(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 March 2006 (02.03.2006)

PCT

(10) International Publication Number
WO 2006/021800 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/GB2005/003337

(22) International Filing Date: 26 August 2005 (26.08.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0419080.7 26 August 2004 (26.08.2004) GB

(71) Applicant *(for all designated States except US)*: **MEDICAL RESEARCH COUNCIL** [GB/GB]; 20 Park Crescent, London W1B 1AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **RABBITTS, Terence, Howard** [GB/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). **FORSTER, Alan** [GB/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB).

(74) Agents: **SOAMES, Candida, Jane** et al.; D Young & Co, 120 Holborn, London EC1N 2DY (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2006/021800 A2

(54) Title: TUMOUR MODEL

(57) Abstract: The invention relates to a method for generating a non-human animal model of a chromosomal rearrangement, comprising creating a transgenic non-human mammal expressing a site-specific recombinase under the control of a cell type-specific promoter, and having sites recognised by the recombinase present in its genome such that a chromosomal rearrangement is catalysed by the recombinase, wherein upon expression of the recombinase gene inversion and the subsequent expression of one or more fusion proteins ocurrs.

TUMOUR MODEL

The present invention relates to a model for tumourigenesis. In particular, the invention relates to an animal model for leukaemia which is based on a chromosomal translocation which creates a fusion from the *Ews* and *ERG* genes or the MLL and AF4 genes.

Introduction

The aetiology of tumours often involves chromosomal translocations in haematopoietic malignancies, sarcomas and epithelial tumours and these can either activate proto-oncogenes or create novel fusion genes which function in tumourigenesis (Rabbitts, 1994). These translocation products act at various levels in afflicted cells, although the translocation genes always encode intracellular proteins located in either the cytoplasm or nuclei. The recurrent chromosomal translocations in haematopoietic malignancies and in sarcomas generally display a tropism in respect of the cell type in which they are found, for instance the *MLL* gene fusion sub-types being restricted to either lymphoid or myeloid malignancies (Downing and Shannon, 2002; Look, 1997; Rowley, 1998). The *MLL* gene is a frequent site of chromosomal translocations in human leukaemias (Corral et al., 1993; Djabali et al., 1992; Domer et al., 1993; Gu et al., 1992; Thirman et al., 1993; Tkachuk et al., 1992), affecting over 30 different chromosomes and resulting in many leukaemia-associated *MLL*-gene fusions (Ayton and Cleary, 2001; Collins and Rabbitts, 2002). Different fusions distinguish distinct leukaemias, for instance the *MLL-AF9* fusion is almost exclusively in myeloid leukaemias while *MLL-AF4* is confined to B-lymphoid leukaemias (Gu et al., 1992; Rowley, 1998). Others, such as *MLL-ENL* can be found in myeloid and lymphoid leukaemias (Ayton and Cleary, 2001).

Mouse models which mimic the effects of chromosomal translocations, or which recapitulate these events, are essential to understand the determinants of frank malignancy and those dictating the restriction of translocations to specific lineages. In addition, with advancement of efforts to devise new reagents which can interfere with

protein function, mouse models of human cancer are crucial to develop and test these rationally devised therapeutic approaches, prior to application in patients. Finally, mouse models of human cancer will be needed for *in vivo* testing of new compounds, which may have promising biochemical properties, in order to accelerate the translation of these into clinical use.

Mouse cancer models involving chromosomal translocation-genes have been either gain of function transgenic models (Adams et al., 1999; Pandolfi, 1998), retroviral transduction models or translocation mimics in which a gene fusion is created by homologous recombination into an endogenous target gene using embryonic stem cells (ES cells) (Corral et al., 1996). For instance, the consequences of *MLL*-associated translocations have been modelled in mice using retroviral transduction of bone marrow progenitors (Lavau et al., 1997; Schulte et al., 2002; Slany et al., 1998) or by homologous recombination knock-in of fusion genes (Corral et al., 1996; Dobson et al., 1999; Dobson et al., 2000). In the knock-in approach, homologous recombination was used to introduce fusion oncogenes into mice, via embryonic stem cells. In this way, a mouse model of the human *MLL-AF9* gene fusion was made in which knock-in mice were found to develop an acute myeloid malignancy similar to that found in human patients with the chromosomal translocation t(9;11) (Corral et al., 1996; Dobson et al., 1999). While the knock-in approach is a close mimic of natural chromosomal translocations, it is limited by the fact that the homologous recombination event generates a knock-in of one allele which is subsequently transferred to all cells of the mouse. Thus the fusion gene will be expressed in all sites where the targeted gene is expressed, for instance as found for the *Mll-AF9* knock-in fusion (Corral et al., 1996). In addition, dominant, embryonic lethal effects of knock-in alleles have been observed (Okuda et al., 1998; Yergeau et al., 1997). These can be overcome by creating conditional knock-in alleles such as the *Aml1-Eto* example, which circumvents the embryonic lethal effects of the fusion, and causes AML to arise in mice (Higuchi et al., 2002).

The most ideal concept is the *de novo* creation of chromosomal translocations which could potentially be achieved by *in vivo* recombination systems, such as the Cre-*loxP* system (Smith et al., 1995; van Deursen et al., 1995).

SSRs of the integrase class such as FLP, from the 2μ episome of S. cereviseae, Cre, from the *E. coli* P1 phage, and R from *Z. rouxii* are able to mediate precise and conservative recombination between their binding sites. These proteins share overall similarity in their primary sequence and attract attention because they have been shown to function efficiently in a broad range of organisms from *E.coli* to mice. Their binding sites are 13bp inverted repeats which are sequence specifically recognised by a monomer per half site. Other SSRs having homologous structures are known.

The use of SSRs is potentially the nearest that could be achieved of a direct recapitulation of the *de novo* occurrence of human cancer-associated chromosomal translocations. The Cre-*loxP* recombination system has been shown to facilitate *de novo* chromosomal translocations in mice (Buchholz et al., 2000; Collins et al., 2000) but no tumour incidence was reported in these initial translocator mice.

Human cancer originates in somatic cells acquiring mutations, either point mutations or more gross changes manifest by chromosomal translocations, inversions or deletions [1]. All tumors have some or all of these changes. Chromosomal translocations are amongst the best studied since they are usually visible by microscopy, facilitating the molecular characterisation of the breakpoint sequences and of the genes (oncogenes) involved [2,3], Differences occur in the nature of the consistent chromosomal translocations in haematopoietic (leukemias and lymphomas) or mesenchymal tumors (sarcomas) and epithelial tumors (carcinomas) since the former two tend to have reciprocal translocations whereas the latter often have non-reciprocal translocations [4]. The main outcomes of the reciprocal chromosomal translocations are either forced oncogene expression as found in lymphoid malignancies or gene fusion found in both leukemias and sarcomas [2,3]. Gene fusion results when the break and re-joining during the translocation occurs within the introns of genes, such that the two

translocated chromosomes have new exon organization, leading to the formation of chimaeric mRNA species and in turn, chimaeric proteins.

Different tumor phenotypes can result from specific versions of related fusion proteins, for instance the Philadelphia t(9;22) translocation (which yields the *BCR-ABL* fusion) occurs in both myeloid- and lymphoid-lineage tumors, dictated largely by the position of the translocation junction within the 5' *BCR* gene [5]. The variability of gene fusions is even greater with some leukemia genes, in particular the *MLL* gene for which more than 30 different fusion partners are known [6,7]. This situation is not confined to haematopoietic malignancies, as the *EWS-FUS* family, typically involved in sarcomas, exhibits similar characteristics [8-11]. The *EWS* gene was identified in Ewing's sarcoma by its association with *FLI1* and *ATF1* [12,13] and subsequently found in subsets of Ewing's with either *ERG* or *ETV1* [14,15]. Further, the *EWS* gene is involved in other sarcomas and with yet different partners, such as *WT1* in desmoplastic small round cell sarcoma [16] or *CHN* in myxoid chondrosarcoma [17]. A further ramification of this infidelity, is that the *FUS* gene, first identified by its fusion with the *CHOP* gene in t(12;16) of malignant myxoid liposarcoma and related to the *EWS* gene [18,19], has also been found in Ewing's sarcoma or acute myeloid leukemia fused with the *ERG* gene [20-22]. Finally the *EWS* gene can also be involved with the *CHOP* gene by chromosomal translocation in malignant myxoid liposarcoma [23], analogous to *FUS-CHOP*. These observations suggest that the EWS/FUS fusion proteins are effective in mesenchymal cells or in haematopoietic cells and their presence may be dictated simply by the cell type in which the relevant chromosomal translocation took place. Conditional mouse models of the *EWS* and *FUS*-associated chromosomal translocations should give insights into whether these fusions influence the malignant phenotype.

Conditional mouse models which mimic chromosomal translocations are important to provide a biology basis to study how oncogenes influence tumorigenicity. The present inventors have previously developed two mouse models mimicking chromosomal translocations using homologous recombination in ES cells to create mutant alleles and transferring these alleles into the mouse germline. The first approach was the knock-in model [24] which involved fusing the *MLL* gene partner *AF9* into the mouse *Mll* gene.

This model resulted in acute myeloid leukemias in mice, preceded by a myeloproliferation in the bone marrow [24,25]. A drawback of this knock-in mouse model is that the fusion is effectively a germ-line mutation and thus cell specific effects cannot be achieved.

This deficiency was obviated in the translocator [26,27] in which *de novo* chromosomal translocations are created in mice [28,29], resulting in cell specific chromosomal translocations and leukemias [30]. The translocator mouse model involves the use of *loxP* sites, introduced by homologous recombination into two target chromosomes, to facilitate site-specific recombination mediated by Cre recombinase [28-30].

In an *Mll-Enl* model which was also developed by the present inventors, myeloid leukemias arose quickly and with high penetrance when Cre was expressed under the control of the haematopoietic stem cell gene *Lmo2* [30]. The translocator approach could, in principle, be used to model any human chromosomal translocation but if the gene orientation on the mouse chromosomes is incompatible with transcription following Cre-*loxP* recombination, a conditional version of the knock-in chromosomal translocation mimic strategy must be invoked. A conditional version of the original knock-in approach, using a *loxP* flanked (floxed) transcription stop, has been used to circumvent embryonic lethality [31] but even this approach cannot always prevent low levels of read-through transcription, to avoid the dominant lethal effect of a fusion protein (MM, AF, RP and THR, unpublished results).

Thus there remains a need in the art to provide a mouse model of human chromosomal rearrangements which overcomes the problems described above.

SUMMARY OF THE INVENTION.

To avoid the problems of the prior art described above, the present inventors have developed a novel version of the knock-in translocation mimic, named **the invertor model**. In this approach, a floxed cassette carrying a cDNA sequence with an intron segment and acceptor splice site is knocked-in to the intron of a target gene but in the

reverse transcription orientation to the latter. Consequently and importantly, no RNA splicing can take place between the inserted cassette until it is inverted by expression of Cre recombinase. The method of the invention ensures that RNA splicing and thus transcription only occurrs in conjunction with the expression of Cre recombinase. This important step therefore ensures that read-through transcription does not ocurr and thus the problems associated with the expression of embryonically lethal fusion proteins is avoided.

Thus in a first aspect the present invention provides a method for generating a non-human animal model of a chromosomal rearrangement, which method comprises creating a transgenic non-human mammal wherein an invertor cassette (comprising a first nucleic acid sequence and one or more introns comprising one or more acceptor splice sites from a gene of interest, wherein the first nucleic acid sequence and the intronic sequence/s are flanked by site/s which recognise a site specific recombinase) is knocked into a target gene of interest with that animal so that the first nucleic acid sequence is in a reverse transcription orientation with respect to that gene of interest and wherein that animal expresses a site-specific recombinase under the control of a cell type-specific promoter, such that a chromosomal inversion is catalysed by the recombinase upon its expression.

The method referred to herein and described above has been named the invertor method.

According to the method of the invention, the first nucleic acid sequence is comprised within a nucleic acid cassette (invertor cassette). Preferably, the nucleic acid is cDNA. Advantageously, the nucleic acid cassette comprises one or more of the following components in addition to a first nucleic acid sequence: a segment of intronic sequence including an acceptor splice site from the **gene of interest,** a poly A site and a segment for the selection of homologous recombinant clones, preferably a MC1-neo-pA seqment. These regions are flanked by one or more site/s which recognise and bind the recombinase. Advantageously, the recognition sites are loxPsites and the recombinase is the Cre recombinase.

Further according to the above aspect of the invention, the nucleic acid cassette (invertor cassette) for introduction of the first nucleic acid sequence into the gene of interest may be introduced into the gene of interest, preferably the mouse gene of interest using any suitable method known to those skilled in the art. Advantageously, it is introduced using gene targetting in suitable cells, for example embryonic stem cells (ES cells). According to the method of the invention, advantageously, the nucleic acid invertor cassette according to the invention is introduced into the gene of interest downstream from a donor splice site. By introducing the invertor cassette 3' to 5' with respect to the orientation of the gene of interest, then a splice from the gene of interest (comprising a splice donor site) to the acceptor site of the invertor cassette is not possible. Thus read-through transcription is prevented using the method of the invention.

Thus in a further aspect the present invention provides an invertor cassette comprising a first nucleic acid sequence and one or more introns comprising one or more acceptor splice sites from a gene of interest, wherein the first nucleic acid sequence and the intronic sequence/s are flanked by site/s which recognise a site specific recombinase.

According to the above aspect of the invention, preferably the nucleic acid is cDNA. Advantageously, the invertor cassette comprises one or more of the following components in addition to those listed above: a poly A site and a segment for the selection of homologous recombinant clones, preferably a MC1-neo-pA seqment. These regions are flanked by one or more site/s which recognise and bind the site-specific recombinase. Advantageously, the recognition sites are loxPsites and the recombinase is the Cre recombinase.

According to the method of the invention, once suitable cells (for example ES) have been targetted so that they contain the invertor cassette within the gene of interest, then chimeras may be generated by injecting such cells into blastocysts or other suitable cells. Heterozygous carrier mammals which possess the Ews-ERG invertor allele may

then be generated from these chimeras. Details are provided in the detailed description of the invention.

When the method of the invention is performed, chromosomal inversion and subsequent transcription may be performed. Those skilled in the art will appreciate that an essential feature of the method of the invention is that transcription of the fusion oncogene cannot occurr until inversion has been performed. In turn inversion step is dependent upon the expression of the specific recombinase. Thus, fusion gene transcription is dependent upon the process of homologous recombination ocurring. This method therefore prevents read through transcription with the possible generation of lethal fusion proteins.

Using the invertor method described above, oncogene fusions were created in several mouse models, in particular the inventors created a mouse model of the Ews-ERG gene fusion which is commonly found in human cancers and which gene fusion has been found by the present inventors to produce lethal fusion proteins when it is attempted to generate such a mouse model by other frequently used techniques such as the translocator technique or the loxP-stop method. Details of the generation of such mouse model according to the method of the invention are provided in Example 1.

.In addition an MLL-AF4 fusion has been created by the present inventors using the invertor model. This mutation has been commonly found in leukemias. Details of the generation of such a mouse model of leukemia using the invertor model described herein are given in Example 2.

Thus, in a further aspect the present invention provides the use of the invertor method according to the invention in the preparation of a non-human model of a chromosomal rearrangement.

In a further aspect still, the present invention provides an Ews-ERG mouse model or an MLL-AF4 mouse model of a human chromosomal gene rearrangement.

In a preferred embodiment of the above aspect of the invention, the Ews-ERG mouse model and/or the MLL-AF4 mouse model is obtainable, preferably obtained, using the invertor method described herein.

According to the above aspects of the invention, the Ews-ERG advantageously represents a mouse model of human sarcoma and/or leukemia. Moreover, the MLL-AF4 model advantageously represents a mouse model of adult and/or infant leukemia.

The present inventors have found that the Ews-ERG fusion protein and the MLL-AF4 fusion protein when expressed in lymphoid tissue in invertor mice prepared according to the methods of the invention plays a role in the onset and or maintenance of leukemogenesis. Specifically, the present inventors found that 25/29 mice which carried both the Ews-ERG and Rag1-|Cre genes developed leukemia associated with thymoma, whereas those displaying Ews-ERG only did not display leukemias. Morever, the inventors consider that the presence of the Ews-ERG fusion protein is crucial for malignancies in the invertor mice since such malignancies only arose in mice with both the Ews-ERG and rag1-Cre genes.

In a final aspect the invention provides the use of a non-human animal tumour model according to the present invention in the testing of potential anti-tumour compounds.

Advantageously, according to the above aspect of the invention, the non-human animal tumour model is the MLL-AF4 mouse model described herein. In an alternative embodiment of the above aspect of the invention the non-human animal tumour model is the Ews-ERG mouse model described herein.

The present inventors tested the invertor approach by modelling the *Ews-ERG* fusion found in human sarcomas and in addition sought to determine if an Ews-ERG fusion can be oncogenic in a cell type in which it is not generally associated in human tumors (specifically in lymphoid cells). The inventors found that clonal T cell leukemias arose in the invertor model when Cre expression was controlled by the lymphocyte-specific *Rag1* gene, showing that Ews-ERG can invoke leukemia in lymphoid cells. The

inventors results show the efficacy of the invertor translocation mimic model as a general method for creating fully conditional mouse models and suggest that in the example studied, that EWS-ERG is promiscuous in its ability to direct the development of malignancies.

According to the above aspect of the invention, the chromosomal rearrangement may be any rearrangement, for example an insertion, deletion, inversion or translocation; advantageously, the chromosomal translocation is a reciprocal translocation.

The chromosomal rearrangement is preferably tumourigenic, that is, gives rise to tumours. The invention thus provides non-human animal models of tumour conditions associated with chromosomal rearrangements.

Preferably, the cell type-specific promoter is the *Rag1* promoter. The *Rag1* gene is known in the art.

Recombinases are known in the art and include Cre, Flp and R recombinases. The sequences of these polypeptides are known: for example, the sequence of Cre recombinase is given in GenBank GI:5690440 and GI:17016300; Flp recombinase is given in GI:173333 and GI:173335; R recombinase, from *Zygosaccharomyces rouxii*, is given in GenBank GI:5260.

The cognate sites for these recombinases are *lox* (or *loxP*) sites for Cre, *frt* sites for Flp and *Rs* sites for R. Any recombinase system may be used in the invention; advantageously, it is selected from Cre-*lox*, Flp-*frt* and R-*Rs*.

Leukaemias which can be modelled by the method of the invention include all those which involve a chromosomal rearrangement; such rearrangements are very common in leukaemias. In particular, the human *Ews* gene is involved in a large number of leukaemias, and *Ews-ERG* fusions are found in human sarcomas and leukaemias. Moreover, MLL translocations ocurr in 5-10% of acute leukemias in children and

adults. Further, MLL-AF4 is the most common abheration in infant leulaemia and is associated with poor treatment outcome.

Sites recognised by the recombinase are inserted into the genome of the animal in positions such that recombination between the sites will give rise to the desired rearrangement. Rearrangements include deletions, inversions, duplications and both reciprocal and non-reciprocal translocations. Advantageously, the rearrangement is a translocation, preferably a reciprocal translocation, in which parts of non-homologous chromosomes exchange positions in the rearrangement.

The recombinase, which is advantageously Cre, Flp or R as described above, is placed under the control of a cell type-specific promoter which is advantageously the *Rag* promoter.

Advantageously, the reciprocal translocation gives rise to a *Ews-ERG* fusion.or an Mll-AF4 fusion protein as herein described.

The animal model according to the invention advantageously has leukaemia (or another tumour). It has surprisingly been found that the *Ews-ERG* fusion protein is crucial for malignancies in 'invertor mice' as malignancies only arose in mice with both the *Ews-ERG* and *Rag1-Cre* genes.

Animal models in accordance with the invention are useful in testing anti-tumour therapeutics, or compounds with a potential anti-tumour effect. Compounds may be tested for effectiveness against any type of tumour, including leukeamias and sarcomas but also other tumours which involve a chromosomal rearrangement.

In a further aspect, therefore, the invention provides a method for testing anti-tumour compounds, comprising exposing an non-human animal model in accordance with the

present invention to said compounds and assessing the incidence of tumours in said model.

According to the above aspect of the invention, advantageously the non-human animal model is either an MLL-AF4 mouse model obtained using the invertor method described herein or the Ews-ERG mouse model obtained using the invertor model described herein.

In a further aspect the present invention provides a method for the testing of an anti-tumour agent which method comprises the steps of:

(a) providing a non-human animal tumour model according to the invention,

(b) administering to that model one or more potential anti-tumour agents; and

(c) testing for a change in one or more characteristics of tumour/s comprising the model according to step (a).

In a preferred embodiment of the above aspect of the invention step (c) (testing for a change in one or more characteristics of tumour/s comprising the model according to step (a)) involving the testing for a change in any one or more characteristics of one or more tumours in the group consisting of the following: size, morphology, presence and/or levels of one or more tumour markers and cytological assays.

DEFINITIONS.

In general, an 'animal model' is one in which pathological processes (for example, cancer) can be investigated, and in which the phenomenon in one or more respects resembles the same phenomenon in humans or other species of animals. An animal model, as referred to herein, is a model based on a whole organism or part thereof such as a tissue which possesses cellular organisation which replicates that seen in the whole organism. Thus, "animal model" does not include cell-based assays performed on single cells or cultures of cells. Preferably, "animal model" refers to complete animals. The animals are non-human animals, and may be mammals, birds, repltiles, fish etc. Mammals are preferred. Rodent models are convenient and within the scope

of the present invention; mice and rats are preferred rodent models. Other models include canine, feline, monkey and primate models.

A "chromosomal rearrangement" is a change to the structure of a chromosome which leads to a genetic change in the subject. Rearrangements include deletions, where DNA is lost from the chromosome, insertions, inversions, duplications and translocations. Deletions can be located within a chromosome (interstitial) or can remove the end of a chromosome (terminal). Deletions can be small (intragenic), affecting only one gene, or can span multiple genes (multigenic). A deletion is a rearrangement that results in an increase in copy number of a particular chromosomal region. In tandem duplications, the duplicated regions lie right next to one another, either in the same order or in reverse order. In non-tandem duplications, the repeated regions lie far apart on the same chromosome or on different chromosomes. An inversion is a rearrangement in which a chromosomal segment is rotated 180 degrees. Inversions in which the rotated segment includes the centromere are called pericentric inversions; those in which the rotated segment is located completely on one chromosomal arm and do not include the centromere are called paracentric inversions. A translocation is a chromosomal rearrangement in which part of one chromosome becomes attached to a non-homologous chromosome (non-reciprocal), or in which parts of two non-homologous chromosomes exchange places (reciprocal).

A "site-specific recombinase" is an enzyme that catalyses site-specific recombination of nucleic acid between paired sets of cognate recognition sequence sites. Examples are set forth above. Cre Recombinase is a Type I topoisomerase from bacteriophage P1 that catalyses the site-specific recombination of DNA between *loxP* sites. Recombination products depend on the location and relative orientation of the *lox* (or *loxP*) sites. Two DNA species containing single *lox* sites will be fused. DNA between directly repeated *lox* sites will be excised in circular form while DNA between opposing *lox* sites will be inverted with respect to external sequences. Members of the Cre family cleave DNA substrates by a series of staggered cuts, during which the protein becomes covalently linked to the DNA through a catalytic tyrosine residue at the carboxy end of the recombinase motif.

A "cell type-specific promoter" is a promoter which directs expression of a nucleic acid in a cell type-specific manner; i.e. such that the said expression essentially occurs in certain tissues only. Many cell type-specific promoters are known in the art, and haematopoietic cell type-specific promoters are also known. The *Rag1* promoter is preferred in the present invention. Cell type-specific promoters may "leak", that is give rise to gene expression in cells other than the target cell type. This is acceptable, as long as sufficient levels of recombinase are achieved in the desired cell type to direct the rearrangement event therein to a sufficient level for the biological effect of the rearrangement to manifest itself.

Sites recognised by a recombinase are the cognate recognition sites. For example, Cre recognises *lox* or *loxP* sites. The *loxP* recognition element is a 34 bp sequence comprised of two 13 bp inverted repeats flanking an 8 bp spacer region which confers directionality. Other recognition sites, such as *frt* and *Rs* sites, are known in the art.

A "tumour", or neoplasm, is a growth of cells resulting from abnormal regulation of cell growth leading to excessive mitotic division. Tumours may be benign, that is slow-growing and usually harmless, or malignant, that is fast growing and with a tendency to metastasise to other tissues. A "tumourigenic" event is an event that leads to or enhances tumour formation or growth.

A 'leukaemia', broadly speaking, is a malignant proliferation of haematopoietic cells, characterised by replacement of bone marrow by neoplastic cells. The leukaemic cells usually are present in peripheral blood, and may infiltrate other organs of the reticuloendothelial system, such as liver, spleen and lymph nodes. Leukaemia is broadly classified into acute and chronic leukaemia, with multiple distinct clinicopathologic entities subclassified in each category. Leukaemias are divided into four main types: acute myelogenous (or myeloid) leukaemia (AML), chronic myelogenous (CML), acute lymphocytic (ALL), and chronic lymphocytic (CLL).

BRIEF DESCRIPTION OF THE TABLES AND FIGURES.

Table 1: Frequency of targeted clones and chimaera generation with *Ews* knock-in ES clones.

Ews knock-in targeted clones were made using homologous recombination in ES (CCB) cells and successful targeting established by growth on G418 (due to expression of the *neo* resistance gene) and gancyclovir (homologous events deleting the *Hsv-tk* gene from the targeting vector).

1A: The number of targeted clones identified from the indicated number of colonies analysed. *Ews* cassette control represents the knock-in of pC2A-neo (Figure 5 online), lacking additional cDNA sequences, into *Ews* exon 7. *Ews-AF9, Fli1, Chop* and *ATF* represent respectively knock-ins of fusion cDNAs as shown in Supplementary Figure 2. 1B: The number of chimaeras (identified by coat colour) resulting from injection of *Ews*targeted ES clones into blastocysts following transfer to recipient females.

Table 2: Tumor characteristics of *Ews-ERG* invertor mice expressing Cre recombinase from the *Rag1* gene.

A cohort of 29 *Ews-ERG; Rag1-Cre* mice was established and monitored for ill health. 25 of these mice developed leukemia. The main post-mortem characteristics (including the age in days to disease manifestation) are shown. The status of the *ERG* cassette was determined by filter hybridization to ascertain that the transcription orientation was 5'->3' (inv) with respect to the *Ews* gene in the thymoma DNA; + = *ERG* cassette inversion found in thymoma DNA and - = *ERG* cassette inversion not found in thymoma DNA. T cell receptor β gene (TCRJβ2$_{36}$) and IgH (heavy chain intron μ enhancer probe $_{37}$) rearrangements were determined by filter hybridization of thymoma DNA and the status of each allele is indicated as R = rearranged; G = germ-line. The cell surface protein expression profile of the cells present in the thymomas was determined using flow cytometry. Pairs of markers were used (Thy1 and B220 or CD4 and CD8) and the data summarised for the latter.

Table 3: Sequences of rearranged TCR Jβ loci in *Ews-ERG* thymoma DNA.

Thymoma DNAs were prepared from *Ews-ERG;Rag1-Cre* mice and the V-D-Jβ2 rearrangements were analysed (Fig. 4 and summary Table 2). The rearranged V-D-Jβ2

segments were amplified by PCR using primer pairs previously described [38,39] and the pools of Vβ primers (as indicated) with the single Jβ2 primer. Amplification was assayed using agarose gel electrophoresis (not shown) and the PCR products were directly sequenced using the Jβ2N primer [39]. Assignment of sequences was accomplished by comparison with those in the IGMT database [40].

V, D or J segments are underlined and nucleotides between represent N-region additional nucleotides. P = productive V-D-J join; NP = non-productive V-D-J join resulting in out of frame joint.

Table 4: shows the Tumour characterisitcs of the Ews-ERG invertor mice expressing Cre Recombinase from the Rag1 gene..

Figure 1: Strategy for generating *Ews-ERG* invertor gene by homologous recombination.

A. A partial restriction map of the mouse *Ews* gene including the XbaI-EcoRI fragment (indicated above the line) cloned from a mouse 129 genomic library and the invertor cassette comprising loxP-ERG-polyA-MC1neopA-loxP (using human *ERG* cDNA sequence, Figure 6) was cloned into the XbaI site just to the 3'side of *Ews* exon 7, to generate a targeting vector. The transcription orientation of the *ERG* cDNA cassette was opposite from that of the endogenous *Ews* gene.

B. A partial restriction map of the targeted *Ews-ERG* invertor allele. Homologous recombination in ES cells using the targeting vector (indicated above the line) resulted in insertion of the inverted *ERG* cassette just 3' of *Ews* exon 7 on one allele. Two probes for detecting homologous recombinants were located 5' and 3' of the *Ews* targeting fragment. d = donor RNA splice site (from *Ews* exon 7); a = acceptor RNA splice site (from the *ERG* invertor cassette, origin *Af4* intron 4; see Supplementary information). In the targeted clone, the transcription orientation for these donor and acceptor sites is 5'>3' and 3'>5' respectively.

.

C. Expression of Cre recombinase causes inversion of the invertor cassette to bring the

transcription orientation of *ERG* to be the same as *Ews* and the acceptor splice of the invertor cassette is thus upstream of *ERG* cDNA, allowing post-transcriptional RNA splicing from the donor site of *Ews* exon 7 to this acceptor site associated with *ERG*.

R = EcoRI; K = KpnI; H = HindIII; RV = EcoRv; X = XbaI; S = SacI; B = BamHI; Bg =BglII; d = donor splice site; a = acceptor splice site

D. Genomic sequence adjacent to the *Ews* exon 7 and the *ERG* invertor cassette (the derived amino acid sequence is shown in the single letter code) following Cre-mediated inversion obtained from DNA of ES cells and of thymus after Cre expression. The boxed sequence (TCTAG/CGAT) corresponds to the ligation of filled-in XbaI (*Ews* genomic) and ClaI (*ERG* invertor cassette) sites. The boxed XbaI site (TCTAGA) corresponds to the position of cloning of the mouse *Af4* intronic sequence (shaded) in the *ERG* invertor cassette. On the 3' side of the fused XbaI-ClaI sites is a *loxP* site of the *ERG* invertor cassette (highlighted in yellow; the second site resides at the 3' end of the *ERG* cassette; see Figure 7) followed by a partial sequence (shaded) of the *Af4* intron 4 (note this sequence is noncontiguous and the dots represent the gap; the full sequence appears in Figure 7) upstream of the *ERG* cDNA sequence, and up to the *Af4* acceptor splice site. This is followed by the *ERG* cDNA sequence from the pC2A transfer vector (see Methods and Figure 7). The NotI site derived from the pC2A vector is boxed.

^ represents splice donor of *Ews* exon 7.

E: Nucleotide sequence, and derived protein sequence, of the junction between *Ews* and *ERG* sequences in the fusion mRNA obtained by RT-PCR of mRNA made from *Ews-ERG* ES cells transiently transfected with a *Cre* expression vector or from spleen RNA of *Ews-ERG; Rag1-Cre* mice. The underlined sequence comes from mouse *Ews*, the highlighted sequence from the cassette used to construct the intron-*ERG*-polyA sequence (see Figs. 5 and 8) and the plain sequence is the human *ERG* cDNA fused to *Ews*. The NotI site derived from the pC2A vector is boxed.

Figure 2: Leukemia incidence and characteristics of *Ews-ERG* invertor mice.

Cohorts of 29 *Ews-ERG;Rag1-Cre* mice and 20 *Ews-ERG* control mice were established and monitored over a period of ~17 months. Individual mice were culled

and post-mortem conducted when signs of ill health were observed. Leukemia/lymphoma was established by various criteria (see Methods).

Top left panel. Survival curve of *Ews-ERG; Rag1-Cre* mice as a function of time (days).

Top right panel. Histology of blood and bone marrow from leukemic mice (M21 and M20 respectively). Blood smears were stained with MGG and photographed at 400X magnification, whilst bone marrow was photographed at 100X.

Bottom panel. Histological sections of tissues from mice with thymoma, illustrating three levels of severity of leukemic infiltration. Tissues were fixed in 10% formalin and wax embedded, followed by sectioning and staining with haematoxylin and eosin. Mouse M9 had the most severe phenotype, with thymus and spleen heavily disrupted by tumor cells and marked peri-vascular deposits of leukocytes in liver and kidney. Mouse M21 had severe alterations to thymus and spleen but little involvement of liver or kidney. M6 has the least severe phenotype, with only its thymus being involved with an abnormal structure.

Sections with 0.4μm and photographed as 400X times.

Figure 3: Cell surface analysis of thymomas from *Ews-ERG* invertor mice.

Thymus tissue was resected from mice with thymoma and FACS analysis performed to determine T cell phenotype (summarised in Table 2). The data in the figure shows representative flow diagrams of three tumor-bearing mice (M2, M3 and M4) compared with a wild type C57BL/6 control, using anti-Thy1 (y axis) plus anti-B220 (x-axis) antibodies or using anti-CD4 (y-axis) plus anti-CD8 (x-axis) antibodies. The three thymomas show a range of CD4 and CD8 co-expression phenotypes characterizing the thymomas generated in the *Ews-ERG; Rag1-Cre* invertor mice.

Figure 4: *Ews-ERG;Rag1-Cre* invertor mice have clonal T cell leukemias.

Genomic DNA was prepared from various tissues of invertor mice with thymomas. Genomic analysis was carried out using filter hybridization to assess the presence of the inverted *ERG* cassette in the tumor cells (A), and to assess whether the lymphocytes involved in the leukemias were clonal T (B) or B cells (C).

A. Inversion hybridization autoradiograph: DNA was prepared from the thymoma and other tissues of M18, cleaved with EcoRI and hybridised to the 5' *Ews* probe (which detects a 5Kb targeted *ERG* cassette fragment or a 6.5Kb Cre-inverted fragment). If the *ERG* cassette is inverted by Cre activity, the size of the EcoRI fragment increases from the initial targeted gene size, as indicated in the maps below the figure. The data shown are for DNA made from spleen (spl), thymus (thy), liver (liv), kidney (kid) or tail of invertor mouse M18 (ES cell DNA is used as a control). The Cre-mediated inverted band (~6.5Kb) is evident in thymus DNA (thymoma) but not the other tissues. The summary of the data from the cohort of invertor mice is in Table 2.

Restriction fragment sizes are represented as GL = germline; inv = inverted allele of *ERG* cassette; Tgt; = initial targeted *Ews* allele. The organisation of the targeted *Ews* allele is indicated underneath the hybridization autoradiograph showing the location of *Ews* exon 7 and the initial targeted orientation (bottom) or inverted orientation of *ERG* invertor cassette after Cre-mediated recombination (top).

a = acceptor splice site

B. Autoradiograph showing rearrangement of T cell receptor β locus: A TCR Jβ2 probe 36 (diagrammatically shown below the autoradiograph) detects a 5Kb germline HindIII Jβ2 band whilst V-D- Jβ2 joins in T cells result in new HindIII sized bands depending on the nature of the rearrangement. Twelve thymoma DNAs were compared and each showed one or two Jβ2 alleles rearranged, signifying that these tumors were clonal T cells. In some thymoma samples, there is almost complete absence of the germline band indicating that the thymuses of these mice is solely comprised of malignant clonal T cells. DNA sequence analysis of the V-D-J junctions of mice M2, M5, M13, M18 showed functional VDJ joins (see Table 3).

C. Autoradiograph showing rearrangement status of immunoglobulin heavy chain genes: A Cμ intron probe 37 (diagrammatically shown below the autoradiograph) was used to hybridize a set of thymoma DNAs for the presence of Igh rearranged bands. Only two samples showed rearrangements.

D. Detection of Ews-Erg fusion protein in thymoma cells. Single cell suspensions were made of T cells from a normal thymus (wt) or from the thymoma of M18. 18μg protein was solubilised in SDS-loading buffer, fractionated on 4-20% acrylamide gel, transferred to nitrocellulose membrane and incubated with anti-Ews or anti-ERG

antibody. Bound antibody was detected with anti-rabbit or anti-goat HRP respectively and exposed to X-ray film by ECL (Amersham).

Figure 5: Targeting cassettes for knock-in or invertor cDNA cloning

A: Partial restriction maps of pC2-neo, pC2A-neo and pC2A-frt-neo-frt transfer vectors for inserting cDNA sequences for knock-in. These vectors were made from pBSpt-C2 (B) or pBSpt-C2A (C). For knock-in constructs, cDNA was PCR amplified using primers with NotI sites added to the ends and cloned into the unique NotI of the vectors, giving the reading frame appropriate for the site of homologous recombination in the gene of interest. The resulting plasmids can be digested with either SfiI (pC2) or BglII (pC2A), for insertion into an appropriate targeting vector made from 129 mouse DNA for homologous recombination in ES cells. Each transfer cassette includes the SV40 polyA sequence (shown in D) cloned between NotI and BamHI sites of pBSpt-C2 or pBSpt-C2A, MC1-neo-pA or PGK-neo-pA fragments, for selection of ES clones in G418, cloned as BamHI fragments. A cassette similar to pC2A-neo is available with the PGK-hygromycin gene sequence as a BglII fragment cloned into the BamHI site as above (thereby destroying both BamHI and BglII sites but leaving the BglII sites at the end for knock-in cloning) instead of the MC1-neo-pA cassette (pC2A-hygro)

pC2Afrt-neo-frt also has a PGK-neo fragment flanked *frt* sites cloned into the BamHI site as above instead of the MC1-neo-pA cassette (pC2-frt-neo-frt). This allows deletion of the neomycin selectable marker gene using the FLP recombinase after homologous recombinants have been identified in ES cells. A similar cassette has *loxP* sites flanking the MC1-neomycin cassette.**B:** Polylinker of intermediate vector pBSpt-C2. The clone was constructed by cloning an oligonucleotide pair into the NotI-BamHI sites of pBSpt (and at the same time, destroying the original NotI site in pBSpt) to create an SfiI-NotI polylinker to give pBSpt-C2.

C: Polylinker of intermediate vector pBSpt-CA2: The clone was made as above but using the inserted oligoncucleotide sequence shown with BglII sites on either side of the NotI-BamHI sites.

D: The *SV40* polyA signal was PCR amplified from SV40 DNA and cloned into the NotIBamHI sites of either cassette pBSpt-C2 or pBSpt-C2A and the MC1-neo-polyA clone 1 cloned into the BamHI site. The sequence of the PCR fragment made from

SV40 DNA is shown comprising the dual polyadenylation signals (highlighted in yellow). The primers used for amplification are also highlighted and included NotI or BamHI sites for cloning into pBSpt-C2 or pBSpt-C2A vectors.

Figure 6: *Ews* knock-in targeting clones

cDNA sequences were cloned as NotI PCR fragments into pC2A-neo (Fig. 1) for generation of targeted knock-in *Ews* genes in ES cells. *Fli1* cDNA was prepared from MEL RNA (A), *AF9* cDNA was prepared from HEL RNA (B), *Chop* from 3T3L1 RNA (C), *ATF* from Jurkatt RNA (D) and *ERG* from COLO320 RNA (E) (latter was used for the invertor cassette). The NotI restriction sites (lower case) at each end of the PCR fragments were added as part of the PCR primers, for cloning into the NotI site of pC2A-neo.

F. The fusion junction of *Ews* exon 7 BamHI site and the BamHI/BglII junction of *Ews* and *Fli1* (as an example)

All the derived protein sequences are shown as single letter code.

Figure 7: Strategy for construction of the invertor cDNA cassette.

A. Organisation of the floxed *ERG* cDNA inversion cassette. The *ERG* invertor cDNA cassette is depicted (middle line) as it would be in the genome following homologous recombination and Cre-mediated inversion. From the 5' end, there is a loxP site (sequence shown in D), the *Af4* intron (sequence shown in E), *ERG* cDNA (sequence shown in Fig S2E), the *SV40* pA (sequence shown in Figure 5D online), MC1-neo-pA 1 and a 3' loxP site (sequence shown in F), The top line of A illustrates the orientation in which the cassette is targeted after homologous recombination, adjacent to *Ews* exon 7 (which is indicated with the *Ews* ORF and the donor splice site location). The bottom line indicates the relationship between the *Ews* exon 7 ORF and donor splice sequences and the *ERG* invertor cDNA cassette ORF and acceptor splice sequences after Cre-dependent inversion. For general use, any invertor cDNA cassette may be constructed in either pC2 or pC2A vector.

B. Sequence of pMG1 vector polylinker. This vector was adapted from pBSpt by replacing the polylinker sequence as indicated to include a single site for the rare cutter SceI restriction endonuclease (TAGGGATAA/CAGGGTAAT).

C. Sequence of pMG2 vector polylinker. This vector was adapted from pBSpt by replacing the polylinker sequence as indicated to include two sites for the SceI restriction endonuclease.

D. Sequence of the 5' *loxP* site of the cDNA invertor cassette. The *loxP* site (highlighted in yellow) was cloned from pPGK-loxP-hygro 2 as a 120bp EcoRI fragment (filled in ends) into the filled-in NotI site of pMG2, thereby removing both sites.

E. Sequence of the mouse *Af4* gene intron 4. The indicated *Af4* gene segment was amplified from 129 DNA using the primers (underlined) including XbaI and BamHI restriction sites for cloning into the *loxP* vector indicated in D.

F. Sequence of the 3' *loxP* site of the cDNA invertor cassette. The *loxP* site (highlighted in yellow) was cloned from pPGK-loxP-hygro 2 as a 120bp EcoRI fragment (filled in ends) into the EcoRV site of the pMG2-derived vector with *loxP-Af4* intron (D, E). This gave the pLOXKI clone (depicted in G).

The *ERG* cDNA transfer cassette (in pC2A and comprising *ERG* cDNA, *SV40* pA and MC1-neo sequences as a BglII fragment, see Figs. 5 and 6) was cloned into the unique BamHI site of the pLOXKI invertor cassette. This final *ERG* invertor cassette was cleaved with SceI and ClaI, end-filled and cloned into the XbaI (end-filled) of *Ews* genomic fragment downstream of exon 7 for homologous recombination (see Fig. 1).

Figure 8: Cassettes for knock-in of the *Cre* recombinase gene into targeting vectors for gene specific expression.

Three vectors are available for the knock-in of *Cre* (or of *FLP*) recombinase genes into target genes of interest (illustrated in D).

A: In each case, the *Cre* cDNA sequence was cloned as a NotI flanked PCR fragment into the NotI of pC2A vectors giving the ORF shown (the single letter code is used for protein and those residues in upper case correspond to the N-terminus of Cre, including the added nuclear localisation signal).

B: To create a *Cre* knock-in into the coding sequence of any gene of interest, a mutation would normally be required just 3' of the initiation codon of the target gene for insertion of a pC2A-Cre clone in frame. In the example here, an in-frame knock-in

of *Cre* into *Rag1* 3 was achieved by mutagenesis of *Rag1* to include a BamHI site adjacent to the *Rag1* initiator converting the sequence as shown.

C: *Cre* was cloned into the pC2A-Cre clone digested with NotI and the cassette cleaved with BglII and cloned into the *Rag1* mutated BamHI site. This yielded in-frame knock-in of *Cre* into *Rag1* which produces a fusion protein comprising two amino-acids of Rag1 and the whole Cre sequence, including the nuclear localisation signal.

D: Three Cre transfer vectors are available as shown. In the case of pC2A-Cre-frt-neo-frt, the frt-neo-frt portion (frt-PGK-neo-frt) was cloned as (XhoI-SalI) from pfrt-neo-frt (a gift from Dr A. McKenzie) into the filled in BamHI site of pC2A (note: a single BglII site in pfrt-neofrt was removed by linearization with BglII, repairing the cohesive ends followed by religation to form circular plasmid). *Cre* was cloned as the NotI PCR fragment. NOTE : If required *FLP* recombinase can be cloned into pC2A-lox-neo-lox using PCR to clone the relevant coding region of *FLP*, facilitating knock-in of *FLP* into genes of interest.

Fig 9: The invertor strategy-An MLL-AF4 fusion gene model.

Fig 10: Leukemia incidence in MLL-AF4 invertor using different Cre expressing genes.

Fig 11: Cre dependent inversion of MLL-AF4 gives a functional MLL-AF4 fusion mRNA.

Fig 12: MLL-AF4 have a myeloid phenotype but have Ig gene rearrangements.

Fig. 13: Mouse models are useful for pre-clinical testing of drugs prior to use in patients.

Fig 14: Strategy for generating *Ews-ERG* invertor gene by homologous recombination.

(A) The method for making invertor mice is described in detail elsewhere [35]. In

summary, an invertor cassette comprising a short intronic region with an acceptor splice site, human *ERG* cDNA sequence, a polyA site and the *MC1neopA* gene all flanked by *loxP* sites (depicted by black triangles) was knocked in, using homologous recombination, into *Ews* intron 8. The transcription orientation of the targeted *ERG* cDNA invertor cassette was opposite from that of the endogenous *Ews* gene after initial homologous recombination. Germline carrier mice of this targeted allele were crossed with *Cre*-expressing mouse strains [36] and, as indicated in the right hand side of the diagram, the invertor cassette is turned around to create a transcription orientation identical with that of *Ews*. Thus after transcription, a pre-mRNA is made with the donor splice site of *Ews* exon 7 adjacent to the acceptor site of the invertor cassette allowing post-transcriptional fusion of *Ews* with *ERG* in an analogous format to that found in human sarcomas with t(21;22).

(B) Genomic sequence adjacent to the *Ews* exon 7 and the *ERG* invertor cassette (the derived amino acid sequence is shown in the single letter code) obtained from DNA of ES cells and of thymus after Cre-mediated inversion. The *Ews* intron 7/8 donor splice site is indicated by an arrow. The boxed sequence (TCTAG/CGAT) corresponds to the ligation of filled-in XbaI (*Ews* genomic) and ClaI (*ERG* invertor cassette). The boxed XbaI site (TCTAGA) corresponds to the position of cloning of the mouse *Af4* intronic sequence (shaded) in the *ERG* invertor cassette. On the 3' side of the fused XbaI-ClaI sites, there is a *loxP* site originating from the *ERG* invertor cassette followed by the *Af4* intron 4 (shaded) upstream of the *ERG* cDNA sequence. The *Af4* acceptor splice site is indicated by an arrow, and is followed by the *ERG* cDNA sequence. A NotI site used for cloning the amplified *ERG* sequence is boxed (note this sequence is non-contiguous and the dots represent a gap;

Figure 15. Incidence of haematological malignancy and characteristics of *Ews-ERG* invertor mice.

Cohorts of 29 *Ews-ERG;Rag1-Cre* mice and 20 *Ews-ERG* control mice were monitored over a period of ~17 months. Mice were culled and post-mortem conducted when signs of ill health were observed. Leukaemia/lymphoma was established by various criteria (see Methods).

Top panel. Survival curve of *Ews-ERG; Rag1-Cre* (+Cre) or *Ews-ERG* (-Cre) mice as

a function of time (days).

Bottom panel. Histology of blood and bone marrow from leukaemic mice (M21 and M20 respectively). Blood smears were stained with MGG and photographed at 400X magnification, whilst bone marrow was photographed at 1000X.

Figure 16. Analysis of cell surface antigens of thymomas from *Ews-ERG* invertor mice using flow cytometry.

Thymus tissue was resected from mice with thymoma and FACS analysis performed to determine T cell phenotype. The data in the figure shows representative flow diagrams of three tumour-bearing mice (M2, M3 and M4) compared with a wild type C57BL/6 control, using anti-Thy1 (y axis) plus anti-B220 (x axis) antibodies or using anti-CD8 (y axis) plus anti-CD4 (x axis) antibodies. The three thymomas show a range of CD4 and CD8 co-expression phenotypes characterizing the thymomas generated in the *Ews-ERG; Rag1-Cre* invertor mice.

Figure 17. Clonality of T cell neoplasias in *Ews-ERG* invertor mice

Genomic DNA was prepared from various tissues of invertor mice with thymomas. Genomic analysis was carried out using filter hybridization to assess the presence of the inverted *ERG* cassette in the tumour cells (**A**), and to assess whether the lymphocytes involved in the tumours were clonal T (**B**) or B cells (**C**).

(A) Inversion hybridization autoradiograph: DNA was prepared from the thymoma and other tissues of M18, cleaved with EcoRI and hybridized to the 5' *Ews* probe (which detects a 5Kb targeted *ERG* cassette fragment or a 6.5Kb Cre-inverted fragment). If the *ERG* cassette is inverted by Cre activity, the size of the EcoRI fragment increases from the initial targeted gene size, as indicated in the maps below the figure. The data shown are for DNA extracted from spleen (spl), thymus (thy), liver (liv), kidney (kid) or tail (ES cell DNA is used as a control). The Cre-mediated inverted band (~6.5Kb) is evident in thymus DNA (thymoma).. Note that despite extensive infiltration of spleen detected by histology, we cannot see the inverted band by Southern blot; this presumably reflects regional clustering of neoplastic cells in spleen.

Restriction fragment sizes are represented as GL = germline; inv = inverted allele of

ERG cassette; Tgt = initial targeted *Ews* allele.

The organization of the targeted *Ews* allele is indicated underneath. The hybridization autoradiograph showing the location of *Ews* exon 7 and the initial targeted orientation (bottom) or inverted orientation of *ERG* invertor cassette after Cre-mediated recombination (top).

a = acceptor splice site

(B) Autoradiograph showing rearrangement of T cell receptor β locus: A TCR Jβ2 probe (diagrammatically shown below the autoradiograph) detects a 5Kb germline HindIII Jβ2 band whilst V-D- Jβ2 joins in T cells result in new HindIII sized bands depending on the nature of the rearrangement. Twelve thymoma DNAs were compared and each showed one or two Jβ2 alleles rearranged, signifying that these tumours were clonal T cells. In some thymoma samples, there is almost complete absence of the germline band indicating that the thymuses of these mice are solely comprised of malignant clonal T cells. DNA sequence analysis of the V-D-J junctions of mice M2, M5, M13, M18 showed functional VDJ joins..

(C) Autoradiograph showing rearrangement status of immunoglobulin heavy chain genes: A Cμ intron probe. (diagrammatically shown below the autoradiograph) was used to hybridize a set of thymoma DNAs for the presence of Igh rearranged bands. Only two samples showed rearrangements.

(D) Detection of Ews-Erg fusion protein in thymoma cells. Single cell suspensions were made of T cells from a normal thymus (wt) or from the thymoma of M6, protein fractionated on 4-20% acrylamide gel, transferred to nylon membranes. Specific proteins were detected with anti-Ews or anti-ERG antibody.

Figure 18. B and T cells express the *Ews-ERG* fusion RNA.

A 96 day old mouse with both *Ews-ERG* and *Rag1-Cre* alleles was used as a source of spleen and thymus cells. Single cell suspensions of spleen cells were labelled with anti-B220 or with anti-thy1.2 and cells purified using a MoFlo preparative flow cytometer. Estimated purities were achieved of >95%. cDNA was prepared from RNA extracted from sorted cells or from aliquots of unsorted populations and RT-PCR (approximately 3400 B220+ or 6400 thy1.2+ cell equivalents per PCR reaction) carried out with specific *Pax5* (A), *CD3* (B) or *Ews-ERG* (C) primers. PCR reaction

products were fractionated on 1% agarose gels and either stained with ethidium bromide and photographed (A, B) or the gel blotted and hybridised with an *Ews-ERG* probe (C).

Figure 19. Invertor strategy for generating *Mll-AF4* mice.

Invertor *Mll-AF4* was made employing the strategy described previously (Codrington et al., 2005; Forster et al., 2005) with a human *AF4* cDNA sequence. In outline, a floxed invertor cassette was constructed comprising a short stretch of intron, an acceptor splice site, the *AF4* eDNA followed by a polyadenylation site (pA) and a neomycin gene for selecting homologous recombination events. This cassette was inserted into an *Mll* targeting vector at a Bg1II site and transfected into ES cells followed by identification of clones in which the invertor cassette had entered the genomic DNA in a 3'->5' orientation with respect to the endogenous *Mll* gene. ES clones were injected into blastocysts, and germline carrier mice made. The floxed invertor cassette was made such that Cre recombinase recognizes the *loxP* sites and inverts the cassette to make the transcription orientation 5'-3', tins allowing transcription of the *Mll-AF4* fusion gene.

(A) Diagram of part of the mouse *Mll* gene and of the *AF4* invertor cassette. A partial restriction map of mouse *Mll* around exons 8-10 is shown in the top line, indicating the Bg1II site into which the *AF4* invertor cassettes were introduced. Two probes (used for detecting targeted alleles) were located 5' and 3' of the respective *Mll* fragment *(Mll* 5' probe, *Mll* 3'probe). The invertor cassette is depicted below the *Mll* map and two forms were used, one flanked by wild type *loxP* sequences *(loxP)* and a second generation by the mutant *loxP* sites *(lox66* and *lox 71* (Albert et aL, 1995)). These sequences are shown below the AF4 invertor cassette.

(B) Organization of the targeted *Mll-AF4* invertor allele before Crc mediated inversion of the cassette. The floxed cassette carrying the *AF4* cDNA sequence was knocked-in to the *Mll* intron 10 but in the opposite orientation for transcription of the *Mll* gene (as depicted by the arrow beneath the map). RNA splicing from the *Mll* exon 10 donor splice site to the acceptor site (a) in front of the *AF4* cDNA insert cannot occur until the cassette is inverted by expression of Cre recombinase.

(C) Organization of the targeted *Mll-AF4* invertor allele after Cre mediated inversion

of the cassette.

Cre recombinase mediated inversion of the floxed cassette allows post-transcriptional RNA splicing from the donor site of *Mll* exon 10 to the acceptor site upstream of *AF4* cDNA, forming a fusion transcript representative of the chimeric mRNA found in human *MLL-AF4* positive leukemia.

Primers located in *Mll* exon 10 (MEF), *Af4* intron 4 (AIR, fluorescence labeled), and *AF4* exon 5 (AER) are indicated by red arrows. The genomic fusion sequence was amplified using primers MEF and AIR. Primer AIR was labeled with 5' fluorescein to quantify the inversion efficiency and Crc specificity by competitive PCR.. Primers MEF and AER were used for RT-PCR to amplify and sequence the expressed fusion mRNA.

(D) *Mll-AF4* fusion transcript and derived protein sequence (written in single letter code) obtained from spleen RNA from *Mll-AF4 lox; Cre* mice.

(E) Fusion transcript and derived protein sequences *of Mll-AF4 1ox66/71; Cre* invertor. Note that the exact sequence at the junction *of Mll* and *AF4* in the two forms of invertor mice differ slightly due to splice variation.

Sp, SphI; Xb, XbaI; R, EcoRI; H, HindIII; K, KpnI; Bg, BgIII; B, BamHI.

chr. = chromosome

Figure 20. *Mlll-AF4* invertor mice develop hematopoietic malignancy

Survival curves of *Mll-AF4* invertor mice. Cohorts of *Mll-AF4lox; Cre* mice and controls lacking either the invertor or *Cre* alleles were established and post-mortem examinations carried out at signs of ill health. Latency and frequency of disease is plotted for invertor mice with Cre recombinase expression under the control of either the *Rag-], Lck* and *CD19* promoters (McCormack et al., 2003). Incidence of lymphoid tumors was 31/36 mice (86%) for *Mll-AF4 loxP; Rag-Cre;* 25/29 (86%) for *Mll-AF4 1ox66/71; Rag-Cre* with a median latency of 466 and .317 days, respectively. Incidence was 11/12 mice (92%, mean latency 416 days) for *Mll-AF4 loxP; Lck-Cre* and 17/18 (94%, mean latency 472 days) for *Mll-AF4 1ox66/71; Lck-Cre.* Incidence was 6/7 mice (86%, mean latency 475 days) for *Mll-AF4 loxP; GD19-Cre* mice and 10/14 (71%, mean latency 460 days) *Mll-AF4 1ox66/71; CDI9-Cre* mice. Black lines represent *Mll-AF4 1ox66/71* strains, grey lines *Mll-AF4 loxP* strains and dashed lines

Mll wt; Cre strains.

Figure 21. Tumor cells infiltrating *Mll-AF4* invertor mouse tissues have gene inversion and fusion transcripts.

(A, C) Filter hybridization analysis of SphI digested genomic DNA from the indicated tissues using the *Mll* 5'probe is shown for one representative individual for each Cre expressor strain *of Mll-AF4 loxP* (panel A) and *Mll-AF4 tox66/71* (panel C) invertor mice.

The 9.7 kb fragment (inv) indicates the inversion of the invertor cassette. Signal intensity depends on the degree of tumor infiltration in the respective organ.

gl = germline allele; tg = targeted allele, inv = inverted allele; Mes. LN = Mesenteric lymph node; Axill LN. = Axillary lymph node

(B, D) Detection of *Mll-AF4* RNA using RT-PCR with primers MEF and AER confirms the presence of *Mll-AF4* fusion transcripts in affected tissues of *Mll-AF4 loxP* (B) and *Mll-AF4 lox66/71* mice (D), respectively. The integrity of mRNA is confirmed by *β-actin* amplification products.

(E) Cell type specific inversion by *Rag-Cre, Lck-Cre,* and *GD19-Cre* expression

Fluorescence labeled PCR products of the genomic fragment following inversion (primers MEF and AIR) and exon 2 of the single copy gene *Lmo2* as a control were quantified and compared in unsorted, Macl+, B220+, or Thyl.2+ sorted cell populations. Columns show mean and SEM of triplicate PCR reactions from representative DNA preparations.

Figure 22. *Mll-AF4* invertor mice develop diffuse large B cell lymphomas.

Fixed tissues from *Mll-AF4; Cre* neoplastic mice were analyzed by histology or by immunohistochemistry.

(A) Hematoxylin and eosin (H&E) stained paraffin sections (100x magnification) show disruption of spleen and liver architecture due to tumor infiltration in a representative *MllAF4 tox66/71; Rag-Cre* mouse compared to a CS7BL/6 mouse (wt). The bottom right panel shows lymphoma cells stained with the B-cell marker B220 (400x magnification).

(B) In the top row showing high magnification (1000x), infiltrates can be identified as

diffuse large cell lymphoma in each of the three different Cre expressor strains. H&E staining shows large vesicular nuclei, clumping chromatin, prominent nucleoli, and mitoses in lymphoma cells. Immunohistochemistry demonstrates the B cell phenotype (B220+) of the lymphoma. Reactive T cells (CD3+) with small regular nuclei are also present within the infiltrated areas. No myeloid cells (as judged by lack of expression of the marker F4/80) are present in the lymphoma, whereas myeloid cells in remaining red pulp of spleens stain with this marker (inset).

Figure 23. Lymphomas of *Mll-AF4* invertors have clonal immunoglobulin gene rearrangements.

Genomic DNA from lymphoma infiltrated spleens or mesenteric lymph nodes were analyzed by filter hybridization using either *Igh* (A), *Tcr Cβ1* (B) or *Tcr Jβ2* (C) probes. Examples of 2 mice of each strain are presented. Positions of the respective germline (gl) fragments are indicated by arrows. Partial restriction maps of the *JH* or *Cβ1 -Jβ2* genomic regions and location of the probe fragments are shown between the respective southern blots filter hybridizations.

(D) Results of 53 mice are summarized in the table. Each box shows the numbers of tumors with *Igh* or *Tcr* rearranged/total number in group.

H, HindIII; X, XbaI; R, EcoRI

Figure 24. *Mll-AF4* invertor tumors are transplantable in *Rag-1* null mice.

A single cell suspension was made from the mesenteric lymph node of an *Mll-AF4 1ox66/71; CD19-Cre* tumor mouse and 5x 10^6 cells injected into the tail vein of a *Rag-1 -/-* mouse. A massive lymphoma infiltration of spleen and mesenteric lymph nodes as well as liver and kidneys of the first recipient mouse occurred within 56 days and the spleen of this mouse was used as a source of cells for flow cytometric analysis, DNA preparation and re-transplantation into four secondary *Rag-1 -/-* recipients. The latter developed tumors after 17-23 days and the spleens from these mice were taken for analysis.

(A) Flow cytometry using anti-B220, anti-Thyl.2 or anti-Mac1 FITC coupled antibodies of mesenteric lymph node cells from the *Mll-AF4 1ox66/71; CDI9-Cre* mouse (top row) or spleen cells from the first Rag-/- (middle row) or one of the 4

secondary Rag-/- recipients (bottom row).

(B) DNA was prepared from the indicted tissues of the primary *Mll-AF4 1ox66/7I; CD19-Cre* mouse and from the spleens of the *Rag-1 -/-* transplant recipient mice. Filter hybridization was carried out using an *Mll* probe to detect inversion of the *AF4* cassette (top panel), an *Igh* probe (middle) or a *Tcr* probe (bottom).

Mes. LN = mesenteric lymph node; inv = inverted allele; gl = germline allele; tg = targeted *Mll* allele; rearr. = rearranged *Igh* allele

DETAILED DESCRIPTION OF THE INVENTION.

General Techniques

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, nucleic acid chemistry, hybridisation techniques and biochemistry). Standard techniques are used for molecular, genetic and biochemical methods (see generally, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. and Ausubel et al., Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc. which are incorporated herein by`\zz reference) and chemical methods. In addition Harlow & Lane., A Laboratory Manual Cold Spring Harbor, N.Y, is referred to for standard Immunological Techniques.

Generation of transgenic animals

Methods for the generation of transgenic animals are generally known to those skilled in the art and are discussed in detail in First N, Haseltine FP., Transgenic animals: Stoneham, UK: Butterworth-Heinmann, 1991; and Grosveld F, Kollias G., Transgenic animals: San Diego, California: Academic Press, 1992. Commonly used methods include the ES cell method and pronuclear microinjection. In the context of the present invention, ES cell techniques are preferred.

In this technique, embryonic stem cells (ES cells) are harvested from the inner cell mass (ICM) of animal blastocysts. They can be grown in culture and retain their full potential to produce all the cells of the mature animal, including its gametes. The cultured cells are exposed to the DNA which is intended to be inserted into the cells so that some cells will incorporate it. The cells are then tested for successful incorporation of the DNA, usually by selecting for the presence of a *neo* marker, and injected into the inner cell mass of an animal blastocyst. A proportion of the cells of the embryo resulting from the blastocyst will incorporate the desired DNA.

A pseudopregnant animal is prepared and the embryos transferred into the uterus. A proportion of the embryos will develop into animals carrying the inserted DNA.

Targeted gene insertion requires some additional steps. The vector used to transfer the DNA typically includes:

the desired gene neo^r, a gene that encodes an enzyme that inactivates the antibiotic neomycin (and its relatives)

tk, a gene that encodes thymidine kinase, an enzyme that phosphorylates the nucleoside analogue gancyclovir. DNA polymerase fails to discriminate against the resulting nucleotide and inserts this non-functional nucleotide into freshly-replicating DNA.

The culture of ES cells is treated with a preparation of vector DNA:

Most cells fail to take up the vector; these cells will be killed if exposed to neomycin.

In a few cells: the vector is inserted randomly in the genome. In random insertion, the entire vector, including the tk gene, is inserted into host DNA. These cells are resistant to neomycin but killed by gancyclovir.

In still fewer cells: homologous recombination occurs. Stretches of DNA sequence in the vector find the homologous sequences in the host genome and the region between these homologous sequences replaces the equivalent region in the host DNA.

Positive selection of transformed cells is achieved by culturing all cells in neomycin. The cells (the majority) that fail to take up the vector are killed. Negative selection of all cells (tk^+) in which the vector was inserted randomly by culturing cells surviving the neomycin selection is performed by culturing the cells in gancyclovir. This step leaves a population of cells transformed by homologous recombination (enriched several thousand fold). These cells are injected into the inner cell mass of animal blastocysts.

Methods for introduction of recombinase specific sites into the genome of a transgenic animal are described in Collins *et al.*, (2000) EMBO Reports 1:127-132.

The insertion of genes encoding a site-specific recombinase under the control of a cell type-specific promoter has been described in the art. For example, Cre recombinase has been placed under the control of promoters for:

αA-crystallin	Eye lens [Lasko *et al.*, (1992) PNAS 89:6232-6236]
Calcium/calmodulin-dependent protein kinase IIα	Forebrain (CA1 pyramidal cells) [Tsien *et al.*, (1996) Cell 87:1317-1326]
P0 gene	Schwann cells [Akagi *et al.*, (1997) NAR 25:1766-1773]
Pro-opiomelanocortin	Pituitary gland (intermediate lobe) [Akagi *et al.*, (1997) NAR 25:1766-1773]
Interphotoreceptor retinoid binding protein	Retina (photoreceptor cells) [Akagi *et al.*, (1997) NAR 25:1766-1773]
wnt-1	Nervous system [Danielian *et al.*, (1998) Curr. Biol. 8:1323-1326]
Engrailed-2	Nervous system [Zinyk *et al.* (1998) Curr Biol 8:665-668]
Proximal lck	T cells [Vooijs *et al.*, (1998) Oncogene 17:1-12]
CD19 (knock in)	B cells [Rickert *et al.*, (1997) NAR 25:1317-1318]
α myosin heavy chain	Heart (ventricular myocytes) [Agah *et al.*, (1997) J Clin Invest 100:169-179]

Myosin light chain 2v (knock in)	Heart (ventricular myocytes) [Chen *et al.*, (1998) Development 25:1943-1949;Chen *et al* ., (1998) J Biol Chem 273:1252-1256]
Muscle creatine kinase	Skeletal muscle, heart [Bruening *et al.*, (1998) Mol Cell 2:559-569; Wang *et al.*, (1999) Nat Genet 21:133-137]
Insulin	Pancreas (β cells) [Ray *et al.*, (1998 BBRC 253:65-69; Postic *et al.*, (1999) J Biol Chem 274:305-315]
Albumin enhancer/promoter	Liver [Postic *et al.*, (1999) J Biol Chem 274:305-315]
Whey acidic protein	Mammary gland, brain [Wagner *et al.*, (1997) NAR 25:4323-4330]
β-lactoglobulin	Mammary gland [Selbert *et al.*, (1998) Transgenic Res 7:387-396]
Adipose protein 2	Adipose tissue [Barlow *et al.*, (1997) NAR 25:2543-2545]
Keratin 5	Skin (basal keratinocytes) [Tarutani *et al.*, (1997) PNAS 94:7400-7405]

The gene encoding the recombinase may be inserted in the same procedure as that used to insert the cognate sites, or by a separate procedure. It can be introduced into the same animal, or a separate animal. For example, a transgenic animal carrying the recombinase gene, introduced by ES cell transformation or pronuclear microinjection, can be crossed with an animal carrying the cognate sites to produce an animal of the desired genotype.

Site-Specific Recombination

Site-specific recombinase enzymes include the Int recombinase of bacteriophage (with or without Xis) (Weisberg, R. *et al.*, in *Lambda II*, (Hendrix, R., *et al.*, Eds.), Cold Spring Harbor Press, Cold Spring Harbor, NY, pp. 211-50 (1983); TpnI and the -lactamase transposons (Mercier, *et al.*, *J. Bacteriol.*, 172:3745-57 (1990)); the Tn3 resolvase (Flanagan & Fennewald *J. Molec. Biol.*, 206:295-304 (1989); Stark, *et al.*, *Cell,* 58:779-90 (1989)); the yeast recombinases (Matsuzaki, *et al.*, *J. Bacteriol* .,

172:610-18 (1990)); the *B. subtilis* SpoIVC recombinase (Sato, *et al., J. Bacteriol.* 172:1092-98 (1990)); the Flp recombinase (Schwartz & Sadowski, *J. Molec. Biol.*, 205:647-658 (1989); Parsons, *et al., J. Biol. Chem.,* 265:4527-33 (1990); Golic & Lindquist, *Cell,* 59:499-509 (1989); Amin, *et al., J. Molec. Biol.,* 214:55-72 (1990)); the Hin recombinase (Glasgow, *et al., J. Biol. Chem.*, 264:10072-82 (1989)); immunoglobulin recombinases (Malynn, *et al., Cell,* 54:453-460 (1988)); and the Cin recombinase (Haffter & Bickle, *EMBO J.,* 7:3991-3996 (1988); Hubner, *et al., J. Molec. Biol.*, 205:493-500 (1989)). Such systems are discussed by Echols (*J. Biol. Chem.* 265:14697-14700 (1990)); de Villartay (*Nature,* 335:170-74 (1988)); Craig, (*Ann. Rev. Genet.,* 22:77-105 (1988)); Poyart-Salmeron, *et al.,* (*EMBO J.* 8:2425-33 (1989)); Hunger-Bertling, *et al.,*(*Mol Cell. Biochem.*, 92:107-16 (1990)); and Cregg & Madden (*Mol. Gen. Genet.,* 219:320-23 (1989)).

Cre recombinase has been purified to homogeneity, and its reaction with the loxP site has been extensively characterised (Abremski & Hess *J. Mol. Biol.* 259:1509-14 (1984); Gopaul and van Duyne, (1999) Curr Opin Struct Biol 9:14-20). Cre protein has a molecular weight of 35,000 and can be obtained commercially from New England Nuclear/DuPont. The *cre* gene (which encodes the Cre protein) has been cloned and expressed (Abremski, *et al., Cell* 32:1301-11 (1983)). The Cre protein mediates recombination between two loxP sequences (Sternberg, *et al., Cold Spring Harbor Symp. Quant. Biol.* 45:297-309 (1981)), which may be present on the same or different DNA molecule. Because the internal spacer sequence of the loxP site is asymmetrical, two loxP sites can exhibit directionality relative to one another (Hoess & Abremski *Proc. Natl. Acad. Sci. U.S.A.* 81:1026-29 (1984)). Thus, when two sites on the same DNA molecule are in a directly repeated orientation, Cre will excise the DNA between the sites (Abremski, *et al., Cell* 32:1301-11 (1983)). However, if the sites are inverted with respect to each other, the DNA between them is not excised after recombina-tion but is simply inverted. Thus, a circular DNA molecule having two loxP sites in direct orientation will recombine to produce two smaller circles, whereas circular molecules having two loxP sites in an inverted orientation simply invert the DNA sequences flanked by the loxP sites. In addition, recombinase action

can result in reciprocal exchange of regions distal to the target site when targets are present on separate DNA molecules.

Modelling of chromosomal rearrangements

Chromosomal rearrangements are commonplace in cancers and are efficiently modelled by the method of the invention. The use of a recombinase-catalysed rearrangement at recombinase-specific sites in the absence of any read-through transcription permits the generation of realistic models of chromosomal rearrangements. Moreover, the method of the invention permits the generation of chromosomal rearrangements commonly found in cancers which cannot be generated using prior art methods such as the translocator method devised by the present inventors and the LoxP-STOP method due to the problem of read-through transcription using these prior art methods and the consequent production of embryonically lethal fusion proteins.

Many chromosomal rearrangements are known in the art, which can be modelled in accordance with the present invention. These include:

1. Chronic Myeloproliferative Diseases (MPD)

Agnogenic myeloid metaplasia
Atypical Chronic Myeloid Leukaemia (aCML)
Chronic Myelomonocytic Leukaemia (CMML)
Chronic myelogenous leukaemia (CML)
Essential thrombocythemia
Idiopathic myelofibrosis
Idiopathic thrombocythemia
Juvenile Chronic Myelogenous Leukaemia (JCML)
Polycytemia vera
del(13q) in myeloid malignancies
del(16)(q22)

del(20q) in myeloid malignancies
del(5q) in myeloid malignancies
del(9q) solely
i(17q) in myeloid malignancies
idic(X)(q13)
ins(3;3)(q26;q21q26)
inv(16)(p13q22)
inv(3)(q21q26)
t(12;13)(p12;q12-14)
t(12;22)(p13;q11-12)
t(16;16)(p13;q22)
t(1;18)(q10;q10)
t(1;3)(p36;p21)
t(1;3)(p36;q21)
t(1;7)(q10;p10)
t(3;3)(q21;q26)
t(5;10)(q33;q21)
t(5;11)(q31;q23)
t(5;12)(q33;p13)
t(7;11)(p15;p15)
t(8;14)(q11;q32)
t(8;22)(p11;q11)
t(9;12)(p24;p13)
t(9;12)(q34;p13)
t(9;22)(q34;q11) in CML
t(Y;1)(q12;q12)

2. Myelodysplastic Syndromes (MDS)

11q23 rearrangements in leukaemia
12p abnormalities in myeloid malignancies

Childhood myelodysplastic syndromes
Chronic Myelomonocytic Leukaemia (CMML)
Congenital leukaemias
Infant leukaemias
Neonatal leukaemias
del(13q) in myeloid malignancies
del(16)(q22)
del(17p) in myeloid malignancies
del(20q) in myeloid malignancies
del(5q) in myeloid malignancies
del(9q) solely
i(17q) in myeloid malignancies
idic(X)(q13)
ins(3;3)(q26;q21q26)
inv(16)(p13q22)
inv(3)(q21q26)
t(11;16)(q23;p13)
t(11;17)(q23;q25)
t(12;13)(p12;q12-14)
t(12;22)(p13;q11-12)
t(16;16)(p13;q22)
t(16;21)(q24;q22)
t(1;16)(q11;q11)
t(1;18)(q10;q10)
t(1;19)(p13;p13.1)
t(1;3)(p36;p21)
t(1;3)(p36;q21)
t(1;7)(q10;p10)
t(2;11)(p21;q23)
t(3;12)(q26;p13)
t(3;21)(q26;q22)
t(3;3)(q21;q26)

t(5;11)(q31;q23)
t(5;12)(q33;p13)
t(5;7)(q33;q11)
t(6;8)(q27;p12)
t(6;9)(p23;q34)
t(7;12)(q36;p13)
t(8;13)(p12;q12)
t(9;12)(q22;p12)
t(Y;1)(q12;q12)

3. Treatment Related Leukaemias (t-ANLL)

11q23 rearrangements in leukaemia
11q23 rearrangements in therapy related leukaemias
12p abnormalities in myeloid malignancies
Biphenotypic Acute Leukaemia (BAL)
Childhood myelodysplastic syndromes
del(16)(q22)
del(17p) in myeloid malignancies
del(5q) in myeloid malignancies
inv(16)(p13q22)
t(10;11)(p12;q23)
t(11;12)(p15;q13)
t(11;15)(q23;q14)
t(11;16)(q23;p13)
t(11;17)(q23;p13)
t(11;17)(q23;q25)
t(11;19)(q23;p13.1)
t(11;19)(q23;p13.3)
t(11;22)(q23;q13)
t(15;17)(q22;q21)
t(16;16)(p13;q22)

t(16;21)(q24;q22)
t(17;21)(q11.2;q22)
t(18;21)(q21;q22)
t(19;21)(q13.4;q22)
t(1;21)(p36;q22)
t(1;3)(p36;p21)
t(1;3)(p36;q21)
t(1;7)(p36;q34)
t(1;7)(q10;p10)
t(3;11)(p21;q23)
t(3;11)(q25;q23)
t(3;11)(q28;q23)
t(3;21)(q26;q22)
t(4;11)(q21;q23)
t(4;12)(q11-q21;p13)
t(5;11)(q31;q23)
t(6;11)(q27;q23)
t(6;9)(p23;q34)
t(8;16)(p11;p13)
t(8;21)(q22;q22)
t(9;11)(p22;q23)

4. Acute Non Lymphocytic Leukaemias (ANLL)

11q23 rearrangements in leukaemia
12p abnormalities in myeloid malignancies
Acute Erythroid leukaemias
Acute basophilic leukaemia
Biphenotypic Acute Leukaemia (BAL)
Congenital leukaemias
Infant leukaemias
M0 acute non lymphocytic leukaemia (M0-ANLL)

Neonatal leukaemias
Systemic mast cell disease (SMCD)
del(13q) in myeloid malignancies
del(16)(q22)
del(17p) in myeloid malignancies
del(20q) in myeloid malignancies
del(5q) in myeloid malignancies
del(9q) solely
idic(X)(q13)
ins(3;3)(q26;q21q26)
inv(14)(q11;q32.1)
inv(16)(p13q22)
inv(3)(q21q26)
t(10;11)(p11.2;q23)
t(10;11)(p12;q23)
t(10;11)(p13;q21)
t(11;16)(q23;p13)
t(11;17)(q13;q21)
t(11;17)(q23;p13)
t(11;17)(q23;q12)
t(11;17)(q23;q21)
t(11;17)(q23;q25)
t(11;19)(q23;p13.1)
t(11;19)(q23;p13.3)
t(11;22)(q23;q11.2)
t(12;13)(p12;q12-14)
t(12;22)(p13;q11-12)
t(14;14)(q11;q32.1)
t(15;17)(q22;q21)
t(16;16)(p13;q22)
t(16;21)(p11;q22)
t(16;21)(q24;q22)

t(17;21)(q11.2;q22)
t(1;11)(q21;q23)
t(1;12)(q25;p13)
t(1;18)(q25;q23)
t(1;19)(p13;p13.1)
t(1;19)(q23;p13)
t(1;22)(p13;q13)
t(1;3)(p36;p21)
t(1;3)(p36;q21)
t(1;7)(p36;q34)
t(1;7)(q10;p10)
t(2;11)(p21;q23)
t(3;12)(q26;p13)
t(3;21)(q26;q22)
t(3;3)(q21;q26)
t(3;5)(q25;q34)
t(4;11)(q21;q23)
t(4;12)(q11-q21;p13)
t(5;11)(q31;q23)
t(5;11)(q35;p15.5)
t(5;14)(q33;q32)
t(5;17)(q35;q21)
t(6;11)(q27;q23)
t(6;8)(q27;p12)
t(6;9)(p23;q34)
t(7;11)(p15;p15)
t(7;12)(q36;p13)
t(8;16)(p11;p13)
t(8;21)(q22;q22)
t(9;11)(p22;q23)
t(9;11)(q34;q23)
t(9;12)(q34;p13)

t(9;22)(q34;q11) in ANLL

t(X;11)(q13;q23)

t(X;6)(p11;q23)

5. B-Cell Acute Lymphocytic Leukaemias (B-ALL)

11q23 rearrangements in leukaemia

12p rearrangements in ALL

9p Rearrangements in ALL

Biphenotypic Acute Leukaemia (BAL)

Congenital leukaemias

High hyperdiploid acute lymphoblastic leukaemia

Infant leukaemias

Near haploid acute lymphoblastic leukaemia

Neonatal leukaemias

Severe hypodiploid acute lymphoblastic leukaemia

del(6q) abnormalities in lymphoid malignancies

dic(9;12)(p11-13;p11-12)

dic(9;20)(p11-13;q11)

ins(5;11)(q31;q13q23)

inv(14)(q11;q32.1)

t(10;11)(p12;q23)

t(10;11)(p13;q21)

t(11;17)(q23;p13)

t(11;17)(q23;q25)

t(11;19)(q23;p13.3)

t(12;13)(p12;q12-14)

t(12;21)(p12;q22)

t(14;14)(q11;q32.1)

t(14;18)(q32;q21)

t(17;19)(q22;p13)

t(18;22)(q21;q11)

t(1;18)(q25;q23)
t(1;19)(q23;p13)
t(1;3)(p36;p21)
t(2;11)(p21;q23)
t(2;14)(p13;q32)
t(2;18)(p11;q21)
t(2;8)(p12;q24)
t(4;11)(q21;q23)
t(4;12)(q11-q21;p13)
t(5;11)(q31;q23)
t(5;14)(q31;q32)
t(5;15)(p15;q11-13)
t(8;14) (q24;q11)
t(8;14)(q11;q32)
t(8;14)(q24;q32)
t(8;22)(q24;q11)
t(9;12)(p24;p13)
t(9;12)(q34;p13)
t(9;22)(q34;q11) in ALL

6. T-Cell Acute Lymphocytic Leukaemias (T-ALL)

11q23 rearrangements in leukaemia
12p rearrangements in ALL
1p32 rearrangements
9p Rearrangements in ALL
Biphenotypic Acute Leukaemia (BAL)
del(6q) abnormalities in lymphoid malignancies
del(9q) solely
dic(9;12)(p11-13;p11-12)
inv(14)(q11;q32.1)
t(10;11)(p13;q21)

t(10;14)(q24;q11)
t(11;14)(p13;q11)
t(11;14)(p15;q11)
t(11;19)(q23;p13.3)
t(12;13)(p12;q12-14)
t(14;14)(q11;q32.1)
t(1;14)(p32;q11)
t(1;19)(q23;p13)
t(1;3)(p32;p21)
t(1;5)(p32;q31)
t(1;7)(p32;q34)
t(2;14)(p13;q32)
t(4;11)(q21;p15)
t(4;11)(q21;q23)
t(5;14)(q35;q32)
t(5;17)(q13;q21)
t(6;11)(q27;q23)
t(7;10)(q34;q24)
t(7;11)(q35;p13)
t(7;9)(q34;q32)
t(8;14) (q24;q11)
t(9;12)(p24;p13)
t(9;22)(q34;q11) in ALL
t(X;11)(q13;q23)

7. Non Hodgkin Lymphomas (NHL)

3q27 rearrangements in non Hodgkin lymphoma
Anaplasic large cell lymphoma (ALCL)
Angioimmunoblastic T-cell lymphoma
Burkitt's lymphoma (BL)
Classification of T-Cell disorders

Diffuse large cell lymphoma
Follicular lymphoma (FL)
Hodgkin's disease
Lymphoepithelioid lymphoma
Lymphoplasmacytic lymphoma
Mantle cell lymphoma (incomplete)
Marginal Zone B-cell lymphoma
Small lymphocytic lymphoma
T-cell large granular lymphocyte leukaemia
Waldenstrom's macroglobulinemia (WM)
del(11q) in non-Hodgkin's lymphoma (NHL)
del(13q) in non-Hodgkin's lymphoma
del(17p) in non-Hodgkin's lymphoma (NHL)
del(6q) abnormalities in lymphoid malignancies
del(7q) in non-Hodgkin's lymphoma (NHL)
dic(9;12)(p11-13;p11-12)
inv(14)(q11;q32.1)
inv(2)(p23q35)
t(10;14)(q24;q11)
t(11;14)(q13;q32)
t(11;18)(q21;q21)
t(14;14)(q11;q32.1)
t(14;18)(q32;q21)
t(14;19)(q32;q13)
t(18;22)(q21;q11)
t(1;13)(q32;q14)
t(1;14)(p22;q32) in non Hodgkin's lymphoma (NHL)
t(1;16)(q11;q11)
t(1;19)(q23;p13)
t(1;2)(q25;p23)
t(1;3)(p36;p21)
t(1;7)(q21;q22)

t(2;14)(p13;q32)
t(2;18)(p11;q21)
t(2;22)(p23;q11.2)
t(2;3)(p12;q27)
t(2;3)(p23;q21)
t(2;5)(p23;q35)
t(2;8)(p12;q24)
t(3;13)(q27;q14)
t(3;14)(q21;q32)
t(3;14)(q27;q32)
t(3;22)(q27;q11)
t(3;4)(q27;p13)
t(3;Var)(q27;Var) in non Hodgkin lymphoma
t(6;8)(q11;q11)
t(6;8)(q27;p12)
t(7;10)(q34;q24)
t(8;13)(p12;q12)
t(8;14) (q24;q11)
t(8;14)(q24;q32)
t(8;22)(q24;q11)
t(9;14)(p13;q32)
t(X;2)(q11;p23)

8. Chronic Lymphoproliferative Diseases (CLD)

1q rearrangements in multiple myeloma
B-cell prolymphocytic leukaemia (B-PLL)
Chronic lymphocytic leukaemia (CLL)
Fibrogenesis imperfecta ossium
Hairy Cell Leukaemia (HCL) and Hairy Cell Leukaemia Variant (HCL-V)
Multiple myeloma
Plasma cell leukaemia (PCL)

Splenic lymphoma with villous lymphocytes
T-cell prolymphocytic leukaemia (T-PLL)
del (13q) in chronic lymphoproliferative diseases
del (6q) in Multiple Myeloma
del(13q) in multiple myeloma
del(6q) abnormalities in lymphoid malignancies
inv(14)(q11;q32.1)
t(11;14)(p11;q32)
t(11;14)(q13;q32)
t(11;14)(q13;q32) in multiple myeloma
t(14;14)(q11;q32.1)
t(14;18)(q32;q21)
t(14;19)(q32;q13)
t(18;22)(q21;q11)
t(1;3)(p36;q21)
t(2;14)(p13;q32)
t(2;18)(p11;q21)
t(4;14)(p16;q32)
t(9;14)(p13;q32)

See also Mitelman's Catalog of Chromosome Aberrations in Cancer '98, John Wiley & Sons, NY, USA; ISBN 0-471-17603-6.

The present inventors consider that the invertor model could be also used as a conditional transgenic approach for making 'normal' (non-fusion) transgenes which are conditional on inversion.Advantageously, the disease modelled by the methods of the invention is leukaemia. Preferably, it is a myeloid or lymphoid leukaemia which produces an Ews-ERG fusion protein. Most advantageously, the leukemia which produces Ews-ERG fusion protein will manifest itself in the form of clonal T-cell malignancies. Further details are provided in **Example 2**.

Sequence from the rearrangement sites is obtained from publicly-available sources, such as GenBank, and used to construct homologous recombination vectors which insert recombinase specific sites, such as *lox* sites, into the regions of interest.

Disease models of chromosomal translocations according to the invention replicate the aetiology of human chromosomal translocations with surprising fidelity. Preferably, the translocation according to the inventions results in an *Ews-ERG* fusion, which is associated with leukaemia in humans. As in humans, the chromosomal translocation between mouse *Ews* and *Erg* results in leukaemia. The data from the inventors study shows that mice with the compound genotype comprising an invertor cassette according to the invention and also Rag1-Cre develop leukaemia only if Cre recombinase is expressed. Further the inventors have shown that the *Ews-ERG* chromosomal translocation results in leukemia in 25/29 mice tested. Thus, the inventors have shown that the Ews-ERG fusion protein is crucial for malignancies in invertor mice as malignancies only arise in mice with both the Ews-ERG and Rag1-Cre genes. (Fig 2).

The presence of thymomas in all afflicted *Ews-ERG*; *Rag1-Cre* mice suggested that the haematological malignancy in these mice comprised T cells. This was confirmed by FACS analysis of T cell surface markers and by determination of T cell or B cell receptor gene rearrangement status. The majority of normal mouse thymic T cells express both the CD4 and CD8 surface markers (double positive or DP cells), as well as the pan-T cell marker Thy1 (CD90) (wt, Fig. 3). The situation with the thymomas of invertor mice varied from mainly CD4+CD8+ DP cells (e.g. M2, Fig. 3), mainly CD8+ SP cells (e.g. M3, Fig. 3) or mainly DN cells (e.g. M4, Fig. 3). Thirteen of the cohort of (25) leukemic invertor mice were analysed for their thymic T cell profile by FACS (summarised in Table 2). While we found either CD8+ SP or CD4+, CD8+ DP thymomas, none showed a CD4+ SP phenotype. Two of the thymomas had a mixed phenotype comprising mainly CD8+ SP cells, with a subpopulation of DP cells (M3 and M6, Table 2).

Detection of tumour formation

The presence of tumours in animals according to the invention is possible by a variety of methods. These include genetic testing, testing for tumour markers, physical observation, cytological assays and the like.

Tumour markers are substances that can be detected in higher than normal amounts in the blood, urine, or body tissues of animals with certain types of cancer. A tumour marker may be produced by the tumour itself, or by the body in response to a cancer presence. When diagnosing cancer, blood and tumour tissue biopsies may be tested. For example, the following markers are known to associated with tumours:

Tumour Marker	**Primary Cancer Site**
Antidiuretic Hormone (ADH)	Small cell lung cancer, adenocarcinoma
Alpha-feto protein (AFP)	Liver, germ cell cancer of ovaries or testis
BTA (Bladder Tumour Antigen)	Bladder
CA15-3 (carbohydrate antigen 15-3)	Breast
CA19-9	Pancreas, colorectal
CA125	Ovarian
Calcitonin	Thyroid medullary carcinoma
Carcinoembryonic antigen (CEA)	Colon, Lung
Creatin-kinase-BB	Breast, ovary, colon, prostate
hCG (human chorionic gonadotropin)	Trophoblastic disease
Lactic dehydrogenase (LDH)	Lymphoma, seminoma, acute leukaemia, metastatic carcinoma
Neuron-specific enolase (NSE)	Neuroblastoma, small cell lung cancer
NMP 22	Bladder

Prostatic acid phosphatase (PAP)	Metastatic cancer of prostate, myeloma, lung cancer, osteogenic sarcoma
Prostate specific antigen (PSA)	Prostate

Tests for tumour markers are available commercially; for example, immunoassays for particular polypeptides may be used to assess their presence and/or amounts in biological samples.

Cytological assays are frequently used in the detection of leukaemia, for example by analysis of various tissues post mortem or analysis of bone marrow tissue obtained by biopsy. Another technology useful for detecting leukaemia, multiparameter immunological detection by flow cytometry, is claimed to be capable of detecting one leukaemic cell among 10,000 normal cells. The principle behind this assay is that leukaemia cells display certain surface, cytoplasmic and nuclear leukocyte antigens which normal cells do not. A fluorescent dye is tagged to an antigen-specific monoclonal antibody, incubated with a sample of bone marrow aspirate from a leukaemic animal, and the cells monitored by flow cytometry for the attachment of fluorescent dyes to the cells. If the cells fluoresce, they possess leukaemia antigens (see Coustan-Smith *et al.*, Lancet 1998; 351: 550-4).

EWS-ERG is a promiscuous fusion protein able to mediate myeloid, lymphoid or mesenchymal tumorigenesis.

Human cancers of mesenchymal origin involves the *EWS* gene with various partner genes in different cell settings, such as in the chromosomal translocation t(21;22)(q22;q21) where *EWS* fuses with *ERG* to encode a novel EWS-ERG fusion protein (14). Thus, the present inventors considered that EWS-fusion proteins can be involved in tumor formation of various cell types and that restrictions in humans are dictated largely by the cells in which the specific chromosomal translocation occurs.

The present inventors tested this hypothesis, and at the same time validated the invertor chromosomal translocation mimic approach, by assessing the possible universality of *EWS-ERG* as an oncogene. The *Ews-ERG* invertor allele was activated in lymphocytes, using the *Rag1-Cre* knock-in mice (35), to determine if lymphoid malignancies would arise.

It was found that invertor *Ews-ERG* mice develop clonal T cell malignancies. The T cells were of varying phenotypes but the majority of cases expressed CD8 on the cell surface and had productive V-D-Jβ rearrangements, indicative of mature T cells. Since Cre-mediated inversion through *loxP* sites recreates the *loxP* sequence, there is potential for the cassette to flip back and forth with continued Cre expression. This had no obvious deleterious effect on tumorigenesis in the *Ews-ERG* invertors as judged by various criteria (tumor penetrance, consistent T cell phenotype, pathology). The potential Cre-mediated flip-flop inversion of the *ERG* cassette may partly explain why predominantly mature T cells in the overt tumors were observed (Table 2 and 3) since CD8+ or CD4+ T cells express lower levels of *Rag1-Cre* 35. Thus the invertor allele of transformed cells may be fixed in the correct orientation for synthesis of *Ews-ERG* mRNA when *Rag-Cre* is no longer effective. On the other hand, the lack of CD4+ (SP) thymomas is difficult to encompass with this explanation, suggesting that other biological mechanisms may be involved in this bias.

The *Rag1-Cre* mouse line used constitutively expresses Cre in lymphocytic cells (35) causing T cell malignancies to arise. This implies that the Ews-ERG fusion in lymphoid cells causes T lymphocyte tumors and that the EWS/FUS fusion proteins could be effective in a spectrum of cell types, outside those normally seen in human malignancies (i.e. sarcomas and myeloid leukemias). But in man, the cell in which an EWS/FUS associated chromosomal translocation occurs could be crucial in determining if a tumor arises or not. In some cells, it may be that EWS/FUS fusions are lethal and those cells acquiring a translocation would die; in others the fusion protein may be tolerated and thus may become tumors. In this respect, the absence of B cell tumors in the *Ews-ERG* invertors is of interest as both B and T cells should

undergo inversion of the *Ews-ERG* cassette, since *Rag1-Cre* is expressed in both cell types (35).

The absence of B cell tumors may reflect toxicity of the fusion protein for B cells. Alternatively, it may mean that Ews-ERG does not function in B cells or that the development of B cell tumors is inhibited by the T cell malignancies which may arise earlier and then dominate the lymphoid compartment.

MLL-AF4 fusion proteins are commonly found in patients suffering from Leukemia.

The recurrent chromosomal translocation t(4;11)(q21;q23) is the most common abnormality in acute leukemias of patients less than 1 year old, and overall the most common MLL fusion protein, giving rise to B lineage tumors invoking a poor prognosis. New methods to combat these leukemias are needed and a mouse model of MLL-AF4 tumorigenesis is needed. Modeling MLL-AF4 effects in mice has been hampered by embryonic lethality due to expression of *Mll-AF4* knock-in alleles, even using *loxP*-STOP conditional alleles. The present inventors have obviated this problem using the invertor mouse conditional technology in which *Mll-AF4* fusion cannot take place until *Cre* expression occurs, giving many options for cell-specific generation of Mll-AF4 fusion. The results show that the *Mll-AF4* invertor mice develop B cell tumors with high penetrance providing an ideal model for studying MLL-AF4 biology.

The invertor technology has allowed the first generation of B cell specific models for MLL-AF4. This is a valuable addition to the repertoire of MLL-models allowing studies of secondary mutations co-operating with MLL-fusions, lineage specific target genes and an important element to development of new therapies to target MLL-leukemia. This last point is of particular clinical relevance as the prognosis of MLL-AF4 associated leukemia is very poor, especially if it occurs before 1 year of age (Pui et al., 2002) and new drugs are needed to treat this leukemia.

More details of this model are provided in Example 5 herein.

Uses of transgenic animals according to the invention

The armoury of approaches to chromosomal translocation mimics includes the knock-in model, the conditional knock-in model, the translocator model (28-30) and the invertor model described here. The new invertor mouse model adds an approach to generation of gene fusions, mimicking the consequence of chromosomal translocations, which is fully dependent of expression of Cre recombinase and inversion of the floxed intron-cDNA cassette. No fusion mRNA can be produced until inversion has taken place as the acceptor splice site is incorrectly orientated prior to inversion. This approach is applicable to any fusion gene using the transfer cassettes described herein. The invertor approach overcomes an inability to generate 'standard' knock-in *Ews* and *MLL* fusion alleles in ES cells and precludes any leakiness due to read through transcription, which can occur with *loxP*-STOP conditional knock-ins, since the invertor approach gives an absolute conditional allele, dependent on Cre recombinase.

Recently, *de novo* translocations have been made in mice using the translocator model which gave rise to leukemias (30). This model is ideal as the abnormal karyotype of the translocator mice is created somatically by Cre-*loxP* recombination. Translocator mice can only be made when the two genes into which *loxP* sites are targeted are orientated on the relevant mouse chromosomes in the same way to generate 5'-3' exon fusion following translocation. The expanding sequence data available on the mouse genome has revealed that the transcription orientation of some genes equivalent to those that fuse in human tumor associated chromosomal translocations, differ in mouse and this would result in dicentric chromosomes in the translocator model. In such situations, the invertor mice can be used as an alternative. The invertor methodology is simple and reliable since no fusion mRNA can be generated until Cre-mediated inversion. Possible effects of re-inversion could be further minimised in future by using inducible, cell-specific Cre expression, such as might be achieved with the Cre-oestrogen receptor-tamoxifen induction methods (42). Alternatively, use of additional

recombination sites such as *frt* sites for FLP recombinase activity, could be used to delete one loxP site and render the invertor cassette stable in one orientation.

The *de novo* chromosomal translocations in mice are the closest to natural, human chromosomal translocations of any model thus far. The finding that inter-chromosomal translocations between *Ews* and *ERG* and *MLL* and *AF4* respectively in mice leads to leukaemia suggests that the *invertor* method of the present invention can be employed, with appropriately specific Cre expressing mice, to any pair of genes as long as the orientation to the centromere does not result in dicentric aberrant chromosomes. Moreover, the advantage of the invertor method described herein is that read-through transcription is prevented in the absence of gene translocation mediated by Cre expression which is induced in adult mice. In this way the production of potentially embryonically lethal fusion proteins generated by read-through transcription of an inserted gene cassette into a gene of interest is avoided in gene-targetted cells such as ES cells is avoided.

The time course of tumorigenicity in the invertor mice suggests that Ews-ERG is necessary but not sufficient for cancer in this model, and thus that 2nd mutational event(s) may have to occur to fix the overt tumor. As a consequence the Ews fusion protein would no longer be needed for tumor viability. It will thus be valuable to evaluate whether the function of the EWS-ERG fusion protein (or for that matter, any of the EWS or FUS fusions) remains essential for the tumor at the time of overt tumor presentation and therefore whether it is a suitable target for cancer therapy.

Engineering mouse models of chromosomal translocations i.e. knock-in, invertor and translocator mice, should be useful as a pre-clinical setting to test new therapeutic approaches. In particular, it is noteworthy that chromosomal translocation products are invariably intracellular (reviewed in 43), anti-cancer reagents must therefore work in the intracellular milieu and drug targeting is an important problem. Thus one motivation for generating mouse models of chromosomal translocations is for potential use as a pre-clinical setting, prior to their translation into use with patients. Several new therapeutic regimes are available but each is at a developmental stage which

requires careful experimental analysis *in vivo* to assess not only efficacy but also possible side effects (toxicity). In addition, there is the all-important issue of delivering reagents, such as for siRNA or intracellular antibodies, to the tumor cells (44). An approach to these crucial points can be made in mouse models, as a precursor to use in humans.

Thus in a further aspect the present invention provides a method for the testing of an anti-tumour agent which method comprises the steps of:

(a) providing a non-human animal tumour model according to the invention,

(b) administering to that model one or more potential anti-tumour agents; and

(c) testing for a change in one or more characteristics of tumour/s comprising the model according to step (a).

In a preferred embodiment of the above aspect of the invention step (c) (testing for a change in one or more characteristics of tumour/s comprising the model according to step (a)) involving the testing for a change in any one or more characteristics of one or more tumours in the group consisting of the following: size, morphology, presence and/or levels of one or more tumour markers and cytological assays.

Drugs used routinely for the treatment of cancer currently include the foloowing:

·**Classic cytotoxics:** these cover alkylating agents, antibiotics including topoisomerase inhibitors, plant alkaloids, taxanes, and antimetabolites; and

· **Sensitisers:** ie chemosensitisers, which are agents that increase the sensitivity of tumour cells to chemotherapeutic agents (radiosensitisers will not be covered in this article).

Classic cytotoxics

Alkylating agents

Alkylating agents act by binding to guanine on the DNA strand and damaging the genetic material of the cancer cell, often by creating cross-links between the guanine and its partner base, cytosine, on the complementary DNA strand. Such cross-linking prevents separation of the DNA strands, an event which must occur for cell division.

Outgrowth of resistant cancer cells is common, and is probably due to the upregulation of enzymes, such as 0(6)-alkylguanine DNA transferase, which break the alkyl bond. Hence the development of therapeutic approaches which seek to downregulate the expression of these resistance-associated enzymes during the course of chemotherapy.

There are over 30 marketed alkylating agents, including drugs such as carmustine, dacarbazine, fomustine, temozolomide, lomustine, streptozotin and procarbazine, as well as the well-known agent cisplatin.

Anti-metabolites

These drugs affect dividing cells; they may be directly incorporated into DNA, thus disrupting the cell's capacity to function, or they may interfere with metabolic pathways necessary for DNA synthesis. Marketed antimetabolites include the 5-FU derivatives doxifluridine and capecitabine, and methotrexate and gemcitabine.

Antimetabolites which fall into the former group include the pyrimidine analogues and the purine analogues (pyrimidines and purines are the two types of base found in the DNA strand). Analogues of pyrimidines include doxifluridine and the well-known 5-fluorouracil, which replace thymine and cytosine bases. Analogues of purines include thioguanine and mercaptopurine, which replace adenine and guanine bases. Antimetabolites which fall into the latter group include the antifolates, which inhibit folic acid metabolism thereby inhibiting DNA synthesis. These include inhibitors of dihydrofolate reductase and methionine transmethylase, such as methotrexate. There are over a dozen marketed antimetabolites.

Plant alkaloids

These are represented by the vinca alkaloids, which bind to microtubules (sub-cellular structures which must function correctly in order for cell division to occur). Interference with microtubules interferes with cell division and therefore interrupts

tumour growth. Marketed alkaloids include vincristine, vinblastine, vindesine and vinorelbine; other microtubule poisoning agents are under development.

Taxanes

Like the vinca alkaloids, taxanes are plant-derived molecules which target the microtubules. There are only two marketed taxanes, paclitaxel and docetaxel. Rather than preventing appropriate microtubule formation, the taxanes promote it, but then prevent microtubule dissociation. This in turn prevents cell division, leading to a kind of paralysis of the cancer cell, rather than to cell death.

The best known of these agents is paclitaxel, which facilitates the assembly of intracellular microtubules which in turn inhibit cell division and therefore prevent the proliferation of tumour cells and growth of tumours. Paclitaxel appears to be effective in ovarian lung and breast tumours. Side effects appear to be relatively minor, being mainly limited to hypersensitivity reactions which can be largely prevented by premedication. Other more rare side effects are peripheral neuropathy and neutropenia.

Anti-angiogenic agents

Tumours cannot grow to any significant size without forming their own blood supply. Agents which prevent the formation of new blood vessels therefore would be expected to prevent the progression of metastatic disease. Many agents of anti-angiogenic potential are proteins or peptides and therefore fall outside the remit of this article. However, there are some small molecules being developed which may inhibit new vessel formation. Combretastatin is a small molecule that is said to target the tumour vasculature; it appears to directly kill angiogenic tumour cells by the induction of apoptosis.

Antibiotics

Antibiotics Some antibiotics have a degree of cancer cell-selectivity in that they either

inhibit enzymes involved in the synthesis of DNA or else they directly interfere with DNA. Anthracycline antibiotics, which include daunorubicin and doxorubicin, are antibiotics which are have the ability to inhibit topoisomerases. Topoisomerases are enzymes which unwind supercoiled DNA in a step that is an essential precursor for DNA translation, and replication / cell division. Topoisomerase inhibitors interfere with this step, for example by preventing the DNA re-associating after the unwinding event (eg etoposide) and hence have a direct action on dividing cells.There are over 10 marketed topoisomerase inhibitors including actinomycin D and mitomycin C.

Chemosensitisers

Many of these agents are aimed at decreasing the capability of tumour cells to develop or maintain resistance to anticancer drugs. Such resistance often is mediated by a single gene (MDR-1), and therefore may in theory be inhibited by any drug which targets the product of that gene. Some drugs which have been used to combat the MDR-1 phenotype to date include cyclosporin and verapamil; several others are in development.

Some drugs have been developed which are sensitive to light of particular wavelengths. These can be administered systemically; subsequent exposure of a localised area of tissue to light results in localised conversion of the drug into a short-lived cytotoxic form (eg by generation of free radicals). This may be a successful route for the treatment of some inoperable solid tumours; however, it is unlikely that photodynamic therapy will be converted into a treatment for disseminated disease.

The testing of these and other agents in cancer treatment may be performed using the cancer models of the invention.

The invention is further described, fur the purpose of illustration, in the following examples:

EXAMPLES.

Example 1.

Methods

Generation of targeting constructs and homologous recombination.

***Ews* knock-ins**

An *Ews* genomic clone was isolated from a 129 λ phage library 45 and a targeting vector made extending from the Xba1 site on the 5' side of exon 7 to the EcoRI site on its 3' side (Fig. 1A). This initial *Ews* targeting vector was made in two steps in pBSpt. First, the 5' XbaI-BamHI 2Kb fragment was cloned into SalI-BamHI vector sites with SalI and XbaI sites filled-in to eliminate the XbaI site. Second, the 3' BamHI-EcoRI *Ews* 4.6Kb fragment was cloned into the BamHI-ClaI sites of pBSpt with ClaI and EcoRI sites filled in to eliminate the EcoRI site. MC1-tk 46,47 was cloned into the EcoRV site for negative selection of the homologous recombinants in the ES cells. The knock-in clones to generate *Ews*- fusions with *AF9, Fli1, CHOP* and *ATF* were made using a transfer vector (clone pC2A-neo, Figure 5) into which respective cDNA sequences were cloned as NotIflanked PCR fragments. The mRNA sources for RT-PCR amplification were from cell lines; the mouse erythroleukemia MEL (*Fli1*), the human erythroleukemia HEL (*AF9*), mouse 3T3L1 fibroblasts (*Chop*) and the human T cell leukemia Jurkatt (*ATF*). The sequences of the cDNA cassettes are given in Figure 6. Each transfer cassette comprised the corresponding fusion part of the *EWS* partners, a polyA signal from SV40 and a selectable marker (neo) for selection of homologous recombinants and the transfer cassettes were cloned into the BamHI site in *Ews* exon 7 (Fig.1).

Knock-in was carried out by homologous recombination as described (45). Two probes used for detecting homologous recombinants in filter hybridizations of ES cell genomic DNA were cloned into pBSpt (Fig. 1); the 5' probe was a 1.4Kb KpnI fragment and the 3' probe was a 0.25Kb EcoRI-BamHI fragment. These were used to detect homologous recombinants in EcoRI or BglII digested genomic DNA for 5' or 3' probes respectively. The targeting vectors also had the HSV *tk* gene for negative selection of non- homologous integrations (46,47).

The cDNA transfer cassettes were cloned into the *Ews* exon 7 BamHI site within the XbaIEcoRI fragment indicated in Figure 1A. These were subsequently transfected into the ES cell line CCB as described, G418-gancyclovir-resistant colonies grown and DNA was prepared and homologous recombination determined using filter hybridization 48. Targeted clones for *Ews-AF9* or the control targeting with the transfer vector only (pC2-neo, Figure 5) were expanded, verified by hybridizing fresh DNA preparations. These were injected into C57BL/6 blastocysts and chimaeras assessed by coat colour. High percentage male chimaeras were bred with C57BL/6 females and the genotypes of pups analysed to determine germline transmission.

Invertor knock-in

The *ERG* inversion knock-in cassette was prepared using the pC2A-neo transfer cassette (Figure 5A). RT-PCR was carried with cDNA made from human colon carcinoma COLO320 mRNA, and primers allowing cloning into the unique NotI site of pC2A-neo. The sequence of the *ERG* segment, corresponding to the part of *ERG* found in *EWS-ERG* fusion (15), was determined (see Figure 6E) and the resulting p-ERG-C2A-neo transfer cassette was cloned as a BglII fragment into the BamHI site of pLOXKI (Figure 7 online) prepared in the pMG2 vector (Figure 7B and C; pMG1 and pMG2 was prepared by mutagenesis of pBSpt to incorporate respectively one or two SceI sites; this restriction enzyme recognizes an 18bp site). An EcoRI 120bp *loxP*-containing fragment, from pPGK-loxP-hygro 29, was blunt end cloned into the filled-in NotI site of pMG2 (Figure 7D; both EcoRI and NotI sites are eliminated), followed by a 530bp XbaI-BamHI fragment of intron 4 from the mouse *Af4* gene, to include the distal region of the intron up to and including the acceptor splice site (this fragment was PCR amplified as an XbaI-BamHI fragment, from mouse 129 genomic DNA, Figure 7E). A further *loxP* fragment was cloned as blunt EcoRI fragment into the EcoRV site of the vector (Figure 7F online), in an orientation compatible with Cre recombinase mediated inversion of the cassette. This yielded the final pLOXKI cloning vector (Figure 7G) for receipt of the *ERG* transfer cassette which was cloned as a BglII fragment into the unique BamHI site of pLOXKI. The overall structure of the *ERG* inversion cassette is shown in Figures 1A and Figure 8A, and this was cloned

into the unique XbaI site of *Ews* (XbaI)-EcoRI targeting vector (described above, Fig. 1A) as a blunt ended SceI-ClaI fragment, leaving an XhoI site for linearization prior to transfection into ES cells. The orientation of the *ERG* inversion cassette was 3'-> 5' with respect to *Ews* and this was determined by restriction site mapping and genomic DNA sequencing (Figure 7A depicts the organisation of the *ERG* inversion cassette before and after Cre-mediated inversion in the *Ews* targeting vector). Homologous recombinant knock-in alleles were detected and isolated as above. These ES cells were injected into C57BL/6 blastocysts, chimaeric mice produced and germline transmission of the inversion knock-in allele obtained. These mice were crossed with the line of mice expressing Cre from a *Rag1* knock-in 35 and specimen spleen DNA prepared and a PCR product made to determine sequence of the Cre-mediated inversion junction (Fig. 1D). The previously published *Rag-Cre* knock-in allele (35) was made as outlined in Figure 8A-C, which also shows relevant junction sequences of three Cre cassettes suitable for knock-in of Cre into any gene of interest (Figure 8D).

Molecular analyses

Genomic DNA was prepared from tissues using proteinase K digestion and phenolchloroform extraction and filter hybridizations was carried out with radiolabelled probes as described (48,49). The probes used to detect antigen receptor gene rearrangements were a heavy chain Ig μ intron enhancer probe (37) and a T cell receptor Jβ2 probe (36). T cell receptor V-D-Jβ junction sequences were obtained as described (38,39) using pools of Vβ primers combined with a single Jβ2 primer to amplify products covering the junction of the joints between Vβ, Dβ and Jβ2, followed by direct sequencing of the PCR product.

Analysis of leukemia/lymphoma

A cohort of mice was established by inter-breeding *Ews-ERG* invertor line with a *Rag1-Cre* line (35) to establish littermates with *Ews-ERG* + *Rag1-Cre* or *Ews-ERG* genotypes. Genotypes were determined using filter hybridization of genomic DNA from a small tail biopsy. The health status of these mice were monitored and if signs of ill-health appeared, mice were sacrificed and a post-mortem was carried out. Tissue samples were removed for single cell preparation for determination of surface protein

expression phenotype, for nucleic acid preparation or for fixation in 10% formalin for histology. Blood smears were mice prepared on glass slides and stained with May-Grünwald-Giemsa (MGG). For histology, fixed tissues were embedded in wax and 0.4μ sections made, stained with haematoxylin and eosin (H&E) after mounting on slides. FACS analysis was conducted using a FACSCalibur with fluorescent antibodies purchased from BD Biosciences. Data was analysed with CellQuest software. Western protein detection was carried out as described 50 using 18μg protein per lane and proteins fractionated on 4-20% SDS-PAGE. The separated proteins were electrotransferred to Protran membranes (Schleicher & Schuell) and specific proteins detecting using anti-Ews antibody (Santa Cruz; detecting the amino terminus of Ews) or anti-ERG antibody (Santa Cruz; detecting the carboxy terminus of ERG). Antibody bound to filter was detected using horse radish peroxidase (HRP) detection and ECL as described by the Manufacturer (Amersham).

Example 2.

Conditional inversion of the *ERG* gene and fusion with *Ews*

The use of homologous recombination to generate oncogene fusions was established with the creation of the *Mll-AF9* mouse model [24]. We employed a similar technology to make *Ews* fusions of the type found in specific human sarcomas and leukemias. Transfer cassettes were made into which RT-PCR fragments of relevant *Ews*-fusion partners were cloned into an *Ews* genomic fragment for selecting homologous recombinants in embryonic stem cells (ES cells) (the details of three transfer cassettes and targeting clones are given in Figs. 5 and 6;). *Ews* fusions with *Chop*, *ATF* and *FliI* were made and an *Ews-AF9* fusion (not having been observed as a human *EWS* fusion genes) made as a control (sequences of these fusions cDNAs are shown in Figure 6). While we were able to obtain homologous recombination knock-in clones with either a vector carrying the transfer cassette only or the *Ews-AF9* control fusion, no clones were obtained with any of the *Ews*-fusions equivalent to those naturally found in human cancers (Table 1A). Furthermore, chimaeras could be made with the *Ews* cassette and the *Ews-AF9* knock-in (Table1B) and the *Ews-AF9* chimaeras were bred to give heterozygous carriers of the knock-in allele (data not shown). We conclude that

Ews-fusions of the type naturally associated with human cancers are lethal in ES cells and therefore that a conditional knock-in model was required for this system.

An alternative approach developed for knock-in technology is the conditional transcription *loxP*-STOP method, developed to overcome embryonic lethality of fusion proteins [31]. We attempted this approach with *Mll* gene fusion knock-ins, but found that even a low level of transcription read-through of the knock-in allele was sufficient for lethality (unpublished data). This finding necessitated a more robust condition form of knock-in, in which transcription of the fusion gene is impossible until a conditional activation step is invoked. *LoxP*-Cre recombination systems are known to facilitate inversion of sequences added into the genome [32-34], suggesting that knocking-in an inverted cassette flanked by *loxP* sites to create an absolute conditional allele in which Cre mediated inversion was a mandatory step for protein fusion. This has the additional advantage that cell-specific Cre expression will allow spatial timing to the generation of the fusion allele.

We tested this approach by knocking-in an *ERG* invertor cassette in the *Ews* gene. The invertor cassette was constructed to include a short segment of intronic sequence (from intron 4 of the mouse *Af4* gene) including the acceptor splice site, the *ERG* cDNA made by RT-PCR of the relevant part of *ERG* usually involved in human translocations, a polyA site and the MC1-neo-pA segment for selection of homologous recombinant clones, all flanked by *loxP* sites with an orientation compatible for inversional recombination by Cre (Fig. 1A and detailed sequences of the components of the inversion cassette are shown in Figs. 5 and 7). Gene targeting in ES cells was used to introduce the invertor cassette, cloned in an *Ews* genomic fragment, into the mouse *Ews* gene downstream of the donor splice site of exon 7 (Fig. 1A). The selected targeted *Ews* allele (identified using the external 5' probe and verified with the external 3' probe, Fig. 1A) had the *ERG* invertor cassette orientated 3' to 5' with respect to *Ews*, thus making a splice from *Ews* exon 7 donor site to the acceptor site of the invertor cassette impossible. After Cre-mediated inversion to give the correct transcription orientation, the correct splice occurred (illustrated in Fig. 1C) as confirmed by transiently transfecting targeted ES cells with a plasmid encoding Cre

recombinase and amplifying the genomic junction of the the *Ews-ERG* allele (sequence shown in Fig. 1D). By sequencing the RT-PCR product from these cells, we also confirmed an in-frame sequence of *Ews* fused to *ERG* (Fig. 1E). The targeted ES cells were injected into blastocysts and these yielded normal numbers of chimaeras which gave rise to heterozygous carriers of the *Ews-ERG* invertor allele (invertor mice).

Example 3: The *Ews-ERG* fusion protein causes T cell leukemia in invertor mice.

We investigated the ability of Ews-ERG fusion protein to contribute to leukemogenesis by causing the inversion of the *ERG*-containing cassette in lymphoid cells using a *Rag1-Cre* knock-in mice [35] (Figure 8 illustrates the plasmid pC2A-neo used for knocking-in the *Cre* gene into the *Rag1* gene; Figure 8C shows the junction sequence of *Rag1* and *Cre*; Figure 8D shows other transfer cassettes suitable for knock-in homologous recombination of *Cre* recombinase into genes of interest). 25 out of the cohort of 29 mice (86%) mice carrying both *Ews-ERG* and *Rag1-Cre* genes developed leukemia associated with thymoma, whereas the *Ews-ERG*-only cohort (20 mice) did not display leukemias (Fig. 2). Blood smears at the time of sacrifice typically contained elevated numbers of leukocytes, with lymphoid morphology of different maturities including lymphoblasts (Fig. 2). In addition, bone marrow sections showed high levels of infiltrating lymphoblasts (Fig. 2). The diseased mice all had large thymomas, often associated with splenomegaly (Table 2), and the histology of spleens showed variously either complete loss of normal architecture (loss of demarcated white and red pulp), partial loss or normal architecture (Fig. 2C, M9, M21 and M6 respectively). Similarly, the level of infiltration of leukemic lymphocytes into liver and kidney varied, with some having large amounts of perivascular deposits and others marginal levels. All mice had thymomas with homogeneous structures of leukemic cells.

The Ews-ERG fusion appears crucial for malignancies in the invertor mice as these only arose in mice with both the *Ews-ERG* and *Rag1-Cre* genes (Fig. 2). The presence of *Ews-ERG* mRNA was confirmed in spleens of these mice identical with that in ES

cells (Fig. 1E and data not shown). In addition, the steady state orientation of the *loxP*-flanked *ERG* cassette was studied using filter hybridization and we found that the *ERG* sequence had become inverted into the 5' to 3'orientation with respect to the *Ews* gene, in each case of thymoma (Fig. 4A and Table 2). The other tissues examined have little evidence of an inverted band. Finally, the presence of the Ews-ERG fusion protein was shown by Western blotting of thymoma proteins using antibodies binding to either Ews or ERG which detected respectively the normal mouse Ews or Erg proteins and the Ews-ERG fusion molecule in the thymoma T cells of an *Ews-ERG; Rag1-Cre* mouse (Fig. 4D).

The presence of thymomas in all afflicted *Ews-ERG; Rag1-Cre* mice suggested that the haematological malignancy in these mice comprised T cells. This was confirmed by FACS analysis of T cell surface markers and by determination of T cell or B cell receptor gene rearrangement status. The majority of normal mouse thymic T cells express both the CD4 and CD8 surface markers (double positive or DP cells), as well as the pan-T cell marker Thy1 (CD90) (wt, Fig. 3). The situation with the thymomas of invertor mice varied from mainly CD4+CD8+ DP cells (e.g. M2, Fig. 3), mainly CD8+ SP cells (e.g. M3, Fig. 3) or mainly DN cells (e.g. M4, Fig. 3). Thirteen of the cohort of (25) leukemic invertor mice wereanalysed for their thymic T cell profile by FACS (summarised in Table 2). While we found either CD8+ SP or CD4+, CD8+ DP thymomas, none showed a CD4+ SP phenotype. Two of the thymomas had a mixed phenotype comprising mainly CD8+ SP cells, with a subpopulation of DP cells (M3 and M6, Table 2).

The clonality of these T cell tumors was assessed by filter hybridization of genomic thymoma DNA with a T cell receptor β probe (*Tcrb*) from the Jβ2 region [36] (diagrammatically shown in Fig. 4B). The probe detects a 5Kb band in non-lymphoid DNA corresponding to the intact *Tcrb* gene; if V-D-J or D-J joins have occurred, new restriction fragments are created giving rise to 'rearranged' bands on the hybridization autoradiograph. Figure 4B shows the rearrangement status of 12 of the *Ews-ERG* thymomas. All except one (M9) showed at least one rearranged allele and several have two rearranged alleles (e.g. M18). Variable amounts of residual germline allele were

present. The *Tcrb2* rearrangement status of the cohort of *Ews-ERG* leukemic mice is summarised in Table 2. In addition, the status of the immunoglobulin locus was examined with a probe from the region between JH and Cμ segments [37] (Fig. 4C). Single *Igh* allele rearrangements were observed in about half of the cohort (Table 2) including ones in which FACS analysis and *Tcrb* clonality hybridizations showed the tumors to be clonal T cell malignancies. The Igh rearrangements are therefore likely to be abortive rearrangements.

Confirmation that the *Tcrb* rearrangements observed in the clonality hybridizations were due to productive V-D-J join, was made by sequence analysis of the *Tcrb* genes in four of the cases. PCR amplification from thymoma DNA was carried out with pools of Vβ primers and a Jβ2 reverse primer (primer sequences from [38,39]) and the product fractionated on agarose to determine the presence of a single amplified band. In turn, each band was eluted and the sequence obtained using the Jβ2N primer [39] and the sequences identified by comparison with the IGMT database [40,41]. Table 3 shows the V-D-J junctional sequences indicating the presence of productive T cell receptor β genes. In all cases, N-region diversity is present between the V-D and D-J junctions. In the thymoma DNA from M2, a nonproductive join has also occurred, in addition to the productive one, resulting in an out-offrame joint.

The invertor model of chromosomal translocations has successfully been applied with the creation of an *Ews-ERG* invertor allele in mice. Our data show that the strategy is effective in producing both cell-specific fusion protein and allowing for the manifestation of the oncogenic phenotype of the protein (in this exemplification clonal T cell malignancies).

This chromosomal translocation mimic is applicable to any fusion gene of interest using the set of vectors described herein.

Example 4: The Ews-ERG Fusion Protein can initiate neoplasia from lineage-committed haematopoietic cells.

To assess whether EWS-ERG can be a tumour initiator in cells other than mesenchymal cells, we have analysed Ews-ERG function in a cellular environment not typical of that found in human cancers, namely committed lymphoid cells. We have used *Ews-ERG* invertor mice having an inverted *ERG* cDNA cassette flanked by *loxP* *sites* knocked in the *Ews* intron 8, and were crossed with mice expressing *Cre* recombinase under the control of the *Rag1* gene to give conditional, lymphoid-specific expression of the fusion protein. Clonal T cell neoplasias arose in these mice. This first conditional *Ews* gene fusion model of tumourigenesis shows that Ews-ERG can cause haematopoietic tumours and the precursor cells are committed cells. Thus Ews-ERG can function in cells that do not have to be pluripotent progenitors nor mesenchymal cells.

Materials and Methods

Generation of targeting constructs and homologous recombination. The *Ews* genomic targeting clone was constructed from λ phage DNA isolated from mouse 129 DNA. The knock-in clones for potential *Ews*- fusions with *AF9, Fli1, Chop* and *ATF1* were made using a transfer vector pC2A-neo. Knock-in was achieved by homologous recombination. The *ERG* inversion knock-in cassette was prepared using cDNA made from human colon carcinoma COLO320 mRNA (the sequence of the *ERG* segment, corresponding to the part of *ERG* found in *EWS-ERG* fusion. A diagram of overall structure of the *ERG* inversion cassette is shown in Figure 14A, indicating the initial orientation of the *ERG* inversion cassette (3'-> 5' with respect to *Ews*) and this was determined by restriction site mapping and genomic DNA sequencing. After identification of targeted ES cells, these were injected into C57BL/6 blastocysts, chimaeric mice produced and germline transmission of the inversion knock-in allele obtained. These mice were bred with mice expressing Cre from a *Rag1* knock-in. Specificity of expression of the *Rag1-Cre* allele has previously been described and was sustained using the ROSA26-R reporter mice.

Molecular analyses. Genomic DNA was prepared from tissues using proteinase K digestion and phenol-chloroform extraction and filter hybridizations was carried out with radiolabeled probes. The probes used to detect antigen receptor gene

rearrangements were a heavy chain Ig μ intron enhancer probe and a T cell receptor Jβ2 probe. T cell receptor V-D-Jβ junction sequences were obtained by PCR amplification from thymoma DNA with pools of Vβ primers and a Jβ2 reverse primer (primer sequences from and the product fractionated on agarose to determine the presence of a single amplified band. In turn, each band was eluted and the sequence obtained using the Jβ2N primer and identified by comparison with the IGMT database

RT-PCR was carried out on cDNA as described.. RNA was prepared using RNeasy (Quiagen) from cells which were labelled with FITC-conjugated antibodies and isolated by flow cytometry. Sequences have been described for *Pax5* and *CD3* RT-PCR primers and the sequences of the *Ews-ERG* primers were 5'-CCACAGGATGGTAACAAGCCTGC-3' (*Ews*) and 5'-CGAACTTGTAGGCGTAGC-3' (*ERG*). The hybridisation probe was a 303bp fragment of *Ews-ERG* residing within the RT-PCR product.

Analysis of leukaemia/lymphoma. A cohort of mice was established by inter-breeding *Ews-ERG* invertor line with a *Rag1-Cre* line to generate littermates with *Ews-ERG* + *Rag1-Cre* or *Ews-ERG* genotypes. Genotypes were determined using filter hybridization of genomic DNA from a small tail biopsy. The health status of these mice was monitored and if signs of ill-health appeared, mice were sacrificed and a post-mortem was carried out. Tissue samples were removed for single cell preparation for determination of surface protein expression phenotype, for nucleic acid preparation or for fixation in 10% formalin for histology. Blood smears were prepared and stained with May-Grünwald-Giemsa (MGG). For histology, fixed tissues were embedded in wax and 0.4μμ sections made, stained with haematoxylin and eosin (H&E) after mounting on slides. FACS analysis was conducted using a FACSCalibur with fluorescent antibodies purchased from BD Biosciences. Data were analyzed with CellQuest software. Western protein detection was carried out as described using 18μg protein per lane and proteins fractionated on 4-20% SDS-PAGE. The separated proteins were electro-transferred to PVDF nylon membranes (Millipore) and specific proteins detected using anti-Ews antibody (Santa Cruz; raised against the amino

terminus of Ews) or anti-ERG antibody (Santa Cruz; raised against the carboxy terminus of ERG). Antibody bound to filter was detected using secondary antibodies and ECL as described by the manufacturer (Amersham).

Results: The Ews-ERG fusion protein causes T cell neoplasia in invertor mice.

We investigated the ability of Ews-ERG fusion protein to contribute to Leukaemogenesis by causing the inversion of the *ERG*-containing cassette in lymphoid cells using *Rag1-Cre* knock-in mice. Twenty five out of the cohort of 29 mice (86%) mice carrying both *Ews-ERG* and *Rag1-Cre* genes developed leukaemia associated with thymoma within 500 days, whereas the *Ews-ERG*-only cohort (20 mice) did not display neoplasia. Blood smears at the time of sacrifice of *Ews-ERG; Rag1-Cre* mice typically showed elevated numbers of large leukocytes, with lymphoid morphology of different maturities including lymphoblasts. In addition, bone marrow sections showed high levels of infiltrating lymphoblasts. The diseased mice all had large thymomas, usually associated with splenomegaly and lymphadenopathy, and the histology of spleens showed variously either complete loss of normal architecture (loss of demarcated white and red pulp), partial loss or normal architecture. Similarly, the level of infiltration of leukaemic lymphocytes into liver and kidney varied, with some having large amounts of peri-vascular deposits and others marginal levels. All mice had thymomas with homogeneous presence of large leukaemic cells. These characteristics, together with evidence of clonal *TCRB* gene rearrangements and expression of TCR-associated CD8 and CD4, permits diagnosis of large cell anaplastic T cell lymphoma according to the Bethesda proposals for lymphoid tumours in mice.

The Ews-ERG fusion appears crucial for malignancies in the invertor mice as disease only arose in mice with both the *Ews-ERG* and *Rag1-Cre* genes. The presence of *Ews-ERG* mRNA was confirmed in spleens of these mice, identical with that in *Ews-ERG* ES cells transfected with *Cre* expression plasmids. In addition, the steady state orientation of the *loxP*-flanked *ERG* cassette was studied using filter hybridization and we found that the *ERG* sequence had become inverted into the 5' to 3'orientation with respect to the *Ews* gene, in each case of thymoma. The other tissues examined have little evidence of an inverted band. Finally, the presence of the Ews-

ERG fusion protein was shown by Western blotting of thymoma proteins using antibodies binding to either Ews or ERG, which detected respectively the normal mouse Ews or Erg proteins and the Ews-ERG fusion molecule in thymoma T cells of an *Ews-ERG; Rag1-Cre* mouse..

Studies of progenitor gene expression have indicated the promiscuous expression of lymphoid markers precedes lineage commitment and thus, the Rag1-Cre allele might be expressing in cells destined to become non-lymphoid. The *Rag1-Cre* knock-in allele was previously shown to be specific for lymphoid cells using a reporter assay dependent on Cre-mediated deletion of a *loxP*-flanked (floxed) segment of the *Lmo2* gene. We have sustained these observations using an additional reporter assay, namely the *ROSA-loxSTOP-lacZ (ROSA26-R)* allele where βgal expression is activated by deletion of a *loxP-pA* site. Haematopoietic populations from mice carrying both *Rag1-Cre* and *ROSA26-R* alleles were stained with fluorescent antibodies binding to various surface markers and flow cytometry carried out for co-detect ion antibody and βgal-derived fluorescence. βgal signal was found in thymus and spleen (almost all cells) and in BM, a population of βgal+ cells exist but did not co-express with CD34, Sca1, Ckit or Ter119 (or Mac1, not shown). This means that the Rag1-Cre is restricted to lymphoid cells as previously shown.

The *Rag1-Cre* gene is expressed in both B and T cells but only T cell tumours have arisen in the *Ews-ERG; Rag1-Cre* invertor line. The possible inversion of the *Ews-ERG* gene in B cells was investigated using RT-PCR analysis of expressed *Ews-ERG* fusion mRNA (Fig 16). RNA was prepared from whole spleen or thymus or from flow sorted B220+ spleen cells (3400 cells) or thy1.2+ spleen cells (6400 cells) and RT-PCR performed. *Pax5* and *CD3* primers were used for specific detection of transcripts B cell and T cells respectively. *Pax5* transcripts were detected in cDNA made from spleen and B220+ sorted cells and *CD3* in the spleen, thymus and thy1.2+ sorted cells. No evidence of *CD3* expression was found in the sorted B220+ cells, nor of *Pax5* in the sorted thy1.2+ cells, showing that the sorted cells were practically free of contaminating T or B cells respectively. The presence of *Ews-ERG* fusion RNA was analysed with RT-PCR primers yielding a product spanning the fusion junction and detected with an internal junction probe. *Ews-ERG* RT-PCR product was detected in the unfractionated spleen and thymus sources, as well as in the purified, sorted B220+

and thy1.2+ cells. Therefore, Cre-mediated inversion of the *Ews-ERG* gene occurs in both T and B cells.

Ews-ERG induces malignancies of mature T cells

The presence of thymomas in all afflicted *Ews-ERG*; *Rag1-Cre* mice suggested that the hematological malignancy in these mice comprised T cells. This was confirmed by FACS analysis of T cell surface markers and by determination of T cell or B cell receptor gene rearrangement status. The majority of normal mouse thymic T cells express both the CD4 and CD8 surface markers (double positive or DP cells), as well as the pan-T cell marker Thy1 (CD90). The situation with the thymomas of invertor mice varied from mainly CD4+CD8+ DP cells mainly CD8+ SP cells, or mainly DN cells. Thirteen of the cohort of 25 leukaemic invertor mice were analyzed for their thymic T cell profile by FACS. While we found either CD8+ SP or CD4+CD8+ DP thymomas, none showed a CD4+ SP phenotype. Two of the thymomas had a mixed phenotype comprising mainly CD8+ SP cells, with a sub-population of DP cells.

The clonality of these T cell tumours was assessed by filter hybridization of genomic thymoma DNA with a T cell receptor β probe (*Tcrb*) from the Jβ2 region. The probe detects a 5Kb band in non-lymphoid DNA corresponding to the intact *Tcrb* gene; if V-D-J or D-J joins have occurred, new restriction fragments are created giving rise to 'rearranged' bands on the hybridization autoradiograph. All except one (M9) showed at least one rearranged Jβ2 allele and several have two rearranged alleles (e.g. M18). Variable amounts of residual germline allele were present. The status of the immunoglobulin locus was examined with a probe from the region between JH and Cμ segments. Single *Igh* allele rearrangements were observed in about half of the cohort including ones in which FACS analysis and *Tcrb* clonality hybridizations showed the tumours to be clonal T cell malignancies. The Igh rearrangements are therefore likely to be abortive rearrangements.

Confirmation that the *Tcrb* rearrangements observed by hybridizations were due to productive V-D-J joins, was made by sequence analysis of the *Tcrb* genes in four of the cases. PCR amplification from thymoma DNA was carried out with pools of Vβ primers and a Jβ2 reverse primer (primer sequences from and sequences identified with the IGMT database. Table 4 shows the V-D-J junctional sequences

indicating the presence of productive T cell receptor β genes. In all cases, N-region diversity is present between the V-D and D-J junctions. In the thymoma DNA from M2, a non-productive join has also occurred, in addition to the productive one, resulting in an out-of-frame joint.

Discussion: Our results show that the Ews-ERG fusion, normally restricted to sarcomas in humans, can initiate T lymphocyte tumours, if conditionally expressed in committed lymphoid cells. Leukaemogenesis in *Ews-ERG* invertor mice thus concurs with the hypothesis that EWS fusions can cause neoplasia arising in various cell types. Further, our data shows that the cellular context of *EWS*-associated chromosomal translocations in humans does not have to be a stem cell or even a multi-potent progenitor, as committed lymphocyte precursors expressing *Rag* recombinase genes are the leukaemic precursors in our invertor model. The same conclusions can be applied to *FUS*-associated chromosomal translocations that seem to be largely interchangeable with EWS, given the spectrum of cancers in which these are found and the relatedness of *FUS* and *EWS* coded proteins. Thus, *EWS* or *FUS* chromosomal translocations probably arise more by virtue of accessibility of the genes to chromosomal translocations than by the precise cellular specificity of the resultant fusion proteins.

Example 5: MLL-AF4 invertor mice develop B-cell tumours.

The recurrent chromosomal translocation t(4;11)(q21;q23) is the most common abnormality in acute leukemias of patients less than 1 year old, and overall the most common MLL fusion protein, giving rise to B lineage tumors invoking a poor prognosis. New methods to combat these leukemias are needed and a mouse model of MLL-AF4 tumorigenesis is needed. Modeling MLL-AF4 effects in mice has been hampered by embryonic lethality due to expression of *Mll-AF4* knock-in alleles, even using *loxP*-STOP conditional alleles. The present inventors have obviated this problem using the invertor mouse conditional technology in which *Mll-AF4* fusion cannot take place until *Cre* expression occurs, giving many options for cell-specific generation of

Mll-AF4 fusion. The results show that the *Mll-AF4* invertor mice develop B cell tumors with high penetrance providing an ideal model for studying MLL-AF4 biology.

Results

***Mll-AF4* fusion is lethal for mouse embryos**

We initially used the knock-in approach to produce an *MLL-AF4* mouse model analogous to the method used for the *Mll-AF9* knock-in (Corral et al., 1996; Dobson et al., 1999). An *AF4* cDNA-polyA cassette was fused into exon 10 of the mouse *Mll* gene in ES cells. Expression of the *Mll-AF4* fusion transcript could be demonstrated in these knock-in ES cells. However, injection of two different clones of targeted ES only yielded one low level chimeric mouse in 334 transferred blastocysts. This chimera was runted and died after 14 days. The failure in chimera production was due to embryonic lethality at around E12.5.

In an attempt to circumvent this embryonic lethal effect, a conditional knock-in version was generated using a floxed transcription stop (*Mll-AF4 loxP*-stop) analogous to that described for *Aml-Eto* fusion (Higuchi et al., 2002). Like the *Mll-AF4* fusion knock-in, we were unable to produce chimeric mice with the *Mll-AF4 loxP*-stop ES cells, whereas control ES cells with only a loxP sequence added at the same intronic position resulted in usual numbers of chimera. We concluded that transcriptional read through of the *loxP*-STOP allowed the embryonic lethal effect of the *Mll-AF4* fusion transcript.

Therefore *Mll-AF4* invertor ES cells were created, injected into blastocysts and germline carriers of the cassette were made. This *Mll-AF4* knock-in model was not subject to embryonic lethality, allowing inter-breeding with lines of mice expressing Cre recombinase in various specific hematopoietic cell types. Two sets of *Mll-AF4* invertor ES cells were made in which the *AF4* cassette was flanked by either wild type or mutant *loxP* sites (Albert et al., 1995; Hoess et al., 1982) (*Mll-AF4 loxP and Mll-AF4 lox66/71* invertors,). Following expression of Cre recombinase in the ES cells, the invertor cassette knocked-in to *Mll* intron 10, became inverted and thus was orientated in the same transcriptional direction to *Mll* and was transcribed into a fusion mRNA of *Mll-AF4*.

***Mll-AF4* invertor mice develop lymphoid tumors.**

Targeted ES cells were used to make *Mll-AF4* invertor mice that were crossed with mice expressing Cre recombinase under control of the lymphocyte specific promoters of *Rag-1*, *Lck*, and *CD19* (McCormack et al., 2003) to generate cohorts that were observed for a period of 18 months (we were unable to generate *Mll-AF4* invertor mice crossed with *Lmo2-Cre* due to embryonic lethality similar to *Mll-AF4* knock-in, unpubl.). High proportions of these mice developed lymphoid tumors in this period, ranging from ~70-90%. Survival curves are shown in Fig. **15** for the six groups of mice studied, demonstrating a similar time course of disease incidence irrespective of the Cre strain of mice used to invert the *AF4* cassette into the *Mll-AF4* fusion mRNA.

Prominent findings in the post-mortem examinations were extreme splenomegaly and mesenteric lymphadenopathy due to infiltration by lymphoma. Macroscopic infiltration of lymphoma into liver and kidney was only seen in late stage disease. The thymus was often slightly enlarged and discrimination from mediastinal lymph nodes was not always possible. Lymphomas were not usually found in other sites. Blood smears of tumor bearing mice showed no striking abnormalities compared to wild-type animals. The number of circulating white blood cells was usually only slightly increased (WBC 4.8 ± 2.8 vs. 3.5 ± 2.0), whilst the red blood cell and platelet numbers were slightly reduced (RBC 7.6 ± 1.3 vs. 9.8 ± 0.5; PLT 893.6 ± 387.0 vs. 1095.1 ± 378.2). In differential white blood cell counts, granulocyte and monocyte numbers were moderately increased (lymphocytes 1.5 ± 0.8 vs. 1.7 ± 1.0; monocytes 0.7 ± 0.4 vs. 0.4 ± 0.2; granulocytes 2.7 ± 1.6 vs. 1.4 ± 1.1).

***Mll-AF4* fusion transcript is present in lymphomas**

Tumor incidence in the *Mll-AF4* cohorts was dependent on the *Cre*-expression allele (Fig. **21**) but there was only a significant difference in time of onset of tumorigenesis between mice with the invertor cassette flanked by wild type or mutant alleles *loxP* in the case of the *Rag-1-Cre* mice. Filter hybridization analysis of genomic DNA was used to establish the occurrence of *AF4* inversion in spleen, thymus, lymph node, liver, and kidney (tail biopsy DNA was used as control to show the germline and knock-in *Mll* alleles) of mice that succumbed to lymphoma. The inverted segment was readily

detected in spleen, thymus and lymph node and the intensity of the inverted allele correlated with the level of tumor infiltration in the respective tissues.

Transcription of the resulting *Mll-AF4* fusion gene was confirmed by RT-PCR with primers positioned in exon 10 of *Mll* and exon 5 of *AF4*.. After inversion of the *AF4* cassette, chimeric mRNA is transcribed and spliced to give *Mll-AF4* fusion mRNA representative of the fusion product found in human leukemia. Due to higher sensitivity of RT-PCR compared to filter hybridization, fusion transcript was also detected in low grade infiltrated tissues such as the liver and kidney. The RT-PCR products were cloned and sequenced to demonstrate the in-frame fusion of *Mll* and *AF4* exons as shown for transcripts in the targeted ES cells..

Specific Cre expression in defined hematopoietic populations has been demonstrated previously using a highly sensitive reporter assay for Cre-mediated deletion (McCormack et al., 2003). In brief, *Rag-1* Cre is highly expressed in both, B- and T-progenitor subsets, but is not detectable in Mac1+ cells. *Lck-Cre* is exclusively expressed in the T-cell compartment from the DN1 stage onwards. There was no evidence of Cre activity in cells sorted for the expression of CD34, Sca1, Ter119, Mac-1 or Gr-1. Further, no Cre activity could be detected after selection of cells expressing B cell markers (CD19, B220, IgM, IgD or IgG). These results with *Lck-Cre* were confirmed by flow cytometry using the ROSA-YFP lox-STOP deletion (Srinivas et al., 2001) as a fluorescent gene reporter assay (L. Drynan & THR unpubl). *CD19-Cre* expression is limited to B-cells, with different levels according to the respective differentiation state. Single and double positive T-cells, Ter119+, and Mac-1+ populations were Cre negative.

Specificity of Cre activity was additionally assessed by differential PCR to estimate the efficiency of inversion in sorted subsets of spleen cells from young asymptomatic invertor mice. *Mll-AF4 lox66/71; Rag-Cre* mice showed inversion of about 65% unsorted spleen cells, whereas in the Mac1+ fraction inverted alleles were only found at low frequency, attributable to the level of contaminating lymphocytes in the sorted populations. Inversion was induced almost equally in B220+ and Thy1.2+ cells. By contrast, *Mll-AF4 lox66/71; Lck-Cre* mice showed inversion exclusively in

Thy1.2+ cells. Efficiency of inversion was higher than in *Mll-AF4 lox66/71; CD19* mice, in which Cre expression was tightly restricted to B220+ lymphocytes.

***Mll-AF4* invertors develop diffuse large B cell lymphoma.**

The post-mortem examination of the three types of invertor mice showed these animals all developed lymphoid neoplasms. Histology and immunohistochemistry classified the tumors, by the Bethesda system (Morse et al., 2002), uniformly as diffuse large cell lymphoma of the centroblastic subtype, typically involving spleen, thymus, and mesenteric lymph nodes, although some mice had only splenomegaly, without mesenteric lymph node enlargement. The infiltration of non-lymphatic organs such as liver and kidney was more variable and dependent on the stage of disease at the time of post-mortem examination. Bone marrow was rarely affected. Enlarged mesenteric lymph nodes were full of lymphoid cells that could be identified as B cells by their staining with the anti-B220 B cell marker. The lymphoma cells were characterized by their large vesicular nuclei, clumping chromatin, and 1 to 5 prominent nucleoli. Numerous mitoses and many apoptotic cells were present in tumor areas.

Immunohistochemistry was used to demonstrate the B-cell phenotype of the lymphoma cells. These were large B cell centroblastic lymphomas (DLBCL) and typically, were infiltrated by small reactive T cells to some level. Myeloid cells were absent in tumor areas, as judged by the absence of cells expressing the F4/80 myeloid marker in the lymphomas. In a minority of tumors, small areas of histiocytic infiltrates were observed in a pattern compatible with histiocytic subtype of DLBCL. There were no differences between *Rag*, *Lck*, or *CD19-Cre* expressing mice in respect of lymphoma morphology and immunophenotype.

The B cell phenotype of the *Mll-AF4* invertor tumors was confirmed by studying the status of the immunoglobulin heavy chain and T cell receptor genes. Overall, about 60% of the lymphoma infiltrated spleens or lymph nodes analyzed showed clonal *Igh* rearrangements of at least one allele. The remaining germline band originated from residual normal cells or reactive T cells in spleens, which were only

partially replaced by lymphoma cells or reactive T-cells in the mesenteric lymphoma tissue. In addition, two samples displayed one *Tcr* rearrangement.

***Mll-AF4* invertor B cell lymphomas are transplantable tumors**

The B cell tumors arising in the *Mll-AF4* invertor mice are invasive tumors, extensively occupying sites outside of the primary lymphoid organs, such as perivascular deposits in liver and kidney. Lymphadenopathy was a characteristic feature found in the mice. This tumor profile suggested that these could be serially transplanted into syngeneic mice. Single cell suspensions of lymphoma infiltrated spleens and mesenteric lymph nodes of *CD19-Cre* were tested in transplantation assays using *Rag-1* null recipient mice, which lack lymphocytes (Mombaerts et al., 1992). Single cell suspensions, purified by density gradient centrifugation (comprising a population of B lymphoma cells together with reactive T-cells and low numbers of Mac-1-positive cells), were injected into a *Rag-1-/-* mouse. FACS analysis of the donor cells from the *Mll-AF4; CD19-Cre* lymphoma showed 50% B220 (B cell), 34% Thy1.2 (T cell) and 12% Mac1 (myeloid cell). A transplanted tumor arose in the recipient *Rag-1 -/-* mouse in just under 2 months and was exemplified by splenomegaly and extensive infiltration of liver and kidneys. FACS analysis of the tumor cell population showed that there had been expansion of only the lymphomatous B cells and very few T and myeloid cells remained. Secondary transfer of these splenic tumor cells into secondary *Rag-1-/-* recipients (four mice were used) gave rise to tumors in 17-23 days and the surface marker phenotype was explicitly that of a B cell neoplasm. The correspondence of the primary tumor with that of the transplanted tumors in the *Rag-1-/-* recipients was investigated by analysis of clonal *Igh* gene rearrangement.. Filter hybridization showed a faint band corresponding to clonal rearranged *Igh* allele in the DNA from the invertor tumor from lymph node and spleen but this rearranged band became prominent in the first and second *Rag-1-/-* recipient mouse tumors. No rearrangement of the *Tcrb* gene was found.

Discussion.

The invertor technology has allowed the first generation of B cell specific models for MLL-AF4. This is a valuable addition to the repertoire of MLL-models allowing studies of secondary mutations co-operating with MLL-fusions, lineage specific target genes and an important element to development of new therapies to target MLL-leukemia. This last point is of particular clinical relevance as the prognosis of MLL-AF4 associated leukemia is very poor, especially if it occurs before 1 year of age (Pui et al., 2002) and new drugs are needed to treat this leukemia.

B cell tumors in *Mll-AF4* invertor mice

In human *MLL-AF4* translocations, cytogenetic studies have indicated that the crucial event lies on the der(11) chromosome. Similar conclusions apply to MLL-AF9 positive leukemia (Huret et al., 2001). In the *Mll-AF9* knock-in mouse model, in which only expression of the Mll-AF9 fusion product occurs, it was sufficient to induce myeloid leukemia (Corral et al., 1996; Dobson et al., 1999). We made an invertor conditional *Mll-AF4* mouse model in which expression of the *Mll-AF4* fusion corresponds to the human der(11) product.

We used three distinct lines of Cre expressing mice to facilitate the inversion of the *AF4* cassette and an identical lymphoma occurred in all three situations with high penetrance. A long latency period was observed in all three cases and little difference was found when the invertor cassette was flanked by wild type *loxP* sites or the mutant *loxP* sites, except for the *Mll-AF4; Rag-1-Cre* in which the mutant *loxP* allele increases the rate of tumor formation somewhat. The long latency suggests, however, that expression of the fusion transcript alone is not sufficient for formation of overt lymphoma, but secondary mutations have to occur. This interpretation is supported by the detection of clonal *Igh* rearrangements in 60% of the cases. As Cre-mediated inversion and expression of the *Mll-AF4* chimaeric mRNA will occur in a large number of cells, oligo- or polyclonal tumors with diverse immunoglobulin gene rearrangements might be expected to occur.

Cellular origin of Mll-AF4 leukemias

In infant MLL-AF4 positive leukemia, the tumors typically show the phenotype of the B-lymphocyte lineage, mainly immature pre/pro-B cell leukemias (Pui et al., 1991).

The tumors in the *Mll-AF4* invertor mice were always more mature B cells.. This difference may reflect the timing with which the *Mll-AF4* fusion is available to induce neoplasia. There is a high percentage of patients with t(4;11) translocations without clonal immunoglobulin heavy chain and T-cell receptor gene rearrangements suggesting that an immature progenitor cell is the target of the chromosomal translocation (Peham et al., 2002), supported by rare cases of lineage switch in some infants with ALL/t(4,11) (Ridge et al., 1995). Recent microarray data showed co-expression of many myeloid and monocytes/macrophage-specific genes in B-lineage ALL with *MLL* aberrations, suggesting that leukemia with *MLL*-translocations might originate from a bi-potential progenitor or perhaps MLL-AF4 protein drives the trans-differentiation of an early lymphocyte progenitor (Armstrong et al., 2002). However, characterization of sorted cells from leukemic bone marrow of childhood high-risk ALL/t(4;11) suggested an origin in a primitive lymphoid-restricted progenitor cell (Hotfilder et al., 2005). Moreover, studies in monozygotic twins with leukemia, carrying an identical *MLL-AF4* chromosomal translocation, but different immunoglobulin rearrangements strongly supported the model that the primary MLL-AF4 target cell was a progenitor cell at a differentiation stage before efficient *RAG* expression (Greaves et al., 2003; Mori et al., 2002). This issue could be addressed using the *Mll-AF4* invertor model and an inducible *Cre* driven by a promoter of a gene, such as *Lmo2,* expressed in bone marrow progenitor cells.

Our data show that Mll-AF4 fusion can be oncogenic in cells restricted to the lymphoid lineage (i.e. not only in uncommitted progenitors). The resultant tumors are transplantable and can infiltrate primary and secondary host tissue aggressively. All the lymphomas showed an identical phenotype, even *Mll-AF4; Lck-Cre* mice. Previous data showed that the *Lck-Cre* expression is tightly restricted to cells in the T cell developmental pathway from the DN1 stage onwards (McCormack et al., 2003) which is just a stage before endogenous Lck activity is apparent (Wolfer et al., 2001; Wolfer et al., 2002).

The intriguing possibility is that the expression of Mll-Af4 in a postulated non-committed T cell progenitor could also cause reassignment of the lineage, in this case to the B cell lineage. While the examination of such developmental alteration by the

Mll-Af4 fusion is outside the scope of the current establishment of a tumor system, second generation models should be able to address this question using Cre-expressing mice with the capability to cause *AF4* inversion at precisely defined times in hematopoiesis.

Generation and characterization of *Mll-AF4* invertor mice

The *AF4* cDNA segment from nt 1504 to nt 4304 (AccNo. NM_005935) was amplified from a human B-ALL sample RNA and cloned into the NotI site of pC2-neo (Forster et al., 2005). The entire fragment from the clone (*AF4-pA-MC1Neo-pA*) was excised as an Sfi1 fragment, the restriction site ends filled in and the fragment was introduced into the blunted BamHI site of pLOXKI (Forster et al., 2005) so that the AF4-pA-MC1Neo-pA sequences were immediately downstream of the short intron sequence and acceptor splice site in pLOXKI and all flanked by *loxP* sites (to give pLOXKI-AF4). The invertor cassette of pLOXKI-AF4 was cloned as a SceI-ClaI blunt fragment into the blunt-ended BglII site of the *Mll* targeting vector (a 5.5Kb EcoRI fragment (Collins et al., 2000)). Finally the HSV *tk* gene was added as a blunt end filled XbaI fragment into the blunted ClaI site of the *Mll* targeting vector. For electroporation into ES cells, the final construct was linearized with an XhoI within the polylinker sequence of the *Mll* targeting vector.

For the *Mll-AF4 lox66/71* version, a vector designated pLOXKI2 was constructed. The *lox66* sequence was introduced as oligonucleotides into SacII-NotI digested pMG2 (Forster et al., 2005) followed by the short segment of the *Af4* intron as an XbaI-BamHI PCR fragment. The *lox71* site was cloned as oligonucleotides into the EcoRV-HindIII sites to give pLOXKI2 (analogous to pLOXKI (Forster et al., 2005)). The AF4 cDNA fragment described above, was cloned as a blunt Sfi1-NotI into the blunt BamHI site of pLOXKI2 and the SV40pA-MC1-neo-pA fragment was blunt end cloned into the polylinker EcoRV site to yield pLOXKI2-AF4. This invertor cassette was cloned as a SceI-ClaI blunt fragment into the blunt-ended BglII site of the genomic *Mll* EcoRI fragment targeting vector. The HSV *tk* gene was added as a blunt XbaI fragment into the blunt ClaI site of the *Mll* targeting vector and linearized with XhoI for transfection into ES cells..

Mice carrying *Mll-AF4* loxP or *Mll-AF4* lox66/71 alleles were bred with those expressing Cre recombinase under the control of the *Rag-1*, *Lck*, or *CD19* promoter (McCormack et al., 2003). Animals were culled at the first signs of ill health. After post-mortem examination, samples for RNA and DNA extraction were snap frozen and tissues for histological analysis were fixed in 10% buffered formalin. Blood smears were stained with May-Grünwald-Giemsa (MGG) stain and an automated differential blood count from EDTA-blood was carried out (Animal Blood Counter, Vet abc). 4 μm sections of paraffin embedded tissues were stained with hematoxylin and eosin.

Tissue sections for immunohistochemistry were dried overnight at 60°C and subsequently deparaffinized in xylene, dehydrated in ethanol, and washed in water. Antigen retrieval was accomplished by placing slides in a pressure cooker at full pressure for 3 minutes in 0.01 M Citric Acid, pH 6.0. Sections were stained using a Dako TechMate 500 automated immunostainer with a primary (rat anti-mouse B220 (BD Pharmingen; 1/25 dilution), CD3 (Dako; 1/50 dilution), and F4/80 (Serotec; 1/50 dilution)) antibody incubation time of 45 minutes and a standard biotinylated secondary (goat anti-rat; 1/250 dilution) antibody for 30 minutes and a streptavidin tertiary complex for 30 minutes with a DAB visualisation system. Sections were counterstained with Mayer's haematoxylin, for a blue nuclear contrast to the DAB-derived brown immunopositive end-product.

Filter hybridization of SphI digested genomic DNA from various tissues was carried out using the *Mll* 5′probe (Collins et al., 2000) (Fig. 16A) to detect the inverted allele in the various organs. To estimate the proportion of inverted versus non-inverted *Mll-AF4* alleles in various cell populations, DNA was prepared (Qiagen Blood Kit) from 10,000 unsorted or FACS sorted cells expressing Mac-1, B220 or Thy1.2 Purity of the sorted populations was ≥97.5%. 1/10 of the DNA solution, primer MEF, the fluorescent primer AIR, and a primer set for the single copy gene *Lmo2* were used in a PCR reaction with the following conditions: 94°C 10 min, 35 cycles each of 94°C 30s, 66°C 30s, 72°C 60s and final extension at 72°C 7 min. PCR products were separated by capillary electrophoresis (ABI Prism 3100 Genetic Analyzer), quantified by

calculation of the area under the curve (Genescan Analysis Software, Applied Biosystems), and the ratios of inverted allele/*Lmo2* allele were calculated.

The *Mll-AF4* fusion transcript resulting from the inversion, post-transcriptional splicing and processing was detected by RT-PCR. Total RNA was extracted from tissues or sorted cells and reverse transcribed. 100 ng of cDNA was amplified in a PCR using nested primers primer sets positioned in *Mll* exon 10 and *AF4* exon 5 (Fig. 16B and C. PCR products were visualized on 2% agarose gels stained with ethidium bromide.

Flow cytometric analysis of cell surface marker expression

Single-cell suspensions were prepared from mouse thymus, spleen, lymph node, and bone marrow. Cells were stained with either PE or FITC conjugated antibodies to CD4, CD8, Thy1.2, Gr-1 (Ly-6G), Mac-1 (CD11b), B220 (CD45R), IgM, IgG, CD43, CD24, c-kit, Sca-1, and CD34, or matching isotope control antibodies (BD Pharmingen) and analyzed using a FACSCalibur flow cytometer and Cell Quest software (Becton-Dickinson).

Immunoglobulin and T-cell receptor gene rearrangement assays

Genomic DNA from various tissues was digested with HindIII (or EcoRI as in Fig. 6) and filter-hybridized with probes for *Igh* (Neuberger and Williams, 1986), *Tcr Jβ2*, and *Tcr Cβ1* (Malissen et al., 1984). A diagrammatic map of the respective regions is shown in Fig. 21A and B.

Cell transplantation into Rag-/- recipients

Single cell suspensions of tumor cells from infiltrated spleens or mesenteric lymph nodes were purified by density gradient centrifugation and injected intravenously at a dose of $5x10^6$ cells into Rag-/- mice (Mombaerts et al., 1992). Recipients were monitored and sacrificed at the first signs of lymphoma. Tissues were tested for the presence of *Igh* rearrangements by filter hybridization and immunophenotype using FACS analysis as described above.

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

CLAIMS

1. A method for generating a non-human animal model of a chromosomal rearrangement, which method comprises creating a transgenic non-human mammal wherein an invertor cassette (comprising a first nucleic acid sequence and one or more introns comprising one or more acceptor splice sites from a gene of interest, wherein the first nucleic acid sequence and the intronic sequence/s are flanked by site/s which recognise a site specific recombinase) is knocked into a target gene of interest with that animal so that the first nucleic acid sequence is in a reverse transcription orientation with respect to that gene of interest and wherein that animal expresses a site-specific recombinase under the control of a cell type-specific promoter, such that a chromosomal inversion is catalysed by the recombinase upon its expression.

2. A method according to claim 1 wherein upon chromosomal inversion and expression of the recombinase a fusion protein is expressed.

3. A method according to claim 2 wherein the fusion protein is *MLL-AF4*.

4. A method according to claim 2 wherein the fusion protein is *Ews-ERG*.

5. A method according to any preceding claim wherein the first nucleic acid sequence is a cDNA sequence.

6. The method according to any preceding claimwherein the site-specific recombinase is *Cre* and wherein the recognition site are *LoxP* sites.

7. A method according to any of claims 1 to 5 wherein the site-specific recombinase is *Flp* and wherein the recognition sites are *FRT* sites.

8. A method according to any preceding claim wherein the cell type specific promoter is *Rag*.

9. A method according to any preceding claim wherein the chromosomal rearrangement is tumourigenic.

10. A method according to claim 9 wherein the tumour is a leukemia.

11. An invertor cassette comprising a first nucleic acid sequence and one or more introns comprising one or more acceptor splice sites from a gene of interest, wherein the first nucleic acid sequence and the intronic sequence/s are flanked by site/s which recognise a site specific recombinase.

12. An invertor cassette according to claim 11, wherein the nucleic acid is cDNA.

13. An invertor cassette according to claim 11 or claim 12 which further comprises a poly A site.

14. An invertor cassette according to any of claims 11 to 13 which further comprises a segment for the selection of homologous recombinant clones, preferably a MC1-neo-pA seqment.

15. An invertor cassette according to any of claims 11 to 14 wherein the recognition sites are *loxP* sites and the recombinase is the *Cre* recombinase.

16. An invertor cassette according to any of claims 11 to 14 wherein the site-specific recombinase is *Flp* and wherein the recognition sites are *FRT* sites.

17. The use of the invertor method according to any of claims 1 to 10 or an invertor cassette according to any of claims 11 to 16 in the preparation of a non-human model of a chromosomal rearrangement.

18.A non-human animal tumour model with a chromosomal rearrangement, wherein the aminal model is obtained by the method of any of claims 1 to 10.

19. A non-human animal tumour model with a chromosomal rearrangement which is an Ews-ERG mouse model.

20. A non-human animal tumour model with a chromosomal rearrangement which is an MLL-AF4 mouse model.

21. A mouse model according to claim 18 or claim 19 or claim 20 obtainable using the method according to any of claims 1 to 10..

22. A non-human animal tumour model according to any of claims 18 to 21 which has leukemia and/or sarcoma.

23. A non-human animal tumour model according to any of claims 18 to 21 which has a T-lymphocyte tumour.

24 A.non-human animal tumour model according to claim 19 which has a T-lymphocyte tumour.

25. A non-human animal tumour model according to claim 20 which has a B-lymphocyte tumour.

26. A non-human animal tumour model according to claim 18 to 25 which has leukemia.

27.The use of a non-human animal tumour model according to any of claims 18 to 26 in the testing of potential anti-tumour compounds.

28. A method for the testing of an anti-tumour agent which method comprises the steps of:

(a) providing a non-human animal tumour model according to any of claims 18 to 26,

(b) administering to that model one or more potential anti-tumour agents; and

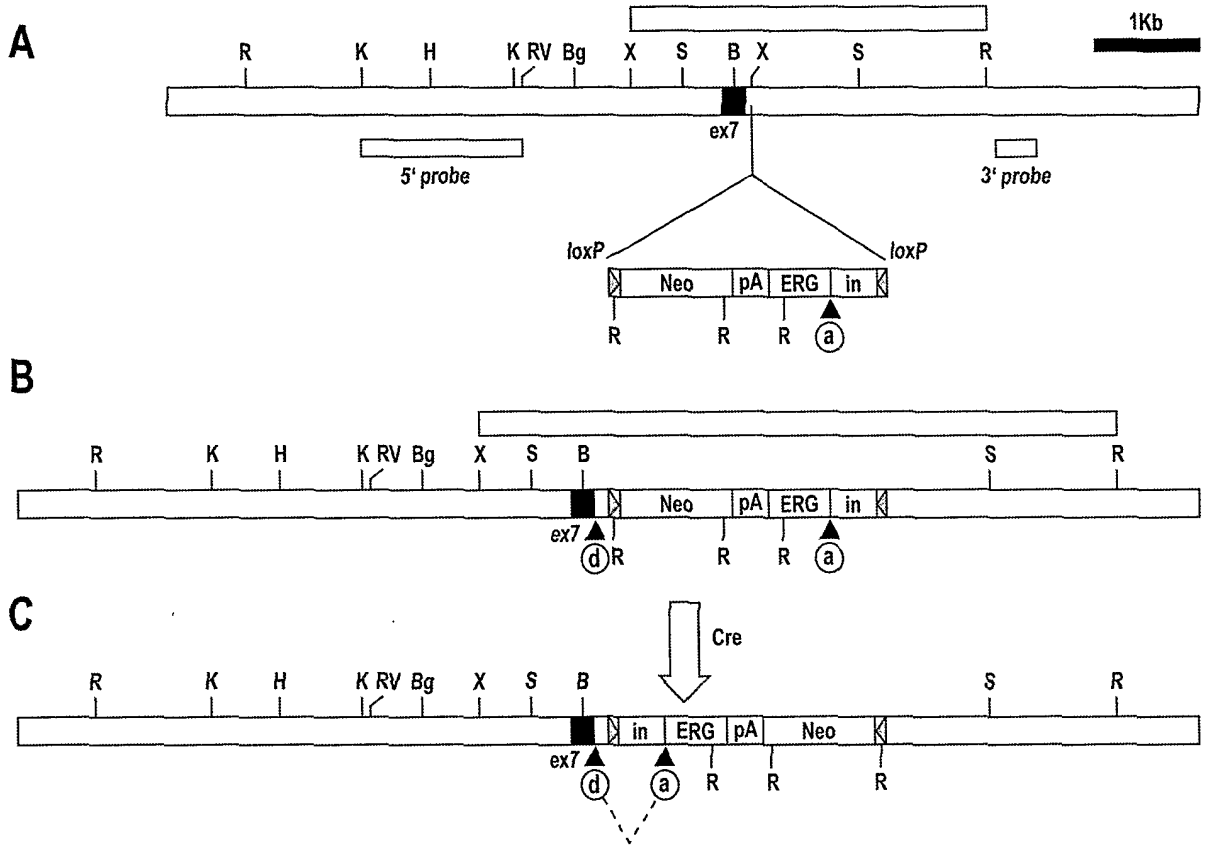
(c) testing for a change in one or more characteristics of tumour/s comprising the model according to step (a).

29. A method according to claim 28 wherein the model is an MLL-AF4 B-cell tumour model prepared according to any of claims 1 to 10.

30. A model according to claim 28 wherein the model is an ERG-Ews T-cell tumour model prepared according to any of claims 1 to 10.

31. A method according to claim 28 wherein step (c) (testing for a change in one or more characteristics of tumour/s comprising the model according to step (a)) involving the testing for a change in any one or more characteristics of one or more tumours in the group consisting of the following: size, morphology, presence and/or levels of one or more tumour markers and cytological assays.

FIG. 1



D

Ews exon 7

Q Q S S S Y G Q Q

AACAGAGCAGCAGCTACGGGCAGCAGAG^TGAGTTGCTAAGAGAGAAAACTAAGTATGAGTGATTTTGCTTGGAGTT
TGTGAGACTGGGGAATGCTTGTTACCACTGTTTGCTTCTGGAGTCTAGCGATAAGCTTGATAATTCATCGGGCCGCA
TCTAGTCGACCTGCAGCCAAGCTAGCTTGGCTGGACGTAAACTCCTCTTCAGACCTAATAACTTCGTATAGCATACA
TTATACGAAGTTATATTAAGGGTTATTGAATATGATCGGAATTGGCCGCTCTAGAGCATATTAATTCGTAAGTGCTA
GAAACCTTGCCTTTAACTCAGATATTAGGGAGTGATTAAAGTTTTGCTTCTTAAGCCTGCTGACGGTGCCGCCAGGG
AAGGGTCAAATCTATTAGAACCCTCCC...TCCCTTCCTTTTCCAAATCTCTTTCTAGGGATCTGCGGGCCGCAGGC

d 1 a a a G

acceptor splice site *ERG*

E

Ews

TACTAGTTACCCCCCTCAGACTGGATCCTACAGCCAGGCTCCAAGTCAATATAGCCAACAGAGCAGCAGCTACGGGCAGCAG

T S Y P P Q T G S Y S Q A P S Q Y S Q Q S S S Y G Q Q

AGGGGATCTGCGGCCGCAGGCAGTGGCCAGATCCAGCTTTGGCAGTTCCTCCTGGAGCTCCTGTCGGACAGCTCCAACTCCAGC

r g s a a a G S G Q I Q L W Q F L L E L L S D S S N S S

ERG

FIG. 2

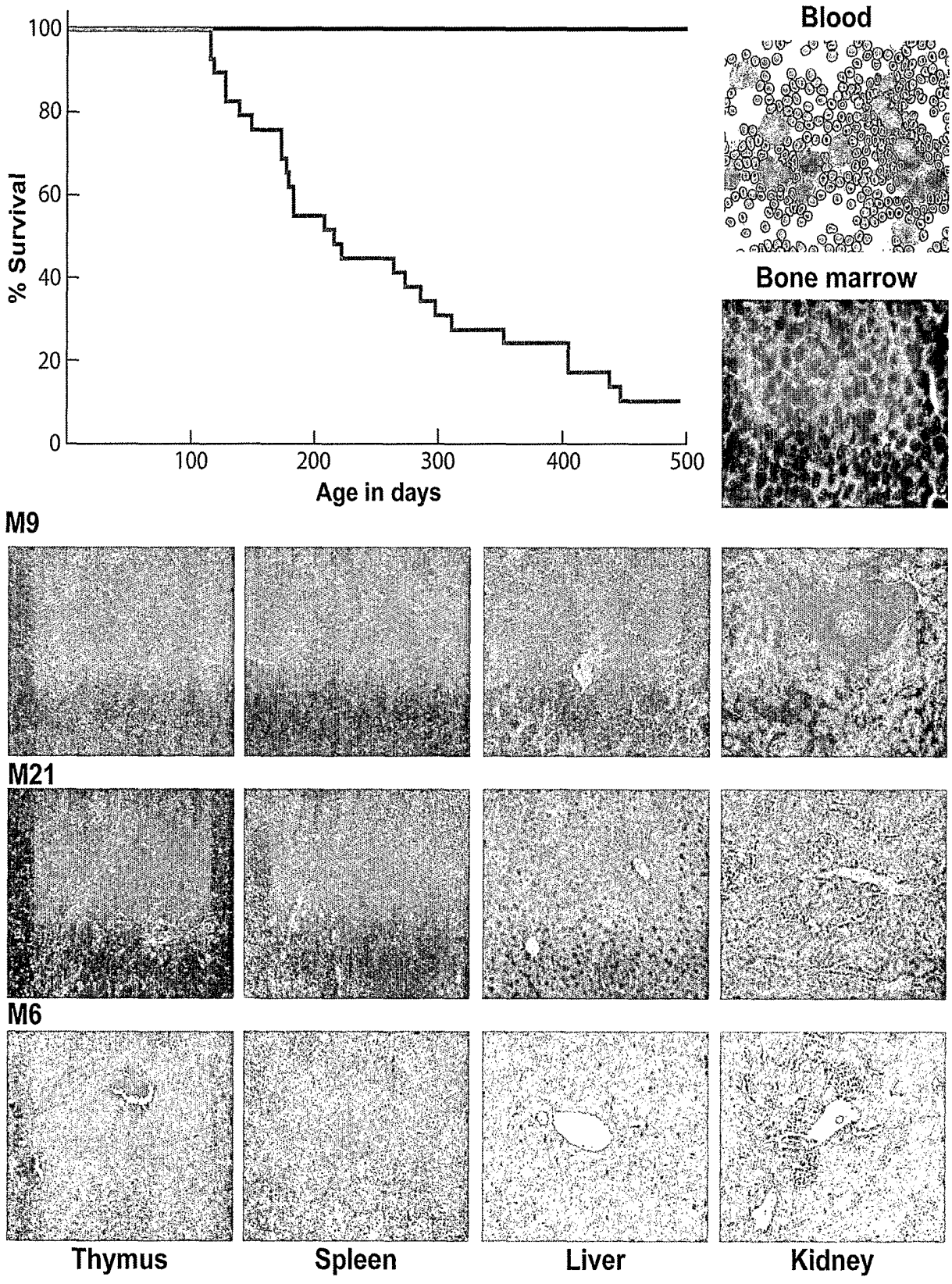

100
80
60
40
20
0
% Survival
100
200
300
400
500
Age in days
Blood
Bone marrow
M9
M21
M6
Thymus
Spleen
Liver
Kidney

FIG. 3

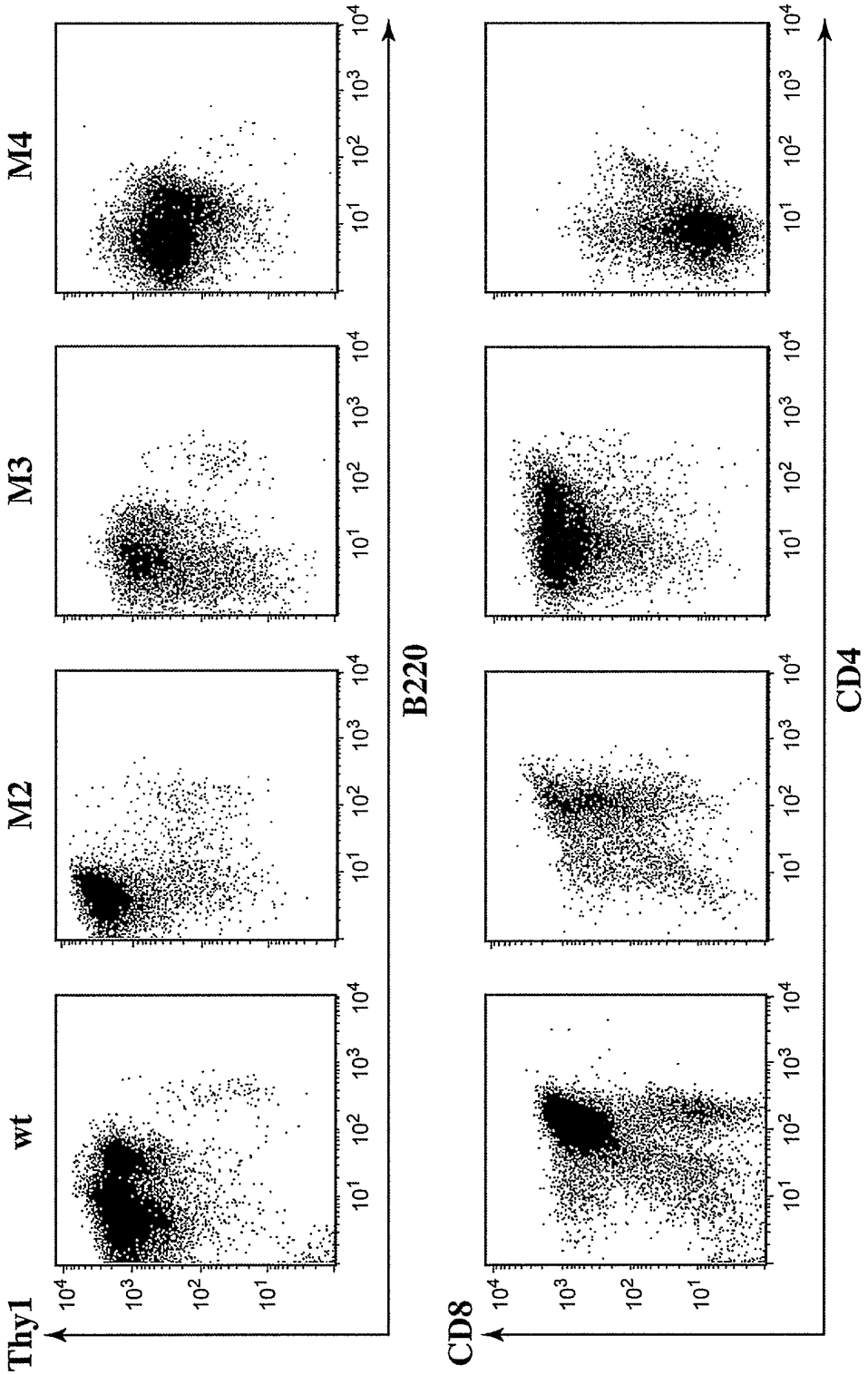


FIG. 4

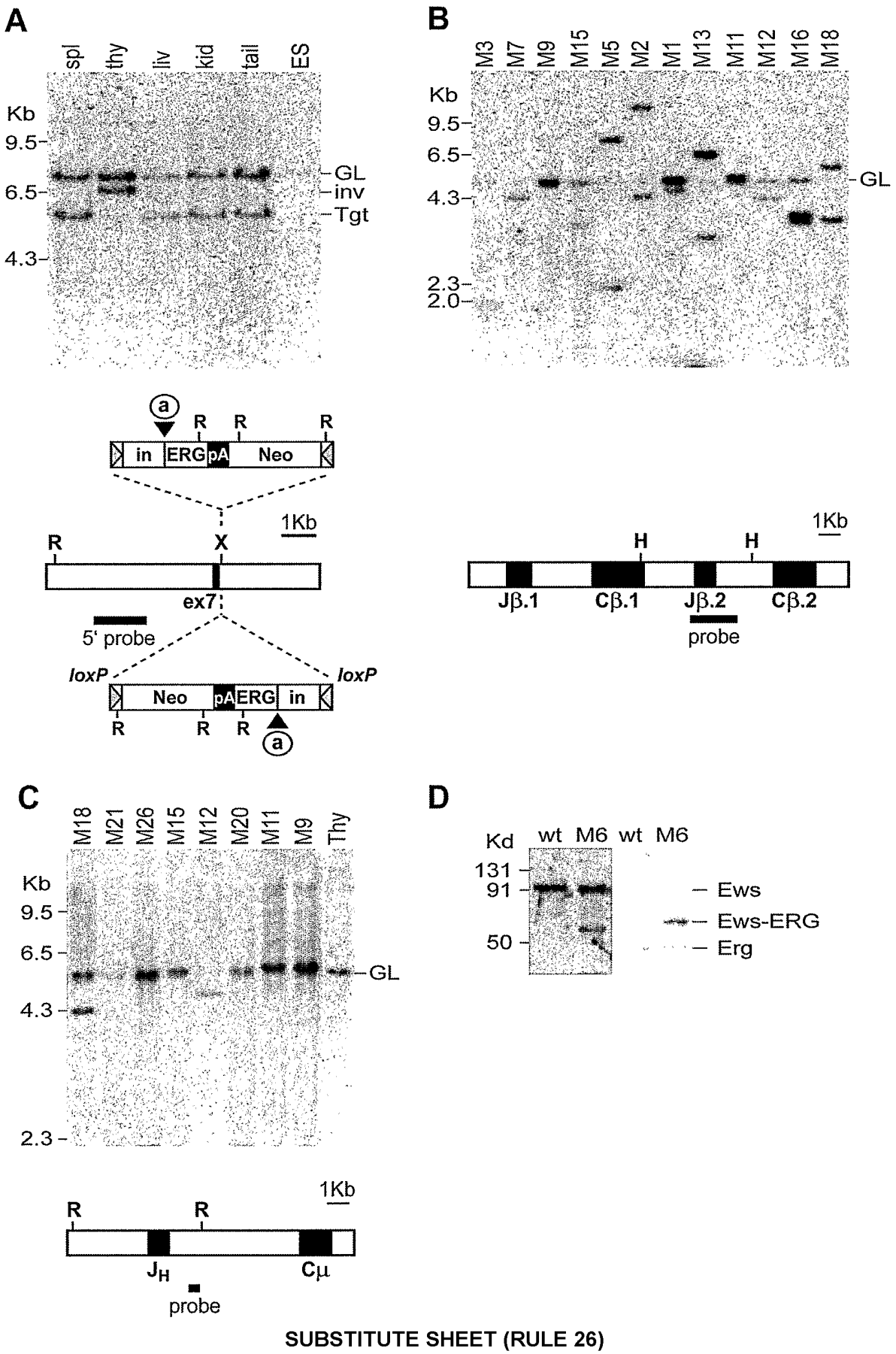
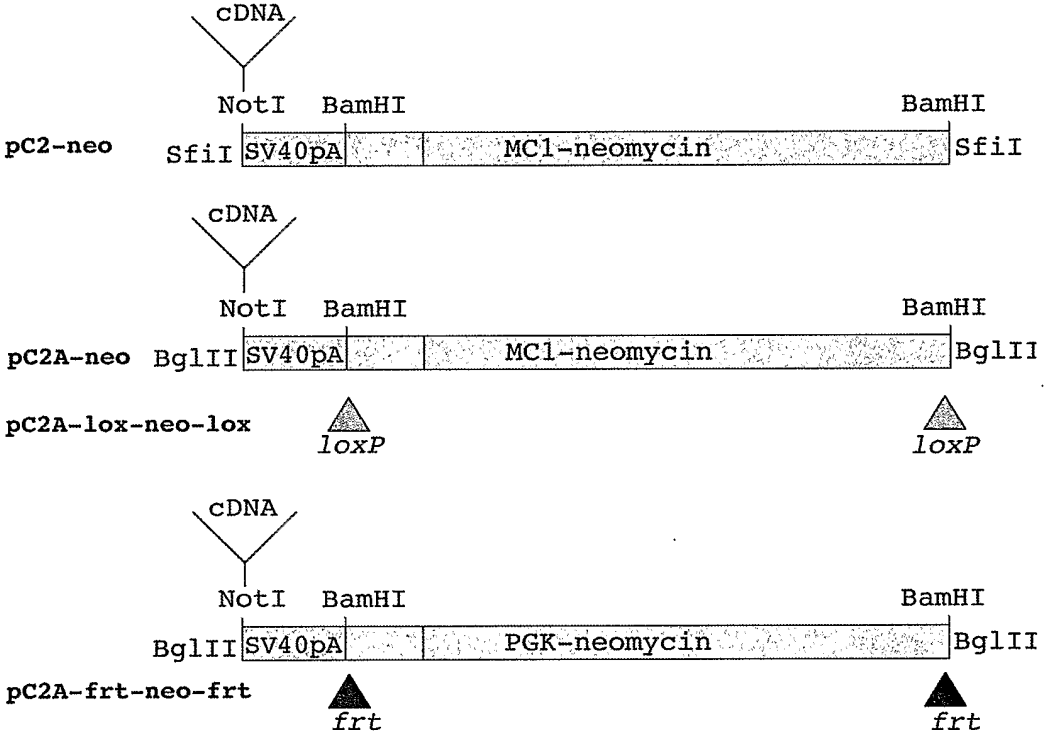

A
spl
thy
liv
kid
tail
ES
Kb
9.5
6.5
4.3
GL
inv
Tgt
B
M3
M7
M9
M15
M5
M2
M1
M13
M11
M12
M16
M18
Kb
9.5
6.5
4.3
2.3
2.0
GL
a
R
R
R
in
ERG
pA
Neo
1Kb
R
X
ex7
5' probe
loxP
loxP
Neo
pA
ERG
in
R
R
R
a
1Kb
H
H
Jβ.1
Cβ.1
Jβ.2
Cβ.2
probe
C
M18
M21
M26
M15
M12
M20
M11
M9
Thy
Kb
9.5
6.5
4.3
2.3
GL
1Kb
R
R
JH
Cμ
probe
D
Kd
wt
M6
wt
M6
131
91
50
Ews
Ews-ERG
Erg

FIG. 5

A. Maps of pC2-neo, pC2A-neo, pC2A-frt-neo-frt



B: Partial sequence of polylinker in pBspt-C2

(NotI) NotI
SfiI BamHI
gcggccatcctggccgcggccgcggatccggccatcctggcc

C: Partial sequence of polylinker in pBSpt-C2A

BglII NotI BamHI BglII
agatctgcggccgcggatccagatct

D: sequence of SV40 polyA

NotI
gcggccgc forward primer->
tctgttcatgatcataatcagccataccacatttgtagaggttttacttgctttaaaaaa 60
agacaagtactagtattagtcggtatggtgtaaacatctccaaaatgaacgaaatttttt

cctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaactt 120
ggagggtgtggaggggggacttggactttgtattttacttacgttaacaacaacaattgaa

gtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataa 180
caaataacgtcgaatattaccaatgtttatttcgttatcgtagtgtttaaagtgtttatt

agcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatca 240
tcgtaaaaaaagtgacgtaagatcaacaccaaacaggtttgagtagttacatagaatagt

BamHI
tgtctggatcc
acagacctagg <- back primer

FIG. 6

A. Sequence of mouse *FliI* RT-PCR fragment (with NotI ends for cloning into cassette 2A)

```
        10        20        30        40        50
gcggccgcAGACCCTTCTTATGACTCTGTCAGGAGAGGAGCATGGAACAA
cggccggcTCTGGGAAGAATACTGAGACAGTCCTCTCCTCGTACCTTGTT
 a  a  a  D  P  S  Y  D  S  V  R  R  G  A  W  N  N

        60        70        80        90        100
TAATATGAACTCTGGCCTCAACAAAAGTCCTCTCCTTGGAGGATCACAGA
ATTATACTTGAGACCGGAGTTGTTTTCAGGAGAGGAACCTCCTAGTGTCT
  N  M  N  S  G  L  N  K  S  P  L  L  G  G  S  Q

       110       120       130       140       150
CCATGGGCAAGAACACTGAGCAGCGGCCCCAGCCAGATCCTTATCAGATC
GGTACCCGTTCTTGTGACTCGTCGCCGGGGTCGGTCTAGGAATAGTCTAG
T  M  G  K  N  T  E  Q  R  P  Q  P  D  P  Y  Q  I

       160       170       180       190       200
CTGGGGCCAACCAGCAGCCGCCTAGCAAACCCTGGGAGTGGGCAGATCCA
GACCCCGGTTGGTCGTCGGCGGATCGTTTGGGACCCTCACCCGTCTAGGT
 L  G  P  T  S  S  R  L  A  N  P  G  S  G  Q  I  Q

       210       220       230       240       250
GCTGTGGCAGTTTCTCCTGGAACTACTGTCCGACAGCGCCAACGCCAGCT
CGACACCGTCAAAGAGGACCTTGATGACAGGCTGTCGCGGTTGCGGTCGA
  L  W  Q  F  L  L  E  L  L  S  D  S  A  N  A  S

       260       270       280       290       300
GTATCACCTGGGAGGGGACCAACGGGGAGTTCAAAATGACGGACCCTGAT
CATAGTGGACCCTCCCCTGGTTGCCCCTCAAGTTTTACTGCCTGGGACTA
C  I  T  W  E  G  T  N  G  E  F  K  M  T  D  P  D

       310       320       330       340       350
GAGGTGGCCAGGCGCTGGGGAGAGCGGAAGAGCAAGCCCAACATGAATTA
CTCCACCGGTCCGCGACCCCTCTCGCCTTCTCGTTCGGGTTGTACTTAAT
 E  V  A  R  R  W  G  E  R  K  S  K  P  N  M  N  Y

       360       370       380       390       400
TGACAAGCTGAGCCGGGCCCTCCGATACTACTATGACAAAAACATTATGA
ACTGTTCGACTCGGCCCGGGAGGCTATGATGATACTGTTTTTGTAATACT
  D  K  L  S  R  A  L  R  Y  Y  Y  D  K  N  I  M

       410       420       430       440       450
CCAAAGTGCATGGCAAAAGGTATGCCTACAAGTTTGACTTCCATGGCATT
GGTTTCACGTACCGTTTTCCATACGGATGTTCAAACTGAAGGTACCGTAA
T  K  V  H  G  K  R  Y  A  Y  K  F  D  F  H  G  I

       460       470       480       490       500
GCCCAGGCCCTGCAGCCACATCCAACAGAGACATCCATGTACAAGTATCC
CGGGTCCGGGACGTCGGTGTAGGTTGTCTCTGTAGGTACATGTTCATAGG
 A  Q  A  L  Q  P  H  P  T  E  T  S  M  Y  K  Y  P

       510       520       530       540       550
CTCTGATATCTCCTACATGCCTTCCTACCATGCCCATCAACAGAAGGTGA
GAGACTATAGAGGATGTACGGAAGGATGGTACGGGTAGTTGTCTTCCACT
  S  D  I  S  Y  M  P  S  Y  H  A  H  Q  Q  K  V
```

FIG. 6 CONT'D

```
       560        570        580        590        600
ACTTTGTCCCGTCTCACCCATCCTCCATGCCTGTCACCTCCTCCAGCTTC
TGAAACAGGGCAGAGTGGGTAGGAGGTACGGACAGTGGAGGAGGTCGAAG
N  F  V  P  S  H  P  S  S  M  P  V  T  S  S  S  F

       610        620        630        640        650
TTTGGAGCAGCATCACAATACTGGACCTCCCCCACTGCTGGGATCTATCC
AAACCTCGTCGTAGTGTTATGACCTGGAGGGGGTGACGACCCTAGATAGG
 F  G  A  A  S  Q  Y  W  T  S  P  T  A  G  I  Y  P

       660        670        680        690        700
AAACCCCAGTGTCCCCCGCCATCCTAACACCCACGTGCCTTCACACTTAG
TTTGGGGTCACAGGGGGCGGTAGGATTGTGGGTGCACGGAAGTGTGAATC
  N  P  S  V  P  R  H  P  N  T  H  V  P  S  H  L

       710        720
GCAGCTACTACTAGcggccgc
CGTCGATGATGATCgccggcg
G  S  Y  Y  *
```

B. Sequence of human *AF9* RT-PCR fragment (NotI ends for cloning into cassette 2A)

```
        10         20         30         40         50
gcggccgcTCTTGAAGTGAAAAGTCCAATAAAGCAAAGCAAATCAGATAA
cgccggcgAGAACTTCACTTTTCAGGTTATTTCGTTTCGTTTAGTCTATT
 a  a  a  L  E  V  K  S  P  I  K  Q  S  K  S  D  K

        60         70         80         90        100
GCAAATAAAGAATGGTGAATGTGACAAGGCATACCTAGATGAACTGGTAG
CGTTTATTTCTTACCACTTACACTGTTCCGTATGGATCTACTTGACCATC
  Q  I  K  N  G  E  C  D  K  A  Y  L  D  E  L  V

       110        120        130        140        150
AGCTTCACAGAAGGTTAATGACATTGAGAGAAAGACACATTCTGCAGCAG
TCGAAGTGTCTTCCAATTACTGTAACTCTCTTTCTGTGTAAGACGTCGTC
E  L  H  R  R  L  M  T  L  R  E  R  H  I  L  Q  Q

       160        170        180        190        200
ATCGTGAACCTTATAGAAGAAACTGGACACTTTCATATCACAAACACAAC
TAGCACTTGGAATATCTTCTTTGACCTGTGAAAGTATAGTGTTTGTGTTG
 I  V  N  L  I  E  E  T  G  H  F  H  I  T  N  T  T

       210        220        230        240        250
ATTTGATTTTGATCTTTGCTCGCTGGACAAAACCACAGTCCGTAAACTAC
TAAACTAAAACTAGAAACGAGCGACCTGTTTTGGTGTCAGGCATTTGATG
  F  D  F  D  L  C  S  L  D  K  T  T  V  R  K  L

       260        270        280        290
AGAGTTACCTGGAAACATCTGGAACATCCTGAgcggccgc
TCTCAATGGACCTTTGTAGACCTTGTAGGACTcgccggcg
Q  S  Y  L  E  T  S  G  T  S  *
```

FIG. 6 CONT'D

C. Sequence of mouse *Chop* RT-PCR fragment (NotI ends for cloning into cassette 2A)

```
        10        20        30        40        50
gcggccgcCGTGTTCCAGAAGGAAGTGCATCTTCATACACCACCACACCT
cgccggcgGCACAAGGTCTTCCTTCACGTAGAAGTATGTGGTGGTGTGGA
 a  a  a  V  F  Q  K  E  V  H  L  H  T  P  P  H  L

        60        70        80        90       100
GAAAGCAGAACCTGGTCCACGTGCAGTCATGGCAGCTGAGTCCCTGCCTT
CTTTCGTCTTGGACCAGGTGCACGTCAGTACCGTCGACTCAGGGACGGAA
  K  A  E  P  G  P  R  A  V  M  A  A  E  S  L  P

       110       120       130       140       150
TCACCTTGGAGACGGTGTCCAGCTGGGAGCTGGAAGCCTGGTATGAGGAT
AGTGGAACCTCTGCCACAGGTCGACCCTCGACCTTCGGACCATACTCCTA
F  T  L  E  T  V  S  S  W  E  L  E  A  W  Y  E  D

       160       170       180       190       200
CTGCAGGAGGTCCTGTCCTCAGATGAAATTGGGGGCACCTATATCTCATC
GACGTCCTCCAGGACAGGAGTCTACTTTAACCCCCGTGGATATAGAGTAG
 L  Q  E  V  L  S  S  D  E  I  G  G  T  Y  I  S  S

       210       220       230       240       250
CCCAGGAAACGAAGAGGAAGAATCAAAAACCTTCACTACTCTTGACCCTG
GGGTCCTTTGCTTCTCCTTCTTAGTTTTTGGAAGTGATGAGAACTGGGAC
  P  G  N  E  E  E  E  S  K  T  F  T  T  L  D  P

       260       270       280       290       300
CGTCCCTAGCTTGGCTGACAGAGGAGCCAGGGCCAACAGAGGTCACACGC
GCAGGGATCGAACCGACTGTCTCCTCGGTCCCGGTTGTCTCCAGTGTGCG
A  S  L  A  W  L  T  E  E  P  G  P  T  E  V  T  R

       310       320       330       340       350
ACATCCCAAAGCCCTCGCTCTCCAGATTCCAGTCAGAGTTCTATGGCCCA
TGTAGGGTTTCGGGAGCGAGAGGTCTAAGGTCAGTCTCAAGATACCGGGT
 T  S  Q  S  P  R  S  P  D  S  S  Q  S  S  M  A  Q

       360       370       380       390       400
GGAGGAAGAGGAGGAAGAGCAAGGAAGAACTAGGAAACGGAAACAGAGTG
CCTCCTTCTCCTCCTTCTCGTTCCTTCTTGATCCTTTGCCTTTGTCTCAC
  E  E  E  E  E  E  Q  G  R  T  R  K  R  K  Q  S

       410       420       430       440       450
GTCAGTGCCCAGCCCGGCCTGGGAAGCAACGCATGAAGGAGAAGGAGCAG
CAGTCACGGGTCGGGCCGGACCCTTCGTTGCGTACTTCCTCTTCCTCGTC
G  Q  C  P  A  R  P  G  K  Q  R  M  K  E  K  E  Q

       460       470       480       490       500
GAGAACGAGCGGAAAGTGGCACAGCTAGCTGAAGAGAACGAGCGGCTCAA
CTCTTGCTCGCCTTTCACCGTGTCGATCGACTTCTCTTGCTCGCCGAGTT
 E  N  E  R  K  V  A  Q  L  A  E  E  N  E  R  L  K

       510       520       530       540       550
GCAGGAAATCGAGCGCCTGACCAGGGAGGTGGAGACCACACGGCGGGCTC
CGTCCTTTAGCTCGCGGACTGGTCCCTCCACCTCTGGTGTGCCGCCCGAG
  Q  E  I  E  R  L  T  R  E  V  E  T  T  R  R  A
```

FIG. 6 CONT'D

```
       560        570        580        590        600
TGATCGACCGCATGGTCAGCCTGCACCAAGCATGAACAGTGGGCATCACC
ACTAGCTGGCGTACCAGTCGGACGTGGTTCGTACTTGTCACCCGTAGTGG
L  I  D  R  M  V  S  L  H  Q  A  *  T  V  G  I  T

       610        620        630        640        650
TCCTGTCTGTCTCTCCGGAAGTGTACCCAGCACCATCGCGCCAGCGCCAA
AGGACAGACAGAGAGGCCTTCACATGGGTCGTGGTAGCGCGGTCGCGGTT
 S  C  L  S  L  R  K  C  T  Q  H  H  R  A  S  A  K

       660        670        680        690        700
GCATGTGACCCTGCACTGCACTGCACATGCTGAGGAGGGGACTGAGGGTA
CGTACACTGGGACGTGACGTGACGTGTACGACTCCTCCCCTGACTCCCAT
  H  V  T  L  H  C  T  A  H  A  E  E  G  T  E  G

       710        720        730
GACCAGGAGAGGGCTCGGCTTGCACATAgcggccgc
CTGGTCCTCTCCCGAGCCGAACGTGTATcgccggcg
R  P  G  E  G  S  A  C  T  *
```

D. Sequence of human *ATF* RT-PCR fragment (NotI ends for cloning into cassette 2A)

```
        10         20         30         40         50
gcggccgcAAAAATTTTGAAAGACTTATCTTCTGAAGATACACGGGGCAG
cgccggcgTTTTTAAAACTTTCTGAATAGAAGACTTCTATGTGCCCCGTC
 a  a  a  K  I  L  K  D  L  S  S  E  D  T  R  G  R

        60         70         80         90        100
AAAAGGAGACGGAGAAAATTCTGGAGTTTCTGCTGCTGTCACTTCTATGT
TTTTCCTCTGCCTCTTTTAAGACCTCAAAGACGACGACAGTGAAGATACA
  K  G  D  G  E  N  S  G  V  S  A  A  V  T  S  M

       110        120        130        140        150
CTGTTCCAACTCCCATCTATCAGACTAGCAGCGGACAGTACATTGCCATT
GACAAGGTTGAGGGTAGATAGTCTGATCGTCGCCTGTCATGTAACGGTAA
S  V  P  T  P  I  Y  Q  T  S  S  G  Q  Y  I  A  I

       160        170        180        190        200
GCCCCAAATGGAGCCTTACAGTTGGCAAGTCCAGGCACAGATGGAGTACA
CGGGGTTTACCTCGGAATGTCAACCGTTCAGGTCCGTGTCTACCTCATGT
 A  P  N  G  A  L  Q  L  A  S  P  G  T  D  G  V  Q

       210        220        230        240        250
GGGACTTCAGACATTAACCATGACAAATTCAGGCAGTACTCAGCAAGGTA
CCCTGAAGTCTGTAATTGGTACTGTTTAAGTCCGTCATGAGTCGTTCCAT
  G  L  Q  T  L  T  M  T  N  S  G  S  T  Q  Q  G

       260        270        280        290        300
CAACTATTCTTCAGTATGCACAGACCTCTGATGGACAGCAGATACTTGTG
GTTGATAAGAAGTCATACGTGTCTGGAGACTACCTGTCGTCTATGAACAC
T  T  I  L  Q  Y  A  Q  T  S  D  G  Q  Q  I  L  V
```

FIG. 6 CONT'D

```
       310        320        330        340        350
CCCAGCAATCAGGTGGTCGTACAAACTGCATCAGGAGATATGCAAACATA
GGGTCGTTAGTCCACCAGCATGTTTGACGTAGTCCTCTATACGTTTGTAT
 P  S  N  Q  V  V  V  Q  T  A  S  G  D  M  Q  T  Y

       360        370        380        390        400
TCAGATCCGAACTACACCTTCAGCTACTTCTCTGCCACAAACTGTGGTGA
AGTCTAGGCTTGATGTGGAAGTCGATGAAGAGACGGTGTTTGACACCACT
  Q  I  R  T  T  P  S  A  T  S  L  P  Q  T  V  V

       410        420        430        440        450
TGACATCTCCTGTGACTCTCACCTCTCAGACAACTAAGACAGATGACCCC
ACTGTAGAGGACACTGAGAGTGGAGAGTCTGTTGATTCTGTCTACTGGGG
M  T  S  P  V  T  L  T  S  Q  T  T  K  T  D  D  P

       460        470        480        490        500
CAATTGAAAAGAGAAATAAGGTTAATGAAAAACAGAGAAGCTGCTCGAGA
GTTAACTTTTCTCTTTATTCCAATTACTTTTTGTCTCTTCGACGAGCTCT
 Q  L  K  R  E  I  R  L  M  K  N  R  E  A  A  R  E

       510        520        530        540        550
ATGTCGCAGAAAGAAGAAAGAATATGTGAAATGCCTGGAAAACCGAGTTG
TACAGCGTCTTTCTTCTTTCTTATACACTTTACGGACCTTTTGGCTCAAC
  C  R  R  K  K  K  E  Y  V  K  C  L  E  N  R  V

       560        570        580        590        600
CAGTCCTGGAAAATCAAAATAAAACTCTAATAGAAGAGTTAAAAACTTTG
GTCAGGACCTTTTAGTTTTATTTTGAGATTATCTTCTCAATTTTTGAAAC
A  V  L  E  N  Q  N  K  T  L  I  E  E  L  K  T  L

       610        620        630
AAGGATCTTTATTCCAATAAAAGTGTTTGAgcggccgc
TTCCTAGAAATAAGGTTATTTTCACAAACTcgccggcg
 K  D  L  Y  S  N  K  S  V  *
```

E. Sequence of human *ERG* RT-PCR fragment (NotI ends for cloning into cassette 2A)

```
10         20         30         40         50
          gcggccgcAGGCAGTGGCCAGATCCAGCTTTGGCAGTTCCTCCTGGAGCT
          cgccggcgTCCGTCACCGGTCTAGGTCGAAACCGTCAAGGAGGACCTCGA
           a  a  a  G  S  G  Q  I  Q  L  W  Q  F  L  L  E  L

                 60         70         80         90        100
          CCTGTCGGACAGCTCCAACTCCAGCTGCATCACCTGGGAAGGCACCAACG
          GGACAGCCTGTCGAGGTTGAGGTCGACGTAGTGGACCCTTCCGTGGTTGC
            L  S  D  S  S  N  S  S  C  I  T  W  E  G  T  N

                110        120        130        140        150
          GGGAGTTCAAGATGACGGATCCCGACGAGGTGGCCCGGCGCTGGGGAGAG
          CCCTCAAGTTCTACTGCCTAGGGCTGCTCCACCGGGCCGCGACCCCTCTC
          G  E  F  K  M  T  D  P  D  E  V  A  R  R  W  G  E

                160        170        180        190        200
          CGGAAGAGCAAACCCAACATGAACTACGATAAGCTCAGCCGCGCCCTCCG
          GCCTTCTCGTTTGGGTTGTACTTGATGCTATTCGAGTCGGCGCGGGAGGC
           R  K  S  K  P  N  M  N  Y  D  K  L  S  R  A  L  R
```

FIG. 6 CONT'D

```
       210        220        230        240        250
TTACTACTATGACAAGAACATCATGACCAAGGTCCATGGGAAGCGCTACG
AATGATGATACTGTTCTTGTAGTACTGGTTCCAGGTACCCTTCGCGATGC
  Y  Y  Y  D  K  N  I  M  T  K  V  H  G  K  R  Y

       260        270        280        290        300
CCTACAAGTTCGACTTCCACGGGATCGCCCAGGCCCTCCAGCCCCACCCC
GGATGTTCAAGCTGAAGGTGCCCTAGCGGGTCCGGGAGGTCGGGGTGGGG
A  Y  K  F  D  F  H  G  I  A  Q  A  L  Q  P  H  P

       310        320        330        340        350
CCGGAGTCATCTCTGTACAAGTACCCCTCAGACCTCCCGTACATGGGCTC
GGCCTCAGTAGAGACATGTTCATGGGGAGTCTGGAGGGCATGTACCCGAG
 P  E  S  S  L  Y  K  Y  P  S  D  L  P  Y  M  G  S

       360        370        380        390        400
CTATCACGCCCACCCACAGAAGATGAACTTTGTGGCGCCCCACCCTCCAG
GATAGTGCGGGTGGGTGTCTTCTACTTGAAACACCGCGGGGTGGGAGGTC
  Y  H  A  H  P  Q  K  M  N  F  V  A  P  H  P  P

       410        420        430        440        450
CCCTCCCCGTGACATCTTCCAGTTTTTTTGCTGCCCCAAACCCATACTGG
GGGAGGGGCACTGTAGAAGGTCAAAAAAACGACGGGGTTTGGGTATGACC
A  L  P  V  T  S  S  S  F  F  A  A  P  N  P  Y  W

       460        470        480        490        500
AATTCACCAACTGGGGGTATATACCCCAACACTAGGCTCCCCACCAGCCA
TTAAGTGGTTGACCCCCATATATGGGGTTGTGATCCGAGGGGTGGTCGGT
 N  S  P  T  G  G  I  Y  P  N  T  R  L  P  T  S  H

       510        520        530        540
TATGCCTTCTCATCTGGGCACTTACTACTAAgcggccgc
ATACGGAAGAGTAGACCCGTGAATGATGATTcgccggcg
  M  P  S  H  L  G  T  Y  Y  *
```

F. Sequence of Ews-pC2A-neo junction

Ews exon 7

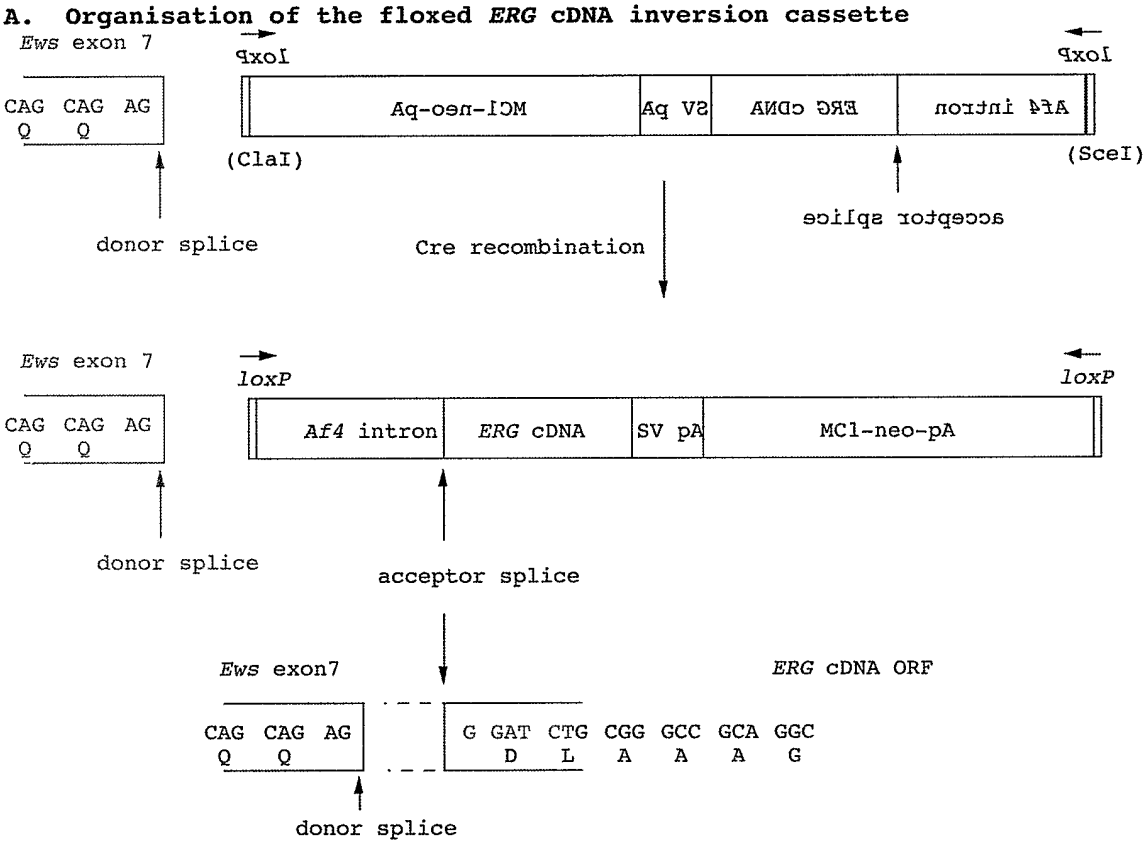
```
ACTAGTTACCCCCCTCAGACTGGATCC
 T  S  Y  P  P  Q  T  G  S
```

Ews exon 7 knock-in with *FliI* in pC2A

```
                    BamHI/BglII NotI
ACTAGTTACCCCCCTCAGACTGGATCTGCGGCCGCAGACCCTTCTTATGAC
 T  S  Y  P  P  Q  T  G  S  a  a  a  D  P  S  Y  D
```

FIG. 7

A. Organisation of the floxed *ERG* cDNA inversion cassette



B. pMG1 polylinker (pBSpt variant with one Sce1 restriction site)

```
                                                    XbaI
 SacI          SceI              SacII    NotI      _____
GAGCTCTAGGGATAACAGGGTAATCACCGCGGTGGCGGCCGCTCTAGAAC
CTCGAGATCCCTATTGTCCCATTAGTGGCGCCACCGCCGGCGAGATCTTG

     BamHI          PstI        EcoRV HindIII ClaI
SpeI_____ SmaI _____ EcoRI_____       _____
TAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATACGA
ATCACCTAGGGGGCCCGACGTCCTTAAGCTATAGTTCGAATAGCTATGCT

               KpnI
TAACAGGGTAATGGTACC
ATTGTCCCATTACCATGG
```

C. pMG2 polylinker (pBSpt variant with two Sce1 restriction sites)

```
                                                    XbaI
 SacI          SceI              SacII    NotI      _____
GAGCTCTAGGGATAACAGGGTAATCACCGCGGTGGCGGCCGCTCTAGAAC
CTCGAGATCCCTATTGTCCCATTAGTGGCGCCACCGCCGGCGAGATCTTG

     BamHI        PstI        EcoRV HindIII ClaI
SpeI_____ SmaI _____ EcoRI_____       _____ SalI
TAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCGATGTCG
ATCACCTAGGGGGCCCGACGTCCTTAAGCTATAGTTCGAATAGCTACAGC
```

FIG. 7 CONT'D

XhoI

SceI KpnI

ACCTCGAGTAGGGATAACAGGGTAATGGTACC

TGGAGCTCATCCCTATTGTCCCATTACCATGG

D. Sequence of *loxP* site junction with 5' end of *Af4* intron 4

LoxP

CTTCAGACCTAATAACTTCGTATAGCATACATTATACGAAGTTATATTAAGGGTTATTGA

XbaI

(EcoRI/NotI) *Af4* intron

ATATGATCGGAATTGGCCGCTCTAGAGCATATTAATTCGTAAGTGCTAGAAACCTTGCCT

E. Sequence of PCR fragment mouse *Af4* intron 4

```
                    10         20         30         40         50
          XbaI
     tctAGAGCATATTAATTCGTAAGTGCTAGAAACCTTGCCTTTAACTCAGATAT

                    60         70         80         90        100
     TAGGGAGTGATTAAAGTTTTGCTTCTTAAGCCTGCTGACGGTGCCGCCAG

                   110        120        130        140        150
     GGAAGGGTCAAATCTATTAGAACCCTCCCGCACGGAGCAAGTGGCCTCAA

                   160        170        180        190        200
     TGACTTGTTTGTTTAAGTCAAACAGGAAAGTTCTTCCTTTGAACTTCAAT

                   210        220        230        240        250
     AATTCCCGGAAGTGCAGTGGAGAAGCTGCAAGGCAGAATGTGCCGAGATG

                   260        270        280        290        300
     GTGGCCTTTCTTGGCTCAGCCTTCCCAGAGCTGCCTGATGCTCCAGGGCC

                   310        320        330        340        350
     CAGAGCTGCACCAGCCTGCCGGCACCTGCTGGAGCGCTTGCGCACTGTGG

                   360        370        380        390        400
     TGTTCTTGCTGGGCCGCTCAGTTTCTAATCTAATGTTCTTAAGCTGCTTT

                   410        420        430        440        450
     TTGTCAGTTTTTTTGTTCAGCTTCCAGTGTATGTAGCCTGGTACAAGTGT

                   460        470        480        490        500
     TGCCCACTTTGTTAGCTGAAAACTAATTCGGATGTTCCACTTACGTCTGC

                              acceptor splice site
                   510        520          |
     CTTCCCTTCCTTTTCCAAATCTCTTTCTAG
           ggaaaaggtttagagaaagatcctagg
                                    BamHI
```

FIG. 7 CONT'D

F. Sequence of *loxP* site junction with 3' end of *Af4* intron 4

Af4 acceptor splice site

Af4 Intron

SmaI PstI (EcoRV/EcoRI)

TTTCTAG **GGATCC**CCCGGGCTGCAGGAATTCGATAATTCCGATCATATTCAATAACC

BamHI (insertion site for inversion cassette ORF)

LoxP site

CTTAATATAACTTCGTATAATGTATGCTATACGAAGTTATTAGTTCTGAAGAGGAGTTTA

(EcoRI/EcoRV)

CGTCCAGCCAAGCTAGCTTGGCTGCAGTCGACTAGATGCGGCCCGATGAATTATCAAGCT

G. Diagram of pLOXKI

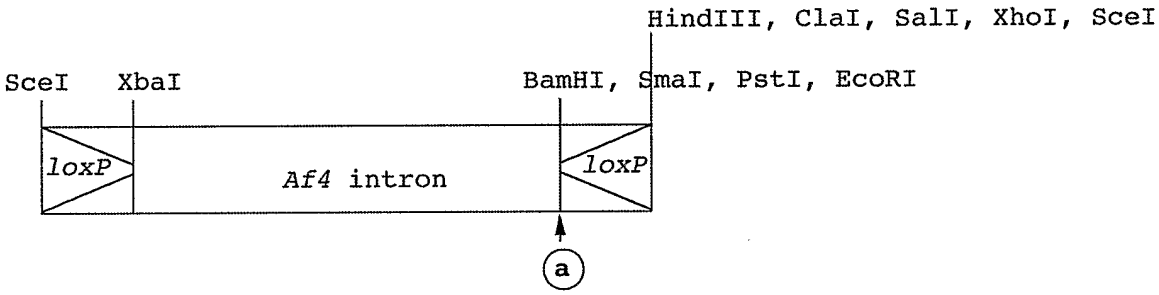


FIG. 8

A.
Cre in pC2A

BglII NotI
ATTCAGATCTGCGGCCGCCATGCCCAAGAAGAAGAGGAAGGTGTCCAATTTACTGACCGT
s a a a M P K K K R K V S N L L T V
nls

B.
Rag1 genomic

GGATCC
GA
ATGGCTGCCTCCTTGCCGTCTACCCTGAGCTTCAGTTCTGCACCCGATGAAATTCAACAC
M A A S L P S T L S F S S A P D E I Q H

C.
Rag1-Cre KI

BamHI/BglII NotI
ATGGCTGGATCTGCGGCCGCCATGCCCAAGAAGAAGAGGAAGGTGTCCAATTTACTGACC
M A G s a a a M P K K K R K V S N L L T
Rag1 linker Cre (from pC2A-neo)

D.

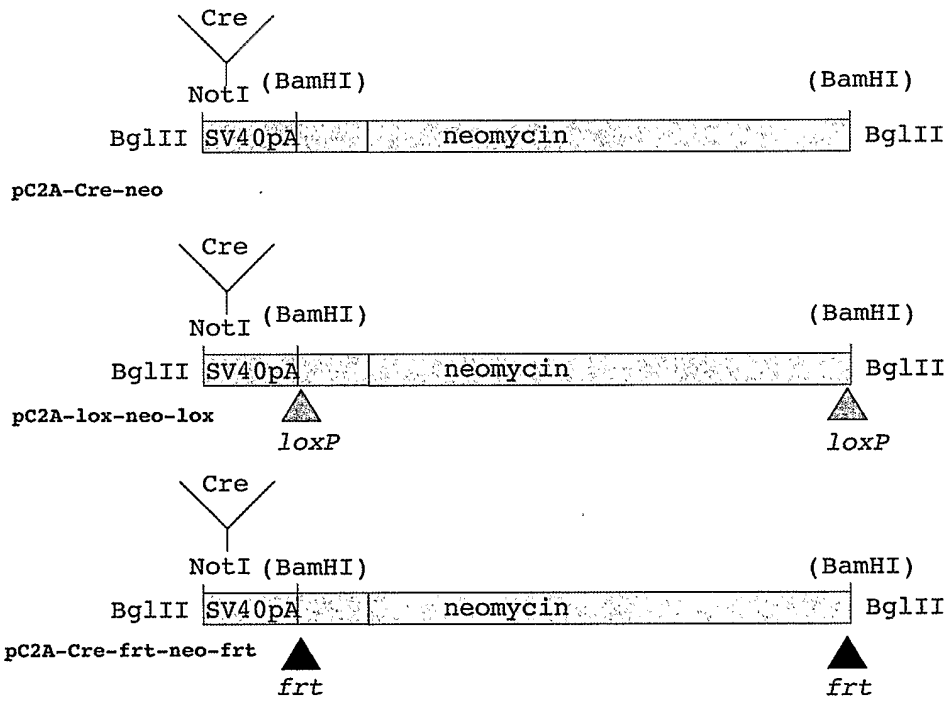


FIG. 9

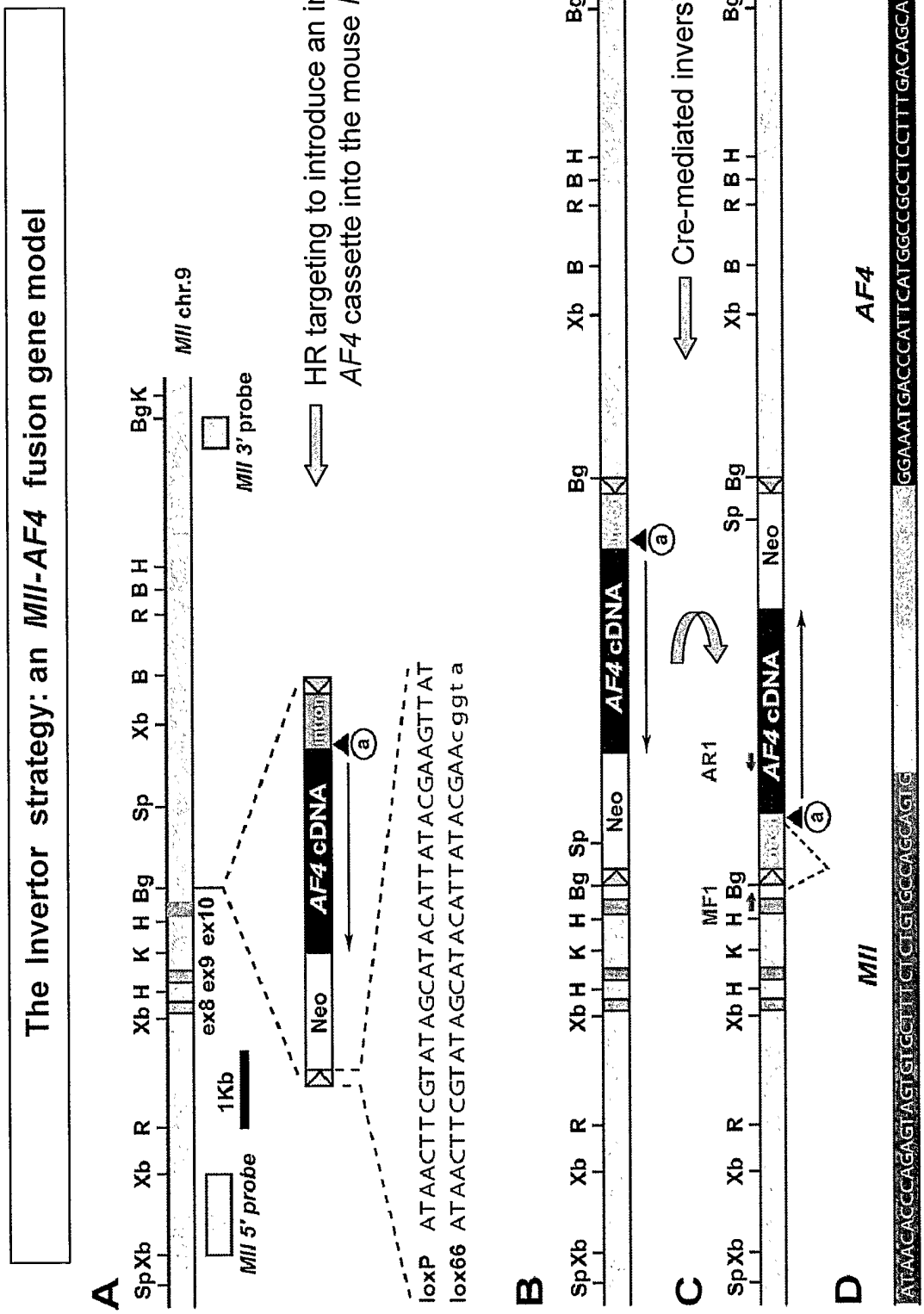

The Invertor strategy: an MII-AF4 fusion gene model
A
Sp Xb Xb R Xb H K H Bg Sp Xb B R B H Bg K
MII chr.9
MII 5' probe
1Kb
ex8 ex9 ex10
MII 3' probe
HR targeting to introduce an inverted AF4 cassette into the mouse MII gene
Neo
AF4 cDNA
loxP ATAACTTCGTATAGCATACATTATACGAAGTTAT
lox66 ATAACTTCGTATAGCATACATTATACGAACggta
B
Sp Xb Xb R Xb H K H Bg Sp Bg Xb B R B H Bg K
Neo
AF4 cDNA
Cre-mediated inversion of AF4
C
Sp Xb Xb R Xb H K H Bg MF1 AR1 Sp Bg Xb B R B H Bg K
Neo
AF4 cDNA
D
MII
AF4
ATAACACCCAGAGTAGTGTGCTTCTCTGTGCCAGCAGG GGAAATGACCCATTCATGGCCGCCTCCTTTGACAGCAATA
I T P R V V C F L C A S S G H V E G I R P E M T H S W P P P L T A I

FIG. 10

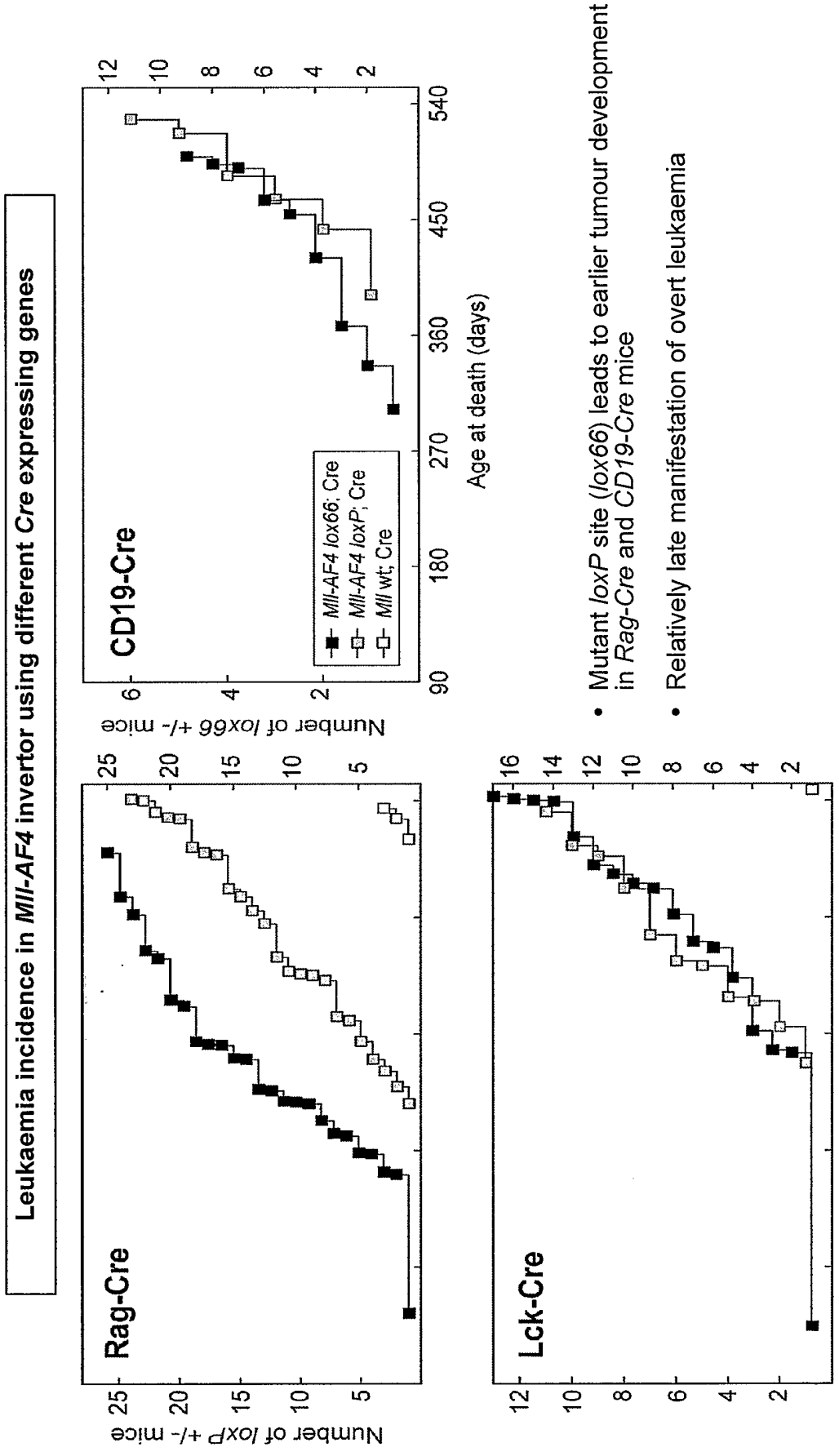

Leukaemia incidence in Mll-AF4 invertor using different Cre expressing genes
Rag-Cre
CD19-Cre
Lck-Cre
Number of loxP +/- mice
Number of lox66 +/- mice
Age at death (days)
Mll-AF4 lox66; Cre
Mll-AF4 loxP; Cre
Mll wt; Cre
• Mutant loxP site (lox66) leads to earlier tumour development in Rag-Cre and CD19-Cre mice
• Relatively late manifestation of overt leukaemia

FIG. 11

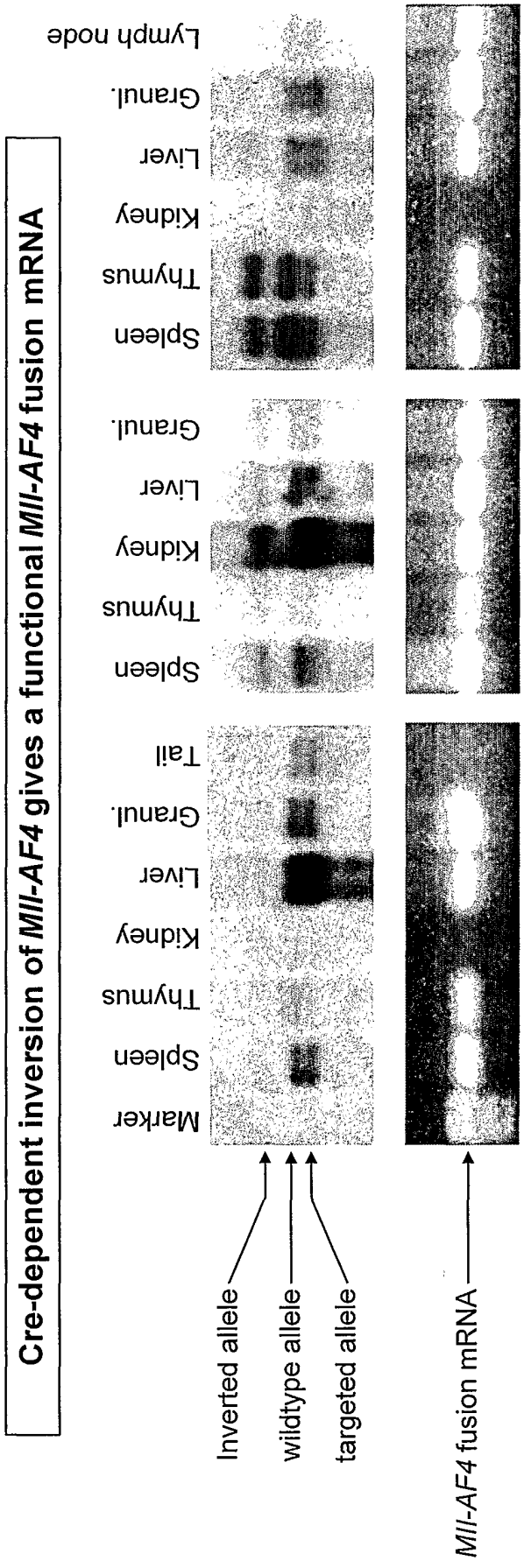

Cre-dependent inversion of *Mll-AF4* gives a functional *Mll-AF4* fusion mRNA
Marker
Spleen
Thymus
Kidney
Liver
Granul.
Tail
Spleen
Thymus
Kidney
Liver
Granul.
Spleen
Thymus
Kidney
Liver
Granul.
Lymph node
Inverted allele
wildtype allele
targeted allele
Mll-AF4 fusion mRNA

FIG. 12

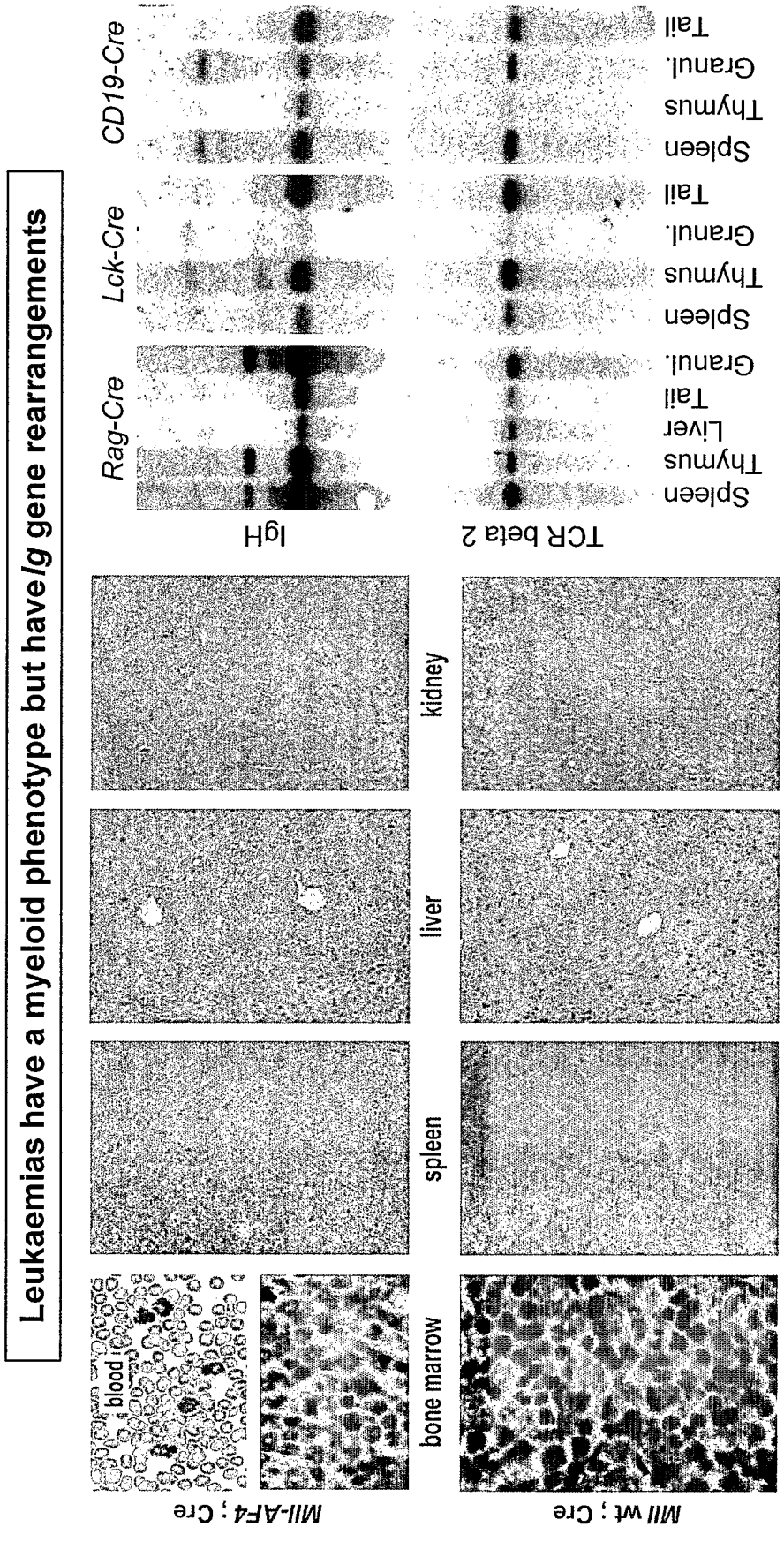

Leukaemias have a myeloid phenotype but have Ig gene rearrangements
Mll wt ; Cre
Mll-AF4 ; Cre
bone marrow
spleen
liver
kidney
blood
IgH
TCR beta 2
Rag-Cre
Lck-Cre
CD19-Cre
Spleen
Thymus
Liver
Tail
Granul.

FIG. 13

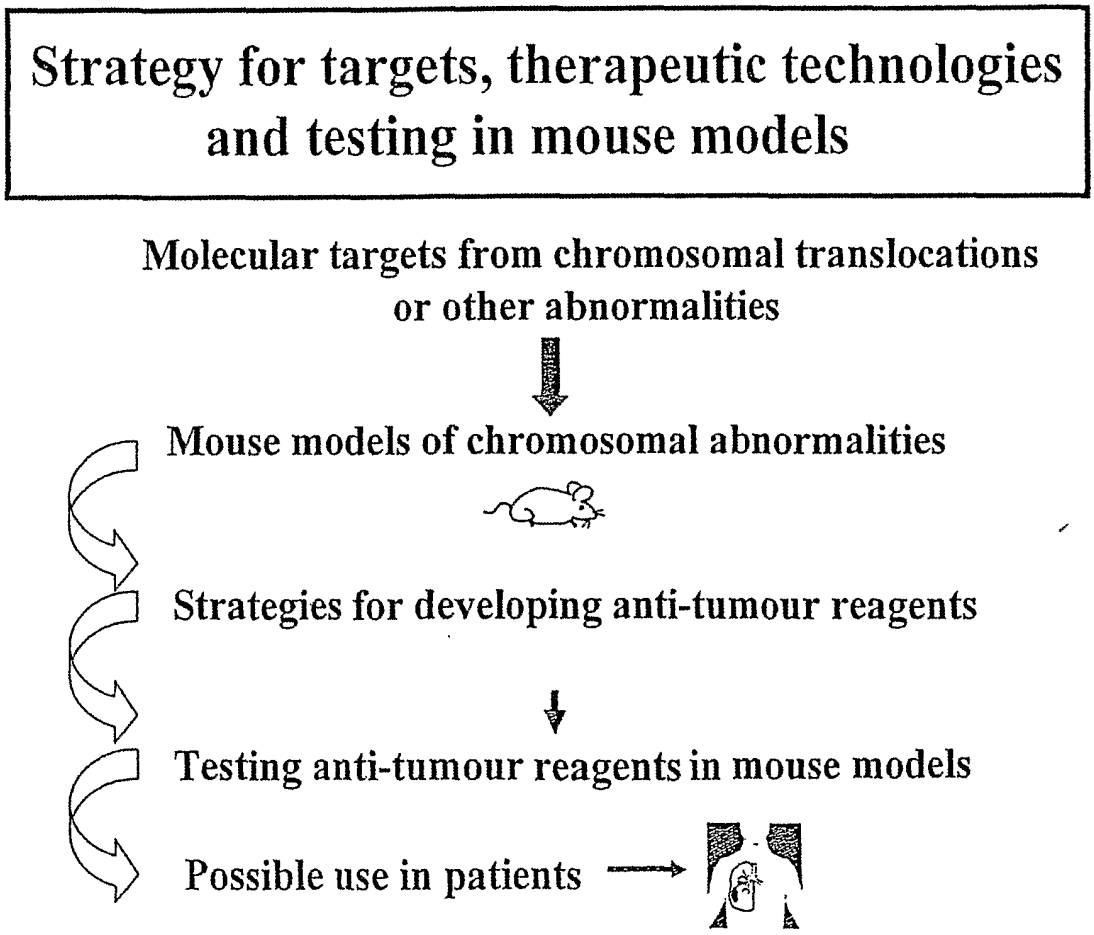

Strategy for targets, therapeutic technologies and testing in mouse models
Molecular targets from chromosomal translocations or other abnormalities
Mouse models of chromosomal abnormalities
Strategies for developing anti-tumour reagents
Testing anti-tumour reagents in mouse models
Possible use in patients

FIG. 14

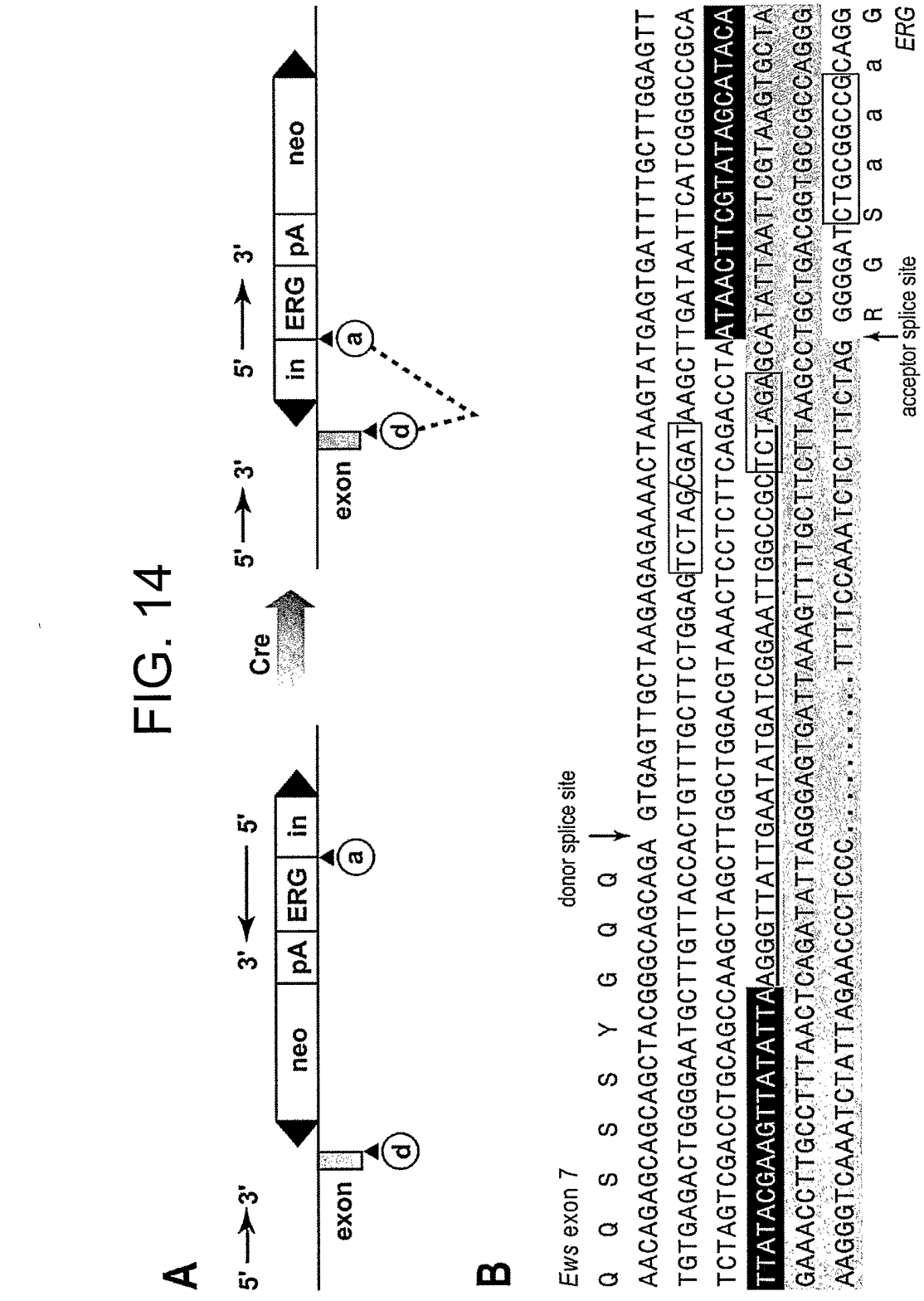
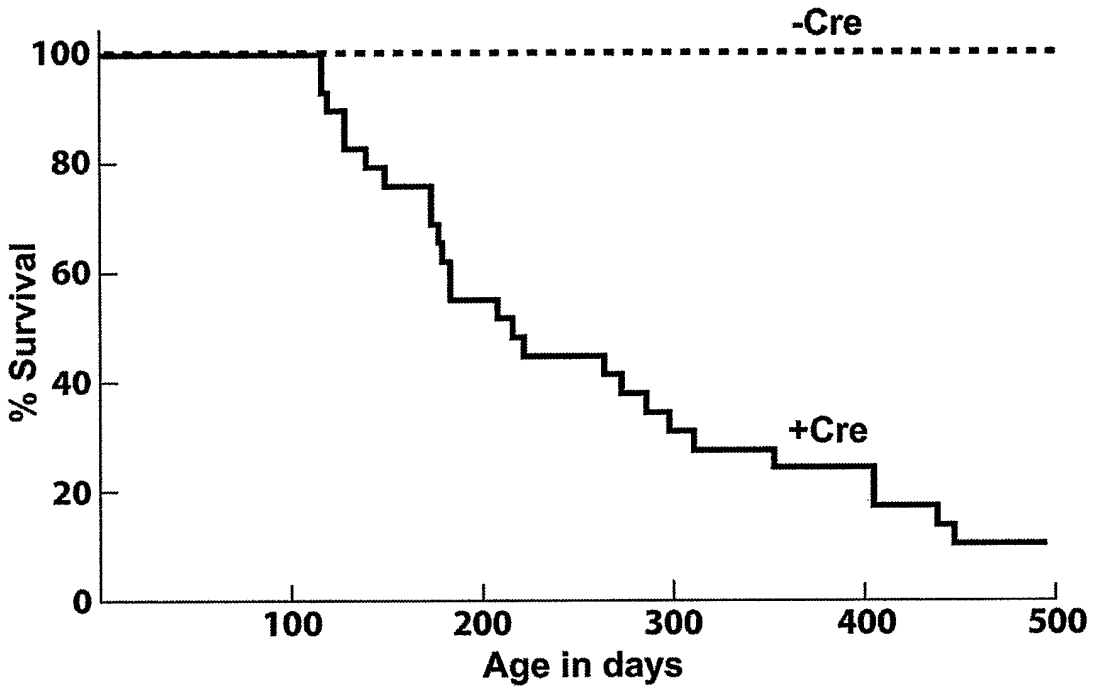

A
5' → 3'
neo
pA
ERG
in
3' ← 5'
a
d
exon
Cre
5' → 3'
in
ERG
pA
neo
5' → 3'
a
d
exon
B
Ews exon 7
donor splice site
Q Q S S S Y G Q Q Q
AACAGAGCAGCAGCTACGGGCAGCAGA GTGAGTTGCTAAGAGAGAAAACTAAGTATGAGTGATTTTTGCTTGGAGTT
TGTGAGACTGGGGAATGCTTGTTACCACTGTTTGCTTCTGGAGTCTAGCGATAAGCTTGATAATTCATCGGGCCGCA
TCTAGTCGACCTGCAGCCAAGCTAGCTTGGCTGGACGTAAACTCCTCTTCAGACCTAATAACTTCGTATAGCATACA
TTATACGAAGTTATATTAAGGGTTATTGAATATGATCGGAATTGGCCGCTCTAGAGCATATTAATTCGTAAGTGCTA
GAAACCTTGCCTTTAACTCAGATATTAGGGAGTGATTAAAGTTTTGCTTCTTAAGCCTGCTGACGGTGCCGCCAGGG
AAGGGTCAAATCTATTAGAACCCTCCC. . . . . TTTTCCAAATCTCTTCTAG GGGGATCTGCGGCCGCAGG
R G S a a a G
acceptor splice site
ERG

FIG. 15


-Cre
+Cre
% Survival
100
80
60
40
20
0
100
200
300
400
500
Age in days

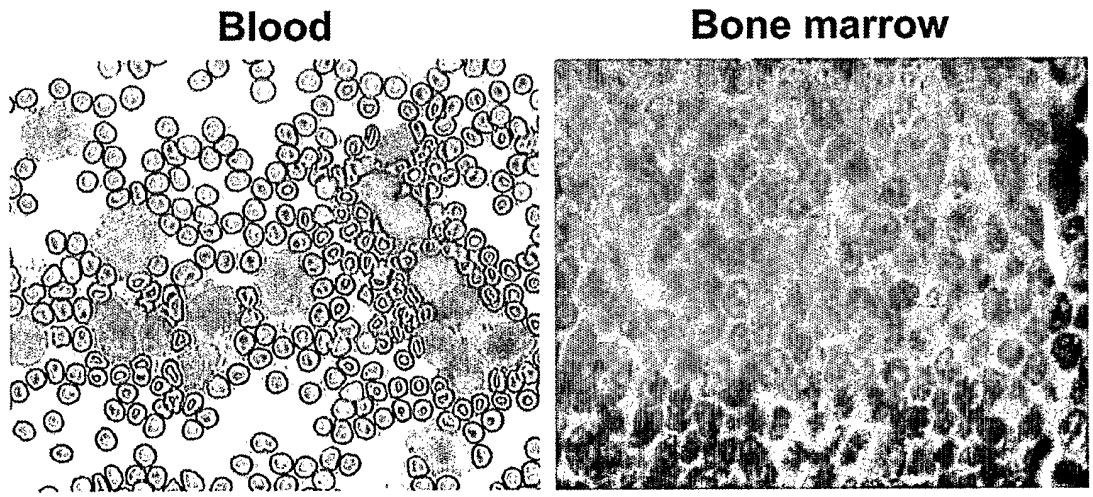

Blood
Bone marrow

FIG. 16

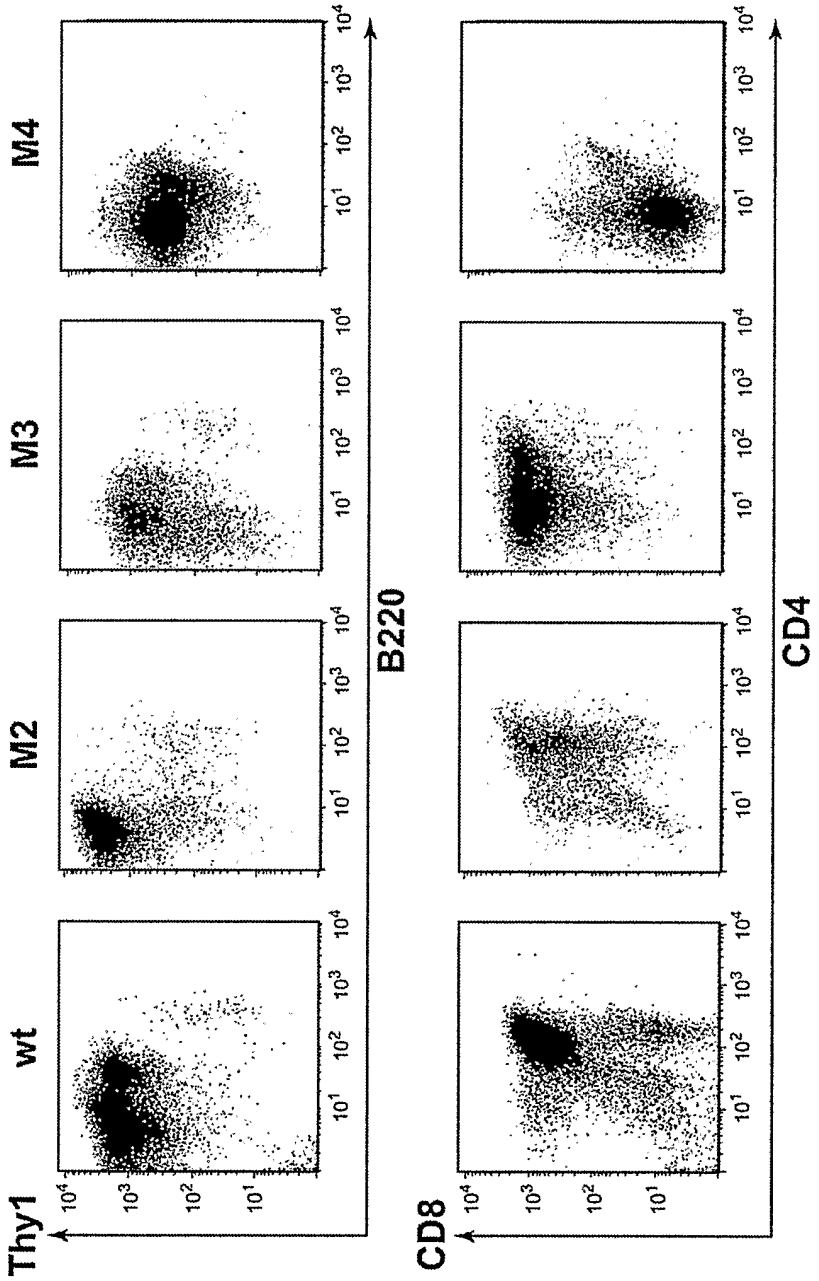


FIG. 17

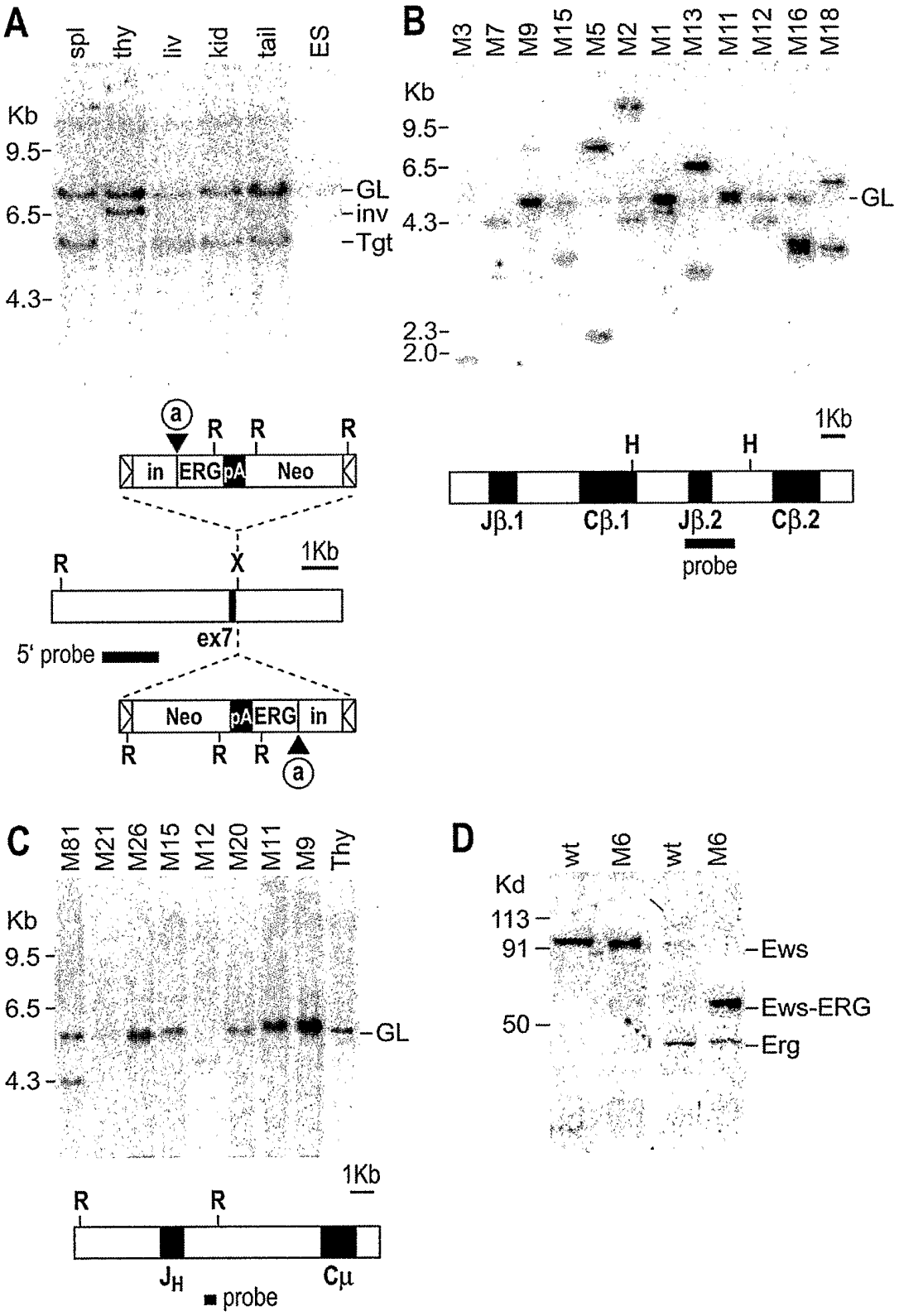

A
spl
thy
liv
kid
tail
ES
Kb
9.5-
6.5-
4.3-
GL
inv
Tgt
B
M3
M7
M9
M15
M5
M2
M1
M13
M11
M12
M16
M18
Kb
9.5-
6.5-
4.3-
2.3-
2.0-
GL
a
R
R
R
in
ERG
pA
Neo
1Kb
H
H
Jβ.1
Cβ.1
Jβ.2
Cβ.2
probe
R
X
1Kb
ex7
5' probe
Neo
pA
ERG
in
R
R
R
a
C
M81
M21
M26
M15
M12
M20
M11
M9
Thy
Kb
9.5
6.5
4.3
GL
1Kb
R
R
JH
Cμ
probe
D
wt
M6
wt
M6
Kd
113
91
50
Ews
Ews-ERG
Erg

FIG. 18

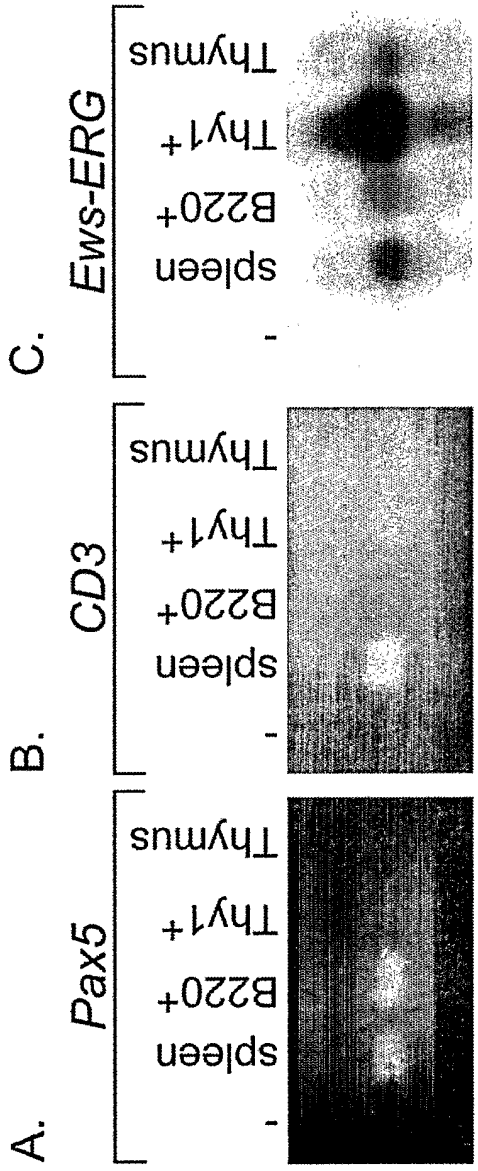

A.
Pax5
-
spleen
B220+
Thy1+
Thymus
B.
CD3
-
spleen
B220+
Thy1+
Thymus
C.
Ews-ERG
-
spleen
B220+
Thy1+
Thymus

FIG. 19

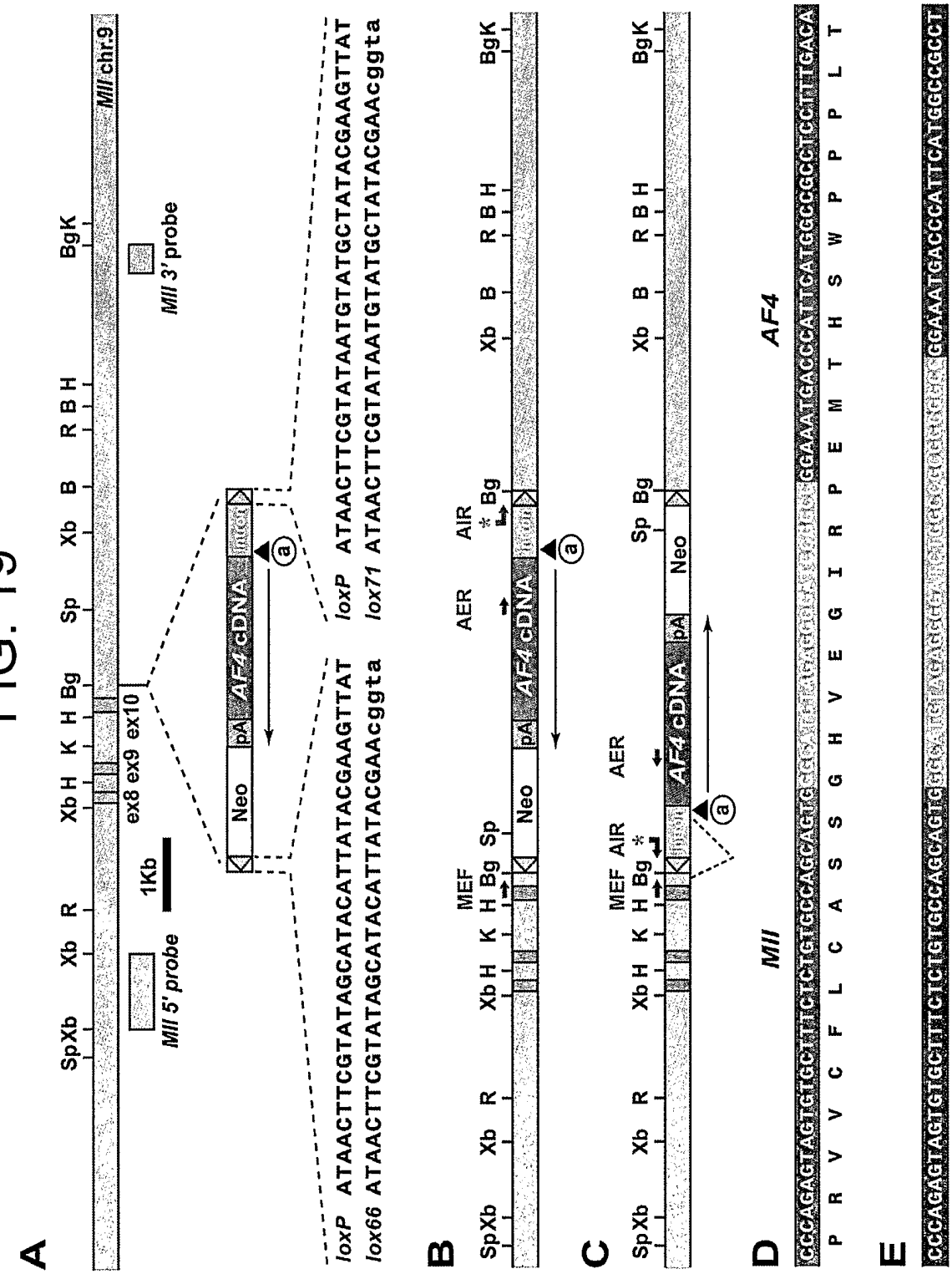

A
Sp Xb Xb R Xb H K H Bg Sp Xb B R B H BgK
MII chr.9
ex8 ex9 ex10
1Kb
MII 5' probe
MII 3' probe
Neo pA AF4 cDNA mef
a
loxP ATAACTTCGTATAGCATACATTATACGAAGTTAT
lox66 ATAACTTCGTATAGCATACATTATACGAAcggta
loxP ATAACTTCGTATAATGTATGCTATACGAAGTTAT
lox71 ATAACTTCGTATAATGTATGCTATACGAAcggta
B
MEF
AER AIR
Sp Xb R Xb H K H Bg Sp Xb B R B H BgK
Neo pA AF4 cDNA
C
MEF AIR AER
Sp Xb R Xb H K H Bg Sp Xb B R B H BgK
Neo pA AF4 cDNA
D
MII
AF4
GCCAGAGTAGTGTGCTTCTCTGTGCCAGCACGTGCGGAAATGACCATTCATGGCGCCTCTTTTGACA
P R V V C F L C A S S G H V E G I R P E M T H S W P P P L T
E
CCCAGAGTAGTGTGCTTCTCTGTGCCAGCACGTGCGGAAATGACCCATTCATGGCGCCTCT
P R V V C F L C A S S G H V E G I W P R R P E M T H S W P P

FIG. 20

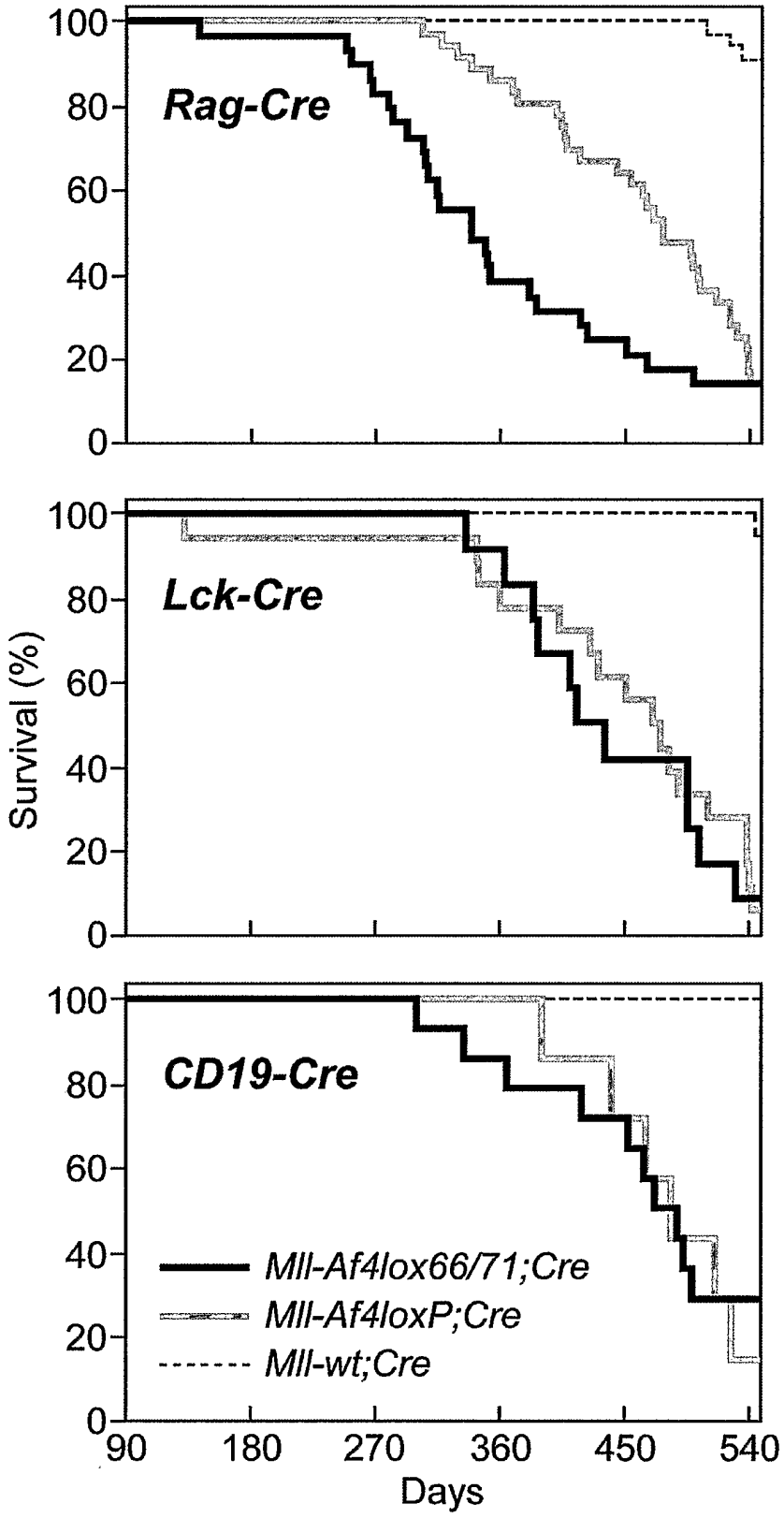

Rag-Cre
Lck-Cre
CD19-Cre
Survival (%)
100
80
60
40
20
0
90
180
270
360
450
540
Days
Mll-Af4lox66/71;Cre
Mll-Af4loxP;Cre
Mll-wt;Cre

FIG. 21

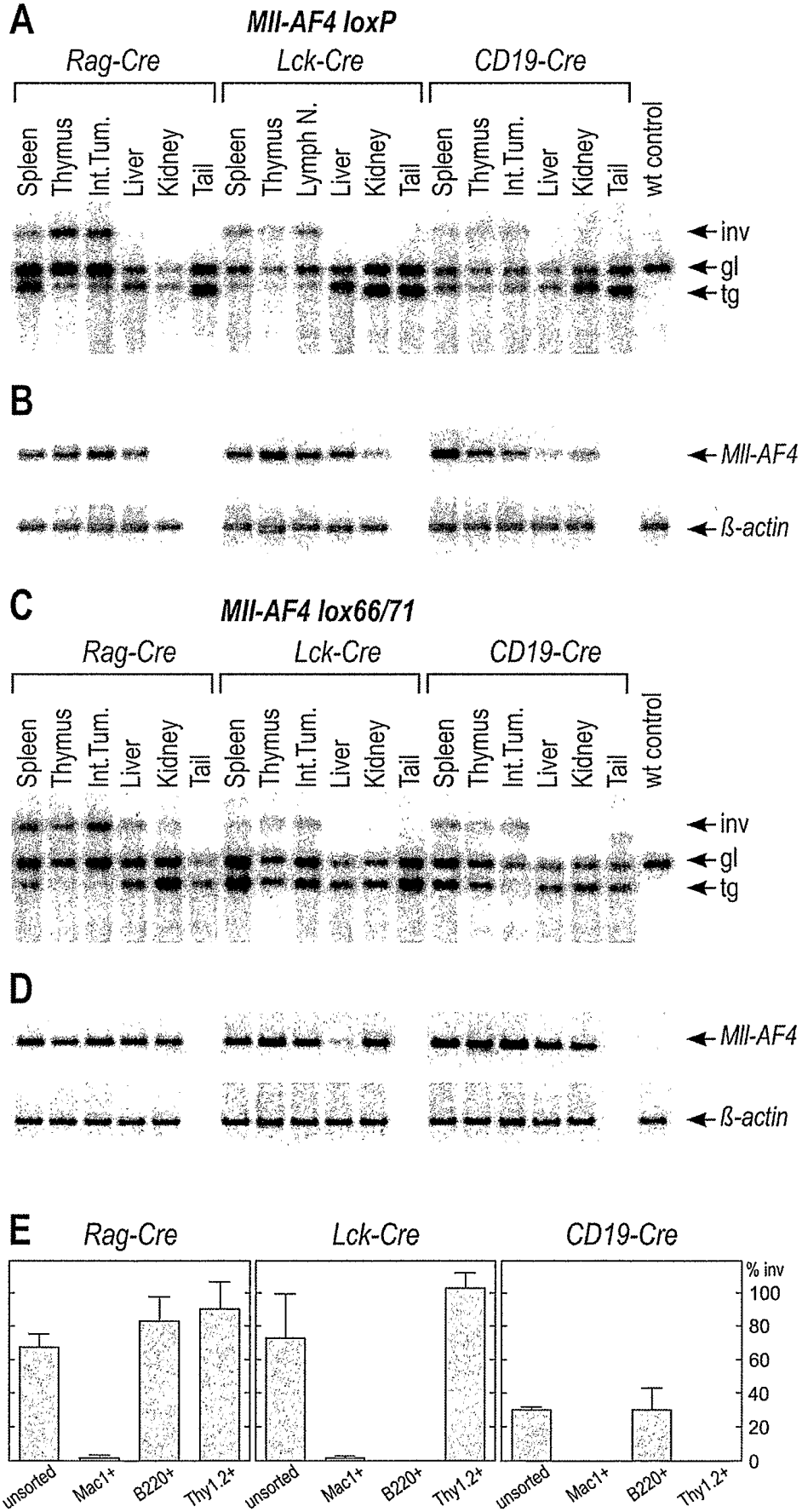

A
Mll-AF4 loxP
Rag-Cre
Lck-Cre
CD19-Cre
Spleen
Thymus
Int.Tum.
Liver
Kidney
Tail
Lymph N.
wt control
inv
gl
tg
B
Mll-AF4
ß-actin
C
Mll-AF4 lox66/71
D
E
% inv
100
80
60
40
20
0
unsorted
Mac1+
B220+
Thy1.2+

FIG. 22

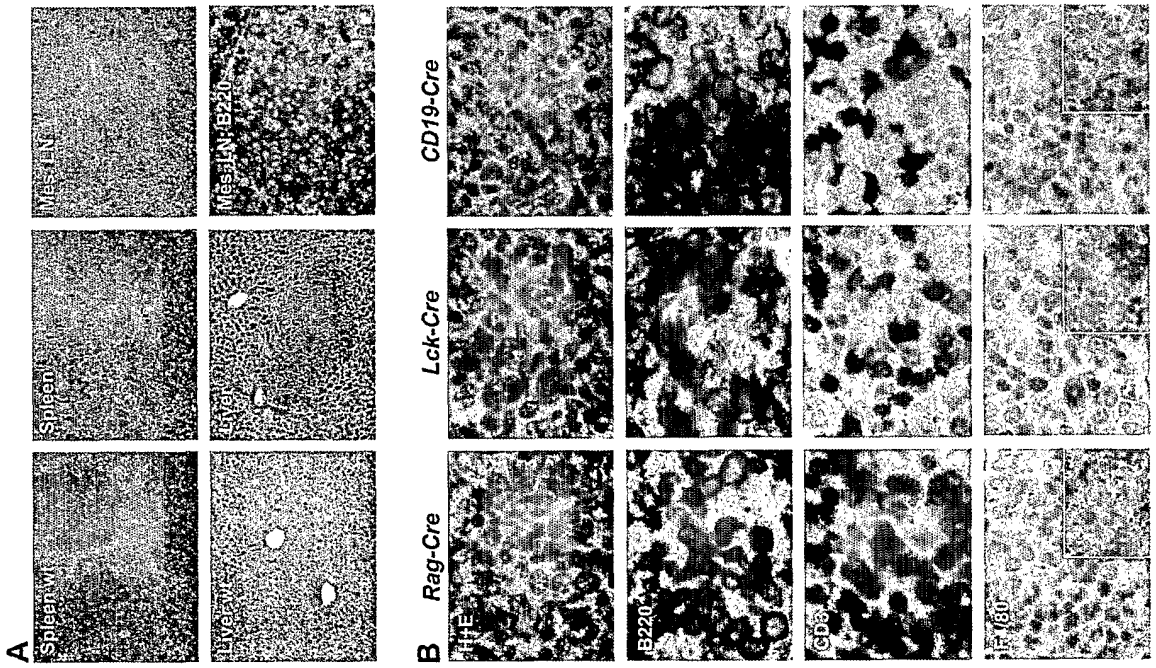
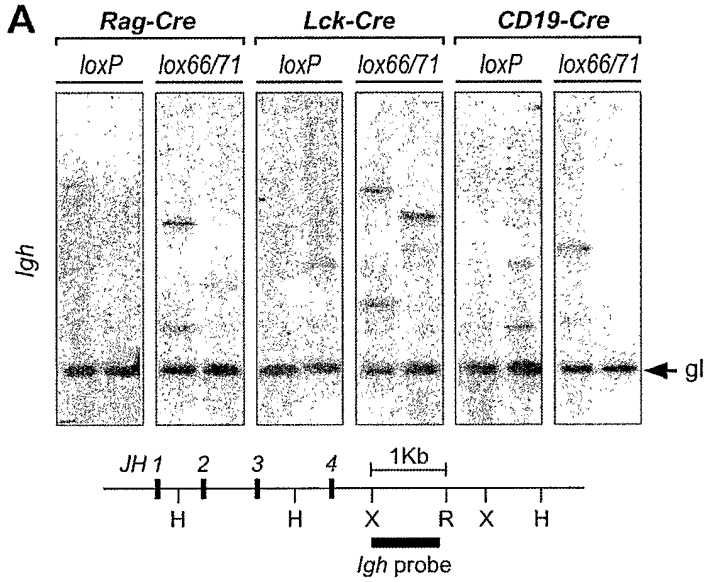
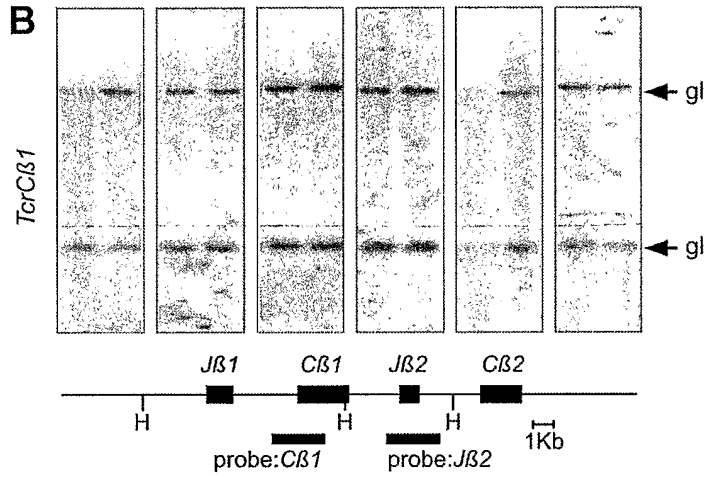
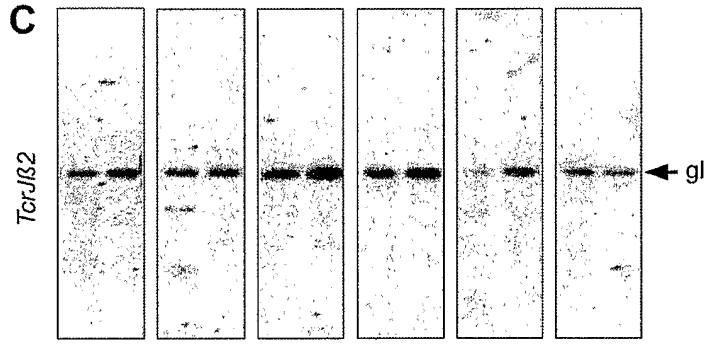

A
Spleen WT
Liver WT
Mes LN
Spleen
Liver
Mes LN: B220
B
Rag-Cre
Lck-Cre
CD19-Cre
H&E
B220
CD3
F4/80
RGB
CMYK

FIG. 23







D

	Rag-Cre		Lck-Cre		CD19-Cre		Total
	loxP	*lox66/71*	*loxP*	*lox66/71*	*loxP*	*lox66/71*	
Igh	7/12 (58%)	8/13 (62%)	3/05 60%)	6/10 (60%)	4/05 (80%)	5/08 (63%)	33/53 (62%)
TcrCß1	0/12 (0%)	0/13 (0%)	0/05 (0%)	1/10 (0%)	0/05 (0%)	0/08 (0%)	1/53 (0.2%)
TcrJß2	0/12 (0%)	1/13 (8%)	0/05 (0%)	0/10 (0%)	0/05 (0%)	0/08 (0%)	1/53 (0.2%)

FIG. 24

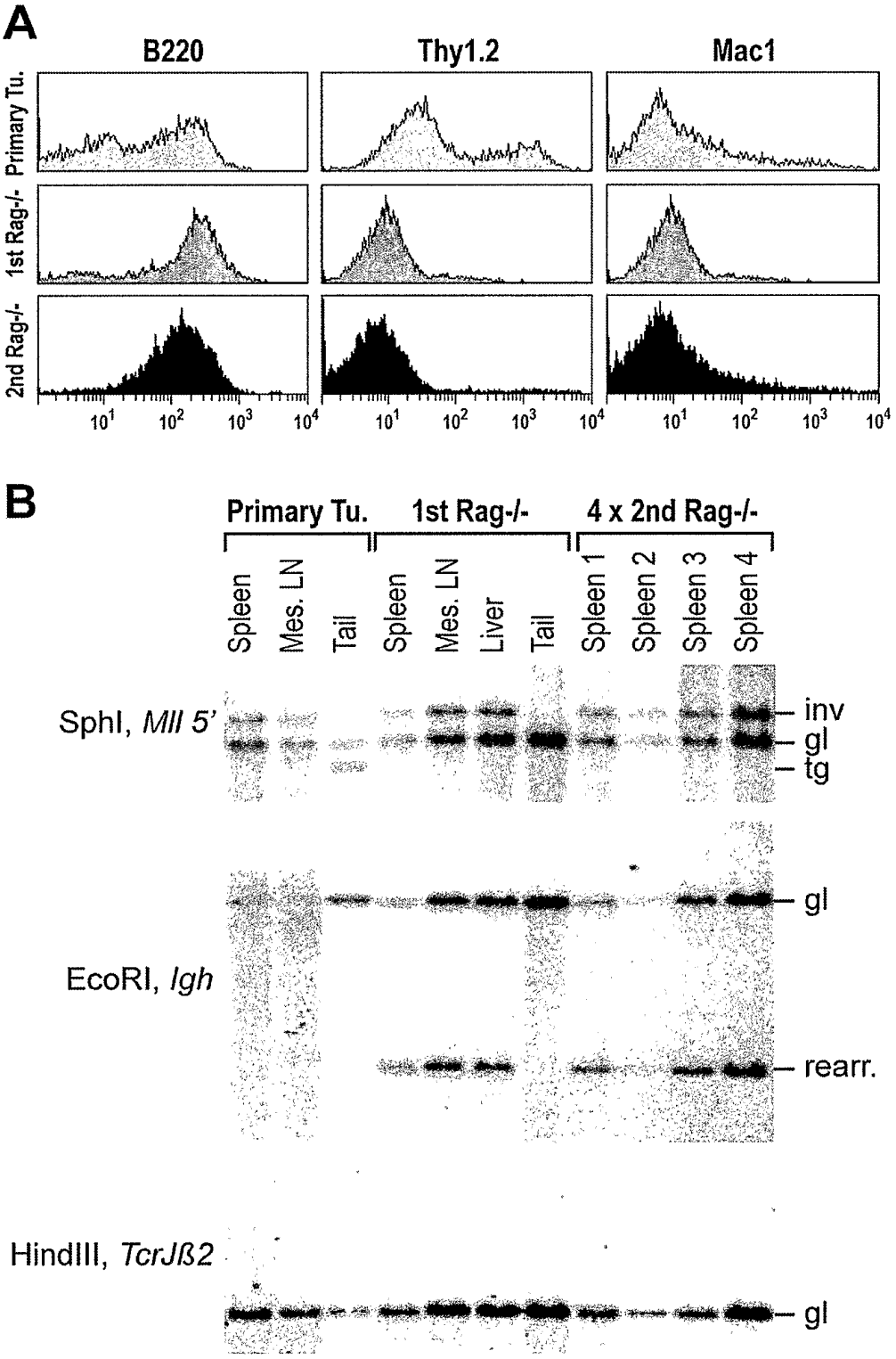

A
B220
Thy1.2
Mac1
Primary Tu.
1st Rag-/-
2nd Rag-/-
10^1 10^2 10^3 10^4
B
Primary Tu.
1st Rag-/-
4 x 2nd Rag-/-
Spleen
Mes. LN
Tail
Spleen
Mes. LN
Liver
Tail
Spleen 1
Spleen 2
Spleen 3
Spleen 4
SphI, Mll 5'
inv
gl
tg
EcoRI, Igh
gl
rearr.
HindIII, TcrJß2
gl

FIG.25

Table 1A

Tgt vector	Colonies analysed	No. tgted clones
Ews-cassette control	64	4
Ews-AF9	95	2
Ews-Fli1	170	0
Ews-Chop	144	0
Ews-ATF	200	0

Table 1B

Tgt ES cell	Blastocysts transferred	No. of chimaeras
Ews-cassette control	143	18
Ews-AF9	78	20

FIG. 26

Table 2

Mouse number	Age	PM details	Inv	TCR Jβ2	IgH	Thymus profile CD4/CD8
1	116	Thymoma; splenomegaly	+	R/G	G/G	
2	119	Thymoma	+	R/R	G/G	$CD4^{HI}$; $CD8^{HI}$
3	128	Thymoma; splenomegaly	+	R/G	G/G	CD4-; $CD8^{HI}$ (85%) + $CD4^{LO}$; $CD8^{HI}$ (15%)
4	128	Thymoma; splenomegaly	+	G/G	R/G	CD4-; CD8-
5	139	Thymoma	+/-	R/R	R/G	CD4-; $CD8^{HI}$
6	149	Thymoma	+	R/R	G/G	CD4-; $CD8^{HI}$ (75%) + $CD4^{LO}$; $CD8^{HI}$ (20%)
7	173	Thymoma	+	R/R	R/R	
8	173	Thymoma	+	R/R	R/G	
9	179	Thymoma; splenomegaly	+	G/G	G/G	
10	183	Thymoma	+	R/R	G/G	
11	183	Thymoma; splenomegaly	+	R/G	G/G	CD4-; $CD8^{LO}$
12	208	Thymoma; splenomegaly	+	R/G	R/-	
13	216	Thymoma	+	R/R	G/G	CD4-; $CD8^{HI}$
14	222	Thymoma; splenomegaly	+	R/R	R/G	CD4-; $CD8^{HI}$
15	264	Thymoma; splenomegaly	+/-	R/G	G/G	CD4-; $CD8^{HI}$
16	273	Thymoma; splenomegaly	+	R/R	R/R	
17	286	Thymoma; splenomegaly	+	R/G	G/G	CD4-; $CD8^{HI}$
18	298	Thymoma	+/-	R/R	R/G	
19	311	Thymoma	+	R/G	R/R	
20	353	Thymoma	+	R/G	G/G	CD4-; $CD8^{HI}$
21	405	Thymoma	+	R/R	G/G	
22	405	Thymoma	+/-	R/G	R/G	CD4-; CD8-
23	438	Thymoma	+	R/G	R/G	
24	447	Thymoma	+	R/G	G/G	CD4-;$CD8^{LO}$
25	494	Thymoma	+/-	R/G	G/G	

FIG. 27

Table 3

Mouse Number	Vβ primer pool	Vβ	Dβ	Jβ2	P/NP	Sequence of V-D-Jβ2 rearrangement
M2	2 + 11 +15	VB16	DB1	JB2.4	P	TGTGCAAGCAGCTGGGACACTAGTCAAAACACCTTGTACTTTGGTGCGGGCACCCGA
M2	2 + 11 +15	?	DB2	JB2.2	NP	TGTGCCAGCATTTCCAGGACTGTG---
M5	5 + 13	VB4	DB1	JB2.5	P	TGTGCCAGCAGTGGCGACAGGGGG CAAGACACCCAGTACTTTGGGCCAGGCACTCGG
M13	4 + 8 + 14	VB8	DB1or 2?	JB2.3	P	TGTGCCAGCGGTGAGAAACCGGGGACAGCTAGTGCAGAAACGCTGTATTTTGGCTCAGGAACCAGA
M18	6 + 12	VB19	DB2	JB2.5	P	TGTGCCAGCAGTCCCCACTGGGGGGGGAACCAAGACACCCAGTACTTTGGGCCAGGCACTCGG

FIG. 28

Table 4

Mouse number	Age	PM details	Inv	TCR Jβ2	IgH	Thymus profile CD4/CD8
1	116	Thymoma; splenomegaly	+	R/G	G/G	
2	119	Thymoma	+	R/R	G/G	$CD4^{HI}$; $CD8^{HI}$
3	128	Thymoma; splenomegaly	+	R/G	G/G	CD4-; $CD8^{HI}$ (85%) + $CD4^{LO}$; $CD8^{HI}$ (15%)
4	128	Thymoma; splenomegaly	+	G/G	R/G	CD4-; CD8-
5	139	Thymoma	+*	R/R	R/G	CD4-; $CD8^{HI}$
6	149	Thymoma	+	R/R	G/G	CD4-; $CD8^{HI}$ (75%) + $CD4^{LO}$; $CD8^{HI}$ (20%)
7	173	Thymoma	+	R/R	R/R	
8	173	Thymoma	+	R/R	R/G	
9	179	Thymoma; splenomegaly	+	G/G	G/G	
10	183	Thymoma	+	R/R	G/G	
11	183	Thymoma; splenomegaly	+	R/G	G/G	CD4-; $CD8^{LO}$
12	208	Thymoma; splenomegaly	+	R/G	R/-	
13	216	Thymoma	+	R/R	G/G	CD4-; $CD8^{HI}$
14	222	Thymoma; splenomegaly	+	R/R	R/G	CD4-; $CD8^{HI}$
15	264	Thymoma; splenomegaly	+*	R/G	G/G	CD4-; $CD8^{HI}$
16	273	Thymoma; splenomegaly	+	R/R	R/R	
17	286	Thymoma; splenomegaly	+	R/G	G/G	CD4-; $CD8^{HI}$
18	298	Thymoma	+*	R/R	R/G	
19	311	Thymoma	+	R/G	R/R	
20	353	Thymoma	+	R/G	G/G	CD4-; $CD8^{HI}$
21	405	Thymoma	+	R/R	G/G	
22	405	Thymoma	+*	R/G	R/G	CD4-; CD8-
23	438	Thymoma	+	R/G	R/G	
24	447	Thymoma	+	R/G	G/G	CD4-;$CD8^{LO}$
25	494	Thymoma	+*	R/G	G/G	