DTU Library

Multiplex editing system

Pedersen, Lasse Ebdrup; Ronda, Carlotta; Kildegaard, Helene Faustrup; Lee, Jae Seong

[Link back to DTU Orbit](#)

(54) Title: MULTIPLEX EDITING SYSTEM

(57) Abstract: The present invention relates to a multiplex editing system. The system allows multiple editing of nucleic acid sequences such as genomic sequences, such as knockins of genes of interest in a genome, knockouts of genomic sequences and/or allele replacement. Also provided herein are a method for editing nucleic acids and a cell comprising a stably integrated endonuclease.

Multiplex editing system

Field of invention

The present invention relates to a multiplex editing system allowing multiple rounds of editing of nucleic acid sequences such as genomic sequences, e.g. knockins of genes of interest in a genome, knockouts of genomic sequences and/or allele replacement. Also provided herein are a method for editing nucleic acids and a cell comprising a stably integrated endonuclease.

Background of invention

The ability to integrate genes into the genome of an organism is of paramount importance to the biotech and biopharmaceutical industries. The purpose is often to stably produce a compound of interest, e.g. a protein, in high quantities. Currently, common methods rely on random insertions of single gene constructs followed by labor intensive screening of single cell clones. If multiple gene inserts are desired, the whole process must be repeated for each gene to be inserted.

The most labor-intensive part of inserting a gene at a specific site in the genome is discovering and validating a target area. The genomic area must facilitate high expression of the inserted gene, but must also contain a target sequence that the endonuclease can target with high specificity and efficacy. However, in the process of integrating a gene of interest at a target site, the target site is destroyed, preventing further use of this desirable location. Thus methods are needed which allow multiple insertions of genes of interest, i.e. multiplex editing of nucleic acids, in particular of genomes, by allowing repeated use of advantageous target locations.

Summary of invention

The present invention relates to a multiplex editing system that solves the above problem by allowing, in principle, unlimited numbers of insertions of genes of interest. The system allows multiple editing of nucleic acid sequences such as genomic sequences, such as knockins of genes of interest in a genome, knockouts of genomic sequences and/or allele replacement.

Thus in one embodiment the invention relates to a multiplex editing system comprising at least two pluralities of Continuously Regenerated Endonuclease Site Cassettes

CRESCs, wherein:

- a first plurality of CRESCs comprises at least one first targeted endonuclease site and at least one other nucleic acid sequence; and

- a second plurality of CRESCs comprises at least one second targeted endonuclease site and at least one other nucleic acid sequence,

wherein the first and second targeted endonuclease sites are different.

The system is based on the presence of at least two continuously regenerated endonuclease site cassettes (CRESCs), wherein each recognizes one of two different targeting sequences (TES) respectively, and each comprises a targeting sequence that it does not itself target. Upon each integration of a CRESC in the target sequence, the previous TES is destroyed and the other TES is integrated, allowing for recognition by the next CRESC. At the next integration, the other TES is destroyed and the previous TES is integrated, allowing for recognition by another CRESC. Thus is also provided a simple method for multiplex editing of e.g. genomic sequences or plasmids, where the same two TES can be used for integrating different genes of interest, thus greatly facilitating existing editing methods and allowing for the same advantageous targeting sites to be reused.

The method comprises the steps of i) introducing a first CRESC with a first TES in a cell capable of expressing an endonuclease, allowing said endonuclease to create a break in a nucleic acid comprised within said cell, thereby allowing integration of the first CRESC or at least a part thereof in said nucleic acid; ii) introducing a second CRESC with a second TES in said cell, allowing said endonuclease to create a break in the first CRESC, thereby allowing integration of the second CRESC or at least a part thereof in the first CRESC; and iii) optionally repeating steps i) and ii) with CRESCs comprising a TES identical to the penultimate TES so that the new CRESC or at least a part thereof is integrated in the previous CRESC. The advantages of the invention lie in the regeneration of advantageous integration sites and in limited off-target effects. The invention also allows multiple editing of nucleic acid sequences with limited cloning efforts. In some embodiments, the method allows for rapid generation of stable cell lines, in that it allows rapid selection of clones wherein targeted gene editing events have occurred. In specific embodiments, no selection marker is integrated at the target site of the host cell, and the engineered cell can thus undergo subsequent additional rounds of gene editing without first having to excise a selection marker.

In another aspect, the invention relates to a cell comprising a stably integrated endonuclease gene such as Cas9 or a variant thereof. Such a cell may provide a convenient way of performing knock outs and/or knockins, as will be explained in detail below.

Description of Drawings

Figure 1. Two plasmids according to the invention.

Figure 2. Principle for excision of a nucleic acid A.

Figure 3. Principle for excision of a selection marker.

Figure 4. Plasmids with CRESCs of the invention.

Figure 5. Principle of the method of the invention for multiplex editing.

Figure 6: HDR mediated targeted integration into COSMC.

Figure 7. Junction PCR analysis of clones.

Figure 8. mCherry expression and growth rate analysis of clones.

Figure 9. Analysis of donor DNA integration in transiently transfected and stable pools of cells. 5' and 3' junction PCR was performed on transiently transfected cells and stable pools of cells.

Figure 10. Indels generated by Mgat1 sgRNAs.

Figure 11: HDR mediated targeted integration into Mgat1.

Figure 12. Analysis of donor DNA integration in transiently transfected and stable pools of cells.

Figure 13. Analysis of Mgat1 disruptions by F-RCA staining.

Figure 14. GFP_2A_Cas9 expression vector. The GFP_2A_Cas9 expression vector was constructed with seamless USER cloning and verified by sequencing. The GFP gene facilitates FACS sorting for cells expressing the Cas9 nuclease.

Figure 15. Percentage of indels generated from Cas9 and GFP_2A_Cas9. The percentage of indels created at FUT8, BAK and BAX loci upon treatment of CHO cells with Cas9 and GFP_2A_Cas9 was analyzed by MiSeq sequencing of target loci. R1 and R2 represent biological replicate 1 and 2 respectively.

Figure 16. FACS Analysis of GFP_2A_Cas9 and triple sgRNA transfected CHO cells. CHO-S cells transiently transfected with GFP_2A_Cas9 and three sgRNAs against FUT8, BAK-1 and BAX simultaneously were analyzed for GFP expression on a BD FACS Jazz cell sorter.

Figure. 17. Analysis of the percentage of indels generated from multiplexing. The

percentage of indels generated in FUT8, BAK1 and BAX loci in cells harvested before and after FACS sorting of GFP positive cells. R1 and R2 describes biological replicate 1 and 2 respectively.

Figure 18. Plasmids for targeted integration. Schematic overview of plasmid expressing Cas9 together with a fluorescent marker, plasmid expressing the sgRNA targeting the integration site and plasmid containing the donor DNA expressing GOI within the homology arms and a fluorescent marker outside the homology arms. Together, these three plasmids are applied in site-specific integration of GOI into genomic DNA.

Figure 19. HDR mediated integration. HDR mediated targeted integration of the donor DNA leaves no selection marker or fluorescent marker gene, only the GOI expression vector, in the targeted genomic site. The desired genomic locus which is targeted by the Cas9 nuclease directed with the selected sgRNA and the donor DNA with homology arms undergo HDR (Homologous recombination-mediated integration). The pure GOI expression cassette is inserted specifically into the desired genomic locus of the targeted integrants without using selection and without integration of the fluorescent gene.

Figure 20. Transient transfection and FACS enrichment. The three vectors expressing GFP 2A linked Cas9, sgRNA against COSMC and donor DNA expressing EPO, Enbrel or Rituximab was co-transfected into CHO-S cells. Cell pools enriched by fluorescent sorting was isolated and cultivated for further analysis.

Figure 21. Site-specific targeted integration into COSMC. 3' and 5' Junction PCR was performed on cell pools upon CRISPR Cas9 mediated targeted integration of EPO, Rituximab and Enbrel into COSMC.

Figure 22. Percentage of indels generated in 2AGFP_Cas9 cell line. The percentage of indels created at FUT8, BAK and BAX loci upon induction was analyzed by MiSeq sequencing of target loci.

Figure 23. FACS data of sorting after transfection. GFP positive population and sub population of medium expressing GFP (Bottom 50%) were sorted to investigate the correlation of Cas9 expression and editing efficiency.

Figure 24. Percentage of indels generated from permanently integrated Cas9 sorted for medium of high GFP expression. The percentage of indels created at FUT8 loci upon treatment of CHO cells with permanently integrated Cas9 and transiently transfected guideRNA was analyzed by MiSeq sequencing of target loci after FACS sorting for GFP.

Figure 25. Map of pcDNA4/TO::2AGFP_Cas9 (SEQ ID NO: 111).

5

Figure 26. Map of pJTI™ R4::2AGFP_Cas9 (SEQ ID NO: 110).

Figure 27. Map of pcDNA™6/TR _ZEO (SEQ ID NO: 112).

Figure 28. Dark field image of CHO-K1 polyclonal population expressing 2AGFP_Cas9.

Data analysis with Celigo S cell cytometer (Nexcelom Bioscience) analyzer 77.8%of

Cas9 expressing cells.

Figure 29. FACS data for the wild type (control).

Figure 30. FACS data for the induced population.

Figure 31. FACS data for the non-induced population.

Figure 32. Workflow for accelerated generation of monoclonal cell line with site-specific

and clean targeted integration of GOI into the genome for stable expression.

Mammalian cells, for example CHO cells, is transfected with GFP 2A labelled Cas9,

sgRNA against integration site and donor DNA expressing mcherry. On day two or

three after transfection, the cells are sorted for mcherry and GFP fluorescence to select

for the most potential pool of cells with targeted integration events. These cells are

single cell sorted to facilitate generation of monoclonal cell lines or bulk sorted to

facilitate analysis and later single cell sorting. The generated monoclonal cell lines are

either screened for fluorescence and non-mcherry expressing cells (potential targeted

integrants) are selected for 3' and 5' junction PCR to identify targeted integrants. In

addition, analysis of the productivity of the monoclonal cell lines generated can be

applied to detect potential integrants as these will show less expression variation

among each other due to the same integration site. All the targeted integrants will be

analyzed for growth and productivity to select the best performing clones.

Figure 33. Structure of MACE CHIP/platform and Loop work.

25

Figure 1. Two plasmids according to the invention. RS: restriction site. TES: targeted

endonuclease site. UTR: untranslated region. (1a) The plasmid carries CRESC1

comprising: a sequence coding for a first gene (gene 1 surrounded by a 5'-UTR and a

3'-UTR), a first selection marker (SM1 surrounded by a 5'-UTR and a 3'-UTR), and two

identical TES1 surrounding SM1. (1b) The plasmid carries CRESC2 comprising: a

sequence coding for a second gene (gene 2 surrounded by a 5'-UTR and a 3'-UTR), a

second selection marker (SM2 surrounded by a 5'-UTR and a 3'-UTR), and two

identical TES2 surrounding SM2.

35

Figure 2. Principle for excision of a nucleic acid A. (2a) The nucleic acid A in its

genomic location, surrounded by two TES1. (2b) The endonuclease creates a break in each TES1 (X), resulting in (2c) excision of the nucleic acid A and formation of an indel in the absence of donor DNA or repair template. If a donor DNA is present, it is inserted instead of the nucleic acid A (not shown).

Figure 3. Principle for excision of a selection marker. (3a) The selection marker in a genomic location, after insertion of a CRESC comprising a first gene 1, a first TES1, a selection marker, a second TES1. (3b) The endonuclease creates a break in each TES1 (X), resulting in (3c) excision of the selection marker and formation of an indel in the absence of donor DNA or repair template. If a donor DNA is present, it is inserted instead of the selection marker (not shown).

Figure 4. Plasmids with CRESCs of the invention. (4a) Plasmid with homology arms (HA) surrounding a first CRESC1 comprising a gene 1, two identical TES1, and a selection marker SM1. (4b) Plasmid with homology arms surrounding a second CRESC2 comprising a gene 2, two identical TES2, a selection marker SM2. (4c) The same plasmid as in 4a but devoid of homology arms.

Figure 5. Principle of the method of the invention for multiplex editing. (5a) A TES present in the genome (TESgen) of a wild type cell is targeted by an endonuclease 0 which creates a break (lightning arrow) at said TESgen. CRESC1, comprising a GOI1 (gene of interest 1), a marker 1 (M1), two identical TES1 (T1) surrounding the marker, and homology arms HA_L1 and HA_R1, homologous to genomic regions surrounding the genomic TES (TESgen), integrates as shown, yielding cell line 1. (5b) An endonuclease 1 now targets the two TES1 (T1), generating two breaks (lightning arrows). CRESC2, comprising a GOI2, a marker 2 (M2), two identical TES2 (T2) surrounding the marker, and homology arms HA_L2 and HA_R1, homologous respectively to the region upstream the first TES1 and the region downstream the second TES1, integrates as shown, yielding cell line 2. The marker 1 (M1) is excised together with the two TES1 (T1). The genome now harbours GOI1, GOI2, TES2 (T2), marker 2 (M2), TES2 (T2). TES2 may be identical to the TES originally present on the genome (TESgen in 5a). (5c) An endonuclease 2, which optionally is the same as in step a, now targets the two TES2 (T2), generating two breaks (lightning arrows). If TES2 and the genomic TES of 5a are identical, the same means for recognising the TES may be used at this step as in step 5a. CRESC3, comprising a GOI3, a marker 1

(M1), two identical TES1 (T1) surrounding the marker, and homology arms HA_L3 and HA_R1, homologous respectively to the region upstream the first TES2 (T2) and the region downstream the second TES2 (T2), integrates as shown, yielding cell line 3. The marker used in this step is preferably marker 1 to facilitate generation of CRESC3. The marker 2 is excised together with the two TES2. The genome now harbours GOI1, GOI2, GOI3, TES1 (T1), marker 1 (M1), TES1 (T1).The next round of editing will be performed with a CRESC4 comprising a gene of interest 4 and two identical TES2 surrounding a marker 2, similar to step 5b.

Detailed description of the invention

Definitions

Allele replacement: The term 'allele replacement' refers to the process of replacing an allele, e.g. a genomic allele, with another allele of the same gene. Thus allele replacement involves both knockin and knockout.

Break: the term 'break' shall be construed as referring to a double stranded break or to a nick or single-stranded break in a DNA strand.

CRESC: CRESC as understood herein refers to a "continuously regenerated endonuclease site cassette", i.e. an endonuclease site cassette allowing regeneration of an endonuclease site, in theory for an unlimited amount of steps. In the present context, the CRESC comprises at least one endonuclease site cassette comprising at least one endonuclease site specifically recognised by a given endonuclease. The CRESC may comprise additional elements, such as a targeting sequence designed to be recognised by an endonuclease differing from said given endonuclease. The CRESC may also comprise elements such as nucleic acid sequences coding for genes of interest or for makers allowing for selection of clones where the CRESC has been successfully knocked in the target nucleic acid.

Donor DNA: the term 'donor DNA' refers to the DNA sequence used as template for repair by homologous recombination.

DSB: A double strand break (DSB) as understood herein refers to a break on both strands of a nucleic acid. DSBs are particularly hazardous to the cell because they can lead to genome rearrangements. Two major mechanisms exist to repair DSBs: non-

8

homologous end joining (NHEJ) and homologous recombination (HR). The choice of pathway depends on parameters such as the organism and the cell cycle phase.

Endonuclease site cassette: an endonuclease site cassette is a nucleic acid sequence which is typically designed by a user to be recognised by a given endonuclease. Such a cassette comprises at least one endonuclease site, which is the nucleic acid sequence specifically recognised by said endonuclease, but may also contain other functional elements.

Enhancers: Enhancers are *cis*-acting elements that can regulate transcription from nearby genes and function by acting as binding sites for transcription factors.
Gene: A gene as understood herein refers to a gene or a putative gene. The gene may code for a selection marker, a protein of interest, a peptide, or it may be a gene resulting in the production of a miRNA, a siRNA, a tRNA, or any gene which can be transcribed and/or translated.

Homologous Recombination (HR or HDR): Homologous Recombination is one of the two major pathways for repairing DSBs. HR is a type of genetic recombination in which nucleotide sequences are exchanged between two similar or identical molecules of DNA. HR involves "copying" information from a donor DNA.

Homology arm: The term covers a stretch of DNA with sequences homologous to the upstream and downstream regions of a region of interest, in particular of a cut site or a targeted endonuclease site.

Indel: an indel refers to a mutation class, resulting in an insertion and/or a deletion of nucleotides, leading to a net change in the total number of nucleotides. The change in the total number of nucleotides is typically in the range of 1 to 5 nucleotides, but may be up to 100 nucleotides or more.

Insulators: Insulators are transcriptional regulation elements in eukaryotes that stop communication between enhancers on one side of it with promoters on the other side. Insulators play an important role in limiting the chromatin region over which enhancers can operate.

9

Knockin: 'Knockin' refers to the process by which genes can be inserted in a genome. The inserted genes may be genes from the same organism or from other species.

Knockout: 'Knockout' refers to the process by which genes can be inactivated in an organism, for example by deletion, mutation, of part of the gene, the whole gene, or of part or all of the elements necessary for the gene to be expressed in a functional protein.

Multiplex editing: The term refers herein to simultaneous or serial editing of multiple nucleic acid sequences. For example, multiplex editing may refer to simultaneous knockins and/or simultaneous knockouts or a combination of simultaneous knockins and knockouts. It may also refer to gene editing performed in two or more consecutive steps. For example, multiplex editing may refer to serial knockins and/or serial knockouts or a combination of serial knockins and knockouts, where each step involves a given gene editing event such as a knockin or a knockout. Multiplex editing also encompasses combinations of simultaneous and serial editing events.

Nick: A nick is a discontinuity in a double-stranded DNA molecule where there is no phosphodiester bond between adjacent nucleotides of one strand.

Non-Homologous End Joining (NHEJ): NHEJ is one of the two major pathways for repairing DSBs. DNA ligase IV, in complex with XRCC4, directly joins the two ends at the break. The ends at the break may be resected prior to repair, which may lead to loss of some nucleotides and improper repair. Thus NHEJ is often error-prone.

Nuclear Localisation Sequence (NLS): A nuclear localisation signal or sequence (NLS) is an amino acid sequence which 'tags' a protein for import into the cell nucleus by nuclear transport. Typically, this signal consists of one or more short sequences of positively charged lysines or arginines exposed on the protein surface. Different nuclear localised proteins may share the same NLS. An NLS has the opposite function of a nuclear export signal, which targets proteins out of the nucleus.

Nucleic acid: The term refers herein to a sequence of nucleotides.

Nucleic acid of Interest: As understood herein, a nucleic acid of interest is a nucleic

acid sequence which comprises at least one gene, and/or at least one 5'-UTR such as a promoter and/or at least one 3'-UTR such as a terminator.

Open reading frame (ORF): As understood herein, the term open reading frame refers to a nucleic acid sequence with long stretches of codons uninterrupted by stop codons. ORFs often comprise genes.

Plurality: By plurality is understood at least two, such as three, such as four or more.

Polynucleotide / Oligonucleotide: The terms "polynucleotide" and "oligonucleotide" as used herein denote a nucleic acid chain. Throughout this application, nucleic acids are designated starting from the 5'-end.

Promoter: A promoter is a DNA sequence near the beginning of a gene (typically upstream) that signals the RNA polymerase where to initiate transcription. Eukaryotic promoters may comprise regulatory elements several kilobases upstream of the gene and typically bind transcription factors involved in the formation of the transcriptional complex. Promoters may be inducible, i.e. their activity may be induced by the presence or absence of a biotic or abiotic compound.

Recognition: As understood herein, the term 'recognition' refers to the ability of a molecule to identify a nucleotide sequence. For example, an enzyme or a DNA binding domain may recognise a nucleic acid sequence as a potential substrate and bind to it. Nucleic acids such as guiding RNAs may recognise a specific sequence to which they are at least partly homologous. Certain enzymes may require the presence of additional recognition means, such as guiding RNAs or DNA binding domains, to efficiently recognise their substrate sequence.

Recombinase: As understood herein, the term 'recombinase' refers to an enzyme that can catalyse directionally sensitive DNA exchange reactions between short (30–40 nucleotides) target site sequences that are specific to each recombinase. These reactions enable four basic functional modules, excision/insertion, inversion, translocation and cassette exchange.

Stable integration: This term refers herein to the stable insertion into a genomic nucleic

acid of another nucleic acid, resulting in permanent integration of the other nucleic acid into a genome. In other words, stably integrated nucleic acids are not readily lost by their host; consequently cell lines with stably integrated nucleic acids generally do not need to be maintained by selective pressure with a marker.

5

Targeted endonuclease sites TES: As understood herein, a targeted endonuclease site (TES) is a nucleic acid which is specifically recognised by a given endonuclease such as Cas9, a zinc-finger nuclease (ZFN) or a transcriptor-activator like effector nuclease (TALEN).

10

Terminator: A terminator is a DNA sequence near the end of a gene (typically downstream) that signals the RNA polymerase where to stop transcription. Eukaryotic terminators are recognized by protein factors and termination is followed by polyadenylation of the mRNA.

15

In one aspect the invention relates to a multiplex editing system comprising at least two pluralities of Continuously Regenerated Endonuclease Site Cassettes CRESCs, wherein:

20
    i. a first plurality of CRESCs comprises at least one first targeted endonuclease site and at least one other nucleic acid sequence; and
    ii. a second plurality of CRESCs comprises at least one second targeted endonuclease site and at least one other nucleic acid sequence,

wherein the first and second targeted endonuclease sites are different.

25

In another aspect the invention relates to a method for editing nucleic acids with the system herein, comprising the steps of:

30
    i. introducing a first CRESC in a cell capable of expressing at least one endonuclease and allowing one of the at least one endonuclease to create a break in a nucleic acid comprised within said cell; thereby allowing integration of the first CRESC or at least a part thereof in said nucleic acid;
    ii. introducing a second CRESC in said cell and allowing one of the at least one endonuclease to create a break in the first CRESC; thereby allowing
35
        integration of the second CRESC or at least a part thereof in the first

CRESC;

iii. optionally, introducing any subsequent CRESC in said cell and allowing
one of the at least one endonuclease to create a break in the previously
integrated CRESC, thereby allowing integration of the subsequent CRESC
or at least a part thereof in the previous CRESC.

In another aspect the invention relates to a cell comprising a stably integrated
endonuclease gene such as Cas9 or a variant thereof.

In another aspect the invention relates to a kit of parts comprising at least two
pluralities of CRESCs, wherein:

i. a first plurality of CRESCs comprises at least one first targeted
endonuclease site; and

ii. a second plurality of CRESCs comprises at least one second targeted
endonuclease site,

wherein the first and second targeted endonuclease sites are different.

Multiplex editing system

Disclosed herein is a multiplex editing system. The system allows multiplex editing of
nucleic acids. The term 'multiplex editing' refers to serial and/or simultaneous knockins
and/or knockouts of a nucleic acid. The multiplex editing system disclosed herein
comprises at least two pluralities of Continuously Regenerated Endonuclease Site
Cassettes CRESCs, wherein:

i. a first plurality of CRESCs comprises at least one first targeted
endonuclease site and at least one other nucleic acid sequence; and

ii. a second plurality of CRESCs comprises at least one second targeted
endonuclease site and at least one other nucleic acid sequence,

wherein the first and second targeted endonuclease sites are different.

The use of different targeted endonuclease sites (TES) in the first and second
pluralities of CRESCs allows precise control over the editing events. A system in which
the first and second targeted endonuclease sites are identical would result in numerous
integration events occurring at once at the same site. Having two different targeted
endonuclease sites ensures that only one CRESC or at least a part thereof will
integrate in the TES that it recognises. Examples of CRESCs are shown in figure 1.

13

Figure 1a shows the structure of the first plurality of CRESCs, wherein gene 1 and SM1 (selection marker 1) may be multiple genes or selection markers and may be replaced by other genes or selection markers for other members of the first plurality. Figure 1b shows the structure of the second plurality of CRESCs, wherein gene 2 and SM2 (selection marker 2) may be multiple genes or selection markers and may be replaced by other genes or selection markers for other members of the second plurality. All members of the first plurality contain the same TES1; all members of the second plurality contain the same TES2.

The nucleic acids to be edited may be on a genome, such as a chromosome, or on a plasmid.

Although the present system is particularly advantageous for performing serial gene editing, it can also be used for simultaneous editing of multiple loci or for a combination of serial and simultaneous editing of multiple loci.

In some embodiments, the multiplex editing system comprises at least three pluralities of CRESCs, wherein the first and second pluralities are as defined above, and the third plurality comprises at least one third targeted endonuclease site and at least one other nucleic acid sequence, said at least one third targeted endonuclease site being either:

    i. identical to the at least one first or second targeted endonuclease sites; or

    ii. different from the at least one first and second targeted endonuclease sites.

TES

The targeted endonuclease site (TES) is a nucleic acid sequence which is capable of being specifically recognised by an endonuclease such as Cas9, a ZFN or a TALEN. Recognition of the TES by the endonuclease may require additional means, such as guiding RNAs for Cas9. Such means may be provided by introducing vectors capable of expressing such means in a cell, or by integrating nucleic acid sequences capable of expressing such means in the genome of a cell.

Upon recognition, the endonuclease creates a double-strand break or a nick within the first TES comprised within a first CRESC. If donor DNA is present, such as a second CRESC comprising homology arms homologous to sequences surrounding the first TES of the first CRESC, repair of the break may facilitate integration of donor DNA

within the first TES.

The at least one TES may be at least two TES, such as at least three TES, such as at least four TES. In a preferred embodiment, the at least one TES is two TES. In some embodiments, the two TES delimit a nucleic acid A, i.e. the nucleic acid comprised between the first and the second TES is a nucleic acid A (figure 2a).

Creation of a DSB by an endonuclease at each TES as shown in figure 2b may result in one of the following outcomes:

i)   formation of an indel at the first TES

ii)   formation of an indel at the second TES

iii)   excision of nucleic acid A and religation of the ends without indel formation

iv)   formation of an indel at both TES

v)   excision of nucleic acid A and religation of the ends with indel formation
        (figure 2c).

Outcomes i), ii) and iii) are the result of partial endonuclease activity and are not desired. Adjustment of the reaction conditions such as the reaction temperature, the duration of the reaction, the concentration of endonuclease (for example by adjusting the expression level from the endonuclease gene) may limit the occurrence of these outcomes. Outcomes iv) and v) are most likely to occur in conditions where the endonuclease activity is optimal. A preferred outcome is outcome v), in which nucleic acid A is excised, the ends are religated and an indel is formed (figure 2c).

In one embodiment, the nucleic acid sequence A comprises at least one gene coding for a selection marker (figure 3). Suitable selection markers include, but are not limited to: antibiotic resistance genes, auxotrophy or prototrophy genes, genes coding for fluorescent proteins to allow for selection of positive clones by fluorescence-activated cell sorting (FACS), genes conferring increased tolerance to toxic compounds, and any gene resulting in a trait that allows selection of cells harbouring the selection marker. In other embodiments, the nucleic acid sequence A comprises at least one gene encoding a protein of interest, such as a mutant protein, a conditional mutant, a truncated mutant, a heterologous gene, a gene coding for a compound of interest to be expressed in a heterologous organism, a gene encoding a chimera protein or an artificial fusion protein, or any gene which it may be relevant to knock in.

In some embodiments, the nucleic acid sequence A comprises at least one selection marker SM and at least one other gene. Upon creation of a DSB by the endonuclease, the selection marker may be excised.

The CRESC may further comprise at least one other open reading frame which is not delimited by the two TES, i.e. the at least one other open reading frame is located upstream of the first TES or downstream of the second TES. In these embodiments, the at least one other open reading frame is not excised upon generation of a break by the endonuclease at the two TES.

In preferred embodiments, the at least one TES is two TES. Thus the endonuclease preferably creates two breaks, one at each TES. Repair of the breaks may be directed to the homologous recombination pathway if a donor DNA is provided. Donor DNA may comprise a gene of interest flanked by homology arms homologous to the regions surrounding the two breaks. In other words, the first homology arm may be homologous to the region immediately upstream of the first TES and the second homology arm may be homologous to the region immediately downstream of the second TES. The donor DNA may be comprised within a CRESC. In embodiments where donor DNA is not provided, repair may occur by non-homologous end joining, wherein the region immediately upstream of the first TES and the region immediately downstream of the second TES may be ligated. This may be accompanied by the formation of an indel.

Within the scope of the present invention are embodiments wherein the at least one open reading frame is at least two open reading frames, such as at least three open reading frames, such as at least four open reading frames, such as at least five open reading frames. The at least one TES is preferably two TES. The extent of the desired excision determines the position of the TES. Preferably, only one of the at least two open reading frames is a selection marker. The TES may be placed so that the excised nucleic acid sequence comprises a selection marker and another gene. This is particularly relevant in cases where transient expression is desirable.

Nucleic acid of interest

Herein are disclosed embodiments in which the CRESC comprises at least one nucleic

acid of interest, said at least one nucleic acid of interest being comprised in the nucleic acid delimited by the two. The at least one nucleic acid of interest may comprise a gene. It may further comprise untranslated regions, such as a 5'-UTR and/or a 3'-UTR. The 3'-UTR is not translated into a protein and may contain elements important for regulating the expression of a given gene; the regulation may be exerted at the transcriptional, translational and/or post-translational level. The 3'-UTR may comprise *cis*-acting and/or *trans*-acting regulatory elements such as terminators, polyadenylation signals, microRNA response elements, and AU-rich elements. The 5'-UTR (Untranslated Region) relates to the sequence giving rise to the 5'-leader or 5'-UTR of a transcript, i.e. the 5'-terminal part of the transcript located upstream the start codon. The 5'-UTR is not translated into a protein and may contain elements important for regulating the expression of a given gene; the regulation may be exerted at the transcriptional, translational and/or post-translational level. The 5'-UTR may comprise *cis*-acting and/or *trans*-acting regulatory elements such as promoters or enhancers.

The term 'gene' is to be understood in a broad meaning referring to a nucleic acid comprising a gene or a putative gene. The nucleic acid of interest may comprise a gene coding for a selection marker, a protein of interest, a peptide, or any gene which can be transcribed and/or translated, or it may result in the production of a miRNA, a siRNA, a tRNA.

Homology arms

In specific embodiments, at least one of the CRESCs further comprises two homology arms HA_L and HA_R, wherein one is 5'-terminal and the other is 3'-terminal, and wherein HA_L and HA_R are homologous to a target nucleic acid sequence, T_L, and a target nucleic acid sequence T_R, respectively, where T_L and T_R delimit a nucleic acid T. This is illustrated in figure 4a and 4b.

The homology arm HA_L is homologous to a target nucleic acid sequence T_L and the homology arm HA_R is homologous to a target nucleic acid sequence T_R. The homology arms are such that they allow homologous recombination between HA_L and T_L and between HA_R and T_R. The homology arms may be between 10 and 1000 bp long, such as between 20 and 800 bp, such as between 50 and 500 bp, such as between 200 and 500 bp, such as between 300 and 500 bp. Homology arms may help facilitate the correct integration of the CRESC into the genome, they are not

necessary. Homology arms may determine in which direction a CRESC is inserted.

Preferably, T_L and T_R are comprised within the same nucleic acid T, which they
delimit. T may be comprised within a genome, such as a chromosome or a plasmid.

5   The nucleic acid T may comprise at least one open reading frame, and/or a 5'-UTR
and/or a 3'-UTR. Integration of the CRESC and subsequent HR between HA_L and
T_L on one hand and between HA_R and T_R on the other hand results in excision of
the nucleic acid T. Thus expression of the gene encoded by the at least one open
reading frame comprised within T or regulated by the 3'-UTR or the 5'-UTR comprised

10   within T is inactivated, or results in expression of a mutated protein, such as a
truncation protein, a conditional mutant protein, a misfolded protein, or a mislocalised
protein.

In preferred embodiments, the nucleic acid T is comprised within a eukaryotic genome.

15   Examples of eukaryotic genomes include, but are not limited to: mammalian genomes,
including Chinese hamster ovary (CHO) genomes, human genomes, murine genomes;
unicellular genomes, including *Saccharomyces cerevisiae* genomes,
Schizosaccharomyces pombe genomes; avian genomes, such as chicken genomes. In
a preferred embodiment, the eukaryotic genome is the CHO genome. In other

20   embodiments, the CRESC does not comprise homology arms (figure 4c). Since the
CRESC comprises all the elements needed to be functional once knocked in the
genome, in some embodiments the direction of integration may not be essential and
homology arms may not be necessary.

25   Within the scope of the invention are CRESCs comprising, in a 5' to 3' direction: a
homology arm HA_L homologous to T_L; a gene of interest with its regulatory,
untranslated regions; a targeting sequence TES1; a selection marker; the targeting
sequence TES1; a homology arm HA_R homologous to T_R, in which T_L and T_R
define a nucleic acid T comprising at least part of an open reading frame or part of its

30   regulatory untranslated regions (figure 4a and 4b). Such a CRESC allows simultaneous
knockin of the gene of interest and selection marker and knockout of the nucleic acid T.

In some embodiments, the CRESC is expressed from a plasmid. In some
embodiments, the CRESC comprises a nucleic acid coding for a selection marker,

35   such as a fluorescent protein, an auxotrophy marker, a resistance marker, or any other

marker known in the art. In specific embodiments, the selection marker comprised within the CRESC is located outside the homology arms. This can be particularly interesting when the selection marker is a fluorescent protein, which can be expressed from the plasmid harbouring the CRESC. In such embodiments, integration of the

5   CRESC into the targeting sequence will result in excision and loss of the selection marker. Thus such a CRESC can be used as follows:

    i)   transfection of cells with a plasmid comprising a CRESC

    ii)   selection of transfected clones by positive selection for the marker comprised within the CRESC;

10   iii)   activation of the endonuclease and insertion of the CRESC at its target site; the selection marker, located outside the homology arms, is lost;

    iv)   selection of clones having undergone genome editing by negative selection for the marker comprised within the CRESC.

15   In such embodiments, the selection marker is not integrated into the target sequence. Only part of the CRESC is integrated into the target sequence. Consequently, the selection marker needs not be excised or counter-selected following gene editing.

In embodiments where the selection marker is a fluorescent protein, the clones of

20   interest may rapidly be selected by fluorescence-activated cell sorting (FACS) first by selecting fluorescing clones in step ii, and non-fluorescing clones in step iii.

In other embodiments, the selection marker is located between the homology arms. In such embodiments, integration of the CRESC into the targeting sequence will be

25   accompanied by integration and expression of the selection marker.

Plasmid

Also disclosed herein is a plasmid comprising at least one CRESC (figure 1). Suitable

30   plasmids are well known in the art and depend on the organism in which the CRESCs are to be integrated. The plasmid comprising the CRESCs may comprise a multiple cloning site. The plasmid may also comprise at least one restriction site outside the CRESC, for example the plasmid may comprise two restriction sites, wherein the first is upstream of the CRESC and the second is downstream of the CRESC. The at least

35   one restriction site allows for linearization of the plasmid, which may be religated. In

specific embodiments, the plasmid comprises two restriction sites. The two restriction sites may allow recognition and restriction by two different enzymes or by the same enzyme. In some embodiments, the ends of the linearised fragments are compatible. In other embodiments, the ends are incompatible. The plasmid may be used for cloning,
5   as is known in the art. Religation after digestion may be prevented by treatment with phosphatase. Preferably, the plasmid comprises two restriction sites allowing isolation of the at least one CRESC. In embodiments where the plasmid comprises more than one CRESC, the plasmid comprises restriction sites allowing isolation of each CRESC individually. For example, in a plasmid comprising two CRESCs, the plasmid comprises
10   at least three restriction sites for three different enzymes, and each CRESC can be isolated upon digestion with two of the three enzymes. The number and choice of restriction sites will be obvious to the skilled man depending on the number of CRESCs comprised within the plasmid. In some embodiments, the restriction site may be recognised by the same endonuclease that can recognise the TES.
15

The restriction sites may be designed so that new plasmids can easily be constructed from existing plasmids, wherein the gene of interest may be replaced by a new one by simple cloning, while the TES and optionally the homology arms remain in the plasmid. Thus such plasmids allow for easy construction of multiple vectors carrying different
20   CRESC constructs wherein only the gene of interest is varying.

Preferred embodiments of the CRESCs of the invention are listed below as illustrative examples, but should not be construed as limiting the scope of the invention.

25   A preferred embodiment relates to a CRESC1 comprised in a plasmid and comprising, from a 5' to 3' direction: i) a first restriction site RS1; ii) a left homology arm HA_L1; iii) one or more genes of interest which may comprise a 5'-UTR and a 3'-UTR; iv) a TES (TES1); v) one or more selection markers which may comprise a 5'-UTR and a 3'-UTR; vi) another TES (TES1); vii) a right homology arm HA_R1; viii) a second restriction site
30   RS2, which optionally may be identical to RS1. The 5'-UTR is in preferred embodiments an inducible promoter, the 3'-UTR may be a polyadenylation signal, the selection marker may be a fluorescent protein, the homology arms may be homologous to a nucleic acid comprised within the targeted nucleic acid and comprising a TES (TES2) (figure 4a).

35

Another preferred embodiment relates to a CRESC2 comprised in a plasmid and comprising, from a 5' to 3' direction: i) a first restriction site RS2; ii) a left homology arm HA_L2; iii) one or more genes of interest which may comprise a 5'-UTR and a 3'-UTR; iv) a TES (TES2); v) one or more selection markers which may comprise a 5'-UTR and

5   a 3'-UTR; vi) another TES (TES2); vii) a right homology arm HA_R2; viii) a second restriction site RS2, which may be identical to RS2. The 5'-UTR may be an inducible promoter, the 3'-UTR may be a polyadenylation signal, the selection marker may be a fluorescent protein, the homology arms may be homologous to a nucleic acid comprised within the targeted nucleic acid and comprising a TES (TES1) (figure 4b).

10   Another preferred embodiment relates to a CRESC comprised in a plasmid and comprising, from a 5' to 3' direction: i) a first restriction site RS1; ii) one or more genes of interest which may comprise a 5'-UTR and a 3'-UTR; iii) a TES (TES1); iv) one or more selection markers which may comprise a 5'-UTR and a 3'-UTR; v) another TES (TES1); vi) a second restriction site RS2, which may be identical to RS1. The 5'-UTR

15   may be an inducible promoter, the 3'-UTR may be a polyadenylation signal, the selection marker may be a fluorescent protein (figure 4c).

Endonucleases and means for targeting

In some embodiments, the cell into which the multiplex editing system is transfected is

20   capable of expressing an endonuclease or a variant thereof from a genomic location. Preferably, the endonuclease gene is stably integrated in the genome of the cell. By 'stably integrated' is understood that the integration is not spontaneously reversible, i.e. that the integration is stable for many generations. Alternatively, the endonuclease or variant thereof may be expressed from a plasmid. The plasmid may be the same as the

25   plasmid comprising the CRESCs. Preferably, the endonuclease gene or variant thereof is under the control of an inducible promoter, so that it is only expressed when multiplex editing is to be performed. In some embodiments, the nucleic acid encoding the endonuclease or variant thereof further comprises a nuclear localisation signal facilitating its import to the nucleus. The endonuclease gene or variant thereof may be

30   codon-optimised as known by the skilled person.

The endonucleases of the present system are selected from the group comprising Cas9, zinc finger nucleases (ZFNs) and Transcriptor-Activator Like Effector Nucleases (TALENs), or variants thereof. Variants thereof are functional homologues, functional

35   mutants, codon-optimised homologues, and any homologue capable of allowing

integration of a CRESC according to the invention.

In some embodiments, the endonuclease or variant thereof is selected from the group consisting of Cas9, a ZFN or a TALEN. The present system may further comprise

5   means for targeting the endonuclease or variant thereof to a TES. In other words, the endonuclease may require means for recognising a TES.

Cas 9 and variants thereof

In preferred embodiments, the endonuclease or variant thereof is Cas9 or a variant

10   thereof, and the targeting means are gRNAs that enable precise targeting of Cas9 to the TES. Cas9 is a CRISPR-associated nuclease originally discovered in *Streptococcus pyogenes*. Cas9 can form a complex with small RNAs as guides (gRNAs) to cleave DNA in a sequence-specific manner upstream of a protospacer adaptor motif (PAM), thus creating a double-strand break. The CRISPR-Cas9 system

15   further comprises guide RNAs known as the crRNA and tracrRNA. The gRNAs may be expressed from a plasmid. In some embodiments, the crRNA and the tracrRNA are expressed from a plasmid. In other, preferred embodiments, the crRNA and the tracrRNA are expressed from two different plasmids. In yet other embodiments, the crRNA and the tracrRNA are expressed from a genomic location. The crRNA and the

20   tracrRNA may be under the control of identical or different promoters.

Suitable variants of Cas9 include the D10A Cas9 mutant and the H40A Cas9 mutant. Either of the two point mutations D10A or H40A results in inactivation of the nuclease catalytic activity of Cas9, which is thereby converted to a nickase mutant, which

25   catalyses the formation of a single-stranded break at the target site, and thereby directs the repair pathway toward HR, resulting in fewer NHEJ-mediated repair events. Thus the use of nickase mutants is believed to lead to fewer off-target editing events.

In some embodiments, the Cas9 variant is a nickase mutant, i.e. a mutant capable of

30   generating a single-stranded break at the target site. In a preferred embodiment, the Cas9 variant is a D10A mutant. In another preferred embodiments, the Cas9 variant is a H40A mutant. In other embodiments, the Cas9 variant is a double D10A, H40A mutant.

35   Another suitable variant of Cas9 is a Cas9 protein or variant thereof tagged with a

fluorescent protein. The fluorescent protein may be any fluorescent protein known in the art, such as Green Fluorescent Protein (GFP), Yellow Fluorescent Protein (YFP), Red Fluorescent Protein (RFP), Cyan Fluorescent Protein (CFP), Tomato Fluorescent Protein, mCherry Fluorescent Protein, as well as enhanced versions thereof. The Fluorescent Protein may be codon-optimised for use in a particular host cell. It will be clear to the skilled person that any fluorescent protein which can be functionally fused to Cas9 or the variant thereof so that the resulting fluorescence can be measured when the fusion protein is expressed can be used. In preferred embodiments, Cas9 is functionally fused to GFP. In some embodiments, Cas9 is functionally fused to ZsGreen1. In some embodiments, Cas9 is functionally fused to YFP. In some embodiments, Cas9 is functionally fused to RFP. In some embodiments, Cas9 is functionally fused to mCherry.

In some embodiments of the invention, a variant of Cas9 is fused to a fluorescent protein as detailed above. Such a variant may be a nuclease mutant as described above, in particular a D10A mutant or a H40A mutant. Thus in some embodiments, D10A-Cas9 is functionally fused to GFP. In some embodiments, D10A-Cas9 is functionally fused to YFP. In some embodiments, D10A-Cas9 is functionally fused to RFP. In some embodiments, D10A-Cas9 is functionally fused to mCherry. In other embodiments, H40A-Cas9 is functionally fused to GFP. In some embodiments, H40A-Cas9 is functionally fused to YFP. In some embodiments, H40A-Cas9 is functionally fused to RFP. In some embodiments, H40A-Cas9 is functionally fused to mCherry.

The fluorescent protein may be fused to Cas9 at its N-terminus or at its C-terminus. It may also be fused internally.

A functional fusion is a fusion of two proteins such as Cas9 and a fluorescent protein which do not jeopardize the function of either protein. In other words, a functional Cas9 fusion to a fluorescent protein leads to a fusion protein comprising Cas9 and said fluorescent protein, where Cas9 is able to fold properly and has essentially the same activity as a non-fused protein, and where the fluorescent protein is able to emit fluorescence at detectable levels. A functional fusion protein may require that a linker is present between the two proteins to be fused. Suitable linkers are known in the art and comprise linkers such as glycine linkers and alanine linkers. The length of the linker may be from 1 to 20 amino acids, such as from 1 to 15 amino acids, such as from 1 to

10 amino acids, such as from 1 to 8 amino acids, such as from 1 to 6 amino acids, such as from a to 4 amino acids, such as from 1 to 2 amino acids. For example, the linker is 2 amino acids long. In some embodiments, the linker is a 2A linker. In a preferred embodiment, the functional fluorescently-tagged Cas9 protein is GFP-2A-

5   Cas9.

In specific embodiments, Cas9 further comprises at least one nuclear localisation signal ensuring that Cas9 is imported to the nucleus of the cell.

10   ZFNs and TALENS
In some embodiments, the endonuclease or variant thereof is a pair of zinc-finger nucleases (ZFNs) or variants thereof. ZFNs are artificial restriction enzymes generated by fusing a zinc finger DNA-binding domain to a DNA-cleavage domain. The non-specific cleavage domain from the type II restriction endonuclease FokI is typically

15   used as the cleavage domain in ZFNs. This cleavage domain must dimerize in order to cleave DNA and thus a pair of ZFNs is required to target non-palindromic DNA sites. In these embodiments, the targeting means are the DNA-binding domains of the individual ZFNs and typically contain between three and six individual zinc finger repeats which each recognize 3 base pairs, thus the targeting means typically

20   recognize between 9 and 18 base pairs. Thus in some embodiments of the invention is provided at least one pair of ZFNs which is capable of targeting at least one TES. Thus in one embodiment is provided a cell capable of expressing an endonuclease or variant thereof, preferably from a genomic location. In preferred embodiments, the endonuclease or variant thereof is a pair of ZFNs or codon-optimised variants thereof,

25   and is under the control of an inducible promoter. The pair of ZFNs is preferably expressed as proteins comprising a nuclear localisation signal. In preferred embodiments, the cell is a Chinese hamster ovary cell (CHO).

In other embodiments, the endonuclease or variant thereof is a pair of Transcription

30   Activator-Like Effector Nucleases (TALENs) or variants thereof. TALENs are artificial restriction enzymes inducing DSBs and generated by fusing a Transcription Activator-Like Effector (TALE) DNA binding domain to a DNA cleavage domain. The DNA cleavage domain is non-specific. In these embodiments, the targeting means is the DNA-binding domains of the individual TALENs which typically contain a repeated

35   highly conserved 33-34 amino acid sequence with the exception of the 12th and 13th

24

amino acids, which are highly variable and show a strong correlation with specific nucleotide recognition. Thus in some embodiments of the invention is provided at least one pair of TALENs which is capable of targeting at least one TES.

5   Suitable endonucleases or variants thereof are ZFNs or TALENs tagged with a fluorescent protein. The fluorescent protein may be any fluorescent protein known in the art, such as Green Fluorescent Protein (GFP), Yellow Fluorescent Protein (YFP), Red Fluorescent Protein (RFP), Cyan Fluorescent Protein (CFP), Tomato Fluorescent Protein, mCherry Fluorescent Protein, as well as enhanced versions thereof. The

10   Fluorescent Protein may be codon-optimised for use in a particular host cell. It will be clear to the skilled person that any fluorescent protein can be used, which can be functionally fused to the ZFN or the TALEN or the variant thereof so that the resulting fluorescence can be measured when the fusion protein is expressed. In some embodiments, a ZFN is functionally fused to GFP. In some embodiments, a ZFN is

15   functionally fused to ZsGreen1. In some embodiments, a ZFN is functionally fused to YFP. In some embodiments, a ZFN is functionally fused to RFP. In some embodiments, a ZFN is functionally fused to mCherry. In some embodiments, a TALEN is functionally fused to GFP. In some embodiments, a TALEN is functionally fused to ZsGreen1. In some embodiments, a TALEN is functionally fused to YFP. In some

20   embodiments, a TALEN is functionally fused to RFP. In some embodiments, a TALEN is functionally fused to mCherry.

Knockout

Disclosed herein is a system for knocking out at least one genomic sequence G by

25   using at least one CRESC as described above. The at least one CRESC comprises at least one targeting sequence TES; at least one sequence encoding a gene of interest, said at least one sequence being optionally surrounded by a 5'-UTR and/or a 3'-UTR sequence; and optionally, two homology arms HA_L and HA_R. In preferred embodiments, the at least one CRESC comprises two homology arms HA_L and

30   HA_R.

In some embodiments, the homology arms HA_L and HA_R, wherein one of HA_L and HA_R is 5'-terminal and the other of HA_L and HA_R is 3'-terminal relative to the CRESC, are homologous to a target nucleic acid sequence G_L and to a target nucleic

35   acid sequence G_R, respectively. G_L and G_R delimit the nucleic acid sequence G to

be knocked out. The nucleic acid G may comprise at least one open reading frame, and/or a 5'-UTR and/or a 3'-UTR. Integration of the CRESC and subsequent HR between HA_L and G_L on one hand and between HA_R and G_R on the other hand results in excision of the nucleic acid G. Thus expression of the gene encoded by the at least one open reading frame comprised within G or regulated by the 3'-UTR or the 5'-UTR comprised within G is inactivated, or results in expression of a mutated protein, such as a truncation protein, a conditional mutant protein, a misfolded protein, or a mislocalised protein.

Preferably, the nucleic acid sequence G is comprised within a eukaryotic genome. Examples of eukaryotic genomes include, but are not limited to: mammalian genomes, including Chinese hamster ovary (CHO) genomes, human genomes, murine genomes; unicellular genomes, including *Saccharomyces cerevisiae* genomes, *Schizosaccharomyces pombe* genomes; avian genomes, such as chicken genomes; preferably, the genome is the CHO genome.

Also disclosed herein is a system for knocking out a genomic nucleic acid sequence G comprising at least two genes, such as at least three genes, such as at least four genes. In such embodiments, the homology arms HA_L and HA_R are homologous to a target nucleic acid sequence G_L and G_R, respectively, wherein G_L and G_R delimit the nucleic acid sequence G.

Knockin
Also provided is a system for knocking in at least one gene of interest. The at least one CRESC comprises at least one targeting sequence TES; at least one sequence encoding a gene of interest, said at least one sequence being optionally surrounded by a 5'-UTR and/or a 3'-UTR sequence; and optionally, two homology arms HA_L and HA_R. In preferred embodiments, the at least one CRESC comprises two homology arms HA_L and HA_R.

According to the invention, the system allows knockin of a nucleic acid K, which comprises at least one gene of interest, and optionally may comprise untranslated 5' and 3' regions. The nucleic acid K may also comprise a selection marker. In some embodiments, two identical targeting sequences TES1 surround the selection marker, so that the first TES1 is located upstream of and the second TES1 is located

downstream of the selection marker. In such embodiments, homologous recombination between the two identical TES1 can lead to loss of the marker. Such homologous recombination events may be induced by methods known in the art, such as by chemicals, or by selective pressure in order to select for cells having lost the selection marker. Methods for such counter-selection of marker loss events are known in the art.

Thus the present invention also relates to a system which allows simultaneous knockin of a nucleic acid K and knockout of a nucleic acid G. Thus the present system allows allele replacement. The selection marker which may be comprised within the nucleic acid K may be excised by inducing HR between HA_L and HA_R.
In some embodiments, the CRESC is at least two CRESCs, such as at least three CRESCs, such as at least four CRESCs, such as at least five CRESCs.

Method for editing nucleic acids

The invention relates to a multiplex editing method, wherein 'editing' refers to knocking in and/or knocking out as described above with a system as described herein, said system comprising at least two Continuously Regenerated Endonuclease Site Cassettes CRESCs, wherein:

i.   a first CRESC comprises at least one first targeted endonuclease site (TES) and at least one nucleic acid sequence encoding a gene of interest; and

ii.   a second CRESC comprises at least one second targeted endonuclease site and at least one nucleic acid sequence encoding a gene of interest; and

iii.   optionally a subsequent CRESC comprises at least one targeted endonuclease site identical to the targeted endonuclease site of the penultimate CRESC.

The method comprises the steps of:

i.   introducing the first CRESC in a cell able to express at least one endonuclease and allowing one of the at least one endonuclease to create a break in a nucleic acid comprised within said cell; thereby allowing integration of the first CRESC in said nucleic acid;

ii.   introducing the second CRESC in said cell and allowing one of the at least one endonuclease to create a break in the first CRESC; thereby allowing integration of the second CRESC in the first CRESC;

iii.   optionally, introducing a subsequent CRESC in said cell and allowing one of

the at least one endonuclease to create a break in the previously integrated CRESC, thereby allowing integration of the subsequent CRESC in the previous CRESC.

5   Thus is provided a method for performing multiple knockins and/or knockout of genes. Thus the method described herein allows serial editing of nucleic acids with as little as two different TES. As explained above, using one TES only would result in uncontrolled, multiple integration events and is not desirable. For each additional editing step after steps i and ii, the sequence of the TES in each new CRESC is

10   determined by the sequence of the TES in the CRESC used two editing steps before (the penultimate TES). For example, if the TES of the first CRESC used in step i is TES1, and the TES of the second CRESC used in step ii is TES2, then the TES of the CRESC used in the subsequent steps will be either TES1 or TES2 but should not be the same as the TES used in the immediately preceding step. So the TES of the third,

15   fifth and all uneven subsequent CRESCs will be TES1, and the TES of the fourth, sixth and all even subsequent CRESCs will be TES2. Upon each integration, the TES of the previous CRESC is destroyed, while a new TES is integrated. The method is illustrated in figure 5.

20   Introduction of a CRESC within a targeted endonuclease site may occur as follows. If the CRESC does not comprise homology arms, its integration by NHEJ only requires creation of a break by the endonuclease at the TES. Thus in such embodiments any CRESC may integrate in any TES as long as it can be recognised by the endonuclease, which then creates a break such as a DSB in said TES. The CRESC

25   may integrate in any direction; statistically, half the integration events occur in one direction and the other half in the other direction. The advantages of such embodiments are their flexibility, the only requirement for the CRESC being that it should preferably comprise a TES different from the one it is to be integrated in, in order to allow for further editing steps. Moreover, the CRESC comprises all the

30   elements needed for its own expression, i.e. expression of the CRESC is not influenced by the direction of its integration.

If the CRESC comprises homology arms, its integration requires:
       -   that one or more breaks be created at one or more TES;

35   -   that one homology arm be homologous to the region upstream of the first TES

and the other homology arm be homologous to the region downstream of the last TES;

so that repair of the one or more breaks is directed toward homologous recombination wherein the CRESC is the donor DNA. In this case integration of the CRESC occurs essentially in one direction only, the direction being determined by the homology arms. Although the design and construction of such CRESCs may be more labor-intensive and time-consuming than CRESCs devoid of homology arms, they may limit off-target integration events.

Also provided herein are embodiments in which more than two TES are used. The invention may be adapted if e.g. a collection of CRESCs involving more than two TES is available.

Endonucleases suitable for performing the present method are capable of cleaving the DNA to create a nick or a double-strand break. The present method can be performed with at least two endonucleases, one for each TES to be recognised. Preferably, there are two endonucleases, each recognising one of the TES. Such endonucleases are Cas9, ZFNs and/or TALENs. Preferably, one of the endonucleases recognises the TES of the first CRESC (and the subsequent CRESCs comprising the same TES), and the other recognises the TES of the second CRESC (and the subsequent CRESCs comprising the same TES). By 'recognise' is understood that the endonuclease is capable of specifically binding to the TES and cleave the DNA at the recognition site.

The cell can further comprise means for targeting the endonuclease to a TES. Said means for targeting may be expressed from a plasmid or from a genomic location, from an inducible promoter or from a constitutive promoter. If Cas9 is the endonuclease, means for targeting Cas9 to the TES as described above are comprised within the cell. For example, guide RNAs (crRNA and tracrRNA) may be expressed from inducible promoters on a plasmid. One set of guide RNAs (i.e. crRNA and tracrRNA) is required for each TES to be recognised by Cas9. Thus if Cas9 is used for targeting both TES, two sets of gRNAs are needed. These two sets may be expressed from the same plasmid, where each set is under the control of a different promoter; or they may be expressed from different plasmids, preferably each set is under the control of a different promoter; or they may be expressed from a genomic location, preferably each set is under the control of a different promoter. If a pair of ZFNs or TALENs is used,

these are designed so that each pair can recognise one of the two TES. The two
endonucleases can be the of the same nature: Cas9 may be provided with different
sets of guide RNAs, wherein each recognises one TES; the ZFNs may be designed so
that each recognises one TES; the TALENs may be designed so that each recognises
one TES. Alternatively, the two endonucleases are of different nature: Cas9 with an
appropriate set of guiding RNAs may recognise one of the TES, while the other may be
recognised by a ZFN or a TALEN designed for that purpose; or one TES is recognised
by a ZFN and the other by a TALEN.

The at least two endonucleases may be expressed from the same plasmid and
preferably under the control of different inducible promoters. Alternatively, the at least
two endonucleases are expressed from a genomic location, preferably under the
control of different inducible promoters. Alternatively, one of the at least two
endonucleases is expressed from a genomic location and the other from a plasmid,
preferably under the control of different inducible promoters. In some embodiments, the
plasmids are counter-selectable. Preferably, expression of the endonuclease is
induced for short periods of time, i.e. the endonuclease is not expressed constitutively.
In some embodiments, expression of the endonuclease is induced from the time at
which the CRESC is introduced within the cell and expression is maintained for a
duration long enough to allow integration of the CRESC. Such durations may be from 1
hour to several days, for example two hours, for example four hours, for example six
hours, for example eight hours, for example ten hours, for examples twelve hours, for
example one day, for example two days, for example three days, for example seven
days.

The genes encoding the endonucleases may further comprise a nuclear localisation
signal to ensure that the endonucleases get imported into the nucleus.

The method described herein can thus be used for genome editing such as allele
replacement, sequential knockins of genes of interest, and/or knockouts of genomic
sequences.

Methods for introducing the CRESCs within cells are methods of transfection or
transformation well known in the art. These methods include, but are not limited to,
nucleofection, chemical-based transfection, electroporation, optical transformation,

sonoporation, protoplast fusion, particle-based methods such as gene gun, magnet-assisted transfection, or heat shock.

In a specific embodiment of the invention is provided a method in which an
5   endonuclease creates a break at a first target site TES1 on a genome (figure 5a). The presence of homology arms in a given CRESC can direct the integration of the CRESC or at least a part thereof to integration by homologous recombination. Transfection of a first CRESC1 comprising a homology arm HA_L1 (homologous to a nucleic acid upstream of the genomic target site TES1), a nucleic acid of interest comprising a gene (GOI1), a target site TES2, a marker gene, another second target site TES2, and a
10   homology arm HA_R1 (homologous to a nucleic acid downstream of the genomic target site TES1). Upon recombination between the homology arms and their homologous genomic sequences, TES1 is excised and CRESC1 is integrated. The genome now contains two target sites TES2, surrounding a selection marker. The cell comprising this genome gives rise to a first cell line. A break may now be created by an
15   endonuclease at the at least one target site TES2 and a second CRESC2 may be provided, comprising a homology arm HA_L2 (homologous to a nucleic acid upstream the first TES1, e.g. homologous to the 3'-terminal part of GOI1), a nucleic acid of interest comprising a gene (GOI2), a target site TES2, a marker gene, another target site TES2, and a homology arm HA_R2 (homologous to a nucleic acid downstream of
20   the second TES1; HA_R2 may be the same as HA_R1, as shown in figure 5b). Upon recombination between the homology arms and their homologous genomic sequences, the sequence comprised between the two TES1 is excised and replaced by CRESC2. The resulting genomic sequence (shown in figure 5b) contains two newly integrated genes and a selection marker. A break may now be created by an endonuclease at the
25   at least one target site T1 (figure 5c). The endonuclease used in this step may be the same as the endonuclease used in the first step (5a). A third CRESC3 may be provided, comprising a homology arm HA_L3 (homologous to a nucleic acid upstream the first TES1, e.g. homologous to the 3'-terminal part of GOI2), a nucleic acid of interest comprising a gene (GOI3), a target site TES1, a marker gene (which may be
30   identical to the marker gene used in the first step), another target site TES1, and a homology arm HA_R3 (homologous to a nucleic acid downstream of the second TES2; HA_R3 may be the same as HA_R2 or HA_R1, as shown in figure 5c). Thus the third CRESC3 may be identical to CRESC1, with the exception of the gene of interest and
35   the left homology arm HA_L3. These editing steps can be repeated as many times as

necessary to obtain the desired cell line. In some embodiments, at least one of the CRESCs does not comprise homology arms.

In particular embodiments, the method is performed with Cas9 or a variant thereof as described above.

Method for fast generation of engineered cell lines

The present invention also provides for a method for fast generation of engineered cell lines.

It is an object of the invention to provide for a method for gene editing as described above, wherein:

     i.   the at least one endonuclease is tagged with a first fluorescent protein;

     ii.   at least one of said CRESCs further comprises two homology arms HA_L and HA_R, wherein one is 5'-terminal and the other is 3'-terminal, and

     iii.   wherein HA_L and HA_R are homologous to a target nucleic acid sequence, T_L, and a target nucleic acid sequence T_R, respectively, where T_L and T_R delimit a target nucleic acid T; and

     iv.   wherein said at least one of said CRESCs functionally expresses a second fluorescent protein from a nucleic acid sequence located outside said homology arms HA_L and HA_R,

said method further comprising the steps of:

     v.   positively selecting for cells functionally expressing said first and second fluorescent protein prior to the steps of allowing said at least one endonuclease to create a double-strand break;

     vi.   negatively selecting for cells expressing said second fluorescent protein subsequent to the steps of allowing said at least one endonuclease to create a break.

It is a further object of the invention to provide for a method for gene editing as described above, wherein:

     i.   the at least one endonuclease is tagged with a first fluorescent protein;

     ii.   at least one of said CRESCs further comprises two homology arms HA_L

and HA_R, wherein one is 5'-terminal and the other is 3'-terminal, and

iii.   wherein HA_L and HA_R are homologous to a target nucleic acid

sequence, T_L, and a target nucleic acid sequence T_R, respectively,

where T_L and T_R delimit a target nucleic acid T; and

iv.   wherein said at least one of said CRESCs functionally expresses a second

fluorescent protein from a nucleic acid sequence located inside said

homology arms HA_L and HA_R,

said method further comprising the steps of:

v.   positively selecting for cells functionally expressing said first and second

fluorescent protein prior to the steps of allowing said at least one

endonuclease to create a double-strand break;

vi.   positively selecting for cells expressing said second fluorescent protein

subsequent to the steps of allowing said at least one endonuclease to

create a double-strand break;

vii.   optionally, excising the nucleic acid sequence encoding said second

fluorescent protein from the target nucleic acid sequence.

Negative and/or positive selection of cells emitting said first and/or said second

fluorescent signal can be achieved by methods known in the art. In a specific

embodiment, the selection is performed using fluorescence-activated cell sorting

(FACS).

In one embodiment, the system of the invention comprises a fluorescently-tagged

endonuclease, such as fluorescently-tagged Cas9, ZFN or TALEN as described above.

Methods for positive selection of clones functionally expressing said fluorescently-

tagged endonuclease are known in the art and can be advantageously used in order to

enrich for cells functionally expressing said fluorescently-tagged endonuclease. Said

endonuclease can be expressed from a plasmid or from a genomic location. In a

preferred embodiment, said endonuclease is GFP-2A-Cas9.

In an advantageous embodiment, is provided a method for editing nucleic acids with

the system described herein, comprising the steps of:

i.   introducing an endonuclease functionally fused to a first fluorescent

protein in a cell;

ii.   positively selecting for cells emitting a first fluorescent signal;

iii.   introducing a first CRESC in said cell;

iv.   allowing said endonuclease to create a break in a nucleic acid
         comprised within said cell; thereby allowing integration of the first
         CRESC in said nucleic acid.

The first CRESC may comprise a fluorescent gene coding for a second fluorescent protein, said second fluorescent protein being different from said first fluorescent signal. Preferably, the first and second fluorescent proteins are chosen so that they emit fluorescent signals at different wavelengths, so that it is possible to discriminate between cells which are positive for only one of the two fluorescent signals and cells which are positive for both fluorescent signals.

The nucleic acid encoding for the second fluorescent protein preferably leads to expression of the second fluorescent protein.

In some embodiments, the nucleic acid encoding the second fluorescent protein is comprised within the region defined by the homology arms, In this case, the method comprises the steps of:

i.   introducing an endonuclease functionally fused to a first fluorescent
         protein in a cell;

ii.   positively selecting for cells emitting a first fluorescent signal;

iii.   introducing a first CRESC in said cell, said CRESC comprising
         homology arms as described above and a nucleic acid coding for a
         second fluorescent protein, said nucleic acid being located within the
         region defined by the homology arms.

iv.   If the second fluorescent protein can be functionally expressed from
         said CRESC, positively selecting for cells emitting a second
         fluorescent signal;

v.   allowing said endonuclease to create a break in a nucleic acid
         comprised within said cell; thereby allowing integration of at least part
         of the first CRESC in said nucleic acid. The nucleic acid encoding the
         second fluorescent protein is integrated in the target nucleic acid;

vi.   positively selecting cells emitting the first fluorescent signal and

positively selecting cells emitting the second fluorescent signal.

It will be understood that if the CRESC is on a vector, the term "region defined by the homology arms" refers to the region of the CRESC which is integrated into the target nucleic acid. In such embodiments, the gene encoding the second fluorescent protein is not integrated in the target nucleic acid.

In one embodiment, the CRESC comprises a nucleic acid coding for a second fluorescent protein, said nucleic acid being located outside the region encompassed between the homology arms, where said region is to be integrated in the target sequence. In such embodiments is provided a method for editing nucleic acids with the system described herein, comprising the steps of:

i.   introducing an endonuclease functionally fused to a first fluorescent protein in a cell;

ii. positively selecting for cells emitting a first fluorescent signal;

iii. introducing a first CRESC in said cell, said CRESC comprising homology arms as described above and a nucleic acid coding for a second fluorescent protein, said nucleic acid being located outside the region defined by the homology arms.

iv. If the second fluorescent protein can be functionally expressed from said CRESC, positively selecting for cells emitting a second fluorescent signal;

v.   allowing said endonuclease to create a break in a nucleic acid comprised within said cell; thereby allowing integration of at least part of the first CRESC in said nucleic acid. The nucleic acid encoding the second fluorescent protein is not integrated in the target nucleic acid;

vi. positively selecting cells emitting the first fluorescent signal and negatively selecting cells emitting the second fluorescent signal. Said cells are expected to be true positives, i.e. recombinant cells where the desired editing events have taken place.

It will be understood that if the CRESC is on a vector, the term "region defined by the homology arms" refers to the region of the CRESC which is integrated into the target nucleic acid. In such embodiments, the gene encoding the second fluorescent protein

is not integrated in the target nucleic acid, but is instead excised and subsequently degraded. Thus the cells in which the desired editing events have taken place no longer emit the second fluorescent signal.

5   Such embodiments have the advantage that no nucleic acid sequence coding for a selection marker is integrated in the target nucleic acid. Thus, the fluorescent markers can be recycled and used repeatedly for serial rounds of genome editing.

Positive and negative selection of cells emitting the first and/or the second fluorescent
10   signal can easily be achieved by methods in the art. In particular, fluorescence-activated cell sorting is a preferred, but not the only, method for rapidly sorting cells with specific fluorescent characteristics.

Particularly preferred embodiments relate to a method comprising the steps of:
15   i.   introducing Cas9 or a variant thereof functionally fused to a first fluorescent protein such as GFP in a cell;
       ii.   positively selecting for cells emitting a first fluorescent signal such as a green fluorescent signal;
       iii.   introducing a first CRESC in said cell, said CRESC comprising
20   homology arms as described above and a nucleic acid coding for a second fluorescent protein such as mCherry, said nucleic acid being located within the region defined by the homology arms.
       iv.   If the second fluorescent protein can be functionally expressed from said CRESC, positively selecting for cells emitting a second
25   fluorescent signal such as an mCherry signal;
       v.   allowing said endonuclease to create a break in a nucleic acid comprised within said cell; thereby allowing integration of at least part of the first CRESC in said nucleic acid. The nucleic acid encoding the second fluorescent protein, e.g. mCherry, is integrated in the target
30   nucleic acid;
       vi.   positively selecting cells emitting the first fluorescent signal such as a green fluorescent signal and positively selecting cells emitting the second fluorescent signal such as an mCherry fluorescent signal.

35   In another preferred embodiment, the method comprises the steps of:

i.   introducing Cas9 functionally fused to a first fluorescent protein such as GFP in a cell;

ii.   positively selecting for cells emitting a first fluorescent signal such as a green fluorescent signal;

5   iii.   introducing a first CRESC in said cell, said CRESC comprising homology arms as described above and a nucleic acid coding for a second fluorescent protein such as mCherry, said nucleic acid being located outside the region defined by the homology arms.

iv.   If the second fluorescent protein such as mCherry can be functionally
10   expressed from said CRESC, positively selecting for cells emitting a second fluorescent signal such as mCherry;

v.   allowing said endonuclease to create a break in a nucleic acid comprised within said cell; thereby allowing integration of at least part of the first CRESC in said nucleic acid. The nucleic acid encoding the
15   second fluorescent protein such as mCherry is not integrated in the target nucleic acid;

vi.   positively selecting cells emitting the first fluorescent signal such as green fluorescent signal and negatively selecting cells emitting the second fluorescent signal such as mCherry signal. Said cells are
20   expected to be true positives, i.e. recombinant cells where the desired editing events have taken place.

The above methods involving selection of cells positively or negatively emitting a first and/or a second fluorescent signal can lead to significant enrichment of cells with
25   desirable characteristics. Such methods can be used to select for cells with a frequency of indel occurrence higher than a given threshold. In some embodiments, indels in a cell population selected with one of the above methods is greater than 10%, such as greater than 15%, such as greater than 20%, such as greater than 25%, such as greater than 30%, such as greater than 35%, such as greater than 40%, such as
30   greater than 45%, such as greater than 50%, such as greater than 55%, such as greater than 60%, such as greater than 65%, such as greater than 70%, such as greater than 75%, such as greater than 80%, such as greater than 85%, such as greater than 90%, such as greater than 95%, such as 100%. Thus in particular embodiments, a fraction of the cell population has undergone at least one gene editing
35   event, said fraction being between 10 and 100%, such as between 10% and 20%, such

as between 20% and 30%, such as between 30% and 40%, such as between 40% and 50%, such as between 50% and 60%, such as between 60% and 70%, such as between 70% and 80%, such as between 80% and 90%, such as between 90% and 100%.

In specific embodiments, the endonuclease is stably expressed from a genomic location. In particular embodiments, the endonuclease is Cas9. The above methods can further comprise a step of excising the stably integrated endonuclease such as Cas9. This can be performed by methods known in the art. For example, Cas9 can be excised using Cas9 itself with means for targeting Cas9. The Cas9 genomic locus can thus undergo self-destruction, generating a clean cell line not expressing Cas9.

In some embodiments, the first CRESC comprises a recognition site for a second CRESC, which in turn comprises a recognition site for a first CRESC, thereby allowing recycling of integration sites as described herein.

Cell

It is an object of the present invention to provide a cell capable of expressing an endonuclease or variant thereof, preferably from a genomic location. The cell may preferably further comprise means for targeting the endonuclease to a specific sequence such as a TES. In preferred embodiments, the endonuclease or variant thereof is Cas9 or a codon-optimised variant thereof, stably integrated in the genome, optionally under the control of an inducible promoter. In other embodiments, the endonuclease or variant thereof comprises a nuclear localisation signal (NLS). In preferred embodiments, the cell is a Chinese hamster ovary cell (CHO). In other embodiments, the variant of Cas9 is a mutant favouring homologous recombination over non-homologous end-joining, such as a nickase mutant selected from the group comprising the D10A and the H840A mutants.

The cell can also further comprise means for targeting the at least one endonuclease to a targeted endonuclease site.

Thus it is an object of the present invention to provide a cell capable of expressing an endonuclease or variant thereof, preferably from a genomic location. In specific embodiments, the endonuclease or variant thereof is a pair of TALENs or codon-

optimised variants thereof, and is under the control of an inducible promoter. The pair of TALENs is preferably expressed as proteins comprising a nuclear localisation signal. In preferred embodiments, the cell is a Chinese hamster ovary cell (CHO).

5   Such a cell can be useful for multiplex editing. For example, in embodiments where the endonuclease is Cas9 or a variant thereof expressed from a genomic location, the cell can be transformed with plasmids comprising guide RNAs suited for inducing a break at a given genomic location. Said break may lead to formation of an indel. Such a cell thus allows easy knockout without having to provide homology arms. This can be
10   particularly advantageous for cells such as mammalian cells, where DNA repair occurs predominantly via NHEJ. Sequential transfection of the cell with various plasmids comprising different sets of guide RNAs thus allows multiple knockouts to be performed in a convenient manner.

15   In other embodiments, a cell according to the invention may be used for knocking in sequences of interest. Upon transfection with guide RNAs a break is created at a given genomic location. Providing a donor DNA may result in insertion of a new sequence at this location. Cells in which HR is predominant over NHEJ are well suited for performing knockins in this way. For example, this method may be advantageous in
20   eukaryotic cells including, but not limited to, *Saccharomyces cerevisiae*. Alternatively, mutants of mammalian cells favouring HR over NHEJ may be used. Examples of such mutants of Cas9 are the D10A and the H840A mutants.

A cell according to the invention may be transfected with a plasmid comprising a
25   CRESC as described and can thus be used in a method for multiplex editing

Kit of parts
The invention further relates to a kit of parts, comprising at least two pluralities of CRESCs, wherein:
30   i.   a first plurality of CRESCs comprises at least one first targeted
                     endonuclease site; and
                ii. a second plurality of CRESCs comprises at least one second targeted
                     endonuclease site,
                wherein the first and second targeted endonuclease sites are different.

35

The kit may further comprise at least one endonuclease which is capable of recognising at least one of the at least two targeted endonuclease sites. The endonuclease may be comprised on a plasmid or within the genome of a cell, and may be under the control of an inducible promoter, as described above. The kit may also comprise means for allowing the at least one endonuclease to recognise the at least two endonuclease sites. For example, if the kit comprises Cas9, which is capable of recognising at least one of the TES, the kit may further comprise guiding RNAs, for example on a plasmid, as described above.

The kit can further comprise a nucleic acid sequence encoding a functional endonuclease capable of recognising at least one of said first and second targeted endonuclease sites.

In some embodiments, the kit comprises a functional endonuclease which is Cas9 or a variant thereof, said variant being selected from the group consisting of fluorescently-tagged Cas9, the D10A Cas9 mutant and the H840A Cas9 mutant.

The kit may also comprise a cell as described above.

Examples

Example 1. Introduction to the Continuously Regenerated Endonuclease Site Cassette (CRESC) system

The ability to insert genes into an organism is of paramount importance to the biotech and biopharmaceutical industries. The purpose is often to stably produce a compound (e.g. a protein) of interest in high quantities. Currently common methods rely on random insertions of a single gene constructs followed by labor intensive screening of single cell clones. If multiple gene inserts are desired, the whole process must be repeated.

The CRESC system is a targeting construct that facilitate one or more rounds of targeted insertion of desired genes into a specific site of a genome that has been preselected to yield highly producing cells. To facilitate this, the CRESC donor DNA will be co-transfected with a targeted nuclease system (TNS) (e.g. Cas9, Zinc finger nucleases (ZFNs), TALE nucleases (TALENS)) that will create a break such as a double stranded break (DSB) at the preselected target site which will engage cellular

repair systems that then insert the CRESC DNA. This greatly reduces the need to screen for stable high producers.

The most labor intensive part of inserting a gene at a specific site is discovering and validating a target area for the TNS to target. The genomic area must facilitate high expression of the inserted gene, but also contain a target sequence that the TNS can target with high specificity and efficacy. With the CRESC system this needs only be done once and can then be reused for multiple genes or new product producing cell lines.

In the process of inserting any DNA at a target site, the target site is destroyed preventing further use of this desirable genomic insertion site. The CRESC DNA therefore carries a new insert site to re-enable the use of that desirable genomic site. The CRESC DNA will carry one of two target sequences for the TNS. So if CRESC A is inserted into target sequence 1, then CRESC A carries target sequence 2. Later CRESC B is inserted into target sequence 2 and carries target sequence 1. This prevents the CRESC from being integrated multiple times and requires only the discovery of two target sequence and accompanying TNS.

Example 2. Introduction to Cas9 cell line

This invention is a mammalian cell line (CHO-S) that has an inducible modified Cas9 permanently integrated into its genome which will facilitate quick, easy and low cost genomic modifications (deletions, insertions, mutations). The Cas9 gene is regulated by inducible promoter and can, if desired after a specific genotype is obtained, be removed by targeting itself. The cell line enables in an easy and high throughput manner genome modification by simply transfecting one or more guide RNA molecules that targets a desired region of the genome which will induce double stranded breaks (DSB's) which depending upon the experimental setup can result in one or more of the afore mentioned genomic modifications. The method can run in multiplex by using more than one guide RNA in one step to target multiple loci in the genome.

By combining the Cas9 cell line with the CRESC system, all that is needed to create a stable high producing mammalian cell line is a gene of interest and a small guide RNA.

41

Example 3. Experiments for CRESC.

1. Construct creation

    a. We are using the USER cloning system to assemble CRESC's. A CRESC is a plasmid that consist of the following components (see figure 2 below):

        i. A plasmid backbone for propagation and selection in bacteria

        ii. One of two (the other is *xiii*) identical restriction sites (RS) for cutting out the part to be integrated into the genome (could be the same as *vii*)

        iii. An optional homology arm (HA) for homologous recombination (HR) rather than non-homologous end joining (NHEJ)

        iv. A 5' UTR for *v* (e.g. a promoter)

        v. A gene of interest (GOI)

        vi. A 3' UTR for *v* (e.g. poly A signal)

        vii. One of two (the other is *xi*) targeted endonuclease sites (TES) for removing selection marker (SM) and create a double stranded break (DSB) to insert the next CRESC into.

        viii. A 5' UTR for *ix* (e.g. a promoter)

        ix. A SM (e.g. a fluorescent protein or an antibiotic resistance gene)

        x. A 3' UTR for *ix* (e.g. poly A signal)

        xi. See *vii*

        xii. See *iii*

        xiii. See *ii*

    b. One CRESC can contain multiple *iv-v-vi* and/or *viii-ix-x* parts. This would enable the user to insert several genes at once and/or several selection markers.

    c. For experiment part 2 we are constructing a number of different CRESC with a fluorescent protein (FP) (e.g. GFP, mCherry, CFP, YFP, mBlueberry, etc.) for SM and a different FP for GOI.

    d. For experiment part 3 we are using the same constructs as those made in c

    e. For experiment part 4 we are constructing several CRESC's each with an FP as SM and a "housekeeping gene" (e.g. GADPH, B2M, etc.) because they are easy/cheap to detect/quantify with RT-qPCR and not likely to cause disturbances in the cells.

2. Transient expression of a number of constructs to determine transient lifetime of FP's and verify expression from both GOI and SM. Construct describe in Experiment part 1,C

   a. The purpose of using an FP as SM is to enable FACS sorting rather than having to rely on antibiotics for selection.

   b. The purpose of determining the transient lifetime is to determine the time before we can confidently state that the majority/all of the SM expression is from permanently integrated CRESC's.

   c. Based on transient lifetimes we select two FP's with short transient life time and good spectral separability.

3. Permanent integration of two constructs. Constructs described in Experiment part 1, D

   a. First one CRESC with an FP (FP1) as GOI and another FP (FP2) as SM. Then we insert a second CRESC into the insert site of the first CRESC with FP3 as GOI and FP4 as SM.

   b. This is done using information from experiment part 2 to determine time from transfection until we can FACS sort.

   c. The insert site is the COSMC gene

   d. The insert method will be NHEJ for constructs without HA and HR for constructs with HA.

4. Multiple insertions with SM altering between FP1 and FP2 and the GOI being different genes that we will detect with RT-qPCR. Constructs described in Experiment part 1,E

   a. This is the final test which proves the system is functional.

Example 4. Design and use of CHO optimized Cas9 and guideRNA to knockout Fut8 and COSMC.

Bacterial cultures and Media

The strain used for cloning was the commercial available E.coli Dh5alfa strain from Invitrogen (#18265-017).

43

It was routinely grown in LB medium at 37°C, 250rpm with the appropriate antibiotic if necessary (100µg/µL Ampicillin and 50µg/µL Kanamycin). After heat shock cells were recovered in SOC rich medium.

5   Plasmid construction and gRNA target design
Cas9 (csn1) gene was directly ordered codon optimized and with the addition of the 3' NLS sequence from DNA 2.0 and it was provided in pJ607 plasmid.

The plasmid was transform in E.coli Dh5 alfa chemically competent cells according to
10   the standard procedure. The subsequent day the Ampicillin positive clones were picked and put in 25 mL of LB supplemented with 100 µg/µL Amp and let them grow ON at 37 C°. Finally the plasmid was Midiprep with the kit NucleoBond® Xtra Midi / Maxi EF (Micherey-Nagel).

15   The target RNA was designed according to Martin Jinek et al. tracrRNA and crRNA chimera by using U6 promoter (Chang at al., 2013). The entire gRNA sequence was synthetized by IDT and directly cloned in the high copy number pRSF-duet vector (Novagen) using KpnI and HindIII. The pRSF-duet :: gRNA plasmids were transforms in E.coli NEB Turbo competent cells (NEB biolabs) according to the standard
20   procedures. Transformant clones were selected in 50 µg/mL Kanamycin LB plates.

U6 promoter (SEQ ID NO: 4)	AAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCT TCATATTTGCATATACGATACAAGGCTGTTAGAGAGATA ATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTA CAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAG TTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATA TGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTT ATATATCTTGTGGAAAGGACGAAACACC
Scaffold (SEQ ID NO: 5)	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG
Terminator (SEQ ID NO: 6)	CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG CTTTTTTT
gRNA	GAAAAGTGTCCTGAACAAGGT

44

Target (SEQ ID NO: 7)	

Transfection of Cas9 and transient gRNA in CHO cell line

Both Cas9 and gRNA expression constructs were transfected simultaneously into CHO-S cells using a Lonza Nuclefector following manufacturers protocol.

Validation of COSMC and Fut8 Knock outs.

Genomic DNA prepaparation: After 1 week transfected CHO cells were harvested.

Cell pellet of CHO-S KO pool for the 4 knocked out exon (1,5,7,8) and 4 Cosmc targets was first lysed by using 50 µL of QuickExtract (Epicenter # QE09050), incubating first at 65 °C for 20 min an then at 98 °C for 5 min. The lysate was kept on ice or frozen at -20 °C for storage.

Locus PCR amplification:

The single targets locus for Fut8 and Cosmc were PCR amplified with Dream Taq (from Thermoscientific-Fermentas).

Products sizes and primers sequence:

Fut8

Exon1 Primers: 1864 + 1865 = 412bp

Exon5 Primers: 1868 + 1869 = 386bp

Exon7 Primers: 1878 + 1879 = 278bp

Exon8 Primers: 1884 + 1885 = 269bp

Primer 1864 (SEQ ID NO: 8): AGGCCCTATTGATCAGGGGA

Primer 1865 (SEQ ID NO: 9): TGGAAGCCCAAATGAAGCAC

Primer 1868 (SEQ ID NO: 10): GGTCGAGCTCCCCATTGTAG

Primer 1869 (SEQ ID NO: 11): GCTCTGCTGCCCTAACTGAA

Primer 1878 (SEQ ID NO: 12): GCCCCCATGACTAGGGATA

Primer 1879 (SEQ ID NO: 13): CCCATACAGAACCACTTGTTG

Primer 1884 (SEQ ID NO: 14): CCCAGAGTCCATGTCAGACG

Primer 1885 (SEQ ID NO:15): GCAACAAGAACCACAAGTTCCC

Cosmc:

GR1, GR2, GR4, GR5 has one set of primers:

Expected size of 318 bp

COSMC primer forward (SEQ ID NO: 16): GGATCCATCGCAGCCTTTCT

COSMC primer reverse (SEQ ID NO: 17): AACCACCCGAACCAGGTAGT

5   The PCR is checked on 1% agarose gel and used for T7 assay and TOPO cloning.

T7 assay

The purpose of this assay is to confirm the presence of small mutations (indels) at the
target site. By melting the PCR products and slowly re-annealing the single DNA

10   strands from the PCR product can re-anneal as wildtype/wildtype, wildtype/mutant or
mutant/mutant. Commonly there is an excess of wildtype, so only very little
mutant/mutant is formed. The re-annealed PCR products are then mixed with the T7
enzyme that recognizes mismatched (e.g. wildtype/mutant) and cuts the product at the
mismatch site which is present in the target site. This can then be visualized and

15   quantified on a gel.

After Re-annealing PCR, add 1 uL T7 enzyme buffer mixture (NEB, cat. M0302L).
Incubate at 37°C for 1 hour and then run on 6% TBE gel.

20   TOPO cloning

The TOPO cloning is performed according to the protocol instructions (Invitrogen
#450030).

So 4 µL of freshly amplified DNA with Dream Taq were added to the Mix:

Reagent	Volume
Salt Solution	1 µL
TOPO vector	1 µL
Final Volume	6 µL

25

The mix is incubated at 22°C for 20 min and then 2 µL of it is transformed in E.coli
Dh5alfa chemically competent cells (Invitrogen #18265-017) according to the
established protocol. After recovery the cells were plated on 50 µg/mL Kanamycin
plates and incubated overnight at 37°C.

30   The day after 12 clones each samples (4 Fut8 and 4 Cosmic) are picked and put in 96
deep wells plates to be incubated in LB at 37°C overnight. The all overnight cultures in

96 wells plates were eventually miniprepped (Nucleospin quick pure) and prepared for sequencing which performed on a sanger sequencer.

This demonstrates that the Cas9+gRNA system works in CHO cells.

Example 5. Stable cell line with inducible Cas9

Bacterial cultures and media as well as gRNA contruction and transfection and verification of indel creation is done in the same way as example 1.

The gene sequence encoding CHO optimized Cas9 is inserted into pcDNA4/TO (invitrogen) generating the plasmid pcDNA4/TO-cas9. A CHO-S cell line expressing the CHO optimized cas9 in a doxycycline responsive manner is isolated by co-transfecting CHO-S cells with pcDNA6/TR and pcDNA4/TO-cas9 constructs and selecting productively transfected cells with zeocin and blasticidin. In a parallel approach, the gene sequence encoding CHO optimized Cas9 is cloned into pTRE3G from Clontech in addition to zsGreen1 in order to prepare a cell line expressing both Cas9 and a green reporter protein from a bidirectional tet-express inducible promoter. A stable tet-express inducible cell line will be prepared by co-transfecting the pTRE3G-Cas9-zsGreen1 with a hygromycin or puromycin marker.

Testing stable cell line

By using the same guideRNAs from example 4, the function of the stable Cas9 CHO cells is verified by repeating the validation experiments in example 4.

This demonstrates that a CHO cell with a permanently integrated inducible Cas9 can be used to easily and cheaply manipulate the genome of CHO cells.

Example 6. Accelerating genome editing in CHO cells using CRISPR Cas9; materials and methods.

Plasmid construction and sgRNA target design

The Cas9 sequence from the S. pyogenes strain M1 GAS genome with a 3' nuclear localization signal was codon-optimized for CHO cells (SEQ ID NO. 2), synthesized and subcloned into the mammalian expression vector pJ607-03 (DNA 2.0). The plasmid was then transformed into DH5α subcloning cells (Invitrogen). Transformant clones were selected on 100 μg/mL Ampicillin (Sigma-Aldrich). The chosen sgRNA target sequences are listed in Table 1.

Table 1: sgRNA genomic target sequences

Target gene	sgRNA name (SEQ ID NO: 18)	GNNNNNNNNNNNNNNNNNNNGG
COSMC	sgRNA1_C (SEQ ID NO: 19)	GAAAAGTGTCCTGAACAAGGTGG
	sgRNA2_C (SEQ ID NO: 20)	GAATATGTGAGTGTGGATGGAGG
	sgRNA3_C (SEQ ID NO: 21)	GAAATATGCTGGAGTATTTGCGG
	sgRNA4_C (SEQ ID NO: 22)	GCAGTCTGCCTGAAATATGCTGG
FUT8	sgRNA1_F (SEQ ID NO: 23)	GAAAGGATCATGAAATCTTAAGG
	sgRNA2_F (SEQ ID NO: 24)	GATCCGTCCACAACCTTGGCTGG
	sgRNA3_F (SEQ ID NO: 25)	GTCAGACGCACTGACAAAGTGGG
	sgRNA4_F (SEQ ID NO: 26)	GGATAAAAAAGAGTGTATCTGG

The sgRNA expression constructs were designed by fusing tracrRNA and crRNA into a
chimeric sgRNA (Jinek et al., 2012) and located immediately downstream of a U6
promoter (Chang et al., 2013). The sequences of the U6 promoter, scaffold and
terminator are shown in Supplementary Materials and Methods. Initially, the sgRNA
expression cassette was synthesized as a gBlock (Integrated DNA Technologies) and
subcloned into the pRSFDuet-1 vector (Novagen, Merck) using KpnI and HindIII
restriction sites. This pRSFDuet-1/sgRNA expression vector was used as backbone in
a PCR-based uracil specific excision reagent (USER) cloning method. This method
was designed to easily and rapidly change the 19 bp-long variable region (N19) of the
sgRNA in order to generate our sgRNA constructs. From the pRSFDuet-1/sgRNA
expression vector, a 4221 bp-long amplicon (expression vector backbone) was
generated by PCR (1x: 98°C for 2 min; 30x: 98°C for 10 s, 57°C for 30 s, 72°C for 4
min 12 s; 1x: 72°C for 5 min) using two uracil-containing primers (sgRNA Backbone_fw
and sgRNA Backbone_rv, Integrated DNA Technologies, Table 2) and the X7 DNA

polymerase. Subsequent to Fastdigest DpnI (Thermo Fisher Scientific) treatment, the amplicon was purified from a 2% agarose TBE gel using the QIAEX II Gel Extraction Kit (Qiagen). In parallel, 54 bp-long and 53 bp-long single stranded oligos, (sense and antisense strand, respectively) comprising the variable region of the sgRNA, were

5   synthesized (TAG Copenhagen, Table 2). The sense and antisense single stranded oligos (100 µM) were annealed in NEBuffer4 (New England Biolabs) by incubating the oligo mix at 95°C for 5 min in a heating block and the oligo mix was subsequently allowed to slowly cool to RT by turning off the heating block. The annealed oligos were then mixed with the gel purified expression vector backbone and treated with USER

10   enzyme (New England Biolabs) according to manufacturer's recommendations. After USER enzyme treatment, the reaction mixture was transformed into E. coli Mach1 competent cells (Life Technologies) according to standard procedures. Transformant clones were selected on 50 µg/mL Kanamycin (Sigma-Aldrich) LB plates. All constructs were verified by sequencing and purified by NucleoBond Xtra Midi EF (Macherey-

15   Nagel) according to manufacturer's guidelines.

Table 2. Primer sequences

Primer name	Purpose	Sequence (5'-3')
Oligo sgRNA1_C_fw (SEQ ID NO: 27)	Sense oligo for sgRNA1_C	GGAAAGGACGAAACACCGAAAAGTGTCC TGAACAAGGGTTTTAGAGCTAGAAAT
Oligo sgRNA1_C_rv (SEQ ID NO: 28)	Antisense Oligo for sgRNA1_C	CTAAAACCCTTGTTCAGGACACTTTTCGG TGTTTCGTCCTTTCCACAAGATAT
Oligo sgRNA2_C_fw (SEQ ID NO: 29)	Sense oligo for sgRNA2_C	GGAAAGGACGAAACACCGAATATGTGAG TGTGGATGGGTTTTAGAGCTAGAAAT
Oligo sgRNA2_C_rv (SEQ ID NO: 30)	Antisense Oligo for sgRNA2_C	CTAAAACCCATCCACACTCACATATTCGG TGTTTCGTCCTTTCCACAAGATAT
Oligo sgRNA3_C_fw (SEQ ID NO: 31)	Sense oligo for sgRNA3_C	GGAAAGGACGAAACACCGAAATATGCTG GAGTATTTGGTTTTAGAGCTAGAAAT
Oligo	Antisense Oligo for	CTAAAACCAAATACTCCAGCATATTTCGG

49

sgRNA3_C_rv (SEQ ID NO: 32)	sgRNA3_C	TGTTTCGTCCTTTCCACAAGATAT
Oligo sgRNA4_C_fw (SEQ ID NO: 33)	Sense oligo for sgRNA4_C	GGAAAGGACGAAACACCGCAGTCTGCCTGAAATATGCGTTTTAGAGCTAGAAAT
Oligo sgRNA4_C_rv (SEQ ID NO: 34)	Antisense Oligo for sgRNA4_C	CTAAAACGCATATTTCAGGCAGACTGCGGTGTTTCGTCCTTTCCACAAGATAT
Oligo sgRNA1_F_fw (SEQ ID NO: 35)	Sense oligo for sgRNA1_F	GGAAAGGACGAAACACCGAAAGGATCATGAAATCTTAGTTTTAGAGCTAGAAAT
Oligo sgRNA1_F_rv (SEQ ID NO: 36)	Antisense Oligo for sgRNA1_F	CTAAAACTAAGATTTCATGATCCTTTCGGTGTTTCGTCCTTTCCACAAGATAT
Oligo sgRNA2_F_fw (SEQ ID NO: 37)	Sense oligo for sgRNA2_F	GGAAAGGACGAAACACCGATCCGTCCACAACCTTGGCGTTTTAGAGCTAGAAAT
Oligo sgRNA2_F_rv (SEQ ID NO: 38)	Antisense Oligo for sgRNA2_F	CTAAAACGCCAAGGTTGTGGACGGATCGGTGTTTCGTCCTTTCCACAAGATAT
Oligo sgRNA3_F_fw (SEQ ID NO: 39)	Sense oligo for sgRNA3_F	GGAAAGGACGAAACACCGTCAGACGCACTGACAAAGTGTTTTAGAGCTAGAAAT
Oligo sgRNA3_F_rv (SEQ ID NO: 40)	Antisense Oligo for sgRNA3_F	CTAAAACACTTTGTCAGTGCGTCTGACGGTGTTTCGTCCTTTCCACAAGATAT
Oligo sgRNA4_F_fw (SEQ ID NO: 41)	Sense oligo for sgRNA4_F	GGAAAGGACGAAACACCGGATAAAAAAGAGTGTATCGTTTTAGAGCTAGAAAT
Oligo sgRNA4_F_rv (SEQ ID NO: 42)	Antisense Oligo for sgRNA4_F	CTAAAACGATACACTCTTTTTTTATCCGGTGTTTCGTCCTTTCCACAAGATAT
sgRNA Backbone_fw	USER PCR primer for sgRNA backbone amplicon	AGCTAGAAAUAGCAAGTTAAAATAAGGC

50

(SEQ ID NO: 43)		
sgRNA Backbone_rv (SEQ ID NO: 44)	USER PCR primer for sgRNA backbone amplicon	ACAAGATAUATAAAGCCAAGAAATCGA
COSMC-fw (SEQ ID NO: 45)	COSMC amplicon for TOPO cloning and T7 assay	GGATCCATCGCAGCCTTTCT
COSMC-rv (SEQ ID NO: 46)	COSMC amplicon for TOPO cloning and T7 assay	ACTACCTGGTTCGGGTGGTT
M13 forward (−20) (SEQ ID NO: 47)	Sequencing primer for pCR4-TOPO vector	GTAAAACGACGGCCAG
gRNA1_C_fw_D (SEQ ID NO: 48)	COSMC amplicon for MiSeq analysis	CGCATTTTCCGCAAATACTCCAG
gRNA1_C_rv_D (SEQ ID NO: 49)	COSMC amplicon for MiSeq analysis	TGAATATGTGAGTGTGGATGGAGG
gRNA2_C_fw_D (SEQ ID NO: 50)	COSMC amplicon for MiSeq analysis	TCCCACCTTGTTCAGGACACT
gRNA2_C_rv_D (SEQ ID NO: 51)	COSMC amplicon for MiSeq analysis	GGATCCATCGCAGCCTTTCTAT
gRNA3_C_fw_D (SEQ ID NO: 52)	COSMC amplicon for MiSeq analysis	ATGAAAAGCCCAACAGATTTGGTA
gRNA3_C_rv_D (SEQ ID NO: 53)	COSMC amplicon for MiSeq analysis	AGACTCAACAGCCTTCTCAGTG
gRNA4_C_fw_D (SEQ ID NO: 54)	COSMC amplicon for MiSeq analysis	TGAAAAGCCCAACAGATTTGGTA
gRNA4_C_rv_D (SEQ ID NO: 55)	COSMC amplicon for MiSeq analysis	AGTGTTCCGGAAAAGTGTCCT
gRNA1_F_fw_D (SEQ ID NO: 56)	FUT8 amplicon for MiSeq analysis	AGTGACCTTAGCACAAATATTGAAA
gRNA1_F_rv_D (SEQ ID NO: 57)	FUT8 amplicon for MiSeq analysis	TGTCTTTGGAGTTCGTTTCCT
gRNA2_F_fw_D	FUT8 amplicon for MiSeq	AGAGTCCATGGTGATCCTGC

(SEQ ID NO: 58)	analysis	
gRNA2_F_rv_D (SEQ ID NO: 59)	FUT8 amplicon for MiSeq analysis	TACTGTTTAAGGGGAGGGGGA
gRNA3_F_fw_D (SEQ ID NO: 60)	FUT8 amplicon for MiSeq analysis	TGCCCCCATGACTAGGGATA
gRNA3_F_rv_D (SEQ ID NO: 61)	FUT8 amplicon for MiSeq analysis	TCTGCGTTCGAGAAGCTGAAA
gRNA4_F_fw_D (SEQ ID NO: 62)	FUT8 amplicon for MiSeq analysis	ACAGAAGCAGCCTTCCATCC
gRNA4_F_rv_D (SEQ ID NO: 63)	FUT8 amplicon for MiSeq analysis	CCCATACAGAACCACTTGTTGG

Cell culture and transfection

CHO-K1 adherent cells obtained from ATCC (#ATCC-CCL-61) were grown in CHO-K1

F-12K medium (ATTC) supplemented with 10% fetal calf serum (Life Technologies)

and 1% Penicillin-Streptomycin (Sigma-Aldrich). Cells were expanded in T-75 cm^2

vented cap tissue culture flasks (SARSTEDT) and experiments were performed in

Advanced TC Cell Culture Multiwell plates (Greiner Bio-one). Cells were released from

plastic ware using trypsin-EDTA (Sigma-Aldrich). Cells were transfected (Day 0) by the

Nucleofector 2b device using the Amaxa Cell Line Nucleofector Kit V (Lonza) according

to manufacturer's guidelines (program U-023). A total of $1\cdot10^6$ cells were transfected

with 1 μg Cas9 plasmid and 1 μg sgRNA plasmid. Cells were incubated at 30 °C, 5%

CO_2 from Day 1 to Day 2 (cold shock) and incubated at 37 °C, 5% CO_2 at all other

times. Two days after transfection (Day 2), cells transfected with the pmaxGFP plasmid

(Lonza) were used to estimate the transfection efficiency by analyzing GFP signal

using a Celigo Imaging Cell Cytometer (Brooks Automation). The transfection efficiency

was calculated as the percentage of GFP positive cells. Five days after transfection

(Day 5), cells were trypsinized and pelleted (200g, 5 min, RT). Genomic DNA was

extracted from the cell pellets using QuickExtract DNA extraction solution (Epicentre,

Illumina) according to manufacturer's instructions and stored at -20 °C.

Selection and phenotypic staining of FUT8 knockout cells

Five days after transfection (Day 5), selection of FUT8 knockout cells was initiated by

supplementing complete media with 50 μg/mL *Lens culinaris* agglutinin (LCA; Vector

Laboratories) from a 5 mg/mL LCA (10 mM Hepes/NaOH, pH 8.5, 0.15 mM NaCl, 0.1 mM CaCl$_2$) stock solution. Bright field images were taken with a Celigo Imaging Cell Cytometer (Brooks Automation). After 7 days of selection (Day 12), genomic DNA was extracted as described above. In parallel, cells were seeded in complete media without

5   LCA. The day after (Day 13), cells were incubated for 45 minutes at RT in complete media containing 20 μg/mL fluorescein-LCA (Vector Laboratories) and 2 droplets NucBlue® Live ReadyProbes (Life Technologies) per mL media. Cells were washed three times with complete media and fluorescence microscopy was performed on a LEAP instrument (Intrexon) using the 2 channel imaging application with the NucBlue

10   stain as target 1 using the blue fluorescence channel and the fluorescein labeled LCA as target 2 using the green fluorescence channel.

T7 endonuclease assay

Genomic regions flanking the CRISPR target site for T7 endonuclease assay were amplified from the genomic DNA extracts using DreamTaq DNA polymerase (Thermo

15   Fisher Scientific) by touchdown PCR for COSMC (95°C for 2 min; 10x: 95°C for 30 s, 69°C-59°C (-1°C/cycle) for 30 s, 72°C for 50 s; 20x: 95°C for 30 s, 59°C for 30 s, 72°C for 50 s; 72°C for 5 min), using PCR primers listed in Table 2. The PCR products were subjected to a re-annealing process to enable heteroduplex formation which is sensitive to T7 digestion: 95°C for 10 min; 95°C to 85°C ramping at -2°C/s; 85°C to

20   25°C at -0.25 °C/s; and 25°C hold for 1 min. Re-annealed PCR products were treated with T7 endonuclease (New England Biolabs) for 30 min at 37°C. T7 digested and undigested samples were analyzed on a 3% TAE gel.

TOPO™ TA cloning and Sanger sequencing

A genomic region of 318 bp covering the four COSMC sgRNA target sites was PCR-

25   amplified from the genomic extracts as described in the T7 endonuclease assay. PCR products were subjected to agarose gel electrophoresis and subsequently gel purified from a 1% agarose TBE gel using the QiaQuick Gel Extraction Kit (Qiagen). Purified PCR products were TOPO-cloned into the pCR4-TOPO vector using the TOPO™ TA cloning kit (Life Technologies) and subsequently transformed into *E. coli* Mach1

30   chemically competent cells (Life Technologies). Transformed Mach1 cells were then plated on LB-ampicillin agar plates and grown at 37°C overnight. Plasmids from single colony 60 μg/ml carbenicillin (Novagen, Merck) 2X YT-cultures were extracted using the Nucleospin 8/96 Plasmid kit (Macherey-Nagel). Each plasmid preparation was sequenced using the M13 forward (−20) primer (table 2) on an AB 3500xL Genetic

Analyzer (Life Technologies) using the BigDye Terminator v3.1 cycle sequencing kit (Life Technologies).

MiSeq library construction and deep sequencing

PCR amplicons were designed to be between 150 bp and 200 bp long and to span the
sgRNA target sequence (Table 2 for primers and Table 3 for amplicon sizes).

Table 3. Primer pairs used to generate PCR amplicons for deep sequencing analysis.

Target gene	sgRNA	Primer pair	Amplicon size (bp)
COSMC	sgRNA1_C	gRNA1_C_fw_D + gRNA1_C_rv_D	178
	sgRNA2_C	gRNA2_C_fw_D + gRNA2_C_rv_D	155
	sgRNA3_C	gRNA3_C_fw_D + gRNA3_C_rv_D	179
	sgRNA3_C	gRNA4_C_fw_D + gRNA4_C_rv_D	160
FUT8	sgRNA1_F	gRNA1_F_fw_D + gRNA1_F_rv_D	198
	sgRNA2_F	gRNA2_F_fw_D + gRNA2_F_rv_D	163
	sgRNA3_F	gRNA3_F_fw_D + gRNA3_F_rv_D	174
	sgRNA4_F	gRNA3_F_fw_D + gRNA3_F_rv_D	178

Amplicons were generated from the genomic DNA extracts using Phusion Hot Start II
HF Pfu polymerase (Thermo Fisher Scientific) by touch-down PCR (95°C for 7 min;
20x: 95°C for 45 s, 69°C-59°C (-0.5°C/cycle) for 30 s, 72°C for 30 s; 35x: 95°C for 45 s,
59°C for 30 s, 72°C for 30 s; 72°C for 7min). Amplicons were purified on 2% agarose
TBE gels and bands with expected fragment sizes were excised and purified using
QIAEX II Gel Extraction Kit (Qiagen). Amplicon concentration was measured on Qubit®
using the dsDNA BR Assay Kit (Life Technologies). Amplicons were pooled in four for

multiplexing (25 ng each, 100 ng in total). Illumina multiplexing adapters were ligated to the pooled amplicons using the TruSeq™ LT DNA Sample Preparation LT kit (Illumina) according to manufacturer's instructions. DNA concentration of the multiplexed libraries was measured with the Qubit® dsDNA BR Assay Kit, and library quality was

5   determined with an Agilent DNA1000 Chip (Agilent Bioanalyzer 2100). Finally, multiplexed libraries were pooled and sequenced on a MiSeq Benchtop Sequencer (Illumina) using the MiSeq Reagent Kit v2 (300 cycles) according to manufacturer's protocol for a 151 bp paired-end analysis.

Deep sequencing data analysis

10   To minimize the number of required indexes, the same index was used on different PCR products (multiplexing) and the identities of the PCR products were found in the data analysis step based on their individual PCR primer sequences. A Python script was developed to process MiSeq data resulting from the targeted re-sequencing of the Cas9 target site regions. The script performs the following tasks: 1. Join paired-end

15   reads. 2. Check if resulting sequences contain correct PCR primer at both beginning and end of the sequences and discard those sequences that fail to do so. 3. Compute output length of PCR product. 4. Compare PCR product length to expected PCR product length. Paired-end reads were joined using fastq-join (http://code.google.com/p/ea-utils). Ends were checked for correct PCR primer using

20   fastx_barcode_splitter (http://hannonlab.cshl.edu/fastx_toolkit/index.html).

Example 7. Accelerating genome editing in CHO cells using CRISPR Cas9; results.

RNA-guided CRISPR Cas9 shows endonuclease activity in CHO

25   In order to test if the RNA-guided CRISPR Cas9 system could be applied for gene disruptions in CHO cells, an expression vector with a CHO codon-optimized version of Cas9 with a C-terminal SV40 nuclear localization signal under the control of a CMV promoter was constructed. To direct Cas9 to disrupt genes of interest, sgRNA expression constructs were generated using the human U6 polymerase III promoter as

30   previously described. Four sgRNAs were designed for each of the two genes; C1GALT1C1 (COSMC) encoding the C1GALT1-specific chaperone 1 and FUT8 encoding fucosyltransferase 8 (Alpha-(1,6)-fucosyltransferase). COSMC is a chaperone essential for correct protein O-glycosylation and FUT8 catalyzes the transfer of fucose from GDP-fucose to N-acetylglucosamine. The four sgRNA

constructs for COSMC target the only exon present in the gene. However, FUT8 consists of 11 exons and the FUT8 sgRNA constructs target exon 5, exon 7, and exon 9. To compare the activity of the designed sgRNAs, adherent CHO-K1 cells were transfected transiently with the CHO codon-optimized Cas9 expression vector and

5   each of the eight sgRNAs to introduce DSBs in the two test genes in two independent experiments (replicate 1 and 2). Initially, a T7 endonuclease assay was performed to analyze the indel frequency at the COSMC loci resulting from Cas9 guided by the four different COSMC-targeting sgRNAs (sgRNA1_C, sgRNA2_C, sgRNA3_C and sgRNA4_C). When assayed 5 days after transfection, genomic indel events were

10   detected for all four sgRNAs. The fragment sizes of the digested amplicons correspond to the expected sizes (Table 4).

Table 4. Predicted fragment sizes after T7 digestion of the 318 bp PCR product amplified from the COSMC gene. DNA mismatches resulting from sgRNA-Cas9 activity were predicted to take place at the 5'-end of the PAM sequence and this position was

15   defined as the T7 nuclease cleavage site: GNNNNNNNNNNNNNNNNNNNN|NGG.

	Large fragment (bp)	Small fragment (bp)
sgRNA1_C	168	150
sgRNA2_C	246	72
sgRNA3_C	218	100
sgRNA4_C	207	111

High indel frequency obtained by all four COSMC-targeting sgRNAs

To further assess the indel frequency achieved with the COSMC sgRNAs, TOPO cloning-based sequencing of gel-purified amplicons from the COSMC genomic site was

20   performed (Table 5). Consistent with the T7 endonuclease assay, Cas9 activity was observed for all four COSMC sgRNAs in two independent experiments (replicate 1 and 2). sgRNA1_C gave rise to the highest indel frequency of 48.0 and 60.0% and sgRNA4_C displayed the lowest indel frequency of 13.8 and 17.9%, in replicate 1 and 2, respectively. Based on GFP fluorescence of cells transfected with GFP-encoding

25   plasmids, transfection efficiency was estimated to be approximately 60% and 65% for replicate 1 and 2, respectively. The indels created by the COSMC sgRNAs predominantly involved a single-base insertion of a thymine or deletions. To analyze the Cas9 activity in greater detail, deep sequencing was performed using the genomic DNA extracts from the two independent experiments. Deep sequencing data

comprising between approximately 200,000-700,000 reads per sgRNA in each of the two replicates correlated well with the sequencing data obtained from TOPO cloning (between 21 to 32 sequences per sgRNA). Both sequence-based methods detected relatively high Cas9-activity for all four sgRNAs. Deep sequencing reported indel

5   frequencies of 47.3% and 44.3% for sgRNA1_C, 45.6% and 40.2% for sgRNA2_C, 36.0% and 27.2% for sgRNA3_C and 15.2% and 13.6% for sgRNA4_C in replicate 1 and 2, respectively. Deep sequencing of control cells transfected only with Cas9-encoding plasmids showed an indel frequency of 0.1 to 0.2%. To examine the fidelity of both sequence-based methods, indel-containing sequences obtained from TOPO-

10   cloning were checked using the deep sequencing data. All indels detected in the TOPO-cloning experiments were also retrieved in the deep sequencing data (data not shown).

Table 5. TOPO cloning-based targeted sequencing of COSMC.

	Target	#Indels	#WT	Indels (%)	WT (%)
	Cas9 only	0	22	0	100
	sgRNA1_C	12	13	48.0	52.0
Replicate 1	sgRNA2_C	13	14	48.1	51.9
	sgRNA3_C	7	25	21.9	78.1
	sgRNA4_C	3	26	10.3	89.7
	Cas9 only	0	26	0	100
	sgRNA1_C	14	7	66.7	33.3
Replicate 2	sgRNA2_C	10	14	41.7	58.3
	sgRNA3_C	8	17	32.0	68.0
	sgRNA4_C	5	23	17.9	82.1

15   *Homozygous knockout of FUT8 in CHO cells generated by CRISPR*

The α1,6-fucosyltransferase FUT8 catalyzes the addition of fucose on IgG1 antibodies produced by CHO cells which can reduce antibody-dependent cell-mediated cytotoxicity. Disruption of the FUT8 gene in CHO cells is therefore attractive in order to achieve highly active and completely nonfucosylated therapeutic antibodies. To expand

20   our knowledge of applying CRISPR Cas9 in CHO cells, the gene disruption efficiency of four FUT8 sgRNAs was investigated by deep sequencing. Genomic regions covering the target site of sgRNA1_F, sgRNA2_F, sgRNA3_F and sgRNA4_F (table 1) were PCR amplified and sequenced. This analysis revealed that all four sgRNAs gave rise to significant Cas9 activity with an indel frequency of 17.6% and 15.1% for sgRNA1_F,

38.7% and 31.2% for sgRNA2_F, 42.5% and 36.0% for sgRNA3_F and 18.9% and

11.1% for sgRNA4_F in replicate 1 and 2, respectively. As previously mentioned,

transfection efficiency was estimated to be approximately 60% and 65% for replicate 1

and 2, respectively. *Lens culinaris* agglutinin (LCA)-based selection was further used to

5   select for FUT8-disrupted CHO cells. LCA binds fucosylated plasma membrane

proteins leading to endocytosis and cell death. This enables selection for homozygous

FUT8 gene disruptions, since LCA can no longer bind to cells devoid of FUT8 and

these cells therefore survive. LCA-treatment was initiated 5 days after transfection and

resulted in non-adherent round-shaped morphology of all control cells. However, many

10   adherent cells were detected in the pool of cells transfected with Cas9 and the four

FUT8 sgRNAs, indicating Cas9-mediated functional knockout of FUT8 in these cells.

Eight days after initiation of selection, LCA selected and non-LCA selected cells

transfected with and without Cas9 and sgRNAs were stained with Fluorescein-labeled

LCA (F-LCA). Cells transfected with Cas9+sgRNAs without LCA selection revealed a

15   fraction of F-LCA negative cells, demonstrating the presence of cells with homozygous

disruption of the FUT8 gene. For Cas9+sgRNA3_F, these F-LCA negative cells

constituted 29.1% of the entire population. Cells transfected with Cas9+sgRNA3_F,

which subsequently had been exposed to LCA treatment revealed that the majority of

cells (98.6%) stained LCA negative. This clearly demonstrates that the LCA treatment

20   efficiently selects for cells devoid of functional FUT8 as previously observed. Indeed,

this observation was confirmed by deep sequencing, since LCA selection significantly

enriched cells with FUT8 disruption to an indel frequency between 98.2% and 99.7%

for the four FUT8-targeting sgRNAs.

The majority of CRISPR-generated indels is single base pair insertions

25   The vast amount of information obtained from deep sequencing led us to investigate

further the indel sizes created by the NHEJ repair mechanism resulting from the

COSMC and FUT8 sgRNAs. The frequency was calculated as an average for both

independent experiments and was based on $11 \cdot 10^5$ reads for COSMC sgRNAs and

$8 \cdot 10^5$ reads for FUT8 sgRNAs. All targets were weighted equally with each target

30   contributing 12.5%. Surprisingly, mainly single base pair insertions were detected with

a frequency of 32.8%. High frequency of single base pair insertions was also observed

in sequences obtained from TOPO cloning. Two and one base pair deletions were the

second and third most frequent indel size with a frequency of 10.3% and 8.7%

respectively. Together, almost half of the identified indels (56.5%) were single or

35   double-base pair indels. Collectively, 85% of the indels observed in this study resulted

in frame shift mutations (+/-1 or +/-2 bp) in the reading frame, which most likely leads to a loss-of-function of the target protein. This finding further underlines CRISPR Cas9 as a powerful tool to disrupt genes of interest in the CHO genome, and prompted us to develop a target design tool that facilitates identification of Cas9 targets.

5

Example 8. CRISPR Cas9 mediated site-specific targeted integration in CHO cells

In order to facilitate site specific integration of a gene of interest (GOI), a targeted integration platform was set up for CHO cells. This platform consists of the CHO codon optimized Cas9, sgRNAs for the integration site and donor plasmids encoding the GOI,

10   antibiotic resistance, fluorescence marker and homology arms to facilitate targeted integration by homologous recombination. The two tested integration sites are the genes C1GALT1-specific chaperone 1-like (COSMC, gene id: 100751243, scaffold: NW_003628455.1) and mannoside acetylglucosaminyltransferase 1 (Mgat1, gene id: 100682529, scaffold: NW_003614027.1). mCherry was selected as GOI to facilitate

15   easy screening of expression. Applying this platform, monoclonal cell lines with targeted integration of GOI were generated. To characterize the potential for future application of the CRISPR Cas9 system in targeted integration in CHO, the integration event was analyzed to elucidate the potential of the cells own homology directed repair (HDR) and non-homologous end-joining (NHEJ) pathways. Integration by HDR

20   facilitates precise integration of the donor DNA (GOI) while integration by the error prone NHEJ can result in unpredictable genome modifications.

Materials and methods

The CHO codon optimized Cas9 expression vector applied in the study is described in example 6. The CRISPy bioinformatic tool (http://staff.biosustain.dtu.dk/laeb/crispy/)

25   was applied for generating two sgRNAs targeting MGAT1 (Mgat1_sgRNA1 and Mgat1_sgRNA5). The sgRNA target sequences in COSMC and Mgat1 are described in table 6.

Table 6: sgRNA target sequences

Name	CRISPy identification number	Target sequence
C1GALT1/COSMC (sgRNA2_C) (SEQ ID NO: 64)	1964768	GAATATGTGAGTGTGGATGGAGG

Mgat1_sgRNA1 (SEQ ID NO: 65)	905598	GCTCACACCCTTACGGCCAAAGG
Mgat1_sgRNA5 (SEQ ID NO: 66)	904917	GTGGAGTTGGAGCGGCAGCGGGG

The sgRNA expression vectors have been constructed as described in example 6. The sgRNA targeting COSMC (sgRNA2_C) is described in example 6. The sgRNAs targeting Mgat1 are constructed with the oligos described in table 7.

5   **Table 7: Oligos for construction of Mgat1 gRNA expression vectors**

Oligo name	Sequence (5' – 3')
Mgat1_sgRNA1_fwd (SEQ ID NO: 67)	GGAAAGGACGAAACACCGCTCACACCCTTACGGCCAAGTTTTAGAGCTAGAAAT
Mgat1_sgRNA1_rev (SEQ ID NO: 68)	CTAAAACTTGGCCGTAAGGGTGTGAGCGGTGTTTCGTCCTTTCCACAAGATAT
Mgat1_sgRNA5_fwd (SEQ ID NO: 69)	GGAAAGGACGAAACACCGTGGAGTTGGAGCGGCAGCGGTTTTAGAGCTAGAAAT
Mgat1_sgRNA5_rev (SEQ ID NO: 70)	CTAAAACCGCTGCCGCTCCAACTCCACGGTGTTTCGTCCTTTCCACAAGATAT

Donor DNA was constructed with USER cloning; The different parts of the vectors were amplified from commercial expression vectors or genomic DNA from CHO-S cells as templates with PCR and purified. The purified PCR fragments were assembled with Uracil-Specific Excision Reagent (USER) cloning and transformed into *E. coli*
10   competent cells and plated on LB media with ampicillin. Colonies were selected and plasmids were harvested. The donor DNA was verified by sequencing.

CHO cells e.g. CHO-S cells from Life Technologies were grown in appropriate medium e.g. CD CHO medium (Life Technologies) supplemented with 8 mM L-Glutamine and
15   cultivated in shake flasks. The cells were incubated at 37°C, 5 % CO_2 with 120 rpm shaking and passaged every 2-3 days. Transfection was performed with expression vectors encoding CHO optimized Cas9, sgRNA targeting the integration site and

corresponding donor DNA with homology arms towards the integration site and encoding mCherry. For each sample, 3×10^6 cells were transfected with a total of 3.75 µg of DNA. 16 hours post transfection, the samples were incubated at 30 °C for 32 hours before transferred back to 37 °C. Stable pools were generated by seeding cells in tissue culture plates on day 3 followed by selecting for G418 (500 µg/ml) resistant clones. During selection, medium was changed every 3-4 days. After 2 weeks of selection, cells were transferred to shake flasks. The stable pool of cells were sorted using Fluorescence-activated cell sorting (FACS)to harvest mCherry positive cells which were also ZsGreen 1 negative to select for cells with potential HDR mediated targeted integration as ZsGreen1 is present outside the homology arms in the plasmid while mCherry is present inside. The FACS sorting was followed by limiting dilution for single cell clone generation. 1 cell was seeded per well in 200 µl medium in 96 well plates. The generated colonies were analyzed by imaging cytometry to identify round shaped colonies expressing mCherry.

Junction PCRs of the limiting dilution derived clones were performed on genomic DNA extracts from harvested cells. Growth analysis of the clones was performed.

Ricinus Communis Agglutinin (RCA) sensitivity test on CHO-S cells was performed with RCA concentrations at 0, 5, 10, 20, 50 and 100 µg/ml. Cells were seeded at 2×10^5 cells/ml. Cells were stained with a cell-permeant nuclear counterstain with and without fluorescein labeled RCA at final concentration of 20 µg/ml. Analysis was performed with imaging cytometry (Celigo).

Targeted integration into COSMC

CHO-S cells were transfected with expression vectors encoding CHO codon optimized Cas9, sgRNA2_C and donor DNA with homology arms specific for the COSMC integration site. Donor DNA with mCherry expression cassette and neomycin resistance cassette inside and ZsGreen1 expression cassette outside homology arms was constructed to facilitate CRISPR Cas9 HDR mediated targeted integration into COSMC (Figure 6). In the first round of selection, G418 was applied to select for cells with the donor DNA integrated into the genome. FACS sorting was applied in the

second round of selection to sort for cells expressing mCherry but not ZsGreen1 to select for HDR mediated targeted integrants.

5   In total, 588 clones were screened with the Celigo. 138 clones showed consistent expression of mCherry. 23 clones were excluded due to very slow growth resulting in 115 clones which were characterized further. Junction PCR was applied to analyze the integration of donor DNA into the integration site. 5' and 3' junction PCR was performed on genomic extract from 115 clones. 79 of the clones (68.7%) were both junction PCR positive indicating HDR mediated targeted integration of the donor DNA

10   into COSMC (Figure 7).

       Junction PCR positive clones showed a homogenous mCherry expression level indicated by mCherry fluorescence in junction PCR positive clones in comparison with Junction PCR negative clones despite similar variations in specific growth rate between

15   the two populations. MCherry expression and relative specific growth rates were measured for 52 junction PCR + and 33 PCR – clones. Both junction PCR + and junction PCR - clones showed almost the same relative specific growth rate but junction PCR + clones showed more stable mCherry expression with less variation compared to the junction PCR – clones indicating more stable and predictive

20   expression from site-specific integration of donor DNA compared to random integration (Figure 8).

       In accordance with the assumption that cells expressing ZsGreen1 green fluorescence are random integrants, not site-specific integrants, cells expressing both mCherry and

25   ZsGreen1 showed incomplete junction PCR positive bands. Integration events were not detected in the transiently transfected pools but were detected in stable selected pools verifying targeted integration of donor DNA into COSMC. Stable mCherry+/Zsgreen1+ showed less clean and intense bands indicating a mixture of integration events as expected (Figure 9).

30

       Targeted integration into MGAT1

The efficiency of the two Mgat1 sgRNAs (sgRNA1 and sgRNA5) was analyzed with a
T7 endonuclease assay. The two sgRNAs generated 6.5 % and 13.3% indels,
respectively. T7 results are shown for the 5 tested sgRNAs of which sgRNA1 and
sgRNA5 was selected for targeted integration. T7 endonuclease assay was performed
5   on genomic DNA from cells transfected with five Mgat1 sgRNAs. The Indels (%)
generated by the sgRNAs were estimated from the intensities of the cut DNA bands
obtained from the T7 endonuclease treatment of PCR fragments. The sgRNAs resulted
in between 5.9 and 13.3% of indels generated at the target site. sgRNA1 and sgRNA5
were chosen for targeted integration (Figure 10).

10

Donor DNAs for each of the two sgRNAs were constructed. Donor DNA with mCherry
expression cassette and neomycin resistance cassette inside and ZsGreen1
expression cassette outside homology arms towards either sgRNA1 or sgRNA5 was
constructed to facilitate CRISPR Cas9 HDR mediated targeted integration into Mgat1
15   (Figure 11).

CHO-S cells were transfected with Cas9, sgRNA and donor DNA to generate two pools
of cells, one for each sgRNA. Cells expressing the neomycin gene were selected using
G418 in the first round of selection. In the second round of selection, mCherry
20   expressing and ZsGreen1 non-expressing cells were FACS sorted as earlier described
followed by limiting dilution cloning. MCherry positive and ZsGreen1 positive cells were
also harvested from the FACS sorting for analysis but these cells were not cloned.

Junction PCR analysis of the generated stable pool of cells were performed and
25   compared to junction PCR analysis on transiently transfected cells. Indeed, 5' and 3'
junction PCR is only positive for the stable pools. Furthermore, similar to the result of
COSMC site, cells expressing both mCherry and ZsGreen1 showed incomplete
junction PCR positive bands. Junction PCR was performed on transiently transfected
cells and stable pools of cells. Integration events was not detected in the transiently
30   transfected pools but was detected in stable selected pools with both 5' and 3' junction
PCR verifying targeted integration of donor DNA into Mgat1. Stable sorted cells
mCherry+/Zsgreen1+ showed less clean and intense bands than sorted

mCherry+/Zsgreen1- indicating a mixture of integration events as expected (Figure 12). Precise gene insertion was confirmed by sequencing of PCR products.

5   The number of Mgat1 knocked out cells in these two pools were estimated by imaging cytometry e.g. Celigo analysis of fluorescein labelled RCA staining. Since Mgat1 adds N-acetylglucosamine to the Man5GlcNAc2 (Man5) N-glycan structure, and *Ricinus communis* agglutinin-I (F-RCA) is a cytotoxic lectin that binds Man5GlcNAc2, it has been reported that Mgat1 knock-out (disrupted) cells show the RCA-resistant. Thus, resistance RCA-I was used to confirm disruption of the Mgat1 locus.

10

Simultaneous introduction of active sgRNAs, Cas9 nucleases, and donor templates targeting Mgat1 locus is likely to induce disruption of Mgat1 locus in the range of 7-10% accompanying co-expression of GOI from donor template or not The mCherry expression and F-RCA stained cells of stable pools generated upon transfections of
15   donor DNA with and without Cas9 and sgRNA for both Mgat1 integration sites were analyzed by Celigo. Q3 contains mCherry - / ZsGreen1 & F-RCA staining – cell population and Q4 contains mCherry +/ZsGreen 1 & F-RCA – cell population. Indeed co-transfections with Cas9 and sgRNA1 or sgRNA5 increased the percentage of Q3 and Q4 populations upon F-RCA staining indicating an increase in Mgat1 disrupted
20   cells and potentially increased HDR mediated integrations (Figure 13).

Donor DNA vectors:

1991   **COSMC-mCherry-HDR-TI donor** (SEQ ID NO: 71)

Features: pJ204_COSMC_5' arm_EF1a_mCherry-BGH_sv40-NeoR-sv40 pA_COSMC_3' arm_CMV-
25   ZsGreen1-BGH
AGTCGGTGTGTAATCCATGGAGGAGTTTCTATAATGTTGCAGTTTCTACCTAATGG
TGACCAAATGCCAGTGAAAGGATTGTAAGAGTACTTGTCACATATACTACTCACCT
CATTTCAAGAATGTGGACCTGCTTTTAAACATTAAGAGCAAATCGTAATTATATAAG
AAATAAGCAAATGAAACTATTAGACTGTTTGAAAAGTCTTTTTCTTTACAGGAAAAA
30   TGCTTTCAGAAAGCAGTTCATTTTTGAAAGGTGTGATGCTTGGAAGCATCTTCTAT
GCCTTGATCACTACGCTAGGCCACATTAGGATTGGGCACAGAAACAGGACACACC
ACCATGAGCATCACCACCTGCAAGCTCCTAACAAAGAAGATATCTCGAAAATCTCA
GCGGCTGAGCGCATGGAGCTCAGTAAGAGCTTCCGGGTATACTGTATAGTTCTTG

TAAAACCCAAAGATGTGAGTCTTTGGGCTGCAGTGAAGGAGACTTGGACCAAACA
CTGTGACAAAGCAGAGTTCTTCAGTTCTGAAAATGTTAAAGTGTTTGAGTCAATTA
ACGTGGACACTGATGACATGTGGTTGATGATGAGGAAAGCTTATAAATATGCCTTT
GATAAATACAAAGAGCAGTACAACTGGTTCTTCCTTGCACGCCCCAGTACTTTTGC
TGTGATTGAAAATCTAAAATATTTTTTGTTAAAAAAGGATCCATCGCAGCCTTTCTA
TCTAGGACACACTGTAAAATCTGGAGACCTTAAGCAGCGTGTGAGGCTCCGGTGC
CCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAG
GGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAA
GTGATGTCGTGTACTGGCTCCGCCTTTTCCCGAGGGTGGGGGAGAACCGTATAT
AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACA
CAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC
CCTTGCGTGCCTTGAATTACTTCCACCTGGCTCCAGTACGTGATTCTTGATCCCGA
GCTGGAGCCAGGGGCGGGCCTTGCGCTTTAGGAGCCCCTTCGCCTCGTGCTTGA
GTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCT
TCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGAC
CTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAGGATCT
GCACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGT
CCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATC
GGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCC
GCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTG
CGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTCCAGGGGGCTCAAAATGGA
GGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAGGG
GCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCG
TCCAGGCACCTCGATTAGTTCTGGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGG
GGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAG
TTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTT
GGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTT
CAGGTGTCGTGAAGACGTCATCGCCACCATGGTGAGCAAGGGCGAGGAGGATAA
CATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGT
GAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCGCCCCTACGAGG
GCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCT
GGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCC
CGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGA
GCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTC
CCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCC

65

CTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGA
GCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGA
AGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCA
AGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCA

5   CCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCC
GCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGACACAGTCTCTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTG
GAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTG
TCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGG

10   GGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGA
GGTCTGAGTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCC
CCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTG
GAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA
GTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCC

15   CAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTATTTATGCAGAGG
CCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGG
AGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCT
GATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACG
CAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAAC

20   AGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCC
CGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGA
GGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGC
TCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGG
GGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGC

25   TGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCAC
CAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTC
GATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTC
GCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGG
CGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATC

30   GACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACC
CGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTT
ACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGA
GTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAAC
CTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCG

35   GAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGC

66

TGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAA
AGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGT
GGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGAGGTCTGAGTGATTGTCTT
AAGCATAGAGTCAATGAAAAGACTCAACAGCCTTCTCAGTGTTCCGGAAAAGTGT
5   CCTGAACAAGGTGGGATGATTTGGAAGATATCTGAAGATAAGCAGCTAGCAGTCT
GCCTGAAATATGCTGGAGTATTTGCGGAAAATGCGGAAGACGCTGATAGAAAGA
TGTATTTAATACCAAATCTGTTGGGCTTTTCATTAAAGAGGCCATGTCTAACCACC
CGAACCAGGTAGTAGAAGGATGCTGTTCCAATATGGCTGTCACTTTTAATGGACTA
ACTCCTAATCAGATGCATGTGATGATGTATGGGGTGTACCGGCTTAGGGCCTTTG
10   GACATGTTTTCAACGATGCGTTGGTTTTCTTACCTCCAAACGGTTCTGATAATGAC
TGACAAAAGCAAGAGCATGCATTTGGTAACCACATTAAGACATGTTATGCTTTCT
AATCGATAATGCATCTAACACAGTAGTGTGTTTCTTTTCCTTATCTGGTCACATTGA
AGTCTACTTGTACATTTTCAAATGGAATGGTATTTTTTTCCCTTAAATCATTTGTGA
GAAATTTTAATGTGTTAGAAATAAATGTTTTAAGAATAGCAATTTTGCAAATAATGTA
15   TTTATAAATATTATATTTATGTGATAAAGACCAAATTATAGACATTAAAATCTGTGAT
GTATCTTTGCCTATTGATTTTAAATGTTTAATGTATCTTTTTAGATTTCAAATATATG
CAAATGAGGACGTCGCTGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAAT
TACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGG
TAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAAT
20   GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTG
GAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAG
TACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAG
TACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGC
TATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTT
25   GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTT
GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGAC
GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTG
GCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATAGTGCGATcgccaccatggc
ccagtccaagcacggcctgaccaaggagatgaccatgaagtaccgcatggagggctgcgtggacggccacaagttc
30   gtgatcaccggcgagggcatcggctacccttcaagggcaagcaggccatcaacctgtgcgtggtggagggcggccc
cttgcccttcgccgaggacatcttgtccgccgccttcatgtacggcaaccgcgtgttcaccgagtaccccaggacatcgt
cgactacttcaagaactcctgccccgccggctacacctgggaccgctccttcctgttcgaggacggcgccgtgtgcatctg
caacgccgacatcaccgtgagcgtggaggagaactgcatgtaccacgagtccaagttctacggcgtgaacttccccgc
cgacggccccgtgatgaagaagatgaccgacaactgggagccctcctgcgagaagatcatccccgtgcccaagcag
35   ggcatcttgaagggcgacgtgagcatgtacctgctgctgaaggacggtggccgcttgcgctgccagttcgacaccgtgta

caaggccaagtccgtgccccgcaagatgcccgactggcacttcatccagcacaagctgacccgcgaggaccgcagc

gacgccaagaaccagaagtggcacctgaccgagcacgccatcgcctccggctccgccttgcccaagcttccgcggag

ccatggcttcccgccggcggtggcggcgcaggatgatggcacgctgcccatgtcttgtgcccaggagagcgggatgga

ccgtcaccctgcagcctgtgcttctgctaggatcaatgtgtagACACAGTCTCTGTGCCTTCTAGTTGCC

AGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCAC

TCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGT

GTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG

GAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGACTTGCGTAGT

GAGTCGAATAAGGGCGACACAAAATTTATTCTAAATGCATAATAAATACTGATAAC

ATCTTATAGTTTGTATTATATTTTGTATTATCGTTGACATGTATAATTTTGATATCAA

AAACTGATTTTCCCTTTATTATTTTCGAGATTTATTTTCTTAATTCTCTTTAACAAACT

AGAAATATTGTATATACAAAAAATCATAAATAATAGATGAATAGTTTAATTATAGGTG

TTCATCAATCGAAAAAGCAACGTATCTTATTTAAAGTGCGTTGCTTTTTTCTCATTT

ATAAGGTTAAATAATTCTCATATATCAAGCAAAGTGACAGGCGCCCTTAAATATTCT

GACAAATGCTCTTTCCCTAAACTCCCCCCATAAAAAAACCCGCCGAAGCGGGTTTT

TACGTTATTTGCGGATTAACGATTACTCGTTATCAGAACCGCCCAGGGGGCCCGA

GCTTAAGACTGGCCGTCGTTTTACAACACAGAAAGAGTTTGTAGAAACGCAAAAA

GGCCATCCGTCAGGGGCCTTCTGCTTAGTTTGATGCCTGGCAGTTCCCTACTCTC

GCCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCG

AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT

AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAA

AAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCAC

AAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACC

AGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGC

TTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG

CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGT

GTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGT

CTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA

ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT

GGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAA

GCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC

GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAG

GATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGA

CGCGCGCGTAACTCACGTTAAGGGATTTTGGTCATGAGCTTGCGCCGTCCCGTCA

AGTCAGCGTAATGCTCTGCTTTTACCAATGCTTAATCAGTGAGGCACCTATCTCAG

CGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACT
ACGATACGGGAGGGCTTACCATCTGGCCCCAGCGCTGCGATGATACCGCGAGAA
CCACGCTCACCGGCTCCGGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCC
ATCGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCT
CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGC
GGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTA
TCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG
ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATA
GCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCA
AGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACT
GATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAACAGGAAGG
CAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATAT
TCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTCAGTGTTACAACCAATT
AACCAATTCTGAACATTATCGCGAGCCCATTTATACCTGAATATGGCTCATAACAC
CCCTTGTTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAA
CTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGACTCCCCATGCGAG
AGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGG
CCTTTCGCCCGGGCTAATTATGGGGTGTCGCCCTTATTCGACTC

2326 **Mgat1-mCherry-HDR-TI donor (sgRNA1)** (SEQ ID NO: 72)

Features: pJ204_Mgat1_5' arm_EF1a_mCherry-BGH_sv40-NeoR-sv40 pA_Mgat1_3' arm_CMV-ZsGreen1-BGH

Mgat1-mCherry HDR-TI vector

AGTCGGTGTTTGCAGCAAATCAGGGAGCATCATGCTTTGTGGAGACAGAGGTGGA
AAGTGCCCACCGTGGCCCCTCCAGCCTGGCCCCGTGTGCCTGCGACCCCCTCAC
CAGCCGTGATCCCCATCCTGGTCATTGCCTGTGACCGCAGCACTGTCCGGCGCT
GCTTGGATAAGTTGTTGCACTATCGGCCCTCAGCTGAGCATTTCCCCATCATTGTC
AGCCAGGACTGCGGGCACGAGGAGACAGCACAGGTCATTGCTTCCTATGGCAGT
GCAGTCACACACATCCGGCAGCCAGACCTGAGTAACATCGCTGTGCCCCCAGAC
CACCGCAAGTTCCAGGGTTACTACAAGATCGCCAGGCACTACCGCTGGGCACTG
GGCCAGATCTTCAACAAGTTCAAGTTCCCAGCAGCTGTGGTAGTGGAGGACGATC

TGGAGGTGGCACCAGACTTCTTTGAGTACTTCCAGGCCACCTACCCACTGCTGAG

AACAGACCCCTCCCTTTGGTGTGTGTCTGCTTGGAATGACAATGGCAAGGAGCAG

ATGGTAGACTCAAGCAAACCTGAGCTGCTCTATCGAACAGACTTTTTTCCTGGCCT

TGGCTGGCTGCTGATGGCTGAGCTGTGGACAGAGCTGGAGCCCAAGTGGCCCAA

GGCCTTCTGGGATGACTGGATGCGCAGACCTGAGCAGCGGAAGGGGCGGGCCT

GTATTCGTCCAGAAATTTCAAGAACGATGAAAGCAGCGTGTGAGGCTCCGGTGCC

CGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGG

GGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAG

TGATGTCGTGTACTGGCTCCGCCTTTTCCCGAGGGTGGGGGAGAACCGTATATA

AGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACAC

AGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCC

CTTGCGTGCCTTGAATTACTTCCACCTGGCTCCAGTACGTGATTCTTGATCCCGAG

CTGGAGCCAGGGGCGGGCCTTGCGCTTTAGGAGCCCCTTCGCCTCGTGCTTGAG

TTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTT

CGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACC

TGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAGGATCTG

CACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTC

CCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCG

GACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCG

CCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGC

GTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTCCAGGGGGCTCAAAATGGAG

GACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAGGGG

CCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGT

CCAGGCACCTCGATTAGTTCTGGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGGG

GGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAGTT

AGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGG

ATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTCA

GGTGTCGTGAAGACGTCATCGCCACCATGGTGAGCAAGGGCGAGGAGGATAACA

TGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAA

CGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCA

CCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGG

GACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCG

CCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGC

GCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCC

TGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCT

CCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAG
CGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAA
GCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAA
GAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCAC
CTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCG
CCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGACACAGTCTCTGTGCCTT
CTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGG
GAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGAG
GTCTGAGTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCC
AGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGA
AAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
CAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCA
GTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTATTTATGCAGAGGCC
GAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAG
GCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGAT
CAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAG
GTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGA
CAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCG
GTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGG
CAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCG
ACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGC
AGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGAT
GCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAA
GCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGAT
CAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCC
AGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGAT
GCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACT
GTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTG
ATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGG
TATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTC
TTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGC
CATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAAT
CGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGA

GTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCA
ATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTT
TGTCCAAACTCATCAATGTATCTTATCATGTCTGAGGTCTGAGTCATGGGCAGTTC
TTTGATCAGCATCTTAAGTTCATCAAGCTGAACCAGCAGTTCGTGTCTTTCACCCA
GTTGGATTTGTCATACTTGCAGCGGGAGGCTTATGACCGGGATTTCCTTGCCCGT
GTCTATAGTGCCCCCCTGCTACAGGTGGAGAAAGTGAGGACCAATGATCAGAAG
GAGCTGGGGGAGGTGCGGGTACAGTACACTAGCAGAGACAGCTTCAAGGCCTTT
GCTAAGGCCCTGGGTGTCATGGATGACCTCAAGTCTGGTGTCCCCAGAGCTGGC
TACCGGGGCGTTGTCACTTTCCAGTTCAGGGGTCGACGTGTCCACCTGGCACCC
CCACAAACCTGGGAAGGCTATGATCCTAGCTGGAATTAGCAGCACCTGCCTTTCC
CTGCTGGATCTGCTTGTCATATCATGAGCTGAGACAGGCCTGCAGTCCCTGAGCT
GTACCATCCTGTCCCTGTTTCCCTCTTGGGTCTATATATCTCCTATTTCTGTGCCC
CCCTCCCTCCCTGTGGCATTCTAGTGCATAAATCATGATGAGGGTTATACTCCTGT
TGTCCAGGGAGTCATCAGGAGAACTATTGTGTGGTGTAGTTGGGGGTATTGAACA
AGAAACCACTGTGTGGTATGGGGAGGCTTGGGCTTGTTGGGGCCAAGGAGTCCT
GAGTTCTCTGGAAGGGCATCGCAGAGAGCTTGGCAACTCGAGCTCTCTTGACCAA
GCCTGTTGACCCTAACCTGGCTCCTACGTCGCTGTTGACATTGATTATTGACTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCG
CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCA
AGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC
GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACAT
CTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGT
CAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACA
ACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA
TAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATA
GTGCGATcgccaccatggcccagtccaagcacggcctgaccaaggagatgaccatgaagtaccgcatggaggg
ctgcgtggacggccacaagttcgtgatcaccggcgagggcatcggctacccccttcaagggcaagcaggccatcaacct
gtgcgtggtggagggcggcccccttgcccttcgccgaggacatcttgtccgccgccttcatgtacggcaaccgcgtgttcac
cgagtacccccaggacatcgtcgactacttcaagaactcctgccccgccggctacacctgggaccgctccttcctgttcg
aggacggcgccgtgtgcatctgcaacgccgacatcaccgtgagcgtggaggagaactgcatgtaccacgagtccaag
ttctacggcgtgaacttccccgccgacggccccgtgatgaagaagatgaccgacaactgggagccctcctgcgagaag
atcatccccgtgcccaagcagggcatcttgaagggcgacgtgagcatgtacctgctgctgaaggacggtggccgcttgc

gctgccagttcgacaccgtgtacaaggccaagtccgtgccccgcaagatgcccgactggcacttcatccagcacaagc

tgacccgcgaggaccgcagcgacgccaagaaccagaagtggcacctgaccgagcacgccatcgcctccggctccg

ccttgcccaagcttccgcgcggagccatggcttcccgccggcggtggcggcgcaggatgatggcacgctgcccatgtcttgt

gcccaggagagcgggatggaccgtcaccctgcagcctgtgcttctgctaggatcaatgtgtagACACAGTCTCT

GTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGA

CCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATC

GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAG

CAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT

CTATGGACTTGCGTAGTGAGTCGAATAAGGGCGACACAAAATTTATTCTAAATGCA

TAATAAATACTGATAACATCTTATAGTTTGTATTATATTTTGTATTATCGTTGACATG

TATAATTTTGATATCAAAAACTGATTTTCCCTTTATTATTTTCGAGATTTATTTTCTTA

ATTCTCTTTAACAAACTAGAAATATTGTATATACAAAAAATCATAAATAATAGATGAA

TAGTTTAATTATAGGTGTTCATCAATCGAAAAAGCAACGTATCTTATTTAAAGTGCG

TTGCTTTTTTCTCATTTATAAGGTTAAATAATTCTCATATATCAAGCAAAGTGACAG

GCGCCCTTAAATATTCTGACAAATGCTCTTTCCCTAAACTCCCCCCATAAAAAAC

CCGCCGAAGCGGGTTTTTACGTTATTTGCGGATTAACGATTACTCGTTATCAGAAC

CGCCCAGGGGGCCCGAGCTTAAGACTGGCCGTCGTTTTACAACACAGAAAGAGT

TTGTAGAAACGCAAAAAGGCCATCCGTCAGGGGCCTTCTGCTTAGTTTGATGCCT

GGCAGTTCCCTACTCTCGCCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGG

TCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC

CACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAA

GGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC

CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGAC

AGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT

GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG

TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCG

CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT

ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTG

GCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA

GAGTTCTTGAAGTGGTGGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTA

TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATC

CGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATT

ACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA

CGCTCAGTGGAACGACGCGCGCGTAACTCACGTTAAGGGATTTTGGTCATGAGCT

TGCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCTTTTACCAATGCTTAATCAGTG

AGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCC
GTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGCGCTGCG
ATGATACCGCGAGAACCACGCTCACCGGCTCCGGATTTATCAGCAATAAACCAGC
CAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATCGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCAT
GTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAG
TTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGT
CATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATA
ATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCG
GGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
CTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGA
GCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAA
TGTTGAATACTCATATTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT
GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTC
AGTGTTACAACCAATTAACCAATTCTGAACATTATCGCGAGCCCATTTATACCTGA
ATATGGCTCATAACACCCCTTGTTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACC
TGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGG
GACTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCA
GTCGAAAGACTGGGCCTTTCGCCCGGGCTAATTATGGGGTGTCGCCCTTATTCGA
CTC

2327 **Mgat1-mCherry-HDR-TI donor (sgRNA5)** (SEQ ID NO: 73)

25 Features: pJ204_Mgat1_5' arm_EF1a_mCherry-BGH_sv40-NeoR-sv40 pA_Mgat1_3' arm_CMV-ZsGreen1-
    BGH

AGTCGGTGTTCACTGTGTTTCCTTACTAAGTCAGGTATGATACTGGATTTTTTTTT
TTTTTTTAATCAGTAATTCTATGGCAGCTTGACTTGAGAAATGGGCGTTAAATGGC
CTTACCTCTCCCTCCTGAACTAGATAGACTCTGGCTGAAAGACACGAGGTGAGAT
30 CTGTCTGGTAGGTACCAGGAGTTGCAGTTTTTCTTTCTTTTTTTTTTTTTTCCCAGT
TCTGTTCCACTAGCATCTTTCCTTGCCTCCTTGTCATACTAACATTGTTTTTCTGAG
GACCCATTAGCTCATTTGAAGTGAAGAGCATGCTGTCCTAACCCTGCAGTCAACT
GCCTTACTTCCCTTTTTCCTTAGTTTCTCACCTGTGCCTATCCTCTTCCCTGTGGCA
TGTGGAATTATTATACATGCTTGAAGTAACAGTTGGTTTTCTCTCCTCTCTTTAG

GTACATGGCTTCTTCCTAGTCCCAGCCCAGAAGAATAAACCCTTAGACTGGGGAA
GTGAGCCAGGCAAGCCAAAGGCAGCCTTGAGCCCTCCCCTTGCCTGCCCTCCCC
TGTGGGGGCCAGGATGCTGAAGAAGCAGTCTGCAGGGCTTGTGCTTTGGGGTGC
TATCCTCTTTGTGGGCTGGAATGCCCTGCTGCTCCTCTTCTTCTGGACACGCCCA
GCCCCTGGCAGGCCCCCCTCAGATAGGGCTATCTCTGATGCCCCTGCCAGCCTC
ACCCGTGAGGTGTTCCGCCTGGCTGAGGACGCTGAGAAGCAGCGTGTGAGGCTC
CGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGG
GGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAAC
TGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTCCCGAGGGTGGGGGAGAA
CCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCG
CCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGG
GTTATGGCCCTTGCGTGCCTTGAATTACTTCCACCTGGCTCCAGTACGTGATTCTT
GATCCCGAGCTGGAGCCAGGGGCGGGCCTTGCGCTTTAGGAGCCCCTTCGCCTC
GTGCTTGAGTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGG
TGGCACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTT
TTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCC
AGGATCTGCACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCC
GTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCG
AGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTC
GCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACC
AGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTCCAGGGGGCTCAAA
ATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGA
AAGGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGG
CGCCGTCCAGGCACCTCGATTAGTTCTGGAGCTTTTGGAGTACGTCGTCTTTAGG
TTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGAC
TGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTG
AGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTC
CATTTCAGGTGTCGTGAAGACGTCATCGCCACCATGGTGAGCAAGGGCGAGGAG
GATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCT
CCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTAC
GAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTT
CGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAG
CACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGT
GGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGAC
TCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAAC

TTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC

TCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAG

GCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAA

GGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGA

CATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGA

GGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGACACAGTCTCT

GTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGA

CCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATC

GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAG

CAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT

CTATGGAGGTCTGAGTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCA

GGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA

GGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATC

TCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAAC

TCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTATTTATG

CAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCT

TTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTC

GGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGAT

TGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGG

CACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGG

GGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCA

GGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAG

CTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAG

TGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCAT

CATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTC

GACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGT

CTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAA

CTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACC

CATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGAT

TCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGG

CTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGT

GCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTT

GACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGC

CCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGG

CTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCT

CATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACA
AATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTA
GTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGAGGTCTGAGTGCT
GTTGCAGCAAATCAGGGAGCATCATGCTTTGTGGAGACAGAGGTGGAAAGTGCC
5   CACCGTGGCCCCTCCAGCCTGGCCCCGTGTGCCTGCGACCCCCTCACCAGCCGT
GATCCCCATCCTGGTCATTGCCTGTGACCGCAGCACTGTCCGGCGCTGCTTGGAT
AAGTTGTTGCACTATCGGCCCTCAGCTGAGCATTTCCCCATCATTGTCAGCCAGG
ACTGCGGGCACGAGGAGACAGCACAGGTCATTGCTTCCTATGGCAGTGCAGTCA
CACACATCCGGCAGCCAGACCTGAGTAACATCGCTGTGCCCCCAGACCACCGCA
10   AGTTCCAGGGTTACTACAAGATCGCCAGGCACTACCGCTGGGCACTGGGCCAGA
TCTTCAACAAGTTCAAGTTCCCAGCAGCTGTGGTAGTGGAGGACGATCTGGAGGT
GGCACCAGACTTCTTTGAGTACTTCCAGGCCACCTACCCACTGCTGAGAACAGAC
CCCTCCCTTTGGTGTGTGTCTGCTTGGAATGACAATGGCAAGGAGCAGATGGTAG
ACTCAAGCAAACCTGAGCTGCTCTATCGAACAGACTTTTTCCTGGCCTTGGCTG
15   GCTGCTGATGGCTGAGCTGTGGACAGAGCTGGAGCCCAAGTGGCCCAAGGCCTT
CTGGGATGACTGGATGCGCAGACCTGAGCAGCGGAAGGGGCGGGCCTGTATTC
GTCCAGAAATTTCAAGAACGATGACCTTTGGCCGTAAGGGACGTCGCTGTTGACA
TTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCC
ATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG
20   CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC
CAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCA
CTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATG
ACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCC
TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT
25   GGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT
CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTT
CCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTAC
GGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTA
CTGGCTTATCGAAATAGTGCGATcgccaccatggcccagtccaagcacggcctgaccaaggagatga
30   ccatgaagtaccgcatggagggctgcgtggacggccacaagttcgtgatcaccggcgagggcatcggctacccccttca
agggcaagcaggccatcaacctgtgcgtggtggagggcggcccccttgcccttcgccgaggacatcttgtccgccgcctt
catgtacggcaaccgcgtgttcaccgagtacccccaggacatcgtcgactacttcaagaactcctgccccgccggctac
acctgggaccgctccttcctgttcgaggacggcgccgtgtgcatctgcaacgccgacatcaccgtgagcgtggaggaga
actgcatgtaccacgagtccaagttctacggcgtgaacttccccgccgacggccccgtgatgaagaagatgaccgaca
35   actgggagccctcctgcgagaagatcatccccgtgcccaagcagggcatcttgaagggcgacgtgagcatgtacctgc

tgctgaaggacggtggccgcttgcgctgccagttcgacaccgtgtacaaggccaagtccgtgccccgcaagatgcccg

actggcacttcatccagcacaagctgacccgcgaggaccgcagcgacgccaagaaccagaagtggcacctgaccg

agcacgccatcgcctccggctccgccttgcccaagcttccgcggagccatggcttcccgccggcggtggcggcgcagg

atgatggcacgctgcccatgtcttgtgcccaggagagcgggatggaccgtcaccctgcagcctgtgcttctgctaggatc

5   aatgtgtagACACAGTCTCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCC

CCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAA

ATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG

GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTG

GGGATGCGGTGGGCTCTATGGACTTGCGTAGTGAGTCGAATAAGGGCGACACAA

10   AATTTATTCTAAATGCATAATAAATACTGATAACATCTTATAGTTTGTATTATATTTT

GTATTATCGTTGACATGTATAATTTTGATATCAAAAACTGATTTTCCCTTTATTATTT

TCGAGATTTATTTTCTTAATTCTCTTTAACAAACTAGAAATATTGTATATACAAAAAA

TCATAAATAATAGATGAATAGTTTAATTATAGGTGTTCATCAATCGAAAAAGCAACG

TATCTTATTTAAAGTGCGTTGCTTTTTTCTCATTTATAAGGTTAAATAATTCTCATAT

15   ATCAAGCAAAGTGACAGGCGCCCTTAAATATTCTGACAAATGCTCTTTCCCTAAAC

TCCCCCCATAAAAAAACCCGCCGAAGCGGGTTTTTACGTTATTTGCGGATTAACGA

TTACTCGTTATCAGAACCGCCCAGGGGGCCCGAGCTTAAGACTGGCCGTCGTTTT

ACAACACAGAAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGGGCCTTCT

GCTTAGTTTGATGCCTGGCAGTTCCCTACTCTCGCCTTCCGCTTCCTCGCTCACT

20   GACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAG

GCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAG

CAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT

TCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG

GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTC

25   CCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTT

CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTT

CGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC

CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA

CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA

30   TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGGCTAACTACGGCTACACTAGA

AGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAG

TTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTT

TGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTT

TTCTACGGGGTCTGACGCTCAGTGGAACGACGCGCGCGTAACTCACGTTAAGGG

35   ATTTTGGTCATGAGCTTGCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCTTTTAC

CAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCAT
AGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCT
GGCCCCAGCGCTGCGATGATACCGCGAGAACCACGCTCACCGGCTCCGGATTTA
TCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACT
5   TTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTC
GCCAGTTAATAGTTTGCGCAACGTTGTTGCCATCGCTACAGGCATCGTGGTGTCA
CGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAG
TTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATC
GTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGC
10   ATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTAC
TCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGG
CGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATT
GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
15   AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAA
GGGCGACACGGAAATGTTGAATACTCATATTCTTCCTTTTTCAATATTATTGAAGCA
TTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATA
AACAAATAGGGGTCAGTGTTACAACCAATTAACCAATTCTGAACATTATCGCGAGC
CCATTTATACCTGAATATGGCTCATAACACCCCTTGTTTGCCTGGCGGCAGTAGC
20   GCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCC
GATGGTAGTGTGGGGACTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAAT
AAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGCCCGGGCTAATTATGGGGT
GTCGCCCTTATTCGACTC

25   **Example 9. CRISPR Cas9 mediated multiplexing in CHO cells**

To accelerate genome engineering of CHO cells, multiplexing of gene editing event is
of high interest. Thus, the capacity of the CRISPR Cas9 system in generating several
gene disruptions in CHO cells simultaneously was investigated.

30   In order to decrease the screening load for the identification of triple gene disrupted
CHO cell lines, a vector expressing GFP with a 2A linkage to Cas9 was constructed.
When transfecting with the GFP_2A_Cas9 vector, the GFP fluorescence from the cells
expressing the vector can be applied for FACS enrichment of the Cas9 expressing
cells e.g. the part of the transfected cell population most likely to have undergone gene
35   editing events. Three sgRNA targeting the FUT8, BAK-1 and BAX loci in CHO-S were

applied for comparison of the GFP_2A_Cas9 activity to the regular Cas9 expression
vector applied in example 6. MiSeq data analysis revelated that the GFP_2A_Cas9
generated the same percentage of indels compared to the regular Cas9 expression
vector.

The applicability of the GFP_2A_Cas9 vector for generating three gene disruptions
simultaneously was tested. For this, CHO-S cells transfected with three sgRNAs were
FACS-sorted on day 2 to enrich for the population of cells expressing GFP e.g. the
Cas9 nuclease. Indeed, FACS sorting allowed enrichment of cells with indels created
at the three loci. The same cell population was furthermore sorted at the single-cell
level and indeed, several of the generated monoclonal cell lines showed three gene
disruption events from the single round of transfection and FACS sorting. These data
highlight the applicability of Cas9 for simultaneous editing of multiple loci.

Construction and validation of GFP 2A linked Cas9 expression vector
The GFP_2A_Cas9 expression vector was constructed with seamless USER cloning of
two PCR products generated from applying primer 6924 and 6925 on plasmid 1182
(CHO codon optimized Cas9 expression vector described in example 6) and primer
6926 and 6927 on plasmid 1787 (template for GFP_2A part). The list of primers is
presented in table 8. The purified PCR fragments were assembled with USER enzyme
(New England Biolabs) and transformed into *E. coli* Mach1 competent cells and plated
on LB with ampicillin. Colonies were selected and plasmids were harvested using
NucleoSpin Plasmid kit (Macherey-Nagel). The plasmids were verified by sequencing.
The vector is illustrated in Figure 14 and the sequences of the generated
GFP_2A_Cas9 (lab id number 2632) expression plasmid are presented below.

Table 8. Primer sequences

Primer name	Purpose	Sequence (5'-3')
6924 (BB_Cas9_2A-link_fw) (SEQ ID NO: 74)	USER PCR primer for Cas9 and plasmid backbone amplification	AGAAGCAUGGACAAGAAATAC TCCATCG
6925 (BB_Cas9_2A-link_rv) (SEQ ID NO: 75)	USER PCR primer for Cas9 and plasmid backbone amplification	ATGCATGGUGGCGGCGCTAG CCAGCTTC
6926 (eGFP_2A_2A-	USER PCR primer for GFP	ACCATGCAUGTGAGCAAGGG

link_fw) (SEQ ID NO: 76)	and 2A amplification	CGAGGAGCTG
6927 (eGFP_2A_2A-link_rv) (SEQ ID NO: 77)	USER PCR primer for GFP and 2A amplification	ATGCTTCUCGATCGTGGGCCA GGATTCTCCTCGACG
7531 (FUT8 gRNA3_F_Nex) (SEQ ID NO: 78)	Nextera PCR primer (miseq) for sequencing of FUT8 target site	TCGTCGGCAGCGTCAGATGTG TATAAGAGACAGTGCCCCAT GACTAGGGATA
7532 (FUT8 gRNA3_R_Nex) (SEQ ID NO: 79)	Nextera PCR primer (miseq) for sequencing of FUT8 target site	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGTCTGCGTT CGAGAAGCTGAAA
8738 (BAK1_1544257_F_Nex) (SEQ ID NO: 80)	Nextera PCR primer (miseq) for sequencing of BAK-1 target site	TCGTCGGCAGCGTCAGATGTG TATAAGAGACAGCAAGGTGGG CTCTCCGTGAT
8739 (BAK1_1544257_R_Nex) (SEQ ID NO: 81)	Nextera PCR primer (miseq) for sequencing of BAK-1 target site	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGCGATGCAA TGGTGCAGTATGAT
8740 (BAX_1345650_F_Nex) (SEQ ID NO: 82)	Nextera PCR primer (miseq) for sequencing of BAX target site	TCGTCGGCAGCGTCAGATGTG TATAAGAGACAGTGTGGATAC TAACTCCCCACG
8741 (BAX_1345650_R_Nex) (SEQ ID NO: 83)	Nextera PCR primer (miseq) for sequencing of BAX target site	GTCTCGTGGGCTCGGAGATGT GTATAAGAGACAGTCCCTGAA CCTCACTACCCC

The CRISPy bioinformatic tool (http://staff.biosustain.dtu.dk/laeb/crispy/) was applied for generating three sgRNAs targeting FUT8, BAK-1 and BAX respectively. The sgRNA target sequences are described in table 9.

5   **Table 9.**

Name	CRISPy identification number	Target sequence
FUT8_681494 (sgRNA3_F) (SEQ ID NO: 84)	681494	GTCAGACGCACTGACAAAGTGGG
BAK-1_1544257 (SEQ ID NO: 85)	1544257	GGAAGCCGGTCAAACCACGTTGG
BAX_1345650 (SEQ	1345650	GCTGATGGCAACTTCAACTGGGG

ID NO: 86)		

The sgRNA expression vectors were constructed as described in example 6. The sgRNA targeting FUT8 (FUT8_681494/sgRNA2_C) is described in example 6. The sgRNAs targeting BAK-1 and BAX were constructed with the oligos described in table 10.

Table 10.

Oligo name	Sequence (5' – 3')
BAK-1_1544257_fwd (SEQ ID NO: 87)	GGAAAGGACGAAACACCGGAAGCCGGTCAAACCACGTGTTTTAGAGCTAGAAAT
BAK-1_1544257_rev (SEQ ID NO: 88)	CTAAAACACGTGGTTTGACCGGCTTCCGGTGTTTCGTCCTTTCCACAAGATAT
BAX_1345650_fwd (SEQ ID NO: 89)	GGAAAGGACGAAACACCGCTGATGGCAACTTCAACTGGTTTTAGAGCTAGAAAT
BAX_1345650_rev (SEQ ID NO: 90)	CCTAAAACCAGTTGAAGTTGCCATCAGCGGTGTTTCGTCCTTTCCACAAGATAT

CHO-S cells from Life Technologies were cultured in CD CHO medium (Life Technologies) supplemented with 8 mM L-Glutamine and cultivated in Corning erlenmeyer shake flasks (Sigma-Aldrich). The cells were incubated at 37°C, 5 % CO2 with 120 rpm shaking and passaged every 2-3 days. To validate the genome editing capacity of the constructed GFP_2A_Cas9 plasmid, CHO-S cells were transfected with expression vectors encoding regular Cas9 or the new GFP_2A_Cas9 construct together with sgRNAs targeting the FUT8, BAK-1 or BAX target site. For this, 3 x 106 cells were transfected with up to 3.75 µg of DNA using FreeStyle MAX reagent together with OptiPRO SFM medium (Life Technologies) according to manufacturer's recommendations. On the day after transfection, anticlumping agent was added to the cells (1.5 µl/ml, Life Technologies). Three days after transfection, 500.000 cells were harvested by centrifugation and the pellet was stored at -20°C. The percentage of

indels generated by Cas9 or GFP_2A_Cas9 expression vectors were analyzed for each of the three FUT8, BAK-1 and BAX sgRNAs by MiSeq sequencing of the target loci (Figure 15). For MiSeq primers, see table 8.

5   When comparing the percentage of indels generated at either FUT8, BAK-1 or BAX target site from co-expression of either Cas9 or GFP_2A_Cas9, no major difference in the activity was observed. It was therefore concluded that GFP 2A labelling of Cas9 did not interfere with the genome editing capacity of the nuclease.

10   *GFP FACS enrichment increases the percentage of indels created in sorted cells*

To facilitate multiplexing and FACS enrichment of genome edited cells, CHO-S cells were transfected with GFP_2A_Cas9 together with sgRNAs targeting FUT8, BAK-1 and BAX site simultaneously. For this, 14×10^6 cells were transfected with up to 17.5 µg of DNA using FreeStyle MAX reagent together with OptiPRO SFM medium (Life

15   Technologies) according to manufacturer's recommendations. Anticlumping agent (1.5 µl/ml, Life Technologies) was added on the day after transfection. On day 2, the transfected cells were analyzed on a BD FACS Jazz cell sorter. 10.3% of the cells transfected with GFP_2A_Cas9 and three different gRNAs simultaneously were GFP positive (Figure 16).

20

To analyze the percentage of indels created at the three sgRNA target sites in the GFP positive cell population, 500.000 cells were harvested by centrifugation both before FACS sorting and after FACS sorting for GFP positive cells. Genomic DNA was prepared using Quick extract as described in example 6 and the three loci were MiSeq

25   sequenced. Indeed, FACS sorting enriched for genome editing activities and up to 68%, 67% and 78% of indels were generated at the FUT8, BAK-1 and BAX sgRNA target sites respectively (Figure 17).

Triple gene disrupted clones generated by multiplexing with CRISPR Cas9

30   CHO-S cells transfected with GFP_2A_Cas9 together with the three sgRNA targeting FUT8, BAK-1 and BAX were furthermore single cells sorted applying the same gates as set in Figure 16. Cells were sorted (1 cell per well) into Corning Falcon 96 U-well plates with 100 ul per well of CD-CHO medium supplemented with 8 mM L-glutamine,

penicillin and streptomycin (1:100) and 20% conditioned medium. After 10 days, 100 µl
CD CHO supplemented with 8 mM L-glutamine, P/S (1:100) and 4 µl/ml anticlumping
agent to reach a final of 2 µl/ml of anticlumping agent were added to each well. 14 days
after single cell sorting, the clones were moved to flat bottom Corning Falcon 96 well

5   plates. The medium was changed twice a week. When reaching densities close to
confluency in the wells, the clones were split in two separate 96 well plates (one for
further cultivation and one for miseq sequencing). When reaching close to confluency,
the clones in the miseq plate was harvested by centrifugation and the pellets stored at -
20°C. Genomic DNA was extracted as in example 6 and the three loci were MiSeq

10   sequenced. Miseq analysis was performed on 96 clones. Of these, only 44 of the
clones showed consistent MiSeq data allowing analysis if the clones showed wild-type,
single gene deletion, double gene deletion and triple gene deletion. See data in Table
11. Of the 44 clones, 25 clones were wild-type, 2 clones had 1 gene disruption, 8
clones had two gene disruptions and 9 clones had three gene disruptions. A few clones

15   were especially impressive with a total of five gene disruptions in the three genes. For
example, clone number 87 have 4 bp and 10 bp deletions in the two FUT8 alleles, 10
bp and 13 bp deletions in the two BAK-1 alleles and 4 bp deletion in the single BAX
allele. In conclusion, we were able to generate monoclonal cell lines with multiple gene
disruptions simultaneously with the CRISPR Cas9 system.

20

Table 11.

Index	Product	nts	WT count	WT	Indels	Knockout	Peaks	Peak 1	Peak 2	Peak 3
1	FUT8_ 681494	18682	18660	99,88%	0,12%		1	0		
1	BAK1_ 1544257	79188	79009	99,77%	0,23%		1	0		
1	BAX_	80493	80224	99,67%	0,33%	0	1	0		

	13456 50									
2	FUT8_ 681494	9979	9965	99,86%	0,14%		1	0		
2	BAK1_ 1544257	45222	45121	99,78%	0,22%		1	0		
2	BAX_ 1345650	45355	45231	99,73%	0,27%	0	1	0		
3	FUT8_ 681494	18085	18065	99,89%	0,11%		1	0		
3	BAK1_ 1544257	77702	77540	99,79%	0,21%		1	0		
3	BAX_ 1345650	79928	79679	99,69%	0,31%	0	1	0		
5	FUT8_ 681494	14595	14585	99,93%	0,07%		1	0		
5	BAK1_ 1544257	70637	70484	99,78%	0,22%		1	0		
5	BAX_ 1345650	69541	69361	99,74%	0,26%	0	1	0		
7	FUT8_ 681494	15332	15314	99,88%	0,12%		1	0		
7	BAK1_ 1544257	75063	74923	99,81%	0,19%		1	0		
7	BAX_ 1345650	92755	92506	99,73%	0,27%	0	1	0		
9	FUT8_ 681494	6381	6375	99,91%	0,09%		1	0		
9	BAK1_ 1544257	32314	32266	99,85%	0,15%		1	0		
9	BAX_ 1345650	26243	26152	99,65%	0,35%	0	1	0		
11	FUT8_ 681494	17004	16986	99,89%	0,11%		1	0		

11	BAK1_ 1544257	84651	84467	99,78%	0,22%		1	0		
11	BAX_ 1345650	84467	84206	99,69%	0,31%	0	1	0		
12	FUT8_ 681494	11851	11832	99,84%	0,16%		1	0		
12	BAK1_ 1544257	94059	93850	99,78%	0,22%		1	0		
12	BAX_ 1345650	98495	98205	99,71%	0,29%	0	1	0		
13	FUT8_ 681494	22836	22812	99,89%	0,11%		1	0		
13	BAK1_ 1544257	103364	102816	99,47%	0,53%		1	0		
13	BAX_ 1345650	61267	60901	99,40%	0,60%	0	1	0		
14	FUT8_ 681494	11380	14	0,12%	99,88 %		2	35	-4	
14	BAK1_ 1544257	47424	159	0,34%	99,66 %		2	62	2	
14	BAX_ 1345650	11550	11409	98,78%	1,22%	2	1	0		
23	FUT8_ 681494	12816	78	0,61%	99,39 %		1	1		
23	BAK1_ 1544257	49419	428	0,87%	99,13 %		1	-4		
23	BAX_ 1345650	42223	41855	99,13%	0,87%	2	1	0		
24	FUT8_ 681494	24594	24565	99,88%	0,12%		1	0		
24	BAK1_ 1544257	91694	276	0,30%	99,70 %		1	-13		
24	BAX_ 1345650	92741	3012	3,25%	96,75 %	2	1	-2		

31	FUT8_68194	23495	23415	99,66%	0,34%		1	0		
31	BAK1_1544257	109343	108881	99,58%	0,42%		1	0		
31	BAX_1345650	80827	3408	4,22%	95,78%	1	1	-37		
32	FUT8_681494	14481	14417	99,56%	0,44%		1	0		
32	BAK1_1544257	72212	71829	99,47%	0,53%		1	0		
32	BAX_1345650	62236	62007	99,63%	0,37%	0	1	0		
33	FUT8_681494	5109	5085	99,53%	0,47%		1	0		
33	BAK1_1544257	27653	27509	99,48%	0,52%		1	0		
33	BAX_1345650	23638	23507	99,45%	0,55%	0	1	0		
34	FUT8_681494	13617	13563	99,60%	0,40%		1	0		
34	BAK1_1544257	63610	63240	99,42%	0,58%		1	0		
34	BAX_1345650	62646	62355	99,54%	0,46%	0	1	0		
36	FUT8_681494	30028	52	0,17%	99,83%		2	1	-8	
36	BAK1_1544257	97591	487	0,50%	99,50%		2	1	-10	
36	BAX_1345650	95504	5425	5,68%	94,32%	3	1	-13		
43	FUT8_681494	7216	31	0,43%	99,57%		1	-2		
43	BAK1_1544257	32049	469	1,46%	98,54%		1	-14		

43	BAX_ 1345650	10100	9894	97,96%	2,04%	2	1	0		
47	FUT8_ 681494	40394	40326	99,83%	0,17%		1	0		
47	BAK1_ 1544257	192086	191536	99,71%	0,29%		1	0		
47	BAX_ 1345650	174340	173707	99,64%	0,36%	0	1	0		
48	FUT8_ 681494	12745	12684	99,52%	0,48%		1	0		
48	BAK1_ 1544257	99759	439	0,44%	99,56 %		2	-14	-8	
48	BAX_ 1345650	49987	1362	2,72%	97,28 %	2	1	1		
51	FUT8_ 681494	37107	48	0,13%	99,87 %		1	1		
51	BAK1_ 1544257	117005	263	0,22%	99,78 %		1	-28		
51	BAX_ 1345650	94985	94485	99,47%	0,53%	2	1	0		
52	FUT8_ 681494	14023	13966	99,59%	0,41%		1	0		
52	BAK1_ 1544257	61089	60707	99,37%	0,63%		1	0		
52	BAX_ 1345650	74698	74330	99,51%	0,49%	0	1	0		
54	FUT8_ 681494	12876	15	0,12%	99,88 %		2	-1	-5	
54	BAK1_ 1544257	80300	174	0,22%	99,78 %		2	87	-14	
54	BAX_ 1345650	108677	4760	4,38%	95,62 %	3	1	71		
56	FUT8_ 681494	33112	23	0,07%	99,93 %		2	117	-1	

56	BAK1_1544257	73097	394	0,54%	99,46%		1	-5		
56	BAX_1345650	64062	6273	9,79%	90,21%	3	1	-2		
57	FUT8_681494	10655	6	0,06%	99,94%		1	-1		
57	BAK1_1544257	40452	109	0,27%	99,73%		1	-13		
57	BAX_1345650	40207	611	1,52%	98,48%	3	1	-1		
64	FUT8_681494	15600	28	0,18%	99,82%		2	-20	-1	
64	BAK1_1544257	51660	51399	99,49%	0,51%		1	0		
64	BAX_1345650	95285	94929	99,63%	0,37%	1	1	0		
65	FUT8_681494	40218	32	0,08%	99,92%		1	59		
65	BAK1_1544257	85525	677	0,79%	99,21%		1	1		
65	BAX_1345650	77133	5397	7,00%	93,00%	3	1	1		
67	FUT8_681494	19155	19100	99,71%	0,29%		1	0		
67	BAK1_1544257	99142	98736	99,59%	0,41%		1	0		
67	BAX_1345650	110305	109900	99,63%	0,37%	0	1	0		
68	FUT8_681494	15867	15802	99,59%	0,41%		1	0		
68	BAK1_1544257	78876	78342	99,32%	0,68%		1	0		
68	BAX_1345650	63627	63274	99,45%	0,55%	0	1	0		

69	FUT8_ 681494	7408	7385	99,69%	0,31%		1	0		
69	BAK1_ 1544257	46556	46427	99,72%	0,28%		1	0		
69	BAX_ 1345650	19289	19175	99,41%	0,59%	0	1	0		
70	FUT8_ 681494	16190	19	0,12%	99,88 %		1	-1		
70	BAK1_ 1544257	84119	497	0,59%	99,41 %		2	1	-2	
70	BAX_ 1345650	61238	1673	2,73%	97,27 %	3	1	-8		
73	FUT8_ 681494	18158	18112	99,75%	0,25%		1	0		
73	BAK1_ 1544257	89109	88831	99,69%	0,31%		1	0		
73	BAX_ 1345650	82668	82392	99,67%	0,33%	0	1	0		
74	FUT8_ 681494	12311	12278	99,73%	0,27%		1	0		
74	BAK1_ 1544257	58106	57958	99,75%	0,25%		1	0		
74	BAX_ 1345650	51038	50875	99,68%	0,32%	0	1	0		
75	FUT8_ 681494	24751	24723	99,89%	0,11%		1	0		
75	BAK1_ 1544257	111244	110932	99,72%	0,28%		1	0		
75	BAX_ 1345650	103628	102727	99,13%	0,87%	0	1	0		
77	FUT8_ 681494	17911	17883	99,84%	0,16%		1	0		
77	BAK1_ 1544257	87181	86904	99,68%	0,32%		1	0		

77	BAX_ 1345650	84167	83925	99,71%	0,29%	0	1	0		
79	FUT8_ 681494	17456	17423	99,81%	0,19%		1	0		
79	BAK1_ 1544257	81727	81470	99,69%	0,31%		1	0		
79	BAX_ 1345650	102928	102634	99,71%	0,29%	0	1	0		
80	FUT8_ 681494	21365	21325	99,81%	0,19%		1	0		
80	BAK1_ 1544257	99746	99450	99,70%	0,30%		1	0		
80	BAX_ 1345650	89009	88730	99,69%	0,31%	0	1	0		
81	FUT8_ 681494	10777	10753	99,78%	0,22%		1	0		
81	BAK1_ 1544257	39288	39146	99,64%	0,36%		1	0		
81	BAX_ 1345650	26199	26090	99,58%	0,42%	0	1	0		
85	FUT8_ 681494	12093	20	0,17%	99,83 %		1	-5		
85	BAK1_ 1544257	81864	465	0,57%	99,43 %		2	-2	98	
85	BAX_ 1345650	85296	4578	5,37%	94,63 %	3	1	-1		
87	FUT8_ 681494	24663	21	0,09%	99,91 %		2	-4	-10	
87	BAK1_ 1544257	109915	352	0,32%	99,68 %		2	-10	-13	
87	BAX_ 1345650	111986	6664	5,95%	94,05 %	3	1	-4		
88	FUT8_ 681494	11829	6	0,05%	99,95 %		1	-10		

88	BAK1_ 1544257	114771	274	0,24%	99,76 %		2	47	-1	
88	BAX_ 1345650	60776	60418	99,41%	0,59%	2	1	0		
89	FUT8_ 681494	21087	19	0,09%	99,91 %		1	-8		
89	BAK1_ 1544257	123857	623	0,50%	99,50 %		1	1		
89	BAX_ 1345650	88189	87830	99,59%	0,41%	2	1	0		
91	FUT8_ 681494	12221	20	0,16%	99,84 %		2	-2	-23	
91	BAK1_ 1544257	56027	370	0,66%	99,34 %		2	-19	-14	
91	BAX_ 1345650	137793	7295	5,29%	94,71 %	3	1	88		
92	FUT8_ 681494	18613	18544	99,63%	0,37%		1	0		
92	BAK1_ 1544257	95209	94622	99,38%	0,62%		1	0		
92	BAX_ 1345650	78199	77813	99,51%	0,49%	0	1	0		

GFP_2A_Cas9 vector (SEQ ID NO: 91)

2632: GFP_2A_Cas9

Features: CMV, GFP, 2A, Cas9, bghpA, sv40, hygR, sv40pA, amp

5   CTCATGACCAAAATCCCTTAACGTGAGTTACGCGCGCGTCGTTCCACTGAGCGTC
       AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAA
       TCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGA
       TCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATA
       CCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGCCCACCACTTCAAGAACTCTGT
10   AGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGT

GGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGG
CGCAGCGGTCGGGCTGAACGGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGA
ACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACG
CTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAAC
5   AGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCC
TGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCC
TTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGAT
AACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACC
10   GAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGGCGAGAGTAGGGAACTGCC
AGGCATCAAACTAAGCAGAAGGCCCCTGACGGATGGCCTTTTTGCGTTTCTACAA
ACTCTTTCTGTGTTGTAAAACGACGGCCAGTCTTAAGCTCGGGCCCCCTGGGCGG
TTCTGATAACGAGTAATCGTTAATCCGCAAATAACGTAAAAACCCGCTTCGGCGG
GTTTTTTTATGGGGGGAGTTTAGGGAAAGAGCATTTGTCAGAATATTTAAGGGCGC
15   CTGTCACTTTGCTTGATATATGAGAATTATTTAACCTTATAAATGAGAAAAAGCAA
CGCACTTTAAATAAGATACGTTGCTTTTTCGATTGATGAACACCTATAATTAAACTA
TTCATCTATTATTTATGATTTTTTGTATATACAATATTTCTAGTTTGTTAAAGAGAATT
AAGAAAATAAATCTCGAAAATAATAAAGGGAAAATCAGTTTTTGATATCAAAATTAT
ACATGTCAACGATAATACAAAATATAATACAAACTATAAGATGTTATCAGTATTTATT
20   ATCATTTAGAATAAATTTTGTGTCGCCCTTAATTGTGAGCGGATAACAATTACGAG
CTTCATGCACAGTGGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATT
ACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGT
AAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAAT
GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTG
25   GAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAG
TACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAG
TACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGC
TATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTT
GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTT
30   GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGAC
GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTG
GCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATA
GGGGAAGCTGGCTAGCGCCGCCACCATGCATGTGAGCAAGGGCGAGGAGCTGT
TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACA
35   AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC

CTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG
ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAG
CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG

5   GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
GGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATA
TCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCAT
CGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAA

10   GCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT
GACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGCCTAGGGGCA
GTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTG
GCCCACGATCGAGAAGCATGGACAAGAAATACTCCATCGGCCTTGATATCGGGAC
CAACAGCGTGGGTTGGGCCGTGATCACTGACGAATACAAAGTGCCATCGAAAAAG

15   TTTAAGGTGCTGGGCAATACTGACAGACATAGCATCAAGAAAAACCTGATTGGAG
CCCTCCTGTTCGACTCCGGGGAAACCGCCGAAGCCACTCGCCTGAAACGGACCG
CACGCCGGAGGTACACTCGCCGGAAGAATAGAATCTGCTACCTTCAAGAGATCTT
TAGCAACGAGATGGCCAAAGTGGACGATTCGTTCTTCCACCGGCTGGAGGAGTC
ATTCCTGGTCGAAGAGGACAAAAAGCATGAGAGACATCCGATCTTCGGTAACATT

20   GTGGATGAGGTCGCCTACCACGAGAAGTACCCCACTATCTACCATTTGCGCAAGA
AGTTGGTCGACTCGACTGATAAGGCCGACCTCAGACTCATCTACCTCGCTTTGGC
ACACATGATCAAGTTTAGGGGTCACTTCCTGATTGAAGGGGATCTGAACCCAGAC
AACTCGGATGTGGATAAACTGTTTATCCAGCTGGTCCAGACTTACAATCAGCTCTT
TGAGGAGAACCCGATCAATGCCTCCGGCGTGGATGCTAAGGCAATCCTGTCCGC

25   ACGCTTGTCAAAGAGCCGGCGGCTCGAAAATCTGATCGCACAGCTTCCGGGAGA
AAAGAAAAACGGACTGTTTGGAAACCTTATCGCCCTCTCGCTGGGACTCACTCCT
AACTTCAAATCCAATTTTGATTTGGCCGAGGATGCCAAGCTGCAGCTGTCGAAGG
ACACCTACGACGATGACCTCGACAATCTTCTGGCCCAGATCGGGGACCAGTACG
CCGATCTGTTCCTCGCGGCCAAAAACCTGTCAGACGCAATCTTGCTGAGCGATAT

30   CTTGAGGGTCAATACCGAAATCACTAAGGCTCCATTGTCAGCATCGATGATCAAG
AGGTACGATGAACACCATCAGGACCTGACCCTGCTCAAGGCACTGGTGCGGCAG
CAACTGCCGGAAAAGTACAAGGAGATCTTCTTCGATCAAAGCAAGAATGGGTACG
CAGGGTACATCGATGGTGGAGCATCACAAGAAGAGTTCTACAAATTCATCAAGCC
CATCCTTGAAAAGATGGACGGGACGGAAGAACTGCTGGTGAAGCTGAATCGGGA

35   AGATCTGCTGCGGAAGCAGCGCACCTTCGACAATGGTTCCATCCCACACCAAATC

94

```
CATCTCGGCGAACTGCACGCGATCCTTCGCCGGCAGGAAGATTTCTACCCGTTCT
TGAAGGATAACAGAGAAAGATCGAGAAAATCCTGACCTTTAGAATCCCGTACTAC
GTGGGCCCGTTGGCTCGCGGAAACTCAAGATTCGCCTGGATGACTAGAAAATCC
GAAGAAACGATTACCCCGTGGAACTTTGAAGAGGTCGTCGACAAGGGAGCCTCG
GCACAGTCGTTCATCGAACGGATGACCAATTTCGACAAGAACCTCCCCAACGAAA
AGGTGCTCCCGAAACACTCGCTGCTGTATGAGTACTTCACGGTCTACAACGAGCT
GACCAAGGTGAAGTACGTGACCGAAGGAATGCGGAAGCCCGCTTTTCTGTCCGG
TGAACAGAAAAAGGCCATCGTGGACCTCCTCTTCAAGACTAACAGAAAGGTGACC
GTGAAACAGCTTAAAGAGGACTACTTCAAGAAGATCGAGTGTTTTGACTCCGTGG
AAATCTCGGGAGTGGAGGATAGGTTTAACGCTTCGCTGGGGACCTACCATGACCT
TCTCAAGATCATCAAGGATAAGGATTTCCTGGACAACGAAGAAACGAGGACATC
CTTGAGGACATCGTGTTGACCTTGACCCTGTTTGAGGATCGGGAAATGATCGAGG
AGAGACTGAAAACTTACGCGCATCTTTTCGACGACAAAGTGATGAAGCAGCTGAA
AAGAAGAAGATACACTGGCTGGGGACGGCTGTCGAGGAAACTGATTAACGGAATT
AGGGATAAACAAAGCGGAAAGACTATCCTCGATTTTCTCAAGTCGGACGGCTTCG
CCAATCGGAACTTTATGCAGCTCATCCACGACGATAGCCTGACCTTCAAGGAAGA
TATCCAAAAAGCCCAAGTGTCCGGCCAGGGAGATAGCTTGCACGAGCACATTGCT
AACCTGGCCGGTTCACCAGCCATTAAGAAGGGAATCCTCCAAACCGTGAAGGTG
GTCGACGAACTCGTCAAGGTGATGGGCCGCCATAAGCCAGAGAACATCGTGATC
GAGATGGCTCGCGAAAATCAGACTACTCAGAAGGGTCAAAAGAACTCCCGCGAG
CGCATGAAGCGCATCGAAGAAGGAATTAAGGAGCTGGGCTCACAAATCCTCAAAG
AACATCCTGTCGAGAACACCCAACTTCAGAACGAGAAACTGTACCTCTACTATCTC
CAAAATGGCCGGGACATGTACGTCGATCAGGAATTGGATATCAACCGCCTCTCAG
ACTACGACGTGGATCATATCGTGCCGCAGTCGTTTCTCAAAGATGATAGCATCGA
CAACAAGGTCCTGACGAGGTCGGATAAGAACCGCGGGAAATCCGACAATGTGCC
TTCGGAAGAGGTGGTGAAAAGATGAAGAACTACTGGAGGCAATTGCTTAATGCC
AAACTGATCACCCAGCGCAAATTCGACAACCTTACTAAGGCCGAGCGCGGAGGA
CTGTCCGAACTCGATAAGGCGGGGTTCATCAAAAGACAATTGGTCGAAACTCGGC
AAATTACCAAGCATGTCGCTCAGATCCTCGACTCGCGCATGAACACTAAGTATGAT
GAGAACGACAAGCTCATTCGCGAAGTGAAAGTGATTACCCTCAAATCAAAGCTGG
TGTCGGACTTCAGAAAAGACTTTCAATTCTACAAGGTGCGGGAGATTAACAACTAC
CACCACGCGCACGACGCCTACCTCAATGCAGTGGTGGGAACCGCCCTCATCAAG
AAGTATCCAAAGCTCGAAAGCGAGTTCGTGTACGGGGACTATAAAGTGTACGATG
TGCGGAAAATGATCGCAAGAGCGAGCAAGAGATCGGCAAAGCAACGGCCAAAT
ACTTCTTCTACAGCAACATTATGAACTTTTTCAAGACCGAGATTACCCTGGCGAAC
```

GGAGAAATCCGGAAGCGGCCTTTGATCGAAACGAATGGGGAAACTGGAGAAATC
GTGTGGGACAAGGGCAGAGACTTTGCGACCGTGAGAAAGGTGCTGTCCATGCCA
CAAGTCAACATCGTGAAGAAACTGAAGTGCAGACTGGAGGATTTTCCAAGGAGT
CAATCCTTCCGAAGCGCAACAGCGACAAGCTGATCGCCAGGAAGAAGGATTGGG
ACCCCAAGAAGTACGGTGGTTTTGATTCACCTACTGTCGCTTACAGCGTGCTGGT
GGTGGCCAAGGTGGAAAAGGGGAAGTCAAAGAAATTGAAGTCGGTGAAGGAATT
GCTGGGGATTACTATCATGGAGAGGAGCAGCTTCGAAAAGAATCCCATCGACTTC
TTGGAGGCCAAGGGATACAAAGAGGTGAAGAAAGACCTGATCATTAAGCTGCCGA
AGTACTCCTTGTTCGAACTGGAAAACGGCAGAAAGCGGATGCTGGCCTCCGCCG
GAGAGCTGCAGAAGGGTAACGAGCTGGCTCTGCCCAGCAAATACGTGAATTTCCT
CTACCTGGCCTCCCATTACGAGAAGCTCAAGGGAAGCCCGGAGGATAATGAGCA
AAAACAACTGTTCGTCGAGCAGCACAAGCACTACCTCGACGAGATCATCGAACAA
ATCTCCGAGTTCTCGAAGCGGGTGATTCTGGCCGACGCAAACCTTGATAAAGTCC
TCAGCGCCTACAACAAGCACCGCGACAAACCGATCAGAGAACAAGCTGAGAACAT
CATCCACCTGTTCACGCTGACTAACTTGGGAGCGCCTGCCGCCTTCAAGTACTTC
GACACCACTATCGATCGGAAACGGTACACTTCCACTAAGGAAGTGCTGGACGCAA
CCCTGATCCATCAGTCCATCACCGGACTTTATGAGACTCGGATCGATCTCAGCCA
GCTTGGCGGTGATTCAAGAGCTGACCCGAAGAAGAAGCGCAAGGTCTAGAAAAT
CAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGT
GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAG
GAAATTGCATCACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGA
TTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACG
CGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCC
CCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTG
GAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA
GTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCC
CAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTATTTATGCAGAGG
CCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGG
AGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCT
GATCAGCACGTGATGAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTC
TGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAG
AATCTCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAA
TAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGG
CCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGA
CCTATTGCATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAAC

CGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGC
GGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGG
TCAATACACTACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGTGTATC
ACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCG

5   ATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACG
CGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCAT
TGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCGCCAACATCTT
CTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCG
GAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCAT

10   TGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTT
GGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGG
CGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAA
GTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAA
TAGCACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCT

15   TCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCA
TGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAA
TAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGT
TGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCT
AGCTAGAGCTTGGCGTAATCATGGTCATTACCAATGCTTAATCAGTGAGGCACCTA

20   TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGCGCTGCGATGATACCG
CGAGAACCACGCTCACCGGCTCCGGATTTATCAGCAATAAACCAGCCAGCCGGA
AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA
ATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGT

25   TGTTGCCATCGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCA
TTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
AAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC
AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAG

30   TGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCG
CCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAA
ACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCA
CCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAAC
AGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATA

35   CTCATATTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGA

GCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTCAGTGTTACA
ACCAATTAACCAATTCTGAACATTATCGCGAGCCCATTTATACCTGAATATGGCTC
ATAACACCCCTTG

5   **Example 10. Accelerated clone generation with clean and precise integration of transgenes**

Traditionally, development of recombinant CHO (rCHO) cell lines relies on random integration of the gene of interests (GOI) into the genome, followed by selection of cells carrying the transgenes. However, lack of control of gene insertion can give rise to

10   unwanted phenotypic heterogeneity due to the varying accessibility of integration sites for gene expression – termed the position effect variation. In addition, gene amplification methods are traditionally used to increase expression. As a result, these cell lines are often unstable and show reduced production over time. Due to this variation in expression and genomic composition, subsequent screening of multiple

15   clones is therefore necessary to select proper clones suitable for high and stable expression of recombinant proteins. Precise targeting of transgenes into specific desirable sites in the CHO genome reduces the variation in expression, generating a uniform population with stable transgene expression.

20   To shorten the time required to generate monoclonal cell lines with improved clonal stability while avoiding the application of antibiotic selection, a new targeted integration method was developed to generate clones with site-specific, marker free (no antibiotic selection needed) and clean (no unwanted DNA present) targeted integration of transgenes into the genome. The method is based on CRISPR Cas9 genome editing

25   and homology-directed DNA repair (HDR) mediated targeted integration to obtain controlled and precise integration of transgenes. Applying this method, a pool of cells with targeted integration events of genes encoding the three therapeutic proteins EPO, Enbrel and Rituximab were identified by 3'/5' junction PCR. These pools of cells were isolated from a simple FACS enrichment applying transiently expressed fluorescent

30   proteins and without the use of antibiotic selection. These data support the hypothesis that the CRISPR Cas9 and homology-directed repair combined with FACS enrichment using transiently expressed fluorescent markers can accelerate the generation of stable cell lines with site-specific integrated transgenes without the use of antibiotic selection.

35

98

The targeted integration system is based on three DNA parts: (a) a vector expressing Cas9 which is 2A-linked to a fluorescent marker (for example GFP); (b) donor DNA with homology arms towards the integration site containing an expression cassette inside the donor arms and a fluorescent marker gene (for example mcherry) outside

5   the homology arms; and (c) a sgRNA targeting the selected integration site (Figure 18).

The fluorescent marker outside the homology arms and the fluorescent marker linked to Cas9 facilitate sorting of cells transfected with both Cas9 and the donor DNA. In this way the unwanted non-fluorescent or single fluorescent cells can be sorted away to

10   enrich for a pool of double positive cells with enriched targeted integration events. The fluorescent marker outside the homology regions of the donor DNA is not integrated if the cells apply HDR mediated DNA repair. Only the part within the homology arms will be inserted; this part could contain an expression cassette for expression of a gene of interest (GOI), for example a biopharmaceutical (Figure 19).

15

The clones expressing the fluorescent marker from the donor DNA are likely to show random integration of the donor plasmid and can be discarded by simple screening. The remaining non-fluorescent clones can be screened by a simple junction PCR to identify targeted integrants.

20

Construction of donor DNA

As a proof of concept, three donor DNA plasmids expressing the therapeutic protein EPO, Rituximab and Enbrel respectively were constructed for integration into COSMC. The fluorescent protein mCherry was cloned outside the homology arm according to

25   Figure 18. The donor DNA was constructed by USER cloning of PCR fragments as described previously (example 9, targeted integration). A list of the PCR fragments assembled is given in table 12 including the primers and templates applied. The sequence of the primers is given in table 13. The purified PCR fragments were assembled with USER enzyme (New England Biolabs) and transformed into E. coli

30   Mach1 competent cells and plated on LB with ampicillin. Colonies were selected and plasmids were harvested using NucleoSpin Plasmid kit (Macherey-Nagel). The plasmids were verified by sequencing.

Table 12: PCR products for USER cloning of donor DNA plasmids

PCR product	Primer pair	Template DNA
5' arm (DB0034)	PR0125 and PR0126	CHO-S genomic DNA
EF1a (DB0036)	PR0121 and PR0122	labID1053_pBudCE4.1
coEPO-BGH (DB0032)	PR0123 and PR0038	labID1593_coEPO-pcDNA3.1(+)
coEnbrel-BGH (DB0064)	PR0603 and PR0038	labID1070_Enbrel-pcDNA3.1(+)
Rituximab HC-BGH (DB0062)	PR0602 and PR0064	labID2699_pLMG3
Rituximab LC-BGH (DB0063)	PR0604 and PR0038	labID1068_Rituximab-pBudCE4.1
3' arm (DB0035)	PR0127 and PR0128	CHO-S genomic DNA
CMV-mCherry-BGH (DB0061)	PR0048 and PR0050	labID1968_mcherry-TI-cosmc
Backbone (DB0018)	PR0023 and PR0043	labID783_pMEV4-GFP

Table 13: Uracil primer sequences

Primer	Name	**Sequence**
PR0125 (SEQ ID NO: 92)	COSMC 5' arm_750bp_LA_fwd	agtcggtgUGTAATCCATGGAGG AGTTTCT
PR0126 (SEQ ID NO: 93)	COSMC 5' arm_750bp_LB_rev	acgctgctUAAGGTCTCCAGATTT TACAGT
PR0121 (SEQ ID NO: 94)	EF-1a_LB_fwd	aagcagcgUGTGAGGCTCCGGT GCCC
PR0122 (SEQ ID NO: 95)	EF-1a_LC_rev	atgacgtcUTCACGACACCTGAA ATGGAA
PR0123 (SEQ ID NO: 96)	kozak_co-EPO_LC_fwd	agacgtcaUCGCCACCATGGGA GTGCACG
PR0038 (SEQ ID NO: 97)	BGH pA_LD_rev	actcagaccUccatagagcccaccgca tcc
PR0603 (SEQ ID NO: 98)	kozak_Enbrel_LC_fwd	agacgtcaUAGCACCATGGCGC CCGTCG
PR0602 (SEQ ID NO: 99)	kozak_Rituximab HC_LC_fwd	agacgtcaUGACACCATGGGCT GGTCCTG
PR0064 (SEQ ID NO: 100)	BGH pA_O2_rev	ATCGCACUccatagagcccaccgc Atcc
PR0604 (SEQ ID NO: 101)	EF-1a_O2_fwd	AGTGCGAUGTGAGGCTCCGG TGCCC

PR0127 (SEQ ID NO: 102)	COSMC 3' arm_750bp_LD_fwd	aggtctgagUGATTGTCTTAAGC ATAGAGTC
PR0128 (SEQ ID NO: 103)	COSMC 3' arm_750bp_O1_rev	AGCGACGUCCTCATTTGCAT ATATTTGAA
PR0023 (SEQ ID NO: 104)	pJ204 backbone_O5_fwd	ACTTGCGUAGTGAGTCGAAT AAGGGCGACACAAA
PR0043 (SEQ ID NO: 105)	pJ204 backbone_LA_rev	acaccgacUGAGTCGAATAAG GGCGACACCCCA

CHO-S cells from Life Technologies were cultured in CD CHO medium (Life Technologies) supplemented with 8 mM L-Glutamine and cultivated in Corning erlenmeyer shake flasks (Sigma-Aldrich). The cells were incubated at 37°C, 5 % CO_2

5   with 120 rpm shaking and passaged every 2-3 days. CHO-S cells were transfected with the expression vectors encoding GFP_2A_Cas9, sgRNAs targeting COSMC and one of the three donor DNA encoding EPO (SEQ ID NO: 107), Enbrel (SEQ ID NO: 108) or rituximab (SEQ ID NO: 109). For this, 3 x 10^6 cells were transfected with up to 3.75 µg of DNA using FreeStyle MAX reagent together with OptiPRO SFM medium (Life

10   Technologies) according to manufacturer's recommendations. Two days after transfection, cells transfected with combinations of GFP_2A_Cas9, sgRNA2_C and donor DNA (EPO, Enbrel or Rituximab) were bulk sorted using a BD FACS Jazz cell sorter for GFP and/or mCherry positive cells. 300,000 to 500,000 cells were recovered in Corning Falcon MD24 plate with a working volume of 1 ml. At the day after bulk

15   sorting (D1) and at six days after bulk sorting (D6), cells were harvested by centrifugation for analysis. An overview of the work flow is given in Figure 20.

Genomic DNA from cell pellets was prepared as described in example 6. 5'/3' junction PCR applying extracted genomic DNA from bulk sorted cells was performed as

20   previously described. From all three targeting events, 3' and 5' junction PCR could be observed six days after the FACS sort and already at day 1 after FACS sort for the EPO construct (Figure 21).

These data indicate targeted integration events in the FACS sorted cells, and show the

25   advantages of the developed method for accelerated generation of rCHO integrants using CRISPR Cas9, homology-directed repair and transiently expressed fluorescent markers.

Donor DNA plasmid with EPO (SEQ ID NO: 107):

LabID: 2826: COSMC-coEPO-HDR-TI donor ver.2

AGTCGGTGTGTAATCCATGGAGGAGTTTCTATAATGTTGCAGTTTCTACCTAATGG

5   TGACCAAATGCCAGTGAAAGGATTGTAAGAGTACTTGTCACATATACTACTCACCT

CATTTCAAGAATGTGGACCTGCTTTTAAACATTAAGAGCAAATCGTAATTATATAAG

AAATAAGCAAATGAAACTATTAGACTGTTTGAAAAGTCTTTTTCTTTACAGGAAAAA

TGCTTTCAGAAAGCAGTTCATTTTTGAAAGGTGTGATGCTTGGAAGCATCTTCTAT

GCCTTGATCACTACGCTAGGCCACATTAGGATTGGGCACAGAAACAGGACACACC

10   ACCATGAGCATCACCACCTGCAAGCTCCTAACAAAGAAGATATCTCGAAAATCTCA

GCGGCTGAGCGCATGGAGCTCAGTAAGAGCTTCCGGGTATACTGTATAGTTCTTG

TAAAACCCAAAGATGTGAGTCTTTGGGCTGCAGTGAAGGAGACTTGGACCAAACA

CTGTGACAAAGCAGAGTTCTTCAGTTCTGAAAATGTTAAAGTGTTTGAGTCAATTA

ACGTGGACACTGATGACATGTGGTTGATGATGAGGAAAGCTTATAAATATGCCTTT

15   GATAAATACAAAGAGCAGTACAACTGGTTCTTCCTTGCACGCCCCAGTACTTTGC

TGTGATTGAAAATCTAAAATATTTTTTGTTAAAAAAGGATCCATCGCAGCCTTTCTA

TCTAGGACACACTGTAAAATCTGGAGACCTTAAGCAGCGTGTGAGGCTCCGGTGC

CCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAG

GGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAA

20   GTGATGTCGTGTACTGGCTCCGCCTTTTCCCGAGGGTGGGGGAGAACCGTATAT

AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACA

CAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC

CCTTGCGTGCCTTGAATTACTTCCACCTGGCTCCAGTACGTGATTCTTGATCCCGA

GCTGGAGCCAGGGGCGGGCCTTGCGCTTTAGGAGCCCCTTCGCCTCGTGCTTGA

25   GTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCT

TCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGAC

CTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAGGATCT

GCACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGT

CCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATC

30   GGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCC

GCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTG

CGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTCCAGGGGGCTCAAAATGGA

GGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAGGG

GCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCG

35   TCCAGGCACCTCGATTAGTTCTGGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGG

102

GGGAGGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAG

TTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTT

GGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTT

CAGGTGTCGTGAAGACGTCATCGCCACCATGGGAGTGCACGAGTGTCCTGCTTG

GCTGTGGCTGCTGCTGTCCCTGCTGTCTCTGCCTCTGGGACTGCCTGTGCTGGG

CGCTCCTCCTAGACTGATCTGCGACTCCCGGGTGCTGGAAAGATACCTGCTGGAA

GCCAAAGAGGCCGAGAACATCACCACCGGCTGCGCCGAGCACTGCTCCCTGAAC

GAGAATATCACCGTGCCCGACACCAAAGTGAACTTCTACGCCTGGAAGCGGATG

GAAGTGGGCCAGCAGGCTGTGGAAGTGTGGCAGGGACTGGCTCTGCTGAGCGA

GGCTGTGCTGAGAGGACAGGCCCTGCTCGTGAACTCCTCCCAGCCTTGGGAACC

CCTGCAGCTGCACGTGGACAAGGCTGTGTCCGGCCTGAGATCCCTGACCACCCT

GCTGAGAGCACTGGGAGCCCAGAAAGAGGCCATCTCTCCACCTGACGCCGCCTC

TGCTGCTCCTCTGAGAACCATCACCGCCGACACCTTCAGAAAGCTGTTCCGGGTG

TACTCCAACTTCCTGCGGGGCAAGCTGAAGCTGTACACCGGCGAGGCTTGCCGG

ACCGGCGATAGATGAGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAG

TCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCC

AGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCAC

TCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGT

GTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG

GAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGAGGTCTGAGTG

ATTGTCTTAAGCATAGAGTCAATGAAAAGACTCAACAGCCTTCTCAGTGTTCCGGA

AAAGTGTCCTGAACAAGGTGGGATGATTTGGAAGATATCTGAAGATAAGCAGCTA

GCAGTCTGCCTGAAATATGCTGGAGTATTTGCGGAAAATGCGGAAGACGCTGATA

GAAAAGATGTATTTAATACCAAATCTGTTGGGCTTTTCATTAAAGAGGCCATGTCTA

ACCACCCGAACCAGGTAGTAGAAGGATGCTGTTCCAATATGGCTGTCACTTTTAAT

GGACTAACTCCTAATCAGATGCATGTGATGATGTATGGGGTGTACCGGCTTAGGG

CCTTTGGACATGTTTTCAACGATGCGTTGGTTTTCTTACCTCCAAACGGTTCTGAT

AATGACTGACAAAAGCAAGAGCATGCATTTGGTAACCACATTAAGACATGTTATG

CTTTCTAATCGATAATGCATCTAACACAGTAGTGTGTTTCTTTTCCTTATCTGGTCA

CATTGAAGTCTACTTGTACATTTTCAAATGGAATGGTATTTTTTTCCCTTAAATCATT

TGTGAGAAATTTTAATGTGTTAGAAATAAATGTTTAAGAATAGCAATTTTGCAAAT

AATGTATTTATAAATATTATATTTATGTGATAAAGACCAAATTATAGACATTAAAATC

TGTGATGTATCTTTGCCTATTGATTTAAATGTTTAATGTATCTTTTTAGATTTCAAA

TATATGCAAATGAGGACGTCGCTGTTGACATTGATTATTGACTAGTTATTAATAGTA

ATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAAC

TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGT
CAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT
GCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTAT
GCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGT
CATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAG
CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTT
TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCA
TTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT
CTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATAGTGCGATcgcca
ccATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCG
CTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGG
CGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGA
CCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTA
CGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCT
GTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGG
CGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAA
GGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAA
GACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCC
TGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGAC
GCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCC
TACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCG
TGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAG
CTGTACAAGTAGACACAGTCTCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGC
CCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGG
GGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCA
TGCTGGGGATGCGGTGGGCTCTATGGACTTGCGTAGTGAGTCGAATAAGGGCGA
CACAAAATTTATTCTAAATGCATAATAAATACTGATAACATCTTATAGTTTGTATTAT
ATTTTGTATTATCGTTGACATGTATAATTTTGATATCAAAAACTGATTTTCCCTTTAT
TATTTTCGAGATTTATTTTCTTAATTCTCTTTAACAAACTAGAAATATTGTATATACA
AAAAATCATAAATAATAGATGAATAGTTTAATTATAGGTGTTCATCAATCGAAAAAG
CAACGTATCTTATTTAAAGTGCGTTGCTTTTTTCTCATTTATAAGGTTAAATAATTCT
CATATATCAAGCAAAGTGACAGGCGCCCTTAAATATTCTGACAAATGCTCTTTCCC
TAAACTCCCCCCATAAAAAACCCGCCGAAGCGGGTTTTTACGTTATTTGCGGATT

104

AACGATTACTCGTTATCAGAACCGCCCAGGGGGCCCGAGCTTAAGACTGGCCGT
CGTTTTACAACACAGAAAGAGTTTGTAGAAACGCAAAAAGGCCATCCGTCAGGGG
CCTTCTGCTTAGTTTGATGCCTGGCAGTTCCCTACTCTCGCCTTCCGCTTCCTCGC
TCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACT
5   CAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACAT
GTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGC
GTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA
AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCG
10   CCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCT
CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTA
AGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCG
AGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGGCTAACTACGGCTACA
15   CTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAA
AGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTG
ATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGACGCGCGCGTAACTCACGTTA
AGGGATTTTGGTCATGAGCTTGCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCT
20   TTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCAT
CCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACC
ATCTGGCCCCAGCGCTGCGATGATACCGCGAGAACCACGCTCACCGGCTCCGGA
TTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGC
AACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTA
25   GTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATCGCTACAGGCATCGTGGT
GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGG
CGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTC
CGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTG
30   AGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTG
CCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTC
ATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGA
GATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGG
35   GAATAAGGGCGACACGGAAATGTTGAATACTCATATTCTTCCTTTTTCAATATTATT

GAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGA
AAAATAAACAAATAGGGGTCAGTGTTACAACCAATTAACCAATTCTGAACATTATC
GCGAGCCCATTTATACCTGAATATGGCTCATAACACCCCTTGTTTGCCTGGCGGC
AGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGT

5   AGCGCCGATGGTAGTGTGGGGACTCCCCATGCGAGAGTAGGGAACTGCCAGGCA
TCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGCCCGGGCTAATTA
TGGGGTGTCGCCCTTATTCGACTC

Donor DNA plasmid with Enbrel (SEQ ID NO: 108):
10 LabID: 2827: COSMC-coEnbrel-HDR-TI donor ver.2
AGTCGGTGTGTAATCCATGGAGGAGTTTCTATAATGTTGCAGTTTCTACCTAATGG
TGACCAAATGCCAGTGAAAGGATTGTAAGAGTACTTGTCACATATACTACTCACCT
CATTTCAAGAATGTGGACCTGCTTTTAAACATTAAGAGCAAATCGTAATTATATAAG
AAATAAGCAAATGAAACTATTAGACTGTTTGAAAAGTCTTTTTCTTTACAGGAAAAA

15 TGCTTTCAGAAGCAGTTCATTTTTGAAAGGTGTGATGCTTGGAAGCATCTTCTAT
GCCTTGATCACTACGCTAGGCCACATTAGGATTGGGCACAGAAACAGGACACACC
ACCATGAGCATCACCACCTGCAAGCTCCTAACAAAGAAGATATCTCGAAAATCTCA
GCGGCTGAGCGCATGGAGCTCAGTAAGAGCTTCCGGGTATACTGTATAGTTCTTG
TAAAACCCAAAGATGTGAGTCTTTGGGCTGCAGTGAAGGAGACTTGGACCAAACA

20 CTGTGACAAAGCAGAGTTCTTCAGTTCTGAAAATGTTAAAGTGTTTGAGTCAATTA
ACGTGGACACTGATGACATGTGGTTGATGATGAGGAAAGCTTATAAATATGCCTTT
GATAAATACAAAGAGCAGTACAACTGGTTCTTCCTTGCACGCCCCAGTACTTTTGC
TGTGATTGAAAATCTAAAATATTTTTTGTTAAAAAAGGATCCATCGCAGCCTTTCTA
TCTAGGACACACTGTAAAATCTGGAGACCTTAAGCAGCGTGTGAGGCTCCGGTGC

25 CCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAG
GGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAA
GTGATGTCGTGTACTGGCTCCGCCTTTTCCCGAGGGTGGGGGAGAACCGTATAT
AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACA
CAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC

30 CCTTGCGTGCCTTGAATTACTTCCACCTGGCTCCAGTACGTGATTCTTGATCCCGA
GCTGGAGCCAGGGGCGGGCCTTGCGCTTTAGGAGCCCCTTCGCCTCGTGCTTGA
GTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCT
TCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGAC
CTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAGGATCT

35 GCACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGT

CCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATC
GGACGGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCC
GCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTG
CGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTCCAGGGGGCTCAAAATGGA
GGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAGGG
GCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCG
TCCAGGCACCTCGATTAGTTCTGGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGG
GGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAG
TTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTT
GGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTT
CAGGTGTCGTGAAGACGTCATAGCACCATGGCGCCCGTCGCCGTCTGGGCCGCG
CTGGCCGTCGGACTGGAGCTCTGGGCTGCGGCGCACGCCTTGCCCGCCCAGGT
GGCATTTACACCCTACGCCCCGGAGCCCGGGAGCACATGCCGGCTCAGAGAATA
CTATGACCAGACAGCTCAGATGTGCTGCAGCAAATGCTCGCCGGGCCAACATGC
AAAAGTCTTCTGTACCAAGACCTCGGACACCGTGTGTGACTCCTGTGAGGACAGC
ACATACACCCAGCTCTGGAACTGGGTTCCCGAGTGCTTGAGCTGTGGCTCCCGCT
GTAGCTCTGACCAGGTGGAAACTCAAGCCTGCACTCGGGAACAGAACCGCATCT
GCACCTGCAGGCCCGGCTGGTACTGCGCGCTGAGCAAGCAGGAGGGGTGCCGG
CTGTGCGCGCCGCTGCGCAAGTGCCGCCCGGGCTTCGGCGTGGCCAGACCAGG
AACTGAAACATCAGACGTGGTGTGCAAGCCCTGTGCCCCGGGGACGTTCTCCAA
CACGACTTCATCCACGGATATTTGCAGGCCCCACCAGATCTGTAACGTGGTGGCC
ATCCCTGGGAATGCAAGCATGGATGCAGTCTGCACGTCCACGTCCCCCACCCGG
AGTATGGCCCCAGGGGCAGTACACTTACCCCAGCCAGTGTCCACACGATCCCAA
CACACGCAGCCAACTCCAGAACCCAGCACTGCTCCAAGCACCTCCTTCCTGCTCC
CAATGGGCCCCAGCCCCCCAGCTGAAGGGAGCACTGGCGACGAGCCCAAATCTT
GTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGAC
CGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGGAC
CCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCAA
GTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCG
GGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCA
CCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCT
CCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACC
ACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACCAGGTCAG
CCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGGACTC

CGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCA
GCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCACTAC
ACACAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGATAAGAATTCTGCAGATATC
CAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAG

5   CCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAA
TTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCA
GGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGG
TGGGCTCTATGGAGGTCTGAGTGATTGTCTTAAGCATAGAGTCAATGAAAAGACT

10   CAACAGCCTTCTCAGTGTTCCGGAAAAGTGTCCTGAACAAGGTGGGATGATTTGG
AAGATATCTGAAGATAAGCAGCTAGCAGTCTGCCTGAAATATGCTGGAGTATTTGC
GGAAAATGCGGAAGACGCTGATAGAAAGATGTATTTAATACCAAATCTGTTGGG
CTTTTCATTAAAGAGGCCATGTCTAACCACCCGAACCAGGTAGTAGAAGGATGCT
GTTCCAATATGGCTGTCACTTTTAATGGACTAACTCCTAATCAGATGCATGTGATG

15   ATGTATGGGGTGTACCGGCTTAGGGCCTTTGGACATGTTTTCAACGATGCGTTGG
TTTTCTTACCTCCAAACGGTTCTGATAATGACTGACAAAAGCAAGAGCATGCATT
TGGTAACCACATTAAGACATGTTATGCTTTCTAATCGATAATGCATCTAACACAGTA
GTGTGTTTCTTTTCCTTATCTGGTCACATTGAAGTCTACTTGTACATTTTCAAATGG
AATGGTATTTTTTTCCCTTAAATCATTTGTGAGAAATTTTAATGTGTTAGAAATAAAT

20   GTTTTAAGAATAGCAATTTTGCAAATAATGTATTTATAAATATTATATTTATGTGATA
AAGACCAAATTATAGACATTAAAATCTGTGATGTATCTTTGCCTATTGATTTTAAAT
GTTTAATGTATCTTTTTAGATTTCAAATATATGCAAATGAGGACGTCGCTGTTGACA
TTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCC
ATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG

25   CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC
CAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCA
CTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATG
ACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCC
TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT

30   GGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT
CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTT
CCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTAC
GGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTA
CTGGCTTATCGAAATAGTGCGATcgccaccATGGTGAGCAAGGGCGAGGAGGATAA

35   CATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGT

GAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGG
GCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCT
GGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCC
CGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGA
5   GCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTC
CCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCC
CTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGA
GCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGA
AGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCA
10   AGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCA
CCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCC
GCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGACACAGTCTCTGTGCC
TTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTG
GAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTG
15   TCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGG
GGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGA
CTTGCGTAGTGAGTCGAATAAGGGCGACACAAAATTTATTCTAAATGCATAATAAA
TACTGATAACATCTTATAGTTTGTATTATATTTTGTATTATCGTTGACATGTATAATT
TTGATATCAAAAACTGATTTTCCCTTTATTATTTTCGAGATTTATTTTCTTAATTCTCT
20   TTAACAAACTAGAAATATTGTATATACAAAAAATCATAAATAATAGATGAATAGTTTA
ATTATAGGTGTTCATCAATCGAAAAAGCAACGTATCTTATTTAAAGTGCGTTGCTTT
TTTCTCATTTATAAGGTTAAATAATTCTCATATATCAAGCAAAGTGACAGGCGCCCT
TAAATATTCTGACAAATGCTCTTTCCCTAAACTCCCCCCATAAAAAAACCCGCCGA
AGCGGGTTTTTACGTTATTTGCGGATTAACGATTACTCGTTATCAGAACCGCCCAG
25   GGGGCCCGAGCTTAAGACTGGCCGTCGTTTTACAACACAGAAAGAGTTTGTAGAA
ACGCAAAAAGGCCATCCGTCAGGGGCCTTCTGCTTAGTTTGATGCCTGGCAGTTC
CCTACTCTCGCCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGG
CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAAT
CAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGA
30   ACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGA
GCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA
AGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCC
TGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTC
TCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG
35   GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAAC

TATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA
CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAA
GTGGTGGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTG
CTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA
ACGACGCGCGCGTAACTCACGTTAAGGGATTTTGGTCATGAGCTTGCGCCGTCCC
GTCAAGTCAGCGTAATGCTCTGCTTTTACCAATGCTTAATCAGTGAGGCACCTATC
TCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGCGCTGCGATGATACCGCG
AGAACCACGCTCACCGGCTCCGGATTTATCAGCAATAAACCAGCCAGCCGGAAG
GGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAAT
TGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTG
TTGCCATCGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATT
CAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAA
AAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCC
GTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG
TATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCC
ACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAAC
TCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACC
CAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAG
GAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACT
CATATTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAG
CGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTCAGTGTTACAA
CCAATTAACCAATTCTGAACATTATCGCGAGCCCATTTATACCTGAATATGGCTCA
TAACACCCCTTGTTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCAT
GCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGACTCCCCA
TGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGA
CTGGGCCTTTCGCCCGGGCTAATTATGGGGTGTCGCCCTTATTCGACTC

Donor DNA plasmid with Rituximab (SEQ ID NO: 109):
LabID: 2828: COSMC-Rituximab-HDR-TI donor ver.2
AGTCGGTGTGTAATCCATGGAGGAGTTTCTATAATGTTGCAGTTTCTACCTAATGG
TGACCAAATGCCAGTGAAAGGATTGTAAGAGTACTTGTCACATATACTACTCACCT

CATTTCAAGAATGTGGACCTGCTTTTAAACATTAAGAGCAAATCGTAATTATATAAG

AAATAAGCAAATGAAACTATTAGACTGTTTGAAAAGTCTTTTTCTTTACAGGAAAAA

TGCTTTCAGAAAGCAGTTCATTTTTGAAAGGTGTGATGCTTGGAAGCATCTTCTAT

GCCTTGATCACTACGCTAGGCCACATTAGGATTGGGCACAGAAACAGGACACACC

5   ACCATGAGCATCACCACCTGCAAGCTCCTAACAAAGAAGATATCTCGAAAATCTCA

GCGGCTGAGCGCATGGAGCTCAGTAAGAGCTTCCGGGTATACTGTATAGTTCTTG

TAAAACCCAAAGATGTGAGTCTTTGGGCTGCAGTGAAGGAGACTTGGACCAAACA

CTGTGACAAAGCAGAGTTCTTCAGTTCTGAAAATGTTAAAGTGTTTGAGTCAATTA

ACGTGGACACTGATGACATGTGGTTGATGATGAGGAAAGCTTATAAATATGCCTTT

10   GATAAATACAAAGAGCAGTACAACTGGTTCTTCCTTGCACGCCCCAGTACTTTGC

TGTGATTGAAAATCTAAAATATTTTTTGTTAAAAAAGGATCCATCGCAGCCTTTCTA

TCTAGGACACACTGTAAAATCTGGAGACCTTAAGCAGCGTGTGAGGCTCCGGTGC

CCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAG

GGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAA

15   GTGATGTCGTGTACTGGCTCCGCCTTTTCCCGAGGGTGGGGGAGAACCGTATAT

AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACA

CAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC

CCTTGCGTGCCTTGAATTACTTCCACCTGGCTCCAGTACGTGATTCTTGATCCCGA

GCTGGAGCCAGGGGCGGGCCTTGCGCTTAGGAGCCCCTTCGCCTCGTGCTTGA

20   GTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCT

TCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGAC

CTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAGGATCT

GCACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGT

CCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATC

25   GGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCC

GCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTG

CGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTCCAGGGGGCTCAAAATGGA

GGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAGGG

GCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCG

30   TCCAGGCACCTCGATTAGTTCTGGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGG

GGGAGGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAG

TTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTT

GGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTT

CAGGTGTCGTGAAGACGTCATGACACCATGGGCTGGTCCTGCATCATCCTGTTTC

35   TGGTGGCTACCGCCACCGGCGTGCACTCTCAGGTGCAGCTGCAGCAGCCTGGC

GCCGAGCTCGTGAAACCTGGCGCCTCCGTGAAGATGTCCTGCAAGGCCTCCGGC
TACACCTTCACCAGCTACAACATGCACTGGGTCAAGCAGACCCCCGGCAGAGGC
CTGGAATGGATCGGCGCTATCTACCCCGGCAACGGCGACACCTCCTACAACCAG
AAGTTCAAGGGCAAGGCCACCCTGACCGCCGACAAGTCCTCTTCCACCGCCTAC
ATGCAGCTGTCCTCCCTGACCTCCGAGGACTCCGCCGTGTACTACTGCGCCCGG
TCTACCTACTACGGCGGCGACTGGTACTTCAACGTGTGGGGCGCTGGCACCACC
GTGACCGTGTCTGCTGCTTCTACCAAGGGCCCCTCCGTGTTCCCTCTGGCCCCTT
CCAGCAAGTCTACCTCTGGCGGCACAGCCGCTCTGGGCTGCCTCGTGAAGGACT
ACTTCCCCGAGCCCGTGACAGTGTCCTGGAACTCTGGCGCTCTGACCAGCGGAG
TGCACACCTTCCCTGCTGTGCTGCAGTCCTCCGGCCTGTACTCCCTGTCCAGCGT
CGTGACTGTGCCCTCCAGCTCTCTGGGCACCCAGACCTACATCTGCAACGTGAAC
CACAAGCCCTCCAACACCAAGGTGGACAAGAAGGTGGAACCCAAGTCCTGCGAC
AAGACCCACACCTGTCCCCCTTGTCCTGCCCCTGAACTGCTGGGCGGACCCAGC
GTGTTCCTGTTCCCCCCAAAGCCCAAGGATACCCTGATGATCTCCCGGACCCCCG
AAGTGACCTGCGTGGTGGTGGATGTGTCCCACGAGGACCCTGAAGTGAAGTTCA
ATTGGTACGTGGACGGCGTGGAAGTGCACAACGCCAAGACCAAGCCTAGAGAGG
AACAGTACAACTCCACCTACCGGGTGGTGTCCGTGCTGACCGTGCTGCACCAGG
ATTGGCTGAACGGCAAAGAGTACAAGTGCAAGGTGTCCAACAAGGCCCTGCCTG
CCCCCATCGAAAAGACCATCTCCAAGGCCAAGGGCCAGCCCCGGGAACCCCAGG
TGTACACACTGCCCCCTAGCAGGGACGAGCTGACCAAGAACCAGGTGTCCCTGA
CATGCCTCGTGAAAGGCTTCTACCCCTCCGATATCGCCGTGGAATGGGAGTCCAA
CGGCCAGCCTGAGAACAACTACAAGACCACCCCCCCTGTGCTGGACTCCGACGG
CTCATTCTTCCTGTACAGCAAGCTGACAGTGGACAAGTCCCGGTGGCAGCAGGG
CAACGTGTTCTCCTGCTCCGTGATGCACGAGGCCCTGCACAACCACTATACCCAG
AAGTCCCTGTCCCTGAGCCCCGGCAAGTGAGGATCCGACACAGTCTCTGTGCCTT
CTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGG
GAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGAG
TGCGATGTGAGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGT
CCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGT
GGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCG
AGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCG
CAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCC
TGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCACCTGGCTC

112

CAGTACGTGATTCTTGATCCCGAGCTGGAGCCAGGGGCGGGCCTTGCGCTTTAG

GAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCCTGGGCGCTGGGGCCGC

CGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTCTC

TAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGTC

5   TTGTAAATGCGGGCCAGGATCTGCACACTGGTATTTCGGTTTTTGGGGCCGCGGG

CGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGC

GAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCT

CTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCT

GGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTG

10   CTCCAGGGGGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGAG

TCACCCACACAAAGGAAAGGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACT

CCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTGGAGCTTTTGGA

GTACGTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACAC

TGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTG

15   GAATTTGCCCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGT

TCAAAGTTTTTTTCTTCCATTTCAGGTGTCGTGAACACGTGGTCGCGGCCGCACCA

TGGGCTGGTCCTGCATCATCCTGTTTCTGGTGGCTACCGCCACCGGCGTGCACT

CCCAGATTGTGCTGTCTCAGTCCCCCGCCATCCTGTCTGCTAGCCCTGGCGAGAA

AGTGACAATGACCTGCCGGGCCTCCTCCTCCGTGTCCTACATCCACTGGTTCCAG

20   CAGAAGCCCGGCTCCAGCCCCAAGCCTTGGATCTACGCCACCTCCAACCTGGCC

TCTGGCGTGCCAGTGCGGTTTTCCGGCTCTGGCTCTGGCACCTCCTACTCCCTGA

CCATCTCTCGGGTGGAAGCCGAGGATGCCGCCACCTACTACTGCCAGCAGTGGA

CCAGCAACCCCCCCACATTTGGCGGAGGCACCAAGCTGGAAATCAAGCGGACCG

TGGCCGCTCCCTCCGTGTTCATCTTCCCACCTTCCGACGAGCAGCTGAAGTCCGG

25   CACCGCTTCTGTCGTGTGCCTGCTGAACAACTTCTACCCCCGCGAGGCCAAGGT

GCAGTGGAAGGTGGACAACGCCCTGCAGTCCGGCAACTCCCAGGAATCCGTGAC

CGAGCAGGACTCCAAGGACAGCACCTACAGCCTGTCCTCCACCCTGACCCTGTC

CAAGGCCGACTACGAGAAGCACAAGGTGTACGCCTGCGAAGTGACCCACCAGGG

CCTGTCTAGCCCCGTGACCAAGTCTTTCAACCGGGGCGAGTGCTGACTCGAGAG

30   ATCTGGCCGGCTGGGCCCGTTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGG

TCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCT

GATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCC

CGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATG

AGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGT

35   GGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGG

113

ATGCGGTGGGCTCTATGGAGGTCTGAGTGATTGTCTTAAGCATAGAGTCAATGAA

AAGACTCAACAGCCTTCTCAGTGTTCCGGAAAAGTGTCCTGAACAAGGTGGGATG

ATTTGGAAGATATCTGAAGATAAGCAGCTAGCAGTCTGCCTGAAATATGCTGGAGT

ATTTGCGGAAAATGCGGAAGACGCTGATAGAAAGATGTATTTAATACCAAATCTG

5   TTGGGCTTTTCATTAAAGAGGCCATGTCTAACCACCCGAACCAGGTAGTAGAAGG

ATGCTGTTCCAATATGGCTGTCACTTTTAATGGACTAACTCCTAATCAGATGCATG

TGATGATGTATGGGGTGTACCGGCTTAGGGCCTTTGGACATGTTTTCAACGATGC

GTTGGTTTTCTTACCTCCAAACGGTTCTGATAATGACTGACAAAAAGCAAGAGCAT

GCATTTGGTAACCACATTAAGACATGTTATGCTTTCTAATCGATAATGCATCTAACA

10   CAGTAGTGTGTTTCTTTTCCTTATCTGGTCACATTGAAGTCTACTTGTACATTTTCA

AATGGAATGGTATTTTTTTCCCTTAAATCATTTGTGAGAAATTTTAATGTGTTAGAAA

TAAATGTTTTAAGAATAGCAATTTTGCAAATAATGTATTTATAAATATTATATTTATGT

GATAAAGACCAAATTATAGACATTAAAATCTGTGATGTATCTTTGCCTATTGATTTT

AAATGTTTAATGTATCTTTTTAGATTTCAAATATATGCAAATGAGGACGTCGCTGTT

15   GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATA

GCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTG

ACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTA

ACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG

CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGT

20   CAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGAC

TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG

GTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA

AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGG

ACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCG

25   TGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACT

GCTTACTGGCTTATCGAAATAGTGCGATcgccaccATGGTGAGCAAGGGCGAGGAG

GATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCT

CCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTAC

GAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTT

30   CGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAG

CACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGT

GGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGAC

TCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAAC

TTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCC

35   TCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAG

GCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAA

GGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGA

CATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGA

GGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAGACACAGTCTCT

5   GTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGA

CCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATC

GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAG

CAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT

CTATGGACTTGCGTAGTGAGTCGAATAAGGGCGACACAAAATTTATTCTAAATGCA

10   TAATAAATACTGATAACATCTTATAGTTTGTATTATATTTTGTATTATCGTTGACATG

TATAATTTTGATATCAAAAACTGATTTTCCCTTTATTATTTTCGAGATTTATTTTCTTA

ATTCTCTTTAACAAACTAGAAATATTGTATATACAAAAAATCATAAATAATAGATGAA

TAGTTTAATTATAGGTGTTCATCAATCGAAAAAGCAACGTATCTTATTTAAAGTGCG

TTGCTTTTTTCTCATTTATAAGGTTAAATAATTCTCATATATCAAGCAAAGTGACAG

15   GCGCCCTTAAATATTCTGACAAATGCTCTTTCCCTAAACTCCCCCCATAAAAAAAC

CCGCCGAAGCGGGTTTTTACGTTATTTGCGGATTAACGATTACTCGTTATCAGAAC

CGCCCAGGGGGCCCGAGCTTAAGACTGGCCGTCGTTTTACAACACAGAAAGAGT

TTGTAGAAACGCAAAAAGGCCATCCGTCAGGGGCCTTCTGCTTAGTTTGATGCCT

GGCAGTTCCCTACTCTCGCCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGG

20   TCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATC

CACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAA

GGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC

CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGAC

AGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT

25   GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG

TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCG

CTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT

ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTG

GCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA

30   GAGTTCTTGAAGTGGTGGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTA

TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATC

CGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGCAAGCAGCAGATT

ACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA

CGCTCAGTGGAACGACGCGCGCGTAACTCACGTTAAGGGATTTTGGTCATGAGCT

35   TGCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCTTTTACCAATGCTTAATCAGTG

AGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCC
GTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGCGCTGCG
ATGATACCGCGAGAACCACGCTCACCGGCTCCGGATTTATCAGCAATAAACCAGC
CAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC

5   AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATCGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTA
TGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCAT
GTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAG
TTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGT

10   CATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATA
ATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCG
GGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
CTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGA

15   GCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAA
TGTTGAATACTCATATTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT
GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTC
AGTGTTACAACCAATTAACCAATTCTGAACATTATCGCGAGCCCATTTATACCTGA
ATATGGCTCATAACACCCCTTGTTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACC

20   TGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGG
GACTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCA
GTCGAAAGACTGGGCCTTTCGCCCGGGCTAATTATGGGGTGTCGCCCTTATTCGA
CTC

25   **Example 11. Permanently integrated CRISPR Cas9 mediated multiplexing in CHO cells**

This example demonstrates the genome editing capability of the constructed homogenous polyclonal cell line, CHO-S GFP_2A_Cas9 with permanently integrated Cas9 nuclease described in example 12. We tested this cell line with multiple gRNAs to

30   demonstrate its genome editing efficiency.

Cell cultivation

CHO-S GFP_2A_Cas9 polyclonal cell line was cultured in CD CHO medium (Life Technologies) supplemented with 8 mM L-Glutamine, 1:500 anticlamping agent and

1:100 Pen/Strep and cultivated in Corning Erlenmeyer shake flasks (Sigma-Aldrich). The cells were incubated at 37°C, 5 % CO_2 with 120 rpm shaking and passaged every 2-3 days.

5   *Multiplex gRNA transfection*

CHO-S GFP_2A_Cas9 (also termed 2AGFP_Cas9) polyclonal cell line was induced with doxycycline (1 µg/µL) and was after 24h transfected with three sgRNA targeting the FUT8, BAK-1 and BAX loci using Nucleofection. 2 million cells were transfected with 1 µg/µL of each gRNA. The cells were then placed in a CO_2 incubator at 37°C in

10   fresh complete CD CHO medium (8 mM L-glutamine, anti-clumping agent 1:500 and 1:100 pen/strep). 16 hours post transfection, the samples were incubated at 30°C for 32 hours before being transferred back to 37°C. At day 3 cells were FACS bulk sorted to harvest GFP positive cells, GFP negative cells and GFP intermediate expressing cells. 500.000 cells were harvested by centrifugation and the pellet was stored at -

15   20°C. Genomic DNA was prepared as described in example 6. The percentage of indels generated in GFP_2A_Cas9 cell line were analyzed for each of the three FUT8, BAK-1 and BAX sgRNAs by targeted resequencing of the target loci (Figure 22) with the primers listed in table 10.

20   The CRISPy bioinformatic tool (http://staff.biosustain.dtu.dk/laeb/crispy/) was applied for generating three sgRNAs targeting FUT8, BAK-1 and BAX respectively. The sgRNA target sequences are described in table 9. The sgRNA expression vectors have been constructed as described in example 6. The sgRNA targeting FUT8 (FUT8_681494/sgRNA2_C) is described in example 6. The sgRNAs targeting BAK-1

25   and BAX are constructed as in example 9.

Results of Multiplexed guideRNA with permanently integrated Cas9.

MiSeq data analysis revealed that the editing efficiency (indels rate) in the 2AGFP_Cas9 cell line (sorted for GFP positive) for loci BAK-1 and BAX was

30   comparable to the one achieved with Cas9 vector_base transient expression (figure 15). Only the percentage of loci with indels in the Fut8 loci appeared to be lower (figure 22). This demonstrates that a cell line with permanently integrated Cas9 is capable of

creating indels comparable to transiently transfected cells and that multiple loci can be modified simultaneously.

GFP FACS enrichment increases the percentage of indels created in sorted cells

5   In order to enrich the pool of population with indels and to investigate whether the expression level of Cas9 in the cell line could affect the editing efficiency, the polyclonal population 2AGFP_Cas9 was FACS-sorted after transfection based on the GFP level which directly correlates with Cas9 expression (Figure 23). As shown in figure 24, the amount of indels created increased when using FACS sorted cells from the high GFP

10   category.

This demonstrates that FACS sorting can be used as a method for increasing the amount of indels created and that the amount of Cas9 protein directly affects the creation of indels.

15

Example 12. Construction of a cell line with integrated inducible Cas9

We generated a cell line with an inducible 2AGFP_Cas9 permanently integrated in the genome. This allows fast and cheap genome modifications (insertions, deletions, etc) just by interchanging different sgRNAs that are first designed with the CRISPy

20   bioinformatic tool (http://staff.biosustain.dtu.dk/laeb/crispy/) and then assembled with a bio-brick approach using USER cloning. Moreover Cas9 activity can be monitored with GFP using 2A peptide for co-transcription.

Vectors construction:

25   The CHO codon optimized Cas9 expression vectors for integration have been designed for both targeted integration and random integration for different cell lines (CHO-K1 and CHO-S). The vectors are shown in figures 25, 26 and 27 (SEQ ID NO: 110, 111, 112) and were constructed from commercial vectors pcDNA™ 4/TO, pJTI™ R4 DEST CMV TO pA and pcDNA™6/TR (Invitrogen).

30

Cell line construction:

CHO-S cells from Life Technologies were grown in CD CHO medium (Life
Technologies) supplemented with 8 mM L-Glutamine and cultivated in Corning
erlenmeyer shake flasks (Sigma-Aldrich). The cells were incubated at 37°C, 5 % CO2
with 120 rpm shaking and passaged every 2-3 days. Jump in CHO-K1 cells from Life
5   Technologies were grown in D-MEM with GlutaMAXTM Supplemented with FBS, MEM
Non-Essential Amino Acids Solution, HEPES buffer (pH7.3), Pen/strep and
Hygromacyn B (before targeting the vector) according to the manufacturing using 175
and 75 cm Corning erlenmeyer tissue flasks .

10   CHO-S cells from Life Technologies were transfected with the vectors encoding
2AGFP_Cas9 (CHO-codon optimized) ,pcDNA4/TO::2AGFP_Cas9 and pcDNA™6/TR
using Nucleofection (Nucleofector™ 2b, Lonza). $2*10^6$ cells were transfected with 2
µg/µL of each plasmid. The cells were then placed in a CO2 incubator at 37°C (without
shaking) in fresh complete CD CHO medium (8 mM L-glutamine, anti-clumping agent
15   1:500 and 1:100 pen/strep). 16 hours post transfection, the samples were incubated at
30°C for 32 hours before transferred back to 37°C. Stable pools were generated by
seeding cells ($9*10^5$ cells) in CELLSTAR 6 well Advanced TC plates (Greiner, Sigma-
Aldrich) on day 3 followed by selecting for Zeocin (final concentration: 400 µg/ml) and
Blasticidin (final concentration: 8 µg/ml) resistance clones. During selection, medium
20   was changed every 3-4 days. After 2 weeks of selection, cells were detached with
TrypLE (Life Technologies) according to manufacturer's recommendations and
transferred to Corning Erlenmeyer shake flasks depending on cell concentrations.
Finally, 5 vials of $10*10^6$ cells were frozen as cell bank stock.

25   The stable pool of cells divided in 2 sub-pools. One pool was induced with doxycycline
(1 µg/µL) and then FACS bulk sorted to harvest GFP positive cells. The other pool was
directly FACS bulk sorted to harvest GFP negative cells. Then, the first induced GFP
positive were cultivated without induction for 2 days and then they were FACS bulk
sorted to harvest GFP negative cells and finally after another round of induction the
30   cells with the best inducibility (high GFP expression only upon induction) were selected
as homogenous polyclonal population. The cell pool not induced and harvested as
GFP negative, were subsequently induced and bulk sorted by FACS to harvest GFP
positive cells. Finally after a second round of induction the tightest preforming cells
were selected as homogenous polyclonal cell population.

35

Jump in CHO_K1 from life technologies were detached with TrypLE (Life

Technologies) according to manufacturer's recommendations and prepared for

transfection. Cells were transfected with the vectors encoding 2AGFP_Cas9 (CHO-

codon optimized), pJTI™ R4::2AGFP_Cas9, and pcDNA™6/TR _ZEO using

5   Nucleofection. The machine (Nucleofector™ 2b, Lonza) was set up to program U-023

and 2*106 were transfected with 2 μg/μL of each plasmid. The cells were then place in

a CO2 incubator at 37°C in fresh complete D-MEM with GlutaMAXTM Supplemented

with FBS, MEM Non-Essential Amino Acids Solution, HEPES buffer (pH7.3). 16 hours

post transfection, the samples were incubated at 30°C for 32 hours before transferred

10   back to 37°C. Stable pools were generated by seeding cells (9*10^5 cells) in

CELLSTAR 6 well Advanced TC plates (Greiner, Sigma-Aldrich) on day 3 followed by

selecting for Zeocin (final concentration: 400 μg/ml) and Blasticidin (final concentration:

8 μg/ml) resistance clones. During selection, medium was changed every 3-4 days.

After 2 weeks of selection, cells were detached with TrypLE (Life Technologies)

15   according to manufacturer's recommendations and transferred in 175cm / 75cm flasks

depending on cell concentrations. Finally 7 vials of 10*10^6 cell were frozen as cell bank

stock. The integration efficiency was measured using Celigo S cell cytometer

(Nexcelom Bioscience) analyzer and was calculated 77.8%. The stable pool of

selected cells (Fig. 28) was divided in two sub-pools. The first pool was induced with

20   doxycycline (1 μg/μL) and then bulk sorted by FACS to harvest GFP positive cells. The

other pool was directly FACS bulk sorted to harvest GFP negative cells. The first

induced GFP positive cells were grown without induction for 2 days and then FACS

bulk sorted to harvest GFP negative cells and finally after a second round of induction

the tightest preforming cells were selected as homogenous polyclonal cell population.

25   The cell pool not induced in the first place and harvested GFP negative were

subsequently induced and FACS bulk sorted to harvest GFP positive cells. Finally after

a second round of induction the cells with the best inducibility (high GFP expression

only upon induction) were selected as homogenous polyclonal cell population.

30   <u>FACS data after first sorting of permanent integrated and inducible Ca</u>

After antibiotic selection the polyclonal population was FACS sorted for ability of

induction. After the first round of sorting, the data from the comparison between the

wild type population and the 2AGFP_Cas9 inducible cell line (Fig. 29-31) shows that

our polyclonal population had about 10% of still active Cas9 and among that 10%

35   about 2% was inducible (compare induced and not induced, fig. 30 and 31). Other

120

sorting rounds have to be applied to achieve final monoclonal Cas9 expressing cell line.

Example 13. Accelerated, site-specific, clean and precise targeted integration of transgenes

This method is based on applying a fluorescent protein A (for example GFP) linked Cas9 together with sgRNA towards the integration site and donor DNA which contains a fluorescent gene B (for example mcherry) so that methods such as FACS can be used to select for the cells transfected with both Cas9 and donor DNA. In this way, the unwanted non-fluorescent (i.e. non-transfected) cells can be removed from the cell population and a much smaller pool of double positive cells expressing fluorescent protein A and B (i.e. transfected cells) is obtained to screen for targeted integration events. The mcherry (fluorescent gene B) is present outside the homology regions of the donor DNA so it is not integrated if the cells apply homology-directed (HDR) DNA repair. Only the part within the homology arms is inserted, and this can contain an expression cassette, for example a biopharmaceutical (gene of interest, GOI). HDR is error-free compared to non-homologous end-joining (NHEJ) which facilitates controlled and precise genome editing event dictated by the donor DNA. After clone generation, the cells not expressing mcherry (fluorescent protein B) are selected since to also get rid of the cells with random integrated donor DNA. In this way, the screening time to get a producing cell line is shortened without using selection and without leaving unwanted DNA parts integrated in the genome of the cells. The non-fluorescent clones can be screened by 3' and 5' junction PCR to identify targeted integrants. Additionally, the level of expressed recombinant protein (GOI) can be analysed to find the HDR mediated targeted integrants as these tend to produce the same amount of protein due to lower clonal expression variations compared to random integrants. This method can be used to support screening of potential integration sites supporting high and stable expression of GOI.

Example 14. Permanently integrated CRISPR Cas9 mediated multiplexing in CHO cells

We propose MACE (Multiplex Automated CHO Engineering) where a genome integrated, inducible 2AGFP_Cas9 cell line can be successfully used for genome modifications either with single or multiple targets in one or multiple round of

transfections, e.g. simultaneous gene disruption and site specific gene integration. The cell line offers the efficacy of the Cas9 endonuclease for genome modification integrated in the genome and co-expressed with GFP allowing monitoring of the activity and fast bulk sorting. Moreover the endonuclease is completely inducible thus creating

5   a homogeneous expression within a monoclonal population, that in turn reflects homogeneity also in the data set. It also permits the transfections of more gRNAs since it is not necessary to introduce a Cas9 expressing vector and allows a flexible protocol that can be easily automated in loop and ultimately in a chip (Fig. 33). The expression of the endonuclease is also ensured and the silencing avoided by the use of

10   euchromatin structure stabilizing genetic elements.

This cell line decreases the screening load for the identification of multiple genes disrupted in CHO cell lines since the Cas9 activity can be easily measured and all the population can use unified parameters for transfection variability since all the cells in a

15   monoclonal cell line are exposed to the same concentration of endonuclease. The only variable parameter would be the amount of gRNAs transfected but that variability could also be overcome by associating a gRNA with a marker that can be selected for or counter-selected.

20   The potential of this cell line can be also extended for recycling integration sites by regenerating targets site in safe harbors of insertion using a set of unique Insertion gRNAs.

In conclusion, not only will this cell line allow development of a high-throughtput, user-

25   friendly platform and give the benefits of homogeneity and reproducibility, recycling elements and possibility of being automated but it is also possible to create an endonuclease-free background in the end of the engineering loop by just knocking out Cas9 itself using Cas9 self-destructing gRNAs.

30   In addition to the flexibility and potential for robotic automation, this cell line can also be resistant to exogenous contamination that can infect the cell line by integration or a vector base of set of constitutively expressed gRNAs targeting specific invaders DNA sequences thus conferring immunity to the cell line.

Claims

1. A multiplex editing system comprising at least two pluralities of Continuously Regenerated Endonuclease Site Cassettes CRESCs, wherein:

   i. a first plurality of CRESCs comprises at least one first targeted endonuclease site and at least one other nucleic acid sequence; and

   ii. a second plurality of CRESCs comprises at least one second targeted endonuclease site and at least one other nucleic acid sequence,

   wherein the first and second targeted endonuclease sites are different.

2. The system of claim 1 comprising at least three pluralities of CRESCs, wherein the first and second pluralities are as defined in claim 1, and the third plurality comprises at least one third targeted endonuclease site and at least one other nucleic acid sequence, said at least one third targeted endonuclease site being either:

   i. identical to the at least one first or second targeted endonuclease sites; or

   ii. different from the at least one first and second targeted endonuclease sites.

3. The system according to claim 1 or 2, wherein at least one of the CRESCs further comprises two homology arms HA_L and HA_R, wherein one is 5'-terminal and the other is 3'-terminal, and wherein HA_L and HA_R are homologous to a target nucleic acid sequence, T_L, and a target nucleic acid sequence T_R, respectively, where T_L and T_R delimit a nucleic acid T.

4. The system according to any of the preceding claims, wherein at least one of the CRESCs comprising at least one targeted endonuclease site comprises two identical targeted endonuclease sites, wherein creation of a break at each targeted endonuclease site by an endonuclease results in loss of the nucleic acid surrounded by the two target endonuclease sites.

5. A method for editing nucleic acids with the system according to any of claims 1 to 4, comprising the steps of:

   i. introducing a first CRESC in a cell capable of expressing at least one

endonuclease;

ii. allowing one of the at least one endonuclease to create a break in a nucleic acid comprised within said cell; thereby allowing integration of the first CRESC or at least a part thereof in said nucleic acid;

iii. introducing a second CRESC in said cell:

iv. allowing one of the at least one endonuclease to create a break in the first CRESC; thereby allowing integration of the second CRESC or at least a part thereof in the first CRESC;

said method optionally further comprising the steps of :

v. introducing any subsequent CRESC in said cell; and

vi. allowing one of the at least one endonuclease to create a break in the previously integrated CRESC, thereby allowing integration of the subsequent CRESC or at least a part thereof in the previous CRESC, wherein any two CRESCs or at least a part thereof may be introduced simultaneously or serially in said cell, thereby allowing for simultaneous and/or serial editing of multiple loci.

6. The method according to claim 5, wherein the editing of nucleic acids is knocking in and/or knocking out of a nucleic acid sequence within a genome or a plasmid.

7. The method according to any of claims 5 to 6, wherein the at least one endonuclease is Cas9, a zinc finger nuclease or a TALEN expressed from the genome or from a plasmid, optionally under the control of an inducible promoter.

8. The method according to any of claims 5 to 7, wherein the at least one endonuclease is at least two endonucleases, and wherein each endonuclease is expressed from a different inducible promoter.

9. The method according to any of claims 5 to 8, wherein the cell further comprises means for targeting the at least one endonuclease to a targeted endonuclease site.

10. The method according to any of claims 5 to 9 with the system of claim 3,

wherein the homology arms direct the integration of the CRESC or at least a part thereof to integration by homologous recombination.

11. The method according to any of claims 5 to 10, wherein:

i.   the at least one endonuclease is tagged with a first fluorescent protein;

ii.   at least one of said CRESCs further comprises two homology arms HA_L and HA_R, wherein one is 5'-terminal and the other is 3'-terminal, and

iii.   wherein HA_L and HA_R are homologous to a target nucleic acid sequence, T_L, and a target nucleic acid sequence T_R, respectively, where T_L and T_R delimit a target nucleic acid T; and

iv.   wherein said at least one of said CRESCs functionally expresses a second fluorescent protein from a nucleic acid sequence located outside said homology arms HA_L and HA_R,

said method further comprising the steps of:

v.   positively selecting for cells functionally expressing said first and second fluorescent protein prior to the steps of allowing said at least one endonuclease to create a double-strand break;

vi.   negatively selecting for cells expressing said second fluorescent protein subsequent to the steps of allowing said at least one endonuclease to create a break.

12. The method according to any of claims 5 to 10, wherein:

i.   the at least one endonuclease is tagged with a first fluorescent protein;

ii.   at least one of said CRESCs further comprises two homology arms HA_L and HA_R, wherein one is 5'-terminal and the other is 3'-terminal, and

iii.   wherein HA_L and HA_R are homologous to a target nucleic acid sequence, T_L, and a target nucleic acid sequence T_R, respectively, where T_L and T_R delimit a target nucleic acid T; and

iv.   wherein said at least one of said CRESCs functionally expresses a second fluorescent protein from a nucleic acid sequence located inside said homology arms HA_L and HA_R,

said method further comprising the steps of:

v.   positively selecting for cells functionally expressing said first and second

fluorescent protein prior to the steps of allowing said at least one
endonuclease to create a double-strand break;

vi.   positively selecting for cells expressing said second fluorescent protein
        subsequent to the steps of allowing said at least one endonuclease to
        create a double-strand break;

vii.   optionally, excising the nucleic acid sequence encoding said second
        fluorescent protein from the target nucleic acid sequence.

13. The method according to any one of claims 11 to 12, wherein said steps of
    positively and/or negatively selecting for cells are performed using
    fluorescence-activated cell sorting (FACS).

14. The method according to any one of claims 11 to 13, wherein the first
    fluorescent protein is selected from the group consisting of GFP, enhanced
    GFP, wherein the sequence coding for the first fluorescence protein may be
    codon-optimised, and wherein the endonuclease tagged with said first
    fluorescent protein further comprises a linker.

15. The method according to any one of claims 11 to 14, wherein the endonuclease
    tagged with a first fluorescent protein is GFP-2A-Cas9.

16. The method according to any one of claims 11 to 15, wherein the endonuclease
    is stably integrated in the genome of said cell.

17. The method according to any one of claims 5 to 13, wherein the at least one
    endonuclease is Cas9 or a variant thereof.

18. The method according to claim 17, wherein said Cas9 variant is selected from
    the group consisting of Cas9, a D10A Cas9 mutant and a H840A Cas9 mutant.

19. A cell comprising a stably integrated endonuclease gene such as Cas9 or a
    variant thereof.

20. The cell of claim 19, wherein the variant is selected from the group consisting of
    fluorescently-tagged Cas9, a D10A Cas9 mutant and a H840A Cas9 mutant, a
    fluorescently-tagged D10A Cas9 mutant and a fluorescently-tagged H40A Cas9

mutant.

21. The cell of any of claims 19 or 20, further comprising means for recognising at least two targeting sequences.

22. The cell of any of claims 19 or 21, wherein the endonuclease comprises a nuclear localisation signal and expression of said endonuclease is regulated by an inducible promoter.

23. A kit of parts comprising at least two pluralities of CRESCs, wherein:

    i. a first plurality of CRESCs comprises at least one first targeted endonuclease site; and

    ii. a second plurality of CRESCs comprises at least one second targeted endonuclease site,

wherein the first and second targeted endonuclease sites are different, said kit of parts further comprising instructions for use.

24. The kit according to claim 23, further comprising a nucleic acid sequence encoding a functional endonuclease capable of recognising at least one of said first and second targeted endonuclease sites.

25. The kit according to claim 24, wherein said functional endonuclease is Cas9 or a variant thereof, said variant being selected from the group consisting of fluorescently-tagged Cas9, the D10A Cas9 mutant and the H840A Cas9 mutant.

26. The kit according to any one of claims 23 to 25, further comprising a cell according to any one of claims 19 to 22.

1a

1b
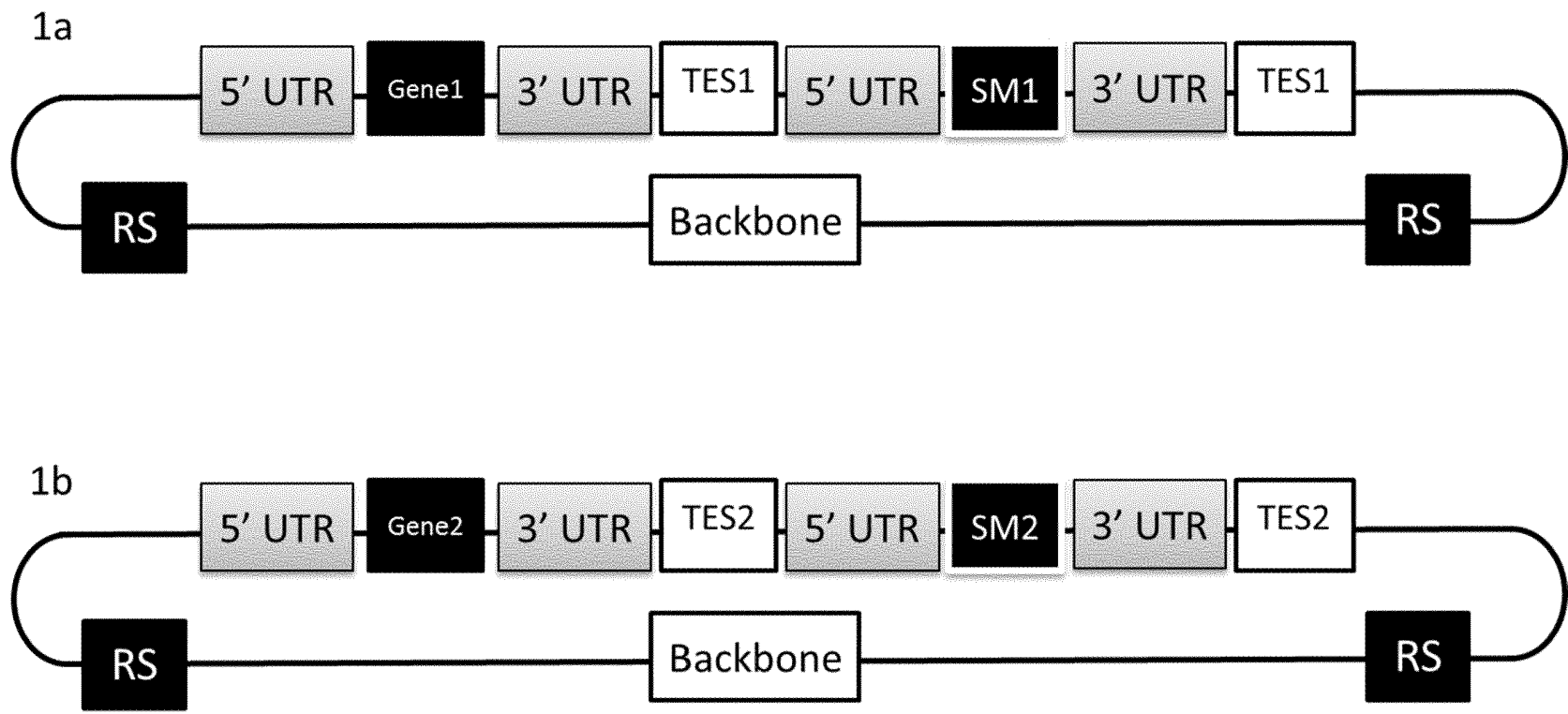
FIG. 1

FIG. 2
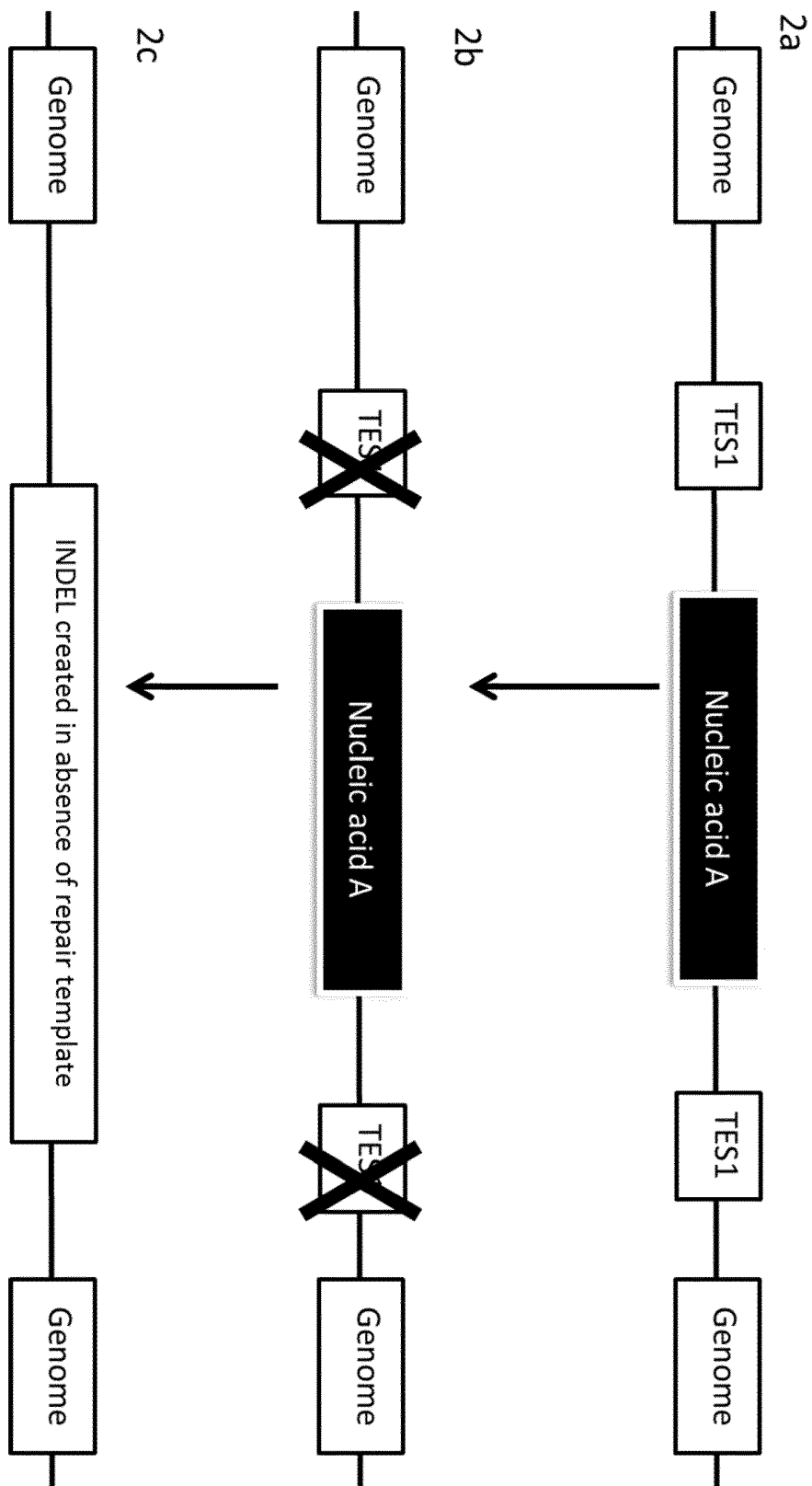
FIG. 3
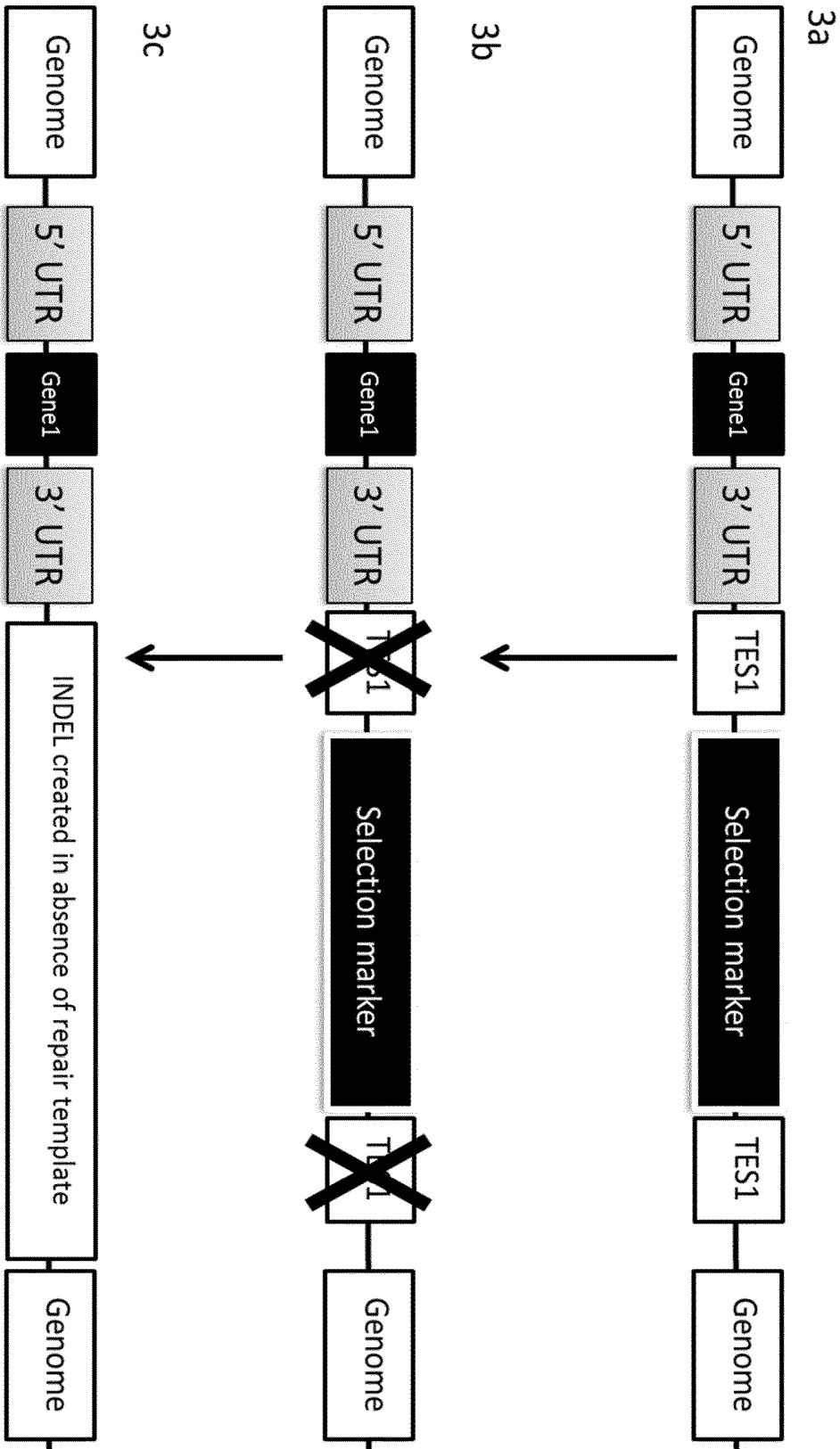
FIG. 4

FIG. 5

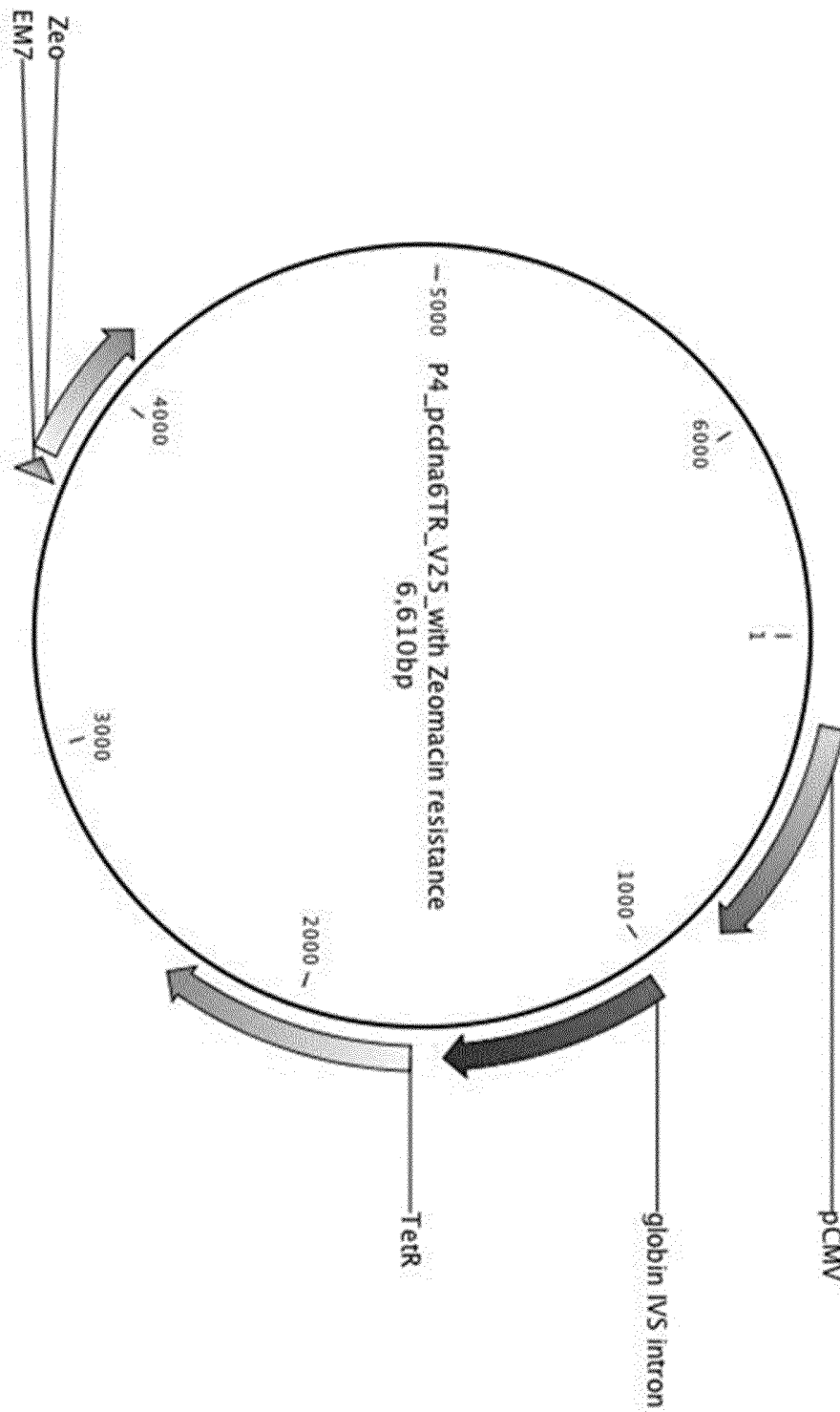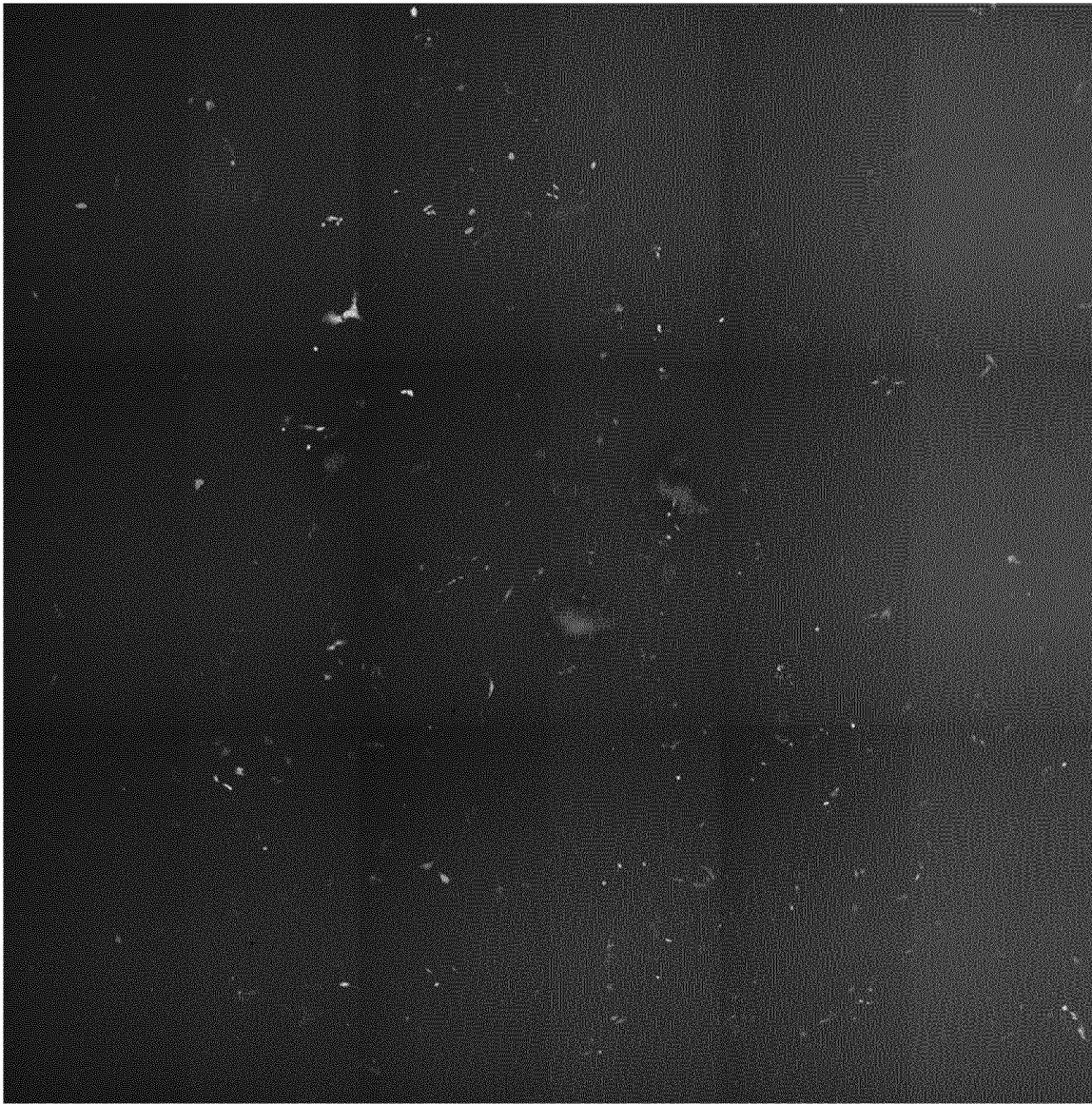FIG. 5 (cont.)

FIG. 5 (cont.)

5c

Target site in transformed cell line no. 2

CRESC3

Transformed cell line no. 3
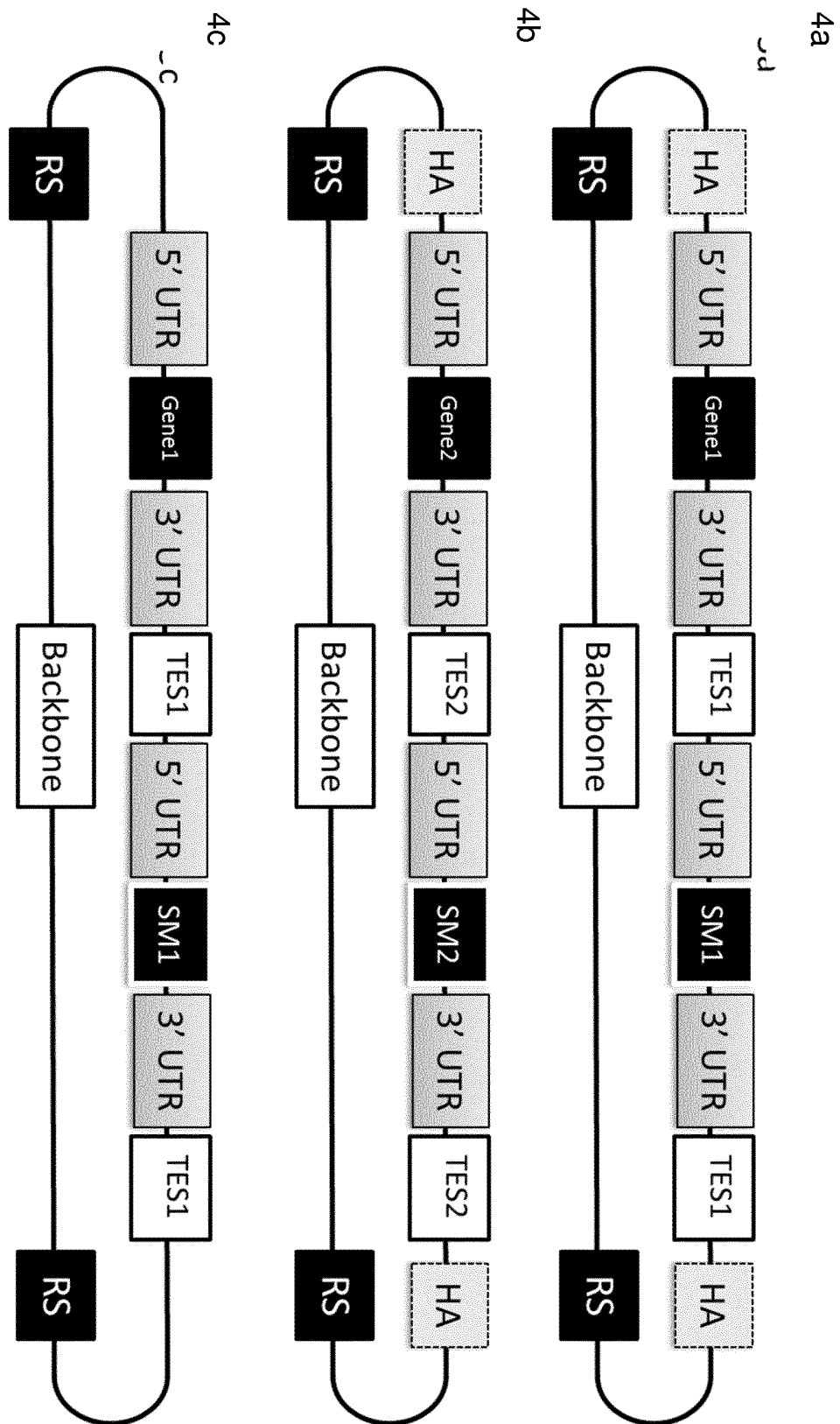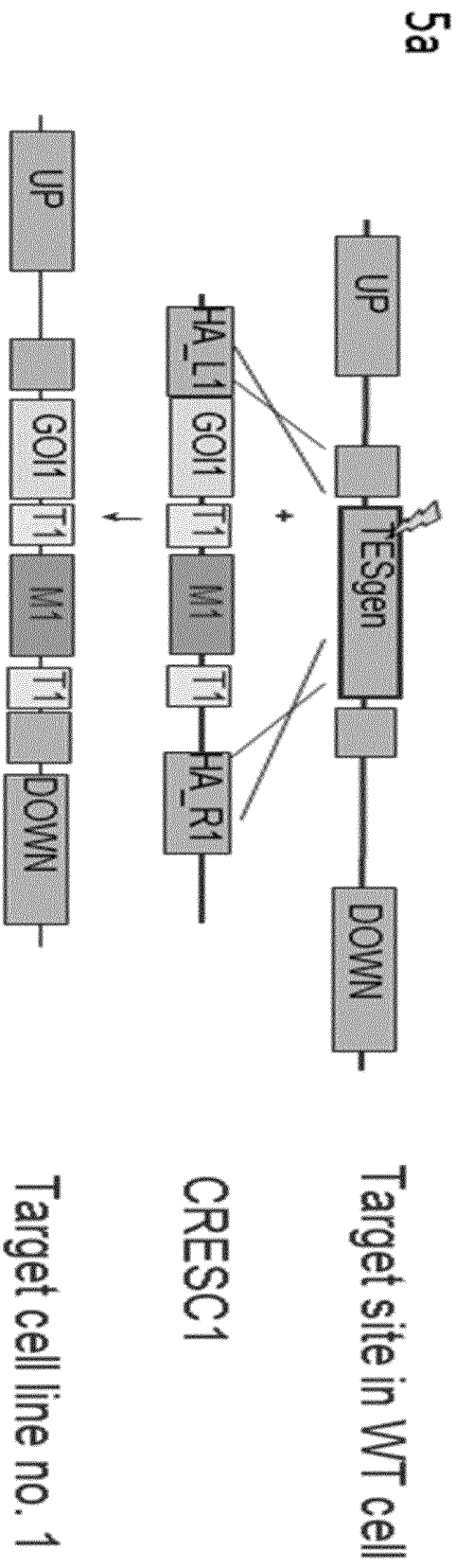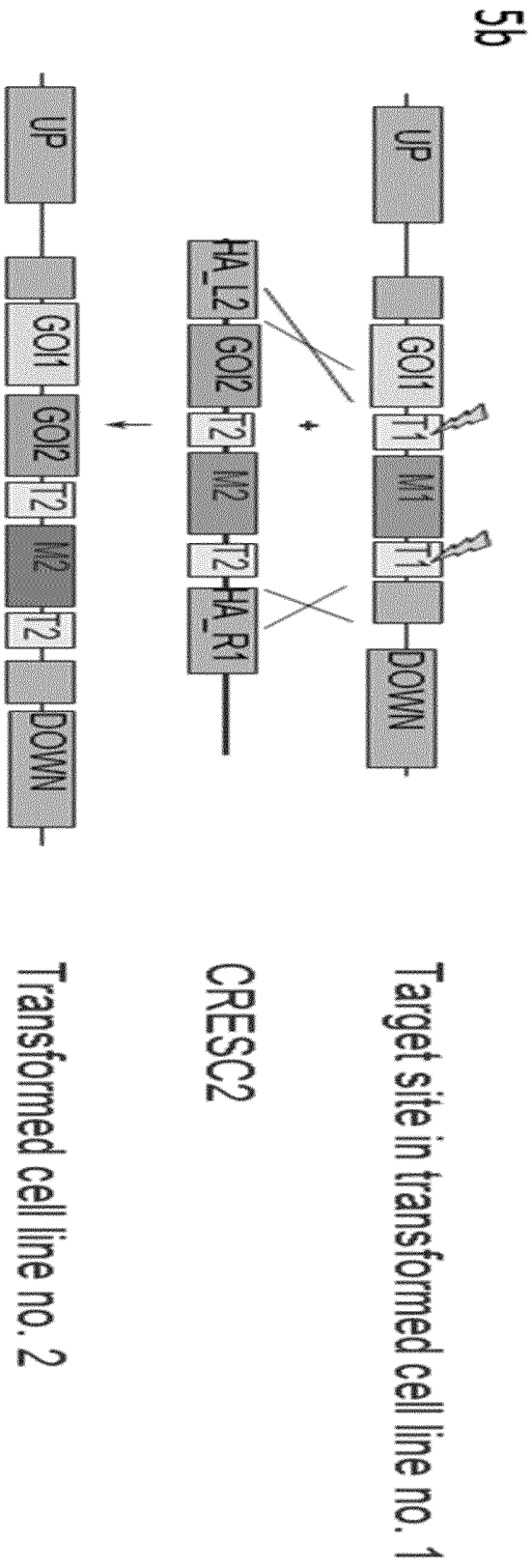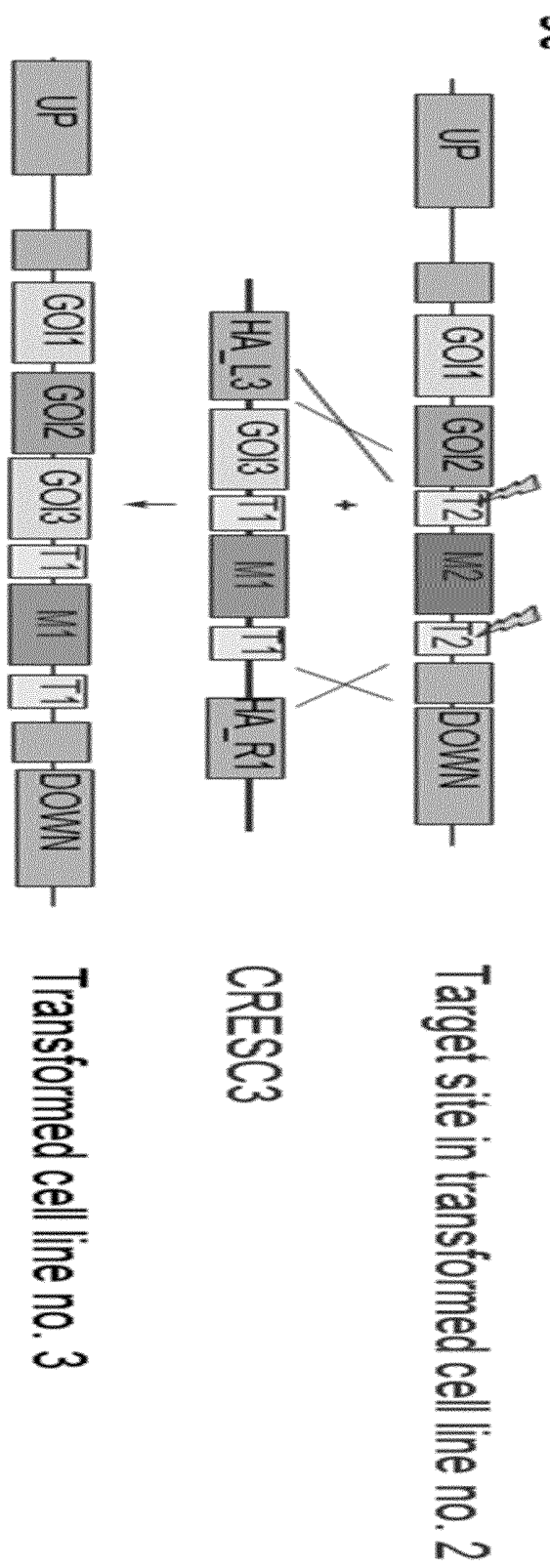
FIG. 6
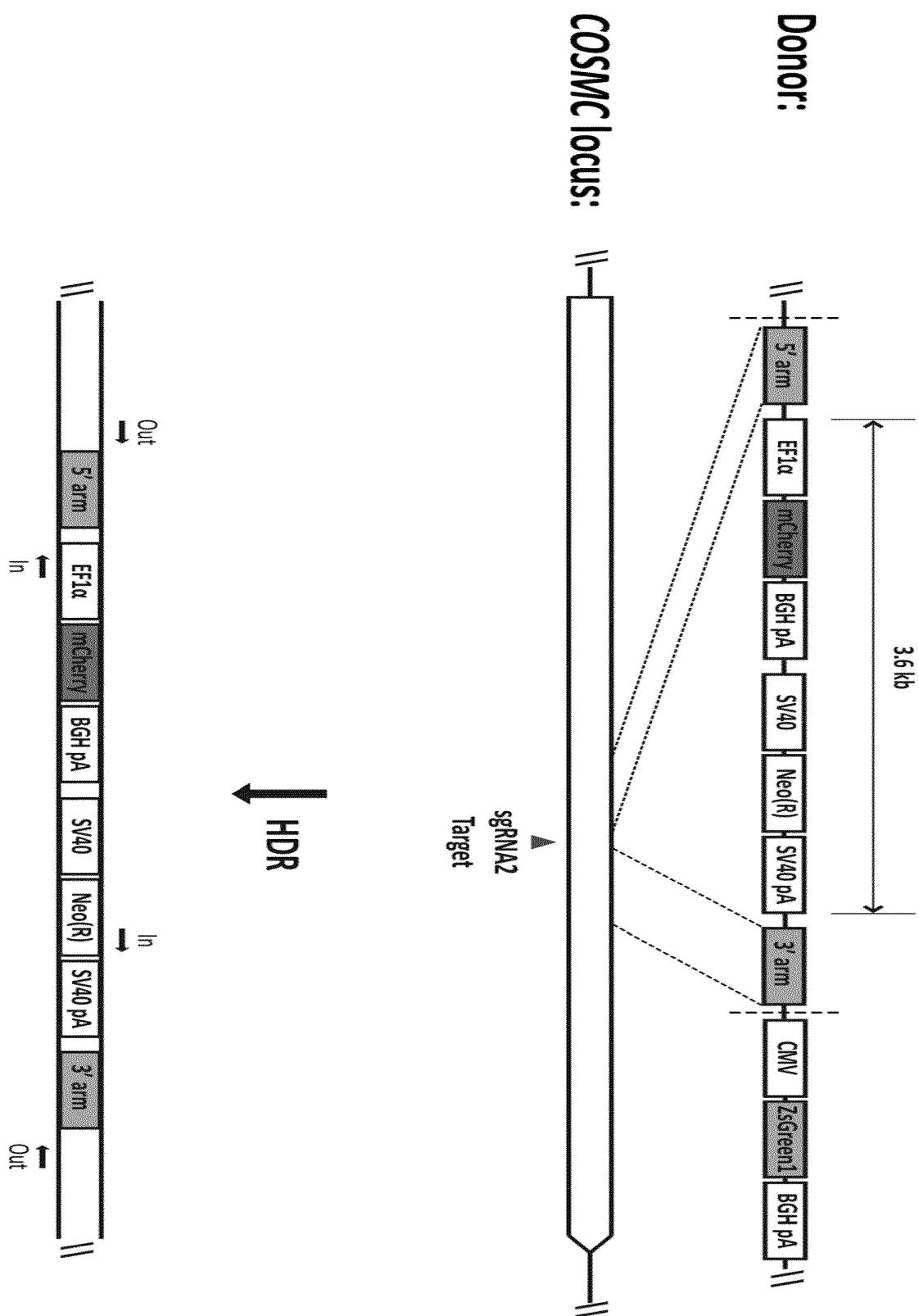
FIG. 7
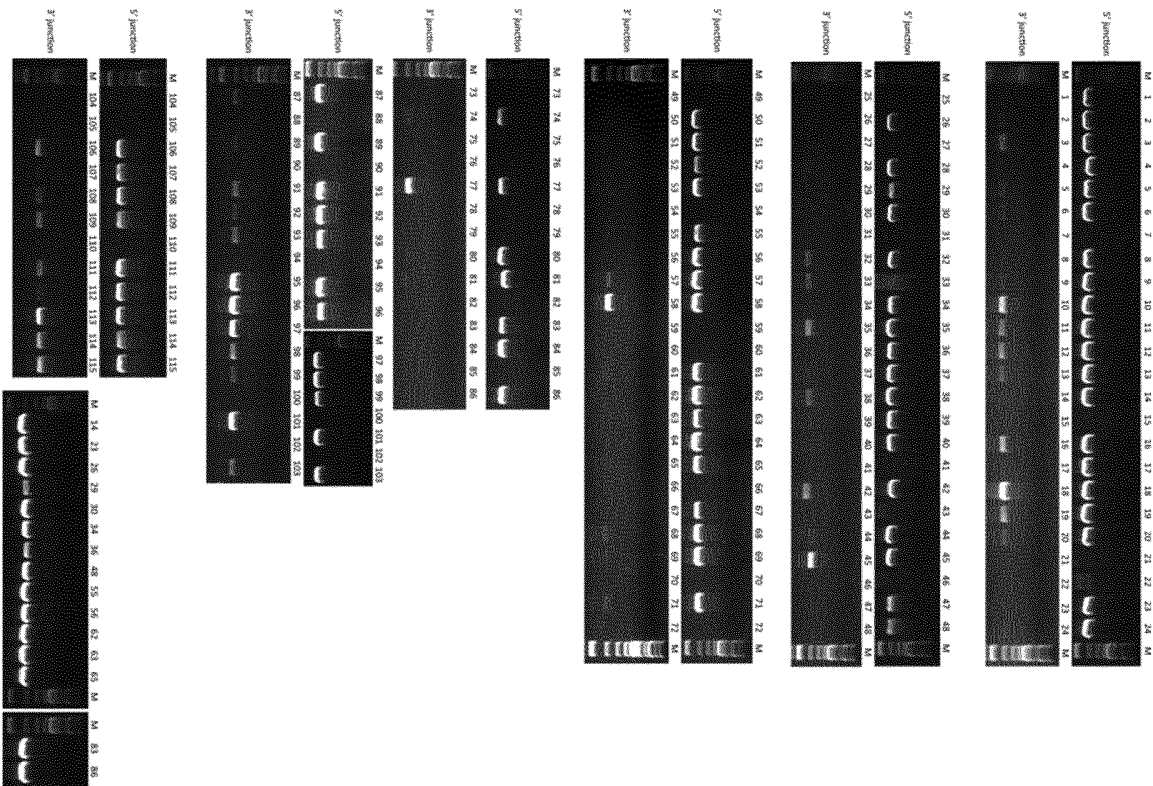
Summary	Clone #	%
Only 5' Junction PCR positive	0	0
Only 3' Junction PCR positive	1	0.9
Both junction PCR negative	35	30.4
Both junction PCR positive	79	68.7
Total	115	100
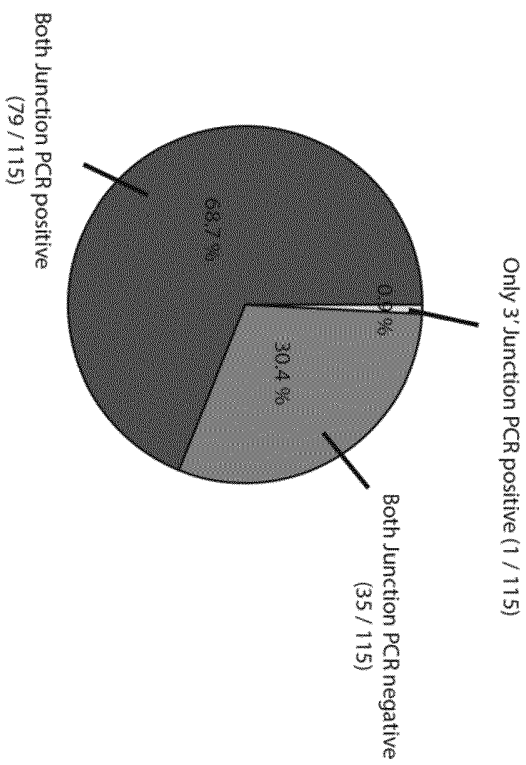
FIG. 8
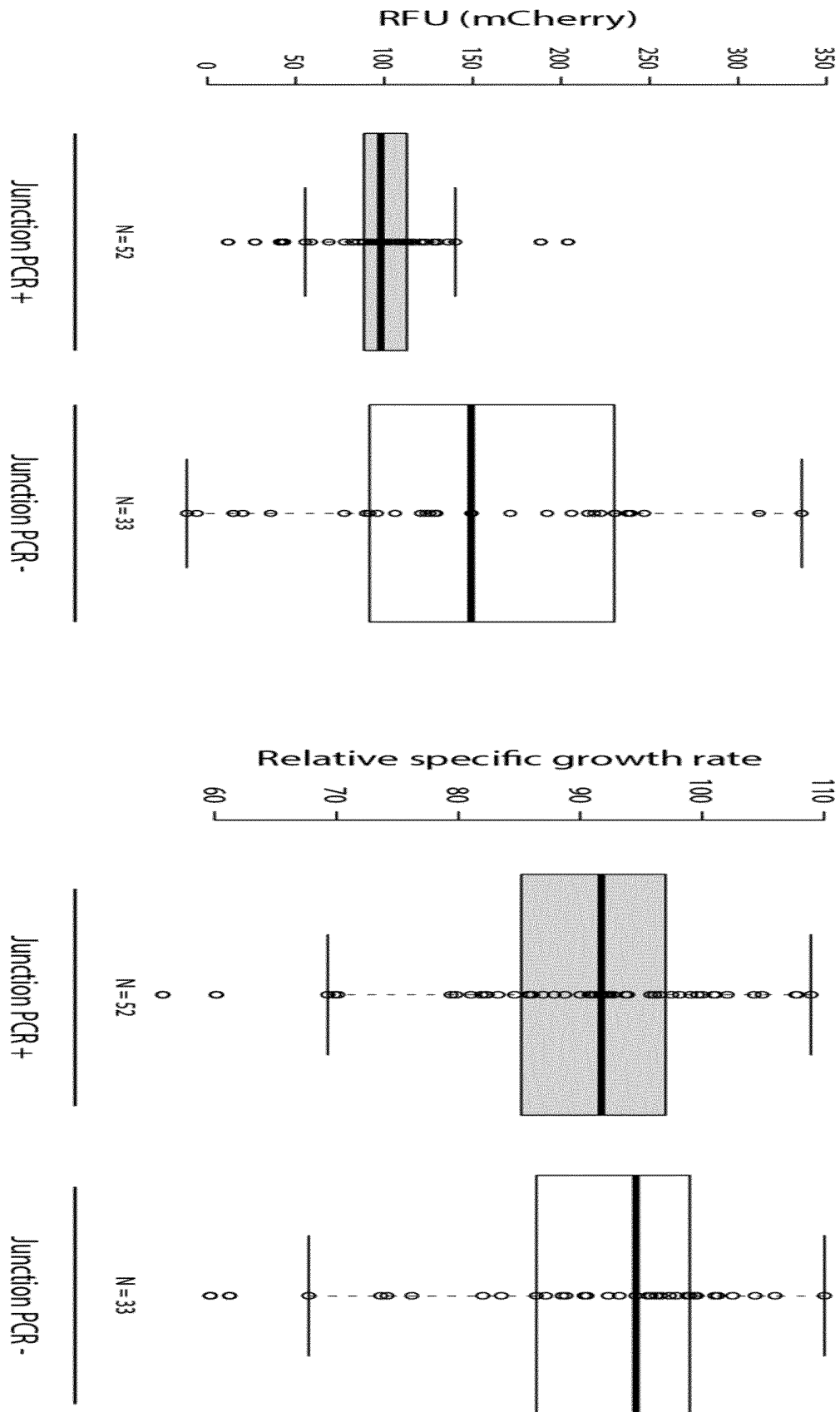
FIG. 9
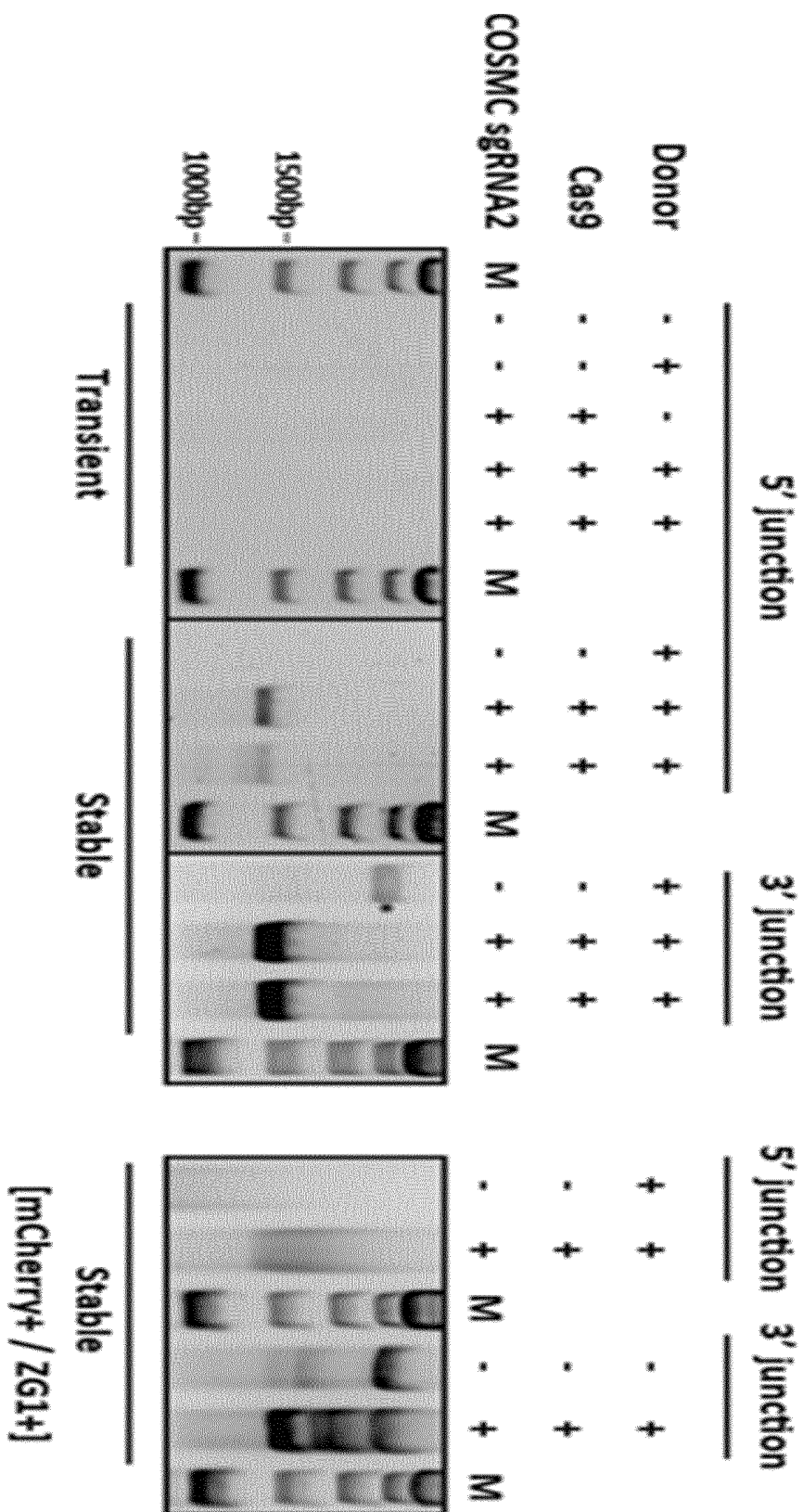
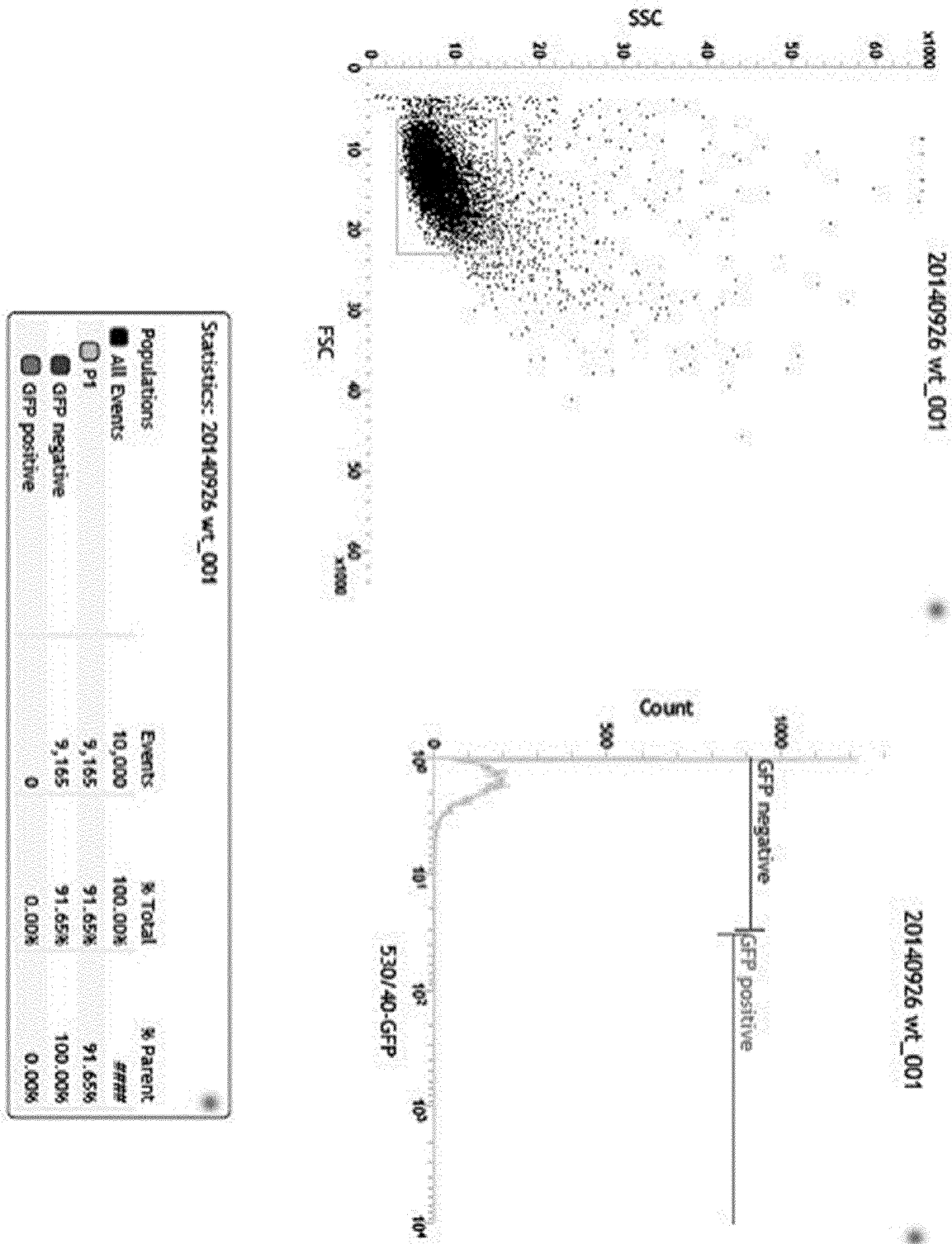FIG. 10
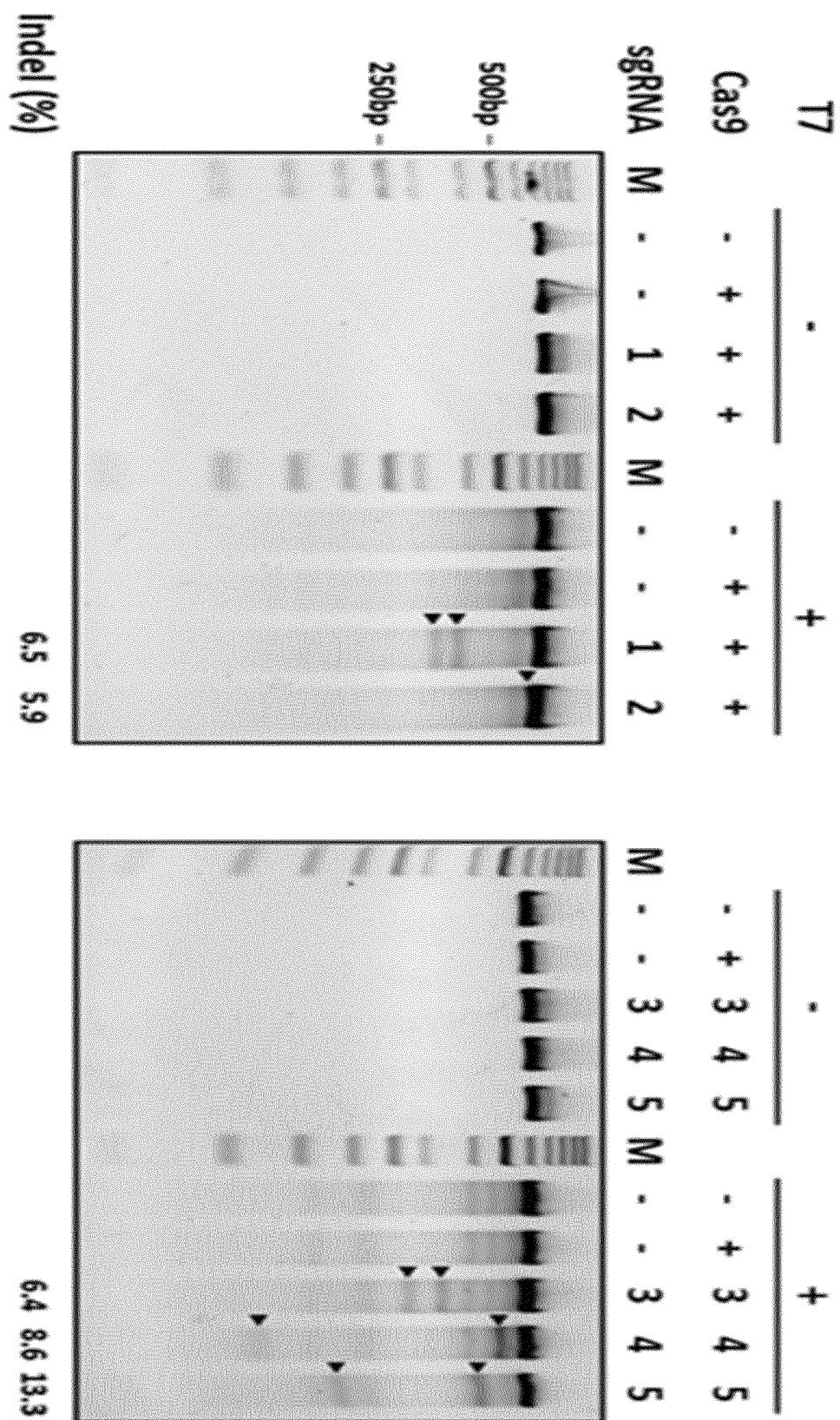
FIG. 11
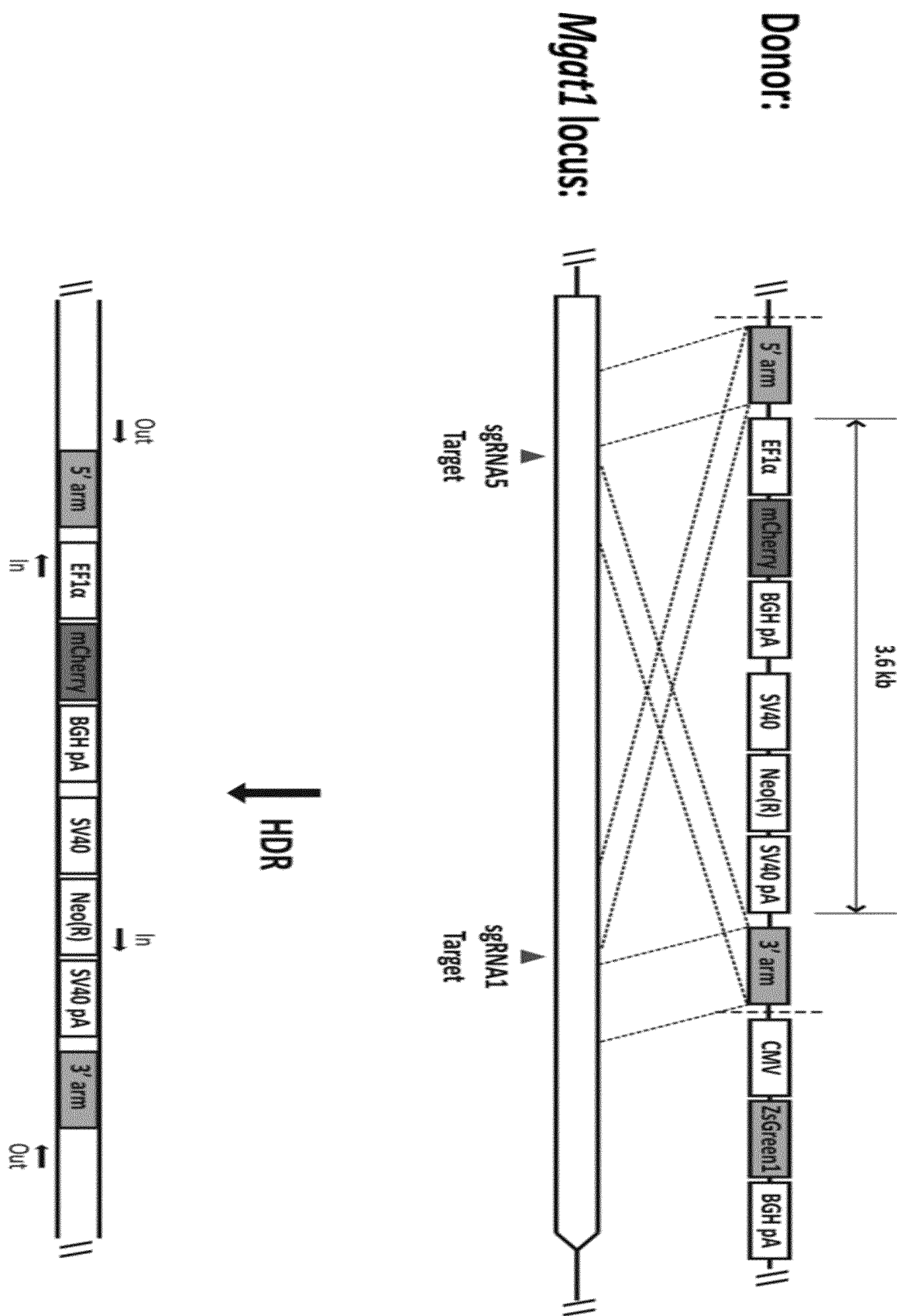
FIG 12
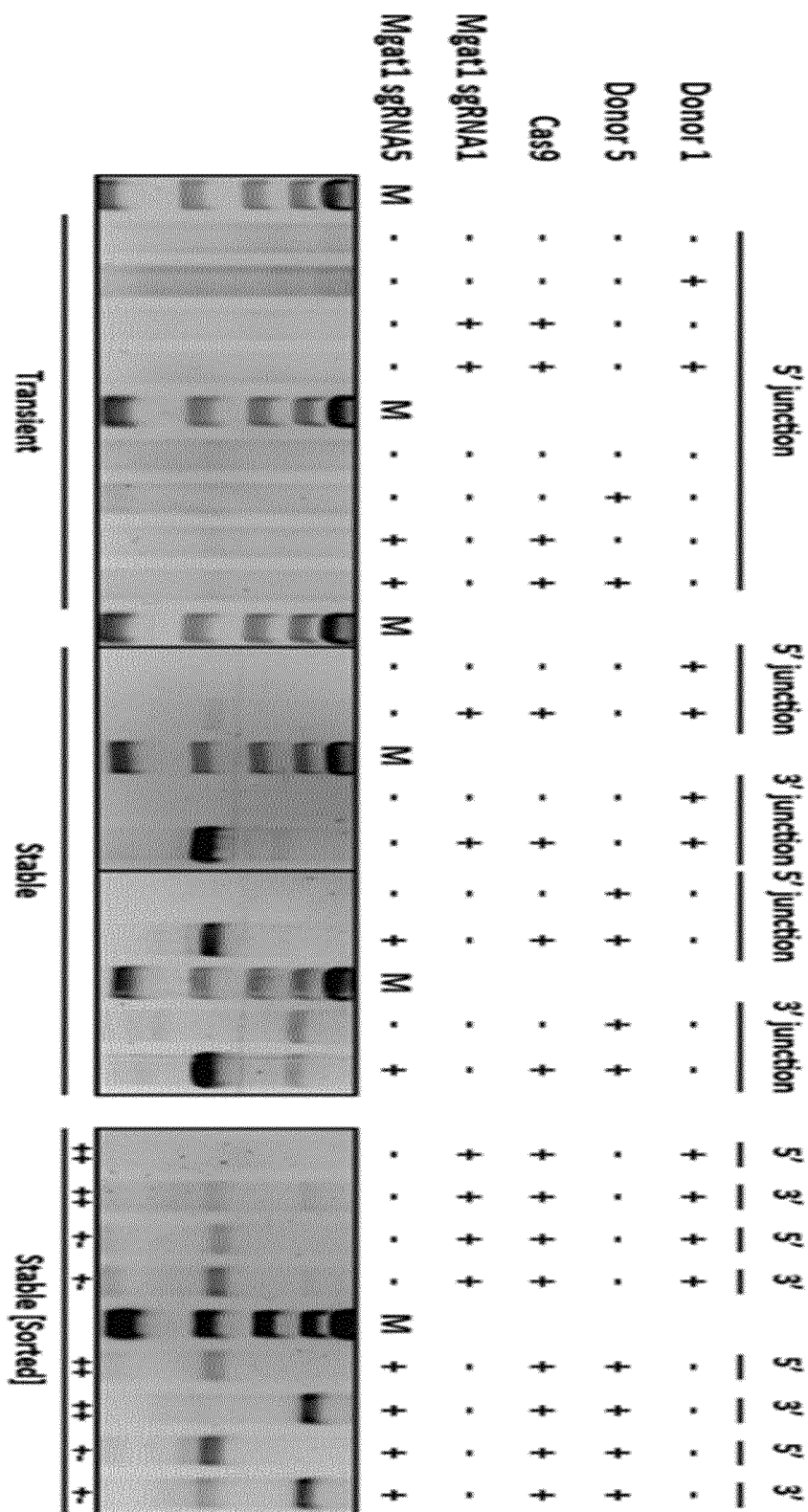
FIG. 13
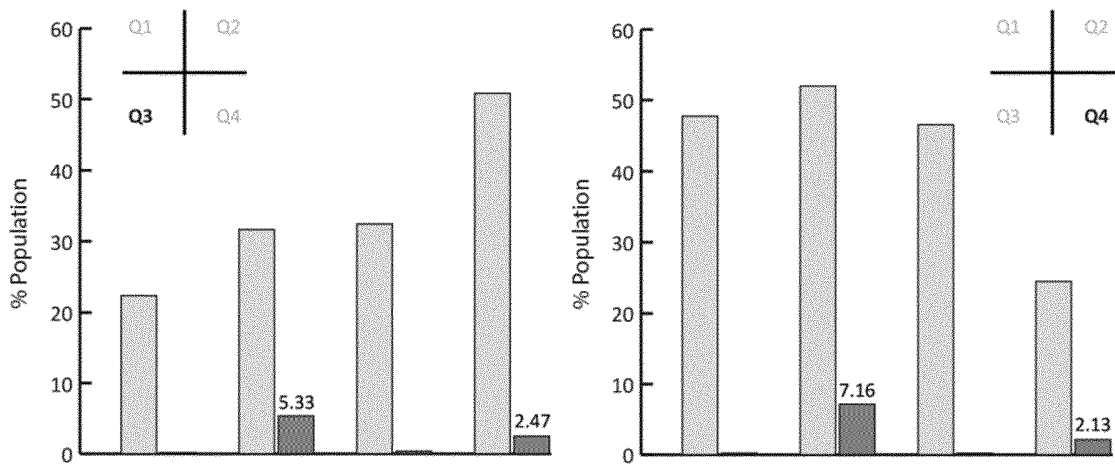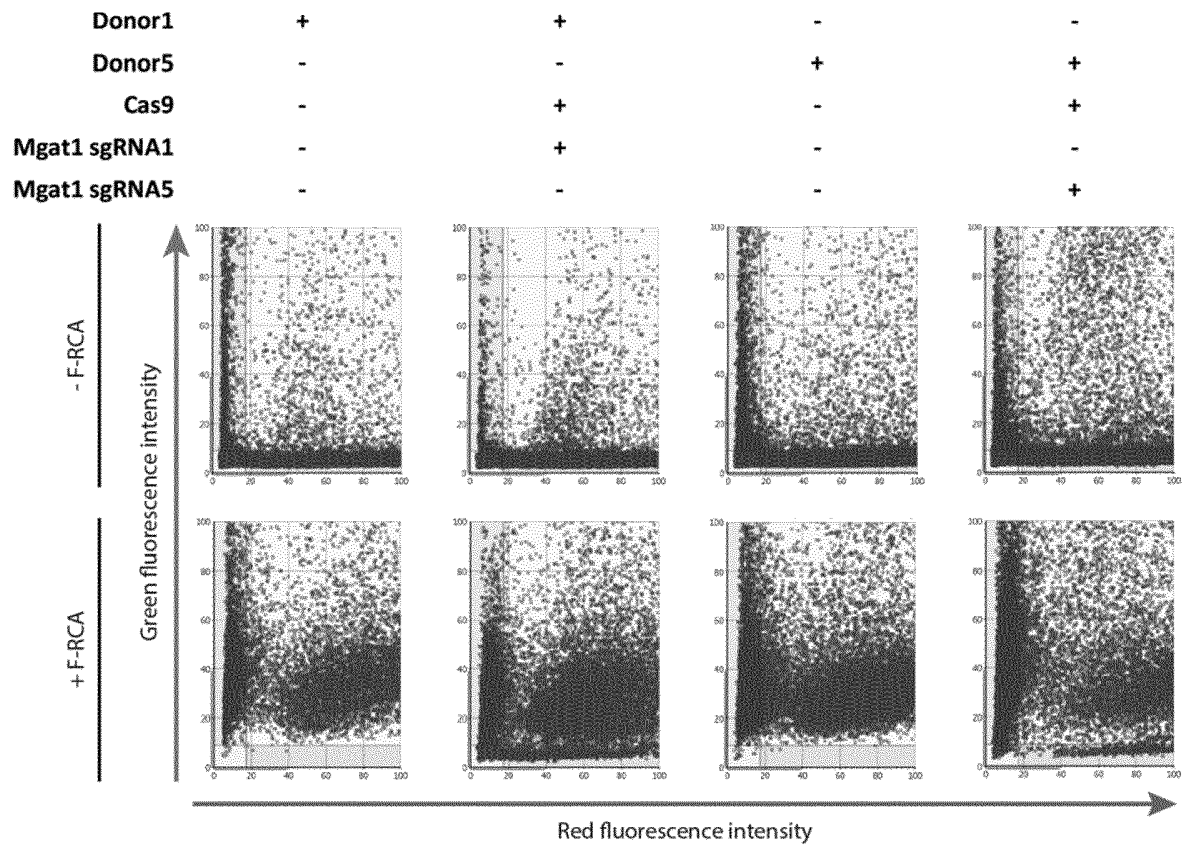
FIG. 14
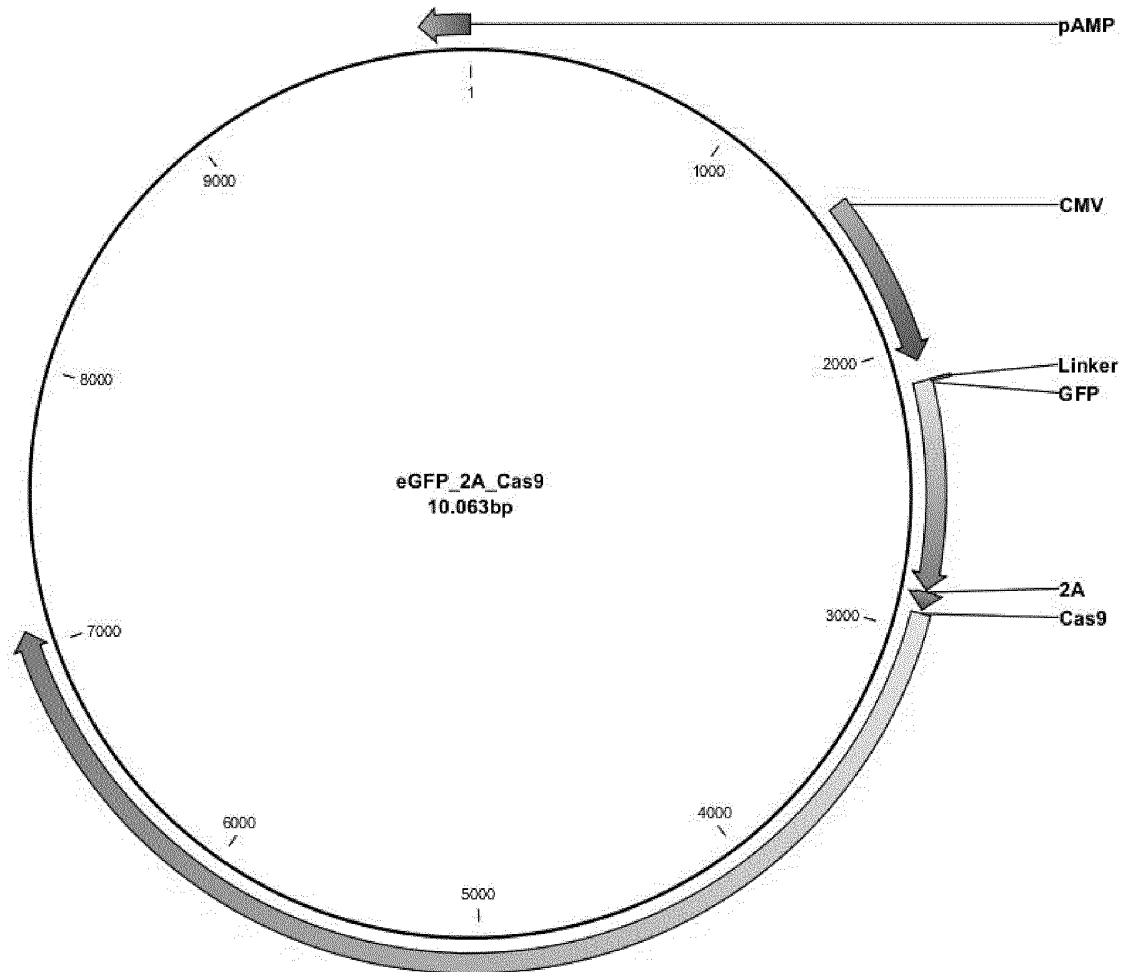
FIG. 15
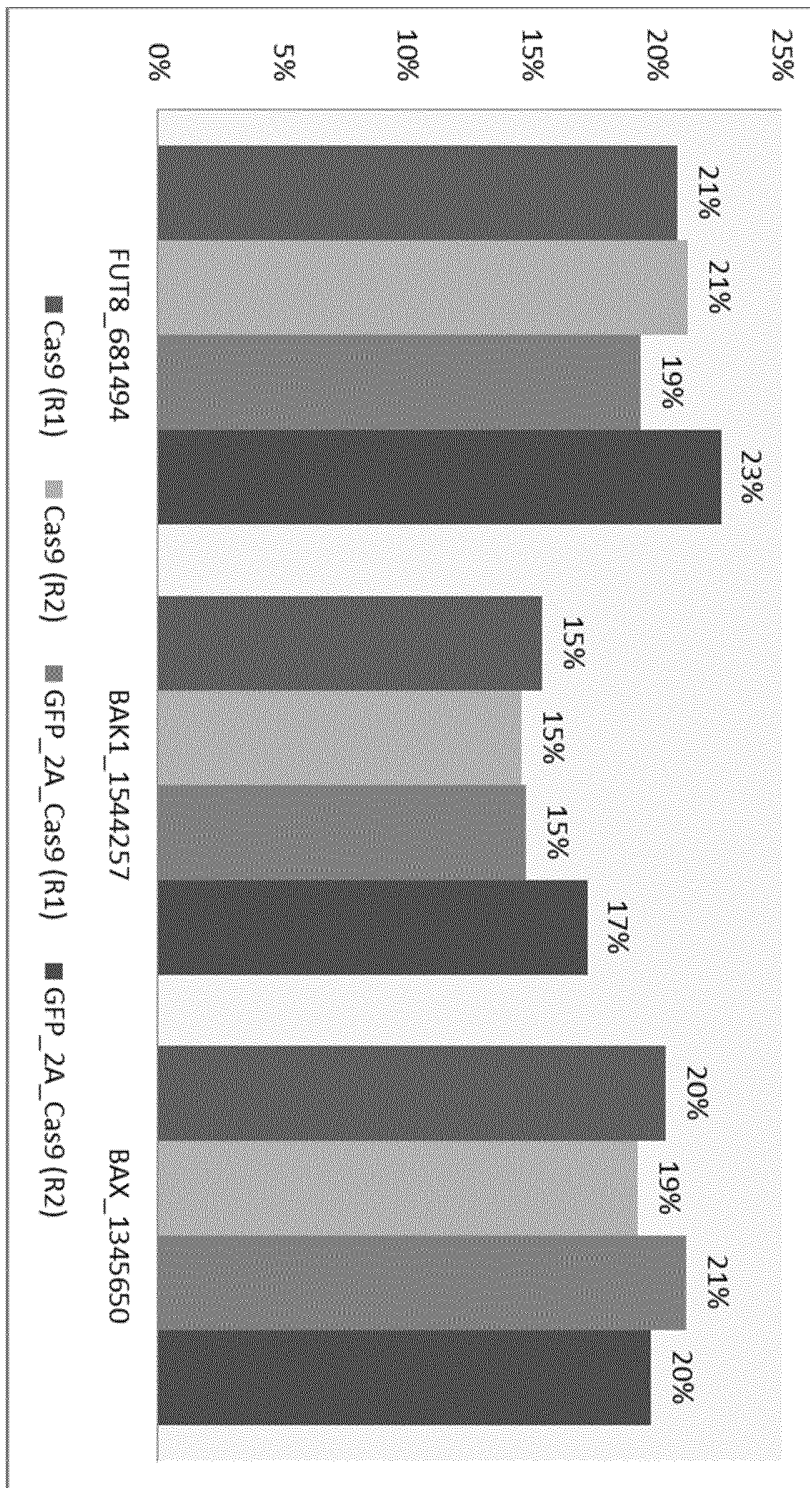
FIG. 16
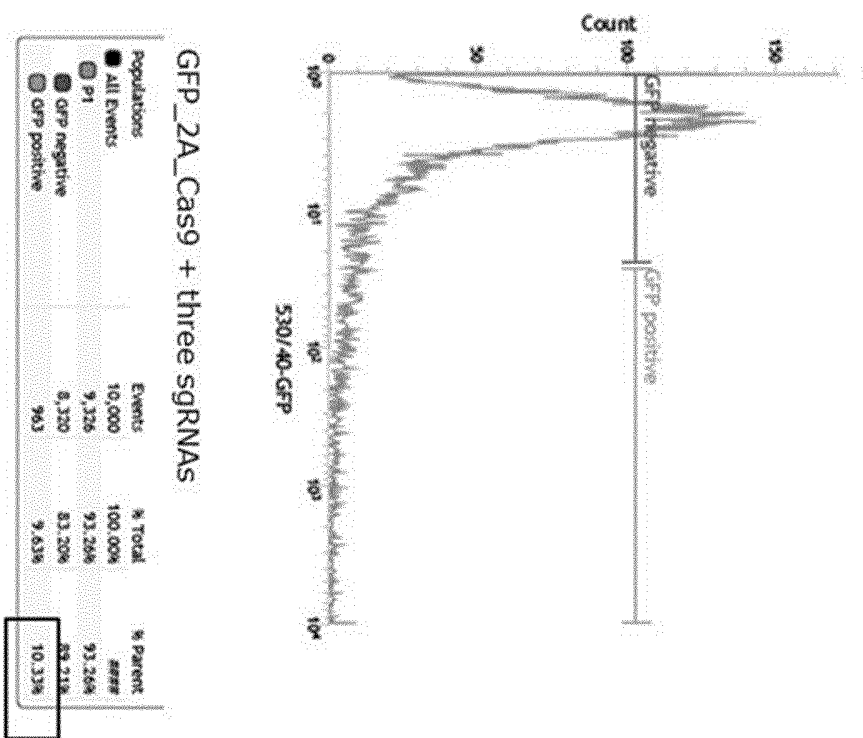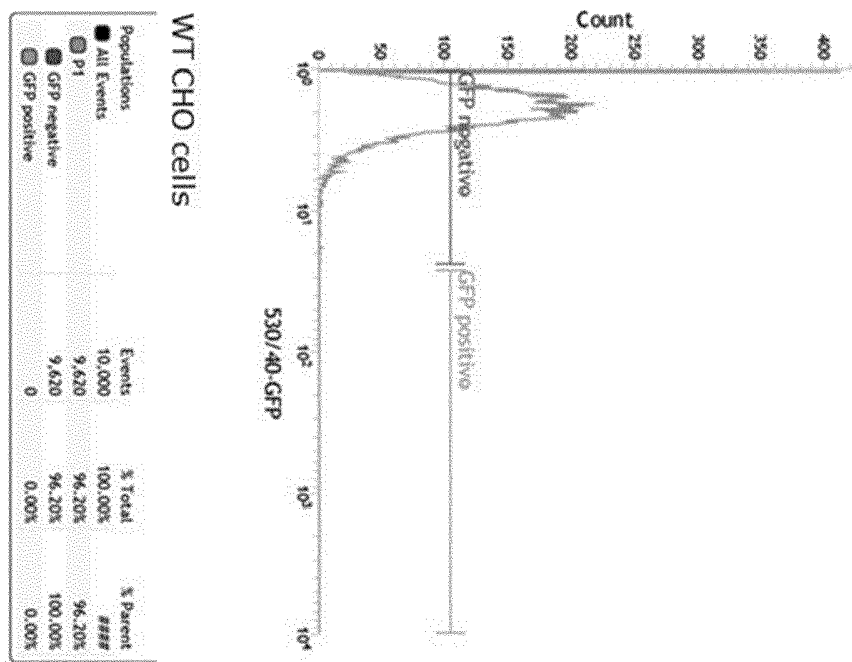
FIG. 17
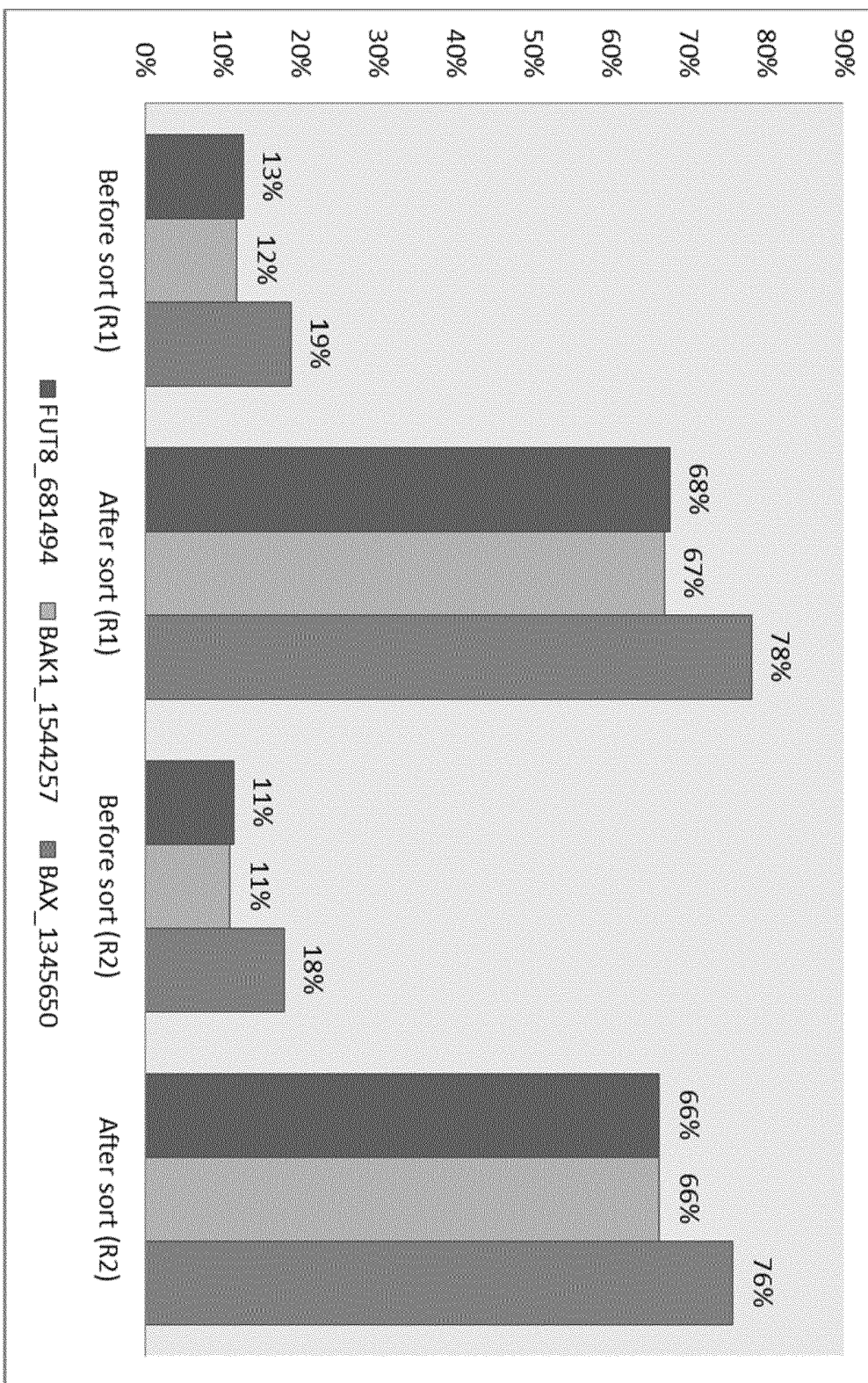
FIG. 18
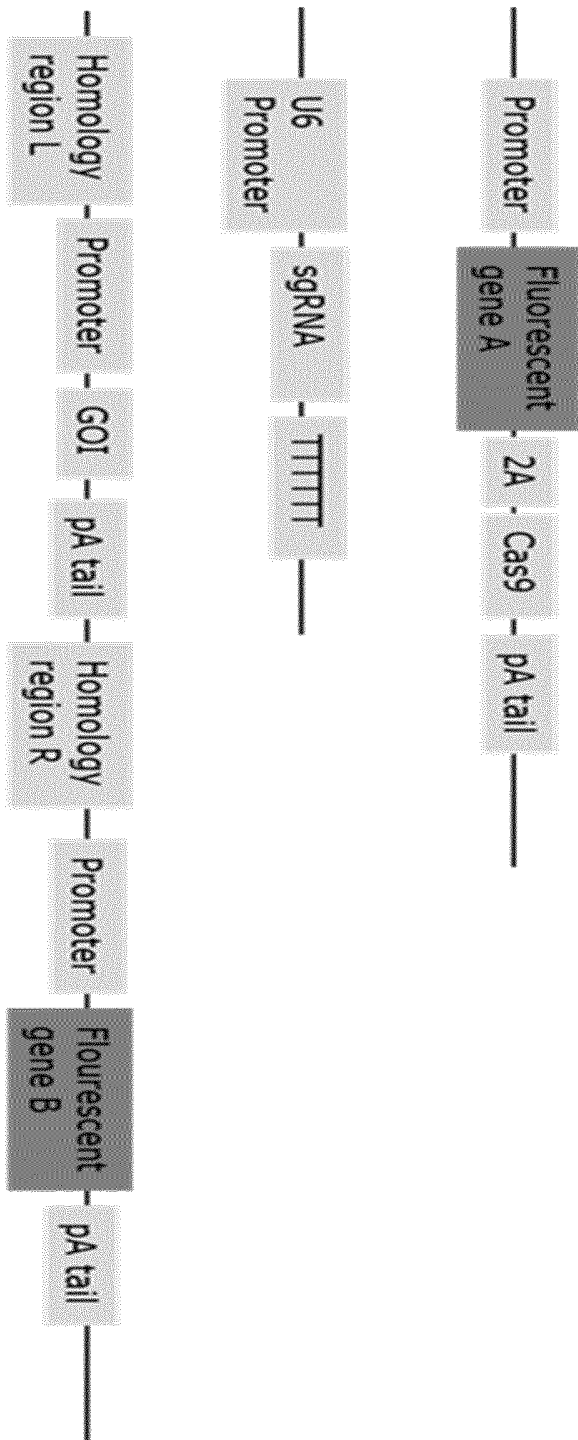
FIG. 19
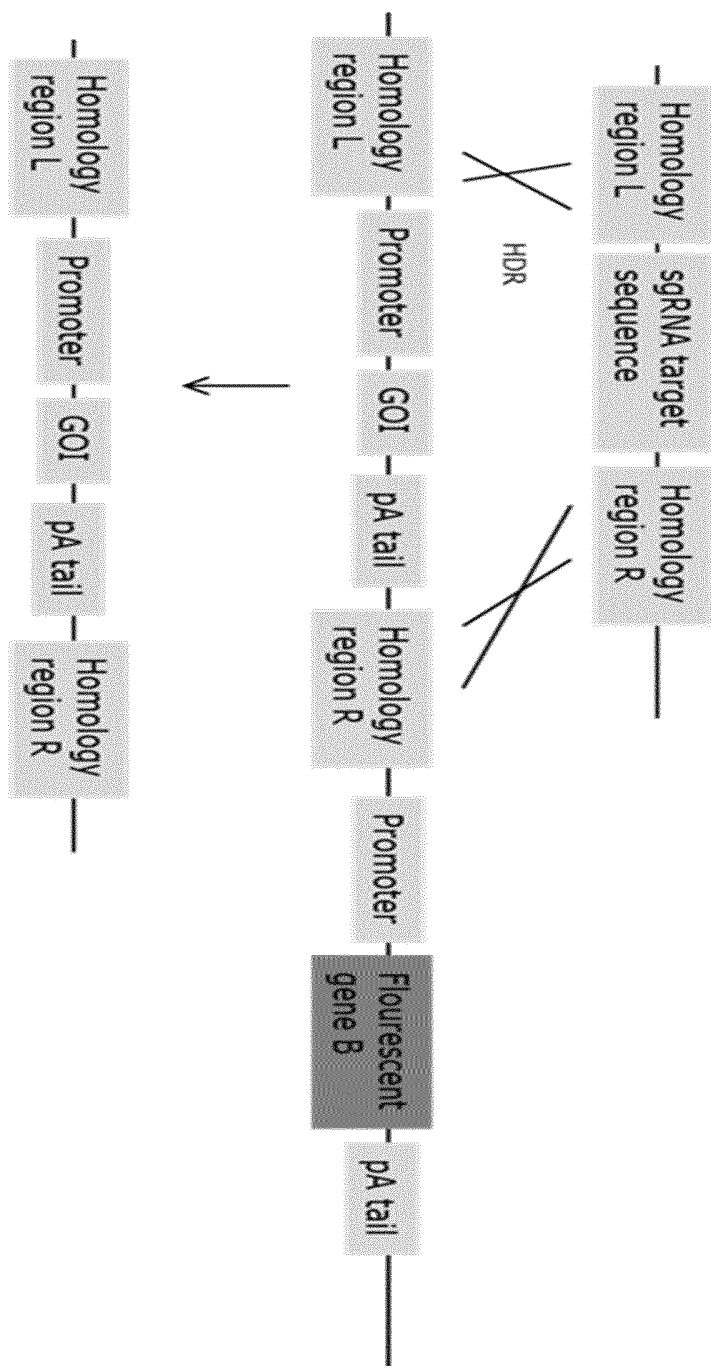
FIG. 20
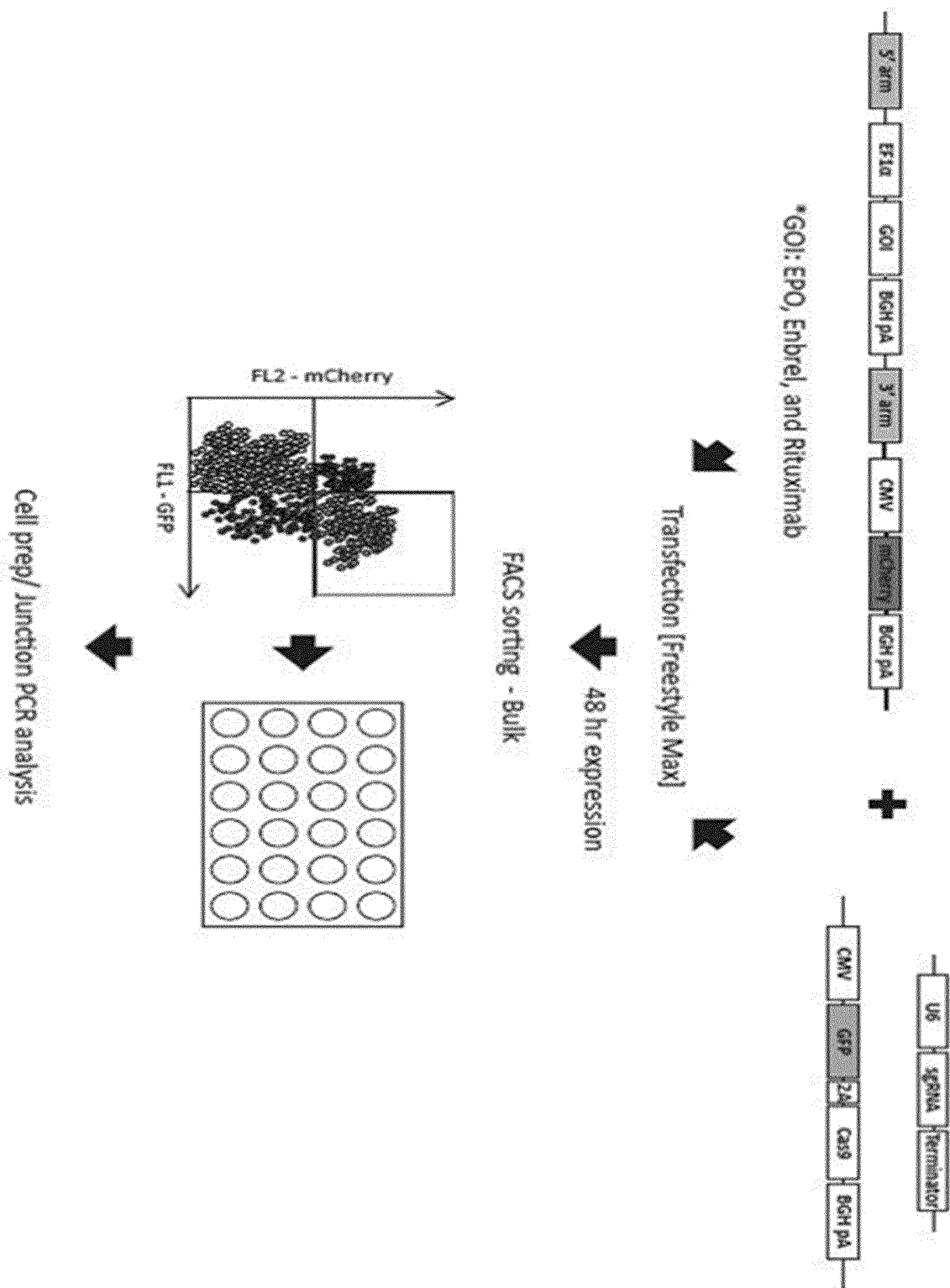
FIG. 21
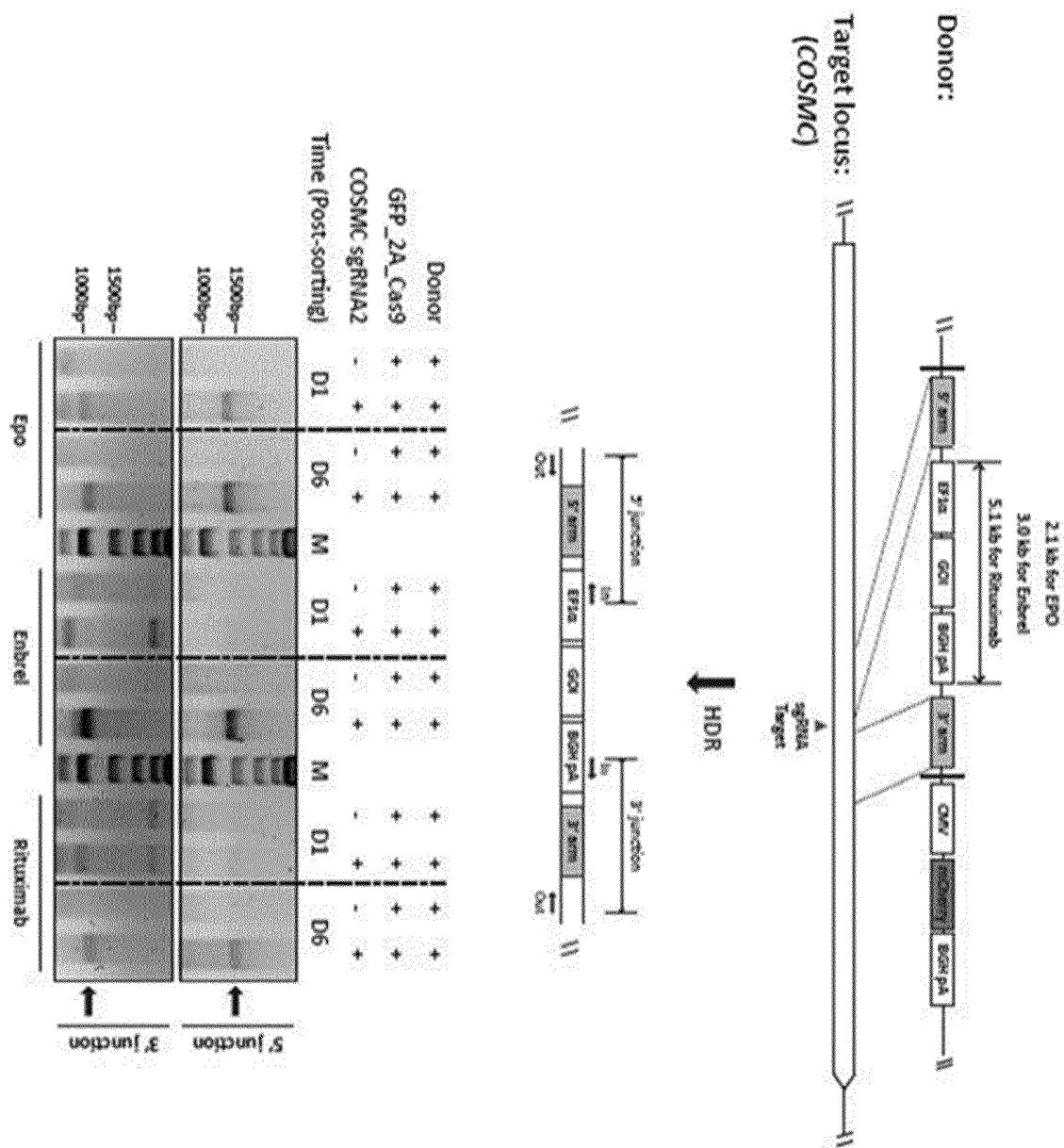
FIG. 22
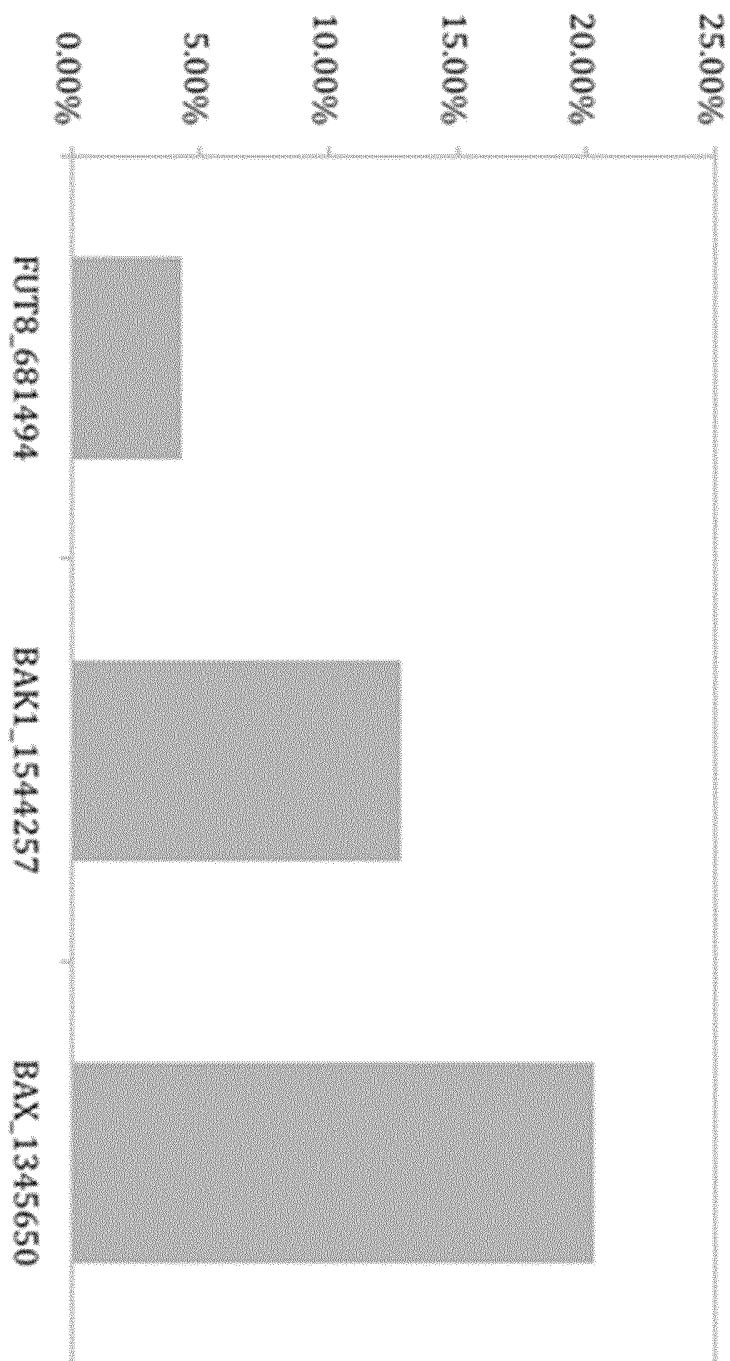
FIG. 23

Statistics: 20140926 transfected_001

Populations		%Total	%Parent
All events	10,000	100.00%	####
P1	4,611	46.11%	46.11%
GFP negative	4,159	41.59%	90.020%
GDP positive	441	4.41%	9.56%
Bottom 50%	229	2.29%	51.93%

FIG. 24

FIG. 25

FIG. 26

HSV TK polyA
adaptor rev

pJT RA4 CMV TO_cas9(with adaptor)(2561 in labcol)
9,944bp

EF1 alfa promoter
CMV promoter

2X teto

adaptor
GFP

2A peptide
Cas9
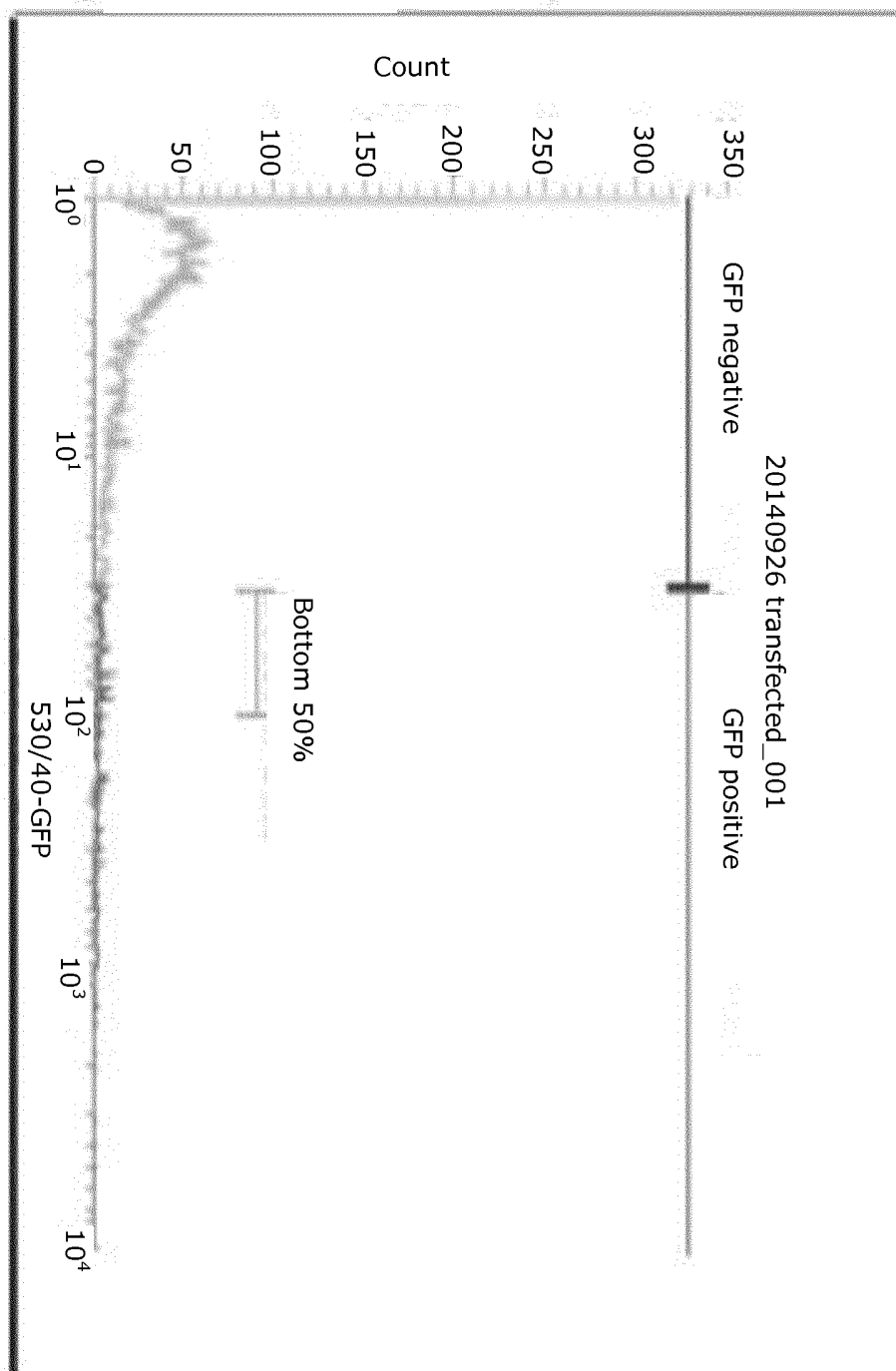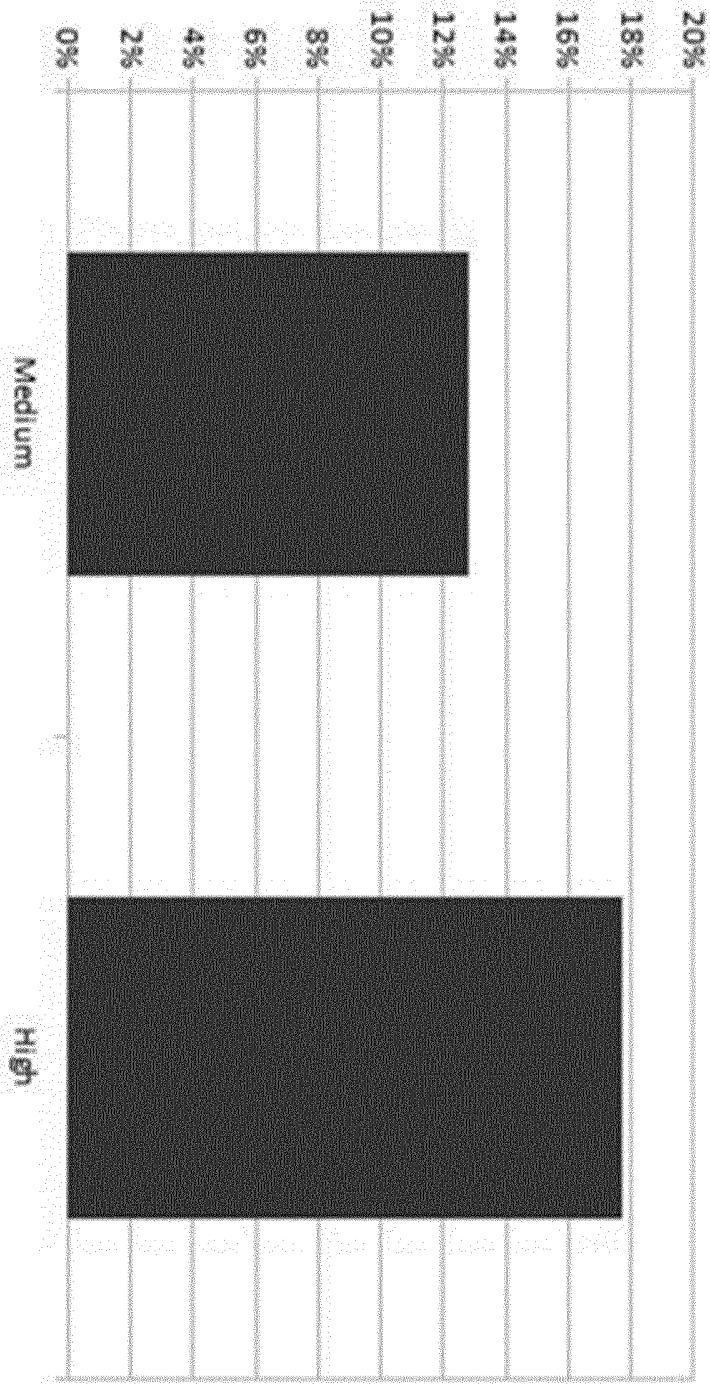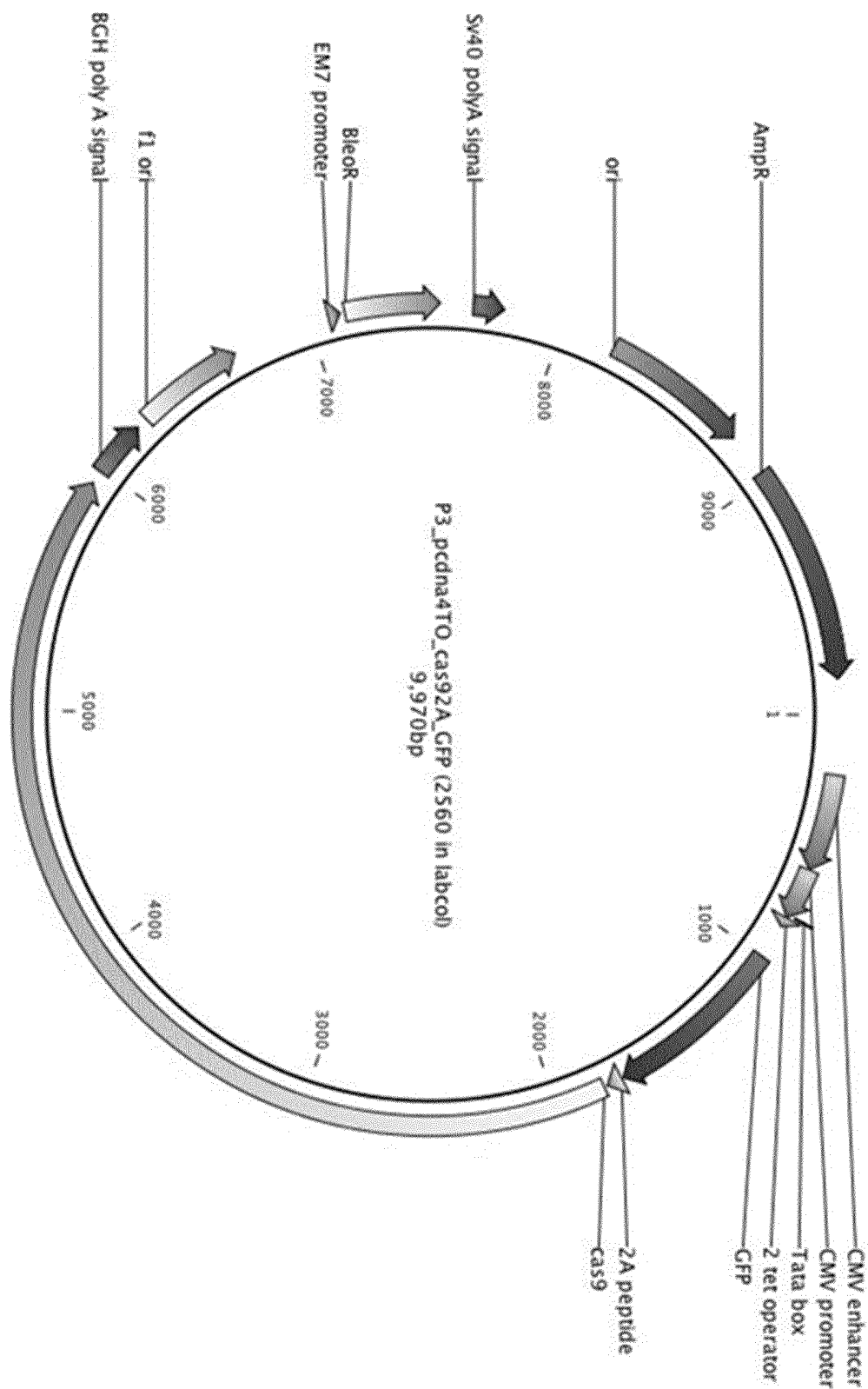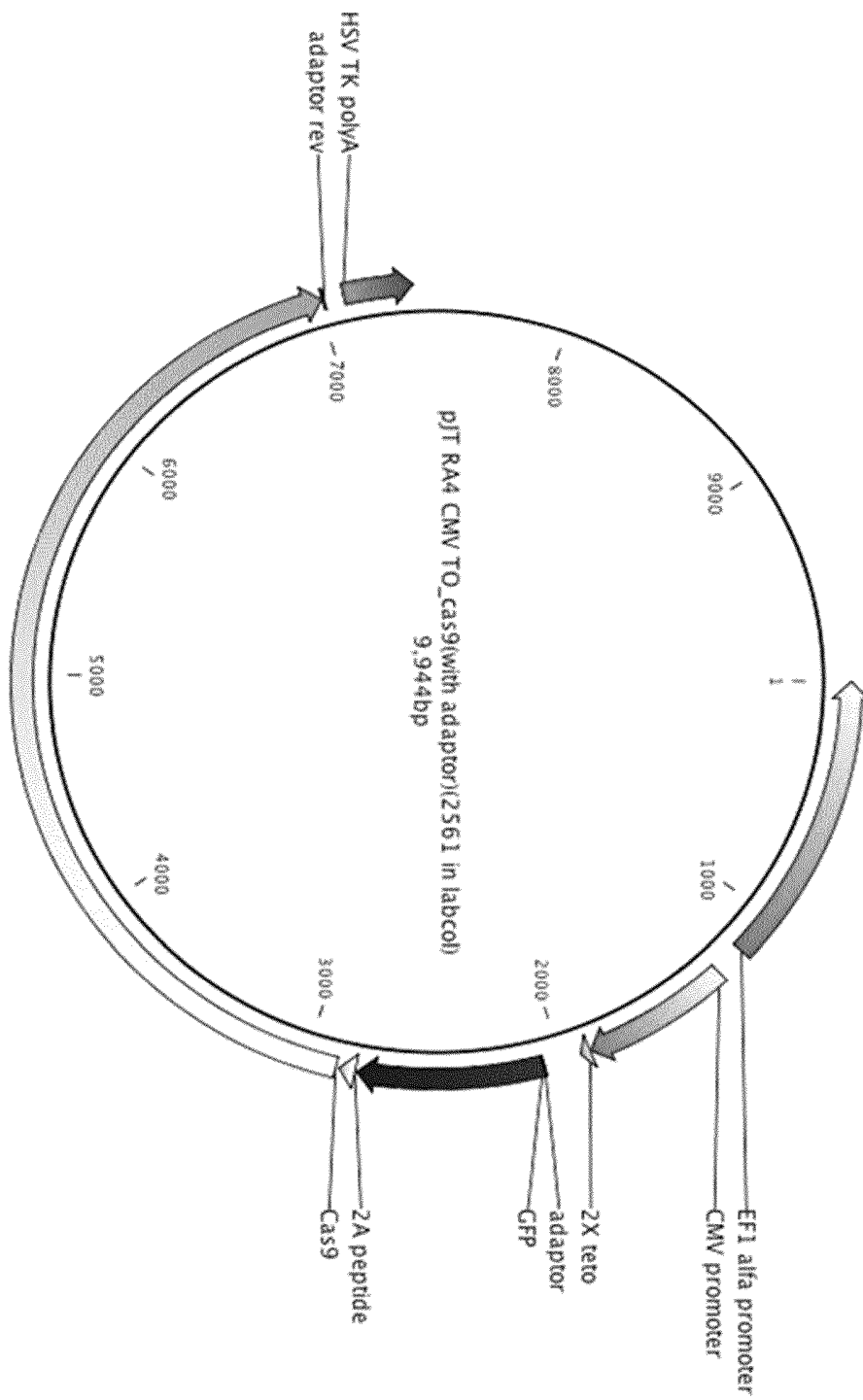
FIG. 27

FIG. 28

FIG. 29

FIG. 30

FIG. 31

FIG. 32

FIG. 33