pedigree shows that the chromosomal phase of alleles at two loci (*cis* or *trans*) can often be ascertained with great precision even in one large pedigree, and that recombinants can be identified if (at least) three generations are available for analysis, and if there are many sibs.

6.1.2.2 Statistical Analysis

In most cases linkage analysis is more difficult. Extensive pedigrees such as that in Fig. 6.2 are exceptional; most available families consist of two parents and their children. Here the problem is that the chromosomal phase is usually unknown: a double heterozygote may be AB/ab *(cis)* or Ab/aB *(trans)*. When the alleles are randomly distributed in the population (linkage equilibrium), the two types are expected in about equal frequencies: an AB/ab person forms germ cells in the ratio:

AB	Ab	aB	ab
$\frac{1-\Theta}{2}$	$\frac{\Theta}{2}$	$\frac{\Theta}{2}$	$\frac{1-\Theta}{2}$

whereas a heterozygote Ab/aB forms germ cells in the ratio:

AB	Ab	aB	ab
$\frac{\Theta}{2}$	$\frac{1-\Theta}{2}$	$\frac{1-\Theta}{2}$	$\frac{\Theta}{2}$

Expectations for germ cells are then in any case:

AB	Ab	aB	ab
$\frac{1-\Theta}{2}$	$\frac{\Theta}{2}$	$\frac{\Theta}{2}$	$\frac{1-\Theta}{2}$

or

$\frac{\Theta}{2}$	$\frac{1-\Theta}{2}$	$\frac{1-\Theta}{2}$	$\frac{\Theta}{2}$

which adds up to:

$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

irrespective of Θ. It even remains true if $\Theta = 0$ (very close linkage).

All four types of germ cells occur with the same frequencies, regardless of the probability of recombination Θ. Linkage does not lead to any association of alleles A,B or a,b in the population (Exception: linkage disequilibrium; Sect. 6.2). Another criterion for linkage must be found, one that is independent of the phase of the double heterozygote.

Such a criterion would be the *distribution* of children within sibships. In mating of AB/ab persons (*cis* phase) most children show the allele combinations of their parents; in matings of Ab/aB (*trans* phase) most children show these alleles in a new combination. How can these deviations from random distribution within sibships be measured and used for establishing linkage and determining the probability of recombination? Bernstein in [6] was the first to develop such a method. It has now been replaced by the method of "logarithm of differences" (LOD) scores as developed by Haldane and Smith [31] and Morton [65–66] and is generally used to assess linkage. Its principle can be described as follows:

The probability P_2 that the observed family data conform to the behavior of two loci under full recombination without any linkage is calculated and similarly, the probability P_1 that the identical family data are the result of two linked loci under a specified recombination fraction (Θ) is estimated for various families. The ratio of these two probabilities is the likelihood ratio and expresses the odds for and against linkage. This ratio $\frac{P_1(F/\Theta)}{P_2(F/({}^1/_2))}$ must be calculated for each family F.

A man may be doubly heterozygous for the gene pairs A,a and B,b. His wife may be homozygous for the two recessive alleles aa, bb. Assume that his two sons, as the father, are doubly heterozygous, i.e., they inherited the dominant alleles A and B from the father. This probability is ${}^1/_2 \times {}^1/_2 = {}^1/_4$ for each son if the genes segregate independently. If the gene loci are closely linked without crossing over, the probability for occurrence of this pedigree may be calculated as follows. Either the genes occur in coupling state: AB/ab, then the possibility for common transmission to each of the two sons is ${}^1/_2$ (transmission of the combination ab would also have a probability of ${}^1/_2$), or the genes occur in repulsion state Ab/aB, where transmission of both dominant alleles to the same son requires crossing over. With close linkage and in absence of crossing over the probability here of common transmission = 0. Hence, the total probability for transmission of the combination aB to either son is ${}^1/_2$ and the likelihood ratio is $P_1/P_2 = ({}^1/_2)/({}^1/_4) = 2$ in favor of close linkage. Likelihood ratios for the various degree of loose linkage can be calculated in the same way.

For convenience the logarithm of the ratio is used, and a LOD score z (meaning "log odds" or "log probability ratio") is used:

$$z = \log_{10} \frac{P(F|\Theta)}{P(F|(^1/_2))} \qquad (6.1)$$

Here, $P(F|\Theta)$ denotes the probability of occurrence for a family F when the recombination fraction is Θ. Using the logarithms instead of the probabilities themselves has the advantage that the score of any newly found family can be added, giving a combined score $z = S\, z_i$ for all families examined.

Equation (6.1) implies an identical recombination fraction for both sexes. Since sex differences in recombination rates have been described [82] (see below), the z score in actual data should be computed separately for the sexes:

$$z = \log_{10} \frac{P(F/\Theta,\Theta')}{P(F/(^1/_2,^1/_2))} \qquad (6.2)$$

where Θ is the recombination fraction in females and Θ' in males.

It follows from the definition of the likelihood ratio that the higher its numerator, the stronger is the deviation in the direction of linkage. In terms of logarithms the higher the z score, the better is the evidence for linkage. A LOD score of 3 or higher is generally considered as proof of linkage. Minor corrections for dominance and for ascertainment of pedigrees with rare traits but are not dealt with here [93].

The score $z(\Theta,\Theta')$ for the entire set of data is the sum of the scores of the separate families. For a first approach $\Theta = \Theta'$ is assumed to simplify the calculations. After linkage has been established, a possible sex difference can be looked for.

Numerous computer programs for detection and estimation of linkage are available (for example, see http://linkage.rockefeller.edu/soft, which lists multiple software tools for genetic linkage analysis of human pedigrees). They also allow for testing whether a part only of the observed families show linkage (=linkage heterogeneity). These programs permit to make optimal use of linkage information even in large and sometimes complicated pedigrees. For a detailed account of reasoning on linkage as well as methods of analysis, see Ott [74].

6.1.2.3 The Use of LOD Scores

The ideal mating for linkage studies involves a double heterozygote, i.e., a person heterozygous for two different traits, with a person homozygous for the two genes. The following types of families do not contribute information regarding linkage:

(a) Families in which neither parent is doubly heterozygous
(b) Families in which there cannot be any observable segregation
(c) Families in which the phases of the parents are unknown and there is only one examined child

Most linkage studies involve analysis of two common markers or of a common gene with a gene for a rare genetic disease. Opportunities to study linkage between two rare genes hardly ever exist. The ideal family for linkage studies is a kindred with at least three generations, many matings, and a large number of offspring. Such families are becoming rare in Western societies. An alternative approach involves testing of many small families. This may even have an advantage if more than one gene locus causes a special phenotype. In these instances the study of a single, large pedigree with linkage may create the impression that this gene locus is the only one whose mutations cause the phenotype in question, whereas analysis of many, smaller pedigrees may point to other loci as well, and hence to genetic heterogeneity.

When linkage has been established and a maximum likelihood estimate of Θ achieved, the question of heterogeneity should be examined. If, for example, linkage between the locus for a genetic polymorphism and a rare dominant condition has been established, linkage analysis can help to prove genetic heterogeneity if only part of the family data shows linkage. This occurs very often [61]; the statistical problem posed by such a situation is tricky. Ott [72] has proposed using the χ^2 statistic to compare hypotheses: linkage without heterogeneity vs. nonlinkage, linkage with heterogeneity vs. linkage without heterogeneity, and linkage with heterogeneity vs. nonlinkage. It is also possible to estimate the proportion of families showing linkage in the data set studied.

The human genome is so saturated with genetic markers that one can estimate linkage not only for two loci but for several markers at once (multipoint linkage). Appropriate computer programs for linkage analysis, for example, the LINKAGE package (http://linkage.rockefeller.edu/), have proven to be very useful. In fact, linkage analysis is now often performed using SNP arrays covering several hundred thousand markers which can easily be analyzed with other software tools such as dChip (http://www.biostat.harvard.edu/complab/dchip/).

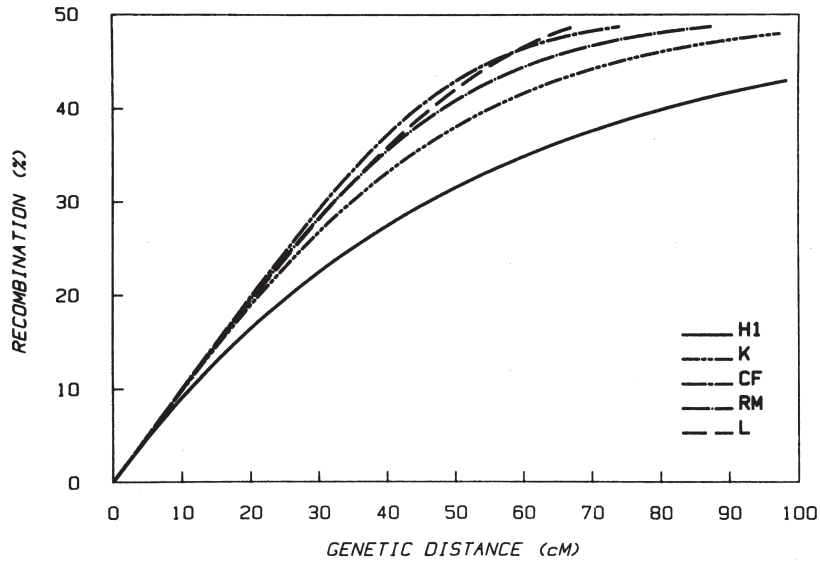

Fig. 6.3 Genetic distance (in centimorgans) in relation to the recombination fraction (in percentage), according to estimates from various authors. *H1* Haldane function with no interference; *K* Kosambi; *CF* Carter and Falconer; *RM* Rao and Morton; *L* Ludwig. From White and Lalouel [110]

6.1.2.4 Recombination Probabilities and Map Distances

Once a number of linkages have been established, the next step is to estimate map distances between these loci. These distances are expressed in morgans and centimorgans, 1 cM (map unit) meaning 1% recombination (Θ=0.01) for small map distances. For larger distances this value must be corrected for double crossing over. Various methods have been proposed. Given a recombination frequency Θ, the map distance (cm) can be read directly from Fig. 6.3.

6.1.2.5 The Sib Pair Method

The use of LOD scores is the ideal method if the mode of inheritance of the two traits to be tested for linkage has been established. Examples include testing of linkage for two genetic markers or for a marker and a clearcut monogenic disease. At least two generations should be available. The analysis becomes more difficult if penetrance of the mutant gene is incomplete, and a definitive phenotype cannot be assigned. While inclusion of a penetrance term in the analysis may be possible, introduction of this and other adjustments can be hazardous since it may lead to false claims, particularly if the data are manipulated in various ways until a "positive" linkage result is obtained.

In general, if the mode of inheritance cannot be established, or when data from only one generation are available, it is preferable to use the sib method first suggested by Penrose in the 1930s [76] (see also [7, 82–84, pp. 90–92]; Fig. 6.4). Its rediscovery has been called by Ott [73] "the cutting of the Gordian knot." (Alexander the Great was challenged to disentangle this knot, which no one had been able to do previously; he cut through it with his sword.) This is because the detection of linkage with this method does not depend on correct assignment of the mode of inheritance but only on the influence of a gene that contributes non-negligibly to the trait and on linkage of this gene with a marker. Such approaches are termed "nonparametric." Penrose explains, "The method is based on the principle that, when pairs of sibs are taken at random, certain types of sibling pairs will be more frequent if there is linkage than if there is free assortment of the characters studied." The method is used as follows: Codominant genetic markers with several alleles are studied in a series of sib pairs (or other pairs of relatives) both of whom are affected with a disease whose linkage relationship to the marker is to be investigated. If there is *no* linkage to the marker, 25% of affected sib pairs share both maternal and paternal alleles of the

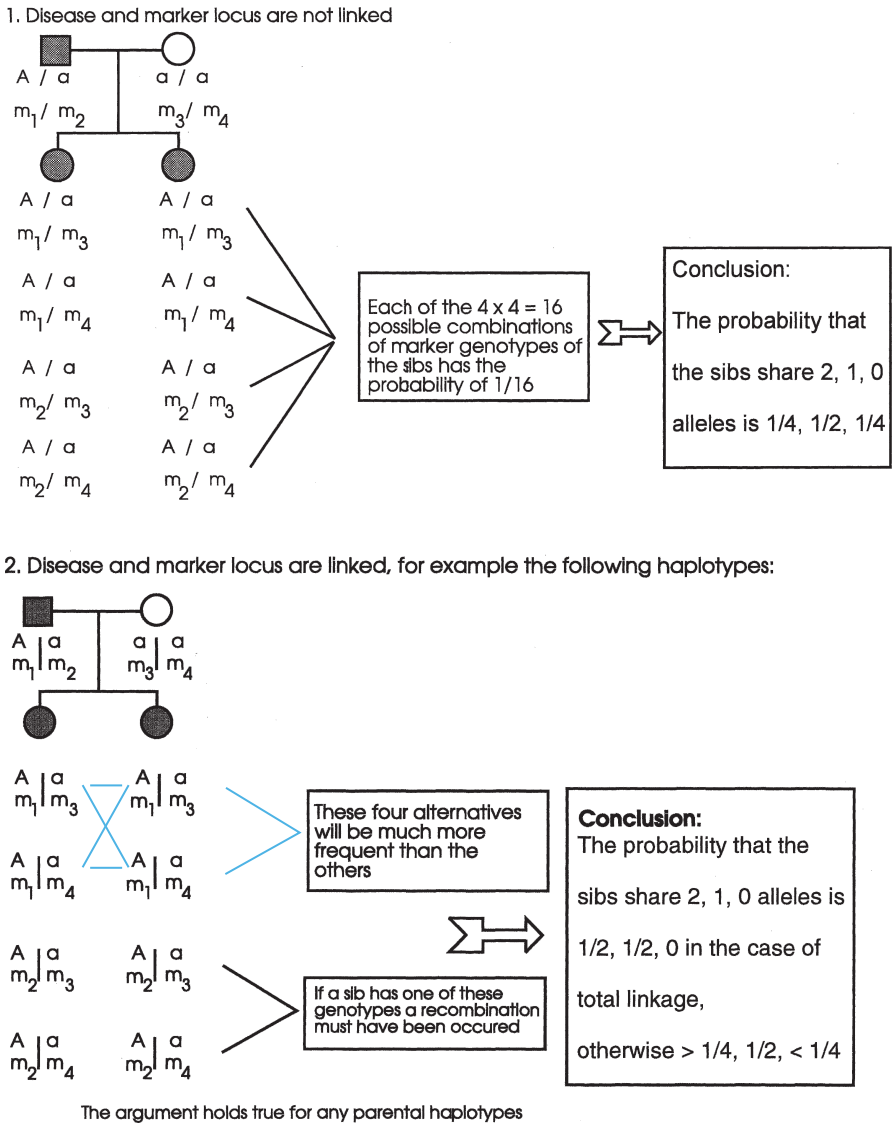

Fig. 6.4 Principle of the sib pair method for finding linkage. Assume a mating between a parent heterozygous for a dominant disease allele A and a parent homozygous for a normal allele a; each of the four parental chromosomes have different marker alleles at a marker locus. The possible alternatives are presented. Blue bars connect the more common sib pairs

marker, 50% share one of the marker alleles, and 25% differ for both marker alleles. If the marker is linked to a gene that contributes to causing the disease, the proportion of sib pairs with the disease who share one or two marker alleles is increased over the expected 25% or 50%, respectively.

The problem is straightforward when the marker status of both parents is known, and the identity by descent of the marker can be established. However, the method can also be used, albeit with more difficulty, if only marker information on the affected sib pairs is available, and the parents are not investigated (see also [13]). Problems arise here if the affected parent is homozygous for the marker, or if the unaffected parent contributes a marker allele to the child that is identical to the marker that cosegregates with the disease gene [67]. This problem can be avoided in either of two ways: (a) *Unaffected* sib pairs in the same families can be studied. Such pairs are expected to be more similar in the alternative marker alleles. However, the disease must have 100% penetrance so that there is certainty that an unaffected sib really does not carry the disease

gene. (b) Another approach is to study haplotypes of several closely linked markers or of multiallelic variable-number tandem repeats (VNTRs) or microsatellites rather than of only a single restriction fragment length polymorphism (RFLP) marker. Under these conditions each parent often has a unique haplotype at the site under study with four haplotypes between the two parents. A child inherits only two haplotypes, one from each parent. With no linkage 25% of sib pairs share two haplotypes, 50% are identical for one, and 25% for none. With linkage, statistically significant increases over the 25% and 50% proportions of shared haplotypes are obtained.

This method has been adapted for pairs of relatives other than sibs as well; computer programs for testing linkage and estimating map distances are available (http://linkage.rockefeller.edu/soft). Risch and Merikangas have shown that the sib pair methods have relatively low power to detect loci with moderate odds ratios [85].

Haplotype analysis is especially useful if general circumstances favor the view that all patients suffer from a genetic disease which can be traced back to one single mutation. (See the discussion of such "founder effects" in Sect. 17.1.2) Here, the time since this mutation occurred may not have been sufficient to randomize marker loci around this mutation by repeated crossing over; it may still exist within the same haplotype in most instances. If the mode of inheritance is autosomal-recessive, patients may be homozygous for this haplotype. In this way, the autosomal-recessive gene for benign recurrent intrahepatic cholestasis (BRIC), an autosomal-recessive condition occurring in the Tyrolian Alps, was mapped to chromosome 18. If its main precondition – evidence for a founder effect – is met, the method is very efficient statistically [35], and the gene can be identified from a single pedigree with a modest number of siblings.

6.1.2.6 Results for Autosomal Linkage, Sex Difference, and Parental Age

The first autosomal linkage in man was found by Mohr [64] between the Lutheran blood groups and the AB0 secretor locus. Some years later linkage between the Rh loci and elliptocytosis (166900) was established and used to detect genetic heterogeneity of elliptocytosis, since not all families with elliptocytosis showed this linkage. A short time later, linkage between the AB0 blood group locus and the dominant nail-patella syndrome (161200) locus was found. This linkage established for the first time a sex difference of recombination probabilities in humans: map distance between these loci was 8 cM for males and 14 cM for females. A great many linkages have since been examined for sex differences; in the majority of instances a higher recombination fraction has been observed in females than in males. The same sex difference had been known for a while in the mouse [80]. It conforms to Haldane's rule [30] that crossing over is generally more frequent in the homogametic than in the heterogametic sex. In humans, however, this rule has exceptions. In the distal portion of 11p, for example, the recombination frequency appears to be higher in males [111]. It appears that such a higher male recombination rate may be characteristic for chromosome parts close to telomeres. There is also good evidence that the absolute exchange rate is higher in chromosome parts close to the telomere [52, 110]. However, the overall recombination rate is definitely higher in females [22, 45]. In *Drosophila,* there is no crossing over at all in the male. Typing of individual sperm shows that the recombination rate in a specific region can be different between individuals [14, 45]. The mechanism controlling this variation in humans has remained elusive. However, recently sequence variants in the 4p16.3 region were identified, which correlated with recombination rates. Interestingly these variants were mapped to the *RNF212* gene. This gene is a putative ortholog of the *ZHP-3* gene that is essential for recombinations and chiasma formation in *Caenorhabditis elegans*. An intriguing finding was that the haplotype formed by two single-nucleotide polymorphisms (SNPs) was associated with the highest recombination rate in males whereas the same haplotype was associated with a low recombination rate in females [46].

There has been considerable discussion as to whether recombination frequency is also influenced by parental age. In the mouse, the data are consistent with decreasing recombination rates with aging in females and increasing rates in males. Weitkamp [109] found a significantly increased incidence of recombinants with increasing birth order in humans for eight closely linked pairs of loci, indicating a parental age effect. There was no difference between males and females for this effect. A similar parental age effect was found for the Lutheran/secretor and Lutheran/myotonic

dystrophy (160900) pairs but not for AB0/nail-patella or Rh/PGD pairs.

In a survey of cytogenetically determined chiasma frequencies from 204 males reported in the literature, little or no linear trend with age was found [53]. No cytogenetic data are available for females. The discrepancy between formal recombination data and chiasma frequencies is unexplained [110].

6.1.2.7 Information from Chromosome Morphology

Pairs or clusters of autosomal loci found to be linked (linkage groups) could not be assigned to specific chromosomes by a formal methodology of family study. The first chromosomal localization was accomplished as follows [21, 81].

The long arms of chromosome 1 frequently show a secondary constriction close to the centromere. In about 0.5 % of the population, this constriction appears much thinner and longer than normal. The variant is dominantly inherited. An uncoiler locus (Un-l) appears to be localized on chromosome 1. Linkage studies show close linkage between the blood group Duffy locus and the Un-l trait; $\Theta = 0.05$. Linkage between Duffy and congenital zonular cataract (116200) had been discovered earlier. Hence, a linkage group with three loci, cataract, Duffy, Un-l could be assigned [21].

Another possibility to localize genes on specific chromosomes was afforded by deletions. If a gene locus whose mutation has a dominant effect is lost by deletion, the absence of that gene may occasionally have a phenotype similar to the dominant mutation. More extensive symptoms may also be present, since more genetic material than a single gene would be expected to be lost. In 1963 a retarded child with bilateral retinoblastoma was found to have a deletion of the long arm of one D chromosome [56]. This chromosome was later identified as no. 13, and this 13q14 deletion has been found in a number of other cases with retinoblastoma and additional anomalies. Patients with retinoblastoma without additional symptoms usually have no deletion. The localization of this gene (RB1) has since been confirmed by DNA marker studies and the gene has been cloned [24, 55] (see Sect. 14.1.2).

Another approach, thought to be more generally useful, is the quantitative examination of enzyme activities in cases with chromosome anomalies. Most enzymes show a clearcut gene dose effect in heterozygotes, i.e., heterozygotes for an enzyme deficiency have approx. 50% of enzyme activity. Therefore a similar gene dose effect might be expected when a gene locus is localized on a chromosome segment that has been lost by deletion.

The results of many early studies of this sort proved disappointing. Later, however, an increasing number of such gene dosage effects have been described in vitro, on trisomic and monosomic cells [48] (Sect. 3.6). To mention only one example, the activity of the enzyme phosphoribosylglycinamide synthetase was studied in several cases of partial monosomy and full and partial trisomy 21, as earlier studies had suggested a gene dosage effect for this enzyme. In regular trisomy 21 an excess was found with a ratio of trisomy 21 to normal of 1.55. A ratio of 0.99 was found in 21q21 → 21pter monosomy; 0.54 in 21q22 → 21qter monosomy; 0.88 in 21q21 → 21pter trisomy; and 1.46 in 21q22.1 trisomy. Therefore the phosphoribosylglycinamide synthetase gene locus could be localized in subband 21q22.1 [15]. Utilization of variants in chromosome morphology (heteromorphisms), such as the secondary constriction on chromosome 1 mentioned above, along with gene dosage studies, slowly opened the way to linkage and gene localization. Another method, using cell fusion, has led to much more rapid progress.

6.1.3 Linkage Analysis in Humans: Cell Hybridization and DNA Techniques

6.1.3.1 First Observations on Cell Fusion

The history of cell fusion is related by Harris [33]. Binucleate cells were observed in 1838 by J. Mueller in tumors, and afterwards by Robin in bone marrow, by Rokitansky in tuberculous tissue, and by Virchow both in normal tissues and in inflammatory and neoplastic lesions. The view that some of these cells were produced by fusion of mononucleate cells derived from the work of de Bary in 1859, who observed that the life cycle of certain myxomycetes involves the fusion of single cells to form multinucleated plasmodia. The earliest reports of multinucleated cells in lesions that can be identified with certainty as being of viral origin appear to be those of Luginbuehl (1873)

and Weigert (1874), who described such cells at the periphery of smallpox pustules.

Following the introduction of tissue culture methods, numerous observations were made on cell fusion in cultures of animal tissue (see [32]). Enders and Peebles (1954) found that the measles virus induces cells in tissue culture to fuse to form multinucleated syncytia. Okada (1958) showed that animal tumor cells in suspension can be fused rapidly to form multinucleated giant cells using high concentrations of hemagglutinating parainfluenza virus (Sendai virus).

In 1960 Barski, identified cells generated by spontaneous fusion in a mixed culture of two different but related mouse tumor cell lines. These cells contained the chromosome complements of both parent cells within a single nucleus. This phenomenon was then examined by Ephrussi et al., who concluded that not only closely related mouse cells could be hybridized; even larger genetic differences did not exclude spontaneous cell fusion. However, it soon became obvious that the frequency of spontaneous cell fusion is very low, and that many cell types never fuse spontaneously. Fusion frequency must be increased in some manner. Furthermore, isolation of hybrid cells was possible only when culture conditions gave these cells a selective advantage.

Both problems were soon solved. Littlefield (1964) isolated the rare products of spontaneous fusion in mixed cultures by a technique adopted from microbial genetics. Fusion of two cells deficient in two different enzymes resulted in hybrids that recovered the complete enzyme set by complementation. Only these cells survived selection against the deficient cells.

Harris and Watkins [33] enhanced the fusion rate of various cells by treatment with UV-inactivated Sendai virus. Along with introduction of this method, they showed that fusion can be induced between cells from widely different species, and that the fused cells are viable. With this work, widespread use of the cell fusion method in various branches of cell biology began.

6.1.3.2 First Observation of Chromosome Loss in Human–Mouse Cell Hybrids and First Assignment of a Gene Locus

Weiss and Green [108] fused a stable, aneuploid mouse cell line, a subline of mouse L cells, with a diploid strain of human embryonic fibroblasts. The mouse cell line was deficient in the thymidine kinase (TK) locus and did not grow in hypoxanthine-aminopterin-thymidine (HAT) medium, a culture medium selective for cells containing the human TK locus (188300).

Cultures were initiated by mixing the two types of cells and growing them on standard medium. After 4 days cultures were placed in the selective HAT medium. This led to degeneration of the mouse cells, leaving a single layer of human cells. After 14–21 days hybrid colonies could be detected growing on the human cell monolayer. A number of these colonies were then isolated, grown for a longer time period, and examined. They turned out to maintain the mouse chromosome complement, but 75–95% of the human chromosomes were lost. One human chromosome, however, was present in almost all cells growing in the HAT medium. This suggested that the locus for thymidine kinase is localized on this chromosome. Therefore control experiments were carried out with a bromodeoxyuridine-containing medium. Bromodeoxyuridine, a base analogue for thymine, is accepted by TK in place of thymine and selects against cells containing this enzyme. A special chromosome described as "having a distinctive appearance" was present in almost all HAT cultures but in none of the bromodeoxyuridine cultures. It was concluded that the TK locus is indeed localized to this chromosome. Shortly thereafter the chromosome bearing the TK locus was found to be no. 17 [63] (Fig. 6.5).

This work led to two principles which were later decisive for the use of cell hybridization in linkage work:

1. Hybrids between mouse and human cells tend to lose many human chromosomes. It was later shown that this loss is random, and therefore examining a great number of hybrids one can expect to find a cell that has kept any one specified chromosome.
2. By using an appropriate selective system it is possible to select cells with a certain enzyme activity and to localize the gene loci specifying this enzyme to a specific chromosome.

Whereas genetics has historically been the science of genetic variability within a species, the hybridization method permits the localization of genes that do not show genetic variability in humans, provided only that the gene products of the human and nonhuman cells can be identified. One means of identification is the use of a selective system.

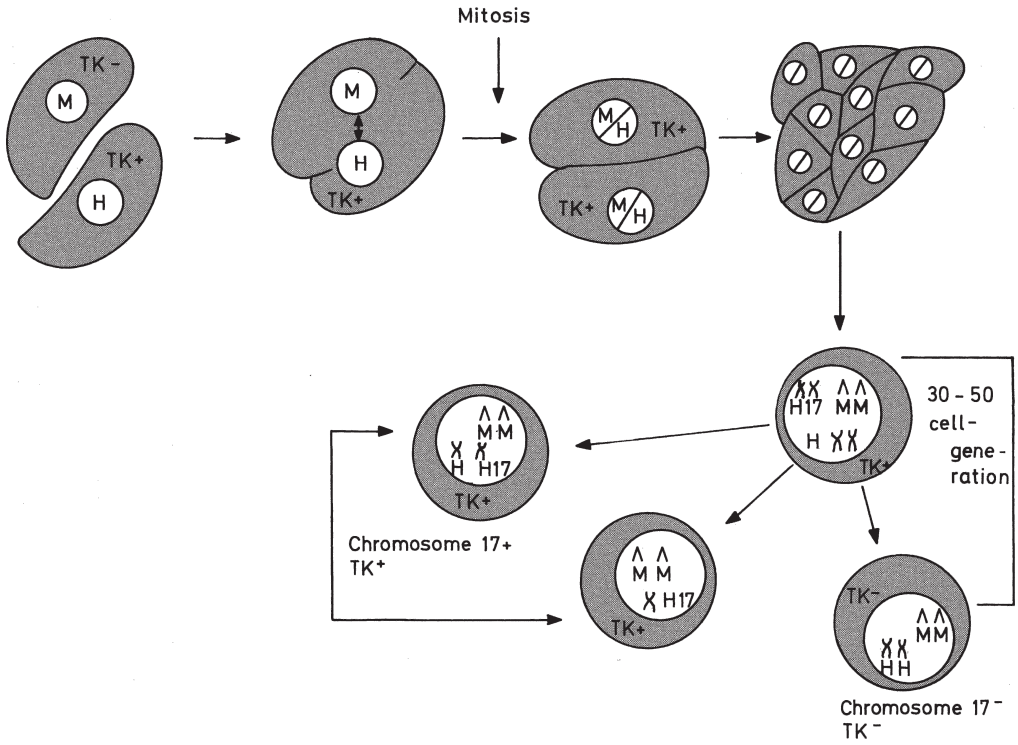

Fig. 6.5 The principle of gene localization on an autosome. Thymidine kinase deficient mouse cells (*M, TK*$^-$) are grown in mixed cell culture with normal human cells (*H, TK*$^+$). The cells are fused spontaneously, chemically, or by Sendai virus. After 30–50 cell generations the cells have lost part of their human chromosomes. Only cells having kept chromosome 17 show thymidine kinase activity (two cells at *left*). Cells without chromosome 17 show no *TK* activity (cell at *bottom right*)

Since 1967 selective systems have been developed for several enzymes. One example uses the hypoxanthine phosphoribosyltransferase locus on the X chromosome. This system can be used for selection not only of other X-linked loci but also of autosomal loci if a part of the autosome has undergone translocation with the X chromosome. It is also possible to assign loci for which no selective system exists, provided that enzymes produced by the two species have recognizable differences such as electrophoretic variation.

In hybrid cultures chromosome breakage and rearrangement are relatively frequent events. This chromosomal behavior made possible the suitable selection of hybrid clones containing identifiable parts of chromosomes, thereby combining the advantages of deletion mapping and hybridization.

Later the use of irradiation of the donor chromosome led to the development of radiation hybrids. These panels contain many chromosome fragments and were used to map genes and RFLP- and PCR-based markers more precisely [18, 54].

6.1.3.3 Other Sources of Information for Gene Localization

In situ hybridization technologies have also been instrumental for the localization of genes on chromosomes (see Sect. 3.4.4).

6.1.3.4 DNA Polymorphisms and Gene Assignment

The detection of a large number of DNA restriction site polymorphisms and other DNA markers opened up an additional approach to mapping of the human genome. In addition to RFLPs, other markers have been detected, among them the minisatellites [40], short DNA sequences distributed in great number over the human genome and occurring with variable numbers of repeats; this number is different in almost every individual. Therefore the information content for linkage studies is very high. Another such system, that of

so-called microsatellites, consists of $(CA)_n$ repeats that occur in great numbers in the genome; the number of repeats per unit is extremely variable, as well. Localization of individual $(CA)_n$ probes in the genome has been achieved by identifying and using specific DNA sequences on both sides of these markers that allows amplification by the polymerase chain reaction (PCR). Genes for many important hereditary diseases could be localized on specific chromosomal sites using such markers.

Immortalized lymphoblastoid cell lines from large three-generation families with known genotypes for many marker loci are available for the study of new markers [111]. These pedigrees, the CEPH families, consist of samples collected in Salt Lake City, Utah, France, and Venezuela [19, 22]. The CEPH families (both parents and a great number of children) have been typed with many genetic markers. DNA from such families is available to the scientific community for further mapping. With increasing numbers of available markers, analysis of linkage relationships not only between two gene loci (e.g., one disease gene and one marker) but also between a disease gene and a set of markers – a haplotype – is routinely carried out. Such haplotypes are used in research studies as well as in genetic counseling and prenatal diagnosis. The proportion of kindreds in which the combination of genotypes is informative for linkage – and hence for prenatal diagnosis – can be enhanced appreciably by using such sets of markers.

6.1.3.5 Gene Symbols to Be Used

In cooperative scientific activities such as mapping of the human genome, certain terminological conventions are necessary. Human gene names are now assigned by the Human Genome Organization (HUGO) gene nomenclature committee. The rules for gene symbols include: only uppercase letters, no hyphens, no more than four or five letters or numbers, etc. For details see http://www.gene.ucl.ac.uk/nomenclature. Standardized nomenclature to describe mutations and genetic variants has also been established [3].

6.1.3.6 Linkage of X-Linked Gene Loci

Assignment of loci to the X chromosome is straightforward when the pedigrees show the typical pattern of X-linked inheritance. The X chromosome is the chromosome most completely saturated with disease loci. Well-known genes for human diseases have been localized. But even with these localizations many X-linked genes need to be mapped; for example, it is estimated that up to 260 genes causing syndromic and nonsydromic X-linked mental retardation (XLMR) may exist. The most common known cause of XLMR is an expansion of a trinucleotide repeat in *FMR1*. Large-scale systematic resequencing of the coding exons of the X chromosome in males with XLMR now allow the identification even of rare, disease-causing sequence variants [101].

6.1.3.7 Genetic and Physical Map of the Homologous Segment of X and Y Chromosomes

When genetic maps of a chromosome were determined, these were complemented by a physical map. The final goal was to identify the DNA sequence of all genes within this area. The following presents an introduction into these methods using the pseudoautosomal region of Xp and Yp as examples.

This region is located in the Giemsa light band Xp22.3 and in Yp11.32 (see Sect. 3.6.3.2). Various authors constructed partial physical maps; and the region comprises about 2,560 kb in its entire length [79]. There is a certain interindividual length variation. A number of families from the CEPH family pool [19] were studied with 11 exactly localized DNA markers; this established very high recombination frequencies in males and lower but clearly elevated frequencies in females. In males, the genetic map is 55 cM long and in females 8–9 cM. Hence the sex difference in this case is opposite to Haldane's rule. Moreover, it is about 20–25 times higher than the average of all chromosomes in the male and 6 times higher in the female. Differences in recombination rates between chromosome regions have also been demonstrated for other chromosomes.

The first step of physical mapping often involved restriction of this area with "rare cutter enzymes" which cut DNA into regions with many CpG islands [49]. As a rule, CpG islands indicate the presence of genes [87]. Closer scrutiny of this terminal area by chromosome jumping revealed at least five regions in which the CpG islands are concentrated. Further analysis, by, constructing of a contig of YAC clones, allowed the first-generation maps to be generated.

6.1.3.8 The Y Chromosome

It is the Y chromosome [105, 113] which determines male sex. Genetic analyses succeeded in localizing specific factors involved in sex determination to certain segments of this chromosome. As in many other instances, analysis of pathological conditions has contributed to understanding of the normal state, such as the study of men with two X but apparently no Y chromosome. As early as 1966, Ferguson-Smith [23] postulated XY translocations which were expected to transfer to the X a small – but for male development decisive – part of the Y chromosome. This expectation has been confirmed by many observations [113]. Since meiotic pairing of X and Y chromosomes occurs in the pseudoautosomal and in adjacent nonhomologous regions, and since pairing errors provide a plausible mechanism for such translocations, the search for the testis-determining factor (TDF) soon concentrated on the short arm. Here the *SRY* (sex reversal gene on Y) gene was finally identified [91]. The mechanism of *SRY* action has been elusive. Recent discoveries shed some light on the role of *SRY*, which is thought to act synergistically with SF1, a nuclear receptor, through an enhancer of *SOX9* to promote Sertoli cell differentiation. *SOX9* is probably the pivotal factor in regulating the gene activity that defines Sertoli cells. Both *SOX9* and SF1 synergize to activate transcription of several downstream genes [88].

The Y chromosome is now known to contain 89 protein-coding genes and at least 27 distinct proteins [92]; www.ensembl.org]. A region within the euchromatic segment of the long arm appears to be important for normal spermatogenesis, since deletions within this region lead to arrested sperm formation either in an early stage, i.e., not even functional spermatogonia are formed, or in postmeiotic stages [105]. The first discovered deletions were so large that they could be recognized by cytogenetic methods [102]. Small deletions were later identified by molecular techniques [105], and their recognition has become important for differential diagnosis of male infertility.

In addition to genes involved in testis development and spermatogenesis, the Y chromosome harbors genes encoding transcription factors, initiation factors, ribosomal proteins, and kinases. Comparisons of the Y chromosomes of human and chimp have provided unique insight into the evolution of Y-chromosome genes [36].

6.1.3.9 DNA Variants in Linkage

The HapMap project has set out to characterize the majority of the SNPs in the human genome and has led to the identification of over 3 million variants [37]. The large number of DNA polymorphisms provide many new markers, and most linkage work is now being carried out with DNA variants often on arrays with 100,000 to 1 million SNPs [59] (Sect. 4.4.2).

Linkage disequilibrium (i.e., failure to demonstrate free assortment; see Sect. 6.2) has frequently been found between the sites of various markers at a given locus. Since these sites are physically very close, crossovers between them are rare, and many generations must pass before linkage equilibrium is reached. Furthermore, current data suggest that recombination rates at closely linked markers may vary considerably between different chromosomal locations. Thus, both "hot" and "cold" spots of recombination appear to exist [26, 68, 69].

6.1.3.10 Practical Application of Results from Linkage Studies

In the past the main interest of linkage studies was theoretical. Practical applications, however, are frequently employed. If, for example, gene A causes a rare hereditary disease manifesting itself later in life, and B is a genetic marker closely linked to A and segregating in the same family, the disease was predicted in a prenatal sample or young individual, and this prediction used in genetic counseling (Sect. 25). Today genetic diagnosis is routinely performed by direct study of the mutant gene itself.

6.1.4 Biology and Statistics of Positional Cloning

For disease loci localized by linkage analysis identifying the gene and mutations involved is termed "positional cloning" [17]. This process involves identifying genes in the interval defined by linkage, and analyzing affected individuals to identify mutations. This process can proceed very rapidly if a gene of obvious biological interest is identified in the interval, and the mutations readily identified. However, the process can

be very time consuming if the interval is large, or the disease gene small or poorly expressed, or if there are few affected individuals for mutation screening. The cystic fibrosis gene was localized to an interval now know to be 1.4 Mb in size [106, 112]. It took over 4 years of intense effort to positionally clone the cystic fibrosis gene [86]. With the completion of the sequence of the human genome, this process is greatly aided as most of the genes in an interval are known. Therefore an investigator can sequence all of the coding exons of the positional candidate genes to identify mutations.

To date, most of the genes responsible for the common Mendelian disorders have been identified. However, identification of genes in less common Mendelian disorders with the aforementioned tools will continue to be important [2]. Nowadays, positional cloning is more often applied to complex diseases. This process is similar – a genetic interval is defined via sib pair linkage or whole-genome association and the relevant gene and "disease-causing variant" need to be identified [20]. This process has resulted in identifying genes for such complex diseases as macular degeneration, diabetes, Parkinson disease, obesity, and others. However, many gene haplotypes associated with common disease do not have coding sequence variants, and regulatory effects are proposed.

6.2 Gene Loci Located Close to Each Other and Having Related Functions

6.2.1 Some Phenomena Observedin Experimental Genetics

6.2.1.1 Closely Linked Loci May Show a Cis-Trans Effect

When series of multiple alleles were analyzed in *Drosophila,* crossing over within these series was observed occasionally, indicating that what had been considered as one "gene" can be subdivided by genetic recombination. Such alleles were termed "pseudoalleles" by McClintock in [60]. In some a so-called *cis-trans* effect was shown. When two mutations were located side by side on the same chromosome (*cis* position), the animal was phenotypically normal, but when they were localized on homologous chromosomes (*trans* position), a phenotypic anomaly was seen [58].

6.2.1.2 Explanation in Terms of Molecular Biology

In fungi, bacteria, and phages, genetic recombination is normally observed within functional genes, i.e., DNA regions carrying information for one polypeptide chain. A *cis-trans* effect is now considered to be typical for two mutations that are not able to complement each other functionally, i.e., that are located within the same structural gene. Complementation between two mutations, by the same token, is regarded as an indication that these mutations are located in different functional genes. A gene has many mutational sites and may be subdivided by recombination. Complementation tests are often used to test genetic, biochemically characterized conditions for heterogeneity.

6.2.1.3 A Number of Genes May Be Closely Linked

Close linkage has frequently been described between mutations affecting closely related functions, which are perfectly able to complement each other functionally and show no *cis-trans* effect. In bacteria such as *E. coli,* gene loci for enzymes acting in one sequence have been found to be closely linked and arranged in the sequence of their metabolic pathway. Their activity is subject to a regulating mechanism by a common operator and promoter [44].

6.2.2 Some Observations in the Human Linkage Map

6.2.2.1 Types of Gene Clusters That Have Been Observed

The first impression when examining the human linkage map and DNA sequence is that while most loci are distributed fairly at random, there are a large number of clusters of closely related genes. Here are a few examples:

(a) The loci for human hemoglobins γ, δ, and β are closely linked.
(b) The immune globulin region comprises a number of loci responsible for synthesis of immunoglobulin

chains. The same is true for genes of the T cell receptor (chromosome 14q11). The major histocompatibility complex (MHC) cluster including various components of complement on chromosome 6.

(c) No less than four gene loci involved in the glycolytic pathway are located on chromosome 1.

(d) A number of genes determining closely related enzymes are closely linked, for example, pancreatic and salivary amylase on chromosome 1, and guanylate kinase 1 and 2 on the same chromosome.

(e) The protan and deutan loci for red–green color blindness are located in the same cluster on the X chromosome.

6.2.2.2 Clusters Not Observed So Far

As mentioned above, functionally related genes in bacteria are frequently closely linked; they are subject to common control within an operon. One might predict that, in humans, such operons would also occur, but functionally related genes are rarely clustered. Two genes linked in the same operon in bacteria are those for galactose-1-phosphate uridyltransferase and galactokinase. In humans these genes are located on chromosomes 3 and 17, respectively. Similarly, the gene for G6PD is located on the X chromosome, and that for 6-PGD, the following enzyme in the shunt pathway, is situated on chromosome 1. Genes belonging to one gene family are sometimes but by no means always located close together. For genes involved in the immune system, including those for immunoglobulin synthesis, T cell receptors, and genes in the MHC system, this location has functional significance.

6.2.3 Why Do Gene Clusters Exist?

6.2.3.1 They Are Traces of Evolutionary History

In some cases clustering is simply left over from the evolutionary history of these genes. Early in evolution there was one locus for a given gene. Then gene duplication occurred and offered the opportunity of functional diversification [70]. The first duplication paved the way for further duplications due to unequal crossing over (Sect. 6.2.8) and hence for further functional specialization.

With no further chromosomal rearrangements the gene clusters remain closely linked. It is unknown whether in these cases close linkage is necessary for orderly function. While it may be so in some cases, this explanation is not needed to explain clustering. Evolutionary explanations are sufficient. For example, the red and green color vision genes appear to have arisen by gene duplication.

6.2.3.2 Duplication and Clustering May Be Used for Improvement of Function

The clustering of genes is without obvious functional significance. It would be surprising, however, if evolution were never to take advantage of this situation, combining products of such gene clusters to form higher functional units. This may be the case for the hemoglobin molecule since in the β cluster the ε, γ, β, and δ genes are arranged in the sequence of their successive activation during individual development (Sect. 11.3). In the immunoglobulins and T cell receptors close linkage of a number of genes, possibly a great many, has become important functionally, as their gene products combine to form various classes of functional molecules. In fact, segmental duplications in the human genome are selectively enriched for genes involved in immunity. In this respect, one of the most fascinating, recent discoveries was the identification of interindividual and interpopulation differences in the copy number of a segmental duplication encompassing the gene encoding CCL3L1 (MIP-1alphaP). This gene is a potent human immunodeficiency virus-1 (HIV-1)-suppressive chemokine and ligand for the HIV coreceptor CCR5. Individuals with a CCL3L1 copy number lower than the population average have a markedly enhanced HIV/acquired immunodeficiency syndrome (AIDS) susceptibility [27].

6.2.4 Blood Groups: Rh Complex (111700), Linkage Disequilibrium

The history of the rhesus blood types provides a fascinating illustration of how science develops. First, a new phenomenon was discovered. Scientists soon

realized that it eludes explanation by conventional concepts. Then a long-lasting scientific controversy arose as to the most appropriate extension of these concepts. In this controversy, a new explanatory principle was created that survived the controversy in this special case, and that could be applied to an increasing number of other observations. Finally, the problem was solved, and the controversy ended – by new methods.

6.2.4.1 History

In 1939 Levine and Stetson [57] discovered a novel antibody in the serum of a woman who had just delivered a macerated stillborn child and had received blood transfusions from her AB0-compatible husband. Of 101 type 0 bloods only 21 showed a negative reaction with this antibody. There was no association with AB0, MN, or P systems.

The following year Landsteiner and Wiener [51], immunizing rabbits with the blood of rhesus monkeys, obtained an immune serum that gave positive reactions with the erythrocytes of 39 of 45 individuals. Later the antibodies were compared with those of Levine and Stetson and thought to give reactions with the same antigens. This was subsequently found to be not quite true, and now the antigen uncovered by the true anti-rhesus antibody is called LW –, in honor of Landsteiner and Wiener. Rh typing in humans is always carried out with sera of human origin, i.e., according to Levine and Stetson's observation. The following discussion relates only to reactions with these human sera.

The great practical importance of the rhesus system became apparent when transfusion accidents were traced to this antibody, and especially when erythroblastosis fetalis, a common hemolytic disease of the newborn, was explained by Rh-induced incompatibilities between mother and fetus. The red blood cells of about 85% of all whites give positive reactions; family examinations showed that Rh-positive individuals are homozygous Rh/Rh or heterozygous Rh/rh, whereas the rh-negative individuals are homozygous rh/rh.

In 1941 Wiener discovered a different antibody that reacted with the cells of 70% of all individuals and was independent of the basic Rh factor (Rh', according to Wiener). A third related factor was discovered in 1943. These three factors were found in all possible combinations with one another, and the combinations were inherited together. Wiener proposed the hypothesis that these serological "factors" are properties of "agglutinogens," and that these agglutinogens are determined by one allele each of a series of multiple alleles. The agglutinogens were thus thought to determine the factors in different combinations. This descriptive hypothesis is so general that it indeed explains all the complexities discovered later.

6.2.4.2 Fisher's Hypothesis of Two Closely Linked Loci

R.A. Fisher developed a more specific hypothesis. At that time another antibody had been detected, anti-Hr. In 1943 Fisher (see [78]) examined a tabulation prepared by Race, containing the data accumulated so far. He recognized that Rh' and Hr were complementary. All humans have either Rh', Hr, or both. Individuals with both antigens never transmit them together to the same child, and a child always receives one of the two. Fisher explained these findings by proposing one pair of alleles for the two antigens. The pair was named C/c. In analogy, an additional pair of alleles D/d was postulated for the original antigens Rh^+ and rh^-, and a third pair of alleles for the third factor that had been discovered. To explain the genetic data close linkage between these three loci was assumed.

Fisher's hypothesis predicted discovery of the two missing complementary antigens d and e. This prediction was later fulfilled for e but not for d. In developing this hypothesis Fisher went one step further. In the British population, there are three classes of frequency of the Rh gene complexes (Figs. 6.6, 6.7). Fisher explained this finding by suggesting that the rare combinations could have originated from the more frequent ones by occasional crossing over. All four combinations of the less common class may have originated from occasional crossing over between the

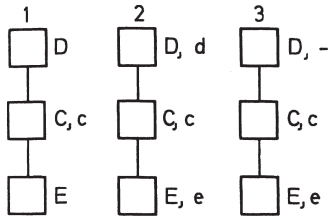

Fig. 6.6 A hypothetical structure of the Rh complex. *1,* On the basis of the evidence known in 1941; *2,* antigens predicted by Fisher and Race; *3,* antigens discovered; antigen *d* was not found

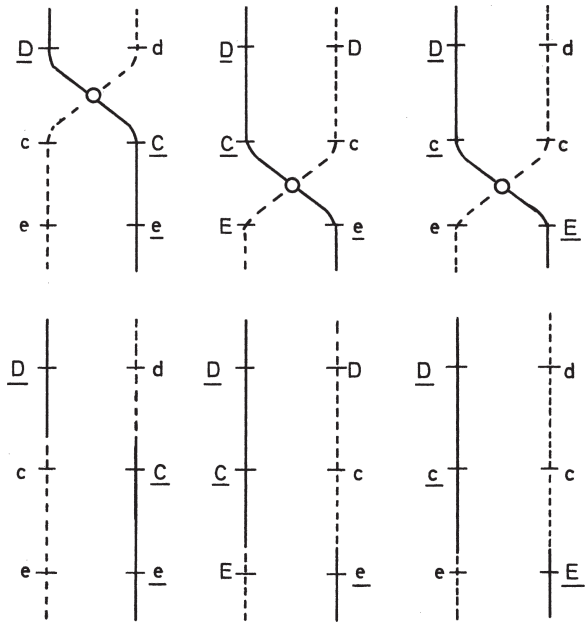

Fig. 6.7 Postulated production of three rare Rh haplotypes from the more common ones by crossing over. Each diagram refers to a different crossing over event. From Race and Sanger [78]

most frequent combinations; not, however, CdE. This complex needs inclusion of a second-order chromosome. Therefore the hypothesis explains why CdE is so rare. Still another prediction is possible. In every crossing over leading, on the one hand, to Cde, CDE, or cdE the complex cDe must also be produced. It follows that the frequency of the three former combinations together should equal the frequency of cDe. Frequencies actually found were: cDe, 0.0257; Cde + cdE + CDE, 0.0241 (in blacks, however, cDe has a high frequency).

Furthermore, Fisher believed the sequence of the three loci to be D-C-E, since cdE – which must have originated by crossing over between D and E from the genotype cDE/cde – is much more frequent in comparison with this genotype than Cde in relation to genotype CDe/cde (crossing over between C and D) and CDE in relation to CDe/cDE (crossing over between C and E).

6.2.4.3 Confirmation and Tentative Interpretation of the Sequential Order

Since Fisher's hypothesis many new observations have been made, the most important for the question of sequence being the combined antigens, for example, ce. These compound antigens were all compatible with the sequence D-C-E, whereas no such antigen suggesting close linkage between D/d and E/e has emerged. Fisher's hypothesis leads to two questions:

1. If the rare types have been formed by occasional crossing over from the more frequent ones, cases of crossing over should occasionally turn up in family studies. One such family has indeed been reported [97]: the father was CDe/cde, the mother cde/cde, four children were cde/cde, and three others CDe/cde, all in concordance with genetic theory. The sixth-born child, however, was Cde/cde. As the discrepancy involved father and child, it could be argued that the child was illegitimate. This, however, was made unlikely by blood and serum groups and by the fact that the family belonged to a religious sect with especially strict rules against adultery.
2. How should we envision the structure of the Rh gene(s) in the light of evidence from molecular genetics? There are two possibilities in principle:

   (a) The Rh complex is one gene with many mutational sites. Mutational changes are expressed as antigenic differences.
   (b) The Rh complex is composed of a number of closely linked genes, possibly three, and the main antigens reflect genetic variability at these genes. One important criterion is the *cis-trans* effect found in mutations affecting the same functional gene. As the ce compound antigen can be found in *cis*-phase CE/ce but not in *trans*-phase Ce/cE, Race and Sanger [77] tentatively concluded that C/c and E/e may be located within the same functional gene.

6.2.4.4 Molecular Basis

The Rh specificities have now been identified as membrane polypeptides. Molecular studies of the gene(s) have shown that in all D^+ individuals, two closely related Rh genes in each haploid genome appear to be present. One of these genes is missing in D^- individuals [16]. The authors concluded that one of the two genes controls the D polypeptide whereas the C/c and E/e specificities are coded by the second

gene, a result confirmed by the molecular cloning [16, 107]. These observations explain at the molecular level why no anti-d serum has been found. They also confirm the sequence D-C-E postulated by Fisher, as well as the above-mentioned conclusion of Race and Sanger [77] that C/c and E/e appear to be located in the same gene product. Thus, both original hypotheses were partially correct: the specificities C,c,E,e are located within the same gene-determined protein, as postulated by Wiener (which does not exclude occasional intragenic crossing over); the D specificity, on the other hand, is located in a second, closely linked gene, as postulated by Fisher. Moreover, attempts at understanding genetics of the Rh system led to the development of a new concept by Fisher that has found widespread application in many fields of human genetics: linkage disequilibrium.

6.2.4.5 Blood Groups: Linkage Disequilibrium

Linkage normally does not lead to association between certain traits in the population (Sect. 6.1.1). Even if initially there is a nonrandom distribution of linkage phases, repeated crossing over randomizes the linkage groups, and in the end the coupling and repulsion phases for two linked loci are equally frequent. There is linkage equilibrium. However, when the population begins with a deviation from this equilibrium, for example, because two populations with different gene frequencies have merged, or because a new mutation has occurred on one chromosome, the time required to reach equilibrium depends on the closeness of linkage: the closer the linkage, the longer the time until equilibrium is reached [12]. It is never reached if certain types have a selective disadvantage.

A selective disadvantage for certain Rh complexes that could lead to their becoming less frequent has not been demonstrated so far; selection works against heterozygotes (Sect. 18.3.3), but this does not mean that a general disadvantage has never existed; neither has a conclusive explanation in terms of population history been postulated. Fisher's hypothesis, by answering some questions, has posed a number of others. However, the concept of linkage disequilibrium proved to be still more important in population genetics and in the genetic analysis of DNA polymorphisms (Sect. 2) and the major histocompatibility (MHC) complex:

6.2.5 Major Histocompatibility Complex [105, 111]

6.2.5.1 History

It had long been known that skin grafts from one individual to another (allotransplants) are usually rejected after a short time. In 1927 K. H. Bauer [5] observed that rejection does not occur when skin is transplanted from one monozygotic twin to the other (isotransplant). Such a transplant is accepted just as a transplant in the same individual (autotransplant). This showed the rejection reaction to be genetically determined. In the following years skin, and later kidney, transplantations between monozygotic twins were occasionally reported. Research on histocompatibility antigens in humans began only when leukocytes were shown to be useful as test cells.

Dausset observed in 1954 that some sera of polytransfused patients contain agglutinins against leukocytes. He later showed that sera from seven such patients agglutinated leukocytes from about 60 % of the French population, but not the leukocytes of the patients themselves. Twin and family investigations soon established that these isoantigens are genetically determined. Other isoantigens (now part of the HLA-B) were discovered by van Rood. Another important achievement was the microlymphocyte toxicity test introduced by Terasaki and McClelland in 1964, which is now the most frequently used method (Figs. 6.8 and 6.9). Subsequently the number of detected leukocyte antigens increased rapidly, and in 1965 it was suggested that most of these antigens were components of the same genetic system. At the histocompatibility workshop in 1967, 16 different teams typed identical samples from Italian families. Here the basic relationships among the different antigens were established. Finally, Kissmeyer-Nielsen [42] proposed the hypothesis of two closely linked loci (now A and B) controlling two series of alleles.

More recently, especially since the PCR technique became available, scientists study MHC genes directly at the DNA level. This has led to a splitting up of serologically defined gene loci, at both the class I and class II antigens (HLA D-DR; see Fig. 6.10) There are over 2,100 alleles described at the MHC locus with the HLA-B gene alone having 728 alleles (http://www.ebi.ac.uk/imgt/hla/stats.html), and several haplotypes have been completely sequenced [34].

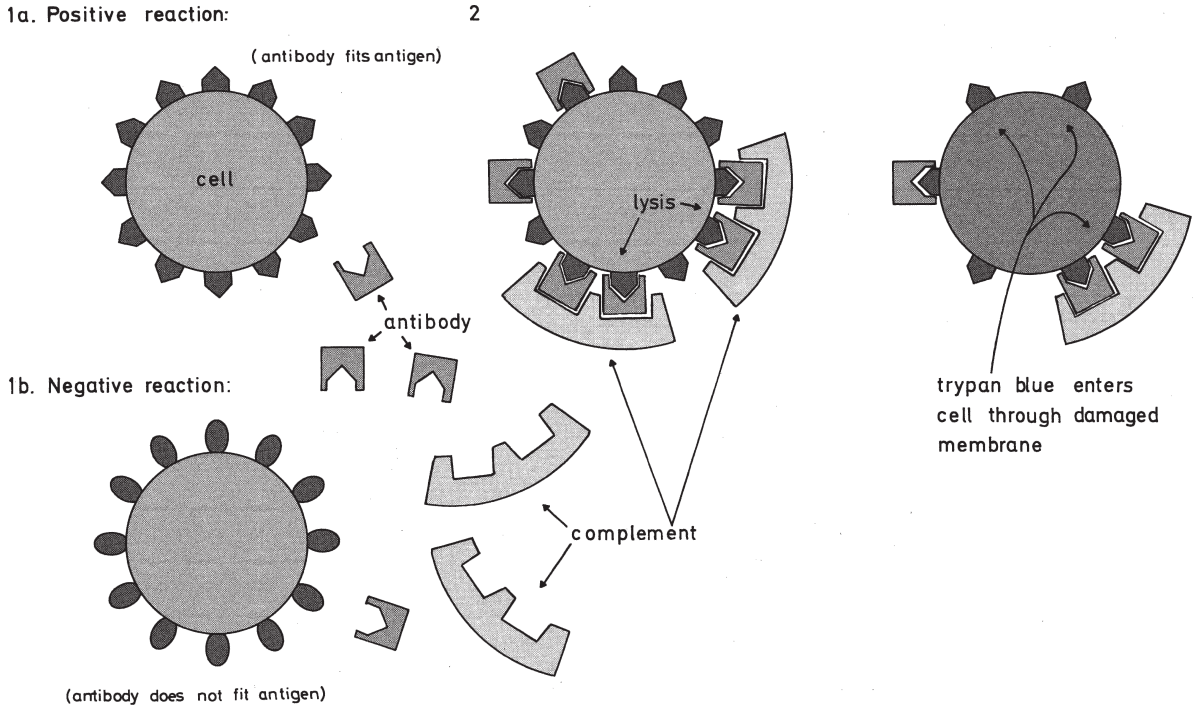

Fig. 6.8 Principle of the lymphocytotoxicity test: A cell having an appropriate antigen reacts with a specific antibody and complement. As a result, trypan blue enters the cell through the damaged membrane and indicates that the cell surface antigen has been recognized by a specific antibody

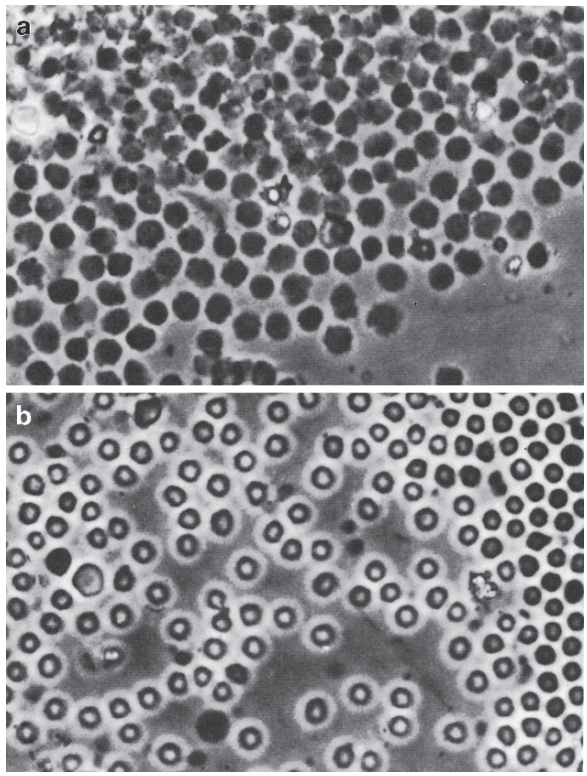

Fig. 6.9 (**a**, **b**) Lymphocytotoxicity test. (**a**) Positive reaction. (**b**) Negative reaction. Positive reaction is indicated by staining of the cells. (Courtesy of Dr. J. Greiner)

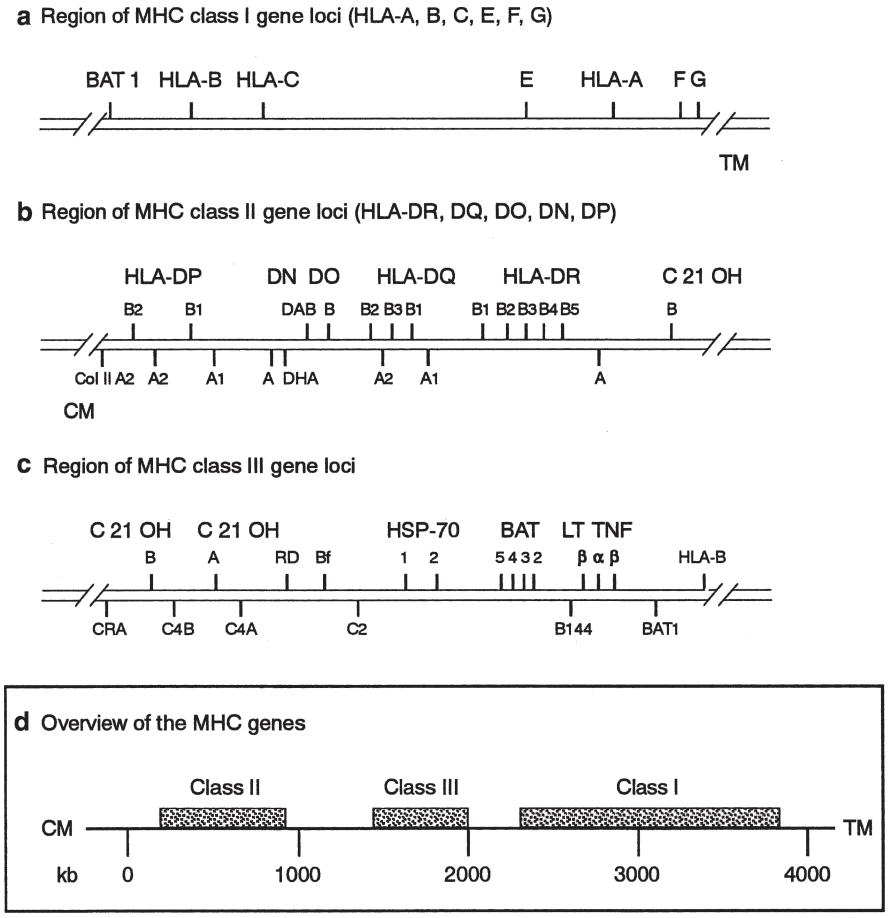

Fig. 6.10 (**a**–**d**) Sequence MHC class I, II, and III genes on the short arm of chromosome 6. (**a**–**c**) Detailed view of the three regions, together with their subregions. (**d**) Overview of approx. 4,000 kb. *TM,* Telomere; CM, centromere. *Letters, numbers, and their combinations,* the genes and their subregions. Genes of class III (HSP-70, DAT, C2/OH, CT) have no direct functional relationship with the immune response. From Albert [1]

6.2.5.2 Main Components of the MHC on Chromosome 6

The linkage group of the MHC is presented in Fig. 6.10. There are now three classes of MHC antigens. As revealed by studies using mainly molecular methods, each class can be subdivided into a great number of subclasses which are not described here (for details see [103]). Class I comprises the HLA-B, HLA-C, and HLA-A loci (in this order). In class II the HLA-D loci are found together with some other related, transcribed areas. Between these two classes a heterogeneous group of genes is located which have been named MHC class III despite the fact that at least some of these genes, such as those for 21-hydroxylase appear to have no functional relationship to the MHC system, and many non-HLA genes are located between the class I genes.

The function of this system has been elucidated; it plays an important role in the immune response. Here only some genetic aspects are considered.

The concept of four series of alleles is based on the following lines of evidence:

(a) No individual possesses more than two antigens from any of the series.
(b) Recombination between these series has been observed, for example, between the loci for A and B series, 40 crossovers among 4,614 meiotic divisions were described up to 1975, giving a combined (♀+♂) recombination frequency of 40/4614 = 0.0087 = 0.87 cM. Ten A-B recombinants

informative for the C series have been reported. In eight of these the C antigen followed B, and in two it followed A. Therefore C is located between A and B, closer to B, a fact confirmed by molecular studies.

(c) When two antigens from the same series are present together in a parent, he or she always transmits one of them – never both or none – to the child. The segregation ratio is 0.5, corresponding to a simple codominant mode of inheritance.

(d) Hardy–Weinberg proportions have been demonstrated for each of the allele series separately in large population samples.

(e) Serological cross reactions occur almost exclusively within the series, not between them. This points to a close biochemical relationship of the antigens within a given series.

(f) Complete sequencing of several MHC haplotypes confirms the presence of four class I genes.

6.2.5.3 Complement Components

Complement consists of a series of at least ten different factors present in fresh serum. The factors are called C1, C2, C3, etc., and C1 is activated by antibodies that react to their corresponding antigens. Then C1 activates C4, this activates C2, and so on. The end result of this "complement cascade" is damage to the cell membrane carrying the antigen and often lysis of the cell. Moreover, activated complement components have a number of other biological properties, such as chemotaxis or histamine release. They are important immune mediators in the body's defense against microbial infection.

The complement system can be activated not only via C1 (the classic pathway) but also via C3 through an alternative pathway involving the "properdin factors." The factor B(BF) acts as "proactivator" for C3.

For some of the complement factors hereditary deficiencies have been described, and polymorphisms are known. BF, C2, C3, and C4, are polymorphic. The loci of C2 and C4 A and B are in class III, together with the properdin factor B with the main alleles BF^F and BF^S. The locus for C3, on the other hand, is located on chromosome 19. Several regulatory factors such as complement factors H and I (CFH, CFI) are also located on autosomes [12, 90].

6.2.5.4 Significance of HLA in Transplantation

One of the main motives for rapid development of our knowledge of HLA antigens has been the hope of improving the survival rate of transplanted organs, primarily kidneys. Indeed, kidneys from HLA-identical and AB0-compatible siblings have a survival rate in the recipient almost equaling that of monozygotic twins. The survival rate is worse in unrelated recipients even if HLA matching is as perfect as possible, and AB0 compatibility is secured. This shows that, apart from the major histocompatibility system – the HLA system – there must be other systems of importance for graft survival. This is not surprising. A great number of such systems are known in the mouse. These systems lead to host-versus-graft reactions in almost all transplantations (Fig. 6.11). These reactions can be managed by immunosuppressive therapy. The chances for survival, and the survival times, of transplanted kidneys have increased substantially. The same is true for transplantation of other organs, such as heart, liver, bone marrow, and pancreas.

Considering the high degree of polymorphism and the low gene frequencies of HLA alleles, successful matching of potential recipients with donor kidneys from others than sibs requires large-scale international organizations. Once kidneys – or other transplantable organs – become available due to the accidental death of an individual, a center is notified in which persons in need of such an organ are registered, together with their HLA status. The donor is typed, and the recipient whose HLA status best fits receives the organ.

6.2.5.5 HLA: Linkage Disequilibrium

One of the most conspicuous properties of the HLA system is that some HLA alleles tend to occur more frequently together than expected by chance. Table 6.1 shows some examples. The A1,B8 haplotype, for example, occurs about five times as often as expected.

Consider two alleles at two linked loci, with frequencies p_1 and p_2. With free recombination between them their combined frequency, i.e., the haplotype frequency h, should be $p_1 \times p_2$. If such a result is obtained, the two loci are said to be in linkage equilibrium. If the haplotype frequency h is higher than expected with free recombination, there is linkage disequilibrium (Δ, deviation from linkage equilibrium), which is often

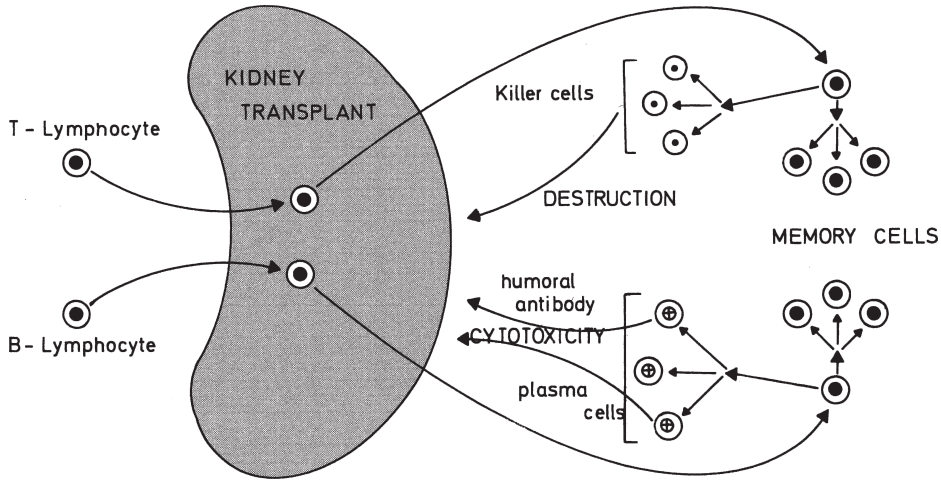

Fig. 6.11 Simplified diagram of the activation of the immune system by a kidney allotransplant. The transplant is recognized as foreign to the host organism by its T and B lymphocytes. This leads to activation of cellular and tumor immune response. From Svejgaard et al. [100]

Table 6.1 Linkage disequilibrium (gametic association; from Svejgaard et al. [100])

Haplotype			Frequency (%)	
A	B	D	Observed	Expected
A1	B8		9.8	2.1
A3	B7		5.4	2.1
	B8	Dw3	8.6	1.4
	B7	Dw2	3.9	1.8

The expected haplotype frequencies were calculated under the assumption of no association

Table 6.2 Association of HLA-A1 and B8 in unrelated Danes (2×2 table; from Svejgaard et al. [100])

	Number of individuals		
	$B8^+$	$B8^-$	Total
$A1^+$	376	235	611
$A1^-$	91	1,265	1,356
Total	467	1,500	1,967

First antigen	Second antigen	+/+ a	+/– b	–/+ c	–/– d	Total *n*
A1	B8	376	235	91	1,265	1,967

where, for example, +/– means number of individuals possessing the first character (A1) and lacking the second (B8). The χ^2 is:

$$\chi_1^2 = \frac{(ad - bc)^2 N}{(a+b)(c+d)(a+c)(b+d)} = 699.4$$

corresponding to the correlation coefficient:

$$r = \sqrt{\chi^2 / n} = \sqrt{699.4 / 1967} = 0.60$$

Gene frequencies for A1 and B8 can be calculated by Bernstein's formula:

$$p = 1 - \sqrt{1 - \alpha}$$

(where a is the antigen frequency) as 0.170 and 0.127, respectively. The Δ value can be calculated by the formula

$$\Delta = \sqrt{\frac{d}{n}} - \sqrt{\frac{(b+d)(c+d)}{n^2}} = 0.077$$

Thus, the frequency of the HLA-A1, B8 haplotype is

$$h_{A1,B8} = p_{A1} p_{B8} + \Delta_{A1,B8} = 0.170 \times 0.127 + 0.077 = 0.099.$$

symbolized as $D = h - p_1 p_2$. Haplotype and gene frequencies can be estimated from family and population data. In families the haplotypes of parents can in most cases be derived from those of their children (Table 6.2).

In the HLA system the deviations from linkage equilibrium are indeed striking. The situation is similar to that encountered with the Rh system (Sect. 6.2.4), but there is one important difference: in the Rh system only one case of recombination has been discovered, whereas many cases are known for the HLA system. Hence, genetic data point to much closer linkage in the Rh system than among the MHC genes. This conclusion has been corroborated by molecular studies in both systems (see above).

The observation of linkage disequilibrium – together with identification of immune response (Ir) genes in the mouse – initiated the investigations of HLA associations with diseases.

Linkage disequilibrium may have either of two main causes:

1. Two populations homozygous for different haplotypes mixed a relatively short time ago, and repeated

crossing over at a low rate has so far not been sufficient to lead to random distribution of alleles.
2. Certain combinations of alleles on closely linked gene loci caused a selective advantage for their bearers and have therefore been preserved.

To be able to decide between these two possibilities Bodmer [8] calculated how long a linkage disequilibrium would need to disappear in a random mating population.

For these calculations he used the work of Jennings [39], according to which Δ decreases to zero at a rate of $1-\Theta$ per generation, where Θ is the recombination fraction between the two loci. Between the HLA-A and HLA-B loci Θ was found to be of the order of magnitude of 0.008. Taking linkage disequilibrium between HLA-A1 and B8 as an example, Δ values of about 0.06–0.1 have been found in European populations. On the other hand, Δ values between 0.01–0.02 are not statistically significant with reasonable sample sizes. Therefore it is meaningful to examine how many generations are needed to reduce Δ from 0.1 by a factor of 5 to 0.02.

Using the above principle of Jennings, we obtain:

$$(1-\theta)^n(1-0.008)^n = 1/5; n \approx 200$$

This means that Δ would be reduced to an insignificant value within about 200 generations of random mating, i.e., 5,000 years, taking a generation as around 25 years.

This period is approximately the length of time since agriculture first came to parts of northern Europe and is certainly a very short time considering the evolutionary life span of the human species. The fact that such a significant Δ could be eroded in so short a time in the absence of selection suggests at least that this particular combination of HLA-A1, B8 is being maintained at its comparatively high frequency by some sort of interactive selection [50]. We consider it likely that selection will also be found to explain some of the other common cases of linkage disequilibrium and that the effect of recent population mixture will be shown to be of minor importance. Certain haplotypes seem to have a selective advantage that keeps them more frequent than others. This selective advantage, on the other hand, cannot be directly related to the diseases for which associations have been shown so far, as they are too rare. Besides, the onset of most of them is usually delayed until after the age of reproduction. Infectious diseases have probably been the most important selective forces for maintaining the MHC polymorphism as well as linkage disequilibrium. This topic is discussed in Sect. 16).

6.2.5.6 The Normal Function of the System

The HLA determinants are localized at the surface of the cell and are strong antigens. They exhibit the most pronounced polymorphism of expressed genes known so far in humans, with abundant linkage disequilibrium. Disease associations have been shown between HLA antigens and diseases for which an autoimmune mechanism had previously been suspected. Furthermore, similar systems are known in all other mammals examined so far (see [1, 43, 114]). Finally, there is close linkage with other loci concerned with the immune response. All this evidence together is very suggestive of a system that regulates the contact of cells with their environment. In recent years, this function has been elucidated in detail. These genes are important mediators of the immune reaction. Such cell recognition mechanisms may be important in embryonic development and differentiation, especially when they are present on only certain cell types. However, such hypothesis would not explain the selective advantage of the high degree of polymorphism in this system.

Another possible function is protection against viral or bacterial infection. Antigenic material of human origin may be incorporated in the outer membrane of the virus, which is thereby made less recognizable to another human host. However, if the virus contains MHC material from a genetically different individual, it is more readily inactivated by the immune system. Such a mechanism would also explain why the extreme polymorphism of the MHC system has a selective advantage. Further elucidation of the MHC will teach us a great deal about how the organism handles its interaction with the environment. This knowledge is important to our understanding of how natural selection has shaped our genetic constitution in the past, and how recent changes in our environment may influence it in the future.

To broaden the empirical basis for such understanding, however, it may be useful to ask whether there are other examples in nature of such gene clusters with related functions? Can their analysis provide us with hints for a better understanding of the MHC cluster? There is indeed one such example that has been analyzed very carefully – mimicry in butterflies. It cannot

be described here for lack of space, since it has no direct relationship with human genetics. But for the reader interested in more general, philosophical aspects of science, it is highly interesting showing how certain general principles may be used by nature in quite different contexts (see also earlier editions of this book).

6.2.6 Unequal Crossing Over

6.2.6.1 Discovery of Unequal Crossing Over

In the early years of work with *Drosophila* some authors observed that the bar mutation, an X-linked dominant character, occasionally reverts to normal, whereas in other cases homozygotes for the allele produce offspring with a new and more extreme allele, later called "double bar." Sturtevant [99] showed that this peculiar behavior is not due to mutations but to unequal crossing over, producing, on the one hand, a chromosome with two bar loci (double bar) and, on the other, a chromosome with no bar locus at all. When the giant salivary chromosomes of *Drosophila* permitted visual testing of genetic hypotheses, Bridges [11] showed that the simple, dominant bar mutation is caused by a duplication of some chromosomal bands. The reversion corresponds to the unduplicated state, whereas double bar is caused by a triplication of that band. Both reversion and triplication can be produced by a single event of unequal crossing over. Bridges did not yet formulate clearly the obvious reason for this event: the mispairing of "structure-homologous" but not "position homologous" chromosome sites (Fig. 6.12).

6.2.6.2 Unequal Crossing Over in Human Genetics

Haptoglobin [10], a transport protein for hemoglobin, is found in the blood serum and shows a polymorphism, the most common alleles being HP^{1F}, HP^{1S} and HP^{2}. Smithies et al. [94] discovered that the allele HP^{2} is almost twice the length of each of the two alleles HP^{1F} and HP^{1S}, as evidenced by the composition of its polypeptide chain. In the HP^{2} chain the amino acid sequence of the HP^{1} chain is repeated almost completely. They concluded that the HP^{2} allele might have been produced by gene duplication. Moreover, they predicted that unequal crossing over might again occur with a relatively high probability between HP^{2} alleles, producing, on the one hand, an allele similar to HP^{1}

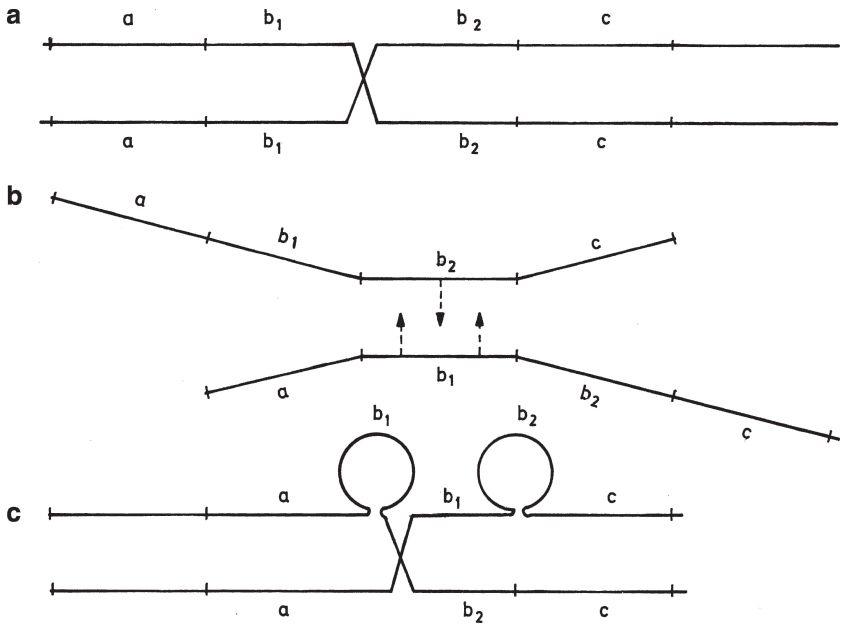

Fig. 6.12 (**a**–**c**) The principle of unequal crossing over. (**a**) Normal pairing and crossing over. The two genes b_1 and b_2 are assumed to have very similar DNA sequences. (**b**) Genes b_1 and b_2 are pairing. This leads to a shift of the two homologous chromosomes relative to each other. (**c**) Such pairing requires formation of two loops in the upper chromosome

and, on the other, an allele comprising the genetic information almost in triplicate. Repeated occurrence of this event might lead to still longer alleles and hence to a polymorphism of allele lengths in the population. Indeed, such alleles have occasionally been observed and are known as Johnson-type alleles [96].

There is an essential difference between the first unique event that produces the almost double-sized gene (for example, HP^2) from a single gene HP^1, and the unequal but homologous crossing over that becomes possible as soon as the first duplicated allele is present in the population [47].

6.2.6.3 First Event

Given a pair of homologous chromosomes, both partner chromosomes consist of largely identical sequences of nucleotides. Normally these partner chromosomes pair at meiosis, and there can be no unequal crossing over. To allow mispairing and thus unequal crossing over, an initial duplication is necessary. Mechanisms for such a duplication are known in cytogenetics, the simplest being two breaks at slightly different sites in adjacent homologous chromatids during meiosis and subsequent crosswise reunion. Another mechanism would be mispairing due to homology of short base sequences in nonhomologous positions. Our present knowledge of the structure of DNA sequences suggests ample opportunities for such a mispairing (slippage).

If the sites of breakage are separated only by the length of one structural gene, this event results in two gametes that do not contain this gene at all, together with two others containing it in duplicate (Fig. 6.13). The gametes containing a relatively large deletion have a high risk of not being transmitted because of lethality of the ensuing embryo. On the other hand, a gamete with the duplication is likely to develop into a diploid

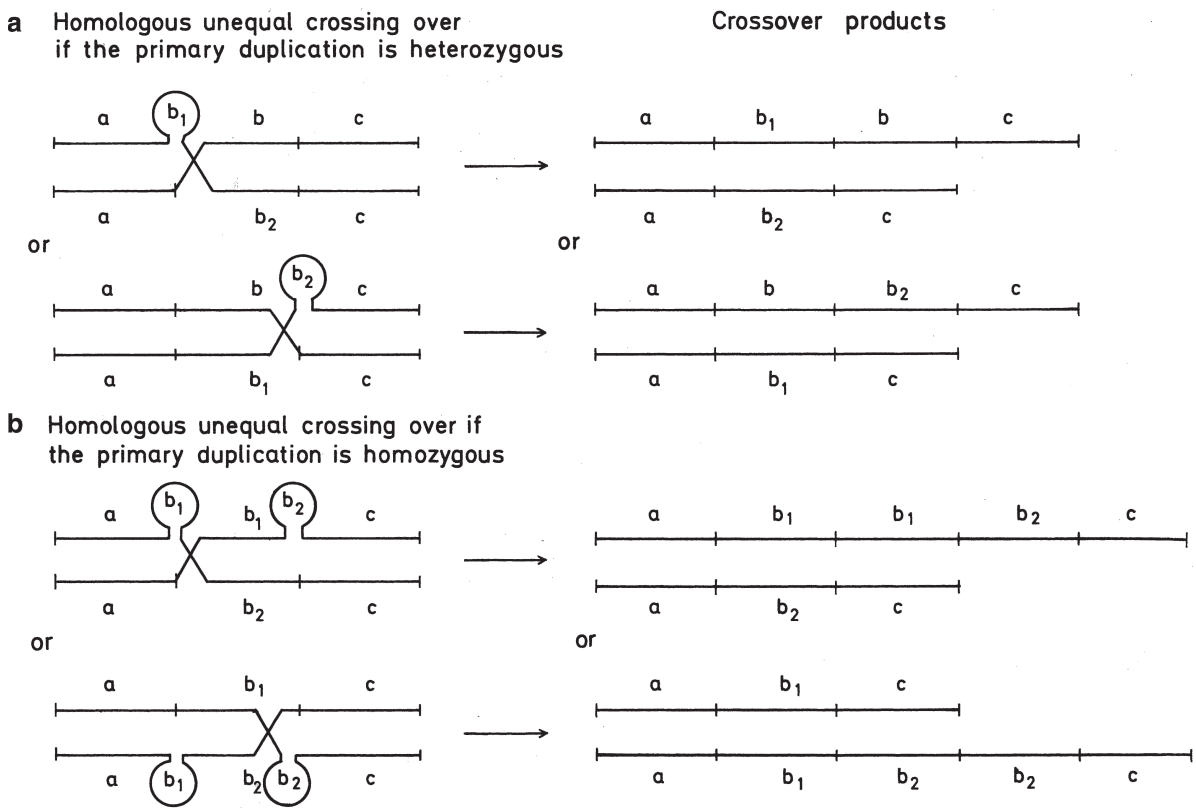

Fig. 6.13 (**a**, **b**) Unequal crossing over between structure-homologous but not position-homologous genes. (**a**) Unequal crossing over always leads to one crossover product with two genes b (b_1b or bb_2) and to another with only one gene (**b**). Formation of larger allele sequences becomes possible if the primary duplication is homozygous. In this case a chromosome with three alleles b (b_1 b_1 b_2 or b_1 b_2 b_2) may be formed. From Krüger and Vogel [47]

individual, providing for the first time a chance for mispairing of homologous sequences and therefore for unequal crossing over.

6.2.6.4 Consequences of Unequal Crossing Over

The consequences are seen in Fig. 6.13. As long as the duplication remains heterozygous, all gametes contain either one or two copies of the duplicated gene. When the duplication becomes homozygous, however, larger allele sequences may be formed. Unequal crossing over may lead, on the one hand, to gametes with only one copy and, on the other, to gametes containing three, and in subsequent generations, more than three copies (Figs. 6.13 and 6.14).

If the probability of unequal crossing over is not too low, high variability is soon found in the number of homologous chromosomal segments that resemble each other in structure but not in position. If selection favors a certain number of such chromosomal segments, which may be as small as a single gene, this number soon becomes the most common one. Selection relaxation leads to an increase in variability in both directions: the proportion of individuals with a very high number of such genes as well as those with a low gene number gradually increases [47]. Another genetic mechanism resembling unequal crossing over in some aspects is gene conversion where nonreciprocal products result.

Other examples besides the haptoglobin genes are the closely linked hemoglobin β- and δ-genes, the color vision pigment locus, and the natural killer cell receptor (KIR) genes [40]. Here the Lepore-type mutants, the X-linked color vision genes, and the diversity of KIR haplotypes are caused by unequal crossing over. Moreover, there are many examples for moderately or highly repetitive DNA sequences within which unequal crossing over should be possible. The presence of short repetitive DNA sequences such as minisatellites (Sect. 2.1.2) provides ample opportunities for pairing "slippage," leading to unequal crossing over. The high mutation rate within such areas (sometimes even a few percent per meiosis (Sect. 3.5) a well as the resulting huge interindividual variability show that this is not merely a theoretical speculation. Other repeated DNA sequences are those coding for the immunoglobulins. Increasing knowledge of the functional significance of repeated DNA sequences will bring a better understanding of the significance of unequal crossing over.

Recently there have been described large DNA segments that are duplicated in tandem or on other chromosomes. These "segmental duplications" can be up to 1 Mb is size or greater. Unequal crossing over between adjacent duplications is the basis for diGeorge/venocardiofacial disorder, Williams syndrome, and several other diseases. This class of diseases has been termed "genomic disorders" [89].

In fact, unequal crossing over, also referred to as "nonallelic homologous recombination," can give rise to numerous, recognizable microdeletion and microduplication syndromes and can significantly contribute to genome plasticity (see also Sect. 3.5.5).

6.2.6.5 Intrachromosomal Unequal Crossing Over

With structure-homologous but not position-homologous genes, such as those found in multigene families (Sect. 3.5.5), unequal crossing over becomes possible not only between homologous chromosomes but also between sister chromatids (intrachromosomal unequal crossing over). Theoretical considerations have shown that this process could have played a role in molecular evolution [41].

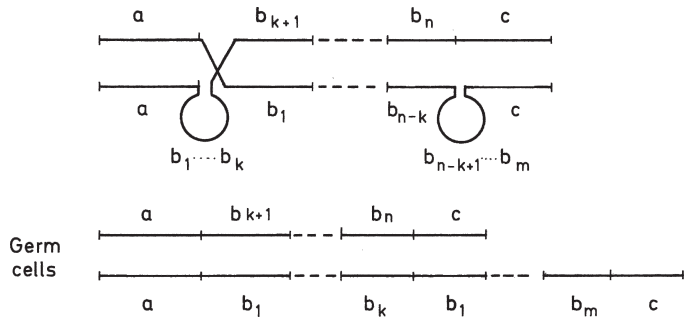

Fig. 6.14 The consequence of unequal crossing over. In subsequent generations chromosomes with (theoretically) unlimited numbers of a alleles may be formed. Unequal crossing over between any of them may lead to still larger (or still shorter) haplotypes. $b_1 \ldots b_k \ldots b_n$ refer to homologous genes

6.3 Conclusions

A few years after the rediscovery of Mendel's laws early in the twentieth century the first exception to Mendel's third law (independent segregation) was discovered: genes located sufficiently close to one other on the same chromosomes often segregate together – they are linked. The frequency of recombination increases with increasing distance between these genes. Genes on the same chromosomes but located far apart from each other, however, may even segregate independently if the distance between them is greater – these are syntenic, but not linked. A great number of genetic markers are available for localizing human genes, and statistical methods for assessing linkage in the human genome and determining the distance between gene loci have been developed. Methods from cell, biochemical, and molecular genetics have helped in localizing genes to specific chromosomes and chromosome segments and led to the molecular isolation of these genes. Such techniques make it possible to localize genes for both normal and abnormal traits and to define the nature of such genes by positional cloning. The identification of genes involved in susceptibilities to common diseases with complex causes by linkage studies remains a major challenge. However, the advent of a human genome sequence, comparative sequence, and a haplotype map, as well as large-scale association studies comparing normal subjects with patients, are leading to some progress in molecular understanding of complex disease.

While genes involved in the same biochemical pathways are seldom located close together, some clusters of closely linked genes exist that have related functions; the genes of the major histocompatibility complex, for example, have been analyzed particularly thoroughly.

References

1. Albert E (1993) Immungenetik. In: Gemsa R (ed) Lehrbuch der Immunologie. Thieme, Stuttgart
2. Antonarakis SE, Beckmann JS (2006) Mendelian disorders deserve more attention. Nat Rev Genet 7:277–282
3. Antonarakis SE (1998) Recommendations for a nomenclature system for human gene mutations. Nomenclature working group. Hum Mutat 11:1–3
4. Bateson W, Saunders PRG (1908) Confirmations and extensions of Mendel's principles in other animals and plants. Report to the Evolution Committee of the Royal Society, London
5. Bauer KH (1927) Homoiotransplantation von Epidermis bei eineiigen Zwillingen. Beitr Klin Chir 141:442–447
6. Bernstein F (1931) Zur Grundlegung der Chromosomentheorie der Vererbung beim Menschen. Z Indukt Abstamm Vererbungsl 57:113–138
7. Bishop MJ (1994) Guide to human genome computing. Academic, London
8. Bodmer WF (1972) Population genetics of the HL-A system: retrospect and prospect. In: Dausset J, Colombani J (eds) Histocompatibility testing. Munksgaard, Copenhagen, pp 611–617
9. 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
10. Bowman BH, Kurosky A (1982) Haptoglobin: the evolutionary product of duplication, unequal crossing over, and point mutation. Adv Hum Genet 12:189–261
11. Bridges CB (1936) The bar "gene", a duplication. Science 83:210
12. Briscoe D, Stephens JC, O'Brien SJ (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. J Hered 85:59–63
13. Cantor RM, Rotter JI (1992) Analysis of genetic data: methods and interpretation. In: King A, Rotter JI, Motulsky AG (eds) The genetic basis of common diseases. Oxford University Press, New York, pp 49–70
14. Carrington M, Cullen M (2004) Justified chauvinism: advances in defining meiotic recombination through sperm typing. Trends Genet 20:196–205
15. Chadefaux B, Allord D, Rethoré MO, Raoul O, Poissonier M, Gilgenkrantz S, Cheruy C, Jérôme H (1984) Assignment of human phosphoribosylglycinamide synthetase locus to region 21 q 22.1. Hum Genet 66:190–192
16. Colin Y, Cherif-Zahar B, LeVankim C et al (1991) Genetic basis of the RhD-positive and the RhD-negative blood group polymorphism as determined by Southern analysis. Blood 78:2747–2752
17. Collins FS (1992) Positional cloning: Let's not call it reverse anymore. Nat Genet 1:3–6
18. Cox DR, Burmeister M, Price ER, Kim S, Myers RM (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245–250
19. Dausset J, Cann H, Cohen D et al (1990) Centre d'études du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. Genomics 6:575–577
20. Dean M (2003) Approaches to identify genes for complex human diseases: lessons from mendelian disorders. Hum Mut 22:261–274
21. Donahue RP, Bias WB, Renwick JH, McKusick VA (1968) Probable assignment of the Duffy blood-group locus to chromosome 1 in man. Proc Natl Acad Sci USA 61:949
22. Donis-Keller H, Green P, Helms C et al (1987) A genetic linkage map of the human genome. Cell 51:319–337
23. Ferguson-Smith MA (1966) X-Y chromosomal interchange in the aetiology of true hermaphroditism and of XX Klinefelter's syndrome. Lancet 2:475–476
24. Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 323:643–646

25. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229
26. Gerhard DS, Kidd KK, Kidd JR, Egeland JA (1984) Identification of a recent recombination event within the human b-globin gene cluster. Proc Natl Acad Sci USA 81:7875–7879
27. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ (2005) Ahuja SK (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307:1434–1440
28. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina C, Wallace MR, Sakaguchi AY, Young AB, Shoulson I, Bonilla E, Martin JB (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234–238
29. Gusella JF, Tanzi RE, Anderson MA, Hobbs W, Gibbons K, Raschtchian R, Gilliam TC, Wallace MR, Wexler NS, Conneally PM (1984) DNA markers for nervous system diseases. Science 225:1320–1326
30. Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. J Genet 12:101–109
31. Haldane JBS, Smith CAB (1947) A new estimate of the linkage between the genes for colour blindness and haemophilia in man. Ann Eugen 14:10–31
32. Harris H (1970) Cell fusion. Clarendon, Oxford
33. Harris H, Watkins JF (1965) Hybrid cells from mouse and man: artificial heterokyryons of mammalian cells from different species. Nature 205:640
34. Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, Almeida J, Forbes S, Gilbert JG, Halls K, Harrow JL, Hart E, Howe K, Jackson DK, Palmer S, Roberts AN, Sims S, Stewart CA, Traherne JA, Trevanion S, Wilming L, Rogers J, de Jong PJ, Elliott JF, Sawcer S, Todd JA, Trowsdale J, Beck S (2008) [Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics 60:1–8
35. Houwen RHJ, Baharloo S, Blankenship K (1994) Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 8:380–386
36. Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, Wilson RK, Page DC (2005) Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature 437:100–103
37. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR et al (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861
38. Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73
39. Jennings S (1917) The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. Genetics 2:97–154
40. Khakoo SI, Carrington M (2006) KIR and disease: a model system or system of models? Immunol Rev 214:186–201
41. Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge
42. Kissmeyer-Nielsen F (ed) (1975) Histocompatibility testing 1975. Munksgaard, Copenhagen
43. Klein J (1986) Natural history of the major histocompatibility complex. John Wiley and Sons, New York
44. Knippers R, Philippsen P, Schafer KP, Fanning E (1990) Molekulare Genetik, 5 th edn. Thieme, Stuttgart
45. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241–247
46. Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, Jonsdottir GM, Gudjonsson SA, Sverrisson S, Thorlacius T, Jonasdottir A, Hardarson GA, Palsson ST, Frigge ML, Gulcher JR, Thorsteinsdottir U, Stefansson K (2008) Sequence variants in the RNF212 gene associate with genome-wide recombination rate. Science 319:1398–1401
47. Krüger J, Vogel F (1975) Population genetics of unequal crossing over. J Mol Evol 4:201–247
48. Kurnit DM (1979) Down syndrome: gene dosage at the transcriptional level in skin fibroblasts. Proc Natl Acad Sci USA 76:2372–2375
49. Kutsenko AS, Gizatullin RZ, Al-Amin AN, Wang F, Kvasha SM, Podowski RM, Matushkin YG, Gyanchandani A, Muravenko OV, Levitsky VG, Kolchanov NA, Protopopov AI, Kashuba VI, Kisselev LL, Wasserman W, Wahlestedt C, Zabarovsky ER (2002) NotI flanking sequences: a tool for gene discovery and verification of the human genome. Nucleic Acids Res 30:316–3170
50. Lander ES et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
51. Landsteiner K, Wiener AS (1940) An agglutinable factor in human blood recognized by immune sera for rhesus blood. Proc Soc Exp Biol Med 43:223
52. Lange K, Boehnke M (1982) How many polymorphic genes will it take to span the human genome? Am J Hum Genet 34:842–845
53. Lange K, Page BM, Elston RC (1975) Age trends in human chiasma frequencies and recombination fractions. I. Chiasma frequencies. Am J Hum Genet 27:410–418
54. Lawrence S, Morton NE, Cox DR (1991) Radiation hybrid mapping. Proc Natl Acad Sci USA 88:7477–7480
55. Lee WH, Bookstein R, Hong F, Young LJ, Shew JY, Lee EY (1987) Human retinoblastoma susceptibility gene: cloning, identification, and sequence. Science 235:1394–1399
56. Lele KP, Penrose LS, Stallard HB (1963) Chromosome deletion in a case of retinoblastoma. Ann Hum Genet 27:171
57. Levine P, Stetson RE (1939) An unusual case of intragroup agglutination. JAMA 113:126–127
58. Lewis EB (1951) Pseudoallelism and gene evolution. Cold Spring Harbor Symp Quant Biol 16:159–174
59. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PV, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77:68–693
60. McClintock B (1944) The relation of homozygous deficiencies to mutations and allelic series in maize. Genetics 29:478–502

61. McKusick VA (1995) Mendelian inheritance in man, 11 th edn. Johns Hopkins University Press, Baltimore
62. Migeon BR, Miller SC (1968) Human-mouse somatic cell hybrids with single human chromosome (group E): link with thymidine kinase activity. Science 162:1005–1006
63. Mittwoch U (1992) Sex determination and sex reversal: genotype, phenotype, dogma and semantics. Hum Genet 89:467–479
64. Mohr J (1954) A study of linkage in man. Munksgaard, Copenhagen Opera ex domo biologiae hereditariae humanae universitatis hafniensis 33
65. Morgan TH (1910) Sex-limited inheritance in drosophila. Science 32:120–122
66. Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318
67. Morton NE (1993) Genetic epidemiology. Annu Rev Genet 27:521–538
68. Murray JC, Mills KA, Demopulos CM, Hornung S, Motulsky AG (1984) Linkage disequilibrium and evolutionary relationships of DNA variants (RFLPs) at the serum albumin locus. Proc Natl Acad Sci USA 81:3486–3490
69. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. Science 310:32–324
70. Ohno S (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. Semin Cell Dev Biol 10:517–522
71. Ohta T (1993) Diversifying selection, gene conversion, and random drift: interactive effects on polymorphism at MHC loci.
72. Ott J (1983) Linkage analysis and family classification under heterogeneity. Ann Hum Genet 47:311–320
73. Ott J (1990) Cutting a Gordian knot in the linkage analysis of complex human traits. Am J Hum Genet 46:219–221
74. Ott J (1991) Analysis of human genetics linkage. Johns Hopkins University Press, Baltimore
75. Page DC, Mosher R, Simpson EM et al (1987) The sex-determining region of the human Y chromosome encodes a finger protein. Cell 51:1091–1104
76. Penrose LS (1934/35) The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. Ann Eugen 6:133–138
77. Race RR, Sanger R (1969) Xg and sex chromosome abnormalities. Br Med Bull 25:99–103
78. Race RR, Sanger R (1975) Blood groups in man, 6 th edn. Blackwell, Oxford
79. Rappold GA (1993) The pseudoautosomal regions of the human sex chromosomes. Hum Genet 92:315–324
80. Reid DH, Parsons PH (1963) Sex of parents and variation of recombination with age in the mouse. Heredity 18:107
81. Renwick JH (1969) Progress in mapping human autosomes. Br Med Bull 25:65
82. Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–238
83. Risch N (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241
84. Risch N (1990) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. Am J Hum Genet 46:242–253
85. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517
86. Rommens JM, Iannuzzi MC, Kerem BS, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole LJ, Kennedy D, Hidaka N, Zsiga M, Buchwald M, Riordan JR, Tsui LC, Collins FS (1989) Identification of the cystic fibrosis gene: Chromosome walking and jumping. Science 245:1059–1065
87. Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet 8:424–436
88. Sekido R, Lovell-Badge R (2009) Sex determination and SRY: down to a wink and a nudge? Trends Genet 25:19–29
89. Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 13:R57–R64
90. Shiang R, Murray JC, Morton CC, Buetow KH, Wasmuth JJ, Olney AH, Sanger WG, Goldberger G (1989) Mapping of the human complement factor I gene to 4q25. Genomics 4:82–86
91. Sinclair AH, Berta P, Palmer MS et al (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240–244
92. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837
93. Smith SM, Penrose LS, Smith CAB (1961) Mathematical tables for research workers in human genetics. Churchill, London
94. Smithies O, Connell GE, Dixon GH (1962) Chromosomal rearrangements and the evolution of haptoglobin genes. Nature 196:232
95. Snell GD, Dausset J, Nathenson S (1977) Histocompatibility. Academic, New York
96. Sørensen H, Dissing J (1975) Association between the $C3^F$ gene and atherosclerotic vascuolar diseases. Hum Hered 25:279–283
97. Steinberg AG (1965) Evidence for a mutation or crossing over at the Rh-locus. Vox Sang 10:721
98. Stern C (1973) Principles of human genetics, 3rd edn. Freeman, San Francisco
99. Sturtevant AH (1925) The effects of unequal crossing over at the bar locus in drosophila. Genetics 10:117
100. Svejgaard A, Hauge M, Jersild C, Platz P, Ryder LP, Staub Nielsen L, Thomsen M (1979) The HLA system. An introductory survey, 2nd edn. Karger, Basel
101. Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C, Dicks E, Menzies A, Stephens P, Blow M, Greenman C, Xue Y, Tyler-Smith C, Thompson D, Gray K, Andrews J, Barthorpe S, Buck G, Cole J, Dunmore R, Jones D, Maddison M, Mironenko T, Turner R, Turrell K, Varian J, West S, Widaa S, Wray P, Teague J, Butler A, Jenkinson A, Jia M, Richardson D, Shepherd R, Wooster R, Tejada MI, Martinez F, Carvill G, Goliath R, de Brouwer AP, van Bokhoven H, Van Esch H, Chelly J,

Raynaud M, Ropers HH, Abidi FE, Srivastava AK, Cox J, Luo Y, Mallya U, Moon J, Parnau J, Mohammed S, Tolmie JL, Shoubridge C, Corbett M, Gardner A, Haan E, Rujirabanjerd S, Shaw M, Vandeleur L, Fullston T, Easton DF, Boyle J, Partington M, Hackett A, Field M, Skinner C, Stevenson RE, Bobrow M, Turner G, Schwartz CE, Gecz J, Raymond FL, Futreal PA, Stratton MR (2009) A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. Nat Genet 41:535–543

102. Tiepolo L, Zuffardi O (1976) Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome. Hum Genet 34:119–124
103. Trowsdale J, Ragoussis J, Campbell JD (1991) Map of the human MHC. Immunol Today 12:429–467
104. Venter JC et al (2001) The sequence of the human genome. Science 291:1304–1351
105. Vogt P, Keil R, Kirsch S (1993) The "AZF" function of the human Y chromosome during spermatogenesis. In: Sumner AT, Chandley AC (eds) Chromosomes today, vol 2. Chapham and Hall, London, pp 227–239
106. Wainwright BJ, Scambler PJ, Schmidtke J, Watson EA, Law HY, Farrall M, Cooke HJ, Eiberg H, Williamson R (1985) Localization of cystic fibrosis locus to human chromosome 7cen-q22. Nature 318:384–385
107. Wagner FF, Flegel WA (2000) RHD gene deletion occurred in the Rhesus box. Blood 95:3662–3668
108. Weiss MC, Green H (1967) Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. Proc Natl Acad Sci USA 58:1104–1111
109. Weitkamp LR (1972) Human autosomal linkage groups. In: Proceedings of the 4 th international congress of human genetics, Paris 1971. Excerpta Medica, Amsterdam, pp 445–460
110. White R, Lalouel J-M (1987) Investigation of genetic linkage in human families. Adv Hum Genet 16:121–228
111. White R, Leppert M, Bishop DT, Barker D, Berkowitz J, Brown C, Callahan P, Holm T, Jerominski L (1985) Construction of linkage maps with DNA markers for human chromosomes. Nature 313:101–105
112. White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel JM, Dean M, Vande Woude G (1985) A closely linked genetic marker for cystic fibrosis. Nature 318:382–384
113. Wolf U, Schempp W, Scherer G (1992) Molecular biology of the human Y chromosome. Rev Physiol Biochem Pharmacol 121:148–213
114. Yuhki N, Beck T, Stephens R, Neelam B, O'Brien SJ (2007) Comparative genomic structure of human, dog, and cat MHC: HLA DLA, and FLA. J Hered 98:390–399