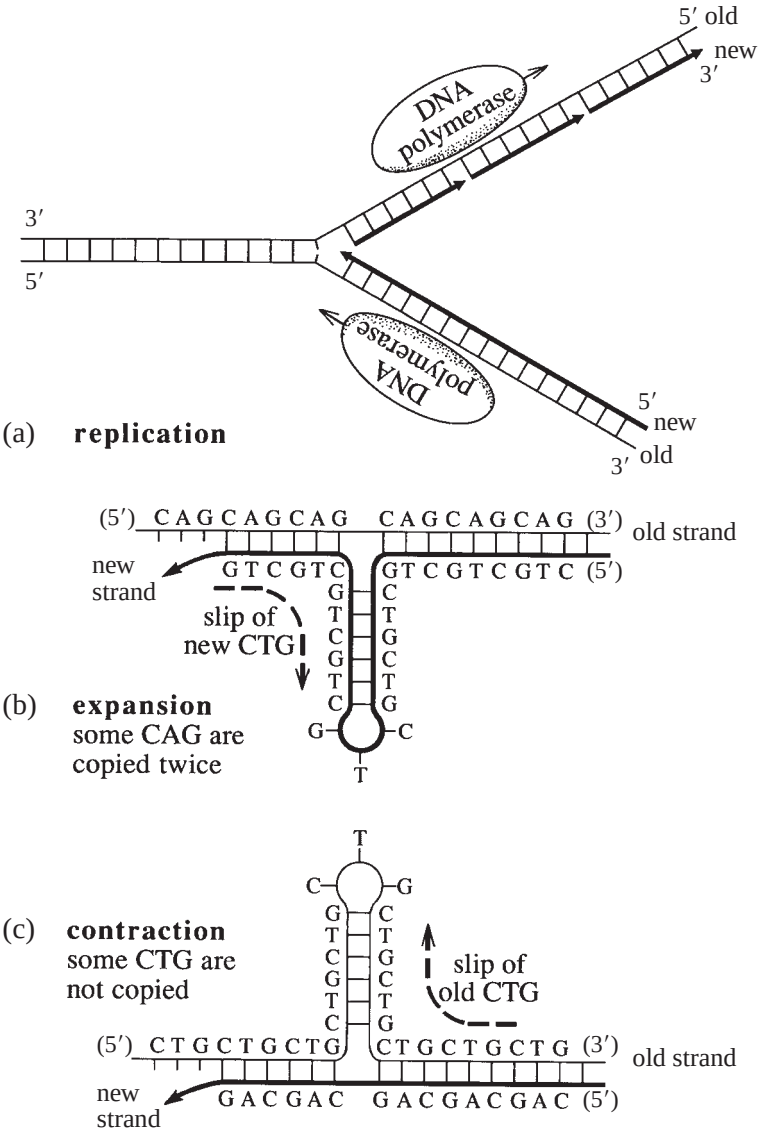

Figure 10.9 (a) During normal replication, a DNA polymerase enzyme can make new strands in a 5′ to 3′direction; hence it can copy one strand continuously, but has to copy the other strand in small pieces of size around 300 base-pairs, which will be joined together subsequently by a ligase enzyme. (b) DNA of repeated sequence $(CAG)_n$ may expand in size during replication, if the new CTG strand slips backward, so that some CAG are copied twice. (c) DNA of repeated sequence $(CTG)_n$ may contract in size during replication, if some CTG on the old strand slip out into a hairpin, so that not all CTG are copied.

Fig. 10.9(b); and the DNA will get longer. But suppose a CTG hairpin forms within the old strand: then some CTG triplets may not be copied if the polymerase skips over them as shown in Fig. 10.9(c); and the DNA will get shorter. Such potential expansion-contraction

events seem especially favored for a cellular DNA polymerase of the kind 'beta' (as opposed to 'alpha'), which 'stalls' at stable hairpins that are long enough for slippage to occur.

This general mechanism for DNA expansion or contraction, which invokes slippage of a double helix to form hairpins during replication, has been supported by many kinds of evidence. For example, sequences of the kind $(CTG)_n$ expand when they replicate in bacteria; whereas other DNA sequences which do not show expansion in humans, do not show expansion in bacteria either. Hence, DNA hairpins as formed from the six triplets which cause human disease – CGG, CCG, CTG, CAG, GAA and TTC – could perhaps be more stable than hairpins made from the other 58 triplet DNA sequences. All of those favored-hairpins would rely on non-Watson–Crick base-pairs for their stabilities. For example, the $(CTG)_5$ hairpins shown in Fig. 10.9(b) and (c) might contain T–T base-pairs of a 'wobble' variety (as for a G–T pair), plus C–G pairs of a Hoogsteen type: see Figs 2.12 and 2.13. A stable hairpin made from $(TTC)_3$ would contain T–T, C–T or C–C pairs; and similarly for $(GAA)_3$ with G–G, G–A or A–A pairs. These would appear to be 'crazy' forms of DNA; yet the available evidence suggests that they exist, at least transiently during replication!

In summary, much has been learned from point mutations and triplet expansions, about the structure and dynamics of DNA in chromosomes. Now let us proceed to the third general part (c) of our survey, which concerns the inheritance of genes in a non-Mendelian fashion, by a poorly-understood mechanism known as 'imprinting'. Gregor Mendel found from his long-term study of plants in 1860, that his specimens could inherit two slightly different genes of each particular kind from its two parents, in an apparently random fashion. We know now that this is because we inherit two slightly different chromosomes for any homologous pair, with slight variations of DNA sequence between them. Some genes when inherited in that way appear to 'dominate' the others, while other genes remain passive or 'recessive'; and only show their corresponding physical attributes ('phenotype') when no dominant gene is present.

Still, it should make no difference by Mendel's scheme whether you inherit a gene from your mother or from your father. A dominant gene from your mother will act in the same way as a dominant gene from your father, according to early 20th-century rules of genetics.

Yet it has been known for a long time that those 'rules' can be at most partly true. For some genes it matters greatly whether you inherit them from your mother or from your father. For example, when you mate a horse with a donkey, the kind of hybrid mule that you get depends greatly on whether it has a horse father and a

donkey mother, or vice-versa. And in humans, the CGG repeat of FMR1 only expands to a large size, causing mental retardation, when it is inherited from your mother in an egg, and not from your father in a sperm. Such non-Mendelian inheritance is called 'imprinting', because one parent or the other seems to place an imaginary mark or imprint on the activities of certain genes, when they pass those genes onto their children in sex cells. For example, an imprinted gene may always be active if received in the sperm from your father, and inactive if received in the egg from your mother, or vice-versa.

About 60 imprinted genes have been found so far in mouse and man, while 200 more are expected to be found in the future. Some of these imprinted genes cause genetic disease – often a dysfunction of a long-distance signalling process (known as 'endocrine') – if the natural imprinting mechanism fails: for example Prader-Willi/Angelman syndromes, Silver-Russell syndrome or Beckwith-Wiedeman syndrome. When studied closely by genetic or biochemical means, all of those diseases seem to be inherited from only one parent, either the mother or the father, but not equally from both.

The molecular mechanism by which genes may be imprinted remains poorly understood. It seems to be related to all of: (a) differences in cytosine methylation between active and inactive genes (see Chapter 11); (b) differences in histone protein methylation or acetylation between active and inactive genes (see Chapter 7); and (c) the synthesis of certain non-coding RNA molecules which adhere to specific chromosomal regions within the nucleus. These long, non-coding transcripts of RNA then repress or 'shut down' genes wherever they adhere, and also for some distance on either side. In other words, imprinted genes are usually found in large clusters along any chromosome: if a single gene becomes imprinted, then its neighbors are likely to become imprinted as well.

Figure 10.10 shows a general 'working hypothesis' that may perhaps be useful, when contemplating the data from imprinted gene systems. In part (a), we see that certain regions of human DNA may contain base sequences which form 'direct repeats'. In other words, those regions show approximately the same sequence of 50 to 200 base-pairs over and over again, repeated for 5 to 20 copies in a tandem array. In part (b), we see that directly-repetitive regions of DNA may 'slip out' to form large staggered loops during DNA replication, just as certain triplet-repeat sequences may slip in steps of three, as shown in Fig. 10.9. Then in part (c), we see that if RNA happens to be transcribed from those directly-repetitive regions of DNA, such RNA molecules can potentially bind to the large staggered loops using Watson–Crick base-pairs.

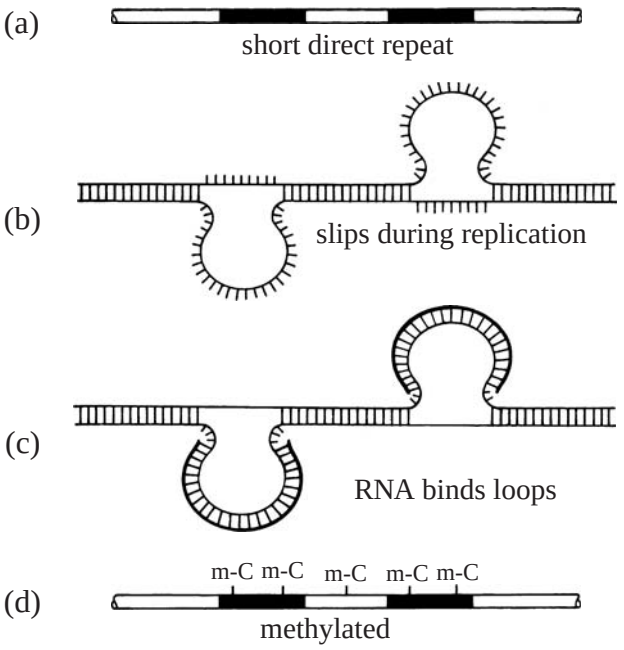

Figure 10.10 A 'working hypothesis' for the combined role of DNA direct repeats, non-coding RNA and cytosine methylation in gene imprinting. As shown in (a), certain regions of chromosomal DNA may be highly repetitive with respect to base sequence in a direct sense. Next (b) some of those regions may potentially 'slip out' to form large staggered loops during DNA replication (cf. the triplets in Fig. 10.9). Then (c) if such directly-repetitive DNA is transcribed into RNA, the RNA molecules so formed may potentially bind to those large staggered loops of DNA using Watson–Crick base-pairs. Finally, as shown in (d), the binding of RNA to DNA direct-repeats seems somehow to trigger cytosine methylation, which is known to silence genes during imprinting.

Finally, in Fig. 10.10(d), we see that the binding of RNA to DNA within such direct repeats may potentially induce a DNA-modifying activity of cellular methylase enzymes, which can convert cytosine 'C' bases to 5-methyl-cytosine or 'm-C'. It is still not known how step (d) is accomplished, but such are the data as typically observed. Perhaps a methyl-dependent, gene-repression activity evolved long ago as a defense against RNA viruses, some of which try to infect mammalian or plant cells by copying their genomes into a host chromosome? If any such insertion were successful, would the host cell want those foreign genes to remain active? It seems worthwhile to note here, that a surprisingly large 45% of human DNA consists of ancient (but now inactive) infectious elements, whether based on RNA or DNA. So we really do need a defense system of that sort!

Many different RNA molecules of the kind shown in Fig. 10.10 (c) are transcribed from DNA inside any cell nucleus, but never leave the

nucleus to make protein – in contrast to the ribosomal and transfer-RNA, which are exported from the nucleus. Instead, they travel within the nucleus to other locations, where they bind to DNA or RNA molecules that they find there, and so effect regulatory functions such as imprinting. We see then that the phenomenon of imprinting could depend critically on which kinds of RNA a cell includes inside itself. Studies to date have only begun to reveal the enormous complexity of genetic inheritance by this novel mechanism, which Mendel could never have foreseen.

Not much is known about RNA-based imprinting, yet in one case a mechanism seems clear. The best-characterized example of an RNA molecule which causes the shutdown of nearby genes is called *Xist* for 'X inactivation specific transcript'. That long RNA molecule is made from the female X chromosome in humans and other mammals as a single strand of 18 000 bases, but it does not code for protein. Instead, *Xist* RNA binds within the general vicinity of its own DNA template; and somehow it causes many genes on the X chromosome to become inactive due to induced cytosine methylation. As its repressive effect spreads widely, through means as yet unknown, eventually the entire X chromosome loses gene activity.

Within *Xist*, the DNA or RNA sequence contains many directly-repeated regions of length 50 to 200 base-pairs. Those regions are called 'A, B, C and D' respectively going in a 5' to 3' direction: the 'C' region in particular contains 14 direct repeats of 110 to 120 bases. Some experiments suggest that the 'C' region of *Xist* mediates its binding to DNA on the X-chromosome. Might the directly-repetitive DNA slip out there into staggered loops as shown in Fig. 10(b) and (c), and thereby facilitate binding of *Xist* RNA? Unfortunately, no one has yet determined a structural mechanism for the specific binding of *Xist* RNA to DNA within living cells. Region 'A' near the 5'-end of *Xist* seems important to trigger gene silencing and perhaps cytosine methylation, once *Xist* binds using 'C' or other sequences.

Now for some simple genetics: it hardly matters whether females lose all gene activity from one of their two X chromosomes, since the other X will continue to function normally. Fortunately the single X chromosome in males remains completely active; and since the single active X in females produces proteins in the same relative amount as for males, the biochemistry of cells may remain the same for both males and females. This levelling-effect on protein synthesis is called 'dosage compensation', and it explains why 'trisomies' – i.e. getting three copies of any non-sex chromosome rather than two – are often harmful.

By a strange twist of Nature, it seems that each individual cell chooses by chance which of its two X chromosomes will be shut

down by *Xist* RNA. Hence, any mature female may contain a 'mosaic' of two kinds of cell, depending on which X chromosome remains active. For example, the variegated colors seen in the coat of a tortoiseshell cat are due to that effect, where hairs of the cat may have several different colors, depending on their choice of X chromosome locally.

In another example of RNA-induced imprinting, a long repetitive RNA molecule known as *'Air'* has been shown to control silencing of genes on a non-sex chromosome or 'autosome'. Also, a protein called *'Eed'* – for 'embryonic ectoderm development' – has been implicated as helping to maintain the repressed state. Changes of DNA methylation are thought to proceed by the joint action of two methylases called 'DNMT3L' and 'Dnmt3a' (see Chapter 11). Finally, the addition of a transcriptional activator called "VP16" to a repressed gene can reverse the effects of imprinting, and cause that gene to work normally again. What a mess! How could any field be more descriptive than the study of gene imprinting at present?

Finally let us proceed to the fourth general part (d) of our survey, which concerns attempts by scientists to correct errors in human DNA – either by adding extra functional DNA into human cells, or else by repairing an undesired mutation within pre-existing DNA. These efforts are known in general as 'gene therapy' or 'gene correction' respectively.

As shown in Fig. 10.11, one may add extra double-stranded DNA to human cells by wrapping it in either: (i) special viruses or else (ii) fat-DNA complexes known as 'liposomes'. One can also repair an undesired mutation in pre-existing DNA, by adding a short single-strand (typically 30 to 500 bases) of the correct DNA sequence, that may be copied into a defective chromosomal region by enzymes such as 'recA' or 'RAD51', which mix or 'recombine' one DNA strand with another. Those recombination enzymes are used normally by the cell to repair defects on one chromosome during cell division, using sequence information found on the other homologous chromosome; but scientists can 'trick them' into correcting general mutations as well.

The double-stranded DNA as added by viruses or liposomes will often carry a functional gene, to supply a necessary protein which the person does not have on account of genetic disease. In other cases, that extra DNA may carry the information for making 'anti-sense' or 'silencing' RNA, which can bind to the 'sense' or messenger-RNA that codes for some defective protein, so as to block its template action at ribosomes. Alternatively, short segments of single-stranded DNA may fix a point or frameshift mutation, which will simply return a person's DNA to the human consensus.

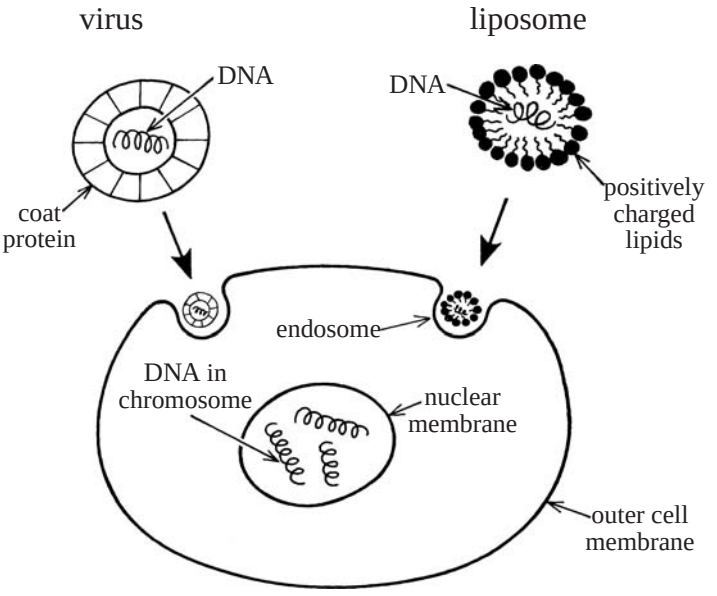

Figure 10.11 Extra DNA may be added to cells for 'gene therapy', after packaging into either viruses or liposomes.

Unfortunately, none of these promising methods works well at present, for several reasons. First, both gene addition and gene correction act much more efficiently on DNA within cultured cells, than on DNA within the whole body of a living animal. Hence one may have to extract bone-marrow or embryonic stem-cells from a donor, in order to carry out such therapies efficiently. Second, the extra double-stranded DNA may cause cancer or have other undesirable effects, if it inserts at many random places within a chromosome. Thirdly, short segments of single-stranded DNA which repair an existing defect, would seem to be much safer than long pieces of double-stranded DNA which add an entirely new gene; however, the efficiency of fixing a point or frameshift mutation by short single-stranded DNA is currently so low, that little therapeutic effect is obtained.

Let us summarize nevertheless some of the standard approaches for DNA delivery to cells or patients, which have been used so far. The liposome method is the most common: there one adds to any long piece of double-stranded DNA (or even short pieces of single-stranded DNA), certain chemicals – 'lipids' – which contain water-insoluble fat on one end, and positively-charged nitrogen N–H groups on the other. Such chemicals, of which there are many varieties, will rapidly enclose a therapeutic DNA molecule inside a protective, fatty coat: see Fig. 10.11.

Other methods for adding DNA within cells, or to humans, animals and plants include: injection into solid tissue; electroporation (i.e. the disruption of membranes by electric fields); the 'gene gun' (i.e. high-velocity particles which may penetrate tissues); ultrasound disruption; positively-charged polymers; and positively-charged peptides.

Those fat-DNA complexes or 'liposomes', once formed, can pass readily through the outer membrane of most cells. Thus our long DNA molecule, once wrapped into a liposome, will typically enter a cell by means of an 'endosome', where a small part of the outer membrane folds into a bubble, as shown in Fig. 10.11. But then trouble begins. Because a typical cell does not wish to take foreign DNA inside of itself, it has developed several defense mechanisms to degrade any DNA which does enter. Almost immediately, the cell begins to pump strong acid and various enzymes which digest protein and DNA into this endosome bubble, so as to depolymerize the molecules inside! The DNA as carried by our liposome has little chance to escape, and so most of it becomes degraded before it can enter the main part or 'cytoplasm' of a cell. By contrast, certain viruses have devised clever mechanisms for escaping from the endosome, for example by making a long 'protein tunnel' through which the DNA can pass.

Next, if some small amount of DNA does escape from the endosome, and proceeds into the general cell fluid or cytoplasm, that extra DNA still has to negotiate many small, tight passages which lead through the nuclear membrane into the cell nucleus, before it can begin to make the messenger-RNA which enables it to synthesize protein. Once again, certain viruses have evolved clever strategies to get their DNA into a cell nucleus: say by attaching DNA to specific viral proteins, which possess the specific recognition codes needed to go in and out. Currently however, only one out of every million DNA molecules that enter the cell by means of a liposome, eventually enters the cell nucleus and makes a desired protein.

Even if we succeed in getting our extra DNA into the cell nucleus by use of a liposome, or else a human virus such as adenovirus or adeno-associated virus, still we have no satisfactory way of adding such extra DNA permanently, and at specific locations, within the DNA of normal chromosomes. The best method so far uses the replication or *rep* protein of adeno-associated virus (AAV), to insert extra DNA by means of a particular base sequence to a specific region on human chromosome 19.

The general mechanism of insertion by AAV seems quite interesting, and is shown schematically in Fig. 10.12. There we can see that the *rep* protein of AAV makes a specific 'nick' within one

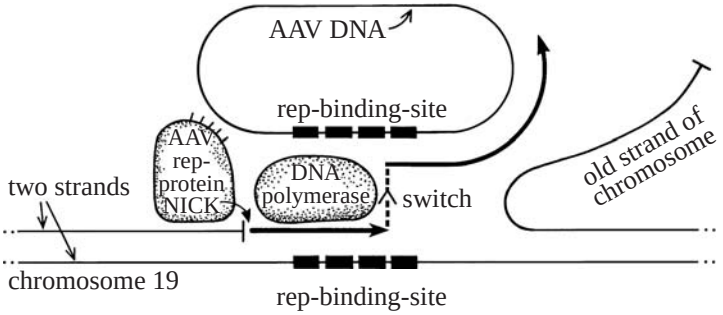

Figure 10.12 A general mechanism for insertion of new genes *via* adeno-associated virus (AAV), specifically at one region along human chromsome 19. First the DNA of chromosome 19 is nicked specifically by the AAV rep protein, which also pulls AAV and chromosome 19 together in space. Next, a cellular DNA polymerase copies from that nick in a rightward sense, until it reaches a repeated sequence known as the rep-binding-site (RBS), typically $(GCTC)_4$. Once stalled at that RBS, the polymerase can easily switch templates; so that it begins to copy from another RBS which lies on AAV nearby. Later, the polymerase returns to the human chromosome.

strand of the human DNA, just at one precise location on chromosome 19, as shown on the left. Next, starting from that nick, a new copy of the DNA is made by DNA polymerase (thick line with arrow), while the old strand is displaced, as indicated on the right in the figure. The DNA polymerase then slows down or stalls once it reaches a repeated sequence known as the *rep-binding-site* ('RBS'). The *rep* protein meanwhile, in addition to making the nick, also pulls the single-stranded viral DNA into close proximity of the chromosomal nick site. While stalled therefore, there is an increased chance that the polymerase will hop onto the nearby viral DNA, and continue to extend the new DNA chain (second thick line with arrow) onwards from the viral RBS, until it returns to the chromosomal DNA somewhat later.

Hence, strand-switching by DNA polymerase, as initiated by a specific *rep* nick, and enhanced by a repetitive sequence RBS, all combine to add new AAV genes specifically at one location in the entire human genome. Other DNA-integrating viruses (e.g. lentoviruses and retroviruses) have thus far performed less well for medical purposes. For example, they have caused cancer in stem-cell trials, by inserting their DNA at many non-specific places within a patient's chromosomes. Further work in this active field will eventually determine which DNA delivery system might be best.

Recall how scientists can add foreign DNA to mouse or fly eggs, by use of a fine needle through which the DNA is injected, directly into the cell nucleus without having to pass through any cell membrane.

That simple technique will not work for gene therapy, however, because the doctor has to add extra DNA to millions of different cells in culture or within a living patient, in order to make enough therapeutic protein to be useful. Then such extra DNA has to be maintained as an active gene which produces protein for many years, without being lost through mutation, or being repressed through imprinting or cytosine methylation.

So scientists today remain highly divided over whether gene addition or gene correction will really work. Still, that has not stopped them from trying! Over the next ten years, we may expect to see many new discoveries which should advance the field substantially.

Further Reading

Transgenic mice, animals and plants

Jimenez-Bermudez, S., Redondo-Nevado, J., Munoz-Blanco, J., Caballer, J.L. *et al.* (2002) Manipulation of strawberry fruit softening by antisense expression of a pectate lyase gene. *Plant Physiology* **128**, 751–9. The genetic design of strawberries which ripen more slowly than normal.

Marquet-Blouin, E., Bouche, F.B., Steinmetz, A., and Muller, C.P. (2003) Neutralizing immunogenicity of transgenic carrot-derived measles virus hemagglutinin. *Plant Molecular Biology* **51**, 459–69. Genetic modification of carrots to carry a measles protein for vaccine purposes.

Murray-Kolb, L.E., Takaiwa, F., Goto, F., Yoshihara, T. *et al.* (2002) Transgenic rice is a source of iron for iron-depleted rats. *Journal of Nutrition* **132**, 957–60. Genetic modification of rice to carry more iron for nutritional purposes.

Nicoll, J.A., Wilkinson, D., Holmes, C., Steart, P., Markham, H., and Weller, R.O. (2003) Neuropathology of human Alzheimer disease after immunization with amyloid-beta peptide. *Nature Medicine* **9**, 448–52. Testing of a therapy for Alzheimer disease in transgenic mice *versus* humans.

Qaim, M. and Zilberman, D. (2003) Yield effects of genetically modified crops in developing countries. *Science* **299**, 900–2. Cotton which has been genetically modified to be pest-resistant assists in its propagation.

Walker, S.C., Shin, T., Zaunbrecher, G.M., Romano, J.E. *et al.* (2002) *Cloning Stem Cells* **4**, 105–12. Technical procedures for reproductive cloning of adult pigs from mature-tissue cells.

DNA amplification technologies

Christian, A.T., Garcia, H.E., and Tucker, J.D. (1999) PCR *in situ* followed by microdissection allows whole chromosome painting probes to be made from single microdissected chromosomes. *Mammalian Genome* **10**, 628–31. The use of semi-random PCR to amplify trace DNA from single copies of a chromosome.