



US 20170051045A1

(19) **United States**

(12) **Patent Application Publication**
BRADLEY et al.

(10) **Pub. No.: US 2017/0051045 A1**
(43) **Pub. Date: Feb. 23, 2017**

(54) **ANIMAL MODELS AND THERAPEUTIC MOLECULES**

(71) Applicant: **Kymab Limited**, Cambridge (GB)

(72) Inventors: **Allan BRADLEY**, Cambridge (GB); **E-Chiang LEE**, Cambridge (GB); **Qi LIANG**, Cambridge (GB); **Wei WANG**, Cambridge (GB); **Glenn FRIEDRICH**, Cambridge (GB)

(21) Appl. No.: **15/251,969**

(22) Filed: **Aug. 30, 2016**

Related U.S. Application Data

(60) Continuation of application No. 14/137,902, filed on Dec. 20, 2013, now Pat. No. 9,434,782, which is a continuation of application No. 13/740,727, filed on Jan. 14, 2013, now Pat. No. 9,505,827, which is a division of application No. 13/310,431, filed on Dec. 2, 2011, which is a continuation-in-part of application No. PCT/GB2010/051122, filed on Jul. 7, 2010, said application No. 13/310,431 is a continuation-in-part of application No. PCT/GB2011/050019, filed on Jan. 7, 2011.

(60) Provisional application No. 61/223,960, filed on Jul. 8, 2009, provisional application No. 61/355,666, filed on Jun. 17, 2010.

(30) **Foreign Application Priority Data**

Jul. 8, 2009 (GB) 0911846.4
Jul. 28, 2009 (GB) 0913102.0

Publication Classification

(51) **Int. Cl.**
C07K 16/00 (2006.01)
A01K 67/027 (2006.01)
(52) **U.S. Cl.**
CPC **C07K 16/00** (2013.01); **A01K 67/0278** (2013.01); **C07K 2317/56** (2013.01); **C07K 2317/14** (2013.01); **C07K 2317/24** (2013.01); **C07K 2317/52** (2013.01); **A01K 2217/072** (2013.01); **A01K 2267/01** (2013.01); **A01K 2227/105** (2013.01)

(57) **ABSTRACT**

The invention discloses methods for the generation of chimaeric human—non-human antibodies and chimaeric antibody chains, antibodies and antibody chains so produced, and derivatives thereof including fully humanised antibodies; compositions comprising said antibodies, antibody chains and derivatives, as well as cells, non-human mammals and vectors, suitable for use in said methods.

Three basic vector backbones

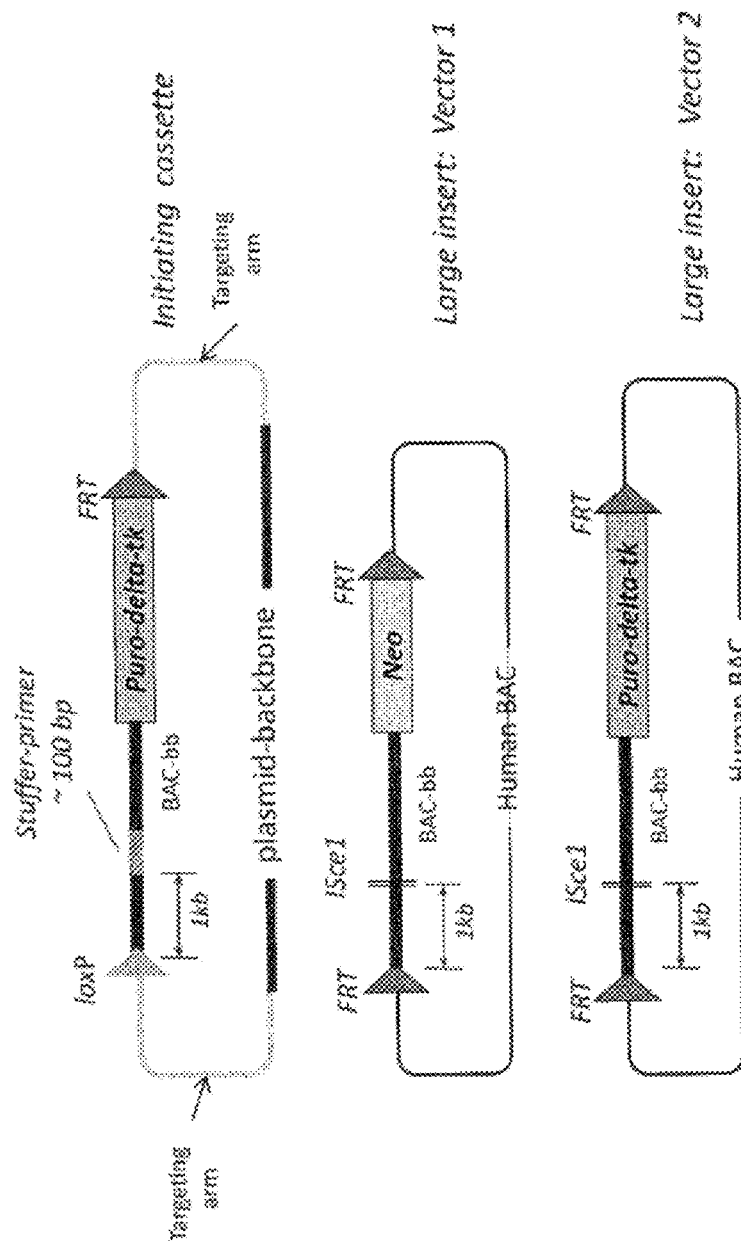
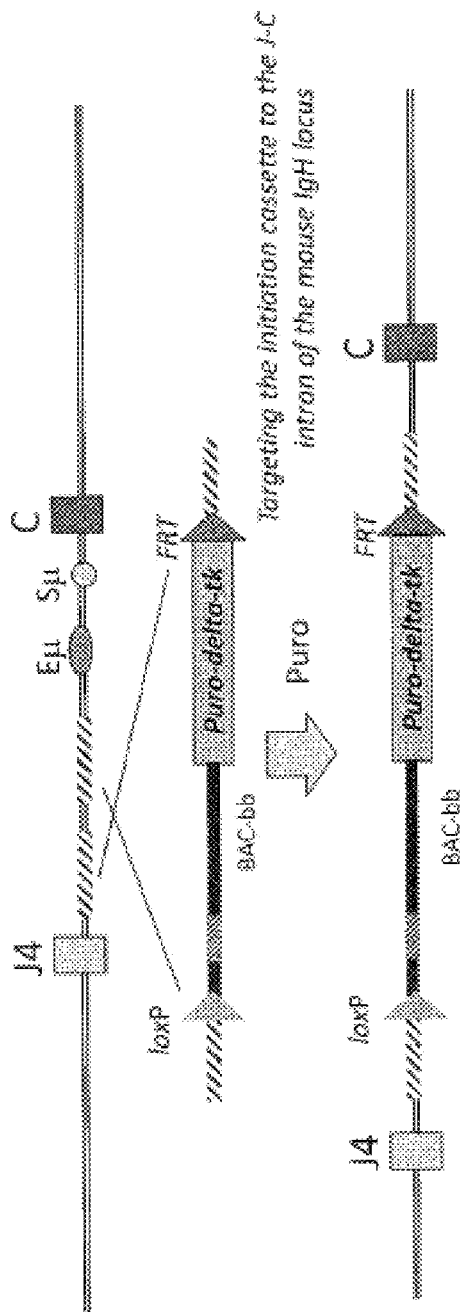


FIG. 1

Entry / Initiation vector: Step 1



Validation:

1. Standard gene-targeting LR-PCR/Southern
2. Make chimaeras & breed in parallel with next step

Fig. 2

Addition vector: Step 2

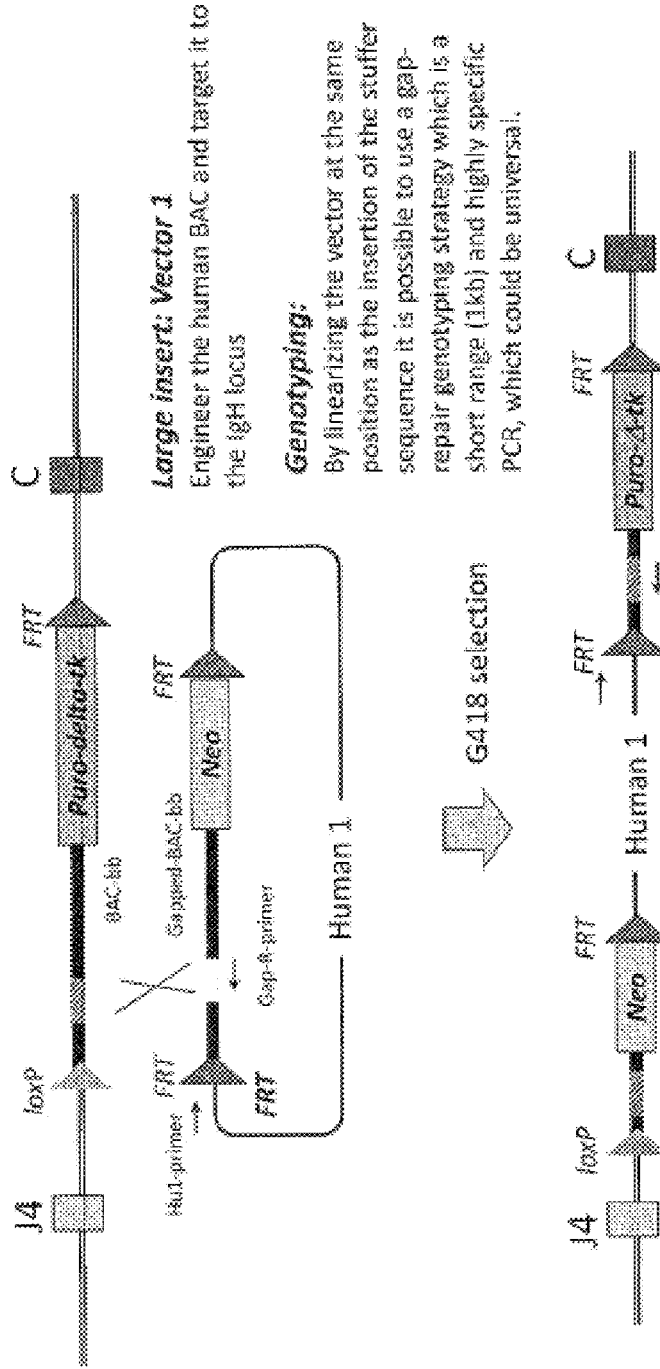


Fig. 3

Addition vector: Step 2

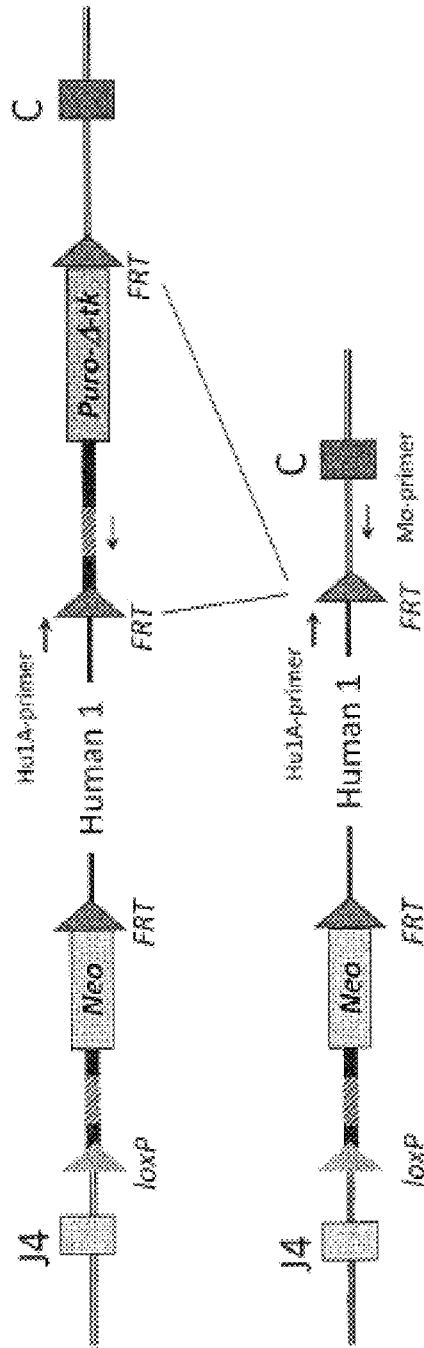


FIG. 4

Addition vector: Step 3A

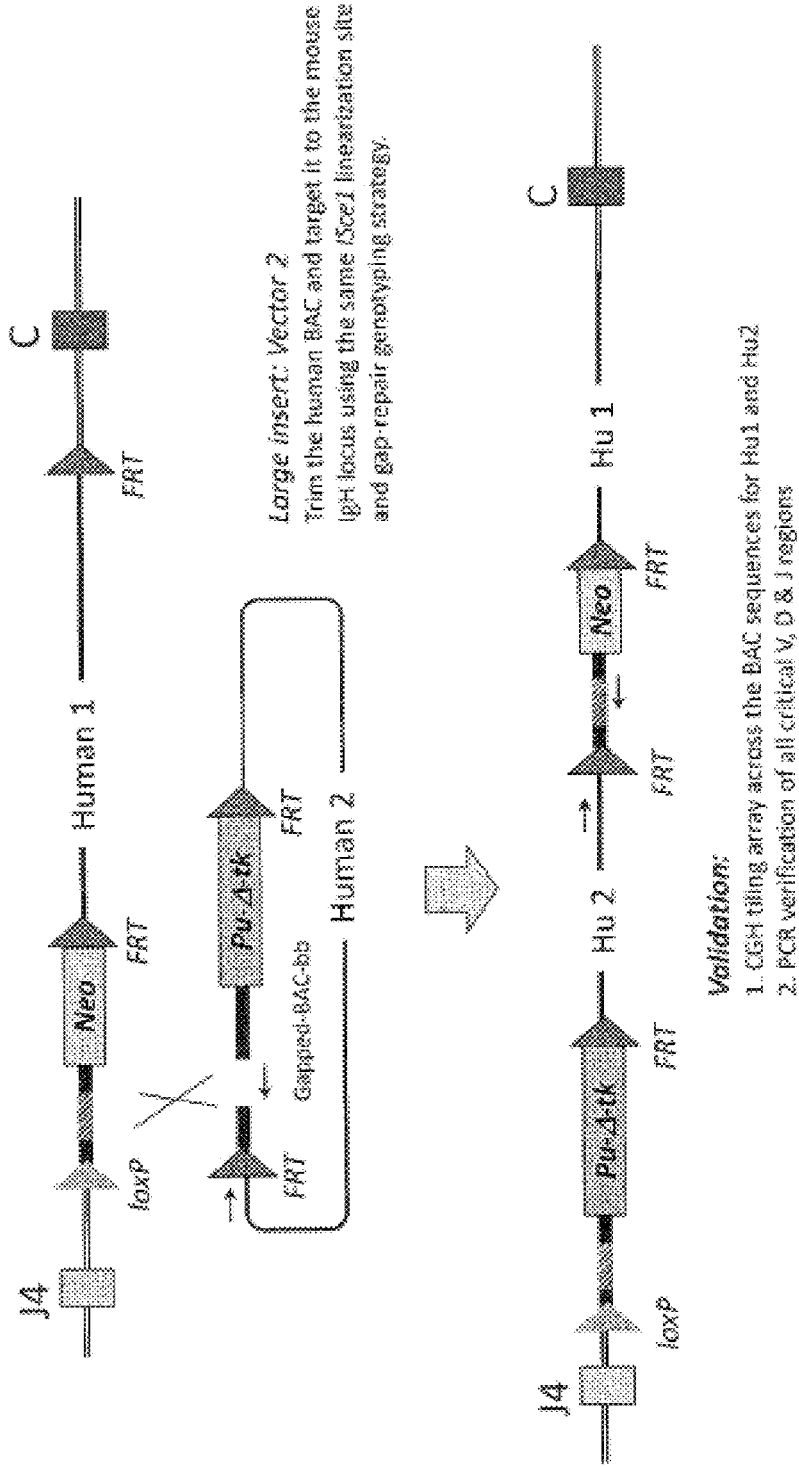


Fig. 5

Addition vector: Step 3B

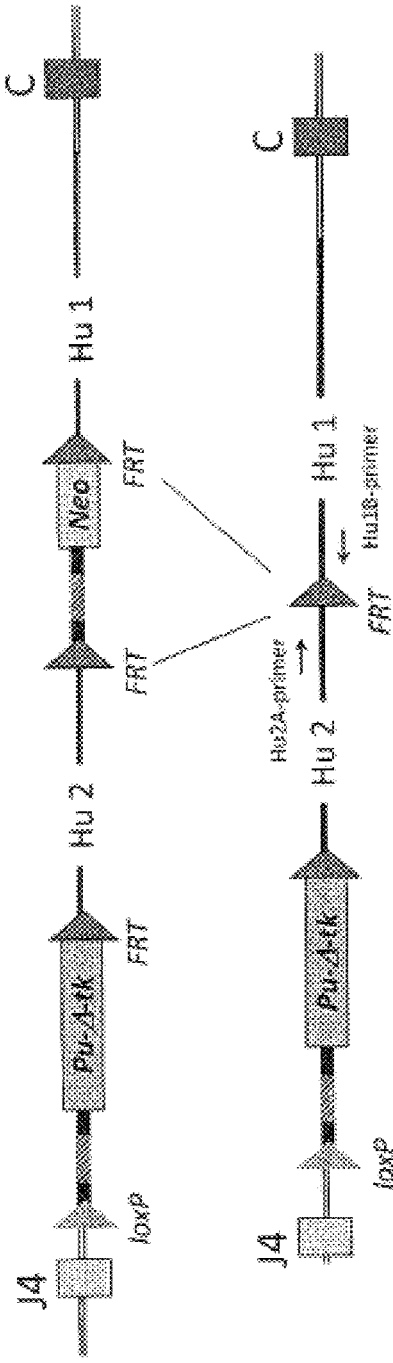


Fig. 6

Addition vector: Step 4

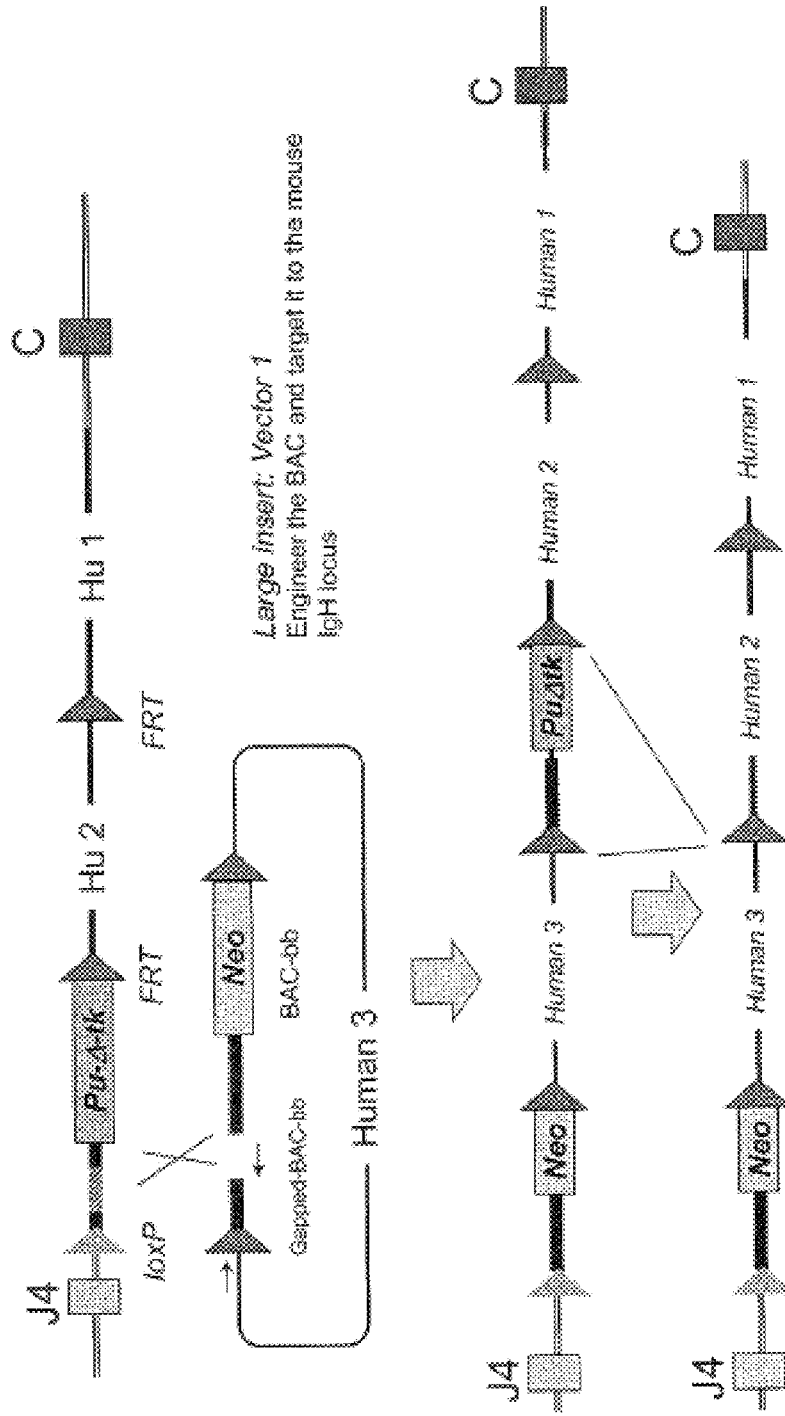


Fig. 7

A humanized mouse IgH locus

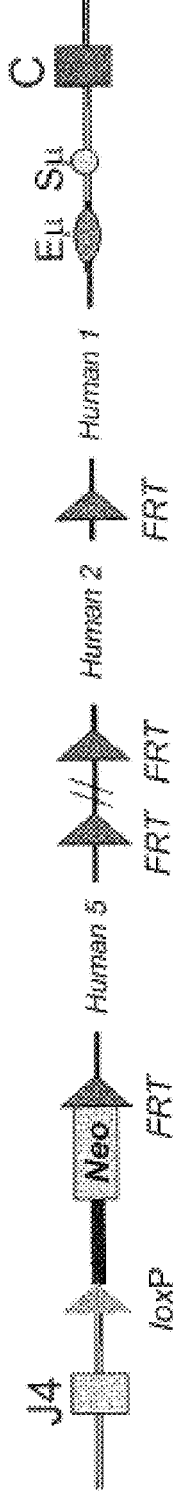


FIG. 8

BAC landing pad inserted at the IgH locus

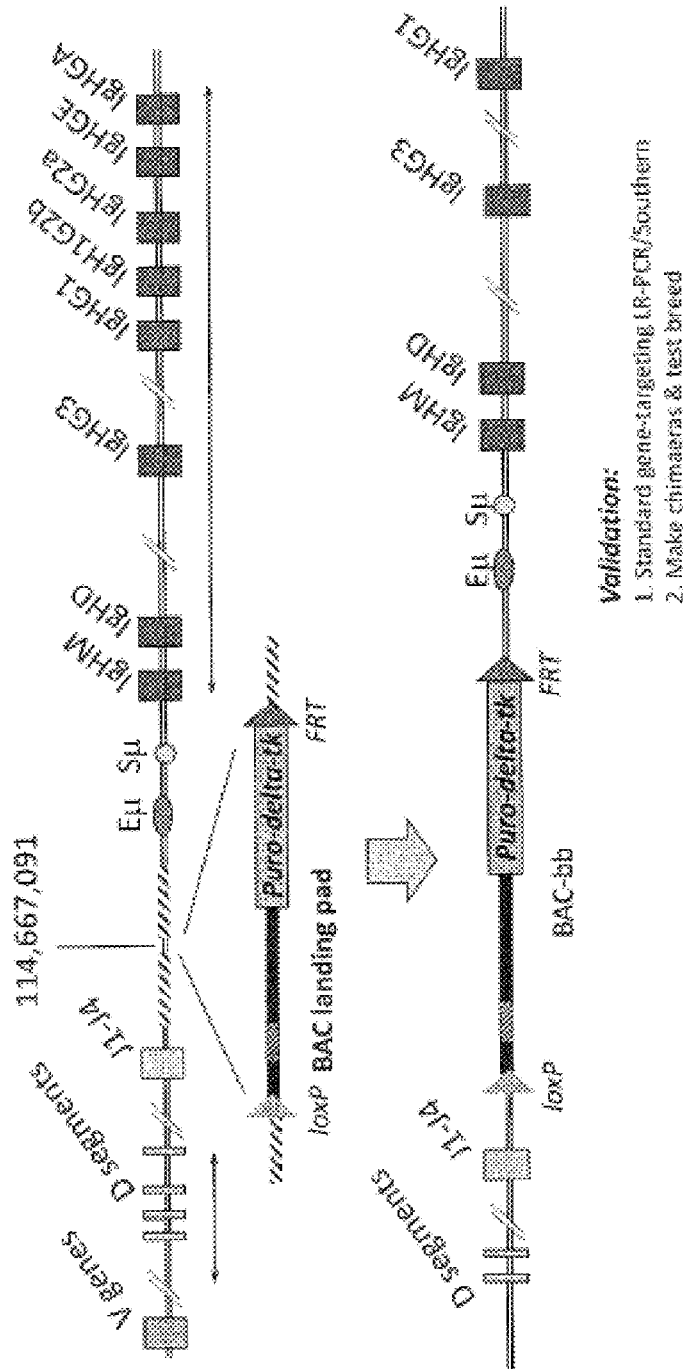


Fig. 9

1st BAC insertion – Step 1

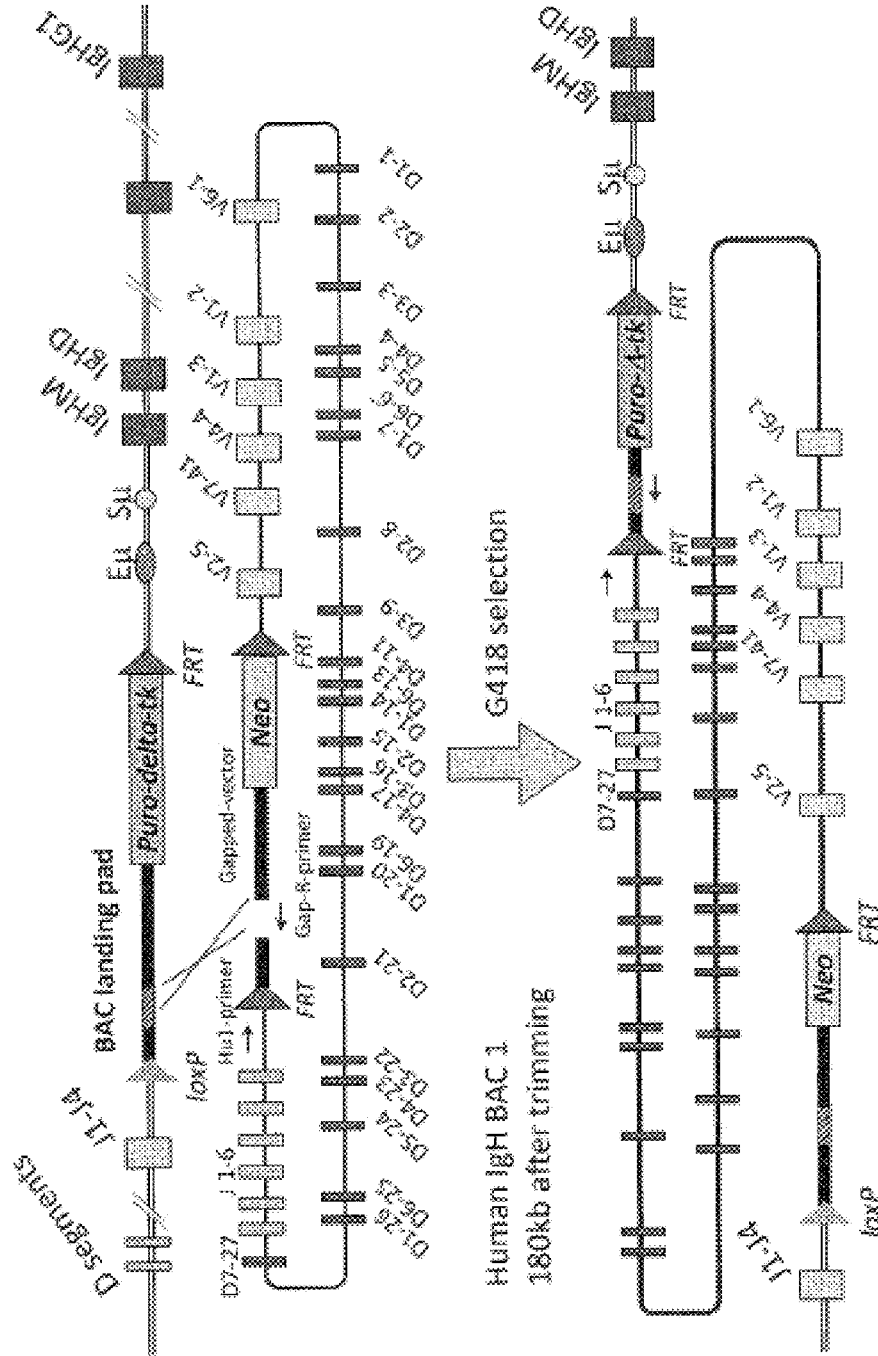


Fig. 10

1st BAC insertion – Step 2

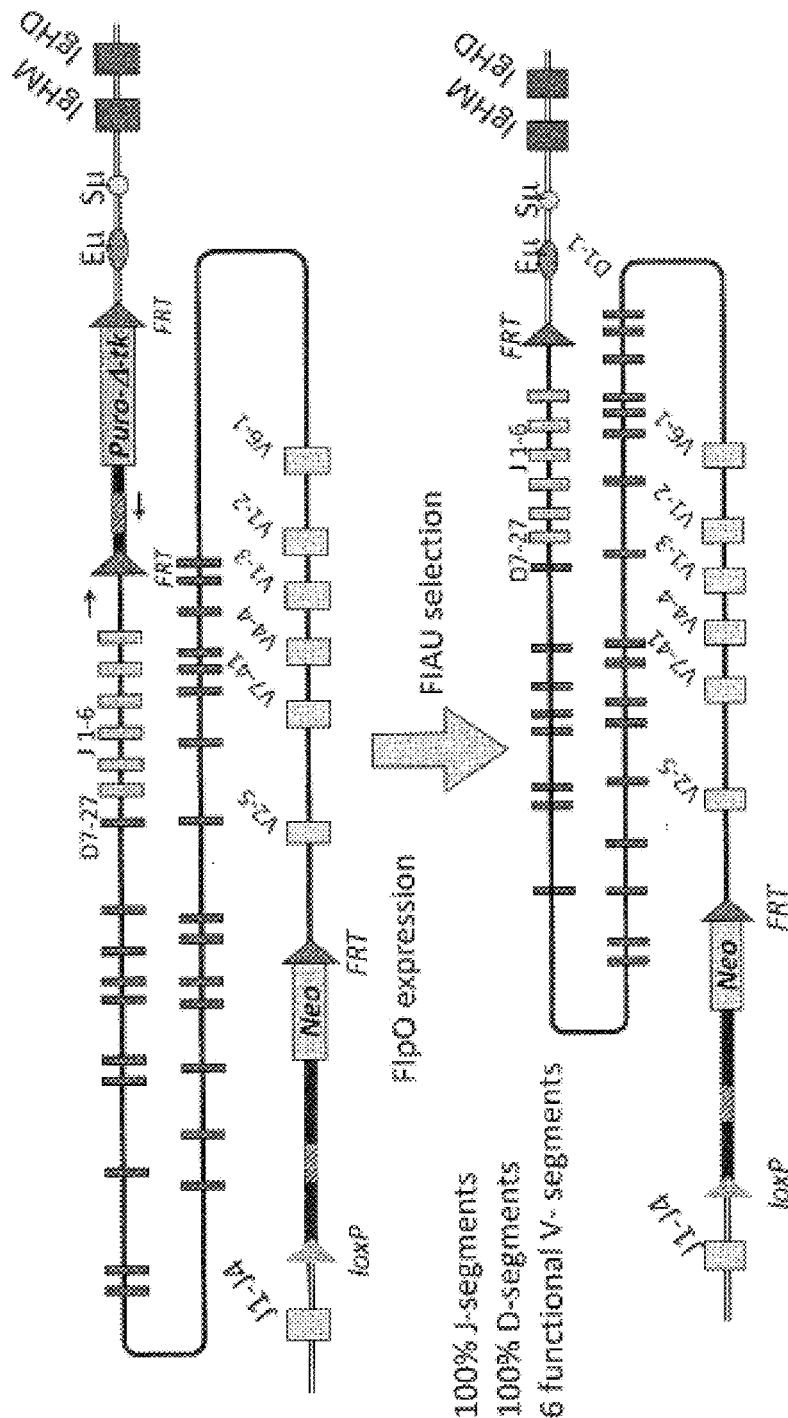


Fig. 11

2nd BAC insertion – Step 3

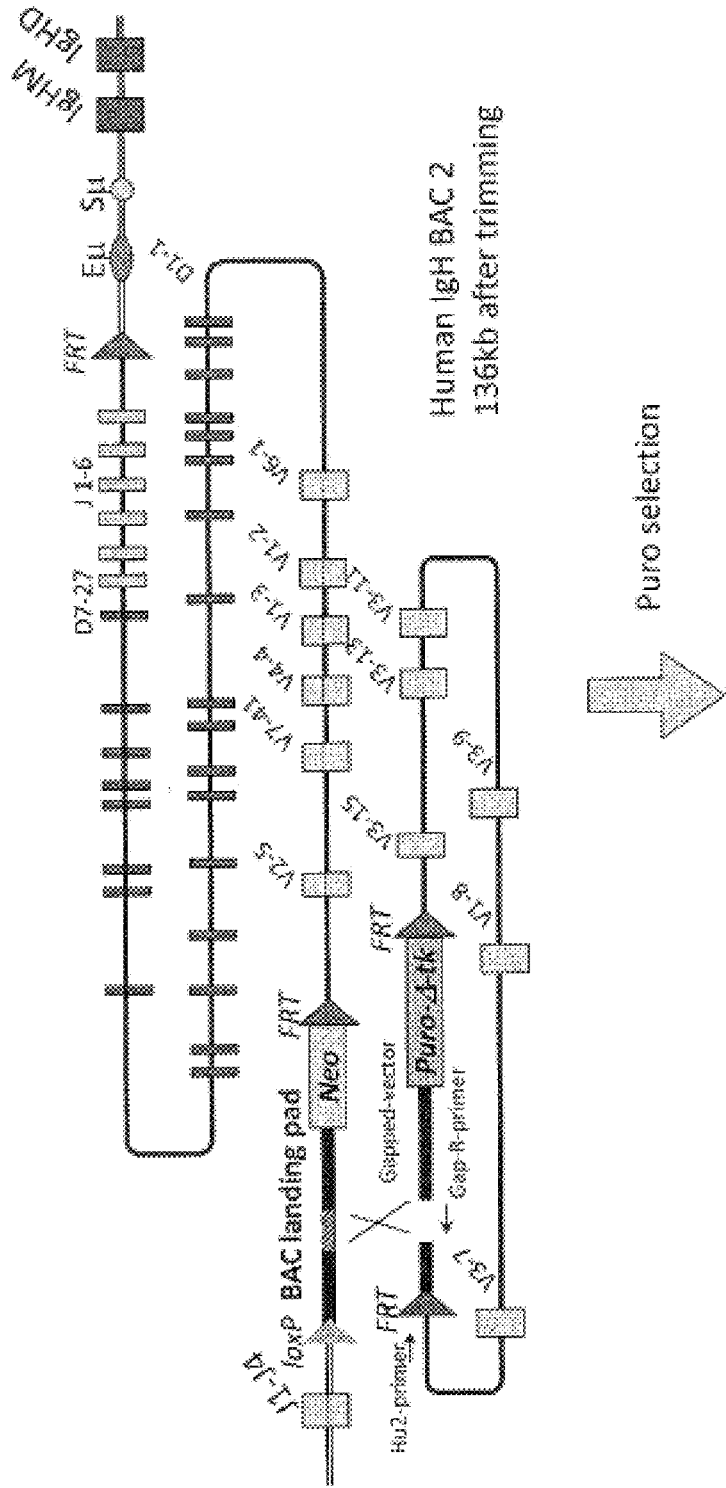


Fig. 12

2nd BAC insertion – Step 4

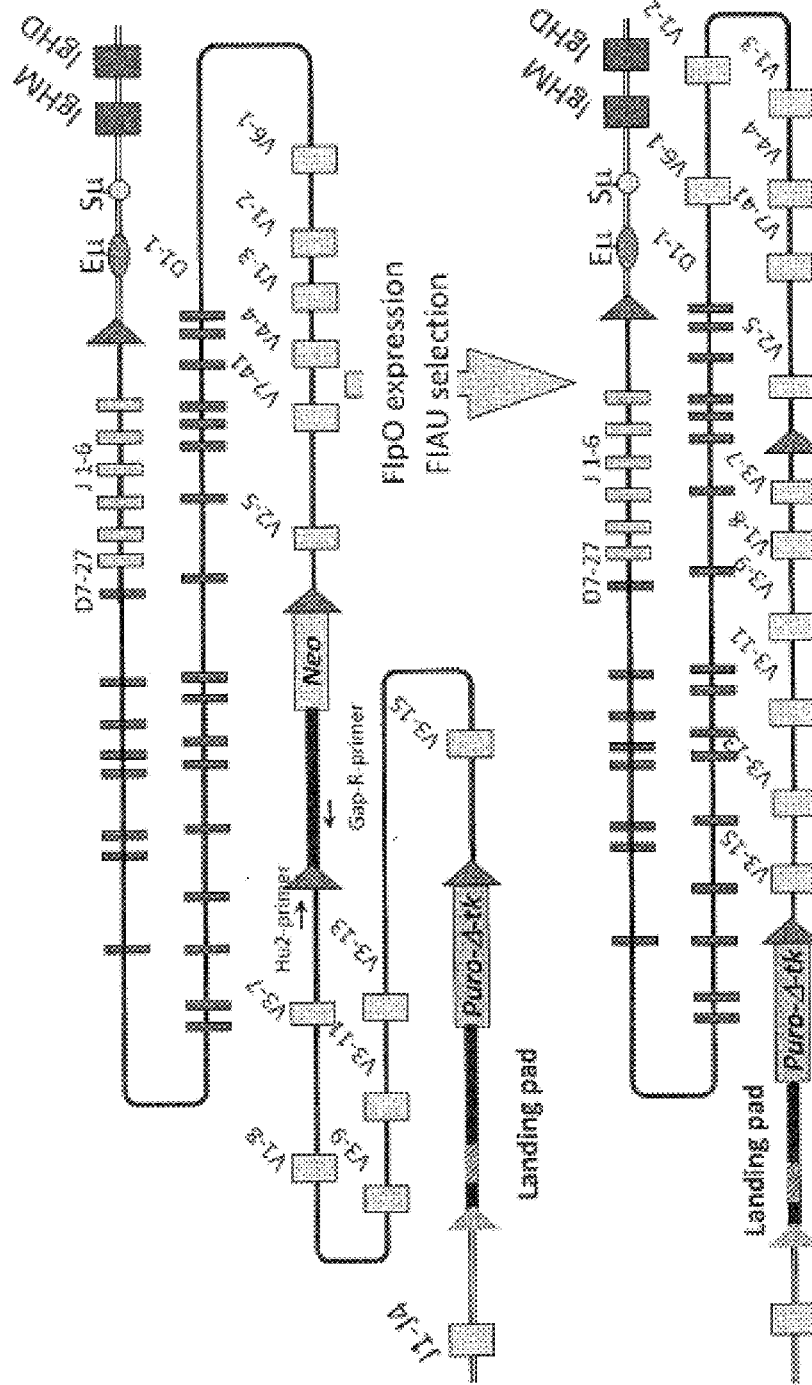


Fig. 13

BAC Landing Pad at Igk Locus

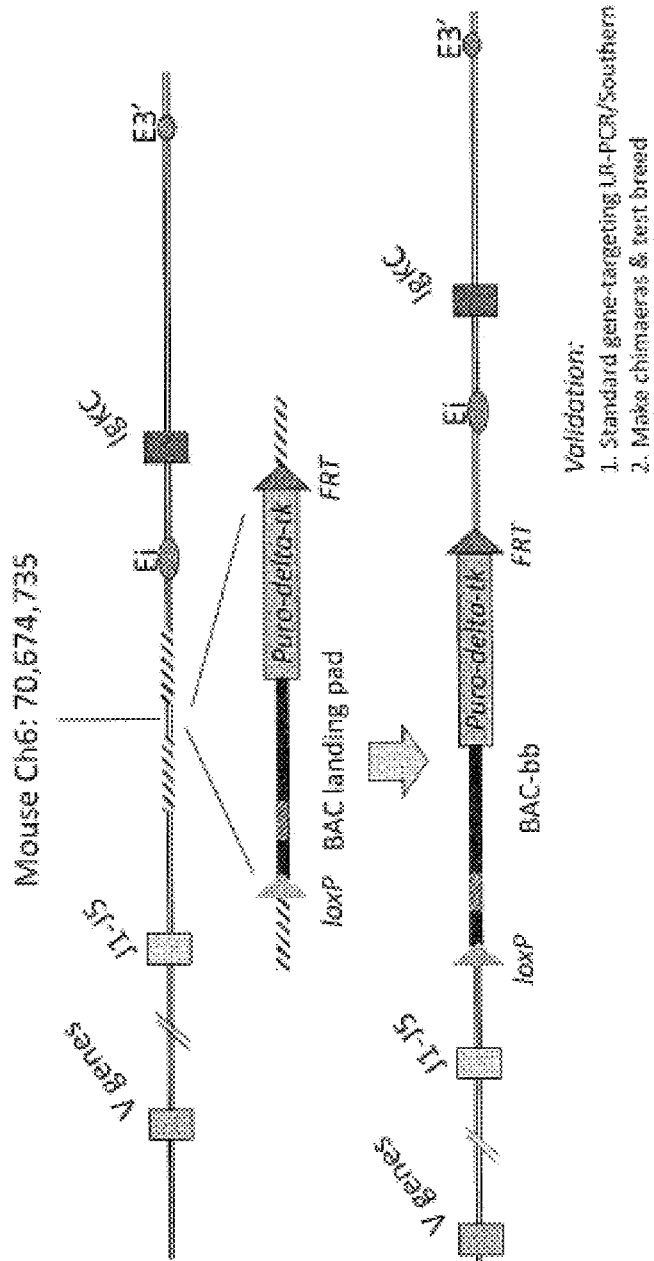


Fig. 14

1st BAC Insertion – Step 1

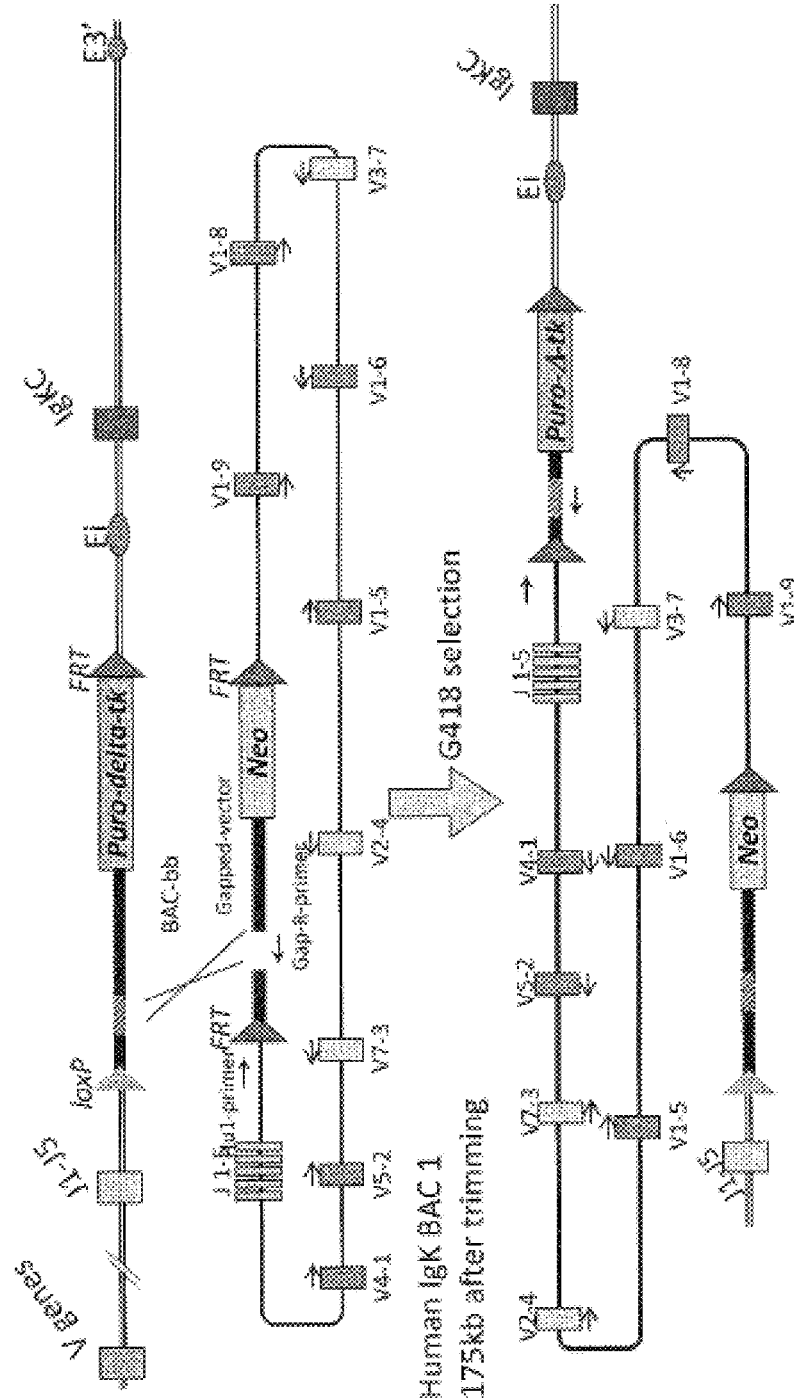


Fig. 15

1st BAC Insertion – Step 2

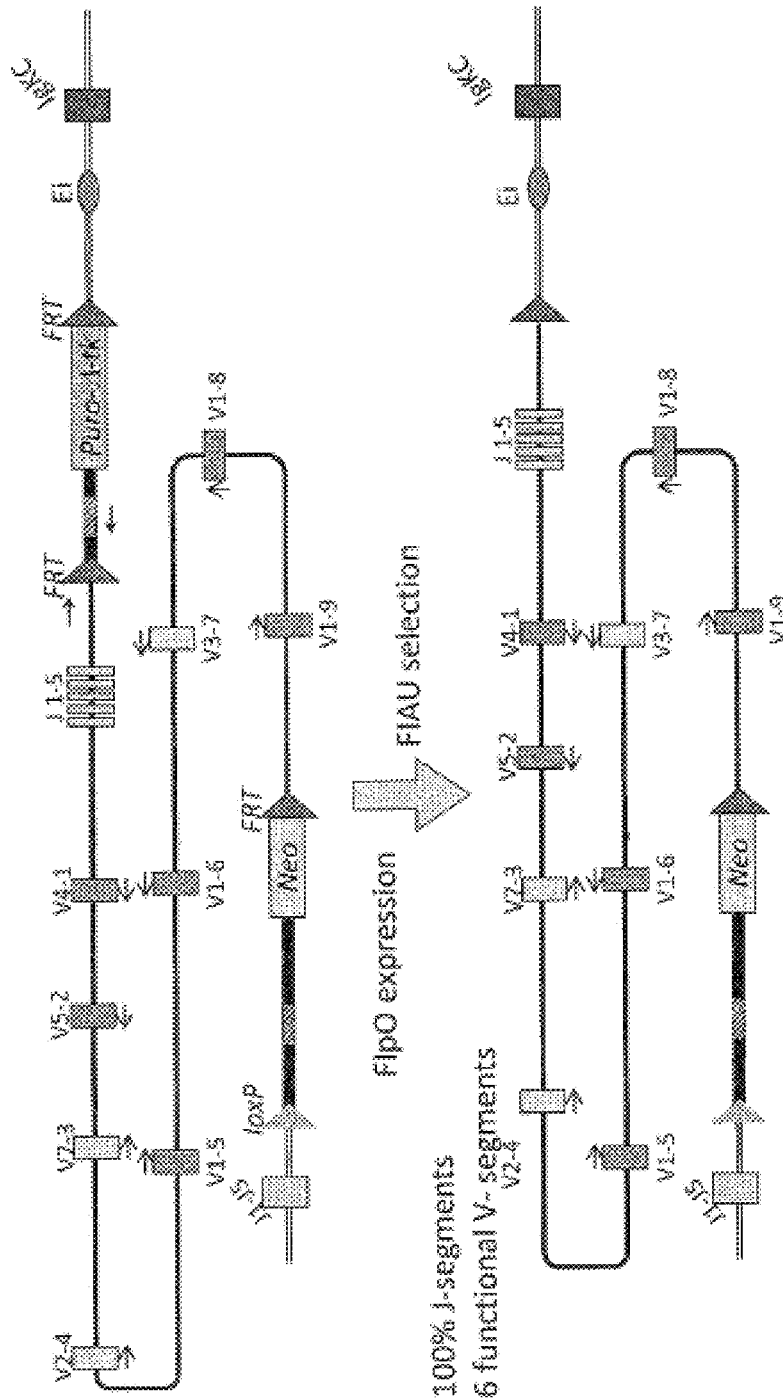


Fig. 16

2nd BAC Insertion – Step 3

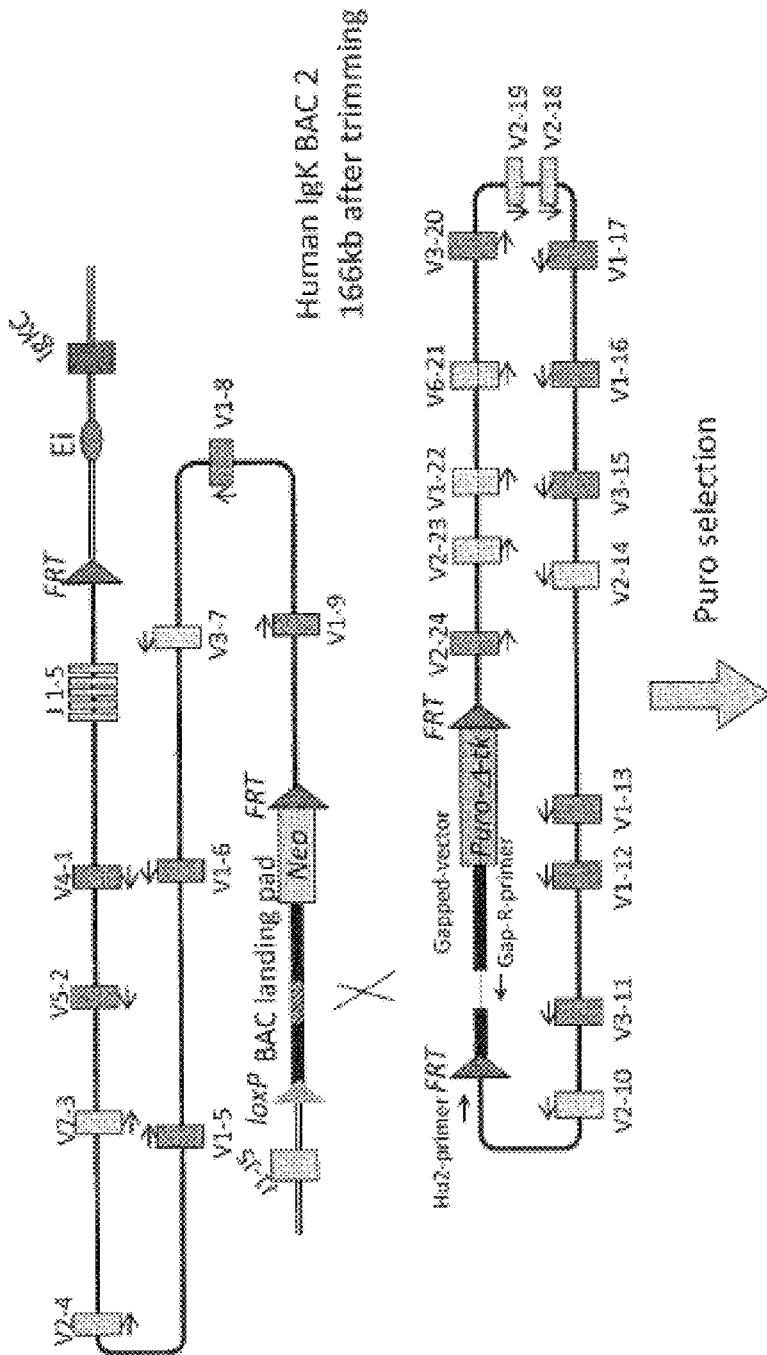


Fig. 17

2nd BAC insertion – Step 4

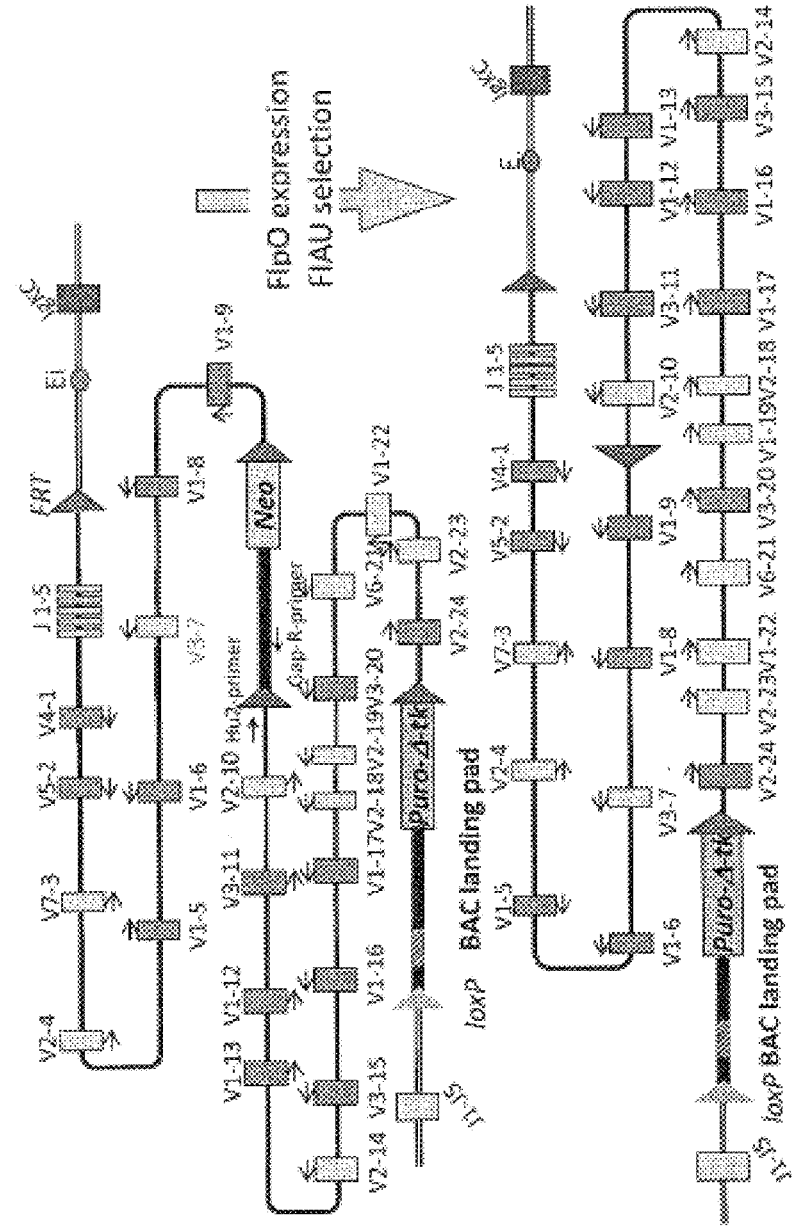


Fig. 18

Rapid Testing of ES Cells with Humanized IgH and IgK Loci

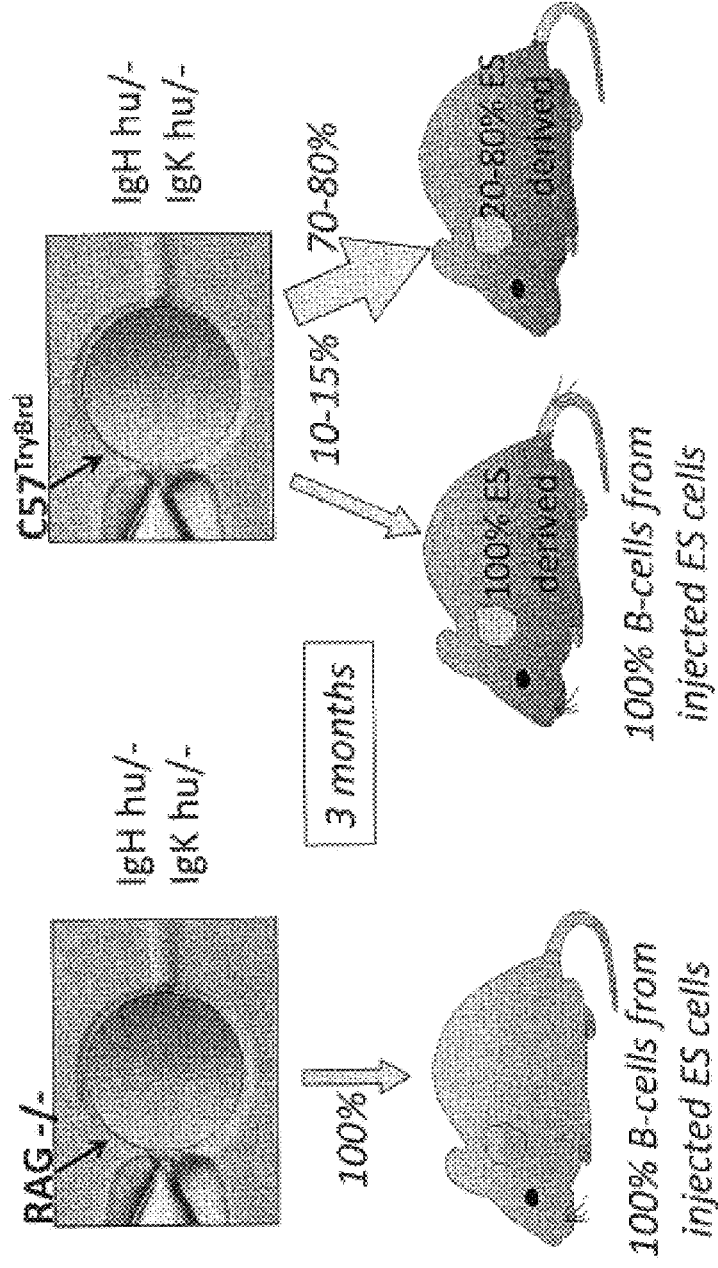


Fig. 19

Antibody Generation in Chimaeric Mice with Humanized IgH and IgK Loci

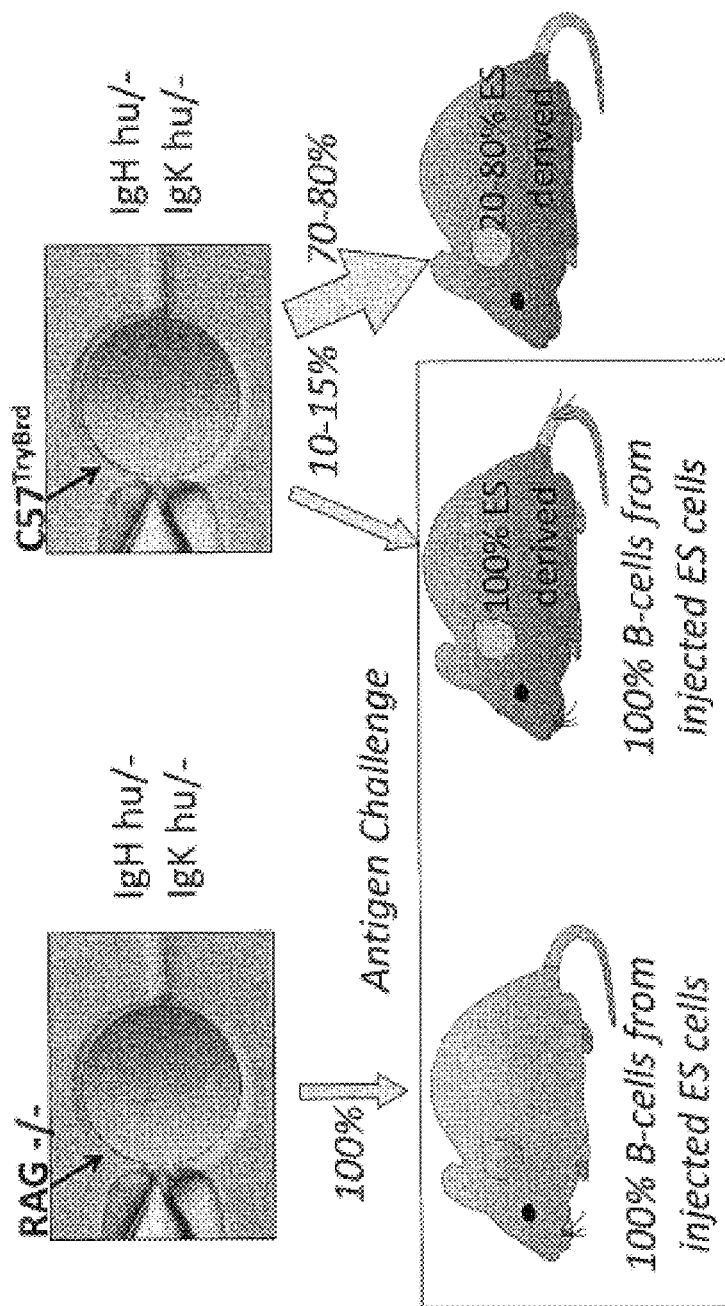


Fig. 20

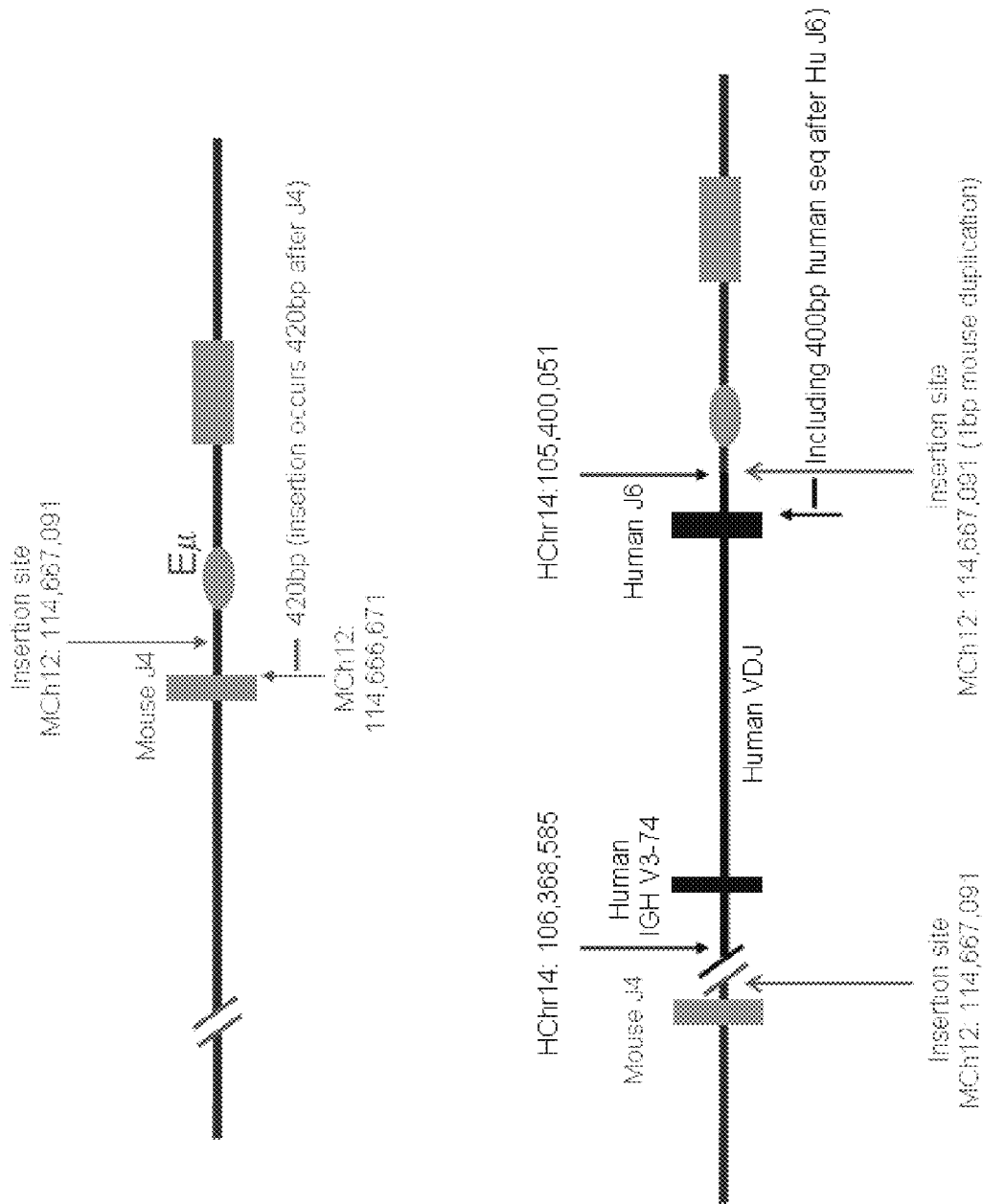


Fig. 21

Targeting of the Landing Pad

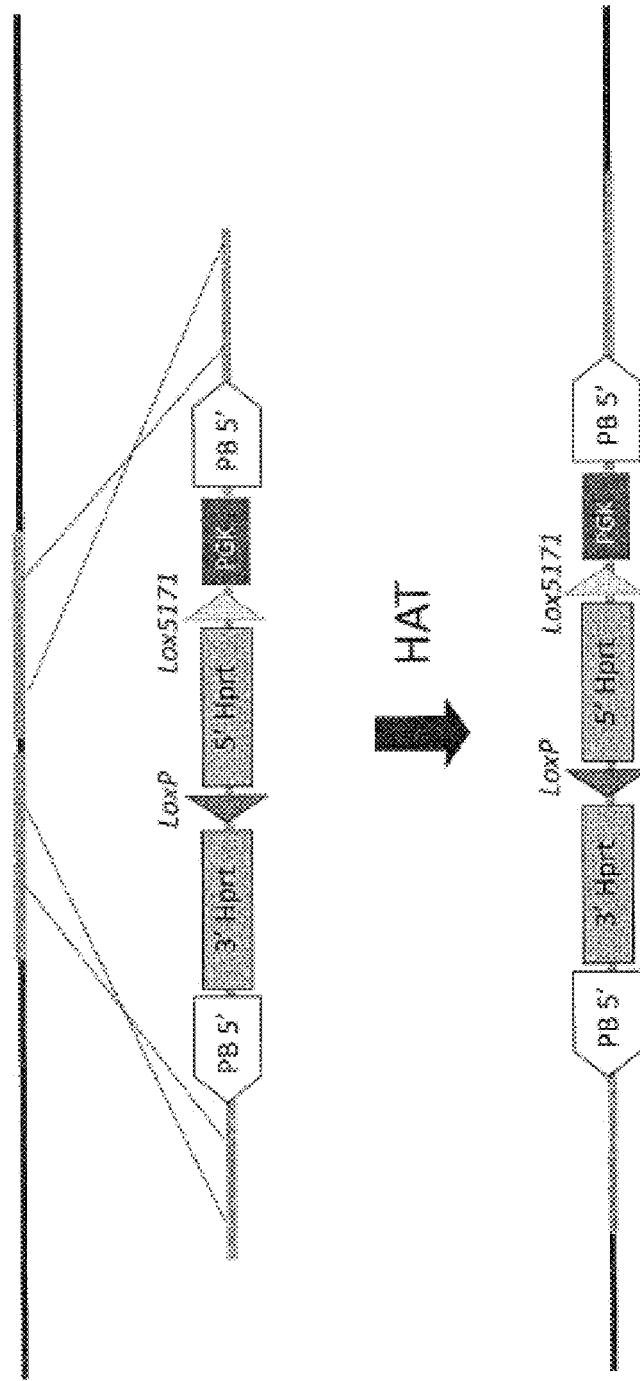


Fig. 22

Insertion of the 1st BAC

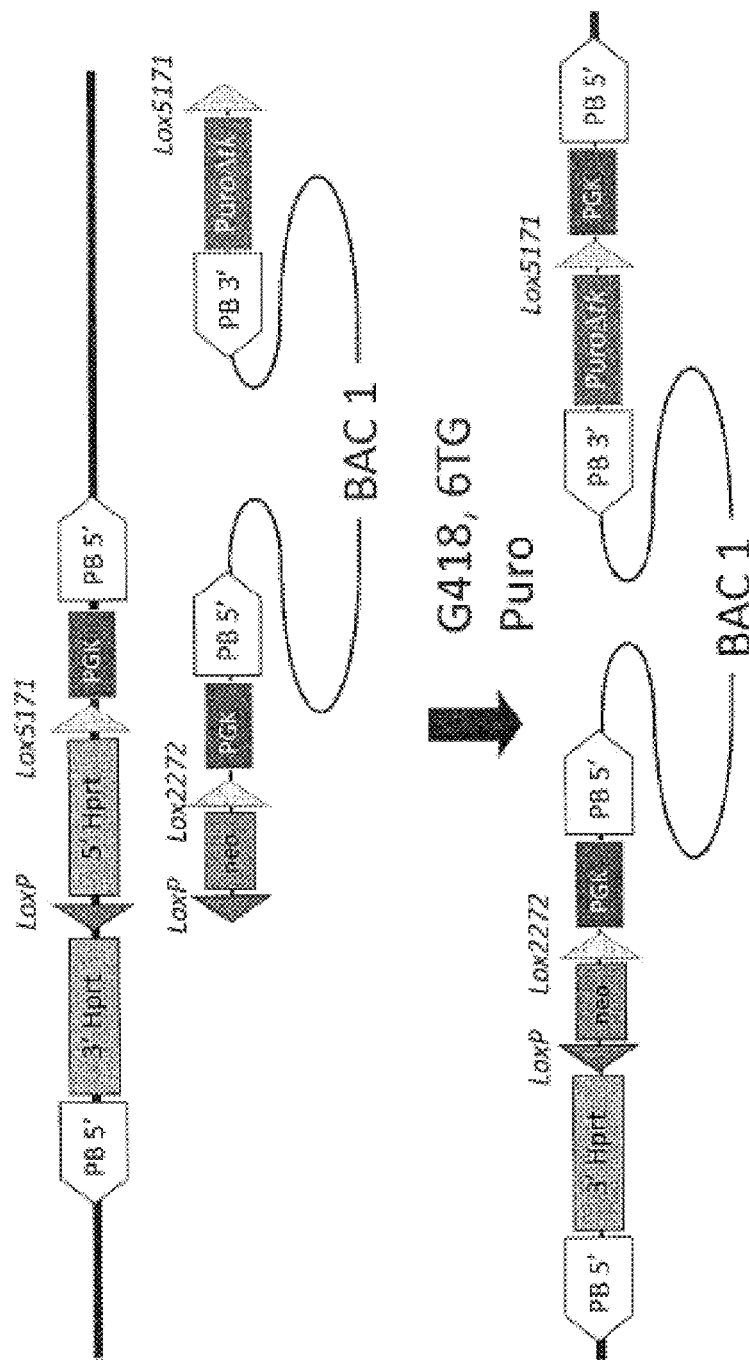


Fig. 23

Curing of the 3' Modification

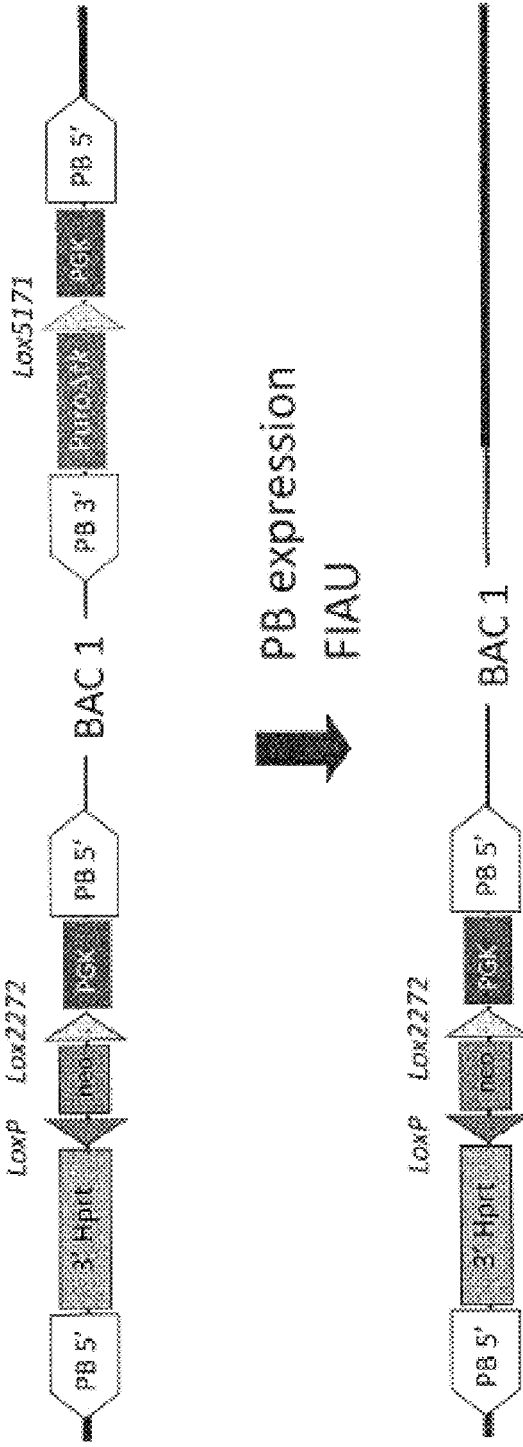


Fig. 24

Insertion of the 2nd BAC

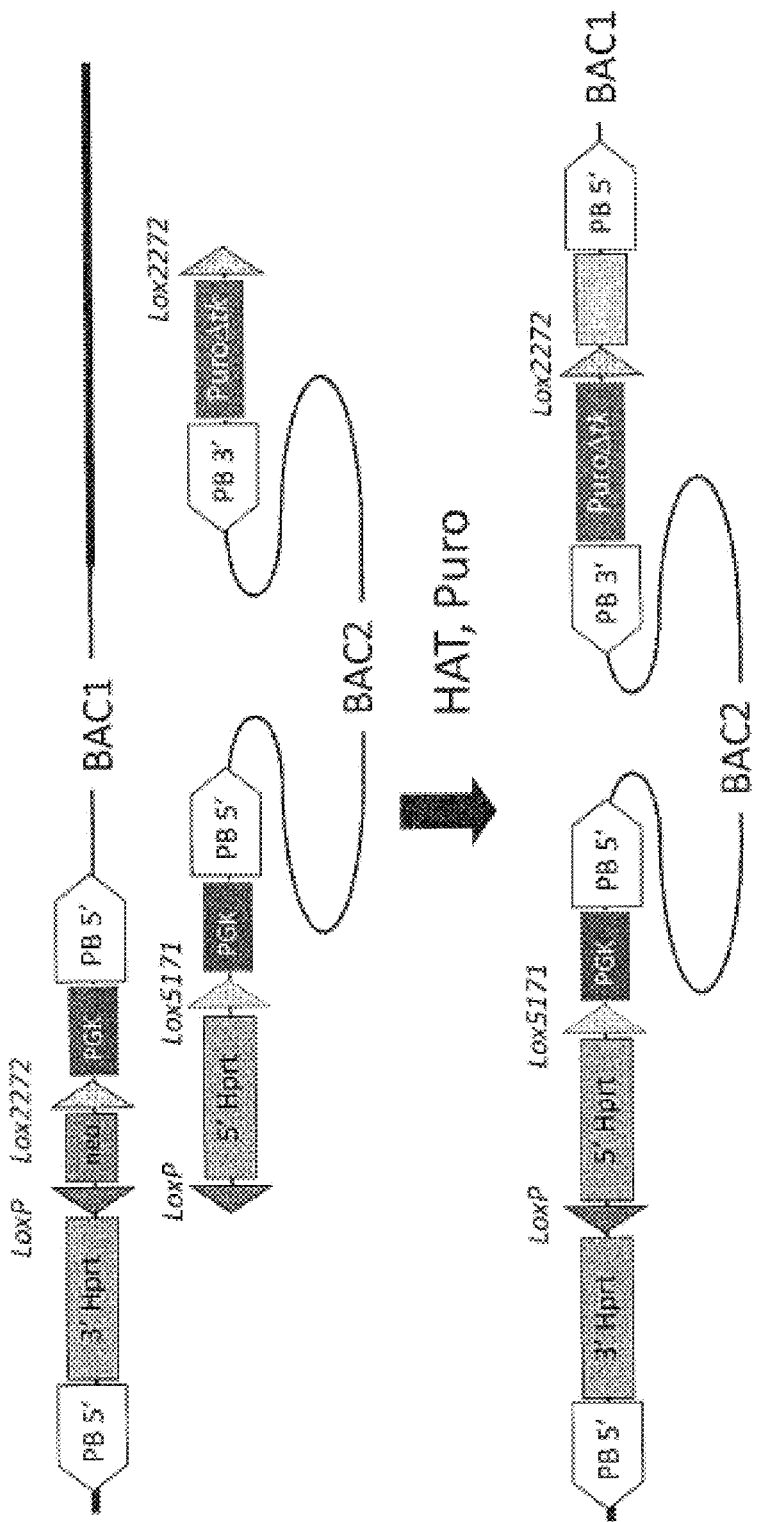


Fig. 25

Removing the 3' Modification

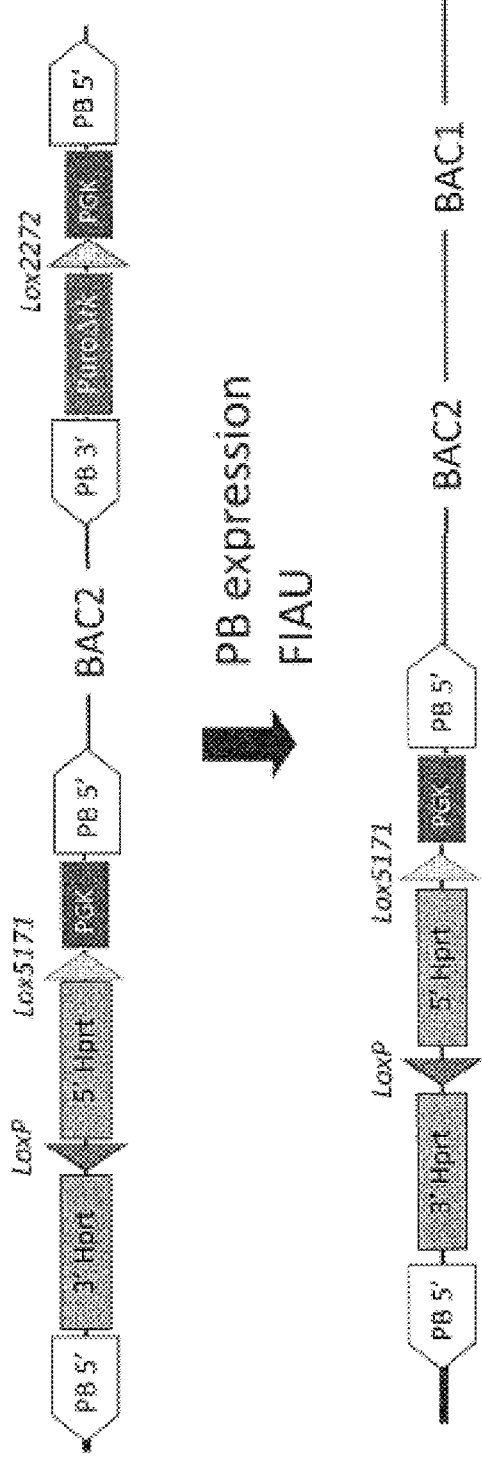


Fig. 26

Targeting the "Flip-Over" Vector

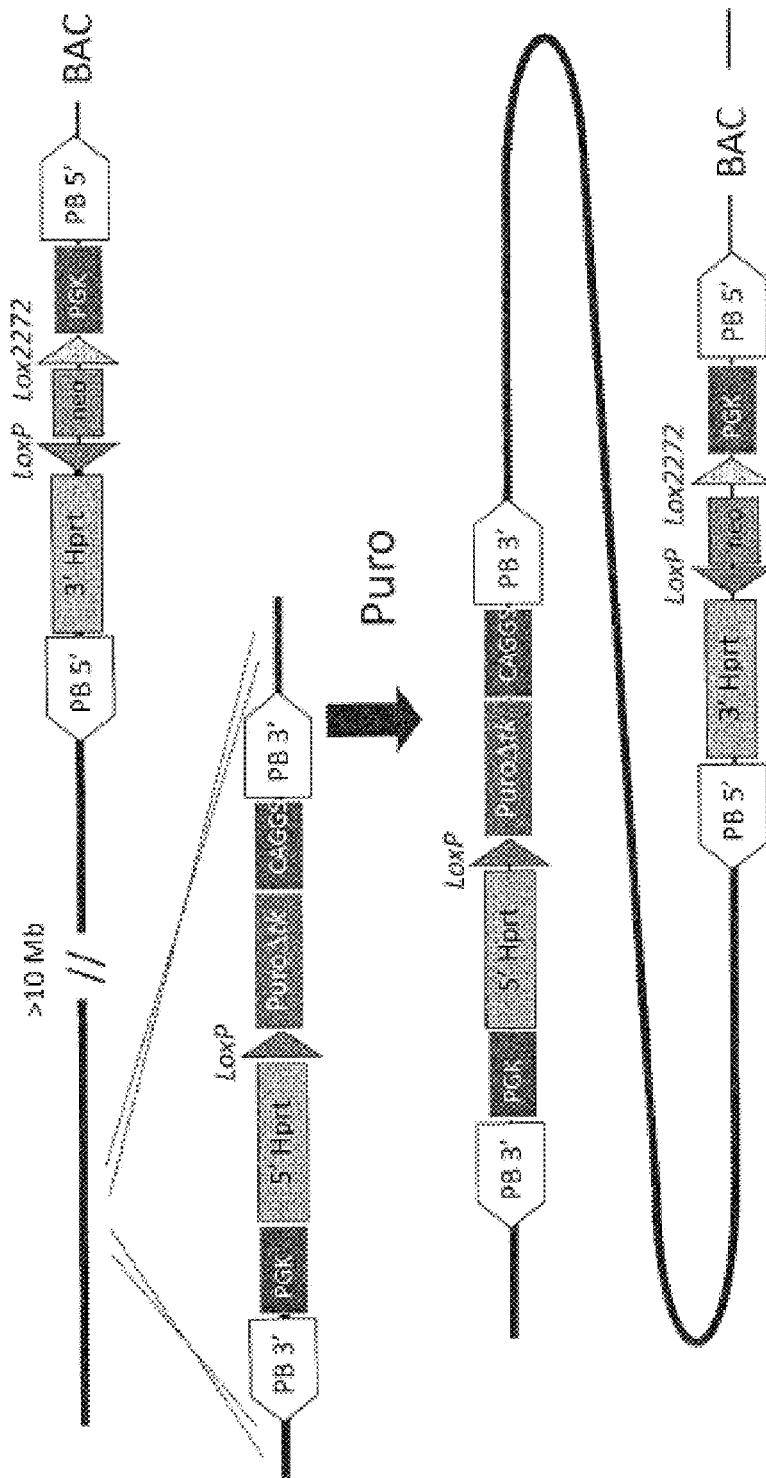


Fig. 27

Inversion

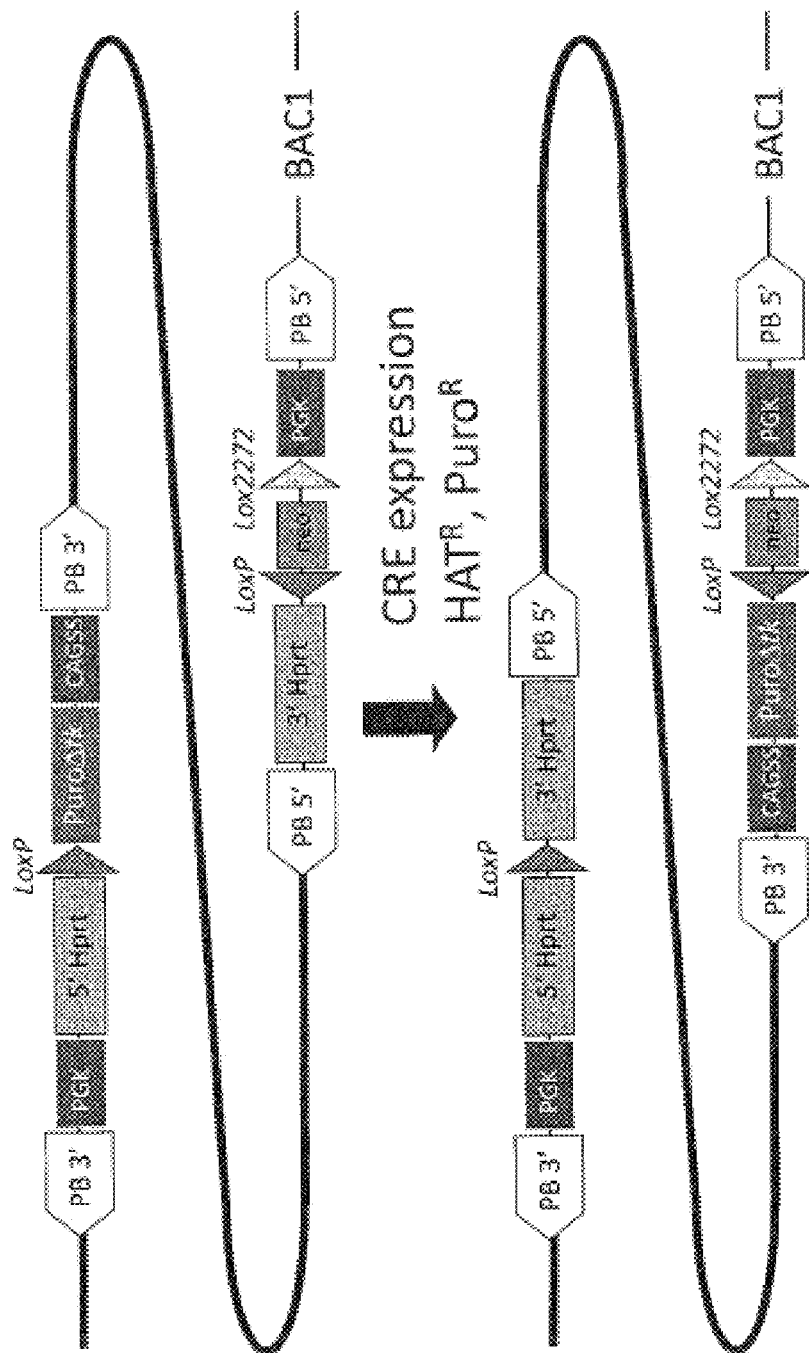


Fig. 28

Removing All Modifications

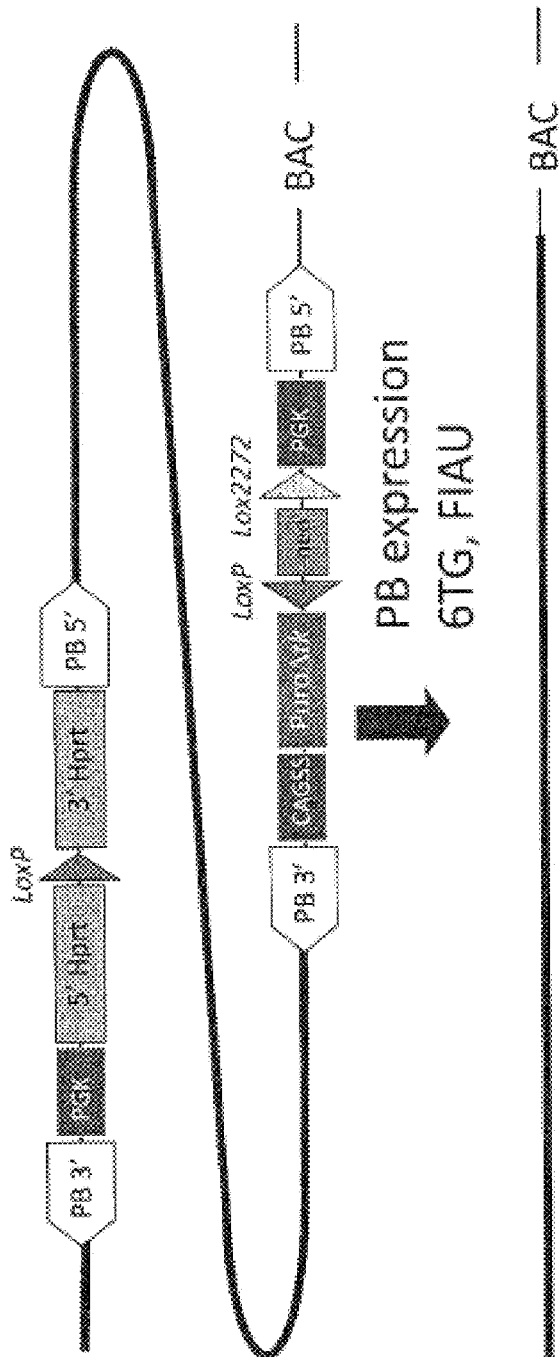
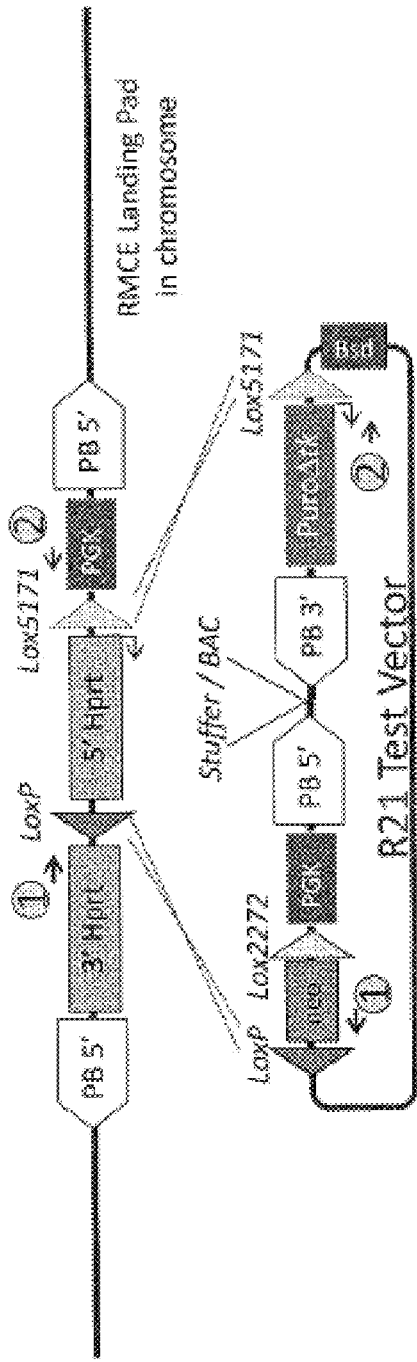


Fig. 29

Sequential RMCE – Integration into Landing Pad



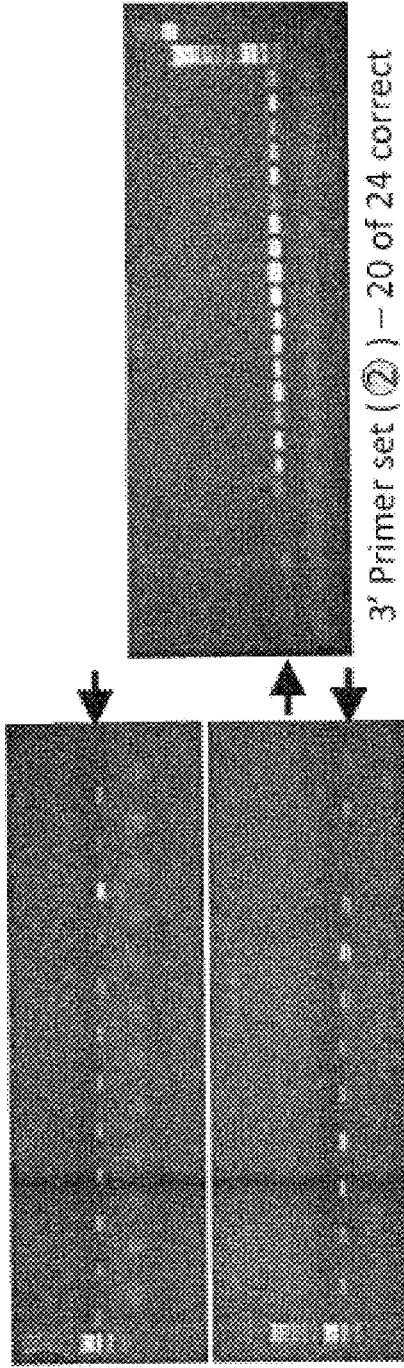
R21 transfected w/ Cre-expression Cassette

Selection:	Puro	Puro + 6TG	G418	G418 + 6TG
# Colonies:	239	220	360	198

Fig. 30

Confirmation of Successful Insertion into Landing Pad

Colonies : 6-TG and Puro - Resistant



Summary:

Colony #	Puro + 6TG		Puro		G418 + 6TG		G418	
	220		239		198		360	
Junction PCR	5'	3'	5'	3'	5'	3'	5'	3'
	23/24	20/24	22/24	19/24	22/24	22/24	11/24	11/24

Fig. 31

PCR Confirmation of 3' End Curing

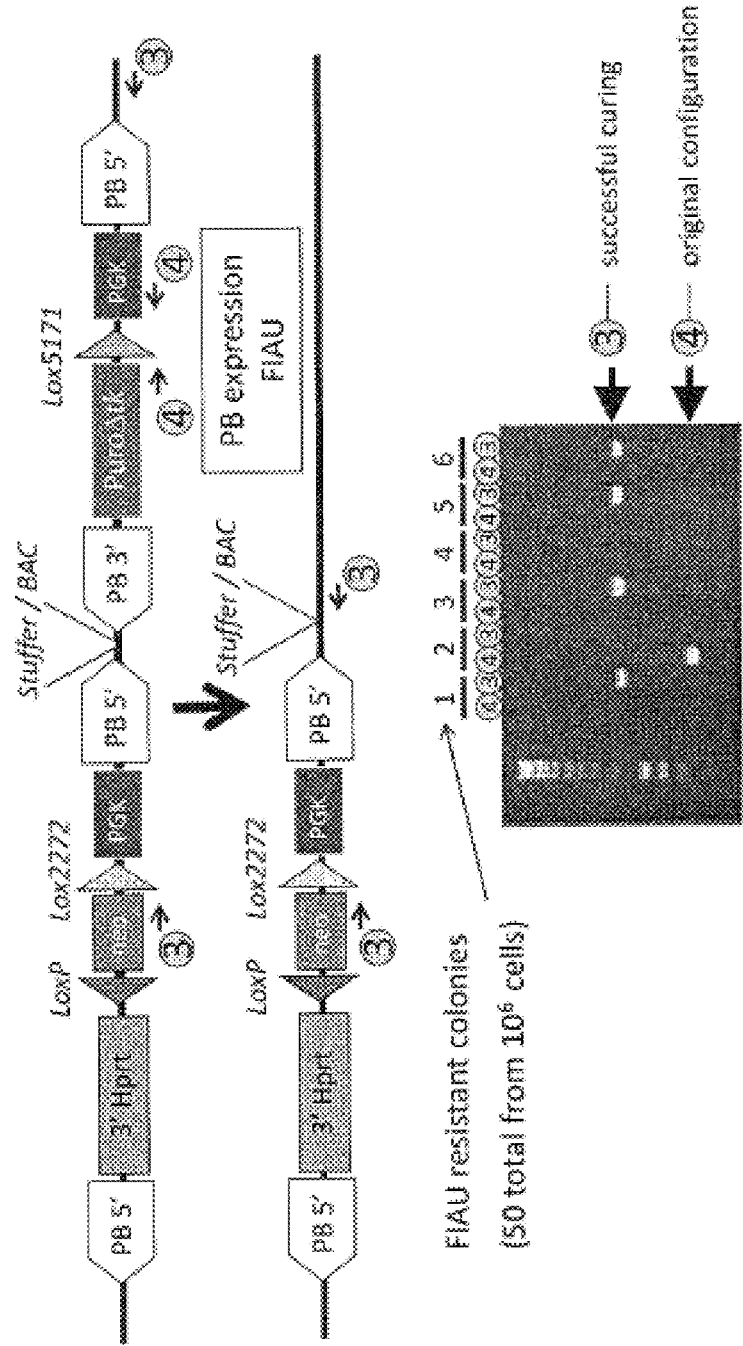


Fig. 32

Insertion of BAC#1 – PCR Diagnostics

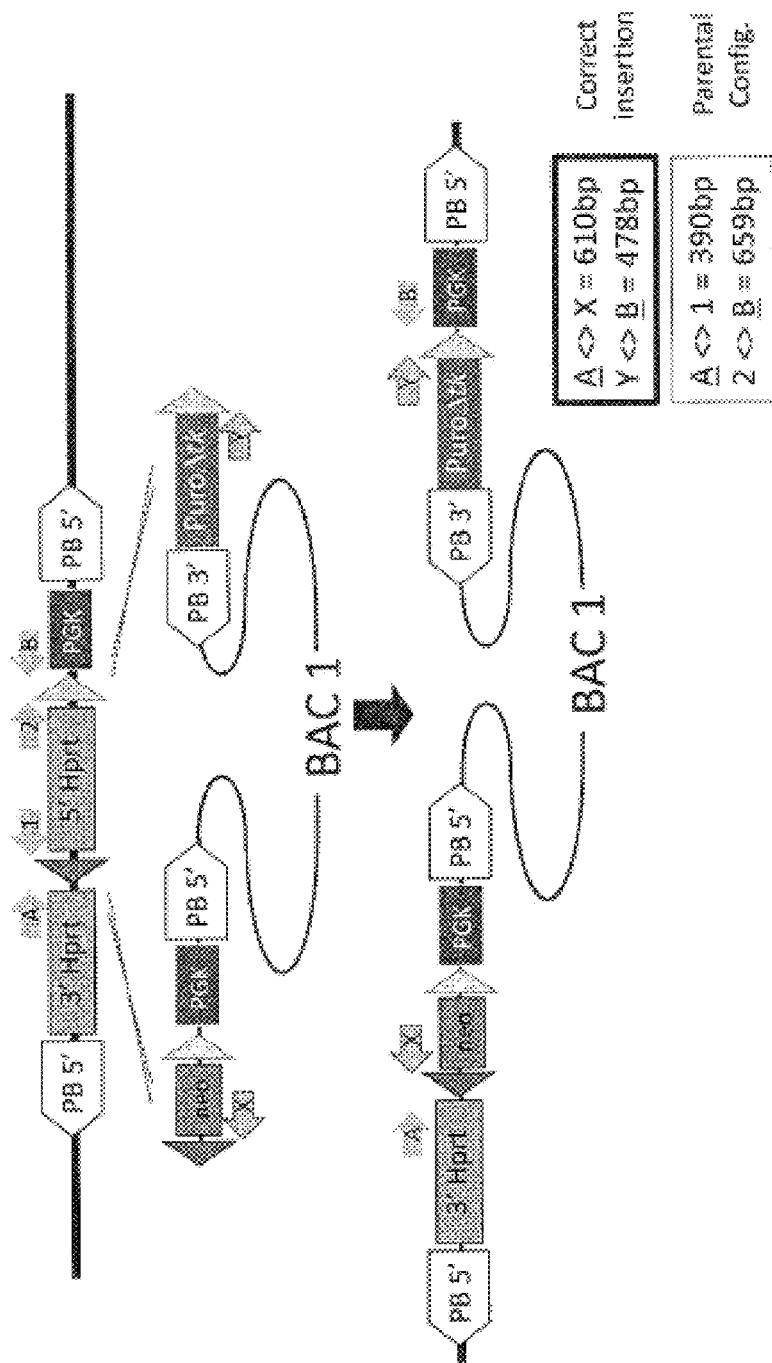
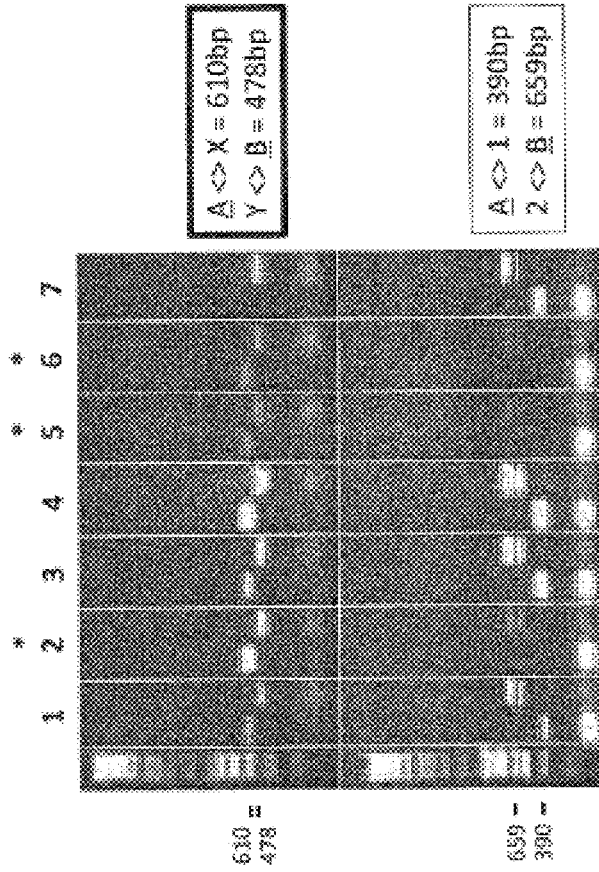


Fig. 33

Insertion of BAC#1 – PCR Diagnostics



*Clones # 2, 5 and 6 have BAC#1 inserted

Clones # 1, 3, 4 and 7 likely have BAC#1 inserted, but the samples are mixed with parental cell lines.

Fig. 34

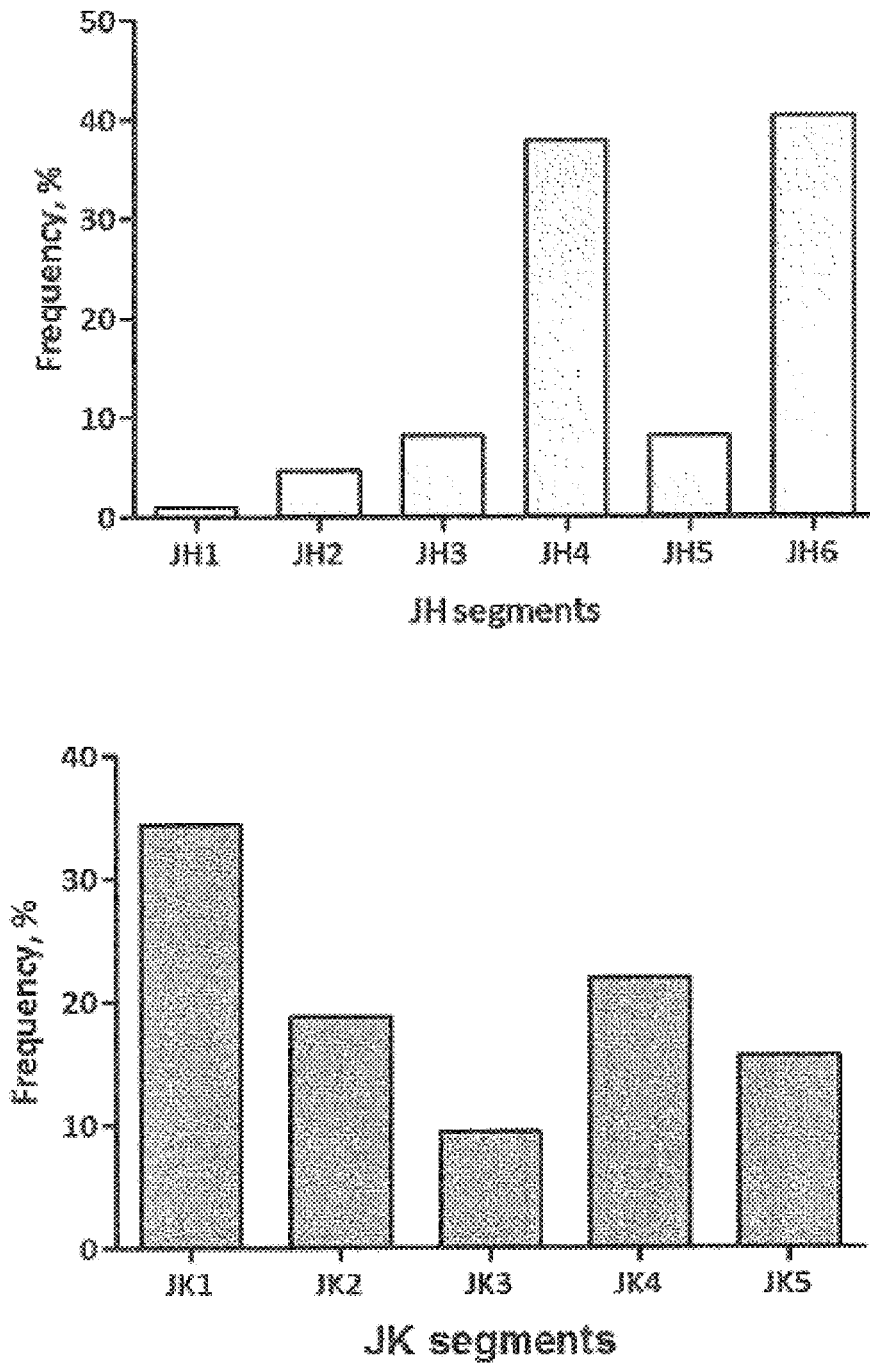


Fig. 35

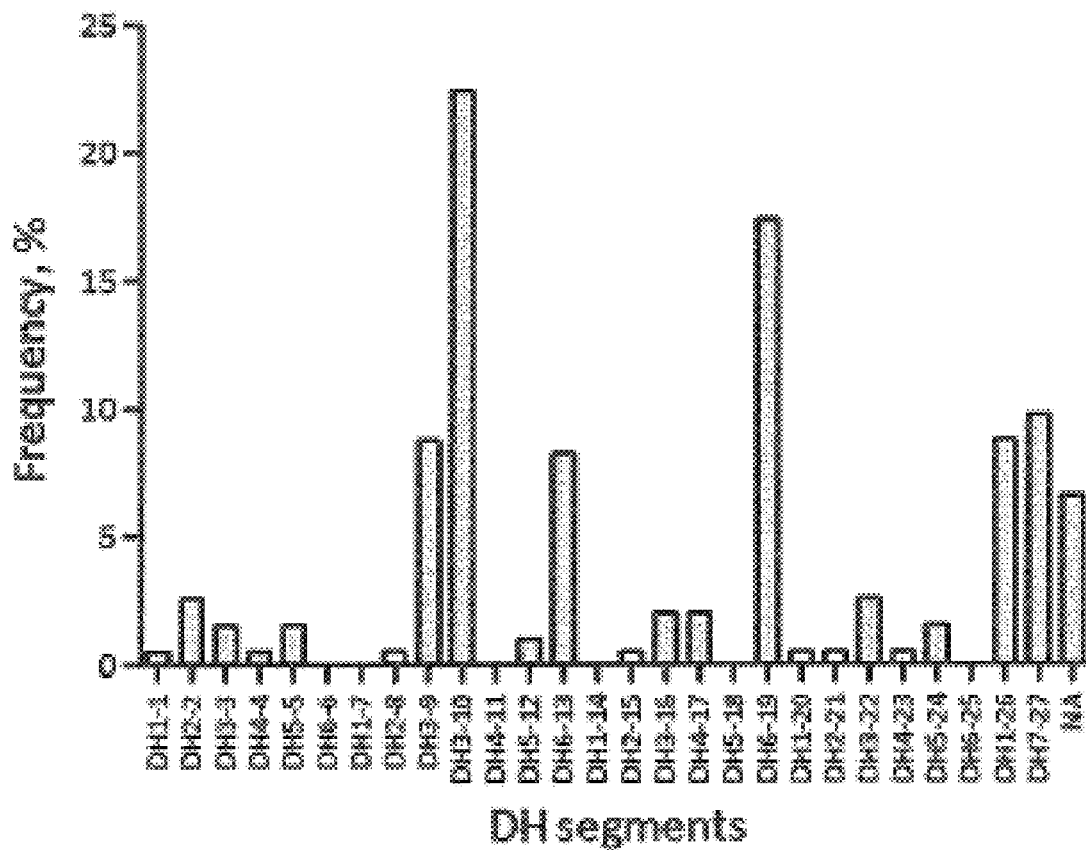


Fig. 36

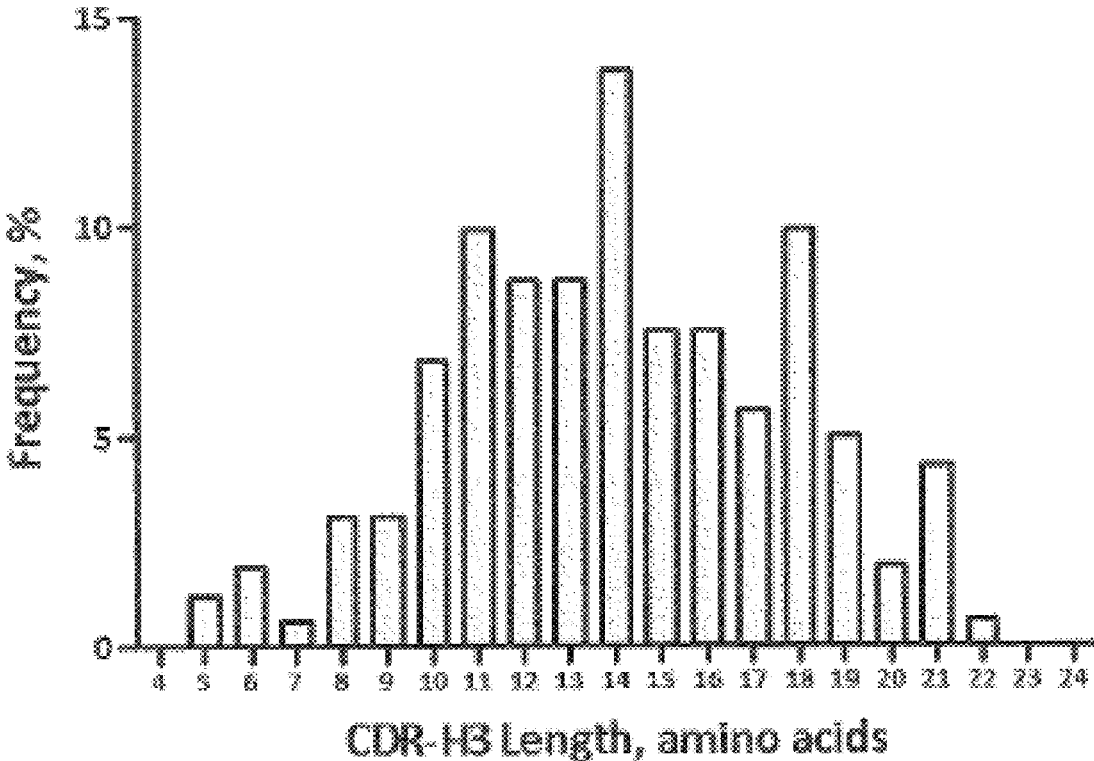


Fig. 37

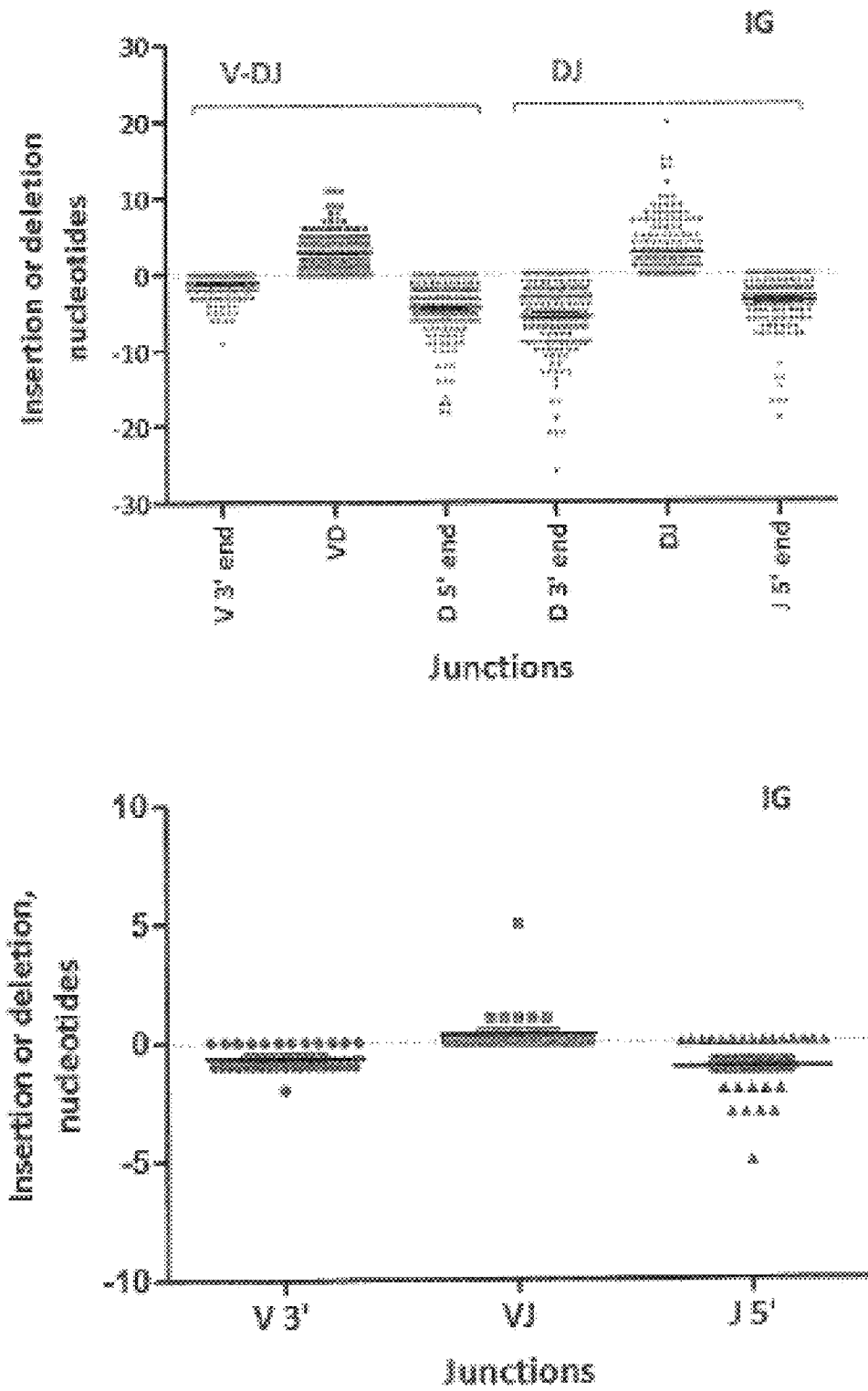


Fig. 38

Distribution of JH Usage Within Each VHs

	JH1	JH2	JH3	JH4	JH5	JH6
V2-5		1	5	5	2	1
V4-4		1	1	1		8
V1-3	1	6	6	49	13	40
V1-2	1			1		1
V6-1		1	4	18	1	29
Total	2	9	16	74	16	79

The data includes 196 independent sequences.

Fig. 39

Nucleotide Gain or Loss at VJ Joints Generates IGK Variants

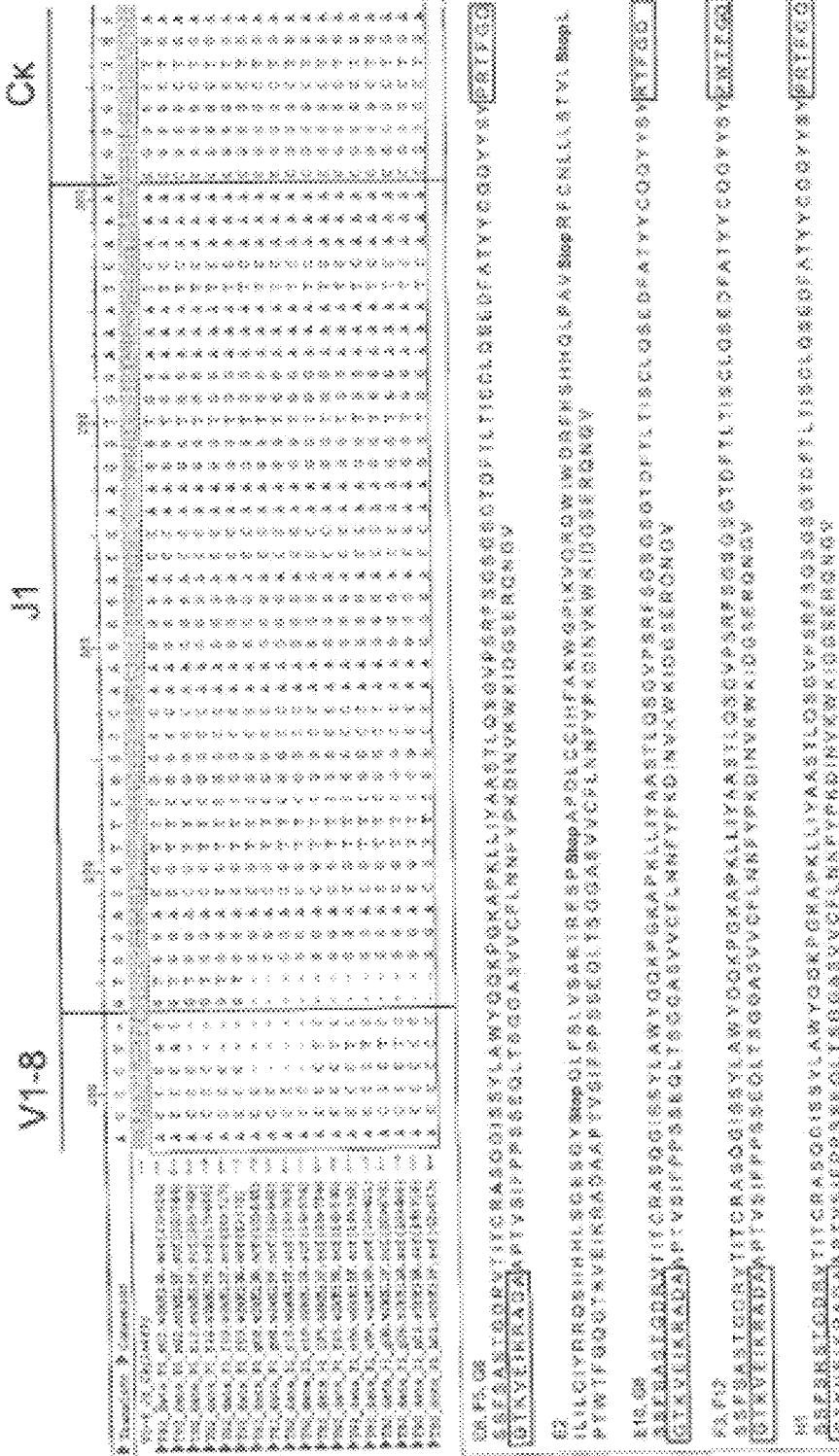


Fig. 41

Hypermutation in J Regions Generates IGK Variants

	V1-6										J1										Ck																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
100	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
101	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
102	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
103	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
104	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
105	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
106	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
107	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
108	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
109	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
110	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
111	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
112	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
113	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
114	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
115	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
116	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
117	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
118	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
119	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
120	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

A1, A2, A6, A10, A11, A12, B3, B5, B7, C3, D3, D6
 KLLIYAASLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCLOGYNHWTFGGGTKVEIKRADAPTVISIFPPSSAEQLTIS
 GDSVYCVFLNRFYPKDINVKWKIDGSERONGV

A4, A5, A7, B4, B9, B11, C3, C4, C7, C12, D7
 KLLIYAASLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCLOGYNHWTFGGGTKVEIKRADAPTVISIFPPSSAEQLTISG
 GDSVYCVFLNRFYPKDINVKWKIDGSERONGV

B12, C6, D3
 KLLIYAASLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCLOGYNHWTFGGGKLEIKRADAPTVISIFPPSSAEQLTISG
 GDSVYCVFLNRFYPKDINVKWKIDGSERONGV

Fig. 42

Joint Diversity Produces Functional CDS

V1-4₁,L₁,C₁
 VAQWPOISRNQCKPLSS~~SS~~SSIRNRLV~~SS~~SSKVVSSMCOSSAAV~~SS~~SSLOQNSLSPSAAC
 SLMLQLITANSIIVILTFGGQTKVEIKRADAAPTYSIFPSSSEQLTSSGASVVCFL
 CFLNNFYPKDINVKWKIDSSERONGV

H12
 SSWLAWYQQKPKAPKLLIYKASSLESQVPSRFSSQSSQTEFTLTISSLQPPDD
 FATYCCQGYNSYSTFGGQTKVEIKRADAAPTYSIFPSSSEQLTSSGASVVCFL
 NNFYPKDINVKWKIDSSERONGV

V1-4₂,J₁,C₁
 A~~SW~~WQI~~SR~~RQCKPLSS~~SS~~SSIRNRLV~~SS~~SSKVVSSMCOSSAAV~~SS~~SSLOQNSLSPSAAC
 L~~ML~~QLITANSIIVILTFGGQTKVEIKRADAAPTYSIFPSSSEQLTSSGASVVCFL
 CFLNNFYPKDINVKWKIDSSERONGV

E1
 SWLAWYQQKPKAPKLLIYKASSLESQVPSRFSSQSSQTEFTLTISSLQPPDD
 ATYCCQGYNSYSTFGGQTKVEIKRADAAPTYSIFPSSSEQLTSSGASVVCFL
 NNFYPKDINVKWKIDSSERONGV

E2
 SWLAWYQQKPKAPKLLIYKASSLESQVPSRFSSQSSQTEFTLTISSLQPPDD
 ATYCCQGYNSYSTFGGQTKVEIKRADAAPTYSIFPSSSEQLTSSGASVVCFL
 NNFYPKDINVKWKIDSSERONGV

E3
 SSWLAWYQQKPKAPKLLIYKASSLESQVPSRFSSQSSQTEFTLTISSLQPPDD
 FATYCCQGYNSYSTFGGQTKVEIKRADAAPTYSIFPSSSEQLTSSGASVVCFL
 LNNFYPKDINVKWKIDSSERONGV

Fig. 43

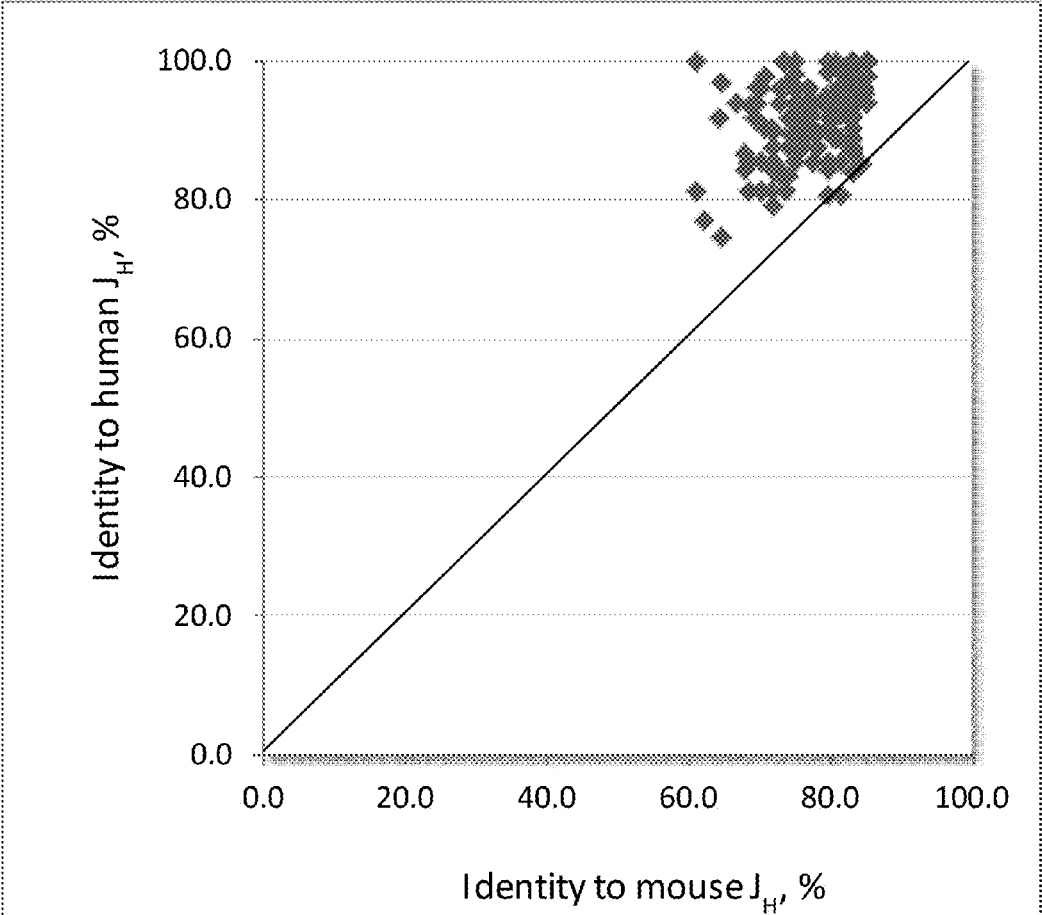


Fig. 44

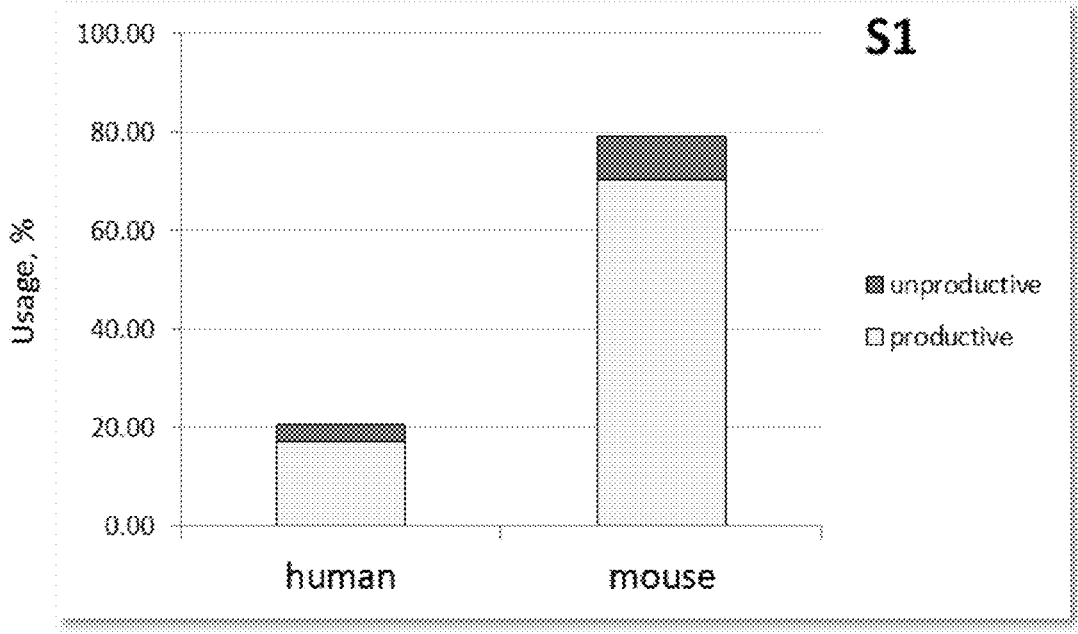


Fig. 45

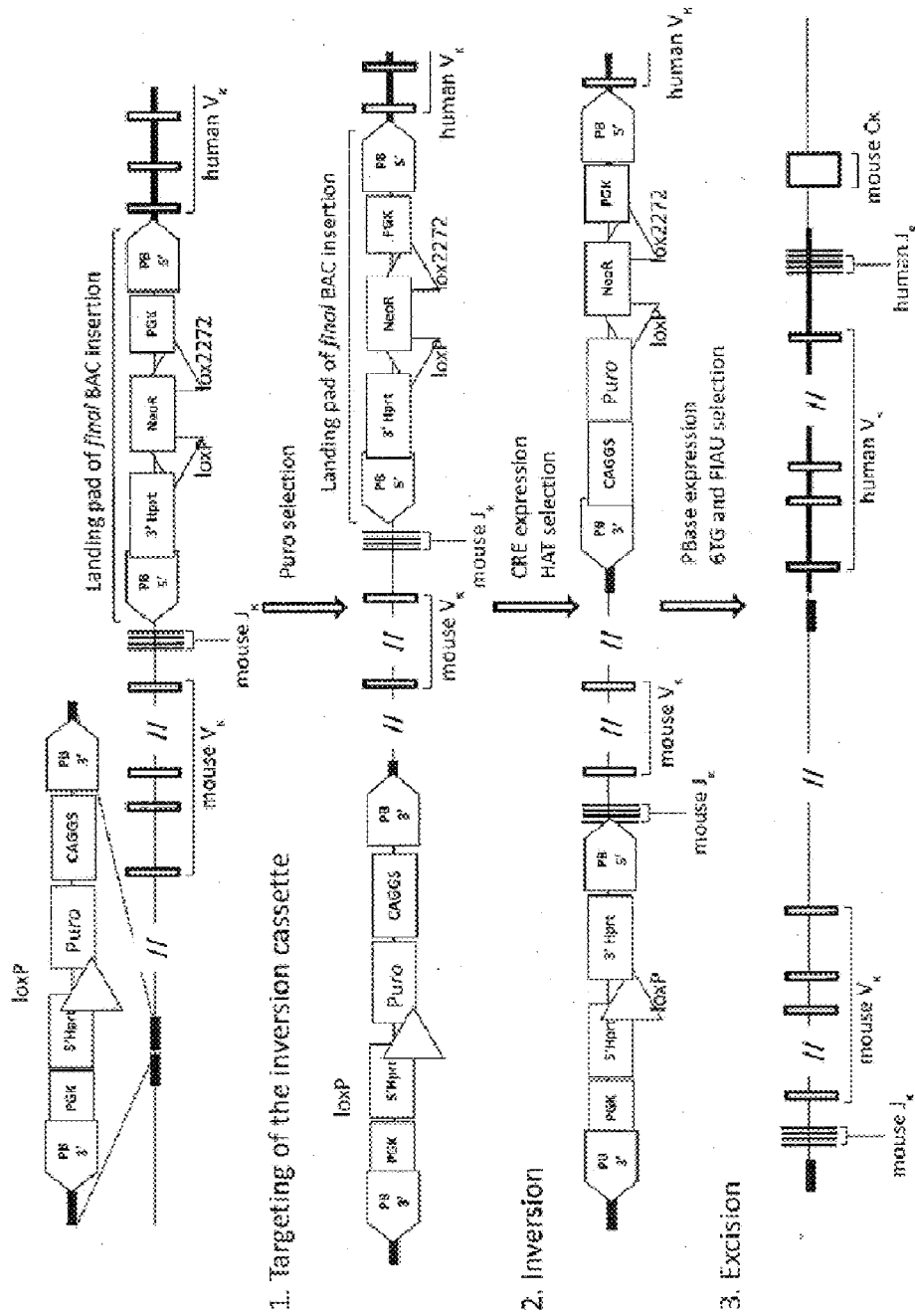


Fig. 46

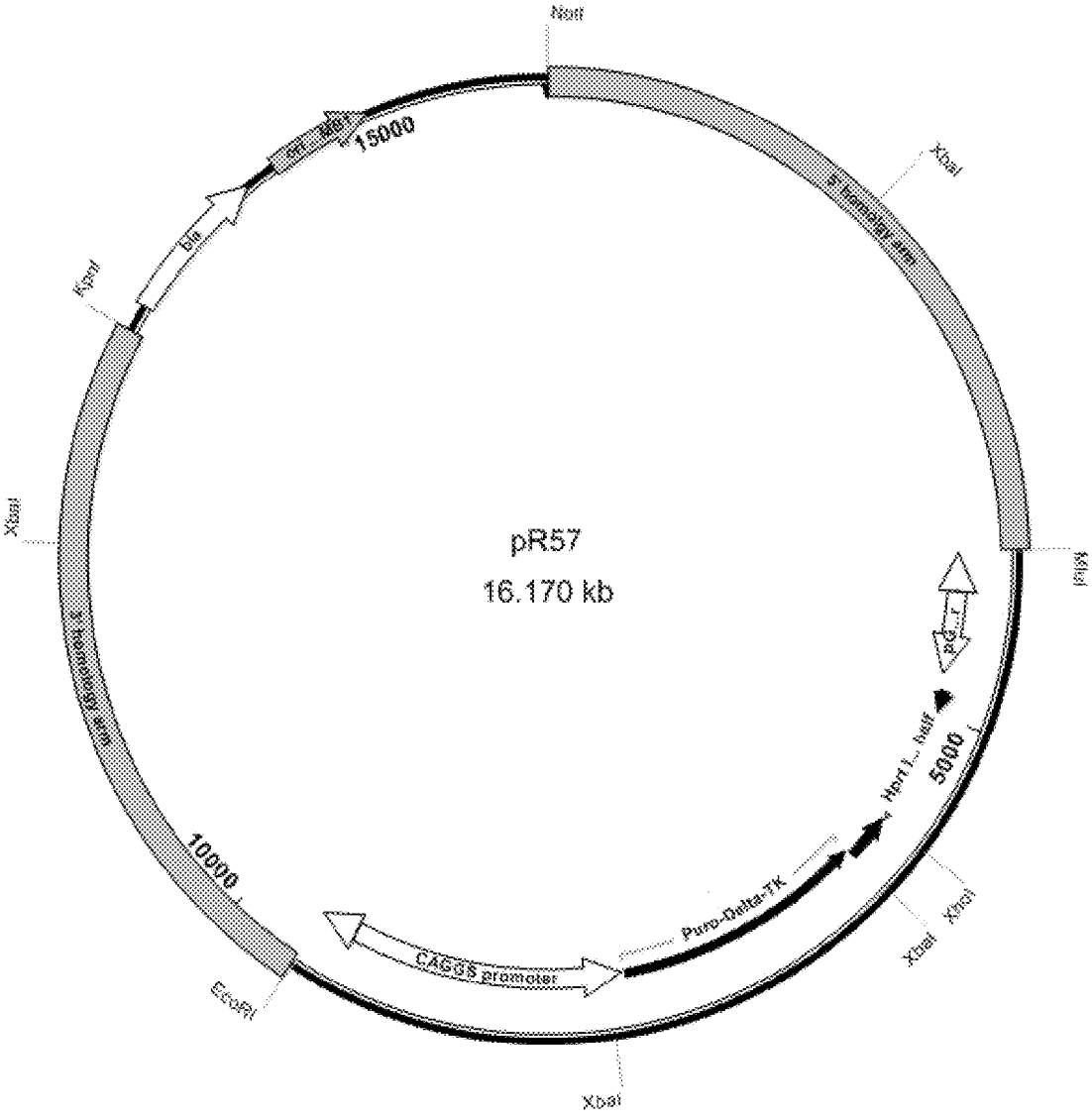


Fig. 47

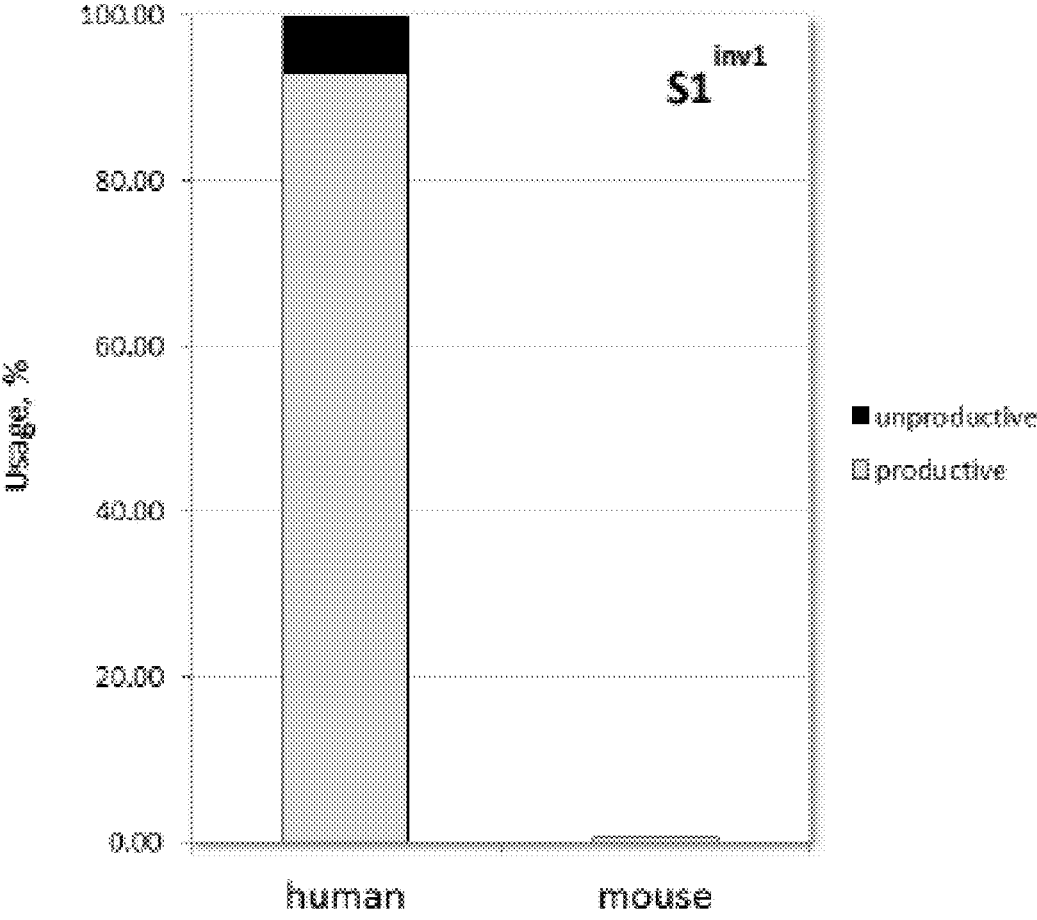


Fig. 48

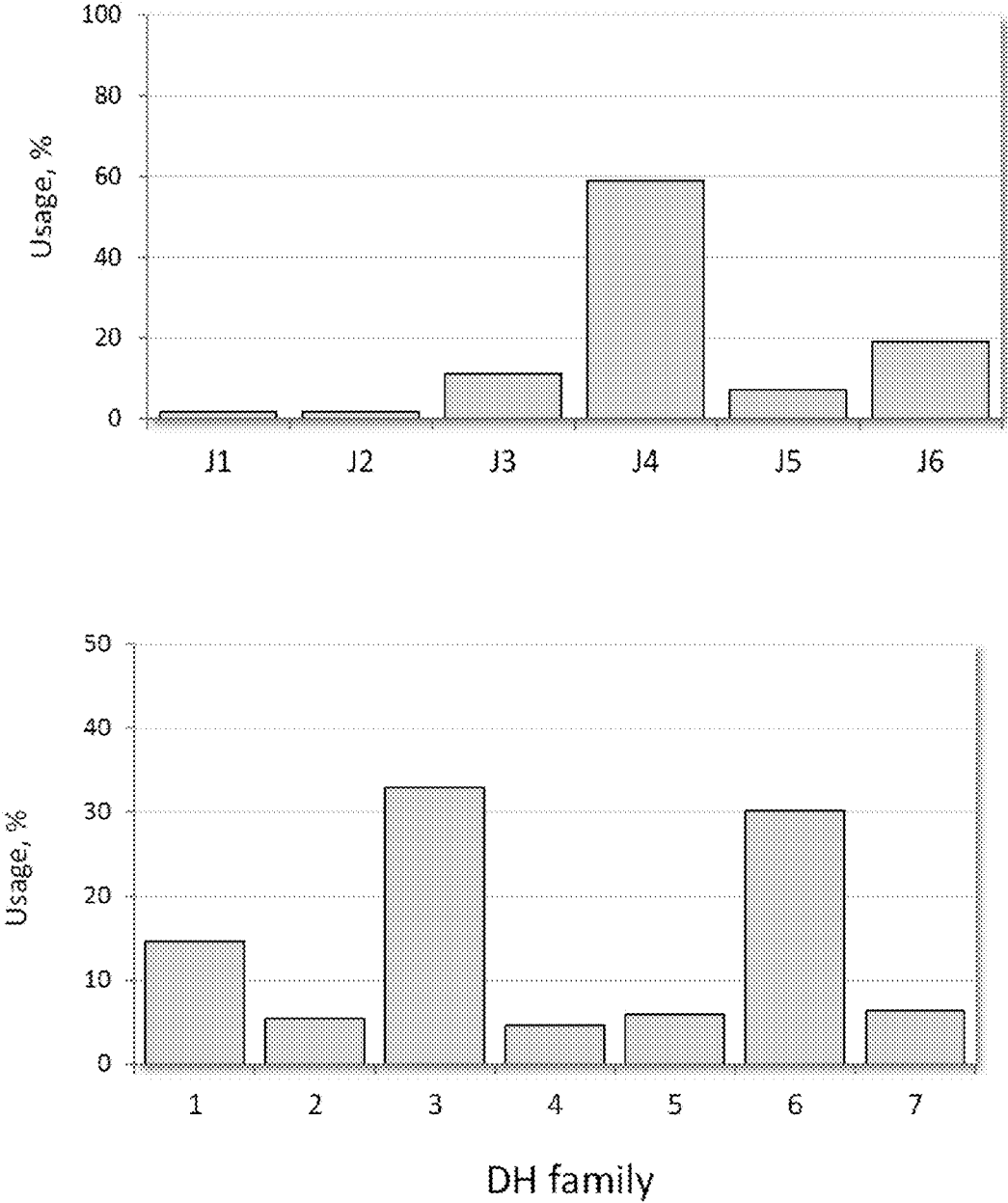


Fig. 49

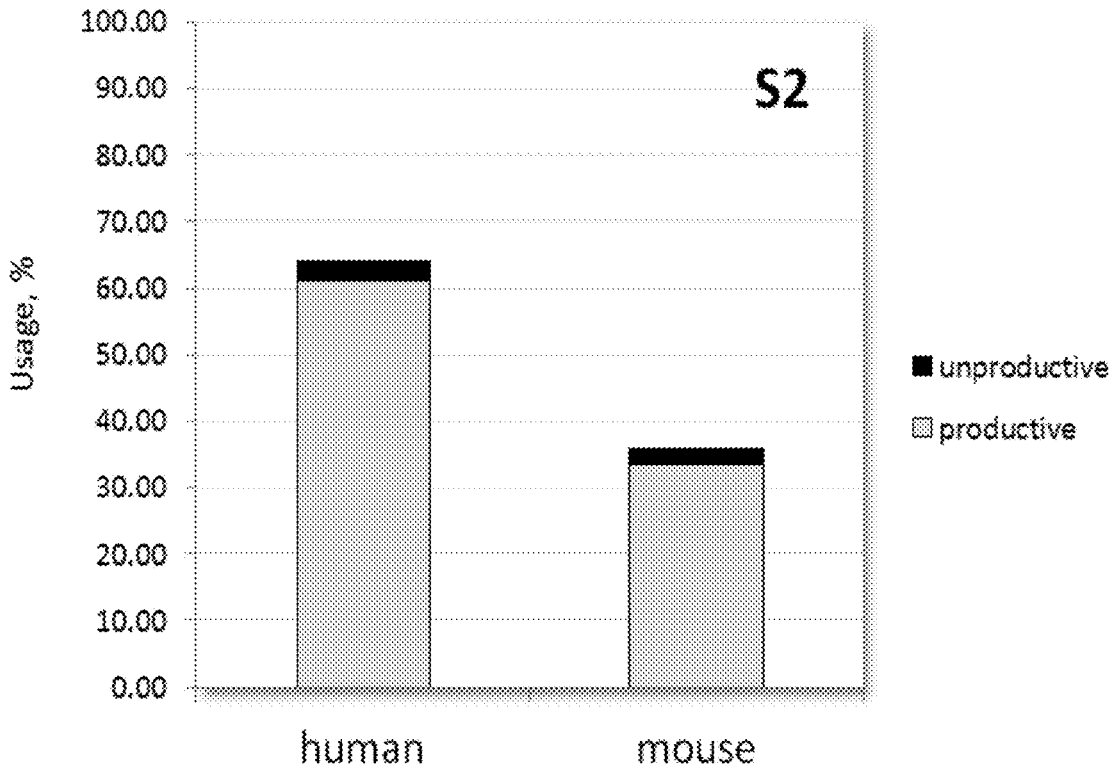


Fig. 50

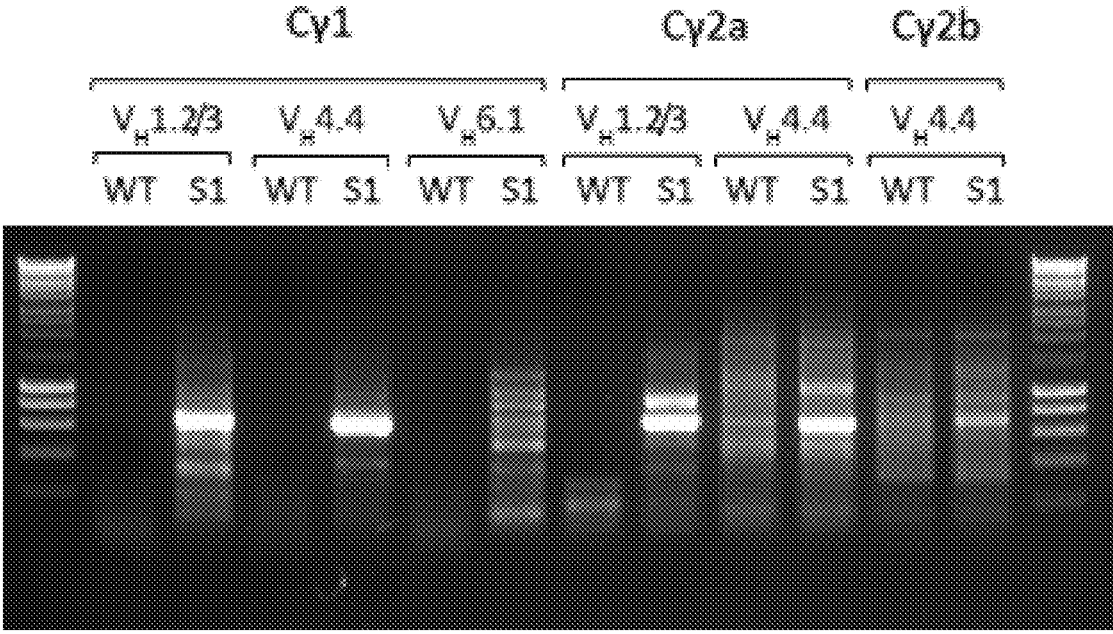


Fig. 51

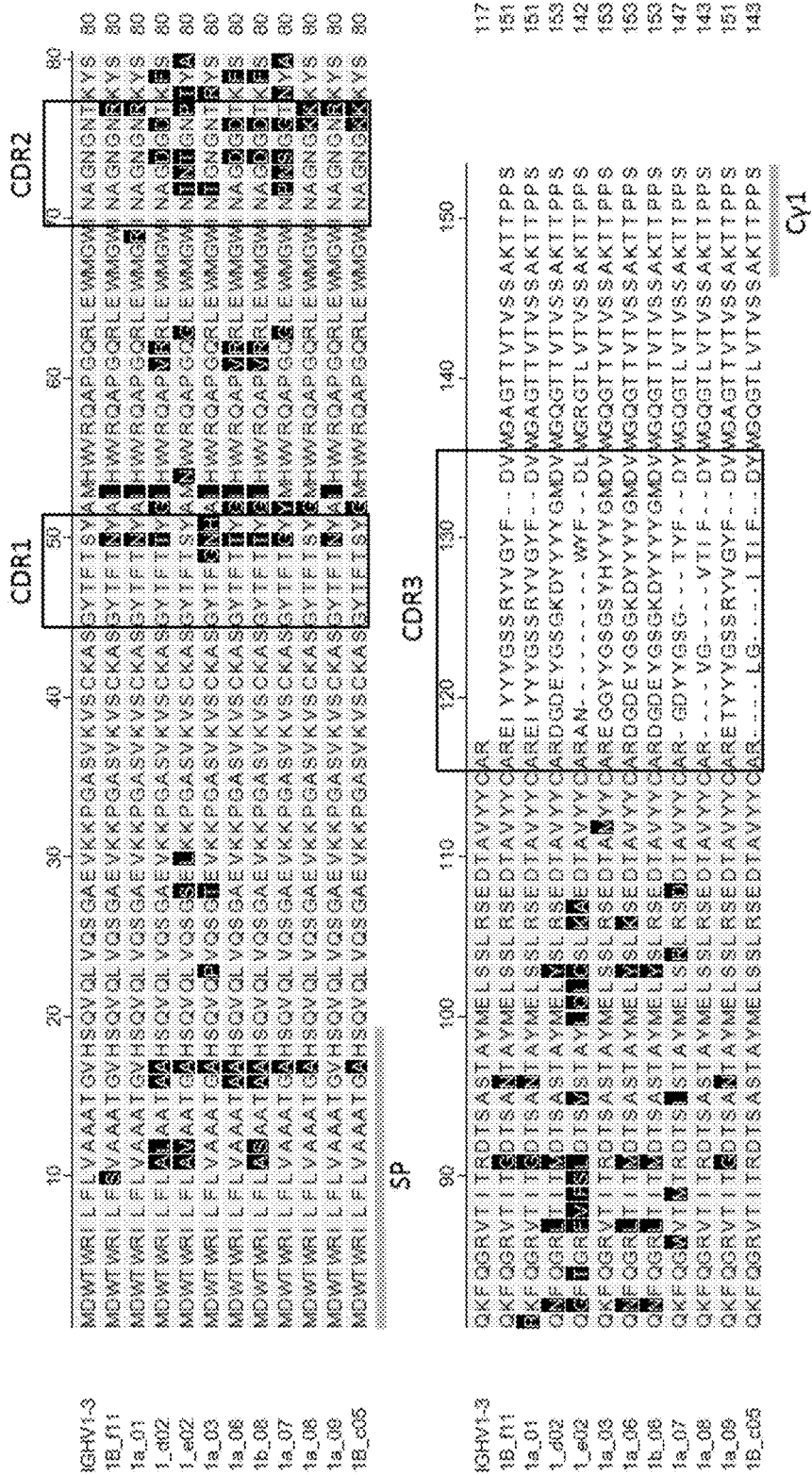


Fig. 52

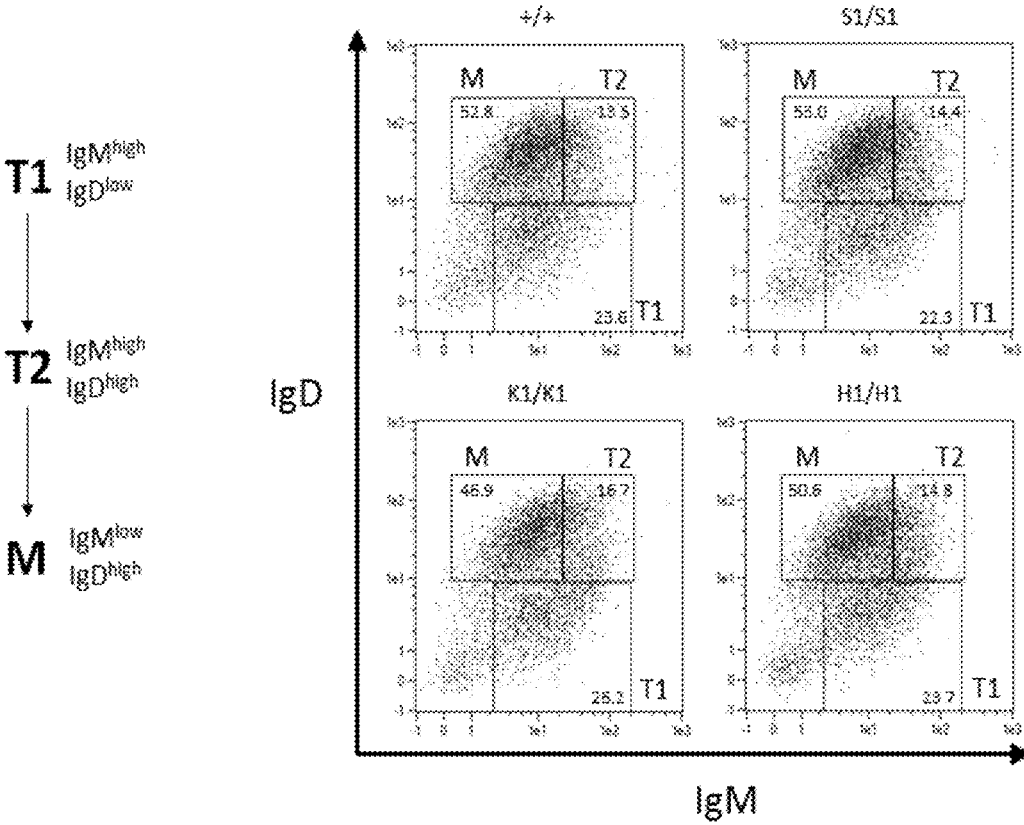
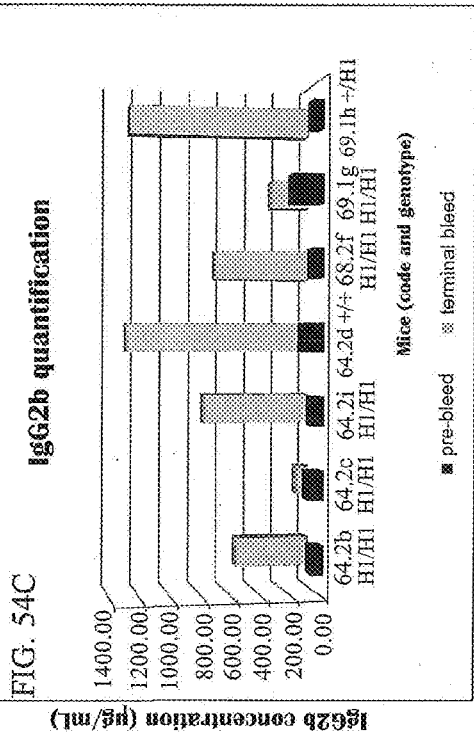
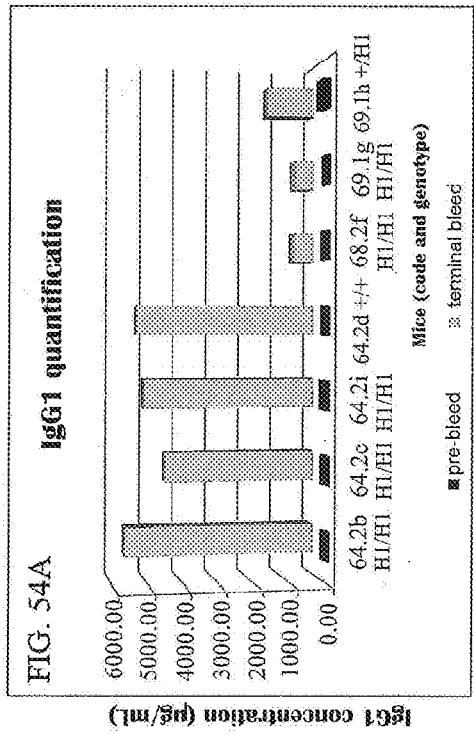
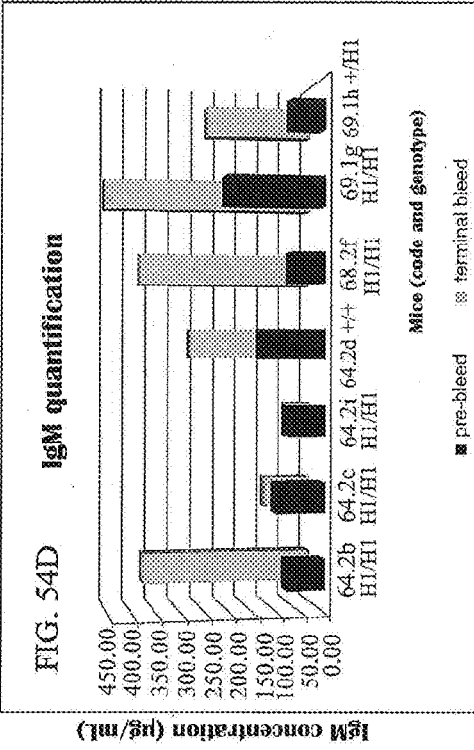
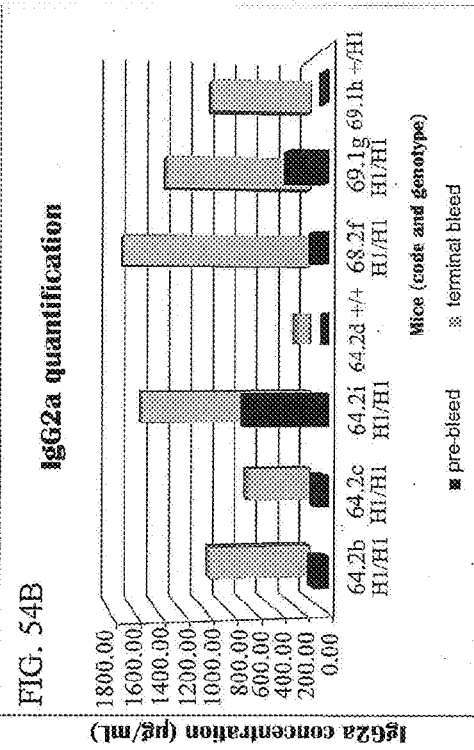
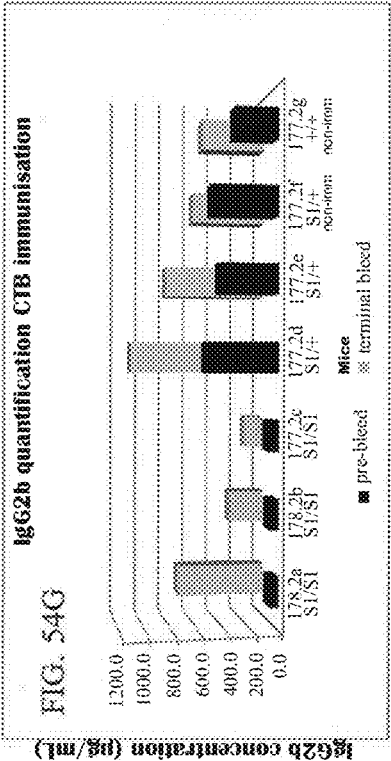
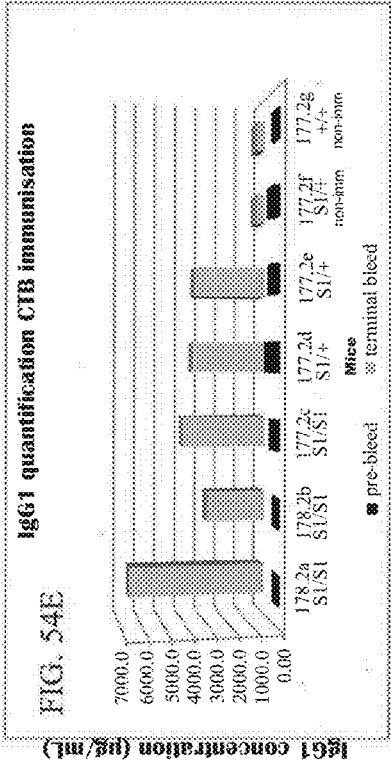
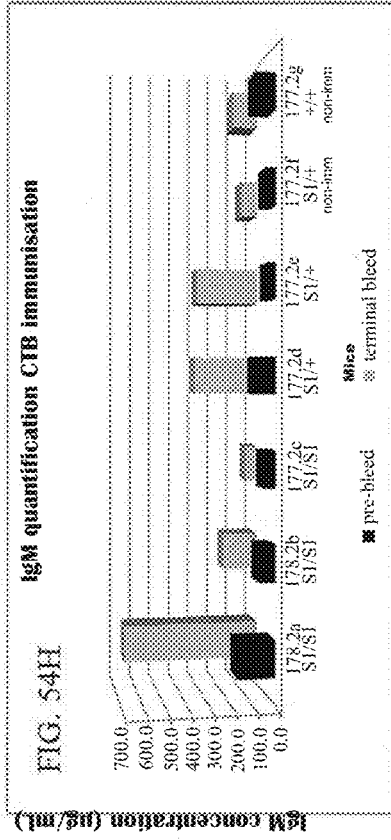
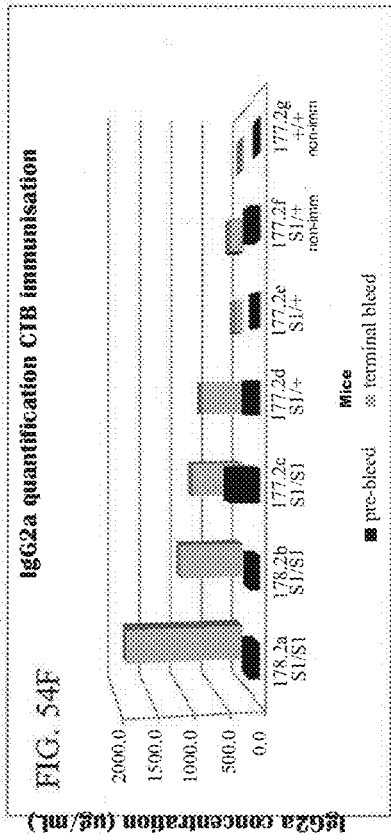


Fig. 53





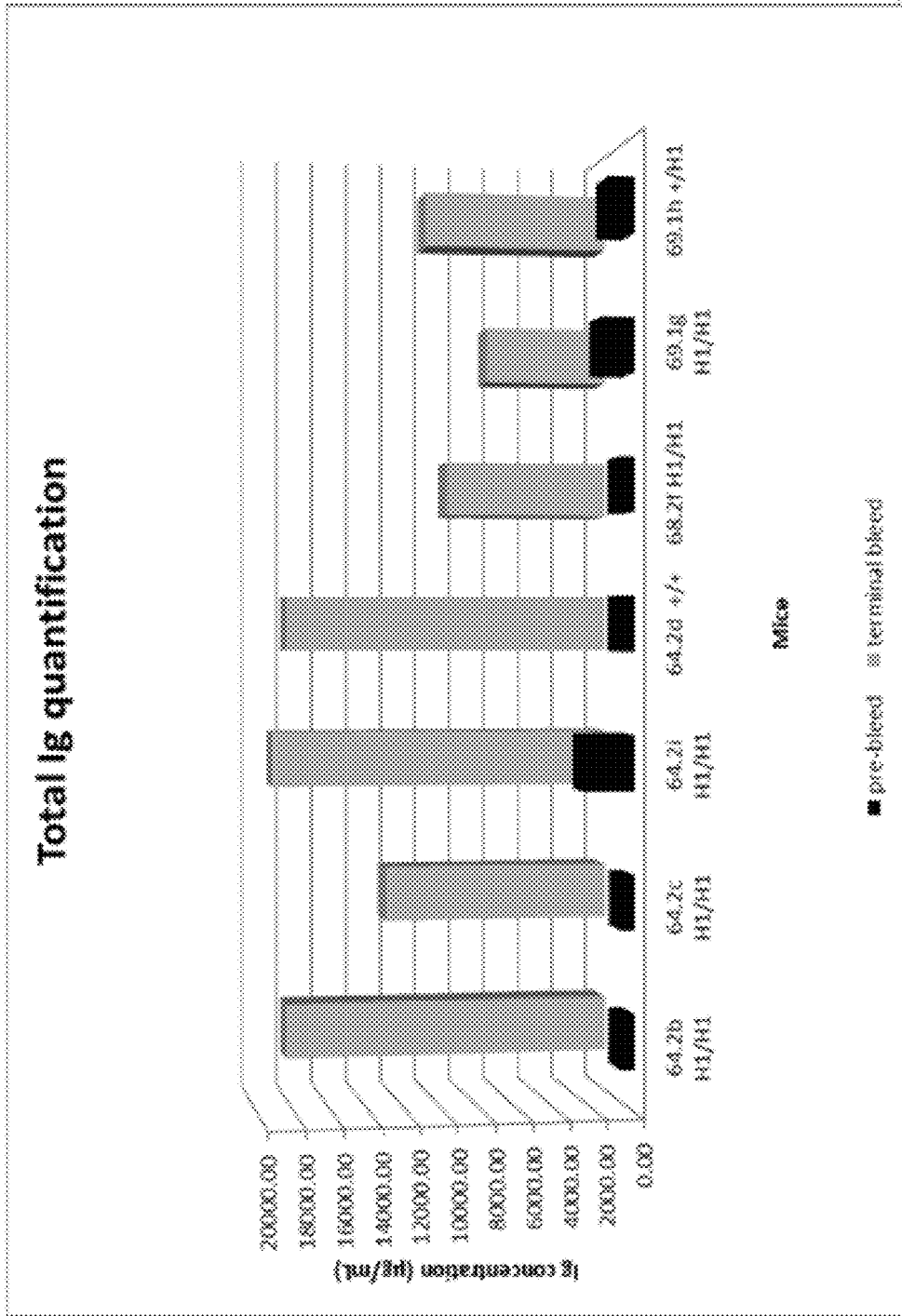


Fig. 55A

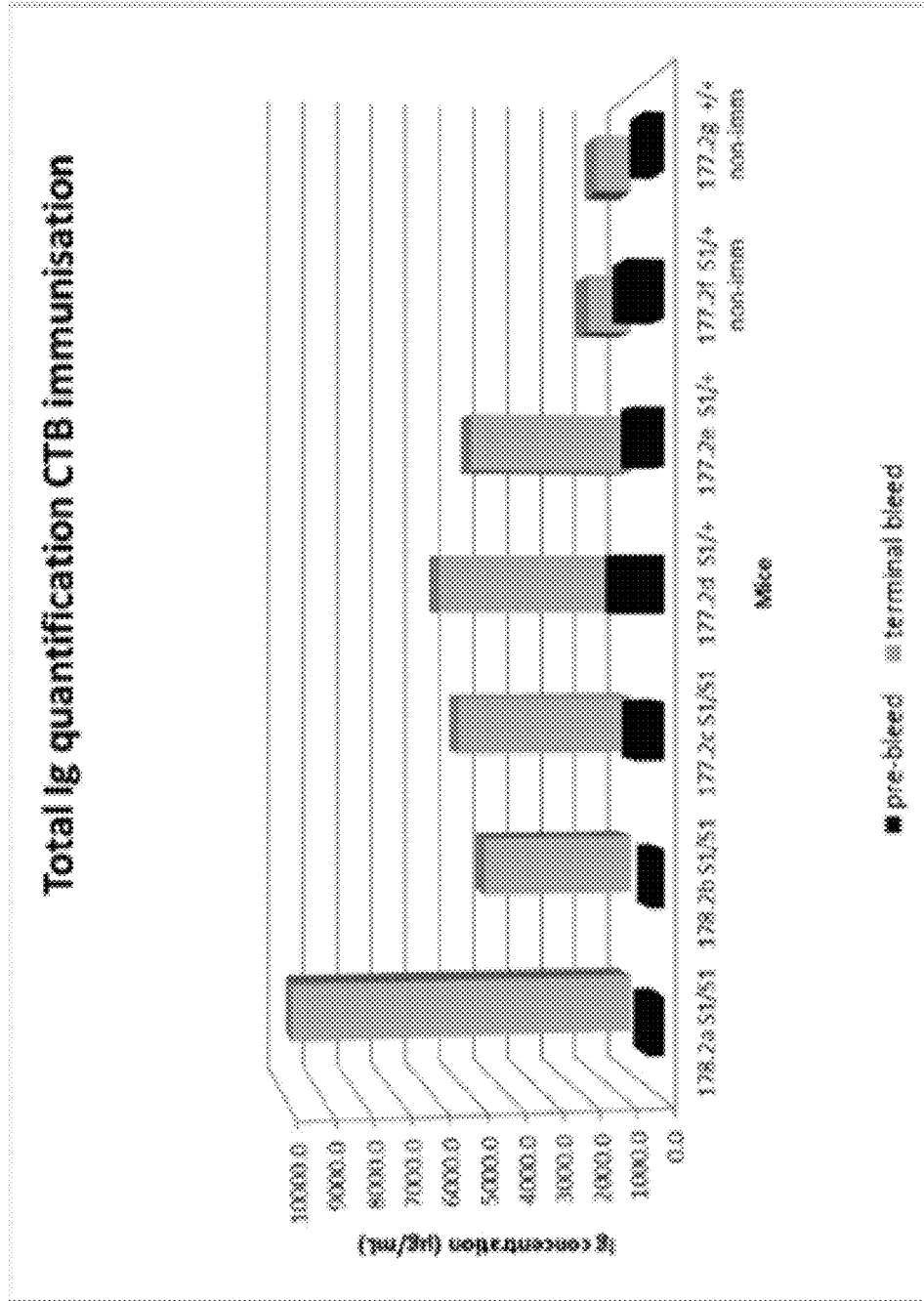


Fig. 55B

ANIMAL MODELS AND THERAPEUTIC MOLECULES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a Continuation application of Ser. No. 14/137,902, filed Dec. 20, 2013, which is a Continuation application of U.S. Ser. No. 13/740,727, filed Jan. 14, 2013, which is a Divisional application of U.S. Ser. No. 13/310,431, filed Dec. 2, 2011, which is a Continuation-in-Part of PCT/GB2010/051122, filed Jul. 7, 2010, which claims the benefit of U.S. Provisional Application No. 61/223,960 filed Jul. 8, 2009; U.S. Provisional Application No. 61/355,666, filed Jun. 17, 2010; GB Patent Application No. 0911846.4, filed Jul. 8, 2009; and GB Patent Application No. 0913102.0, filed Jul. 28, 2009. U.S. application Ser. No. 13/310,431 is also a Continuation-in-Part of PCT/GB2011/050019, filed Jan. 7, 2011. The entire contents of the above-referenced applications are incorporated herein by reference.

BACKGROUND

[0002] The present invention relates inter alia to non-human animals and cells that are engineered to contain exogenous DNA, such as human immunoglobulin gene DNA, their use in medicine and the study of disease, methods for production of non-human animals and cells, and antibodies and antibody chains produced by such animals and derivatives thereof.

[0003] In order to get around the problems of humanizing antibodies a number of companies set out to generate mice with human immune systems. The strategy used was to knockout the heavy and light chain loci in ES cells and complement these genetic lesions with transgenes designed to express the human heavy and light chain genes. Although fully human antibodies could be generated, these models have several major limitations:

(i) The size of the heavy and light chain loci (each several Mb) made it impossible to introduce the entire loci into these models. As a result the transgenic lines recovered had a very limited repertoire of V-regions, most of the constant regions were missing and important distant enhancer regions were not included in the transgenes.

(ii) The very low efficiency of generating the large insert transgenic lines and the complexity and time required to cross each of these into the heavy and light chain knockout strains and make them homozygous again, restricted the number of transgenic lines which could be analysed for optimal expression.

(iii) Individual antibody affinities rarely reached those which could be obtained from intact (non-transgenic) animals.

WO2007117410 discloses chimaeric constructs for expressing chimaeric antibodies.

WO2010039900 discloses knock in cells and mammals having a genome encoding chimaeric antibodies.

[0004] The present invention provides, inter alia, a process for the generation in non-human mammals of antibodies that comprise a human Ig variable region, and further provides non-human animal models for the generation of such antibodies.

SUMMARY OF THE INVENTION

[0005] All nucleotide co-ordinates for the mouse are those corresponding to NCBI m37 for the mouse C57BL/6J strain, e.g. April 2007 ENSEMBL Release 55.37h, e.g. NCBI37 July 2007 (NCBI build 37) (e.g. UCSC version mm9 see World Wide Web (www) genome.ucsc.edu and World Wide Web (www) genome.ucsc.edu/FAQ/FAQreleases.html) unless otherwise specified. Human nucleotides coordinates are those corresponding to GRCh37 (e.g. UCSC version hg 19, World Wide Web (www) genome.ucsc.edu/FAQ/FAQreleases.html), February 2009 ENSEMBL Release 55.37, or are those corresponding to NCBI36, Ensemble release 54 unless otherwise specified. Rat nucleotides are those corresponding to RGSC 3.4 Dec. 2004 ENSEMBL release 55.34w, or Baylor College of Medicine HGSC v3.4 Nov. 2004 (e.g., UCSC rn4, see World Wide Web (www) genome.ucsc.edu and World Wide Web (www) genome.ucsc.edu/FAQ/FAQreleases.html) unless otherwise specified.

[0006] In the present invention, methods are disclosed for constructing a chimaeric human heavy and light chain loci in a non-human mammal, for example a mouse. Reference to work in mice herein is by way of example only, and reference to mice is taken to include reference to all non-human mammals unless otherwise apparent from the disclosure, with mice being preferred as the non-human mammal.

[0007] In one aspect the invention relates to a non-human mammal whose genome comprises:

[0008] (a) a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region; and

[0009] (b) optionally one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region;

wherein the non-human mammal is able to produce a repertoire of chimaeric antibodies, or chimaeric light or heavy chains, having a non-human mammal constant region and a human variable region.

[0010] In one aspect the invention relates to non-human mammal whose genome comprises

[0011] (a) a plurality of human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or a plurality of human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region; and

[0012] (b) optionally one or more human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region;

wherein the non-human mammal is able to produce a repertoire of chimaeric antibodies, or chimaeric light or heavy chains, having a non-human mammal constant region and a human variable region.

[0013] In one aspect the invention relates to non-human mammalian cell whose genome comprises

- [0014]** (a) a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region and
- [0015]** (b) optionally one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region.
- [0016]** In one aspect the invention relates to a non-human mammalian cell whose genome comprises
- [0017]** (a) a plurality of human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or a plurality of human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region; and
- [0018]** (b) optionally one or more human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region;
- [0019]** In a further aspect the invention relates to a method for producing a non-human cell or mammal comprising inserting into a non-human mammal cell genome, such as an ES cell genome;
- [0020]** (a) a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region; and
- [0021]** (b) optionally one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region; respectively, the insertion being such that the non-human cell or mammal is able to produce a repertoire of chimaeric antibodies having a non-human mammal constant region and a human variable region, wherein steps (a) and (b) can be carried out in either order and each of steps (a) and (b) can be carried out in a stepwise manner or as a single step. Insertion may be by homologous recombination.
- [0022]** In a further aspect the invention relates to a method for producing an antibody or antibody chain specific to a desired antigen the method comprising immunizing a transgenic non-human mammal as disclosed herein with the desired antigen and recovering the antibody or antibody chain.
- [0023]** In a further aspect the invention relates to a method for producing a fully humanised antibody comprising immunizing a transgenic non-human mammal as disclosed herein with the desired antigen, recovering the antibody or cells producing the antibody and then replacing the non-human mammal constant region with a human constant region, for example by protein or DNA engineering.
- [0024]** In a further aspect the invention relates to humanised antibodies and antibody chains produced according to the present invention, both in chimaeric (for example, mouse-human) and fully humanised form, as well as fragments and derivatives of said antibodies and chains, and use of said antibodies, chains and fragments in medicine, including diagnosis.
- [0025]** In a further aspect the invention relates to use of a non-human mammal as described herein as a model for the testing of drugs and vaccines.
- [0026]** In one aspect the invention relates to a non-human mammal whose genome comprises:
- [0027]** (a) a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region; and
- [0028]** (b) optionally one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region;
- wherein the non-human mammal is able to produce a repertoire of chimaeric antibodies or antibody chains having a non-human mammal constant region and a human variable region.
- [0029]** In a further aspect the invention relates to a non-human mammal whose genome comprises:
- [0030]** (a) a plurality of human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or a plurality of human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region; and
- [0031]** (b) optionally one or more human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant;
- wherein the non-human mammal is able to produce a repertoire of chimaeric antibodies having a non-human mammal constant region and a human variable region.
- [0032]** Optionally the non-human mammal genome is modified to prevent expression of fully host-species specific antibodies.
- [0033]** In one aspect the inserted human DNA comprises at least 50% of the human heavy chain variable (V) genes, such as at least 60%, at least 70%, at least 80%, at least 90%, and in one aspect all of the human V genes.
- [0034]** In one aspect the inserted human DNA comprises at least 50% of the human heavy chain diversity (D) genes, such as at least 60%, at least 70%, at least 80%, at least 90%, and in one aspect all of the human D genes.
- [0035]** In one aspect the inserted human DNA comprises at least 50% of the human heavy chain joining (J) genes, such as at least 60%, at least 70%, at least 80%, at least 90%, and in one aspect all of the human J genes.
- [0036]** In one aspect the inserted human DNA comprises at least 50% of the human light chain Variable (V) genes, such as at least 60%, at least 70%, at least 80%, at least 90%, and in one aspect all of the human light chain V genes.
- [0037]** In one aspect the inserted human DNA comprises at least 50% of the human light chain joining (J) genes, such as at least 60%, at least 70%, at least 80%, at least 90%, and in one aspect all of the human light chain J genes.

[0038] The inserted human genes may be derived from the same individual or different individuals, or be synthetic or represent human consensus sequences.

[0039] Although the number of V D and J regions is variable between human individuals, in one aspect there are considered to be 51 human V genes, 27 D and 6 J genes on the heavy chain, 40 human V genes and 5 J genes on the kappa light chain and 29 human V genes and 4 J genes on the lambda light chain (Janeway and Travers, Immunobiology, Third edition)

[0040] In one aspect the human heavy chain locus inserted into the non-human mammal contains the full repertoire of human V, D and J regions, which in the genome is in functional arrangement with the non-human mammal constant regions such that functional chimaeric antibodies can be produced between the human variable and non-human mammal constant regions. This total inserted human heavy chain genetic material is referred to herein as the human IgH VDJ region, and comprises DNA from a human genome that encodes all the exons encoding human V,D and J portions and suitably also the associated introns. Similarly, reference to the human Ig light chain kappa V and J regions herein refers to human DNA comprising all the exons encoding V and J regions and suitably also the associated introns of the human genome. Reference to the human Ig light chain lambda V and J regions herein refers to human DNA comprising all the exons encoding V and J regions and suitably also the associated introns of the human genome.

[0041] Human variable regions are suitably inserted upstream of a non-human mammal constant region, the latter comprising all of the DNA required to encode the full constant region or a sufficient portion of the constant region to allow the formation of an effective chimaeric antibody capable of specifically recognising an antigen.

[0042] In one aspect the chimaeric antibodies or antibody chains have a part of a host constant region sufficient to provide one or more effector functions seen in antibodies occurring naturally in a host mammal, for example that they are able interact with Fc receptors, and/or bind to complement.

[0043] Reference to a chimaeric antibody or antibody chain having a host non mammal constant region herein therefore is not limited to the complete constant region but also includes chimaeric antibodies or chains which have all of the host constant region, or a part thereof sufficient to provide one or more effector functions. This also applies to non-human mammals and cells and methods of the invention in which human variable region DNA may be inserted into the host genome such that it forms a chimaeric antibody chain with all or part of a host constant region. In one aspect the whole of a host constant region is operably linked to human variable region DNA.

[0044] The host non-human mammal constant region herein is preferably the endogenous host wild-type constant region located at the wild type locus, as appropriate for the heavy or light chain. For example, the human heavy chain DNA is suitably inserted on mouse chromosome 12, suitably adjacent the mouse heavy chain constant region.

[0045] In one aspect the insertion of the human DNA, such as the human VDJ region is targeted to the region between the J4 exon and the Cp locus in the mouse genome IgH locus, and in one aspect is inserted between co-ordinates 114,667,090 and 114,665,190, or at co-ordinate 114,667,091, after 114,667,090. In one aspect the insertion of the

human DNA, such as the human light chain kappa VJ is targeted into mouse chromosome 6 between co-ordinates 70,673,899 and 70,675,515, suitably at position 70,674,734, or an equivalent position in the lambda mouse locus on chromosome 16.

[0046] In one aspect the host non-human mammal constant region for forming the chimaeric antibody may be at a different (non endogenous) chromosomal locus. In this case the inserted human DNA, such as the human variable VDJ or VJ region(s) may then be inserted into the non-human genome at a site which is distinct from that of the naturally occurring heavy or light constant region. The native constant region may be inserted into the genome, or duplicated within the genome, at a different chromosomal locus to the native position, such that it is in a functional arrangement with the human variable region such that chimaeric antibodies of the invention can still be produced.

[0047] In one aspect the human DNA is inserted at the endogenous host wild-type constant region located at the wild type locus between the host constant region and the host VDJ region.

[0048] Reference to location of the variable region upstream of the non-human mammal constant region means that there is a suitable relative location of the two antibody portions, variable and constant, to allow the variable and constant regions to form a chimaeric antibody or antibody chain in vivo in the mammal. Thus, the inserted human DNA and host constant region are in functional arrangement with one another for antibody or antibody chain production.

[0049] In one aspect the inserted human DNA is capable of being expressed with different host constant regions through isotype switching. In one aspect isotype switching does not require or involve trans switching. Insertion of the human variable region DNA on the same chromosome as the relevant host constant region means that there is no need for trans-switching to produce isotype switching.

[0050] As explained above, the transgenic loci used for the prior art models were of human origin, thus even in those cases when the transgenes were able to complement the mouse locus so that the mice produced B-cells producing fully human antibodies, individual antibody affinities rarely reached those which could be obtained from intact (non-transgenic) animals. The principal reason for this (in addition to repertoire and expression levels described above) is the fact that the control elements of the locus are human. Thus, the signalling components, for instance to activate hyper-mutation and selection of high affinity antibodies are compromised.

[0051] In contrast, in the present invention, host non-human mammal constant regions are maintained and it is preferred that at least one non-human mammal enhancer or other control sequence, such as a switch region, is maintained in functional arrangement with the non-human mammal constant region, such that the effect of the enhancer or other control sequence, as seen in the host mammal, is exerted in whole or in part in the transgenic animal.

[0052] This approach above is designed to allow the full diversity of the human locus to be sampled, to allow the same high expression levels that would be achieved by non-human mammal control sequences such as enhancers, and is such that signalling in the B-cell, for example isotype switching using switch recombination sites, would still use non-human mammal sequences.

[0053] A mammal having such a genome would produce chimaeric antibodies with human variable and non-human mammal constant regions, but these could be readily humanized, for example in a cloning step. Moreover the in vivo efficacy of these chimaeric antibodies could be assessed in these same animals.

[0054] In one aspect the inserted human IgH VDJ region comprises, in germline configuration, all of the V, D and J regions and intervening sequences from a human.

[0055] In one aspect 800-1000 kb of the human IgH VDJ region is inserted into the non-human mammal IgH locus, and in one aspect a 940, 950 or 960 kb fragment is inserted. Suitably this includes bases 105,400,051 to 106,368,585 from human chromosome 14.

[0056] In one aspect the inserted IgH human fragment consists of bases 105,400,051 to 106,368,585 from chromosome 14. In one aspect the inserted human heavy chain DNA, such as DNA consisting of bases 105,400,051 to 106,368,585 from chromosome 14, is inserted into mouse chromosome 12 between the end of the mouse J4 region and the E μ region, suitably between co-ordinates 114,667,090 and 114,665,190, or at co-ordinate 114,667,091, after 114,667,090. In one aspect the insertion is between co-ordinates 114,667,089 and 114,667,090 (co-ordinates refer to NCBI m37, for the mouse C57BL/6J strain), or at equivalent position in another non-human mammal genome.

[0057] In one aspect the inserted human kappa VJ region comprises, in germline configuration, all of the V and J regions and intervening sequences from a human. Suitably this includes bases 88,940,356 to 89,857,000 from human chromosome 2, suitably approximately 917 kb. In a further aspect the light chain VJ insert may comprise only the proximal clusters of V segments and J segments. Such an insert would be of approximately 473 kb. In one aspect the human light chain kappa DNA, such as the human IgK fragment of bases 88,940,356 to 89,857,000 from human chromosome 2, is suitably inserted into mouse chromosome 6 between co-ordinates 70,673,899 and 70,675,515, suitably at position 70,674,734. These co-ordinates refer to NCBI36 for the human genome, ENSEMBL Release 54 and NCBI37 for the mouse genome, relating to mouse strain C57BL/6J.

[0058] In one aspect the human lambda VJ region comprises, in germline configuration, all of the V and J regions and intervening sequences from a human.

[0059] Suitably this includes analogous bases to those selected for the kappa fragment, from human chromosome 2.

[0060] A cell or non-human mammal of the invention, in one embodiment, comprises an insertion of human heavy chain variable region DNA between co-ordinates 114, 666, 183 and 114, 666, 725, such as between 114 666 283 and 114 666 625, optionally between co-ordinates 114,666,335 and 114,666,536, optionally between 114,666,385 and 114,666, 486, or between 114,666,425 and 114,666,446, or between 114,666,435 and 114,666,436 of mouse chromosome 12 with reference to NCBI37 for the mouse genome, relating to mouse strain C57BL/6J or an equivalent position of mouse chromosome 12 from a different mouse strain or an equivalent position in the genome of another non-human vertebrate, e.g., a rat. The insertion between co-ordinates 114,666,435 and 114,666,436 relating to mouse strain C57BL/6J is equivalent to an insertion between co-ordinates 1207826 and 1207827 on chromosome 12 with reference to

the 129/SvJ genomic sequence of the GenBank® access number NT114985.2. An insertion may be made at equivalent position in another genome, such as another mouse genome. In an example of this embodiment, the cell or mammal of the invention comprises a human IgH VDJ region which comprises or consists of nucleotides 106,328, 851-107,268,544, such as nucleotides 106,328,901-107,268, 494, such as nucleotides 106,328,941-107,268,454, such as nucleotides 106,328,951-107,268,444 of human Chromosome 14, with reference to the GRCH37/hg19 sequence database, or insertion of equivalent nucleotides relating to chromosome 14 from a different human sequence or database. The human insertion may be made between the regions indicated above.

[0061] A cell or mammal of the invention, in one embodiment, comprises an insertion of the human kappa VJ region, suitably comprising or consisting of, in germline configuration, all of the V and J regions and intervening sequences from a human, the insertion of the human DNA being made between co-ordinates 70,673,918-70,675,517, such as between co-ordinates 70, 674,418 and 70 675, 017, such as between co-ordinates 70,674, 655-70,674,856, such as between co-ordinates 70,674, 705-70,674,906, such as between co-ordinates 70,674, 745-70,674,766, such as between co-ordinates 70,674,755 and 70,674,756 of mouse chromosome 6, numbering with reference to NCBI37 for the mouse genome, relating to mouse strain C57BL/6J, or an insertion at an equivalent position in another genome, such as another mouse genome. In an example of this embodiment, a cell or mammal of the invention comprises an insertion of nucleotides 89,159,079-89,630,437 and/or 89,941,714-90,266,976 of human chromosome 2 with reference to the GRCH37/hg19 sequence database (or equivalent nucleotides relating to chromosome 2 from a different human sequence or database), such as an insertion of these 2 discrete fragments without the intervening sequence, or an insertion of the complete 89,159,079-90,266,976 region.

[0062] The insertion may comprise, or consist, of:

[0063] (i) nucleotides 89,158,979-89,630,537, such as 89,159,029-89,630,487, such as 89,159,069-89,630, 447, such as 89,159,079-89,630,437, optionally in addition to fragment (ii) below

[0064] (ii) nucleotides 89,941,614-90,267,076, such as 89,941,664-90,267,026, such as 89, 941,704-90,266, 986, such as 89,941,714-90,266,976; optionally in addition to fragment (i)

[0065] (iii) nucleotides 89,158,979-90,267,076, such as nucleotides 89,159,079-90,266,976.

[0066] The human insertion may be made between the regions indicated above.

[0067] In an embodiment, a cell or mammal of the invention comprises an insertion of a human lambda region which comprises at least one human J λ region (eg, a germline region) and at least one human C λ region (eg, a germline region), optionally C λ 6 and/or C λ 7. For example, the cell or mammal comprises a plurality of human J λ regions, optionally two or more of J λ 1, J λ 2, J λ 6 and J λ 7, optionally all of J λ 1, J λ 2, J λ 6 and J λ 7. In an example, the cell or mammal comprises at least one human J λ -C λ cluster, optionally at least J λ 7-C λ 7.

[0068] In one aspect the human JC cluster is inserted 3' of the last endogenous J lambda or is inserted 3' of the last

endogenous J kappa region, suitably immediately 3' of these sequences, or substantially immediately 3' of these sequences.

[0069] In one aspect the insertion into the mouse lambda locus is made downstream of the endogenous C1 gene segment, for example where there is a 3' J1C1 cluster, suitably immediately 3' of the C1 segment, or substantially immediately 3' of the segment.

[0070] In one aspect (e.g. cell or non-human mammal) a human JC cluster is inserted into a kappa locus and any resulting cell or animal is heterozygous at that locus, such that the cell has one chromosome with human lambda DNA inserted into the kappa locus, and another chromosome with human kappa DNA at the endogenous kappa locus.

[0071] In an embodiment, a cell or mammal of the invention comprises a human EA enhancer.

[0072] A cell or mammal may of the invention comprise an inserted human lambda VJ region, suitably comprising or consisting of, in germline configuration, all of the V and J regions and intervening sequences from a human, the inserted region comprises or consisting of nucleotides 22,375,509-23,327,984, such as nucleotides 22,375,559-23,327,934, such as nucleotides 22,375,599-23,327,894, such as nucleotides 22,375,609-23,327,884 from human Chromosome 22, with reference to the GRCH37/hg19 sequence database, or equivalent DNA from another human sequence or database. The insertion into the mouse genome may be made between co-ordinates 19,027,763 and 19,061,845, such as between co-ordinates 19,037,763 and 19,051,845, such as between co-ordinates 19,047,451 and 19,047,652, such as between co-ordinates 19,047,491 and 19,047,602, such as between co-ordinates 19,047,541 and 19,047,562, such as between co-ordinates 19,047,551 and 19,047,552 of mouse Chromosome 16 (with reference to NCBIM37 for the mouse genome, relating to mouse strain C57BL/6J, equivalent to co-ordinates 1,293,646-1,293,647 of the 129 SvJ genomic sequence in the sequence file of NT_039630.4), or may be an insertion at an equivalent position in other genome, such as another mouse genome. The insertion of the human lambda nucleic acid into the mouse genome may alternatively be made between co-ordinates 70,673,918 and 70,675,517, such as between co-ordinates 70,674,418 and 70,675,017, such as between co-ordinates 70,674,655 and 70,674,856, such as between co-ordinates 70,674,705 and 70,674,806, such as between co-ordinates 70,674,745 and 70,674,766, such as between co-ordinates 70,674,755 and 70,674,756 of mouse Chromosome 6 (with reference to NCBIM37 for the mouse genome, relating to mouse strain C57BL/6J) or equivalent in another genome. The human insertion may be made between the regions indicated above.

[0073] All specific human fragments described above may vary in length, and may for example be longer or shorter than defined as above, such as 500 bases, 1 KB, 2K, 3K, 4K, 5 KB, 10 KB, 20 KB, 30 KB, 40 KB or 50 KB or more, which suitably comprise all or part of the human V(D)J region, whilst preferably retaining the requirement for the final insert to comprise human genetic material encoding the complete heavy chain region and light chain region, as appropriate, as described above.

[0074] In one aspect the 5' end of the human insert described above is increased in length. Where the insert is generated in a stepwise fashion then the increase in length is generally in respect of the upstream (5') clone.

[0075] In one aspect the 3' end of the last inserted human gene, generally the last human J gene to be inserted is less than 2 kb, preferably less than 1 KB from the human-mouse join region.

[0076] In one aspect the non-human mammal comprises some or all of the human light chain kappa VJ region as disclosed herein but not the human light chain lambda VJ region.

[0077] In one aspect the cell or non-human mammal comprises a fully human lambda locus (lambda VJC regions from a human), a chimaeric kappa locus (human kappa VJ regions operatively linked to a host kappa constant region) and a chimaeric heavy chain locus, having a human VDJ region operatively linked to a host heavy chain constant region.

[0078] In a further aspect the genome comprises an insertion of V, D (heavy chain only) and J genes as described herein at the heavy chain locus and one light chain locus, or at the heavy chain locus and both light chain loci. Preferably the genome is homozygous at one, or both, or all three loci.

[0079] In another aspect the genome may be heterozygous at one or more of the loci, such as heterozygous for DNA encoding a chimaeric antibody chain and native (host cell) antibody chain. In one aspect the genome may be heterozygous for DNA capable of encoding 2 different antibody chains of the invention, for example, comprising 2 different chimaeric heavy chains or 2 different chimaeric light chains.

[0080] In one aspect the invention relates to a non-human mammal or cell, and methods for producing said mammal or cell, as described herein, wherein the inserted human DNA, such as the human IgH VDJ region and/or light chain V, J regions are found on only one allele and not both alleles in the mammal or cell. In this aspect a mammal or cell has the potential to express both an endogenous host antibody heavy or light chain and a chimaeric heavy or light chain.

[0081] In a further aspect of the invention the human VDJ region, or light chain VJ region, is not used in its entirety, but parts of the equivalent human VDJ or VJ region, such as the exons, from other species may be used, such as one or more V, D, or J exons from other species, or regulatory sequences from other species. In one aspect the sequences used in place of the human sequences are not human or mouse. In one aspect the sequences used may be from rodent, or, primate such as chimp. For example, 1, 2, 3, 4, or more, or all of the J regions from a primate other than a human may be used to replace, one, 2, 3, 4, or more or all of the human J exons in the VDJ/VJ region of the cells and animals of the invention.

[0082] In a further aspect the inserted human DNA, such as the human IgH VDJ region, and/or light chain VJ regions, may be inserted such that they are operably linked in the genome with a mu constant region from a non-human, non-mouse species, such as a rodent or primate sequence, such as a rat sequence.

[0083] Other non-human, non-mouse species from which DNA elements may be used in the present invention include rabbits, llamas, dromedary, alpacas, camels and sharks.

[0084] In one aspect the inserted human DNA, such as the human VDJ or VJ region, is not operably linked to the endogenous host mu sequence but rather to a non-host mu sequence.

[0085] Operable linkage suitably allows production of an antibody heavy or light chain comprising the human variable region.

[0086] In one aspect the inserted human DNA, such as the human IgH VDJ region (and/or light chain VJ regions) may be inserted into the host chromosome together with mu constant region nucleic acid which is not host mu constant region nucleic acid, and preferably is a mu constant region from a non-mouse, non-human species. Suitably the inserted human DNA, such as the human VDJ region (and/or light chain VJ regions) is operably linked to a non-human, non-mouse mu, and is able to form a chimaeric antibody heavy or light chain. In another aspect a non-mouse, non-human mu may be inserted into the host chromosome on a separate genetic element to that of the human variable region, or at a different location in the genome, suitably operably linked to the variable region such that a chimaeric antibody heavy or light can be formed.

[0087] In an additional aspect the invention relates to a non-human mammal or a cell whose genome comprises a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of a host non-human mammal light chain constant region, arranged such that the cell or mammal is able to express a chimaeric antibody chain. The invention also relates to a non-human mammal or a cell whose genome additionally or alternatively comprises a plurality of human Ig light chain V regions, and one or more human J regions upstream of a host non-human mammal heavy chain constant region, such that the cell or mammal is able to express a chimaeric antibody chain. The cell or mammal may be able to express an antibody having both heavy and light chains, including at least one chimaeric antibody chain, as disclosed above.

[0088] The inserted human heavy chain variable regions may be any of those described herein, and may be inserted at the positions described above for insertion 5' of the lambda and kappa constant regions. Likewise the inserted human light chain variable regions may be those described above, and may be inserted at the positions described above for insertion 5' of the heavy chain constant region.

[0089] For example, the genome or the cell or non-human mammal of the invention may encode an antibody comprising an antibody chain having a human heavy chain variable region upstream of a mouse light chain constant region, or an antibody chain having a human light chain variable region upstream of a mouse heavy chain constant region, in combination with one of: a fully human antibody light chain;

[0090] a fully human antibody heavy chain;

[0091] a non-human vertebrate (e.g., mouse or rat) antibody light chain;

[0092] a non-human vertebrate (e.g., mouse or rat) antibody heavy chain;

[0093] a chimaeric non-human vertebrate (e.g., mouse or rat)—human antibody chain;

[0094] an antibody chain having a human heavy chain variable region upstream of a non-human vertebrate (e.g., mouse or rat) light chain constant region;

[0095] an antibody chain having a human light chain variable region upstream of a non-human vertebrate (e.g., mouse or rat) heavy chain constant region.

[0096] The invention also relates to a transgene encoding a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of a host non-human mammal light chain constant region, optionally comprised within a vector.

[0097] The invention also relates to a transgene encoding a plurality of human Ig light chain V regions, and one or

more human light chain J regions upstream of a host non-human mammal heavy chain constant region, optionally comprised within a vector.

[0098] In one aspect the invention relates to a cell, or non-human mammal, the genome of which comprises: one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of all or part of the human kappa constant region.

[0099] In another aspect the invention relates to a cell, or non-human mammal, the genome of which comprises: one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of all or part of the human lambda constant region.

[0100] Suitably the light chain VJ and C regions are able to form antibody chains in vivo capable of specifically reacting with an antigen.

[0101] In one aspect of the invention there is no non-human coding sequence in the inserted light chain region.

[0102] In such aspects a human kappa and/or lambda region is inserted into the genome, in combination with insertion of the heavy chain VDJ region or part thereof, upstream of the host heavy chain constant region as disclosed herein.

[0103] The cell or non-human mammal of the invention may comprise:

[0104] (a) a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region; and

[0105] (b) one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of all or part of the non-human kappa constant region, wherein the non-human mammal is able to produce a repertoire of antibodies having an antibody chain comprising non-human mammal constant region and a human variable region.

[0106] The cell or non-human mammal of the invention may comprise

[0107] (a) a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region; and

[0108] one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region;

wherein the non-human mammal is able to produce a repertoire of antibodies having an antibody chain comprising a non-human mammal constant region and a human variable region.

[0109] Suitably the insertion of the human VJC light chain DNA, or part thereof as disclosed above, is made at the equivalent mouse locus. In one aspect the human light chain kappa VJC DNA, or part thereof, is inserted immediately upstream or downstream of the mouse kappa VJC region. In one aspect, the human light chain lambda VJC region or part thereof is inserted immediately upstream or downstream of the mouse lambda VJC region. In one aspect only the human kappa VJC locus is inserted and not the human lambda VJC locus. In one aspect only the human lambda VJC locus is inserted and not the human kappa VJC locus. Insertions may be made using the techniques disclosed herein, and suitably do not remove the host sequences from the genome. In one aspect the non-human mammal host VJC sequences may be

inactivated in some way, by mutation, or inversion, or by insertion of the human variable region DNA, or by any other means. In one aspect the cell or non-human mammal of the invention may comprise an insertion of the complete VJC human region.

[0110] The human kappa variable region DNA might be inserted into the genome in functional arrangement with a lambda constant region, for example inserted upstream of a lambda constant region. Alternatively human lambda region variable DNA might be inserted in functional arrangement with a kappa constant region, for example inserted upstream of a kappa constant region.

[0111] In one aspect one or more non-human mammal control sequences such as the enhancer sequence(s) is maintained upstream of the nonhuman mammal Mu constant region, suitably in its native position with respect to the distance from the constant region.

[0112] In one aspect one or more non-human mammal control sequences such as an enhancer sequence(s) are maintained downstream of the nonhuman mammal Mu constant region, suitably in its native position with respect to the distance from the constant region.

[0113] In one aspect a non-human mammal switch sequence, suitably the endogenous switch sequence, is maintained upstream of the non-human mammal Mu constant region, suitably in its native position with respect to distance from the constant region.

[0114] In such location the host enhancer or switch sequences are operative in vivo with the host constant region sequence(s).

[0115] In one aspect a switch sequence is neither human, nor native in the non-human mammal, for example in one aspect a non-human mammal switch sequence is not a mouse or human switch sequence. The switch sequence may be, for example, a rodent or primate sequence, or a synthetic sequence. In particular the switch sequence may be a rat sequence where the non-human mammal is a mouse. By way of example, a mouse or human constant mu sequence may be placed under the control of a switch sequence from a rat, or chimp, or other switch sequence, suitably capable of allowing isotype switching to occur in vivo.

[0116] In one aspect the switch sequence of the invention is a switch sequence comprising 3, 4, 5, 6 or more (up to 82) contiguous repeats of the repeat sequence GGGCT (SEQ ID no 46-50), such as a rat switch sequence. By "rat switch" herein it is meant that the switch is a wild-type switch corresponding to a switch from a rat genome or derived from such a switch.

[0117] In one aspect the switch sequence of the invention is a rat switch sequence comprising the following repeats: GAGCT (296 repeats; SEQ ID No 18), GGGGT (50 repeats; SEQ ID No 19), and GGGCT (83 repeats; SEQ ID No 20).

[0118] In one example the rat switch sequence comprises or consists of the sequence of SEQ ID no 1.

[0119] In these embodiments, and where the non-human mammal is a mouse or the cell is a mouse cell, the switch is optionally a rat switch as described herein.

[0120] Alternatively, the switch sequence present in cells or mammal of the invention is a mouse switch, eg, is from a mouse such as a mouse 129 strain or mouse C57 strain, or from a strain derived therefrom, optionally comprising or consisting of the sequence of SEQ ID no 4 or 5. By "mouse switch" herein it is meant that the switch is a wild-type switch corresponding to a switch from a mouse genome or

derived from such a switch. In this embodiment, and where the non-human mammal is a mouse or the cell is a mouse cell, the mouse switch sequence is optionally the endogenous switch or is a mouse switch from another mouse strain.

[0121] The cell or mammal of the invention may therefore comprise a human or non-human mammal switch sequence and a human or non-human mammal enhancer region or regions. They may be upstream of a human or non-human mammal constant region. Preferably the control sequences are able to direct expression or otherwise control the production of antibodies comprising a constant region with which they are associated. One combination envisaged is a rat switch with mouse enhancer sequences and mouse constant regions in a mouse cell.

[0122] In one aspect the invention relates to a cell, preferably a non-human cell, or non-human mammal comprising an immunoglobulin heavy chain or light chain locus having DNA from 3 or more species. For example, the cell or animal may comprise host cell constant region DNA, one or more human V, D or J coding sequences and one or more non-human, non-host DNA regions that are able to control a region of the immunoglobulin locus, such as a switch sequence, promoter or enhancer which are able to control expression or isotype switching in vivo of the Ig DNA. In one aspect the cell or animal is a mouse and comprises additionally human DNA from the human Ig locus and additionally a non-mouse DNA sequence, such as a rat DNA sequence, capable of regulation of the mouse or human DNA.

[0123] In another aspect the invention relates to a cell, preferably non-human cell, or non-human mammal comprising an immunoglobulin heavy chain or light chain locus having DNA from 2 or more different human genomes. For example, it could comprise heavy chain V(D)J sequences from more than one human genome within a heavy or light chain, or heavy chain VDJ DNA from one genome and light chain VJ sequences from a different genome.

[0124] In one aspect the invention relates to a DNA fragment or cell or non-human mammal comprising an immunoglobulin heavy chain or light chain locus, or part thereof, having DNA from 2 or more species, where one species contributes a non-coding region such as a regulatory region, and the other species coding regions such as V, D, J or constant regions.

[0125] In one aspect the human promoter and/or other control elements that are associated with the different human V, D or J regions are maintained after insertion of the human VDJ into the mouse genome.

[0126] In a further aspect one or more of the promoter elements, or other control elements, of the human regions, such as the human V regions, are optimised to interact with the transcriptional machinery of a non-human mammal.

[0127] Suitably a human coding sequence may be placed under the control of an appropriate non-human mammal promoter, which allows the human DNA to be transcribed efficiently in the appropriate non-human animal cell. In one aspect the human region is a human V region coding sequence, and a human V region is placed under the control of a non-human mammal promoter.

[0128] The functional replacement of human promoter or other control regions by non-human mammal promoter or control regions may be carried out by use of recombinering, or other recombinant DNA technologies, to insert a part of

the human Ig region (such as a human V region) into a vector (such as a BAC) containing a non-human Ig region. The recombinering/recombinant technique suitably replaces a portion of the non-human (e.g. mouse) DNA with the human Ig region, and thus places the human Ig region under control of the non-human mammal promoter or other control region. Suitably the human coding region for a human V region replaces a mouse V region coding sequence. Suitably the human coding region for a human D region replaces a mouse D region coding sequence. Suitably the human coding region for a human J region replaces a mouse J region coding sequence. In this way human V, D or J regions may be placed under the control of a non-human mammal promoter, such as a mouse promoter.

[0129] In one aspect the only human DNA inserted into the non-human mammalian cell or animal are V, D or J coding regions, and these are placed under control of the host regulatory sequences or other (non-human, non-host) sequences. In one aspect reference to human coding regions includes both human introns and exons, or in another aspect simply exons and no introns, which may be in the form of cDNA.

[0130] It is also possible to use recombinering, or other recombinant DNA technologies, to insert a non-human-mammal (e.g. mouse) promoter or other control region, such as a promoter for a V region, into a BAC containing a human Ig region. A recombinering step then places a portion of human DNA under control of the mouse promoter or other control region.

[0131] The approaches described herein may also be used to insert some or all of the V, D and J regions from the human heavy chain upstream of a light chain constant region, rather than upstream of the heavy chain constant region. Likewise some or all of the human light chain V and J regions may be inserted upstream of the heavy chain constant region. Insertion may be at the endogenous constant region locus, for example between the endogenous constant and J region, and may be of some, or all, of the V, D or J genes alone, excluding promoter or enhancer sequences, or may be of some, or all, of the V, D or J genes with one or more or all respective promoter or enhancer sequences. In one aspect the full repertoire of V, D or J fragments in germline orientation may be inserted upstream and in functional arrangement with a host constant region.

[0132] Thus the present invention allows V and/or D and/or J regions from a human, or any species, to be inserted into a chromosome of a cell from a different species that comprises a constant region, allowing a chimaeric antibody chain to be expressed.

[0133] In one aspect the invention requires only that some human variable region DNA is inserted into the genome of a non-human mammal in operable arrangement with some, or all, of the human heavy chain constant region at the region of the endogenous heavy chain constant region locus such that an antibody chain can be produced. In this aspect of the invention and where human light chain DNA is additionally inserted, the light chain DNA insertion can be in the form of a completely human construct, having both human variable DNA and human constant region DNA, or have human variable region DNA and constant region DNA from a non-human, non-host species. Other variations are also possible, such as insertion of both of the light chain human variable region and host genome constant region. In addition the insertion of said light chain transgenes need not be at the

equivalent endogenous locus, but may be anywhere in the genome. In such a scenario the cell or mammal may produce chimaeric heavy chains (comprising human variable region DNA and mouse constant region DNA) and light chains comprising human variable and human constant region DNA. Thus in one aspect of the invention the lambda and or kappa human variable region DNA can be inserted upstream of the endogenous locus, or downstream, or indeed on a different chromosome to the endogenous locus, and inserted with or without constant region DNA.

[0134] As well insertion of human light chain DNA upstream of the host non-human mammal constant region, a further aspect of the invention relates to insertion of one or both light chain human variable regions downstream of the equivalent endogenous locus constant region, or elsewhere in the genome.

[0135] Generally, insertion of human variable region DNA at or close to the equivalent endogenous locus in the recipient genome is preferred, for example within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 kb of the boundary (upstream or downstream) of a host immunoglobulin locus.

[0136] Thus in one aspect the invention can relate to a cell or non-human mammal whose genome comprises:

[0137] (a) a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region; and

[0138] (b) one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions, and/or, one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions;

wherein the non-human mammal is able to produce a repertoire of chimaeric antibodies, or chimaeric light or heavy chains, having a non-human mammal constant region and a human variable region.

[0139] In one particular aspect the genome of the cell or non-human mammal comprises:

[0140] a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region;

[0141] one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region, and one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions downstream of the host non-human mammal lambda constant region,

[0142] optionally in which the human lambda variable region may be inserted upstream or downstream of the endogenous host lambda locus in operable linkage with a human lambda constant region, such that the non-human mammal or cell can produce fully human antibody light chains and chimaeric heavy chains.

[0143] In a further, different, aspect of the invention, the use of the methods of the invention allows a locus to be built up in a stepwise manner by sequential insertions, and thus allows for the insertion of human variable DNA together with human or non-human constant region DNA at any suitable location in the genome of a non-human host cell. For example, methods of the invention can be used to insert human immunoglobulin variable region DNA together with constant region DNA from the host genome anywhere in the

genome of a non-human host cell, allowing a chimaeric antibody chain to be produced from a site other than the endogenous heavy region. Any human heavy chain or light chain DNA construct contemplated above can be inserted into any desired position into the genome of a non-human host cell using the techniques described herein. The present invention thus also relates to cells and mammals having genomes comprising such insertions.

[0144] The invention also relates to a vector, such as a BAC, comprising a human V, D or J region in a functional arrangement with a non-human mammal promoter, or other control sequence, such that the expression of the human V, D or J region is under the control of the non-human mammal promoter in a cell of the non-human mammal, such as an ES cell, in particular once inserted into the genome of that cell.

[0145] The invention also relates to cells and non-human mammals containing said cells, which cells or mammals have a human V, D or J region in a functional arrangement with a non-human mammal promoter, or other control sequence, such that the expression of the human V, D or J region is under the control of the non-human mammal promoter in the cells or mammal.

[0146] Generally, one aspect of the invention thus relates to a non-human mammal host cell capable of expression of a human V, D or J coding sequence under the control of a host promoter or control region, the expression capable of producing a humanised antibody having a human variable domain and non-human mammal constant region.

[0147] In one aspect the invention relates to a cell, such as a non mammalian cell, such as an ES cell, the genome of which comprises

[0148] (a) a plurality of human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region; and

[0149] (b) optionally one or more human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or one or more human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region;

[0150] In another aspect the invention relates to a cell, such as a non-human mammal cells, such as ES cells whose genome comprises

[0151] (a) a plurality of human Ig light chain kappa V regions and one or more human Ig light chain kappa J regions upstream of the host non-human mammal kappa constant region and/or a plurality of human Ig light chain lambda V regions and one or more human Ig light chain lambda J regions upstream of the host non-human mammal lambda constant region; and

[0152] (b) optionally one or more human IgH V regions, one or more human D regions and one or more human J regions upstream of the host non-human mammal constant region

[0153] In one aspect the cell is an ES cell is capable of developing into a non-human mammal able to produce a repertoire of antibodies which are chimaeric, said chimaeric antibodies having a non-human mammal constant region and a human variable region. Optionally the genome of the cell is modified to prevent expression of fully host-species specific antibodies.

[0154] In one aspect the cell is an induced pluripotent stem cell (iPS cell).

[0155] In one aspect cells are isolated non-human mammalian cells.

[0156] In one aspect a cell as disclosed herein is preferably a non-human mammalian cell.

[0157] In one aspect the cell is a cell from a mouse strain selected from C57BL/6, M129 such as 129/SV, BALB/c, and any hybrid of C57BL/6, M129 such as 129/SV, or BALB/c.

[0158] The invention also relates to a cell line which is grown from or otherwise derived from cells as described herein, including an immortalised cell line. The cell line may comprise inserted human V, D or J genes as described herein, either in germline configuration or after rearrangement following in vivo maturation. The cell may be immortalised by fusion (eg, electrofusion or using PEG according to standard procedures.) to a tumour cell (eg, P3X63-Ag8.653 (obtainable from LGC Standards; CRL-1580), SP2/0-Ag14 (obtainable from ECACC), NSI or NSO), to provide an antibody producing cell and cell line, or be made by direct cellular immortalisation.

[0159] The present invention also relates to vectors for use in the invention. In one aspect such vectors are BACs (bacterial artificial chromosomes). It will be appreciated that other cloning vectors may be used in the invention, and therefore reference to BACs herein may be taken to refer generally to any suitable vector.

[0160] In one aspect BACs used for generation of human DNA to be inserted, such as the VDJ or VJ regions are trimmed so that in the final human VDJ or VJ region or part thereof in the non-human mammal, no sequence is duplicated or lost when compared to the original human genomic sequence.

[0161] In one aspect the invention relates to a vector comprising an insert, preferably comprising a region of human DNA from some of the human VDJ or VJ locus, flanked by DNA which is not from that locus. The flanking DNA may comprise one or more selectable markers or one or more site specific recombination sites. In one aspect the vector comprises 2 or more, such as 3, heterospecific and incompatible site specific recombination sites. In one aspect the site specific recombination sites may be loxP sites, or variants thereof, or FRT sites or variants thereof. In one aspect the vector comprises one or more transposon ITR (inverted terminal repeat) sequences.

[0162] In one aspect the non-human animals of the invention suitably do not produce any fully humanised antibodies. In one aspect this is because there is no DNA inserted from the human constant region. Alternatively there is no human constant region DNA in the genome capable of forming an antibody in conjunction with the inserted human variable region DNA component, for example due to mutation within any human constant region DNA or distance from any constant region human DNA and human variable region DNA.

[0163] In one aspect human light chain constant region DNA may be included in the cell genome, such that a fully human lambda or kappa human antibody chain might be generated, but this would only be able to form an antibody with a chimaeric heavy chain, and not produce a fully human antibody having human variable and constant regions.

[0164] In one aspect the non-human mammal genome is modified to prevent expression of fully host-species specific

antibodies. Fully host species specific antibodies are antibodies that have both variable and constant regions from the host organism. In this context the term 'specific' is not intended to relate to the binding of the antibodies produced by the cells or animals of the invention but rather to the origin of the DNA which encodes those antibodies.

[0165] In one aspect the non-human mammal genome is modified to prevent expression of the native (fully host species specific) antibodies in the mammal by inactivation of all or a part of the host non-human mammal Ig loci. In this context, inactivation or prevention of endogenous antibody or gene segment usage (using any inactivation technique described herein) is, for example, substantially complete inactivation or prevention (substantially 100%, ie, essentially none (eg, less than 10, 5, 4, 3, 2, 1 or 0.5%) of the endogenous antibody chain (eg, no endogenous heavy chains) is expressed). This can be determined, for example, at the antibody chain (protein) level by assessing the antibody repertoire produced by the non-human vertebrate, mammal or at the nucleotide level by assessing mRNA transcripts of antibody chain loci, eg, using RACE. In an embodiment, inactivation is more than 50% (ie, 50% or less of the antibodies or transcripts are of an endogenous antibody chain), 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99%. For example, in an embodiment, endogenous heavy chain expression is substantially inactivated such that no more than 85%, 90%, 95%, 96%, 97%, 98% or 99% of the heavy chain repertoire of the vertebrate (mammal) is provided by endogenous heavy chains. For example, endogenous heavy chain expression is substantially inactivated such that substantially none of the heavy chain repertoire of the vertebrate (mammal) is provided by endogenous heavy chains. For example, in an embodiment, endogenous heavy chain expression is substantially inactivated such that no more than 85%, 90%, 95%, 96%, 97%, 98% or 99% of the kappa chain repertoire of the vertebrate (mammal) is provided by endogenous kappa chains. For example, endogenous kappa chain expression is substantially inactivated such that substantially none of the kappa chain repertoire of the vertebrate (mammal) is provided by endogenous kappa chains. For example, in an embodiment, endogenous heavy chain expression is substantially inactivated such that no more than 85%, 90%, 95%, 96%, 97%, 98% or 99% of the lambda chain repertoire of the vertebrate (mammal) is provided by endogenous lambda chains. For example, endogenous lambda chain expression is substantially inactivated such that substantially none of the lambda chain repertoire of the vertebrate (mammal) is provided by endogenous lambda chains.

[0166] In one aspect this is achieved by inversion of all or part of the non-human mammal VDJ region, or VJ region, optionally by insertion of one or more site specific recombination sites into the genome and then use of these sites in recombination-mediated excision or inversion of all or a part of the non-human mammal Ig locus. In one aspect a double inversion, may be employed, the first to move the V(D)Js away from the endogenous locus and then a more local inversion which puts them in the correct orientation. In one aspect a single loxP site is used to invert the non-human mammal VDJ region to a centromeric locus or telomeric locus.

[0167] In one example, a mouse or mouse cell of the invention comprises inverted endogenous heavy chain gene segments (eg, VH, D and JH, such as the entire endogenous

heavy chain VDJ region) that are immediately 3' of position 119753123, 119659458 or 120918606 on an endogenous mouse chromosome 12. Optionally, the genome of the mouse or cell is homozygous for said chromosome 12.

[0168] The invention also provides:—

[0169] A cassette for inversion and inactivation of endogenous non-human vertebrate (eg, mouse or rat) antibody chain gene segments, the segments being part of an antibody chain locus sequence on a chromosome of a non-human vertebrate (eg, mouse or rat) cell (eg, ES cell) wherein the sequence is flanked at its 3' end by a site-specific recombination site (eg, lox, rox or frt), the cassette comprising a nucleotide sequence encoding an expressible label or selectable marker and a compatible site-specific recombination site (eg, lox, rox or frt) flanked by a 5' and a 3' homology arm, wherein the homology arms correspond to or are homologous to adjacent stretches of sequence in the cell genome on a different chromosome or on said chromosome at least 10, 15, 20, 25, 30, 35, 40, 45 or 50 mb away from the endogenous gene segments.

[0170] The invention also provides:—

[0171] A cassette for inversion and inactivation of endogenous mouse antibody heavy chain gene segments, the segments being part of a heavy chain locus sequence on chromosome 12 of a mouse cell (eg, ES cell) wherein the sequence is flanked at its 3' end by a site-specific recombination site (eg, lox, rox or frt), the cassette comprising a nucleotide sequence encoding an expressible label or selectable marker and a compatible site-specific recombination site (eg, lox, rox or frt) flanked by a 5' and a 3' homology arm, wherein the homology arms correspond to or are homologous to adjacent stretches of sequence in the mouse cell genome on a different chromosome or on chromosome 12 at least 10, 15, 20, 25, 30, 35, 40, 45 or 50 mb away from the endogenous gene segments.

[0172] The invention provides:—

[0173] A cassette for inversion and inactivation of endogenous mouse antibody heavy chain gene segments, the segments being part of a heavy chain locus sequence on chromosome 12 of a mouse cell (eg, ES cell) wherein the sequence is flanked at its 3' end by a site-specific recombination site (eg, lox, rox or frt), the cassette comprising a nucleotide sequence encoding an expressible label or selectable marker and a compatible site-specific recombination site (eg, lox, rox or frt) flanked by a 5' and a 3' homology arm, wherein (i) the 5' homology arm is mouse chromosome 12 DNA from coordinate 119753124 to coordinate 119757104 and the 3' homology arm is mouse chromosome 12 DNA from coordinate 119749288 to 119753123; or (ii) the 5' homology arm is mouse chromosome 12 DNA from coordinate 119659459 to coordinate 119663126 and the 3' homology arm is mouse chromosome 12 DNA from coordinate 119656536 to 119659458; or (iii) the 5' homology arm is mouse chromosome 12 DNA from coordinate 120918607 to coordinate 120921930 and the 3' homology arm is mouse chromosome 12 DNA from coordinate 120915475 to 120918606.

[0174] Embodiment (i) results in an inversion of mouse chromosome 12 from coordinate 119753123 to coordinate 114666436.

[0175] Embodiment (ii) results in an inversion of mouse chromosome 12 from coordinate 119659458 to coordinate 114666436

[0176] Embodiment (iii) results in an inversion of mouse chromosome 12 from coordinate 12091806 to coordinate 114666436.

[0177] Thus, the invention provides a mouse or mouse cell whose genome comprises an inversion of a chromosome 12, wherein the inversion comprises inverted endogenous heavy chain gene segments (eg, VH, D and JH, such as the entire endogenous heavy chain VDJ region); wherein the mouse comprises a transgenic heavy chain locus comprising a plurality of human VH gene segments, a plurality of human D segments and a plurality of human JH segments operably connected upstream of an endogenous constant region (eg, C mu) so that the mouse or cell (optionally following differentiation into a B-cell) is capable of expressing an antibody comprising a variable region comprising sequences derived from the human gene segments; and wherein the inversion is (i) an inversion of mouse chromosome 12 from coordinate 119753123 to coordinate 114666436; (ii) an inversion of mouse chromosome 12 from coordinate 119659458 to coordinate 114666436; or (iii) an inversion of mouse chromosome 12 from coordinate 12091806 to coordinate 114666436.

[0178] In one embodiment, the endogenous gene segments are from a 129-derived mouse cell (eg, segments from an AB2.1 cell) and the homology arms are isogenic DNA (ie, identical to 129-derived endogenous sequences demarcated by the respective coordinates stated in (i) to (iii) above). Thus, no new sequence is created by homologous recombination using these homology arms. In another embodiment, the arms are from a mouse strain that is different from the endogenous strain. The site-specific recombination sites are mutually compatible and mutually inverted such that, on expression of an associated recombinase enzyme (eg, Cre, Dre or Flp), recombination between the site in the inserted inversion cassette and the site flanking the endogenous gene segments is carried out, thereby inverting and moving the endogenous gene segments far upstream (5') of their original location in the heavy chain locus. This inactivates endogenous heavy chain expression. Similarly, light chain inactivation can be performed by choosing the homology arms of the inversion cassette with reference to a chromosomal region spaced at least 10, 15, 20, 25, 30, 35, 40, 45 or 50 mb away from the endogenous light chain locus, the latter comprising a site-specific recombination site that is compatible with the site in the inversion cassette.

[0179] In one embodiment, the expressible label is a fluorescent label, eg, GFP or a variant thereof (eg, YFP, CFP or RFP). Thus, a label is used instead of a selection marker, such as one that confers resistance to allow for selection of transformants.

[0180] The invention provides a method of inactivating gene segments of an endogenous antibody locus, the method comprising

[0181] (i) Providing a non-human vertebrate cell (eg, an ES cell, eg, a mouse ES cell) whose genome comprises an antibody chain locus comprising endogenous variable region gene segments;

[0182] (ii) Targeting a site-specific recombination site to flank the 3' of the 3'-most of said endogenous gene segments;

[0183] (iii) Targeting a second site-specific recombination site at least 10 mb away from said endogenous gene segments, the second site being compatible with the first site inverted with respect to the first site;

[0184] (iv) Expressing a recombinase compatible with said sites to effect site-specific recombination between said sites, thereby inverting and moving said gene segments away from said locus, wherein the endogenous gene segments are inactivated; and

[0185] (v) Optionally developing the cell into a progeny cell or vertebrate (eg, mouse or rat) whose genome is homozygous for the inversion.

[0186] The genome of the progeny cell or vertebrate can comprise transgenic heavy and/or light chain loci, each capable of expressing antibody chains comprising human variable regions. Optionally, endogenous heavy and kappa light chain expression is inactivated by inverting endogenous heavy and kappa variable region gene segments according to the method of the invention. Optionally, endogenous lambda chain expression is also inactivated in this way.

[0187] In an alternative to the method and inversion cassettes of the invention, instead of inverting and moving variable region gene segments only, other parts of the endogenous locus can alternatively or additionally be inverted and moved to effect inactivation. For example, one or more endogenous regulatory elements (eg, Smu and/or Emu) and/or one or more endogenous constant regions (eg, Cmu and/or Cgamma) can be inverted and moved.

[0188] Sites that "flank" in the above contexts of the invention can be provided such that a site-specific recombination site immediately flanks the endogenous sequence or is spaced therefrom, eg, by no more than 250, 200, 150, 100, 50 or 20 kb in the 3' direction.

[0189] In one aspect the non-human mammal genome into which human DNA is inserted comprises endogenous V, (D) and J regions, and the endogenous sequences have not been deleted.

[0190] The invention comprises a method for insertion of multiple DNA fragments into a DNA target, suitably to form a contiguous insertion in which the inserted fragments are joined together directly without intervening sequences. The method is especially applicable to the insertion of a large DNA fragment into a host chromosome which can be carried out in a stepwise fashion.

[0191] In one aspect the method comprises insertion of a first DNA sequence into a target, the sequence having a DNA vector portion and a first sequence of interest (X1); insertion of a second DNA sequence into the vector portion of the first sequence, the second DNA sequence having a second sequence of interest (X2) and a second vector portion; and then excising any vector sequence DNA separating X1 and X2 to provide a contiguous X1X2, or X2X1 sequence within the target. There is optionally insertion of a further one or more DNA sequences, each DNA sequence having a further sequence of interest (X3, . . .) and a further vector portion, into the vector portion of the preceding DNA sequence, to build up a contiguous DNA fragment in the target.

[0192] The DNA target for insertion of the first DNA sequence may be a specific site or any point in the genome of a particular cell.

[0193] The general method is described herein in relation to the insertion of elements of the human VDJ region, but is applicable to insertion of any DNA region, from any organism, and in particular insertion of large DNA fragments of >100 kb, such as 100-250 kb, or even larger, such as that of

the TCR or HLA. Features and approaches described herein in respect of the VDJ insertion may be equally applied to the any of the methods disclosed

[0194] In one aspect the inserted DNA is human DNA, such as the human VDJ or VJ region, is built up in the genome of a cell, such as an ES cell, in a stepwise manner using 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20 or more separate insertions for each heavy chain or light chain region. Fragments are suitably inserted at the same or substantially the same cell locus, e.g. ES cell locus, one after another, to form the complete VDJ or VJ region, or part thereof. The present invention also relates to cells and non-human animals comprising intermediates in the process whose genomes may comprise only a partial VDJ region, such as only human variable region DNA.

[0195] In a further aspect the method for producing a transgenic non-human mammal comprises the insertion of human VDJ or VJ regions upstream of the host non-human mammal constant region by step-wise insertion of multiple fragments by homologous recombination, preferably using an iterative process. Suitably fragments of approximately 100 KB from the human VDJ and VJ locus are inserted, suitably to form part of, or a complete, VDJ or VJ region after the final iteration of the insertion process, as disclosed herein.

[0196] In one aspect the insertion process commences at a site where an initiation cassette has been inserted into the genome of a cell, such as an ES cell, providing a unique targeting region. In one aspect the initiation cassette is inserted in the non-human mammal heavy chain locus, for use in insertion of human heavy chain DNA. Similarly an initiation cassette may be inserted in the non-human mammal light chain locus, for use in insertion of human light chain VJ DNA. The initiation cassette suitably comprises a vector backbone sequence with which a vector having a human DNA fragment in the same backbone sequence can recombine to insert the human DNA into the cell (e.g. ES) cell genome, and suitably a selection marker, such as a negative selection marker. Suitably the vector backbone sequence is that of a BAC library, to allow BACs to be used in the construction of the ES cells and mammals. The vector backbone sequence may however be any sequence which serves as a target site into which a homologous sequence can insert, for example by homologous recombination, for example RMCE, and is preferably not DNA encoding any of the VDJ or constant region.

[0197] In one aspect the insertion of the first DNA fragment into an initiation cassette is followed by insertion of a second DNA fragment into a portion of the first DNA fragment, suitably a part of the vector backbone of the second DNA fragment. In one aspect an inserted DNA fragment comprises a part of the human VDJ region flanked by 5' and/or 3' sequences that are not from the human VDJ region. In one aspect the 5' and/or 3' flanking sequences may each contain one or more selectable markers, or be capable of creating a selectable system once inserted into the genome. In one aspect one or both flanking sequences may be removed from the genome in vitro, or in vivo, following insertion. In one aspect the method comprises insertion of a DNA fragment followed by selection of both 5' and 3' ends of the inserted fragment flanking the human VDJ DNA. In one aspect the iterative insertion is made by insertion of DNA fragments at the 5' end of the previous inserted fragment, and in this aspect there may be deletion in vivo of

the vector DNA which separates the inserted human DNA sequences, to provide a contiguous human DNA sequence.

[0198] In one aspect insertion of human VDJ DNA into a genome may be achieved without leaving any flanking DNA in the genome, for example by transposase mediated DNA excision. One suitable transposase is the Piggybac transposase.

[0199] In one aspect the first human variable region fragment is inserted by homologous recombination at the initiation cassette backbone sequence and then the DNA of any negative selection marker and initiation cassette are subsequently removed by recombination between recombinase target sequences, such as FRT using in this example, FLPase expression. Generally repeated targeted insertions at the (e.g. BAC) backbone initiation sequence and subsequent removal by rearrangement between recombinase target sequences are repeated to build up the entire human VDJ region upstream of the host non-mammal constant region.

[0200] In one aspect a selectable marker or system may be used in the method. The marker may be generated upon insertion of a DNA fragment into a genome, for example forming a selectable marker in conjunction with a DNA element already present in the genome.

[0201] In one aspect the cell (e.g. ES) cell genome does not contain 2 identical selectable markers at the same time during the process. It can be seen that the iterative process of insertion and selection can be carried out using only 2 different selection markers, as disclosed in the examples herein, and for example the third selectable marker may be identical to the first marker, as by the time of insertion of the third vector fragment the first vector fragment and the first marker has been removed.

[0202] In one aspect a correct insertion event, is confirmed before moving to the next step of any multistep cloning process, for example by confirmation of BAC structure using high density genomic arrays to screen ES cells to identify those with intact BAC insertions, sequencing and PCR verification.

[0203] Initiation cassette (also called a "landing pad")

[0204] The invention also relates to a polynucleotide 'landing pad' sequence, the polynucleotide comprising nucleic acid regions homologous to regions of a target chromosome to allow for insertion by homologous recombination into the target chromosome, and comprising a nucleic acid site which permits recombinase-driven insertion of nucleic acid into the landing pad. The invention also relates to vectors, cells and mammals of the invention comprising a landing pad as disclosed herein inserted into the genome of the cell.

[0205] The landing pad optionally comprises a non-endogenous S-mu, e.g. a rat S-mu switch

[0206] The landing pad optionally comprises (in 5' to 3' orientation) a mouse Eμ sequence, a non-human, non-mouse (e.g. rat) Switch p and at least a portion of a mouse Cp or the entire mouse Cp.

[0207] The rat switch sequence optionally comprises or consists of SEQ ID NO 1.

[0208] The landing pad optionally comprises the 5' homology arm of SEQ ID NO 6.

[0209] The landing pad optionally has the sequence of SEQ ID 2 or SEQ ID NO 3.

[0210] In one embodiment, the landing pad comprises an expressible label. For example the label is a fluorescent label, eg, GFP or a variant thereof (eg, YFP, CFP or RFP).

Thus, a label is used instead of a selection marker (such as one that confers resistance to allow for selection of transformants).

[0211] In an embodiment, the landing pad comprises 5' and 3' homology arms for insertion into the cell genome using homologous recombination. The homology arms can be isogenic DNA (eg, identical to 129-derived endogenous sequences of when a 129-derived ES cell is used). Thus, no new sequence is created by homologous recombination using these homology arms. In another embodiment, the arms are from a mouse strain that is different from the endogenous strain (ES cell strain).

[0212] The methods of the invention include methods wherein the landing pad sequence comprises any of the configurations or sequences as disclosed herein.

[0213] Another method of the invention comprises the step of insertion of the landing pad into a mouse chromosome by homologous recombination between mouse J1-4 and mouse C mu sequences.

[0214] Another method of the invention comprises the step of insertion of the landing pad into the mouse chromosome 12 by homologous recombination between mouse J1-4 and E mu.

[0215] In one aspect the method uses site specific recombination for insertion of one or more vectors into the genome of a cell, such as an ES cell. Site specific recombinase systems are well known in the art and may include Cre-lox, and FLP/FRT or combinations thereof, in which recombination occurs between 2 sites having sequence homology.

[0216] Additionally or alternatively to any particular Cre/Lox or FLP/FRT system described herein, other recombinases and sites that may be used in the present invention include Dre recombinase, rox sites, and PhiC31 recombinase.

[0217] Suitable BACs are available from the Sanger centre, see "A genome-wide, end-sequenced 129Sv BAC library resource for targeting vector construction". Adams D J, Quail M A, Cox T, van der Weyden L, Gorick B D, Su Q, Chan W I, Davies R, Bonfield J K, Law F, Humphray S, Plumb B, Liu P, Rogers J, Bradley A. *Genomics*. 2005 December; 86(6):753-8. Epub 2005 Oct. 27. The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK. BACs containing human DNA are also available from, for example, Invitrogen™. A suitable library is described in Osoegawa K et al, *Genome Research* 2001. 11: 483-496.

[0218] In one aspect a method of the invention specifically comprises:

[0219] (1) insertion of a first DNA fragment into a non-human ES cell, the fragment containing a first portion of human VDJ or VJ region DNA and a first vector portion containing a first selectable marker;

[0220] (2) optionally deletion of the a part of the first vector portion;

[0221] (3) insertion of a second DNA fragment into a non-human ES cell containing the first DNA fragment, the insertion occurring within the first vector portion, the second DNA fragment containing a second portion of the human VDJ or VJ region and a second vector portion containing a second selectable marker,

[0222] (4) deletion of the first selectable marker and first vector portion, preferably by a recombinase enzyme action;

[0223] (5) insertion of a third DNA fragment into a non-human ES cell containing the second DNA frag-

ment, the insertion occurring within the second vector portion, the third DNA fragment containing a third portion of the human VDJ or VJ region and a third vector portion containing third selectable marker,

[0224] (6) deletion of the second selectable marker and second vector portion; and

[0225] (7) iteration of the steps of insertion and deletion, as necessary, for fourth and further fragments of the human VDJ or VJ human regions, as necessary, to produce an ES cell with a part or all of the human VDJ or VJ region inserted as disclosed herein, and suitably to remove all the vector portions within the ES cell genome.

[0226] In another aspect the invention comprises

[0227] (1) insertion of DNA forming an initiation cassette into the genome of a cell;

[0228] (2) insertion of a first DNA fragment into the initiation cassette, the first DNA fragment comprising a first portion of a human DNA and a first vector portion containing a first selectable marker or generating a selectable marker upon insertion;

[0229] (3) optionally removal of part of the vector DNA

[0230] (4) insertion of a second DNA fragment into the vector portion of the first DNA fragment, the second DNA fragment containing a second portion of human DNA and a second vector portion, the second vector portion containing a second selectable marker, or generating a second selectable marker upon insertion;

[0231] (5) optionally, removal of any vector DNA to allow the first and second human DNA fragments to form a contiguous sequence; and

[0232] (6) iteration of the steps of insertion of human VDJ DNA and vector DNA removal, as necessary, to produce a cell with all or part of the human VDJ or VJ region sufficient to be capable of generating a chimeric antibody in conjunction with a host constant region, wherein the insertion of one, or more, or all of the DNA fragments uses site specific recombination.

[0233] In one aspect the non-human mammal is able to generate a diversity of at least 1×10^6 different functional chimeric immunoglobulin sequence combinations.

[0234] In one aspect the targeting is carried out in ES cells derived from the mouse C57BL/6N, C57BL/6J, 129S5 or 129Sv strain.

[0235] In one aspect non-human animals, such as mice, are generated in a RAG-1-deficient or a RAG-2-deficient background, or other suitable genetic background which prevents the production of mature host B and T lymphocytes.

[0236] In one aspect the non-human mammal is a rodent, suitably a mouse, and cells of the invention, are rodent cells or ES cells, suitably mouse ES cells.

[0237] The ES cells of the present invention can be used to generate animals using techniques well known in the art, which comprise injection of the ES cell into a blastocyst followed by implantation of chimeric blastocysts into females to produce offspring which can be bred and selected for homozygous recombinants having the required insertion. In one aspect the invention relates to a chimeric animal comprised of ES cell-derived tissue and host embryo derived tissue. In one aspect the invention relates to genetically-altered subsequent generation animals, which include animals having a homozygous recombinants for the VDJ and/or VJ regions.

[0238] In a further aspect the invention relates to a method for producing an antibody specific to a desired antigen the method comprising immunizing a transgenic non-human mammal as above with the desired antigen and recovering the antibody (see e.g. Harlow, E. & Lane, D. 1998, 5th edition, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Lab. Press, Plainview, N.Y.; and Pasqualini and Arap, *Proceedings of the National Academy of Sciences* (2004) 101:257-259). Suitably an immunogenic amount of the antigen is delivered. The invention also relates to a method for detecting a target antigen comprising detecting an antibody produced as above with a secondary detection agent which recognises a portion of that antibody.

[0239] In a further aspect the invention relates to a method for producing a fully humanised antibody comprising immunizing a transgenic non-human mammal as above with the desired antigen, recovering the antibody or cells expressing the antibody, and then replacing the non-human mammal constant region with a human constant region. This can be done by standard cloning techniques at the DNA level to replace the non-human mammal constant region with an appropriate human constant region DNA sequence—see e.g. Sambrook, J and Russell, D. (2001, 3^d edition) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, N.Y.).

[0240] In a further aspect the invention relates to humanised antibodies and antibody chains produced according to the present invention, both in chimaeric and fully humanised form, and use of said antibodies in medicine. The invention also relates to a pharmaceutical composition comprising such antibodies and a pharmaceutically acceptable carrier or other excipient.

[0241] Antibody chains containing human sequences, such as chimaeric human-non-human antibody chains, are considered humanised herein by virtue of the presence of the human protein coding regions region. Fully humanised antibodies may be produced starting from DNA encoding a chimaeric antibody chain of the invention using standard techniques.

[0242] Methods for the generation of both monoclonal and polyclonal antibodies are well known in the art, and the present invention relates to both polyclonal and monoclonal antibodies of chimaeric or fully humanised antibodies produced in response to antigen challenge in non-human mammals of the present invention.

[0243] In a yet further aspect, chimaeric antibodies or antibody chains generated in the present invention may be manipulated, suitably at the DNA level, to generate molecules with antibody-like properties or structure, such as a human variable region from a heavy or light chain absent a constant region, for example a domain antibody; or a human variable region with any constant region from either heavy or light chain from the same or different species; or a human variable region with a non-naturally occurring constant region; or human variable region together with any other fusion partner. The invention relates to all such chimaeric antibody derivatives derived from chimaeric antibodies identified according to the present invention.

[0244] In a further aspect, the invention relates to use of animals of the present invention in the analysis of the likely effects of drugs and vaccines in the context of a quasi-human antibody repertoire.

[0245] The invention also relates to a method for identification or validation of a drug or vaccine, the method

comprising delivering the vaccine or drug to a mammal of the invention and monitoring one or more of: the immune response, the safety profile; the effect on disease.

[0246] The invention also relates to a kit comprising an antibody or antibody derivative as disclosed herein and either instructions for use of such antibody or a suitable laboratory reagent, such as a buffer, antibody detection reagent.

[0247] The invention also relates to a method for making an antibody, or part thereof, the method comprising providing:

[0248] (i) a nucleic acid encoding an antibody, or a part thereof, obtained according to the present invention; or

[0249] (ii) sequence information from which a nucleic acid encoding an antibody obtained according to the present invention, or part thereof, can be expressed to allow an antibody to be produced.

[0250] The present invention also relates to a chimaeric antibody comprising a human variable region and a non-human vertebrate or mammal (optionally a rat or mouse) constant region (optionally a C gamma or C mu), wherein the antibody is encoded by a nucleotide sequence corresponding to the nucleotide sequence of a chimaeric heavy chain locus of a cell (optionally a B-cell, ES cell or hybridoma), the locus comprising a non-human vertebrate constant region nucleotide sequence and a rearranged VDJ nucleotide sequence produced by the in vivo rearrangement of a human V region, a human D region and a human J region, the V region being selected from one of a V1-3 region, V2-5 region, V4-4 region, V1-2 region or V6-1 region, and optionally a V1-3 or V6-1 segment. Optionally, the J region is any of JH1, JH2, JH3, JH4, JH5 or JH6, and in one aspect is JH4 or JH6. The D region is, in one aspect, any D3-9, D3-10, D6-13 or D6-19. In one example, rearranged VDJ nucleotide sequence is produced by the in vivo rearrangement of human V1-3 and JH4 (optionally with D3-9, D3-10, D6-13 or D-19); or V1-3 and JH6 (optionally with D3-9, D3-10, D6-13 or D-19); or V6-1 and JH4 (optionally with D3-9, D3-10, D6-13 or D-19) and JH6 (optionally with D3-9, D3-10, D6-13 or D-19). In one example the rearranged VDJ nucleotide sequence is produced by the in vivo rearrangement of human V6-1 DH3-10, V1-3 DH3-10, V1-3 DH6-19, V1-3 Dh3-9 or V6-1 DH6-19. In one aspect the antibody comprises any combination exemplified in the Examples and Figures herein. Optionally, the in vivo rearrangement is in a cell (eg, B cell or ES cell) derived from the same non-human vertebrate species as the constant region sequence (eg, a mouse B cell or ES cell). The invention also relates to a non-human vertebrate or mammal cell (eg, a B-cell or ES cell or hybridoma) whose genome comprises a chimaeric heavy chain locus as described above in this paragraph. The invention also relates to a non-human vertebrate or mammal (eg, a mouse or rat) whose genome comprises a chimaeric heavy chain locus as described above in this paragraph.

[0251] The present invention also relates to a non-human vertebrate or mammal having a genome encoding a chimaeric antibody, the chimaeric antibody comprising a human variable region and a non-human vertebrate or mammal (optionally a rat or mouse) constant region (optionally a C gamma or C mu), the mammal:

[0252] expressing more V1-3 antibodies than V2-5, V4-4, V1-2 or V6-1 antibodies; and/or

[0253] expressing more V1-3 JH4 or V1-3 JH6 antibodies than any of, individually, V1-3 JH1, V1-3 JH2, V1-3 JH3 or V1-3 JH5 antibodies, and/or

[0254] expressing more V6-1 JH4 or V6-1 JH6 antibodies than any of, individually, V6-1 JH1, V6-1 JH2, V6-1 JH3 or V6-1 JH5 antibodies and/or

[0255] expressing a greater number of V1-3 DH3-10 antibodies than antibodies V1-3 with any other D region. Expression of antibodies can be assessed by methods readily available to the skilled person and as conventional in the art. For example, expression can be assessed at the mRNA level as shown in the examples below.

[0256] The invention also relates to a chimaeric antibody comprising a human variable region and a non-human vertebrate or mammal (optionally a rat or mouse) constant region (optionally a light chain constant region), wherein the antibody is obtainable from a mammal (optionally a rat or mouse) whose genome comprises an antibody chain locus comprising a germline human kappa V1-8 and germline human kappa J1 sequence, and wherein the antibody is obtainable by in vivo recombination in said mammal of the V1-8 and J1 sequences and wherein the antibody has a variable region sequence which is different from that which is encoded by germline human kappa V1-8 and germline human kappa J1 sequences. Thus, in this aspect of the invention the human germline sequences are able to undergo productive rearrangement to form a coding sequence which, in conjunction with the non-human constant region sequence, can be expressed as a chimaeric antibody chain having at least a complete human variable region and a non-human constant region. This is in contrast (as the examples show below) to the combination of the germline human kappa V1-8 and germline human kappa J1 sequences per se, which do not provide for an antibody coding sequence (due to the inclusion of stop codons). In one aspect the rearranged sequence of the chimaeric antibody is a result of somatic hypermutation. In one aspect the antibody is a kappa antibody; in another aspect the antibody comprises a non-human heavy chain constant region (eg, a rat or mouse C gamma or C mu). The antibody sequence optionally comprises a X_1X_2 T F G Q, where X_1X_2 =PR, RT, or PW (SEQ ID No 21); optionally a X_1X_2 T F G Q G T K V E I K R A D A (SEQ ID No 22) motif. Such motifs are not found in the equivalent position in the germline sequence as shown in the examples. The invention also relates to a non-human vertebrate or mammal cell (eg, a B-cell or ES cell or hybridoma) whose genome comprises a chimaeric antibody chain locus as described above in this paragraph. The invention also relates to a non-human vertebrate or mammal (eg, a mouse or rat) whose genome comprises a chimaeric antibody chain locus as described above in this paragraph.

[0257] The invention also relates to a chimaeric antibody comprising a human variable region and a non-human vertebrate or mammal (optionally a rat or mouse) constant region (optionally a light chain constant region), wherein the antibody is obtainable from a mammal (optionally a rat or mouse) whose genome comprises an antibody chain locus comprising a germline human kappa V1-6 and germline human kappa J1 sequence, and wherein the antibody is obtainable by in vivo recombination in said mammal of the V1-6 and J1 sequences and wherein the antibody has a variable region sequence which is different from that which

is encoded by germline human kappa V1-6 and germline human kappa J1 sequences. Thus, in this aspect of the invention the human germline sequences are able to undergo productive rearrangement to form a coding sequence which, in conjunction with the non-human constant region sequence, can be expressed as a chimaeric antibody chain having at least a complete human variable region and a non-human constant region. This is in contrast (as the examples show below) to the combination of the germline human kappa V1-6 and germline human kappa J1 sequences per se, which do not provide for an antibody coding sequence (due to the inclusion of stop codons). In one aspect the rearranged sequence of the chimaeric antibody is a result of somatic hypermutation. In one aspect the antibody is a kappa antibody; in another aspect the antibody comprises a non-human heavy chain constant region (eg, a rat or mouse C gamma or C mu). The antibody sequence optionally comprises a X_3X_4 T F G Q, where X_3X_4 =PR or PW (SEQ ID No 23); optionally a X_3X_4 T F G Q G T K V E I K R A D A (SEQ ID No 24) motif. Such motifs are not found in the equivalent position in the germline sequence as shown in the examples. The invention also relates to a non-human vertebrate or mammal cell (eg, a B-cell or ES cell or hybridoma) whose genome comprises a chimaeric antibody chain locus as described above in this paragraph. The invention also relates to a non-human vertebrate or mammal (eg, a mouse or rat) whose genome comprises a chimaeric antibody chain locus as described above in this paragraph.

[0258] The invention also relates to a chimaeric antibody comprising a human variable region and a non-human (optionally a rat or mouse) constant region (optionally a C gamma or C mu or a C kappa), wherein the antibody is obtainable from a mammal (optionally a rat or mouse) whose genome comprises an antibody chain locus comprising a germline human kappa V1-5 and germline human kappa J1 sequence, and wherein the antibody is obtainable by in vivo recombination in said mammal of the V1-5 and J1 sequences. The invention also relates to a non-human vertebrate or mammal cell (eg, a B-cell or ES cell or hybridoma) whose genome comprises a chimaeric antibody chain locus as described above in this paragraph. The invention also relates to a non-human vertebrate or mammal (eg, a mouse or rat) whose genome comprises a chimaeric antibody chain locus as described above in this paragraph.

[0259] The invention also relates to a chimaeric antibody comprising a human variable region and a non-human (optionally a rat or mouse) constant region (optionally a C gamma or C mu or a C kappa), wherein the antibody is obtainable from a mammal (optionally a rat or mouse) whose genome comprises an antibody chain locus comprising a germline human kappa V1-5 and germline human kappa J4 sequence, and wherein the antibody is obtainable by in vivo recombination in said mammal of the V1-5 and J4 sequences. The invention also relates to a non-human vertebrate or mammal cell (eg, a B-cell or ES cell or hybridoma) whose genome comprises a chimaeric antibody chain locus as described above in this paragraph. The invention also relates to a non-human vertebrate or mammal (eg, a mouse or rat) whose genome comprises a chimaeric antibody chain locus as described above in this paragraph.

[0260] Antibodies of the invention may be isolated, in one aspect being isolated from the cell or organism in which they are expressed.

[0261] A non-human mammal whose genome comprises:

[0262] (a) the human IgH VDJ region upstream of the host non-human mammal constant region; and

[0263] (b) the human Ig light chain kappa V and J regions upstream of the host non-human mammal kappa constant region and/or the human Ig light chain lambda V and J regions upstream of the host non-human mammal lambda constant region;

wherein the non-human mammal is able to produce a repertoire of chimaeric antibodies having a non-human mammal constant region and a human variable region, and optionally wherein the non-human mammal genome is modified to prevent expression of fully host-species specific antibodies.

[0264] A non-human mammal ES cell whose genome comprises:

[0265] (a) the human IgH V, D and J region upstream of a non-human mammal constant region; and

[0266] (b) the human Ig locus light chain kappa V and J regions upstream of the host non-human mammal kappa constant region, and/or the human Ig locus light chain lambda V and J regions upstream of the host non-human mammal lambda constant region

wherein the ES cell is capable of developing into a non-human mammal, being able to produce a repertoire of antibodies which are chimaeric, having a non-human mammal constant region and a human variable region.

[0267] A method for producing a transgenic non-human mammal able to produce a repertoire of chimaeric antibodies, the antibodies having a non-human mammal constant region and a human variable region, the method comprising inserting by homologous recombination into a non-human mammal ES cell genome

[0268] (a) the human IgH VDJ region upstream of the host non-human mammal heavy chain constant region, and

[0269] (b) the human IgL VJ region for lambda or kappa chains upstream of the host non-human mammal lambda or kappa chain constant region, respectively

such that the non-human mammal is able to produce a repertoire of chimaeric antibodies having a non-human mammal constant region and a human variable region, wherein steps (a) and (b) can be carried out in either order and each of steps (a) and (b) can be carried out in a stepwise manner or as a single step.

[0270] In one aspect the insertion of human VDJ or VJ regions upstream of the host non-human mammal constant region is accomplished by step-wise insertion of multiple fragments by homologous recombination.

[0271] In one aspect the step-wise insertions commence at a site where an initiation cassette has been inserted into the genome of an ES cell providing a unique targeting region consisting of a BAC backbone sequence and a negative selection marker.

[0272] In one aspect the first human variable region fragment is inserted by homologous recombination at the initiation cassette BAC backbone sequence and said negative selection marker and initiation cassette are subsequently removed by recombination between recombinase target sequences.

[0273] In one aspect repeated targeted insertions at the BAC backbone initiation sequence and subsequent removal of the backbone by rearrangement between recombinase

target sequences is repeated to build up the entire human VDJ region upstream of the host non-mammal constant region.

Insertion of Human Variable Region Gene Segments Precisely within the Endogenous Mouse JH4-Cmu Intron

[0274] There is further provided a cell or non human mammal according to the invention wherein the mammal is a mouse or the cell is a mouse cell and wherein the insertion of the human heavy chain DNA is made in a mouse genome between coordinates 114,667,091 and 114,665,190 of mouse chromosome 12.

[0275] There is further provided a cell or non human mammal according to the invention wherein the insertion of the human heavy chain DNA is made at coordinate 114,667,091.

[0276] There is further provided a cell or non human mammal according to the invention wherein the human IgH VDJ region comprises nucleotides 105,400,051 to 106,368,585 from human chromosome 14 (coordinates refer to NCBI36 for the human genome).

[0277] There is further provided a method, cell or non human mammal according to the invention wherein a human coding region DNA sequence is in a functional arrangement with a non-human mammal control sequence, such that transcription of the human DNA is controlled by the non-human mammal control sequence. In one example, the initiation cassette is inserted between the mouse J4 and C alpha exons. There is further provided an initiation cassette suitable for use in the method comprising a vector backbone sequence and a selection marker.

[0278] The invention provides the following aspects (starting at aspect number 103):—

[0279] 103. A cell or non human mammal according to any one of the above configurations, examples, embodiments or aspects, wherein the mammal is a mouse or the cell is a mouse cell and wherein the insertion of the human heavy chain DNA is made in a mouse genome between coordinates 114,667,091 and 114,665,190 of mouse chromosome 12.

[0280] 104. A cell or non human mammal according to any one of the above configurations, examples, embodiments or aspects, wherein the insertion of the human heavy chain DNA is made at coordinate 114,667,091.

[0281] 105. A cell or mammal according to any one of the above configurations, examples, embodiments or aspects, wherein the human IgH VDJ region comprises nucleotides 105,400,051 to 106,368,585 from human chromosome 14 (coordinates refer to NCBI36 for the human genome).

[0282] 106. A method, cell or mammal according to any one of the above configurations, examples, embodiments or aspects, wherein a human coding region DNA sequence is in a functional arrangement with a non-human mammal control sequence, such that transcription of the human DNA is controlled by the non-human mammal control sequence.

[0283] 107. A method according to aspect 106 wherein the initiation cassette is inserted between the mouse J4 and C alpha exons.

[0284] 108. An initiation cassette suitable for use in the method of aspect 107 comprising a vector backbone sequence and a selection marker.

Inactivation of Endogenous Antibody Chain Expression by Insertion of Human Antibody Variable Region Gene Segments

[0285] 109. A non-human vertebrate (optionally a mouse or rat) or non-human vertebrate cell (optionally a mouse or rat cell) having a genome that

[0286] (i) comprises a transgenic antibody chain locus capable of expressing an antibody chain comprising a human variable region (optionally following antibody gene rearrangement); and

[0287] (ii) is inactivated for endogenous non-human vertebrate antibody chain expression; wherein the transgenic locus comprises

[0288] (iii) a DNA sequence comprising a plurality of human antibody variable region gene segments inserted between endogenous antibody variable region gene segments and an endogenous antibody constant region, whereby endogenous antibody chain expression is inactivated.

[0289] The transgenic locus is a heavy chain or light chain locus.

[0290] Inactivation of endogenous heavy chain expression in non-human vertebrates such as mice and rats has involved the deletion of all or part of the endogenous heavy chain VDJ region (including sequences between gene segments). The ADAM6 genes are present in the endogenous mouse VDJ region. In mouse, there are two copies of ADAM6 (ADAM6a, ADAM6b) located between the VH and D gene segments in the IgH locus of chromosome 12 (in the intervening region between mouse VH5-1 and D1-1 gene segments). These two adjacent intronless ADAM6 genes have 95% nucleotide sequence identity and 90% amino acid identity. In human and rat, there is only one ADAM6 gene. Expression pattern analysis of mouse ADAM6 shows that it is exclusively expressed in testis [1]. Although ADAM6 transcripts can be detected in lymphocytes, it is restricted to the nucleus, suggesting that the transcription of ADAM6 gene in particular was due to transcriptional read-through from the D region rather than active messenger RNA production [2]. In rat, ADAM6 is on chromosome 6.

[0291] Mature ADAM6 protein is located on the acrosome and the posterior regions of sperm head. Notably, ADAM6 forms a complex with ADAM2 and ADAM3, which is required for fertilization in mice [3]. Reference [4] implicates ADAM6 in a model where this protein interacts with ADAM3 after ADAM6 is sulphated by TPST2, sulphation of ADAM6 being critical for stability and/or complex formation involving ADAM6 and ADAM3, and thus ADAM6 and ADAM3 are lost from Tpst2-null sperm. The study observes that Tpst2-deficient mice have male infertility, sperm motility defects and possible abnormalities in sperm-egg membrane interactions.

[0292] Thus, the maintenance of ADAM6 expression in sperm is crucial for fertility. Thus, it is thought that transgenic male mice and rats in which ADAM6 genes have been deleted are not viably fertile. This hampers breeding of colonies and hampers the utility of such mice as transgenic antibody-generating platforms. It would be desirable to provide improved non-human transgenic antibody-generating vertebrates that are fertile.

[0293] [1]. Choi I, et. al., Characterization and comparative genomic analysis of intronless Adams with testicular gene expression. *Genomics*. 2004 April; 83(4):636-46.

[0294] [2]. Featherstone K, Wood A L, Bowen A J, Corcoran A E. The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. *J Biol Chem*. 2010 Mar. 26; 285(13):9327-38. Epub 2010 Jan. 25.

[0295] [3]. Han C, et. al., Comprehensive analysis of reproductive ADAMs: relationship of ADAM4 and ADAM6 with an ADAM complex required for fertilization in mice. *Biol Reprod*. 2009 May; 80(5):1001-8. Epub 2009 Jan. 7.

[0296] [4]. Marcello et al, Lack of tyrosylprotein sulfotransferase-2 activity results in altered sperm-egg interactions and loss of ADAM3 and ADAM6 in epididymal sperm, *J Biol Chem*. 2011 Apr. 15; 286(15):13060-70. Epub 2011 Feb. 21.

[0297] According to aspect 109 of the invention, inactivation does not involve deletion of the VDJ region or part thereof including endogenous ADAM6, but instead inactivation by insertion allows for the preservation of endogenous ADAM6 and thus does not risk infertility problems.

[0298] The final mouse resulting from the method (or a mouse derived from a cell produced by the method) is in one embodiment a male, so that the invention improves upon the prior art male transgenic mice that are infertile as a result of genomic manipulation. Fertile mice produce sperm that can fertilise eggs from a female mouse. Fertility is readily determined, for example, by successfully breeding to produce an embryo or child mouse. In another embodiment, the method of the invention makes a final female mouse. Such females are, of course, useful for breeding to create male progeny carrying ADAM6 and which are fertile.

[0299] In one embodiment of aspect 109, the genome is homozygous for the transgenic locus. For example, the genome is homozygous for endogenous ADAM6 genes.

[0300] In one embodiment of the vertebrate of aspect 109, the genome is inactivated for expression of endogenous heavy and kappa (and optionally also lambda) chains.

[0301] In one embodiment, in part (iii) of aspect 109 said DNA comprises human VH, D and JH gene segments or human VL and JL gene segments (eg, Vk and Jk gene segments). In an example, the DNA comprises a landing pad having a selectable marker, eg, a HPRT gene, neomycin resistance gene or a puromycin resistance gene; and/or a promoter.

[0302] In one embodiment, in part (iii) of aspect 109 the endogenous gene segments are the entire endogenous VDJ region of a heavy chain locus and/or the endogenous constant region is a Cmu or Cgamma.

[0303] In one embodiment, in part (iii) of aspect 109 the endogenous gene segments are the entire endogenous VJ region of a kappa chain locus and/or the endogenous constant region is a Ckappa

[0304] In one embodiment, in part (iii) of aspect 109 the endogenous gene segments are the entire endogenous VJ region of a lambda chain locus and/or the endogenous constant region is a Clambda.

- [0305] The non-human vertebrate cell can be a hybridoma, B-cell, ES cell or an IPS cell. When the cell is an ES cell or IPS cell, the endogenous antibody chain expression is inactivated following differentiation of the cell into a progeny B-cell (eg, in a B-cell in a non-human vertebrate).
- [0306] The invention further provides:—
- [0307] 110. The vertebrate or cell according to aspect 109, wherein said plurality of human antibody gene segments comprises at least 11 human V segments and/or at least 6 human J segments, eg at least 11 human VH gene segments and at least 6 human JH segments and optionally also at least 27 human D segments; optionally with the human inter-gene segment intervening sequences. In an embodiment, the human antibody gene segments are provided by a stretch of DNA sequence of human chromosome 14, comprising the gene segments and intervening sequences in germline configuration.
- [0308] 111. The vertebrate or cell according to aspect 109 or 110, wherein said inserted DNA sequence comprises a human nucleotide sequence comprising said antibody gene segments, wherein the nucleotide sequence is at least 110, 130, 150, 170, 190, 210, 230, 250, 270 or 290 kb. In an embodiment, the nucleotide sequence corresponds to a stretch of DNA sequence of human chromosome 14, comprising the gene segments and intervening sequences in germline configuration, eg, at least a sequence corresponding to the nucleotide sequence from coordinate 106328951 to coordinate 106601551 of a human chromosome 14, eg, a sequence in the GRCH37/hg19 sequence database.
- [0309] 112. The vertebrate or cell according to aspect 109, wherein the transgenic locus is a light chain kappa locus and the human antibody gene segments are between the 3'-most endogenous Jk gene segment and endogenous Ck; optionally wherein the human antibody gene segments comprise five functional human $J\lambda$ - $C\lambda$ clusters and at least one human $V\lambda$ gene segment, eg, at least a sequence corresponding to the nucleotide sequence from coordinate 23217291 to 23327884 of a lambda locus found on a human chromosome 22.
- [0310] 113. The vertebrate or cell according to any one of aspects 109 to 112, wherein the transgenic locus is a heavy chain locus and the human antibody gene segments are between the 3'-most endogenous JH gene segment (eg, JH4 in a mouse genome) and endogenous Cmu.
- [0311] 114. The vertebrate or cell according to any one of aspects 109 to 113, wherein the genome is homozygous for said transgenic locus.
- [0312] 115. A mouse or mouse cell or a rat or rat cell according to any one of aspects 109 to 114.
- [0313] 116. A method of making a non-human vertebrate cell (optionally a mouse or rat cell), the method comprising
- [0314] (a) providing a non-human ES cell whose genome comprises an endogenous antibody chain locus comprising endogenous antibody variable region gene segments and an endogenous antibody constant region; and
- [0315] (b) making a transgenic antibody chain locus by inserting into said endogenous locus a DNA sequences comprising a plurality of human antibody variable region gene segments between said endogenous antibody variable region gene segments and said endogenous constant region, so that the human antibody variable region gene segments are operably connected upstream of the endogenous constant region,
- [0316] whereby a non-human vertebrate ES cell is produced that is capable of giving rise to a progeny cell in which endogenous antibody expression is inactivated and wherein the progeny is capable of expressing antibodies comprising human variable regions; and
- [0317] (c) optionally differentiating said ES cell into said progeny cell or a non-human vertebrate (eg, mouse or rat) comprising said progeny cell.
- [0318] 117. The method according to aspect 116, wherein said plurality of human antibody gene segments comprises at least 11 human V segments.
- [0319] 118. The method according to aspect 116 or 117, wherein said plurality of human antibody gene segments comprises at least 6 human J segments.
- [0320] 119. The method according to aspect 116, 117 or 118, wherein a human nucleotide sequence is inserted in step (b), the nucleotide sequence comprising said antibody gene segments, wherein the nucleotide sequence is at least 110 kb.
- [0321] 120. The method according to any one of aspects 110 to 113, wherein the endogenous locus is a heavy chain locus and the human antibody gene segments are between the 3'-most endogenous JH gene segment and endogenous Cmu.
- [0322] 121. The method according to any one of aspects 116 to 120, wherein the progeny cell is homozygous for said transgenic locus.
- [0323] In one embodiment of the method of aspect 116, the method comprises inactivating the genome for expression of endogenous heavy and kappa (and optionally also lambda) chains.
- [0324] In one embodiment of the method of aspect 116, in part (b) said DNA sequence comprises human VH, D and JH gene segments or human VL and JL gene segments (eg, VK and JK gene segments). In an example, the DNA comprises a landing pad having a selectable marker, eg, a HPRT gene, neomycin resistance gene or a puromycin resistance gene; and/or a promoter.
- [0325] In one embodiment, in part (b) of aspect 116 the endogenous gene segments are the entire endogenous VDJ region of a heavy chain locus and/or the endogenous constant region is a Cmu or Cgamma.
- [0326] In one embodiment, in part (b) of aspect 116 the endogenous gene segments are the entire endogenous VJ region of a kappa chain locus and/or the endogenous constant region is a Ckappa
- [0327] In one embodiment, in part (b) of aspect 116 the endogenous gene segments are the entire endogenous VJ region of a lambda chain locus and/or the endogenous constant region is a Clambda.
- [0328] The non-human vertebrate cell can be a hybridoma, B-cell, ES cell or an IPS cell. When the cell is an ES cell or IPS cell, the endogenous antibody chain expression is inactivated following differentiation of the cell into a progeny B-cell (eg, in a B-cell in a non-human vertebrate).
- [0329] The invention further provides:—
- [0330] The method according to aspect 116, wherein said inserted DNA sequence comprises a human nucleotide sequence comprising said human antibody gene segments,

wherein the nucleotide sequence is at least 110, 130, 150, 170, 190, 210, 230, 250, 270 or 290 kb. In an embodiment, the nucleotide sequence corresponds to a stretch of DNA sequence of human chromosome 14, comprising the gene segments and intervening sequences in germline configuration, eg, at least a sequence corresponding to the nucleotide sequence from coordinate 106328951 to coordinate 106601551 of a human chromosome 14, eg, a sequence in the GRCH37/hg19 sequence database.

[0331] The method according to aspect 116, wherein the transgenic locus is a light chain kappa locus and the human antibody gene segments are between the 3'-most endogenous Jk gene segment and endogenous Ck; optionally wherein the human antibody gene segments comprise five functional human JA-CA clusters and at least one human Vλ gene segment, eg, at least a sequence corresponding to the nucleotide sequence from coordinate 23217291 to 23327884 of a lambda locus found on a human chromosome 22.

[0332] The method according to aspect 116, wherein the transgenic locus is a heavy chain locus and the human antibody gene segments are inserted between the 3'-most endogenous JH gene segment (eg, JH4 in a mouse genome) and endogenous Cmu.

[0333] 122. The method according to any one of aspects 116 to 121, comprising making the genome of the progeny homozygous for said transgenic locus.

[0334] Isolating Antibodies from Transonic Non-Human Vertebrates of the Invention & Useful Antigen-Specific Antibodies of Therapeutically-Relevant Affinities

[0335] 123. A method of isolating an antibody that binds a predetermined antigen, the method comprising

[0336] (a) providing a vertebrate (optionally a mammal; optionally a mouse or rat according to any one of the above configurations, examples, embodiments or aspects;

[0337] (b) immunising said vertebrate with said antigen (optionally wherein the antigen is an antigen of an infectious disease pathogen);

[0338] (c) removing B lymphocytes from the vertebrate and selecting one or more B lymphocytes expressing antibodies that bind to the antigen;

[0339] (d) optionally immortalising said selected B lymphocytes or progeny thereof, optionally by producing hybridomas therefrom; and

[0340] (e) isolating an antibody (eg, and IgG-type antibody) expressed by the B lymphocytes.

[0341] 124. The method of aspect 123, comprising the step of isolating from said B lymphocytes nucleic acid encoding said antibody that binds said antigen; optionally exchanging the heavy chain constant region nucleotide sequence of the antibody with a nucleotide sequence encoding a human or humanised heavy chain constant region and optionally affinity maturing the variable region of said antibody; and optionally inserting said nucleic acid into an expression vector and optionally a host.

[0342] 125. The method of aspect 123 or 124, further comprising making a mutant or derivative of the antibody produced by the method of aspect 122 or 123.

[0343] As demonstrated by the examples below, the non-human vertebrates of the invention are able to produce antigen-specific antibodies of sub-50 nM affinity with human sequences in their CDR3 regions. Thus, the invention further provides:—

[0344] 126. An antibody or fragment (eg, a Fab or Fab₂) thereof comprising variable regions that specifically bind a predetermined antigen with a sub-50 nM affinity (optionally sub-40, 30, 20, 10, 1, 0.1 or 0.01 nM) as determined by surface plasmon resonance, wherein the antibody is isolated from a non-human vertebrate (optionally a mammal; optionally a mouse or rat) according to any one of the above configurations, examples, embodiments or aspects and comprises heavy chain CDR3s (as defined by Kabat) encoded by a rearranged VDJ of said vertebrate, wherein the VDJ is the product of rearrangement in vivo of a human JH gene segment of a heavy chain locus of said vertebrate with D (optionally a human D gene segment of said locus) and VH gene segments.

[0345] In one embodiment, the surface plasmon resonance (SPR) is carried out at 25° C. In another embodiment, the SPR is carried out at 37° C.

[0346] In one embodiment, the SPR is carried out at physiological pH, such as about pH7 or at pH7.6 (eg, using Hepes buffered saline at pH7.6 (also referred to as HBS-EP)).

[0347] In one embodiment, the SPR is carried out at a physiological salt level, eg, 150 mM NaCl.

[0348] In one embodiment, the SPR is carried out at a detergent level of no greater than 0.05% by volume, eg, in the presence of P20 (polysorbate 20; eg, Tween-20™) at 0.05% and EDTA at 3 mM.

[0349] In one example, the SPR is carried out at 25° C. or 37° C. in a buffer at pH7.6, 150 mM NaCl, 0.05% detergent (eg, P20) and 3 mM EDTA. The buffer can contain 10 mM Hepes. In one example, the SPR is carried out at 25° C. or 37° C. in HBS-EP. HBS-EP is available from Teknova Inc (California; catalogue number H8022).

[0350] In an example, the affinity of the antibody is determined using SPR by

[0351] 1. Coupling anti-mouse (or other relevant non-human vertebrate) IgG (eg, Biacore BR-1008-38) to a biosensor chip (eg, GLM chip) such as by primary amine coupling;

[0352] 2. Exposing the anti-mouse IgG (non-human vertebrate antibody) to a test IgG antibody to capture test antibody on the chip;

[0353] 3. Passing the test antigen over the chip's capture surface at 1024 nM, 256 nM, 64 nM, 16 nM, 4 nM with a 0 nM (i.e. buffer alone); and

[0354] 4. And determining the affinity of binding of test antibody to test antigen using surface plasmon resonance, eg, under an SPR condition discussed above (eg, at 25° C. in physiological buffer). SPR can be carried out using any standard SPR apparatus, such as by Biacore™ or using the ProteOn XPR36™ (Bio-Rad®).

[0355] Regeneration of the capture surface can be carried out with 10 mM glycine at pH1.7. This removes the captured antibody and allows the surface to be used for another interaction. The binding data can be fitted to 1:1 model inherent using standard techniques, eg, using a model inherent to the ProteOn XPR36™ analysis software.

[0356] The invention also relates to an scFv, diabody or other antibody fragment comprising a VH and VL domain from an antibody or fragment of aspect 126 (optionally following affinity maturation, eg, by phage display).

- [0357] In one embodiment, the antigen is a serpin, eg, ovalbumin, antithrombin or antitrypsin. Serpins are a group of proteins with similar structures that were first identified as a set of proteins able to inhibit proteases. The acronym serpin was originally coined because many serpins inhibit chymotrypsin-like serine proteases (serine protease inhibitors). The first members of the serpin superfamily to be extensively studied were the human plasma proteins antithrombin and antitrypsin, which play key roles in controlling blood coagulation and inflammation, respectively.
- [0358] Initially, research focused upon their role in human disease: antithrombin deficiency results in thrombosis and antitrypsin deficiency causes emphysema. In 1980 Hunt and Dayhoff made the surprising discovery that both these molecules share significant amino acid sequence similarity to the major protein in chicken egg white, ovalbumin, and they proposed a new protein superfamily.
- [0359] 127. An antibody or fragment that is identical to an antibody of aspect 126 or a derivative thereof (optionally a derivative whose constant regions are human and/or an affinity matured derivative) that specifically binds said antigen with a sub-50 nM affinity as determined by surface plasmon resonance.
- [0360] 128. A pharmaceutical composition comprising an antibody or fragment of aspect 126 or 127 and a pharmaceutically-acceptable diluent, excipient or carrier.
- [0361] 129. A nucleotide sequence encoding a heavy chain variable region of an antibody or fragment of aspect 126 or 127, optionally as part of a vector (eg, an expression vector).
- [0362] 130. The nucleotide sequence of aspect 129, wherein the sequence is a cDNA derived from a B-cell of the vertebrate from which the antibody of aspect 126 is isolated, or is identical to such a cDNA.
- [0363] 131. An isolated host cell (eg, a hybridoma or a CHO cell or a HEK293 cell) comprising a nucleotide sequence according to aspect 129 or 130.
- [0364] 132. A method of isolating an antibody that binds a predetermined antigen, the method comprising
- [0365] (a) providing a vertebrate (optionally a mammal; optionally a mouse or rat according to any one of the above configurations, examples, embodiments or aspects;
- [0366] (b) immunising said vertebrate with said antigen;
- [0367] (c) removing B lymphocytes from the vertebrate and selecting a B lymphocyte expressing an antibody that binds to the antigen with sub-nM affinity, wherein the antibody is according to aspect 126;
- [0368] (d) optionally immortalising said selected B lymphocyte or progeny thereof, optionally by producing hybridomas therefrom; and
- [0369] (e) isolating an antibody (eg, and IgG-type antibody) expressed by the B lymphocyte.
- [0370] 133. The method of aspect 132, comprising the step of isolating from said B lymphocyte nucleic acid encoding said antibody that binds said antigen; optionally exchanging the heavy chain constant region nucleotide sequence of the antibody with a nucleotide sequence encoding a human or humanised heavy chain constant region and optionally affinity maturing the variable region of said antibody; and optionally inserting said nucleic acid into an expression vector and optionally a host.
- [0371] 134. The method of aspect 132 or 133, further comprising making a mutant or derivative of the antibody produced by the method of aspect 132 or 133.
- [0372] Inactivation by Inversion of Endogenous VDJ to Qenome Desert Regions
- [0373] 135. A mouse or mouse cell comprising inverted endogenous heavy chain gene segments (eg, VH, D and JH, such as the entire endogenous heavy chain VDJ region) that are immediately 3' of position 119753123, 119659458 or 120918606 on an endogenous mouse chromosome 12, wherein the mouse comprises a transgenic heavy chain locus comprising a plurality of human VH gene segments, a plurality of human D segments and a plurality of human JH segments operably connected upstream of an endogenous constant region (eg, C μ) so that the mouse or cell (optionally following differentiation into a B-cell) is capable of expressing an antibody comprising a variable region comprising sequences derived from the human gene segments.
- [0374] 136. The mouse or cell of aspect 135, wherein the genome of the mouse or cell is homozygous for said chromosome 12.
- [0375] 137. A cassette for inversion and inactivation of endogenous non-human vertebrate (eg, mouse or rat) antibody chain gene segments, the segments being part of an antibody chain locus sequence on a chromosome of a non-human vertebrate (eg, mouse or rat) cell (eg, ES cell) wherein the sequence is flanked at its 3' end by a site-specific recombination site (eg, lox, rox or frt), the cassette comprising a nucleotide sequence encoding an expressible label or selectable marker and a compatible site-specific recombination site (eg, lox, rox or frt) flanked by a 5' and a 3' homology arm, wherein the homology arms correspond to or are homologous to adjacent stretches of sequence in the cell genome on a different chromosome or on said chromosome at least 10 mb away from the endogenous gene segments.
- [0376] 138. A cassette for inversion and inactivation of endogenous mouse antibody heavy chain gene segments, the segments being part of a heavy chain locus sequence on chromosome 12 of a mouse cell (eg, ES cell) wherein the sequence is flanked at its 3' end by a site-specific recombination site (eg, lox, rox or frt), the cassette comprising a nucleotide sequence encoding an expressible label or selectable marker and a compatible site-specific recombination site (eg, lox, rox or frt) flanked by a 5' and a 3' homology arm, wherein (i) the 5' homology arm is mouse chromosome 12 DNA from coordinate 119753124 to coordinate 119757104 and the 3' homology arm is mouse chromosome 12 DNA from coordinate 119749288 to 119753123; (ii) the 5' homology arm is mouse chromosome 12 DNA from coordinate 119659459 to coordinate 119663126 and the 3' homology arm is mouse chromosome 12 DNA from coordinate 119656536 to 119659458; or (iii) the 5' homology arm is mouse chromosome 12 DNA from coordinate 120918607 to coordinate 120921930 and the 3' homology arm is mouse chromosome 12 DNA from coordinate 120915475 to 120918606.
- [0377] 139. A method of inactivating gene segments of an endogenous antibody locus, the method comprising
- [0378] (i) Providing a non-human vertebrate cell (eg, an ES cell, eg, a mouse ES cell) whose genome comprises

- an antibody chain locus comprising endogenous variable region gene segments;
- [0379] (ii) Targeting a site-specific recombination site to flank the 3' of the 3'-most of said endogenous gene segments;
- [0380] (iii) Targeting a second site-specific recombination site at least 10 mb away from said endogenous gene segments, the second site being compatible with the first site inverted with respect to the first site;
- [0381] (iv) Expressing a recombinase compatible with said sites to effect site-specific recombination between said sites, thereby inverting and moving said gene segments away from said locus, wherein the endogenous gene segments are inactivated; and
- [0382] (v) Optionally developing the cell into a progeny cell or vertebrate (eg, mouse or rat) whose genome is homozygous for the inversion.
- [0383] 140. A mouse or mouse cell whose genome comprises an inversion of a chromosome 12, wherein the inversion comprises inverted endogenous heavy chain gene segments (eg, VH, D and JH, such as the entire endogenous heavy chain VDJ region); wherein the mouse comprises a transgenic heavy chain locus comprising a plurality of human VH gene segments, a plurality of human D segments and a plurality of human JH segments operably connected upstream of an endogenous constant region (eg, C μ) so that the mouse or cell (optionally following differentiation into a B-cell) is capable of expressing an antibody comprising a variable region comprising sequences derived from the human gene segments; and wherein the inversion is (i) an inversion of mouse chromosome 12 from coordinate 119753123 to coordinate 114666436; (ii) an inversion of mouse chromosome 12 from coordinate 119659458 to coordinate 114666436; or (iii) an inversion of mouse chromosome 12 from coordinate 12091806 to coordinate 114666436.
- [0384] Other Aspects Include:
- [0385] A method for producing an antibody specific to a desired antigen the method comprising immunizing a non-human mammal as disclosed herein with the desired antigen and recovering the antibody or a cell producing the antibody.
- [0386] A method for producing a fully humanised antibody comprising immunizing a non-human mammal as disclosed herein and then replacing the non-human mammal constant region of an antibody specifically reactive with the antigen with a human constant region, suitably by engineering of the nucleic acid encoding the antibody.
- [0387] A method, cell or mammal as disclosed herein wherein a human coding region DNA sequence is in a functional arrangement with a non-human mammal control sequence, such that transcription of the DNA is controlled by the non-human mammal control sequence. In one aspect the human coding region V, D or J region is in a functional arrangement with a mouse promoter sequence.
- [0388] The invention also relates to a humanised antibody produced according to any methods disclosed herein and use of a humanised antibody so produced in medicine.
- [0391] FIGS. 19 and 20 show the principles behind antibody generation in chimaeric mice
- [0392] FIG. 21 shows a possible insertion site for the human DNA in a mouse chromosome
- [0393] FIGS. 22-26 disclose an alternative iterative process for insertion of a series of human BACs into a mouse Ig locus
- [0394] FIGS. 27-29 illustrate a mechanism for inversion of the host VDJ region
- [0395] FIG. 30 illustrates proof of principle for insertion of a plasmid using an RMCE approach
- [0396] FIG. 31 illustrates sequential RMCE—Integration into Landing Pad
- [0397] FIG. 32 illustrates confirmation of Successful Insertion into Landing Pad
- [0398] FIG. 33 illustrates PCR Confirmation of 3' End Curing
- [0399] FIG. 34 illustrates insertion of BAC#1 and PCR Diagnostics
- [0400] FIG. 35 illustrates JH and JK usage
- [0401] FIG. 36 illustrates DH usage
- [0402] FIG. 37 illustrates the distribution of CDR-H3 length in human VDJC μ transcripts from chimera mice
- [0403] FIG. 38 illustrates the distribution of nucleotide numbers of deletion and insertion in IGH-VDI or IGK-VJ junctions
- [0404] FIG. 39 illustrates Distribution of JH Usage Within Each VHs
- [0405] FIG. 40 illustrates Distribution of DH Usage Within Each VHs
- [0406] FIG. 41 illustrates Nucleotide Gain or Loss at VJ Joints Generates IGK Variants
- [0407] FIG. 42 illustrates Hypermutation in J Regions Generates IGK Variants
- [0408] FIG. 43 illustrates Joint Diversity Produces Functional CDS
- [0409] FIG. 44 illustrates a plot of identity of J_H gene segment use a 5'-RACE Cp-specific library generated from the splenic B lymphocytes of transgenic mice according to the invention in which endogenous gene segment use has been inactivated by inversion
- [0410] FIG. 45 illustrates the ratio of mouse V_H to human V_H usage as determined from antibody sequences from splenic B lymphocytes of transgenic mice according to the invention in which endogenous gene segment use has been inactivated by inversion
- [0411] FIG. 46 illustrates inversion strategy schematic
- [0412] FIG. 47 illustrates targeting construct R57 for inversion
- [0413] FIG. 48 illustrates sequence analysis from a Cp-specific 5'-RACE library of splenic B lymphocytes of S1^{mv1} (one human IGH BAC (ie, multiple human VH, all functional human D and JH) with an inverted endogenous IGH locus) mouse shows that practically all the transcripts came from rearranged human V_H-D-J_H gene segments
- [0414] FIG. 49 illustrates that the S1^{mv1} mouse shows a similar usage of both D and J_H gene segments to human
- [0415] FIG. 50 illustrates that mouse V_H usage is further significantly reduced following insertion of the 2nd human BAC into the endogenous heavy chain locus
- [0416] FIG. 51 illustrates a gel showing that normal class-switching (to IgG-type) was observed in transcripts from mice of the invention. The rearranged transcripts were detected using RT-PCR with human VH-specific and mouse

BRIEF DESCRIPTION OF THE FIGURES

- [0389] FIGS. 1-8 show an iterative process for insertion of a series of human BACs into a mouse Ig locus
- [0390] FIGS. 9-18 show in more detail the process of FIGS. 1-8 for the IgH and kappa locus

C γ -specific primers for amplification from peripheral blood cells of immunized transgenic mice

[0417] FIG. 52 illustrates sequence analysis amplified fragments demonstrate hypermutation occurred within the human variable regions of these IGH chains from mice of the invention

[0418] FIG. 53 illustrates Flow cytometric analysis showing normal B-cell compartments in transgenic mice of the invention

[0419] FIGS. 54A-54D illustrate normal IgH isotypes in transgenic mice (H1) immunised with 100 μ g Cholera Toxin B subunit. FIGS. 54E-54H illustrate normal IgH isotypes in transgenic mice (S1) immunised with 100 μ g Cholera Toxin B subunit.

[0420] FIG. 55A and FIG. 55B illustrate normal IgH isotypes and serum levels are obtained in transgenic H1 and S1 animals, respectively, of the invention following immunisation with antigens.

SEQUENCES

[0421] SEQ ID No 1 is a Rat switch sequence
SEQ ID No 2 is a landing pad targeting vector (long version)
SEQ ID No 3 is a landing pad targeting vector (shorter version)

SEQ ID No 4 is the mouse strain 129 switch
SEQ ID No 5 is the mouse strain C57 switch
SEQ ID No 6 is the 5' homology arm of a landing pad

SEQ ID No 7 is oligo HV2-5

SEQ ID No 8 is oligo HV4-4

SEQ ID No 9 is oligo HV1-3

SEQ ID No 10 is oligo HV1-2

SEQ ID No 11 is oligo HV6-1

SEQ ID No 12 is oligo Cp

SEQ ID No 13 is oligo KV1-9

SEQ ID No 14 is oligo KV1-8

SEQ ID No 15 is oligo KV1-6

SEQ ID No 16 is oligo KV1-5

SEQ ID No 17 is oligo CK

SEQ ID Nos 18-20 are rat switch sequences

SEQ ID No 21 is X_1X_2 T F G Q, where X_1X_2 =PR, RT, or PW

SEQ ID No 22 is X_1X_2 T F G Q G T K V E I K R A D A, where X_1X_2 =PR, RT, or PW;

SEQ ID No 23 is X_3X_4 T F G Q, where X_3X_4 =PR or PW

SEQ ID No 24 is X_3X_4 T F G Q G T K V E I K R A D A, where X_3X_4 =PR or PW

SEQ ID No 25 is Primer E1554

SEQ ID No 26 is Primer E1555

SEQ ID No 27 is Primer ELP1352_C γ 1

SEQ ID No 28 is Primer ELP1353_C γ 2b

SEQ ID No 29 is Primer ELP1354_C γ 2a

SEQ ID No 30 is Primer ELP1356_VH4-4

SEQ ID No 31 is Primer ELP1357_VH1-2,3

SEQ ID No 32 is Primer ELP1358_VH6-1

[0422] SEQ ID No 33 is Primer mlgG1_2 rev

SEQ ID No 34 is Primer mlgG2b rev

SEQ ID No 35 is Primer mlgG2a_2 rev

SEQ ID No 36 is Primer mCH1 unirev

SEQ ID No 37 is Primer mCH1 unirev_2

SEQ ID Nos 38-45 are CDRH3 sequences

SEQ ID Nos 46-50 is 3, 4, 5, 6 or more (up to 82) repeats of GGGCT

SEQ ID NOs 51-55 are heavy chain CDR1 sequences against CTB (cloned and reference)

SEQ ID NOs 56-60 are heavy chain CDR2 sequences against CTB (cloned and reference)

SEQ ID NOs 61-63 are heavy chain CDR3 sequences against CTB (cloned and reference)

SEQ ID NOs 64-68 are J Region sequences against CTB (cloned and reference)

DETAILED DESCRIPTION OF THE INVENTION

[0423] It will be understood that particular embodiments described herein are shown by way of illustration and not as limitations of the invention. The principal features of this invention can be employed in various embodiments without departing from the scope of the invention. Those skilled in the art will recognize, or be able to ascertain using no more than routine study, numerous equivalents to the specific procedures described herein. Such equivalents are considered to be within the scope of this invention and are covered by the claims.

[0424] The use of the word “a” or “an” when used in conjunction with the term “comprising” in the claims and/or the specification may mean “one,” but it is also consistent with the meaning of “one or more,” “at least one,” and “one or more than one.” The use of the term “or” in the claims is used to mean “and/or” unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and “and/or.” Throughout this application, the term “about” is used to indicate that a value includes the inherent variation of error for the device, the method being employed to determine the value, or the variation that exists among the study subjects.

[0425] As used in this specification and claim(s), the words “comprising” (and any form of comprising, such as “comprise” and “comprises”), “having” (and any form of having, such as “have” and “has”), “including” (and any form of including, such as “includes” and “include”) or “containing” (and any form of containing, such as “contains” and “contain”) are inclusive or open-ended and do not exclude additional, unrecited elements or method steps

[0426] The term “or combinations thereof” as used herein refers to all permutations and combinations of the listed items preceding the term. For example, “A, B, C, or combinations thereof” is intended to include at least one of: A, B, C, AB, AC, BC, or ABC, and if order is important in a particular context, also BA, CA, CB, CBA, BCA, ACB, BAC, or CAB. Continuing with this example, expressly included are combinations that contain repeats of one or more item or term, such as BB, AAA, MB, BBC, AAAB-CCCC, CBBAAA, CABABB, and so forth. The skilled artisan will understand that typically there is no limit on the number of items or terms in any combination, unless otherwise apparent from the context.

[0427] As a source of antibody gene segment sequences, the skilled person will also be aware of the following

available databases and resources (including updates thereof) the contents of which are incorporated herein by reference:

- [0428] The Kabat Database (G. Johnson and T. T. Wu, 2002; World Wide Web (www.kabatdatabase.com). Created by E. A. Kabat and T. T. Wu in 1966, the Kabat database publishes aligned sequences of antibodies, T-cell receptors, major histocompatibility complex (MHC) class I and II molecules, and other proteins of immunological interest. A searchable interface is provided by the Seq-huntIII
- [0429] tool, and a range of utilities is available for sequence alignment, sequence subgroup classification, and the generation of variability plots. See also Kabat, E. A., Wu, T. T., Perry, H., Gottesman, K., and Foeller, C. (1991) *Sequences of Proteins of Immunological Interest*, 5th ed., NIH Publication No. 91-3242, Bethesda, Md., which is incorporated herein by reference, in particular with reference to human gene segments for use in the present invention.
- [0430] KabatMan (A. C. R. Martin, 2002; World Wide Web (www.bioinf.org.uk/abs/simkab.html). This is a web interface to make simple queries to the Kabat sequence database.
- [0431] IMGT (the International ImMunoGeneTics Information System®; M.-P. Lefranc, 2002; World Wide Web (www.imgt.cines.fr). IMGT is an integrated information system that specializes in antibodies, T cell receptors, and MHC molecules of all vertebrate species. It provides a common portal to standardized data that include nucleotide and protein sequences, oligonucleotide primers, gene maps, genetic polymorphisms, specificities, and two-dimensional (2D) and three-dimensional (3D) structures. IMGT includes three sequence databases (IMGT/LIGMDB, IMGT/MHC-DB, IMGT/PRIMERDB), one genome database (IMGT/GENE-DB), one 3D structure database (IMGT/3Dstructure-DB), and a range of web resources (“IMGT Marie-Paule page”) and interactive tools.
- [0432] V-BASE (I. M. Tomlinson, 2002; World Wide Web (www.mrc-cpe.cam.ac.uk/vbase). V-BASE is a comprehensive directory of all human antibody germline variable region sequences compiled from more than one thousand published sequences. It includes a version of the alignment software DNAPLOT (developed by Hans-Helmar Althaus and Werner Müller) that allows the assignment of rearranged antibody V genes to their closest germline gene segments.
- [0433] Antibodies-Structure and Sequence (A. C. R. Martin, 2002; World Wide Web (www.bioinf.org.uk/abs). This page summarizes useful information on antibody structure and sequence. It provides a query interface to the Kabat antibody sequence data, general information on antibodies, crystal structures, and links to other antibody-related information. It also distributes an automated summary of all antibody structures deposited in the Protein Databank (PDB). Of particular interest is a thorough description and comparison of the various numbering schemes for antibody variable regions.
- [0434] AAAAA (A Ho’s Amazing Atlas of Antibody Anatomy; A. Honegger, 2001; World Wide Web (www.unizh.ch/~antibody). This resource includes tools for structural analysis, modeling, and engineering. It adopts a unifying scheme for comprehensive structural alignment of antibody and T-cell-receptor sequences, and includes Excel macros for antibody analysis and graphical representation.
- [0435] WAM (Web Antibody Modeling; N. Whitelegg and A. R. Rees, 2001; World Wide Web (www.antibody.bath.ac.uk). Hosted by the Centre for Protein Analysis and Design at the University of Bath, United Kingdom. Based on the AbM package (formerly marketed by Oxford Molecular) to construct 3D models of antibody Fv sequences using a combination of established theoretical methods, this site also includes the latest antibody structural information.
- [0436] Mike’s Immunoglobulin Structure/Function Page (M. R. Clark, 2001; World Wide Web (www.path.cam.ac.uk/~mrc7/mikeimages.html). These pages provide educational materials on immunoglobulin structure and function, and are illustrated by many colour images, models, and animations. Additional information is available on antibody humanization and Mike Clark’s Therapeutic Antibody Human Homology Project, which aims to correlate clinical efficacy and anti-immunoglobulin responses with variable region sequences of therapeutic antibodies.
- [0437] The Antibody Resource Page (The Antibody Resource Page, 2000; World Wide Web (www.antibodyresource.com). This site describes itself as the “complete guide to antibody research and suppliers.” Links to amino acid sequencing tools, nucleotide antibody sequencing tools, and hybridoma/cell-culture databases are provided.
- [0438] Humanization bY Design (J. Saldanha, 2000; World Wide Web (www.people.cryst.bbk.ac.uk/~ubcg07s). This resource provides an overview on antibody humanization technology. The most useful feature is a searchable database (by sequence and text) of more than 40 published humanized antibodies including information on design issues, framework choice, framework back-mutations, and binding affinity of the humanized constructs.
- [0439] See also Antibody Engineering Methods and Protocols, Ed. Benny K C Lo, Methods in Molecular Biology™, Human Press. Also at World Wide Web (www.blogsua.com/pdf/antibody-engineering-methods-and-protocolsantibody-engineering-methods-and-protocols.pdf
- [0440] Any part of this disclosure may be read in combination with any other part of the disclosure, unless otherwise apparent from the context.
- [0441] All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.
- [0442] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the

invention, and are not intended to limit the scope of what the inventors regard as their invention.

EXAMPLES

Example 1

BAC Recombineering

[0443] Overall Strategy:

[0444] A mouse model of the invention can be achieved by inserting ~960 kb of the human heavy chain locus containing all the V, D and J-regions upstream of the mouse constant region and 473 kb of the human kappa region upstream of the mouse constant region. Alternatively, or in tandem, the human lambda region is inserted upstream of the mouse constant region. This insertion is achieved by gene targeting in ES cells using techniques well known in the art.

[0445] High fidelity insertion of intact V-D-J regions into each locus in their native (wild-type) configuration is suitably achieved by insertion of human bacterial artificial chromosomes (BACs) into the locus. Suitably the BACs are trimmed so that in the final locus no sequence is duplicated or lost compared to the original. Such trimming can be carried out by recombineering.

[0446] The relevant human BACs, suitably trimmed covering these loci are on average 90 kb in size.

[0447] In one approach the full complement of human D and J-elements as well as seven or eight human V-regions are covered by the first BACs to be inserted in the experimental insertion scheme described below. The first BACs to be inserted in the IgH and IgK loci may contain the following V-regions. IgH: V6-1, VII-1-1, V1-2, VIII-2-1, V1-3, V4-4, V2-5 and IgK: V4-1, V5-2, V7-3, V2-4, V1-5, V1-6, V3-7, V1-8.

[0448] Suitably the performance of each locus is assessed after the first BAC insertion using chimaeric mice and also after each subsequent BAC addition. See below for detailed description of this performance test.

[0449] Nine additional BAC insertions will be required for the IgH locus and five for IgK to provide the full complement of human V-regions covering all 0.96 Mb and 0.473 Mb of the IgH and IgK loci, respectively.

[0450] Not all BACs retain their wild-type configuration when inserted into the ES cell genome. Thus, high density genomic arrays were deployed to screen ES cells to identify those with intact BAC insertions (Barrett, M. T., Scheffer, A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P. S., et al. (2004). Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. Proceedings of the National Academy of Sciences of the United States of America 101, 17765-17770.). This screen also enables one to identify and select against ES clones in which the ES cell genome is compromised and thus not able to populate the germ line of chimeric animals. Other suitable genomic tools to facilitate this assessment include sequencing and PCR verification.

[0451] Thus in one aspect the correct BAC structure is confirmed before moving to the next step.

[0452] It is implicit from the description above that in order to completely engineer the loci with ~90 kb BACs, it is necessary to perform a minimum of 10 targeting steps for IgH and 5 steps for the IgK. Mice with an IgL locus can be generated in a similar manner to the IgK locus. Additional

steps are required to remove the selection markers required to support gene targeting. Since these manipulations are being performed in ES cells in a step-wise manner, in one aspect germ line transmission capacity is retained throughout this process.

[0453] Maintaining the performance of the ES cell clones through multiple rounds of manipulation without the need to test the germ line potential of the ES cell line at every step may be important in the present invention. The cell lines currently in use for the KOMP and EUCOMM global knockout projects have been modified twice prior to their use for this project and their germ line transmission rates are unchanged from the parental cells (these lines are publicly available, see World Wide Web (www) komp.org and World Wide Web (www) eucomm.org). This cell line, called JM8, can generate 100% ES cell-derived mice under published culture conditions (Pettitt, S. J., Liang, Q., Rairdan, X. Y., Moran, J. L., Prosser, H. M., Beier, D. R., Lloyd, K. C., Bradley, A., and Skarnes, W. C. (2009). Agouti C57BL/6N embryonic stem cells for mouse genetic resources. Nature Methods.). These cells have demonstrated ability to reproducibly contribute to somatic and germ line tissue of chimaeric animals using standard mouse ES cell culture conditions. This capability can be found with cells cultured on a standard feeder cell line (SNL) and even feeder-free, grown only on gelatine-coated tissue culture plates. One particular sub-line, JM8A3, maintained the ability to populate the germ line of chimeras after several serial rounds of sub-cloning. Extensive genetic manipulation via, for example, homologous recombination—as would be the case in the present invention—cannot compromise the pluripotency of the cells. The ability to generate chimeras with such high percentage of ES cell-derived tissue has other advantages. First, high levels of chimerism correlates with germ line transmission potential and provide a surrogate assay for germ line transmission while only taking 5 to 6 weeks. Second, since these mice are 100% ES cell derived the engineered loci can be directly tested, removing the delay caused by breeding. Testing the integrity of the new Ig loci is possible in the chimera since the host embryo will be derived from animals that are mutant for the RAG-1 gene as described in the next section.

[0454] Another cell line that may be used is an HPRT-ve cell line, such as AB2.1, as disclosed in Ramirez-Solis R, Liu P and Bradley A, "Chromosome engineering in mice," Nature, 1995; 378; 6558; 720-4.

[0455] RAG-1 Complementation:

[0456] While many clones will generate 100% ES derived mice some will not. Thus, at every step mice are generated in a RAG-1-deficient background. This provides mice with 100% ES-derived B- and T-cells which can be used directly for immunization and antibody production. Cells having a RAG-2 deficient background, or a combined RAG-1/RAG-2 deficient background may be used, or equivalent mutations in which mice produce only ES cell-derived B cells and/or T cells.

[0457] In order that only the human-mouse IgH or IgK loci are active in these mice, the human-mouse IgH and IgK loci can be engineered in a cell line in which one allele of the IgH or IgK locus has already been inactivated. Alternatively the inactivation of the host Ig locus, such as the IgH or IgK locus, can be carried out after insertion.

[0458] Mouse strains that have the RAG-1 gene mutated are immunodeficient as they have no mature B- or T-lym-

phocytes (U.S. Pat. No. 5,859,307). T- and B-lymphocytes only differentiate if proper V(D)J recombination occurs. Since RAG-1 is an enzyme that is crucial for this recombination, mice lacking RAG-1 are immunodeficient. If host embryos are genetically RAG-1 homozygous mutant, a chimera produced by injecting such an embryo will not be able to produce antibodies if the animal's lymphoid tissues are derived from the host embryo. However, JM8 cells and AB2.1 cells, for example, generally contribute in excess of 80% of the somatic tissues of the chimeric animal and would therefore usually populate the lymphoid tissue. JM8 cells have wild-type RAG-1 activity and therefore antibodies produced in the chimeric animal would be encoded by the engineered JM8 ES cell genome only. Therefore, the chimeric animal can be challenged with an antigen by immunization and subsequently produce antibodies to that antigen. This allows one skilled in the art to test the performance of the engineered human/mouse IgH and IgK loci as described in the present invention. See FIGS. 19 and 20.

[0459] One skilled in the art would use the chimeric animal as described to determine the extent of antibody diversity (see e.g. Harlow, E. & Lane, D. 1998, 5th edition, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Lab. Press, Plainview, N.Y.). For example, the existence in the chimeric animal's serum of certain antibody epitopes could be ascertained by binding to specific anti-idiotypic antiserum, for example, in an ELISA assay. One skilled in the art could also sequence the genomes of B-cell clones derived from the chimeric animal and compare said sequence to wild-type sequence to ascertain the level of hypermutation, such hypermutation indicative of normal antibody maturation.

[0460] One skilled in the art would also use said chimeric animal to examine antibody function wherein said antibodies are encoded from the engineered Ig loci (see e.g. Harlow, E. & Lane, D. 1998, 5th edition, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Lab. Press, Plainview, N.Y.). For example, antisera could be tested for binding an antigen, said antigen used to immunize the chimeric animal. Such a measurement could be made by an ELISA assay. Alternatively, one skilled in the art could test for neutralization of the antigen by addition of the antisera collected from the appropriately immunized chimeric animal.

[0461] It is well known to those skilled in the art that positive outcomes for any of these tests demonstrate the ability of the engineered Ig loci, the subject of the instant invention, to encode antibodies with human variable regions and mouse constant regions, said antibodies capable of functioning in the manner of wild-type antibodies.

[0462] Experimental Techniques: Recombineering for the production of vectors for use in homologous recombination in ES cells is disclosed in, for example, WO9929837 and WO0104288, and the techniques are well known in the art. In one aspect the recombineering of the human DNA takes place using BACs as a source of said human DNA. Human BAC DNA will be isolated using QIAGEN®, BAC purification kit. The backbone of each human BAC will be modified using recombineering to the exact same or similar configuration as the BAC already inserted into the mouse IgH region. The genomic insert of each human BAC will be trimmed using recombineering so that once the BACs are inserted, a seamless contiguous part of the human V(D)J genomic region will form at the mouse IgH or IgK locus. BAC DNA transfection by electroporation and genotyping

will be performed accordingly to standard protocols (Prosser, H. M., Rzadzinska, A. K., Steel, K. P., and Bradley, A. (2008). "Mosaic complementation demonstrates a regulatory role for myosin VIIa in actin dynamics of stereocilia." *Molecular and Cellular Biology* 28, 1702-1712; Ramirez-Solis, R., Davis, A. C., and Bradley, A. (1993). "Gene targeting in embryonic stem cells." *Methods in Enzymology* 225, 855-878.). Recombineering will be performed using the procedures and reagents developed by Pentao Liu and Don Court's laboratories (Chan, W., Costantino, N., Li, R., Lee, S. C., Su, Q., Melvin, D., Court, D. L., and Liu, P. (2007). "A recombineering based approach for high-throughput conditional knockout targeting vector construction." *Nucleic Acids Research* 35, e64).

[0463] These and other techniques for gene targeting and recombination of BAC-derived chromosomal fragments into a non-human mammal genome, such as a mouse are well-known in the art and are disclosed in, for example, in World Wide Web (www.eucomm.org/information/targeting) and World Wide Web (www.eucomm.org/information/publications).

[0464] Cell culture of C57BL/6N-derived cell lines, such as the JM8 male ES cells will follow standard techniques. The JM8 ES cells have been shown to be competent in extensively contributing to somatic tissues and to the germline, and are being used for large mouse mutagenesis programs at the Sanger Institute such as EUCOMM and KOMP (Pettitt, S. J., Liang, Q., Rairdan, X. Y., Moran, J. L., Prosser, H. M., Beier, D. R., Lloyd, K. C., Bradley, A., and Skarnes, W. C. (2009). "Agouti C57BL/6N embryonic stem cells for mouse genetic resources." *Nature Methods*). JM8 ES cells (1.0×10^7) will be electroporated (500 μ F, 230V; Bio-Rad®) with 10 μ g I-SceI linearized human BAC DNA. The transfectants will be selected with either Puromycin (3 μ g/ml) or G418 (150 μ g/ml). The selection will begin either 24 hours (with G418) or 48 hours (with Puromycin) post electroporation and proceed for 5 days. 10 μ g linearized human BAC DNA can yield up to 500 Puromycin or G418 resistant ES cell colonies. The antibiotic resistant ES cell colonies will be picked into 96-well cell culture plates for genotyping to identify the targeted clones.

[0465] Once targeted mouse ES cell clones are identified, they will be analyzed by array Comparative Genomic Hybridization (CGH) for total genome integrity (Chung, Y. J., Jonkers, J., Kitson, H., Fiegler, H., Humphray, S., Scott, C., Hunt, S., Yu, Y., Nishijima, I., Velds, A., et al. (2004). "A whole-genome mouse BAC microarray with 1-Mb resolution for analysis of DNA copy number changes by array comparative genomic hybridization." *Genome research* 14, 188-196. and Liang, Q., Conte, N., Skarnes, W. C., and Bradley, A. (2008). "Extensive genomic copy number variation in embryonic stem cells." *Proceedings of the National Academy of Sciences of the United States of America* 105, 17453-17456.). ES cells that have abnormal genomes do not contribute to the germline of the chimeric mice efficiently. BAC integrity will be examined by PCR-amplifying each known functional V gene in the BAC. For example, in one approach the first human BAC chosen for the IgH locus has 6 functional V genes. To confirm the integrity of this BAC for the presence of these 6 IGH V genes, at least 14 pairs of PCR primers will be designed and used to PCR-amplify genomic DNA from the targeted ES cells. The human wild-type size and sequence of these fragments will ensure that the inserted BAC has not been rearranged. More

detailed CGH will also confirm the integrity of the inserted BACs. For example, one skilled in the art could use an oligo aCGH platform, which is developed by Agilent Technologies, Inc. This platform not only enables one to study genome-wide DNA copy number variation at high resolution (Barrett, M. T., Scheffer, A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P. S., et al. (2004). "Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA." *Proceedings of the National Academy of Sciences of the United States of America* 101, 17765-17770.), but permit examination of a specific genome region using custom designed arrays. Comparing the traditional aCGH techniques which rely on cDNA probes or whole BAC probes, the 60-mer oligonucleotide probes can ensure specific hybridization and high sensitivity and precision that is needed in order to detect the engineered chromosome alterations that were made. For example, oligos designed to hybridize at regular intervals along the entire length of the inserted BAC would detect even quite short deletions, insertions or other rearrangements. Also, this platform provides the greatest flexibility for customized microarray designs. The targeted ES cell genomic DNA and normal human individual genomic DNA will be labelled separately with dyes and hybridized to the array. Arrays slides will be scanned using an Agilent Technologies DNA microarray scanner. Reciprocal fluorescence intensities of dye Cy5 and dye Cy3 on each array image and the log 2 ratio values will be extracted by using Bluefuse software (Bluegenome). Spots with inconsistent fluorescence patterns ("confidence"<0.29 or "quality"<0) will be excluded before normalizing all log 2 ratio values. Within an experiment, Log 2 ratio between -0.29 and +0.29 for the signal from any oligo probe are regarded as no copy number change. The log 2 ratio threshold for "Duplication" is usually >0.29999, and for deletion is <0.29999.

[0466] Once the first human BAC is inserted into the mouse IgH locus and confirmed to be in its intact, native configuration, the FRT-flanked BAC backbone will be excised by using Flp site-specific recombinase. If regular Flp-catalyzed FRT recombination is not high enough, one can use Flo, an improved version of Flpo recombinase which in certain tests is 3-4 times more efficient than the original Flp in ES cells. After the BAC backbone is excised, ES cells will become sensitive to Puromycin (or G418) and resistant to FIAU (for loss of the TK cassette). The excision events will be further characterized by PCR amplification of the junction fragment using human genomic DNA primers. These FRT-flanked BAC backbone-free ES cells will be used for the next round of human BAC insertion and for blastocyst injection.

[0467] Targeting of the genome of an ES cell to produce a transgenic mouse may be carried out using a protocol as explained by reference to the attached FIGS. 1-18.

[0468] FIG. 1 illustrates three basic backbone vectors; an initiating cassette and 2 large insert vectors 1 and 2 respectively. The initiating cassette comprises sequences homologous to the desired site of insertion into the mouse genome, those sites flanking a selectable marker and stuffer primer sequence for PCR based genotyping to confirm correct insertion of BACs. The Stuffer-primer sequence provides the basis for genotyping each BAC addition step. This sequence is considered to provide a robust well validated sequence

template for PCR primer and may be located at the ISceI site, ideally ~1 kb from the BAC insert.

[0469] The large insert vectors comprise human DNA on plasmids with selectable markers and a unique restriction site for linearisation of the plasmid to aid in homologous recombination into the genome of the ES cell.

[0470] FIG. 2 illustrates insertion of an initiating cassette into the mouse genome by Homologous recombination between the mouse J4 and C alpha exons. Puromycin selection allows identification of ES cells with insertion of the cassette. pu(Delta)tk is a bifunctional fusion protein between puromycin N-acetyltransferase (Puro) and a truncated version of herpes simplex virus type 1 thymidine kinase (DeltaTk). Murine embryonic stem (ES) cells transfected with pu(Delta)tk become resistant to puromycin and sensitive to 1-(2-deoxy-2-fluoro-1-beta-D-arabinofuranosyl)-5-iodouracil (FIAU). Unlike other HSV1 tk transgenes, puDeltatk is readily transmitted through the male germ line. Thus pu(Delta)tk is a convenient positive/negative selectable marker that can be widely used in many ES cell applications.

[0471] FIG. 3 illustrates targeting of the large insert vector 1 to the mouse ES cell genome. Linearisation of the vector is made at the same position as the stuffer primer sequence which allows for a gap repair genotyping strategy, well known in the art—see Zheng et al NAR 1999, Vol 27, 11, 2354-2360. In essence, random insertion of the targeting vector into the genome will not 'repair' the gap whereas a homologous recombination event will repair the gap. Juxtaposition of appropriate PCR primer sequences allows colonies to be screened individually for a positive PCR fragment indicating proper insertion. Positive selection using G418 allows for identification of mouse ES cells containing the neo selection marker. PCR verification can be made of all critical V, D and J regions. Array comparative genomic hybridization can be used to validate the BAC structure.

[0472] FIG. 4 illustrates the puro-delta-tk cassette and the BAC plasmid backbone is deleted using Flpe and select in FIAU. Since Flpe works inefficiently in mouse ES cells (5% deletion with transient Flpe expression), it is expected that in most cases, the recombination occurs between the two FRT sites flanking the BAC backbone. Flpo can also be tested to find out the recombination efficiency between two FRT sites that are 10 kb away.

[0473] Given that the FRT deletion step is selectable it is possible to pool FIAU resistant clones and proceed immediately to the next step in parallel with clonal analysis. Alternatively it may be desirable to show by short range PCR that the human sequences are now adjacent to those of the mouse as shown (Hu-primer 1 and Mo-primer) At this stage a 200 kb human locus will have been inserted.

[0474] FIG. 5 illustrates a second large insert vector is targeted into the ES cell chromosome. The human BAC is targeted to the mouse IgH locus using the same initiation cassette insertion followed by ISceI BAC linearization, BAC targeting to the initiation cassette and gap-repair genotyping strategy. Verification of the BAC insertion is carried out as before.

[0475] FIG. 6 illustrates the FRTY flanked BAC backbone of large insert vector 2 and the neo marker are deleted via Flpo. Note that this is not selectable, thus it will be necessary for clonal analysis at this point. This will enable confirma-

tion of the juxtaposition of the human 2 insert with human 1 and other validation efforts.

[0476] At this stage a ~200 kb human locus will have been inserted.

[0477] FIG. 7 illustrates the next large insert vector targeted to the mouse IgH locus. The pu-delta TK cassette is then removed, as for FIG. 4. The process can be repeated to incorporate other BACs.

[0478] FIG. 8 illustrates the final predicted ES cell construct.

[0479] FIGS. 9-18 provide a further level of detail of this process.

Example 2

Site-Specific Recombination

[0480] In a further method of the invention site specific recombination can also be employed. Site-specific recombination (SSR) has been widely used in the last 20-years for the integration of transgenes into defined chromosomal loci. SSR involves recombination between homologous DNA sequences.

[0481] The first generation of SSR-based chromosomal targeting involved recombination between (i) a single recombination target site (RT) such as loxP or FRT in a transfected plasmid with (ii) a chromosomal RT site provided by a previous integration. A major problem with this approach is that insertion events are rare since excision is always more efficient than insertion. A second generation of SSR called RMCE (recombinase-mediated cassette exchange) was introduced by Schlake and Bode in 1994 (Schlake, T.; J. Bode (1994). "Use of mutated FLP-recognition-target-(FRT)-sites for the exchange of expression cassettes at defined chromosomal loci". *Biochemistry* 33: 12746-12751). Their method is based on using two hetero-specific and incompatible RTs in the transfected plasmid which can recombine with compatible RT sites on the chromosome resulting in the swap of one piece of DNA for another—or a cassette exchange. This approach has been successfully exploited in a variety of efficient chromosomal targeting, including integration of BAC inserts of greater than ~50 kb (Wallace, H. A. C. et al. (2007). "Manipulating the mouse genome to engineering precise functional synthetic replacements with human sequence". *Cell* 128: 197-209; Prosser, H. M. et al. (2008). "Mosaic complementation demonstrates a regulatory role for myosin Vila in actin dynamics of *Stereocilia*". *Mol. Cell. Biol.* 28: 1702-12).

[0482] The largest insert size of a BAC is about ~300-kb and therefore this places an upper limit on cassette size for RMCE.

[0483] In the present invention a new SSR-based technique called sequential RMCE (SRMCE) was used, which allows continuous insertion of BAC inserts into the same locus.

[0484] The method comprises the steps of

[0485] 1 insertion of DNA forming an initiation cassette (also called a landing pad herein) into the genome of a cell;

[0486] 2 insertion of a first DNA fragment into the insertion site, the first DNA fragment comprising a first portion of a human DNA and a first vector portion containing a first selectable marker or generating a selectable marker upon insertion;

[0487] 3 removal of part of the vector DNA;

[0488] 4 insertion of a second DNA fragment into the vector portion of the first DNA fragment, the second DNA fragment containing a second portion of human DNA and a second vector portion, the second vector portion containing a second selectable marker, or generating a second selectable marker upon insertion;

[0489] 5 removal of any vector DNA to allow the first and second human DNA fragments to form a contiguous sequence; and

[0490] 6 iteration of the steps of insertion of a part of the human V(D)J DNA and vector DNA removal, as necessary, to produce a cell with all or part of the human VDJ or VJ region sufficient to be capable of generating a chimaeric antibody in conjunction with a host constant region, wherein the insertion of at least one DNA fragment uses site specific recombination.

[0491] In one specific aspect the approach utilizes three heterospecific and incompatible loxP sites. The method is comprised of the steps as follows, and illustrated in FIGS. 22-26:

[0492] 1. Targeting a landing pad into the defined locus. An entry vector containing an HPRT mini-gene flanked by inverted piggyBac (PB) ITRs is targeted into defined region (for example: a region between IGHJ and E μ or IGKJ and E κ or IGLC1 and E λ 3-1) to serve as a landing pad for BAC targeting. The HPRT mini-gene is comprised of two synthetic exons and associated intron. The 5' HPRT exon is flanked by two heterospecific and incompatible loxP sites (one wild-type and the other a mutated site, lox5171) in inverted orientation to each other (FIG. 22). These two loxP sites provide recombination sites for the BAC insertion through RMCE.

[0493] 2. Insertion of the 1st modified BAC into the targeted landing pad. The 1st BAC has a length of DNA to be inserted into the genome flanked by engineered modifications. The 5' modification (loxP—neo gene—lox2272—PGK promoter—PB 5'LTR) and 3' modification (PB3'LTR—puro Δ TK gene—lox5171) is depicted in FIG. 23 along with the relative orientations of the lox sites and PB LTRs. With transient CRE expression from a co-electroporated vector, the DNA sequence would be inserted into the defined locus through RMCE. The cells in which a correct insertion has occurred can be selected as follows: (i) Puromycin-resistance (the puro Δ TK gene has acquired a promoter—"PGK"—from the landing pad), (ii) 6TG-resistance (the HPRT mini-gene has been disrupted), and (iii) G418-resistance (selects for any insertion via the 5' region PGK-neo arrangement). Any combination of these selection regimes can be used. G418- and 6TG-resistance select for correct events on the 5' end while puro-resistance selects for correct events on the 3' end.

[0494] 3. Curing (removing) the 3' modification of the 1st insertion. A properly inserted 1st BAC results the 3' end having a puro Δ TK gene flanked by inverted PB LTRs (FIG. 24)—essentially a proper transposon structure. This transposon can then be removed by the transient expression of the piggyBac transposase (from an electroporated vector). Cells with the correct excision event can be selected by FIAU resistance—ie, no thymidine kinase activity from the puro Δ TK gene. This completely removes the 3' modification leaving no trace nucleotides.

- [0495]** 4. Insertion of a 2nd modified BAC into the 5' end of 1st insertion. The 2nd BAC has a length of DNA to be inserted into the genome (usually intended to be contiguous with the DNA inserted with the 1st BAC) flanked by engineered modifications. The 5' modification (loxP—HPRT mini gene 5' portion—lox5171—PGK promoter—PB3'LTR) and 3' modification (PB3'LTR—puroΔTK—lox2272) is depicted in FIG. 25 along with the relative orientations of the lox sites and PB LTRs. With transient CRE expression from a co-electroporated vector, the DNA sequence would be inserted into the defined locus through RMCE. The cells in which a correct insertion has occurred can be selected as follows: (i) HAT-resistance (the HPRT mini-gene is reconstituted by a correct insertion event, ie: the 5' and 3' exon structures are brought together), and (ii) puromycin-resistance (puroΔTK gene has acquired a promoter—"PGK"—from the landing pad).
- [0496]** 5. Curing (removing) the 3' modification of the 2nd insertion. A properly inserted 2nd BAC results the 3' end having a puroΔTK gene flanked by inverted PB LTRs (FIG. 26)—essentially a proper transposon structure, exactly analogous to the consequence of a successful 1st BAC insertion. And therefore this transposon can likewise be removed by the transient expression of the piggyBac transposase (from an electroporated vector). Cells with the correct excision event can be selected by FIAU resistance—ie, no thymidine kinase activity from the puroΔTK gene. This completely removes the 3' modification leaving no trace nucleotides.
- [0497]** 6. After curing of the 3' modification of the 2nd BAC insertion, the landing pad becomes identical to the original. This entire process, steps 2 through 5, can be repeated multiple times to build up a large insertion into the genome. When complete, there are no residual nucleotides remaining other than the desired insertion.
- [0498]** With the insertion of an odd number of BACs into the Ig loci, the endogenous VDJ or VJ sequences can be inactivated through an inversion via chromosomal engineering as follows (see FIGS. 27-29):
- [0499]** 1. Targeting a "flip-over" cassette into a 5' region 10 to 40 megabases away from the endogenous VDJ or VJ. The flip-over vector (PB3'LTR—PGK promoter—HPRT mini gene 5' portion—loxP—puroΔTK—CAGGS promoter—PB3'LTR) is depicted in FIG. 27 along with the relative orientations of the lox sites and PB LTRs.
- [0500]** 2. Transient CRE expression will result in recombination between the loxP site in the "flip-over" cassette and the loxP site in the 5' modification. This 5' modification is as described in Steps 2 and 3 above—essentially the modification resulting from insertion of an odd number of BACs, after the 3' modification has been cured. The loxP sites are inverted relative to one another and therefore the described recombination event results in an inversion as depicted in FIG. 28. Cells with the correct inversion will be HAT-resistance since the HPRT mini-gene is reconstituted by a correct inversion.
- [0501]** 3. A correct inversion also leaves two transposon structures flanking the "flip-over" cassette and the 5' modification. Both can be excised with transient piggyBAC transposase expression, leaving no remnant of

either modification (FIG. 29). Cells with the correct excisions can be selected as follows: (i) 6TG-resistance (the HPRT mini-gene is deleted) and (ii) FIAU-resistance (the puroΔTK gene is deleted). An inversion as described in the Ig loci would move the endogenous IGH-VDJ or IGK-VJ region away from the Eμ or Ek enhancer region, respectively, and lead to inactivation of the endogenous IGH-VDJ or IGK-VJ regions.

[0502] The methods of insertion of the invention suitably provide one or more of:

- [0503]** Selection at both 5' and 3' ends of the inserted DNA fragment;
- [0504]** Efficient curing of the 3' modification, preferably by transposase mediated DNA excision;
- [0505]** Inactivation of endogenous IGH or IGK activity through an inversion; and
- [0506]** Excision of modifications, leaving no nucleotide traces remaining in the chromosome.

Example 3

Insertion of a Test Vector into the Genome at a Defined Location

[0507] Proof of concept of the approach is disclosed in FIG. 30. In FIG. 30 a landing pad as shown in FIG. 22 was inserted into the genome of a mouse by homologous recombination, followed by insertion of the R21 plasmid into that landing pad via cre-mediated site specific recombination. The insertion event generated a number of general insertion events, 360 G418 resistant colonies, of which ~220 were inserted into the desired locus, as demonstrated by disruption of the HRPT minilocus.

[0508] The R21 vector mimicks the 1st BAC insertion vector at the 5' and 3' ends, including all selection elements and recombinase target sites. In place of BAC sequences, there is a small 'stuffer' sequence. This vector will both test all the principals designed in the invention and allow easy testing of the results in that PCR across the stuffer is feasible and therefore allows both ends of the insertion to be easily tested. R21 was co-electroporated with a cre-expressing vector into the ES cells harbouring the landing pad in the IGH locus. Four sets of transformed cells were transfected in parallel and then placed under different selection regimes as indicated in FIG. 30. G418 selection (neo gene expression) resulted in the largest number of colonies due to there being no requirement for specific landing-pad integration. Any integration of R21 into the genome will provide neo expression leading to G418-resistance. Puro selection resulted in a similar colony number to Puro+6TG or G418+6TG, suggesting that the stringency of Puro selection is due to the PuroΔTK lacking a promoter in the vector. Puro expression is only acquired when an integration occurs near a promoter element—in this design most likely specifically in the landing pad. These conclusions are supported by the results from junction PCR which is shown in FIG. 31.

[0509] The next step in the invention is to 'cure' the 3' end of the integrated BAC vector, leaving a seamless transition between the insertion and the flanking genome. This curing was demonstrated by expanding an individual clone from above (R21 inserted into the landing pad) and expressing piggyBac recombinase in this clone via transfection of an expressing plasmid. FIAU was used to select colonies in which the 3' modification was excised—ie, through loss of the 'PGK-puroΔTK' element between the piggyBac termi-

nal repeats. Fifty such clones resulted from a transfection of 10^6 cells; of these six were tested for the expected genomic structure.

[0510] Successful curing resulted in positive PCR between the primer set labelled "3" in FIG. 32. Of the 6 clones, 4 had correct excisions, 1 clone remained in the original configuration and 1 other had a deletion.

[0511] These data demonstrate iterative insertion of DNA into a landing pad at a defined genomic locus using the approaches outlined above.

Example 4

Insertion of Large Parts of the Human IG Loci into Defined Positions in the Mouse Genome

[0512] Example 3 demonstrated that the design of the claimed invention was capable of providing for the insertion of a test vector into the genome at a defined location, in this case the R21 vector into the mouse IGH locus. The use of the appropriate selection media and the expression of cre-recombinase resulted in a genomic alteration with the predicted structure.

[0513] The same design elements described in this invention were built into the 5' and 3' ends of a BAC insert. Said insert comprised human sequences from the IGH locus and was approximately ~166-kb. This engineered BAC was electroporated along with a cre-expressing plasmid DNA into mouse ES cells harbouring the landing pad at the mouse IGH locus. The transfected cell population was grown in puro-containing media to select for appropriate insertion events.

[0514] Seven resulting clones were isolated and further analysed. The expected recombination event and resulting structure are depicted in FIG. 33. Based upon data from the R21 experiment outlined in Example 3, a stringent selection for correct clones was expected when the transfected population was selected in puro-containing media. This is because the puro-coding region requires a promoter element and this is preferentially supplied by the landing pad after recombination. Accordingly, the majority of the 7 isolated clones had inserted correctly into the genome at the landing pad as determined by the diagnostic PCR. The primers for diagnosing a correct insertion are depicted in FIG. 33. Correct junctions are present in the genome if a 610-bp fragment is amplified between primers 'A' and 'X' and a 478-bp fragment is amplified between primers 'Y' and 'B' (FIGS. 33 and 34). Note that there are amplified fragments between 'A' and '1' primers and '2' and 'B' primers indicating the presence of parental genome (that is, the landing pad alone). These result from parental cells present internally in the cell colonies under puro-selection that escape the selection due to the geometry of a colony. After passaging the colony through puro-containing media, these parental junction fragments disappear indicating that the parental cells are removed from the population. In addition, all the clones were shown to be resistant to 6-TG as expected if the HPRT gene is inactivated by the correct insertion event.

[0515] These data indicate that the disclosed strategy for inserting large parts of the human IG loci into defined positions in the mouse genome will enable the construction of a mouse with a plurality of the variable regions of human IG regions upstream of the mouse constant regions as described.

Example 5

Inserted Loci are Functional in Terms of Gene Rearrangement, Junctional Diversity as Well as Expression

[0516] Bacterial artificial chromosomes (BACs) were created, wherein the BACs had inserts of human Ig gene segments (human V, D and/or J gene segments). Using methods described herein, landing pads were used in a method to construct chimaeric Ig loci in mouse embryonic stem cells (ES cells), such that chimaeric IgH and IgK loci were provided in which human gene segments are functionally inserted upstream of endogenous constant regions. To test if the human IgH-VDJ or IgK-VJ gene segments in the chimaera mice derived from human BAC-inserted ES cell clones appropriately rearrange and express, RT-PCR was performed for the RNA samples of white blood cells from those mice with the primer pairs of human variable(V) region and mouse constant(C) region. The sequences of oligos are shown as follows (Table 1). Each V oligo is paired with C oligo (HV with C μ ; KV with C κ) for PCR reaction.

TABLE 1

Oligo	Sequence
HV2-5	AGATCACCTTGAAGGAGTCTGGTCC (SEQ ID NO 7)
HV4-4	TGGTGAAGCCTTCGGAGACCCTGTG (SEQ ID NO 8)
HV1-3	CACTAGCTATGCTATGCATTGGGTG (SEQ ID NO 9)
HV1-2	ATGGATCAACCCCTAACAGTGGTGGC (SEQ ID NO 10)
HV6-1	GGAAGGACATACTACAGGTCCAAGT (SEQ ID NO 11)
C μ	TAGGTACTTGCCCCCTGTCTCTCAGT (SEQ ID NO 12)
KV1-9	AGCCCAGTGTGTTCCGTACAGCCTG (SEQ ID NO 13)
KV1-8	ATCCTCATCTCTGCATCTACAGGA (SEQ ID NO 14)
KV1-6	GGTAAGGATGGAGAACAACCTGGCAGT (SEQ ID NO 15)
KV1-5	TTAGTAGCTGGTTGGCCTGGTATCA (SEQ ID NO 16)
C κ	CTTTGCTGTCTGTATCAGTCCAAC (SEQ ID NO 17)

[0517] Using the one-step formulation of SuperScriptTM III One-Step RT-PCR System with Platinum[®] Taq High Fidelity (InvitrogenTM; World Wide Web ([www](http://www.invitrogen.com/site/us/en/home/References/protocols/nucleic-acid-amplification-and-expression-profiling/pcr-protocols/super-script-3-one-step-rt-pcr-system-with-platinum-taq-high-fidelity.html#prot3)) invitrogen.com/site/us/en/home/References/protocols/nucleic-acid-amplification-and-expression-profiling/pcr-protocols/super-script-3-one-step-rt-pcr-system-with-platinum-taq-high-fidelity.html#prot3), both cDNA synthesis and PCR amplification were achieved in a single tube using gene-specific primers and target RNAs.

[0518] The RT-PCR results showed most of the human IGH-VDJ or IGK-VJ gene segments appropriately rearrange and express in the chimaera mice. To investigate the details about the diversity generated from VDJ/VJ rearrangement, those specific RT-PCR fragments were cloned into a common vector for sequencing.

[0519] Sequencing results indicate that JH, DH, and JK usages (FIG. 35 and FIG. 36) are similar to human results. In addition, the results from the IGH-VDJCp transcripts show that the range and mean of CDR-H3 length (FIG. 37) are similar to that observed in human. The junctional diver-

sity generated from exonuclease and nucleotide addition activities (FIG. 38) was also observed. The IGH rearrangement possessed a higher frequency of these activities compared to the IGK one. These data suggest that the inserted loci are functional in terms of gene rearrangement, junctional diversity as well as expression.

Example 6

[0520] Productive VJ Rearrangement and Somatic Hypermutation can be Obtained FIG. 41 shows an analysis of kappa mRNA from mice B-cells bearing rearranged VJ, the VJ having been rearranged from human germline kappa V1-8 and J1, and demonstrates that both that productive VJ rearrangement and somatic hypermutation can be obtained, the latter as seen from the changes in antibodies encoded by mRNA with respect to the germline sequences. The same is displayed for V1-6 and J1 in FIG. 42. Importantly, the recombination eliminates stop codons that are encoded by the combination of (unmutated) human germline gene segments, thereby allowing for antibody-encoding mRNA sequences. FIG. 43 demonstrates that inserted human kappa V1-5 J1 and V1-5 J4 can produce functional coding sequences in vivo and junctional diversity.

Example 7

[0521] Inactivation of Use of Endogenous IGHV Gene Segments for Expressed Rearranged Heavy Chain by Inversion

Introduction

[0522] A 5'-RACE C μ -specific library was generated from the splenic B lymphocytes of transgenic mice, denoted S1 mice. These mice comprise transgenic heavy chain loci, each locus containing the six most 3' functional human V_H gene segments (V_H2-5, 7-4-1, 4-4, 1-3, 1-2, 6-1), and all the human D and J_H gene segments inserted into the endogenous heavy chain locus between endogenous IGHJ4 and E μ (mouse chromosome 12: between coordinates 114666435 and 114666436). The human DNA was obtained from a bacterial artificial chromosome (BAC) containing the sequence of human chromosome 14 from coordinate 106328951 to coordinate 106494908. Further details on the construction of transgenic antibody loci using sRMCE is given elsewhere herein and in WO2011004192 (which is incorporated herein by reference). 4x96-well plates of clones were randomly picked for sequencing to determine the usage of the gene segments. All detected immunoglobulin heavy chains were rearranged from mouse V_H or human V_H with human D-J_H. No mouse D and J_H segments were detected in rearranged products (FIG. 44).

[0523] This result indicates that insertion of human V_H-D-J_H gene segments into an endogenous locus between the last endogenous J region (in this case, J_{H4}) and the E μ enhancer effectively inactivates the use of endogenous D and J_H gene segments for expressed rearranged immunoglobulin heavy chains.

[0524] The ratio of mouse V_H to human V_H usage was around 3 to 1 (FIG. 45). To completely eliminate mouse V_H use for antibody generation, the endogenous mouse V_H-D-J_H was inverted and moved to a distant region of the same chromosome. The rearrangement of mouse V_HS to human D-J_H segments was totally blocked by effects of inversion and distance from the heavy chain locus.

[0525] The inversion strategy included three steps: (a) targeting of an inversion cassette, (b) inversion of endogenous VDJ and (c) excision of markers (FIG. 46).

[0526] (a) Targeting of the Inversion Cassette:

[0527] The inversion cassette consists of four components: a CAGGS promoter-driven puromycin-resistant-delta-thymidine kinase (puro Δ tk) gene, a 5' HPRT gene segment under the PGK promoter control, a loxP site between them and inversely oriented to another loxP site already in the heavy chain locus, and two flanking piggyback LTRs (PB3'LTRs). The inversion targeting cassette was inserted to a region that is 5' and distant to the endogenous IGH locus at chromosome 12 as shown in FIG. 46. The targeted ES clones were identified and confirmed by PCR.

[0528] (b) Inversion:

[0529] Following the insertion, transient expression of cre from a transfected plasmid resulted in inversion of a section of chromosome 12 fragment including the endogenous V_H-D-J_H locus and intervening sequences through recombination of two inverted loxP sites, ie, those in the inversion cassette and the landing pad for the BAC insertion respectively. The invertants were selected by HAT and confirmed by junction PCRs cross the two recombined loxP sites.

[0530] (c) Excision of Markers:

[0531] The inversion rearranged the relative orientation of the PB3'LTRs from the inversion cassette and PB5'LTR from the landing pad to generate two piggyBac transposon structures flanking the inverted region. With transient expression of piggyBac transposase (PBase), these two transposons were excised from the chromosome (and thus the mouse cell genome). The cured ES clones were selected by 1-(2-deoxy-2-fluoro-1-b-D-arabinofuranosyl)-5-iodouracil (FIAU) and 6TG, and confirmed by junction PCRs cross the excised regions.

[0532] Methods Tissue culture: The procedures for ES cell culture, electroporation and drug selection have been described previously (Ramirez-Solis, R., A. C. Davis, and A. Bradley. 1993. Gene targeting in mouse embryonic stem cells. *Methods Enzymol.* 225:855-878).

[0533] Targeting of the locus for inversion: Briefly, S1 cell line (S1.11.1) was cultured in M15 medium (Knockout™ DMEM supplemented with 15% fetal bovine serum, 2 mM glutamine, antibiotics, and 0.1 mM 2-mercaptoethanol). Targeting construct R57 (FIG. 47) was linearized outside the region of homology by NotI. A total of 20 μ g of the linearized construct was electroporated into S1 cell lines (AB2.1-derived) with a Bio-Rad® Gene Pulser™, and 107 cells were plated onto three 90-mm-diameter SNL76/7 feeder plates containing M15 medium. At 24 h after electroporation, M15 containing puromycin (3 μ g of the active ingredient per ml) was added to each 90-mm-diameter plate, and the cells were maintained under selection for 9 days. 96 puromycin-resistant clones were then picked and expanded in 96-well plates. The targeting events were identified by long-range PCR.

[0534] Cre-loxP mediated inversion: 12 positive clones were pooled together and cultured in a 6-well tissue culture plate with M15 medium. The cells were transfected with 10 μ g of pCAGGS-Cre plasmid for the inversion of mouse endogenous locus and then plated onto three 90-mm-diameter SNL76/7 feeder plates containing M15 medium. At 24 h after electroporation, M15 containing 1XhAT (hypoxanthine-aminopterin-thymidine) was added to each 90-mm-diameter plate, and the cells were maintained under selection

for 7 days and then treated with 1XHT (hypoxanthine-thymidine) for 2 days. 48 HAT resistant colonies were picked and genotyped by PCR amplification of the junctions after Cre-loxP mediated inversion. HyPBBase-mediated marker excision: 12 positive clones were pooled together and cultured in 6-well tissue culture plate using M15 medium. The cells were transfected with 5 μ g of HyPBBase plasmid to activate the PB transposon LTRs flanking two selection markers (Hprt-mini gene and PGK-puroAtk gene) and plated onto one 90-mm-diameter SNL76/7 feeder plates containing M15 medium. At 72 h after electroporation, a serial dilution of the cells was then plated onto three 90-mm-diameter SNL76/7 feeder plates containing M15 supplemented with 1-(α -2-deoxy-2-fluoro-1- β -D-arabino-furanosyl)-5-iodouracil (FIAU). Cells were maintained under selection for 10 days, and FIAU-resistant colonies were counted, picked, and expanded in 96-well plates. Positive clones were identified by PCR amplification of the junctions after excision of the selection markers.

[0535] Positive clones were then expanded for blastocyst microinjection.

[0536] Generation of chimera and breeding: Mouse chimaeras were generated by microinjection of ES cells into C57/BL6 blastocysts and transferred into pseudopregnant recipients. Male chimaeras were test-crossed with C57/BL6 mice. Agouti F1 offspring were genotyped by S1 3' junction PCR.

[0537] Test-cross positive heterozygotes were further intercrossed to generate homozygotes.

[0538] Determination of V_H-D-J_H usage by rapid amplification of 5'-cDNA ends (5' RACE) PCR: Total RNA was extracted from the spleen of Slinv1 mouse (KMSF30.1d) with TRIzol® Reagent (Invitrogen™, Life Technologies Ltd™) and treated with DNase I. Rapid amplification of 5'-cDNA ends (5' RACE) PCR was performed using 5'/3' RACE kit (2nd Generation, Roche) following the protocol supplied by the manufacturer. The first-strand cDNA was synthesised using primer E1554 (5'-ATGACTTCAGTGTTGTTCTGGTAG -3'; SEQ ID No 25) which is located at the mouse endogenous Cp region. The synthesised first cDNA strand was purified using High Pure PCR Product Purification Kit (Roche). Poly(A) tail was added following the protocol supplied with the 5'/3' RACE kit (2nd Generation, Roche). The 5' end of the V_H-D-J_H rearranged transcript was amplified by nested PCR with forward primers Oligo dT, which is included in the kit, and nested Cp-specific reverse primers E1555 (5'-CACCAGATTCTATCAGAC-3'; SEQ ID No 26).

[0539] Following reaction, the 5' RACE PCR product was checked on a 1% agarose gel and purified using QIAquick® Gel Extraction Kit (QIAGEN) as the protocol supplied with the kit, then cloned into pDrive vector using QIAGEN PCR Cloning Kit (QIAGEN) for sequencing analysis.

[0540] Results

[0541] The sequence analysis from a Cp-specific 5'-RACE library of splenic B lymphocytes of S1 (one human IGH BAC (ie, multiple human VH, all functional human D and JH) with an inverted endogenous IGH locus version 1) mouse shows that practically all the transcripts came from rearranged human V_H-D-J_H gene segments (FIG. 48). Mouse V_H usage was rarely detected (0.4%), and no mouse D and J_H usage was detected. Human V_H usage was 99.6% and only human D and J_H were used; it was hypothesized

that the rare mouse V_H usage was due to trans-switching with another chromosome and not due to use of mouse V_H from the inverted sequences.

[0542] The inversion resulted in complete inactivation of the endogenous V_H use.

[0543] This Result Indicates that Inversion is an Effective Way to Inactivate the Rearrangement of Endogenous V_H Gene Segments.

[0544] The S1^{inv1} mouse also shows a similar usage of both D and J_H gene segments to human (FIG. 49) (Link, J M et al. Mol. Immunol. 2005. 42, 943-955). Thus, a mouse was produced that comprises a transgenic heavy chain locus that expresses heavy chains comprising human variable regions, but no mouse variable regions, and furthermore the human variable regions demonstrated a normal, human sequence distribution corresponding to human D and J usage observed in humans.

Example 8

[0545] Inactivation of Use of Endogenous IGHV Gene Segments for Expressed Rearranged Heavy Chain By Insertion Of Human IgH Genomic DNA

Introduction

[0546] Insertion of human BACs with V_H-D-J_H gene segments into an endogenous mouse heavy chain locus between J_H4 and E μ in chromosome 12 allows human V_H-D-J_H gene segments to effectively use mouse E μ and 3' enhancers and rearrange to generate chimeric antibody with human variable region and mouse constant region. Meanwhile, the endogenous V_H-D-J_H gene segments are pushed away from endogenous enhancers and constant regions. This distance effect results in inactivation of mouse D and J_H use for expressed rearranged antibody products. As the distance increases by stepwise BAC insertion, it is expected that the mouse VH usage would be significantly reduced.

[0547] Results

[0548] Insertion of human DNA from a 1st human BAC (BAC comprising a the sequence of mouse Chromosome 14 from coordinate 106328951 to coordinate 106494908; containing six most 3' functional V_H gene segments (V_H2-5, 7-4-1, 4-4, 1-3, 1-2, 6-1), and all the human D and J_H gene segments) into the heavy chain endogenous locus of a AB2.1 ES cell genome between endogenous IGHJ4 and E μ (at mouse chromosome 12: between coordinates 114666435 and 114666436) effectively inactivates the use of endogenous D and J_H gene segments for expressed rearranged immunoglobulin heavy chain (FIG. 44). The rearranged transcripts with mouse V_H gene segments are reduced in the resulting S1 mouse. The proportion of transcripts using mouse V_H is around 75% of all observed sequences (FIG. 45).

[0549] Following the 1st BAC DNA insertion, human DNA from a 2nd human BAC (Chr14: 106494909-106601551) (BAC comprising a the sequence of mouse Chromosome 14 from coordinate 106494909 to coordinate 106601551; containing 5 more functional VH gene segments (V_H3-13, 3-11, 3-9, 1-8, 3-7)) was inserted into the landing pad left behind after curing following the 1st BAC insertion (see, eg, FIG. 24). The mouse V_H usage is further significantly reduced following this insertion of the 2nd BAC

into the locus. The proportion of transcripts using mouse VH was further reduced to 35% of all observed sequences (FIG. 50).

[0550] This result indicate that the endogenous V_{H-D-J_H} gene segments could be inactivated (ie, not used for expressed rearranged heavy chains) through insertion of human VDJ sequences from one or more BACs. As the distance increases by stepwise BAC insertion, it is expected that the mouse VH usage would be significantly reduced.

Example 9

Normal Class Switch and Hypermutation in Transgenic Mice of the Invention

Introduction

[0551] The B cell arm of the immune system has evolved to produce high affinity, antigen-specific antibodies in response to antigenic challenge. Antibodies are generated in B lymphocytes by a process of gene rearrangement in which variable (V), diversity (D; for the IGH locus) and joining (J) gene segments are recombined, transcribed and spliced to a Cp (for IGH) or a CK or CA (for IGL) constant region gene segment to form an IgM antibody. Depending on the stage of B cell development, IgM is either located on the cell surface or secreted. The recombination process generates a primary antibody repertoire with sufficient germ line diversity to bind a wide range of antigens. However, it is usually not large enough to provide the high affinity antibodies that are required for an effective immune response to an antigen such as an infectious agent. Therefore, the immune system adopts a two-stage diversification process to increase diversity further. When challenged with antigens, B cells undergo selection and maturation by a process called somatic mutation. B cells expressing antibodies which bind to antigen undergo multiple rounds of diversification, clonal expansion and antigen selection in the germinal centres (GCs) of the secondary lymphoid organs. During this process, the rearranged variable regions of the immunoglobulin genes acquire somatic hypermutation through nucleotide substitution, addition or deletion. This stepwise process creates a secondary repertoire from the weak binders selected originally from the primary repertoire and combines rapid proliferation of antigen-reactive B cells with intense selection for quality of binding, eventually giving rise to high affinity antibodies with broad epitope coverage. During this process, antibodies undergo class switching in which the Cp constant region is replaced by Cy, Ca or CE to produce respectively IgG, A or E classes of antibody with different effector functions.

[0552] Insertion of 1st human BAC (Chr14: 106328951-106494908) containing six most 3' functional V_{H} gene segments (V_{H2-5} , 7-4-1, 4-4, 1-3, 1-2, 6-1), and all the D and J_H gene segments into the locus between endogenous IGHJ4 and E μ (Chr12: 114666435 and 114666436) produces transgenic mice that generate chimeric immunoglobulin heavy chains containing human variable and mouse constant regions. This result demonstrates that human immunoglobulin gene segments are able to be rearranged and expressed in mice. Here, RT-PCR experiments and sequence analysis were performed to further demonstrate that immunized transgenic mice have proper class switch and hypermutation for generated antibodies.

[0553] Methods

[0554] RT-PCR and sequence analysis: Wild type or S1 chimera mice at 6-8 weeks of age were primed by intraperitoneal injection of 10⁶ sheep RBCs suspended in phosphate buffer saline (PBS). The immunized mice were boosted twice with the same amount of sheep RBCs two and four weeks after priming. Four days after the last boost, peripheral blood cells were collected from the immunized mice. Total RNA was isolated from peripheral blood cells with TRIzol[®] reagent (Invitrogen[™]) and treated with DNase I. Reverse transcription polymerase chain reaction (RT-PCR) was performed using SuperScript[®] III First-Strand Synthesis System (Invitrogen[™]) following the protocol supplied by the manufacturer. The 1st strand cDNA was synthesized with the specific $C\gamma$ primers ($C\gamma 1$, $C\gamma 2a$, $C\gamma 2b$), following by PCR with specific human V primers (VH1-2,3, VH4-4, VH6-1) and $C\gamma$ primers (Table 2). Following reaction, the RT-PCR product was checked on a 1% agarose gel and purified using QIAquick[®] Gel Extraction Kit (QIAGEN) as the protocol supplied with the kit, then cloned into pDrive vector using QIAGEN PCR Cloning Kit (QIAGEN) for sequencing analysis.

TABLE 2

ELP1352_ $C\gamma 1$	5' -AGAGCGGCCGCTGGGCA ACGTTGCAGGTGACGGTC-3'	SEQ ID No 27
ELP1353_ $C\gamma 2b$	5' -AGAGCGGCCGCTTTGTC CACCGTGGTCTGCTGG-3'	SEQ ID No 28
ELP1354_ $C\gamma 2a$	5' -AGAGCGGCCGCACATTG CAGGTGATGGACTGGC-3'	SEQ ID No 29
ELP1356_ VH4-4	5' -AGGACGCGTGAACACC TGTGGTTCTTCTCTCTGC-3'	SEQ ID No 30
ELP1357_ VH1-2,3	5' -AGGACGCGTCACCATGG ACTGGACCTGGAGGAT-3'	SEQ ID No 31
ELP1358_ VH6-1	5' -AGGACGCGTATGTCTGT CTCCTTCTCATCTTCC-3'	SEQ ID No 32

[0555] Results

[0556] The rearranged transcripts were detected using RT-PCR with human VH-specific and mouse $C\gamma$ -specific primers for amplification from peripheral blood cells of immunized transgenic mice (FIG. 51). Further sequence analysis of these amplified fragments demonstrated hypermutation happened within the human variable regions of these I γ chains (FIG. 52). These results indicate that loci of the invention comprising insertion of human IGH BAC containing V_H , D and J_H gene segments into the locus between endogenous IGHJ4 and E μ regions has normal class switching and hypermutation functionality (IgM to IgG) following antigen challenge.

Example 10

Normal B Cell Compartments in Transgenic Mice of the Invention

Introduction

[0557] In mice, about 2×10^7 bone marrow immature B cells are produced daily. Among them, only 10-20% of these cells survive to exit the bone marrow and enter the spleen. The immature splenic B cell population is divided into two distinct subsets: transitional 1 (T1) and transitional 2 (T2) B

cells. In vivo experiments indicate that T1 cells give rise to T2 cells, whereas T2 cells can further differentiate into mature (M) B cells. In contrast to immature B cells (3-4 days old), mature B cells are long-lived (15-20 weeks old) and are ready to respond to antigens (Pillai S et al; Immunol. Reviews. 2004. 197: 206-218). Thus, the component of mature B cell population is directly linked to the efficiency of humoral immune response.

[0558] The T1, T2 and M cell populations can be categorized by their cell surface IgM and IgD levels. A normal phenotype of splenic B cell compartment is required to mount a robust immune response.

Methods

[0559] Flow cytometric analysis of mature B lymphocytes: To obtain a single cell suspension from spleen, the spleens of mice listed below were gently passaged through a 30 μ m cell strainer. Single cells were resuspended in PBS supplemented with 3% heat inactivated foetal calf serum (FCS; Gibco®). The following antibodies were used for staining:

[0560] Antibody against B220/CD45R conjugated with allophycocyanin (APC) (eBioscience, clone RA3-6B2), antibody against IgD receptor conjugated with phycoerythrin (PE) (eBioscience, clone 11-26) and IgM receptor conjugated with fluorescein isothiocyanate (FITC) (eBioscience, clone 11/41).

[0561] 5×10^6 cells were used for each staining. To each vial containing splenocytes a cocktail of antibodies was added consisting of: IgD (PE) (eBioscience, clone 11-26), IgM (FITC) and B220/CD45R (APC). Cells were incubated at 6° C. for 15 minutes, washed to remove excess of unbound antibodies and analysed using a fluorescence-activated cell sorting (FACS) analyser from Miltenyi Biotech. B-cells were gated as B220⁺IgM⁺IgD⁻ for T1 population, B220⁺IgM⁻IgD⁺ for T2 population and B220⁺IgM⁻IgD⁺ for M population. Percentage of cells was calculated using gating system.

[0562] Results

[0563] Four different genotypes of mice were generated:—

[0564] Wild type (WT);

[0565] A transgenic mouse homozygous for a heavy chain transgene comprising insertion of the 1st BAC human DNA noted above in which there are 6 human VH, all functional human D and JH gene segments (S1/S1);

[0566] A transgenic mouse homozygous for a heavy chain transgene comprising insertion of a human VH, all functional human D and JH gene segments (H1/H1); and

[0567] A transgenic mouse homozygous for a kappa chain transgene comprising insertion of 6 functional human VK and 5 functional JK gene segments (K1/K1).

[0568] Spleens from these naïve mice were collected and analysed for their B cell compartments. The number and percentages of T1, T2 and M cells among those mice are similar (FIG. 53), indicating that genetic manipulation of endogenous IG loci in transgenic mice according to the invention do not compromise their B cell development. These data help to establish that animals according to the invention provide a robust platform for antibody discovery.

Example 11

[0569] Normal IgH Isotypes & Serum Levels in Transgenic Animals of the Invention Transgenic mice (H1) carrying all human JH, all human DH and human Vh2-5 under control of a rat switch region or mice (S1) carrying all human JH, all human DH and human Vh2-5, Vh7-41, Vh4-4, Vh1-3, Vh1-2 and Vh6-1 under control of a mouse switch region were immunised with 100 μ g Cholera Toxin B subunit (CTB; Sigma-Aldrich® C9903) emulsified in Complete Freund's Adjuvant CFA; Sigma-Aldrich® F 5881). At least three animals were injected sc or ip and then boosted with 25 μ g antigen in Incomplete Freund's Adjuvant (IFA; Sigma-Aldrich® F 5506) at (i) 14 days and 21 days or (ii) 28 days after priming. Blood was taken before priming at day “-1” (pre-bleeds) and on the day the spleens were taken (usually 4d after last boost). Serum was analysed by ELISA using an antigen independent assessment of Ig isotypes. This assay detects total serum antibodies of all species. Specific detection for mouse IgG1, IgG2a, IgG2b and IgM was used ((Anti-mouse IgG1 HRP AbD Serotec STAR132P, Anti-mouse IgG2a HRP AbD Serotec STAR133P, Anti-mouse IgG2b HRP AbD Serotec STAR134P, Anti-mouse IgM HRP Abcam® ab97230) and concentrations were read off a standard curve produced for each isotype using polyclonal isotype controls (IgG1, Kappa murine myeloma Sigma-Aldrich® M9269, IgG2a, Kappa murine myeloma Sigma-Aldrich® M9144, IgG2b, Kappa from murine myeloma Sigma-Aldrich® M8894, IgM, Kappa from murine myeloma Sigma-Aldrich® M3795). Results (FIGS. 54 & 55 for H1 homozygous and S1 homozygous and heterozygous mice) showed that even with these relatively short immunisation regimes mice showed an increase in overall IgG levels after immunisation over pre-bleeds. In cases where control mice (+/+) not carrying any human immunoglobulin genes were included and immunised, these mice showed comparable changes in total observed Ig levels (FIG. 54). Individual isotype levels were more variable between animals possibly showing various stages of class switching. IgM levels never exceeded 800 pg/ml whereas IgG levels reached more than 6 mg/ml in some animals. Non-immunised controls showed no such increases in switched isotype Ig levels.

[0570] These results demonstrate that mice comprising multiple human VDJ gene segments under the control of a rat Sp rat or mouse switch are able to undergo productive recombination and class switching in response to antigen challenge and that the mice produce antibody levels that are broadly comparable to unmodified mice The transgenic mice are able to produce antibodies of each of the IgG1, IgG2a, IgG2b and IgM isotypes after immunisation. Titers for CTB-specific Ig in pre-bleeds and terminal bleeds were determined and all immunised animals showed at CTB-specific titres of at least 1/100 000.

Example 12

Generation of Anti-Ovalbumin Antibodies with Sub-50 nm Affinities from Animals of the Invention

[0571] Transgenic mice carrying all human JH, all human DH and human Vh2-5 under control of a rat Sp switch region were immunised with 25 μ g ovalbumin (OVA; Sigma-Aldrich® A7641) in Sigma-Aldrich® adjuvant (Sigma Adjuvant System® S6322) ip and then boosted with the

same amount of OVA in adjuvant at day 14 and day 21. Spleenocytes were taken 4 days later and fused using 1 ml polyethyleneglycol (PEG Average MW1450; Sigma-Aldrich® P7306) with a myeloma line. Fused hybridoma cells were plated on 5 96-well plates and after selection with hypoxanthine-aminopterin-thymidine (HAT) wells tested for expression of OVA-specific antibodies by ELISA. Clones positive by ELISA were re-tested by surface plasmon resonance (SPR) and binding kinetics determined using the ProteOn™ XPR36 (Bio-Rad®). Briefly, anti-mouse IgG (GE Biacore™ BR-1008-38) was coupled to a GLM biosensor chip by primary amine coupling, this was used to capture the antibodies to be tested directly from tissue culture supernatants.

[0572] Ovalbumin was used as the analyte and passed over the captured antibody surface at 1024 nM, 256 nM, 64 nM, 16 nM, 4 nM with a 0 nM (i.e. buffer alone) used to double reference the binding data. Regeneration of the anti-mouse IgG capture surface was by 10 mM glycine pH1.7, this removed the captured antibody and allowed the surface to be used for another interaction. The binding data was fitted to 1:1 model inherent to the ProteOn™ XPR36 analysis software. The run was carried out 1xHBS-EP (10 mM Hepes, 150 mM NaCl, 3 mM EDTA, 0.05% polysorbate, pH7.6 (Teknova H8022)) used as running buffer and carried out at 25° C. For 8 positive clones, heavy chain V-regions were recovered by RT-PCR (Access RT-PCR System, A1250, Promega) using forward primers specific for Ig signal sequences (Wardemann et al Science 301, 1374 (2003)) and the following reverse primers for the constant regions of mouse IgG (Table 3):

TABLE 3

Primer Name	Sequence	bp	
mIgG1_2 rev	GGGGCCAGTGG ATAGACAGAT	21	SEQ ID No 33
mIgG2b rev	CAGTGGATAGA CTGATGG	18	SEQ ID No 34
mIgG2a_2 rev	CAGTGGATAGA CCGATGG	21	SEQ ID No 35
mCH1 unirev	KCAGGGGCCAG TGGATAGAC	20	SEQ ID No 36
mCH1 unirev_2	TARCCYTTGAC MAGGCATCC	20	SEQ ID No 37

[0573] RT-PCR products were either directly sequenced using the same primer pairs or cloned in to TA plasmids (TOPO® TA Cloning® Kit for Sequencing, K4595-40, Invitrogen™) and submitted for plasmid sequencing. Results (Table 4, below) show that CDRH3 sequences had variable CDRs except for two identical clones (16C9 and 20B5) that also had near identical KD kinetic values.

[0574] The determined equilibrium binding constant KD ranged from 0.38 nM to 40.60 nM, as determined by SPR at 25° C.

[0575] These results demonstrate that mice comprising multiple human VDJ gene segments under the control of a rat Cp switch are able to undergo productive recombination and produce high affinity antigen-specific antibodies whose

CDR3 regions have sequences encoded by human gene segments (human JH was separately identified by V-Quest, IMGT).

TABLE 4

KD [nM]	clone code	CDR3 and FR4 (underlined) according to Kabat definition	
0.38	16C9	QEVINYYYYGMDV <u>WGQGTTVTVSS</u>	SEQ ID No 38
0.52	20B5	QEVINYYYYGMDV <u>WGQGTTVTVSS</u>	SEQ ID No 39
5.89	19F4	LEMATINYYYYGM <u>DVWGQGTMTVTVSS</u>	SEQ ID No 40
39.70	19E1	QEFGNYYYYGMDV <u>WGQGTTVTVSS</u>	SEQ ID No 41
3.10	19G8	QEDGNPYFFGMDF <u>WGQGTTVTVSS</u>	SEQ ID No 42
8.95	20H10	GSSYYDGMVDVW <u>QGTTTVTVSS</u>	SEQ ID No 43
4.46	18D10	LENDYGYYYYYGMD <u>VWGQGTTVTVSS</u>	SEQ ID No 44
40.60	16F2	RGGLSPLYGMDVW <u>GQGTTVTVSS</u>	SEQ ID No 45

Example 13

Generation of Anti-Cholera Toxin B Antibodies With Human Vh Regions from Animals of the Invention

[0576] Transgenic mice carrying all human JH, all human DH and human Vh2-5, Vh7-41, Vh4-4, Vh1-3, Vh1-2 and Vh6-1 under control of a mouse Sp switch region were immunised and fused as described in Example 11. Fused hybridoma cells were plated on 5 96-well plates and after selection with hypoxanthine-aminopterin-thymidine (HAT) or G418 (Gibco® Cat No 10131-027, Lot 503317) and wells tested for expression of CTB-specific antibodies by ELISA. Clones positive by ELISA were re-tested by surface plasmon resonance SPR and binding kinetics determined using the ProteOn XPR36™ (Bio-Rad®).

[0577] Briefly, anti-mouse IgG (GE Biacore™ BR-1008-38) was coupled to a GLM biosensor chip by primary amine coupling, this was used to capture the antibodies to be tested directly from tissue culture supernatants. Cholera toxin B was used as analyte and passed over the captured antibody surface at 256 nM, 64 nM, 16 nM, 4 nM and 1 nM, with a 0 nM (i.e. buffer alone) used to double reference the binding data. Regeneration of the anti-mouse IgG capture surface was by 10 mM glycine pH1.7, this removed the captured antibody and allowed the surface to be used for another

interaction. The binding data was fitted to 1:1 model inherent to the ProteOn XPR36™ analysis software. The run was carried out 1×HBS-EP (10 mM Hepes, 150 mM NaCl, 3 mM EDTA, 0.05% polysorbate, pH7.6 (Teknova H8022)) used as running buffer and carried out at 37° C.

[0578] From the clones initially identified by ELISA, binding to CTB was confirmed by SPR. However, due to the pentameric nature of the cholera toxin B, the majority of fits to the 1:1 model were poor and the equilibrium binding constant KDs could not be accurately determined. Where fits were acceptable, equilibrium binding constant KDs determined ranged from 0.21 nM to 309 nM but due to the pentameric nature of cholera toxin B these are likely to be the result of multimeric interactions and therefore apparent affinities with possible avidity components.

[0579] Clones identified by SPR for binding to CTB were subjected to RT-PCR as described in Example 12 to recover the Vh regions. RT-PCR products were directly sequenced using the same primer pairs. Results were obtained for only 14 clones presumably because the human primers described in Wardemann et al were not designed to amplify mouse Vh regions and therefore may have failed to amplify certain mouse Vh classes. Results showed that 3 of the 14 CTB-specific recovered heavy chain V-region sequences were human V, D and J regions as identified by V-Quest, IMGT (Table 5).

OTHER EMBODIMENTS

[0580] From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

[0581] The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

[0582] All publications and patent applications mentioned in the specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

TABLE 5

Alignment of heavy chain CDRs and J-region of 3 clones identified as binding to CTB and preferentially matching with human reference sequences from IMGT database; note that the KD values given here are apparent values due to the avidity of the CTB-antibody interaction						
Vh	Clone	Sequence (Kabat definitions)				KD
region	Name	CDR1	CDR2	CDR3	J-regions	[nM]
IGHV4- 4*02	—	SSNWWS (SEQ ID NO 51)	EIYHSGSTNYNPSL KS (SEQ ID NO 56)	n/a	IGHJ2*01 YWYFDLWGRGTLVT VSS (SEQ ID NO 64)	—
	12D10	SGNWWS (SEQ ID NO 52)	EIYHSGTNYNPSL KS (SEQ ID NO 57)	GPLTGEKYYFDL (SEQ ID NO 61)	-YYFDLWGRGTLVT VSS (SEQ ID NO 65)	0.27
	1283	RSNWWS (SEQ ID NO 53)	EIYHSGSTNYNPSL KS (SEQ ID NO 58)	IGDWYFDL (SEQ ID NO 62)	-WYFDLWGRGTLVT VSS (SEQ ID NO 66)	0.85
IGHV6- 1*01	—	SNSAAWN (SEQ ID NO 54)	RTYYRSKWYNDYAV SVKS (SEQ ID NO 59)	n/a	IGHJ3*01 DAFDVWGQGTMTVT SS (SEQ ID NO 67)	—
	4A12	SNSAAWN (SEQ ID NO 55)	RTYYRSKWYNDYKV SVKS (SEQ ID NO 60)	EGSHSGSGWYLDAP DI (SEQ ID NO 63)	DAFDIWGQGTKVTV SS (SEQ ID NO 68)	1.61

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 68

<210> SEQ ID NO 1

<211> LENGTH: 4272

<212> TYPE: DNA

<213> ORGANISM: Rattus norvegicus

<400> SEQUENCE: 1

```
agatctgccc atctcaggct agttaaatta gttcatccca gtttgccca acttacccca    60
tctagagtag cgaaactaat ctgagcctag ctaagtccag ttagttaa ttagtccag    120
cttgccacag gctaatacat actgctacag tttggtctag cctaccctaa ttaagctgat    180
ccaggcctgg gttagacctag ctcatctcag cccagtttag gttatccagt acatctcttt    240
ccagttcagc tcaggttacc ataccttatc tcaattcagc tcagctagtg taattcatct    300
tagttcatcc cctaccctc tagactcct gttgatctta actcagttta gacatggcca    360
acaaagcctg gccaaactca ggccagggta gtgtagctca gcataagcag tctagccttg    420
ctcagcttag ctacccttc ctcatctaaa ttcaactcag ctatgccggc cctgcagcag    480
gtcccctcag ctacccaag tccaaccagt tcagctctggc tcatttaagt cttgacaatc    540
cccaattcat cccagctcag cttagcataa ctcaggcagt ccattcttag cccaaccag    600
tttagcccag tttatcccag ttcactctgg ctgactcag tgcaactcga ttcagttct    660
cccaggccac ctacgccag ttcagtgtag ctcatctgag cccaacttat cccagctcat    720
cccaaacccac ctacccaag cctgctcag cctagctcat ctgagcctag ttcacactct    780
ctcatcctgc cagctagccc agtttagtcc acatcatctt gcaaagctca accagcccaa    840
gtcagccggg tccagctcat tcagttcaca accagctcag tcagctcat cctaactcag    900
cctcaccatc atccacatca gctagcccag ttcagctgag ctcatcccag cccacttcaa    960
tcacagctca ttaagtaca gctcaccoca gctctattta gctcaagctca gcttatttag   1020
cctacttcat cccagctcag cccagccaac tcaactcacc ctgctcagc taaacctctgc   1080
tcagctcacc caagcaaac tgactccaac ccagatcctt tcagctcagc tcaccagct    1140
caggccagct caccatccc agctcaccca gcttagctca cccagcccag ctacagcccag   1200
ctcaccagc ccagctcagc ccagctcacc cagcccagct cagcccagct cagctcagct   1260
cagctcagct cagctcagct caccagctc agctcagcca gctcagctca cccagctca    1320
gtccagctca gttcagctca cccagctca gctcaccoca ctcagctcacc tcaactcagc   1380
tcacccaact cagctcagct cagttcacc agctcagctc accagccca gcacagctca    1440
tctcaccag ctacagctcacc cagcccagc tcaccccagc tcaccccagc tcagctcagc   1500
tcaccccagc tcagcccagc tcagctcacc cagctcagct caccacaact agctcagctc   1560
agttcaccca gctcagctca cccagcccag cacagctcat ctaccccagc ccagctcacc   1620
ccagctcacc ccagctcagc tcagctcagc ccagctcacc cagctcagct cagctcacc   1680
cagctcagct caccagctc agctcaccca gccagctca gctcagctca cccagctca    1740
gcccagctca gctcaccag ctacagctcacc cagcccagc tcaccccagc tcaccccagc   1800
tcagttcagc tcagttcagc tcaccccagct cagctcacc aactcaactc agctcagttc   1860
accagctca gctcagctca cccagctca cccagctca cccagctcag ttcagctcacc   1920
cccagctcag ttcaccccag tcagctcacc cagcccagct cagcccagct caccagct    1980
```

-continued

cagctcaacc agatcagctc agcccagctc accctagttt agttcaccca gccagctca	2040
ccccagctca gctcacccca actcagctca cccagctcat cccagctcag ccagctaate	2100
ccagctcagc tcacccagc tcagctcacc cagctcagct cacccactc agctcacccc	2160
agctcacccc agctcatccc agctcatccc agttcagacc tggtcagctc atctcacccc	2220
agctcagctc accccagttc agctcaccta gcccaactca cccagctca gtccagctca	2280
gttcagctca ccccactca tctcacccag ctcagctcac cccagctcat cccagctcag	2340
ctcacccag ttcagccctg ttcagctcat ctcacccagc tcagctcacc cagcccagct	2400
cacccagct cacccagct cagtcagct cagttcagct caccagctc agctcaccca	2460
actcaactca gctcagttca cccagctcag ctcagctcac cccagctcac ccagctcagt	2520
tcagctcacc ccagctcagt tcacccagct cagctcacc agcccagctc aaccagatca	2580
gctcagccca gctcaccta gtttagttca cccagcccag ctcacccag ctcagctcac	2640
cccactcag ctcacctagc tcacccagc tcagctcacc ccagctcagc tcacccagc	2700
tcattcacc ccagctcagc tcacccagct catcccagct cagctcagcc cagctcatcc	2760
cagccctgct catcccagct cagctcagct cagcccagct cagcccagct cagcccagct	2820
cagcccagct cagcccagct cagctcaacc cagctcagct caccagccc agctcagccc	2880
agctcaccca gctcagctca cccagctca gctcacccca gctcatctca cccagctcag	2940
ctcacccagc tcagcccagc tcagctcagc tcacccagct catctcacc agctcagctc	3000
acccagctc atcccagctc agctcacccc agttcagccc tggtcagctc atctcacccc	3060
agctcagctc acccagttca gctcatocca gcccatocca gctcagctca gccagctca	3120
gccagctca gcccagccca gcccagccca gctcagctca gccagctca gccagctca	3180
gtccagctca gcttagccca gcccagctca gctcagccca gctcagccca gctcagccca	3240
gctcagctca cccagctcac cccagctcag cccagctcag cccagctcag ctcacccagc	3300
tcacccacc ccagctcacc ccagttcagc ccagctcagc ccagctcagc ccagcccagc	3360
ccagcccagc ccagcccagc tcagcccagc tcagctcagc ccagcccagc tcagctcagc	3420
ccagctcagc ccagctcacc ccagctcagc tcacccagc tcagcccagc tcagcccagc	3480
tcagctcacc cagctcacc caccagctc caccagctt cagcccagct catccagctc	3540
agctcacccc cagctctgct caccagctc agctcagctt acccagctca gctcaactca	3600
cccagctcag ctcacccagc tcagctcagc tcacccagc ccagctcagc tcagctcacc	3660
ccagctctgc tcacccagct cagctcagct cacctcagct ctgctcacc agctcagctc	3720
aaccacctca ggtcagccca gctcacccca gcttaoccca gctcacccag ctcagctcag	3780
ctcacccagc tcagctcacc cagctcagct caccagctc taccagctc caccagctc	3840
cagctaacc agctcagctc acccagctca gctcacccag ctcagctcat cccagctcac	3900
cccagctacc acagagtagc tcagctcagc tcagctcacc ccagcacaac acagcccac	3960
acagctcagt tcagagcagt ccagtagagt ttagctccaa tcagcccaga tcaagacaat	4020
tcattccaat ttggctatct tggtaagtc agcctagttt agcttagccg gcctagctca	4080
attcagctca ttgagctca cctcgctcct gctcaagtc agctttggct acctcagagt	4140
aatcatctca gcttagcaca tttttgaagg gctcagggaa gcctacacat ctcagtccaa	4200
ctgtgcttaa ctagagccta gcttcctagc caggtgtca acctgttca ctaaattttg	4260

-continued

ctcagcaagc tt 4272

<210> SEQ ID NO 2
 <211> LENGTH: 22190
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Tragetting Vector (long version)

<400> SEQUENCE: 2

gcggccgcaa cctgggcaaa tgggagctta gcaacaatgt agggggctgg acctagactt 60
 cctacacatt tgtagcagat gtgcagcttg gtcttcatgt gtgtattacc ctaacatttg 120
 gagcaggagc tgtctctgac tctgttgctt gccattggat ccccttcccc tgettgggct 180
 gccttgtttg gccttagtag gaaaggatgt gcttagtctt gctgtgactt gatgtcccta 240
 ggcaaatga taccccaggg gggctcccca tctctgagga gatgggcaaa gggtaatgg 300
 tggagggact tgtgaggctg ggactgggag gagagaaagg agacagctgt aactggaatg 360
 atgttaagtg aacaaatgaa tggatagatt agatagacag atagacagac agacagacag 420
 acagacagac agacagacag acagatagaa agatagatag ataaggggaa aaagaaacgt 480
 agctgagcaa gccagagaga gcaagccaaa taagcagcat tcctccatga cttttccttc 540
 agctcctgcc tatgagtctg ccttgacttc cctcagtgat tggttgtaag ttaaaagggtg 600
 aaataaacc tttctttgac aagtgtcttt tggttctgat tttatcaca gcaagagaaa 660
 atcaaaactag aacaacatg tatttttctt ggcacatgtc catagtaagg cagaaatgat 720
 cttcagacct agaccataga tactacagag agcagaagtg tagatagggt gacttactgt 780
 atgattgtaa tccaagtaaa tctacatagc tagagagcta gaggaaaggc caaagcttcc 840
 tctgggagggt cagatcctgt cgcactgtag ccaataaggc atattgcac acaggaaagg 900
 actaagacc aggctggcaa tagtgtctgt atcttaacta gacctctcta gtgagtgagg 960
 aaggaaagtt gtgagagccc agactgtggg ctcggaagg acctgccatg cccctgttag 1020
 taactgagta ctacagcagg agcagggtgt ctctagaaag cctgagacaa ctctacttct 1080
 tctctcaaga gaccacctaa tacaggctg agagaacaga ctctgaaat agatgggact 1140
 taaggagcta agatctagag ctcatctaca gagcagaatc ccagccaaga gaacaaagaa 1200
 tactggctct ctctcctggt cctactcct agagttctaa aacacactat agggaaagga 1260
 gcctctagac ctccgtccat tccccatctt gctcattcca tcttccatg tccccaggctc 1320
 tccaagccac agacactacc tttcctatc acccaccttt ctgtgtccct aggtccccag 1380
 gccatagtca cctcccccca cacacacccc actcacctg ccccatctat gcccttagat 1440
 gcttacttac cagagtcttt tgtctgacgt ggggctacaa gcatctatgc tccctaagca 1500
 cctactgctg acctgtagga cccagctctg aaccaactca tataagtaaa tacagactct 1560
 cccctgtctt aggatggcct cctggatcag gaggagacca ctgccaaaga accttctctc 1620
 agagcaactga actcctcccc tgtaccactt aggcagacac tgagacctat tattactgat 1680
 taccagagct ctggcagtga ccacggagga gatagggtcca ccctggacac aggaaacaca 1740
 gcagcagaga tactgctcca tcacaacagt agagtgcacac tttagacttt aatttgggtc 1800
 actttcctgc tgcagagggt ggatcagaaa gcaagagca gtatgagtgc ctgataggca 1860
 cccaagtaca ctatagagta ctcatggtga ataaggtacc tccatggctt cccagggagg 1920

-continued

ggcactgccc cccccccacc atcacagacc tttctccata gttgataact cagacacaag 1980
tgaatgacag atggacctcc atctactctt attttaaaaa gaagacaaac cccacaggct 2040
cgagaacttt agcgactggt ttgagagaaa tcattggtcc ctgactcaag agatgactgg 2100
cagattgggg atcagaatac ccatactctg tggctagtgt gaggtttaag cctcagagtc 2160
cctgtggtct ctgactggtg caaggttttg actaagcggg gcaccacagt gctaactggg 2220
accacgggta cacgtggctc aacaaaaacc ttctgtttg agctctccag gggcagcctg 2280
agctatgagg aagtagagag gcttgagaaa tctgaggaag aaaagagtag atctgagagg 2340
aaaggtagct ttctggaggc caggagacag tgcagagaag aacgagttac tgtggacagg 2400
tcttagatgg ggaagaatg agcaaatgca agcatcagaa ggggtgatgc aatgtcctgc 2460
caaggactta ccaagaggat ccccgacag agcaggcagg tggagttgac tgagaggaca 2520
gggtagggtc aggtccctct ctgcttctct ttctctctct cctgttctct tcctctcttg 2580
tcacagggtc cactatgcta gccaaaggta gectgaaaga ttaccatcct acagatgggc 2640
ccatccagtt gagttaaggc ggagatctct ccaaacatct gagtttctga ggcttgatg 2700
ccactgggga cgccaaggga ctttgggctg ggtttgggtg gcccagatg aagggtact 2760
tcaactgggc tataattact ctgatgteta ggaccagggg gctcaggtea ctcaggctag 2820
gtgagtcctg catctgggga ctgtggggtt caggtgtcct aaggcaggat gtggagagag 2880
ttttagtata ggaacagagg cagaacagag actgtgctac tggctactcg atgtctgggg 2940
cgcagggacc acggtcaccg tctcctcagg taagctggct tttttcttcc tgcacattcc 3000
attctgaaat gggaaaagat attctcagat ctccccatgt caggccatct gccacactct 3060
gcatgctgca gaagcttttc tgaaggata gggctctcac tcccaggaaa agaggcagtc 3120
agaggctagc tgctgtgga acagtgacaa tcatggaaaa taggcattta cattgttagg 3180
ctacatgggt agatggggtt ttgtacacc actaaagggg tctatgatag tgtgactact 3240
ttgactactg gggccaaggc accactctca cagtctctc aggtgagtc ttacaacctc 3300
tctctctat tcagcttaaa tagattttac tgcatttgtt gggggggaaa tgtgtgtatc 3360
tgaatttcag gtcatgaagg actagggaca ccttgggagt cagaaagggt cattgggagc 3420
cctggctgat gcagacagac atcctcagct cccagacttc atggccagag atttataggg 3480
atcctggcca gcattgccc taggtccctc tctctctatgc tttctttgct cctcactggc 3540
ctccatctga gatcatcctg gagccctagc caaggatcat ttattgtcag gggctctaatc 3600
attgttgtea caatgtgcct ggtttgctta ctggggccaa gggactctgg tcaactgtctc 3660
tgcagggtgag tcctaacttc tcccattcta aatgcatggt ggggggatc tgagccttca 3720
ggaccaagat tctctgcaaa cggaatcaa gattcaacct ctttgtccca aagttgagac 3780
atgggtctgg gtcagggact ctctgectgc tggctctggt tgacattaga actgaagtat 3840
gatgaaggat ctgccagaac tgaagcttga agtctgaggc agaactctgt ccagggtcta 3900
tcggactctt gtgagaatta ggggctgaca gttgatgggtg acaatttcag ggtcagtgac 3960
tgtctggttt ctctgagggt aggctggaat ataggtcacc ttgaagacta aagaggggtc 4020
caggggcttc tgcacaggca ggaacagaa tgtggaacaa tgacttgaat ggttgattct 4080
tgtgtgacac caggaattgg cataatgtct gagttgccc ggggtgatc tagtcagact 4140
ctggggtttt tgcgggtat agaggaaaaa tccactattg tgattactat gctatggact 4200

-continued

actggggca aggaacctca gtcaccgtct cctcaggtaa gaatggcctc tccaggtctt	4260
tatttttaac ctttggtatg gagttttctg agcattgcag actaatcttg gatatttgc	4320
cctgaggag cggctgaga gaagtggga aataaactgt ctagggatct cagagccttt	4380
aggacagatt atctccacat ctttgaaaaa ctaagaatct gtgtgatggt gttggtggag	4440
tccctggatg atgggatagg gactttggag gctcatttga gggagatgct aaaacaatcc	4500
tatggctgga gggatagttg gggctacgcg tttttaacct tagaaagata gtctgcgtaa	4560
aattgacgca tgcattcttg aaatattgct ctctctttct aaatagcgcg aatccgtcgc	4620
tgtgcattta ggacatctca gtcgccgtt ggagctccc tgaggcgtgc ttgtcaatgc	4680
ggtaagtgc actgatcttg aactataacg accgcgtgag tcaaaatgac gcatgattat	4740
cttttacgtg acttttaaga ttaactcat acgataatta tattgttatt tcatgttcta	4800
cttacgtgat aacttattat atatatttt tcttgttata gatatcgcta gtggatccgg	4860
ctggtctctt ccgctcaga agtactttt tttttttttt tttttttttt tttttttttt	4920
ttttttttt ttttttttt ttttttaaat ttttgggaat ttattgattt gcatttaaaa	4980
gggaactgct gacaaagatt cactggtaat aatttgaaca agttggaaaa tacagtcaac	5040
attactgaaa cactactaaa ataattocag gacagaacaa aacttcttag atgctgtctt	5100
tgatgtgaaa attgactgct tcttactttt ctaacacacg gtggtataat taacaatatt	5160
caatcacttc tattctttcc tgcataata aaaaataaaa taccaatata aaaactaata	5220
tatctctctt ttattcttta cagatagag ttcaatgttt cactcaatag tgctgtggtt	5280
taagagaatt ttttcattta caagttaac aacaatccgc ccaaagggaa ctgatagtct	5340
ataggctcat agtgcaata aacagtttag gaatgcagca actgacattt ctaaagtaca	5400
aaacagataa aattcttaga agatacatgc aaaaagctct actaagcaga tggccacaga	5460
actagaacat tgataatttt actggcgtg tcaataggac tccagatgtt tccaaactca	5520
acttgaactc tcactctagg ctttgtattt tgcttttcca gtttcaacta tgacacaaac	5580
atgattcaaa tccctgaagt attcattata gtcaagggca tatcctacaa caaacttgtc	5640
tggaatttca aatccaacaa agtctggctt atatccaaca cttcgtgggg tccttttcac	5700
cagcaagctt gcgacctga ccatctttgg attatactgc ctgaccaagg aaagcaaagt	5760
ctgcattggt ttgccagtgt caattatctc tccacaatc aagacattct tccagttaa	5820
agttgagaga tcactctccac caattacttt tatgtcccct gttgactggt cattacaata	5880
gctcttcagt ctgataaaat ctacagtcac aggaatggat ctatcactat ttctattcag	5940
tgctttgatg taatccagca ggtcagcaaa gaatttatag ccccccttga gcacacagag	6000
ggctacaatg tgatggcctc ccatctcctt catcacatct cgagcaagac gttcagtcct	6060
acagaaataa aatcaggaat ttaatagaaa gtttcataca ttaaacttta taacaaacac	6120
ctcttagtca ttaaacttcc acaccaacct gggcaatata gtgagacccc atgctgcaa	6180
aaaaaaaaa attagccagg catggttagca tgtacctgta gtcccagcta cttgagaggt	6240
gaggtgggaa aatcacttta gtgcaggatg ttgaggctgg agtgaactgt gattgtgcca	6300
ctgcactcca gcctggacaa tagagcaaga ccttgtctca aaaaaatgca taaaaattt	6360
tttttaaatc tccacgtaa cacatccttt gccctcatgt ttcataaggt aaaaaattg	6420
ataccttcaa aaaaaccaag cataccacta tcataatttt ttttaaatgc aaataaaaac	6480

-continued

aagataccat	tttcacctat	cagactggca	ggttctgatt	aatgaaatt	tcttgataa	6540
tatacaatat	taagagagac	tgtagaaact	gggccagtgg	ctcatgcctg	taatcccagc	6600
actttgggag	gctgggtaac	atggcgaacc	ctgtttctac	aaaataaaaa	tattagctgg	6660
gagtgggtggc	gcacacctat	agtcccagct	actcaggagg	ctgaggtgga	aggatcgctt	6720
gaaccacagga	ggttgagact	gcagtgaact	gtgatcattc	tgctgcactg	caccccagcc	6780
tgggcaacag	agacctgtc	tcaaaaaaaaa	aaaaaaaaaga	gacaaattgt	gaagagaaa	6840
gtactctcat	ataacatcag	gagtataaaa	tgattcaact	tcttagagga	aaatttggca	6900
ataccaaaat	attcaataaa	ctctttcccc	ttgaccaga	aattccactt	gaataaagct	6960
gaacaagtac	caaacatgta	aaagaatgtt	tcttctagta	cagtcggtaa	gaacaaaata	7020
gtgtctatca	atagtggact	ggttaaatca	gttatggtat	ctccataaga	cagaatgcta	7080
tgcaaccttt	aaaatatatt	agatagctct	agacagtggg	tcccctcgag	ggaccttaata	7140
acttcgtata	gcatacatta	tacgaagtta	tattaagggg	tattgaaat	gtcgactaga	7200
cacactaata	ttaaaagtgt	ccaataacat	ttaaaactat	actcatacgt	taaaataata	7260
atgtatatat	gtacttttgc	atatagtata	catgcatagc	cagtgcttga	gaagaaatgt	7320
gtacagaagg	ctgaaaggag	agaactttag	tcttctgtt	tatggcctcc	atagttagaa	7380
tattttataa	cacaaatatt	ttgatattat	aattttaaaa	taaaacaca	gaatagccag	7440
acatacaatg	caagcattca	ataccaggta	aggtttttca	ctgtaattga	cttaacagaa	7500
aattttcaag	ctagatgtgc	ataataataa	aaatctgacc	ttgccttcat	gtgattcagc	7560
cccagtccat	tacctgttt	aggactgaga	aatgcaagac	tctggctaga	gttcttctt	7620
ccatctccct	tcaatgttta	ctttgttctg	gtccctacag	agtcccacta	taccacaact	7680
gatactaagt	aattagtaag	gccctcctct	tttattttta	ataaagaaga	ttttagaaa	7740
catcagttat	ttaataagtt	ggcctagttt	atgttcaaat	agcaagtact	cagaacagct	7800
gctgatgttt	gaaattaaca	caagaaaaag	taaaaaacct	cattttaaga	tcttacttac	7860
ctgtccataa	ttagtccatg	gggaataaac	accctttcca	aatcctcagc	ataatgatta	7920
ggtatgcaaa	ataaatcaag	gtcataacct	ggttcatcat	cactaatcac	gacgccaggg	7980
ctgcgggtcg	ccataacgga	gccggccggc	gcgcgggctg	aataacttcg	tataatgtgt	8040
actatacgaa	gttatttgtt	caggaggagg	aagccggtgg	gggagcagag	gaggaggcgg	8100
aggcgcagca	agaccccccc	ccccctgcag	gtcgaaggc	ccggagatga	ggaagaggag	8160
aacagcgcgg	cagacgtgcg	cttttgaagc	gtgcagaatg	ccgggcctcc	ggaggacctt	8220
cgggcgcgcc	ccccgcctct	gagccccgcc	ctgagccgc	ccccggaccc	acccttccc	8280
agcctctgag	cccagaaagc	gaaggagcaa	agctgctatt	ggccgctgcc	ccaaaggcct	8340
accgccttcc	attgctcagc	ggtgctgtcc	atctgcacga	gactagtggag	acgtgctact	8400
tccatttgtc	acgtcctgca	cgacgcgagc	tgcggggcgg	gggggaactt	cctgactagg	8460
ggaggagtag	aaggtggcgc	gaaggggcca	ccaaagaacg	gagccggttg	gcgcctaccg	8520
gtggatgtgg	aatgtgtgcg	aggccagagg	ccacttgtgt	agcgcacaagt	gccagcggg	8580
gctgctaaag	cgcatgctcc	agactgcctt	gggaaaagcg	cctcccctac	ccggtagata	8640
tctataacaa	gaaaatatat	atataataag	ttatcacgta	agtagaacat	gaaataacaa	8700
tataattatc	gtatgagtta	aatcttaaaa	gtcacgtaaa	agataatcat	gcgtcatttt	8760

-continued

gactcacgcg	gtcgttatag	ttcaaaatca	gtgacactta	ccgcattgac	aagcacgcct	8820
cacgggagct	ccaagcggcg	actgagatgt	cctaaatgca	cagcgacgga	ttcgcgctat	8880
ttagaaagag	agagcaatat	ttcaagaatg	catgcgtcaa	ttttacgcag	actatctttc	8940
tagggttaaa	agaattcgat	atcaagctta	tcgatgtagt	tggagatfff	cagtttttag	9000
aataaaagta	ttagttgtgg	aatatacttc	aggaccacct	ctgtgacagc	atttatacag	9060
tatccgatgc	atagggacaa	agagtggagt	ggggcacttt	cttttagatt	gtgaggaatg	9120
ttccgcacta	gattgtttaa	aacttcattt	gttgaagga	gagctgtctt	agtgattgag	9180
tcaagggaga	aaggcatcta	gcctcggctc	caaaagggta	gttgctgtct	agagaggctc	9240
ggtggagcct	gcaaaagtcc	agctttcaaa	ggaacacaga	agtatgtgta	tggaatatta	9300
gaagatgttg	cttttactct	taagttggtt	cctaggaaaa	atagttaaat	actgtgactt	9360
taaaatgtga	gagggttttc	aagtactcat	ttttttaa	gtccaaaatt	tttgtcaatc	9420
aatttgaggt	cttgttgtg	tagaactgac	attacttaaa	gtttaaccga	ggaatgggag	9480
tgaggctctc	tcatacccta	ttcagaactg	acttttaaca	ataataaatt	aagtttaaaa	9540
tatttttaaa	tgaattgagc	aatgttgagt	tggagtcaag	atggccgatc	agaaccagaa	9600
cacctgcagc	agctggcagg	aagcagggtca	tgtggcaagg	ctatttgggg	aagggaaaat	9660
aaaaccacta	ggtaaacttg	tagctgtggt	ttgaagaagt	ggttttgaaa	caactctgtcc	9720
agccccacca	aaccgaaagt	ccaggctgag	caaacacca	cctgggtaat	ttgcatttct	9780
aaaataagtt	gaggattcag	ccgaaactgg	agaggtcctc	ttttaactta	ttgagttcaa	9840
ccttttaatt	ttagcttgag	tagttctagt	ttccccaaac	ttaagtttat	cgacttctaa	9900
aatgtattta	gaattcattt	tcaaaattag	gttatgtaag	aaattgaagg	acttttagtgt	9960
ctttaatttc	taatatat	agaaaacttc	ttaaaattac	tctattat	ttccctctga	10020
ttattggtct	ccattcaatt	cttttccaat	accggaagca	tttacagtga	ctttgttcat	10080
gatctttttt	agttggttgt	tttgccttac	tattaagact	tgacattctc	ggtcaaaacg	10140
gcttcacaaa	tctttttcaa	gaccactttc	tgagtattca	ttttaggaga	aatacttttt	10200
ttttaaatga	atgcaattat	ctagacttat	ttcagttgaa	catgctgggt	ggtggttgag	10260
aggacactca	gtcagtcagt	gacgtgaagg	gcttctaagc	cagtccacat	gctctgtgtg	10320
aactccctct	ggcctctgct	attggtgaat	gggccaagg	tctgagacca	ggctgctgct	10380
gggtaggcct	ggactttggg	tctcccacc	agacctggga	atgtatgggt	gtggcttctg	10440
ccaccatcc	acctggctgc	tcattggacca	gccagcctcg	gtggctttga	aggaacaatt	10500
ccacacaaa	actctggacc	tctccgaaac	caggcaccgc	aaatggtaag	ccagaggcag	10560
ccacagctgt	ggctgctgct	cttaaagctt	gtaaaactgtt	tctgcttaag	agggactgag	10620
tcttcagtca	ttgctttagg	gggagaaaga	gacatttgtg	tgtcttttga	gtaccgttgt	10680
ctgggtcact	cacatttaac	tttccctgaa	aaactagtaa	aagaaaaatg	ttgctgttta	10740
accaataatc	atagagctca	tggtattttg	aggaaatctt	agaaaacgtg	tatacaattg	10800
tctggaatta	tttcagttaa	gtgtattagt	tgaggtactg	atgctgtctc	tacttcagtt	10860
atacatgtgg	gtttgaattt	tgaatctatt	ctggctcttc	ttaagcagaa	aatttagata	10920
aaatggatc	ctcagtggtt	tttaatggtg	ggtttaatat	agaaggaatt	taaattggaa	10980
gctaatttag	aatcagtaag	gagggacca	ggctaagaag	gcaatcctgg	gattctggaa	11040

-continued

gaaaagatgt	ttttagtttt	tatagaaaac	actactacat	tcttgatcta	caactcaatg	11100
tggtttaatg	aatttgaagt	tgccagtaaa	tgtacttcct	ggttgttaa	gaatggtatc	11160
aaaggacagt	gcttagatcc	aaggtgagtg	tgagaggaca	ggggctgggg	tatggatacg	11220
cagaaggaag	gccacagctg	tacagaattg	agaaagaata	gagacctgca	gttgaggcca	11280
gcaggtcggc	tggactaact	ctccagccac	agtaatgacc	cagacagaga	aagccagact	11340
cataaagctt	gctgagcaaa	atttagtgaa	caaggttgac	agcctggcta	ggaagctag	11400
ctctagttaa	gcacagtgg	actgagatgt	gtaggcttcc	ctgagccctt	caaaaatgtg	11460
ctaagctgag	atgattactc	tgaggtagcc	aaagctggac	ttgagcagga	acgaggtaga	11520
ctgcaatgag	ctgaattgag	ctaggccggc	taagctaaac	taggctgact	taaccaagat	11580
agccaaatg	gaatgaattg	tcttgatctg	ggctgattgg	agctaaactc	tactggactg	11640
ctctgaactg	agctgtgttg	ggctgtgttg	tgctggggtg	agctgagcta	gcatgagcta	11700
ctctgtggta	gctggggtga	gctgggatga	gctgagctgg	gtgagctgag	ctgggtgagc	11760
tgagctgggt	tagctgagct	ggggtgagct	ggggtaaact	ggggtgagct	gagctgggtg	11820
agctgagctg	ggtgagctga	gctgagctgg	gtgagctggg	gtaagctggg	gtgagctggg	11880
ctgacctgag	gtggttgagc	tgagctgggt	gagcagagct	gaggtgagct	gagctgagct	11940
gggtgagcag	agctggggtg	agctgagctg	agctgggctg	gggtgagctg	agctgagctg	12000
ggtgagctga	gctgggtgag	ttgagctgag	ctgggtaaac	tgagctgagc	tgggtgagca	12060
gagctggggg	tgagctgagc	tggatgagct	gggctgaact	ggggtgagct	ggggtggggt	12120
gagctgggtg	agctgagctg	ggctgagctg	ggctgagctg	gggtgagctg	agctgggatg	12180
agctgggctg	agctgggctg	agctgagctg	ggctgggctg	agctgagctg	ggctgagctg	12240
ggctgggctg	ggctgggctg	ggctgggctg	agctgggctg	agctgggctg	aaactggggtg	12300
agctggggtg	gggtgagctg	ggtgagctga	gctgggctga	gctgggctga	gctggggtga	12360
gctgggtgag	ctgagctggg	ctgagctggg	ctgagctggg	ctgagctgag	ctgggctggg	12420
ctaagctgag	ctggactgag	ctgggctgag	ctgggctgag	ctgagctggg	ctgggctggg	12480
ctgggctgag	ctgggctgag	ctgggctgag	ctgagctggg	atgggctggg	atgagctgaa	12540
ctgggtgagc	tgagctgggg	tgagatgagc	tgaacagggc	tgaactgggg	tgagctgagc	12600
tgggatgagc	tggggtgagc	tgagctgggt	gagatgagct	gggtgagctg	agctgagctg	12660
ggctgagctg	ggtgagctga	gctgggtgag	atgagctggg	gtgagctgag	ctggggtgag	12720
ctgagctggg	tgagctgggc	tgagctgggc	tgggtgagct	gagctgggtt	gagctgagct	12780
gggctgagct	gggctgagct	gggctgagct	gggctgagct	gggctgagct	gagctgagct	12840
gggatgagca	gggctgggat	gagctgggct	gagctgagct	gggatgagct	gggtgagctg	12900
agctggggtg	agatgagctg	gggtgagctg	agctggggtg	agctgagctg	ggatgagcta	12960
ggtgagctga	gttggggtga	gctgagctgg	ggtgagctgg	gctgggtgaa	ctaaactagg	13020
gtgagctggg	ctgagctgat	ctggttgagc	tgggctgggt	gagctgagct	gggtgaaactg	13080
agctggggtg	agctgaaactg	agctgggtga	gctggggtga	gctgagctga	gctgggtgaa	13140
ctgagctgag	ttgagttggg	tgagctgagc	tgggtgagct	gaaactgagct	ggactgagct	13200
ggggtgagct	ggggtgagct	gggctggatg	agctgagctg	ggtgagatga	gctgaaacag	13260
gctgaaactg	ggtgagctga	gctgggatga	gctggggtga	gctgagctgg	gtgagatgag	13320

-continued

ttggggtgag ctgaactgag ctggactgag ctggggtgag ttgggctagg tgagctgaac 13380
tggggtgagc tgagctgggg tgagatgagc tgaacaggtc tgaactggga tgagctggga 13440
tgagctgggg tgagctgggg tgagctgagt tgggtgagct gagctgggtg agctgagctg 13500
gggtgagctg agctgggatt agctggctga gctgggatga gctgggtgag ctgagttggg 13560
gtgagctgag ctggggtgag ctgggctggg tgaactaac tagggtgagc tgggctgagc 13620
tgatctgggt gagctgagct ggggtgagct gggctgagct gggctgggtg agctgagctg 13680
ggtgaactga gctggggtga gctgaactga gctgggtgag ctggggtgag ctggggtgag 13740
ctgagctgag ctgggtgaac tgagctgagt tgagttgggt gagctgagct ggggtgagctg 13800
aactgagctg gactgagctg ggggtgagctg ggggtgagctg ggctgggtga gctgagctgg 13860
gtgagctgag ctgggctgag ctggggtgag ctgagctgag ctgggctggg tgagctgagc 13920
tgggtgagct gagctggggt gagctgagct gagctgggtg agctgggctg agctgagctg 13980
agctggggtg agctggggtg agctgggctg ggtgagatga gctgtgctgg gctgggtgag 14040
ctgagctggg tgaactgagc tgagctgagt tgggtgagct gagctgggtg agctgagctg 14100
ggctgagctg ggggtgagctg agctgagctg ggggtgagctg ggggtgagctg ggctgggtga 14160
gctgagctgg gtgagatgag ctgtgctggg ctgggtgagc tgagctgggt gaactgagct 14220
gagctgagtt ggggtgagctg agttgagtga gctgagttgg gtgagctgag ctggggtgag 14280
ctgaactgag ctggactgag ctggggtgag ctgagctggc tgagctgagc tgggtgagct 14340
gagctgagct gagctgagct gagctgagct gggctgagct gggctgggtg agctgggctg 14400
agctgggctg ggtgagctgg gctgagctgg gctgggtgag ctaagctggg tgagctggga 14460
tgggtgagct ggctgagct ggggtgagctg agctgaaagg atctgggttg gagtcagctt 14520
tgcttgggtg agctgagcag ggtttagctg agctaggatg agttgagttg gctgggctga 14580
gctgggatga agtaggctaa ataagctagc ttgagctaaa tagagctggg gtgagctgta 14640
cttaaatgag ctgtgattga agtgggctgg gatgagctca gctgaactgg gctagctgat 14700
gtggatgatg gtgaggctga gttaggatga gcatgactga gctgggttgg acatgaatga 14760
gctggaccog gctgacttgg gctggttgg ctttgcaaga tgatgtggac taaactgggc 14820
tagctggcag gatgagagag gttgaactag gctcagatga gctaggtga gcaggctta 14880
gggtagggtg tttgggatga gctgggataa gttgggctca gatgagctac catgaactgg 14940
gctgaggtgg cctgggagaa catgaatcga gttgactga gtacagccag gatgaactgg 15000
gataaactgg gctaaactgg gttgggctaa gaatggactg cctgagttat gctaagctga 15060
gctgggatga attggggatt gtaagactt aaatgagcca gactgaactg gttggacttg 15120
ggtgagctga ggggacctgc tgacgggccc gcatagctga gttgaattta gatgaggaag 15180
ggtgagctag actgagcaag gctagactgc ttatgctgag ctacactaac ctggcctgag 15240
ttgggccagg ctttgttggc catgtctaaa ctgagttaa atcaacaggg agtctagagg 15300
ggtaggggat gaactaagat gaattacact agctgagctg aattgagata aggtatggta 15360
acctgagctg aactggaaag agatgtactg gataacctta actgggctga gatgagctag 15420
gtctaccag gcctggatca gcttaattag ggtaggctag accaaactgt agcagtatgt 15480
attagcctgt gccaaactgg gctacattaa actaaactgg acttagctag gctcagatta 15540
gttctgctac tctagatggg gtaagttggg ccaaactggg atgaactaat ttaactagcc 15600

-continued

tgagatgggc agatctgaat gagcagagct gggatgaact gaatgagttt caccaggcct 15660
ggaccagtta ggctaggacc tcgttctata gaggcagact gtgtgctaca gtggagtttc 15720
aagatgattc catgagtoct ccccgcccc aacataacct accttctcc taccctaaa 15780
gcctgtctgg tgtgtaaate ccagctttgt gtgctgatac agaagcctga gccctcccc 15840
cacctocacc tacctattac tttgggatga gaatagttct ccagccaggt gtctcagagg 15900
gaagccaagc aggacaggcc caaggctact tgagaagcca ggatctagge ctctccctga 15960
gaacgggtgt tcatgcccct agagttggct gaagggccag atccacctac tctagaggca 16020
tctctcccctg tctgtgaagg cttccaaagt cacgttccctg tggctagaag gcagctccat 16080
agccctgctg cagtttctgc ctgtatacca ggttcaceta ctaccatate tagccctgcc 16140
tgccctaaga gtagcaaaa gaaatagca ggggttagag ggatctcctg tctgacagga 16200
ggcaagaaga cagattctta cccctccatt tctcttttat ccctctctgg tcctcagaga 16260
gtcagtcctt cccaaatgct tccccctcg tctctgcca gagccccctg tetgataaga 16320
atctgggtgc catgggtgc ctggcccggg acttctctgc cagcaccatt tccttcacct 16380
ggaactacca gaacaacact gaagtcattc agggatcag aaccttccca aactgagga 16440
cagggggcaa gtacctagcc acctcgagg tgttctctgc tccaagagc atccttgaag 16500
gttcagatga atacctggta tgcaaaatcc actacggagg caaaaacaaa gatctgcatg 16560
tgcccatcc aggtgaagaac caaacctcc cagcaggggt gccaggccc aggcattggcc 16620
cagagggagc agcgggtgg ggettaggcc aagctgagct cacacctga cctttcattc 16680
cagctgtcgc agagatgaac cccaatgtaa atgtgttctg cccaccacgg gatggcttct 16740
ctggcccctgc accacgcaag tctaaactca tctgcgaggc cacgaacttc actccaaaac 16800
cgatcacagt atcctggcta aaggatggga agctcgtgga atctggcttc accacagatc 16860
cggtgaccat cgagaacaaa ggatccacac cccaaacctc caaggtcata agcacactta 16920
ccatctctga aatcgactgg ctgaacctga atgtgtacac ctgcccgtgtg gatcacaggg 16980
gtctcacctt cttgaagaac gtgtctctca catgtgtctc cagtgagttg cctgggataa 17040
gcccattgcc tagccctccc agattagga agtcctccta caattatggc caatgccacc 17100
cagacatggt catttctcc ttgaactttg gctccccaga gtggccaagg acaagaatga 17160
gcaataggca gttaggggtg gagaatcagc tggaaggacc agcatcttc cttaagtagg 17220
tttgggggat ggagactaag cttttttcca acttcacaac tagatatgtc ataacctgac 17280
acagtgttct cttgactgca ggtccctcca cagacatcct aaccttcacc atccccct 17340
cctttgccga catcttctc agcaagtccg ctaacctgac ctgtctggtc tcaaacctgg 17400
caacctatga aacctgaat atctcctggg cttctcaaag tgggtgaacca ctggaaacca 17460
aaattaaat catggaaagc catcccaatg gcacctcag tgctaagggt gtggctagt 17520
tttgtgtgga agactggaat aacaggaagg aatttgtgtg tactgtgact cacagggatc 17580
tgcttcacc acagaagaaa ttcactctca aacctaatgg taggtatccc ccttccctt 17640
ccccccaat tgcaggacc ttcctgtacc tcataggggg ggcaggtcct cttccaccct 17700
atcctcacta ctgtctctat ttacagaggt gcacaaacat ccacctgctg tgtacctgct 17760
gccaccagct cgtgagcaac tgaacctgag ggagtcagcc acagtacct gcctggtgaa 17820
gggcttctct cctgcagaca tcagtgtgca gtggctcag agagggcaac tcttgccca 17880

-continued

```

agagaagtat gtgaccagtg ccccgatgcc agagcctggg gccccagget tctactttac 17940
ccacagcatc ctgactgtga cagaggagga atggaactcc ggagagacct atacctgtgt 18000
tgtaggccac gaggcctgc cacacctggt gaccgagagg accgtggaca agtccactgg 18060
taaaccaca ctgtacaatg tctccctgat catgtctgac acaggcggca cctgctattg 18120
accatgctag cgctcaacca ggcaggocct ggggtgccag ttgctctgtg tatgcaaact 18180
aaccatgtca gagtgagatg ttgcatttta taaaaattag aaataaaaaa aatccattca 18240
aacgtcactg gtttgatta tacaatgctc atgcctgctg agacagtgtg gttttgcttg 18300
ctctgcacac accctgcata cttgcctcca ccttggcctc tcctctacct tgccagtttc 18360
ctcctgtgtg gtgaactcag tcaggcttac aacagacaga gtatgaacat gcgattcctc 18420
cagctacttc tagatatatg gctgaaagct tgcctaacct ggtgcaggca gcattcaggc 18480
acatatatag acacacatgc atttatacat agatatatag gtacacatgt gtagacacat 18540
acatgaatgt gtattcatgg acacacagac aaaggtacac atatatacac atgagttcat 18600
gcgcacacac atgcatggac acttacaac gccttcagag acaaataggc atagacacac 18660
aaccactcac agaaacagat accaatatgc atggctcctg gtacacagaa acagactata 18720
ggcaaatata cacaaataaa ctatatagat acaaagatat gcatatacac acatgtacag 18780
aaacatcttc acatgtgtac actaacatgt ggacaggat agcacacaga tacacctgga 18840
ctctgaccag ggctgtaatc tccaaggctc acggctcaga gagcctacac taggctgggt 18900
cactgatact cctcaggagc ccactctatg attgggagag ataaccocag gtacaaagta 18960
tgcctatctg tctcaacacc atggggcaga agatactcca ctaaccaccc atgacagaaa 19020
gttagccttg gctgtgtctc cattaataga acacctcaga agaccaatgt gaaattgcct 19080
aaccactca caccaccct gatctccagt tcaaaatgca gaaaacataa tgcagttgtc 19140
caaaagatgc cccaaccaca cacacacaca cacacacaca cacacacaca cacacacaca 19200
cacacataca cacacacaca ccatcaagga gcctctgtaa ggagtcca cccaataaca 19260
ctgcctcttt gggctcatat cctggacatt ctctcatatc atatccattt ggggcctagg 19320
ctttagatag cccaagggc tcacttttac agggatcaga gatcccaata aatgccctgg 19380
tcccacagcc tcctcaggt atctgtctgt ttatctcttg gtacctttct tagacgttag 19440
gtggcacttt tcggggaaat gtgcgcgaa cccctatttg tttatttttc taaatacatt 19500
caaatatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa 19560
ggaagagtat gagtatcaa catttccgtg tcgccttat tcctttttt gcggcatttt 19620
gccttctctg ttttgcctac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt 19680
tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt 19740
ttcgccccga agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg 19800
tattatcccg tgttgacgcc gggcaagagc aactcggctg ccgcatacac tattctcaga 19860
atgacttggg tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa 19920
gagaattatg cagtgtctgc ataaccatga gtgataacac tgcggccaac ttacttctga 19980
caacgatcgg aggaccgaag gagctaaccg cttttttgca caacatgggg gatcatgtaa 20040
ctgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca 20100
ccacgatgcc tgcagcaatg gcaacaacgt tgcgcaaac attaactggc gaactactta 20160

```

-continued

```

ctctagcttc ccgcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac 20220
ttctgcgctc ggcccttccg gctggctggt ttattgctga taaatctgga gccggtgagc 20280
gtgggtctcg cggtatcatt gcagcactgg gccagatgg taagccctcc cgtatcgtag 20340
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgtgaga 20400
taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt 20460
agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 20520
atctcatgac caaaatccct taacgtgagt ttctgttcca ctgagcgtca gaccccgtag 20580
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaaatctgc tgcttgcaaa 20640
caaaaaaac accgctacca gcggtggtt gtttgccgga tcaagagcta ccaactcttt 20700
ttccgaaggt aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc 20760
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa 20820
tctgttacc agtggctgct gccagtgccg ataagtcgtg tcttaccggg ttggactcaa 20880
gacgatagtt accggataag gcgcagcggc cgggctgaac ggggggttcg tgcacacagc 20940
ccagcttggg gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa 21000
gcccacgct tcccgaaggg agaaggcggg acaggtatcc ggtaagcggc agggctcgaa 21060
caggagagcg cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg 21120
ggtttcgcca cctctgactt gagcgtogat ttttgtgatg ctcgtcaggg gggcggagcc 21180
tatgaaaaa cgcacgaac gcggcctttt tacggttcct ggcttttgc tggccttttg 21240
ctcacatggt ctttctgcg ttatccctg attctgtgga taaccgtatt accgccttg 21300
agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcagatca gtgagcgagg 21360
aagcggaaga gcgcctgatg cggtatcttc tccttacgca tctgtgcggg atttcacacc 21420
gcatatggtg cactctcagt acaatctgct ctgatgccgc atagttaagc cagtatacac 21480
tccgctatcg ctacgtgact gggtcatggc tgcgccccga caccgccea caccgcctga 21540
cgcgccctga cgggcttgte tgcctccggc atccgcttac agacaagctg tgaccgtctc 21600
cgggagctgc atgtgtcaga ggttttacc gtcacaccg aaacgcgca ggcagctgcg 21660
gtaaagctca tcagcgtggt cgtgaagcga ttcacagatg tctgcctggt catccgcgtc 21720
cagctcgttg agtttctoca gaagcgttaa tgtctggctt ctgataaagc gggccatggt 21780
aagggcgggt ttttctggt tggctactga tgcctccgtg taaggggat ttctgttcat 21840
gggggtaatg ataccgatga aacgagagag gatgctcagc atacgggta ctgatgatga 21900
acatgccccg ttactggaac gttgtgaggg taaacaactg gcggtatgga tgcggcggga 21960
ccagagaaaa atcactcagg gtcaatgccg gcgcttcggt aatacagatg taggtgttcc 22020
acagggtagc cagcagcatc ctgcatgca gatccggaac ataaggtgc agggcctga 22080
cttccgctt tccagacttt acgaaacacg gaaaccgaag accattcatg ttgttctca 22140
ggtcgcagac gttttgcagc agcagtcgct tcacgttcgc tcgctatcg 22190

```

<210> SEQ ID NO 3

<211> LENGTH: 14130

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Targetting vector (short version)

-continued

<400> SEQUENCE: 3

gcgggcgcgcaa cctgggcaaaa tgggagctta gcaacaatgt agggggctgg acctagactt	60
cctacacatg tgtaacagat gtgcagcttg gtcttcctgt gtgtattacc ctaacatttg	120
gagcaggagc tgtctctgac tctgttgect gccattggat ccccttcccc tgcttgggct	180
gccttgtttg gccttagtag gaaaggatgt gcttagtctt gctgtgactt gatgtcccta	240
ggcagaatga taccaccagg gggctcccca tctctgagga gatgggcaaa gggtaatgg	300
tggagggact tgtgaggctg ggactgggag gagagaaagg agacagctgt aactggaatg	360
atgttaagtg aacaaatgaa tggatagatt agatagacag atagacagac agacagacag	420
acagacagac agacagacag acagacagat agaaagatag atagataagg ggaaaaagaa	480
acgtagctga gcaagccaga gagagcaagc caaataagca gcattcctcc atgacttttc	540
cttcagctcc tgcctatgag tctgccttga cttccctcag tgattggttg taagttaaaa	600
gggtaataaa accctttcct tgacaagtgt cttttggttc tgatttttat cacagcaaga	660
gaaaaataaa ctagaacaaa catgtatatt tcttggcaca tgtccatagt aaggcagaaa	720
tgatcttcag acctagacca tagatactac agagagcaga agttagata ggtggactta	780
ctgtatgatt gtaatccaag taaacttaca tagctagaga gctagaggaa aggccaaagc	840
ttcctctggg aggtcagatc ctgtcgcact gtagccaata aggcatttg catcacagga	900
aaggactaag acccaggctg gcaatagtgt ctgtatctta actagatctc tctagtgagt	960
gaggaagtaa atttgtgaga gccccagctg tgggctcgga aggtacctgc catgcccctg	1020
ttagtaactg agtactacag caggagcagg tgttctctag aaagcctgag acaactctac	1080
ttcttctctc aagagaccac ctaatacagg cctgagagaa cagactctgg aaatagatgg	1140
gacttacgga gctaagatct agagctcacc tacagagcag aatcccagcc aagagaacaa	1200
agaatactga ctctctcctg ttccctactc cttaggttct aaaacacact atagggagg	1260
gagcctctag acctccgctc attccccacc ttgctcattc catcttccca tgtcccagc	1320
tctccaagcc acagacacca cctttcctat tcacccacct ttctgtgtcc ctaggteccc	1380
agccatagt cacctcccc cacacccgc tcacctgcc ccatctatgc ccctagatgc	1440
ttacttacca gagtcttttg tctgacgtgg ggctacaagc atctatgctc cctaagcacc	1500
tactgctgac ctgtaggacc cagctctgaa ccaactcata taagtaata cagactctcc	1560
cctgtcttag gatggcccc tgggtcagga ggagaccact gccaaagaa cttctcttag	1620
agcactgaac tcctcccctg taccacttag gacagaccct agacctatta ttactgatta	1680
ccagagctct ggcagtgacc acggaggaga tagatccacc ctggacacag gaaacacagc	1740
accagagata ctgcttcacc acaacagtag agtgacactt tagactttaa tttgggtcac	1800
tttctgctg tagaggtggg atcagaaagc aaagagcagt atgagtgcct gataggcacc	1860
caagtacact atagagtact catgggtgaat aaggtacctc catggcttcc cagggagggg	1920
cactgcccc cccccaccat cacagacctt tctccatagt tgataactca gacacaagtg	1980
aatgacagat ggacctccat ctgctcttat tttaaaaaga agacaaaacc cacaggctcg	2040
agaactttag cgactgtttt gagagaaatc attggctcct gactcaagag atgactggca	2100
gattggggat cagaataccc atactctgtg gctagtgtga ggtttaagcc tcagagtccc	2160
tgtggtctct gactggtgca aggttttgac taagcggagc accacagtgc taactgggac	2220

-continued

cacggtgaca	cgtggctcaa	caaaaacctt	ctgtttggag	ctctccaggg	gcagcctgag	2280
ctatgaggaa	gtagagaggc	ttgagaaatc	tgaggaagaa	aagagtagat	ctgagaggaa	2340
aggtagcttt	ctggaggcca	ggagacagtg	cagagaagaa	cgagttactg	tggacaggtc	2400
ttagatgggg	aaagaatgag	caaatgcaag	catcagaagg	gtggatgcaa	tgtcctgcca	2460
aggacttacc	aagaggatcc	ccggacagag	caggcagggtg	gagttgactg	agaggacagg	2520
ataggtgcag	gtccctctct	tgtttccttt	ctccttctcc	tgtttccttc	ttctcttgtc	2580
acaggtctca	ctatgctagc	caaggctagc	ctgaaagatt	accatcctac	agatgggccc	2640
atccagttga	attaagggtg	agatctctcc	aaacatctga	gtttctgagg	cttggatgcc	2700
actggggacg	ccaagggact	ttgggatggg	tttggttggc	cccagatgaa	gggctacttc	2760
actgggtcta	taattactct	gatgtctagg	accagggggc	tcaggtcact	caggtcagggt	2820
gagtcctgca	tctggggact	gtggggttca	ggtggcctaa	ggcaggatgt	ggagagagtt	2880
ttagtatagg	aacagaggca	gaacagagac	tgtgctactg	gtacttcgat	gtctggggca	2940
cagggaccac	ggtcaccgtc	tcctcaggta	agctggcttt	tttctttctg	cacattccat	3000
tctgaaacgg	gaaaagatat	tctcagatct	ccccatgcca	ggccatctgc	cacactctgc	3060
atgctgcaga	agcttttctg	taaggatagg	gtcttcactc	ccaggaaaag	aggcagtcag	3120
aggctagctg	cctgtggaac	agtgacaatc	atggaaaata	ggcatttaca	ttgttaggct	3180
acatgggtag	atgggttttt	gtacacccac	taaaggggtc	tatgatagtg	tgactacttt	3240
gactactggg	gccaaaggc	cactctcaca	gtctcctcag	gtgagtcctt	acaacctctc	3300
tcttctattc	agcttaaata	gattttactg	catttggtgg	gggggaaatg	tgtgatctg	3360
aatttcagggt	catgaaggac	tagggacacc	ttgggagcca	gaaagggcca	ttgggagccc	3420
tggtgcacgc	agacagacat	cctcagctcc	catacttcat	ggccagagat	ttatagggat	3480
cctggccagc	attgccgcta	ggtccctctc	ttctatgctt	tctttgtccc	tcactggcct	3540
ccatctgaga	tcactcctgga	gcctagcca	aggatcattt	attgtcaggg	gtctaactcat	3600
tgttgtcaca	atgtgctctg	tttcttact	ggggccaagg	gactctggtc	actgtctctg	3660
caggtgagtc	ctaacttctc	ccattctaaa	tgcatgttgg	ggggattctg	ggccttcagg	3720
accaagattc	tctgcaaacg	ggaatcaaga	ttcaacctct	ttgtcccaaa	gttgagacat	3780
gggtctgggt	cagggactct	ctgcctgctg	gtctgtgggtg	acattagaac	tgaagtatga	3840
tgaaggatct	gccagaactg	aagcttgaag	tctgaggcag	aatcttgtcc	agggtctatc	3900
ggactcctgt	gagaattagg	ggctgacagt	tgatggtgac	aatttcaggg	tcagtgactg	3960
tctggtttct	ctgagggtgag	gctggaatat	aggtcacctt	gaagactaaa	gaggggtcca	4020
ggggcttctg	cacaggcagg	gaacagaatg	tggaaacaatg	acttgaatgg	ttgattcttg	4080
tgtgacacca	ggaattggca	taatgtctga	gttgcccagg	ggtgattcta	gtcagactct	4140
ggggtttttg	tcgggtatag	aggaaaaatc	cactattgtg	attactatgc	tatggactac	4200
tgggtcaag	gaacctcagt	caccgtctcc	tcaggtaaga	atggcctctc	caggtcttta	4260
tttttaacct	ttgttatgga	gttttctgag	cattgcagac	taatcttggg	tatttgtccc	4320
tgagggagcc	ggctgagaga	agtggggaaa	taaactgtct	agggatctca	gagcctttag	4380
gacagattat	ctccacatct	ttgaaaaact	aagaatctgt	gtgatgggtg	tgggtggagtc	4440
cctggatgat	gggatagggg	ctttggaggc	tcatttgaag	aagatgctaa	aacaatccta	4500

-continued

tggctggagg gatagttggg gctacgcggt tttaacccta gaaagatagt ctgcgtaaaa	4560
ttgacgcgatg cattottgaa atattgctct ctctttctaa atagcgcgaa tccgctgctg	4620
tgcatttagg acatctcagt cgcgccttg agctcccgtg aggcgtgctt gtcaatgcgg	4680
taagtgtcac tgattttgaa ctataacgac cgcgtgagtc aaaatgacgc atgattatct	4740
tttacgtgac ttttaagatt taactcatac gataattata ttgttatctt atgttctact	4800
tacgtgataa cttattatat atatatcttc ttgttataga tategctagt ggatcctggt	4860
tctttccgoc tcagaaggta cttttttttt tttttttttt tttttttttt tttttttttt	4920
tttttttttt tttttttttt taaatttttg ggaattttat gatttgcatt taaaagggaa	4980
ctgctgacaa agattcactg gtaataattt gaacaagttg gaaaatacag tcaacattac	5040
tgaaacacta ctaaaataat tccaggacag aacaaaactt cttagatgct gtctttgatg	5100
tgaaaattga ctgcttctta cttttctaac acacgggtgt ataattaaca atattcaatc	5160
acttctatct tttcctgcat atataaaaa taaaatacca attaaaaaac taatatatct	5220
tctctttatt tcttacagat atgagttcaa tgtttcactc aatagtgctg tggtttaaga	5280
gaatttttct atttacaagt taaacaacaa tccgcccaaa gggaaactgat agtctatag	5340
ctcatagtgc aaataaacag tttaggaatg cagcaactga catttctaaa gtacaaaaca	5400
gataaaattc ttagaagata catgcaaaaa gctctactaa gcagatggcc acagaactag	5460
aacattgata attttactgg cgatgtcaat aggactccag atgtttccaa actcaacttg	5520
aactctcatc ttaggctttg tattttgctt ttccagtttc actaatgaca caaacatgat	5580
tcaaatccct gaagtattca ttatagtcaa gggcatatcc tacaacaaac ttgtctggaa	5640
tttcaaatcc aacaaagtct ggcttatatc caacacttcg tggggctcct ttcaccagca	5700
agcttgcgac cttgaccatc tttggattat actgctgac caaggaaagc aaagtctgca	5760
ttgttttgcc agtgtcaatt atactttcca caatcaagac attctttcca gttaaagtgt	5820
agagatcatc tccaccaatt acttttatgt cccctgttga ctggtcatta caatagctct	5880
tcagtctgat aaaatctaca gtcataggaa tggatctatc actatttcta ttcagtgtt	5940
tgatgtaatc cagcagggtca gcaaagaatt tatagcccc cttgagcaca cagagggcta	6000
caatgtgatg gcctcccac ccttccatca catctcgagc aagacgttca gtcctacaga	6060
aataaaatca ggaatttaat agaaagtctc atacattaaa ctttataaca aacacctctt	6120
agtcattaaa cttccacacc aacctgggca atatagttag accccatgcc tgcaaaaaaa	6180
aaaaaattag ccaggcatgg tagcatgtac ctgtagtccc agctacttga gaggtgaggt	6240
gggaaaaatca ctttagtgca ggatgttgag gctggagtga actgtgattg tgccactgca	6300
ctccagcctg gacaatagag caagacctg tctcaaaaa atgcattaaa aattttttt	6360
aatcttcca cgtaacacat cctttgccct catgtttcat aaggtaaaaa atttgatacc	6420
ttcaaaaaaa ccaagcatac cactatcata atttttttta aatgcaaata aaaacaagat	6480
accattttca cctatcagac tggcagggtc tgattaaatg aaatttcttg gataatatac	6540
aatattaaga gagactgtag aaactgggcc agtggctcat gcctgtaatc ccagcacttt	6600
gggaggctgg gtaacatggc gaacctgtt tctacaaaat aaaaatatta gctgggagtg	6660
gtggcgcaca cctatagtcc cagctactca ggaggctgag gtggaaggat cgcttgaacc	6720
caggaggttg agactgcagt gaactgtgat cattctgctg cactgcaccc cagcctgggc	6780

-continued

aacagagacc	ttgtctcaaa	aaaaaaaaa	aaagagacaa	attgtgaaga	gaaaggtact	6840
ctcatataac	atcaggagta	taaaatgatt	caacttctta	gaggaaaatt	tggcaatacc	6900
aaaatattca	ataaactott	tccccttgac	ccagaaattc	cacttgaata	aagctgaaca	6960
agtaccaaac	atgtaaaaga	atgtttcttc	tagtacagtc	ggtaagaaca	aaatagtgtc	7020
tatcaatagt	ggactgggta	aatcagttat	ggtatctcca	taagacagaa	tgctatgcaa	7080
cctttaaaat	atattagata	gctctagaca	gtggatcccc	tcgagggacc	taataaacttc	7140
gtatagcata	cattatacga	agttatatta	agggttattg	aatatgtcga	ctagacacac	7200
taatattaaa	agtgctcaat	aacattttaa	actatactca	tacgttaaaa	tataaatgta	7260
tatatgtact	tttgcataata	gtatacatgc	atagccagtg	cttgagaaga	aatgtgtaca	7320
gaaggctgaa	aggagagaac	tttagtcttc	ttgtttatgg	cctccatagt	tagaatattt	7380
tataacacaa	atattttgat	attataattt	taaaataaaa	acacagaata	gccagacata	7440
caatgcaagc	attcaatacc	aggtaaagtt	tttcaactgta	attgacttaa	cagaaaattt	7500
tcaagctaga	tgtgcataat	aataaaaatc	tgaccttgcc	ttcatgtgat	tcagccccag	7560
tccattaccc	tgtttaggac	tgagaaatgc	aagactctgg	ctagagtcc	ttcttccatc	7620
tcccctcaat	gtttactttg	ttctggctcc	tacagagtcc	cactatacca	caactgatac	7680
taagtaatta	gtaaggccct	cctcttttat	ttttaataaa	gaagatttta	gaaagcatca	7740
gttatttaat	aagttggcct	agtttatgtt	caaatagcaa	gtactcagaa	cagctgctga	7800
tgtttgaat	taacacaaga	aaaagtaaaa	aacctcattt	taagatctta	cttacctgtc	7860
cataattagt	ccatggggaa	taaacaccct	ttccaaatcc	tcagcataat	gattaggtat	7920
gcaaaaataa	tcaaggcoat	aacctgggtc	atcatcacta	atcacgacgc	cagggctgcg	7980
ggctgccata	acggagccgg	ccggcgcgcg	ggctgaataa	cttcgtataa	tgtgtactat	8040
acgaagttat	ttgttcagga	ggaggaagcc	ggtggcggag	cagaggagga	ggcggaggcg	8100
cagcaagacc	ccccccccc	tcaggtcga	aaggcccga	gatgaggaag	aggagaacag	8160
cgcggcagac	gtgcgctttt	gaagcgtgca	gaatgccggg	cctccggagg	accttcgggc	8220
gcccccccg	cccctgagcc	cgcccctgag	cccgccccg	gaccoacccc	ttcccagcct	8280
ctgagcccag	aaagcgaag	agccaaaagct	gctattggcc	gctgccccea	aggcctaccc	8340
gcttccattg	ctcagcggtg	ctgtccatct	gcacgagact	agtgagacgt	gctacttcca	8400
tttgtcacgt	cctgcacgac	gagagctgcg	ggcgggggg	gaacttcctg	actaggggag	8460
gagtagaagg	tggcgcgaag	gggccaccaa	agaacggagc	cggttggcgc	ctaccggtgg	8520
atgtggaatg	tgtcggaggc	cagaggccac	ttgtgtagcg	ccaagtgcc	agcggggctg	8580
ctaaagcgca	tgctccagac	tccttggga	aaagcgcctc	ccctaccggg	tagatatcta	8640
taacaagaaa	atatatatat	aataagttat	cacgtaagta	gaacatgaaa	taacaatata	8700
attatcgtat	gagttaaatc	ttaaaagtca	cgtaaaagat	aatcatgcgt	cattttgact	8760
cacgcggtcg	ttatagttca	aaatcagtga	cacttaccgc	attgacaagc	acgcctcacg	8820
ggagctccaa	gcggcgactg	agatgtccta	aatgcacagc	gacggattcg	cgctatttag	8880
aaagagagag	caatatttca	agaatgcotg	cgtcaatttt	acgcagacta	tctttctagg	8940
gttaaaagaa	ttcgtagttg	gagattttca	gtttttagaa	taaaagtatt	agctgcggaa	9000
tatacttcag	gaccacctct	gtgacagcat	ttatacagta	tccgatgcac	agggacaaaag	9060

-continued

agtgagtg	ggcactttct	ttagatttgt	gaggaatggt	ccacactaga	ttgtttaaaa	9120
cttcatttgt	tggaaaggaga	gctgtcttag	tgattgagtc	aagggagaaa	ggcatctagc	9180
ctcgtctca	aaagggtagt	tgtgtcttag	agaggtctgg	tggagcctgc	aaaagtcag	9240
ctttcaaagg	aacacagaag	tatgtgtatg	gaatattaga	agatgttct	ttactctta	9300
agttggttcc	taggaaaaat	agttaaatac	tgtgacttta	aaatgtgaga	gggttttcaa	9360
gtactcattt	ttttaaagt	ccaaaatttt	tgtcaatcaa	tttgaggct	tgtttgta	9420
gaactgacat	tacttaaagt	ttaaccgagg	aatgggagtg	aggctctctc	ataccctatt	9480
cagaactgac	ttttaacaat	aataaattaa	gtttaaaata	ttttaaatg	aattgagcaa	9540
tgttgagttg	gagtcaagat	ggccgatcag	aaccagaaca	cctgcagcag	ctggcaggaa	9600
gcaggtcatg	tggcaaggct	atttgggaa	gggaaaataa	aaccactagg	taaacttgta	9660
gctgtggttt	gaagaagtgg	ttttgaaaca	ctctgtccag	ccccaccaa	ccgaaagtcc	9720
aggtgagca	aaacaccacc	tgggtaattt	gcatttctaa	aataagtga	ggattcagcc	9780
gaaactggag	aggtcctctt	ttaacttatt	gagttcaacc	ttttaatttt	agcttgagta	9840
gttctagttt	ccccaaactt	aagtttatcg	acttctaaaa	tgtatttaga	attcattttc	9900
aaaattaggt	tatgtaagaa	attgaaggac	tttagtgtct	ttaatttcta	atatatttag	9960
aaaacttctt	aaaattactc	tattattctt	ccctctgatt	attggtctcc	attcaattct	10020
tttccaatc	ccgaagcatt	tacagtgact	ttgttcatga	tcttttttag	ttgtttgttt	10080
tgcttacta	ttaagacttt	gacattctgg	tcaaaaacggc	ttcacaatc	tttttcaaga	10140
ccactttctg	agtattcatt	ttaggagaaa	tacttttttt	ttaaataaat	gcaattatct	10200
agacttattt	cgggtgaaca	tgtctggttg	tgggtgagag	gacactcagt	cagtcagtg	10260
cgtgaagggc	ttctaagcca	gtccacatgc	tctgtgtgaa	ctccctctgg	ccctgcttat	10320
tgttgaatgg	gccaaaggtc	tgagaccagg	ctgctgctgg	gtaggcctgg	actttgggtc	10380
tcccaccag	acctgggaat	gtatggttgt	ggcttctgcc	acccatccac	ctggctgctc	10440
atggaccagc	cagcctcggg	ggctttgaag	gaacaattcc	acacaaagac	tctggacctc	10500
tccgaaacca	ggcaccgcaa	atggaagcc	agaggcagcc	acagctgtgg	ctgctgctct	10560
taaagcttgt	aaactgtttc	tgttaagag	ggactgagtc	ttcagtcatt	gctttagggg	10620
gagaaagaga	catttgtgtg	tcttttgagt	accgttgtct	gggtcactca	catttaactt	10680
tccttgaaaa	actagtaaaa	gaaaaatggt	gcctgttaac	caataatcat	agagctcatg	10740
gtattttgag	gaaatcttag	aaaacgtgta	tacaattgtc	tggaattatt	tcagttaagt	10800
gtattagtgt	aggtactgat	gctgtctcta	cttcagttat	acatgtgggt	ttgaattttg	10860
aatctattct	ggctcttctt	aagcagaaaa	tttagataaa	atggatacct	cagtggtttt	10920
taatgggtggg	tttaatatag	aaggaattta	aattggaagc	taatttagaa	tcagtaagga	10980
gggaccag	ctaagaaggc	aatcctggga	ttctggaaga	aaagatgttt	ttagttttta	11040
tagaaaaac	tactacattc	ttgatctaca	actcaatgtg	gtttaatgaa	tttgaagttg	11100
ccagtaaatg	tacttctctg	ttgttaaaga	atggtatcaa	aggacagtgc	ttagatccaa	11160
ggtgagtggt	agaggacagg	ggctggggta	tggatacga	gaaggaaggc	cacagctgta	11220
cagaattgag	aaagaataga	gacctgcagt	tgaggccagc	aggtcggctg	gactaactct	11280
ccagccacag	taatgacca	gacagagaag	gccagactca	taaagcttta	tcgataaccg	11340

-continued

cgacctcgag ggggggcccg gtacctttct tagacgtcag gtggcacttt tgggggaaat 11400
gtgcgcgga cccctatttg tttatttttc taaatacatt caaatatgta tccgctcatg 11460
agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa 11520
catttccgtg tgcaccttat tccctttttt gcggcatttt gccttctgt ttttgetcac 11580
ccagaaacgc tgggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac 11640
atcgaaactg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt 11700
ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccc tgttgaocgc 11760
gggcaagagc aactcggctc cgcatacac tattctcaga atgacttggg tgagtactca 11820
ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgtctgc 11880
ataacctga gtgataaac tgcggccaac ttacttctga caacgatcgg aggaccgaag 11940
gagctaaccg cttttttgca caacatgggg gatcatgtaa ctgccttga tggttgggaa 12000
ccggagctga atgaagccat accaaaacgc gagcgtgaca ccacgatgac tgcagcaatg 12060
gcaacaacgt tgcgcaaac attaactggc gaactactta ctctagcttc ccggcaaaa 12120
ttaatagact ggatggaggc ggataaagt gcaaggaccac ttctgcgctc ggcccttccg 12180
gctggctggg ttattgtgta taaatctgga gccggtgagc gtgggtctcg cggtatcatt 12240
gcagcaactg gccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 12300
caggcaacta tggatgaacg aaatagacag atcgtgaga taggtgcctc actgattaag 12360
cattggtaac tgtcagaaca agtttactca tatatacttt agattgattt aaaaactcat 12420
ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaatccct 12480
taacgtgagt tttcgttoca ctgagcgtca gaccccgtag aaaagatcaa aggatcttct 12540
tgagatcctt tttttctgcg cgtaatctgc tgcctgcaaa caaaaaaac accgctacca 12600
gcgggtggtt gtttgcgga tcaagagcta ccaactcttt ttccgaaggc aactggcttc 12660
agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg ccaccacttc 12720
aagaactctg tagcaccgac tacatactc gctctgctaa tcctgttacc agtggctgct 12780
gccagtggcg ataagctgtg tcttacggg ttggactcaa gacgatagtt accggataag 12840
gcgcagcggc cgggctgaac ggggggttcg tgcacacagc ccagccttga gcgaacgacc 12900
tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgcccagct tcccgaaggg 12960
agaaaggcgg acaggtatcc ggtaaagcgc agggtcggaa caggagagcg cacgaggag 13020
cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca cctctgactt 13080
gagcgtcgat ttttgtgatg ctgcctcaggg gggcggagcc tatggaaaaa cgcagcaac 13140
gcggcctttt tacggttcct ggccttttgc tggccttttg ctacatggt ctttctgctg 13200
ttatccccctg attctgtgga taaccgtatt accgcctttg agtgagctga tacgctcgc 13260
cgcagccgaa cgaccgagc cagcaggtca gtgagcaggg aagcgggaaga gcgcctgatg 13320
cggatatttc tccttacgca tctgtcgggt atttcacacc gcatatgggt cactctcagt 13380
acaactctgt ctgatccgc atagttaagc cagtatacac tccgctatcg ctacgtgact 13440
gggtcatggc tgcgccccga cccccgcaa caccgctga cgcgccctga cgggcttgtc 13500
tgctccccgc atccgcttac agacaagctg tgaccgtctc cgggagctgc atgtgtcaga 13560
ggttttcacc gtcatacccg aaacgcgca ggcagctgag gtaaagctca tcagcgtggg 13620

-continued

```

cgtgaagcga ttcacagatg tctgctgtt catccgcgtc cagctcgttg agtttctcca 13680
gaagcgtaa tgtctggctt ctgataaac gggccatggt aaggcggtt ttttctgtt 13740
tggctactga tgctccgtg taaggggat tctgttcat ggggtaatg ataccgatga 13800
aacgagagag gatgctcacg ataccgggta ctgatgatga acatgcccg ttaactggaac 13860
gttgtgaggg taaacaactg gcggtatgga tgcggcggga ccagagaaaa atcaactcag 13920
gtcaatgccg gcgcttctgt aatacagatg taggtgttcc acagggtagc cagcagcatc 13980
ctgcgatgca gatccggaac ataatggtgc agggcgtgta cttccgcgtt tccagacttt 14040
acgaaacacg gaaaccgaag accattcatg ttgttctca ggctgcagac gttttgcagc 14100
agcagtcgct tcacgttgc tcgctatcg 14130

```

```

<210> SEQ ID NO 4
<211> LENGTH: 3551
<212> TYPE: DNA
<213> ORGANISM: Mus musculus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1399)..(1498)
<223> OTHER INFORMATION: n is a, c, g, or t

```

<400> SEQUENCE: 4

```

aagcttgctg agcaaaatta agggaacaag gttgagagcc ctagtaagcg aggctctaaa 60
aagcatggct gagctgagat ggtggtgctt ctctgagcgc ttctaaaatg cgctaaactg 120
aggtgattac tctgaggtaa gcaaaactgg gcttgagcca aaatgaagta gactgtaatg 180
aactggaatg agctgggccc ctaagctaaa ctaggctggc ttaaccgaga tgagccaaac 240
tggaatgaac ttcattaatc taggttgaat agagctaaac tctactgcct aactgggact 300
gttctgagct gagatgagct ggggtgagct cagctatgct acgctgtgtt ggggtgagct 360
gatctgaaat gagctactct ggagtagctg agatggggtg agatggggtg agctgagctg 420
ggctgagctg gactgagctg agctaggggtg agctgagctg ggtgagctga gctaaactgg 480
ggtgagctga gctgagcttg actgagctag ggtgagctgg actgagctgg ggtgagctga 540
gctgagctgg ggtgagctgg gatgagctgg ggtgagctga gctgagctgg agtgagctga 600
gctgggctga gctggggtga gctgggctgg gctgagctgg ggtgagctgg gctgagctgg 660
ggtgagctga gctggggtga gctgagctga gctggggtga gctgagctga gctggggtga 720
gctgagctgg ggtgagctga gctgagctgg gctgagctga ggtgagctga gctggggtga 780
gctgagctgg ggtgagctga gctgagctgg ggtgagctgg gatgagctgg ggtgagctga 840
gctgagctgg agtgagctga gctgggctga gctgggctga gctggggtga gctgagctgg 900
ggtgagctga gctgagctgg gctgagctga ggtgagctga gctggggtga gctgagctga 960
gctggggtga gctgagctga gctggggtga gctgagctgg ggtgagctga gctggggtga 1020
gctgagctga gctggggtga gctgagctgg ggtgagctga gctgagctgg ggtgagctga 1080
gctgagctgg ggtgagctga gctgagctga gctggggtga gctgagctga gctggggtga 1140
gctgagctga gctggggtga gctgagctgg ggtgagctgg gctgagctga gctggggtga 1200
gctgagctga gctgagctga gctggggtga gctgagctgg gctgagctgg ggtgagctgg 1260
gctgagctgg ggtgagctga gctggggtga gctgagctga gctggggtga gctgagctga 1320
gctggggtga gctgagctgg ggtgagctga gctgagctgg gctgagctga gctgagctgg 1380

```

-continued

ggtgagctga gctgagctnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	1440
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnag	1500
ctgagctgag ctgagctgag ctgagctggg gtgagctggg gtgagctgag ctggggtgag	1560
ctgagctgag ctggggtgag ctgagctgag ctgagctgag ctgagctgag ctggggtgagc	1620
tgagctgagc tgagctgggg tgagctgagc tgggggtgagc tgagctgagc tgggggtgagc	1680
tgagctgggg tgagctgagc tgagctgggg tgagctgggg tgagctgggg tgagctgggg	1740
tgagctgagc tgaactgggg tgagctgggc tgagctgggg tgagctgagc tgagctgggc	1800
tgagctgggg tgagctgggg tgagctgggg tgagctgagc tgagctaggg tgagctgagc	1860
tgagctaggg tgagctgagc tgagctgggg tgagctgagc tgagctgggg tgagctgagc	1920
tgagctgggg tgagctgagc tgagctgggg tgagctgagc tgagctgggg tgagcttggc	1980
tgagctgggg tgagctgggg tgagctgagc tgggggtgagc tggggtgagc tgagctgagc	2040
tgggggtgagc tgagctgagc tgggggtgagc tgggggtgagc tgagctgagc tgagctgggt	2100
gatctgagct gagctgagct ggggtgagctg agctgagctg agctgggtga gctgagctga	2160
gctgagctga gctgggtgag ctgagctgag ctgagctgag ctgagctgag ctggggtgag	2220
ctgggctgag ctgagctgag ctggggtgag ctgagctgag ctgagctgag ctggggtgag	2280
ctgggctgag ctggggtgag ctgggctgag ctgagctggg tgagctgagc tgaactgagc	2340
tgagctgggt gagctgagct gagctgagct ggggtgagctg agctgggctg agctgagctg	2400
gggtgagctga gctgaactga gctgagctgg gtgagctgag ctgagctgag ctgggtgagc	2460
tgagctgggg tgagctgagc tgagctgggg tgagctgagc tgagctgagc tgggtgagct	2520
gagctgggggt gagctgagct gagctgggggt gagctgagct gagctgggggt gagctgagct	2580
gagctgggggt gagctgagct gagctgggggt gagctgagct aggggtgaact gggctgggtg	2640
agctggagtg agctgagctg aggtgaactg ggggtgagccg ggatggtttg agttgagctg	2700
gggtaagatg agctgaactg gggtaagatg ggatgagctg tgggtgagggg agctggattg	2760
aactgagctg tgtgagctga gctgggggtca gctgagcaag agtgagtaga gctggctggc	2820
cagaaccaga atcaattagg ctaagtgagc cagattgccc tgggatcagc tgtactcaga	2880
tgagctggga tgaggtaggc tgggatgagc tgggctagct gacatggatt atgtgaggct	2940
gagctagcat gggctggcct agctgatgag ctaagcttga atgaacgggg ctgagctgga	3000
ctcagatgtg ctgagctgag ctgtactgga tgatctggtg taggggtgac tggactcaac	3060
tgggctggct gatgggatgc cccaggttga actaggctca gataagttag gctgagtagg	3120
gcctgggttga gatgggtcgg gatgagctgg gaaaagatgg actgggacca tgaactgggc	3180
tgagctgggt tgggagacca tgaattgagc tgaactgagt gcagctggga taaactgggt	3240
tgagctaaga atagactacc tgaattgtgc caaactgggc tgggatcaat tggaaattat	3300
caggatattag atgagccgga ctaaaactatg ctgagctgga ctggttggat gtgttgaact	3360
ggcctgctgc tgggctggca tagctgagtt gaacttaaat gaggaaggat gagcaaggct	3420
agcctgcttg catagagctg aactttagcc tagcctgagc tggaccagcc tgagctgagt	3480
aggtctaaac tgagttaaaa atcaacaggg ataatttaac agctaattta acaagcctga	3540
ggtctgagat t	3551

-continued

<211> LENGTH: 2484

<212> TYPE: DNA

<213> ORGANISM: Mus musculus

<400> SEQUENCE: 5

aagcttgctg agcaaaatta agggaacaag gttgagagcc ctagtaagcg aggctctaaa 60
aagcacagct gagctgagat ggggtggctt ctctgagtgc ttctaaaatg cgctaaactg 120
agggtattac tctgaggtaa gcaaagctgg gcttgagcca aaatgaagta gactgtaatg 180
aactggaatg agctgggocg ctaagctaaa ctaggctggc ttaaccgaga tgagccaaac 240
tggaatgaac ttcattaatc taggttgaat agagctaaac tctactgcct aactgggact 300
gttctgagct gagatgagct ggggtgagct cagctatgct acgctgtgtt ggggtgagct 360
gatctgaaat gagatactct ggagtagctg agatggggtg agatggggtg agctgagctg 420
ggctgagcta gactgagctg agctaggggtg agctgagctg ggtgagctga gctaagctgg 480
ggtgagctga gctgagcttg gctgagctag ggtgagctgg gctgagctgg ggtgagctga 540
gctgagctgg ggtgagctgg gatgagctgg ggtgagctga gctgagctgg agtgagctga 600
gctgggctga gctggggtga gctgggctga gctgggctga gctgggctga gctggggtga 660
gctgagctgg ggtgagctga gctgagctgg ggtgagctga gctgagctgg ggtgagctgg 720
ggtgagctga gctggggtga gctgagctga gctggggtga gctgagctgg ggtgagctga 780
gctgagctgg ggtgagctga gctgagctga gctgagctga gctggggtga gctgagctga 840
gctgagctgg ggtgagctgg ggtgagctga gctgagctgg agtgagctga gctgggctga 900
gctggggtga gctgggctga gctggggtga gctgagctga gctgagctga gctggggtga 960
gctgagctga gctggggtga gctgagctgg ggtgagctgg gctgagctga gctgagctga 1020
gctgagctga gctgagctga gctgagctga gctgagctga gctgagctga gctgagctga 1080
gctgagctgg ggtgagctga gctgagctgg gctgagctgg ggtgagctgg gctgagctgg 1140
gctgagctgg gctgagctgg ggtgagctga gctggggtga gctgagctga gctgggctga 1200
gctgagctga gctggggtga gctgagctga gctggggtga gctgagctga gctgagctgg 1260
ggtgagctga gctgggctga gcaggctga gctggggtga gctgagctga gctggggtga 1320
gctgggctga gctgggctga gctgagctga gctgggctga gctgggctga gctgggctga 1380
gctgggctga gctgggctga gctggggtga gctgagctga gctggggtga gctggggtga 1440
gctgagctgg ggtgagctga gctggggtga gctgagctga gctggggtga gctgagctgg 1500
ggtgagctga gctgagctgg ggtgagctga gctgagctgg ggtgagctga gctagggtga 1560
actgggctgg gtgagctgga gtgagctgag ctgaggtgaa ctggggtgag ccgggatggt 1620
ttgagttgag ctggggtgaa atgagctgaa ctggggtaaa ctgggatgag ctgtggtgag 1680
cggagctgga ttgaaactgag ctgtgtgagc tgagctgggg tcagctgagc aagagtgagt 1740
agagctggct ggccagaacc agaatcaatt aggctaagtg agccagattg tgctgggatc 1800
agctgtactc agatgagctg ggatgaggta ggctgggatg agctgggcta gctgacatgg 1860
attatgtgag gctgagctag catgggctgg cctagctgat gagctaagct tgaatgagcg 1920
gggctgagct ggactcagat gtgctagact gagctgtact ggatgatctg gtgtagggtg 1980
atctggactc aactgggctg gctgatggga tgcgccaggt tgaactaggc tcagataagt 2040
taggctgagt agggcctggt tgagatggtt cgggatgagc tgggaaaaga tggactcggg 2100

-continued

ccatgaactg ggctgagctg ggttgggaga ccatgaattg agctgaactg agtgcagctg	2160
ggataaactg ggttgagcta agaatagact acctgaattg tgccaaactc ggctgggatc	2220
aattggaaat taccagatt tagatgagcc ggactaaact atgctgagct ggactggtt	2280
gatgtgttga actggcctgc tctgggctg gcatagctga gttgaactta aatgaggaag	2340
gctgagcaag gctagcctgc ttgcatagag ctgaacttta gcctagcctg agctggacca	2400
gctgagctg agtaggtcta aactgagta aaaatcaaca gggataattt aacagctaat	2460
ttaacaagcc tgaggtctga gatt	2484

<210> SEQ ID NO 6

<211> LENGTH: 4515

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 5' homology arm of targeting vector

<400> SEQUENCE: 6

aacctgggca aatgggagct tagcaacaat gtagggggct ggacctagac ttcctacaca	60
tgtgtaacag atgtgcagct tggctctcat gtgtgtatta ccctaacatt tggagcagga	120
gctgtctctg actctgttgc ctgccattgg atccccctcc cctgcttggg ctgcttgtt	180
tggccttagt aggaaaggat gtgcttagtc ctgctgtgac ttgatgtccc taggcagaat	240
gataccccag gggggctccc catctctgag gagatgggca aagggtaatg gttggagggg	300
cttgtgaggc tgggactggg aggagagaaa ggagacagct gtaactggaa tgatgttaag	360
tgaacaaatg aatggataga ttagatagac agatagacag acagacagac agacagacag	420
acagacagac agacagacag atagaaagat agatagataa ggggaaaaag aaacgtagct	480
gagcaagcca gagagagcaa gccaaataag cagcattcct ccatgacttt tccttcagct	540
cctgcctatg agtctgcctt gacttccctc agtgattggt tgtaagttaa aagggtgaaat	600
aaacccttcc tttgacaagt tcttttgggt tctgattttt atcacagcaa gagaaaaatca	660
aactagaaca aacatgtatt tttcctggca catgtccata gtaaggcaga aatgatcttc	720
agacctagac catagatact acagagagca gaagtgtaga taggtggact tactgtatga	780
ttgtaatcca agtaaatcta catagctaga gagctagagg aaaggccaaa gcttcctctg	840
ggaggtcaga tcctgtcgcg ctgtagccaa taaggcatat tgcacacag gaaaggacta	900
agaccaggc tggcaatagt gtctgtatct taactagatc tctctagtga gtgaggaagt	960
aaatttgtga gagcccagac tgtgggctcg gaaggtacct gccatgcccc tgttagtaac	1020
tgagtactac agcaggagca ggtgttctct agaaagcctg agacaactct acttcttctc	1080
tcaagagacc acctaatata ggcctgagag aacagactct ggaatatagat gggacttacg	1140
gagctaagat ctgagactca tctacagagc agaatcccag ccaagagaac aaagaatact	1200
gactctctcc tgttccctac tcttagagtt ctaaaaacaca ctatagggaa gggagcctct	1260
agacctccgt ccattccccca tcttgetcat tccatcttcc catgtcccca ggtctccaag	1320
ccacagacac caccttctct attcaccacc ctttctgtgt ccctagggtcc ccaggccata	1380
gtcacctccc cccacacccc gctcacctcg ccccatctat gccctagat gcttacttac	1440
cagagtcttt tgtctgacgt ggggctacaa gcatctatgc tccttaagca cctactgctg	1500
acctgtagga cccagctctg aaccaactca tataagtaaa tacagactct cccctgtctt	1560

-continued

aggatggccc	cctgggtcag	gaggagacca	ctgccaagga	accttctctt	agagcactga	1620
actcctcccc	tgtaccactt	aggacagacc	tgagacctat	tattactgat	taccagagct	1680
ctggcagtga	ccacggagga	gatagatcca	ccctggacac	aggaaacaca	gcaccagaga	1740
tactgcttca	tcacaacagt	agagtgacac	ttagactttt	aatttgggtc	actttcctgc	1800
tgtagagggtg	ggatcagaaa	gcaaagagca	gtatgagtgc	ctgataggca	cccaagtaca	1860
ctatagagta	ctcatgggta	ataaggtacc	tccatggctt	cccagggagg	ggcactgccc	1920
cacccccacc	atcacagacc	tttctccata	gttgataact	cagacacaag	tgaatgacag	1980
atggacctcc	atctgctctt	attttaaaaa	gaagacaaac	cccacaggct	cgagaacttt	2040
agcgactggt	ttgagagaaa	tcattggtcc	ctgactcaag	agatgactgg	cagattgggg	2100
atcagaatac	ccatactctg	tggctagtgt	gaggtttaag	cctcagagtc	cctgtggtct	2160
ctgactgggtg	caagggtttg	actaagcggg	gcaccacagt	gctaactggg	accacgggtga	2220
cacgtggctc	aacaaaaacc	ttctgtttgg	agctctccag	gggcagcctg	agctatgagg	2280
aagtagagag	gcttgagaaa	tctgaggaag	aaaagagtag	atctgagagg	aaaggtagct	2340
ttctggaggt	caggagacag	tgacagagaag	aacgagttac	tgtggacagg	tcttagatgg	2400
ggaaagaatg	agcaaatgca	agcatcagaa	gggtggatgc	aatgtctctc	caaggactta	2460
ccaagaggat	ccccggacag	agcaggcagg	tggagttgac	tgagaggaca	ggatagggtc	2520
aggteccctc	cttgtttcct	ttctccttct	cctgtttcct	tcttctcttg	tcacaggtct	2580
cactatgcta	gccaaaggcta	gcttgaagaa	ttaccatcct	acagatgggc	ccatccagtt	2640
gaattaaggt	ggagatctct	ccaaacatct	gagtttctga	ggcttggatg	ccactgggga	2700
cgccaaggga	ctttgggatg	ggtttggtg	gccccagatg	aagggtact	tcactgggtc	2760
tataattact	ctgatgtcta	ggaccagggg	gctcagggtca	ctcaggtcag	gtgagtctg	2820
catctgggga	ctgtggggtt	caggtggcct	aaggcaggat	gtggagagag	ttttagtata	2880
ggaacagagg	cagaacagag	actgtgctac	tggactctcg	atgtctgggg	cacagggacc	2940
acggtcaccg	tctcctcagg	taagctggct	tttttcttct	tgcacattcc	attctgaaac	3000
gggaaaagat	attctcagat	ctccccatgt	caggccatct	gccacactct	gcatgctgca	3060
gaagcttttc	tgtaaggata	gggtcttcac	tcccaggaaa	agaggcagtc	agaggctagc	3120
tgctgtgga	acagtgacaa	tcatggaaaa	taggcattta	cattgttagg	ctacatgggt	3180
agatgggttt	ttgtacacc	actaaagggg	tctatgatag	tgtgactact	ttgactactg	3240
gggccaaggc	accactctca	cagtctcctc	aggtagtcc	ttacaacctc	tctctctat	3300
tcagcttaaa	tagattttac	tgcatttgtt	gggggggaaa	tgtgtgtatc	tgaatttcag	3360
gtcatgaagg	actagggaca	ccttgggagt	cagaaaggtt	cattgggagc	cctggctgac	3420
gcagacagac	atcctcagct	cccatacttc	atggccagag	atztataggg	atcctggcca	3480
gcattgcccg	taggtccctc	tctctatgc	ttcttttgc	cctcactggc	ctccatctga	3540
gatcatctcg	gagccctagc	caaggatcat	ttattgtcag	gggtctaate	attgtgttca	3600
caatgtgctc	ggtttgttta	ctggggccaa	gggactctgg	tactgtctc	tgacaggtgag	3660
tcctaaactc	tcccattcta	aatgcatggt	ggggggattc	tgggccttca	ggaccaagat	3720
tctctgcaaa	cggaatcaaa	gattcaaccc	ctttgtccca	aagttgagac	atgggtctgg	3780
gtcagggact	ctctgctg	tggtctgtgg	tgacattaga	actgaagtat	gatgaaggat	3840

-continued

```

ctgccagaac tgaagcttga agtctgaggc agaattctgt ccagggctca tggactctt 3900
gtgagaatta ggggctgaca gttgatggtg acaatttcag ggtcagtgac tgtctggttt 3960
ctctgaggtg aggctggaat ataggtcacc ttgaagacta aagaggggtc caggggcttc 4020
tgcacaggca gggaacagaa tgtggaacaa tgacttgaat gggtgattct tgtgtgacac 4080
caggaattgg cataatgtct gagttgocca ggggtgattc tagtcagact ctggggtttt 4140
tgtcgggtat agaggaaaaa tccactattg tgattactat gctatggact actgggggtca 4200
aggaacctca gtcaccgtct cctcaggtaa gaatggctc tccaggtctt tatttttaac 4260
ctttgttatg gagttttctg agcattgcag actaatcttg gatatttgtc cctgagggag 4320
ccggctgaga gaagttggga aataaactgt ctagggatct cagagccttt aggacagatt 4380
atctccacat ctttgaaaaa ctaagaatct gtgtgatggg gttgggtggag tccctggatg 4440
atgggatagg gactttggag gctcatttga agaagatgct aaaacaatcc tatggctgga 4500
gggatagttg gggct 4515

```

```

<210> SEQ ID NO 7
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide HV2-5

```

```

<400> SEQUENCE: 7

```

```

agatcacctt gaaggagtct ggtcc 25

```

```

<210> SEQ ID NO 8
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide HV4-4

```

```

<400> SEQUENCE: 8

```

```

tggatgaagcc ttcggagacc ctgtc 25

```

```

<210> SEQ ID NO 9
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide HV1-3

```

```

<400> SEQUENCE: 9

```

```

cactagctat gctatgcatt gggtg 25

```

```

<210> SEQ ID NO 10
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide HV1-2

```

```

<400> SEQUENCE: 10

```

```

atggatcaac ctaacagtg gtggc 25

```

```

<210> SEQ ID NO 11
<211> LENGTH: 25
<212> TYPE: DNA

```

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide HV6-1

<400> SEQUENCE: 11

ggaaggacat actacaggtc caagt 25

<210> SEQ ID NO 12
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide C

<400> SEQUENCE: 12

taggtacttg cccctgtcc tcagt 25

<210> SEQ ID NO 13
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide KV1-9

<400> SEQUENCE: 13

agcccagtgt gttccgtaca gcctg 25

<210> SEQ ID NO 14
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide KV1-8

<400> SEQUENCE: 14

atcctcattc tctgcatcta cagga 25

<210> SEQ ID NO 15
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide KV1-6

<400> SEQUENCE: 15

ggtaaggatg gagaacactg gcagt 25

<210> SEQ ID NO 16
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide KV1-5

<400> SEQUENCE: 16

ttagtagctg gttggcctgg tatca 25

<210> SEQ ID NO 17
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Oligonucleotide Ck

-continued

<400> SEQUENCE: 17

ctttgctgtc ctgatcagtc caact 25

<210> SEQ ID NO 18
 <211> LENGTH: 10
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Rat Switch
 <220> FEATURE:
 <221> NAME/KEY: repeat_region
 <222> LOCATION: (1)..(10)
 <223> OTHER INFORMATION: Repeated 148 times

<400> SEQUENCE: 18

gagctgagct 10

<210> SEQ ID NO 19
 <211> LENGTH: 10
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Rat Switch
 <220> FEATURE:
 <221> NAME/KEY: repeat_region
 <222> LOCATION: (1)..(10)
 <223> OTHER INFORMATION: Repeat 25 times

<400> SEQUENCE: 19

gggggtggggt 10

<210> SEQ ID NO 20
 <211> LENGTH: 10
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Rat Switch
 <220> FEATURE:
 <221> NAME/KEY: repeat_region
 <222> LOCATION: (1)..(5)
 <223> OTHER INFORMATION: Repeat 82 times

<400> SEQUENCE: 20

gggctgggct 10

<210> SEQ ID NO 21
 <211> LENGTH: 6
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Antibody Sequence
 <220> FEATURE:
 <221> NAME/KEY: MISC_FEATURE
 <222> LOCATION: (1)..(2)
 <223> OTHER INFORMATION: Xaa Xaa = PR, RT, or PW

<400> SEQUENCE: 21

Xaa Xaa Thr Phe Gly Gln
 1 5

<210> SEQ ID NO 22
 <211> LENGTH: 17
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Antibody Sequence

-continued

```

<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(2)
<223> OTHER INFORMATION: Xaa Xaa= PR, RT, or PW

<400> SEQUENCE: 22

Xaa Xaa Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Ala Asp
1           5           10           15

```

Ala

```

<210> SEQ ID NO 23
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Antibody Sequence
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(2)
<223> OTHER INFORMATION: Xaa Xaa = PR or PW

```

<400> SEQUENCE: 23

```

Xaa Xaa Thr Phe Gly Gln
1           5

```

```

<210> SEQ ID NO 24
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Antibody Sequence
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(2)
<223> OTHER INFORMATION: Xaa Xaa = PR or PW

```

<400> SEQUENCE: 24

```

Xaa Xaa Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys Arg Ala Asp
1           5           10           15

```

Ala

```

<210> SEQ ID NO 25
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer E1554

```

<400> SEQUENCE: 25

atgacttcag tgtgttctg gtag

24

```

<210> SEQ ID NO 26
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer E1555

```

<400> SEQUENCE: 26

caccagattc ttatcagac

19

```

<210> SEQ ID NO 27
<211> LENGTH: 35
<212> TYPE: DNA

```

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: ELP1352_Cy1

<400> SEQUENCE: 27
agagcggcgcg ctgggcaacg ttgcaggtga cggtc 35

<210> SEQ ID NO 28
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: ELP1353_Cy2b

<400> SEQUENCE: 28
agagcggcgcg ctttgtccac cgtggtgctg ctgg 34

<210> SEQ ID NO 29
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: ELP1354_Cy2a

<400> SEQUENCE: 29
agagcggcgcg cacattgcag gtgatggact ggc 33

<210> SEQ ID NO 30
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: ELP1356_VH4-4

<400> SEQUENCE: 30
aggacgcgtg aaacacctgt ggttcttcct cctgc 35

<210> SEQ ID NO 31
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: ELP1357_VH1-2,3

<400> SEQUENCE: 31
aggacgcgtc accatggact ggacctggag gat 33

<210> SEQ ID NO 32
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: ELP1358_VH6-1

<400> SEQUENCE: 32
aggacgcgta tgtctgtctc cttctcctc ttcc 34

<210> SEQ ID NO 33
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: mIgG1.2 rev

-continued

<400> SEQUENCE: 33

ggggccagtg gatagacaga t

21

<210> SEQ ID NO 34

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: mIgG2b rev

<400> SEQUENCE: 34

cagtgatag actgatgg

18

<210> SEQ ID NO 35

<211> LENGTH: 18

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: mIgG2a_2 rev

<400> SEQUENCE: 35

cagtgatag accgatgg

18

<210> SEQ ID NO 36

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: mCH1 unirev

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (1)..(1)

<223> OTHER INFORMATION: K = G or T

<400> SEQUENCE: 36

kcaggggcca gtggatagac

20

<210> SEQ ID NO 37

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: mCH1 unirev_2

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (3)..(3)

<223> OTHER INFORMATION: R = A or G

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (6)..(6)

<223> OTHER INFORMATION: Y = C or T

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (12)..(12)

<223> OTHER INFORMATION: M = A or C

<400> SEQUENCE: 37

tarcyyttga cmaggcatcc

20

<210> SEQ ID NO 38

<211> LENGTH: 24

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 16C9

-continued

<400> SEQUENCE: 38

Gln Glu Val Ile Asn Tyr Tyr Tyr Tyr Gly Met Asp Val Trp Gly Gln
1 5 10 15
Gly Thr Thr Val Thr Val Ser Ser
 20

<210> SEQ ID NO 39

<211> LENGTH: 24

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 20B5

<400> SEQUENCE: 39

Gln Glu Val Ile Asn Tyr Tyr Tyr Tyr Gly Met Asp Val Trp Gly Gln
1 5 10 15
Gly Thr Thr Val Thr Val Ser Ser
 20

<210> SEQ ID NO 40

<211> LENGTH: 26

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 19F4

<400> SEQUENCE: 40

Leu Glu Met Ala Thr Ile Asn Tyr Tyr Tyr Tyr Gly Met Asp Val Trp
1 5 10 15
Gly Gln Gly Thr Met Val Thr Val Ser Ser
 20 25

<210> SEQ ID NO 41

<211> LENGTH: 24

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 19E1

<400> SEQUENCE: 41

Gln Glu Phe Gly Asn Tyr Tyr Tyr Tyr Gly Met Asp Val Trp Gly Gln
1 5 10 15
Gly Thr Thr Val Thr Val Ser Ser
 20

<210> SEQ ID NO 42

<211> LENGTH: 24

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 19G8

<400> SEQUENCE: 42

Gln Glu Asp Gly Asn Pro Tyr Tyr Phe Gly Met Asp Phe Trp Gly Gln
1 5 10 15
Gly Thr Thr Val Thr Val Ser Ser
 20

<210> SEQ ID NO 43

<211> LENGTH: 22

<212> TYPE: PRT

-continued

<213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: 20H10

<400> SEQUENCE: 43

Gly Ser Ser Tyr Tyr Tyr Asp Gly Met Asp Val Trp Gly Gln Gly Thr
 1 5 10 15
 Thr Val Thr Val Ser Ser
 20

<210> SEQ ID NO 44
 <211> LENGTH: 25
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: 18D10

<400> SEQUENCE: 44

Leu Glu Asn Asp Tyr Gly Tyr Tyr Tyr Tyr Gly Met Asp Val Trp Gly
 1 5 10 15
 Gln Gly Thr Thr Val Thr Val Ser Ser
 20 25

<210> SEQ ID NO 45
 <211> LENGTH: 23
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: 16F2

<400> SEQUENCE: 45

Arg Gly Gly Leu Ser Pro Leu Tyr Gly Met Asp Val Trp Gly Gln Gly
 1 5 10 15
 Thr Thr Val Thr Val Ser Ser
 20

<210> SEQ ID NO 46
 <211> LENGTH: 15
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Rat Switch

<400> SEQUENCE: 46

gggctgggct gggct 15

<210> SEQ ID NO 47
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Rat Switch

<400> SEQUENCE: 47

gggctgggct gggctgggct 20

<210> SEQ ID NO 48
 <211> LENGTH: 25
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Rat Switch

-continued

<400> SEQUENCE: 48
 gggctgggct gggctgggct gggct 25

<210> SEQ ID NO 49
 <211> LENGTH: 30
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Rat Switch

<400> SEQUENCE: 49
 gggctgggct gggctgggct gggctgggct 30

<210> SEQ ID NO 50
 <211> LENGTH: 10
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Rat Switch
 <220> FEATURE:
 <221> NAME/KEY: repeat_region
 <222> LOCATION: (1)..(5)
 <223> OTHER INFORMATION: Repeat 6 to 81 times

<400> SEQUENCE: 50
 gggctgggct 10

<210> SEQ ID NO 51
 <211> LENGTH: 6
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: IGHV4-4*02 CDR1

<400> SEQUENCE: 51
 Ser Ser Asn Trp Trp Ser
 1 5

<210> SEQ ID NO 52
 <211> LENGTH: 6
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: 12D10 CDR1

<400> SEQUENCE: 52
 Ser Gly Asn Trp Trp Ser
 1 5

<210> SEQ ID NO 53
 <211> LENGTH: 6
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: 1283 CDR1

<400> SEQUENCE: 53
 Arg Ser Asn Trp Trp Ser
 1 5

<210> SEQ ID NO 54
 <211> LENGTH: 7
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:

<223> OTHER INFORMATION: IGHV6-1*01 CDR1

<400> SEQUENCE: 54

Ser Asn Ser Ala Ala Trp Asn
 1 5

<210> SEQ ID NO 55

<211> LENGTH: 7

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 4A12 CDR1

<400> SEQUENCE: 55

Ser Asn Ser Ala Ala Trp Asn
 1 5

<210> SEQ ID NO 56

<211> LENGTH: 16

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: IGHV4-4*02 CDR2

<400> SEQUENCE: 56

Glu Ile Tyr His Ser Gly Ser Thr Asn Tyr Asn Pro Ser Leu Lys Ser
 1 5 10 15

<210> SEQ ID NO 57

<211> LENGTH: 16

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 12D10 CDR2

<400> SEQUENCE: 57

Glu Ile Tyr His Ser Gly Asn Thr Asn Tyr Asn Pro Ser Leu Lys Ser
 1 5 10 15

<210> SEQ ID NO 58

<211> LENGTH: 16

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 1283 CDR2

<400> SEQUENCE: 58

Glu Ile Tyr His Ser Gly Ser Thr Asn Tyr Asn Pro Ser Leu Lys Ser
 1 5 10 15

<210> SEQ ID NO 59

<211> LENGTH: 18

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: IGHV6-1*01 CDR2

<400> SEQUENCE: 59

Arg Thr Tyr Tyr Arg Ser Lys Trp Tyr Asn Asp Tyr Ala Val Ser Val
 1 5 10 15

Lys Ser

-continued

<210> SEQ ID NO 60
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 4A12 CDR2

<400> SEQUENCE: 60

Arg Thr Tyr Tyr Arg Ser Lys Trp Tyr Asn Asp Tyr Lys Val Ser Val
1 5 10 15

Lys Ser

<210> SEQ ID NO 61
<211> LENGTH: 12
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 12D10 CDR3

<400> SEQUENCE: 61

Gly Pro Leu Thr Gly Glu Lys Tyr Tyr Phe Asp Leu
1 5 10

<210> SEQ ID NO 62
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 1283 CDR3

<400> SEQUENCE: 62

Ile Gly Asp Trp Tyr Phe Asp Leu
1 5

<210> SEQ ID NO 63
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 4A12 CDR3

<400> SEQUENCE: 63

Glu Gly Ser His Ser Gly Ser Gly Trp Tyr Leu Asp Ala Phe Asp Ile
1 5 10 15

<210> SEQ ID NO 64
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: IGHJ2*01 J-region

<400> SEQUENCE: 64

Tyr Trp Tyr Phe Asp Leu Trp Gly Arg Gly Thr Leu Val Thr Val Ser
1 5 10 15

Ser

<210> SEQ ID NO 65
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: 12D10 J-region

-continued

<400> SEQUENCE: 65

Tyr Tyr Phe Asp Leu Trp Gly Arg Gly Thr Leu Val Thr Val Ser Ser
 1 5 10 15

<210> SEQ ID NO 66

<211> LENGTH: 16

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 1283 J-Region

<400> SEQUENCE: 66

Trp Tyr Phe Asp Leu Trp Gly Arg Gly Thr Leu Val Thr Val Ser Ser
 1 5 10 15

<210> SEQ ID NO 67

<211> LENGTH: 16

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: IGHJ3*01 J-Region

<400> SEQUENCE: 67

Asp Ala Phe Asp Val Trp Gly Gln Gly Thr Met Val Thr Val Ser Ser
 1 5 10 15

<210> SEQ ID NO 68

<211> LENGTH: 16

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: 4A12 J-region

<400> SEQUENCE: 68

Asp Ala Phe Asp Ile Trp Gly Gln Gly Thr Lys Val Thr Val Ser Ser
 1 5 10 15

1. A method for obtaining an antigen specific antibody or antigen binding fragment thereof, said antigen-specific antibody comprising a human immunoglobulin heavy (IgH) chain comprising a human IgH chain variable region and a human IgH chain constant region, said antigen-binding fragment comprising a human IgH chain variable region, the method comprising:

expressing the antibody from a cell comprising nucleic acid encoding said human IgH chain variable region and said human IgH chain constant region of said antibody, or expressing said fragment from a cell comprising nucleic acid encoding said human IgH constant region,

wherein said human IgH chain variable region is of a transgenic mouse contacted with said antigen;

wherein the germline of said transgenic mouse comprises a homozygous chimeric immunoglobulin heavy chain (IgH) locus comprising unrearranged human IgH variable region gene segments at an endogenous mouse heavy chain locus upstream of an enhancer and constant (C) region comprising an endogenous IgH C gene segment, wherein said human variable region gene segments in said chimeric IgH locus are operably

linked to said C region at a human/mouse chimeric junction within the JC intron of said chimeric IgH locus,

wherein said IgH locus comprises, in 5' to 3' transcriptional orientation,

(i) unrearranged human immunoglobulin heavy chain (IgH) variable region (VH) DNA comprising human IgH V gene segments, human D gene segments and human JH gene segments comprising a human-3' JH gene segment,

(ii) a chimeric J/C intron comprising human DNA downstream of and naturally contiguous with said human 3' JH gene segment, which is contiguous with mouse JC intronic DNA, and

and

(iii) said enhancer and said constant region

wherein said human 3'JH is less than 2 kb upstream of said chimeric junction,

wherein DNA between said chimeric junction and said enhancer comprises mouse JC intronic DNA;

wherein said transgenic mouse is functional to form rearranged human VH, DH, and JH gene segments and to express chimeric immunoglobulin heavy chain polypeptide comprising a human VH region and a mouse C region, and

wherein said transgenic mouse is capable, upon stimulation with antigen, of producing antibody specific for said antigen comprising a chimeric Ig heavy chain comprising a human IgH variable region and said C region; and

wherein the germline of said transgenic mouse comprises all or part of the endogenous mouse IgH variable region, and is capable of breeding to produce subsequent generation mice having in their germline an IgH locus comprising unrearranged human IgH variable region gene segments upstream of an IgH constant C region comprising an endogenously-located CH gene segment.

2. The method of claim 1, said method further comprises the step of:

recovering said antibody or antigen-binding fragment thereof, said cell expressing said antibody or antigen-binding fragment there, and/or recovering nucleic acid encoding said human VH region.

3. The method of claim 1,

wherein said cell is a first cell, and wherein said nucleic acid encoding said human IgH chain variable region is obtained from a second cell comprising nucleic acid encoding said human IgH chain variable region, and wherein said second cell is of a transgenic mouse contacted with said antigen.

4. The method of claim 1, wherein in said mouse germline said human 3' JH is less than 1 kb upstream of said chimeric junction.

5. The method of claim 1, wherein in said mouse germline said unrearranged human heavy chain V region DNA comprises gene segments V6-1, V11-1-1, V1-2, V111-2-1, V1-3, V4-4 and V2-5 gene segments.

6. The method of claim 1, wherein in said mouse germline said unrearranged human IgH VH gene segments comprise all of human V, D and J gene segments.

7. The method of claim 1, wherein in said mouse germline said mouse JC intron comprises mouse 129 strain DNA.

8. The method of claim 2, further comprising the step of: expressing the antibody from said nucleic acid, wherein said nucleic acid encodes an antibody comprising said human variable region.

9. The method of claim 3, wherein said second cell is selected from the group consisting of: a B cell of a transgenic mouse contacted with said antigen; a cell comprising

nucleic acid encoding said human IgH chain variable region and a mouse IgH constant region; a hybridoma expressing said human IgH chain variable region; a plurality of cells comprising nucleic acid encoding human IgH chain variable regions; and an immortalised cell comprising nucleic acid encoding said human IgH chain variable region.

10. The method of claim 3, wherein all or part of said endogenous mouse IgH locus is inactivate.

11. The method of claim 10, wherein expression of IgH heavy chains comprising endogenous mouse variable and constant regions in said transgenic mouse is reduced or prevented.

12. The method of claim 1, wherein said mouse is functional to produce antibody isotypes IgM and IgG specific for said antigen each isotype comprising a human heavy chain variable region.

13. The method of claim 1, wherein in said mouse germline said human DNA downstream of and naturally contiguous with said human 3' JH gene segment comprises a truncated human JC intron and said mouse intronic DNA comprises a truncated mouse JC intron.

14. The method of claim 1, further comprising providing a pharmaceutical composition comprising said antigen-specific antibody or antigen-binding fragment thereof and a pharmaceutically acceptable carrier or excipient.

15. The method of claim 1, further comprising providing said antibody or antigen-binding fragment thereof to a human subject.

16. The method of claim 1, said antigen-specific antibody or antigen binding fragment thereof is a monoclonal antibody, a domain antibody or a neutralizing antibody, or antigen binding fragment thereof.

17. The method of claim 1, wherein in said mouse germline said enhancer comprises a E μ enhancer.

18. The method of claim 17, wherein in said mouse germline said mouse enhancer is a mouse E μ enhancer.

19. The method of claim 1, wherein in said mouse germline said 3' JH gene segment comprises a human JH6 gene segment.

20. The method of claim 18, wherein said mouse E μ enhancer comprises mouse 129 strain DNA.

21. The method of claim 1, wherein in said mouse germline said C region comprises a plurality of CH gene segments endogenous to said transgenic mouse.

* * * * *