

(51) International Patent Classification:
C12Q 1/68 (2006.01)(21) International Application Number:
PCT/EP2016/065932(22) International Filing Date:
6 July 2016 (06.07.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
15176251.5 10 July 2015 (10.07.2015) EP(71) Applicant: DEUTSCHES KREBSFORSCHUNGSZEN-
TRUM [DE/DE]; Im Neuenheimer Feld 280, 69120
Heidelberg (DE).(72) Inventors: RODEWALD, Hans-Reimer; Brückenkopf-
strasse 8, 69120 Heidelberg (DE). FEYERABEND, Thor-
sten B.; Schlüsselweg 27, 69221 Dossenheim (DE). PEI,
Weike; Im Neuenheimer Feld 679, 00-07, 69120 Heidel-
berg (DE).(74) Agent: HERZOG FIESSER & PARTNER PAT-
ENTANWÄLTE PARTG MBB; Dudenstraße 46, 68167
Mannheim (DE).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: GENETIC RANDOM DNA BARCODE GENERATOR FOR IN VIVO CELL TRACING

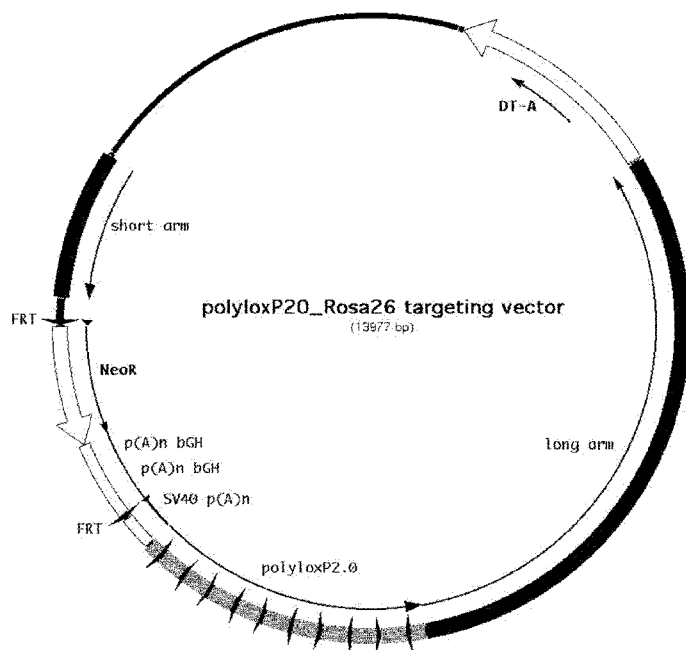


Fig. 12

(57) Abstract: The present invention relates to a barcoding polynucleotide comprising at least five, preferably at least ten core structures, wherein each of said core structures comprises two uniquely identifiable sequences separated by a recombinase recognition sequence, wherein said recombinase is a recombinase capable of (i) mediating excision or (ii) mediating excision or inversion of a polynucleotide flanked by two recognition sequences of said recombinase; and to a vector, a host cell, a host organism and an experimental animal comprising said barcoding polynucleotide. Furthermore, the present invention relates to kits, methods, and uses related said barcoding polynucleotide. In particular, the means and methods of the invention are useful in barcoding of cell lines or of cells within an organism.

Genetic random DNA barcode generator for in vivo cell tracing

The present invention relates to a barcoding polynucleotide comprising at least five, preferably at least ten core structures, wherein each of said core structures comprises two uniquely identifiable sequences separated by a recombinase recognition sequence, wherein said recombinase is a recombinase capable of (i) mediating excision or (ii) mediating excision or inversion of a polynucleotide flanked by two recognition sequences of said recombinase; and to a vector, a host cell, a host organism and an experimental animal comprising said barcoding polynucleotide. Furthermore, the present invention relates to kits, methods, and uses related to said barcoding polynucleotide. In particular, the means and methods of the invention are useful in barcoding of cell lines or of cells within an organism.

Methods for determining cell fate in multicellular tissues, organs, or individuals have been regarded as highly desirable, in order to understand developmental processes, but also to understand e.g. carcinogenesis. For the relatively simple organism of *Caenorhabditis elegans*, which comprises 959 somatic cells, this was accomplished by direct observation. However, in higher organisms, comprising higher cell numbers and being much less deterministic in development, this method was not feasible. As alternatives, various clonal assays were developed to determine development of cell lineages, e.g. the one disclosed in US 7,917,306 B2.

Recombinases have been used extensively to manipulate and modify DNA. In particular, Cre, a site-specific recombinase, was found to be particularly useful in genome editing, e.g. inducible and/or tissue-specific excision of DNA fragments from chromosomal DNA, by inverting and/or depleting DNA fragments (Branda et al. (2004), *Dev Cell* 6(1): 7; PMID: 14723844).

More recently, methods for differentially labeling cells within cell populations or within cell lineages were devised making use of random excision and/or inversion of DNA fragments by recombinases. E.g., genes encoding three fluorescent proteins were flanked by mutually incompatible loxP sites, such that the coding genes are randomly inverted or excised, giving

rise to cells expressing various ratios of the three fluorescence proteins and, thus, fluorescing in different shades. A similar strategy was used by Clevers and colleague to obtain cells with single color labels (Snippert et al., 2010; Schepers et al., 2012). However, the diversity that could be created by such attempts was limited by the limited availability of fluorescent colors. For this reason, use of polynucleotides, i.e. individual labeling of cells with unique DNA tags, was considered advantageous. Zador and colleague developed a novel method of cellular labeling in vivo with a unique DNA tag generated by Rci, a site-specific DNA invertase (Peikon et al. (2014), NAR, online publication, doi: 10.1093/nar/gku604). However, this method could not be broadly applied currently because only few Rci-sfx systems are available; moreover, since Rci is an invertase, only shuffling of fragments can be obtained, which limits the obtainable diversity.

There is, thus, still a need in the art for improved means and methods for individually labeling cells. This problem is solved by the means and methods provided herein.

Accordingly, the present invention relates to a barcoding polynucleotide comprising at least five, preferably at least ten core structures, wherein each of said core structures comprises two uniquely identifiable sequences separated by a recombinase recognition sequence, wherein said recombinase is a recombinase capable of (i) mediating excision or (ii) mediating excision or inversion of a polynucleotide flanked by two recognition sequences of said recombinase.

As used in the following, the terms “have”, “comprise” or “include” or any arbitrary grammatical variations thereof are used in a non-exclusive way. Thus, these terms may both refer to a situation in which, besides the feature introduced by these terms, no further features are present in the entity described in this context and to a situation in which one or more further features are present. As an example, the expressions “A has B”, “A comprises B” and “A includes B” may both refer to a situation in which, besides B, no other element is present in A (i.e. a situation in which A solely and exclusively consists of B) and to a situation in which, besides B, one or more further elements are present in entity A, such as element C, elements C and D or even further elements.

Further, as used in the following, the terms "preferably", "more preferably", "most preferably", "particularly", "more particularly", "specifically", "more specifically" or similar terms are used in conjunction with optional features, without restricting alternative

possibilities. Thus, features introduced by these terms are optional features and are not intended to restrict the scope of the claims in any way. The invention may, as the skilled person will recognize, be performed by using alternative features. Similarly, features introduced by "in an embodiment of the invention" or similar expressions are intended to be optional features, without any restriction regarding alternative embodiments of the invention, without any restrictions regarding the scope of the invention and without any restriction regarding the possibility of combining the features introduced in such way with other optional or non-optional features of the invention. Moreover, the term "about", if not noted otherwise, relates to the indicated value $\pm 20\%$.

The term "barcoding polynucleotide" as used in accordance with the present invention relates to a DNA polynucleotide comprising the sequence features as described herein below. Preferably, the barcoding polynucleotide has the activity that an appropriate recombinase excises or inverts a sequence flanked by two recombinase recognition sequences. Suitable assays for measuring said activity are described in the accompanying examples or in Abremski & Hoess (1984), JBC 259(3):1509, PMID: 6319400).

A barcoding polynucleotide having the aforementioned activity has been obtained in accordance with the present invention as described in the examples. Preferably, the barcoding polynucleotide comprises at least one, preferably at least two, more preferably at least three, still more preferably at least four, most preferably at least five of the nucleic acid sequences shown in SEQ ID NO: 1 to 30. More preferably, the barcoding polynucleotide comprises at least one, preferably at least two, more preferably at least three, still more preferably at least four, most preferably at least five of the nucleic acid sequences shown in SEQ ID NO: 1 to 10. Still more preferably, the barcoding polynucleotide comprises or consists of the nucleotide sequence of SEQ ID NO: 31 (PolyloxP2.0 w/o restriction sites) or 32 (PolyloxP2.0 with restriction sites). Most preferably, the barcoding polynucleotide comprises or consists of the nucleotide sequence of SEQ ID NO: 33. Preferably, the term "barcoding polynucleotide" as used in accordance with the present invention further encompasses variants of the aforementioned specific polynucleotides, provided that said variants still are polynucleotides having the activity as specified above. The barcoding polynucleotide variants, preferably, comprise a nucleic acid sequence characterized in that the sequence can be derived from the aforementioned specific nucleic acid sequences by at least one nucleotide substitution, addition and/or deletion. Variants also encompass polynucleotides comprising a nucleic acid

sequence which is capable of hybridizing to the aforementioned specific nucleic acid sequences, preferably, under stringent hybridization conditions. These stringent conditions are known to the skilled worker and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N. Y. (1989), 6.3.1-6.3.6. A preferred example for stringent hybridization conditions are hybridization conditions in 6x sodium chloride/sodium citrate (= SSC) at approximately 45°C, followed by one or more wash steps in 0.2x SSC, 0.1% SDS at 50 to 65°C. The skilled worker knows that these hybridization conditions differ depending on the type of nucleic acid and, for example when organic solvents are present, with regard to the temperature and concentration of the buffer. For example, under "standard hybridization conditions" the temperature differs depending on the type of nucleic acid between 42°C and 58°C in aqueous buffer with a concentration of 0.1 to 5x SSC (pH 7.2). If organic solvent is present in the abovementioned buffer, for example 50% formamide, the temperature under standard conditions is approximately 42°C. The hybridization conditions for DNA:DNA hybrids are preferably for example 0.1x SSC and 20°C to 45°C, preferably between 30°C and 45°C. The hybridization conditions for DNA:RNA hybrids are preferably, for example, 0.1x SSC and 30°C to 55°C, preferably between 45°C and 55°C. The abovementioned hybridization temperatures are determined for example for a nucleic acid with approximately 100 bp (= base pairs) in length and a G + C content of 50% in the absence of formamide. The skilled worker knows how to determine the hybridization conditions required by referring to textbooks such as the textbook mentioned above, or the following textbooks: Sambrook et al., "Molecular Cloning", Cold Spring Harbor Laboratory, 1989; Hames and Higgins (Ed.) 1985, "Nucleic Acids Hybridization: A Practical Approach", IRL Press at Oxford University Press, Oxford; Brown (Ed.) 1991, "Essential Molecular Biology: A Practical Approach", IRL Press at Oxford University Press, Oxford. Alternatively, barcoding polynucleotide variants are obtainable by PCR-based techniques. Oligonucleotides suitable as PCR primers as well as suitable PCR conditions are described in the accompanying Examples. As a template, DNA or cDNA from bacteria, fungi, plants or animals may be used. Further, variants include polynucleotides comprising nucleic acid sequences which are at least 70%, preferably at least 80%, more preferably at least 90%, still more preferably at least 95%, most preferably at least 98% to at least one, preferably at least two, more preferably at least three, still more preferably at least four, most preferably at least five of the nucleic acid sequences shown in SEQ ID NO: 1 to 10. The percent identity values are, preferably, calculated over the entire nucleic acid sequence region. More preferably, the percent identity values are calculated over the recombinase recognition sequences, i.e. preferably, neglecting the uniquely identifiable

sequences of the polynucleotide. A series of programs based on a variety of algorithms is available to the skilled worker for comparing different sequences. In this context, the algorithms of Needleman and Wunsch or Smith and Waterman give particularly reliable results. To carry out the sequence alignments, the program PileUp (J. Mol. Evolution., 25, 351-360, 1987, Higgins et al., CABIOS, 5 1989: 151-153) or the programs Gap and BestFit (Needleman and Wunsch (J. Mol. Biol. 48; 443-453 (1970)) and Smith and Waterman (Adv. Appl. Math. 2; 482-489 (1981)), which are part of the GCG software packet (Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 (1991)), are to be used. The sequence identity values recited above in percent (%) are to be determined, preferably, using the program GAP over the entire sequence region with the following settings: Gap Weight: 50, Length Weight: 3, Average Match: 10.000 and Average Mismatch: 0.000, which, unless otherwise specified, shall always be used as standard settings for sequence alignments. The barcoding polynucleotides of the present invention either essentially consist of the aforementioned nucleic acid sequences or comprise the aforementioned nucleic acid sequences. Thus, they may contain further nucleic acid sequences as well. The barcoding polynucleotide of the present invention shall be provided, preferably, either as an isolated polynucleotide (i.e. isolated from its natural context) or in genetically modified form. The term encompasses single as well as double stranded polynucleotides. Moreover, comprised are also chemically modified polynucleotides including naturally occurring modified polynucleotides such as glycosylated or methylated polynucleotides or artificially modified ones such as biotinylated polynucleotides.

The term "uniquely identifiable sequence" is understood by the skilled person. Preferably, the uniquely identifiable sequence is a sequence unique, i.e., preferably, occurring only once, within the barcoding polynucleotide of the present invention. More preferably, a uniquely identifiable sequence is a sequence not occurring within a host cell; i.e., preferably, occurring only once within a host cell after introducing the barcoding polynucleotide of the invention into the host cell. Most preferably, a uniquely identifiable sequence is a sequence not occurring within a host organism; i.e., preferably, occurring only once within a host organism after introducing the barcoding polynucleotide of the invention into the host organism. For the avoidance of doubt, it is envisaged by the present invention that the uniquely identifiable sequences separated by a recombinase recognition sequence in a core structure of the present invention are uniquely identifiable sequences differing in sequence. Preferably, the uniquely identifiable sequence has a length of at least 5 nucleotides, preferably at least 10 nucleotides,

more preferably at least 25 nucleotides, even more preferably at least 50 nucleotides, most preferably at least 75 nucleotides. The barcoding polynucleotide of the present invention comprises at least five, preferably at least seven, more preferably at least nine, most preferably at least 10 core structures; accordingly, the barcoding polynucleotide of the present invention comprises at least ten, preferably at least 14, more preferably at least 18, most preferably at least 20 uniquely identifiable sequences. Preferably, each uniquely identifiable sequence comprised in a barcoding polynucleotide of the present invention differs from all other uniquely identifiable sequences present in said barcoding polynucleotide by at least one deletion, insertion, or, preferably, substitution.

The term "recombinase", as used herein, relates to a DNA recombinase capable of (i) mediating excision or (ii) mediating excision or inversion of a polynucleotide flanked by two recognition sequences of said recombinase. The term "excision" is known to the skilled person and relates to the removal of a polynucleotide fragment from a polynucleotide and re-establishing covalent bonding for the remaining polynucleotide, i.e., preferably, re-sealing of the double-strand break formally generated by removing the polynucleotide fragment. The term "inversion" of a polynucleotide fragment is known to the skilled person and relates to inverting the polynucleotide fragment relative to the surrounding nucleic acid sequence. Preferably, said inversion is mediated by excision of said polynucleotide fragment and its re-insertion in the inverse orientation. Suitable recombinases known in the art are listed in Table 1. As used herein, the term "recombinase" includes variants of the aforesaid recombinases, wherein said variants are active in excising polynucleotide fragments flanked by recombinase recognition sequences of said recombinase. Preferably, activity is established in the excision assay as specified elsewhere herein.

Table 1: Suitable recombinases

Recombinase	Recognition site	Source	Reference
Cre	loxP	Bacteriophage P1	
Vika	vox	Bacteriophage (Vibrio coralliilyticus)	Karimova et al., Nucleic Acids Research, 2012
VCre	VloxP	Bacteria (Vibrio)	Suzuki and Nakayama, Nucleic Acids Research, 2011
SCre	SloxP	Bacteria (Shewanella)	Suzuki and Nakayama, Nucleic Acids Research, 2011
Tre	loxLTR	modified from Cre	Sarkar et al., Science, 2007
Dre	rox	Bacteriophage P1 related phages	Sauer and McDermott, Nucleic Acids Research, 2004

Flp	frt	Yeast (<i>Saccharomyces cerevisiae</i>)	
R	RSRT	Yeast (<i>Zygosaccharomyces</i>)	Nern et al., PNAS, 2011
KD	KDRT	Yeast (<i>Kluyveromyces drosophilum</i>)	Nern et al., PNAS, 2011
B2	B2RT	Yeast (<i>Zygosaccharomyces bailii</i>)	Nern et al., PNAS, 2011
B3	B3RT	Yeast (<i>Zygosaccharomyces bisporus</i>)	Nern et al., PNAS, 2011
β	six	bacteria	Diaz et al., JBC, 2001
$\gamma\delta$	res	bacterial $\gamma\delta$ transposon	Schwikardi and Droge, FEBS Letters, 2000
PhiC31	attB/attP	Bacteriophage	Thomson et al., BMC Biotechnology, 2010
λ	attB/attP	Bacteriophage λ	Christ and Droge, Genesis, 2002
HK022	attB/attP	Bacteriophage HK022	Kolot and Yagil, Biotech Bioeng, 2003
CinH	RS2	Acetinetobacter	Thomson and Ow, Genesis, 2006
ParA	MRS		Thomson and Ow, Genesis, 2006
Tn1721	res	transposon Tn1721	Thomson and Ow, Genesis, 2006
Tn5053	res	transposon Tn5053	Thomson and Ow, Genesis, 2006
Bxb1	attB/attP	Mycobacteriophage Bxb1	Thomson and Ow, Genesis, 2006
TP901-1	attB/attP	Bacteriophage TP901-1	Thomson and Ow, Genesis, 2006
U153	attB/attP	Bacteriophage U153	Thomson and Ow, Genesis, 2006

Preferably, the recombinase is Cre, Dre, or Flp. More preferably, the recombinase is Cre.

The term "recombinase recognition sequence" is, in principle, known to the skilled person. Specific recognition sequences for particular recombinases can, e.g., preferably be obtained from the references cited in Table 1. Moreover, the term, preferably, also includes variants of the naturally occurring recombinase recognition sequence, provided that said variants still are functional in mediating excision as specified herein above. Preferably, said variant comprises at least one point mutation relative to the naturally occurring recombinase recognition sequence, more preferably at least one substitution mutation. As will be understood by the skilled person, e.g. a loxP site, which is a recombinase recognition sequence for Cre recombinase, comprises a 13 base pair 5' recognition sequence, an 8 base pair spacer region, and a 13 base pair 3' recognition sequence, wherein the 3' recognition sequence is the inverse complement of the 5' recognition sequence. Since the 8 base pair spacer region is less

conserved in loxP sequences, mutations are preferably located in said spacer region. However, functional mutations in a 5' or 3' recognition sequence are also known, e.g. from Missirlis et al. (BMC Genomic(2006), 7:73). Preferably, activity of a variant recombinase recognition sequence in mediating excision is established by incubating 1 µg of a linearized polynucleotide comprising two of said variant recombinase recognition sequences at a distance of 500 base pairs for 12 hours at 37°C with 1 unit of recombinase, wherein 1 unit of recombinase is the amount of recombinase polypeptide mediating excision of at least 20% of a corresponding wild type loxP-flanked polynucleotide within 1 hour under the same conditions. Preferably, a variant recombinase recognition sequence is considered active if at least 10 % of fragments flanked by said variant recombinase recognition sequence were excised in said assay ("excision assay"). Preferably, activity of a variant recombinase recognition sequence in mediating inversion is established by incubating 1 µg of a linearized polynucleotide comprising two of said variant recombinase recognition sequences at a distance of 500 base pairs for 12 hours at 37°C with 1 unit of recombinase, wherein 1 unit of recombinase is the amount of recombinase polypeptide mediating excision of at least 20% of a corresponding wild type loxP-flanked polynucleotide within 1 hour under the same conditions. Preferably, a variant recombinase recognition sequence is considered active if in at least 10 % of plasmids thus treated, the fragment flanked by said variant recombinase recognition sequences is inverted in said assay ("inversion assay"), wherein inversion is preferably tested for by digesting said plasmid with one or more appropriate restriction enzyme(s). As will be understood by the skilled person, catalysis of both inversion and excision is, in general, a property of the recombinase, not of the recombinase recognition sequence; accordingly, preferably, either the excision assay or the inversion assay is performed with respect to a given recombinase recognition sequence, and catalysis of both reaction types, or not, on said recombinase recognition sequence is deduced from the result of said assay.

As will be understood by the skilled person, the expression "recognition sequence of a recombinase" relates to a recombinase recognition sequence corresponding to said recombinase, i.e., to a recombinase recognition sequence recognized, preferably specifically recognized, by said recombinase. Preferably, each recombinase recognition sequence is separated from its neighboring recombinase recognition sequence or recombinase recognition sequences by at least 10, 25, or 50, more preferably at least 82, even more preferably at least 94, most preferably at least 178 nucleotides. Preferably, the recombinase recognition

sequences comprised in the barcoding polynucleotide of the present invention are recombinase recognition sequences of the same recombinase. As is understood by the skilled person, a recombinase may recognize recognition sequences differing in sequence by one or more nucleotide exchanges; e.g. Cre recombinase recognizes sequences corresponding to the generic consensus of SEQ ID NO:34. As is also known to the skilled person, a specific recombinase may be compatible, i.e. preferably, mediate excision and/or inversion, if it is combined with a recombinase recognition sequence of the same nucleic acid sequence, but not if it is combined with a recombinase recognition sequence of the same recombinase, but with a different sequence. Accordingly, the barcoding polynucleotide may comprise various recombinase recognition sequences of a recombinase, varying in nucleic acid sequence, wherein, preferably, said recombinase recognition sequences with different sequences are incompatible with each other. However, more preferably, the recombinase recognition sequences comprised in the barcoding polynucleotide of the present invention are mutually compatible recombinase recognition sequences, wherein, preferably, a first and a second recombinase recognition sequence are mutually compatible if they are active in the aforesaid excision assay, in which the fragment to be excised is flanked by the first and second recombinase recognition sequence. Most preferably, all recombinase recognition sequences comprised in the barcoding polynucleotide of the present invention are identical. Preferably, the recombinase recognition sequence comprises the sequence of SEQ ID NO: 34; more preferably, the recombinase recognition sequence comprises the sequence of SEQ ID NO: 35.

The term "core structure", as used herein, relates to a polynucleotide comprising two uniquely identifiable sequences as specified herein above, separated by a recombinase recognition sequence as specified herein above. Preferably, the core structure comprises exactly one recombinase recognition sequence per recombinase, i.e. preferably, per recombinase, one recombinase recognition sequence intervenes between the two uniquely identifiable sequences. More preferably, the core structure comprises exactly one recombinase recognition sequence. Preferably, the one recombinase recognition sequence comprised in said core structure is a Cre recombinase recognition sequence, more preferably comprising or having the sequence of SEQ ID NO: 34, more preferably of SEQ ID NO: 35.

During recombination, the sequences intervening between two recombinase recognition sequences are inverted or deleted randomly; Thereby, structures comprising two uniquely identifiable sequences separated by a recombinase recognition sequence are generated, which

either correspond to a core structure originally present in the barcoding polynucleotide of the present invention, or which represent a new combination of two uniquely identifiable sequences separated by a recombinase recombination sequence. Both types of structures are referred to herein as "barcode sequence" or "barcode". An example of how barcodes are generated according to the present invention is shown in Fig. 1.

The barcoding polynucleotide of the present invention comprises at least five of the aforesaid core structures. Preferably, the barcoding polynucleotide of the present invention comprises at least seven, more preferably at least nine, most preferably at least 10 of the aforesaid core structures. As will be appreciated by the skilled person, the upper limit of core structures is essentially only determined by practical aspects, e.g. size of the barcoding polynucleotide, and/or stability of the barcoding polynucleotide in cloning procedures and during amplification in, e.g. a plasmid. As detailed elsewhere herein, the diversity detectable from the products of the barcoding polynucleotide of the present invention depends on the method of detection. Accordingly, the barcoding polynucleotide of the present invention, preferably, comprises at most 250 core structures, more preferably at most 50 core structures, even more preferably at most 20, most preferably at most 15 core structures, preferably in case the barcoding polynucleotide shall be used in a method comprising cleaving of the product(s) of recombination with a restriction enzyme. Also preferably, the barcoding polynucleotide of the present invention comprises at most 100 core structures, more preferably at most 25 core structures, even more preferably at most 17, most preferably at most 15 core structures, preferably in case the barcoding polynucleotide shall be used in a method comprising sequencing the product(s) of recombination on the whole (in toto). Preferably, the orientation of the recombinase recognition sequence in at least one core structure is inverted relative to the orientation of the neighboring recombinase recognition sequences or the neighboring recombinase recognition sequence. More preferably, the orientation of the recombinase recognition sequence in at least two core structures is inverted relative to the orientation of the neighboring recombinase recognition sequences or the neighboring recombinase recognition sequence. Most preferably, in the barcoding polynucleotide of the present invention, the orientation of each recombinase recognition sequence is inverted relative to the orientation of the neighboring recombinase recognition sequences or the neighboring recombinase recognition sequence; i.e., preferably, the orientations of the recombinase recognition sequences are alternating. Preferably, the recombinase recognition sequence comprised in at least one core structure as determined starting from the starting nucleotide of the 5' uniquely

identifiable sequence is shifted by at least one nucleotide as compared to at least one further core structure comprised the barcoding polynucleotide of the invention. More preferably, the recombinase recognition sequence of each core structure of the barcoding polynucleotide of the invention as determined starting from the starting nucleotide of the 5' uniquely identifiable sequence is shifted by at least one nucleotide as compared to any further core structure comprised in said barcoding polynucleotide.

As will be understood, the barcoding polynucleotide of the present invention may comprise additional sequence elements, e.g. preferably, one or more restriction enzyme recognition sequences and/or sequencing primer annealing sites as specified elsewhere herein. In such case, preferably, the restriction enzyme recognition sequences are spaced such that sequencing of the complete fragments arising from cleaving the barcoding polynucleotide with an appropriate restriction enzyme is feasible with the selected sequencing method; more preferably, the restriction enzyme recognition sequences are arranged between the core structures of the barcoding polynucleotide; most preferably, all core structures of the barcoding polynucleotide are flanked by at least one restriction enzyme recognition site on both sides. Also in such case, preferably, the sequencing primer annealing sites are spaced such that sequencing of the complete barcoding polynucleotide is feasible with the selected sequencing method; more preferably, the sequencing primer annealing sites are arranged between the core structures of the barcoding polynucleotide; most preferably, all core structures of the barcoding polynucleotide are flanked by a sequencing primer annealing site. Preferably, the core structures are separated by at most 1000 base pairs, more preferably at most 100 base pairs in the barcoding polynucleotide of the invention. Most preferably, the core structures are directly linked in the barcoding polynucleotide of the invention.

The terms "restriction enzyme" and "restriction enzyme recognition sequence" are known to the skilled person. Preferably, the restriction enzyme is a Type II or Type III restriction enzyme. More preferably, the restriction enzyme is a Type IIS restriction enzyme, i.e. a restriction enzyme cleaving DNA at a defined distance outside of its own recognition sequence, but, preferably, not more than 50 base pairs away from said recognition sequence. Type IIS restriction enzymes are also known to the skilled person as "outside cutters". More preferably, the restriction enzyme is BsgI and/or BciVI. In principle, a restriction enzyme recognition sequence may be present at least once in a barcoding polynucleotide of the present invention. Preferably, the restriction enzyme recognition sequence is present at least

once in each core structure of the barcoding polynucleotide of the present invention. More preferably, the restriction enzyme recognition sequence is present at essentially the same location relative to the other elements in each core structure of the barcoding polynucleotide. Still more preferably, each core structure is flanked at its 5' side and at its 3' side by at least one restriction enzyme recognition sequence. Most preferably, each core structure is flanked at its 5' side and at its 3' side by two restriction enzyme recognition sequences. Preferably, the restriction enzyme recognition sequence is arranged such that the restriction enzyme cleavage site lies between the restriction enzyme recognition sequence and the most proximal uniquely identifiable sequence. More preferably, two restriction enzyme recognition sequences are positioned next to each other and oriented such that the restriction enzyme recognition sequences are located in between the restriction enzyme cleavage sites; i.e., preferably, the restriction enzyme recognition sequence of the second restriction enzyme lies between the restriction enzyme cleavage site and the restriction enzyme recognition sequence of the first enzyme, and the restriction enzyme recognition sequence of the first restriction enzyme lies between the restriction enzyme cleavage site and the restriction enzyme recognition sequence of the second enzyme. Said latter arrangement is particularly advantageous in certain applications, since by cleaving with both restriction enzymes, either simultaneously or consecutively, the restriction enzyme recognition sequences can be removed from the core structures. As will be understood by the skilled person, in directly adjacent core structures, a restriction enzyme recognition sequence or restriction enzyme recognition sequences are, preferably, present only once per core structure border. Preferably, the restriction enzyme recognition site is a restriction enzyme recognition site not present in any core element of the barcoding polynucleotide of present invention; more preferably, the restriction enzyme recognition site is a restriction enzyme recognition site not naturally present in the barcoding polynucleotide of the present invention.

The term "sequencing primer annealing site" is also known to the skilled person. Preferably, the sequencing primer annealing site is a stretch of nucleotides with known sequence, more preferably allowing annealing of an oligonucleotide at a temperature suitable for PCR and/or sequencing reactions. Preferably, each core element comprises a sequencing primer annealing site. More preferably, only the first and the last core element of a consecutive series of core elements comprise a sequencing primer annealing site. Preferably, in such case, the sequencing primer annealing sites are oriented such that amplification and/or sequencing of the core elements comprised between said sequencing primer annealing sites is possible. As

will be understood by the skilled person, in the latter case, the sequencing primer annealing site comprised in the first core structure and the sequencing primer annealing site comprised in the last core structure are, preferably, non-identical.

Advantageously, it was found in the work underlying the present invention, that recombinases like Cre do not act processively, but randomly excise and/or invert DNA fragments flanked by appropriate recognition sites from polynucleotides comprising a multitude of recognition sites. Accordingly, appropriate constructs can be used for generating random combinations of barcodes, which can be used to uniquely label cells. Moreover, it was found that, after removal of the recombinase, the random barcode combinations are stable and are suitable to track the progeny of labeled cells.

The definitions made above apply *mutatis mutandis* to the following. Additional definitions and explanations made further below also apply for all embodiments described in this specification *mutatis mutandis*.

The present invention further relates to a vector comprising a barcoding polynucleotide according to the present invention.

The term “vector”, as used herein, encompasses phage, plasmid, and viral (including, e.g. retroviral) vectors, as well as artificial chromosomes, such as bacterial, yeast, and mammalian artificial chromosomes. Moreover, the term also relates to targeting constructs which allow for random or site-directed integration of the targeting construct into genomic DNA. Such targeting constructs, preferably, comprise DNA of sufficient length for either homologous or heterologous recombination. The vector encompassing the barcoding polynucleotide of the present invention, preferably, further comprises a selectable marker for propagation and/or selection in a host cell. The vector may be incorporated into a host cell by various techniques well known in the art. For example, a plasmid vector can be introduced in a precipitate such as a calcium phosphate precipitate or rubidium chloride precipitate, or in a complex with a charged lipid or in carbon-based clusters, such as fullerenes. Alternatively, a plasmid vector may be introduced by heat shock or electroporation techniques. Should the vector be a virus, it may be packaged *in vitro* using an appropriate packaging cell line prior to administration to host cells. Suitable vectors are known in the art. Preferably, in case the vector is a plasmid vector, said plasmid has an origin of replication providing for medium or, preferably, medium

copy number in a bacterial host cell. Preferably, the vector is a gene transfer or targeting vector. Methods which are well known to those skilled in the art can be used to construct recombinant vectors; see, for example, the techniques described in Sambrook, *Molecular Cloning A Laboratory Manual*, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y. (1994). Most preferably, the vector is a vector comprising or consisting of the nucleotide sequence of SEQ ID NO: 47 (PolyloxP2.0_Rosa26 targeting vector, pWP-AG).

The present invention further relates to a host cell comprising a barcoding polynucleotide according to the present invention and/or a vector according to the present invention.

The term "host cell", as used herein, relates to any bacterial, archeal, or eukaryotic cell. Preferably, the cell is bacterial cell, more preferably an *Escherichia* cell (e.g. *E. coli*) or a *Bacillus* cell (e.g. *B. subtilis*). Preferably, the host cell is a eukaryotic cell; more preferably, the eukaryotic cell is a fungal cell, most preferably a yeast cell, e.g. a cell of *Saccharomyces cerevisiae*. Preferably, the eukaryotic cell is a mammalian cell, more preferably a cell of an experimental animal, e.g. preferably, of a mouse, rat, guinea pig, or hamster; most preferably, the host cell is a human cell. Preferably, the cell is a proliferation competent cell; more preferably, the cell is a cultured cell, preferably a cell line, more preferably a tumor cell line. It is, however, also envisaged that the host cell is a primary cell, preferably a primary tumor cell. Even more preferably, the host cell is a stem cell, most preferably an embryonic or somatic stem cell. Preferably, in case the host cell is a human embryonic stem cell, said human embryonic stem cell was obtained by blastocyst biopsy. Preferably, the host cell is not a human embryonic stem cell.

Moreover, the present invention relates to a host organism comprising a barcoding polynucleotide according to the present invention, a vector according to the present invention, and/or a host cell according to the present invention.

As used herein, the term "host organism" relates to a multicellular organism, preferably an experimental animal. Preferred experimental animals are: insects, preferably of the genus *Drosophila*, more preferably *D. melanogaster*; nematodes, preferably of the genus *Caenorhabditis*, more preferably *C. elegans*; Amphibia, preferably of the genus *Xenopus*, more preferably *X. laevis*; fishes, preferably of the genus *Danio*, more preferably *D. rerio*;

mammals, of which dogs, cats, horses, sheep, goats, and cattle are preferred. More preferably, the experimental animal is a rat, mouse, guinea pig, pig, or hamster. Preferably, the host organism is non-human.

The present invention also relates to a method of labeling a host cell, comprising introducing a barcoding polynucleotide according to the present invention into said host cell, contacting said barcoding polynucleotide with a corresponding recombinase in said host cell, and thereby labeling a host cell.

The method of labeling a host cell of the present invention, preferably, is an in vitro method. It is, however, understood by the skilled person, that the method may also be performed on a cell comprised in an organism; i.e. the method of labeling a host cell may also be an in vivo method. Moreover, the method may comprise steps in addition to those explicitly mentioned above. For example, further steps may relate, e.g., to detecting the combination of uniquely identifiable sequences after propagation of the cell. Moreover, one or more of said steps may be performed by automated equipment.

Methods for introducing a polynucleotide into a host cell are known in the art and are disclosed elsewhere herein. Preferably, introducing a barcoding polynucleotide is stably introducing said barcoding polynucleotide, i.e. preferably, eliciting integration of said barcoding polynucleotide into the genome of the host cell, preferably by homologous recombination or other integration mechanisms known to the skilled person. As will be understood by the skilled person, the barcoding polynucleotide of the present invention is preferably flanked by sequences of sufficient length homologous to the intended insertion site in case insertion by homologous recombination is to be obtained. Preferably, the recombinase recognition sequence comprised in said barcoding polynucleotide is a loxP sequence as specified elsewhere herein and wherein said corresponding recombinase is Cre as specified elsewhere herein.

Methods for contacting a barcoding polynucleotide with a recombinase in a host cell are known to the skilled person. E.g. preferably, the recombinase protein is introduced into the host cell. More preferably, a polynucleotide comprising an expression construct for said recombinase is introduced into the host cell. Preferably, said contacting a barcoding polynucleotide with a recombinase in a host cell comprises transient transfection of an

expression construct for said recombinase into the host cell. More preferably, said contacting a barcoding polynucleotide with a recombinase in a host cell comprises stably introducing a regulable expression construct for said recombinase, or comprises stably introducing an expression construct for a regulable variant of said recombinase. As will be understood by the skilled person, stable introduction is, preferably, achieved by transfecting said expression construct into a host cell under conditions which allow for integration of the expression construct into the genome of the host cell. It is, however, also envisaged by the present invention, that the expression construct is present in the genome of an experimental animal and the expression construct is introduced into said host cell by cross-breeding. Preferably, in such case, a first experimental animal stably carrying the barcoding polynucleotide of the present invention and a second experimental animal stably carrying the, preferably regulable, expression construct for the recombinase, or the expression construct for the regulable variant of the recombinase, are crossed. Preferably, the barcoding polynucleotide is contacted with the recombinase at least once each on at least two or, more preferably, at least three, preferably consecutive, days. Preferably, contacting a barcoding polynucleotide with a recombinase in a host cell comprises providing recombinase activity within the host cell for at least three, more preferably at least 6, most preferably, at least 12 hours. Preferably, contacting a barcoding polynucleotide with a recombinase in a host cell comprises providing recombinase activity within the host cell for a period of from at most seven days, preferably at most five days, still more preferably at most four days, most preferably, at most three days. Preferably, contacting a barcoding polynucleotide with a recombinase in a host cell comprises providing recombinase activity within the host cell for of from three hours to seven days, more preferably of from six hours to five days, still more preferably of from twelve hours to four days, most preferably of from one day to three days. As the skilled person will understand, the aforesaid values may require adjustment depending on the activity of the recombinase within the specific host cell; and adjustment can be accomplished according to the methods as provided in the examples elsewhere herein.

Methods of providing a regulable expression construct are known to the skilled person and, preferably, comprise cloning of a gene encoding a polypeptide comprising the recombinase in appropriate orientation downstream of a regulable promotor. Regulable promotors are well known in the art and include, e.g. tetracyclin-inducible or tetracyclin-repressible promotors.

Methods for obtaining regulable variants of the recombinase are also known in the art. Preferably, the regulable variant is fusion polypeptide comprising the recombinase and a polypeptide inhibiting recombinase activity and/or preventing the recombinase from contacting its substrate DNA. More preferably, the regulable variant of the recombinase comprises a polypeptide preventing the recombinase from accessing the nucleus in a eukaryotic, preferably mammalian, host cell. Even more preferably, the regulable variant of the recombinase comprises an estrogen receptor polypeptide, most preferably an estrogen receptor polypeptide as specified herein in the Examples.

Preferably, the barcoding polynucleotide introduced in the method further comprises restriction enzyme recognition sequences as specified herein above, and the method further comprises cleaving said barcoding polynucleotide at said restriction enzyme recognition sequences prior to sequencing the uniquely identifiable sequences.

Moreover, the present invention relates to a method of identifying a host cell, comprising labeling said cell according to the method of labeling a host cell of the present invention and further comprising sequencing uniquely identifiable sequences comprised in said barcoding polynucleotide after contacting said barcoding polynucleotide with said recombinase.

Advantageously, by incubating a barcoding polynucleotide of the present invention with a corresponding recombinase, a vast diversity of sequences is generated, comprising random deletions and inversions of sequences intervening two recombinase recognition sequences, of which each, if the number of core elements is selected appropriately, is unique for a specific cell, even if all cells of an organism are barcoded by the method of the present invention. Accordingly, the polynucleotide remaining from the barcoding polynucleotide of the present invention after recombination is also referred to as "barcoded polynucleotide". As will be understood, the diversity created lies with the specific barcodes remaining or newly created, as well as the relative arrangement of the barcodes to each other. Accordingly, a method comprising sequencing the barcoded polynucleotide as such (on the whole) will preserve the complete complexity, whereas a method comprising cleaving the barcoded polynucleotide with a restriction enzyme may only reveal complexity as far as the barcodes remaining or newly created are concerned, but not their relative arrangement. Accordingly, if high complexity is required, a method comprising sequencing the barcoded polynucleotide on the whole (in toto) is preferred. On the other hand, a method comprising cleaving the barcoded

polynucleotide with a restriction enzyme is technically less demanding and amenable to current high throughput sequencing equipment. As is understood by the skilled person, a barcoded polynucleotide comprising restriction enzyme recognition sequences may also be used for single molecule sequencing.

Thus, preferably, the barcoded polynucleotide remaining after recombinase action is sequenced in its entirety, preferably by long-range sequencing over the whole polynucleotide, or, also preferably, by sequencing overlapping fragments preferably using, e.g. uniquely identifiable sequences as sequencing primer binding sites, or by providing dedicated sequencing primer binding sites within the barcoding polynucleotide of the invention. Also preferably, restriction enzyme recognition sites intervening the core sequences are included in the barcoding polynucleotide of the invention as specified herein above and, after isolating DNA comprising the barcode, said DNA is cleaved by a restriction enzyme active on said restriction enzyme recognition sequences. Preferably, in such case, sequencing adapters comprising primer binding sites are ligated to the fragments generated by restriction enzyme digestion and fragments are sequenced, more preferably by a high-throughput method.

As will be appreciated by the skilled person, the method of labeling a host cell of the present invention enables several methods requiring barcoding of cells:

Accordingly, the present invention relates to a method for providing an experimental animal comprising a labeled population of cells, comprising labeling at least one host cell in said experimental animal, preferably a stem cell, more preferably a non-embryonic stem cell, according to the method of labeling a host cell of the present invention and keeping said experimental animal under conditions allowing proliferation of said host cell, thereby providing an experimental animal comprising a labeled population of cells. Preferably, labeling at least one host cell in an experimental animal comprises introducing a barcoding polynucleotide of the present invention into a cell of said experimental animal, e.g., preferably, by transfection or infection with a viral vector. Also preferably, labeling at least one host cell in an experimental animal comprises introducing a barcoding polynucleotide of the present invention into a cultured cell, preferably a cultured cell capable of proliferating in an experimental animal, and introducing said host cell into said experimental animal. Preferably, at least ten, more preferably at least 1000, most preferably at least a million, host

cells are labeled. It is, however, also envisaged that all cells of a specific tissue or organ or that all cells of a host organism are labeled.

Further, the present invention also relates to a method for barcoding host cells to study population dynamics in cell populations or in bacteria in vivo or, preferably, in vitro, comprising the method of labeling a host cell of the present invention. In said method, preferably, host cells are labeled and population growth dynamics is analyzed, preferably under various conditions like, preferably antibiotic treatment, niche competition, or limited nutrients.

Moreover, the present invention also relates to a method for barcoding of cells and tissues in mice. Preferably, within the hematopoietic system, hematopoietic stem cells (HSCs) barcoded by breeding knockin mice comprising the barcoding polynucleotide of the present invention, preferably stably integrated into the genome of at least one cell, with mice comprising a, preferably inducible, recombinase expression sequence. Preferably, said inducible recombinase expression sequence comprises an HSC specific promoter (e.g. Tie2). Preferably, said method is used to study clonal dynamics of native hematopoiesis and lineage relationships in diverse cell compartments (i.e., preferably, T- and B-cells, macrophages, granulocytes, dendritic cells etc.). Furthermore, using said method, development and maintenance of essentially any organ for which recombinase, preferably, Cre-specific inducible drivers exist can be examined. In addition to these studies aiming at developing tissue development maps under physiological conditions, the method can also be applied to address population dynamics of body cells during pathological conditions, e.g. inflammation, degeneration, regeneration, aging or functional adaptation.

Also, the present invention relates to a method for barcoding for studies of cancer development, progression and metastasis. Models of cancer development are known to the skilled person, e.g. using switching on oncogenes in a tissue- and time-defined manner. At the same time, in mice bearing the barcoding polynucleotide of the present invention together with an inducible recombinase, preferably under the control of tumor specific markers, tumor cells are tagged initially by the barcodes. Subsequently, the population dynamics of tumor development, progression and/or metastasis is studied. Preferably, e.g. relationships of metastasis-forming cells are established.

The present invention also relates to a kit comprising the barcoding polynucleotide according to the present invention, the vector according to the present invention, the host cell according to the present invention, or the experimental animal according to the present invention; and (a) a corresponding recombinase (i) mediating excision or (ii) mediating excision and inversion of a polynucleotide flanked by recognition sequences comprised in said barcoding polynucleotide; and/or (b) a polynucleotide encoding the corresponding recombinase of (a). Moreover, the present invention relates to a kit comprising (a) the barcoding polynucleotide according to the present invention or the vector according to the present invention; and (b) means for introducing said barcoding polynucleotide or vector into a host cell.

The term “kit”, as used herein, refers to a collection of the aforementioned components. Preferably, said components are combined with additional components, preferably within an outer container. The outer container, also preferably, comprises instructions for carrying out a method of the present invention. Examples for such components of the kit as well as methods for their use have been given in this specification. The kit, preferably, contains the aforementioned components in a ready-to-use formulation. Preferably, the kit may additionally comprise instructions, e.g., a user’s manual for applying the components with respect to the applications provided by the methods of the present invention. Details are to be found elsewhere in this specification. Additionally, such user’s manual may provide instructions about correctly using the components of the kit. A user’s manual may be provided in paper or electronic form, e.g., stored on CD or CD ROM. The present invention also relates to the use of said kit in any of the methods according to the present invention.

The present invention further relates to a host cell, obtained or obtainable by the method of labeling a host cell of the present invention. The present invention also relates to an experimental animal, obtained or obtainable by a method comprising the method of labeling a host cell of the present invention or by the method for providing an experimental animal comprising a labeled population of cells of the present invention.

Moreover, the present invention relates to the use of a barcoding polynucleotide according to the present invention and/or the vector according to the present invention for labeling a host cell.

In view of the above, the following embodiments are preferred:

1. A barcoding polynucleotide comprising at least five, preferably at least ten core structures, wherein each of said core structures comprises two uniquely identifiable sequences separated by a recombinase recognition sequence, wherein said recombinase is a recombinase capable of (i) mediating excision or (ii) mediating excision or inversion of a polynucleotide flanked by two recognition sequences of said recombinase.
2. The barcoding polynucleotide of embodiment 1, wherein said recombinase is Flp and wherein said recombinase recognition sequence is an *frt* sequence or, preferably, wherein said recombinase is Cre and wherein said recombinase recognition sequence is a *loxP* sequence.
3. The barcoding polynucleotide of embodiment 1 or 2, wherein each of said uniquely identifiable sequences differs from all other uniquely identifiable sequences present in said barcoding polynucleotide by at least one deletion, insertion, or, preferably, substitution.
4. The barcoding polynucleotide of embodiment 3, wherein the orientation of at least one recombinase recognition sequence is inverted relative to the orientation of the neighboring recombinase recognition sequences or the neighboring recombinase recognition sequence.
5. The barcoding polynucleotide of any one of embodiments 1 to 4, wherein the orientation of each recombinase recognition sequence is inverted relative to the orientation of the neighboring recombinase recognition sequences or the neighboring recombinase recognition sequence.
6. The barcoding polynucleotide of any one of embodiments 1 to 5, wherein said core structures are flanked by restriction enzyme recognition sequences.
7. The barcoding polynucleotide of any one of embodiments 1 to 6, wherein each of said core structures is flanked at its 5' side and at its 3' side by at least one, preferably two, restriction enzyme recognition sequence(s).
8. The barcoding polynucleotide of any one of embodiments 1 to 7, wherein said restriction enzyme recognition sequence is or wherein said restriction enzyme recognition sequences are identical for all core structures comprised in said barcoding polynucleotide.

9. The barcoding polynucleotide of any one of embodiments 1 to 8, wherein said restriction enzyme recognition sequences are recognition sequences of type IIS restriction enzymes, preferably BsgI and/or BciVI restriction enzyme recognition sequences.
10. The barcoding polynucleotide of any one of embodiments 1 to 9, wherein said recombinase recognition sequence comprised in at least one core structure is shifted relative to the starting nucleotide of the 5' uniquely identifiable sequence by at least one nucleotide as compared to at least one further core structure comprised in said barcoding polynucleotide.
11. The barcoding polynucleotide of any one of embodiments 1 to 10, wherein the recombinase recognition sequence of each core structure is shifted relative to the starting nucleotide of the 5' uniquely identifiable sequence by at least one nucleotide as compared to any further core structure comprised in said barcoding polynucleotide.
12. The barcoding polynucleotide of any one of embodiments 1 to 11, wherein said uniquely identifiable sequence has a length of at least 5 nucleotides, preferably at least 25 nucleotides.
13. The barcoding polynucleotide of any one of embodiments 1 to 12, wherein each of said recombinase recognition sequences is separated from any neighboring recombinase recognition site or recombinase recognition sites by at least 10, 25, or 50, preferably at least 82, more preferably at least 94 nucleotides.
14. The barcoding polynucleotide of any one of embodiments 1 to 13, wherein said core structures are separated by at most 100 nucleotides, preferably wherein said core structures are directly linked.
15. The barcoding polynucleotide of any one of embodiments 1 to 14, wherein said barcoding polynucleotide comprises the nucleotide sequence of at least one of SEQ ID NO: 1 to 30, preferably comprises the nucleotide sequence of at least one of SEQ ID NO: 1 to 10, more preferably comprises or consists of SEQ ID NO: 31 or 32, most preferably comprises or consists of SEQ ID NO:33.

16. A vector comprising a barcoding polynucleotide according to any one of embodiments 1 to 15.
17. A host cell comprising a barcoding polynucleotide according to any one of embodiments 1 to 15 or a vector according to embodiment 16.
18. A host organism comprising a barcoding polynucleotide according to any one of embodiments 1 to 15, a vector according to embodiment 16, or a host cell according to embodiment 17.
19. A method of labeling a host cell, comprising introducing a barcoding polynucleotide according to any one of embodiments 1 to 15 into said host cell, contacting said barcoding polynucleotide with a corresponding recombinase in said cell, and thereby labeling a host cell.
20. The method of embodiment 19, wherein introducing a barcoding polynucleotide is stably introducing said barcoding polynucleotide.
21. The method of embodiments 19 or 20, wherein the recombinase recognition sequence comprised in said barcoding polynucleotide is a loxP sequence and wherein said corresponding recombinase is Cre.
22. The method of any one of embodiments 19 to 21, wherein contacting a barcoding polynucleotide with a recombinase in a host cell comprises expressing a gene encoding said corresponding recombinase in said host cell.
23. The method of any one of embodiments 19 to 22, wherein expressing a gene encoding said corresponding recombinase is regulably expressing a gene encoding said corresponding recombinase; or is expressing a gene encoding a polypeptide comprising a regulable variant of said recombinase.
24. The method of any one of embodiments 19 to 23, wherein said barcoding polynucleotide is contacted with said recombinase at least once each on at least two or, preferably, at least three, preferably consecutive, days.

25. A method of identifying a host cell, comprising labeling said cell according to the method of any one of embodiments 19 to 23 and further comprising sequencing uniquely identifiable sequences comprised in said barcoding polynucleotide after said barcoding polynucleotide was contacted with said recombinase.
26. The method of embodiment 25, wherein, in said barcoding polynucleotide introduced, the core structures further comprise restriction enzyme recognition sequences, and wherein the method further comprises cleaving the resulting barcoded polynucleotide at said restriction enzyme recognition sequences prior to sequencing the uniquely identifiable sequences.
27. A method for providing an experimental animal comprising a labeled population of cells, comprising labeling at least one host cell in said experimental animal, preferably a stem cell, more preferably a non-embryonic stem cell, according to the method of any one of embodiments 19 to 24 and keeping said experimental animal under conditions allowing proliferation of said host cell, thereby providing an experimental animal comprising a labeled population of cells.
28. The method of embodiment 27, wherein said host cell of said animal is a host cell comprised in an experimental animal or is a host cell capable of proliferating in said experimental animal.
29. The method of embodiment 27 or 28, wherein at least ten, preferably at least 1000, more preferably at least a million host cells are labeled.
30. A kit comprising the barcoding polynucleotide according to any one of embodiments 1 to 15, the vector of embodiment 16, the host cell of embodiment 17, or the experimental animal of embodiment 18; and
- (a) a corresponding recombinase (i) mediating excision or (ii) mediating excision and inversion of a polynucleotide flanked by recognition sequences comprised in said barcoding polynucleotide; and/or
 - (b) a polynucleotide encoding the corresponding recombinase of (a).

31. A kit comprising (a) the barcoding polynucleotide according to any one of embodiments 1 to 15 or the vector according to embodiment 16; and (b) means for introducing said barcoding polynucleotide or vector into a host cell.
32. A labeled host cell, obtained or obtainable by the method of any one of embodiments 19 to 24.
33. An experimental animal, obtained or obtainable by a method comprising the method of any one of embodiments 19 to 24 or by the method of any one of embodiments 27 to 29.
34. Use of a barcoding polynucleotide according to any one of embodiments 1 to 15 and/or the vector of embodiment 16 for labeling a host cell.
35. A method for obtaining a barcoded polynucleotide, comprising contacting a barcoding polynucleotide according to any one of embodiments 1 to 15 or a vector according to embodiment 16 with a corresponding recombinase.
36. The method of embodiment 35, wherein said method is an in vivo method, preferably performed in a host cell or in a non-human animal.
37. The method of embodiment 35 or 36, wherein said method is a method for obtaining a barcoded polynucleotide comprising at least two, preferably at least four, more preferably at least six, even more preferably at least eight, most preferably at least ten barcode sequences.
38. A barcoded polynucleotide comprising at least two, preferably at least four, more preferably at least six, even more preferably at least eight, most preferably at least ten barcode sequences, obtained or obtainable by the method according to any one of embodiments 35 to 37.
39. A cell comprising a barcoded polynucleotide at least two, preferably at least four, more preferably at least six, even more preferably at least eight, most preferably at least ten barcode sequences, obtained or obtainable by the method according to any one of embodiments 19 to 24.

40. An experimental animal comprising at least one cell according to embodiment 39 and/or obtained or obtainable by the method according to any one of embodiments 27 to 29.

All references cited in this specification are herewith incorporated by reference with respect to their entire disclosure content and the disclosure content specifically mentioned in this specification.

Figure legends:

Figure 1: The PolyloxP2.0 barcoding cassette

A) Schematic drawing of the PolyloxP2.0 barcoding cassette. Triangles represent loxP recombination recognition sequences. Uniquely identifiable sequences are numbered from 1 to 20. Two examples for Cre-mediated recombination are depicted. Cre recombinase randomly selects two loxP recognition sites and the DNA sequence in between is either deleted (e.g. in recombination between the first and the third loxP site) or inverted (e.g. in recombination between the first and the second loxP site), depending on whether the selected loxP sites were in parallel or opposite orientation, respectively. As a result, new arrangements of the numbered uniquely identifiable sequences are formed. B) Schematic drawing of a core structure comprised in a barcoding polynucleotide. The dotted lines indicate specific restriction sites (BsgI/BciVI) for the fragmentation of the barcoding polynucleotide into barcodes for "next generation sequencing". C) List of the 90 theoretically possible new combinations of uniquely identifiable sequences in a barcoded polynucleotide derived from the PolyloxP2.0 barcoding cassette, together with the 10 unrecombined barcodes.

Figure 2: Work flow description

Reporter mice are obtained by combining an allele of the PolyloxP2.0 barcoding cassette and a tissue specific, inducible Cre recombinase gene. The promoter controlling Cre expression determines which tissue or type of cells will initially be barcoded. Treatment with tamoxifen enables the translocation of Cre into the nucleus, leading to the random recombination of the barcoding cassette in the cell population of interest. The specific barcodes of these cells will be inherited to their daughter cells and descending lineages. Different cell populations are isolated e.g. by FACS sorting and the genomic DNA is extracted. The barcoded cassette is

amplified by PCR. The occurrence of Cre-mediated deletion events can be visualized by gel electrophoresis. For full-resolution analysis of the barcoded polynucleotide, PCR products from individual cells are analyzed by traditional Sanger sequencing or the pool of sequences is subjected to third generation sequencing e.g. on the PacBio platform. Alternatively, the barcoded polynucleotides are digested into barcode sequences and then analyzed for recombination pairs of unique DNA elements via Illumina sequencing (e.g. 250 bp paired-end MiSeq analysis).

Figure 3: Gene targeting in ES cells and generation of ROSA26(PolyloxP2.0) knockin mice.

A) Strategy for targeting the PolyloxP2.0 cassette into the ROSA26 locus by homologous recombination. The first two exons of the wild-type ROSA26 locus are depicted together with the targeting vector and the targeted ROSA26 locus. The localization of the Southern blot probe and the BamHI restriction sites are indicated. SA, short arm of targeting vector (1.08kb). LA, long arm of targeting vector (4.32 kb). Neo, neomycin resistance gene for positive selection. PGK-DTA, diphtheria toxin subunit A with PGK promoter for negative selection. B) Example of data from the PCR screen for site-directed vector integration in ES cells. H₂O, water control. B6, genomic DNA from wild-type B6 ES cells used as PCR negative control. RosaRFP, genomic DNA from RosaRFP-targeted ES cells used as positive control. C) Southern blot on targeted ES cell clones used for the generation of ROSA26(PolyloxP2.0) knockin mice. Genomic DNA was digested with BamHI and blots were hybridized with a locus specific external probe (see also Fig.3A). Expected band sizes are 5.8 kb for the wild-type and 4.8 kb for the targeted ROSA26(PolyloxP2.0) allele. D) Genotyping of the offspring from chimeric mice for detecting germline transmission of the PolyloxP2.0 cassette. Genomic DNA from PolyloxP2.0-targeted ES cells and Rosa26RFP knock-in mice were used as PCR controls. The arrow indicates the lane with the PCR positive sample from the founder animal of the B6.ROSA26(PolyloxP2.0) mouse line.

Figure 4: In vitro recombination of PolyloxP2.0 on plasmid DNA.

A) Workflow. PolyloxP2.0 plasmid DNA was double digested with SacII and AscI, and then incubated with Cre recombinase for in vitro recombination and subsequent sequence analysis. B) Cre-mediated PolyloxP2.0 recombination products were separated by gel electrophoresis. The five different size fragments that are obtained by Cre-mediated excision are boxed. C) Recombination products were amplified by PCR, digested into barcodes and sequenced. The obtained frequencies of individual barcodes relative to the total number of barcodes

sequenced are depicted in a heatmap. Each row is representing an individual barcode and different frequencies are coded by different gray scales. 'WT' lists the ten different unrecombined barcodes and 'Recombination' the 90 recombined barcodes. The overall frequencies of recombined or unrecombined barcodes are depicted in the table below.

Figure 5: In vitro recombination of PolyloxP2.0 on genomic DNA.

A) Workflow. Genomic DNA was extracted from PolyloxP2.0-targeted ES cells and incubated with Cre recombinase, in vitro. Recombination products were amplified by PCR, separated by gel electrophoresis, and digested barcodes were sequenced. B) PolyloxP2.0-specific PCR on genomic DNA templates. Without Cre treatment (-Cre) one full-length wild-type band was obtained. After Cre treatment (+Cre) a ladder of five different size fragments was amplified by PCR. Genomic DNA from non-targeted ES cells was used as negative control (Neg.) for the PCR. C) PCR-amplified recombination products were digested into barcodes and sequenced. The obtained frequencies of individual barcodes relative to the total number of barcodes sequenced are depicted in a heatmap. Each row is representing an individual barcode and different frequencies are coded by different gray scales. 'WT' lists the ten different unrecombined barcodes and 'Recombination' the 90 recombined barcodes. The overall frequencies of recombined or unrecombined barcodes are depicted in the table below.

Figure 6: Tamoxifen-induced recombination of the PolyloxP2.0 cassette in ES cells.

A) Workflow. PolyloxP2.0-targeted ES cells, stably transfected with a MerCreMer expression plasmid, were treated once per day with different concentrations of 4-hydroxy-tamoxifen (4-OHT). The MerCreMer fusion protein requires binding of 4-OHT for its translocation into the nucleus and enfolding of Cre activity. During a four-day treatment period, aliquots of the cells were taken at the indicated time points for genomic DNA preparation, PCR amplification and subsequent sequencing analysis. B) Dose- and time-dependent Cre recombination in ES cells. MerCreMer-transfected PolyloxP2.0 ES cells were treated with different doses of 4-OHT and genomic DNA was prepared at the indicated time points. The barcoded PolyloxP2.0 cassette was amplified by PCR and products were separated by gel electrophoresis. Water (H₂O) and wild-type ES cell DNA (E14) were used as PCR negative control. Genomic DNA from PolyloxP2.0 ES cells treated with Cre in vitro served as PCR positive control. Negative control for the 4-OHT treatment were vehicle (EtOH) treated MerCreMer-transfected PolyloxP2.0 ES cells. C) Heat map of the barcode distribution in 4-OHT treated ES cells. Recombination of the PolyloxP2.0 cassette by MerCreMer in ES cells treated with 100 nM 4-

OHT was analyzed by Illumina 250 bp paired-end high-throughput MiSeq sequencing and frequencies of individual barcodes are depicted in a heat map. Each lane represents a treated ES cell sample, each row an individual barcode. 'WT' lists the ten different unrecombined barcodes and 'Recombination' the 90 recombined barcodes. Frequencies are coded by different gray scales.

Figure 7: Pulse-chase recombination in MerCreMer-transfected PolyloxP2.0 ES cells.

ES cells were induced for 3 hours with a single pulse of 100 nM 4-OHT, then washed and genomic DNA was prepared from aliquots of the cells at the indicated time points for PCR amplification of the PolyloxP2.0 cassette. Recombination could already be detected by PCR after 3 hours of incubation (0 days) and was even more pronounced after 3 days of chase. From that time on, no further recombination towards smaller size PCR products could be observed.

Figure 8: Full-length sequencing of the recombined PolyloxP2.0 cassette (barcoded polynucleotide) from single cells.

MerCreMer-transfected PolyloxP2.0 ES cells were induced with 100nM 4-OHT for one day and sorted on day seven into PCR tubes for subsequent single cell analysis. A) The barcoded PolyloxP2.0 cassette was amplified from individual cells by nested PCR. B) PCR products were analyzed by Sanger sequencing. The observed barcode patterns from 18 individual cells are listed.

Figure 9: Tamoxifen-induced recombination of the PolyloxP2.0 cassette in Rosa26^{ERT2-Cre/PolyloxP2.0} mice.

A) Mice carrying in the Rosa26 locus the inducible ERT2-Cre allele and the PolyloxP2.0 cassette (Rosa26^{ERT2-Cre/PolyloxP2.0}) were treated with 1 dose or 4 daily doses of Tamoxifen or with oil control as indicated. On day 5, mice were sacrificed and organs were removed for preparation of genomic DNA. B) Cre-mediated PolyloxP2.0 recombination on tissues samples was analyzed using locus-specific PCR. C) Heatmap presentation of the recombination products in different organs analyzed by Illumina MiSeq 250bp paired-end sequencing. Each lane represents a different DNA sample. Each row represents an individual barcode. 'WT' lists the ten different unrecombined barcodes and 'Recombination' the 90 recombined barcodes. Frequencies are coded by different gray scales.

Figure 10: Barcode probability in PCR sampling controls

In all subfigures, each data point represents one barcode and its abundance in either of two PCR samples and is plotted on its respective axis. Unrecombined “WT” barcodes are depicted as open squares, recombined barcodes are shown as closed black circles. A) Genomic DNA from the lung of one non-induced Rosa26^{ERT2-Cre/PolyloxP2.0} mouse was extracted. Recombination products of the PolyloxP2.0 cassette were amplified in two separate PCR reactions and analyzed by Illumina sequencing. The abundance of individual barcodes in each of the two PCR reactions is plotted against each other. B) Aliquots of genomic DNA from the lung of a 1-dose-treated Rosa26^{ERT2-Cre/PolyloxP2.0} mouse were amplified in two separate PCR reactions, sequenced and the abundance of barcodes was plotted. C) Aliquots of genomic DNA from the lung of a 4-dose-treated Rosa26^{ERT2-Cre/PolyloxP2.0} mouse were amplified in two separate PCR reactions, sequenced and the abundance of barcodes was plotted.

Figure 11: Barcode probability plots from different samples.

In each subfigure, each data point represents one barcode and its abundance in either of two PCR samples and is plotted on its respective axis. Unrecombined “WT” barcodes are depicted as open squares, recombined barcodes are shown as closed black circles. A) Genomic DNA from the spleen and lung of a 1-dose-treated Rosa26^{ERT2-Cre/PolyloxP2.0} mouse was extracted. Recombination products of the PolyloxP2.0 cassette were amplified by PCR and sequenced. The abundance of the barcodes in the two different organs is plotted against each other. B) Genomic DNA from the spleen of a 1-dose and a 4-dose-treated Rosa26^{ERT2-Cre/PolyloxP2.0} mouse was extracted. Recombination products of the PolyloxP2.0 cassette were amplified and sequenced. The abundance of the barcodes upon different treatments is plotted against each other.

Figure 12: Plasmid map of the PolyloxP2.0_Rosa26 targeting vector.

The plasmid has the sequence of SEQ ID NO: 47.

Figure 13: PolyloxP2.0 full-length read coding.

A) Schematic drawing of the PolyloxP2.0 barcoding cassette in its original unrecombined conformation. For reference, terminology of the unique DNA barcode elements used in restriction digest-based Illumina sequencing experiments (upper row, numbers 1-20 in boxes) used in Figures 1-11 is compared to the simplified terminology (lower row, letters A-I in boxes) used for the DNA segments obtained by full-length read coding (FRC) based PacBio

sequencing experiments (Example 8, Table 3). B) PolyloxP2.0 unrecombined FRC starting code. C-E) Examples of PolyloxP2.0 FRC recombinations and the nomenclature used: C) FRC obtained after excision of the segments A and B. D) FRC originating from an inversion of the entire Polylox cassette. E) FRC corresponding to the product of two recombination events: excision of segments B and C plus and inversion of segment H.

The following Examples shall merely illustrate the invention. They shall not be construed, whatsoever, to limit the scope of the invention.

Example 1: Design of a PolyloxP2.0 cassette and cell-free recombination assay on plasmid DNA

We designed a specific substrate for random Cre mediated DNA recombination and cellular DNA barcoding in vivo. The PolyloxP2.0 recombination cassette consists of 10 loxP sites with alternating orientation and a fixed distance of 178 bp between each two neighboring loxP sites. This distance was calculated according to the optimal closest distance of 94bp reported by Hoess et al. (1985), Gene 40(2-3):325 plus an extension of eight DNA windings of 10.5bp.

Diversity, a prerequisite for efficient DNA barcoding, is achieved by partial Cre-mediated recombination of the PolyloxP2.0 cassette. In an embodiment, the alternating orientation of the loxP sites allows an increased recombination product diversity upon Cre activity compared to an “all-same-direction” construct, because it enables both: excisions and inversions of DNA segments between loxP sites with parallel or opposite directions, respectively.

In order to generate a synthetic PolyloxP2.0 cassette with a DNA content that is distinct to the mouse genome but still displays a maximized physiological nucleotide distribution, we chose, with slight modifications, natural DNA segments from a plant gene, namely the cell wall synthase gene *CesA9* of *Arabidopsis thaliana* for the 20 unique DNA fragments flanking the loxP sites (see Fig. 1A). The modifications are the deletion of two nucleotides at the start and end of *CesA9* exons and introns to remove all possible RNA splicing sites within the cassette, and the deletion of T/A repeats, as well as the erasure (*Bam*HI, *Nsi*II, *Bci*VI) and introduction (*Spe*I) of specific restriction sites.

To enable the identification and read-out of recombined PolyloxP2.0 products by second generation sequencing, e.g. on the Illumina MiSeq platform, in an embodiment, we also included specific restriction sites in the PolyloxP2.0 construct. Illumina MiSeq provides a highly efficient sequencing technology, since, with an average of 25 mio. reads per flow cell, it even allows for the simultaneous analysis of several multiplexed sample libraries in a single run. However, with a current maximal read length of 500-600 bp (300 bp paired-end) it cannot provide the analysis of the entire PolyloxP2.0 cassette in one piece. Instead, it requires the analysis of specific fragments of it. Therefore we placed a certain restriction cassette between all of the loxP sites and positioned them centered between two neighboring loxP sites, allowing for the digestion of the PolyloxP2.0 cassette into so-called “core” fragments. Each core fragment has a size of 192 bp and consists of one loxP site as well as one upstream and one downstream uniquely identifiable sequence. As shown in Fig. 1, the combinations of these uniquely identifiable sequences are indicative for specific recombination events.

For the digestion of the PolyloxP2.0 cassette into its core fragments, we chose type IIS (shifted cleavage) endonucleases, namely BsgI and BciVI. Their recognition sequences were placed into the construct as a 9 bp restriction cassette in such a way that the double-digest generates the core fragments and at the same time removes the entire recognition sequences (see Fig. 1B). This removal is advantageous, since it ensures that all the different core fragments start with unique DNA sequences and don't share the common restriction recognition sequences, simplifying Illumina based sequencing.

A second feature that was considered for the same technical reason was a positional +1 bp shift of the loxP sites in comparison to the position of the loxP site in the previous core fragment. Like this, we give each of the 20 unique DNA elements a different length but at the same time maintain the equal distance between all loxP sites. Hence, we ensure that the common loxP sequence is reached in a different sequencing cycle for each of the 20 unique DNA elements.

For the cell-free recombination assay, PolyloxP2.0 targeting vector (1µg) carrying the PolyloxP2.0 cassette was linearized by SacII and AscI (NEB), and then incubated with purified Cre recombinase (NEB, M0298L) for overnight at 37°C. The reaction was terminated by heating at 70°C for 10 min. Afterwards the mixture was sequentially digested by BsgI and BciVI, and the digestion products were purified by gel extraction (QIAquick Gel Extraction

Kit, QIAGEN, 28706). Finally, extracted DNA was quantified (Qubit 2.0) and analyzed for its purity (Agilent 2100 Bioanalyzer) before being used for subsequent library preparation and DNA sequencing.

Example 2: Cell-free recombination assay on genomic DNA

From cell pools larger than 10^4 cells (e.g. PolyloxP2.0 targeted ES cells or mouse organs) genomic DNA was first purified by phenol-chloroform extraction and isopropanol precipitation. Purified genomic DNA was then incubated with Cre recombinase (NEB, M0298L) for overnight at 37°C and the reaction was terminated by heating at 70°C for 10 min. Afterwards, the PolyloxP2.0 cassette (together with some flanking Rosa26 sequence) was amplified by PCR with the Expand Long Template PCR System (Roche, 11759060001).

PCR conditions were as follows:

Step1: 5 min 95°C, Step2: 30s 95°C, Step3: 30s 62°C, Step4: 5 min 72°C, repeat Steps 2-4 for 35 times, continue to final Step5: 10 min 72°C.

Forward primer #493 = 5'-GCAAGCACGTTTCCGACTTGAG-3' (SEQ ID NO: 36)

Reverse primer #2427 = 5'-CATACCTTAGAGAAAGCCTGTCTGAG-3' (SEQ ID NO: 37).

The PCR products were first purified by PCR purification kit (QIAGEN, 28106), digested by BsgI and BciVI, and finally the 200 bp “core-fragments” were purified from the digestion mixture by gel extraction (QIAquick Gel Extraction Kit, QIAGEN, 28706).

Example 3: Induction and analysis of intracellular recombination

The pANMerCreMerpuromycin expression plasmid carrying the tamoxifen-inducible MerCreMer under the control of human beta-actin promoter (ACTB) was transfected into PolyloxP2.0 targeted ES cells to enable the induction of intracellular recombination. Transfected ES cells were selected with puromycin and screened for stable integration by PCR using the following conditions:

Expand Long Template PCR System (Roche, 11759060001), Step1: 5 min 95°C; Step2: 30s 95°C, Step3: 30s 54°C, Step4: 4 min 72°C, repeat Steps 2-4 for 35 times, continue to final Step5: 10 min 72°C.

Forward Primer #2456: 5'-CCATGGGAGATCCACGAAATG-3' (SEQ ID NO:38)

Reverse Primer #2457: 5'-CCTGGTATCTTTATAGTCCTG-3' (SEQ ID NO:39).

For the induction of cellular recombination, a clone of MerCreMer-transfected PolyloxP2.0 ES cells was induced with 4-hydroxytamoxifen (4-OHT) at the indicated doses and time points. Stock solution (800 μ M) was prepared by dissolving 25 mg 4-OHT (Sigma) in 80.6 mL absolute ethanol and was stored at -20°C until use. Preparation of genomic DNA and amplification of the PolyloxP2.0 cassette for high-throughput sequencing were done as described under “Cell-free recombination assay on genomic DNA”.

Example 4: Induction and analysis of PolyloxP2.0 recombination in mice

PolyloxP2.0 targeted ES cells were used to generate B6-Gt(ROSA)26Sor^{tm1(PolyloxP)Hrr} mice (short name Rosa26^{PolyloxP/+}). Those mice were subsequently crossed to Rosa26^{Cre-ERT2} mice (B6.129-Gt(ROSA)26Sortm1(cre/ESR1)Tyj/J). Mice were injected intraperitoneally with 1 mg tamoxifen once or daily on 4 consecutive days. Stock solution was prepared by dissolving 1 g tamoxifen (Sigma T5648) in 36 mL peanut oil (Sigma P2144) and 4 mL absolute ethanol at 55°C. Afterwards, genomic DNA was prepared from thymus, spleen and lung and amplified for subsequent sequencing as described under “Cell-free recombination assay on genomic DNA”.

Example 5: DNA library preparation and multiplexing for Illumina sequencing

Sequencing libraries from individual DNA samples (10ng) were prepared using NEBNext High-Fidelity 2X PCR Master Mix (NEB, E6240). End-repair, dA-tailing and adaptor ligation were done according to the manufacturer's standard protocol. Afterwards, adaptor-ligated DNA was directly used for PCR enrichment (10 PCR cycles) and indexing without size selection. Index primers were provided in NEBNext Multiplex for Illumina (NEB E7335). Distinct DNA libraries were normalized to prepare an equimolar multiplex (10 nM) and mixed together with 5% PhiX carrier DNA. Sequencing was performed using Illumina Miseq V2: 250bp paired-end sequencing platform.

Example 6: Single Cell sequencing

PolyloxP2.0 targeted ES cells carrying inducible Cre (pANMerCreMerpuro transfected clone) were treated with 100nM 4-hydroxytamoxifen for one day, then washed and cultured for 6 days. Next, single cells were sorted into individual PCR tubes by FACS sorting (FACSARIA III, BD Biosciences) and digested by Proteinase K treatment at 55°C for 1 h. The reaction was terminated by heating at 95°C for 10 min and the mixture was directly used for the amplification of PolyloxP2.0 recombination products by nested PCR.

First PCR round:

Expand Long Template PCR System (Roche, 11759060001)

Step1: 5 min 95°C; Step2: 30s 95°C, Step3: 30s 56°C, Step4: 5 min 72°C, repeat Steps 2-4 for 35 times, continue to final Step5: 10 min 72°C.

Forward Primer #2450: 5'-TGTGGTATGGCTGATTATGATCAG-3' (SEQ ID NO: 40)

Reverse Primer #494: 5'-AGCTACAGCCTCGATTTGTGGTG-3' (SEQ ID NO: 41).

Second PCR round:

From first round, 1 µL PCR reaction was used as template in a 50 µL PCR reaction using the same PCR program as in the first round but with the following nested primers:

Forward primer #2426: 5'-CGACGACACTGCCAAAGATTTC-3' (SEQ ID NO: 42) and

Reverse primer #2427 (SEQ ID NO: 37).

PCR products were purified by PCR purification kit (QIAGEN, 28106) and sent to Sanger sequencing using oligos #2426 and #2427 as sequencing primers.

Example 7: Gene targeting

Murine embryonic stem (ES) cells, clone JM8A3 derived from C57BL/6N (Pettitt et al., Nat Meth. 2009, PMID 19525957), were cultured on neomycin-resistant embryonic fibroblasts in Knockout DMEM (GIBCO) supplemented with 10 % FCS (Hyclone), 2mM GlutaMAX (GIBCO), 1 mM sodium pyruvate (GIBCO), 0.1 mM nonessential amino acids (GIBCO), 100 U/ml penicillin, 100 µg/ml streptomycin (GIBCO), 25 µM 2-mercaptoethanol (GIBCO) and 1000 U/ml LIF (Chemicon). ES cells were electroporated with 30 µg of the linearized targeting vector (pWP-AG, PolyloxP2.0_Rosa26 targeting vector) at 500 µF, 0.24 kV (Gene Pulser Xcell, BioRad). Starting one day after electroporation, cells were selected by adding 150 µg/mL Geneticin (G418, GIBCO) to the culture medium. On day 10 of selection, clones

were randomly picked, expanded, and screened for correct homologous recombination by PCR using one external forward primer HL16: 5'-CCTAAAGAAGAGGCTGTGCTTTGG-3' (SEQ ID NO:43) and a vector specific reverse primer HL15: 5'-AAGACCGCGAAGAGTTTGTCC-3' (SEQ ID NO:44) (Lucas et al. (2007), EJI 37(1):43, PMID: 17171761). Correct homologous recombination was confirmed by Southern blot using a biotinylated 822 bp probe located upstream of the first exon and the short arm of homology. The probe was obtained by PCR amplification (Oligo2424: 5'-GCAAAGGCGCCCGATAGAATAA-3' (SEQ ID NO: 45) and Oligo2425: 5'-CCGGGGGAAAGAAGGGTCAC-3' (SEQ ID NO: 46) of genomic C57BL/6 DNA and was labeled by random prime labeling (North2South Biotin Random Prime Labeling Kit, Thermo Scientific).

Generation of the ROSA26^{PolyloxP2.0} knockin mouse line.

Cells from one ROSA26^{PolyloxP2.0} knockin ES clone were injected into day3.5 C57BL/6 blastocysts and implanted into pseudopregnant foster mothers. Chimeric males were backcrossed to C57BL/6N females to transmit the ROSA26^{PolyloxP2.0} knockin allele through the germline. Offspring were tested by PCR and one positively tested male was selected as founder of the ROSA26^{PolyloxP2.0} knockin mouse line.

Example 8: Exploiting the full combinatorial potential of the PolyloxP2.0 cassette by third generation sequencing.

The PolyloxP2.0 cassette can undergo recombination events at multiple sites within the same molecule. To decipher the entire barcode in its full complexity, the recombined cassette preferably is sequenced on the whole as a single molecule. We achieved this by third generation sequencing, using the Single Molecule Real Time (SMRT®) technology on the PacBio RS platform (Pacific Biosciences). With this method, DNA molecules of several thousand base pairs can be sequenced. In order to decipher and uniquely identify full-length read codes (FRCs) of such complex PolyloxP2.0 recombination events, we annotated a 9-digit code to the recombination products representing the position and orientation of the DNA segments between the loxP sites of the recombined PolyloxP2.0 cassette. Based on the unrecombined PolyloxP2.0 cassette, the nine DNA segments spacing the ten loxP sites are denominated with the capital letters A-I. For global orientation of the FRC, the DNA sequence before the first loxP site (barcode 1) defines the 5' end of the code, the DNA

sequence after the last loxP site (barcode 20) defines its 3' end. In case of a sequence inversion relative to this 5'-3' orientation, the affected DNA segments are labeled with their respective small letters a-i. In Fig. 13 there are four exemplary FRCs depicted: the unrecombined full-length code 5'-ABCDEFGHI-3' (Fig.13B), excision of the first two segments 5'-CDEFGHI-3' (Fig.13C), inversion of the entire cassette FRC 5'-ihgfedcba-3' (Fig.13D), and a combination of an excision of segments B+C with an inversion of the second last segment H, FRC 5'-ADEFgHI-3'. DNA sequences of the respective DNA fragments are shown in Table 2 and in SEQ ID NOS: 48 to 65. The theoretical combinatorial diversity of the PolyloxP2.0 FRCs reaches 1.86 million different codes, which is multiple orders of magnitude higher than the barcodes obtained by the restriction digest-based Illumina read-out method.

In order to test the SMART sequencing based identification of full-length barcodes in vivo, we analyzed granulocytes from tamoxifen-induced Rosa26^{Cre-ERT2/PolyloxP2.0} mice. In brief, mice were treated with tamoxifen (1 mg i.p.) and bone marrow cells were isolated several weeks after induction. Cell suspensions were stained with antibodies against CD4 (Invitrogen, RM4.5), CD8 (Pharmingen, 53-6.7), CD11b (eBioscience, M1/70), CD19 (Pharmingen, 1D3) and Gr-1 (Pharmingen, RB6-8C5). Granulocytes were isolated as Gr-1⁺CD11b⁺CD4⁻CD8⁻CD19⁻ cells on a FACS Aria III (BD Biosciences) cell sorter. Genomic DNA extraction and PCR amplification from sorted granulocytes were in principle done as described under Example 2 "Cell-free recombination assay on genomic DNA".

Specific PCR conditions were as follows:

Step1: 5 min 95°C, Step2: 30s 95°C, Step3: 30s 56°C, Step4: 5 min 72°C, repeat Steps 2-4 for 35 times, continue to final Step5: 10 min 72°C.

Forward primer #2450 = 5'-TGTGGTATGGCTGATTATGATCAG-3' (SEQ ID NO: 40)

Reverse primer #2427 = 5'-CATACCTTAGAGAAAGCCTGTGAG-3' (SEQ ID NO: 37).

PCR products were purified and size selected using the Agencourt AMPure PB beads system according to manufacturer's protocol. During this procedure, PCR products were split into a "small fragments" and a "large fragments" fraction by extraction with 0.9x and 0.4x AMPure beads, respectively. Both fractions underwent library preparation procedure according to the manufacturer's protocol (SMRTbell Template Prep Kit, Pacific Biosciences 100-259-100) and were sequenced on the PacBio RS platform (Pacific Biosciences) using standard protocols. The "small fragments" library was loaded by diffusion mode, the "large fragments"

library by MagBead mode. The obtained circular consensus sequence (CCS) reads of both fragment libraries were merged and the PolyloxP2.0 FRC barcode for each of the reads was determined. A summary of the codes detected in the experiment is listed in Table 3.

Table 2: Nucleotide sequences relevant for this specification

SEQ ID NO	Name	Sequence (5'→3')
1	core structure 1	caacaggaatgaattcgttctcattaacgccgacgacactgccaaagatttctcaatttgttctttgagaaac aagaatgataacttcgtatagcatacattatacgaagttattatcttctgtttggatcgatctgggttggagattg aatctgcgttttctaaaacagaagtctttatggaaatttcacctgctgctc
2	core structure 2	atattgcatgattatagctgtcggatagatcaatgagacaaaagttagtgtcagttcagactcatctccatgc ggttataacttcgtataatgtatgctatacgaagtatgtcagctgttgagtattaaattcgttgtcatttgtgtg ggattccaagaagttgcatacacagtattgttaattaatggaaagt
3	core structure 3	actttatgaagtttgttaagaactagatcaggagcttgttaaatgatataataagattgatcttgtttgttacta tataacttcgtatagcatacattatacgaagttatgaatgaatgtggcaatgtgtacacgatcagcagaagaac tgagtgggcaaacttgtaaaatctgtagagatgagattgagttaacaga
4	core structure 4	gtgagccatttatcgcttgcaacgaatgtgctttccctacgtgtagaccgtgctatgagtatgaaagaagaga aggaaaataacttcgtataatgtatgctatacgaagttatccaagctgtcctcagtgcggaacccgatacaa gcgtattaaaagagacattttcttctgtatttttatctcaacacttgctttgag
5	core structure 5	tggattgatttatgtttagtccaagggtcgaaggggatgaagaagatgatgacattgatgatctggagcatg agttataacttcgtatagcatacattatacgaagttattatggaatgggtcctgaacatgttactgaagctgcgc tctattatgcgtcttaacactggctgtggtactgatgaagtgtcacacttg
6	core structure 6	agcttcaccaggctctgaggttcctcttttgacgtattgtgatgcagacttcatgatcgattttctgtgtacattaa ataacttcgtataatgtatgctatacgaagttattgtcgaaattcctacgggtgtgatgattctttgatatacatgtt ctgatattgattcggatagacatgctcttatcgtgccccatctac
7	core structure 7	tggggaacagagtccatcatgtgccatttacagattctttgcatcatgtgttactcataagctctatctcaaaat ataacttcgtatagcatacattatacgaagttatgctatcggtttgtatgaagttgcctcctgatgaagaaaccca catgttggttaccacacaagacccatgggtcctcagaaggatcttacagtt
8	core structure 8	atatggaagtgttgcattggaagaccgtatggaagtatggaagaaacaacaatagaaaagctacaagtcg ttaataacttcgtataatgtatgctatacgaagttatgaatgaaagagtaaatgatgggtgatggcgatggattc attgttgacgagctcgatgatcccggtaccacaaagttgaaaccttttactgcatt
9	core structure 9	taaagaattgtgtatatgtattttctgacatttctttgacgtggttctggatgaaggaaagacaacctctgtcaata acttcgtatagcatacattatacgaagttatcgaaagctaccattcgttcgagcagaataaaccttatagga tgttgattttctgtcggctcgccattcttggttcttccattatagga
10	core structure 10	catccggtgaatgatcgattcggctgtggttaacatcggttatatgcgaaatatgggtcgtgtgtcttggaaat aacttcgtataatgtatgctatacgaagttattcttgatcagttccaaaatgggtataccatcgaaagagaaac atatctcgacaggctttcttaaggtatgagaagggaagggaagccatcagag
11	barcode 1	caacaggaatgaattcgttctcattaacgccgacgacactgccaaagatttctcaatttgttctttgagaaac

		aagaatg
12	barcode 2	tatcttctgtttggatcgatctgggttggagattgaatctgcgttttctaaacagaagtccttatggaaattcac ctgctgtc
13	barcode 3	atattgcatgattatatacgtgtcggatatgatcaatgagacaaaagttagtgtagttcagactcatctccatgc ggttt
14	barcode 4	gtcagcttgttgagtatttaaattcgtttgtcatttgttggattccaagaagttgcatacacagtattgttaatta atggaaagt
15	barcode 5	actttatgaagttttgttaagaactagatcaggagctttgttaaataatgatataataagattgatcttgtttgttacta t
16	barcode 6	gaatgaatgtggcaatgtgtacacgatcagcagaagaactgagtgaggcaaacctgtaaaatctgtagagat gagattgagttacaga
17	barcode 7	gtgagccatttatcgcttgcaacgaatgtgctttccctacgtgtagaccgtgctatgagtatgaaagaagaga aggaaa
18	barcode 8	ccaagcttgcctcagtcggaacccgatacaagcgtattaaaagagacattttcttctgtattttatctcaa cacttgctttgag
19	barcode 9	tggattgattatgtttagtccaagggtcgaaggggatgaagaagatgatgacattgatgatctggagcatg agttt
20	barcode 10	tatggaatggttctgaacatgttactgaagctgcgctctattatatgcgtcttaacactggctggtgactgat gaagtgtcacacttg
21	barcode 11	agcttcaccaggctctgaggttctcttttgacgtattgtgatgcagactcatgatcgattttctgtgtacattaa
22	barcode 12	tgtcgaaattcctacggtgtgatgattctttgatatacatgtttctgatatgtattcggatagacatgctcttatcg tgcccccatctac
23	barcode 13	tggggaacagagtcctcatgtgccatttacagattctttgcatcatgtgttactcataagctctatctcaaaat
24	barcode 14	gctatcgttttgtatgaagttgcctcctgatgaagaaccacatgttgttaccacacaagaccatggttcctc agaaggatcttacagtt
25	barcode 15	atatggaagtgttgcatggaaagaccgtatggaagtatggaagaaacaacaataagaaaagctacaagtcg ttaa
26	barcode 16	gaatgaaagagtaaataatgatggtgatggcgatggattcattgttgacgagctcgatgatccgggctacaaa gttgaaaccttttactgcatt
27	barcode 17	taaagaattgtgtatatgtattttcttgacatttctttgacgtggttctggatgaaggaagacaacctctgtca
28	barcode 18	cgaaagctaccattcgttcgagcagaataaaccttataggatgttgattttctgtcggctcgccattcttggc ttgttcttcattatagga
29	barcode 19	catccggtgaatgatcgattcggcttgtggttaacatcggttatatgcgaaatatggttcgctgtgtcttgga
30	barcode 20	ttcttgatcagttcccaaatggtataccatcgaaagagaacatatctcgacaggctttcttaagggtatgag aaggaagggaagccatcagag
31	PolyloxP2 .0 w/o restriction sites	caacaggaatgaattcgttctcattaacgccgacgacactgccaaagatttctcaattttgttcttttgaaac aagaatgataacttcgtatagcatacattatacgaagttattatcttctgtttggatcgatctgggttggagattg aatctgcgttttctaaacagaagtccttatggaaattcacctgctgtcatattgcatgatttatacgtgtcggat atgatcaatgagacaaaagttagtgtagttcagactcatctccatgcggtttataacttcgtataatgtatgct atacgaagtattgtcagcttgttgagtatttaaattcgtttgtcatttgttggattccaagaagttgcatacaca gtattgttaattaatggaaagtactttatgaagttttgttaagaactagtatcaggagctttgttaaataatgatataata agattgatcttgtttgttactatataaacttcgtatagcatacattatacgaagttatgaatgaatgtggcaatgtgt acacgatcagcagaagaactgagtgaggcaaacctgtaaaatctgtagagatgagattgagttacagagtg agccatttatcgcttgcaacgaatgtgctttccctacgtgtagaccgtgctatgagtatgaaagaagagaagg

		aaaataacttcgtataatgtatgctatagcaagttatccaagcttgctcagtcggaacccgatacaagcgt attaaaagagacattttcttctgtattttatctcaacacttgcttgagtggtgatttatgttgtagtccaagg gtcgaaggggatgaagaagatgatgacattgatgatctggagcatgagttataacttcgtatagcatacatt atacgaagttattatggaatggtcctgaacatgttactgaagctgcgctctattatatgcgtcttaacactggtc gtggtactgatgaagtgtcacacttgagcttcaccaggctctgaggttcctctttgacgtattgtgatgcagac ttcattgatcgattttctgtgtacattaaataacttcgtataatgtatgctatagcaagttattgtcgaaattcctacg gtgtgatgattctttgatatacatgtttctgatattcgatagacatgctcttatcggtcccatctactgg ggaacagagtccatcatgtgccatttacagattctttgcatcatgtgttactcataagctctatctcaaaatata acttcgtatagcatacattatacgaagttatgctatcggtttgtatgaagttgcctcctgatgaagaaacccat gttgttaccacacaagaccatggttcctcagaaggatcttacagttatatggaagttgcatggaaagacc gtatggaagtatggaagaaacaacaatagaaaagctacaagtcgttaataacttcgtataatgtatgctat acgaagttatgaatgaaagagtaaatgatggtgatggcgatggattcattgttgacgagctcgatgatcccg ggctaccaaaagttgaaacctttactgcatttaagaattgtgtatatgtatttcttgacattctttgacgtggttc tggatgaaggaagacaacctctgtcaataacttcgtatagcatacattatacgaagttatcgaaagctacccat tcgttcgagcagaataaaccttataggatgttgattttctgtcggctcgccattcttggttgttcttcattata ggacatccggtgaatgatcgattcggctgtgttaacatcggttatatgcgaatatggttcgctgtgtcttgg aataacttcgtataatgtatgctatagcaagttatttctgatcagttcccaaatggtataccatcgaaagagaa acatatctcgacaggctttcttaaggtatgagaagggaagggaagccatcagag
32	PolyloxP2 .0 with restriction sites	gtgcaggataccaacaggaatgaattcgttctattaacgccgacgacactgccaaagatttctcaattttgtt cttttgagaacaagaatgataacttcgtatagcatacattatacgaagtattatcttctgtttggatcgatctgg gtttgagattgaatctgcgttttctaaaacagaagctttatggaaatttcacctgctgctgcaggatacata ttgcatgatttatacgtgtcggatgatcaatgagacaaaagttagtgtagttcagactcatctccatgcgg ttataacttcgtataatgtatgctatagcaagttatgtcagctgttgagtatttaaatcgtttgtcatttgttgg attccaagaagttgcatacacagtattgttaattaatggaaagttgtcaggatacactttatgaagttttgttaag aactagtatcaggagctttgttaaatgatataataagattgatctgtttgttactatataacttcgtatagcatac attatacgaagttatgaatgaatgtggcaatgtgtacacgatcagcagaagaactgagtgggcaaaactgtga aaatctgtagagatgagattgagttaacagagtgaggatagcgtgagccattatcgcttgcaacgaatgtgc ttccctacgtgtagaccgtgctatgagtatgaaagaagagaaggaaaataacttcgtataatgtatgctatag gaagttatccaagcttgctcagtcggaacccgatacaagcgtattaaaagagacattttcttctgtatttt tatctcaacacttgcttgagggtgcaggatactggattgatttatgttgtagtccaagggcgaaggggatgaa gaagatgatgacattgatgatctggagcatgagttataacttcgtatagcatacattatacgaagtattatgg aatggttctgaacatgttactgaagctgcgctctattatatgcgtcttaacactggctggtgactgatgaagt gtcacacttggtgcaggatacagcttcaccaggctctgaggttcctctttgacgtattgtgatgcagactcat gatcgattttctgtgtacattaaataacttcgtataatgtatgctatagcaagttattgtcgaaattcctacgggt gatgattctttgatatacatgtttctgatattcgatagacatgctcttatcggtcccatctacgtgcag gatactggggaacagagtcattcatgtgccatttacagattctttgcatcatgtgttactcataagctctatctc aaaataaacttcgtatagcatacattatacgaagttatgctatcggtttgtatgaagttgcctcctgatgaagaa acccatggtgttaccacacaagaccatggttcctcagaaggatcttacagttgtgcaggatacatatgga agtgttgcatggaaagaccgtatggaagtatggaagaaacaacaatagaaaagctacaagtcgttaata acttcgtataatgtatgctatagcaagttatgaatgaaagagtaaatgatggtgatggcgatggattcattgtg acgagctcgatgatccgggtaccaaagttgaaacctttactgcattgtgcaggatactaaagaattgtgt atatgtattttcttgacattctttgacgtggttctggatgaaggaagacaacctctgtcaataacttcgtatagca tacattatacgaagttatcgaaagctacccattcggttcgagcagaataaaccttataggatgttgattttctgtc ggctcgccattcttggttcttccattataggagtgaggataccatccggtgaatgatcgattcggcttgt

		ggtaacatcgggtatatgcgaaatatggtcgtgtgtcttgaataacttcgtataatgtatgctatacgaagt tatttcttgatcagttcccaaatggtataccatcgaaagagaaacatatctcgacaggctttcttaagggtatg agaaggaagggaagccatcagaggtgcaggatac
33	PolyloxP2 .0 with flanking restriction sites	ggatccgtgcaggataccaacaggaatgaattcgttctcattaacgccgacgacactgccaaagatttctca atattgtcttttgagaaacaagaatgataacttcgtatagcatacattatacgaagtattatcttctgtttggatc gatctgggtttggagattgaatctgcgttttctaaaacagaagctttatggaaattcacctgctgctgcag gatacatattgcatgatttatacgtgtcggatatgatcaatgagacaaaagtagtgcagtttcagactcatctc catgcggtttataacttcgtataatgtatgctatacgaagtattgtcagcttggtgagtatttaaattcgtttgtcatt tgttgggtttccaagaagttgcatacacagattgttaattaatggaaagtggtgcaggatacacattatgaagtt ttgttaagaactagtatcaggagctttgttaaatgatataataagattgatctgtttgttactatataacttcgtat agcatacattatacgaagttatgaatgaatgtggcaatgtgtacacgatcagcagaagaactgagtgggcaa acttgtaaaatctgtagagatgagattgagttaacagagtgaggatacgtgagccatttatcgttgcacg aatgtgctttccctacgtgtagaccgtgctatgagtatgaaagaagagaaggaaataacttcgtataatgtat gctatacgaagttatccaagcttgctcctcagtcggaacccgatacaagcgtattaaaagagacattttcttct ctgtattttatctcaacacttgctttgaggtgcaggatactggattgattatgtttagtccaagggtcgaagg ggatgaagaagatgatgacattgatgatctggagcatgagttataacttcgtatagcatacattatacgaagt tattatggaatggttctgaacatgttactgaagctgcgctctattatagcgtcttaacactggctggtgactg atgaagtgtcacacttggtgcaggatacagcttcaccaggctctgaggttcctcttttgacgtattgtatgca gacttcattgatcgattttctgtgtacattaaataacttcgtataatgtatgctatacgaagtattgtcgaaattcct acggtgtgatgattctttgatatacatgtttctgatatgtattcggatagacatgctcttatcgtgcccccatcta cgtgcaggatactggggaacagagtcacatgtgccatttacagattctttgcatcatgtgttactcataagc tctatctcaaaatataacttcgtatagcatacattatacgaagttatgctatcgttttgtatgaagttgcctcctgat gaagaaaccacatgttgttaccacacaagaccatggttcctcagaaggatcttacagttgtgcaggatac atatggagtggtgcattgaaagaccgtatggagatggagaaacaacaataagaaaagctacaagtcg ttaataacttcgtataatgtatgctatacgaagttatgaatgaaagagtaaatgatgggtgatggcgatggattc attgttgacgagctcgtatgatccgggctaccaaagttgaaaccttttactgcattgtgcaggatactaaaga attgtgtatatgtatttcttgacatttcttgacgtggttctggatgaagggaagacaacctctgtcaataacttcgt atagcatacattatacgaagttatcgaaagctaccattcgttcgagcagaataaaccttataggatgttgatt ttctgtcggtcgccattcttggttcttccattataggagtgaggataccatccggtgaatgatcgattc ggcttgggttaacatcggttatatgcgaaatatggttcgctgtgtcttggaaataacttcgtataatgtatgctata cgaagttatttcttgatcagttcccaaatggtataccatcgaaagagaaacatatctcgacaggctttctctaa ggatgagaaggaagggaagccatcagaggtgcaggatacggcgcgcc
34	loxP (generic)	ataacttcgtatannntannntatacgaagttat
35	loxP	ataacttcgtatagcatacattatacgaagttat
36	Forward Primer #493	gcaagcacgtttccgacttgag
37	Reverse Primer #2427	cataccttagagaaagcctgtcgag
38	Forward Primer #2456	ccatgggagatccacgaaatg

39	Reverse Primer #2457	cctggtatctttatagtcctg
40	Forward Primer #2450	tgtggtatggctgattatgatcag
41	Reverse Primer #494	agctacagcctcgatttggtg
42	Forward Primer #2426	cgacgacactgccaagatttc
43	Forward Primer HL16	cctaaagaagaggctgtgcttgg
44	Reverse Primer #HL15	aagaccggaagagttgtcc
45	Oligo2424	gcaaaggcgcccgatagaataa
46	Oligo2425	ccgggggaaagaagggtcac
47	PolyloxP2 .0_Rosa26 targeting vector	cacctgacgcgcctgtagcggcgccattaagcgcggcggtgtggtgggttacgcgcagcgtgaccgcta cacttgccagcgccctagcgcgcctcttctgcttcttcccttcttctcgcacgttcgcccgttcccc gtcaagctctaaatcgggggctcccttaggggtccgatttagtgctttacggcacctcgacccccaaaaactt gattaggggtgatgggtcacgtagtgggccatcgccctgatagacgggttttcgcccttgacgttgagtgcca cggtctttaatagtgactctgttccaaactggaacaactcaacctatctcggtctattcttttgattataag ggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaatttaacaaaat attaacgcttacaatttgccattcgccattcaggctgcgcaactgttgggaaggcgcatcggtgcgggcctct tcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggttttc ccagtcacgacgttgtaaacgacggccagtgaattgtaatacgactcactatagggcgaaattgggtaccg ggccccccctcgagccccagctgggtcttccgcctcagaagccatagagcccaccgcacccccagcatg cctgctattgtcttcccaatcctcccccttgcgtcctgccccacccccacccccagaatagaatgacaccta ctcagacaatgcatgcaatttcctcattttattaggaaggacagtgggagtggtcaccttcagggtcaag gaaggcacgggggagggggcaacaacagatggctggcaactagaaggcacagtcgagggtgatcagc gagctctaggatctgcattccaccactgctccattcatcagttccataggttggaatctaaatacacaaaca attagaatcagtagtttaacacattatacacttaaaaattttatattaccttagagctttaaatctctgtaggtagt tgtccaattatgtcacaccacagaagtaagggttccttcacaagagatcgccctgacacgatttctgcacaggc ttgagccatatactcatacatcgcatcttgccacgttttcacgggtttcaaaattaatctcaagttttacgctta acgctttcgctgttcccagttattaatatattcgacgctagaactcccctcggaagggaaggctgagcac tacacgcgaagcaccatccccgaaccttttgataaactcttctgttcgactgtccatcaacgggtcagtga gacttaaacctaactctttcttaatagtttcggcattatccacttttagtgcgaggaccttcgtcagtcctggatac gtcactttgaccacgtctccagcttttcagagagcgggttttcgttatctacagagtatcccgacgcgtcgtat ttattgtcggtactataaaacctttccaatcatcgatcataatttctgtgtaccagattttggctttgtataccttt ttgaatggaatctacataaccaggttagtcccgtgtgtacgaagaaaagtttccatcacaaaagattagaag aatcaacaacatcatcaggatccatggcgaggacctgcagggtcgaaaggcccgagatgaggaaggagg

	agaacagcgcggcagacgtgcgcttttgaagcgtgcagaatgccgggctccggaggaccttggggcgc ccgccccgcccctgagcccgcccctgagcccgcccccgaccaccccttcccagcctctgagcccaga aagcgaaggagcaaagctgctattggccgctgccccaaaggcctaccgcttcattgtcagcggtgct gtccatctgcacgagactagtgcagcgtgctacttccatttgcacgtcctcagcgcgagctgcggggcg ggggggaaacttctgactaggggaggagwagaaggtggcggaagggggccaccaagaacggagcc ggttggcgccctaccggtggatgtggaatgtgtgcgaggccaraggccacttgtgtagcgccaagtgccag cggggctgctaagcgcagctcagactgccttggaaaagcgctccctaccggtagaattcgatatcaa gcttatcgatgtcgacggtttcgataagctagntgccctgaaatatcttatcagaacaaggtagtataaagct ggtaggtatacaaaaacgctagactagtcttatccctgaccttaatctgctagtataccgtaggaagtgtctt aagtgccactagtaccaacagcctctgttagagtgaacagtagatttaataagggccaataaagaccactt aagactactgactgaattaaaaatccaccagtgaggatgtattggttatagttatatacaaaaatatccaact tagccaggcggtggtggtgcagcctttaatccggcactcaggaggcagaggcaggcggatttctgagttc gaggccagcctggtctacaaagtgagttccaggacagccaggctatacagaaaacccgtctcgaaaaa ccaaaaaaaaaaaaaaaaaggaaaaaataatggcttgaaaacatgctattgttgtaaaaaggagatccaa acaagtatttcttgaacaaaaattatgtagaaaacagggtgataaagaacaagggcttttaaaaaaggcca gcattatttgagatatacaagtactccaattcaagcctaaaggccactcaatgctcactaacagtgtcaatca gaatctcttctggataaacatgtcctccaacatataaactagtatacactaacaaaacgtctcaactcaagggt gaaatgcttgactcctagactgtgaccagcaaagtgtcttataggtagggttactaggtcaragagtcttg cctgcaaacacaaattatacaattccaccaaatgtaagacatgtaaaaatgactatttagctcaggttttga attaaaaatcaaacccacaaagacctaactttcacattaagtacaaatgtttaacatatataacatgtttgcctt gatatatgaaatcatatgacatcacctgacaggaaacttaagttatttgaacaacaatcagcctaaggtaggc ctagcacatgatcaggatagtgagggaaacccaaagaagtgtctctgagataaaccaagagacctga aatagcagctttgttctgtatctcatgggctttaactatcacttgaaaacaatttcacaaatcacaaaaccaat ttgtttagtttcaactctgcctattcaccggagaaatccatgtcccaatttgatgggggaaaacttgaatgaa tattatttggaaatttttaatatccctgtgtatagcagatgggttaagacaggtaaccttaactgtagtcttgggt tgtctgccttgtttttgaaatatgctctcaccaggagcctgccaagtaactactctgtgtgctcagcaagtct agggatcccttaagcatgctctaacaggcctggcttttttttaaaaagtaggttgcggtgataataattcaagt ttcatcacttgaaacacattttaccaactatcacccaagctcccttagtgtgtcccctataaaagaatctgacct gcaagttccaaaagtattactacattttaaccttagtattcaggagacaaaagcaggcagaactttagtctaca tatcttgttgagatcaggcagagatacatagtgcacctgcctcaggaaaagctatcattttatattcaca gtaagcaaattaacattaaaaagtcagaaattcttaattgactttggctgtgaagaatttggattctactggattg aaagctnagaaaatctgattaagcgtgcaacagtttctgtatactaaaactctaagacccttgggtctaaaga taccacatttaaaaaaaagttaagtataagacaaaaagccttaatagctcttttataggggagggactcattt aatattagtcacctcactcctataactattttatgaggtttattatataaccatttttaaaaaaaaattaaag gacaatttttagtgtttgaaagatttcccaacccaccctggaaatcaggctgcaaatctcagcagctgccctta gaaggcaaaaagccatgccatttaagccatgggaagttagtagcaacaagagaccacatggcaggaag gaggctttaagaaaagcccacagtgtcaaaagaccaaccaacagcagagacagaaaaataataacca aatctcatcagaaaggtagacggatttagccacatccatagtggctcattagggaatgctcaattataatttta tgaaatctcagagccccctcccaacagggtttctgtgtgtattcctggctatcctagatagaacttgatgtgt agaccaggctgggtctaaactcagagatctgcctgtctctgcctccagagtgtgggattaaaggcataaa ggcaagcaccaccactggctggctaaactctggcctacattttgaaatccaaaattatcattactgtcttca tgtttagtaatggctgcagacttagctttcagctttgtatagatgaagcacaaaagttaagtaaaagctagcagtt taaacaaaataggatactagaagtcaattttatttcttattgttagtgaagagaaaagacattaaagtctgca ttggtcttctgtatcccacaagtctgcagttatggctcctctgtccacagttacacttcaataaaaaattagttcct
--	---

	tggtaacatttccagtgccttaaagagtaatttgattattggctaccatattggaacaaacacaaagtatttcatta cttcagtaaattctactggagacagtattctcttataaaaaaaaaaaaaagaagaaggaggattttatttt aatgtgaatrcacttgttggtcttcagacacaccagaagagggcatcagatcccattacagatgggtgtgagc caccatgtgggtgctgggaattgaactcaggacttctggaagagcagctgtgctcttgaactccgagccac ctctccagcccctggatagattattatagagcacaagcacacacaacacaacacctaanaaaccagcttc acaatacatttgttgcttaattctttacttgtgtttggctacaagtcaagcaaaattataggctcctgaagaagct tggcaaaatcacatttagaccagcaataacgtgtagaatgccatgagtcaagccagtcgaagagaagcac tgcttaaatagaaaaatacaagctcacaagaccttaggtcaggaaagacaacaaacacctgaactttgcat tccaaaaggaaccaccttttacagatgtgtacaatgacaaaatttagtaagcagtaatacatcatgtggct caataatgaaataccttttaattgtcttaaaatatcagcctggcaatatgtaagatacatcaggtaataattggg ggaggagacatccacctggaaaccattaatggftaatagctcagttataaatggagaaaaaggagagagg cattcatgggagtggaagtaagcttttaggttgattctcaatacatctattgttgtaaaactccctggactga gaataggcccaaatgtggaacaccacctgacgggagaggtgatagacactgaaattaaggatcaaggc aaaggatcaacaaaaagtgtactaaggagttataaaagaactgcagtgttgaggcccagctacagcctcg atttgtggtgtatgtaactaatctgtctggtttcatgagtcacagacttctaagatcaggaaagggaaaatgcc aatgctctgtctaggggttgataagccagtaataaatgaaatggggctaaaatgagtgtctaaaatac cttttgataaggctgcagaaggagcgggagaaatggatatgaagtactgggctcttataaaatgattaaaatt ctgcttacatagtctaactcgcgacactgtaatttcatactgtagtaaggatctcaagcaggagagtataaac tcgggtgagcatgtcttaattctacctcgtatggaaaatactccgaggcggatcacaagcaataaacctgta gttttgctgcataaaaacccagatgactacctatcctccattttcttatttgcccctattaaaaaacttcccgc aaaaccgaaaatctgtgggaagtcttgcctccaattttacacctgttcaattcccctgcaggacaacgccc acacaccaggttagcctttaagcctgcccagaagactcccggccatgcttctagtggcgcgccgtatcctgc acctctgatggcttcccttcttctacaccttagagaaagcctgtcgagatatgtttctcttctgatggatataca ttttgggaactgatcaagaaataacttcgtatagcatacattatacgaagtattccaagacacagcgaacat atttcgcatataaccgatgttaaccacaagccgaatcgatcattcaccggatggatctctgcactcctataatg gaagaacaagccaagaatggcgagccgacagaaaatcaacatcctataagggttattctgtctgaacgaa tgggtagctttcgataacttcgtataatgtatgtatatacgaagtattgacagaggtgtcttcttcatccagaa ccacgtcaaagaaatgtcaagaaaatacatatacacaattcttttagtatcctgcacaatgcagtaaaaggtttc aactttggtagcccgggcatcagctcgtcaacaatgaatccatgccatcaccatcatttactctttcattc ataacttcgtatagcatacattatacgaagtatttaacgactgtagcttttctatttgtgtttcttcatactcca tacggctttccatgcaacacttccatattgtatcctgcacaactgtaagatcctctgaggaacctatgggtcttg tgtgtaacaacatgtgggtttcttcatcaggaggcaacttcatacaaacgatagcataacttcgtataatgt atgtatacgaagttatatttgagatagagcttatgagtaacacatgatgcaaaagaatctgtaaatggcaca tgatggactctgttcccagtatcctgcacgtatagggggcacgataagagcatgtctatccgaatacatat cagaaacatgtatatcaaaagaatcatcacaccgtaggaatttcgacaataacttcgtatagcatacattatac gaagtatttaatgtacacagaaaatcgatcatgaagtctgcatcacaatcgtcaaaagaggaacctcaga gcctggtagaagctgtatcctgcaccaagtgtgacacttcacagaccagcagtgtaagacgcataata atagagcgcagcttcagtaacatgttcaggaaccattccataataacttcgtataatgtatgtatatacgaagtta taaactcatgtccagatcatcaatgtcatcatcttcttcatcccttcgaccttggactacaacataaatcaat ccagtatcctgcacctcaaaagcaagtgtgagataaaaaatacagagaagaaaatgtctcttttaatacgttgt atcgggttccgactgaggacaagcttgataacttcgtatagcatacattatacgaagtattttcttctcttc tttcactcatagcacggctacacgtagggaagcacattcggtgcaagcgataaatggctcacgtatcct gcactctgttaactcaatctcatctacagattttcaagtttcccactcagttcttctgtgatcgtgtacaca ttgccacattcattcataacttcgtataatgtatgtatatacgaagttatagtaacaaaacaagatcaatcttatt
--	--

	atatcatttaacaaagctcctgatactagttcttaacaaaacttcataaagtgtatcctgcacactttccattaatt aacaatactgtgtatgcaacttcttggatccacaacaaatgacaaacgaatttaaatactcaacaagctgac ataacttcgtatagcatatactatacgaagtataaaccgcatggagatgagtctgaaactgacactaactttg tctcattgatcatatccgacacgtataaatcatgcaatatgtatcctgcacgacagcaggtgaaattccataaa gacttctgttttagaaaacgcagattcaatctcaaaccagatcgatccaaacagaagataataacttcgtat aatgtatgctatacgaagttatcattcttgtttctcaaaaagaacaaaattgagaaaatttggcagtgctgcggc gttaatgagaacgaattcattcctgttggtatcctgcacggatccaccggatctagataactgatcataatcag ccataccacattttagagggtttacttgccttaaaaaacctcccacacctccccctgaacctgaacataaaat gaatgcaattgttgttgaactgtttattgcagcttataatgggtacaaataaagcaatagcatcacaatttca caaataaagcattttttcactgcattctagtgtggttgcctcaaacatcaatgtatcttaacgcgtgaagtc ctatactttctagagaataggaactcatgcatcgagccccagctgggtcttccgcctcagaagccatagag cccaccgcacccccagcatgctgtattgtcttcccaatcctcccccttgcgtgctgccccaccccc ccagaatagaatgacacctactcagacaatgcatgcaatttcctcattttattaggaaaggacagtgggagt ggacacttcagggtcaaggaaggcacgggggaggggcaaacaacagatggctggcaactagaaggc acagtcgaggctgatcagcgagctctagaggatcgagccccagctgggtcttccgcctcagaagccatag agccccgcacccccagcatgctgtattgtcttcccaatcctcccccttgcgtgctgccccaccccc cccagaatagaatgacacctactcagacaatgcatgcaatttcctcattttattaggaaaggacagtggga gtggcaccttcagggtcaaggaaggcacgggggaggggcaaacaacagatggctggcaactagaagg gcacagtcgaggctgatcagcgagctctagcgctcgagatccgaacaaacgaccaacaccctgctgttt attctgtcttttattgccgatccccctcagaagaactcgtaagaaggcgatagaaggcgatgcgctgcgaat cgggagcggcgataccgtaagcacgaggaagcggctcagccccattcgccccaagctcttcagcaatat cacgggtagccaacgctatgtcctgatagcggctcgccacacccagccggccacagtcgatgaatccaga aaagcggccattttccaccatgatattcggaagcaggcatcgccatgggtcacgacgagatcctcgccgt cgggcatgcgcgcttgagcctggcgaacagttcggtggcgagccccctgatgctcttcgtccagatc atcctgatcgacaagaccggcttccatccgagtagctgctcgtcgtgatgcatggttcgcttggtggtcgaat gggcaggtagccggatcaagcgtatgcagccgcccattgcatcagccatgatggatacttttcggcag gagcaaggtgagatgacaggagatcctgccccggcacttcgccaatagcagccagtccttcccgttc agtgacaacgtcgagcacagctgcgcaaggaacgcccgtcggtggccagccacgatagccgcgtgcct cgtcctgcagttcattcagggcaccggacaggtcggtcttgacaaaaagaaccggcgccccctgcgtga cagccggaacacggcgccatcagagcagccgattgtctgttgcccagtcatagccgaatagccttcca ccaagcggccggagaacctgcgtgcaatccatcttgttcaatggccgatcccatggcgccacagatgaat tcgaagttcctatactttctagagaataggaacttcaccggtaagcttatcgaatccgctgacccactggaa agaccggaagagttgtcctcaaccgcgagctgtggaaaaaaaggacaggataagtatgacatc aaggaaaccttgactactgcgcctacagatctgcagccgggttaattaactagaaagactggagttgc agatcacgaggggaagagggggaagggttctccaggcccaggcggtcctcagaagccaggaggca gcagagaactcccagaaggtattgcaacactccccctccccctccggagaagggtgcggccttctcccc gcctactccactgcagctcccttactgataacactcagagcgactttgggagagcaagtgttctgctcc aaaacagcccaactgagccctcgtccttccctccactccccggagtgcgcgatggaggtctggctcagc acgcccccttgaggcaactcaagtcggaaacgtgcttgacccgccccgcagccgctcagccctactgc ccgtccccgccccagcgcgcgttctgccacgttgcgcaggggagcgcggccagactctgcggcgc ggggcccaggggagggccggaacctgggagcgcctcctcgccgccccgctggccggcgatggact caactgcacgaacacgagccaatggcaaggggcagtttctgggccccgagagccaatcagacgacga ggccccggcgggcggtaaaacgactccccagagggaaggggaggggtggcgggcgctcgcgcg cgagctacttctgctgacctccctccccctccccgccccgagaggccgaccgcgccccgcagctccag
--	--

		ctgcctcaccacacctacctcccgccccaccagtgggcagagcgaggctgccggcggtgcgcactc cggctgccgttaactgacaggcgcttacgccaacaaaacacgccattgtgtttcacacacggcggga ggaaaagaagccaatcagcgacgagacgtcggcggaagcgctcctccgctgcccccccccccgag ccatggccgctccggtggagactttccgctcccttctccctccccctccggttgctgcaggcggaaccgc attcctgcccaccaccgcttgccccttcagcgctcagactcgtaccggctgtctacagaacgggtcca ccacgctcggagggcctgccgcggtggagctccagcttttgtccctttagttaggggttaatttcgagcttgg cgtaatcatggctatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacagagccgga agcataaagtgtaaagcctgggggtgcctaatagtgagctaaactacattaattgcgttgcgctcactgccc gctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcgccaacgcgcggggagaggcggtt gcgtattggcgctcttccgcttctcgtcactgactcgtgcgctcggtcgttcggctgcggcgagcggt atcagctcactcaaaggcggttaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtga gcaaaaggccagcaaaaggccaggaaccgtaaaaggccggttgctggcggttttccataggtccgcc cccctgacgagcatcacaataacgacgctcaagtcagaggtggcgaaaccgacaggactataaagat accaggcggttccccctggaagctccctcgtgcgctcctgttccgacctgccgcttaccggatacctgtc cgcttttctccctcggaagcgtggcgcttttctcatagctcacgctgtaggtatctcagttcggttaggtcg ttcgtccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccgtaactatc gtcttgagtccaaccggtaagacacgacttatccactggcagcagccactggtaacaggattagcaga gcgaggtatgtaggcggtgtacagagttctgaagtgggtggcctaactacggctacactagaagaacagt atttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaacaa accaccgctggtagcggtgggtttttgttgaagcagcagattacgcgcagaaaaaaaggatctcaagaa gatcctttgatctttctacggggtctgacgctcagtggaacgaaaactcacgttaagggttttggtcatgag attatcaaaaaggatcttcacctagatccttttaataaaaaatgaagtttaaatcaatctaaagtatatatgagt aaacttggtctgacagttaccaatgcttaatacgtgagggacctatctcagcgatctgtctatttcgttccat agttgcctgactccccgctgtagataactacgatacgggaggggttaccatctggccccagtgctgcaat gataccgcgagaccacgctcaccggctccagatttatcagcaataaaccagccagccggaaggggccga gcgcagaagtggctctgcaactttatccgctccatccagctctattaattgttgcgggaagctagagtaagt agttgccagttaatagtttgcgcaacgttggtgccattgctacaggcatcgtgggtgcagctcgtcgttgggt atggcttcattcagctccggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcg gttagctccttcggtcctccgatcgttgtagaagtaagttggccgagtggtatcactcatggttatggcagc actgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattct gagaatagtgtatgcggcgaccgagttgctcttgcggcgctcaatacgggataataccgcgccacatagc agaactttaaaagtgtcatcattggaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttga gatccagttcgtatgaaccactcgtgcaccaactgatcttcagcatctttactttcaccagcgtttctgggt gagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataaggcgacacggaaatgttgaatactc atacttctcttttcaatattatgaagcatttatcagggtattgtctcatgagcggatacatattgaatgtattta gaaaaataacaaataggggtccgcgcacatttccccgaaaagtgc
48	DNA Segment A	TATCTTCTGTTTGGATCGATCTGGGTTTGGAGATTGAATCTGCGT TTTCTAAAACAGAAGTCTTTATGGAAATTTACCTGCTGTCGTGC AGGATACATATTGCATGATTTATACGTGTCGGATATGATCAATG AGACAAAAGTTAGTGTCAGTTTCAGACTCATCTCCATGCGGTTT
49	DNA Segment B	GTCAGCTTGTTGAGTATTTAAATTCGTTTGTCAATTTGTTGTGGATT CCAAGAAGTTGCATACACAGTATTGTTAATTAATGGAAAGTGTG CAGGATACACTTTATGAAGTTTTGTTAAGAAcTAGTATCAGGAGC TTTGTTAAATGATATAATAAGATTGATCTTGTGTTTGTACTAT

50	DNA Segment C	GAATGAATGTGGCAATGTGTACACGATCAGCAGAAGAACTGAGT GGGCAAACCTTGTAATAATCTGTAGAGATGAGATTGAGTTAACAGA GTGCAGGATACGTGAGCCATTTATCGCTTGCAACGAATGTGCTTT CCCTACGTGTAGACCGTGCTATGAGTATGAAAGAAGAGAAGGA AA
51	DNA Segment D	CCAAGCTTGTCTCAGTGCAGGAACCCGATACAAGCGTATTAAAA GAGACATTTTCTTCTCTGTATTTTTATCTCAACACTTGCTTTGAGG TGCAGGATACTGGATTGATTTATGTTGTAGTCCAAGGGTCAAG GGGATGAAGAAGATGATGACATTGATGATCTGGAGCATGAGTTT
52	DNA Segment E	TATGGAATGGtTCCTGAACATGTTACTGAAGCTGCGCTCTATTAT ATGCGTCTTAACACTGGTCGTGGTACTGATGAAGTGTCACACTTG GTGCAGGATACAGCTTCACCAGGCTCTGAGGTTCTCTTTTGACG TATTGTGATGCAGACTTCATGATCGATTTTCTGTGTACATTAA
53	DNA Segment F	TGTCGAAATTCCTACGGTGTGATGATTCTTTTGATATACATGTTT CTGATATGTATTCGGATAGACATGCTCTTATCGTGCCCCCATCTA CGTGCAGGATACTGGGGAACAGAGTCCATCATGTGCCATTTACA GATTCTTTTGCATCATGTGTTACTCATAAGCTCTATCTCAAAT
54	DNA Segment G	GCTATCGTTTTGTATGAAGTTGCCTCCTGATGAAGAAACCCACAT GTTGTTACCACACAAGACCCATGGTTCCTCAGAAGGATCTTACA GTTGTGCAGGATACATATGGAAGTGTGTCATGGAAAGACCGTAT GGAAGTATGGAAGAAACAACAATAGAAAAGCTACAAGTCGTT AA
55	DNA Segment H	GAATGAAAGAGTAAATGATGGTGATGGCGATGGATTCAATTGTTG ACGAGCTCGATGATCCCGGGCTACCAAAGTTGAAACCTTTTACT GCATTGTGCAGGATACTAAAGAATTGTGTATATGTATTTTCTTGA CATTTCTTTGACGTGGTTCTGGATGAAGGAAGACAACCTCTGTCA
56	DNA Segment I	CGAAAGCTACCCATTTCGTTTCGAGCAGAATAAACCCCTTATAGGAT GTTGATTTTCTGTCTGGCTCGCCATTCTTGGCTTGTTCTTCCATTAT AGGAGTGCAGGATACCATCCGGTGAATGATcgATTTCGGCTTGTTGG TTAACATCGGTTATATGCGAAATATGGTTCGCTGTGTCTTGGA
57	DNA Segment a	AAACCGCATGGAGATGAGTCTGAAACTGACACTAACTTTTGTCT CATTGATCATATCCGACACGTATAAATCATGCAATATGTATCCTG CACGACAGCAGGTGAAATTTCCATAAAGACTTCTGTTTTAGAAA ACGCAGATTCAATCTCCAAACCCAGATCGATCCAAACAGAAGAT A
58	DNA Segment b	ATAGTAACAAAACAAGATCAATCTTATTATATCATTTAAACAAAG CTCCTGATACTAgTCTTAACAAAACCTTCATAAAGTGTATCCTGCA CACTTTCCATTAAATTAACAATACTGTGTATGCAACTTCTTGGAAT CCACAACAATGACAAACGAATTTAAATACTCAACAAGCTGAC
59	DNA Segment c	TTTCCTTCTCTTCTTTCATACTCATAGCACGGTCTACACGTAGGG AAAGCACATTTCGTTGCAAGCGATAAATGGCTCACGTATCCTGCA CTCTGTAACTCAATCTCATCTCTACAGATTTTACAAGTTTGCCC ACTCAGTTCTTCTGCTGATCGTGTACACATTGCCACATTCATTC
60	DNA	AAACTCATGCTCCAGATCATCAATGTCATCATCTTCTTCATCCCC

	Segment d	TTCGACCCTTGGACTACAACATAAATCAATCCAGTATCCTGCACC TCAAAGCAAGTGTTGAGATAAAAATACAGAGAAGAAAATGTCT CTTTAATACGCTTGTATCGGGTTCGCACTGAGGACAAGCTTGG
61	DNA Segment e	TTAATGTACACAGAAAATCGATCATGAAGTCTGCATCACAATAC GTCAAAGAGGAACCTCAGAGCCTGGTGAAGCTGTATCCTGCAC CAAGTGTGACACTTCATCAGTACCACGACCAGTGTTAAGACGCA TATAATAGAGCGCAGCTTCAGTAACATGTTTCAGGAaCCATTCCAT A
62	DNA Segment f	ATTTTGAGATAGAGCTTATGAGTAACACATGATGCAAAAGAATC TGTAATGGCACATGATGGACTCTGTTCCCCAGTATCCTGCACGT AGATGGGGGCACGATAAGAGCATGTCTATCCGAATACATATCAG AAACATGTATATCAAAAGAATCATCACACCGTAGGAATTTTCGAC A
63	DNA Segment g	TTAACGACTTGTAGCTTTTCTATTTGTTGTTTCTTCCATACTTCCA TACGGTCTTTCCATGCAACACTTCCATATGTATCCTGCACAACCTG TAAGATCCTTCTGAGGAACCATGGGTCTTGTGTGGTAACAACAT GTGGGTTTCTTCATCAGGAGGCAACTTCATACAAAACGATAGC
64	DNA Segment h	TGACAGAGGTTGTCTTTCCTTCATCCAGAACCACGTCAAAGAAAT GTCAAGAAAATACATATACACAATTCTTTAGTATCCTGCACAAT GCAGTAAAAGGTTTCAACTTTGGTAGCCCGGGATCATCGAGCTC GTCAACAATGAATCCATCGCCATCACCATCATTTACTCTTTCATT C
65	DNA Segment i	TCCAAGACACAGCGAACCATATTTTCGCATATAACCGATGTTAAC CACAAGCCGAATcgATCATTACCGGATGGTATCCTGCACTCCTA TAATGGAAGAACAAGCCAAGAATGGCGAGCCGACAGAAAATCA ACATCCTATAAGGGTTTATTCTGCTCGAACGAATGGGTAGCTTTC G

Table 3: Frequency of individual barcodes identified by FRC after in vivo recombination

Full-length read codes	Events (n)	Full-length read codes	Events (n)	Full-length read codes	Events (n)
5'-I-3'	12922	5'-a-d-G-f-e-H-I-3'	20	5'-A-D-C-B-E-F-G-H-I-3'	6
5'-A-3'	5206	5'-c-b-a-D-G-H-I-3'	20	5'-A-b-E-F-i-h-g-3'	6
5'-G-H-I-3'	3480	5'-i-D-A-3'	20	5'-A-f-G-3'	6
5'-A-B-C-D-E-F-G-H-I-3'	2982	5'-i-H-g-3'	20	5'-A-f-e-d-c-B-I-h-g-3'	6
5'-A-B-I-3'	2286	5'-i-h-g-f-E-d-c-b-a-3'	20	5'-A-f-i-3'	6
5'-E-F-G-H-I-3'	1284	5'-A-B-e-d-G-3'	19	5'-A-h-c-d-I-3'	6
5'-A-B-G-H-I-3'	1213	5'-A-D-e-F-G-H-I-3'	19	5'-C-D-E-F-i-h-g-3'	6
5'-A-H-I-3'	1210	5'-A-d-E-F-I-h-g-3'	19	5'-C-D-G-H-I-b-E-F-a-3'	6
5'-A-B-C-3'	1085	5'-A-d-g-B-I-3'	19	5'-C-D-G-b-a-H-I-3'	6
5'-C-3'	1052	5'-A-f-e-b-G-H-I-3'	19	5'-C-D-i-h-e-3'	6
5'-c-3'	988	5'-A-h-g-f-e-d-c-3'	19	5'-I-d-c-b-a-h-g-f-e-3'	6
5'-i-3'	823	5'-C-D-E-h-g-B-a-3'	19	5'-a-D-G-f-e-H-I-3'	6
5'-a-3'	722	5'-G-H-C-D-a-f-I-3'	19	5'-c-D-E-3'	6

5'-E-3'	689	5'-G-H-I-d-c-b-a-f-e-3'	19	5'-c-b-a-D-E-3'	6
5'-G-3'	669	5'-G-f-e-d-I-3'	19	5'-c-b-a-d-E-F-G-H-I-3'	6
5'-A-F-G-H-I-3'	653	5'-e-d-c-h-I-3'	19	5'-c-b-a-d-E-F-i-h-g-3'	6
5'-A-B-C-D-I-3'	648	5'-A-B-C-f-I-3'	18	5'-e-d-A-B-C-H-I-3'	6
5'-C-D-E-F-G-H-I-3'	616	5'-A-B-E-h-g-f-I-3'	18	5'-g-D-I-3'	6
5'-g-3'	582	5'-A-B-e-d-c-F-I-3'	18	5'-i-B-a-h-g-3'	6
5'-A-D-E-F-G-H-I-3'	514	5'-A-B-i-f-c-3'	18	5'-i-b-A-3'	6
5'-A-B-C-D-G-H-I-3'	478	5'-C-D-a-3'	18	5'-i-f-G-3'	6
5'-e-3'	389	5'-E-h-i-3'	18	5'-i-h-A-B-G-3'	6
5'-A-D-G-H-I-3'	340	5'-I-d-c-3'	18	5'-i-h-g-f-A-B-E-3'	6
5'-A-B-C-D-E-F-I-3'	337	5'-a-F-G-H-I-3'	18	5'-A-B-C-D-E-F-g-H-I-3'	5
5'-i-h-g-3'	332	5'-g-f-C-h-I-3'	18	5'-A-B-C-d-e-F-G-H-I-3'	5
5'-A-D-I-3'	314	5'-g-f-c-3'	18	5'-A-B-E-d-C-F-G-H-I-3'	5
5'-A-B-C-D-E-3'	298	5'-i-d-c-b-a-3'	18	5'-A-B-g-f-e-d-C-H-I-3'	5
5'-A-B-C-H-I-3'	294	5'-A-B-C-D-i-h-g-f-E-3'	17	5'-A-B-g-h-I-3'	5
5'-A-B-E-F-G-H-I-3'	268	5'-A-B-C-h-I-3'	17	5'-A-H-I-B-e-d-c-F-G-3'	5
5'-E-F-I-3'	241	5'-A-F-g-H-I-3'	17	5'-A-b-c-d-E-F-i-3'	5
5'-a-H-I-3'	230	5'-A-b-I-h-c-3'	17	5'-A-d-i-h-g-f-e-3'	5
5'-C-D-I-3'	214	5'-A-d-E-F-I-3'	17	5'-A-f-E-h-g-d-C-B-I-3'	5
5'-A-b-I-3'	208	5'-A-d-i-h-g-3'	17	5'-C-D-E-b-a-F-I-3'	5
5'-A-D-E-3'	207	5'-A-f-I-d-c-b-e-3'	17	5'-C-D-i-h-g-3'	5
5'-c-b-a-3'	188	5'-E-b-I-3'	17	5'-E-F-G-H-I-B-A-3'	5
5'-g-H-I-3'	177	5'-E-f-A-3'	17	5'-E-F-a-d-c-H-I-B-g-3'	5
5'-C-H-I-3'	175	5'-G-b-a-d-e-3'	17	5'-G-D-E-3'	5
5'-A-h-I-3'	172	5'-g-D-E-F-I-3'	17	5'-G-H-I-D-C-3'	5
5'-A-B-c-3'	156	5'-g-d-i-3'	17	5'-G-H-I-d-c-3'	5
5'-E-H-I-3'	147	5'-i-B-C-D-E-F-G-H-a-3'	17	5'-G-H-a-f-e-d-c-b-I-3'	5
5'-c-H-I-3'	146	5'-A-d-i-b-c-3'	16	5'-a-B-g-f-C-d-E-H-I-3'	5
5'-A-B-C-F-I-3'	145	5'-A-f-e-d-g-H-I-3'	16	5'-c-F-G-H-I-3'	5
5'-G-h-I-3'	130	5'-A-h-g-D-E-F-i-B-C-3'	16	5'-c-F-e-d-I-3'	5
5'-A-d-I-3'	128	5'-C-D-E-F-G-3'	16	5'-c-b-a-D-E-F-g-3'	5
5'-A-B-G-3'	122	5'-C-d-E-F-G-H-I-3'	16	5'-c-b-a-F-G-H-E-D-I-3'	5
5'-A-B-i-h-g-3'	120	5'-e-B-C-d-a-F-G-H-I-3'	16	5'-c-h-g-f-e-3'	5
5'-A-F-G-3'	117	5'-e-D-c-b-a-F-G-H-I-3'	16	5'-e-b-a-3'	5
5'-C-D-G-H-I-3'	116	5'-g-f-E-b-a-3'	16	5'-e-d-c-b-G-H-I-3'	5
5'-e-H-I-3'	109	5'-i-h-e-3'	16	5'-e-f-A-B-c-d-I-3'	5
5'-A-B-C-D-E-F-G-3'	107	5'-i-h-g-d-c-3'	16	5'-g-f-e-d-C-H-I-3'	5
5'-c-b-a-D-E-F-G-H-I-3'	103	5'-A-B-C-D-g-f-e-H-I-3'	15	5'-g-h-A-B-i-3'	5
5'-i-h-a-3'	103	5'-A-B-e-F-G-H-I-3'	15	5'-i-B-C-3'	5
5'-A-B-C-D-E-h-g-f-I-3'	98	5'-A-h-e-D-I-3'	15	5'-i-h-c-b-a-3'	5
5'-A-h-g-f-I-3'	97	5'-C-d-e-H-I-3'	15	5'-i-h-g-B-A-3'	5
5'-A-b-C-3'	96	5'-E-b-a-h-c-F-G-3'	15	5'-A-B-C-h-g-D-E-F-I-3'	4
5'-A-f-G-H-I-3'	96	5'-E-h-g-d-I-3'	15	5'-A-B-E-F-G-H-c-d-I-3'	4
5'-A-B-C-F-G-H-I-3'	93	5'-E-h-g-f-I-3'	15	5'-A-B-E-H-G-3'	4
5'-i-h-g-b-a-3'	92	5'-G-H-c-f-e-d-I-3'	15	5'-A-B-I-h-e-F-g-3'	4
5'-A-f-I-3'	86	5'-G-d-e-3'	15	5'-A-B-c-f-I-3'	4

5'-g-f-e-H-I-3'	84	5'-e-D-c-3'	15	5'-A-B-e-d-i-h-c-F-G-3'	4
5'-C-D-E-F-I-3'	83	5'-e-d-a-H-I-3'	15	5'-A-B-i-h-g-f-e-3'	4
5'-e-d-I-3'	82	5'-e-d-c-b-a-h-G-3'	15	5'-A-D-E-F-G-H-c-b-I-3'	4
5'-g-f-e-d-I-3'	82	5'-g-b-a-D-c-3'	15	5'-A-D-i-h-g-3'	4
5'-A-B-c-H-I-3'	81	5'-i-h-A-3'	15	5'-A-H-G-F-I-3'	4
5'-A-B-C-d-E-F-G-H-I-3'	79	5'-A-B-e-H-I-3'	14	5'-A-d-c-b-E-h-g-f-I-3'	4
5'-A-B-e-3'	79	5'-A-B-i-d-C-3'	14	5'-C-D-E-h-I-3'	4
5'-a-B-g-3'	79	5'-A-D-E-F-G-3'	14	5'-C-D-E-h-g-f-I-3'	4
5'-E-F-i-h-g-3'	77	5'-A-D-E-h-g-3'	14	5'-C-h-g-B-e-d-I-3'	4
5'-A-B-C-D-E-H-I-3'	76	5'-A-b-e-3'	14	5'-E-F-G-H-i-b-a-D-c-3'	4
5'-A-B-E-3'	73	5'-A-h-C-b-I-3'	14	5'-E-f-C-h-g-3'	4
5'-a-B-G-H-I-3'	72	5'-C-B-A-D-E-3'	14	5'-G-H-c-b-a-f-e-d-I-3'	4
5'-C-F-G-H-I-3'	71	5'-C-B-A-h-E-F-G-d-I-3'	14	5'-G-f-A-3'	4
5'-e-d-c-3'	71	5'-E-F-G-H-c-b-a-d-I-3'	14	5'-I-b-a-h-g-3'	4
5'-g-f-e-d-c-b-a-H-I-3'	71	5'-E-F-G-d-a-H-I-3'	14	5'-a-F-I-3'	4
5'-A-B-C-f-e-d-G-H-I-3'	70	5'-E-F-G-h-I-3'	14	5'-a-b-C-H-I-3'	4
5'-A-B-i-3'	70	5'-G-H-a-3'	14	5'-a-b-I-3'	4
5'-e-F-G-H-I-3'	69	5'-I-B-A-3'	14	5'-c-d-A-B-I-3'	4
5'-a-D-I-3'	68	5'-I-H-A-F-g-3'	14	5'-c-h-g-f-e-d-I-3'	4
5'-A-B-g-H-I-3'	66	5'-a-B-C-D-E-F-G-H-I-3'	14	5'-e-B-C-D-a-F-G-H-I-3'	4
5'-C-D-e-3'	66	5'-a-B-E-3'	14	5'-i-D-G-3'	4
5'-A-B-e-d-c-3'	65	5'-c-b-g-f-e-d-a-H-I-3'	14	5'-i-h-g-d-e-3'	4
5'-A-D-g-f-e-H-I-3'	65	5'-e-D-G-H-I-3'	14	5'-A-B-C-D-E-F-i-3'	3
5'-A-H-g-3'	65	5'-e-d-C-H-I-3'	14	5'-A-B-C-D-E-f-G-3'	3
5'-A-f-e-d-G-H-I-3'	63	5'-e-h-g-B-C-D-I-3'	14	5'-A-B-C-D-G-H-i-3'	3
5'-A-f-e-d-c-3'	63	5'-g-D-a-H-I-3'	14	5'-A-D-I-B-c-3'	3
5'-A-h-g-f-e-d-c-b-I-3'	62	5'-g-b-a-d-c-f-e-3'	14	5'-A-b-g-H-I-3'	3
5'-i-h-c-3'	61	5'-g-f-e-B-C-3'	14	5'-A-d-I-h-g-f-e-3'	3
5'-c-D-I-3'	60	5'-i-H-c-3'	14	5'-A-f-e-d-i-3'	3
5'-C-D-i-h-g-f-e-3'	59	5'-i-f-g-D-E-3'	14	5'-A-h-g-f-E-B-C-D-I-3'	3
5'-A-B-i-d-c-3'	58	5'-A-B-C-D-i-h-g-3'	13	5'-A-h-g-f-c-b-I-3'	3
5'-A-d-c-b-I-3'	57	5'-A-F-g-3'	13	5'-C-B-a-D-I-3'	3
5'-e-d-G-H-I-3'	57	5'-A-d-C-3'	13	5'-C-D-g-H-I-3'	3
5'-A-d-E-F-G-H-I-3'	56	5'-A-f-E-3'	13	5'-C-D-i-h-g-B-e-3'	3
5'-A-d-c-b-E-F-G-H-I-3'	56	5'-C-F-I-3'	13	5'-C-F-G-3'	3
5'-C-D-E-3'	56	5'-C-d-E-3'	13	5'-C-f-e-h-I-3'	3
5'-C-D-i-f-e-3'	56	5'-E-D-i-3'	13	5'-E-F-a-3'	3
5'-A-f-e-d-c-b-G-H-I-3'	55	5'-E-F-a-D-I-3'	13	5'-G-D-I-3'	3
5'-A-B-E-H-I-3'	54	5'-E-d-c-F-G-H-I-3'	13	5'-a-B-C-D-G-H-I-3'	3
5'-A-f-e-3'	54	5'-E-h-I-3'	13	5'-a-D-E-H-I-3'	3
5'-g-b-a-H-I-3'	54	5'-G-H-a-b-I-3'	13	5'-a-b-G-H-I-3'	3
5'-i-F-a-3'	54	5'-G-f-e-3'	13	5'-a-b-c-H-I-3'	3
5'-A-D-e-3'	53	5'-G-f-e-H-I-3'	13	5'-c-b-a-D-I-3'	3
5'-A-d-c-3'	53	5'-a-f-G-h-I-3'	13	5'-c-f-e-d-A-h-g-3'	3
5'-a-D-E-F-G-H-I-3'	52	5'-c-D-G-F-e-H-I-3'	13	5'-e-D-A-B-C-H-I-3'	3
5'-A-B-C-D-G-3'	51	5'-c-b-E-3'	13	5'-e-b-a-F-I-3'	3

5'-A-b-c-3'	51	5'-g-h-l-3'	13	5'-i-F-G-H-e-3'	3
5'-a-d-l-3'	51	5'-i-h-G-f-e-3'	13	5'-i-F-G-b-a-3'	3
5'-c-b-G-H-l-3'	51	5'-A-B-C-D-i-h-G-3'	12	5'-i-f-c-3'	3
5'-e-d-c-H-l-3'	51	5'-A-B-C-F-i-h-g-3'	12	5'-A-B-C-D-E-F-G-h-l-3'	2
5'-E-F-G-d-c-3'	50	5'-A-B-C-d-E-f-l-3'	12	5'-A-B-C-D-i-f-e-3'	2
5'-A-B-C-D-E-3'	49	5'-A-B-C-f-e-H-l-3'	12	5'-A-B-C-h-g-f-e-3'	2
5'-A-h-g-b-l-3'	49	5'-A-B-E-D-l-3'	12	5'-A-B-l-F-E-3'	2
5'-A-B-C-h-g-3'	48	5'-A-B-E-f-G-H-l-3'	12	5'-A-B-l-H-E-3'	2
5'-A-h-C-D-l-3'	48	5'-A-B-G-H-c-f-E-3'	12	5'-A-B-c-D-E-H-l-3'	2
5'-c-b-a-F-i-3'	48	5'-A-B-i-F-e-3'	12	5'-A-B-e-d-C-3'	2
5'-i-d-c-3'	48	5'-A-D-G-3'	12	5'-A-B-e-d-l-3'	2
5'-A-B-E-F-l-3'	47	5'-A-F-i-h-g-3'	12	5'-A-D-E-F-i-3'	2
5'-A-b-C-D-l-3'	47	5'-A-h-E-3'	12	5'-A-F-e-d-c-H-l-3'	2
5'-E-b-a-f-l-3'	47	5'-A-h-g-d-c-3'	12	5'-A-H-l-f-e-d-G-3'	2
5'-g-f-l-3'	47	5'-C-h-a-3'	12	5'-A-H-i-b-g-f-e-d-c-3'	2
5'-A-B-g-f-e-d-c-H-l-3'	46	5'-E-d-l-3'	12	5'-A-b-E-F-l-3'	2
5'-c-F-l-3'	46	5'-G-B-e-3'	12	5'-A-f-e-d-c-b-g-H-l-3'	2
5'-A-f-g-H-l-3'	45	5'-G-H-C-3'	12	5'-A-f-i-h-g-B-E-3'	2
5'-A-f-e-H-l-3'	44	5'-a-B-E-F-G-H-l-3'	12	5'-C-D-E-F-G-H-a-3'	2
5'-c-D-E-F-G-H-l-3'	44	5'-a-B-e-d-l-3'	12	5'-C-H-g-3'	2
5'-i-H-a-3'	44	5'-a-h-E-F-G-d-l-3'	12	5'-C-d-i-h-g-3'	2
5'-A-h-g-f-e-d-l-3'	43	5'-c-b-i-3'	12	5'-E-B-G-3'	2
5'-e-d-a-3'	43	5'-e-d-c-F-l-3'	12	5'-E-d-c-b-a-F-G-H-l-3'	2
5'-A-F-l-3'	42	5'-i-h-g-B-E-F-C-3'	12	5'-G-H-l-f-e-d-a-3'	2
5'-A-b-G-3'	42	5'-i-h-g-f-c-b-a-3'	12	5'-G-H-c-3'	2
5'-E-F-G-3'	42	5'-A-B-C-f-G-H-l-3'	11	5'-G-f-e-d-a-H-l-3'	2
5'-i-h-g-f-e-3'	42	5'-A-B-E-d-c-H-l-3'	11	5'-l-B-C-F-a-h-g-3'	2
5'-A-f-e-d-l-3'	41	5'-A-B-c-d-E-3'	11	5'-l-f-e-H-G-3'	2
5'-l-b-a-3'	41	5'-A-D-E-H-l-3'	11	5'-a-B-i-3'	2
5'-a-D-E-F-l-3'	41	5'-A-d-c-h-g-f-e-B-l-3'	11	5'-a-b-E-F-g-H-l-3'	2
5'-A-B-g-f-l-3'	40	5'-A-h-C-F-G-b-l-3'	11	5'-c-b-a-F-l-3'	2
5'-A-h-g-3'	40	5'-E-D-l-3'	11	5'-c-b-g-H-l-3'	2
5'-C-d-l-3'	40	5'-l-H-a-3'	11	5'-e-d-i-3'	2
5'-a-D-E-3'	40	5'-l-b-a-D-c-3'	11	5'-e-f-G-H-l-3'	2
5'-g-d-l-3'	40	5'-a-h-g-3'	11	5'-g-f-e-h-A-B-l-3'	2
5'-A-B-E-F-G-3'	39	5'-c-B-E-F-l-3'	11	5'-i-F-G-H-e-D-c-b-a-3'	2
5'-A-B-G-h-l-3'	39	5'-c-b-a-F-G-H-l-3'	11	5'-i-H-e-d-c-b-a-3'	2
5'-c-b-l-3'	39	5'-c-d-A-3'	11	5'-A-B-C-D-E-F-l-H-G-3'	1
5'-e-d-c-b-a-F-G-H-l-3'	39	5'-e-d-A-B-C-F-l-3'	11	5'-A-B-C-D-E-h-l-3'	1
5'-g-F-A-H-l-3'	39	5'-e-d-c-b-a-3'	11	5'-A-B-C-D-g-H-l-3'	1
5'-C-f-G-H-l-3'	38	5'-g-B-a-3'	11	5'-A-B-C-D-g-f-l-3'	1
5'-i-B-C-D-E-F-a-3'	38	5'-g-d-a-3'	11	5'-A-B-C-D-g-f-e-h-l-3'	1
5'-i-h-g-d-a-3'	38	5'-g-h-C-D-E-F-l-3'	11	5'-A-B-C-f-e-d-G-3'	1
5'-A-B-C-D-E-F-i-h-g-3'	37	5'-i-B-a-3'	11	5'-A-B-C-f-i-h-g-D-E-3'	1
5'-A-B-C-D-i-h-g-f-e-3'	37	5'-i-h-g-B-C-3'	11	5'-A-B-E-F-C-D-G-H-l-3'	1
5'-A-d-G-3'	37	5'-i-h-g-F-a-3'	11	5'-A-B-E-F-i-h-g-3'	1

5'-A-B-C-D-e-3'	35	5'-A-B-l-h-g-3'	10	5'-A-B-E-H-c-d-l-3'	1
5'-A-B-C-h-g-f-l-3'	35	5'-A-B-i-H-g-f-e-d-c-3'	10	5'-A-B-E-h-g-3'	1
5'-A-D-c-3'	35	5'-A-F-G-H-C-3'	10	5'-A-B-G-H-C-3'	1
5'-A-b-G-h-l-3'	35	5'-C-h-A-B-l-3'	10	5'-A-B-e-d-c-F-G-3'	1
5'-e-d-a-F-G-H-l-3'	35	5'-E-d-A-B-g-3'	10	5'-A-B-g-F-l-3'	1
5'-A-B-g-d-c-H-l-3'	34	5'-G-b-a-H-l-3'	10	5'-A-B-g-F-c-H-l-3'	1
5'-e-d-c-b-l-3'	34	5'-G-h-C-3'	10	5'-A-B-g-f-e-d-l-3'	1
5'-A-B-C-d-G-H-l-3'	33	5'-l-f-e-d-c-3'	10	5'-A-D-E-F-G-H-C-3'	1
5'-A-b-G-H-l-3'	33	5'-c-B-l-3'	10	5'-A-D-E-F-i-B-C-3'	1
5'-A-d-g-3'	33	5'-c-b-a-h-g-F-l-3'	10	5'-A-F-G-H-c-b-l-3'	1
5'-A-d-i-f-c-h-g-B-E-3'	33	5'-c-h-e-d-i-3'	10	5'-A-F-G-d-c-3'	1
5'-C-B-l-3'	33	5'-g-f-c-b-a-H-l-3'	10	5'-A-b-C-d-l-3'	1
5'-c-h-l-3'	33	5'-g-f-e-d-c-b-a-3'	10	5'-A-b-i-d-c-3'	1
5'-A-B-C-d-E-3'	32	5'-i-B-A-h-g-3'	10	5'-A-d-c-H-l-3'	1
5'-A-B-C-h-g-f-e-d-l-3'	32	5'-i-D-e-F-G-b-a-3'	10	5'-A-d-c-b-i-h-g-f-e-3'	1
5'-E-F-g-H-l-3'	32	5'-i-f-a-3'	10	5'-A-d-c-h-g-3'	1
5'-a-F-G-3'	32	5'-i-h-C-b-a-3'	10	5'-A-d-g-f-e-B-C-H-l-3'	1
5'-c-b-a-H-l-3'	32	5'-i-h-g-f-C-D-E-3'	10	5'-A-d-i-b-E-F-G-H-C-3'	1
5'-A-H-c-b-l-3'	31	5'-A-B-C-D-E-h-G-f-l-3'	9	5'-A-f-G-H-C-3'	1
5'-C-f-l-3'	30	5'-A-B-c-F-l-3'	9	5'-A-f-e-B-g-H-l-3'	1
5'-A-B-i-h-g-f-e-d-c-3'	29	5'-A-B-g-f-e-d-c-3'	9	5'-A-f-e-b-i-h-g-3'	1
5'-A-b-i-3'	29	5'-A-D-E-F-l-3'	9	5'-A-h-C-3'	1
5'-C-h-g-B-E-3'	29	5'-A-F-i-3'	9	5'-A-h-G-f-e-d-l-3'	1
5'-A-H-i-3'	28	5'-A-b-E-H-l-3'	9	5'-C-B-A-D-E-F-G-H-l-3'	1
5'-A-h-G-f-e-d-l-b-c-3'	28	5'-A-d-c-b-G-H-l-3'	9	5'-C-B-A-D-l-3'	1
5'-A-h-g-f-e-3'	28	5'-C-h-g-d-l-3'	9	5'-C-B-A-h-g-F-i-3'	1
5'-C-D-E-H-l-3'	28	5'-E-d-A-B-c-3'	9	5'-C-B-e-d-l-3'	1
5'-A-B-C-D-g-f-e-3'	27	5'-G-B-C-D-E-F-i-h-a-3'	9	5'-C-D-E-F-a-3'	1
5'-A-B-C-f-e-3'	27	5'-G-H-a-f-c-3'	9	5'-C-D-E-F-i-3'	1
5'-A-D-i-3'	27	5'-l-b-a-f-e-d-c-3'	9	5'-C-D-E-f-G-H-l-3'	1
5'-C-D-a-b-l-3'	27	5'-a-d-c-b-E-F-G-H-l-3'	9	5'-C-D-G-3'	1
5'-a-D-i-3'	27	5'-c-f-e-b-a-D-G-3'	9	5'-C-D-g-f-e-H-l-3'	1
5'-g-b-l-3'	27	5'-g-F-a-H-l-3'	9	5'-C-H-l-d-G-f-e-3'	1
5'-g-f-c-b-l-3'	27	5'-g-f-e-D-a-H-l-3'	9	5'-C-f-e-d-G-b-a-H-l-3'	1
5'-A-B-C-d-l-3'	26	5'-g-f-i-h-A-B-C-3'	9	5'-C-h-g-f-e-d-l-3'	1
5'-A-D-g-3'	26	5'-A-B-C-F-G-h-e-3'	8	5'-E-F-G-H-C-D-a-3'	1
5'-A-D-i-h-g-f-e-3'	26	5'-A-B-C-H-l-d-e-f-g-3'	8	5'-E-F-G-H-c-b-l-3'	1
5'-A-b-C-D-E-F-G-H-l-3'	26	5'-A-B-i-h-g-f-e-d-C-3'	8	5'-E-F-i-3'	1
5'-A-h-c-b-l-3'	26	5'-A-F-E-D-G-H-l-3'	8	5'-E-b-a-3'	1
5'-C-d-G-3'	26	5'-A-f-C-D-E-b-G-H-l-3'	8	5'-E-d-A-B-C-H-l-3'	1
5'-G-H-A-D-E-3'	26	5'-A-f-e-B-g-d-c-H-l-3'	8	5'-G-F-e-H-l-3'	1
5'-a-h-l-3'	26	5'-A-h-g-F-i-3'	8	5'-G-H-c-b-l-3'	1
5'-g-f-e-d-a-H-l-3'	26	5'-C-b-a-D-E-F-l-3'	8	5'-G-H-e-3'	1
5'-i-f-e-3'	26	5'-E-F-G-H-l-d-A-B-C-3'	8	5'-G-b-A-3'	1
5'-i-h-g-f-e-d-c-3'	26	5'-E-F-c-b-a-3'	8	5'-G-d-a-H-l-3'	1
5'-A-B-e-d-c-F-G-H-l-3'	25	5'-G-B-a-H-l-3'	8	5'-G-d-c-3'	1

5'-A-B-g-3'	25	5'-G-H-l-b-a-3'	8	5'-G-d-c-f-e-b-a-H-l-3'	1
5'-A-B-g-d-l-3'	25	5'-G-H-l-f-e-d-c-3'	8	5'-l-F-E-3'	1
5'-A-D-E-f-l-3'	25	5'-G-b-a-3'	8	5'-l-b-c-3'	1
5'-A-f-e-d-c-b-l-3'	25	5'-G-d-A-B-C-f-e-H-l-3'	8	5'-l-d-c-b-a-h-g-3'	1
5'-a-B-C-D-l-3'	25	5'-a-d-e-F-G-H-l-3'	8	5'-l-f-a-3'	1
5'-c-d-l-3'	25	5'-c-B-G-H-l-3'	8	5'-a-B-C-f-e-d-G-H-l-3'	1
5'-g-f-a-H-l-3'	25	5'-c-b-A-H-l-3'	8	5'-a-D-E-F-G-3'	1
5'-g-f-e-d-c-3'	25	5'-c-b-a-h-g-3'	8	5'-a-D-i-h-g-f-e-3'	1
5'-i-b-a-3'	25	5'-c-h-g-f-e-D-l-3'	8	5'-a-H-c-f-e-d-l-3'	1
5'-A-B-C-D-E-F-g-3'	24	5'-e-b-a-D-G-H-l-3'	8	5'-a-b-E-F-G-H-l-3'	1
5'-A-B-C-D-i-3'	24	5'-e-d-C-3'	8	5'-a-f-e-d-c-B-l-3'	1
5'-A-b-E-F-G-H-l-3'	24	5'-g-B-C-H-l-3'	8	5'-c-B-E-3'	1
5'-A-d-G-H-l-3'	24	5'-g-B-l-3'	8	5'-c-D-E-F-l-3'	1
5'-C-D-E-f-G-3'	24	5'-g-H-e-3'	8	5'-c-H-i-3'	1
5'-C-D-e-F-G-H-l-3'	24	5'-g-H-e-d-l-3'	8	5'-c-b-E-F-G-H-l-3'	1
5'-e-d-A-B-l-3'	24	5'-g-f-a-D-E-H-l-3'	8	5'-c-b-a-D-E-F-i-h-g-3'	1
5'-e-d-c-F-G-H-l-3'	24	5'-i-F-g-3'	8	5'-c-b-a-d-e-3'	1
5'-g-d-a-H-l-3'	24	5'-i-H-c-b-a-3'	8	5'-c-b-a-h-g-f-e-d-l-3'	1
5'-i-D-c-3'	24	5'-i-h-E-F-G-d-c-b-a-3'	8	5'-c-b-g-f-e-3'	1
5'-i-h-G-f-e-d-c-b-a-3'	24	5'-i-h-G-3'	8	5'-c-d-A-h-l-3'	1
5'-A-B-C-h-e-d-l-3'	23	5'-i-h-g-B-C-d-E-F-a-3'	8	5'-c-f-G-H-l-3'	1
5'-A-B-C-D-E-F-G-H-l-3'	23	5'-A-B-C-h-g-d-l-3'	7	5'-c-f-e-3'	1
5'-A-F-c-3'	23	5'-A-B-E-F-G-H-l-d-c-3'	7	5'-c-h-g-B-l-3'	1
5'-C-d-G-H-l-3'	23	5'-A-B-g-f-c-H-l-3'	7	5'-e-B-C-d-a-3'	1
5'-G-H-i-3'	23	5'-A-B-i-h-c-3'	7	5'-e-D-l-3'	1
5'-g-b-a-3'	23	5'-A-b-C-f-g-h-l-3'	7	5'-e-F-g-3'	1
5'-i-h-g-f-a-3'	23	5'-A-d-c-b-g-f-e-H-l-3'	7	5'-e-d-A-B-C-f-G-H-l-3'	1
5'-A-f-c-b-l-3'	22	5'-A-d-e-h-l-3'	7	5'-e-d-a-B-G-H-l-3'	1
5'-C-f-e-d-G-H-l-3'	22	5'-A-f-e-d-c-H-l-3'	7	5'-e-d-c-b-a-H-l-3'	1
5'-G-f-c-3'	22	5'-C-D-E-b-a-3'	7	5'-e-h-l-3'	1
5'-a-B-C-3'	22	5'-C-D-a-b-i-h-g-3'	7	5'-e-h-g-B-c-3'	1
5'-c-D-E-B-l-3'	22	5'-C-D-e-F-a-b-G-H-l-3'	7	5'-g-F-e-d-c-H-l-3'	1
5'-c-D-e-3'	22	5'-E-F-G-H-l-d-c-b-a-3'	7	5'-g-d-c-h-i-3'	1
5'-e-F-l-3'	22	5'-G-b-l-3'	7	5'-g-f-e-H-l-d-c-b-a-3'	1
5'-e-b-l-3'	22	5'-G-h-l-D-E-3'	7	5'-g-f-e-d-A-B-l-3'	1
5'-e-d-a-F-l-3'	22	5'-a-B-C-D-E-3'	7	5'-g-f-e-d-c-b-l-3'	1
5'-A-B-E-F-G-d-c-H-l-3'	21	5'-a-f-e-d-c-B-G-H-l-3'	7	5'-g-f-e-d-c-b-a-h-l-3'	1
5'-E-F-l-H-G-3'	21	5'-c-H-a-D-l-3'	7	5'-g-f-e-h-A-3'	1
5'-G-d-c-H-l-3'	21	5'-c-h-g-f-e-d-A-B-l-3'	7	5'-g-f-i-h-A-B-c-3'	1
5'-G-f-l-3'	21	5'-e-F-c-3'	7	5'-g-h-i-D-E-F-c-3'	1
5'-c-D-g-F-l-3'	21	5'-e-d-c-h-g-f-A-B-l-3'	7	5'-i-B-A-3'	1
5'-c-b-g-f-A-H-l-3'	21	5'-g-D-E-F-c-b-a-H-l-3'	7	5'-i-B-A-h-c-3'	1
5'-e-d-i-h-g-f-A-3'	21	5'-g-d-c-b-a-H-l-3'	7	5'-i-B-C-D-E-3'	1
5'-g-f-e-3'	21	5'-g-h-i-D-E-F-c-b-a-3'	7	5'-i-d-C-3'	1
5'-i-D-E-3'	21	5'-i-B-e-3'	7	5'-i-f-e-d-C-3'	1
5'-i-f-e-d-c-b-a-3'	21	5'-i-d-G-3'	7	5'-i-h-C-D-E-F-G-b-a-3'	1

5'-A-B-C-D-e-F-G-H-I-3'	20	5'-A-B-C-D-E-F-i-H-g-3'	6	5'-i-h-g-d-c-b-a-3'	1
5'-A-B-i-F-C-D-E-3'	20	5'-A-B-C-H-I-f-E-d-g-3'	6	5'-i-h-g-f-A-3'	1
5'-A-h-g-f-e-b-I-3'	20	5'-A-B-I-D-C-3'	6	5'-i-h-g-f-E-3'	1
5'-E-F-a-b-G-d-I-3'	20	5'-A-B-I-h-c-3'	6		
5'-E-f-I-3'	20	5'-A-B-i-h-g-d-c-3'	6		

Claims

1. A barcoding polynucleotide comprising at least five, preferably at least ten core structures, wherein each of said core structures comprises two uniquely identifiable sequences separated by a recombinase recognition sequence, wherein said recombinase is a recombinase capable of (i) mediating excision or (ii) mediating excision or inversion of a polynucleotide flanked by two recognition sequences of said recombinase.
2. The barcoding polynucleotide of claim 1, wherein each of said core structures is flanked at its 5' side and at its 3' side by at least one, preferably two, restriction enzyme recognition sequence(s).
3. The barcoding polynucleotide of claim 1 or 2, wherein each of said uniquely identifiable sequences differs from all other uniquely identifiable sequences present in said barcoding polynucleotide by at least one deletion, insertion, or, preferably, substitution.
4. The barcoding polynucleotide of any one of claims 1 to 3, wherein said recombinase is Flp and wherein said recombinase recognition sequence is an frt sequence or, preferably, wherein said recombinase is Cre and wherein said recombinase recognition sequence is a loxP sequence.
5. The barcoding polynucleotide of any one of claims 1 to 4, wherein the orientation of at least one recombinase recognition sequence is inverted relative to the orientation of the neighboring recombinase recognition sequences or the neighboring recombinase recognition sequence.
6. The barcoding polynucleotide of any one of claims 1 to 5, wherein each of said recombinase recognition sequences is separated from any neighboring recombinase recognition site or recombinase recognition sites by at least 10 nucleotides.
7. The barcoding polynucleotide of any one of claims 1 to 6, wherein said barcoding polynucleotide comprises the nucleotide sequence of at least one of SEQ ID NO: 1 to 30, preferably comprises the nucleotide sequence of at least one of SEQ ID NO: 1 to

- 10, more preferably comprises or consists of SEQ ID NO: 31 or 32, most preferably comprises or consists of SEQ ID NO:33.
8. A vector comprising a barcoding polynucleotide according to any one of claims 1 to 7.
 9. A host cell comprising a barcoding polynucleotide according to any one of claims 1 to 7 or a vector according to claim 8.
 10. A host organism comprising a barcoding polynucleotide according to any one of claims 1 to 7, a vector according to claim 8, or a host cell according to claim 9.
 11. A method of labeling a host cell, comprising introducing a barcoding polynucleotide according to any one of claims 1 to 7 into said host cell, introducing a corresponding recombinase into said cell, and thereby labeling a host cell.
 12. A method for providing an experimental animal comprising a labeled population of cells, comprising labeling at least one host cell in said experimental animal, preferably a stem cell, more preferably a non-embryonic stem cell, according to the method of claim 11 and keeping said experimental animal under conditions allowing proliferation of said host cell, thereby providing an experimental animal comprising a labeled population of cells.
 13. A kit comprising the barcoding polynucleotide according to any one of claims 1 to 7, the vector of claim 8, the host cell of claim 9, or the experimental animal of claim 10; and
 - (a) a corresponding recombinase (i) mediating excision or (ii) mediating excision and inversion of a polynucleotide flanked by recognition sequences comprised in said barcoding polynucleotide; and/or
 - (b) a barcoding polynucleotide encoding the corresponding recombinase of (a).
 14. A labeled host cell, obtained or obtainable by the method of claim 11.
 15. Use of a barcoding polynucleotide according to any one of claims 1 to 7 and/or the vector of claim 8 for labeling a host cell.

A

1 ► 2 3 ◄ 4 5 ► 6 7 ◄ 8 9 ► 10 11 ◄ 12 13 ► 14 15 ◄ 16 17 ► 18 19 ◄ 20

↓ +Cre

Excision between the first and third loxP

1 ► 6 7 ◄ 8 9 ► 10 11 ◄ 12 13 ► 14 15 ◄ 16 17 ► 18 19 ◄ 20

Inversion between the first and second loxP

1 ► 5 4 ◄ 3 2 ► 6 7 ◄ 8 9 ► 10 11 ◄ 12 13 ► 14 15 ◄ 16 17 ► 18 19 ◄ 20

B

1 ► 2 3 ◄ 4 5 ► 6 7 ◄ 8 9 ► 10 11 ◄ 12 13 ► 14 15 ◄ 16 17 ► 18 19 ◄ 20

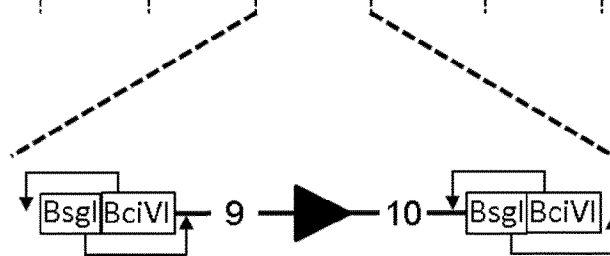


Fig. 1

C**Recombined barcodes**

1-3	1-6	1-7	1-10	1-11	1-14	1-15	1-18	1-19	2-4
2-5	2-8	2-9	2-12	2-13	2-16	2-17	2-20	3-5	3-8
3-9	3-12	3-13	3-16	3-17	3-20	4-6	4-7	4-10	4-11
4-14	4-15	4-18	4-19	5-7	5-10	5-11	5-14	5-15	5-18
5-19	6-8	6-9	6-12	6-13	6-16	6-17	6-20	7-9	7-12
7-13	7-16	7-17	7-20	8-10	8-11	8-14	8-15	8-18	8-19
9-11	9-14	9-15	9-18	9-19	10-12	10-13	10-16	10-17	10-20
11-13	11-16	11-17	11-20	12-14	12-15	12-18	12-19	13-15	13-18
13-19	14-16	14-17	14-20	15-17	15-20	16-18	16-19	17-19	18-20

Unrecombined barcodes

1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20
-----	-----	-----	-----	------	-------	-------	-------	-------	-------

Fig. 1 (continued)

3/17

**Mice carrying stem cell-specific,
inducible Cre and polyloxP reporter**

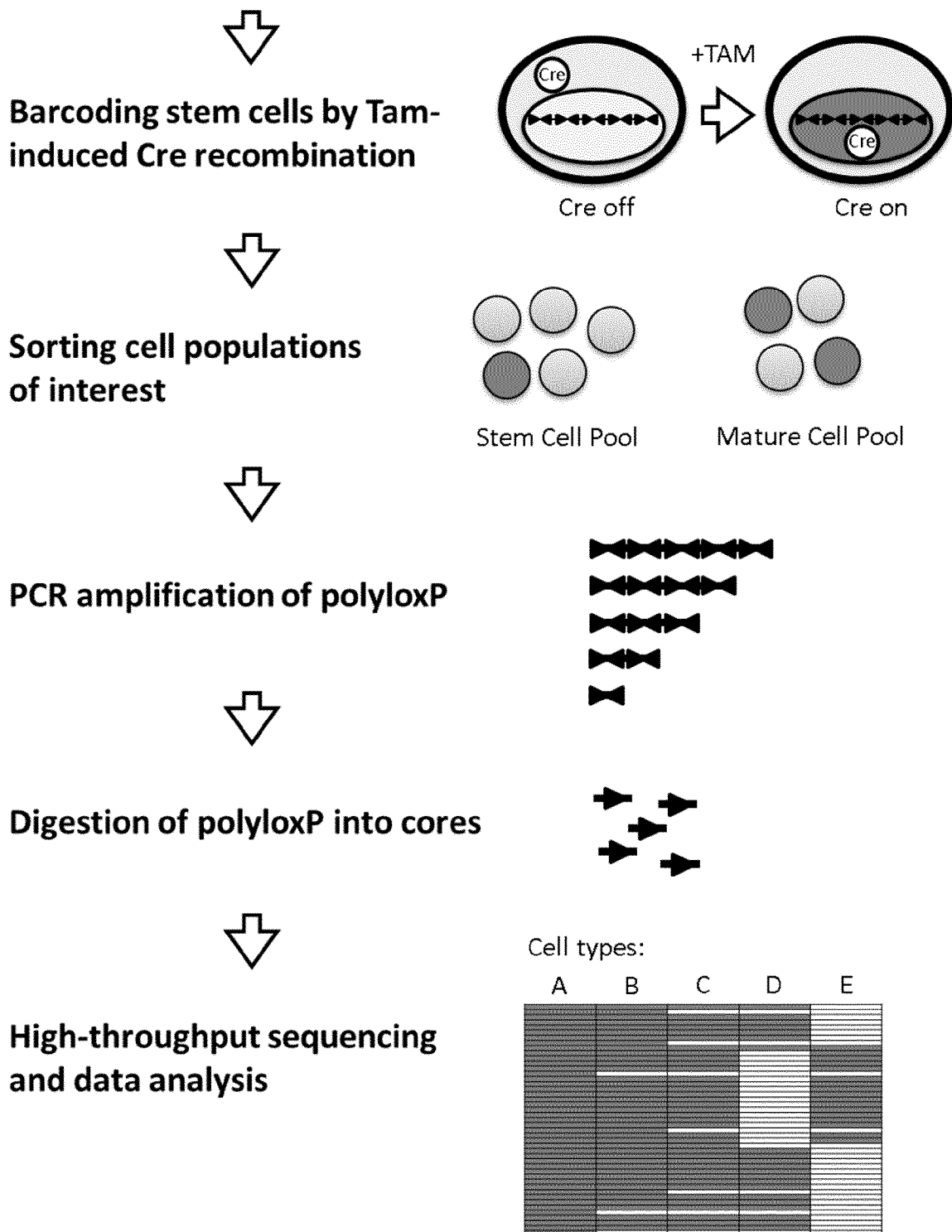


Fig. 2

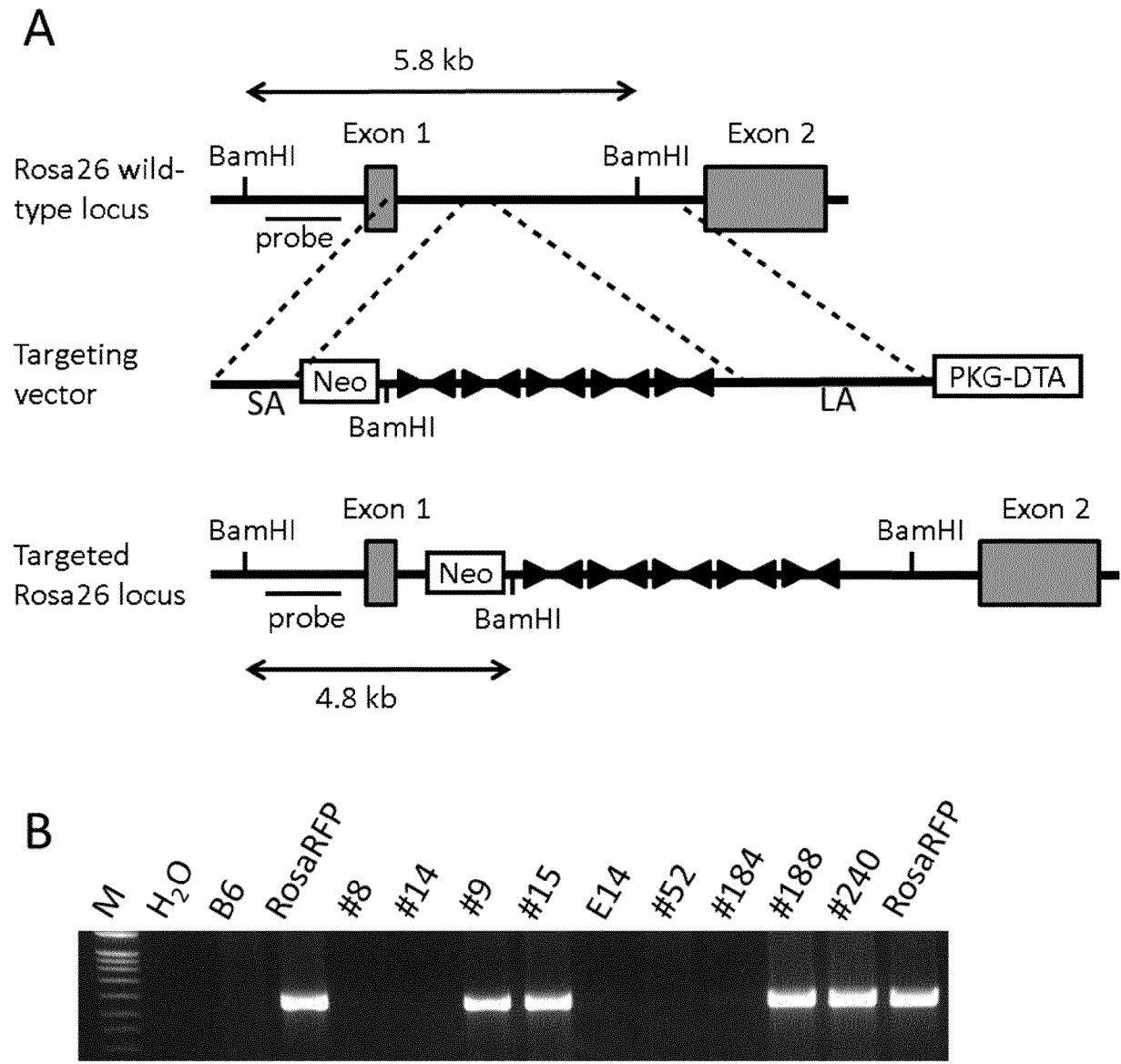


Fig. 3

5/17

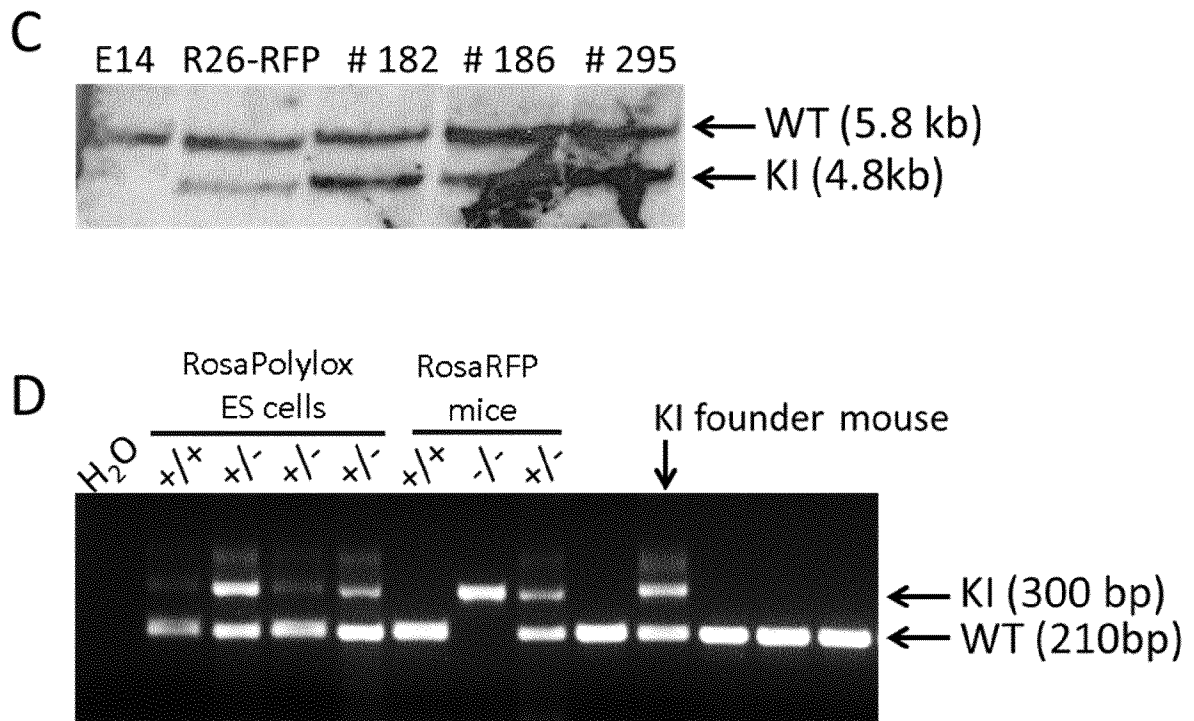


Fig. 3 (continued)

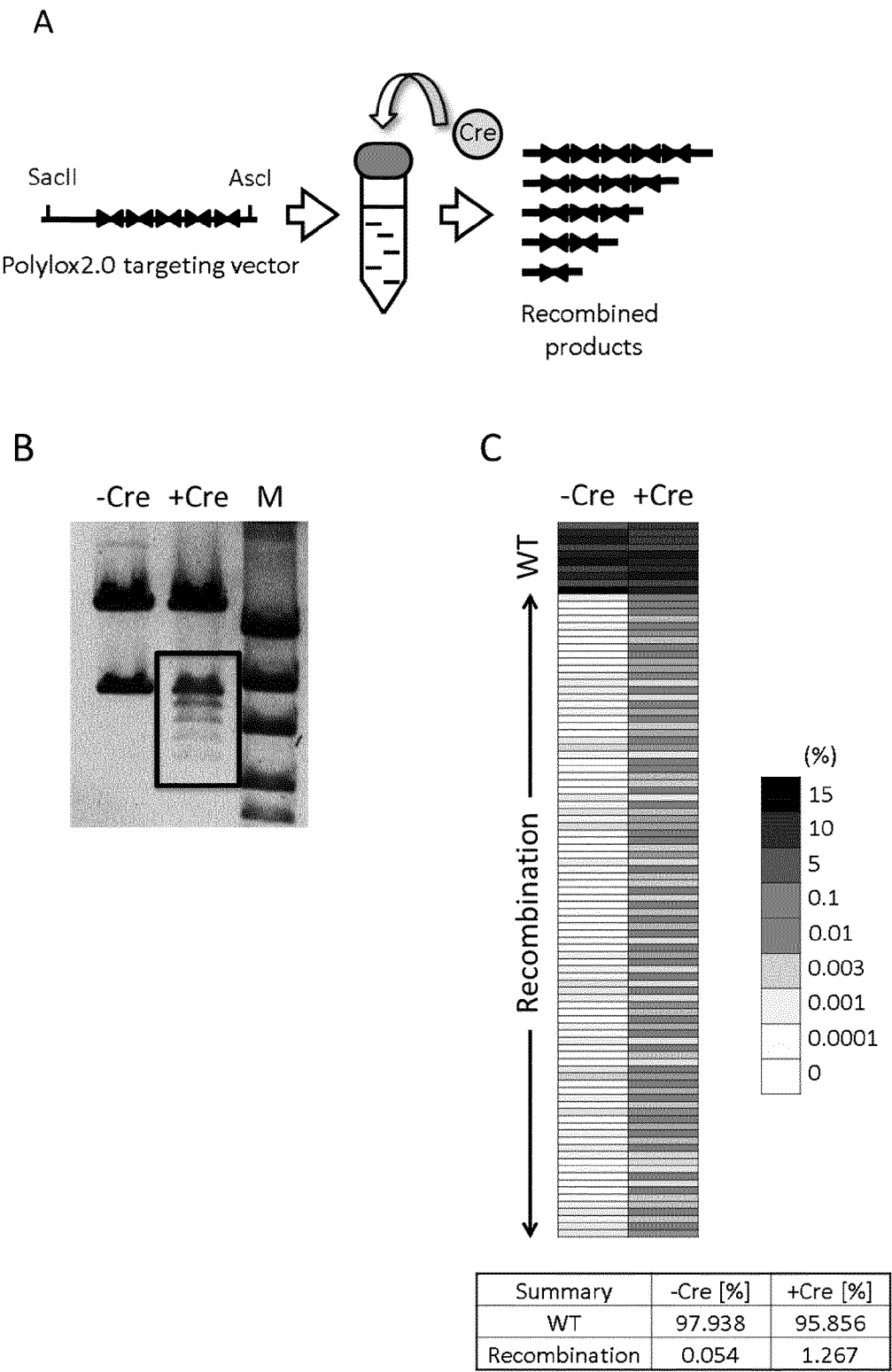


Fig. 4

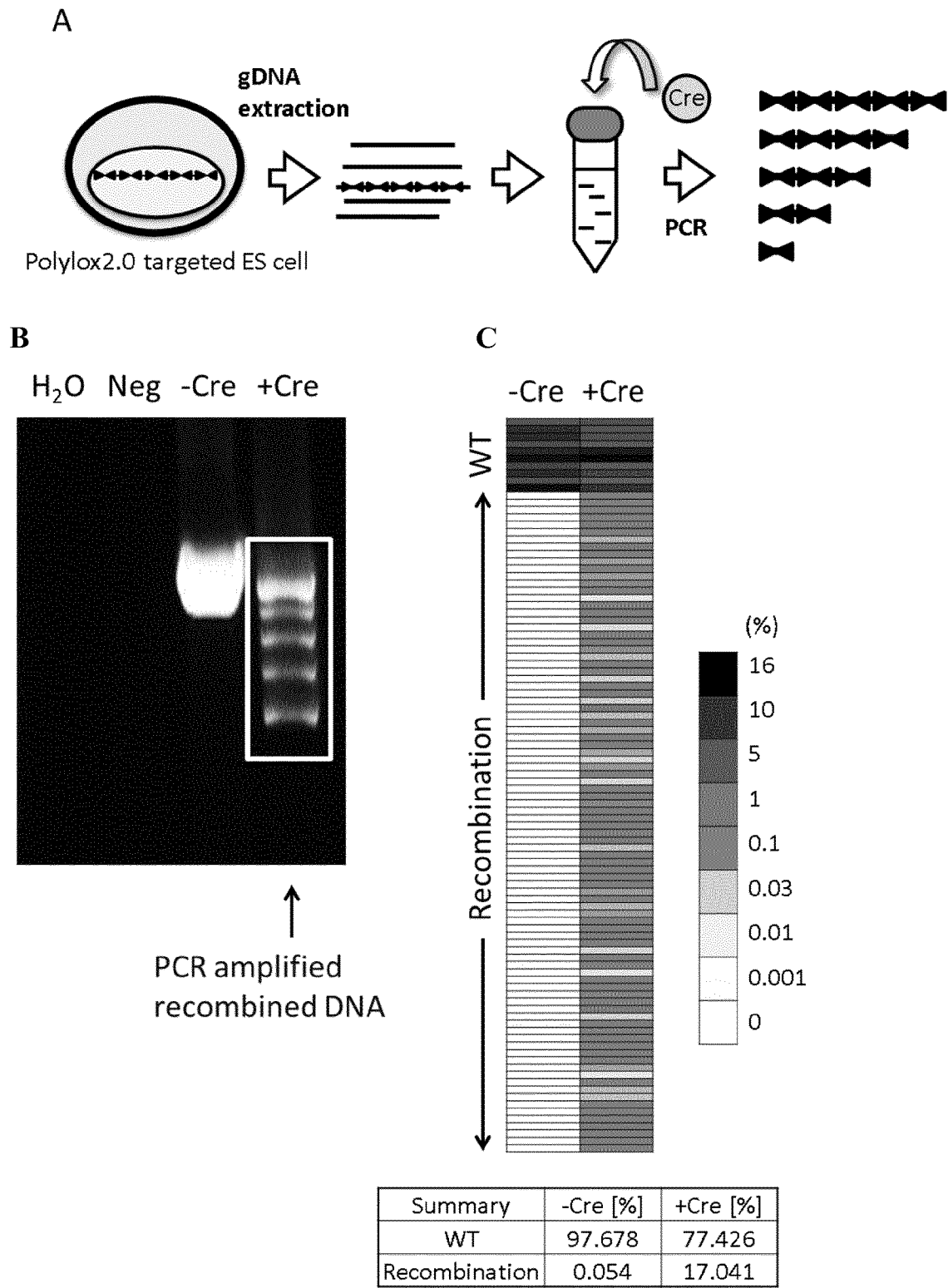


Fig. 5

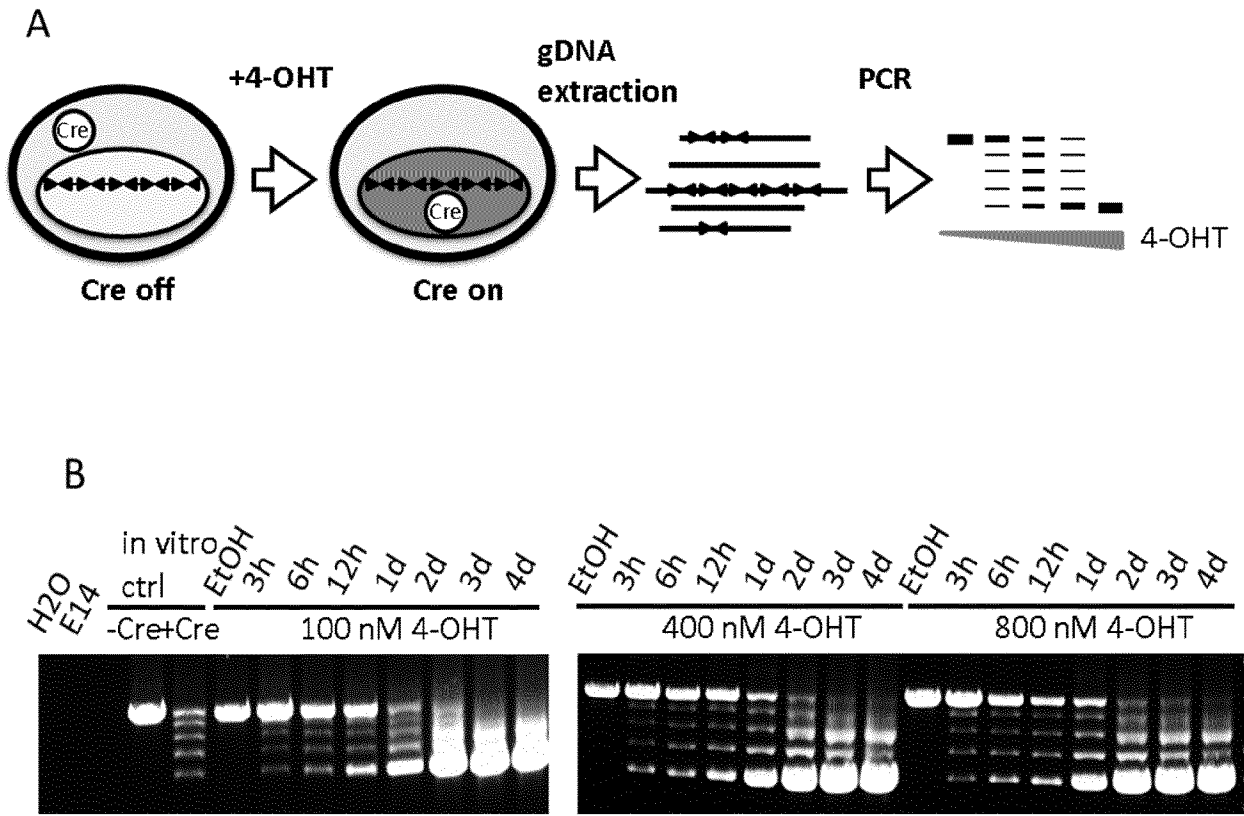


Fig. 6

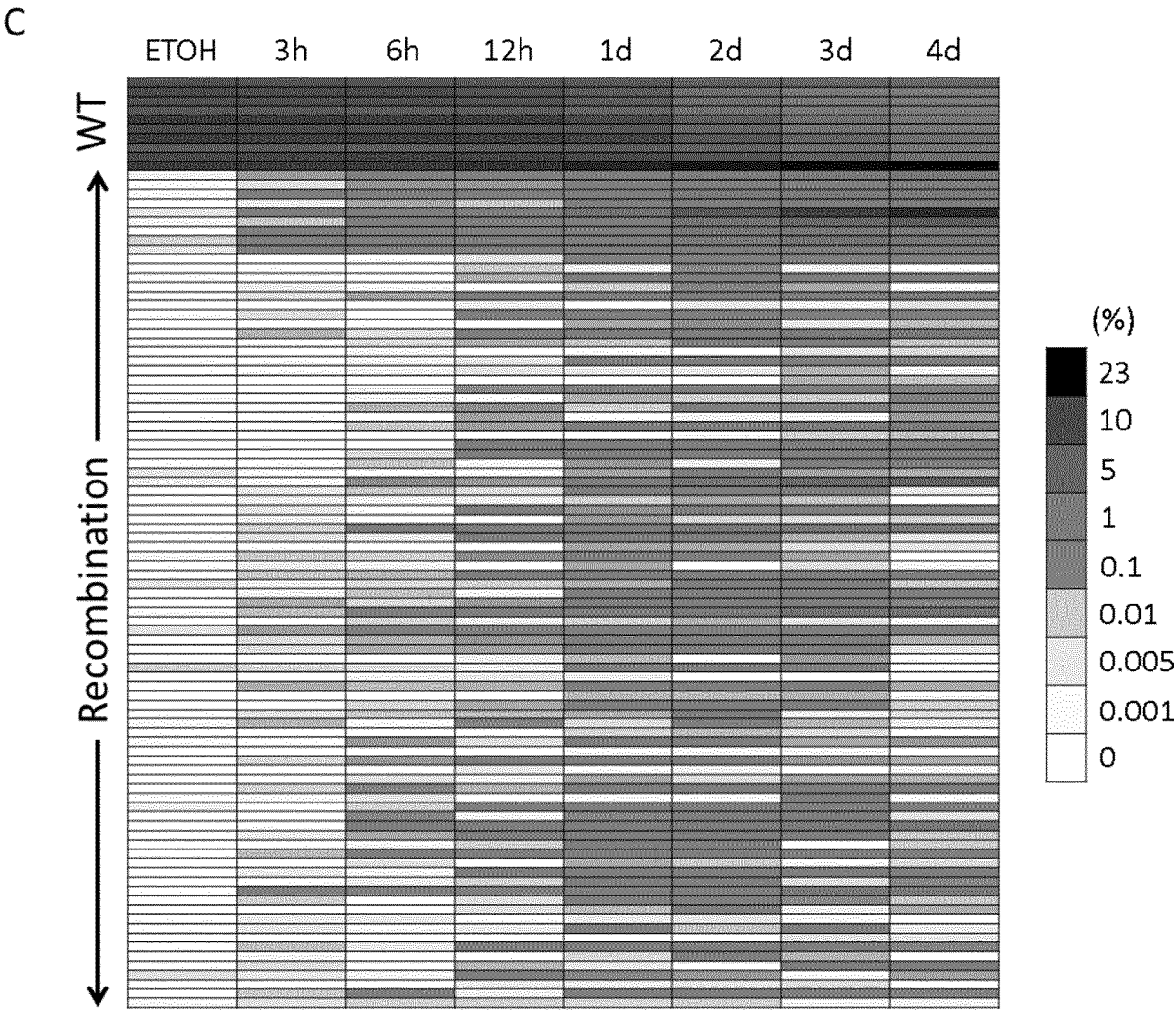


Fig. 6 (continued)

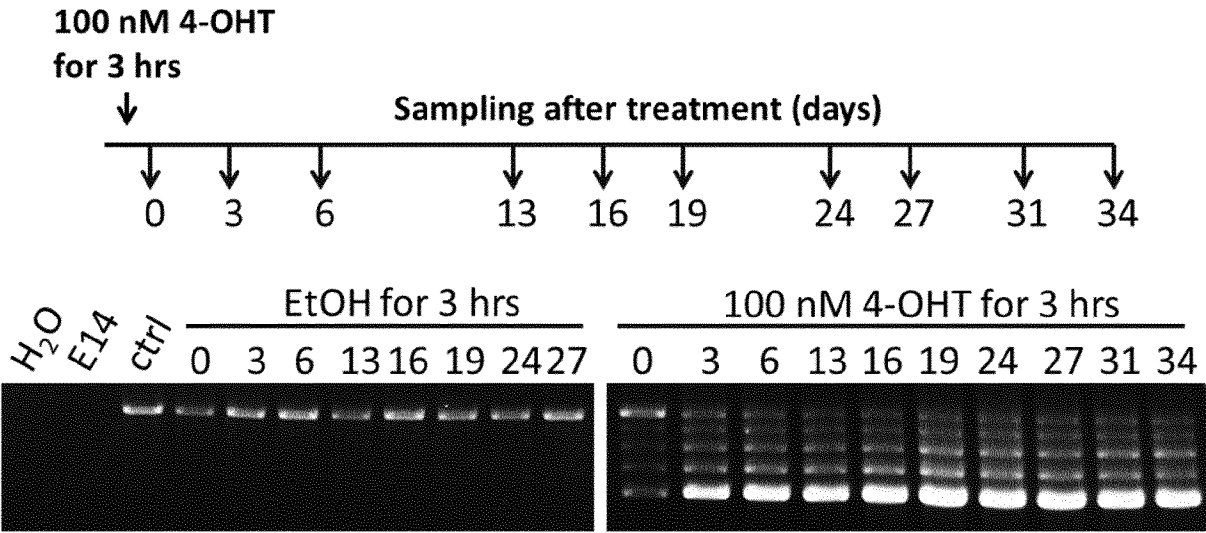
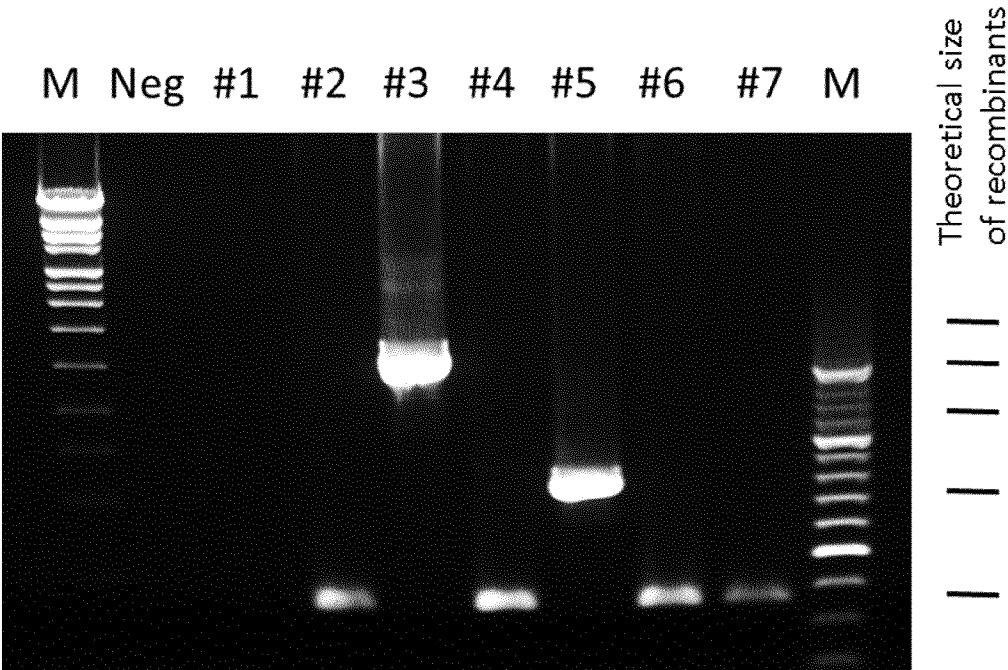


Fig. 7

A



B

Number of clones	Observed PolyloxP barcoded sequences
4	1-2-3-20
1	1-3-2-20
6	1-18-19-20
2	1-19-18-20
1	1-3-2-16-17-18-19-20
1	1-14-15-16-17-18-19-20
1	1-2-3-4-5-10-11-12-13-14-15-16-17-18-19-20
1	1-2-3-8-9-10-11-12-13-14-15-16-17-18-19-20
1	1-6-7-17-16-15-14-13-12-11-10-9-8-18-19-20

Fig. 8

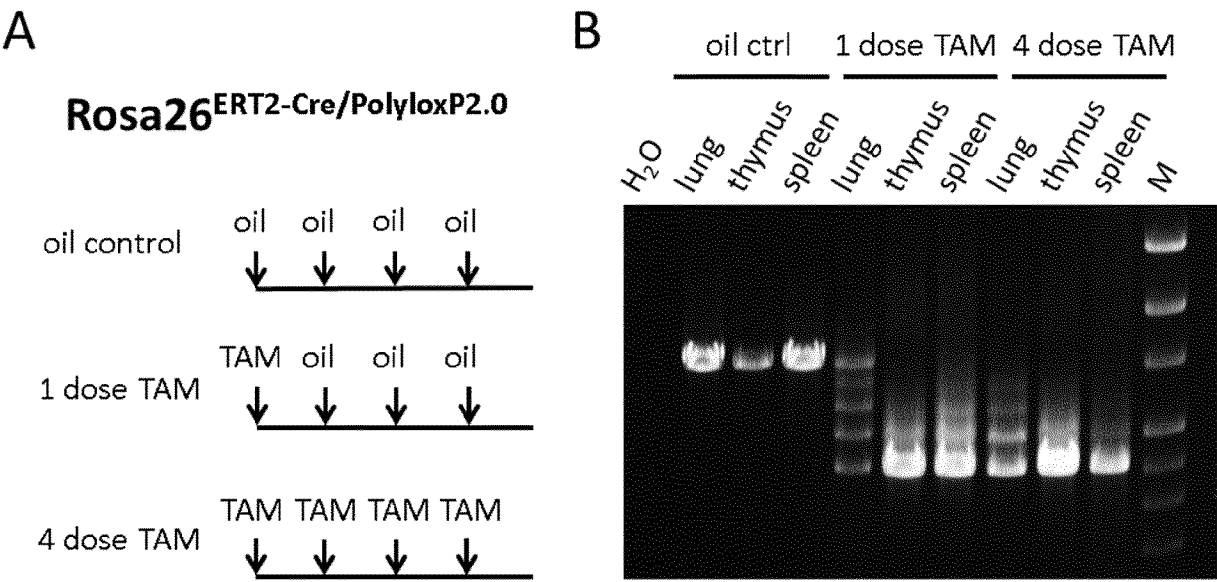


Fig. 9

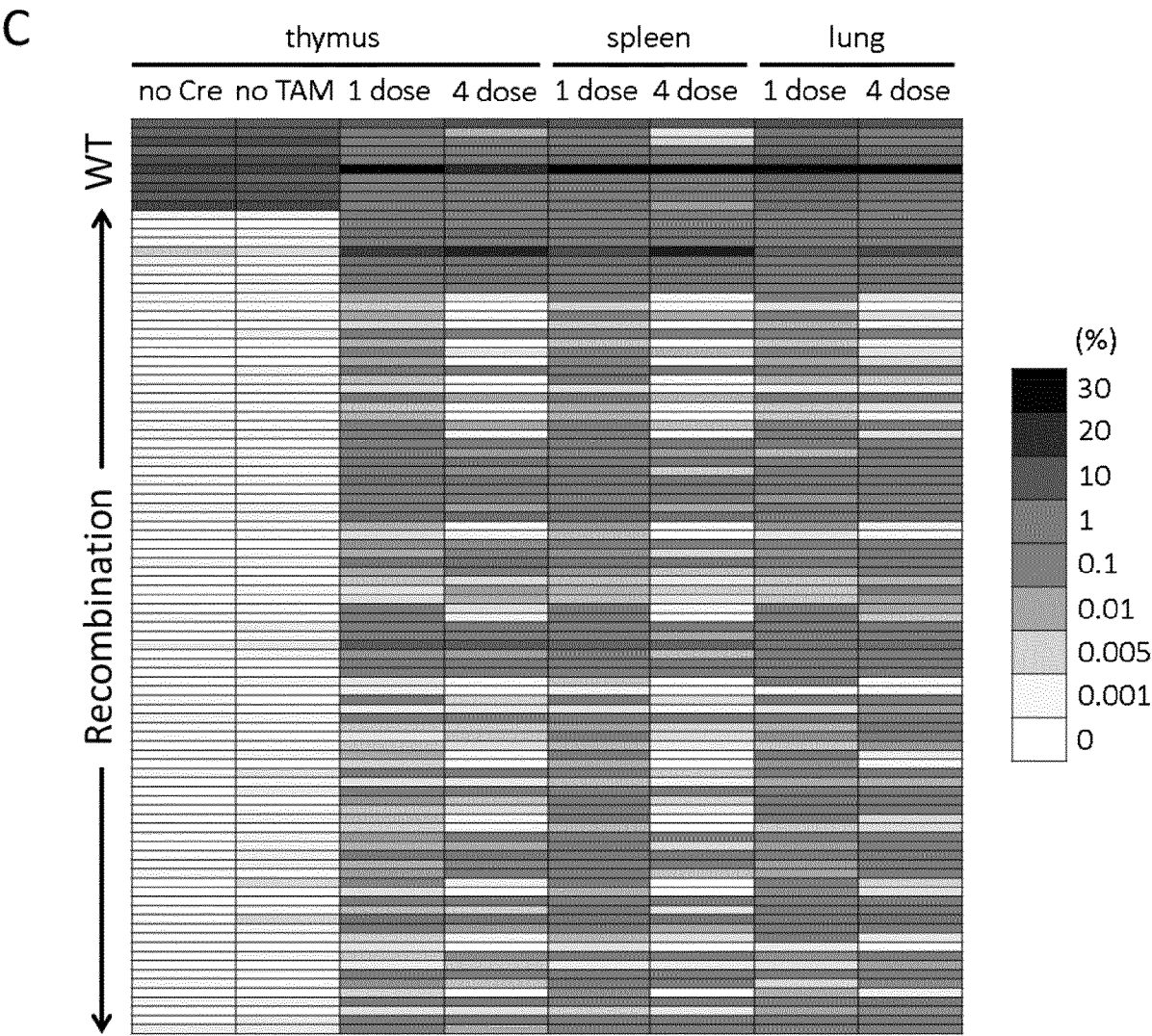


Fig. 9 (continued)

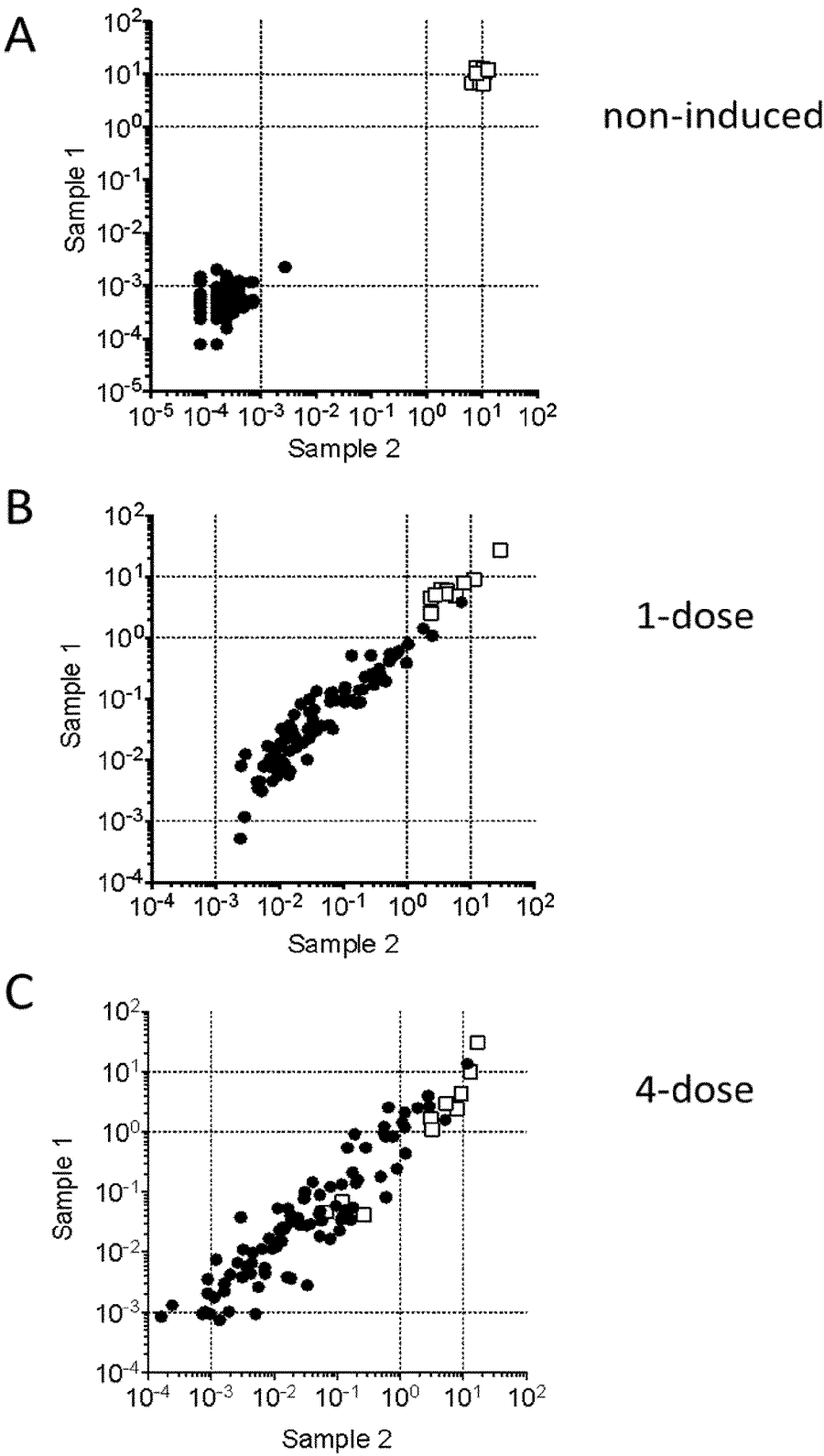


Fig. 10

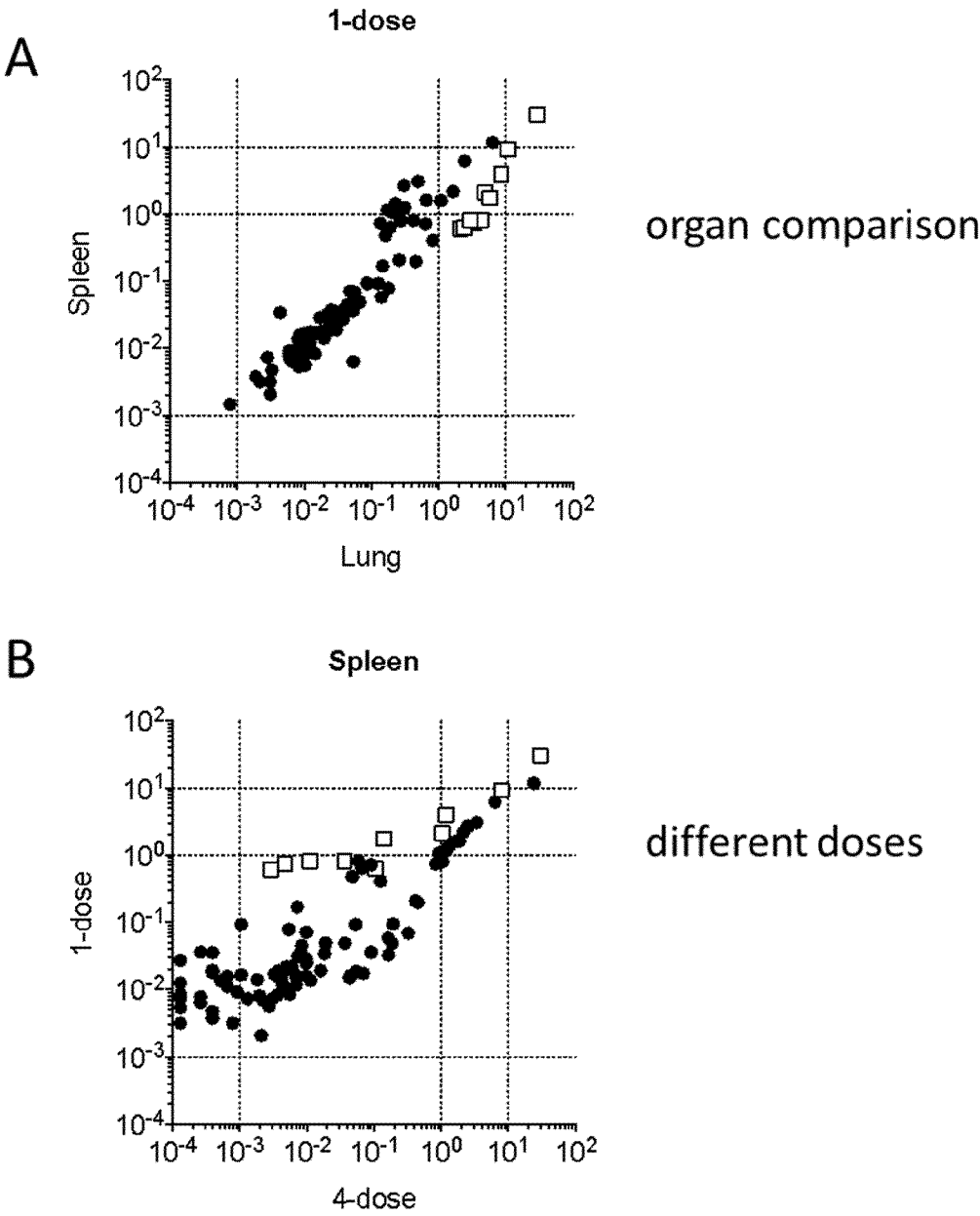


Fig. 11

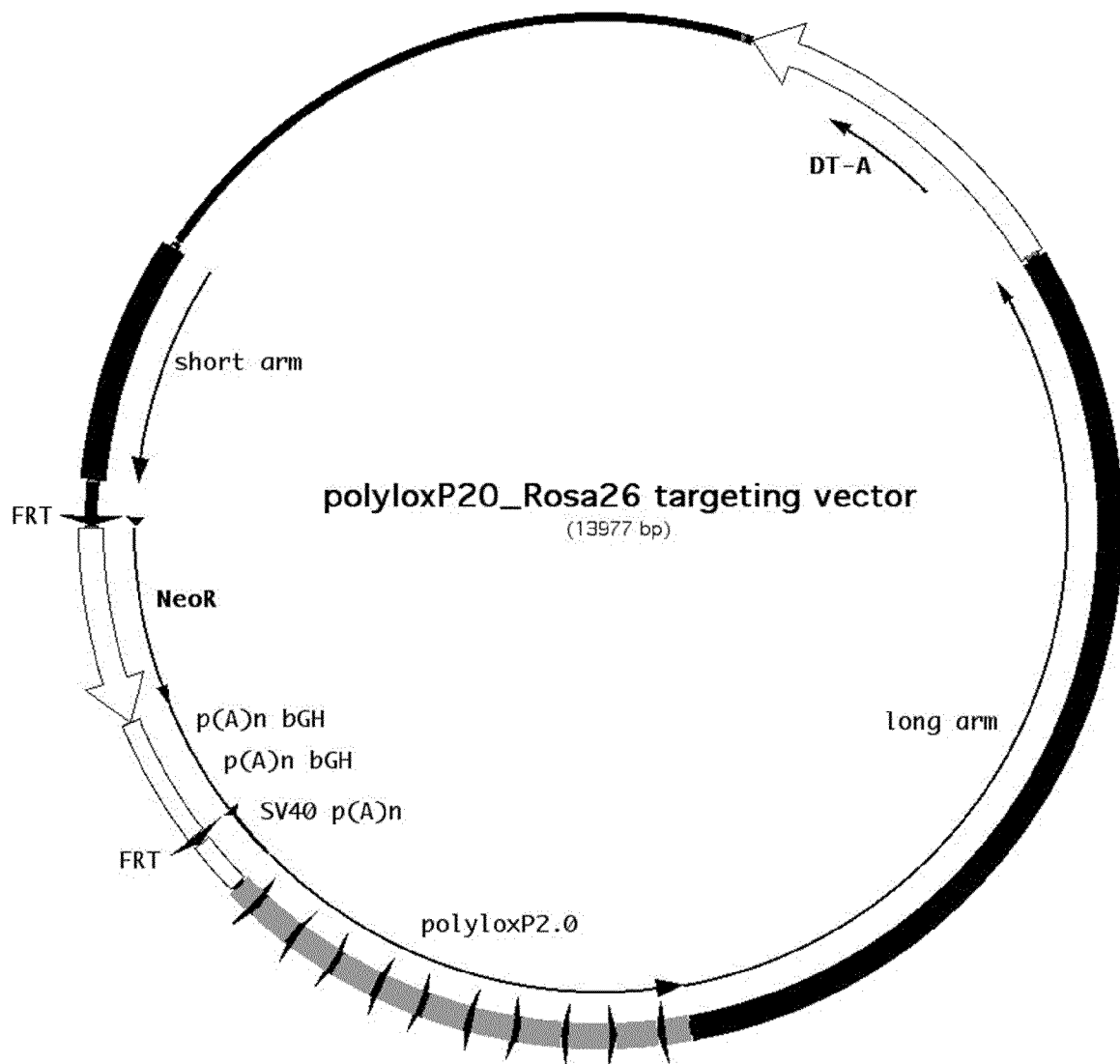
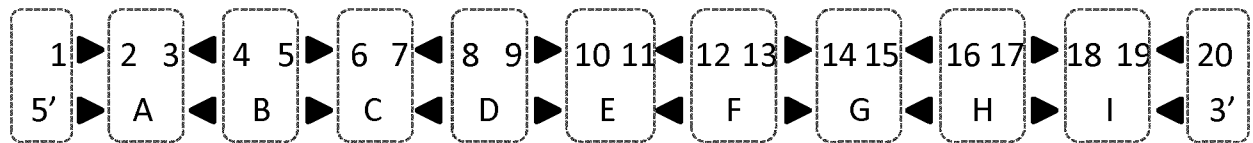
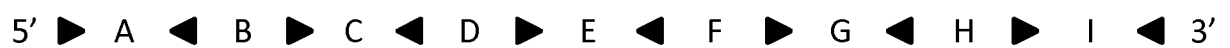


Fig. 12

A Terminology of PolyloxP2.0 DNA segments



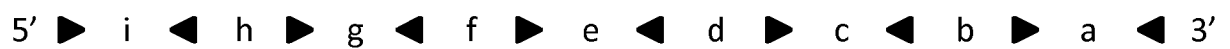
B Unrecombined PolyloxP2.0



C Excision of segments A and B



D Inversion of the segments A to I



E Excision of segments B and C plus inversion of Element



Fig. 13

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/065932

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	S. D. COLLOMS ET AL: "Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination", NUCLEIC ACIDS RESEARCH, vol. 42, no. 4, 12 November 2013 (2013-11-12), pages e23-e23, XP055304294, GB ISSN: 0305-1048, DOI: 10.1093/nar/gkt1101 abstract page 8, column 2, paragraph 2 - page 9, column 1; figures 1-5	1,3,5,6, 8-10
Y	US 2014/272988 A1 (ZADOR ANTHONY M [US] ET AL) 18 September 2014 (2014-09-18) abstract ----- -/-	4-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

21 September 2016

Date of mailing of the international search report

04/10/2016

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Mossier, Birgit

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2016/065932

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	I. D. PEIKON ET AL: "In vivo generation of DNA sequence diversity for cellular barcoding", NUCLEIC ACIDS RESEARCH, vol. 42, no. 16, 15 September 2014 (2014-09-15), pages e127-e127, XP055304273, GB ISSN: 0305-1048, DOI: 10.1093/nar/gku604 cited in the application abstract page 2, column 1 page 3, column 2, paragraph 2 - page 5, column 1, paragraph 2 page 7, column 2, paragraph 2; figure 1 -----	4-15
A	ANTHONY M. ZADOR ET AL: "Sequencing the Connectome", PLOS BIOLOGY, vol. 10, no. 10, 23 October 2012 (2012-10-23), page e1001411, XP055280511, DOI: 10.1371/journal.pbio.1001411 abstract page 4 - page 6; figures 2-5 -----	1-15
A	JONATHAN D. POLLOCK ET AL: "Molecular neuroanatomy: a generation of progress", TRENDS IN NEUROSCIENCE., vol. 37, no. 2, 1 February 2014 (2014-02-01), pages 106-123, XP055304020, NL ISSN: 0166-2236, DOI: 10.1016/j.tins.2013.11.001 abstract page 118, column 2, paragraph 2 - page 120; figures 7-9 -----	1-15
A	BRANDA C S ET AL: "Talking about a revolution: The impact of site-specific recombinase on genetic analyses in mice", DEVELOPMENTAL CELL, CELL PRESS, CAMBRIDGE, MA, US, vol. 6, no. 1, 1 January 2004 (2004-01-01), pages 7-28, XP002994211, ISSN: 1097-4172, DOI: 10.1016/S1534-5807(03)00399-X cited in the application the whole document ----- -/--	1-15

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/065932

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ALICE GERRITS ET AL: "Cellular barcoding tool for clonal analysis in the hematopoietic system", BLOOD, vol. 115, 1 January 2010 (2010-01-01), pages 2610-2618, XP055136898, DOI: 10.1182/blood-2009-06-abstract -----	1-15
T	TOM S. WEBER ET AL: "Site-specific recombinatorics: in situ cellular barcoding with the Cre Lox system", BMC SYSTEMS BIOLOGY, vol. 7, no. 3, 30 June 2016 (2016-06-30), page 33529, XP055304267, DOI: 10.1186/s12918-016-0290-3 the whole document -----	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2016/065932

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014272988	A1	18-09-2014	NONE
