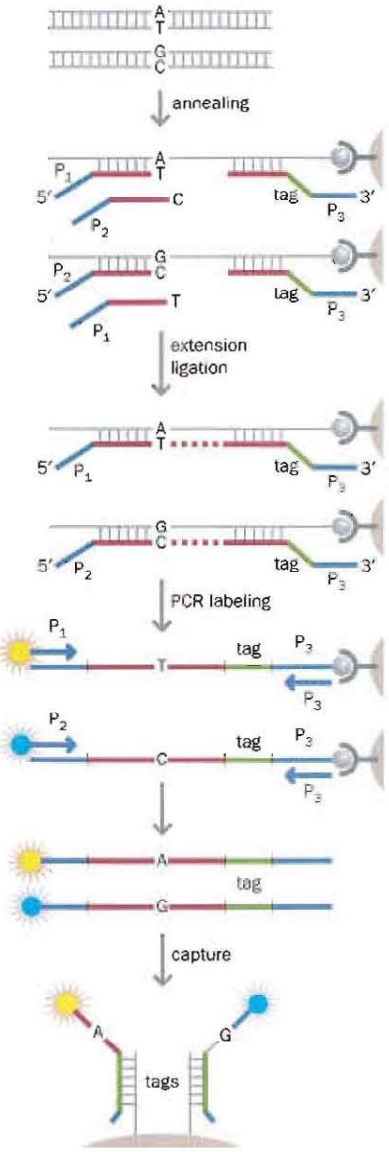


Figure 18.14 SNP genotyping by the Illumina GoldenGate® method. Two allele-specific (P_1 and P_2) oligonucleotides and one locus-specific (P_3) oligonucleotide are used, with a combination of allele-specific primer extension and DNA ligation to provide reliable discrimination of alleles. As with the MIP system (see Figure 18.13), ligation creates a PCR substrate that is amplified with universal dye-labeled primers, and the products are identified through hybridization of the tag sequence to an array—in this case, a bead array. [From Syvanen AC (2005) *Nat. Genet.* 37 (Suppl), S5–S10. With permission from Macmillan Publishers Ltd.]

amplified with one of two universal dye-labeled primers (**Figure 18.14**). The LSO carries a tag sequence that is used to hybridize the PCR product to a specific bead in a bead array.

18.4 SOME SPECIAL TESTS

The techniques described previously are applicable to the detection of almost any variant in almost any DNA sequence. Some particular situations, however, either require special techniques or have special features that can be exploited to make testing easier or more effective. In this section, we describe several such cases.

Testing for whole-exon deletions and duplications requires special techniques

Generally, introns are much bigger than the exons they flank, and so random breakpoints in a gene will most probably lie within an intron. Thus, kilobase-scale partial deletions or duplications of a gene sequence, if they do not lie wholly within one intron, will most probably remove or duplicate one or more whole exons. Such changes will not be apparent when genomic DNA is amplified exon by exon. In heterozygous deletions of whole exons, the mutant allele will give no product. The test sees only the normal allele, and the result looks entirely normal. Similarly, duplications will not show up by these methods. PCR in its normal form is not quantitative, so any extra yield of product would not be noticed. RNA testing would solve the problem, but RNA may be difficult to obtain.

This is a problem that is particular to kilobase-scale deletions or duplications. Small deletions or duplications that lie wholly within an exon will be apparent when the exon is sequenced. Large-scale copy-number changes that might be anywhere in the genome are efficiently detected by array-CGH. However, this is an expensive technique for routine diagnosis, and the normal BAC arrays have low sensitivity for changes involving only a kilobase or so of DNA. When the candidate gene is already known, the technique of choice for detecting whole-exon deletions or duplications is **multiplex ligation-dependent probe amplification (MLPA)**.

The multiplex ligation-dependent probe amplification (MLPA) test

MLPA is a development of the oligonucleotide ligation assay (OLA) described above. As in the OLA, pairs of oligonucleotides hybridize to adjacent locations on the test DNA, and DNA ligase seals the gap between them to produce a single molecule. In the OLA the gap is positioned at a suspected variant nucleotide and ligation will occur only if there is an exact match; in MLPA the gap is positioned at a (hopefully) invariant nucleotide in the test DNA, so ligation will always take place if the template sequence is present. Ligation creates a PCR-amplifiable molecule (see Figure 18.9). The presence of a PCR product signals the presence of the appropriate matching sequence in the test DNA.

MLPA is a multiplex procedure with up to 45 probe pairs combined in some of the available kits. The individual PCR products are distinguished by length, so as to generate a series of peaks when the products are run on a DNA sequencer (**Figure 18.15**). Although the PCR is not truly quantitative (absolute amounts of product vary between probes), the relative amounts of a particular product in two samples should reflect the relative amounts of the template in the two samples. When results from test and control samples are compared, the relative peak heights give a direct readout of the dosage of each sequence in the test DNA relative to the control DNA. Normally, one ligation is used for each exon of a gene, allowing whole-exon deletions and duplications to be detected.

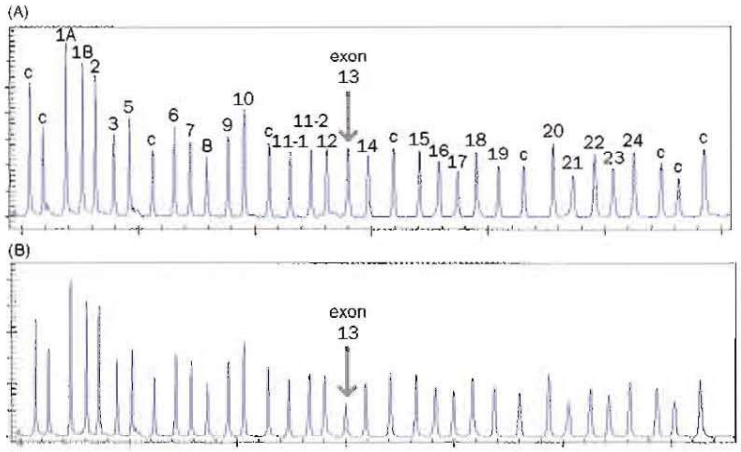


Figure 18.15 Using multiplex ligation-dependent probe amplification (MLPA) to identify an exon 13 deletion in the *BRCA1* gene. (A) Results from a control sample. Numbered peaks represent products from each exon; peaks labeled c are control probes. (B) The same analysis on DNA from a patient with breast cancer. In comparison with the control sample, the exon 13 peak is only half the size. (Courtesy of MRC Holland.)

Several other techniques can similarly identify whole-exon deletions and duplications in a candidate gene, but MLPA is the most widely used. The laboratory procedure is simple and the method has generally proved robust. Rare sequence variants that interfere with hybridization or ligation of the probes could lead to an exon being falsely scored as deleted. These could be checked by sequencing any suspect exon. MLPA probes are quite complex constructs (the stuffer fragment may be several hundred nucleotides long), and each individual probe needs to be carefully designed, so laboratories would normally buy commercial kits for their gene of interest. These are available for many of the genes in which deletions commonly cause clinical problems. Real-time PCR (see Box 8.5) would be an alternative for a gene for which no MLPA kit is available.

Dystrophin gene deletions in males

About 60% of DMD mutations are deletions of one or more exons of the dystrophin gene located on the X chromosome. MLPA is needed to detect female carriers of such deletions, but testing in affected males is simple: the deleted exon(s) will not amplify from the patient's DNA. In affected males, two multiplex PCR reactions (that shown in Figure 18.16 and one testing exons in the 5′ part of

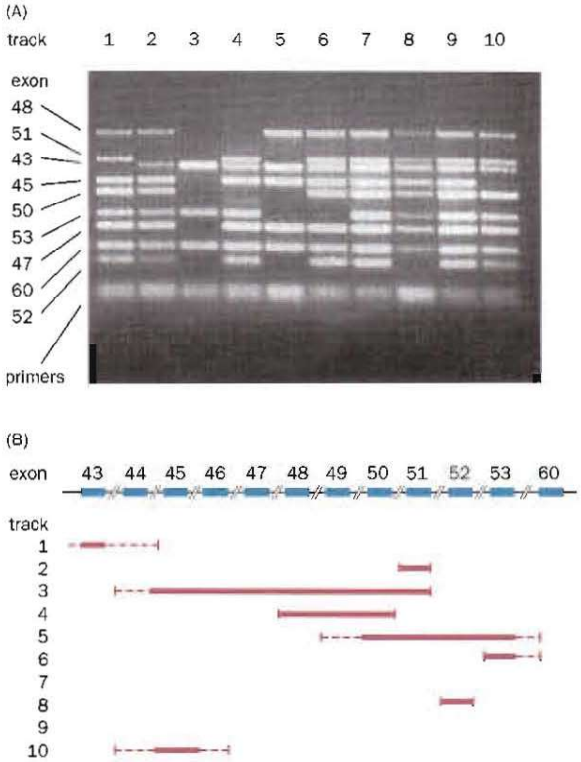


Figure 18.16 Multiplex screen for dystrophin deletions in males. (A) Products of multiplex PCR amplification of nine exons, using samples from 10 unrelated boys with Duchenne/Becker muscular dystrophy. PCR primers have been designed so that each exon, with some flanking intron sequence, gives a different-sized PCR product. (B) Interpretation of the results in (A). Solid lines show exons definitely deleted, dotted lines show the possible extent of deletion running into untested exons. No deletion is seen in samples 7 and 9; these patients may have point mutations, or deletions of exons not examined in this test. Exon sizes and spacing are not to scale. Compare with Figure 13.15. [(A) courtesy of R. Mountford, Liverpool Women's Hospital.]

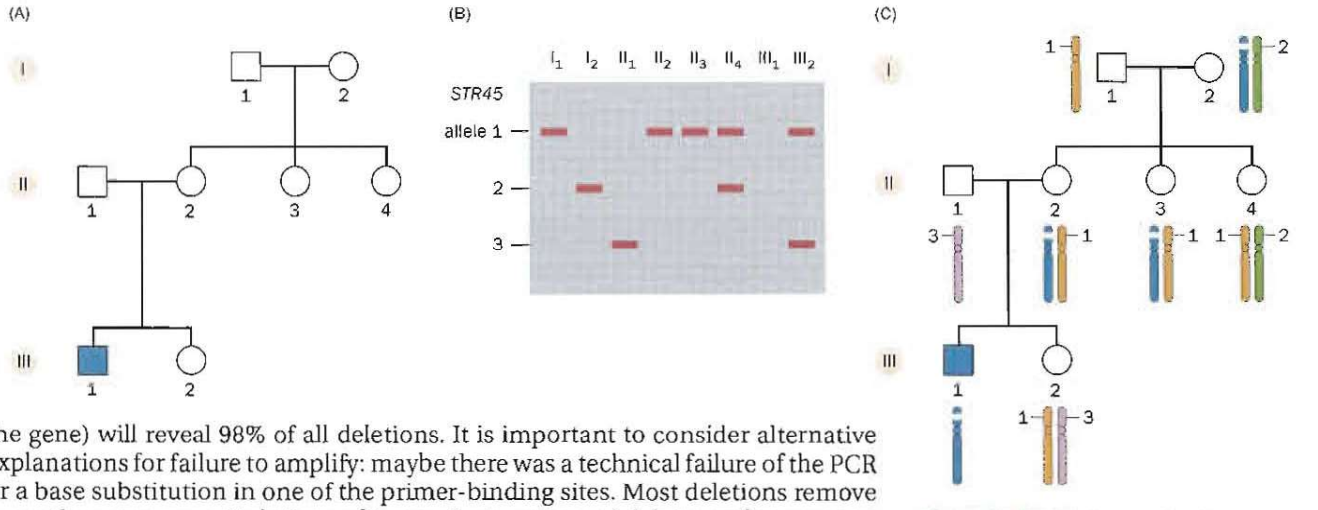


Figure 18.17 Deletion carriers in a DMD family revealed by apparent non-maternity. Pedigree (A) and results (B) of genotyping with the intragenic marker *STR45*. The affected boy, III_1, has a deletion that includes *STR45* (his lane on the gel is blank). His mother, II_2, and his aunt, II_3, inherited no allele of *STR45* from their mother, I_2, showing that the deletion is being transmitted in the family. I_2 is apparently homozygous for this highly polymorphic marker (lane 2), but in fact is hemizygous as a result of the deletion. The boy's other aunt, II_4, and his sister, III_2, are heterozygous for the marker and therefore do not carry the deletion. (C) Pedigree showing how the marker genotypes of individuals I_2, II_2, and II_3 are hemizygous as a result of the deletion.

the gene) will reveal 98% of all deletions. It is important to consider alternative explanations for failure to amplify: maybe there was a technical failure of the PCR or a base substitution in one of the primer-binding sites. Most deletions remove more than one exon. Deletions of just a single exon and deletions that seem to affect non-contiguous exons need confirming. Deletions can be confirmed by using alternative PCR primers or by MLPA. About 5% of dystrophin mutations are duplications of one or more exons, and detecting these requires MLPA in males as well as in females.

Apparent non-maternity in a family in which a deletion is segregating

If a deletion is segregating in a DMD family, genotyping females for microsatellites that map within the deletion may reveal *apparent non-maternity*, in which a mother has transmitted no marker allele to her daughter because of the deletion (Figure 18.17). In such families, non-maternity proves that a woman is a carrier, whereas heterozygosity (in the daughter or sister of a deletion carrier) proves that a woman is not a carrier. Several markers suitable for this purpose have been identified in the introns at deletion hotspots. The method works best in families in which there is an affected male in whom the deletion can first be defined. In principle, the same approach could be applied to tracking any other deletion, whether X-linked or autosomal, through a family. However, for any autosomal deletion, a quantitative method would be needed to identify any heterozygous deletion in the first place, and if such a method were available it would be simpler to use the same method to test other family members.

A quantitative PCR assay is used in prenatal testing for fetal chromosomal aneuploidy

For most diagnostic laboratories, a substantial part of the workload is prenatal testing for fetal chromosome anomalies. Women might request testing because they have had a previous chromosomally abnormal baby, because they (or their partner) have a balanced chromosomal abnormality that predisposes to fetal abnormality (see Chapter 2), or because of a suspect finding on an ultrasound scan. For all these cases, a full karyotype is usually prepared by standard cytogenetic techniques. Fetal cells are obtained by amniocentesis, usually at 14–20 weeks of gestation, or chorionic villus biopsy at 9–12 weeks (the exact date ranges vary from country to country).

However, the commonest single indication for fetal chromosome analysis is some combination of maternal age and maternal serum biochemistry that indicates an above-average risk of a fetal trisomy. The screening methods that identify these high-risk pregnancies are described in Chapter 19. For these women, fetal cells are not usually karyotyped. From the woman's point of view it takes far too long, and from the laboratory viewpoint it is far too expensive. The main risk is specifically for the three autosomal trisomies that are compatible with survival to term, namely trisomies 13, 18, and 21. These are usually checked by *QF-PCR*, a quantitative fluorescence-labeled multiplex PCR test. A multiplex PCR is performed with dye-labeled primers for several highly polymorphic microsatellite markers on each of chromosomes 13, 18, and 21. If a trisomy is present, the markers for that chromosome show either three peaks or two peaks in a 2:1 size ratio,

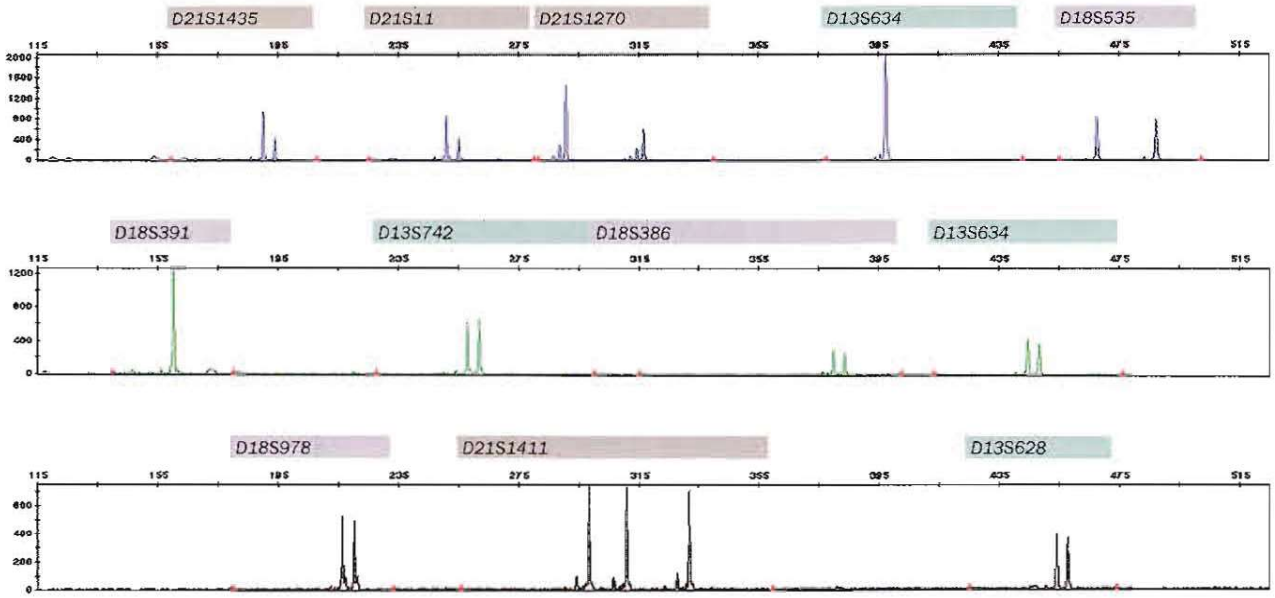


Figure 18.18 Prenatal diagnosis of trisomy 21 with QF-PCR. Fetal DNA, obtained by chorionic villus biopsy or amniocentesis, was amplified in a multiplex PCR reaction using dye-labeled primers for four highly polymorphic microsatellites from each of chromosomes 13, 18, and 21. The markers from chromosomes 13 and 18 each show two equal-sized peaks (if heterozygous) or one larger peak (if homozygous). Markers from chromosome 21 all show three peaks (*D21S1411*) or two peaks in a 2:1 size ratio (*D21S11*, *D21S1270*, and *D21S1435*). (Courtesy of Susan Hamilton, St Mary's Hospital, Manchester.)

depending on the informativeness of the marker (**Figure 18.18**). If there is any suspicion that the test DNA might be derived from contaminating maternal material, rather than from the fetus, this can be checked by comparing the genotypes with genotypes of DNA extracted from the mother's blood.

QF-PCR gives results in one day, in contrast with an average of 14 days for karyotyping cells from amniotic fluid. Mosaicism with the minor component less than about 15% is not reliably detected; nor, of course, are any chromosome abnormalities other than the specific trisomies targeted. There has been much debate about whether this specific focus on trisomies 13, 18, and 21 is an advantage or a disadvantage of the method. Some abnormalities will undoubtedly be missed; however, advocates of the QF-PCR approach point out that few of these are likely to result in a liveborn abnormal baby.

Some triplet repeat diseases require special tests

The expanded repeats that cause various neurological diseases (see Table 13.1) involve a special set of mutation-specific tests (**Figure 18.19**). The number of repeats is different in normal and in affected people, with a threshold for pathogenicity. For the polyglutamine repeat diseases such as Huntington disease, a single PCR reaction makes the diagnosis. Fragile X syndrome is more of a challenge. Normal (fewer than 50 repeats) and pre-mutation (50–200 repeats) alleles give clean PCR products, but full mutations have hundreds or thousands of CGG repeats and do not readily amplify by PCR, especially because of the high GC content. A full mutation in a male is easily recognized because the PCR product shows only a vague smear rather than a discrete band, but in a female the problem is more difficult. If a female sample gives only one band, this may be because she is homozygous for a particular repeat size in the normal range, or it may be because only one of her alleles amplifies, the other being a full expansion. This can only be settled by Southern blotting. Similarly, in people (male or female) who are mosaic for a full mutation and a pre-mutation, the full mutation would be missed without Southern blotting. Additionally, unlike the polyglutamine diseases, the fragile X expansion causes disease by loss of function, and occasional affected patients have deletions or point mutations that would be missed by testing just the repeat. Some of the other very large pathogenic expansions listed in Table 13.1 can be similarly difficult to identify reliably by PCR.

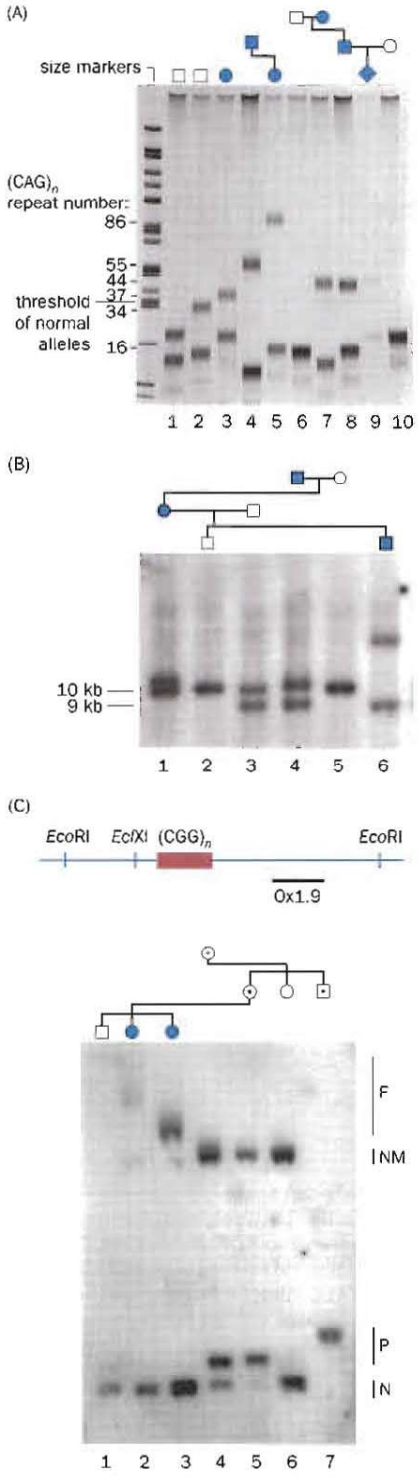


Figure 18.19 Laboratory diagnosis of trinucleotide repeat diseases. (A) Huntington disease. A fragment of the gene containing the $(CAG)_n$ repeat has been amplified by PCR and run out on a polyacrylamide gel. Bands are revealed by silver staining. The scale shows numbers of repeats; those greater than 36 are pathogenic. Lanes 1, 2, 6, and 10 are from unaffected people; lanes 3, 4, 5, 7, and 8 are from affected people. Lane 5 is a juvenile-onset case; her father (lane 4) had 55 repeats but she has 86. Lane 9 is an affected fetus, diagnosed prenatally. (B) Myotonic dystrophy. Southern blot of genomic DNA digested with *Eco*RI and hybridized to a labeled probe consisting of part of the *DMPK* gene. Bands of 9 kb or 10 kb are normal variants, the result of a nonpathogenic insertion-deletion polymorphism. The grandfather (lane 4) has cataracts but no other sign of myotonic dystrophy. His 10 kb band seems to be very slightly expanded compared with the same band from the normal male in lane 3, but this is not definite on the evidence of this gel alone. His daughter (lane 1) has one normal and one definitely expanded 10 kb band; she has classical adult-onset myotonic dystrophy. Her son (lane 6) has a massive expansion and the severe congenital form of the disease. (C) Fragile X. The DNA of the inactivated X chromosome in a female, and of any X chromosome carrying the full mutation, is methylated. Genomic DNA is digested with a combination of *Eco*RI and the methylation-sensitive enzyme *Ecl*XI, Southern blotted, and hybridized to Ox1.9 or a similar probe. The X chromosome in a normal male (lane 1) and the active normal X chromosome in a female (lanes 2, 3, 4, and 6) give a small fragment (labeled N). Unmethylated pre-mutation alleles (P) give a slightly larger band in lanes 4 and 5 (female pre-mutation carriers) and lane 7 (a normal transmitting male). Methylated (inactive) X-chromosome sequences do not cut with *Ecl*XI and give a much larger band (NM), and the fully expanded and methylated sequence gives a very large smeared band (F) because of somatic mosaicism. [(A) courtesy of Alan Dodge, St Mary's Hospital, Manchester. (B) and (C) courtesy of Simon Ramsden, St Mary's Hospital, Manchester.]

The mutation screen for some diseases must take account of geographical variation

The population genetics of recessive diseases is often dominated by founder effects or the effects of heterozygote advantage (see Chapter 3, p. 89). The resulting limited diversity of mutations in a population can make genetic testing much easier. β-Thalassemia and cystic fibrosis are good examples. For both these conditions, a very large number of different mutations in the relevant gene have been described, but in each case a handful of mutations account for most cases in any particular population. With β-thalassemia, DNA testing is not needed to diagnose carriers or affected people—orthodox hematology does this perfectly well—but it is the method of choice for prenatal diagnosis. Different mutations are predominant in different populations (Table 18.6). Provided that one has DNA samples from the parents and knows their ethnic origin, the parental mutations can often be found by using only a small cocktail of specific tests, after which the fetus can be readily checked.

In cystic fibrosis, the p.F508del mutation is the commonest in all European populations and is believed to be of ancient origin. However, the proportion of all mutations that are p.F508del varies, being generally high in the north and west of Europe and lower in the south. Testing for cystic fibrosis mutations divides into two phases. First, a limited number of specified mutations, always including p.F508del but otherwise population-specific, are sought with the methods described in Section 18.3. As Table 18.7 shows, there is no obvious natural cutoff in terms of diminishing returns on testing for specific mutations. If this phase fails to reveal the mutations, then, if resources allow, a screen for unknown mutations may be instituted, using the methods described in Section 18.2. Alternatively, gene tracking (see below) may be used. The impact of this diversity on proposals for population screening is discussed in Chapter 19.

Surprisingly often, when a recessive disease is particularly common in a certain population, it turns out that more than one mutation is responsible. An example is Tay–Sachs disease among Ashkenazi Jews, where there are two common *HEXA* mutations (see Table 18.5). It is difficult to explain this situation except by assuming there has been a long-continuing heterozygote advantage, favoring the accumulation of mutations in the gene.

TABLE 18.6 THE MAIN β-THALASSEMIA MUTATIONS IN DIFFERENT COUNTRIES

Population	Mutation	Frequency (%)	Clinical effect
Sardinia	codon 39 (C>T)	95.7	β^0
	codon 6 (delA)	2.1	β^0
	codon 76 (delC)	0.7	β^0
	intron 1-110 (G>A)	0.5	β^+
	intron 2-745 (C>G)	0.4	β^+
Greece	intron 1-110 (G>A)	43.7	β^+
	codon 39 (C>T)	17.4	β^0
	intron 1-1 (G>A)	13.6	β^0
	intron 1-6 (T>C)	7.4	β^+
	intron 2-745 (C>G)	7.1	β^+
China	codon 41/42 (delTCTT)	38.6	β^0
	intron 2-654 (C>T)	15.7	β^0
	codon 71/72 (insA)	12.4	β^0
	-28 (A>G)	11.6	β^+
	codon 17 (A>T)	10.5	β^0
Pakistan	codon 8/9 (insG)	28.9	β^0
	intron 1-5 (G>C)	26.4	β^+
	619 bp deletion	23.3	β^+
	intron 1-1 (G>T)	8.2	β^0
	codon 41/42 (delTCTT)	7.9	β^0
US black African	-29 (A>G)	60.3	β^+
	-88 (C>T)	21.4	β^+
	codon 24 (T>A)	7.9	β^+
	codon 6 (delA)	0.8	β^0

In each country, certain mutations are frequent because of a combination of founder effects and selection favoring heterozygotes. Data courtesy of J Old, Institute of Molecular Medicine, Oxford. β^0, complete absence of β-globin chains; β^+, β-globin present but in insufficient quantity. The nomenclature of mutations used here is nonstandard (see Box 13.2, p. 407), but is widely used for thalassemia.

Testing for diseases with extensive locus heterogeneity is a challenge

As sequencing becomes cheaper and faster, diagnostic laboratories are able to do more and more individual tests on a patient. With careful optimization of protocols, even genes with many exons can be fully analyzed in a cost-effective way. The focus is now moving on to conditions that are often Mendelian but can be caused by mutations in any one of a large number of genes. Profound mental retardation and congenital profound hearing loss are examples. Identifying the causative mutation in a patient may involve testing dozens of genes. The key question here is whether certain specific mutations, or maybe defects in one particular gene, account for a significant proportion of cases. Hearing loss is one

TABLE 18.7 DISTRIBUTION OF *CFTR* MUTATIONS IN 300 CF CHROMOSOMES FROM THE NORTHWEST OF ENGLAND

Mutation	Exon	Frequency (%)	Cumulative frequency (%)
p.F508del	10	79.9	79.9
p.G551D	11	2.6	82.5
p.G542X	11	1.5	84.0
p.G85E	3	1.5	85.5
p.N1303K	21	1.2	86.7
c.621+1G>T	4	0.9	87.6
c.1898+1G>A	12	0.9	88.5
p.W1282X	21	0.9	89.4
p.Q493X	10	0.6	90.0
c.1154insTC	7	0.6	90.6
c.3849+10 kb (C>T)	intron 19	0.6	91.2
p.R553X	10	0.3	91.5
p.V520F	10	0.3	91.8
p.R117H	4	0.3	92.1
p.R1283M	20	0.3	92.4
p.R347P	7	0.3	92.7
p.E60X	3	0.3	93.0
Unknown/private	–	7.0	100

p.F508del and a few of the other relatively common mutations are probably ancient and spread through selection favoring heterozygotes; the other mutations are probably recent, rare, and highly heterogeneous. Cystic fibrosis is more homogeneous in this population than in most others. See Box 13.2 for nomenclature of mutations. Data courtesy of Andrew Wallace, St Mary's Hospital, Manchester.

case in which this does happen. Worldwide, 20–50% of children with autosomal recessive profound congenital hearing loss have mutations in the *GJB2* gene that encodes connexin 26. Different specific mutations are common in different populations—c.30delG in Europe, c.235delC in East Asia, c.167delT in Ashkenazi Jews. A simple PCR test therefore provides the answer in a good proportion of cases. For the remainder, it would be necessary first to sequence the whole *GJB2* gene and then, if resources allowed, to examine a large number of other genes.

In many other cases, no one gene accounts for a significant proportion of cases. Learning difficulties present the ultimate challenge in this respect. Custom chips are being developed to allow a large panel of genes to be screened; alternatively, the new exon capture and ultra-fast sequencing technologies may allow dozens of genes to be sequenced at a reasonable cost. Technically, the challenge is identical to the problem of screening a person's DNA for large numbers of variants that confer susceptibility or resistance to common complex diseases. However, data interpretation presents different problems. For susceptibility screening, the problem is to know the combined risk from many variants, each of which modifies risk to only a small degree. Risks cannot be simply added or multiplied; combining them requires a quantitative model of the effect of each variant on overall cell biology. For heterogeneous Mendelian conditions this is not a problem: one mutation is responsible for the condition. The problem is the large number of unclassified variants that will undoubtedly be identified.

18.5 GENE TRACKING

Gene tracking was historically the first type of DNA diagnostic method to be widely used. It uses knowledge of the map location of the disease locus, but not knowledge about the actual disease gene. Most of the Mendelian diseases that form the bulk of the work of diagnostic laboratories went through a phase of gene tracking when the disease gene had been mapped but not yet cloned. Once the gene had been identified, testing moved on to direct gene analysis. Huntington disease, cystic fibrosis, and myotonic dystrophy are familiar examples. However, gene tracking may still have a role even when a gene has been cloned. In the setting of a diagnostic laboratory, it is not always cost effective to search all through a large multi-exon gene to find every mutation. Moreover, there are always cases in which the mutation cannot be found. In these circumstances, gene tracking using linked markers is the method of choice. The prerequisites for gene tracking are:

- The disease should be well mapped, with no uncertainty about the map location, so that markers can be used that are known to be tightly linked to the disease locus.
- The pedigree structure and sample availability must allow the determination of phase (see below).
- There must be unequivocally confirmed clinical diagnoses and no uncertainty as to which locus is involved in cases in which there is locus heterogeneity.

Gene tracking involves three logical steps

Box 18.1 illustrates the essential logic of gene tracking. This logic can be applied to diseases with any mode of inheritance. There has to be at least one parent who

BOX 18.1 THE LOGIC OF GENE TRACKING

Shown here are three stages in the investigation of a late-onset autosomal dominant disease where, for one reason or another, direct testing for the mutation is not possible.

- Individual III_2 (arrow), who is pregnant, wishes to have a presymptomatic test to show whether she has inherited the disease allele. The first step is to tell her mother's two chromosomes apart. A marker, closely linked to the disease locus, is found for which II_3 is heterozygous (2–1).
- Next we must establish phase—that is, work out which marker allele in II_3 is segregating with the disease allele. The maternal grandmother, I_2, is typed for the marker (2–4). Thus, II_3 must have inherited marker allele 2 from her mother, which therefore marks her unaffected chromosome. Her affected chromosome, inherited from her dead father, must be the one that carries marker allele 1.
- By typing III_2 and her father, we can work out which marker allele she received from her mother. If she is 2–1 or 2–3, it is good news: she inherited marker allele 2 from her mother, which is the grandmaternal allele. If she types as 1–1 or 1–3 it is bad news: she inherited the grandpaternal chromosome, which carries the disease allele.

Note that it is the segregation pattern in the family, and not the actual marker genotype, that is important: if III_2 has the same marker genotype, 2–1, as her affected mother, this is good news, not bad news, for her.

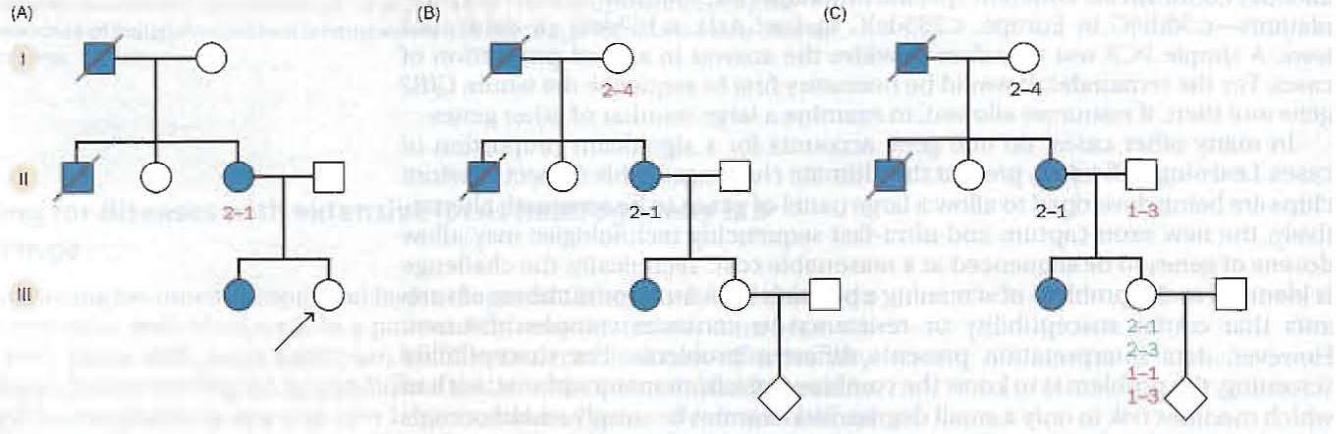


Figure 1 Using gene tracking to predict the risk of inheriting an autosomal dominant disease.

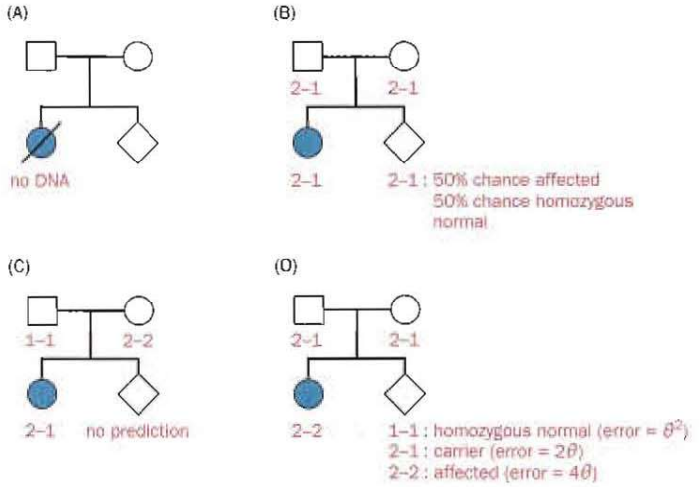


Figure 18.20 Gene tracking for prenatal diagnosis of an autosomal recessive disease. Four families each have a child affected with a recessive disease. Direct mutation testing is not possible (either because the gene has not been cloned or because the mutations could not be found). (A) No diagnosis is possible if there is no sample from the affected child. (B) If everybody has the same heterozygous genotype for the marker, the result is not clinically useful. (C) If the parents are homozygous for the marker, no prediction is possible with this marker. (D) Both parents are heterozygous carriers, and the genotype of the affected child shows that in each parent the pathogenic mutation is on the chromosome that carries allele 2 of the marker, allowing a successful prediction to be made. The error rates shown are the risk of predicting an unaffected pregnancy when the fetus is affected, or vice versa, if the marker used shows a recombination fraction θ with the disease locus. These examples emphasize the need for both an appropriate pedigree structure (DNA must be available from the affected child) and informative marker types.

could have passed on the disease allele to the proband, and who may or may not have actually done so. The process always follows the same three steps:

1. Distinguish the two chromosomes in the relevant parent(s)—that is, find a closely linked marker for which they are heterozygous.
2. Determine phase—that is, work out which chromosome carries the disease allele.
3. Work out which chromosome the proband received.

Figure 18.20 shows gene tracking for an autosomal recessive disease. The pedigrees emphasize the need for both an appropriate pedigree structure (DNA must be available from the affected child) and informative marker types. Even if the affected child is dead, if the Guthrie card used for neonatal screening can be retrieved, sufficient DNA for PCR typing can usually be extracted from the remnants of the dried blood spot. Informativeness of the marker should not be a big problem. With more than 20,000 highly polymorphic microsatellites mapped across the human genome, it should always be possible to find informative markers that map close to the disease locus.

Recombination sets a fundamental limit on the accuracy of gene tracking

Because the DNA marker used for gene tracking is not the sequence that causes the disease, there is always the possibility that recombination may separate the disease allele and the marker. This would lead to an erroneous prediction. The recombination fraction, and hence the error rate, can be estimated from family studies by standard linkage analysis (see Chapter 14). With almost any disease there should be a good choice of markers showing less than 1% recombination with the disease locus. This follows from the observations that 1 nucleotide in 300 is polymorphic, and that loci 1 Mb apart show roughly 1% recombination (see Chapter 14, p. 446). Ideally, one uses an intragenic marker, such as a microsatellite within an intron.

Recombination between marker and disease can never be completely ruled out, even for very tightly linked markers, but the error rate can be greatly reduced by using two marker loci situated on opposite sides of the disease locus. With such *flanking* or *bridging markers*, a recombination between either of the markers and the disease locus will also produce a marker–marker recombinant, which can be detected (e.g. III_1 in Figure 18.21). If a marker–marker recombinant is seen in the consultand, then no prediction can be made about inheritance of the disease, but at least a false prediction has been avoided. Provided that no marker–marker recombinant is seen, the only residual risk is that of double recombinants. As we saw in Chapter 14, the true probability of a double recombinant is very low because of interference (see p. 445). Thus, the risk of an error due to unnoticed recombination is much smaller than the risk of a wrong prediction due to human error in obtaining and processing the DNA samples. Perhaps

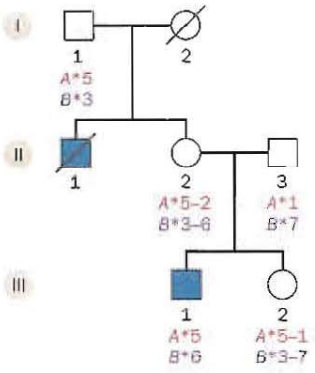


Figure 18.21 Gene tracking in Duchenne muscular dystrophy with flanking markers. The family has been typed for two polymorphisms, *A* and *B*, that flank the dystrophin locus. Individual III_1 has a recombination between marker locus *A* and DMD, but this does not confuse the prediction for III_2 because we know unambiguously that the disease allele in her mother is carried on a chromosome bearing marker alleles *A**2 and *B**6. III_2 can have inherited DMD only if she has a double recombination, one between marker *A* and DMD and another between DMD and marker *B*. If the recombination fractions are θ_A and θ_B, respectively, then the probability of a double recombinant is of the order $\theta_A\theta_B$, which typically will be well under 1%.

a greater risk is unexpected locus heterogeneity, so that the true disease locus in the family is actually different from the locus being tracked.

Calculating risks in gene tracking

Gene tracking can be used for Mendelian diseases with any mode of inheritance; however, unlike direct mutation testing, gene tracking always involves a calculation. Factors to be taken into account in assessing the final risk include:

- The probability of disease–marker and marker–marker recombination.
- Uncertainty due to imperfect pedigree structure or limited informativeness of the markers, about who transmitted which marker allele to whom (see Figure 14.9C for an example).
- Uncertainty as to whether somebody in the pedigree carries a newly mutant disease allele (see Figure 3.21B for an example of this problem in DMD).

Two alternative methods are available for performing the calculation: Bayesian calculations and linkage analysis.

Bayesian calculations

Bayes's theorem provides a general method for combining independent probabilities into a final overall probability. The theory, procedure, and a sample calculation are shown in **Box 18.2**.

BOX 18.2 USE OF BAYES'S THEOREM FOR COMBINING PROBABILITIES

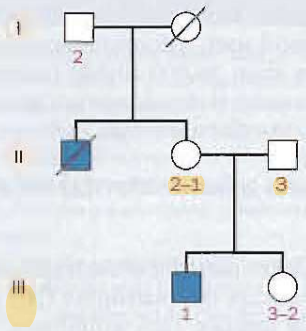
A formal statement of Bayes's theorem is

$$P(H_i|E) = P(H_i) \times P(E|H_i) / \Sigma[P(H_i) \times P(E|H_i)]$$

$P(H_i)$ means the probability of the ith hypothesis, and the vertical line means *given*, so that $P(E|H_i)$ means the probability of the evidence (E), given hypothesis H_i. An example will probably make this clearer (Figure 1).

The steps in performing a Bayesian calculation are:

- Set up a table with one column for each of the alternative hypotheses. Cover all the alternatives.
- Assign a *prior probability* to each alternative. The prior probabilities of all the hypotheses must sum to 1. It is not important at this stage to worry about exactly what information you should use to decide the prior probability, as long as it is consistent across the columns. You will not be using all the information (otherwise there would be no point in doing the calculation because you would already have the answer), and any information not used in the prior probability can be used later.
- Using one item of information not included in the prior probabilities, calculate a *conditional probability* for each hypothesis. The conditional probability is the probability of the information, given the hypothesis, namely $P(E|H_i)$ [*not* the probability of the hypothesis given the information, $P(H_i|E)$]. The conditional probabilities for the different hypotheses do not necessarily sum to 1.
- The previous step can be repeated as many times as necessary until all information has been used once and once only. The end result is a number of lines of conditional probabilities in each column.
- Within each column, multiply together the prior and all the conditional probabilities. This gives a *joint probability*, $P(H_i) \times P(E|H_i)$. The joint probabilities do not necessarily sum to 1 across the columns.
- If there are just two columns, the joint probabilities can be used directly as odds. Alternatively, the joint probabilities can be scaled to give final probabilities, which do sum to 1. This is done by dividing each joint probability by the sum of all the joint probabilities, $\Sigma[P(H_i) \times P(E|H_i)]$.



hypothesis: III_2 is	a carrier	not a carrier
prior probability	1/2	1/2
conditional (1): DNA result	0.05	0.95
conditional (2): CK data	0.7	1
joint probability	0.0175	0.475
final probability	0.0175/0.4925	0.475/0.4925
	= 0.036	= 0.964

Figure 1 Calculating the risk that III_2 is a carrier of DMD. Individual III_2 wishes to know her risk of being a carrier of DMD, which affected her brother III_1 and uncle II_1. Serum creatine kinase (CK) testing (an indicator of subclinical muscle damage common in DMD carriers) gave carrier–non-carrier odds of 0.7:1. A DNA marker that shows on average 5% recombination with DMD gave the types shown in red. These form two conditional probabilities and allow the risk to be calculated, following the guidelines in this Box, to give her overall risk of being a carrier as 3.6%.

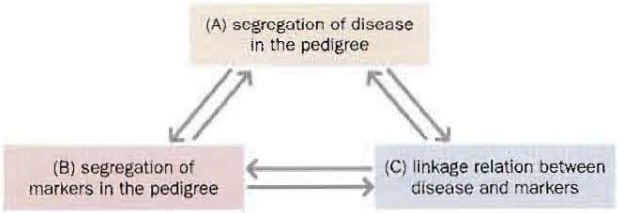


Figure 18.22 Use of linkage analysis programs for calculating genetic risks. Given information on any two of these subjects, the program can calculate the third. For linkage analysis, the program is given (A) and (B), and calculates (C). For calculating genetic risks, the program is given (B) and (C), and calculates (A).

Bayesian calculations give a quick answer for simple pedigrees, but the calculations can get very elaborate for more complex ones. Few people feel fully confident of their ability to work through a complex pedigree correctly, although the attempt is a valuable mental exercise for teasing out the factors contributing to the final risk. An alternative is to use a linkage analysis program.

Using linkage programs for calculating genetic risks

At first sight it may seem surprising that a program designed to calculate lod scores can also calcnlate genetic risks, but in fact the two are closely related (Figure 18.22). As described in Chapter 14, linkage analysis programs are general-purpose engines for calculating the likelihood of a pedigree, given certain data and assumptions. For calculating the likelihood of linkage we calculate the ratio

$$\frac{\text{likelihood of data | linkage, recombination fraction } \theta}{\text{likelihood of data | no linkage } (\theta = 0.5)}$$

For estimating the risk that a proband carries a disease gene, we calculate the ratio

$$\frac{\text{likelihood of data | proband is a carrier, recombination fraction } \theta}{\text{likelihood of data | proband is not a carrier, recombination fraction } \theta}$$

As in Box 18.2, the vertical line | means *given that*.

The special problems of Duchenne muscular dystrophy

Duchenne muscular dystrophy (DMD) poses a remarkably wide range of problems for the diagnostic laboratory. Fortunately, two-thirds of mutations in this X-linked disease are deletions, which are easily identified in males (see Figure 18.16), although more difficult in females. Duplications are hard to spot in either sex without using MLPA or some similar technique, and are undoubtedly underdiagnosed. The 30–35% of point mutations pose major problems. Scanning the DNA for point mutations requires all 79 exons to be individually amplified and sequenced, and therefore gene tracking is often used. However, DMD presents special problems for gene tracking because there is an extremely high recombination frequency across the gene. Even intragenic markers show an average of 5% recombination with the disease. It is therefore prudent to use flanking markers, as in Figure 18.21.

The problems do not end here. There is a high frequency of new mutations. The mutation-selection equilibrium calculations in Chapter 3 (p. 88) show that for any lethal X-linked recessive condition (fitness = 0), one-third of cases are fresh mutations. Therefore, the mother of an isolated DMD boy has only a two-thirds chance of being a carrier. This has two unfortunate consequences:

- It greatly complicates the risk calculations that are necessary for interpreting gene tracking results.
- As shown in Figure 3.21, the first mutation carrier in a DMD pedigree is very often a mosaic (male or female). This raises yet more problems, both for risk estimation and for interpretation of the results of direct testing.

These factors, together with the particularly distressing clinical course of the disease, the high recurrence risk within families, and the high frequency of DMD in the population, mean that DMD remains perhaps the most challenging of all diseases for genetic service providers.