PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00	A2	(11) International Publication Number: **WO 99/28449**
		(43) International Publication Date: 10 June 1999 (10.06.99)

(21) International Application Number: PCT/GB98/03558

(22) International Filing Date: 27 November 1998 (27.11.98)

(30) Priority Data:

9725311.6	28 November 1997 (28.11.97)	GB
9725313.2	28 November 1997 (28.11.97)	GB
9806072.6	20 March 1998 (20.03.98)	GB
9824275.3	5 November 1998 (05.11.98)	GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: SHEN, Sanbing [CN/GB]; 118/2F2 Thirlestane Road, Edinburgh EH9 1AS (GB). SCHEDL, Andreas [DE/DE]; Dianastrasse 31, D–13469 Berlin (DE). HARMAR, Anthony, John [GB/GB]; 56 Liberton Place, Edinburgh EH16 6NA (GB).

(74) Agents: HARDING, Charles, Thomas et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: VECTORS

(57) Abstract

YAC vectors are described which may be used individually or in combination to prepare a YAC. A first YAC vector comprises a nucleotide sequence wherein the nucleotide sequence comprises the sequence presented as SEQ ID No.1 or SEQ ID No.4 or a variant, homologue or derivative thereof.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

VECTORS

The present invention relates to vectors, in particular vectors that are suitable for use as or with or in the preparation of a yeast artificial chromosome.

A yeast artificial chromosome - otherwise known as a YAC - comprises the structural components of a yeast chromosome into which it is possible to clone very large pieces of DNA. By way of example, it is generally possible to clone into a YAC stretches of DNA that are up to about 1000kb long - which are much larger than when compared with the stretches of DNA that can be cloned into other cloning vectors such as plasmids (typically up to 20kb stretches of DNA), bacteriophage λ (typically up to 25kb stretches of DNA), cosmids (typically up to 45kb stretches of DNA) and the P1 vector (typically up to 100kb stretches of DNA) (see Lodish *et al* 1995 Molecular Cell Biology 3rd Edition, Pub. Scientific American Books, page 233).

YACs were initially proposed by Burke *et al* (Burke, D.T., G.F. Carle and M.V. Olson (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-812). General introductory teachings on YACs have been presented by T.A. Brown 1995 (Gene Cloning, An Introduction, 3rd Edition, page 325, Pub. Chapman & Hall, pages 139-142).

Typically, a YAC contains the following essential functional elements: a centromere, two telomeres and one or more origins of replication. The centromere is required to correctly distribute the chromosome to daughter cells during cell division; the telomeres are required to ensure correct replication; and the replication origin(s) is (are) present to ensure initiation of DNA replication. The origins of replication are sometimes referred to as ARS elements.

YACs are typically prepared from YAC vectors. These vectors are typically circular. When they are needed to be used to prepare the YAC they are then linearised - such as by use of specific restriction enzymes.

To date, a number of YAC vectors have been proposed in the literature. Examples of such vectors are pYAC2 and pYAC4 which are discussed in US-A-4889806. Other YAC vectors are disclosed in WO-A-95/03400. Other YAC vectors include pYAC3 and pYAC5.

YAC vectors - such as pYAC3, pYAC4 and pYAC5 - are essentially a pBR322 plasmid into which a number of yeast genes have been inserted. These genes include a yeast centromere region (called *CEN4*), two telomere regions (called *TEL*), and two selectable marker genes (called *URA3* and *TRP1*). The *TEL* sequences do not correspond to the full genomic telomere sequences. Nevertheless these partial sequences still function as telomeric sequences. One replication origin (called *ori*, such as *ARS1*) is positioned intermediate *CEN4* and *TRP1*. When used for cloning, the YAC vectors are cut with *BamH*I and a second restriction enzyme (*Sma*I for pYAC3, *Eco*RI for pYAC4 and *Not*I for pYAC5) to produce two vector arms. The fragments are then ligated with a nucleotide sequence of interest (which for ease of reference shall be called "NOI") which has been digested with corresponding restriction enzymes.

The resultant linear structure (which is not drawn to scale) is presented as Figure 17. This resultant linear structure - which is a YAC - comprises in the correct orientation the essential functional features of a chromosome.

YACs have been - and are still being - used to map the human genome. In this case, the NOI is a gene or fragment thereof whose full sequence (or even function) may not have been determined. In these studies, YACs are used to create genomic libraries which are then screened. By way of example, the physical map of the human Y chromosome and the long arm of chromosome 21 have been determined through analysis of long segments of human DNA cloned into YACs by *inter alia* sequence tagged sites. This work is summarised in Lodish *et al* 1995 (*ibid*, page 285).

However, the use of YACs is by no means limited to mapping of the human genome. For example, the use of YACs has led to the preparation of physical maps of the *Drosophila* X chromosome containing the shibire (*shi*) locus (Bliek and Meyerowitz 1991 Nature 351 441). YACs have also been proposed to map plant genomes such as the *Arabidopsis* genome.

In addition to their usage in genome mapping, YACs have now another important use. In this regard, it has been recently found that under certain conditions YACs can be introduced into mammalian cells (such as murine cells), whereupon they can behave functionally the same as (or very similar to) endogenous chromosomes. In this regard, it is possible to deliver - such as by use of cell fusion techniques, microinjection techniques or transfection techniques - YACs containing large fragments of human DNA (i.e. the NOI) into the mouse germline. This was initially achieved with a 670kb YAC containing a human X-chromosome fragment (Jacobovitis *et al* 1993 Nature 362 pages 255-258). By way of further example, WO-A-94/23049 reports on a YAC containing the gene coding for β-amyloid precursor protein.

Hence, YACs now enable workers to analyse an or the *in vivo* functional behaviour of a NOI, such as a human NOI. For these studies, prior knowledge of the sequence and/or function of the NOI need not be necessary.

A review of these *in vivo* functional studies has been presented by Jacobovitis (Current Biology 1994 vol 4 No 8 pages 761-763), who states:

> *"The ability to replace mouse genes with their human equivalents using yeast artificial chromosome technology provides a powerful new technique for studying the regulation and function of human genes"*.

Other reviews and details on these studies and techniques have been presented by Larin and Lehrach (Genet. Res. Camb 1990 56 pp 203-208), Schedl *et al* (Nucleic Acids Research vol 20 No. 20 pages 3073-3077), and Montoliu *et al* (Reprod. Fert. Dev. 1994 6 577-584).

YACs may even be used to study the functional behaviour of mutant genes. In this regard, WO-A-95/14769 reports on a method of producing a mouse that expresses human mutant protein sequences that utilises the "pop-in/pop-out" method in combination with YAC technology to insert mutations into YACs and thereby derive stem cells capable of being used in the development of transgenic mice. This particular method comprises obtaining a gene contained within a YAC, introducing a predetermined mutant human DNA sequence (which is the NOI) into the YAC by homologous recombination, utilising transgenics to insert the mutant gene into embryonic stem cells, and injecting the stem cells into blastocysts to derive a transgenic mouse that expresses the mutant protein sequences. The "pop-in/pop-out" method is described by Rothstein (1991 Methods In Enzymology vol 194, Guide To Yeast Genetics and also in Molecular Biology. Eds. Gutherie *at al*, San Diego: Academic Press. Pages 281-301) and McCormick *et al* (TCM Vol 6 No. 1 1996 pages 16-24).

According to Schedl *et al* (reference 28): the transfer of YAC DNA into mammalian cells came into focus of interest soon after the first report of yeast artificial chromosomes. In a successfully transgenic cell the genes contained on the YAC are embedded in an almost natural chromosomal context, which should ensure regulation of expression comparable to their endogenous counterparts. Therefore, such a method should allow a more rapid identification of genes by complementation analysis as well as detailed studies of function of genes and their regulation *in vivo*. Schedl *et al* (*ibid*) describe a method for the isolation of purified and concentrated YAC DNA as well as protocols for microinjection into somatic cells in culture and fertilized mouse oocytes.

Despite the fact that YACs have many important applications, there are nevertheless problems associated with their production and their usage.

For example, the low efficiency with which transgenic animals are produced using YAC DNA (1-5%) compared to DNA from conventional vectors (approximately 10%) is probably caused by the low concentration of YAC DNA available for injection. Assuming that 2 pl of 500kb YAC at a concentration of 1 ng/μl is injected into a pronucleus, a fertilised egg receives only 1 molecule of YAC DNA. Amplification of YACs in yeast should provide a possible method for the isolation of more concentrated YAC DNA which should lead to more successful generation of YAC transgenic animals (i.e. a transgenic mammal that comprises a YAC). However, to date, there has not been a totally acceptable solution to this problem.

For example, Smith *et al* proposed use of the YAC vector pCGS990 (Smith *et al* Mammalian Genome 1993 4 pages 141-147). Even though that YAC vector goes some way to overcoming this problem, it nevertheless comprises the *TK* gene (i.e. the herpes simplex virus thymidine kinase gene) as a selectable marker. This is problematic as expression of this gene can cause male sterility in transgenic animals.

Alternatively, or in addition, with current techniques it is not currently possible to readily monitor the *in vivo* expression pattern of a NOI that has been introduced into an organism - such as a mouse. Current techniques - such as *in situ* hydridisation, Polymerase Chain Reaction (PCR) and Northern Blotting - are laborious to carry out.

In addition, and by way of further example, none of the earlier reported studies has reported on the co-expression of a NOI and a reporter gene in a YAC.

Thus there are problems associated with the known vectors for preparing YACs.

The present invention seeks to improve upon the existing techniques associated with the preparation of and usage of YACs.

In this regard, the present invention seeks to provide two types of vectors that can be used on their own or in combination with each other and thereby overcome at least one of the above-mentioned problems.

According to a first aspect of the present invention there is provided one or more of the following embodiments, which for ease have been presented as numbered paragraphs:

1. A YAC vector comprising an IRES.

2. A YAC vector comprising a reporter gene wherein the expression product of the reporter gene is capable of producing a visually detectable signal.

3. A YAC vector comprising a reporter gene wherein the expression product of the reporter gene is capable of producing an immunologically detectable signal.

4. A YAC vector comprising an IRES and a reporter gene wherein the expression product of the reporter gene is capable of producing a visually detectable signal.

5. A YAC vector comprising an IRES and a reporter gene wherein the expression product of the reporter gene is capable of producing an immunologically detectable signal.

6. pYIV1.

7. pYIV2.

8. pYIV3.

9. pYIV4.

10. A YAC prepared by the vector (such as the insertion thereof) according to any one of the above-mentioned embodiments.

11. Use of an IRES to modify a YAC vector or a YAC.

According to a second aspect of the present invention there is provided one or more of the following embodiments, which for ease have been presented as numbered paragraphs:

1. A YAC vector comprising a nucleotide sequence wherein the nucleotide sequence comprises the sequence presented as SEQ ID No.1 or SEQ ID No. 4 a variant, homologue or derivative thereof.

2. A vector capable of modifying a YAC or a YAC vector wherein the vector comprises a nucleotide sequence wherein the nucleotide sequence comprises the sequence presented as SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof.

3. pYAM4.

4. A YAC prepared by the vector according to any one of the above-mentioned embodiments.

5. Use of a nucleotide sequence comprising the sequence presented SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof in a vector to prepare a YAC vector or a YAC.

6. Use of a nucleotide sequence comprising the sequence presented SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof in a vector to increase the expression efficiency of one or more NOIs within a YAC vector or a YAC.

For convenience, the vector of the first aspect of the present invention is sometimes referred to herein as being an insertion vector; whereas the vector of the second aspect of the present invention is sometimes referred to herein as being an amplification vector.

However, the term "vector" as used herein in the general sense means the vector of the first aspect of the present invention and/or the vector of the second aspect of the present invention.

According to a third aspect of the present invention there is provided the combination of at least any one of the embodiments of the first aspect of the present invention and at least any one of the embodiments of the second aspect of the present invention.

With the third aspect of the present invention - namely the combination of the vector of the first aspect of the present invention and the vector of the second aspect of the present invention - the term "combination" means that the resultant YAC or transgenic/transformed cell, organ or organism is prepared by use of both the vector of the first aspect of the present invention and the vector of the second aspect of the present invention.

The third aspect of the present invention is not limited to preparative techniques wherein both the vector of the first aspect of the present invention and the vector of the second aspect of the present invention have to be used at the same time when preparing the YAC, let alone the transformed/transgenic cell, organ or organism. In this regard, it is sometimes advantageous for the vector of the first aspect of the present invention to be used at a different stage than the vector of the second aspect of the present invention during the preparation of the YAC, or even the transformed/transgenic cell, organ or organism.

According to a further aspect of the present invention there is provided a YAC vector or a YAC comprising a selection gene, wherein that selection gene is specifically removable from the YAC vector or the YAC.

Other aspects of the present invention include:

A YAC transgenic mammal co-expressing an NOI and a reporter gene wherein the expression pattern of the NOI can be determined by measuring a detectable signal

(such as a visually or an immunologically detectable signal) produced by the expression product of the reporter gene.

A YAC transgenic mammal expressing a reporter gene under the control of a regulatory sequence from a human NOI.

Use of a YAC transgenic mammal to test for potential pharmaceutical and/or veterinary agents.

An assay method for identifying an agent that can affect the expression pattern of an NOI or the EP ("expression product") activity thereof,

the assay method comprising

administering an agent to a YAC transgenic mammal according to the present invention;

determining whether the agent modulates (such as affects the expression pattern or activity) the NOI or the EP by means of the detectable signal.

An assay method according to the present invention wherein the assay is to screen for agents useful in the treatment of disturbances in any one of: circadian function, sleep disorders, eating disorders, pre-menstural syndrome, autoimmune disorders, birth defects in women and/or sexual dysfunction.

An agent identified by the method according to the present invention.

A process comprising the steps of:

(a) performing the assay according to the present invention;

(b) identifying one or more agents that affect the expression pattern of the NOI or the EP activity thereof;

(c) preparing a quantity of those one or more identified agents.

A process comprising the steps of:

(a) performing the assay according to the present invention;

(b) identifying one or more agents that affect the expression pattern of the NOI or the EP activity thereof;

(c) preparing a pharmaceutical composition comprising one or more identified agents.

A process comprising the steps of:

(a) performing the assay according to the present invention;

(b) identifying one or more agents that affect the expression pattern of the NOI or the EP activity thereof;

(c) modifiying one or more identified agents to cause a different effect on the expression pattern of the NOI or the EP activity thereof.

Use of a YAC according to the present invention or any one of the vectors according to the present invention to screen for agents capable of affecting the expression pattern of an NOI or the EP activity thereof in a transgenic mammal.

Use of an agent in the preparation of a pharmaceutical composition for the treatment of a disorder or condition associated with the expression pattern of an NOI or the EP activity thereof, the agent having an effect on the expression pattern of the NOI

or the EP activity thereof when assayed *in vitro* by the assay according to the present invention.

Use of an agent identified by an assay according to the present invention in the manufacture of a medicament which affects the expression pattern of an NOI or the EP activity thereof.

Use of an agent identified by an assay according to the present invention in the manufacture of a medicament which affects the expression pattern of an NOI or the EP activity thereof.

In accordance with the present invention, at least part of the assay can be carried out in living tissue.

A preferred - but non-limiting - example of an NOI is the human serotonin transporter (SERT), preferably the the human serotonin transporter (SERT) presented as SEQ ID No. 2 or a variant, homologue or derivative thereof.

The terms "variant", "homologue", or “derivative” in relation to this aspect of the present invention include any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) nucleic acid from or to the sequence providing the resultant expression product of the nucleotide sequence has the same activity as the expression product of SEQ ID No. 2, preferably having at least the same level of activity of the expression product of SEQ I.D. No. 2. In particular, the term "homologue" covers identity with respect to structure and/or function providing the resultant nucleotide sequence has promoter activity. With respect to sequence identity (i.e. similarity), preferably there is at least 75%, more preferably at least 85%, more preferably at least 90% sequence identity. More preferably there is at least 95%, more preferably at least 98%, sequence identity. These terms also encompass allelic variations of the sequences.

Another preferred - but non-limiting - example of an NOI is the VIP_2 receptor (VIPR2). The VIP2 receptor is referred to interchangeably throughout the text as the VIP2 receptor, VIPR2 or VPAC2R (Harmar *et al* 1998 Pharmacological Reviews 50: 265-270). Preferably, the VIPR2 is the VIP_2 receptor (VIPR2) presented as SEQ ID No. 3 or a variant, homologue or derivative thereof.

The terms "variant", "homologue", or "derivative" in relation to this aspect of the present invention include any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) nucleic acid from or to the sequence providing the resultant expression product of the nucleotide sequence has the same activity as the expression product of SEQ ID No. 3, preferably having at least the same level of activity of the expression product of SEQ I.D. No. 3. In particular, the term "homologue" covers identity with respect to structure and/or function providing the resultant nucleotide sequence has promoter activity. With respect to sequence identity (i.e. similarity), preferably there is at least 75%, more preferably at least 85%, more preferably at least 90% sequence identity. More preferably there is at least 95%, more preferably at least 98%, sequence identity. These terms also encompass allelic variations of the sequences.

The present invention also encompasses modified YACs comprising these aspects of the present invention, transformed cells comprising these aspects of the present invention, transgenic organisms comprising these aspects of the present invention, processes for making all of these aspects, and methods of expressing all of these aspects.

The term "affects" includes any one or more of: treats, prevents, suppresses, alleviates, restores, modulates, influences or to otherwise alter an existing state.

The term "agent" includes any entity (such as one or more chemical compounds, including peptide sequunces and variants/homologues/derivatives/fragments thereof) which is capable of affecting the expression pattern of the NOI or the EP activity thereof. It also includes mimics and equivalents and mutants thereof. It also includes agonists and antagonists and antibodies. Non-limiting antibodies include: polyclonal,

monoclonal, chimeric, single chain, Fab fragments, fragments produced by a Fab expression library and humanised monoclonal antibodies.

The term "expression product" or "EP" means the expressed protein *per se* but also includes fusion proteins comprising all or part of same. The EP may be the same as the naturally occuring form or is a variant, homologue, fragment or derivative thereof.

The term "disturbances in circadian function" means disorders which may lead to impaird physical and mental well-being that can occur through extremes in work patterns such as shift work, in normal ageing, when travelling through time zones (jet lag), and in dementia.

There are a number of advantages associated with the present invention.

For example, with the use of the amplification vector it is possible to obtain high copy numbers of YACs.

For example, with the use of the insertion vector it is possible to readily monitor the *in vivo* expression pattern of a NOI.

Also, with the use of the insertion vector, it is possible to express - such as over-express - a NOI and a reporter gene contained within or as a YAC.

It provides a means for producing YAC transgenic animals and the analysis of these animals in terms of expression, regulation and function of NOIs present in YAC DNA.

It facilitates the determination of sites/regions where an NOI is expressed and the identification of agents which may affect the expression pattern of the NOI or the EP activity thereof.

Other advantages associated with the present invention will be apparent from the following text.

With the present invention, the vector may comprise at least one NOI.

With the present invention, the term NOI (i.e. nucleotide sequence of interest) includes any suitable nucleotide sequence, which need not necessarily be a complete naturally occuring DNA sequence. Thus, the DNA sequence can be, for example, a synthetic DNA sequence, a recombinant DNA sequence (i.e. prepared by use of recombinant DNA techniques), a cDNA sequence or a partial genomic DNA sequence, including combinations thereof. The DNA sequence need not be a coding region. If it is a coding region, it need not be an entire coding region. In addition, the DNA sequence can be in a sense orientation or in an anti-sense orientation. Preferably, it is in a sense orientation. Preferably, the DNA is or comprises cDNA.

The vectors of the present invention may be used to prepare a modified YAC or a modified YAC vector. The modified YAC or the modified YAC vector can be used, for example, for expression and/or regulation and/or functional studies of the NOI. In addition, the vectors of the present invention can be used to prepare modified YACs or modified YAC vectors that can be used for expression and/or regulation and/or functional studies of the NOI in combination with other entities, such as other NOIs, compounds or compositions. In addition, the modified YAC or modified YAC vector can be used to test potential pharmaceutical agents (including veterinary agents).

Here the term "modified YAC or modified YAC vector" means a modified YAC or modified YAC vector having a modified genetic structure. With the present invention, the YAC or YAC vector has a modified genetic structure since part or all of the vector according to the present invention has been incorporated into the YAC or YAC vector.

The vectors of the present invention may be used to prepare transformed cells that can be used, for example, for functional studies of the NOI. In addition, the vectors of the present invention can be used to prepare transformed cells that can be used for functional studies of the NOI in combination with other entities, such as other NOIs, compounds or compositions. In addition, the transformed cells can be used to test potential pharmaceutical agents (including veterinary agents).

The vectors of the present invention may be used to prepare transformed cells that comprise mutated genes - such as by use of the pop-in/pop-out technique (mentioned above).

Here the term "transformed cell" means a cell having a modified genetic structure. With the present invention, the cell has a modified genetic structure since a vector according to the present invention has been introduced into the cell.

The term "cell" includes any suitable organism. In a preferred embodiment, the cell is a mammalian cell. In a highly preferred embodiment, the cell is a murine cell.

The cell can be an isolated cell or a collection of cells. The cell or cells may even be part of a tissue or organ or an organism (including an animal).

The cell can be transformed *in vivo* or *in vitro*, or combinations thereof.

Typically, the cell will be transformed by any one of the following methods: transfection, microinjection, electroporation or microprojectile bombardment, including combinations thereof.

Preferably, the cell will be transformed by, or by at least, transfection.

For some applications, the transformed cells may be prepared by use of the modified YAC according to the present invention.

The vectors of the present invention can also be used to prepare transgenic organisms that can be used for functional studies of the NOI. In addition, the vectors of the present invention can be used to prepare transgenic organisms that can be used, for example, for functional studies of the NOI in combination with other entities, such as other NOIs, compounds or compositions. In addition, the transgenic organisms can be used to test potential pharmaceutical agents (including veterinary agents).

Here the term "transgenic organism" means an organism comprising a modified genetic structure. With the present invention, the organism has a modified genetic structure since a vector according to the present invention has been introduced into the organism.

The term "organism" includes any suitable organism. In a preferred embodiment, the organism is a mammal. In a highly preferred embodiment, the organism is a mouse.

For some applications, the transgenic organisms may be prepared by use of the transformed cells of the present invention.

In a preferred embodiment, the insertion vector of the present invention is itself a YAC vector.

For some applications, it may be advantageous that the amplification vector of the present invention is itself a YAC vector.

With the present invention, the vector of the present invention may additionally comprise one or more selection genes to enable the vector and any resultant entity comprising the same or made from the same (such as a modified YAC vector, a YAC or a specific yeast strain comprising any one of the same) to be selectively grown and/or screened. These selection genes can be chosen from suitable selection genes that are available. Examples of suitable selection genes include *LYS2* (see Barnes and Thorner 1986, Mol and Cell Biol 6: pp 2828-2838), *LEU2* (see Beach and Nurse 1981 Nature vol 290 pp 140-142), and *ADE2* (see Stotz and Linder 1990 Gene 95 pp 91-98).

In a preferred aspect any one or more of the selection gene is specifically removable from the vector, the modified YAC vector and the modified YAC according to the present invention. In this regard, the term "specifically removable" means being able to remove the one or more selection gene without disrupting any other region in the vector, the modified YAC vector and the modified YAC according to the present invention. For example, the selection gene may be flanked by unique restriction sites. Alternatively, the selection gene may be flanked by a *LoxP* element which is removable by use of *Cre*

recombinase. Teachings on the use of the *LoxP* element and *Cre* recombinase have been published by Deursen *et al* (1995 PNAS Vol 93 pages 7376-7380), Kuhn *et al* (1995 Science Vol 269 pages 1427-1429) and Araki *et al* (1995 PNAS Vol 92 pages 160-164). By way of example, the selection gene flanked by the *LoxP* element may therefore be removed prior to or after formation of the transgenic animal stem cell. Removal of the selection gene is highly desirable as it means that the transgenic organism is not expressing the selection gene and so there can be no affect of that gene on the organism or even on the expression of the NOI being studied. In addition, removal of the selection gene means that the NOI is nearer to any 3' regulatory regions that may be present on the YAC.

With the present invention, the YAC vector may additionally comprise one or more NOIs. The NOI need not be of known function and/or structure. Preferably, the NOI is of human origin.

In accordance with one aspect of the present invention, the YAC vector comprises an internal ribosomal entry site (i.e. an IRES).

A review on IRES is presented by Mountford and Smith (TIG May 1995 vol 11 No. 5 pages 179 - 184). A suitable IRES has also been disclosed by Mountford *et al* (Mountford *et al* 1994 PNAS 91 pages 4303-4307).

IRES sequences are also mentioned in WO-A-93/03143, WO-A-97/14809, WO-A-94/24301, WO-A-95/32298, and WO-A-96/27676. These references do not disclose or suggest the use of an IRES unit in preparing or being a part of a YAC.

According to WO-A-97/14809, IRES sequences act on improving translation efficiency of RNAs in contrast to a promoter's effect on the transcription of DNAs. A number of different IRES sequences are known including those from encephalomyocarditis virus (EMCV) (Ghattas, I.R., *et al.*, *Mol. Cell.* Biol., 11:5848-5859 (1991); BiP protein [Macejak and Sarnow, Nature 353:91 (1991)]; the *Antennapedia* gene of drosphilia (exons d and e) [Oh, *et al.*, *Genes & Development*, 6:1643-1653 (1992)] as well as those

in polio virus [Pelletier and Sonenberg, *Nature 334:* 320-325 (1988); see also Mountford and Smith, TIG 11, 179-184 (1985)].

According to WO-A-97/14809, IRES sequences are typically found in the 5' non-coding region of genes. In addition to those in the literature they can be found empirically by looking for genetic sequences that affect expression and then determining whether that sequence affects the DNA (i.e. acts as a promoter or enhancer) or only the RNA (acts as an IRES sequence).

Thus the present invention is not intended to be limited to a specific IRES sequence. Instead, the sequence to be used can be any sequence that is capable of acting as an IRES sequence - i.e. it is capable of improving translation efficiency of an RNA.

A preferred IRES sequence is that presented as SEQ ID No. 1 or a variant, homologue, derivative or fragment thereof.

The terms "variant", "homologue", "derivative" or "fragment" in relation to this aspect of the present invention include any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) nucleic acid from or to the sequence providing the resultant nucleotide sequence has IRES activity, preferably having at least the same activity of the IRES shown as SEQ I.D. No. 1. In particular, the term "homologue" covers identity with respect to structure and/or function providing the resultant nucleotide sequence has promoter activity. With respect to sequence identity (i.e. similarity), preferably there is at least 75%, more preferably at least 85%, more preferably at least 90% sequence identity. More preferably there is at least 95%, more preferably at least 98%, sequence identity. These terms also encompass allelic variations of the sequences.

Another preferred IRES sequence is that presented as SEQ ID No. 4 or a variant, homologue, derivative or fragment thereof.

The terms "variant", "homologue", "derivative" or "fragment" in relation to this aspect of the present invention include any substitution of, variation of, modification of, replacement

of, deletion of or addition of one (or more) nucleic acid from or to the sequence providing the resultant nucleotide sequence has IRES activity, preferably having at least the same activity of the IRES shown as SEQ I.D. No. 4. In particular, the term "homologue" covers identity with respect to structure and/or function providing the resultant nucleotide sequence has promoter activity. With respect to sequence identity (i.e. similarity), preferably there is at least 75%, more preferably at least 85%, more preferably at least 90% sequence identity. More preferably there is at least 95%, more preferably at least 98%, sequence identity. These terms also encompass allelic variations of the sequences.

Sequence identity with respect to any of SEQ ID 1-4 can be determined by a simple "eyeball" comparison (i.e. a strict comparison) of any one or more of the sequences with another sequence to see if that other sequence has, for example, at least 75% sequence identity to the sequence(s).

Relative sequence identity can also be determined by commercially available computer programs that can calculate % identity between two or more sequences using any suitable algorithm for determining identity, using for example default parameters. A typical example of such a computer program is CLUSTAL. Advantageously, the BLAST algorithm is employed, with parameters set to default values. The BLAST algorithm is described in detail at http://www.ncbi.nih.gov/BLAST/blast_help.html, which is incorporated herein by reference. The search parameters are defined as follows, can be advantageously set to the defined default parameters.

Advantageously, "substantial identity" when assessed by BLAST equates to sequences which match with an EXPECT value of at least about 7, preferably at least about 9 and most preferably 10 or more. The default threshold for EXPECT in BLAST searching is usually 10.

BLAST (Basic Local Alignment Search Tool) is the heuristic search algorithm employed by the programs blastp, blastn, blastx, tblastn, and tblastx; these programs ascribe significance to their findings using the statistical methods of Karlin and Altschul (see http://www.ncbi.nih.gov/BLAST/blast_help.html) with a few enhancements. The

BLAST programs were tailored for sequence similarity searching, for example to identify homologues to a query sequence. For a discussion of basic issues in similarity searching of sequence databases, see Altschul *et al* (1994) Nature Genetics 6:119-129.

The five BLAST programs available at http://www.ncbi.nlm.nih.gov perform the following tasks:

blastp - compares an amino acid query sequence against a protein sequence database.

blastn - compares a nucleotide query sequence against a nucleotide sequence database.

blastx - compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.

tblastn - compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands).

tblastx - compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

BLAST uses the following search parameters:

HISTOGRAM - Display a histogram of scores for each search; default is yes. (See parameter H in the BLAST Manual).

DESCRIPTIONS - Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 100 descriptions. (See parameter V in the manual page).

EXPECT - The statistical significance threshold for reporting matches against database sequences; the default value is 10, such that 10 matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the

statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Fractional values are acceptable. (See parameter E in the BLAST Manual).

CUTOFF - Cutoff score for reporting high-scoring segment pairs. The default value is calculated from the EXPECT value (see above). HSPs are reported for a database sequence only if the statistical significance ascribed to them is at least as high as would be ascribed to a lone HSP having a score equal to the CUTOFF value. Higher CUTOFF values are more stringent, leading to fewer chance matches being reported. (See parameter S in the BLAST Manual). Typically, significance thresholds can be more intuitively managed using EXPECT.

ALIGNMENTS - Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 50. If more database sequences than this happen to satisfy the statistical significance threshold for reporting (see EXPECT and CUTOFF below), only the matches ascribed the greatest statistical significance are reported. (See parameter B in the BLAST Manual).

MATRIX - Specify an alternate scoring matrix for BLASTP, BLASTX, TBLASTN and TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff, 1992). The valid alternative choices include: PAM40, PAM120, PAM250 and IDENTITY. No alternate scoring matrices are available for BLASTN; specifying the MATRIX directive in BLASTN requests returns an error response.

STRAND - Restrict a TBLASTN search to just the top or bottom strand of the database sequences; or restrict a BLASTN, BLASTX or TBLASTX search to just reading frames on the top or bottom strand of the query sequence.

FILTER - Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity

internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

Low complexity sequence found by a filter program is substituted using the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNN") and the letter "X" in protein sequences (e.g., "XXXXXXXXX").

Filtering is only applied to the query sequence (or its translation products), not to database sequences. Default filtering is DUST for BLASTN, SEG for other programs.

It is not unusual for nothing at all to be masked by SEG, XNU, or both, when applied to sequences in SWISS-PROT, so filtering should not be expected to always yield an effect. Furthermore, in some cases, sequences are masked in their entirety, indicating that the statistical significance of any matches reported against the unfiltered query sequence should be suspect.

NCBI-gi - Causes NCBI gi identifiers to be shown in the output, in addition to the accession and/or locus name.

Most preferably, sequence comparisons are conducted using the simple BLAST search algorithm provided at http://www.ncbi.nlm.nih.gov/BLAST.

Other computer program methods to determine identify and similarity between the two sequences include but are not limited to the GCG program package (Devereux *et al* 1984 Nucleic Acids Research 12: 387) and FASTA (Atschul *et al* 1990 J Molec Biol 403-410).

In some aspects of the present invention, no gap penalties are used when determining sequence identity.

The present invention also encompasses nucleotide sequences that are complementary to the sequences presented herein, or any fragment or derivative thereof. If the sequence is complementary to a fragment thereof then that sequence can be used as a probe to identify similar promoter sequences in other organisms etc.

The present invention also encompasses nucleotide sequences that are capable of hybridising to the sequences presented herein, or any fragment or derivative thereof.

Hybridization means a "process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY) as well as the process of amplification as carried out in polymerase chain reaction technologies as described in Dieffenbach CW and GS Dveksler (1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY).

Also included within the scope of the present invention are nucleotide sequences that are capable of hybridizing to the nucleotide sequences presented herein under conditions of intermediate to maximal stringency. Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Maximum stringency typically occurs at about Tm-5°C (5°C below the Tm of the probe); high stringency at about 5°C to 10°C below Tm; intermediate stringency at about 10°C to 20°C below Tm; and low stringency at about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a maximum stringency hybridization can be used to identify or detect identical nucleotide sequences while an intermediate (or low) stringency hybridization can be used to identify or detect similar or related nucleotide sequences.

In a preferred aspect, the present invention covers nucleotide sequences that can hybridise to the nucleotide sequences of the present invention under stringent conditions (e.g. 65 °C and 0.1xSSC).

The present invention also encompasses nucleotide sequences that are capable of hybridising to the sequences that are complementary to the sequences presented herein, or any fragment or derivative thereof. Likewise, the present invention encompasses nucleotide sequences that are complementary to sequences that are capable of hybridising to the sequence of the present invention. These types of nucleotide sequences are examples of variant nucleotide sequences. In this respect, the term "variant" encompasses sequences that are complementary to sequences that are capable of hydridising to the nucleotide sequences presented herein. Preferably, however, the term "variant" encompasses sequences that are complementary to sequences that are capable of hydridising under stringent conditions (eg. 65°C and 0.1xSSC {1xSSC = 0.15 M NaCl, 0.015 Na_3 citrate pH 7.0}) to the nucleotide sequences presented herein.

Insertion of the IRES into a YAC by use of the insertion vector of the present invention - and thus forming a modified YAC - enables the modified YAC to express, in particular over-express, at least two nucleotide sequences. Of these two nucleotide sequences one may be a NOI, such as a NOI of human origin. The other of those nucleotide sequences may be another NOI. Alternatively, and in a preferred aspect, the other nucleotide sequence is a reporter gene according to the present invention.

The present invention also encompasses a vector or a YAC obtained therefrom comprising more than one IRES. In this embodiment, the vector or the YAC obtained therefrom preferably comprises more than NOI.

In a preferred aspect of the first aspect of the present invention, the YAC vector comprises a reporter gene whose expression product is capable of producing a visually detectable signal. Examples of such reporter genes include: *LacZ* (see Mansour *et al* 1990 PNAS vol 87 pp 7688-7692), green fluorescent protein (see Chiocchetti *et al* 1997 *Biochim Biophys Acta* 1352: pp 193-202; and Chalfie *et al* 1994 Science vol 263 pp 802-

805), chloroamphenicol acetyl transferase (see Gorman *et al* 1982 Mol Cell Biol 2(9) pp 1044-1051; and Frebourg and Brison 1988 Gene vol 65 pp 315-318), or luciferase (see de Wet *et al* 1987 Mol Cell Biol 7(2) pp 725-737; and Rodriguez *et al* 1988 PNAS vol 85 pp 1667-1671).

In another preferred embodiment of the first aspect of the present invention, the YAC vector comprises a reporter gene whose expression product is capable of producing, or being detected by an agent capable of providing, an immunologically detectable signal.

In a preferred aspect, the reporter gene when fused to the NOI leads to the production of a fusion protein that can be detected by commercially available antibodies, such as a haemagglutinin tag (see Pati 1992 Gene 15; 114(2): 285-288), a *c-myc* tag (see Emrich *et al* 1993 Biocem Biophys Res Commun 197(1): 214-220), or the FLAG epitope (Ford *et al* 1991 Protein Expr Purif Apr; 2(2):95-107).

By using a reporter gene according to the present invention it is possible to readily observe the functionality of NOIs contained within YAC libraries, such as YAC human DNA libraries.

For example, if the NOI in the YAC has an expression regulatory role (such as a promoter) then expression of the reporter gene according to the present invention by a transgenic organism according to the present invention enables workers to readily determine in or at which sites or regions that expression regulatory element is active. In addition, workers will be able to readily test agents etc. that may affect the expression ability or pattern of that regulatory element.

By way of further example, if the NOI in the YAC has a functional role other than an expression regulatory role then by use of the insertion vector according to the invention workers can fuse (either directly or indirectly such as by means of one or more spacing nucleotide sequences) the NOI to the reporter gene according to the present invention. Thus, if the NOI is fused to the reporter gene according to the present invention and is present in a transgenic organism according to the present invention, then workers can

readily determine which sites or regions that NOI is expressed. In addition, workers will be able to readily test agents etc. that may affect the expression pattern of that NOI. A further advantage is that by being able to readily monitor the expression pattern or level of the NOI enables workers to determine the phenotype.

One aspect of the present invention concerns the use of SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof. Here, the term "variant, homologue or derivative thereof" includes any addition, substitution or deletion of one or more nucleic acids providing the resultant entity can still function as an IRES.

For a preferred aspect of the present invention it is envisaged that any variant, homologue or derivative of the IRES sequence comprises at least 100 bp of SEQ ID No.1 or SEQ ID No. 4. Preferably, any variant, homologue or derivative comprises at least 200 bp of SEQ ID No.1 or SEQ ID No. 4. Preferably, any variant, homologue or derivative comprises at least 300 bp SEQ ID No.1 or SEQ ID No. 4. Preferably, any variant, homologue or derivative comprises at least 400 bp of SEQ ID No.1 or SEQ ID No. 4. Preferably, any variant, homologue or derivative comprises at least 500 bp of SEQ ID No.1 or SEQ ID No. 4. Preferably there is at least 80% sequence identity, preferably at least 85% sequence identity, preferably at least 90% sequence identity, preferably at least 95% sequence identity, more preferably there is at least 98% sequence identity with the sequence shown as SEQ ID No.1 or SEQ ID No. 4.

Preferably, the nucleotide sequence is the sequence presented as SEQ ID No.1 or SEQ ID No. 4.

As indicated, the YAC vector of the present invention comprises the nucleotide sequence presented as SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof. Here the nucleotide sequence increases the expression efficiency of one or more NOIs within a YAC vector or a YAC.

The YAC vector may additionally comprise one or more marker genes. These genes can be chosen from suitable marker genes that are available. An example of a suitable marker gene is *PGK-Hyg* (see Nara *et al* 1993 Curr Genet 23(2): pp 134-140).

The nucleotide sequence of the present invention can be used to modify a YAC or a YAC vector, such as pYAC1, pYAC2, pYAC3 or pYAC4 etc.

The present invention also encompasses combinations of the above-mentioned aspects.

The following samples were deposited by the MRC Brain Metabolism Unit, Royal Edinburgh Hospital, Morningside Park, Edinburgh, EH10 5HF in accordance with the Budapest Treaty at the recognised depositary The National Collections of Industrial and Marine Bacteria Limited (NCIMB) at 23 St. Machar Drive, Aberdeen, Scotland, United Kingdom, AB2 1RY on 24 November 1997

JM109 pYIV1 - deposit number NCIMB 40907
JM109 pYIV2 - deposit number NCIMB 40908
JM109 pYIV3 - deposit number NCIMB 40909
JM109 pYIV4 - deposit number NCIMB 40910
JM109 pYAM4 - deposit number NCIMB 40906

The present invention will now be described only by way of example, in which reference shall be made to the following Figures:

Figure 1 which is a diagrammatic representation of pYIV1;

Figure 2 which is a diagrammatic representation of pYIV2;

Figure 3 which is a diagrammatic representation of pYIV3;

Figure 4 which is a diagrammatic representation of pYIV4;

Figure 5 which is a photographic image;

Figure 6 which is a diagrammatic representation of pYAM4;

Figure 7 which is a picture of a gel;

Figure 8 which is a photographic image;

Figure 9 which is a photographic image;

Figure 10 which presents a nucleotide sequence;

Figure 11 which is a picture of a gel;

Figure 12 which is a PCR map of the integrated SERT 35D8/D6 YAC DNA;

Figure 13 which is a PCR map of the integrated VIPR2 HSC7E526/V12 YAC DNA;

Figure 14 which is a photographic image;

Figure 15 which is a photographic image;

Figure 16 which is a graphical representation of β-Gal enzyme activity determined using a chemiluminescent reporter assay system; and

Figure 17 which is a schematic diagram.

In slightly more detail:

Figure 7 shows amplication of YAC DNA by pYAM4. The endogenous chromosomal DNA from *S. cerevisiae* is shown in lane 3 and sizes in kb at the left. All other lanes were loaded with DNA plugs form Lys+ YAC clones and cultured in medium with galactose but lacking lysine after retrofitting with pYAM4. The migration position of YAC DNA in each clone is indicated with an arrow. Levels of amplification of the YAC DNA are based on comparison of ethidium

bromide staining of YAC DNA to that of endogeous chromosomes of similar size. Lane 1, 350 kb, 8-fold, lane 2, 630 kb, 3-fold, lane 4, 615 kb, 3-fold, lane 5, 500 kb, 4-fold, lane 6, 200 kb, 5-fold, lane 7, 200 kb, 6-fold; lane 8, 230 kb, 2-fold; lane 9, 150 kb, 3-fold, lane 10/11/12, 230 kb, 5-fold.

Figure 10 shows the IRES sequence which is derived from Encephalomycocarditis virus. The sequence has been genetically modified at the 3' end to introduce a *HindIII* restriction site.

Figure 14 shows the immunohistochemical staining of the *LacZ* reporter gene in the suprachiasmatic nuclei of transgenic mice expressing a YAC containing the human VPAC2R gene. The single cell resolution obtainable with the immunohistochemical approach is worthy of note.

Figure 15 illustrates the histochemical staining of β-galactosidase activity in transgenic mice containing the YAC HSC7E526/V12.

Figures 15a and 15b show staining patterns for β-galactosidase activity in a coronal slice from the brain of a transgenic mouse for whom mouse A108.2 was the father. In (a) the stained suprachiasmatic nuclei are indicated with arrowheads, in (b) an enlarged view is shown. Figure 15c shows staining in the pancreas from the same transgenic mouse (tg) and in a wild type (wt) littermate.

Figure 16 illustrates the tissue distribution of β-galactosidase activity in transgenic mice containing the YAC HSC7E526/V12.

β-galactosidase (*Lac*Z) activity was determined in tissue extracts from control mice and two independent lines expressing the hVPAC$_2$R-HA-lacZ transgene. Enzyme activity was determined using a chemiluminescent reporter assay system (Galacto-Light Plus, Tropix). ND indicates tissues in which no β-galactosidase activity was detected.

For ease of reference, some parts of the following text has been split according to the vector of the first aspect of the present invention or according to the vector of the second aspect of the present invention.

MATERIALS AND METHODS

INSERTION VECTOR

Construction of YAC Insertion vectors

For genetic manipulation of YAC DNA, we have generated a series of modification cassettes which can be inserted into any YAC DNA:

pYIV1

A cassette containing a *lacZ* reporter gene flanked by IRES and the *LEU2* selective marker. The vector can be used for YACs which have been introduced into a Leu$^-$ yeast strain. The details of the construction procedure were as follows:

A 3.5 kb cassette containing the *lacZ* gene and polyadenylation sequences was isolated from pMRß-*lacZ*-PA (23) by *Sal*I (complete) and *EcoR*I (partial) digestion and inserted into the *EcoR*I-*Sal*I sites of pBluescript SK$^-$, generating pSK$^-$*lacZ*-PA. The IRES was introduced into pSK$^-$*lacZ*-PA by replacing the 1.1 kb *Xba*I (in the polylinker)-*EcoR*V (in the *lacZ* sequence) fragment with a 1.7 kb *Xba*I-*EcoR*V fragment (IRES-5'-*lacZ*) from pIRES-b*geo* (24), resulting in pIRES-*lacZ*-PA. A 2.2 kb *Xho*I-*Sal*I fragment containing the *LEU2* gene was isolated from pDB248 (25) and inserted into the compatible SalI site in pIRES-*lacZ*-PA in the same orientation as the *lacZ* gene, resulting a plasmid (pYIV1) containing the IRES-*lacZ*-PA-L*EU2* cassette in pBluescript SK$^-$ (Fig. 1). Five unique restriction sites flank the cassette (*Sac*II, *Not*I and *Xba*I between the T3 primer and the IRES, *Sal*I and *Xho*I sites between *LEU2* gene and the T7 primer) which allow insertion of genomic DNA on both sides of the cassette.

pYIV2

pYIV2 is similar to pYIV1 except that the *LEU2* gene is replaced by the *ADE2* gene so that the plasmid can be directly introduced into the conventional yeast strain (AB1380) used for the construction of most YAC libraries. A 2.5 kb *Bgl*II fragment containing the

ADE2 gene was isolated from pASZ11, filled in and inserted into the filled in *Hind*III site of pBluescript SK$^-$ in an orientation such that the T7 promoter is adjacent to the 5' end of the *ADE2* gene (pSK$^-$ADE2). The IRES-*lacZ*-PA cassette was released from pIRES-*lacZ*-PA by *Sal*I digestion, filled in and then cut with *Not*I. The fragment was inserted into the *EcoR*I (filled in) and *Not*I sites of pSK$^-$ADE2, resulting a plasmid (pYIV2) containing and IRES-*lacZ-ADE2* cassette (Fig.4). Four unique sites (*Sac*II, *Not*I, *Sal*I and *Xho*I) are available for cloning.

pYIV3

For many genes, there is no antibody available to detect the encoded protein. We introduced a haemagglutinin (HA) epitope tag into pYIV2 so that a commercially available antibody can be used to localise the distribution of the protein product of a YAC transgene within cells. The HA tag was introduced as follows:

The translation stop codon of the human VIP_2 receptor cloned in the vector pcDNA3 was converted into an *Xho*I site by PCR-based mutagenesis. A linker encoding the HA epitope tag flanked by *Xho*I-*Xba*I sites

```
5'-TC GAG TAC CCA TAC GAT GTT CCA GAT TAC GCC TCC CTC TAG-3'

    3'-ATG GGT ATG CTA CAA GGT CTA ATG CGG AGG GAG ATC AGA TC-5'
```

was cloned into the *Xho*I-*Xba*I sites at the end of the VIP_2 receptor. The VIP_2 receptor-HA fragment was released from the pcDNA3 vector by *BamH*I and *Xba*I (filled-in) digestion and cloned into pBluescript SK$^-$ (in which the *Xho*I site was removed by filling in) at the *BamH*I-*EcoR*V sites generating pSK$^-$VIP2R-HA.

The IRES-*lacZ-ADE2* cassette was isolated from pYIV2 with *Not*I (filled in) and *Sal*I restriction enzymes and inserted into the *Hind*III (filled in) and *Sal*I sites of pSK$^-$VIP2R-HA, yielding a plasmid containing VIP2R-HA-IRES-*lacZ-ADE2*. The HA-IRES-*lacZ*-

ADE2 cassette was isolated by *Xho*I-*Sal*I digestion and inserted into pGEM11Z at the *Xho*I-*Sal*I sites, resulting in pYIV3 (Fig.3). In pYIV3, the HA-IRES-*lacZ-ADE2* cassette is flanked by *Not*I and *Xho*I restriction sites at the 5' side and by *Sal*I and *Sfi*I sites at the 3' side, facilitating the cloning of genomic fragments of interest for YAC manipulation.

pYIV4

pYIV4 is similar to pYIV3 except that the orientation of the *ADE2* gene is opposite to that in pYIV3 and two *lox*P elements, in same orientation, were introduced, one at the *Bgl*II site between the *lacZ* and PA, and another following the *ADE2* gene.

The *loxP* sequence from pBG was cloned into pBluescript SK$^-$ at *EcoR*I and *Sal*I sites (pSK$^-$*loxP*). The *ADE2* gene was excised from pSK$^-$*ADE2* by *EcoR*I and *Cla*I (filled in) digestion and cloned into pSK$^-$*loxP* at the *EcoR*I and *Sma*I sites (pSK$^-$loxP-*ADE2*). The *EcoR*I-*Sal*I *loxP* fragment from pSK$^-$*loxP* was isolated, blunt ended with Klenow and inserted into the *Bgl*II (filled in) site between the *lacZ* and poly A sequences of pSK$^-$IRES-*lacZ*-PA generating pSK$^-$IRES-*lacZ-loxP*-PA. The IRES-*lacZ-loxP*-PA cassette was isolated by *Not*I and *Sal*I (filled in) and cloned into pSK$^-$*loxP-ADE2* at the *Not*I and *BamHI* (filled in) sites, yielding a construct pYIV4 (Fig.4). The pYIV4 vector permits deletion of the SV40 PA and ADE2 gene in the YAC transgenic animals using *Cre* recombinase so that a transgene and *lacZ* reporter gene can be followed by its own 3'-untranslated region.

Yeast DNA preparation

Yeast DNA was isolated with the combined methods of Schedl *et al.* (26) and Bellis *et al.* (27). Clones were inoculated into 15 ml of medium (Ura$^-$/Lys$^-$) with 2% of galactose instead of glucose as the carbon source. When cells had grown to late log phase after 2-4 days, plugs were taken and subjected to novozyme digestion for 4-6 hours as described

by Schedl *et al.* (28). Then, plugs were washed in 50 mM EDTA (2 x 30 min) and digested with proteinase K (2mg/ml) in a buffer containing 0.5 M NaCl, 0.125 M Tris pH 8.0, 0.25 M Na2EDTA, 1% Lithium sulphate, and 0.5 M β-mercaptoethanol at 55°C overnight. Plugs then were washed with TE and stored at 4°C in 0.5 M EDTA.

Pulsed-field gel electrophoresis

DNA plugs were washed in TE (3 x 30min), loaded on a 1% agarose gel and sealed with 1% agarose in 0.5 x TBE buffer. Gels were run in 0.5 x TBE buffer at 6V/cm for 24 hours at 14°C with 60 sec. switch time. After running, gels were stained with ethidium bromide and photographed.

AMPLIFICATION VECTOR

Construction of pYAM4

pYAM4 was constructed using pYAC4, pBluescript SK$^-$ and pBG. pBG is a modification of pCGS990 in which the *Sal*I site has been converted to a *Not*I site and a PGK-Hyg-*loxP* cassette has been introduced between the LYS2 and TK genes in pCGS990.

pBG was constructed as follows: The unique *Sal*I site in pCGS990 was converted into *Not*I with the *Xho*I-*Not*I-*Xho*I linker

```
5'-TCGAGCGGCCGC-3'
       ||||||||
   3'-CGCCGGCGAGCT-5'
```

resulting in pCGS990N. A 2044 bp *Pst*I fragment containing the chloramphenicol resistance gene (cm) flanked by two *loxP* sites was excised from pUC9*lox*2cm, blunt ended with T4 DNA polymerase and ligated to the filled-in *Sph*I site of pHA58 which

contains a Hygromycin B resistance gene flanked by the mouse *Pgk1* (phosphoglycerate kinase 1) promoter and polyadenylation signal. The resulting plasmid (pHA58*lox*2cm.1) was digested with *Bgl*II, a 3.5 kb *Bgl*II fragment harbouring a *loxP*-cm-*loxP*-Hyg cassette was isolated, filled in with Klenow and inserted into the *EcoR*I site (filled in) between the *TK* and *LYS2* genes in pCGS990N, obtaining pCGS990N-Hyg*lox*2cm. The chloramphenicol resistance gene was removed from pCGS990N-Hyg*lox*2cm with purified *Cre* recombinase. DNA was purified, transformed into *E. Coli* XL-1 Blue and chloramphenicol sensitive colonies were selected, leading to the production of pBG.

A 572 bp (*Sma*I-*Cla*I) fragment between the cloning site and CEN4 in pYAC4 was cloned into *Sma*I-*Cla*I sites of pBluescript SK$^-$ vector. The fragment was excised with *Cla*I and *Not*I and inserted into the *Cla*I-*Not*I sites of pBG, resulting in pYAM3.

A 705 bp *Xho*I-*BamH*I fragment of the telomere (*TEL*) from pYAC4 was blunt ended with Klenow and inserted into the filled in *Sac*I-*Sac*II sites of pBluescript SK$^-$ vector. The orientation of the *TEL* in the resulting plasmid (pSK$^-$*TEL*) was confirmed by sequencing with T7 and reverse primers. pSK$^-$*TEL* was digested with *Not*I-*Sal*I restriction enzymes and replaced the corresponding region (pBR322-*TEL*-*TK*-LoxP) in pYAM3, leading to the generation of pYAM4 (Fig. 6).

Transformation

pYAM4 was linearised with *Not*I. Yeast were inoculated into 10 ml SD medium lacking uracil and tryptophan. When yeast had grown to 2 x 10^7 cells/ml, they were harvested and washed with 5 ml of LTE (0.1M LiOAc, 10mM Tris pH7.5, and 1 mMNa_2EDTA). After resuspension in 100 μl of LTE, cells were incubated at 30°C for one hour with regular inversion. One μg of linearised pYAM4 and 5 μl of carrier DNA (salmon sperm DNA,10 mg/ml) were added to the cells and mixed. After incubation at 30°C for 30 minutes, 0.7 ml of PEG/LTE (40% PEG 3,300 in LTE) was mixed with the cells and incubated at 30°C for 30 minutes. Then cells were heat-shocked at 42°C for 5 minutes. Cells were spun down for 2 min at 500 x g and washed twice with TE. After

resuspension in 400 μl TE, cells were plated out on SD medium lacking uracil and lysine. Plates were kept at 30°C for 3-5 days. Single colonies were picked and plated onto dishes lacking lysine or tryptophan. Clones which could not grow on the medium lacking tryptophan were selected for further analysis.

Yeast DNA preparation

Yeast DNA was isolated with the combined methods of Schedl *et al.* (26) and Bellis *et al.* (27). Clones were inoculated into 15 ml of medium (Ura^-/Lys^-) with 2% of galactose instead of glucose as the carbon source. When cells had grown to late log phase after 2-4 days, plugs were taken and subjected to novozyme digestion for 4-6 hours as described by Schedl *et al.* (28). Then, plugs were washed in 50 mM EDTA (2 x 30 min) and digested with proteinase K (2mg/ml) in a buffer containing 0.5 M NaCl, 0.125 M Tris pH 8.0, 0.25 M Na_2EDTA, 1% Lithium sulphate, and 0.5 M β-mercaptoethanol at 55°C overnight. Plugs then were washed with TE and stored at 4°C in 0.5 M EDTA.

Pulsed-field gel electrophoresis

DNA plugs were washed in TE (3 x 30min), loaded on a 1% agarose gel and sealed with 1% agarose in 0.5 x TBE buffer. Gels were run in 0.5 x TBE buffer at 6V/cm for 24 hours at 14°C with 60 sec. switch time. After running, gels were stained with ethidium bromide and photographed.

COMBINATION OF INSERTION VECTOR AND AMPLIFICATION VECTOR

Combination of the pYIV3 and pYAM4 for YAC transgenic study

The vector pYIV3 was used to introduce the haemagglutinin (HA) tag and *lacZ* reporter gene into two YAC clones,35D8 (500kb) and HSC7E526 (630kb), which contain the human serotonin transporter (SERT) and VIP2 receptor (VIPR2) genes respectively. In order to integrate the *lacZ* reporter gene into each of the YAC clones by homologous

recombination, genomic DNA sequences (at least a few hundred bp) flanking the stop codon of each gene were introduced on either side of the HA-IRES-*lacZ*-Ade2 sequences of the pYIV3 vector.

Construction of the pLacZVIPR2$^+$ vector

The *Xho*I site in the polylinker of pBluescript SK$^-$ was removed by digesting the vector with *Xho*I and filling in the recessed 3' termini with Klenow fragment of *E. coli* DNA polymerase I, generating pSKX. A *BamH*I-*Xba*I fragment containing the human VIPR2 cDNA with the HA tag at the C-terminus of the coding sequence (see McDonald *et al* 1997 Biochem Soc 25:442S) was subcloned from the pcDNA3 vector into pSKX at the *EcoRV* site in an orientation such that the 5' end of the cDNA was adjacent to the T3 primer in the pSKX vector, generating pSK-VIPR2-HA. A *Not*I-*Pst*I fragment of pSK$^-$ VIPR2-HA containing VIPR2 cDNA sequences was then replaced with a 1.2 kb *Not*I-*Pst*I fragment of VIPR2 genomic DNA (*Pst*I cuts in the last coding exon of the human VIPR2 gene), generating pVHA. The IRES-*lacZ*-PA-Ade2 cassette was excised from pYIV2 as a *Sal*I - *Not*I (blunt ended) fragment and inserted into the *Sal*I-*Hind*III (blunt ended) sites of pVHA, resulting in pVHAIZA.

By PCR using primers

32366 (5'-CAA ACG GAG ACC TCG GTC CTC GAG CCC CAC-3')

and

32496 (5'-CGG GTA CCA AAA TGG TGG GTT GTT CTG TAA-3')

*Xho*I and *Kpn*I restriction sites were introduced at ends of a 1.6 kb fragment of genomic DNA 3' of the stop codon of the human VIPR2 gene. The fragment was subcloned into the *Xho*I and *Kpn*I sites of the pSK$^-$ vector, generating p3'VIPR2.

The *Not*I-*Sal*I fragment of pVHAIZA which contains VIPR2-HA-IRES-*lac*Z-PA-Ade2 was ligated into *Not*I-*Xho*I digested p3'VIPR2, generating a final construct, pLacZVIPR2+. For efficient homologous recombination in yeast, genomic DNA sequences at least a few hundred bp in length must flank the stop codon of the target gene either side of the HA-IRES-*lac*Z-Ade2 sequences of pYIV3. In pLacZVIPR2+ there is 1.3 kb of VIPR2 genomic sequence upstream and 1.6 kb of VIPR2 genomic sequence downstream of the HA-IRES-*lac*Z-Ade2 cassette.

Construction of the pLacZSERT+ vector

The *Xho*I site in the polylinker of pBluescriptSK- was removed by digesting the vector with *Xho*I and filling in the recessed 3' termini with Klenow fragment of *E.coli* DNA polymerase, generating pSKX. A *BamH*I-*Xba*I fragment containing the human VIPR2 cDNA with the HA tag at the C-terminus of the coding sequence was subcloned from the pcDNA3 vector into pSKX at the *EcoRV* site, in an orientation that the 5'end of the cDNA is adjacent to the T3 primer in the pSKX vector, generating pSK-VIPR2-HA. A 5 kb human genomic DNA fragment contaning intron 13 and exon 14 of the SERT gene was cloned into *Not*I-*Xho*I sites of pBluescriptSK- using PCR primers

32365 (5'ACT GCA TAG CGG CCG CAT CTT TCA TTT GCA TCC CC 3')

and

32853 (5' TGT GCT CGA GAG CAT TCA AGC GGA TGT 3')

generating pIn13. To introduce the HA tag into the C-terminus of the SERT gene product, the 5 kb SERT intron 13 sequence (*Not*I-*Xho*I fragment) was used to replace the *Not*I-*Xho*I fragment in pSK-VIPR2-HA, generating pIn13-HA. The intron 13 sequence and the HA tag were isolated as a *Sac*II - *Cla*I (blunt ended) fragment and inserted into *Sac*II and (blunt ended) *Not*I sites of pYIV2, generating pIn13-HA-IZA. The sequences downstream of the stop codon in exon 14 of the SERT gene were isolated by PCR of human genomic DNA using primers

32358 (5' CTC CTC GAG AGG AAA AAG GCT TCT 3')

and

32359 (5' TAG GTA CCC TGT TCT CTC CTA CGC AGT TT 3')

and cloned into the *Xho*I-*Kpn*I sites of pBluescriptSK$^-$ generating p3'SERT. Finally, the intron13-HA-IRES-*lacZ*-Ade2 fragment was isolated by *Not*I and *Sal*I digestion of pIn13-HA-IZA and inserted into the *Not*I and *Xho*I sites of p3'SERT, resulting in pLacZSERT+

Introduction of the Insertion Constructs into the YAC DNA

The pLacZVIPR2$^+$ and pLacZSERT$^+$ constructs were linearised with *Not*I and introduced into the YAC clones HSC7E526 and 35D8 respectively. The transformants which incorporated HA-*lacZ*-Ade2 sequences into YAC DNA by homolgous recombination were selected by growing on plates lacking uracil, tryptophan and adenine. The integration of the HA-IRES-*lacZ*-Ade2 sequence into the YAC DNA was confirmed by Southern hybridization with an Ade2 probe.

Amplification of Modified YAC DNA

YAC subclones which incorporated the HA-*lacZ*-Ade2 cassette were transformed with *Not*I linearised pYAM4. Recombinants were isolated on selective medium lacking uracil, adenine and lysine and replica plated on plates lacking uracil, adenine and tryptophan. Successful replacement of the YAC left arm (containing TRP1 gene) by pYAM4 would result in yeast capable of growth on medium lacking uracil, adenine and lysine but not on the counter selection medium lacking tryptophan.

Tryptophan sensitive clones were cultured in selective medium (Ura$^-$/Ade$^-$/Lys$^-$) with galactose as carbon source instead of glucose. In such medium, the GAL1 promoter adjacent to the CEN4 in the pYAM4 vector will be induced. Activation of transcription

from the GAL1 promoter interferes with the function of the CEN4 leading to non-segregation of the YAC DNA and a consequent increase the YAC DNA copy number per cell.

TRANSFORMED CELLS/TRANSGENIC ORGANISMS

A YAC according to the present invention may be transfered into mammalian cells by appropriately adapting the teachings of Schedl *et al* (reference 28). By the term "adapting" we mean following the teachings but using the vectors of the present invention where appropriate. It is to be understood that other techniques may be used to transfer a YAC according to the present invention in mammalian cells and these other techniques are well documented in the art (e.g. for example see WO-A-95/14769 and/or Gietz *et al* 1995 Yeast vol 11 No. 4, pp 355-360).

According to Schedl *et al* (*ibid*), possibly the most straight forward approach to generate transgenic cell lines is the transfer of YACs by sphaeroblast fusion (this is a technique disclosed in the previous chapter of Reference 28). This method, however, normally leads to integration of the entire yeast genome in addition to the YAC, which might obscure the results of some experiments. In contrast, direct microinjection of DNA into the nucleus of a recipient cell allows purification of a YAC prior to the transfer.

Some teachings of Schedl *et al* (*ibid*) are as follows:

Materials

1. SE: 1M Sorbitol, 20mM EDTA (pH8.0).

2. TENPA: 10mM Tn's-HCI (pH 7.5), ImM EDTA (pH8.0), 100mM NaCl, 30μM spermine, 70μM spermidine.

3. IB: 10mM Tris-HCI (pH7.5), 0.1mM EDTA (pH8.0), 100mM NaCl, 30μM spermine, 70μM spermidine.

4. LiDS: 1% lithium-dodecyl-sulphate, 100mM EDTA (pH8.0).

5. Zymolyase, ICN Biomedicals Inc., Costa Mesa, CA, USA.

6. Nusieve low melting point (LMP) agarose, FMC, Rockland, ME, USA.

7. Seaplaque low melting point (LMP) agarose, FMC, Rockland, ME, USA.

8. Automatic Injection System, Zeiss, Germany.

9. Femptotips, Eppendorf,

10. Insert moulds (plug formers), Pharmacia, Uppsala, Sweden.

11. CHEF-DR 11, pulsed-field-gel-electrophoresis (PFGE) system, BIO-RAD Laboratories, Richmond, CA, USA.

12. β-mercaptoethanol (14M stock).

{Note: Heavy metal ions present in buffers even in traces will lead to degradation of the YAC DNA during the agarase treatment. Make sure to use water of highest quality for the preparation of buffers.}

Methods

Preparation of high density agarose plugs for preparative Pulsed Field Gel Electrophoresis (PFGE)

1. Inoculate 500 ml of of SD medium (-URA) with your yeast clone and grow the culture to late log phase.

2. Prepare a solution of 1% Seaplaque LMP agarose in SE buffer containing 14mM β-mercaptoethanol and keep at 45°C until use.

3. Spin down cells at 4000 rpm for 5min (Sorvall RT6000). and resuspend the pellet in 50ml SE. Transfer the cell suspension into a 50ml Falcon tube.

4. Seal the bottom of Pharmacia plug formers (insert moulds) with strips of tape and place them on ice.

5. Wash cells twice with SE (4000 rpm, 5min.).

6. After the last washing step, discard the supernatant, and carefully remove all liquid by cleaning the inside of the tube with a paper towel. The cell pellet should be about 1 to 1.5ml.

7. Add 200 µl of SE buffer. With a cut off tip try to resuspend the pellet. The suspension will be very thick and difficult to pipet.

8. Transfer 0.5ml aliquots of the cell suspension into 2ml Eppendorf tubes and keep at 37°C.

9. Just before use dissolve 10mg Zymolyase in 2.5ml of the LMP agarose solution.

{Note: Zymolyase does not completely dissolve at this concentration. Weigh in the required amount and work with a protein suspension.}

10. Transfer 0.5ml of this solution to the yeast cell suspension and mix thoroughly the agarose with the cells by pipetting up and down using a blue cut off tip.

{Note: Only a completely homogenous mixture will yield in high quality plugs with even distribution of DNA.}

Keep the solution at 45°C at all times to avoid setting of the agarose.

11. Using a cut off yellow tip pipet 80µl aliquots of the mixture into plug formers kept on ice. Leave for 10 min to allow the agarose to set.

12. Transfer the plugs into SE buffer containing 14mM β-mercaptoethanol and lmg/ml Zymolyase. Incubate at 37°C for 4 to 6h.

13. Replace the buffer with LiDS buffer using at least 0.5ml/plug and incubate at 37°C with gentle rocking. After 1 h refresh the LiDS buffer and continue incubation overnight.

14. Wash plugs extensively in TE pH8.0 until no more bubbles can be seen. Store plugs in 0.5M EDTA at 4°C until use.

{Note: DNA plugs prepared this way can be stored without degradation for at least one year.

Isolation of intact YAC DNA for microinjection

1. Cast a gel using 0.25xTAE, 1% agarose using a comb of which you have sealed several teeth of the comb with tape to obtain a preparative lane of approximately 5cm.

{Note: To ensure straight bands it is recommended to run the preparative lane in the center of the gel. Bands of preparative lanes bigger than 5cm, may show "smiling" effects, which leads to imprecise excision of the DNA and, hence, a lower yield/ final concentration.}

If the DNA will be concentrated by a second gel run standard agarose can be used. Otherwise use an LMP agarose.

2. Wash the high density plugs for 4x15min in TE pH8 with gentle shaking on a rocking platform.

3. Load the plugs next to one another into the preparative lane.

{Note: Best results are achieved using rectangular plugs (such as produced in Pharmacia plug formers), which can be loaded next to one another without intervening spaces. Use 90ml of the 1% gel for a small BioRad casting chamber (14cm x 12.7cm). The plugs should occupy the entire height of the gel. Therefore, when casting the gel, make sure that the comb is touching the bottom of the casting chamber. Make sure that the casting chamber as well as the PFGE-chamber are absolutely leveled to avoid any loss of DNA during the gel run.}

Then seal the slot with 1% LMP agarose (0.25xTAE).

4. Run the PFGE in a cooled buffer (0.25xTAE) using conditions optimized to separate the YAC from the endogenous chromosomes.

{Note: Best separation results from endogenous chromosomes are achieved using a single pulse time instead of a time ramp for the entire run. It is worth while to test out several conditions before starting the isolation procedure}.

5. After the gel run cut off marker lanes on either side including about 0.5cm of the preparative lane and stain them in 0.25xTAE buffer containing 0.5µg/ml ethidium bromide. Mark the position of the YAC band under UV light using a sterile scalpel blade.

6. Reassemble the gel and excise the part of the preparative lane coresponding to the YAC-DNA. Cut also slices above and below the YAC DNA to serve as marker lanes for the second gel run.

7. Position the gel slices on a minigel tray with the YAC slice in the middle and cast a 4% LMP agarose (Nusieve, FMC) gel 0.25xTAE around them.

8. Run the gel at a 90° angle to the PFGE run for approximately 6-8h at 4V/cm in 0.25xTAE (circulating buffer). The running time depends on the size of the gel slice as well as on the agarose used for the PFGE run.

9. Cut off and stain the two marker lanes to localize the DNA.

{Note: If the DNA has not yet completely run into the Nusieve LMP gel continue the electrophoresis. Since it is impossible to digest normal agarose with the enzyme agarase, it is important to excise only LMP material.}

10. Excise the concentrated DNA from the corresponding position of the YAC DNA lane.

11. Equilibrate the gel slice on a rocking platform in 20ml of TENPA buffer for at least 1.5h.

12. Transfer the gel slice into a 1.5ml Eppendorf tube and remove all additional buffer using a fine tipped pipette.

13. Melt the agarose for 3min at 68°C, centrifuge for 10s to bring down all of the molten agarose and incubate for an additional 5min at 68°C.

14. Transfer the tube to 42°C for 5min. Add 2U of agarase (New Englands BioLabs) per 0.1 ml of molten gel slice.

{Note: Do not add agarase directly from the -20°C freezer, which can lead to setting of part of the LMP agarose. Load the enzyme into the tip and allow to warm up for a few secoonds by placing into the molten agarose. Carefully release the enzyme while stiring slowly with the tip. Mixing can be achieved by releasing air bubbles into the solution}.

Incubate for further 3h at 42°C.

15. Dialyse the resulting DNA solution for 2h on a floating dialysis membrane (Millipore, pore size 0.05μm) against microinjection buffer (10mM Tris:HCl, pH7.5, 0.1mM EDTA, 100mM NaCl, 30μM spermine, 70μM spermidine).

16. To determine the DNA concentration, check 2μl on a thin 0.8% agarose gel with very small slots, using λ DNA of known concentration as a standard.

{Note: It is useful to prepare a 2ng/μl stock solution of DNA. Loading of 2, 5, 10 and 20ng of this standard should allow a relatively accurate determination of the YAC DNA concentration}.

17. The integrity of the DNA can be checked running 20μl of the preparation on a PFGE gel (use a comb with small slots).

Injection into cultured cells

The Zeiss Automatic Injection System (AIS) can be used for rapid injection of large numbers of cells growing on cell culture dishes. A digital camera attached to a microsope transmits an image to the computer screen. An interactive computer program is then used to position the pipette above a "reference cell" and to mark the tip of the needle on the screen. This position is stored by the computer and serves as a reference point for the rest of the injections. Nuclei of other cells visible on the screen can now be marked by clicking on them with a computer mouse and injections are performed automatically by the computer. The amount of DNA injected can be regulated by the injection time as well as the pressure set at the Eppendorf injection system. High pressures result in higher efflux of the DNA containing solution. The pressure to be set depends on the viscosity of the DNA solution and the size of the needle opening and, therefore, has to be adjusted individually in each experiment. The pressure in a standard experiment will vary between 20 and 150 hectopascal. Almost confluent dishes are best to inject. A too low cell density allows only a few cells to be injected per frame,

whereas cells on confluent plates do not grow in one plane making it impossible to inject all cells into the nuclei. The efficiency of microinjection will depend greatly on the cell type. Best results are achieved using cells with big and easily visible nuclei.

1. Grow cells on a 5cm dish to 80 % confluency in the medium required by the cell type.

2. Immediately before injection cover the cells with 5ml of fresh medium and top layer the dish with 5ml of liquid paraffin. Liquid paraffin is preferably used to prevent contamination of the cells as well as evaporation of the medium during injections.

3. Switch on computer, microscope, monitor, Eppendorf microinjector and pump and place the culture dish on the stage.

4. Choose the command STAGE from the main menue to select a region of the dish which is almost confluent but in which the cells are still growing in one plane. The stage can be moved by clicking (always use the top/yellow button) onto the crossed double arrows. The direction of the arrow indicates the direction in which the stage will move. The distance from the center of the cross determines the speed with which the stage moves.

5. Return to the main menu and select MARK/INJECT. A new menu will appear which allows to choose from the following options:

STORE DATA: Allows to generate a file in which the positions of the injected cells will be stored. To use this option the bottom of the dish has to be marked to give the machine left and right hand references (scratch crosses at either side). Find the marks after the plate has been place on the stage and click cursor on the appropriate box to record the references. If you generate a file you must enter an operator and a sample name.

APPEND: Allows you to go back to a previous file to find the cells which have been microinjected.

NO FILE: This option does not record the cells that are injected and is sufficient for most applications.

6. Select the number of frames you want to inject by filling in numbers lower than 10 for X and Y values. A frame is the window visible on the screen and, therefore, represents the field in which cells can be marked and injected at a time. Each frame has specific X and Y coordinates. The computer moves along the x-axis first. An array of 5 x 10 frames will allow you to inject more than 1000 cells depending on the confluency of the plate.

7. Click on DATA OK.

8. Load 1.5µl of DNA solution into an Eppendorf microloader and insert it into an Eppendorf Femptotip placing it at the very bottom of the tip. Slowly release the DNA solution trying to avoid the introduction of air bubbles, which can block the needle.

9. Twist the tip carefully to remove it from its cover and load the needle by screwing it into the injection needle holder at the microscope.

10. Choose the option adjust from the menu. Use the mouse to lower the needle by clicking onto the arrow in the center of the screen. The distance from the center determines the speed of the movement. Start with high speed and slow down when you approach the surface of the dish. Once the needle touches the medium find it in low power magnification and use the micrometer screws on the pipette holder to center the needle in the frame. Change the lens to higher magnification. Focus on a plane intermediate between the cells and the needle and bring the needle down into focus. Repeat this procedure until the tip of the needle is pressing down on a cell. This will result in a small halo surrounding the needle tip as you press down on the cell.

11. The following options are available to adjust the position of the needle

STEP DOWN: Lowers the needle in the smallest possible increment.

MARK TIP: Allows to set the reference point for the computer software. To adjust click on the very tip of the injection needle.

INJECTION TIME: Determines the time the needle remains within the cell and is, therefore, one parameter for the volume delivered to the nucleus. This time has to be varied depending on the pressure, tip size etc. A time of 0.2s is a good value to start with.

MOVE STAGE: Allows to move the stage directed with the mouse.

RESTART: Takes you back to the main menu and you can reset any of the parameters.

HOME: Takes the needle back to the original position.

POSITION OK: Click on this when you are ready to start injecting.

12. To perform the injections click on MARK NEXT. This will allow to direct the computer to the nuclei of cells to be injected. Click on MARK and subsequently onto the nuclei. To start the injections click on INJECT. The computer will perform the injections into the marked cells. Successfully injected cells can be identified by a temporary dramatic swelling of the nucleus. If no change of cells can be observed after a number of injections check the following possibilities:

- The injection needle is blocked: Use the high pressure button (P3) at the injection machine to release DNA. If this does not help the needle has to be replaced.
- The computer is injecting in the wrong plane: Stop the injections by pressing the yellow button and try lowering or lifting the needle in single step increments. Be careful not to break the needle on the surface of the dish by lowering it too much.

Too low pressure: Increase the pressure for P1. Be aware that too high pressure will result in bursting of the cells.

Press the mouse button at any time during irjections to adjust the needle height or remark the tip of the needle to inject the whole frame again press RESTART. Alternatively you can carry on with CONTINUE.

13. To finish the injections press RESTART, MARKJINJECT, HOME, EXIT.

Pronuclear injections into fertilised mouse oocytes

The procedure of generating transgenic mice includes isolation of fertilized oocytes from superovulated females, microinjection of DNA into pronuclei and the transfer of injected oocytes into pseudopregnant foster mothers. A detailed description of these steps can be found in for example Hogan, Murphy and Carter (1993 Transgenesis in the mouse in "Transgenesis Techniques", Methods in Molecular Biology vol. 18 Ed. Murphy and Carter, pp 109-176. Humana Press, Totowa, New Jersey) and reference 28 (subsequent chapter) - the contents of each of which are incorporated herein by reference).

Preparation of DNA constructs for injection normally involves a filtration step in which the DNA is passed through a membrane with 0.2μm pore size. This step is recommended to avoid blocking of the injection needle by dust particles in the DNA solution. YAC DNA preparations should not be subjected to filtration, because of shearing forces occuring during this step. We have found that blockage of the needle is a relatively infrequent event if the agarose digestion was successful. In some cases it might be preferable to centrifuge the DNA for 5min (12000rpm Eppendorf centrifuge) to remove undigested gel pieces. However, since small particles of agarose can trap DNA we would strongly recommend to determine the DNA concentration after the centrifugation step.

Some DNA preparations are very sticky, which is probably due to incomplete agarose digestion. In these cases a higher proportion of injected oocytes will be found to lyse

and the injection needle has to be exchanged more frequently. Prepare a new batch of DNA for the next injection day and take care to digest all agarose. Even more than with normal constructs, try to avoid touching the pronuclei during injections. Once touched they will stick to the needle and being pulled out. If that happens replace the microinjection pipet immediately. The percentage of lysed oocytes should not be markedly higher, when compared with normal constructs. Injected oocytes can be either transfered on the same day to the oviduct of pseudopregnant foster mothers or incubated overnight at 37°C in M16 buffer. Normal survival rates (20 to 30%) of transfered embryos even at DNA concentrations as high as 10ng/μl should be obtained.

Transgenic animals can be identified by PCR or Southern blot analysis with DNA isolated from tail tips. With 250kb constructs about 10 to 20% of the offspring should have YAC DNA incorporated. Once a transgenic line has been established it is important to confirm the integrity of the integrated construct. This can be achieved by conventional PFGE mapping with several probes scattered over the YAC, which, however, requires a detailed knowledge of the restriction map of the construct. Alternatively, the RecA approach can be used to release the entire YAC from the mouse genome.

RESULTS

INSERTION VECTOR

General

The analysis in transgenic animals of genes expressed in YACs can be greatly facilitated by the use of a reporter gene for the accurate and sensitive detection of cellular sites of transcription. We have constructed a series of YAC modification vectors (pYIV1, pYIV2, pYIV3 and pYIV4) which can be inserted into YACs after the translation initiation or stop codon. The common feature of all of these vectors is that they contain a *lacZ* reporter gene downstream of a viral internal ribosome entry site (IRES), together with selective markers. In transgenic animals expressing these constructs, the *lacZ* reporter gene will be expressed in the same pattern as the transgene so that the

expression, regulation and function of the transgene can be analysed using simple histochemical staining procedures. This approach may provide a more complete picture of the pattern of expression of the transgene than standard procedures such as *in situ* hybridisation. The pYIV1 vector can be used for YACs which have been introduced into a Leu$^-$ yeast strain, while pYIV2, pYIV3, and pYIV4 can be directly introduced into the yeast strain (AB1380) which was used for construction of most YAC libraries. One of the vectors (pYIV3) permits the HA epitope tag sequence (from influenza hemagglutinin) to be fused to the carboxyl terminus of the expression product of the gene of interest, so that the protein product of the transgene can be detected using the commercially available 12CA5 monoclonal antibody. pYIV4 contains *loxP* elements flanking the SV40 polyadenylation signal and the *ADE2* gene. In transgenic animals generated using pYIV4, the polyA sequence and the *ADE2* gene sequences can be deleted using *Cre* recombinase so that the transgene and *lacZ* reporter gene are flanked by the authentic 3'-untranslated region of the transgene. A comparison between animals containing the SV40 polyadenylation signal and the *ADE2* gene with those in which these sequences have been removed will reveal the function of the 3' sequence of the transgene.

Expression of the IRES-lacZ from pYIV1 in YAC transgenic mice

We have examined the expression of the *lacZ* reporter gene in transgenic mice expressing a YAC clone modified using the pYIV1 insertion vector. Insulin-like growth factor II (IGF2) cDNA was introduced 3' of the Wilm's tumor (wt1) gene promoter (isolated from a 480 kb human YAC clone). Then the promoter and cDNA were inserted at the *Not*I-*Xba*I sites 5' of the IRES-*lacZ-LEU2* cassette in the pYIV1 vector. The first intron of the wt1 gene was cloned 3' of the *LEU2* gene. The construct was introduced to the 480 kb YAC by homologous recombination. After amplification, the modified YAC DNA was isolated and micro-injected into fertilised eggs. Eight lines of transgenic mice were produced, 5 of which expressed the *lacZ* reporter gene. All expressing lines produced an X-Gal (it is understood that the terms X-Gal, β-Gal and *LacZ* are synonymous) staining pattern (Fig.5) identical to that of the human gene from

which the promoter was derived. These data demonstrated that the IRES-*lacZ* reporter gene is functional in the YAC insertion vectors.

AMPLIFICATION VECTOR

The improved homologous recombination efficiency of pYAM4

To assess the homologous recombination efficiency of pYAM4, the plasmid was linearised with *Not*I and introduced into a variety of YAC clones from the ICI, ICRF and chromosome-7-specific YAC libraries. Recombinants were isolated on selective medium lacking uracil and lysine and replica plated on plates lacking uracil and tryptophan. Successful replacement of the left arm (containing *TRP1* gene) by pYAM4 would result in yeast capable of growth on medium lacking uracil and lysine but not on the counter selection medium lacking tryptophan.

Of a total of 1266 Lys^+ clones analysed, 167 clones could not grow on medium lacking tryptophan (Table 1). That is, the homologous recombination leading to the loss of the TRP1 gene occurred in 167 clones. The retrofitting efficiency of pYAM4 overall is 13.3% which is much higher than pCGS990 and pCGS966 (0.5-2.5%) (11, 12).

Table I. Retrofitting efficiency of pYAM4 in a variety of YAC clones from different YAC libraries

YAC clone	Insert Size	Ura^-/Lys^-	Ura^-/Trp^- (not growing)
ICI YAC Clones			
1 (16FC9)		51	4
2. (27FE5)	320	55	3
3. (30FH70	350	105	33
4. (3HG4)	300	46	2
5. (36AH3)	300	77	16
6. (12GG6)	300	141	31
Subtotal		475	89 (18.7%)

ICRF YAC clones			
1. (49A9)	340	37	3
2. (35D8)	500	103	8
3. (132C6)	630	133	19
Subtotal		274	30 (10.9%)
Chromosomal YAC clones			
HSC7E526	550	107	8 (7.5%)
E145A7	200	129	11 (8.5%)
ywss922	350	57	8
ywss1545	420	59	8
ywss2056	400	15	2
ywss3844	300	151	11
Subtotal		282	29 (10.3%)
Total	1266	167 (13.3%)	

The increased efficiency of retrofitting is probably due to the introduction of a 572 bp *Cla*I-*Sma*I fragment from pYAC4 adjacent to the *CEN4*. When the *Sma*I-*Cla*I fragment of pYAM4 was deleted by *Not*I-*Cla*I digestion, or pYAM4 was linearised with *Cla*I, the frequency of tryptophan sensitive clones was not significantly different from that obtained with pCGS990.

Amplification of YAC DNA by pYAM4

Tryptophan sensitive clones were cultured in selective medium (Ura^-/Lys^-) with galactose as carbon source instead of glucose. In such medium, the GAL1 promoter adjacent to the *CEN4* in the pYAM4 vector will be induced. Activation of transcription from the GAL1 promoter should interfere with the CEN4 leading to non-segregation of the YAC DNA therefore increase the YAC DNA copy number per cell.

As shown in Fig 7, depending on the size and nature of the insert, human YAC DNA was amplified 3 to 5 fold. Although the amplification is not as high as that achieved with pCGS990, it helps to isolate more concentrated YAC DNA for transgenics.

Introduction of an additional conditional promoter such as ADH2 adjacent to the *CEN4*, or of an additional selection gene, might improve the amplification further.

We have introduced the bacterial hygromycin B resistance gene under the control of the mouse *Pgk*1 promoter into pYAM4. After retrofitting with pYAM4, isolated YAC DNA can be introduced into mammalian cells such as embryonic stem (ES) cells, which is an alternative approach to microinjection of YAC DNA for making YAC transgenic animals (29).

If a genomic fragment (which is present in a YAC) is cloned into the *Cla*I-*Not*I site of pYAM4, the truncation of a large YAC and amplification of the shortened YAC DNA can be achieved in a single step.

COMBINATION OF INSERTION VECTOR AND AMPLIFICATION VECTOR

The results are presented in Figures 8 and 9. In this regard, Figures 8 and 9 are photographs of gels.

The amplification of the YAC DNA can be seen from Fig 8 and 9. In this regard: Lane 1 is un-amplified YAC DNA as present in original 35D8 YAC clone, lane 2 is un-amplified YAC DNA in another YAC clone (132C6) containing the SERT gene, and lane 3 is amplified YAC DNA in 35D8/D6 subclone. The blot was hybridized with genomic probes downstream (Fig.8) and upstream (Fig.9) of the SERT gene.

AMPLIFICATION AND PURIFICATION OF YAC DNA

The integration of the amplification vector pYAM4 into YAC clones 35D8 and HSC7E526 greatly increased the yields of YAC DNA obtained. These results are presented in Figure 11. In this regard, Figure 11 is a photograph of a gel prepared by a Southern blot and hybridised with a ^{32}P-labelled pBR322 probe to detect YAC sequences. In each lane, the hybridising band corresponding to YAC DNA is arrowed

Lane 1 is DNA from yeast containing YAC clone 35D8; Lane 2 is DNA from yeast containing YAC clone 35D8/D6, in which the YAC has been modified by integration of the amplification vector pYAM4 and the insertion vector pYIV3; Lane 3 is the isolated YAC DNA from clone 35D8/D6 prior to microinjection. Lane 4 is DNA from yeast containing YAC clone HSC7E526/V12, in which the YAC HSC7E526 has been modified by integration of the amplification vector pYAM4 and the insertion vector pYIV3. Lane 5 is the isolated YAC DNA from clone HSC7E526/V12 prior to microinjection.

GENERATION OF YAC TRANSGENIC MICE

Modified YAC DNA was excised from a 1% pulse-field agarose gel in 0.25 x TAE buffer and concentrated into 4% low melting point agarose. The gel slice containing YAC DNA was equilibrated with microinjection buffer (TE pH 7.0 with 0.1 M NaCl) and digested with gelase. YAC DNA was dialysed against the microinjection buffer for 2 hours before injection into fertilised oocytes.

Two hundred and ninety-eight fertilised oocytes were injected with 35D8/D6 YAC DNA and 364 with HSC7E526/V12. After transfer of injected oocytes into oviducts of pseudopregnant female mice, a total of 190 mice (28.7%) were born of which 26 (13.7%) carried YAC DNA as determined by PCR as shown below:

Table II. Survival rate of transfered oocytes and YAC transgenic mice

YAC Construct	Size (kb)	NEIT	Born	Died	Transgenic
35D8/D6	500	298	97	6	18
HSC7E526/V12	630	364	93	1	8
Total		662	190 (28.7%)	7	26 (13.7%)

N.E.I.T: Number of oocytes injected and transferred

PCR determination of the size of the integrated construct

The size of the integrated YAC 35D8/D6 and YAC HSC7E526/V12 constructs in each transgenic founder animal were determined using two pairs of PCR primers (A and H: Table III) to detect the two YAC vector arms and a series of PCR primer pairs spanning the SERT (B to G: Table III) and VIPR2 (I to L: Table III) genes respectively.

Table III. Primer Pairs Used For PCR

Primer Pair	Forward Primers (5' - 3')	Reverse Primers (5' - 3')
A	54773 TTGACTGGAGCGAGGCGATGTTCG	54774 TCTACACAGCCATCGGTCCAGACG
B	43084 GCGTCTAGGTGGCACCAGAATC	43085 TCGCGCTTGTGTTCCCAGCTAC
C	44771 CTAGTGACTGACATTGCCTGG	44772 TGTCCAGTCTATCTGCACATG
D	26375 AGTTCTGATGAGGCACGC	26376 TTCATCACCTCCATCCACATCC
E	38127 TGGCATGCAATTGTAGTCTC	38128 TTCTTTCCTTACTAAGTTGAGAACG
F	38125 GAATACCAGGTCACCACATGG	38126 AAACCTTGCACAGGGTCTTG
G	38130 TGGCTTTAGTATTTTTCTTTGTTTT	38131 GAAAAGTACTCCTCAGTAGGTTGAA
H	53537 TCTCCGAACAGAAGGAAGAACG	53538 TGTTACTTCTTCTGCCGCCTGC
I	53525 CCACATACAGACTGATGAATC	53526 AGCTGGAATTGGAACTCAGC.
J	53527 AGTGCCTAAAGGGTTATG	53528 AAAGATGCTACACCTGAG
K	53531 CTCAATATCACACAACAGTG	53532 GCTGAAAGGAAAACTGATTG
L	53533 CATCCCAGATTTGTATAG	53534 ATAATCCCAAGAGGCAAG

The results are presented in Figures 12 and 13. In this regard, Figure 12 shows the size of the integrated YAC DNA in transgenic founder animals carrying 35D8/D6 YAC DNA. Figure 13 shows the size of integrated YAC DNA in transgenic founder animals carrying HSC7E526/V12 YAC DNA. For each founder animal, the probable extent of the transgene is indicated as a shaded bar, with pale circles indicating presence of markers, as determined by PCR. The location of these markers is indicated in the schematic diagram of the 35D8/D6 YAC construct (Figure 12) and HSC7E526/V12 YAC (Figure 13) construct which are drawn above the markers.

Three independent transgenic founder mice carrying the intact YAC 35D8/D6 (A102.3, A102.5, A105, Figure 12) and six carrying the intact YAC HSC7E526/V12 (A108, A108.1, A108.2, A108.3, A108.5, A110: Figure 13) were identified. Thus, a beneficial number of mice born in this study carried intact YAC DNA.

GERM LINE TRANSMISSION

An important feature of pYAM4 is that it does not necessarily have to contain a thymidine kinase (*TK*) gene which may cause male infertility. In the absence of the *TK* gene, we found that YAC transgenes were transmitted into the next generation from both male and female founders as determined by primer pair A, which is derived from the hygromycin resistance gene within pYAM4. Thus, preferably, the *TK* gene is not present in the vector of the present invention.

Immunocytochemistry for beta -galactosidase

Animals were perfused transcardially with 4% paraformaldehyde in 0.1M PBS. Brains were postfixed overnight in the same solution, then transferred into PBS next day and kept in that solution until cutting. Brains were infiltrated with 30% sucrose overnight, then 25μm thick sections were cut on a cryostat. Sections were collected in multiwell plates containing PBS and then processed for immunocytochemistry. Sections were treated with 0.1% Triton X-100 and 0.02% H_2O_2 solution for 30 min, then rinsed two times for 5 min with PBS. This was followed by a blocking step with 2% solution of normal donkey

serum, for another half an hour. Sections were incubated in anti beta-galactosidase primary antibody (5'Prime3'Prime, Inc., diluted at 1:20000) for 48 hours. Biotinylated donkey anti rabbit IgG (from Jackson, diluted at 1:1000, for 60 min.), ABC Elite Kit (from Vector, diluted at 1:1000, for 60 min.) and NiDAB as chromogen were used to visualise the immunoreactive areas.

LacZ staining of adult transgenic mice expressing the YAC construct HSC7E526/V12

Mice were anaesthetised with a lethal dose of sodium pentobarbitone and briefly perfused through the heart with 0.9% sodium chloride solution to remove blood followed by a longer perfusion of the ice-cold fixative solution (4% paraformaldehyde in 0.1M sodium phospahate buffer, pH 7.4). After perfusion with approximately 150-200 ml of the fixative solution, the brains and internal organs were removed rapidly and postfixed in the same fixative for 2-4 hours at 4°C. Subsequently 2-5 mm coronal slices of the brain were cut and the brain and other organs were washed twice at room temperature in a detergent wash solution consisting of 2 mM Magnesium Chloride, 0.01% Sodium Deoxycholate, and 0.02% NP40 in phosphate-buffered saline (PBS), pH 7.4. After washing all tisues were transferred to a solution containing 1mg/ml X-Gal in 5mM potassium ferrocyanide, 5mM potassium ferricyanide, 2mM Magnesium chloride, 0.01% Sodium deoxycholate, and 0.02% NP40 in PBS, pH 7.4 and incubated overnight at 30°C. After staining, tissues were washed in PBS and cleared in 40% and 80% glycerol (v/v) in PBS and photographed.

The distribution of β-galactosidase activity in transgenic mice was consistent with the published distribution of VIPR2 mRNA (Cagampang *et al.*, 1998; Inagaki *et al.*, 1994; Sheward *et al.*, 1995; Usdin *et al.*, 1994) and of binding sites for the selective VIP_2 receptor agonist Ro25-1553(Vertongen *et al.*, 1997). In particular, high levels of expression of β-galactosidase were detected:

(i) In the suprachiasmatic nucleus (Figure 15a, 15b), where VIP and/or PACAP,

acting through the VIP_2 receptor, may play a role in the control of circadian rhythms. There is circadian variation of VIP immunoreactivity (Takahashi *et al.*, 1989), prepro VIP mRNA (Albers *et al.*, 1990; Glazer and Gozes, 1994; Gozes *et al.*, 1989) and of VIP_2 receptor mRNA (Cagampang *et al.*, 1998) in the SCN and VIP (Piggins *et al.*, 1995), VIP antagonists (Gozes *et al.*, 1995) and VIP antisense oligodeoxynucleotides (Scarbrough *et al.*, 1996) have been shown to disrupt circadian function.

(ii) In the pancreas (Figure 5c), where VIP and PACAP stimulate insulin release by interaction with the VIP_2 receptor on the beta cell (Straub and Sharp, 1996).

Chemiluminescent assay for β-galactosidase in mouse tissues using the Tropix Galacto-Light Plus kit.

Tissues from mice were dissected, frozen on dry ice and stored at -70°C. They were thawed and homogenised immediately in 100-400μl of cold lysis buffer (as supplied in the kit, with 0.2mM PMSF and 5μg/ml leupeptin added just before use). After homogenisation, samples were centrifuged at 12000g for 10 min at 4°C. An aliquot of the supernatant was stored at -70°C for measurement of protein concentration and the rest was incubated at 48°C for 60 minutes to inactivate the endogenous β-galactosidase. After centrifugation for 5 min at room temperature 20μl of each sample were used in the assay. 200μl of Galacto-Light reaction buffer was added, inbubated for 60 min at room temperature and then 300μl of Accelerator was added and the sample counted in a TD-20/20 luminometer (for 20 sec after 5 sec delay). Protein concentrations were determined using the Bio- Rad assay and activity was expressed as light units/min/mg of protein.

DISCUSSION

INSERTION VECTOR

Almost all YAC transgenic animals described to date have been made with unmodified YAC DNA.

Expression of the transgenes in these animals can only be assessed by *in situ* hybridisation, Northern blotting, PCR and/or immunohistochemistry. Introduction of a reporter gene into the YAC DNA would simplify procedures for the detection of transgene expression.

We have constructed a series of YAC modification vectors (pYIV1, pYIV2, pYIV3 and pYIV4) which can be inserted into YACs after the translation initiation or stop codon. The vectors contain a *lacZ* reporter gene downstream of a viral internal ribosome entry site (IRES), so that a simple histochemical staining procedure can be used to examine the tissue distribution and regulation of the transgene. This approach provides a more complete picture of the pattern of expression of the transgene than standard procedures such as *in situ* hybridisation. One of the vectors (pYIV3) permits the HA epitope tag sequence (from influenza hemagglutinin) to be fused to the carboxyl terminus of the gene product of interest, so that the protein product of the transgene can be detected using the commercially available 12CA5 monoclonal antibody. pYIV4 contains *loxP* elements flanking the SV40 polyadenylation signal and the ADE2 gene. In transgenic animals generated using pYIV4, polyA and *ADE2* sequences can be deleted using *Cre* recombinase so that the transgene and *lacZ* reporter gene are flanked by the authentic 3'-untranslated region of the transgene.

In summation therefore, transgenic technology has played an important role in the understanding of gene function and regulation *in vivo*, and in creating animal models of human genetic diseases. The development of yeast artificial chromosome (YAC) technology has permitted the cloning of DNA segments thousands of kilobases in size

between two YAC vector arms, facilitating transfer of a whole gene and most (if not all) of the elements required for its faithful regulation into transgenic animals.

The analysis in transgenic animals of genes expressed in YACs can be greatly facilitated by the use of a reporter gene in accordance with the present invention for the accurate and sensitive detection of cellular sites of transcription.

AMPLIFICATION VECTOR

Importance of YAC technology and YAC transgenesis

The development of the yeast artificial chromosome (YAC) technology has permitted the cloning of DNA segments thousands of kilobases in size between two YAC vector arms. YACs can be used to clone the complete sequences of large genes or gene complexes that exceed the size limit for cloning in conventional bacterial cloning vectors such as plasmids (10 kb), bacteriophage (15 kb), and cosmids (50 kb). Cloning of such large DNA fragments is essential for physical genome mapping (1) and to isolate large genes relevant to human genetic disease (2, 3). Although bacterial artificial chromosome (BAC) (4) and P1 artificial chromosome vectors (5) have a large cloning capacity (up to 200 kb), it is relatively difficult to perform genetic manipulation in these vectors. The high efficiency of homologous recombination in yeast permits genetic manipulations of genes cloned in YAC vectors to be performed easily.

Transgenic technology has played an important role in the understanding of gene function and regulation *in vivo*, and in creating animal models of human genetic diseases. It is well recognised that transgenes containing genomic DNA with introns and essential regulatory sequences are expressed more appropriately *in vivo* than cDNA based constructs (6-10). The use of YAC constructs to produce transgenic animals facilitates the presence and transfer of most (if not all) elements required for the faithful regulation of a gene and may avoid position effects related to the integration site, which may lead to low levels and, in some cases, aberrant patterns of gene expression in transgenic

animals. However, the efficiency with which transgenic animals are produced using YAC DNA (1-5%) is lower than that using conventional vectors (approx. 10%)

Importance of YAC amplification

YAC vectors (for example pYAC4) contain a yeast centromere, two telomeres, and two selective markers (*TRP1* and *URA3*). After incorporation of YAC DNA, yeast can grow in medium lacking the uracil and tryptophan. Under these conditions of selection, YAC clones are replicated along with the endogenous host chromosomes; only one copy of YAC DNA is produced per cell. Using standard protocols, YAC DNA at a concentration of 1 μg/ml can be isolated. However, there is a substantial increase in copy number if the YAC centromere is inactivated by induced transcription from a *GAL1* or *ADH2* promoter. This increase is thought to reflect the segregation bias of the YAC for the mother cells and loss of the daughter cells without the YAC from the population under selection.

The low efficiency with which transgenic animals are produced using YAC DNA (1-5%) compared to DNA from conventional vectors (approx. 10%) is probably caused by the low concentration of YAC DNA available for injection. Also, conventional YACs are replicated in yeast at one copy per cell.

Therefore, assuming that 2 pl of 500 kb YAC at a concentration of 1 ng/μl is injected into a pronucleus, a fertilised egg would only receive 1 molecule of YAC DNA. Amplification of YACs in yeast therefore provides a possible method for the isolation of more concentrated YAC DNA which should lead to more successful generation of YAC transgenic animals.

Advantages and disadvantages of the existing YAC amplification vectors pCGS966 and pCGS990

To date two vectors which can be used to amplify YAC DNA have been reported: pCGS966 (11) and pCGS990 (12). Both vectors include a conditional centromere and a

heterologous (herpes simplex virus) thymidine kinase (*TK*) gene. YAC DNA of less than 600 kb is amplified efficiently (3 to 11 copies/cell). pCGS966 has been used recently to construct a number of new YAC libraries (13-15). However, when using this vector for the modification (retrofitting) of existing YACs, replacement of the left arm occurs with very low frequency (0.5-2.5%) (12). Most importantly, the expression of the *TK* gene in the testes of transgenic mice interferes with spermatogenesis and causes male infertility (16-22). This complication makes these vectors unsuitable for transgenic studies.

Features of the novel YAC amplification vector pYAM4.

We have constructed a YAC amplification vector which has a number of advantages:

1) it amplifies YAC DNA 3 to 5 fold.

2) unlike existing vectors, it does not contain the herpes simplex virus thymidine kinase (*TK*) gene, which causes male infertility in transgenic mice.

3) it has much higher homologous recombination efficiency (13%) than existing YAC amplification vectors.

4) it contains a selectable marker (hygromycin B resistance) which facilitates the transfer of YAC DNA into embryonic stem cells and other cell lines.

5) it can be used for targeted deletion of sequences cloned in YAC vectors.

6) Fusion of a NOI (such as the SERT gene) to a reporter gene (such as *LacZ*) facilitates the determination of the sites/regions where the NOI is expressed and the testing of agents which may affect the expression pattern of the NOI.

7) Transgenic mice overexpressing the human VIPR2 gene together with the β-galactosidase reporter gene will facilitate the development of agents capable of influencing the activity of the VIP_2 receptor in man. Two classes of agent might be

identified: (i) agents regulating the expression of the human VIP_2 receptor, which could be identified by their ability to influence β-galactosidase activity in transgenic mice and (ii) agonists and antagonists of the human VIP_2 receptor, for which the transgenic mice in which the human VIPR2 gene is expressed will provide an animal model. Optimally, YAC transgenic animals, such as VIP_2 receptor null ("knock out") mice, could be bred with a view to generating "humanised" animals in which the VIP_2 receptor displays identical pharmacology to that seen in man.

By way of example, agents, acting in the suprachiasmatic nucleus, which are capable of influencing the activity of the VIP_2 receptor, may prove useful in the treatment of the disturbances in circadian function. Such disturbances, which may lead to impaired physical and mental well-being, can occur through: (i) extremes in work patterns (shift work); (ii) travelling through many time zones (jet lag); (iii) in normal ageing; and (iv) in dementia. Such agents may also prove useful in the treatment of sleep disorders, seasonal affective disorder (SAD), eating disorders and pre-menstrual syndrome.

By way of further example, agents, acting in the pancreas, which are capable of regulating the expression of the human VIP_2 receptor or agents acting as agonists and antagonists of the receptor may be useful in the treatment of diabetes.

In summation introduction of the *lacZ* reporter gene by using pYIVs, together with amplification of YAC DNA by using pYAM4 should greatly facilitate production of YAC transgenic animals and analysis of these animals in terms of expression, regulation and function of genes present in YAC DNA.

SUMMARY

In these studies we present the construction of a series of YAC modification vectors (such as pYIV1, pYIV2, pYIV3 and pYIV4) that contain a reporter gene (such as the *lacZ* reporter gene) to facilitate the study of the tissue distribution and regulation of YAC

transgenes. These vectors are likely to find widespread application in transgenic research.

In these studies we also present the construction of a YAC amplification vector (such as pYAM4) which has a number of advantages over previous vectors and is suitable for the amplification of YAC DNA for the creation of transgenic mice.

In these studies we also present the combined use of the above-mentioned modification vectors (such as pYIV1, pYIV2, pYIV3 and pYIV4) and the above-mentioned YAC amplification vector (such as pYAM4). The combined use of these vectors is likely to find widespread application in transgenic research.

For example, the vectors of the present invention - in particular the insertion vectors of the present invention - may be used to prepare other artificial chromosomes (i.e. artificial chromosomes other than YACs), which may in turn be used to prepare transgenic organisms (including animals). In this alternative embodiment, the above mentioned statements of invention and description are still applicable but wherein the term YAC represents any suitable artificial chromosome, preferably a yeast artificial chromosome.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

REFERENCES

1 Cohen, D., I. Chumakov and J. Weissenbach (1993) *Nature* **366**, 698-701.

2 Nobile, C., F. Galvagni, J. Marchi, R. Roberts and I. Vitiello (1995) *Genomics* **28**, 97-100.

3 Couch, F.J., L.H. Castilla, J.Z. Xu, K.J. Abel, P. Welcsh, S.E. King, L.H. Wong, P.P. Ho, S. Merajver, L.C. Brody, G.Y. Yin, S.T. Hayes, L.M. Gieser, W.L. Flejter, T.W. Glover, L.S. Friedman, E.D. Lynch, J.E. Meza, M.E. King, D.J. Law, L. Deaven, A.M. Bowcock, F.S. Collins, B.L. Weber and S.C. Chandrasekharappa (1995) *Genomics* **25**, 264-273.

4 Shizuya, H., B. Birren, U.J. Kim, V. Mancino, T. Slepak, Y. Tachiiri and M. Simon (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8794-8797.

5 Shepherd, N.S., B.D. Pfrogner, J.N. Coulby, S.L. Ackerman, G. Vaidyanathan, R.H. Sauer, T.C. Balkenhol and N. Sternberg (1994) *Proceedings Of the National Academy Of Sciences Of the United States Of America* **91**, 2629-2633.

6 Gordon, J.W. (1993) *Meth. Enzymol.* **225**, 747-771.

7 Brinster, R.L., J.M. Allen, R.R. Behringer, R.E. Gelinas and R.D. Palmiter (1988) *Proc. Natl. Acad. Sci. USA* **85**, 836-840.

8 Choi, T., M. Huang, C. Gorman and R. Jaenisch (1991) *Mol. Cell. Biol.* **11**, 3070-3074.

9 Jaenisch, R. (1988) *Science* **240**, 1468-1474.

10 Lamb, B.T. and J.D. Gearhart (1995) *Curr. Opin. Genet. Develop.* **5**, 342-358.

11 Smith, D.R., A.P. Snyth and D.T. Moir (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8242-8246.

12 Smith, D.R., A.P. Smyth, W.M. Strauss and D.T. Moir (1993) *Mamm. Genome* **4**, 141-147.

13 Toye, A.A., L. Schalkwyk, H. Lehrach and N. Bumstead (1997) *Mamm. Genome* **8**, 274-276.

14 Leeb, T., G. Rettenberger, H.B. Hameister, G. and B. Brenig (1995) *Mamm. Genome* **6**, 37-41.

15 Cai, L., L.C. Schalkwyk, A. Schoeberlein-Stehli, R.Y. Zee, A. Smith, T. Haaf, M. Georges, H. Lehrach and K. Lindpaintner (1997) *Genomics* **39**, 385-392.

16 Al-Shawi, R., J. Burke, H. Wallace, C. Jones, S. Harrison, D. Buxton, S. Maley, A. Chandley and J.O. Bishop (1991) *Mol. Cell. Biol.* **11**, 4207-4216.

17 Al-Shawi, R., J. Burke, C.T. Jones, J.P. Simons and J.O. Bishop (1988) *Mol. Cell. Biol.* **8**, 4821-4828.

18 Braun, R.E., D. Lo, C.A. Pinkert, G. Widera, R.A. Flavell, R.D. Palmiter and R.L. Brinster (1990) *Biol. Reprod.* **43**, 684-693.

19 Dzierzak, E., B. Daly, P. Fraser, L. Larsson and A. Muller (1993) *Int. Immunol.* **5**, 975-984.

20 Heyman, R.A., E. Borrelli, J. Lesley, D. Anderson, D.D. Richman, S.M. Baird, R. Hyman and R.M. Evans (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2698-2702.

21 Iwakura, Y., M. Asano, Y. Nishimune and Y. Kawade (1988) *EMBO J.* **7**, 3757-3762.

22 Neznanov, N., I.S. Thorey, G. Cecena and R.G. Oshima (1993) *Mol. Cell. Biol.* **13**, 2214-2223.

23 Shen, S., F.A. Kruyt, J. den Hertog, P.T. van der Saag and W. Kruijer (1991) *DNA Sequence* **2**, 111-119.

24 Mountford, P., B. Zevnik, A. Duwel, J. Nichols, M. Li, C. Dani, M. Robertson, I. Chambers and A. Smith (1994) *Proceedings Of the National Academy Of Sciences Of the United States Of America* **91**, 4303-4307.

25 Beach, D. and P. Nurse (1981) *Nature* **290**, 140-142.

26 Schedl, A., L. Montoliu, G. Kelsey and G. Schutz (1993) *Nature* **362**, 258-261.

27 Bellis, M., M. Pages and G. Roizes (1987) *Nucl. Acids Res.* **15**, 6749-6749.

28 Schedl, A., B. Grimes and L. Montoliu (1995) In: Yeast artificial chromosomes (YAC) protocols eds. D. Markie. Methods in Molecular Biology **54**, 293-306 (Humana Press, Totowa, NJ).

29 Peterson, K.R., C.H. Clegg, Q.L. Li and G. Stamatoyannopoulos (1997) *Trends Genet.* **13**, 61-66.

30. Albers, H. E., Stopa, E. G., Zoeller, R. T., Kauer, J. S., King, J. C., Fink, J. S., Mobtaker, H., and Wolfe, H. (1990). *Brain Res Mol Brain Res* **7:** 85-9.

31. Cagampang, F. R. A., Sheward, W. J., Harmar, A. J., Piggins, H. D., and Coen, C. W. (1998). *Molecular Brain Research* **54:** 108-112.

32. Glazer, R., and Gozes, I. (1994). *Brain Res* **644:** 164-7.

33. Gozes, I., Lilling, G., Glazer, R., Ticher, A., Ashkenazi, I. E., Davidson, A., Rubinraut, S., Fridkin, M., and Brenneman, D. E. (1995). *J Pharmacol Exp Ther* **273:** 161-7.

34. Gozes, I., Shani, Y., Liu, B., and Burbach, J. P. H. (1989). *Neuroscience Research Communications* **5:** 83-86.

35. Inagaki, N., Yoshida, H., Mizuta, M., Mizuno, N., Fujii, Y., Gonoi, T., Miyazaki, J., and Seino, S. (1994). *Proc Natl Acad Sci U S A* **91:** 2679-83.

36. Piggins, H. D., Antle, M. C., and Rusak, B. (1995). Neuropeptides phase shift the mammalian circadian pacemaker. *J Neurosci* **15:** 5612-22.

37. Scarbrough, K., Harney, J. P., Rosewell, K. L., and Wise, P. M. (1996). *Am J Physiol* **270:** R283-8.

38. Sheward, W. J., Lutz, E. M., and Harmar, A. J. (1995). *Neuroscience* **67:** 409-18.

39. Straub, S. G., and Sharp, G. W. (1996). *J Biol Chem* **271:** 1660-8.

40. Takahashi, Y., Okamura, H., Yanaihara, N., Hamada, S., Fujita, S., and Ibata, Y. (1989). *Brain Res* **497:** 374-7.

41. Usdin, T. B., Bonner, T. I., and Mezey, E. (1994). *Endocrinology* **135:** 2662-80.

42. Vertongen, P., Schiffmann, S. N., Gourlet, P., and Robberecht, P. (1997). *Peptides* **18:** 1547-1554.

SEQUENCE

SEQ ID. NO. 1

```
1   ATCGATAAGC TTTAATGCGG TAGTTTATCA CAGTTAAATT GCTAACGCAG TCAGGCACCG
61  TGTATGAAAT CTAACAATGC GCTCATCGTC ATCCTCGGCA CCGTCACCCT GGATGCTGTA
121 GGCATAGGCT TGGTTATGCC GGTACTGCCG GGCCTCTTGC GGGATATCGT CCATTCCGAC
181 AGCATCGCCA GTCACTATGG CGTGCTGCTA GCGCTATATG CGTTGATGCA ATTTCTATGC
241 GCACCCGTTC TCGGAGCACT GTCCGACCGC TTTGGCCGCC GCCCAGTCCT GCTCGCTTCG
301 CTACTTGGAG CCACTATCGA CTACGCGATC ATGGCGACCA CACCCGTCCT GTGGATCAAT
361 TCCCTTTAGT ATAAATTTCA CTCTGAACCA TCTTGGAAGG ACCGGTAATT ATTTCAAATC
421 TCTTTTTCAA TTGTATATGT GTTATGTTAT GTAGTATACT CTTTCTTCAA CAATTAAATA
481 CTCTCGGTAG CCAAGTTGGT TTAAGGCGCA AGACTTTAAT TTATCACTAC GGAATTCCGT
541 AATCTTGAGA TCGGGCGTTC GATCGCCCCG GG
```

SEQ ID. NO. 2

```
LOCUS       SERT.DOC      2889 BP DS-DNA              ENTERED   11/24/98
BASE COUNT      672 A    753 C    722 G    742 T      0 OTHER

COMMENT       1 -   85 exon 1A
             86 -  182 exon 1B (some splice variants do not include this exon)
            183 -  648 exon  2
            649 -  783 exon  3
            784 - 1003 exon  4
           1004 - 1142 exon  5
           1143 - 1277 exon  6
           1278 - 1381 exon  7
           1382 - 1509 exon  8
           1510 - 1622 exon  9
           1623 - 1754 exon 10
           1755 - 1854 exon 11
           1855 - 1955 exon 12
           1956 - 2123 exon 13
           2124 - 2889 exon 14
            180 - 2198 open reading frame encoding SERT
ORIGIN      -
      1 ACAGCCAGCG CCGCCGGGTG CCTCGAGGGC GCGAGGCCAG CCCGCCTGCC CAGCCCGGGA
     61 CCAGCCTCCC CGCGCAGCCT GGCAGGTCTC CTGGAGGCAA GGCGACCTTG CTTGCCCTCT
    121 ATTGCAGAAT AACAAGGGGC TTAGCCACAG GAGTTGCTGG CAAGTGGAAA GAAGAACAAA
    181 TGAGTCAATC CCGACATATC AATCCCGACG ATAGAGAGCT CGGAGGTGAT CCACAAATCC
    241 AAGCACCCAG AGATCAATTG GGATCCTTGG CAGATGGACA TCAGTGTCAT TTACTAACCA
    301 GCAGGATGGA GACGACGCCC TTGAATTCTC AGAAGCAGCT ATCAGCGTGT GAAGATGGAG
    361 AAGATTGTCA GGAAAACGGA GTTCTACAGA AGGTTGTTCC CACCCCAGGG GACAAAGTGG
    421 AGTCCGGGCA AATATCCAAT GGGTACTCAG CAGTTCCAAG TCCTGGTGCG GGAGATGACA
    481 CACGGCACTC TATCCCAGCG ACCACCACCA CCCTAGTGGC TGAGCTTCAT CAAGGGGAAC
    541 GGGAGACCTG GGGCAAGAAG GTGGATTTCC TTCTCTCAGT GATTGGCTAT GCTGTGGACC
    601 TGGGCAATGT CTGGCGCTTC CCCTACATAT GTTACCAGAA TGGAGGGGGG GCATTCCTCC
    661 TCCCCTACAC CATCATGGCC ATTTTTGGGG GAATCCCGCT CTTTTACATG GAGCTCGCAC
```

```
 721 TGGGACAGTA CCACCGAAAT GGATGCATTT CAATATGGAG GAAAATCTGC CCGATTTTCA
 781 AAGGGATTGG TTATGCCATC TGCATCATTG CCTTTTACAT TGCTTCCTAC TACAACACCA
 841 TCATGGCCTG GGCGCTATAC TACCTCATCT CCTCCTTCAC GGACCAGCTG CCCTGGACCA
 901 GCTGCAAGAA CTCCTGGAAC ACTGGCAACT GCACCAATTA CTTCTCCGAG GACAACATCA
 961 CCTGGACCCT CCATTCCACG TCCCCTGCTG AAGAATTTTA CACGCGCCAC GTCCTGCAGA
1021 TCCACCGGTC TAAGGGGCTC CAGGACCTGG GGGGCATCAG CTGGCAGCTG GCCCTCTGCA
1081 TCATGCTGAT CTTCACTGTT ATCTACTTCA GCATCTGGAA AGGCGTCAAG ACCTCTGGCA
1141 AGGTGGTGTG GGTGACAGCC ACCTTCCCTT ATATCATCCT TTCTGTCCTG CTGGTGAGGG
1201 GTGCCACCCT CCCTGGAGCC TGGAGGGGTG TTCTCTTCTA CTTGAAACCC AATTGGCAGA
1261 AACTCCTGGA GACAGGGGTG TGGATAGATG CAGCCGCTCA GATCTTCTTC TCTCTTGGTC
1321 CGGGCTTTGG GGTCCTGCTG GCTTTTGCTA GCTACAACAA GTTCAACAAC AACTGCTACC
1381 AAGATGCCCT GGTGACCAGC GTGGTGAACT GCATGACGAG CTTCGTTTCG GGATTTGTCA
1441 TCTTCACAGT GCTCGGTTAC ATGGCTGAGA TGAGGAATGA AGATGTGTCT GAGGTGGCCA
1501 AAGACGCAGG TCCCAGCCTC CTCTTCATCA CGTATGCAGA AGCGATAGCC AACATGCCAG
1561 CGTCCACTTT CTTTGCCATC ATCTTCTTTC TGATGTTAAT CACGCTGGGC TTGGACAGCA
1621 CGTTTGCAGG CTTGGAGGGG GTGATCACGG CTGTGCTGGA TGAGTTCCCA CACGTCTGGG
1681 CCAAGCGCCG GGAGCGGTTC GTGCTCGCCG TGGTCATCAC CTGCTTCTTT GGATCCCTGG
1741 TCACCCTGAC TTTTGGAGGG GCCTACGTGG TGAAGCTGCT GGAGGAGTAT GCCACGGGGC
1801 CCGCAGTGCT CACTGTCGCG CTGATCGAAG CAGTCGCTGT GTCTTGGTTC TATGGCATCA
1861 CTCAGTTCTG CAGGGACGTG AAGGAAATGC TCGGCTTCAG CCCGGGGTGG TTCTGGAGGA
1921 TCTGCTGGGT GGCCATCAGC CCTCTGTTTC TCCTGTTCAT CATTTGCAGT TTTCTGATGA
1981 GCCCGCCACA ACTACGACTT TTCCAATATA ATTATCCTTA CTGGAGTATC ATCTTGGGTT
2041 ACTGCATAGG AACCTCATCT TTCATTTGCA TCCCCACATA TATAGCTTAT CGGTTGATCA
2101 TCACTCCAGG GACATTTAAA GAGCGTATTA TTAAAAGTAT TACCCCAGAA ACACCAACAG
2161 AAATTCCTTG TGGGGACATC CGCTTGAATG CTGTGTAACA CACTCACCGA GAGGAAAAAG
2221 GCTTCTCCAC AACCTCCTCC TCCAGTTCTG ATGAGGCACG CCTGCCTTCT CCCCTCCAAG
2281 TGAATGAGTT TCCAGCTAAG CCTGATGATG GAAGGGCCTT CTCCACAGGG ACACAGTCTG
2341 GTGCCCAGAC TCAAGGCCTC CAGCCACTTA TTTCCATGGA TTCCCCTGGA CATATTCCCA
2401 TGGTAGACTG TGACACAGCT GAGCTGGCCT ATTTTGGACG TGTGAGGATG TGGATGGAGG
2461 TGATGAAAAC CACCCTATCA TCAGTTAGGA TTAGGTTTAG AATCAAGTCT GTGAAAGTCT
2521 CCTGTATCAT TTCTTGGTAT GATCATTGGT ATCTGATATC TGTTTGCTTC TAAAGGTTTC
2581 ACTGTTCATG AATACGTAAA CTGCGTAGGA GAGAACAGGG ATGCTATCTC GCTAGCCATA
2641 TATTTTCTGA GTAGCATATA GAATTTTATT GCTGGAATCT ACTAGAACCT TCTAATCCAT
2701 GTGCTGCTGT GGCATCAGGA AAGGAAGATG TAAGAAGCTA AAATGAAAAA TAGTGTGTCC
2761 ATGCAAGCTT GTGAGTCTGT GTATATTGTT GTTTCAGTGT ATTCTTATCT CTAGTCCAAT
2821 ATTTTGGGCC CATTACAAAT ATATGAATTC CCCAAATTTT TCTTACATTA ACAAATTCTA
2881 CCAACTCAA
```

SEQ ID. NO. 3

```
LOCUS       hVIP2.DOC    3944 BP DS-DNA              ENTERED   11/23/98
BASE COUNT      795 A   1159 C   1154 G    836 T      0 OTHER
COMMENT       1 -  238 exon  1
            239 -  338 exon  2
            339 -  446 exon  3
            447 -  544 exon  4
            545 -  642 exon  5
            643 -  784 exon  6
            785 -  935 exon  7
            936 -  996 exon  8
            997 - 1066 exon  9
           1067 - 1158 exon 10
           1159 - 1288 exon 11
           1289 - 1330 exon 12
           1331 - 3944 exon 13
            188 - 1504 open reading frame encoding VIP2 receptor
ORIGIN      -
      1 GTGCATTGAG CGCGCTCCAG CTGCCGGGAC GGAGGGGGCG GCCCCCGCGC TCGGGGCGCT
     61 CGGCTACAGC TGCGGGGCCC GAGGTCTCCG CGCACTCGCT CCCGGCCCAT GCTGGAGGCG
    121 GCGGAACCGC GGGGACCTAG GACGGAGGCG GCGGGCGCTG GGCGGCCCCC GGCACGCTGA
    181 GCTCGGGATG CGGACGCTGC TGCCTCCCGC GCTGCTGACC TGCTGGCTGC TCGCCCCCGT
```

241 GAACAGCATT CACCCAGAAT GCCGATTTCA TCTGGAAATA CAGGAGGAAG AAACAAAATG
301 TGCAGAGCTT CTGAGGTCTC AAACAGAAAA ACACAAAGCC TGCAGTGGCG TCTGGGACAA
361 CATCACGTGC TGGCGGCCTG CCAATGTGGG AGAGACCGTC ACGGTGCCCT GCCCAAAAGT
421 CTTCAGCAAT TTTTACAGCA AAGCAGGAAA CATAAGCAAA AACTGTACGA GTGACGGATG
481 GTCAGAGACG TTCCCAGATT TCGTCGATGC CTGTGGCTAC AGCGACCCGG AGGATGAGAG
541 CAAGATCACG TTTTATATTC TGGTGAAGGC CATTTATACC CTGGGCTACA GTGTCTCTCT
601 GATGTCTCTT GCAACAGGAA GCATAATTCT GTGCCTCTTC AGGAAGCTGC ACTGCACCAG
661 GAATTACATC CACCTGAACC TGTTCCTGTC CTTCATCCTG AGAGCCATCT CAGTGCTGGT
721 CAAGGACGAC GTTCTCTACT CCAGCTCTGG CACGTTGCAC TGCCCTGACC AGCCATCCTC
781 CTGGGTGGGC TGCAAGCTGA GCCTGGTCTT CCTGCAGTAC TGCATCATGG CCAACTTCTT
841 CTGGCTGCTG GTGGAGGGGC TCTACCTCCA CACCCTCCTG GTGGCCATGC TCCCCCCTAG
901 AAGGTGCTTC CTGGCCTACC TCCTGATCGG ATGGGGCCTC CCCACCGTCT GCATCGGTGC
961 ATGGACTGCG GCCAGGCTCT ACTTAGAAGA CACCGGTTGC TGGGATACAA ACGACCACAG
1021 TGTGCCCTGG TGGGTCATAC GAATACCGAT TTTAATTTCC ATCATCGTCA ATTTTGTCCT
1081 TTTCATTAGT ATTATACGAA TTTTGCTGCA GAAGTTAACA TCCCCAGATG TCGGCGGCAA
1141 CGACCAGTCT CAGTACAAGA GGCTGGCCAA GTCCACGCTC CTGCTTATCC CGCTGTTCGG
1201 CGTCCACTAC ATGGTGTTTG CCGTGTTTCC CATCAGCATC TCCTCCAAAT ACCAGATACT
1261 GTTTGAGCTG TGCCTCGGGT CGTTCCAGGG CCTGGTGGTG GCCGTCCTCT ACTGTTTCCT
1321 GAACAGTGAG GTGCAGTGCG AGCTGAAGCG AAAATGGCGA AGCCGGTGCC CGACCCCGTC
1381 CGCGAGCCGG GATTACAGGG TCTGCGGTTC CTCCTTCTCC CGCAACGGCT CGGAGGGCGC
1441 CCTGCAGTTC CACCGCGGCT CCCGCGCCCA GTCCTTCCTG CAAACGGAGA CCTCGGTCAT
1501 CTAGCCCCAC CCCTGCCTGT CGGACGCGGC GGGAGGCCCA CGGTTCGGGG CTTCTGCGGG
1561 GCTGAGACGC CGGCTTCCTC CTTCCAGATG CCCGAGCACC GTGTCGGGCA GGTCAGCGCG
1621 GTCCTGACTC CGTCAAGCTG GTTGTCCACT AAACCCCATA CCTGGAATTG GAGTCGTGTT
1681 GTCATTGACT CGATTTAAAC TCCAGCATTT AGATAATCTT GTGCAAAATG TGTTTCAGCC
1741 GTATAGTGGA TCCACTTTTT TTTTTTTTTT TTTTTGAGAC GGAGTCTTGC TCTGTCGCCC
1801 AGGCTGGAGT GCAGTGGCCT GATCTCTGCT CCCTGCAAGC TCCGCCTCCC GGGTTCACGC
1861 CATTCTCCTG CCTCAGCCTC CCATAGCTGG GACTACAGGC GCCCGCCAAC ACGCCTGGCT
1921 AATTTTTTGT ATTTTTAGTA GAGACAGGGT TTCACCATGT TAGCCAGGAT GGTCTCGATC
1981 TCCTGACCTC GTGATGGGCC CGCCTCGGCC TCCCAAAGTG CTGGGATTAA GGCGTGAGCC
2041 ACTGCGCCCG GCCCAAGAGA ATAGGGGAGC CAAGGAGGAA ATGTGGAAAC GCAGTTGTGT
2101 GGCCCAGCAC GAGCCTGGGC GACCACCGGG TGACATCCGT CCCACATCAG GGCGGCCTCC
2161 CAGGTCCCAT AAGGGTAGCC CCCTCATCTG CAGGACAGAG GGAAGCCAGT CAGGGCCCCC
2221 CCGGACGTTA GGACCAGGAG AAATCAACAG GAGGGCAGCC CGTCCTCTCT CTTGGGGCGC
2281 CCACCCGGCC CGGCTGAGCC CTGCCCCACC CAACTCCACA GGGCTGTTTT GCCTCCCCAC
2341 GGAAGGCGGG CTGAGGAGAC AACCAGATCA GGAGAGCAAG GTCATGAAGG AGGGGACCTC
2401 TCCACACAGG TGTTCCGTGG GACCCTCAGC AGCTCTGGCT CTGCCTCAGG AGGTCACCTG
2461 CCGCCCTGTG GGAGCCGCAG AGCCTGACGC TCAGCCCCAG GCCAGCTGCG GCCAGGCCTG
2521 CGGGCCCCTG GTGATGGGGT TACGTGGGGT GCGGGATACA GCTGAGTGGG AACCGGAAAC
2581 CTATTCTCTT TTTAACAAAA ATAATCTTAG GATAAGAATT ATTTTAACAA CATATAAAAC
2641 TGTTTCAAGC CCTCCTCCCC AGAGCTGGCG CTCAGCAGCC CTAGCGGCTG CTCCTTCAGG
2701 CGAAGGGTGG TTTGCAGATG TGGGGAGGGT GTCTGGGGAC GTTGCTGAGC TGGCTGCAGA
2761 AGGGTGGGGA TATCAGGGCA CAGTCTCCAT GTGTGTGCCA AGCCCTGGCC CCCACAGCGC
2821 TCGATGGACC TCAGCAAGCT GCCCAGCCCT GGCCCAGGTG CCCCGACTGT GGGACTCAGT
2881 TGTTCTGAGC ACATTTGACT CCACTTTTCC TTTAAAAATG AATGTCTTGT TCCTGTGCAT
2941 TGGTGGCATC ACAGACCCCA GCTGGGGCGC GATGTCAAAG GTCGGGACAG CTGTGCCGGG
3001 AGGCAGCCAC AGGGAAGCTC ACACATCCTG TCAGTGTCAC CTTGGTTTGC AAAACCCATA
3061 TCCCCGGTAA AATGAGGCCG GACAGAGGGG CTGTTAGGAC AGCAAAGCAG CAGTGTCCAG
3121 AGACCCCTCA ATCCCCAAAG GTCCGCACCC TGTCCTGCAC ACCCTGGGCC ACGCCGGCCA
3181 CACCCCTCTG CTGCAACAAG CTCATCCCTG GACTTCTGGG AGAATGAACC CGAGGTTGGT
3241 TTGGGGAGAC AGGTGAGGCG GTTGGATCTA CAGAACAACC CACCATTTCT GGGGGCCGCA
3301 GAGGATCCAT CACAGACGGA TACTGGGGAG TAAACGGCCC AGGCCAGGTG CCCAGGAAAG
3361 GACGGCTGAG CATGTGGAGC GAGAGGGAGG CAGGTGGACG CTGCAGACCC CAGGTTCAGT
3421 GCGGCCCCTC GGCTGTTCCT CCCCTGTAGG GTTTGGACAG ACCCACCCCC AGCCTTGCCC
3481 AGCTTTCAAA GGACAAAAGG GAGCATCCCC CACCTACTCT CAGGTTTTTG AGGAAACAAA
3541 GATTTGTGGT AACTGAAGGT GTTGGGTCAG TGGCCAGGTG CCGACACTGA GCTGTGACCC
3601 AGAGGGGACG CTGAGGAAGT GGGCGTGAGT GGACATGTCA GGTGGTTACC AGGCACTGGT
3661 TGTTGATGGT CGGTGGTTGG GTGTGGGCAG TCATCAGTCA TCAGGTGTGC TCAGGGGACA
3721 ATCTCCCCTC AACCGCACAT GTGCCACTGT TCAGCGGAGC TGACTGGTTT CTCCTGGTAG
3781 AGGGCCGGCT GTATCCTGAC AGATGCCTGG TGAGCAGGGG AAGCAGGACC CAGTGGTCAA
3841 CAGGTGTCTT TAACTGTCAT TGTGTGTGGA ATGTCGCAGA CTCCTCCACG TGGCGGGAAT
3901 GAGCTGTGTA AATACTTCAA TAAAGCCTGA TCTCACATCT GCAA

SEQ ID NO. 4

GAATTCCGCC CCTCTCCCTC CCCCCCCCCT AACGTTACTG GCCGAAGCCG
CTTGGAATAA GGCCGGTGTG CGTTTGTCTA TATGTTATTT TCCACCATAT
TGCCGTCTTT TGGCAATGTG AGGGCCCGGA AACCTGGCCC TGTCTTCTTG
ACGAGCATTC CTAGGGGTCT TTCCCCTCTC GCCAAAGGAA TGCAAGGTCT
GTTGAATGTC GTGAAGGAAG CAGTTCCTCT GGAAGCTTCT TGAAGACAAA
CAACGTCTGT AGCGACCCTT TGCAGGCAGC GGAACCCCCC ACCTGGCGAC
AGGTGCCTCT GCGGCCAAAA GCCACGTGTA TAAGATACAC CTGCAAAGGC
GGCACAACCC CAGTGCCACG TTGTGAGTTG GATAGTTGTG GAAAGAGTCA
AATGGCTCTC CTCAAGCGTA TTCAACAAGG GGCTGAAGGA TGCCCAGAAG
GTACCCCATT GTATGGGATC TGATCTGGGG CCTCGGTGCA CATGCTTTAC
ATGTGTTTAG TCGAGGTTAA AAAACGTCTA GGCCCCCCGA ACCACGGGGA
CGTGGTTTTC CTTTGAAAAA CACGATGATA AGCTTGCCAC AACCATG

(this sequence is also presented as Figure 10).

Applicant's or agent's file reference number	P/3330.WO CTH	International applicati

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 27, line 15

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St Machar Drive
Aberdeen
AB2 1RY
United Kingdom

Date of deposit	Accession Number
24 November 1997	NCIMB 40907

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will only be made available either until the publication of the mention of the grant of the patent or after twenty years from the date of filing if the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC)

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application 27 NOVEMBER 1998	☐ This sheet was received by the International Bureau on:
Authorized officer P. J. TREEN PJTreen 3.12.98	Authorized officer

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	P3330.WO CTH	International application

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 27, line 16

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet: ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St Machar Drive
Aberdeen
AB2 1RY
United Kingdom

Date of deposit	Accession Number
24 November 1997	NCIMB 40908

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet: ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will only be made available either until the publication of the mention of the grant of the patent or after twenty years from the date of filing if the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC)

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application 27 NOVEMBER 1998	☐ This sheet was received by the International Bureau on:
Authorized officer: PJTreen 3.12.98	Authorized officer:

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	P/3330.WO CTH	International application

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 27, line 17

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St Machar Drive
Aberdeen
AB2 1RY
United Kingdom

Date of deposit	Accession Number
24 November 1997	NCIMB 40909

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will only be made available either until the publication of the mention of the grant of the patent or after twenty years from the date of filing if the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC)

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application 27 NOVEMBER 1998	☐ This sheet was received by the International Bureau on:
Authorized officer: DJTilden 3.12.98	Authorized officer

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	P/3330.WO CTH	International application

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 27, line 18

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St Machar Drive
Aberdeen
AB2 1RY
United Kingdom

Date of deposit	Accession Number
24 November 1997	NCIMB 40910

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will only be made available either until the publication of the mention of the grant of the patent or after twenty years from the date of filing if the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC)

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application 27 NOVEMBER 1998	☐ This sheet was received by the International Bureau on:
Authorized officer PJTreen 3.12.98	Authorized officer

Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	P/3330.WO CTH	International application

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 27, line 19

B. IDENTIFICATION OF DEPOSIT — Further deposits are identified on an additional sheet ☐

Name of depositary institution

The National Collections of Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St Machar Drive
Aberdeen
AB2 1RY
United Kingdom

Date of deposit	Accession Number
24 November 1997	NCIMB 40906

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* — This information is continued on an additional sheet ☐

In respect of those designations in which a European patent is sought, and any other designated state having equivalent legislation, a sample of the deposited microorganism will only be made available either until the publication of the mention of the grant of the patent or after twenty years from the date of filing if the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC)

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only

☑ This sheet was received with the international application
27 NOVEMBER 1998

Authorized officer: PJTeeten 3.12.98

For International Bureau use only

☐ This sheet was received by the International Bureau on:

Authorized officer:

Form PCT/RO/134 (July 1992)

CLAIMS

1. A YAC vector comprising an IRES.

2. A YAC vector comprising a reporter gene wherein the expression product of the reporter gene is capable of producing a visually detectable signal.

3. A YAC vector comprising a reporter gene wherein the expression product of the reporter gene is capable of producing an immunologically detectable signal.

4. A YAC vector comprising an IRES and a reporter gene wherein the expression product of the reporter gene is capable of producing a visually detectable signal.

5. A YAC vector comprising an IRES and a reporter gene wherein the expression product of the reporter gene is capable of producing an immunologically detectable signal.

6. pYIV1.

7. pYIV2.

8. pYIV3.

9. pYIV4.

10. A YAC vector or a YAC comprising a selection gene, wherein that selection gene is specifically removable from the YAC vector or the YAC.

11. A YAC vector comprising a nucleotide sequence wherein the nucleotide sequence comprises the sequence presented as SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof.

12. A vector capable of modifying a YAC or a YAC vector wherein the vector comprises a nucleotide sequence wherein the nucleotide sequence comprises the sequence presented as SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof.

13. pYAM4.

14. A YAC prepared by one or more of the vectors according to any one of the preceding claims.

15. Use of an IRES to modify a YAC vector or a YAC.

16. Use of a nucleotide sequence comprising the sequence presented SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof in a vector to prepare a YAC vector or a YAC.

17. Use of a nucleotide sequence comprising the sequence presented as SEQ ID No.1 or SEQ ID No. 4 or a variant, homologue or derivative thereof in a vector to increase the expression efficiency of one or more NOIs within a YAC vector or a YAC.

18. A YAC transgenic mammal co-expressing an NOI and a reporter gene wherein the expression pattern of the NOI can be determined by measuring a detectable signal (such as a visually or an immunologically detectable signal) produced by the expression product of the reporter gene.

19. A YAC transgenic mammal expressing a reporter gene under the control of a regulatory sequence from a human NOI.

20. Use of a YAC transgenic mammal to test for potential pharmaceutical and/or veterinary agents.

21. An assay method for identifying an agent that can affect the expression pattern of an NOI or the EP activity thereof,

the assay method comprising

administering an agent to a YAC transgenic mammal as defined in claim 18;

determining whether the agent modulates (such as affects the expression pattern or activity) of the NOI or the EP by means of the detectable signal.

22. An assay method according to claim 21 wherein the assay is to screen for agents useful in the treatment of disturbances in any one of: circadian function, sleep disorders, eating disorders, pre-menstural syndrome, autoimmune disorders, birth defects in women and/or sexual dysfunction.

23. An agent identified by the method of claim 22.

24. A process comprising the steps of:

(a) performing the assay according to claim 21 or claim 22;

(b) identifying one or more agents that affect the expression pattern of the NOI or the EP activity thereof;

(c) preparing a quantity of those one or more identified agents.

25. A process comprising the steps of:

(a) performing the assay according to claim 21 or claim 22;

(b) identifying one or more agents that affect the expression pattern of the NOI or the EP activity thereof;

(c) preparing a pharmaceutical composition comprising one or more identified agents.

26. A process comprising the steps of:

(a) performing the assay according to claim 21 or claim 22;

(b) identifying one or more agents that affect the expression pattern of the NOI or the EP activity thereof;

(c) modifying one or more identified agents to cause a different effect on the expression pattern of the NOI or the EP activity thereof.

27. Use of a YAC according to claim 2 or claim 3 or claim 4 or claim 5 or any one of the vectors described in claims 6 to 9 or claim 13 (including combinations thereof) to screen for agents capable of affecting the expression pattern of an NOI or the EP activity thereof in a transgenic mammal.

28. Use of an agent in the preparation of a pharmaceutical composition for the treatment of a disorder or condition associated with the expression pattern of an NOI or the EP activity thereof, the agent having an effect on the expression pattern of the NOI or the EP activity thereof when assayed *in vitro* by the assay method according to claim 21 or claim 22.

29. Use of an agent identified by an assay according to claim 21 or claim 22 in the manufacture of a medicament which affects the expression pattern of an NOI or the EP activity thereof.

30. Use of an agent in the preparation of a pharmaceutical composition for the treatment of a disorder or condition associated with the expression pattern of an NOI or the EP activity thereof, the agent having an effect on the expression pattern of the NOI

or the EP activity thereof when assayed *in vitro* by the assay method accoding to claim 21 or 22.

31. Use of an agent identified by an assay according to claim 21 or 22 in the manufacture of a medicament which affects the expression pattern of an NOI or the EP activity thereof.

32. A vector substantially as described herein and with reference to the accompanying Figures.

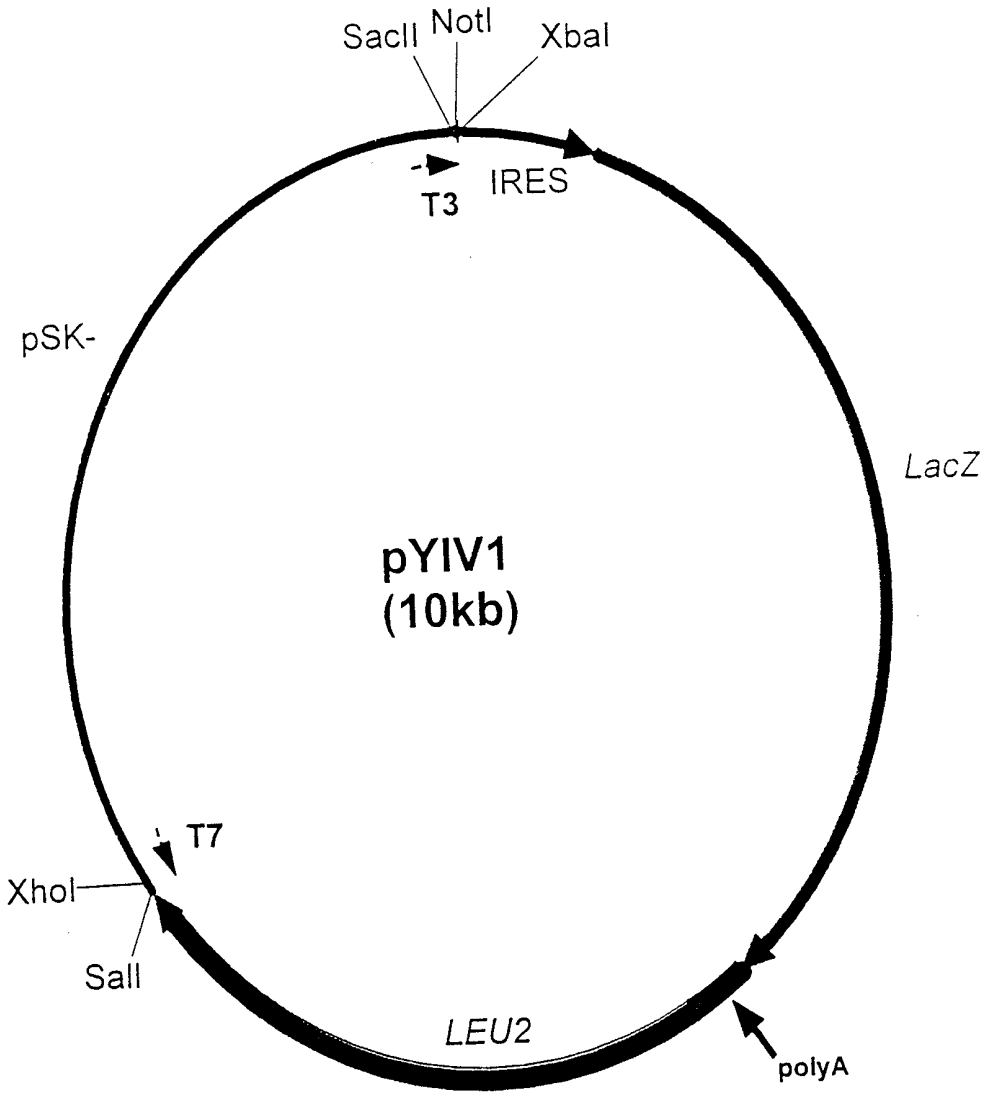

SacII
NotI
XbaI
T3
IRES
pSK-
LacZ
pYIV1
(10kb)
T7
XhoI
SalI
LEU2
polyA

FIGURE 1

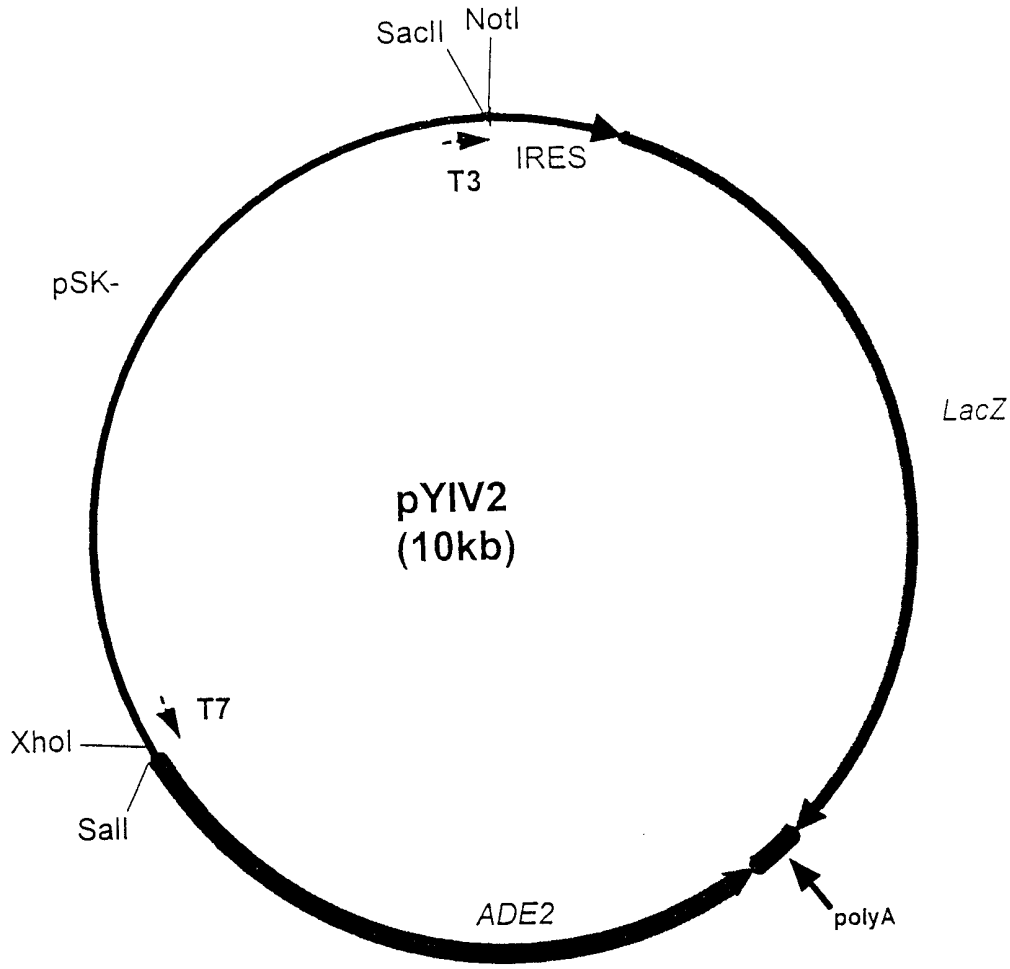

SacII
NotI
IRES
T3
pSK-
LacZ
pYIV2
(10kb)
T7
XhoI
SalI
ADE2
polyA

FIGURE 2

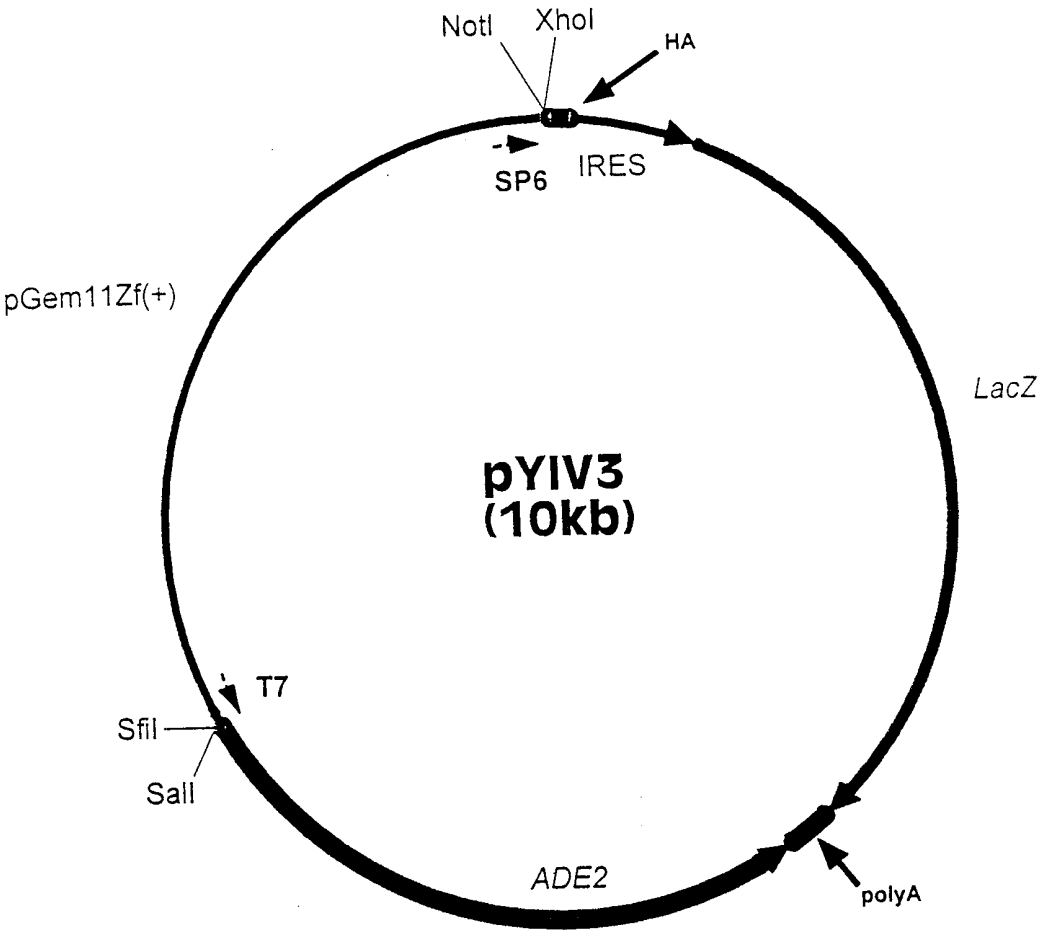

NotI
XhoI
HA
SP6
IRES
pGem11Zf(+)
LacZ
pYIV3
(10kb)
T7
SfiI
SalI
ADE2
polyA

FIGURE 3

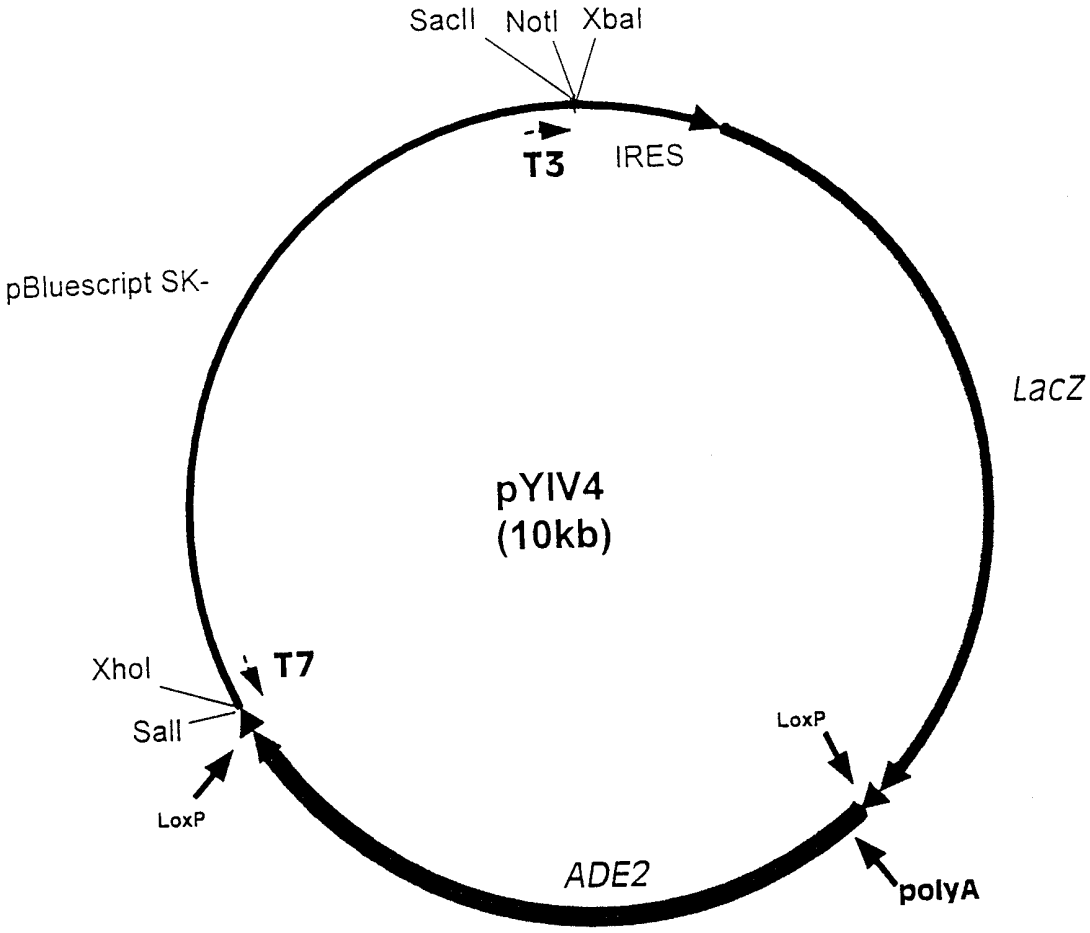

SacII
NotI
XbaI
T3
IRES
pBluescript SK-
LacZ
pYIV4
(10kb)
XhoI
SalI
T7
LoxP
LoxP
ADE2
polyA

FIGURE 4

Fig.5 X-gal staining of a 12.5-day transgenic embryo carrying the IGF2-IRES-lacZ-LEU2 (from pYIV1) under control of the wt1 promoter in a 480 kb YAC DNA .

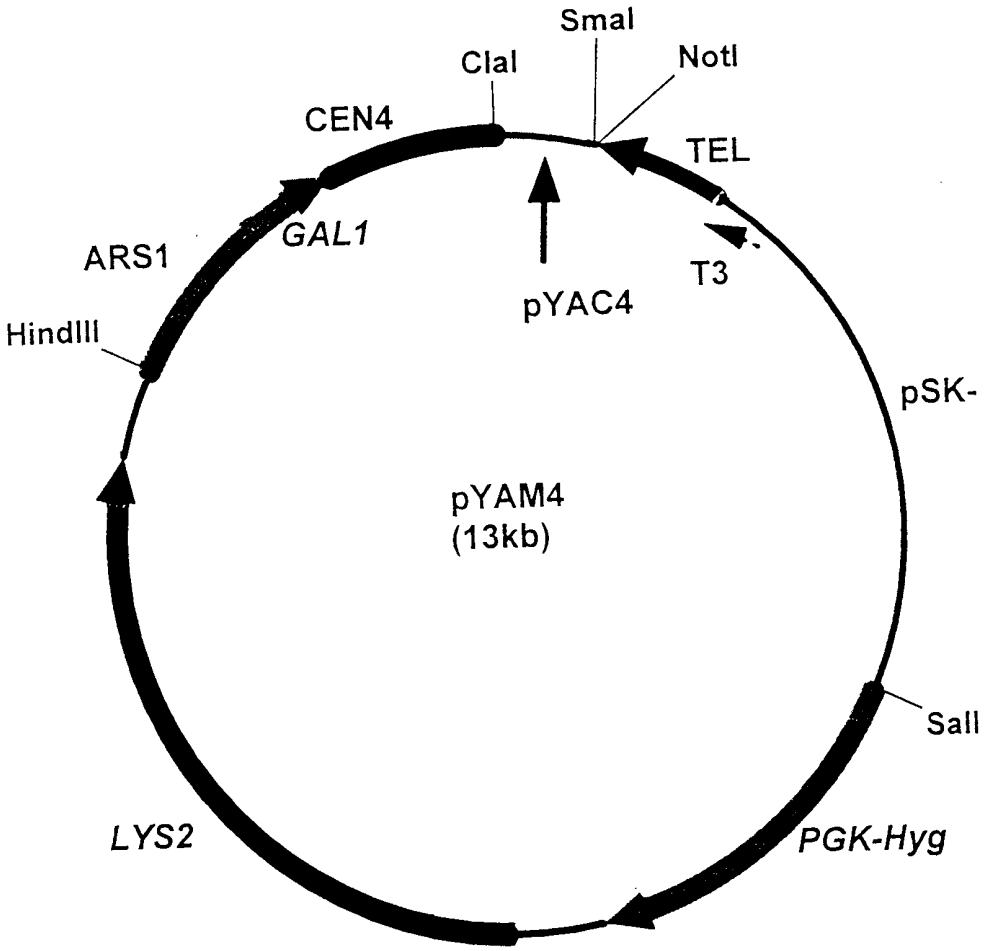

SmaI
ClaI
NotI
CEN4
TEL
GAL1
ARS1
T3
pYAC4
HindIII
pSK-
pYAM4
(13kb)
SalI
LYS2
PGK-Hyg

Figure 6

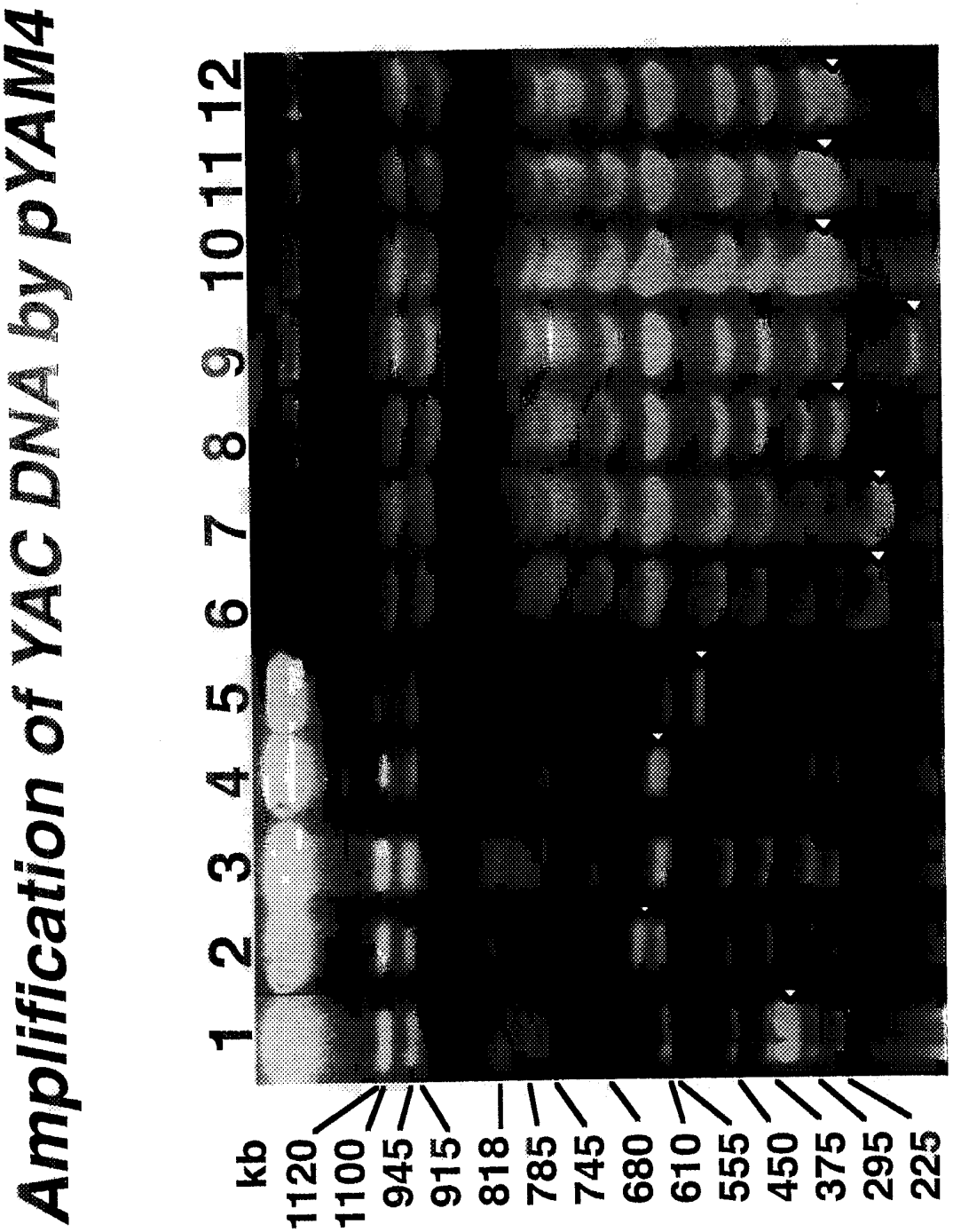

Amplification of YAC DNA by pYAM4
1 2 3 4 5 6 7 8 9 10 11 12
kb
1120
1100
945
915
818
785
745
680
610
555
450
375
295
225

Figure 7

Figure 8

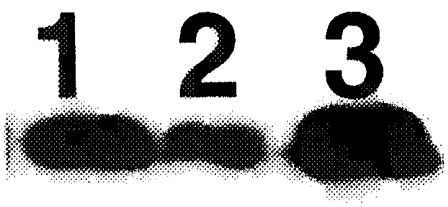


Figure 9

Figure 10

GAATTCCGCC CCTCTCCCTC CCCCCCCCCT AACGTTACTG GCCGAAGCCG
CTTGGAATAA GGCCGGTGTG CGTTTGTCTA TATGTTATTT TCCACCATAT
TGCCGTCTTT TGGCAATGTG AGGGCCCGGA AACCTGGCCC TGTCTTCTTG
ACGAGCATTC CTAGGGGTCT TTCCCCTCTC GCCAAAGGAA TGCAAGGTCT
GTTGAATGTC GTGAAGGAAG CAGTTCCTCT GGAAGCTTCT TGAAGACAAA
CAACGTCTGT AGCGACCCTT TGCAGGCAGC GGAACCCCCC ACCTGGCGAC
AGGTGCCTCT GCGGCCAAAA GCCACGTGTA TAAGATACAC CTGCAAAGGC
GGCACAACCC CAGTGCCACG TTGTGAGTTG GATAGTTGTG GAAAGAGTCA
AATGGCTCTC CTCAAGCGTA TTCAACAAGG GGCTGAAGGA TGCCCAGAAG
GTACCCCATT GTATGGGATC TGATCTGGGG CCTCGGTGCA CATGCTTTAC
ATGTGTTTAG TCGAGGTTAA AAAACGTCTA GGCCCCCCGA ACCACGGGGA
CGTGGTTTTC CTTTGAAAAA CACGATGATA AGCTTGCCAC AACCATG

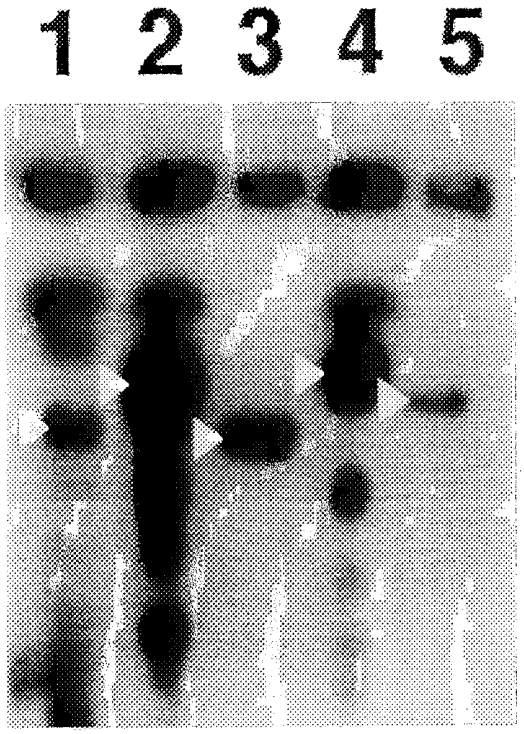


Figure 11

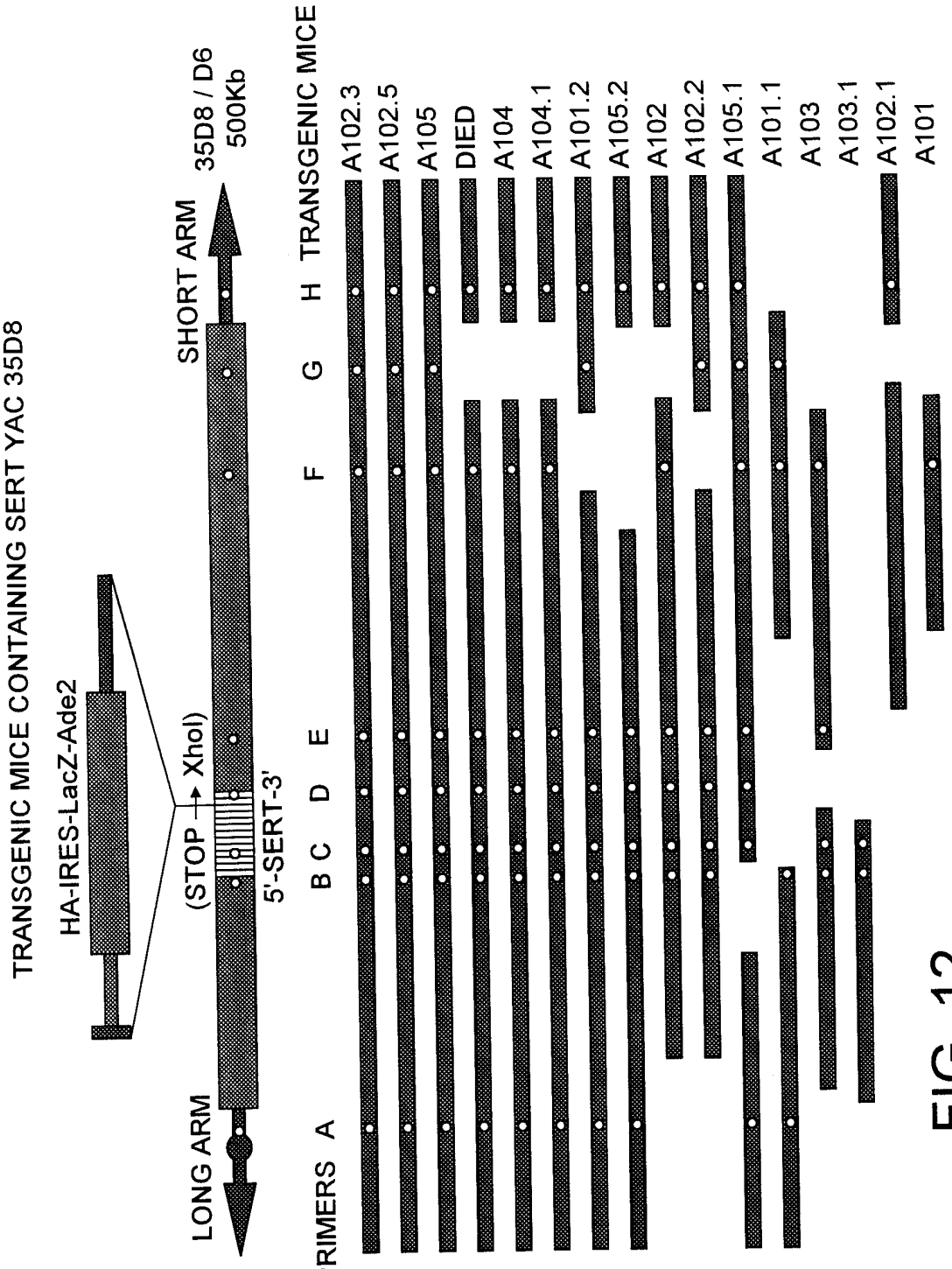

TRANSGENIC MICE CONTAINING SERT YAC 35D8
HA-IRES-LacZ-Ade2
(STOP → XhoI)
5'-SERT-3'
LONG ARM
SHORT ARM
35D8 / D6
500Kb
PRIMERS A B C D E F G H TRANSGENIC MICE
A102.3
A102.5
A105
DIED
A104
A104.1
A101.2
A105.2
A102
A102.2
A105.1
A101.1
A103
A103.1
A102.1
A101

FIG. 12

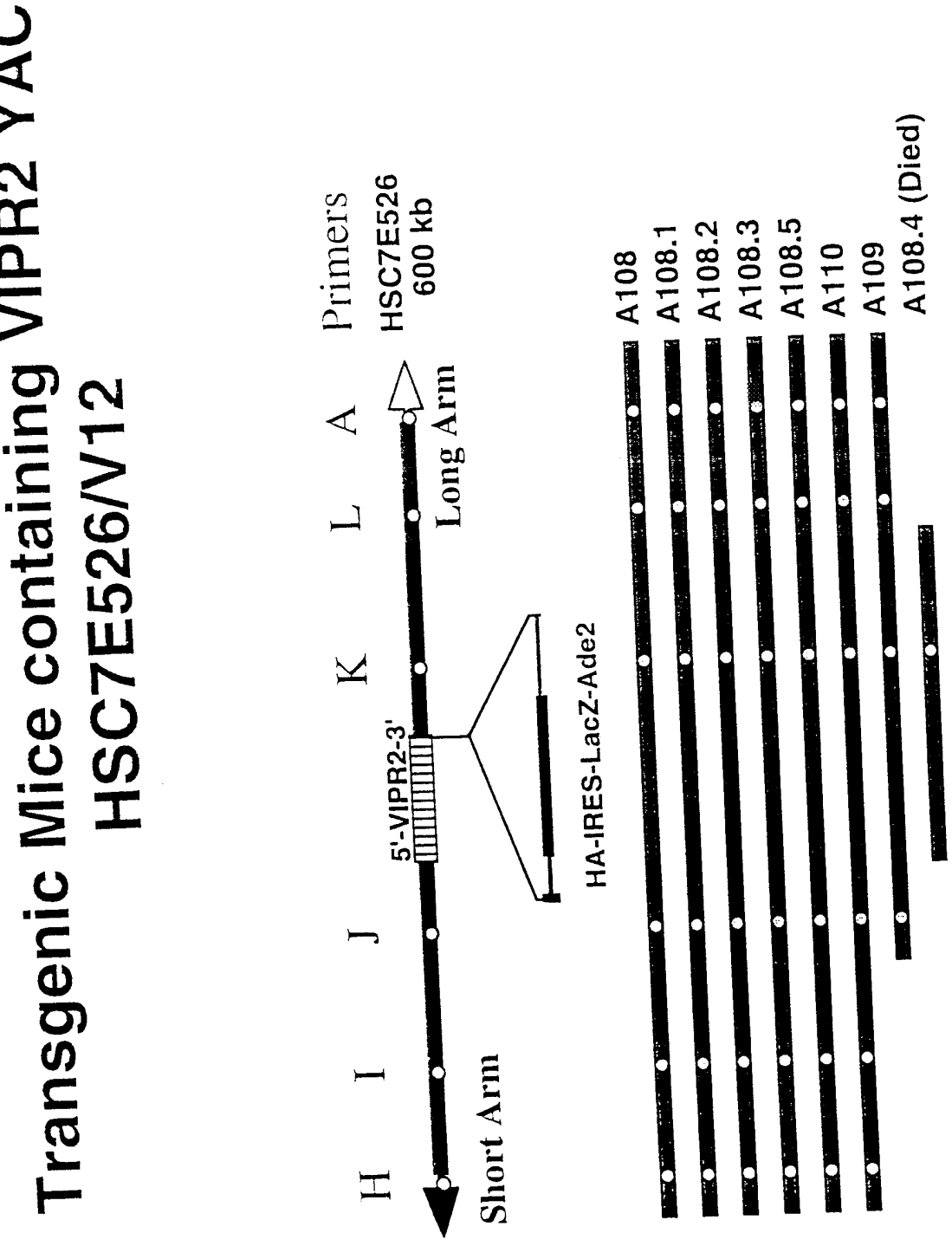
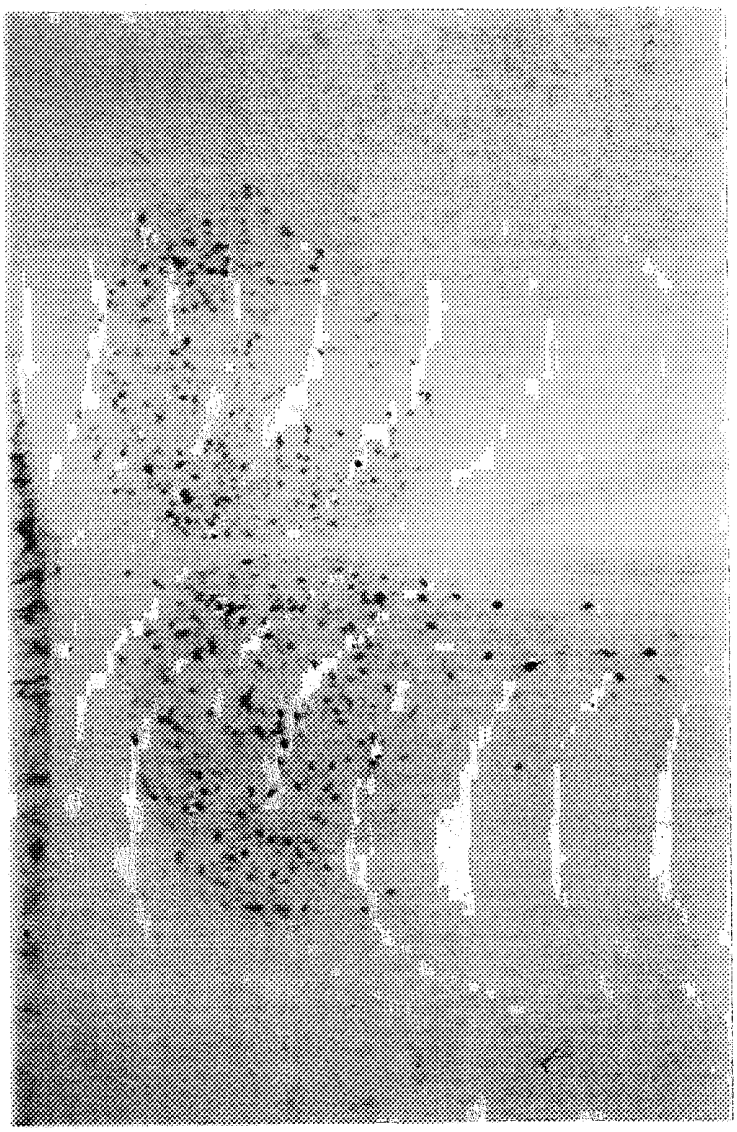

Transgenic Mice containing VIPR2 YAC
HSC7E526/V12
H
I
J
K
L
A
Primers
HSC7E526
600 kb
Long Arm
5'-VIPR2-3'
HA-IRES-LacZ-Ade2
Short Arm
A108
A108.1
A108.2
A108.3
A108.5
A110
A109
A108.4 (Died)

FIGURE 13

Figure 14



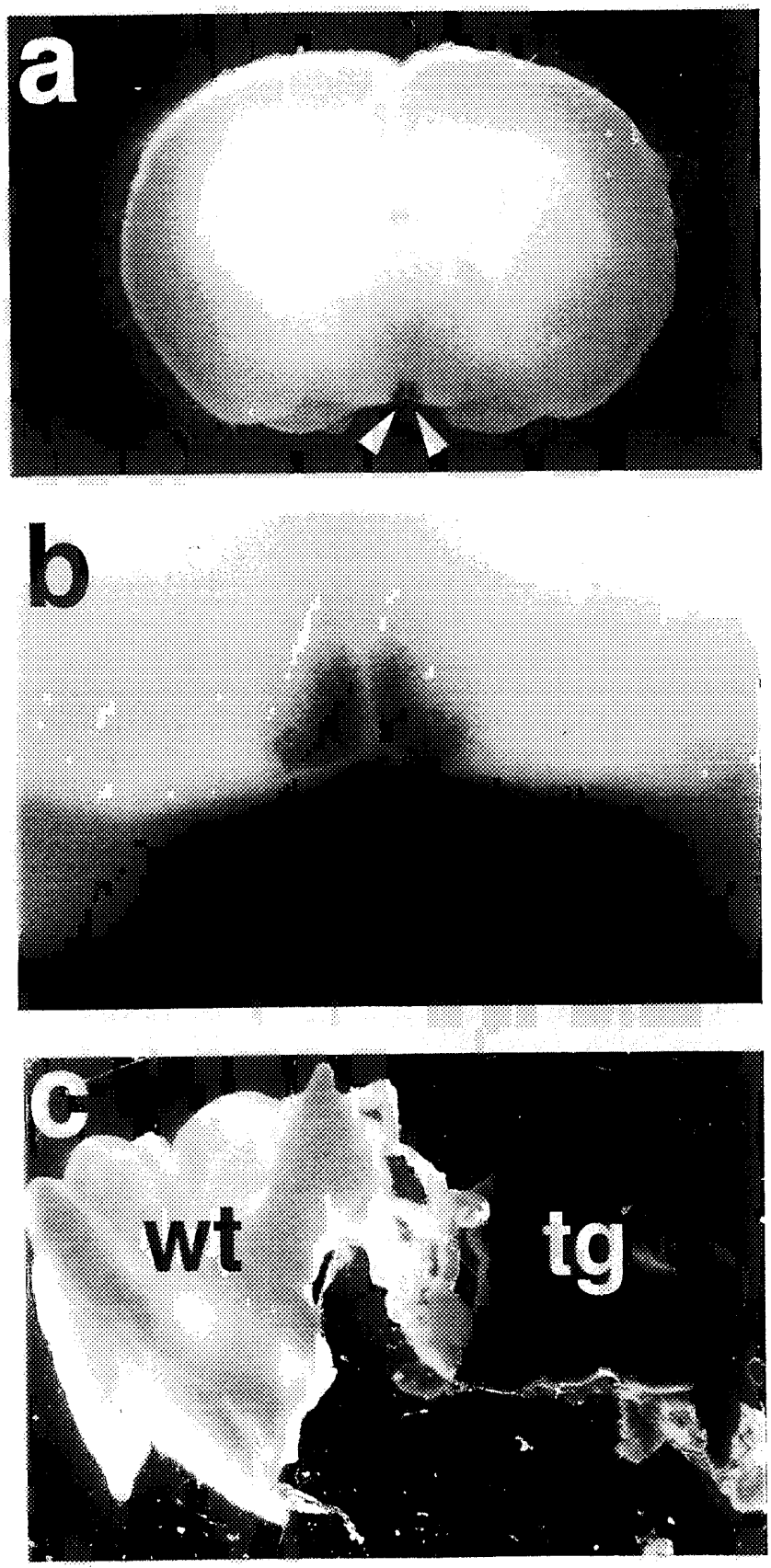


Figure 15

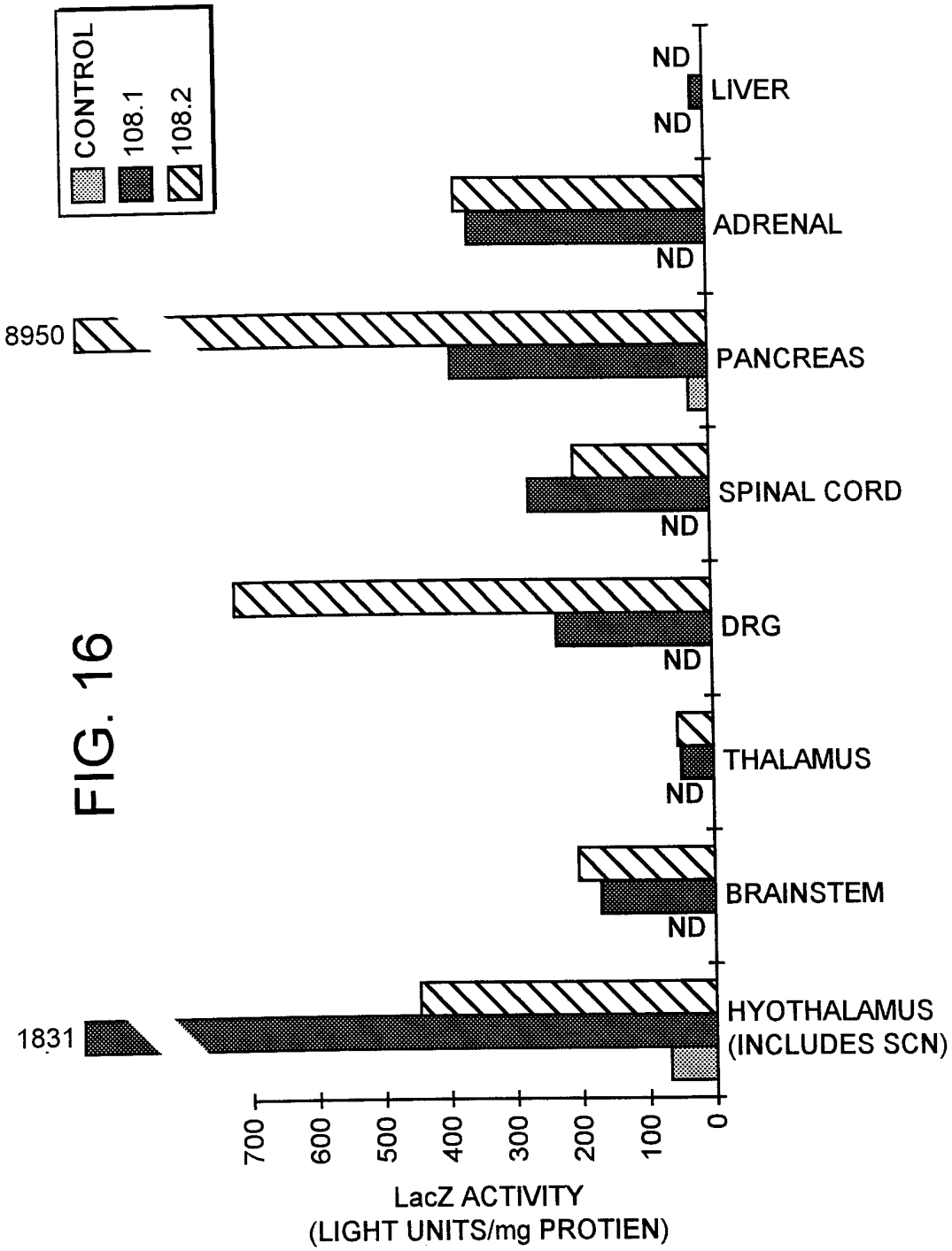

CONTROL
108.1
108.2
ND
LIVER
ND
ADRENAL
ND
8950
PANCREAS
SPINAL CORD
ND
DRG
ND
THALAMUS
ND
BRAINSTEM
ND
1831
HYOTHALAMUS
(INCLUDES SCN)
700
600
500
400
300
200
100
0
LacZ ACTIVITY
(LIGHT UNITS/mg PROTIEN)

FIG. 16

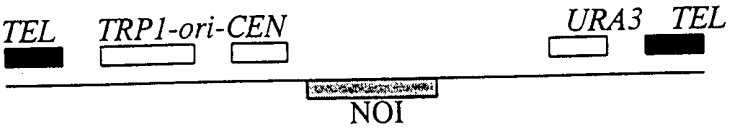



Figure 17