The invention relates to a method for the specific replacement of a copy of a gene present in the genome of a recipient eukaryotic organism, by the integration of a gene different from the inactivated gene. Preferably, the recipient gene will be present in at least 2 copies in the transfected host cell. The recipient gene is defined as the gene where the insertion of the different gene will take place.

More especially, the invention relates to the production of transgenic animals in which the foreign gene has been introduced in a targeted manner in order to permit both maintenance of the animal's normal genetic functions and expression of the foreign gene under the control of endogenous promoters.

"Different or foreign gene" is understood to mean any nucleotide sequence corresponding to the whole or to a portion of a gene which is "foreign to or different from" the recipient gene as this sequence is normally found in the genome (RNA or DNA), or the sequence also corresponds to an artificially modified sequence of the normal gene or alternatively to a fragment of this sequence.

The invention also relates to the method for the production of these transgenic animals.

In the production of transgenic animals, the conventional methods used for the introduction of heterologous DNA sequences into the germ cell line do not enable the integration site of the foreign gene in the genome, or the number of copies thereby introduced, to be controlled. Integration of the foreign gene takes place at random and, in general, several copies of the gene integrate at the same time, sometimes in the form of a head-to-tail tandem, the integration site and the number of copies integrated varying from one transgenic animal to another.

The situation may hence arise where endogenous cellular genes located at the insertion point are thereby inactivated, without this being readily detectable on account of the large number of random insertions. If the

product of these genes is important for the animal's development, the latter will be seriously disturbed. Moreover, the random insertion of the foreign gene may take place at a site which is unsuitable for the expression of the gene. Furthermore, the fact of there being variation in the site and number of insertions from animal to animal makes the interpretation of expression studies extremely difficult.

A major problem encountered in the production of transgenic animals is the obtaining of the expression of the foreign gene. Generally speaking, two types of experiment have been carried out in mice.

The genes introduced into the germ line are:
- either "complete" genes, comprising coding sequences flanked by their own regulatory sequences;
- or composite genes, formed from the coding sequence of a gene fused to the promoter sequence of another gene, the two fragments sometimes even belonging to two different animal species.

It has been possible to confirm in this way that the specificity of the expression of the genes in some particular tissue is determined by their regulatory sequence(s).

The choice of suitable promoter for the expression of the foreign gene in the transgenic animal is hence of the utmost importance.

Furthermore, the site-directed mutagenesis of murine genes in embryonic stem cells has recently been carried out by making use of a "genetic targeting" (gene targeting) technique (Thomas et al., 1987; Thompson et al., 1989).

In the first case the murine HPRT gene was mutated by insertion and replacement, and in the second case a mutated HPRT gene was corrected. Thompson et al. extended their experiments until they obtained chimeric mice, and established that the genetic modification passed into the germ cell line.

In each of the documents cited, the exact site of the integration was targeted by homologous recombination

between, on the one hand exogenous sequences containing the mutation or correction included in a vector, under the control of an exogenous promoter, and on the other hand their genomic homologue. This being the case, it should be pointed out that the above authors carried out their experiments on a specific gene (HPRT) whose activation by mutation was accompanied by a detectable phenotype. The targeted mutation described by Thomas et al. had the effect of inactivating the HPRT gene, and consequently of causing the detectable phenotype normally associated with HPRT to disappear. The Neo^R selectable gene, under the control of a TK promoter, was hence incorporated in the insertion DNA in order to permit selection of the transformants. It should be noted that the experiments described in the prior art involved a selection either by the recipient gene (e.g. HPRT) or by the insertion gene (e.g. Neo^R). The insertion site and/or the type of gene inserted is hence limited to genes conferring a selectable character.

In addition, in the prior art, the exogenous sequences on the vector hence serve both to target the integration site and to introduce the modification. Following homologous recombination, the modified gene is still in its normal genetic environment.

We may recall that a problem which arises during the production of transgenic animals is the danger of inactivating an endogenous cellular gene which lies at the insertion point of the foreign gene.

Depending on the function of the product of the inactivated gene, such an inactivation may lead to substantial physiological or morphological disorders in the transgenic animal, or could even prevent its survival.

In contrast, the inactivation of a gene could be considered advantageous if the gene in question coded for a receptor for a virus or other infectious agent.

The inventors studied the possibility of avoiding the drawbacks described above and which are associated, in some cases, with the possible inactivation of one or

more endogenous cellular genes having an important function during the production of transgenic animals.

The subject of the invention is a method for specific replacement, in particular by targeting a DNA, termed insertion DNA, consisting of a portion of a gene capable of being rendered functional, or whose functioning may be rendered more effective, when it is recombined with a complementary DNA to yield then a complete recombinant gene in the genome of a eukaryotic cell, characterized in that

- the insertion site is in a chosen gene, termed the recipient gene, containing the complementary DNA, and in that
- eukaryotic cells are transfected with a vector containing an insert itself comprising the insertion DNA and two so-called "flanking" sequences on each side of the insertion DNA, which are homologous, respectively, to two genomic sequences which adjoin the desired insertion site in the recipient gene,
- the insertion DNA being heterologous with respect to the recipient gene, and
- the flanking sequences being chosen from those which constitute the abovementioned complementary DNA and which permit, by homologous recombination with corresponding sequences of the recipient gene, the reconstitution of a complete recombinant gene in the genome of the eukaryotic cell, on condition that, when the insertion DNA consists of a coding sequence, it is not the coding sequence of a gene coding for a selectable agent.

The invention also relates to a method for the production of transgenic animals, characterized in that ES cells are transfected under the conditions described above and selected for the homologous recombination event, namely the correct integration of the foreign gene, the transfected cells are injected into embryos at a stage at which they are capable of integrating the transfected cells (for example at the blastocyst stage), the embryos are then reimplanted into a surrogate mother and the chimeric individuals obtained at the end of

gestation are mated. If the ES cells have colonized the chimeric animal's germ line, transgenic animals which are heterozygous for the replaced gene will be obtained by mating (F1) in the progeny.

It is also possible to insert the gene carried by the vector of the invention into the egg shortly after fertilization (that is to say less than 24 hours). In this way, insertion is performed while the egg is in the unicellular state.

The invention also relates to a plasmid capable of performing the targeted insertion of a recombinant gene, termed insertion gene, into the genome of a eukaryotic cell, characterized in that it contains an insert itself comprising the insertion gene and two so-called "flanking" sequences on each side of the insertion gene, which are homologous, respectively, to two genomic sequences which adjoin the desired insertion site in the recipient gene.

Hence the method of the invention makes it possible, as a result of the phenomenon of homologous recombination, to insert foreign genes, especially coding sequences lacking the promoter which is normally associated with them, in a targeted manner into the genome of a eukaryotic organism at a site which permits their expression under the control of the endogenous promoter of the gene where the insertion takes place and consequently to inactivate the targeted endogenous gene.

According to a preferred embodiment of the invention, the targeted recipient gene is a gene which is present in the genome in at least two copies. The use of the electroporation technique (Ref. 11) effects the introduction of only one copy of the foreign gene.

According to this variant of the invention, the targeted insertion of the gene of interest (that is to say the so-called insertion gene) has the effect of inactivating only the copy of the endogenous cellular gene where the insertion takes place, and leaves intact and functional the other copy or copies of this gene.

In this way, the genetic functioning of the

transgenic animal undergoes little or no disturbance through the introduction of the foreign gene, even if the insertion inactivates a single copy of a recipient essential gene for the animal's development. Hence either its development will not be affected by the insertion of the foreign gene, or the possible minor disturbances in the case of the inactivation of a critical gene would probably not be lethal for the animal. The effects of insertion of the foreign gene in the homozygous state could be of any kind, and would be observed in the 2nd generation (F2) after crosses of heterozygous (F1) individuals with one another.

If, on the other hand, the inactivation of all the copies of a gene is desired, for example in the case where the gene codes for a receptor for an infectious agent, multiple copies of the foreign gene are introduced. Control of the amount introduced may be effected by making use of known techniques.

Targeted insertion of the foreign gene hence enables it to be introduced into a site where its expression is under the control of the regulatory sequences of the endogenous gene where the insertion takes place.

The method of the invention thus enables the foreign gene to be inserted behind an endogenous promoter which has the desired functions (for example specificity of expression in some particular tissue), and does so, where appropriate, without inactivating the other copies of the recipient gene.

According to an especially preferred embodiment of the invention, the insertion DNA contains, between the flanking sequences, on the one hand a DNA sequence intended to be recombined with the complementary DNA in the recipient gene to yield a recombinant gene, and on the other hand a sequence coding for a selective agent permitting selection of the transformants and a promoter permitting expression of the selective agent, the recipient gene and the recombinant gene coding for expression products which do not confer a selectable

phenotype.

In this way, the selection of the transformants is completely independent of the nature of the recipient gene and of the inserted gene, contrary to the methods described hitherto, in which the inserted gene or the recipient gene had of necessity to code for an expression product permitting selection of the transformants. The system developed by the inventors permits total flexibility as regards the nature of the recipient gene and of the inserted gene, or of the gene formed by homologous recombination. The inventors found, surprisingly, that the insertion of large-sized sequences (for example of approximately 7.5 kb) does not affect the frequency of homologous recombination.

The effect which may result from the insertion of the DNA sequence according to this aspect of the invention includes, depending on the type of sequence inserted, for example replacement of a coding sequence, replacement of a regulatory sequence, inactivation or reactivation of a gene by mutation or improvement in the level of expression of a gene. It is possible, according to the invention, to replace a coding frame or a part of a coding frame by a heterologous sequence which begins at the initiation codon of the replaced gene in order that the expression of the inserted gene completely replaces the expression of the replaced gene. This avoids the formation of fusion proteins, which could be undesirable in a transgenic animal.

According to this embodiment of the invention, the insertion DNA can contain, between the flanking sequences, a heterologous coding sequences lacking a promoter, the coding sequence being other than a gene coding for a selectable agent. The insertion DNA can contain, in addition, downstream of the coding sequence and still between the flanking sequences, gene coding for a selectable agent, associated with a promoter permitting its expression in the target cell.

In this way, the heterologous coding sequence may be inserted behind an endogenous promoter which has the

desired properties, for example a certain specificity of expression or transcription pattern, and the like, the selectability of the transformed cells being completely independent of the expression of the heterologous coding sequence. This type of construction makes it possible, for example, to select the transformants even if the gene replaced by the heterologous coding sequence is not normally expressed in the target cells. This is especially important in the production of transgenic animals from ES cells (embryonic stem cells), since a large proportion of the genes remains inactive until a more advanced stage of the animal's development. The Hox-3.1 gene is an example of this type of gene. Moreover, if the coding sequence codes for a readily detectable protein, for example b-Gal [sic], the development of the transcription pattern of the replaced endogenous gene may be followed. The vector pGN is an example of this type of construction.

According to another embodiment of the invention, the insertion DNA can contain a foreign regulatory sequence. The insertion site, and consequently the flanking sequences, are chosen in accordance with the desired aim, namely either the insertion of the foreign regulatory sequence to give a "double promoter" effect with the endogenous regulatory sequence, or else the replacement of an endogenous promoter by the foreign promoter. The coding sequence which is under the control of the regulatory sequence may be endogenous.

Another possibility would be the targeted insertion of a foreign DNA which contains both a regulatory sequence and a coding sequence. It is possible for the regulatory sequence to be the one which is naturally associated with the coding sequence.

The method of the invention employs a vector containing two "flanking" sequences on each side of the foreign gene. These flanking sequences have at least 150 base pairs and are preferably smaller than the length of the recipient gene. It is essential for these two flanking sequences to be homologous to the two genomic

sequences which adjoin the desired insertion site. The flanking sequence of the vector which lies upstream of the foreign gene to be introduced is normally homologous to the genomic sequence which is located on the 5′ side of the insertion site. Likewise, the flanking sequence of the vector which lies downstream of the foreign gene is normally homologous to the genomic sequence which is located on the 3′ side of the insertion site.

It is possible to introduce "interposing" sequences between one or other of the flanking sequences and the foreign gene, for example sequences permitting selection of the transformants, markers, sequences permitting cloning of the vector, and the like.

The position of these interposing sequences with respect to the foreign gene must, however, be chosen so as not to prevent the expression of the foreign gene, especially of the foreign coding DNA sequence under the control of the endogenous promoter or, conversely, the endogenous DNA coding sequence under the control of foreign regulatory elements supplied by the insertion sequence.

In spite of the presence of the flanking sequences, which encourages a homologous recombination, it is possible for a number of integrations to take place at random. In order to verify that the targeted insertion has indeed taken place in the targeted site and not somewhere else, the "polymerase chain reaction" (PCR) technique (see Ref. 10) is used to amplify the DNA sequence of the locus where the insertion should have taken place. In this way, only the clones transformed following a homologous recombination are selected.

The flanking sequences of the vector are, quite obviously, chosen in accordance with the insertion site which is desired in order that homologous recombination may take place. Where appropriate, the flanking sequences can contain sequences which are replicas of the endogenous promoter and/or modifications to the sequences which precede the initiation codon in order to improve the level of translation (upstream sequences), and

sequences which are replicas of the termination sequences, in particular polyadenylation sites (downstream sequences).

The insertion gene can be any gene of interest. There may be mentioned, as non-limiting examples, the lacZ gene (as in the model described below), the genes coding for interleukin or interferon, genes for a receptor, for example for the retinoic acid or beta$_3$-adrenergic or HIV receptor, and genes known to be linked to certain disorders, for example myopathy, and the like.

According to a preferred variant of the invention, the eukaryotic cells are embryonic stem cells (see Ref. 14 and 15).

In effect, a mutated ES cell may be injected into an early embryo which, after reimplantation, will be able to be born in a chimeric form. If the germ line is colonized by the mutated cell, the chimeric animal will transmit the mutation to its progeny. Thereafter, it will be possible to observe the effects of this mutation, in the homozygous state in some individuals, on their development, behaviour, metabolism, pathology and the like.

Figure 1 shows the plasmid pGN,

Figures 2a and b show the molecules pGMA and pGMD constructed, respectively, from plasmid pGN in relation to the Hox-3.1 gene. These plasmids are mutagenesis plasmids. The two portions of the coding frame of the Hox-3.1 gene are shown on chromosome 15 with the "homeo" box in black. The corresponding sequences of Hox-3.1 were cloned into plasmid pGN.

(A : polyadenylation signal: Enh/Pro : enhancer-promoter).

07 and 08 note the two oligonucleotides used in the PCR.

Figures 3 to 6 show the plasmids used in the construction of pGN.

Figure 7 illustrates the detection of homologous recombination with the polymerase chain reaction (PCR) technique on transfected ES cells.

Figure 8 (a) and (b) shows Southern analyses of

positive individual clones (L5 and F2) and ES (CCE) cells.

The method of the invention is of very broad industrial application and can vary according to the nature of the foreign gene introduced.

The genetics of mammals is going to progress considerably as a result of the recent possibility of specifically mutagenizing any gene, thereby enabling its role to be better defined. By this technology, which involves homologous recombination and ES cells, valuable information will be provided about oncogenes, growth factors, transcription factors, and the like, genes which relate to very topical subjects in basic research or applied research. An important prospect for medical research is the possibility of reproducing a human disorder whose genetic determination is known (certain human disorders with a pathological basis, such as Duchenne's myopathy) in order to study its mechanisms better and to seek a therapy.

By applying the method of the invention, a gene known to be responsible for a certain disorder is inserted in a targeted manner into the genome of an ES cell. The transgenic animal which is produced as a result presents a useful model of this disorder.

If necessary, and as described above, the normal genetic functionings can be substantially maintained in spite of the insertion of the foreign gene.

Another application of the method of the invention consists in inserting an insertion gene which is readily detected, e.g. the LacZ gene, and which can hence play the part of a cell marker. In this way, filiation studies, e.g. in competition animals, are facilitated, and the breed can be monitored.

Insertion of the lacZ gene as insertion gene also makes promoter studies possible. As a result of the possibility of detecting β-galactosidase activity, the activity and specificity of different endogenous promoters may be studied by targeting different sites in the same type or different types of cells. The same

studies may be performed on a whole organism, during development or in the adult state, using chimeric or transgenic animal techniques.

The inventors found, surprisingly, that the frequency of homologous recombination is not affected by the insertion of large-sized fragments, for example Lac Z.

This observation suggested to the inventors that the homologous recombination technique would be well suited to the insertion of other heterologous genes which are of large size.

As a result of the possibility of modifying an animal's genome, the method of the invention may also be used as "gene therapy". The most obvious uses would consist in inactivating the genes for receptors for infectious agents (viruses or bacteria) or toxic agents. Should this mutagenesis prove lethal, it would be necessary to re-establish the lost function without re-establishing the sensitivity to harmful agents. A modified gene coding for such a receptor could be reintroduced into the mutated cell, unless the modification might be brought about by homologous recombination. This modification of the genotype would confer on the animal an immunity against the disorder in question.

This protocol may also occur in the context of autografting. Sick or healthy cells taken from a patient could be treated and immunized, then reimplanted into the same individual.

The technique of the invention also lends itself to studies of activity of pharmaceutical products presumed to have an activity with respect to the expression products of a pathological gene linked to a disorder. In this case, the insertion gene consists of the pathological gene, and the pharmaceutical product is administered to the transgenic animal with a view to evaluating its activity with respect to the disorder.

The invention will be illustrated by making reference to plasmid pGN and its use in the targeted insertion of a foreign gene (lacZ, coding for the

β-galactosidase enzyme of E. Coli) into the genome of a mouse ES cell. The lacZ gene was chosen on account of the fact that its expression may be readily detected, and is simply by way of illustration.

The coding frame of the β-galactosidase enzyme of E. Coli (lacZ; 1-3057), fused with a genomic sequence (7292-3) of the murine Hox.3-1 gene (Ref. 1), begins with the initiation codon of this gene. In effect, the sequence which precedes the initiation codon of Hox-3.1 is identical to the consensus sequence observed in vertebrates (Ref. 2), thereby permitting a better level of translation of β-galactosidase in vertebrate cells. The lac.Z gene is followed by a polyadenylation signal, for example of the SV 40 virus, like most eukaryotic genes, so as to stabilize the messenger RNAs.

The activity of E. Coli β-galactosidase, which is functional in eukaryotic cells, may be detected in different ways. Cells expressing the lacZ gene take on a blue coloration, after fixation, in the presence of X-Gal, which is a substrate for β-galactosidase (Ref. 3). A new substrate, FDG (fluoroscein di-β-galacto-pyranoside), enables the β-gal. activity to be detected and assayed while keeping the cells alive (Ref. 4). Cells expressing lacZ accumulate a fluorescent product and may be isolated using a cell sorter or FACS (fluorescence-activated cell sorter).

The transcription unit of the neomycin-resistance gene comes in large part from the plasmid pRSV neo (Ref. 5). The Rous sarcoma virus LTR (long terminal repeat) procures activator and promoter sequences which are very potent in many eukaryotic cells (Ref. 6). From the bacterial transposon Tn5 come a promoter which is active in E. Coli and the coding frame of the phosphotransferase enzyme (Ref. 7), which is followed by the SV40 virus polyadenylation signal. The same gene under the dual control of the RSV and Tn5 promoters can confer resistance to neomycin or kanamycin on bacteria and resistance to G418 on eukaryotic cells.

Through the effect of a single point mutation,

the B unit of the activator (enhancer) sequences of polyoma virus strain PyEC F9.1 became much more active in different cell types, and especially in embryonic carcinoma (EC) cells (Ref. 8). Two copies of this Py F9.1 enhancer were inserted in tandem into plasmid pGN, upstream of the RSV LTR, and in the "late promoter" orientation of the polyoma regulatory region.

In order to improve the level of translation of the phosphotransferase, the sequence preceding the initiation codon was modified in an oligonucleotide-induced mutagenesis. Thus, the sequence T T Ć G C A U G became G C A C C A U G, corresponding much better to the translation initiation consensus sequence in vertebrates (Ref. 2).

The improvements made to the transcription unit of the neomycin-resistance gene could be estimated by transfecting mouse embryonic stem (ES) cells. At equal plasmid molarity, a construction with the Py F9.1 enhancer produced 7.5x as many G418-resistant clones as pRSV neo, and 2 to 3x as many as the pMC1 Neo described by Capecchi et al. (ref. 13). Again, the number of clones was increased 60x, equivalent to 450x relative to pRSV neo, by modifying the translation initiation sequence. Homologous recombination can be a fairly rare event, depending on the experimental conditions applied (e.g. 1/1000 for HPRT, ref. 13). A vector displaying a high selection efficiency is hence very useful, all the more so for the fact that the electroporation conditions mainly give rise to the integration of only one copy.

Plasmid pGN contains, in addition, a bacterial origin of replication of the colE1, pBR322 type, which permits clonings and preparations in E. Coli.

Lastly, a multiple cloning site (MCS), synthesized in vitro, which contains only unique cleavage sites in pGN, was inserted upstream of lacZ in order to facilitate the uses of this plasmid.

The plasmid "flanking" sequences which cause homologous recombination are added to the ends of plasmid pGN after linearization of the plasmid upstream of lacZ,

through a site of the MCS (see Fig. 2). In this instance, the chosen flanking sequences are homologous to the chromosomal sequences originating from Hox-3.1 which are subsequently to participate in the homologous recombination.

Figure 2 places the molecule constructed from plasmid pGN in relation to the Hox-3.1 gene. In this case, a recombination between the plasmid sequences and chromosomal sequences of Hox-3.1 would result in an insertion at the beginning of the coding frame of this gene, and hence in its total inactivation.

Plasmid pGN brings together several advantages for this methodology, which is applicable to any gene. Since the homologous recombination event can be fairly rare (of the order of 1 per 1000 non-homologous integrations), it is necessary to be able to analyse a large number of clones whose resistance to G418 is sufficiently strong to be expressed in any part of the genome. The modifications made to the phosphotransferase transcription unit are the perfect answer to these problems. The method of mutagenesis by homologous recombination is equivalent to inactivating a gene by an insertion or a substitution, but plasmid pGN has the further advantage of being able to substitute the expression of β-galactosidase for that of the mutated gene. Lastly, the MCS facilitates the clonings of genomic fragments.

EXAMPLES:

I - Construction of plasmid pGN

The intermediate plasmids are numbered according to their step.

1st step:

Insertion of an Xho I site into the Bgl I site of pRSV neo

Insertion of an Xho I linker into the Bgl I site of pRSV neo, filled in with the Klenow fragment of E. Coli DNA polymerase.

2nd step:

Insertion of a Cla I site into the Nde I site of

plasmid p1

Insertion of a Cla I linker into the Nde I site of p1, filled in with Klenow polymerase.

3rd step:

Insertion of Py F9.1 enhancer into the Cla I site of plasmid p2

Insertion of Pvu II-Pvu II Py F9.1 enhancer, isolated through a unique site, Acc I, into the Cla I site of p2. Selection of a clone containing two enhancers oriented in the "late promoter" direction.

4th step:

Sma I-Hpa I deletion of plasmid p3

The two enzymes, giving "blunt-ended" ends, may be ligated directly. This deletion removes the intron of the SV 40 t antigen, which is not very useful, and decreases appreciably the size of the phosphotransferase transcription unit.

5th step:

Insertion of an Xho I site into the Bam HI site of pCH110

Insertion of an Xho I linker into the Bam HI site of plasmid pCH 110 (Pharmacia), filled in with Klenow polymlerase [sic].

6th step:

Insertion of 3' lac.Z-SV 40 polyA into plasmid p4

The 3' portion of the β-galactosidase coding frame, followed by the SV 40 virus polyadenylation signal, is isolated from plasmid p5 through the Xho I-Aat II sites and cloned into plasmid p4 through the same sites.

7th step:

Insertion of 5' lac.Z into the vector KS-

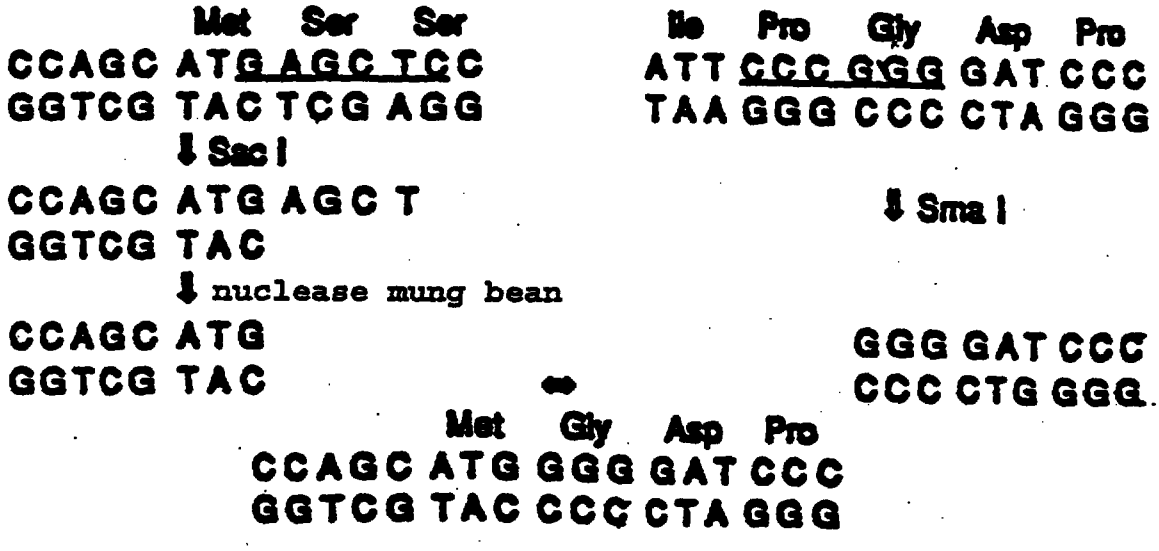
The 5' portion of the β-galactosidase coding frame is isolated from plasmid pMC 1871 (Pharmacia) through the Pst I-Sac I sites and cloned into the vector KS- (Stratagene) through the same sites.

8th step:

Fusion of a Hox-3.1 genomic sequence with 5' lac.Z

A genomic sequence of the Hox-3.1 gene, cloned into the

vector KS-, is purified by successive digestions with the enzyme Sac I, then with mung bean nuclease and lastly with the enzyme Apa I. This insert is fused with the 5' portion of the β-galactosidase coding frame by cloning into plasmid p7 digested with Apa I-Sma I. The protein thus fused contains the translation initiation codon of the Hox-3.1 gene followed by the β-galactosidase coding frame (thereafter verified by sequencing).



9th step:

Insertion of Hox-3.1-5' lac.Z into plasmid p6

The Hox-3.1-5' lac.Z fusion is isolated from plasmid p8 through the Apa I-Sac I sites and cloned into plasmid p6 through the same sites. This cloning has the effect of re-forming the β-galactosidase coding frame in its entirety.

10th step:

Insertion of the Neo® gene into the vector KS+

The neomycin-resistance gene (bacterial promoter and phosphotransferase coding frame) is isolated from pRSV neo through the Hind III-Eco RI sites and cloned into the vector KS+ (Stratagene).

11th step:

Mutagenesis of the Neo® initiation sequence in p10

The phosphotransferase translation initiation sequence is modified so as to be identical to the

consensus sequence observed in vertebrates, and thus to permit a higher level of translation initiation, hence an enhanced resistance to G418 for mammalian cells. The modification also creates an Apa LI site which enables the efficiency of the mutagenesis to be monitored.



An oligonucleotide (CTTGTTCAATCATGGTGCACGATCCTCA) containing a region of mispairing with the pRSV neo sequence (underlined) is synthesized (Gene Assembler, Pharmacia) and then phosphorylated with bacteriophage T4 polynucleotide kinase. A single-stranded template of plasmid p10 is prepared with the aid of the f1 origin of plasmid KS+ and hybridized with the mutagenesis oligonucleotide. The second strand is synthesized and repaired with Klenow polymerase and the bacteriophage T4 DNA ligase. After transformation of bacteria, mutated clones are screened using the oligonucleotide labelled with ^{32}P. The mutagenesis was verified by digesting with Apa LI as well as by sequencing.

12th step:

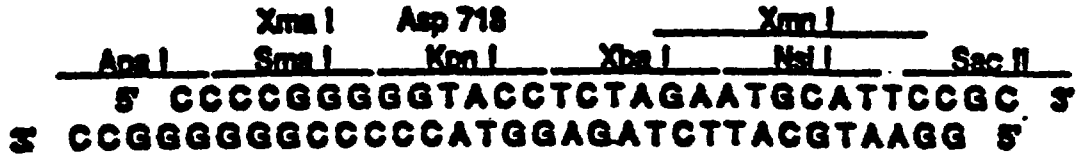
Replacement of the initiation sequence in plasmid p9

A fragment containing the modified translation initiation sequence of the neomycin-resistance gene is isolated from plasmid p11 with the enzymes Hind III-Eag I and cloned into plasmid p9 through the same sites.

13th step:

Insertion of the multiple cloning site into plasmid p12

Two complementary oligonucleotides are synthesized (Gene Assembler, Pharmacia) and then phosphorylated. After pairing, the MCS is cloned into the Apa I-Sac II sites of plasmid p12 by means of its cohesive ends



The multiple cloning site was also verified by sequencing.

II - Addition of the "flanking" sequences to the ends of plasmid pGN linearized upstream of lac.Z' through a site of the MCS

The flanking sequences used were chosen in accordance with the desired insertion site (for example, Hox-3.1, see Fig. 2 a and b pGMA and pGMD).

In the construction of the mutagenesis plasmid pGMD, two arms of DNA homologous at the Hox-3.1 locus were cloned at the Apa I-Nsi I and Nsi I-Sac II sites of the vector pGN. The 5' arm begins at the Sac II site (CCGCGG) at nucleotide 219 of the Hox-3.1 c21 cDNA. This fragment extends over 6.8 kb on the 5' side up to the first BamHI site. The 3' arm begins at the Apa I site (GGGCCC) at nucleotide 885 of the c21 cDNA. This fragment extends over 1.5 kb on the 3' side up to the first PstI site. An NsiI linker was inserted into the BamHI site of the 5' fragment and into the PstI site of the 3' fragment. The 5' and 3' arms were cloned into the vector pGN in the Nsi I-Sac II and Apa I-Nsi I sites, respectively. The sequence of c21 Hox-3.1 cDNA is published (ref. 1).

The mutagenesis plasmid is linearized by digestion with Nsi I before electroporation of ES cells. Its ends are formed from two genomic arms cloned at the Apa I-Nsi I and Nsi I-Sac II sites of the vector pGN.

Plasmid pGMD does not possess a polyadenylation signal after the resistance gene but, on the contrary,

possesses a sequence rich in AU, responsible for a selective degradation of mRNA, inserted into the Hox-3.1 intron sequence of the plasmid.

Another mutagenesis plasmid, pGMA, possesses the same structure as pGMD but contains the polyadenylation and transcription termination signals of SV40 and does not possess the AU sequence for degradation of mRNA downstream of the Neo^r gene. The object of these modifications was to reduce the level of Neo^r transcripts in clones arising from random integration. On the other hand, clones arising from homologous recombination events between pGMD and a Hox-3.1 locus should have unimpaired growth during selection with G418, the AT sequence for degradation of mRNA being removed by the process of recombination itself, or spliced with the Hox-3.1 intron.

In the experimental steps which follow, the protocol described by Thompson et al. 1989 was followed for the production of chimeric animals.

III - <u>Transformation of mouse embryonic cells</u>

The method described by Thompson et al. 1989 was used to transfect mouse embryonic cells. The use of the electroporation technique effects the introduction of only one copy of the foreign gene (lac.Z) per cell. After transfection, several clones expressing β-galactosidase were isolated.

The mutagenesis plasmids pGMD and pGMA were linearized and introduced by electroporation into ES cells so as to promote the insertion of only one copy into the genome (ref. 11).

The initial transfections were performed in order to compare the efficiency of targeting of Hox-3.1 of plasmids pGMA and pGMD (see Table I).

Homologous recombination in the Hox-3.1 gene				
Exp.	Mutagenesis plasmid	No. of the set analysed	No. of clones forming the set	No. of positive PCR results
I	pGMA	3	600	0(2)
II	pGMD	5	250	3(5)
III	pGMD	84	2-3	5(5)

Table I

The ES cell line "CCE" (ref. 16) was maintained continuously on nutrient fibroblast layers (ref. 17). For experiments I and II, 1.5×10^7 ES cells in 1.5 ml HeBS were electroporated (ref. 11) at 200 V with 40 mg of linearized plasmid, and then plated out on four culture dishes (diameter 100 mm). For experiment III, shock was performed under the same conditions, but a quarter of the cells were plated out on four 24-well plates. The following day, 250 μg ml^{-1} G418 were added. Each transfection gave rise to approximately 2400 clones with pGMA and approximately 1000 clones with pGMD.

The average number of clones of G418-resistant ES cells in each set is shown in Table 1, together with the number of sets giving a positive result with the PCR technique. A positive result means that a 1.6-Kb band could be observed on ethidium bromide-stained agarose gel (see Fig. 7). The number of sets giving a positive signal after a Southern analysis of the PCR mixture and hybridization of a specific probe which did not contain sequences of the primers is shown in brackets (Fig. 8).

Detection of homologous recombination with PCR

PCR was performed on 10^5 cells of a set of 250 clones from the transfection II (see lane D of Fig. 7). In the other lanes, four sets from the transfection III were analysed together by mixing approximately 4 x 5000 cells. The primers 07 and 08 used in the PCR surround the 3' Hox-3.1 sequence of the mutagenesis plasmid (Fig. 2). The 1.6-Kb fragment overlapping this 3' sequence can be

amplified only in the case of a homologous recombination. Lanes 2, 3 and D illustrate positive results.

The DNA of ES clones was prepared at the time of making replica filters using the "boiling-proteinase K digestion boiling" method (ref. 18). 40 amplification cycles (40 seconds at 94°C, 1 minute at 60°C, 7 minutes at 72°C) were performed in a reaction mixture of 100 µl containing 67 mM Tris-HCL [sic] (pH 8.6), 16.7 mM $(NH_4)_2SO_4$, 6.7 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.01% (w/v) gelatin, 200 µM dATP, dTTP and dCTP, 100 µM dGTP, 100 µM 7-deaza-dGT, 600 ng of each primer (07 : AACTTCCCTCTCTGCTATTC and 08 : CAGCAGAAACATACAAGCTG) and 3U Taq polymerase (Perkin Elmer Cetus), covered with 100 µl paraffin. Half of the reaction mixture was applied to a 0.7% agarose gel stained with ethidium bromide. The size marker is an Eco RI + Hind III digest of lambda DNA.

Southern analyses

Three independent clones of ES cells containing the mutated Hox-3.1 (identified by PCR) were isolated from the positive sets using pipettes. Their DNA was examined by Southern analysis after digestion with the restriction enzymes indicated in Figure 8, so as to confirm the specific targeting and to make a distinction between the recombinant and wild-type loci. Two different probes were used in the analysis of the 3' end of the Hox-3.1 loci in the mutated clones and in the unmutated ES cells acting as control (Fig. 8c). The first probe "a" was contained in the Hox-3.1 sequences of the mutagenesis plasmid, and demonstrated the number of integrations of vector and their physical linkages. One of the three recombinant clones contained, in addition, a copy of the plasmid integrated at random (Fig. 8a, clone F2). The second probe "b", which was not contained in the mutagenesis vector, made a distinction between the recombinant and wild-type Hox-3.1 alleles (Fig. 8b). The recombinant Hox-3.1 locus displayed, with the two probes, the picture of hybridization expected from the restriction maps of the mutagenesis vector and of the intact locus. In addition, the existence of two

recombination domains in the 3′ arm of the vector was confirmed by the presence or absence of the AT sequence in the recombinant Hox-3.1 locus (for example Fig. 8, clone L5). The 5′ end of the Hox-3.1 locus was also analysed for the homologous recombination event. Restriction enzymes not possessing sites in the 6.8-Kb 5′ Hox-3.1 sequence of the mutagenesis vector were used in the digestion of the DNAs of the recombinant clones. These DNAs were then subjected to pulsed-field electrophoresis in order to differentiate the fragments of high molecular weight. A Southern analysis of this gel also showed the alleles which had recombined correctly and the wild-type Hox-3.1 alleles, using a probe possessing a sequence upstream of the mutagenesis plasmid.

The Southern analyses demonstrated that an allele of the Hox-3.1 gene was recombined as predicted. The homologous recombination was equivalent to a double crossing-over between the genomic arms of the mutagenesis plasmid and the homologous chromosomal sequences (Fig. 2).

In the recombinant clones, the lac Z gene was placed under the control of the Hox-3.1 promoter and regulatory sequences upstream of the AUG codon, but the 3′ signals for maturation of mRNA originated from SV40. In these recombinant clones, the expression of lac Z was not detectable by staining with β-Gal, which is consistent with the absence of transcription of Hox-3.1 in ES cells, determined by analysis of protection of RNase. β-Gal activity could be induced in some cells after 3 or 4 days of culture in the presence of 5×10^{-7}M retinoic acid, conditions known to induce the transcription of Hox-3.1 (ref. 19).

Using the mutagenesis vector pGMA, which displays a total homology of 8.3 Kb DNA with the Hox-3.1 locus, a 120-bp fragment was replaced by an insertion of 7.2 Kb. The frequency of this targeted replacement (1/900) is comparable to that obtained recently (1/1000) with HPRT (ref. 13) or with En-2 (1/260) (ref. 20), the inserted heterologous fragment being, however, in these latter

cases much smaller in size (1.1 and 1.5 Kb, respectively). Surprisingly, it was found that a very high frequency of homologous recombination (1/40) could be obtained with the vector pGMD. The removal of the 3' signals from maturation of mRNA and the addition of the sequence for degradation of mRNA to the neomycin-resistance gene had the effect of reducing the total number of clones resistant to G418 by 2.4 (Table I). The specific targeting ratio was almost 10 times as high (900/40). The actual mechanism of homologous recombination must have been affected in the experiments with pGMD. A possible explanation of these results would be that an AT sequence of 51 bp could furnish, in vivo, an open loop in the mutagenesis plasmid on account of its lower melting temperature. If the neighbouring Hox-3.1 sequences of pGMD can be influenced by this opening, on each side of the AT region, they could react more efficiently in the single-stranded state with the Hox-3.1 chromosomal locus. The model of mitotic recombination in yeast suggests that it would be initiated by such an exchange of strands, although the mechanism of homologous recombination remains unknown in more complex eukaryotes.

Figure 8 shows the results of Southern analysis performed on positive individual clones (L5 and F2) and ES (CCE) cells.

The probes used hybridize only with Hox-3.1 sequences included in the vector (a) or excluded from the mutagenesis vector (b). The picture of hybridization of the recombinant Hox-3.1 locus (open triangles) is clearly distinguished from the wild-type locus (black triangles). Stars indicate the hybridization bands of a copy of the plasmid which has been integrated at random. The size marker is an Eco RI + Hind III digest of lambda DNA.

Figure 8(c) shows the restriction maps of the recombinant (rec.) and wild-type (wt) Hox-3.1 alleles. The portions of the mutagenesis vector and of the Hox-3.1 locus are indicated with the same symbols as those used in Figure 2. In this case, the AT sequence has been integrated by homologous recombination. The vertical

arrow indicates the 3' end of the mutagenesis plasmid. The localization of the probes "a" and "b" used in the Southern analysis is also indicated. The abbreviations used in Figure 8 are as follows: B, Bam HI; D, Dra I, E, Eco RI; H, Hind III; S, Sal I; X, Xho I.

IV - Production of chimeric embryos

A microinjection into blastocysts was performed with two recombinant ES clones containing an intact Hox-3.1 allele and a recombinant allele; these clones did not contain any other copy of the mutagenesis plasmid. The karyotypes of these cells were normal.

Ten to fifteen mutated cells were microinjected per blastocyst. After reimplantation in surrogate mothers, the embryos were collected at 9.5, 10.5 and 12.5 days p.c. and analysed for expression of lac Z. The transcription pattern of Hox-3.1 at these stages had been determined beforehand by in situ hybridization analysis (ref. 1). The Hox-3.1 transcripts are detectable for the first time at the late gastrulation stage, and are distributed throughout all the tissues of the posterior portion of the animal. Later, the distribution becomes gradually limited in space and tissue-specific. At the stage of 12.5 days p.c., transcription is localized in the cervical region of the neural tube, at the level of the heart. During embryogenesis, the distribution of the transcription of Hox-3.1 hence undergoes modification. The 10.5 days p.c. stage seems to be a period of transition, transition taking place both in the two hind regions and in the cervical neural tube.

In 9.5- and 10.5-day p.c. chimeric embryos, the portion caudal to the posterior bud displayed intense β-Gal activity, while the marker was never detected in the anterior thoracic region or the head (Fig. 9a). In the posterior region, cells stained with β-Gal were observed in all the tissues and all the embryonic layers. Between the two buds which give rise to the limbs, stained cells were distributed in restricted areas, in the superficial ectoderm (Fig. 9b), as well as in the hind regions (Fig. 9c) and, in the form of straight lines

or stripes, in the neural tube (Fig. 9b). These stripes displayed an irregular and asymmetric distribution on the wall of the neural tube. Transcription of Hox-3.1 was not detected in the thin layer of cells towards the closure of the neural tube. Possibly, these cells did not withstand the treatment supplied during the in situ hybridization. It was observed that cells of the neural ectoderm form part, very early, of different portions of the nervous system and migrate in a radial direction, following narrow lateral movements (ref. 21). Hence these results are consistent with this observation.

The expression of Lac Z has hence illustrated correctly the first part of the transcription of the Hox-3.1 homeo gene, that is to say in all the tissues of the caudal regions of the 9.5- and 10.5-day p.c. embryos, and furnished new information regarding the mode of transcription of Hox-3.1.

On the other hand, the expression of Lac Z was not observed in the cervical regions of the neural tube of 12.5-day chimeric embryos, or in the anterior region of 10.5-day embryos; this was not the result expected from the in situ hybridization studies. The subsequent phase of transcription of Hox-3.1, observed from day 10.5 in the very localized areas of the neural tube, was not revealed by β-Gal activity. A possible explanation for this result would be that, although the expression of Lac [sic] Z is under the control of the Hox-3.1 promoter, the 3' sequences of Hox-3.1 are absent in the reporter gene. It is possible that some 3' sequences of the Hox-3.1 AUG initiation codon have an influence on the late expression of Hox-3.1 in the anterior domain. A "gene dosage" effect might also explain this result. The self-activation of several homeo genes in Drosophila has been demonstrated genetically or suggested by the formation of complexes between the DNA and the proteins of the homeo boxes.

If the late component of the transcription of Hox-3.1 in the neural tube is maintained by a similar mechanism, the inactivation of an allele could have a

dominant effect in cells of the neural ectoderm. Since one allele only would produce the Hox-3.1 protein, the activation signal would be diluted on the two promoters. The reduction of self-inactivation in both loci could thus lead to a total stoppage of the initiation of transcription. This might explain why no expression of Lac Z was detected in the cervical region of the neural tube of 10.5- and 12.5-day embryos.

V - <u>Passage of the modification into the germ cell line: production of transgenic animals</u>

The effects in F_1 and in F_2 of the modification made by the targeted insertion were observed after reproduction of the chimera. It was established that the modification passed into the germ cell line.

BIBLIOGRAPHY

1. Le Mouellic, H., Condamine, H. and Brûlet, P. (1988). Genes & Development. 2, 125-135.

2. Cavenir, D.R. (1987). Nucleic Acids Res. 15, 1353-1361.

3. Sanes, J.R. Rubenstein, J.L.R. and Nicolas, J.F. (1986) EMBO J. 5, 3133-3142.

4. Nolan, G.P., Fiering, S., Nicolas, J.F. and Herzenberg, L.A. (1988) Proc. Natl. Acad. Sci. USA 85, 2603-2607.

5. Gorman, C., Padmanabhan, R. and Howard, B.H. (1983). Science. 221, 661-553.

6. Gormann, C.M., Merlino, G.T. Willingham, M.C. Pastan, I. and Howard, B.H. (1982) Proc. Natl. Acad. Sci. USA 79, 6777-6781.

7. Southern, P.J. and Berg, P. (1982) J. Mol. Appl. Genet 1, 327-341.

8. Herbomel, P., Bourachot, B. and Yaniv, M. (1984). Cell. 39, 653-662.

9. Robertson, E.J. (1987). Teratocarcinomas and embruonic [sic] stem cells: A practical approach. IRL Press, Oxford.

10. Kahn, A. (1988). Médecine/Sciences 4, 515-518.

11. G. Chu, Hayakawa H., Berg. P. (1987) Nucleic Acid Research 15, No. 3, 1311-1326.

12. Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., Melton, D.W., (1989) Cell, 56, 313-321.

13. Thomas, K.R., Capecchi, M.R., (1987) Cell, 51 503-512.

14. Evans, M.J., Kaufmann, M.H. (1981) Nature, 292, 154-155,

15. Robertson, E.J., (1986) Trends in Genetics, 9-13

16. Robertson, E., Bradley, A., Kuehn, M. & Evans, M. Nature 323, 445-448 (1986).

17. Robertson, E.J. in Teratocarcinomas and Embryonic Stem Cells (ed. Robertson, E.J.) 71-112 (IRL, Oxford, 1987).

18. Kim, H.S. & Smithies, O. Nucleic Acids Res. 16, 8887-8903 (1988).

19. Breier, G., Bucan, M., Francke, U., Colberg-Poley, A.M. & Gruss, P. EMBO J.5, 2209-2215 (1986).

20. Joyner, A.L., Skarnes, W.C. & Rossant, J. Nature 338, 153-156 (1989).

21. McKay, R.D.G. Cell 58, 815-821 (1989).

CLAIMS

1. Method for the specific replacement or insertion of a gene, in particular by targeting a DNA, termed insertion DNA, consisting of the whole of a gene or of a portion of a gene capable of being rendered functional, or whose functioning may be rendered more effective, when it is recombined with a complementary DNA to yield then a complete recombinant gene in the genome of a eukaryotic cell, characterized in that

- the insertion site is in a chosen gene, termed chosen recipient gene, containing the complementary DNA, and in that

- eukaryotic cells are transfected with a vector containing an insert itself comprising the insertion DNA and two so-called "flanking" sequences on each side of the insertion DNA, which are homologous, respectively, to two genomic sequences which adjoin the desired insertion site in the recipient gene,

- the insertion DNA being heterologous with respect to the recipient gene, and

- the flanking sequences being chosen from those which constitute the abovementioned complementary DNA and which permit, by homologous recombination with corresponding sequences of the recipient gene, the reconstitution of a complete recombinant gene in the genome of the eukaryotic cell, on condition that, when the insertion DNA consists of a coding sequence, it is not the coding sequence of a gene coding for a selectable agent.

2. Method for the specific replacement or insertion of a gene, in particular by targeting a DNA, termed insertion DNA, consisting of the whole of a gene or of a portion of a gene capable of being rendered functional, or whose functioning may be rendered more effective, when it is recombined with a complementary DNA to yield then a complete recombinant gene in the genome of a eukaryotic cell, characterized in that

- the insertion site is in a chosen gene, termed chosen recipient gene, containing the complementary DNA, and in that

- eukaryotic cells are transfected with a vector containing an insert itself comprising the insertion DNA and two so-called "flanking" sequences on each side of the insertion DNA, which are homologous, respectively, to two genomic sequences which adjoin the desired insertion site in the recipient gene,
- the insertion DNA being heterologous with respect to the recipient gene, and
- the flanking sequences being chosen from those which constitute the abovementioned complementary DNA and which permit, by homologous recombination with corresponding sequences of the recipient gene, the reconstitution of a complete recombinant gene in the genome of the eukaryotic cell,
- the recipient gene being a gene which is not normally transcribed in the eukaryotic cell.

3. Method according to Claim 1 or 2, characterized in that the said insertion DNA contains either a coding sequence or a regulatory sequence,
- the flanking sequences being chosen so as to permit, by homologous recombination, depending on the case, either the expression of the coding sequence of the entire insertion DNA under the control of the regulatory sequences of the recipient gene, or the expression of a coding sequence of the recipient gene under the control of regulatory sequences of the insertion DNA.

4. Method according to any of Claims 1, 2 and 3, characterized in that the insertion DNA contains a coding sequence lacking a regulatory element, in particular a promoter which is distinctive to it.

5. Method according to Claim 3, characterized in that the insertion DNA contains a regulatory sequence.

6. Method according to Claim 5, characterized in that the endogenous promoter of the recipient gene is inactive in the transfected eukaryotic cells.

7. Method for the specific replacement of a regulatory sequence of a gene, termed recipient gene, by targeting a DNA, termed insertion DNA, the said insertion DNA containing a regulatory sequence, characterized in

that

- eukaryotic cells are transfected with a vector containing an insert itself comprising the insertion DNA and two so-called "flanking" sequences on each side of the insertion DNA, which are homologous, respectively, to two genomic sequences which adjoin the desired insertion site in the recipient gene,
- the insertion DNA being heterologous with respect to the recipient gene, and
- the flanking sequences being chosen so as to permit, by homologous recombination, the expression of a coding sequence of the recipient gene under the control of regulatory sequences of the insertion DNA.

8. Method according to Claim 7, characterized in that the endogenous promoter of the recipient gene is normally inactive in the transfected eukaryotic cells.

9. Method according to any one of the preceding claims, characterized in that the recipient gene is present in the genome of the eukaryotic cell in at least two copies.

10. Method according to any one of the preceding claims, characterized in that each of the flanking sequences has a length greater than 150 base pairs, and smaller than the length of the recipient gene.

11. Method according to any one of Claims 1 to 10, characterized in that the eukaryotic cells are embryonic stem (ES) cells.

12. Method according to one of the preceding claims, characterized in that the insertion DNA is a gene which is heterologous or a portion of a gene which is heterologous with respect to the transfected species.

13. Method according to one of the preceding claims, characterized in that the vector contains sequences interposed between the insertion gene and the flanking sequences.

14. Method according to Claim 13, characterized in that the interposing sequences contain a sequence coding for a selective agent permitting selection of the transformants and, where appropriate, a marker gene, for

example the LacZ [sic] gene.

15. Method according to one of the preceding claims, characterized in that the transfection is performed by electroporation.

16. Method according to one of the preceding claims, characterized in that the polymerase chain reaction (PCR) technique is used to amplify the DNA sequence of the locus where the insertion takes place in order to verify that the insertion has taken place in the desired site.

17. Method according to any one of the preceding claims, characterized in that the insertion DNA contains a portion of the interferon or interleukin gene, of genes for receptors such as the β_3-adrenergic retinoic acid receptor or the HIV receptor, or of a gene linked to a disorder, or of the lacZ gene.

18. Method according to any one of the preceding claims, characterized in that the insertion DNA contains, between the flanking sequences, on the one hand a DNA sequence intended to be recombined with the complementary DNA in the recipient gene to yield a recombinant gene, and on the other hand a sequence coding for a selective agent permitting selection of the transformants and a promoter permitting expression of the selective agent.

19. Method for the production of transgenic animals, characterized in that ES cells are transfected by the method according to one of Claims 1 to 18 and screened for the homologous recombination event, namely the correct integration of the foreign gene, the cells are injected into embryos at a stage at which they are capable of integrating the transfected cells, for example at the blastocyst stage, the embryos are then reimplanted in a surrogate mother, and the chimeric individuals obtained at the end of gestation, and in which colonization by ES cells of the germ line is observed, are mated to obtain transgenic animals which are heterozygous for the replaced gene.

20. Nucleic acid sequence, termed insertion sequence, comprising:

i) a sequence comprising all or part of a gene

capable of being rendered functional, or whose functioning may be rendered more effective, when it is recombined with a complementary DNA to yield then a complete recombinant gene in the genome of a eukaryotic cell, on condition that, when the portion of the gene consists of a coding sequence, it is a coding sequence other than that of a gene coding for a selectable agent,

ii) on each side of the sequence i), two so-called "flanking" sequences homologous, respectively, to the two genomic sequences which adjoin the desired insertion site in the recipient gene.

21. Nucleic acid sequence according to Claim 20, characterized in that it comprises, in addition, between the flanking sequences, a sequence coding for a selective agent permitting selection of the transformants and a promoter permitting expression of the selective agent.

22. Nucleic acid sequence according to Claim 20 or 21, characterized in that it comprises, in addition, a sequence removed by the homologous recombination process, and enabling the transformants originating from homologous recombination events to be selected.

23. Vector containing the nucleic acid sequence according to either of Claims 20 and 21.

24. Nucleic acid sequence containing plasmid pGN as illustrated in Figure 1 and, on each side of the pGN sequence, two flanking sequences homologous, respectively, to the two genomic sequences which adjoin a desired insertion site in a recipient gene.

25. Eukaryotic cells transformed by the method of Claim 1, 2 or 7.

26. Cells according to Claim 25, characterized in that they are ES cells.

27. Use of the nucleic acid sequence according to Claim 20 for the modification of eukaryotic cells for the purpose of rendering the expression of the said sequence, or of an endogenous sequence, functional or more effective.

28. Nucleic acid sequence comprising:

i) a regulatory sequence intended to be recombined

* C L M S *

with a complementary DNA to yield then a complete recombinant gene in the genome of a eukaryotic cell, and

ii) on each side of the sequence i), two so-called "flanking" sequences homologous, respectively, to the two genomic sequences which adjoin the desired insertion site in a recipient gene.