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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 May 2018 (31.05.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/096356 A1

(51) International Patent Classification:
C12N 15/63 (2006.01) *C12N 15/90* (2006.01)
C12N 15/85 (2006.01)

(21) International Application Number:
PCT/GB2017/053546

(22) International Filing Date:
27 November 2017 (27.11.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1620091.7 28 November 2016 (28.11.2016) GB
1717695.9 27 October 2017 (27.10.2017) GB

(71) Applicant: **HORIZON DISCOVERY LIMITED** [GB/GB]; 8100 Cambridge Research Park, Waterbeach, Cambridge Cambridgeshire CB25 9TL (GB).

(72) Inventors: **BUERCKSTUEMMER, Tilmann**; Horizon Genomics GmbH, Campus Vienna Biocenter 3, 1030 Vienna (AT). **COLLIN, Philippe**; Horizon Discovery Ltd, 8100 Cambridge Research Park, Waterbeach, Cambridge Cambridgeshire CB25 9TL (GB). **SANTOS, Rodrigo**; Horizon Discovery Ltd, 8100 Cambridge Research Park, Waterbeach, Cambridge Cambridgeshire CB25 9TL (GB). **GUZZARDO, Paloma**; Horizon Genomics GmbH, Campus Vienna Biocenter 3, 1030 Vienna (AT). **RASHKOVA, Christina**; Horizon Genomics GmbH, Campus Vienna Biocenter 3, 1030 Vienna (AT).

(74) Agent: **STRATAGEM INTELLECTUAL PROPERTY MANAGEMENT LIMITED**; Meridian Court, Comberton Road, Toft, Cambridge Cambridgeshire CB23 2RY (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

(54) Title: METHODS FOR CONDITIONAL GENE KNOCK-OUT

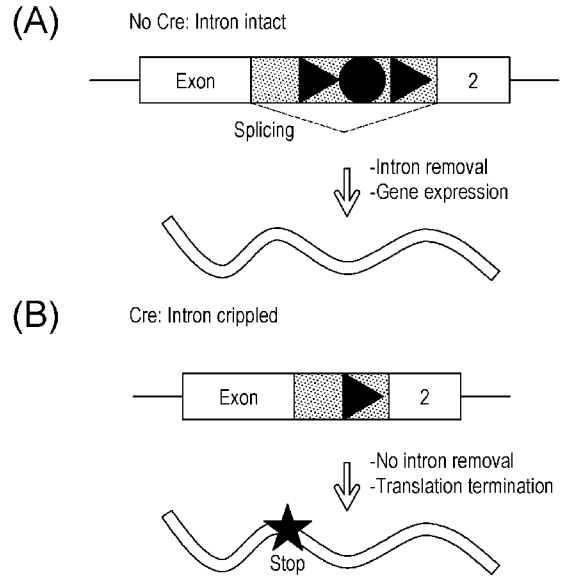


Fig . 5

(57) Abstract: This invention relates to a system and method for generating conditional gene knock-out cells. The system includes the use of an artificial intron to modify a target allele to permit conditional knock-out of the allele by disruption of the artificial intron.In particular, the invention relates to a method for generating an allele for conditional gene knock-out in a cell comprising a target gene, the method comprising: introducing an artificial intron sequence into an exon of the target gene. The invention also relates to cells containing an artificial intron introduced according to the method. The approach has broad utility for conditional gene activation and could be used to study the loss-of-function phenotypes of essential genes.

WO 2018/096356 A1

[Continued on next page]

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

METHODS FOR CONDITIONAL GENE KNOCK-OUT

Field of the Invention

This invention relates to a system and method for generating conditional gene knock-out cells. The system includes the use of an artificial intron to modify a target allele to permit conditional knock-out of the allele by disruption of the artificial intron.

Background of the Invention

Genome editing has been revolutionized by the discovery of the CRISPR/Cas technology that enables the targeted introduction of double strand breaks into the human genome at unprecedented precision and efficiency[1-4]. Once the double-strand break has been established, the cell will repair the break by one of two common pathways: in the presence of a homology template, the cell will use homology-directed repair or homologous recombination[5]. In the absence of homologous sequences, the cell will employ non-homologous end joining to repair the break[6].

As non-homologous end joining is by nature quite imprecise, it triggers the accumulation of small insertions or deletions (so called indels)[6]. If such indels lie within the coding sequence of a gene, they often disrupt the reading frame and thus abrogate the translation (or transcription; as a consequence of nonsense-mediated decay) of the corresponding gene. This has been exploited to create cell lines or whole organisms (e.g. in mice) bearing so called gene knockouts.

Gene knockouts are strains containing a frameshift mutation or, ultimately, a stop codon in coding exons and thus represent complete loss-of-function mutants. They are fundamental for our understanding of gene function as they enable the study of the corresponding loss-of-function phenotype.

Unfortunately, not all human genes can be studied by the means of establishing gene knockouts. Certain human genes cannot be inactivated because inactivation leads to the death of the cell or the whole organism. These genes are called essential genes as they are essential for life[7]. By definition, essential genes are key regulators and it turns out that many of them represent important drug targets as well. As a consequence, there is a large unmet need for the study of essential genes.

Historically, essential genes have been explored using conditional knockouts. These were established by assembling a large homology donor containing both a selection cassette to enrich for the targeting event and loxP sites that enable the Cre-mediated excision of an exon that lies within the gene of interest[8]. The homology donor was then used to establish genetically engineered cells by spontaneous homologous recombination (e.g. in mouse embryonic stem cells) or, more recently, by CRISPR-assisted homologous recombination. Unfortunately, none of these approaches are easily scalable because donor assembly is cumbersome and the overall efficiency of the process is low and hence requires the screening of hundreds if not thousands of clones by Southern blotting or PCR.

With the advent of CRISPR, it became more feasible to tag endogenous genes with drug-regulatable domains that are called degrons. If such a domain is fused to a gene of interest, the resulting gene product can be regulated in its abundance by adding or withdrawing a small molecule. Popular degrons include the auxin/ mAID system[9], the HaloPROTAC3/ HaloTag system[10] and the HCV protease inhibitor/ SMAShTag system[11]. However, the use of degrons has two major limitations: (i) The bi-allelic tagging of genes occurs at very low frequency and thus, it is difficult to obtain a cell line or an organism in which all copies of a gene have been conditionally inactivated and (ii) The insertion of foreign sequences within the coding region of a gene often disturbs protein folding or protein function. As a consequence, degron tagging may be less predictable than desired.

Very recently, an approach referred to as CRISPR-FLIP was introduced (see http://biorxiv.org/content/early/2016/06/01/056549 for details). It is based on an invertible intronic cassette (FLIP), similar to COIN[12]. In the non-mutagenic orientation, the FLIP cassette expresses the puromycin resistance gene (puroR) to select for correct nuclease-assisted targeting into the exon of one allele. Upon exposure to Cre recombinase the FLIP cassette is inverted to a mutagenic configuration that activates a cryptic splice acceptor and polyadenylation signal (pA) and disrupts the initial splice acceptor resulting in the loss of gene function. While this system is attractive, it has several shortcomings: (i) The cassette that is used as a donor is large (>2kb; the exact size not evident from the publication) and hence, the assembly of targeting constructs with homology arms is costly, cumbersome and time-consuming; (ii) The cassette contains an antibiotic resistance gene which is not always desired and (iii) Expression of the gene of interest is disrupted by a splice acceptor whose efficiency

can vary depending on genomic context. Accordingly, there is a need for an alternative conditional gene knock out method.

The present invention is intended to address the disadvantages of the prior art. In particular, the inventors have determined a method which is intended to significantly reduce the size of the artificial intron that is used and ensure that gene expression can be rapidly and efficiently abrogated in a wide range of mammalian genes.

Described herein is a system that overcomes the above mentioned limitations of the prior art and provides an artificial intron that adheres to the following principles: (i) The artificial intron size may be small so that homology donors can be easily assembled by gene synthesis; (ii) the approach does not require the use of antibiotic selection; (iii) the cassette disrupts gene expression completely, i.e. reveals a stop codon upon cassette activation; (iv) the cassette has a good "window of opportunity", so that protein levels are strongly decreased as a consequence of cassette activation; and (v) the approach is scalable and success is predictable for most human genes without too much prior insight into protein function or folding.

Described herein is an artificial intron cassette that may contain as few as about 200 nucleotides. Insertion of this cassette into the coding exon of a gene does not disrupt gene function as the cassette is based on a natural intron sequence and the intron is thus removed by splicing. Addition of Cre recombinase (e.g. by addition of a recombinant adenovirus bearing Cre recombinase) leads to the deletion of critical elements within the intron that are required to maintain splicing, most importantly: the branch site. As a consequence, the cassette will no longer be removed and the ribosome will run into a stop codon, thus disrupting the expression of the gene of interest upon addition of Cre recombinase. The resulting cell line contains a frameshift and thus represents a complete gene knockout.

Described herein is a system that allows the generation of conditional gene knockouts on a large scale with the following features: (i) engineering is straightforward in diploid or even polyploid cells; (ii) the approach does not require any prior insight into the structure/ function of the targeted gene product; (iii) in the "off-state" of the cassette, the activity of the target gene product is largely unaffected; (iv) the cassette activation (transition to "on-state") triggers a profound decrease in gene expression. Based on these requirements, we designed a genetic approach named DECAI (DEgradation via Cre-regulated Artificial Intron), where

conditional gene knockouts can be generated by inserting a small intron cassette into the coding sequence of an exon.

Summary of the Invention

The present inventors have now developed a new conditional gene knock-out system that overcomes the above mentioned limitations. The system uses an artificial intron sequence which, when introduced into the coding exon of a target gene, permits normal gene function because the artificial intron is based on a natural intron sequence and is thus removed by splicing. In order to permit disruption of the target gene, the artificial intron includes one or more stop codons or a series of stop codons (one for each of the three possible reading frames) and an excisable or disruptable intron element. When this intron element is excised, intron function is disrupted and normal splicing of the intron is prevented. The ribosome will then run into the stop codon, thus disrupting the expression of the target gene.

Provided herein is a method for generating an allele for conditional gene knock-out in a cell comprising a target gene, the method comprising: introducing an artificial intron sequence into an exon of the target gene, the artificial intron sequence comprising:

a) a splice donor sequence;
b) a first nuclease or recombinase site;
c) a branch point sequence;
d) a second nuclease or recombinase site;
e) a splice acceptor sequence; and
f) a stop codon positioned 5' to or within the first recombinase or nuclease site.

The method may comprise introducing into the cell a sequence-specific nuclease that cleaves a sequence within the target gene thereby introducing the artificial intron sequence into the target gene.

To inactivate the introduced intron, the method may also include the step of introducing or activating a recombinase or nuclease in the cell thereby excising or disrupting the branch point and abrogating splicing of the artificial intron sequence.

Alternatively, the intron may be disrupted by applying two sequence-specific nucleases (e.g. Cas9 with two guide RNAs targeting two sites flanking the branch point) to excise the branch point and thus abrogate splicing. Alternatively, the guide RNA recognition sites flanking the branch point may also be identical, thus only requiring Cas9 along with a single guide RNA.

In another aspect, artificial introns are provided. The artificial introns may be as described herein and may be used in any of the methods described herein. The sequence of the artificial intron may comprise:

a) a splice donor sequence;
b) a first nuclease or recombinase site;
c) a branch point sequence;
d) a second nuclease or recombinase site;
e) a splice acceptor sequence; and
f) a stop codon positioned 5' to or within the first recombinase or nuclease site.

The sequence of the artificial intron may comprise two stop codons, more preferably three stop codons, which may be positioned in a series of three or more stop codons 5' to the branch point.

Advantageously, the artificial intron may contain a single branch point and may also lack related sequences that can take over branch point function once the single branch point has been removed.

The sequence-specific nuclease used to cleave the target gene may be an RNA guided nuclease, or a DNA guided nuclease, a Zinc finger nuclease or a TALEN. The RNA-guided nuclease may be Cas9 or an RNA-guided Argonaute. The sequence-specific nuclease may be introduced as a protein, mRNA, or cDNA.

The first and second recombinase sites may be loxP sites or FRT sites.
In some embodiments the first (5') loxP site comprises at least one stop codon.

The first recombinase or nuclease site is positioned adjacent to the splice donor site.

Advantageously, the artificial intron sequence may be small, for example, about 200-250, 200-300, 200-400, 200-500, 200-600, 200-700 or 200-800 or 200-1000 or 200 to 1500 nucleotides in length.

In some embodiments the artificial intron sequence comprises a nucleotide sequence at least 80%, 85%, 90%, 95%,96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1 .

In another aspect, the invention provides cells containing any of the artificial introns as described herein. Also provided are cells containing an artificial intron introduced according to any of the methods described herein.

In some embodiments the cells comprise an artificial intron in an exon of a target gene, the artificial intron having at least one stop codon positioned 5' of or within the first recombinase or nuclease site, and wherein the single branch site is flanked by a pair of recombinase or nuclease sites, the recombinase or nuclease sites being arranged such that the single branch site may be excised or disrupted, thereby abrogating splicing of the artificial intron.

In some embodiments provided herein is a cell comprising an artificial intron in an exon of a target gene, the artificial intron having at least one stop codon positioned 5' to or within a first recombinase or nuclease site, and wherein a single branch site is flanked by a pair of recombinase or nuclease sites, the recombinase or nuclease sites being arranged such that the single branch site may be excised or disrupted, thereby abrogating splicing of the artificial intron, and wherein the artificial intron sequence also comprises a splice donor sequence and a splice acceptor sequence.

In some embodiments provided herein is a cell comprising an artificial intron in an exon of a target gene, the artificial intron sequence comprising:

a) a splice donor sequence;
b) a first nuclease or recombinase site;
c) a branch point sequence;
d) a second nuclease or recombinase site;
e) a splice acceptor sequence; and
f) a stop codon positioned 5' to or within the first recombinase or nuclease site.

The cells used in the methods herein, or comprising the artificial intron sequences or nucleic acid constructs described herein, may be eukaryotic cells, including mammalian cells such as haploid cells (e.g. the HAP1 cell-line), human cells, human induced pluripotent cells, human induced pluripotent stem cells, embryonic stem cells, or CHO cells or a T-cell (such as those for use in human therapy).

The cells described herein may be used in therapy.

The cells described herein may be used to generate transgenic animals.

In another aspect, there is provided a nucleic acid construct comprising an artificial intron sequence as described herein and which may be used in any of the methods described herein. In some embodiments, for example where the artificial intron is to be introduced into the target gene by homology directed repair, the nucleic acid construct further comprises 5'

and 3' homology arms comprising sequences homologous to corresponding target gene sequences. Suitably, the nucleic acid constructs described herein may be part of a plasmid.

In another aspect, there is provided a vector comprising the artificial intron sequence or nucleic acid constructs as described herein.

Some embodiments of the methods and cells provided herein are discussed further below.

Generation of conditional Knockout (KO) mice and rats

In the past 20 years, the Cre/loxP system has been used to create a variety of conditional KO mice. In most of these, a coding exon of the target gene was flanked with loxP sites. In the absence of Cre, gene expression is unaffected. Upon addition or activation of Cre, the intervening exon is deleted or inverted, resulting in a gene KO.

Cre could be applied by transient transfection or virus infection. Alternatively, mice could already carry a Cre transgene, often coupled to a promoter that warrants expression in a specific tissue or at a certain stage of development. Mice harbouring Cre are then crossed with strains harbouring alleles flanked with loxP sites. In that way, target gene expression can be disrupted in a tissue or developmental stage-specific manner, depending on which Cre mouse line was used for the intercross.

The approach to introduce artificial introns described herein can be used in a very similar manner, i.e. to create rodents harbouring conditional KO alleles. In a second step, such models can be crossed with rodents strains harbouring Cre, thereby giving rise to rodent models that delete the target gene wherever Cre is activated. This allows researchers to study gene KO phenotypes in specific cell types or developmental stages.

Thus provided herein are non-human transgenic animals whose genome comprises the artificial intron sequence described herein

Establishment of conditional Knockout (KO) cell lines

Many academic researchers wish to study the loss-of-function phenotype of essential genes. Yet, establishing a gene KO in an essential gene is not possible because the resulting cell line is not viable.

In mice, this problem is often solved by creating a heterozygous conditional allele (e.g. by flanking an exon with loxP sites) and backcrossing heterozygous mice to homozygosity. This is unfortunately not possible in cell lines as they do not undergo sexual reproduction and hence, the conditional KO allele cannot be separated from the wild-type allele.

This problem is solved in an elegant way using the approach described herein: By targeting a coding exon with a gRNA and introducing an artificial intron on one allele, the second allele is often disrupted as a consequence of Cas9 cleavage and repair by non-homologous end joining. Hence, the resulting cell line will bear one allele harbouring the artificial intron and a second allele harbouring a frameshift mutation. For many essential genes, such lines will be viable. Hence, this solution alleviates the need for backcrossing and allows conditional KO model generation in one-go.

Cell Therapy

In cell-based therapy, one may encounter a scenario where a gene Knockout(KO) is only desirable in a certain subtype or cells or at a certain developmental stage. In that case, it is feasible to engineer the total population of cells (e.g. all T cells) with the conditional KO cassette described herein. In a second step, Cre is then applied to a specific subset (e.g. the CD4-positive T cells) to inactivate the target gene selectively in this sub-population of cells.

iPSC differentiation

Stem cell differentiation is often heterogenous or incomplete. One way to improve the quality and outcome of the process would be inactivate certain genes at a specific point during the process and thereby steer cell differentiation in a certain direction. For instance, one could introduce a conditional Knockout (KO) allele into a pluripotency factor (e.g. Oct4/POU5F1; as shown in Example 2) and thereby create an iPSC line in which pluripotency can be terminated at will. Then, this line will be subjected to differentiation by applying the appropriate cocktail of factors or by transducing with one or more master regulators as is known in the art.Once cells are starting to differentiate, one activates the conditional Oct4 allele (e.g. by addition of Cre recombinase). As a consequence, cells that have not undergone differentiation will be eliminated from culture and the resulting cell population will be more pure.

Thus in an embodiment of the methods and cells provided herein the cell may be an induced pluripotent stem cell and the target gene is Oct4.

In a similar fashion, the conditional KO approach can be used to create more homogeneous differentiation protocols. Genes that drive and control the maintenance of intermediate states of differentiation (e.g. like pancreatic) could be engineered with the conditional KO cassette at the iPSC stage. Once the differentiation into the final desired cell types (e.g. b-cells) is accomplished, the conditional KO cassette can be activated, leading to the death of progenitor cells that contaminate the culture.

Reduction of tumorigenesis by suicide switches

Uncontrollable proliferation of cells (i.e. tumorigenesis) is one of the most challenging safety concerns of cell and stem cell treatments. One way to tackle this problem is the development of "suicide switches" that can be activated to trigger cell death at any given stage post-transplantation. Another approach consists of the development of "proliferation block switches", whose activation leads to blocking of the proliferation capacity of the transplanted cells. A lack of proliferation is of essence for long term cell grafts, such as the transplant of dopaminergic cells for treatment of Parkinson's disease.

The conditional KO cassette described herein can be used to create a proliferation switch by targeting a gene involved in cell cycle (e.g. Cyclin-dependent Kinases). Blocking of proliferation can be carried out upon cell expansion to the required amounts for transplantation, or after stem cell differentiation into the desired cell type, and can be performed both prior and post-transplantation (different strategies can used for Cre delivery or Cre-activation).

Generation of RNA fusions for reference standards

Certain fusion transcripts (e.g. BCR-ABL) are commonly produced by cancer cells, and detection of these constitutes the diagnostic of certain cancer types. Reference standards are materials that can be used as positive or negative controls during the diagnostic of patient material. The conditional KO cassette described herein can be used to create transcript fusions to be used as reference material by allowing the hiding of a selection cassette (e.g. antibiotic resistance such as puromycin). This feature enables the generation and selection of fusion transcripts composed of an endogenous gene and a transgene, without any scar at the RNA level, as the synthetic intron is removed by slicing.

Hiding an exogenous sequence in the genome

In some instances, one may wish to introduce a sequence into the genome that does not affect gene expression of the resulting cell line. One could obviously target non-coding regions of the genome, but this is risky as one never knows whether a given sequence does or does not have an impact on gene expression. For instance, an element could represent an enhancer that is poorly described or annotated and altering this enhancer could have a profound impact on gene expression.

In another aspect, provided herein is a method which solves this problem: As the intron is actively removed by splicing, it is not present at the mRNA level and does therefore not

impact transcription. Modifying the intron to contain an exogenous sequence (e.g. a molecular barcode) could represent an elegant way of "hiding" this sequence in the genome without leaving a trace in the transcriptome.

Figures

Figure 1: *NanoLuc expression from constructs containing synthetic intron configurations*
HAP1 cells were transfected with a pcDNA3.1 plasmid expressing NanoLuc under the control of a CMV promoter. Cells were lysed 24h post transfection and analysed using the Nano-Glo Luciferase Assay System (Promega) according to manufacturer's instructions. Median luminescence across 24 replicates is shown for each condition. As a control, expression levels from a plasmid containing the NanoLuc gene or the NanoLuc with the chimeric intron with no loxP sites were measured.. For sequence details of the different artificial introns see below SEQ ID NO:2 to SEQ ID NO:11.

Figure 2: *NanoLuc expression from constructs with chimeric intron at a different site*
HAP1 cells were transfected with a pcDNA3.1 plasmid expressing NanoLuc under the control of a CMV promoter. Cells were lysed 24h post transfection and analysed using the Nano-Glo Luciferase Assay System (Promega) according to manufacturer's instructions. Median Luminescence across 24 replicates is shown for each condition. As a control, expression levels from a plasmid containing the NanoLuc gene or the NanoLuc with the artificial intron were measured.

Figure 3: *Genotyping CDK4-intron and METTL16-intron cell lines after Cre recombination*
Cells expressing CDK4-intron or METTL16-intron were infected with a recombinant retrovirus encoding Cre-ERT2. Upon enrichment of transduced cells, Cre was activated by adding 1μM 4-hydroxitamoxifen (4-OHT) for three days. Genomic DNA from CDK4-intron and METTL16-intron cell lines treated with Cre recombinase and/or 4-OHT was used for a PCR amplifying the region around the insertion site of the artificial intron. Genomic DNA from wild type HAP1 cells was used as a control. The expected sized of the PCR products are as follows. For CDK4, HAP1 WT: 210bp, CDK4-intron (before recombination): 411bp, CDK4-intron (after recombination): 264bp. For METTL16, HAP1 WT: 250bp, METTL16-intron (before recombination): 451bp, METTL16-intron (after recombination): 304bp.

***Figure 4**: Cell viability of CDK4-intron and METTL16 intron after Cre recombination*
Cells described in Figure 3 were treated with 4-OHT for six days, fixed with methanol and stained with crystal violet to assess cell viability.

***Figure 5:** Approach for conditional gene inactivation using artificial introns.*
(A) For the purpose of conditional gene inactivation, an artificial intron is introduced into a coding exon of a gene. In the artificial intron (marked in blue), the branch point (in yellow) is flanked by two parallel loxP sites (black triangles). Under steady-state conditions, the intron gets removed by splicing and leaves the mRNA of the target gene intact. (B) Upon recombination with Cre, the branch point is excised and the intron is inactivated. Consequently, the intron is no longer removed and the ribosome reads through the intronic sequence, running into one of three stop codons that were added to the artificial intron cassette. (C) DNA sequence of the artificial intron that was used for most of the experiments presented in Example 2. Important features are highlighted: splice donor in green, loxP sites are in lower case letters, stop codons in three reading frames in red italics, branch point in yellow, polypyrimidine tract is underlined and splice acceptor in blue.

***Figure 6.** Identification of the artificial intron configuration that enables conditional gene* inactivation. (A) HAP1 cells were transfected with NanoLuc without intron ("no intron") or one harbouring the intron without further modification ("intron"). In addition, a set of cassette variants (Var 1-5) were tested in which loxP sites positioning relative to the branch point varied. Note that we included two designs for each cassette, one in which the cassette was intact ("no Cre") and one in which the sequence between the loxP sites had been removed ("Cre"). NanoLuc levels were measured 24h post transfection using the Nano-Glo Dual assay. This data is also shown in Figure 1. (B) Variant 4 identified in Figure 6A was placed into various sequence contexts within the NanoLuc gene, reflecting one of four possible insertion sites: CAG-G, AAG-G, CAG-A and AAG-A. For each of these, two sites were evaluated. HEK293 cells were transiently transfected with these various constructs and NanoLuc luciferase activity was measured using the NanoGlo Dual assay 48h hours after transfection.

***Figure 7.** Cre-recombination leads to branch point excision.* (A) Genomic DNA was isolated from cells transfected with Cre (for CD46) or cells transduced with Cre-ERT2 and treated with 4-hydroxitamoxifen (4-OHT; for METTL16). Samples were genotyped using PCR primers specific for CD46 or METTL16. The right hand side of Figure 7A, regarding METTL16, is also

shown in the right hand side of Figure 3. (B) mRNA was isolated from cells expressing Cre, reverse transcribed using oligo (dT) and analysed with primers specific for the CD46 cDNA.

Figure 8. *Artificial intron cassette activation leads to conditional gene inactivation.* (A and B) HAP1 cells bearing the artificial intron in CD46 were transfected with Cre recombinase. For Western blotting, cells were analysed using a CD46-specific antibody. For FACS, cells were stained with a CD46-specific antibody and analysed by flow cytometry. (C) METTL16-DECAI cells were transduced with Cre-ERT2, treated with 4-OHT as indicated and stained with Crystal Violet. The data in Figure 8C is also shown in Figure 4, right hand side.

Figure 9. *Generation of Oct4-conditional knockout human iPS cells.* (A) Wild-type cells (+/+) or cells bearing the artificial intron in exon 1 of Oct4 (Oct4-DECAI #1 and Oct4-DECAI #2, created with two independent gRNAs) were transfected with Cre-ERT2 recombinase and 4-OHT or left untreated. Cells were fixed in 4% PFA and stained for Oct4 and SSEA4 using the specific antibodies. DAPI was used for nuclear staining. Images were acquired with an Olympus IX83 microscope (10X magnification, 100μm scale bars). For bright field images see Figure 14. (B) Flow cytometry analysis of the cells before (No Cre) and after Cre treatment (+Cre) stained with Oct4 and SSEA4 specific antibodies. (C) qRT-PCR analysis was performed before and after induction of Cre recombination. Expression levels of Oct4, two additional pluripotency markers (Nanog and Sox2), and two differentiation markers (Cdx2 and Eomes) are shown. qRT-PCR values are normalized to GAPDH and shown as relative to the highest value of that target across all samples.

Figure 10.*Homology donor templates for cell line generation.* Homology donor templates used to engineer the CD46-DECAI, METTL16-DECAI and OCT4-DECAI cell lines are shown. The donors contain ~400bp right and left homology arms flanking the 201bp artificial intron cassette (underlined).

Figure 11. *Analysis of CD46 expression in CD46-DECAI and WT Hap1 cells.* HAP1 wild-type and CD46-DECAI cells were transfected with a plasmid expressing Cre recombinase. Cells were then stained with a CD46-specific antibody and analysed by flow cytometry.

Figure 12. *Isolation of Cre-ERT2 expressing METTL16-DECAI clones.* Crystal violet staining of single cell clones isolated from two METTL16-DECAI cell lines (Clone B5 and Clone G4). 12 clones of each cell line were isolated following transduction with Cre-ERT2 and these potentially Cre-ERT2 expressing clones were treated with 4-OHT to trigger cassette activation.

Clones in which 4-OHT triggered cassette activation, and hence, cell death, are marked with a box.

Figure 13. *Cre-recombination leads to a change in morphology of Oct4-DECAI cells.* Bright field microscopy images of WT, Oct4-DECAI#1 and Oct4-DECAI#2 after transfection with Cre-ERT2 and 4-OHT treatment. A clear change in cell morphology is observed only in the Oct4-DECAI#1 and Oct4-DECAI#2 cells.

Figure 14. *Cre-recombination leads to branch point excision in Oct4-DECAI cells.* Genomic DNA was isolated from WT, Oct4-DECAI#1 and Oct4-DECAI#2 cells that had been transfected with Cre-ERT2 and treated with 4-OHT, and from cells not transfected with Cre-ERT2. Samples were genotyped using PCR primers specific to either the cassette insertion site of Oct4-DECAI#1 or Oct4-DECAI#2.

Detailed Description

Introns are non-coding or intervening polynucleotide sequence of varying length, normally present within most eukaryotic genes and which are removed from a newly transcribed mRNA precursor by the process of splicing. In general, the process of splicing requires that the 5' and 3' ends of the intron be correctly cleaved and the resulting ends of the mRNA be accurately joined, such that a mature mRNA having the proper reading frame for protein synthesis is produced. Many splice donor and splice acceptors sites, meaning the sequences immediately surrounding the exon-intron- and intron-exon-boundaries, have been characterized and described and are known to the skilled artisan.

Normally, the sequence of mammalian introns begins with GT and ends with AG with a few minor introns starting with AT and ending with AC.

The artificial introns described herein comprise all the intron elements necessary for normal intron function. However, the artificial introns described herein have been modified by the inclusion of recombinase or nuclease sites so that intron function may be disrupted and gene function knocked-out.

The artificial intron described herein may be based on a natural intronic sequence with the intron elements that are necessary for normal intron function, including (i) a 5’ splice donor, branch point and a 3’ splice acceptor (ii) at least one of these intron elements, typically the branch point, is flanked by a pair of recombinase or nuclease sites; (iii) the 5’end of the

cassette comprises at least one stop codon (preferably two stop codons, more preferable a series of three stop codons; one for each reading frame) which will be in frame with the target gene. As a consequence, treatment with the recombinase or the nuclease will excise or invert the intron element, e.g. branch site, between the recombinase or nuclease sites, abrogate splicing of the intron causing the ribosome to run into a stop codon and terminate translation.

An artificial intron sequence derived from natural mammalian intron sequences may be used, for example based on a well-characterized chimeric intron found in the pCI-neo mammalian expression vector from Promega (Cat. No. E1841). This artificial intron is created by combining the 5' splice donor site from an intron of the human β-globin gene with the branch and 3´splice acceptor site of the intron of an immunoglobulin gene heavy chain variable region (SEQ ID NO: 1).

An artificial intron useful in this invention may be an intron which leads to comparable expression of the target gene of interest when placed in an exon of the target gene. Ideally, following disruption of the inserted intron, by recombinase addition or the like, gene expression should be significantly or even completely abrogated.

The artificial intron can be a) derived from a natural intron but may be modified by nucleotide substitutions, deletions and/or insertions, b) a chimeric intron composed of different intron sequences derived from one or more natural intron sequences of the gene of interest and/or of different genes, c) a de novo designed synthetic intron or d) any combination of the above.

Artificial intron sequences as described herein contain at least three conserved intron elements which are found in natural introns and are essential for splicing: a 5 ' splice site (also known as splice donor site), a branch point (or branch site) and a 3 ' splice site (also known as a splice acceptor) adjacent to a run of pyrimidines called a polypyrimidine tract. Recognition of these sites by the splicing machinery is followed by the excision of an intron.

Splice donor site, splice acceptor site and branch point sequences are well known in the art and any may be utilized in the present invention. These elements can be found, inter alia, in the art or derived from consensus sequences, either empirically by inserting, deleting or substituting nucleotides, or by using software capable of predicting splicing sequences.

Preferred splice donor and splice acceptor sites include the consensus splice donor and splice acceptor sites mentioned below and the sites shown in SEQ ID NO. 1. However, other sequences with sufficient splicing efficiency can be used as well. Efficient splice donor and acceptor sites suitable for this invention can be readily determined using techniques for measuring the efficiency of splicing. Intron splicing efficiency is readily determined by quantifying the spliced transcripts versus the full-length, unspliced transcripts that contain the intron(s), using methods known in the art.

The consensus sequence for a splice donor site consists of the sequence MAG:GTRAGT (with M = C or A, R = A or G and the colon denoting the site of exon-intron boundary = cleavage site) (Ohshima et al, J. Mol. Biol. 195, 247 - 259, 1987).

The consensus sequence for a splice acceptor is composed of a polypyrimidine tract (CnTn) that extends for 10-15 nucleotides followed by NCAG:G (with N = any nucleotide and the colon denoting the site of intron-exon- boundary = cleavage site) (Seif et al, Nucleic Acids Res. 6(10), 3387-98, 1979).

As mentioned above in the description of splice donor and splice acceptor site, the consensus sequences include sequences in the exon as well as the intron. The sequence necessary in the exon is MAG-G (where M = C or A and the hyphen denotes the intron).

The splice sites as described herein include naturally occurring, engineered or synthetic, consensus or cryptic splice sites.

The branch point or branch site is usually located approximately between 10 and 60 nucleotides upstream of the splice acceptor and forms, during the splicing process via its conserved adenosine residue, a lariat structure with the splice donor. The consensus sequence for the branch point is reported as YNYYRAY (with Y = T or C, R = A or G, N = any base; the conserved adenosine involved in the lariat formation is underlined). In mammalian genomes one preferred branch point sequence is TACTAAC (Zhuang et al, PNAS 86, 2752-2756, 1989). The branch point may be adjacent to or within the polypyrimidine tract of the splice acceptor site.

Preferably, the artificial intron used in the constructs and methods of the invention has one or more stop codons in all possible 3 reading frames and/or has a nucleotide sequence length which is not dividable by 3 to prevent a complete read through of the intron sequence in case of a non-splicing event. The artificial intron sequence may include stop codons present in the original intron sequence but may also include additional stop codon(s) positioned between the splice donor site and a first recombinase site. The 5'end of the artificial intron preferably contains at least one stop codon, or two or three stop codons, that will be in frame with the target gene following successful recombination (and intron destruction). As a consequence, treatment with the recombinase or the nuclease will destroy the intron and thus the ribosome will run into a stop codon and terminate translation.

Cre and Flp are the recombinases that are most commonly used in molecular biology. Cre can recognize LoxP sites (or variants thereof); Flp recognizes FRT sites (or variants thereof). If a given site of interest is flanked by two parallel recombinase sites, then the addition of the cognate recombinase will trigger the excision of the intervening sequence with high efficiency and precision.

In terms of designing the artificial intron of the present invention, the key challenge was to add the recombinase sites (e.g. loxP sites as shown in the Examples herein) in a way that did not interfere with splicing. At the same time, the recombinase sites or nuclease sites should be positioned to flank an intron element, e.g. branch site, that is of critical importance to maintain splicing, such that when this intron element is excised or disrupted, by recombinase or nuclease addition, intron function is disrupted. The latter was particularly difficult as the branch point sequence, which was the prime candidate sequence for excision or disruption, is not very well defined and the sequence requirements are quite degenerate. As a consequence, deleting one branch point sequence might lead to an adjacent sequence taking over branch point activity. Hence, it was important to empirically define a cassette architecture in which splicing is maintained in the presence of the loxP sites and splicing is completely abrogated upon deletion of sequences that lie between the loxP sites (so that no other sequences can take over branch point function).

The first loxP site of the pair flanking the branch site, which, typically, remains in the target gene following recombinase treatment, may be placed close to the 5' splice donor site, for example.

The first lox P site may comprise one or more stop codons. Variants of loxP sites are known to the person skilled in the art, for example, as shown here: https://en.wikipedia.org/wiki/Cre-Lox_recombination.

Many of the loxP site variants that have been described contain one or more stop codons.

The 5' loxP site may be placed in close proximity to the splice donor. The 3' loxP site may also be placed in close proximity to the splice acceptor whilst retaining a functional polypyrimidine tract. Functionality can be assessed by common methods to measure splicing activity. The branch site should be contained between the two loxP sites.

If the loxP sites are in parallel orientation, the sequence between the two sites is excised.

The present invention provides an artificial intron that may, advantageously, be small in size, for example about 200-250, 200-300, 200-400, 200-500, 200-600, 200-700 or 200-800 or 200-1000 nucleotides in length. Insertion of this cassette into the coding exon of a gene does not significantly disrupt gene function as the cassette is based on natural intron sequences and is thus removed by splicing.

As mentioned above in the description of splice donor and splice acceptor site, the consensus sequences include sequences in the exon as well as the intron. The sequence necessary in the exon is MAG-G (where the hyphen denotes the intron).

Gene tagging can be conducted based on two concurring DNA repair mechanisms: Non-homologous end joining or homology-direct repair. In the case of homology-directed repair, a donor nucleic acid construct is used that includes, in addition to the artificial intron sequence, homology arms of 100bp-5kb length. Importantly, both the 5' and the 3'end of the artificial intron needs to be flanked with homology arms. If such a nucleic acid construct is applied together with a nuclease, such as Cas9, then the artificial intron can be integrated into the exon of a target gene at reasonable efficiencies.

Similar results can be obtained with a donor nucleic acid construct that includes gRNA recognition sites[13] and another gRNA that cuts the target gene at the site where intron integration is meant to occur. While HDR is more precise and more efficient, it does require custom donor nucleic acid construct assembly (due to the homology arms). The NHEJ

approach, in contrast, can be applied across a variety of genomic loci and does not require any adaptation.

In either of the two scenarios, the resulting cell line will only represent a conditional knockout if all alleles in that cell are tagged. This is usually difficult as the efficiency of bi-allelic gene tagging is very low. This represents a limitation for the approach presented here, but even more so for competing approaches such as the degron tagging approach. Consequently, competing approaches have not been applied at scale.

We envisage that the method presented here offers an attractive compromise to solve this problem, i.e. by targeting one allele with the conditional KO cassette and destroying the second allele with CRISPR/Cas cleavage, followed by the introduction of a frameshift mutation through erroneous NHEJ. As a consequence, one would end up with one allele containing a frameshift mutation and the other one containing a conditional allele that could be inactivated upon recombinase addition.

Such alleles would be usually referred to as “hypomorphic alleles” because they only retain gene function partially. Large scale mutagenesis in yeast has suggested that many essential genes tolerate such hypomorphic alleles.

Recombinase treatment/delivery

Recombinases can be applied in many ways that are known to the skilled person. For instance, one could transiently transfect cells using lipofection or electroporation with an expression plasmid encoding the recombinase. Second, one can infect cells with a recombinant virus (retrovirus, lentivirus, adenovirus are most commonly used) encoding the recombinase. Third, one could create a cell line that stably and inducibly expresses the recombinase. Fourth, one could envisage applying the purified recombinase protein, either by itself (following electroporation) or as fusion with a cell-penetrating peptide (e.g. the arginine-rich motif of the HIV-1 TAT).

Guide RNA

Where the term “guide RNA” or “gRNA” is used herein, a person skilled in the art would understand this could mean synthetic gRNA, gRNA obtained by in vitro transcription or gRNA

expressed in cells from plasmid or PCR product. In some instances herein, "T" is used in place of "U" in guide RNA sequences, as would be readily apparent to the skilled person.

Various further aspects and embodiments of the present invention will be apparent to those skilled in the art in view of the present disclosure.

All documents mentioned in this specification are incorporated herein by reference in their entirety.

"and/or" where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. For example "A and/or B" is to be taken as specific disclosure of each of (i) A, (ii) B and (iii) A and B, just as if each is set out individually herein.

Unless context dictates otherwise, the descriptions and definitions of the features set out above are not limited to any particular aspect or embodiment of the invention and apply equally to all aspects and embodiments which are described.

Certain aspects and embodiments of the invention will now be illustrated by way of example and with reference to the figures and sequences described herein.

Examples

EXAMPLE 1

The sequence of the intron cassette used in the examples herein is based on a well-characterized chimeric intron found in the pCI-neo mammalian expression vector from Promega (Cat. No. E1841). This chimeric intron is created by combining the 5' splice donor site from an intron of the human β-globin gene with the branch and 3´splice acceptor site of the intron of an immunoglobulin gene heavy chain variable region (SEQ ID NO:1).

Chimeric intron sequence (SEQ ID NO:1)

The chimeric intron is shown with important splicing motifs highlighted. 5' and 3' splice sites are shown in bold, the branch site is double underlined and the polypyrimidine tract is underlined.

SEQ ID NO:1

GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGTCGAGACAGAGAA GACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCA**CAG**

LoxP sites are 34 base pair sequences that are recognized by Cre recombinase to trigger a specific recombination event depending on their location and orientation. If the loxP sites are in parallel orientation, the sequence between the two sites is excised. To enable deletion of the branch site, two loxP sites were added to the cassette in parallel orientation flanking the branch site. Cre recombination will cause excision of the branch site and thus abrogate removal of the intron by the splicing machinery. As a consequence, the ribosome will read into the remaining cassette and encounter a premature stop codon that disrupts the gene of interest. In order to ensure that a stop codon would be encountered in all three reading frames, we mutated a single nucleotide near the beginning of the chimeric intron. To be certain that efficient splicing of the intron was maintained with these modifications, we created five different versions of the cassette with the loxP sites at varying locations (SEQ ID NOs: 2,4,6,8 and 10). To simulate the maximal impact that Cre recombination could have on these cassettes, we also included the respective derivatives that would arise following removal of the sequence between the loxP site (SEQ ID NOs: 3,4,7,9 and 11). It was important to find a configuration where the addition of the loxP sites did not alter splicing of the intron but would result in a loss of target gene expression upon loxP site excision.

Five versions of the synthetic intron with loxP sites were constructed, the sequences and the corresponding expected sequence after Cre treatment are shown below with the important features highlighted.

5' and 3' splice sites are shown in bold, the branch site is double underlined and the polypyrimidine tract is underlined. loxP sites are in lower case letters and the stop codons in all three reading frames are shown in italics. The single nucleotide that was mutated to create a stop codon is in brackets. For each version, the expected sequence after Cre treatment and loxP site excision is also shown.

Version 1: Artificial intron with loxP sites (SEQ ID NO:2)

G*TAAG*TATCAAGGT*TA[G]*AAGACAGGTT*TAA*GGAGACCAATAGAAACTGGGCTTataacttcgtatagcata cattatacgaagttatGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACataacttcgtatagcatacattat acgaagttatCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCA**CAG**

Version 1: Expected sequence after Cre treatment (SEQ ID NO:3)

G*TAAG*TATCAAGGT*TA[G]*AAGACAGGTT*TAA***GGAGACCAATAGAAACTGGGCTTataacttcgtatagcata cattatacgaagttatCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCA**CAG**

Version 2: Synthetic intron with loxP sites (SEQ ID NO:4)

G*TAAG*TAa*taacttcgta*tag*catacattatacgaagttatTCAAGGTTA[G]AAGACAGGTTTAAGGAGACCAATA GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACataacttcgtatagcatacattat acgaagttatCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCA**CAG**

Version 2: Expected sequence after Cre treatment (SEQ ID NO:5)

G*TAAG*TAa*taacttcgta*tag*catacattatacgaagttatCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTC CA**CAG**

Version 3: Synthetic intron with loxP sites (SEQ ID NO:6)

G*TAAG*TAa*taacttcgta*tag*catacattatacgaagttatTCAAGGTTA[G]AAGACAGGTTTAAGGAGACCAATA GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACAT CCACTTTGCCTTTCTCTCCAataacttcgtatagcatacattatacgaagttat**CAG**

Version 3: Expected sequence after Cre treatment (SEQ ID NO:7)

G*TAAG*TAa*taacttcgta*tag*catacattatacgaagttat**CAG**

Version 4: Synthetic intron with loxP sites (SEQ ID NO:8)

G*TAAG*TAa*taacttcgta*tag*catacattatacgaagttatTCAAGGTTA[G]AAGACAGGTTTAAGGAGACCAATA GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACAT CCACTTTGCCataacttcgtatagcatacattatacgaagttatTTTCTCTCCA**CAG**

Version 4: Expected sequence after Cre-treatment (SEQ ID NO:9)

G*TAAG*TAa*taacttcgta*tag*catacattatacgaagttatTTTCTCTCCA**CAG**

Version 5: Synthetic intron with loxP sites (SEQ ID NO:10)

G*TAAG*TAa*taacttcgta*tag*catacattatacgaagttatTCAAGGTTA[G]AAGACAGGTTTAAGGAGACCAATA GAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACAa taacttcgtatagcatacattatacgaagttatTCCACTTTGCCTTTCTCTCCA**CAG**

Version 5: Expected sequence after Cre treatment (SEQ ID NO:11)

G*TAAG*TAataacttcgta*tagc*atacattatacgaagttatTCCACTTTGCCTTTCTCTCCA**CAG**

As a proof of concept, we tested the effect of the synthetic intron and modified synthetic introns (with loxP sites) on the expression of the NanoLuc Luciferase gene. To this end, we inserted the various cassettes mentioned above into the coding sequence of the NanoLuc gene. When selecting the site to insert the artificial intron, we ensured that it conformed to the exon sequence requirements (MAG-G; M = C or A and hyphen denotes intron) to ensure efficient splicing. To ensure that deletion of the loxP sequence would result in loss of expression, we also cloned expression plasmids containing the NanoLuc gene with the expected sequence of the artificial introns after Cre treatment (shown above in SEQ ID NOs: 3,5,7, 9 and 11). The ideal cassette should not alter expression of the NanoLuc reporter, but should abolish expression after Cre treatment due to the premature stop codons encoded in the sequence.

NanoLuc sequence with intron insertion sites marked (SEQ ID NO: 12)

NanoLuc sequence is shown below. For the experiments shown in Figure 1 the synthetic intron was inserted in the underlined site between the CAG and G. For the experiments shown in Figure 2, the intron was inserted at the double underlines site between the AAG and G.

ATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAGT
CCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGA
TTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGATATCCATGTCATCATCCCGTATGAAGGTCTGAGC
GGCGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAA
GGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATTTCGGAC
GGCCGTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGG
CAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACG
GAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGTAA

We electroporated HAP1 cells with the NanoLuc expression plasmids and assessed luciferase levels after 24 hours using the NanoGlo Luciferase Assay system (Promega) (Figure 1). Comparing NanoLuc and NanoLuc + intron conditions demonstrates that the presence of the chimeric intron in the NanoLuc gene does not have a profound effect on NanoLuc expression. Importantly, loss of NanoLuc expression in one of the 5 constellations, version 3, highlights

that the position of the loxP sites is not trivial and insertion of loxP sites can completely abrogate splicing and thus expression of NanoLuc. Overall, construct number 4 (SEQ ID NO:8 and SEQ ID NO:9) of the synthetic intron is the configuration that had no effect on NanoLuc expression with the cassette, but caused a profound ~30 fold reduction with the post-Cre configuration (compare NanoLuc + intron loxP v4 and NanoLuc + intron loxP post-Cre v4 conditions in Figure 1). Given these results, we decided to proceed with the version 4 cassette (SEQ ID NO:8) for all further experiments.

To rule out positional effects on splicing of the synthetic intron, we inserted version 4 cassette into a secondary location with the NanoLuc gene (See *SEQ ID NO: 12* for details). The only requirement that had to be adhered to when selecting an insertion site was to have the necessary exonic splicing motifs present (the MAG-G sequence consensus).

The NanoLuc expression plasmids with the version 4 cassette inserted at the new site were electroporated in HAP1 cells and NanoLuc expression was assayed after 24 hours (Figure 2).

Insertion into the alternative site had no effect on expression of NanoLuc, but decreased NanoLuc expression 70 fold upon induction, indicating that synthetic intron function is not sensitive to the sequence context and works robustly at different sites in the NanoLuc gene.

To expand on these results, we set out to test this strategy on endogenous genes in HAP1 cells. As test cases, we selected one essential gene, METTL16, and one non-essential gene, CDK4[7]. Essentiality for all genes in HAP1 cells has been determined experimentally using extensive gene-trapping followed by next generation sequencing analysis[7]. By comparing the number of disruptive versus non-disruptive gene-trap integrations in each gene, the authors were able to confidently predict which genes are essential for viability in HAP1 cells. The artificial intron was inserted into the genes by homology directed repair. To this end, HAP1 cells were transfected with Cas9, a gRNA targeting the locus where the synthetic intron would be inserted and a homology donor (Table 1, SEQ ID NO:13 and SEQ ID NO:14). The site where the artificial intron would be inserted was selected following the exonic sequence requirements for splicing (SEQ ID NO:19 and SEQ ID NO:20). The homology donor comprised the synthetic intron and 400bp homology arms upstream and downstream the insertion site (SEQ ID NO:21, SEQ ID NO:22). Single cell clones containing the synthetic intron were isolated,

and genotyped by PCR and Sanger sequencing (primers: Table 1, SEQ ID NO:15 to SEQ ID NO:18 and Sanger sequencing: SEQ ID NO:23, SEQ ID NO:24).

Table 1: Guide RNA and primer sequences

Name	Sequence
CDK4 guide RNA	CTGATGCGCCAGTTTCTAAG (SEQ ID NO:13)
METTL16 guide RNA	TCTGACGTGTACTCTCCTAA (SEQ ID NO:14)
CDK4_genotyping_primer_fwd (used for sequencing)	TGAGAGGTGGATTGGGACCT (SEQ ID NO:15)
CDK4_genotyping _primer _rev	CCAGCTTGACTGTTCCACCA (SEQ ID NO:16)
METTL16_genotyping _primer _fwd (used for sequencing)	TTCTGCCTGTTTGCCGTAGA (SEQ ID NO:17)
METTL16_genotyping _primer _rev	TTGTCAGAATCCTGGTGACCG (SEQ ID NO:18)

Targeted genomic region with guide RNAs

Genomic region surrounding the artificial intron insertion site is shown below. Exonic sequence is shown in lower case letters and intronic sequence in uppercase letter. The exonic splicing sequence motif is shown in bold. The guide RNA used during the cell line engineering is underlined.

CDK4 (SEQ ID NO:19)

AGGTGGGGTGTGATGATCTGTAGAGAAGTGGGGACCCTGAGGAAATAATGAGAGGCCATGTTGGGT
TAAAGGGGATTGAAAAGTGAGCATTTACTCTGGTCAGgctgatggacgtctgtgccacatcccgaactgaccgggag
atcaaggtaaccctggtgtttgagcatgtagaccaggacctaaggacatatctggacaaggcacccccaccaggcttgccagccgaaa
cgatcaagGTGAGTGGGGTTGGTAGGCATTGAGAGGTGGATTGGGACCTTTGTAGTAGAACCTTCTGGG
ATTTCAGGTATGGTGCCTAGTTTCCAGTGCATCTGTACCTCCCCTTTGAAACTAGgatctgatgcgccagtttct
aagaggcctagatttccttcatgccaattgcatcgttcaccgagatctgaagccagagaacattctggtgacaagtggtggaacagtca
agctggctgactttggcctggccagaatctacagctaccagatggcacttacacccgtgGTCAGTAGAAAGATGGTACCAAA
ATGGGTTCTGGTTGGGAGTAGGAGAGTGATTGCCCGTAGCAATTGAGAAGTCATGTGCTTCATGTGTT
CAGTCAAGCAAGTTGTGTTTCATGGTAACCCATGGGGTCCCCATCCATTCTTCCTATTCCCTTTAGgttgtt
acactctggtaccgagctcccgaagttcttctgcagtccacatatgcaacacctgtggacatgtggagtgttggctgtatctttgcagag
atgtttcgtcgaaaGTATGGGACCCACATACCCTGGACTACCTTGAATTCCCCAAATCGCTTGTTCATAAACC
ACATCCATACCT

METTL16 (SEQ ID NO:20)

ACTTGTTCAAAGTCACACAGTGAGGAGACGACGAGACAGGAATTTGAGCCGAGTCATTTAATCTCTGT
GCTCCATATACGGATATGCAGTGTTTTTGGTGGAGGATTGTACTGATTTTCATAAGCTGTGACTGTGTG
AAACACTCGTGTATGTGCCACTGGGATGAGTGCTCTGTGTAACTGTGTGTGATCTCCCTGGGAGGTTC
GGATGGGTATTTGGTTCTGCCTGTTTGCCGTAGAGCTTGTACAACATAGTGAATCTTACTAATCAGTTTT
ACTTCCTTTTCTTTATTCCTCAGccttaattttaaagaccccgaagcagtcagagctctgacgtgtactctcct**aagg**gaagatt
ttggactttctattgatattccattggagagactaattcccacagttcccttgagactcaactatattcactgggtagaagatctgatcggt
caccaggattctgacaaaagtactctccgaagaggaattgacatagGTATATCATTTTAAATTCTTTTTGGCTAAACAG
TTTTCATAAGTTTTGCGAGATCAAATTCTTTGTAAACTTTTCTACTGGGTATTTGTTGTTGTACAAAAGA
AGCTGGCTACAGATACAGGTTGAGCATCCTTAATCCGAAAATCCGAAATGCTTCAAAATCTGCAACCTT

TGAGCACTGATGTGTTGCTCAAAGGCAATGTGCGTTGGAGCGTTCCAGATTTGGGGGTTAGGGATGTT
CAGCCTGTAAATATAGTGCACATATTTCAAAATCTGAAAAAAAAAATTGAAATCCAAAACAC

Donor templates

The sequence used as a donor template with ~400bp homology arms flanking the artificial intron is shown below. For the artificial intron: 5' and 3' splice sites are shown in bold, the branch site is double underlined and the polypyrimidine tract is underlined. LoxP sites are in lower case letters and the stop codons in all three reading frames are shown in italics. The single nucleotide that was mutated to create a stop codon is in brackets.

CDK4 (SEQ ID NO:21)

AGGTGGGGTGTGATGATCTGTAGAGAAGTGGGGACCCTGAGGAAATAATGAGAGGCCATGTTGGGT
TAAAGGGGATTGAAAAGTGAGCATTTACTCTGGTCAGgctgatggacgtctgtgccacatcccgaactgaccgggag
atcaaggtaaccctggtgtttgagcatgtagaccaggacctaaggacatatctggacaaggcacccccaccaggcttgccagccgaaa
cgatcaagGTGAGTGGGGTTGGTAGGCATTGAGAGGTGGATTGGGACCTTTGTAGTAGAACCTTCTGGG
ATTTCAGGTATGGTGCCTAGTTTCCAGTGCATCTGTACCTCCCCTTTGAAACTAGgatctgatgcgccagtttct
aag**G*TAAG***TAa*taacttcgta*tagc*atacattatacgaagttatTCAAGGTTA[G]AAGACAGGTTTAAGGAGACCA
ATAGAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTG
ACATCCACTTTGCCataacttcgtatagcatacattatacgaagttatTTTCTCTCCA**CAG**aggcctagatttccttcatgcca
attgcatcgttcaccgagatctgaagccagagaacattctggtgacaagtggtggaacagtcaagctggctgactttggcctggccaga
atctacagctaccagatggcacttacacccgtgGTCAGTAGAAAGATGGTACCAAAATGGGTTCTGGTTGGGAGT
AGGAGAGTGATTGCCCGTAGCAATTGAGAAGTCATGTGCTTCATGTGTTCAGTCAAGCAAGTTGTGTT
TCATGGTAACCCATGGGGTCCCCATCCATTCTTCCTATTCCCTTTAGgttgttacactctggtaccgagctcccgaa
gttcttctgcagtccacatatgcaacacctgtggacatgtggagtgttggctgtatctttgcagagatgtttcgtcgaaaGTATGGGA
CCCACATACCCTGGACTACCTTGAATTCCCCAAATCGCTTGTTCATAAACCACATCCATACCT

METTL16 (SEQ ID NO:22)

ACTTGTTCAAAGTCACACAGTGAGGAGACGACGAGACAGGAATTTGAGCCGAGTCATTTAATCTCTGT
GCTCCATATACGGATATGCAGTGTTTTTGGTGGAGGATTGTACTGATTTTCATAAGCTGTGACTGTGTG
AAACACTCGTGTATGTGCCACTGGGATGAGTGCTCTGTGTAACTGTGTGTGATCTCCCTGGGAGGTTC
GGATGGGTATTTGGTTCTGCCTGTTTGCCGTAGAGCTTGTACAACATAGTGAATCTTACTAATCAGTTTT
ACTTCCTTTTCTTTATTCCTCAGccttaattttaaagaccccgaagcagtcagagctctgacgtgtactctcctaag**G*TAAG***T
*Aa*taacttcgta*tagc*atacattatacgaagttatTCAAGGTTA[G]AAGACAGGTTTAAGGAGACCAATAGAAACT
GGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTT
TGCCataacttcgtatagcatacattatacgaagttatTTTCTCTCCA**CAG**ggaagattttggactttctattgatattccattgga
gagactaattcccacagttcccttgagactcaactatattcactgggtagaagatctgatcggtcaccaggattctgacaaaagtactct
ccgaagaggaattgacatagGTATATCATTTTAAATTCTTTTTTGGCTAAACAGTTTTCATAAGTTTTGCGAGA
TCAAATTCTTTGTAAACTTTTCTACTGGGTATTTGTTGTTGTACAAAAGAAGCTGGCTACAGATACAGGT
TGAGCATCCTTAATCCGAAAATCCGAAATGCTTCAAAATCTGCAACCTTTGAGCACTGATGTGTTGCTC
AAAGGCAATGTGCGTTGGAGCGTTCCAGATTTGGGGGTTAGGGATGTTCAGCCTGTAAATATAGTGCA
CATATTTCAAAATCTGAAAAAAAAAATTGAAATCCAAAACAC

Clonal cell line sequencing results

Sanger sequencing results from genotyping PCRs of CDK4-intron and METTL16-intron are shown below

METTL16-intron (SEQ ID NO:23)

TCTTTATTCCTCAGCCTTAATTTTAAAGACCCCGAAGCAGTCAGAGCTCTGACGTGTACTCTCCTAAGGT
AAGTAATAACTTCGTATAGCATACATTATACGAAGTTATTCAAGGTTAGAAGACAGGTTTAAGGAGAC
CAATAGAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACT
GACATCCACTTTGCCATAACTTCGTATAGCATACATTATACGAAGTTATTTTCTCTCCACAGGGAAGATT
TTGGACTTTCTATTGATATTCCATTGGAGAGACTAATTCCCACAGTTCCCTTGAGACTCAACTATATTCA
CTGGGTAGAAGATCTGATCGGTCACCAGGATTCTGACAAA

CDK4-intron (SEQ ID NO:24)

GGTGCCTAGTTTCCAGTGCATCTGTACCTCCCCTTTGAAACTAGGATCTGATGCGCCAGTTTCTAAGGT
AAGTAATAACTTCGTATAGCATACATTATACGAAGTTATTCAAGGTTAGAAGACAGGTTTAAGGAGAC
CAATAGAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACT
GACATCCACTTTGCCATAACTTCGTATAGCATACATTATACGAAGTTATTTTCTCTCCACAGAGGCCTAG
ATTTCCTTCATGCCAATTGCATCGTTCACCGAGATCTGAAGCCAGAGAACATTCTGGTGACAAGTGGTG
GAACAGTCAAGCTG

The METLL16 and CDK4 clones containing the artificial intron (from now on referred to as METTL16-intron and CDK4-intron), were transduced with a retrovirus expressing Cre-ERT2 to induce recombination. Cre-ERT2 encodes a Cre recombinase fused to a mutant estrogen ligand-binding domain (ERT2) that requires tamoxifen or 4-Hydroxytamoxifen (4-OHT) for activity[14]. Six days after transduction, both cell lines were treated with 4-OHT to induce Cre recombination. After three days of 4-OHT treatment, genomic DNA was isolated to verify presence or absence of the artificial intron by amplifying the insertion site by PCR (Figure 3). As expected, Cre recombination and addition of 4-OHT triggered the deletion of the sequence between the two loxP sites (the branch site) in both METTL16-intron and CDK4-intron cells.

Genomic DNA from CDK4-intron and METTL16-intron cell lines treated with Cre and/or 4-OHT was used for a PCR amplifying the region around the insertion site of the artificial intron. The expected sized of the PCR products are as follows. For CDK4-intron (before recombination): 411bp, CDK4-intron (after recombination): 264bp. For METTL16-intron (before recombination): 451bp, METTL16-intron (after recombination): 304bp. The PCR fragment size corresponding to the loci after Cre recombination was only observed when cells were transduced with Cre and treated with 4-OHT. Based on this observation, it could be concluded that Cre recombinase faithfully triggered the deletion of the sequence between the loxP sites. This should lead to the inactivation of the intron and should thus abrogate splicing activity.

Having confirmed that recombination is occurring at the genomic locus, we next wanted to verify that this was leading to knock-out of the targeted gene. METLL16 is a highly essential gene in HAP1 cells and absence of this protein can be easily assessed by measuring cell viability. To this end, equal numbers of cells for METTL16-intron and CDK4-intron were plated in six-well dishes following Cre-ERT2 transduction and either treated or not treated with 4-OHT for a total of 6 days. Cells were then stained with crystal violet to assess total cell number as an approximation of cell viability (Figure 4). One can clearly observe that METTL16-intron in the presence of Cre and 4-OHT shows reduced viability when compared to other conditions. This suggests that METTL16 gene is no longer being expressed due to the deletion of the synthetic intron. As expected, CDK4-intron shows no changes in viability since this gene is known to be non-essential in HAP1. This highlights that the impact of the cassette is specific and likely to be a consequence of METTL16 inactivation.

We show here that the cassette we developed can be inserted into a reporter gene (NanoLuc) and insertion does not interfere with NanoLuc expression. The cassette contains parallel loxP sites and removal of the sequences between the loxP sites triggers a ~30fold decrease in NanoLuc expression. These findings were confirmed by inserting the cassette in endogenous gene in HAP1 cells. Cell lines containing this synthetic intron show recombination at the genomic level in the presence of Cre and show the expected phenotype upon knock-out of the endogenous gene. This suggests that the cassette is functional and able to disrupt gene expression upon transduction with Cre recombinase.

Addition of Cre recombinase (e.g. by addition of a recombinant adenovirus bearing Cre recombinase) leads to the deletion of critical elements within the intron that are required to maintain splicing, most importantly: the branch site. As a consequence, the cassette will no longer be removed and the ribosome will run into a stop codon, thus disrupting the expression of the gene of interest upon addition of Cre recombinase. The resulting cell line contains a frameshift and thus represents a complete gene knockout.

EXAMPLE 2:

Materials and Methods

Cassette sequences

Five different cassettes (Variants 1-5 here, referred to as Versions 1-5 above) were tested with varying locations of the loxP sites within the synthetic intron sequences. For NanoLuc trials, a 'recombined' version of each cassettes, with the predicted sequence after treatment with Cre recombinase, was tested alongside.

Table 2: Cassette sequences of variants 1-5

Name	Cassette sequence	Cassette sequence after Cre recombination
Variant 1	Version 1 (SEQ ID NO: 2)	Version 1 (SEQ ID NO: 3)
Variant 2	Version 2 (SEQ ID NO: 4)	Version 2 (SEQ ID NO: 5)
Variant 3	Version 3 (SEQ ID NO: 6)	Version 3 (SEQ ID NO: 7)
Variant 4	Version 4 (SEQ ID NO:. 8)	Version 4 (SEQ ID NO: 9)
Variant 5	Version 5 (SEQ ID NO:10)	Version 5 (SEQ ID NO: 11)

gRNA sequences

The gRNAs in Table 3 were used to insert the cassette.

Table 3: gRNAs used to insert the cassette

Targeted gene	**gRNA sequence**
CD46	GAUAAGGGUUUUUACCUCGA (SEQ ID NO: 25)
METTL16	UCUGACGUGUACUCUCCUAA (SEQ ID NO: 26)
Oct4/ Pou5F1 site#1	GCACUAGCCCCACUCCAACC (SEQ ID NO: 27)
Oct4/ Pou5F1 site#2	ACCACCUGGAGGGGGCGAGA (SEQ ID NO: 28)

NanoLuc assay

HAP1 or HEK293 cells were transfected with a pcDNA3.1 plasmid encoding NanoLuc or NanoLuc containing artificial intron variants (see above) and pGL4.53[luc2/PGK] vector (Promega) for normalization. Cells were harvested 24h-48h post transfection and analysed using the Nano-Glo Dual Luciferase Assay System (Promega) according to manufacturer's instructions.

Generation of cell lines

HAP1 cells were transfected with TurboFectin transfection reagent (OriGene). In brief, $0.8x10^6$ cells were seeded in a 6-well plate and transfected on the following day with 1.5µg of Cas9 expression plasmid (#48137 from Addgene) , 1µg of U6 driven gRNA expression plasmid (Horizon Discovery), 0.5µg of PCR product encoding the donor and 0.2µg of plasmid encoding a blasticidin resistance gene. 24h post transfection, cells were treated with 20µg/ml of blasticidin for 24h to eliminate untransfected cells. Single clones were obtained by limiting dilution. Positive clones were then identified by PCR screening.

iPS cells cultured in TeSR-E8 and Vitronectin XF (both STEMCELL Technologies) were transfected with the P3 Primary Cell solution using the CM-138 pulse in the 4D-Nucleofector (Lonza). In brief, $1.0x10^6$ cells were transfected with 2.5µg of Cas9 expression plasmid, 2.5µg of gRNA expression plasmid, 2.5µg of plasmid encoding the donor and 0.5µg of plasmid

encoding a blasticidin resistance gene. Cells were plated in culture media supplemented with 10μM ROCKi (Y-27632, Abcam) for 24h, and then treated with 10μg/ml of blasticidin for 24h to eliminate untransfected cells. Recovered colonies were single cell plated after treatment with StemPro Accutase (Thermo Fisher Scientific) and clones were manually picked into 96wells for screening.

Genomic DNA isolation and PCR

Genomic DNA was isolated using the QIAamp DNA Mini Kit (Qiagen) according to manufacturer's instructions. PCRs where done with the primers below using the GoTaq DNA Polymerase (Promega) following manufacturer's instructions.

The primers in Table 4 were used for PCR.

Table 4: Primers used for PCR

Name	Forward primer	Reverse primer
METTL16-DECAI genomic DNA	TTCTGCCTGTTTGCCGTAGA (SEQ ID NO. 17)	TTGTCAGAATCCTGGTGACCG (SEQ ID NO. 18)
CD46-DECAI genomic DNA	TGCATTCCATTCCTTGTCTCTG (SEQ ID NO: 29)	AAGACACTTTGGAACTGGGGG (SEQ ID NO: 30)
Oct4-DECAI#1	CTCTGAGGTGTGGGGGATT (SEQ ID NO:31)	TGCTCCAGCTTCTCCTTCTC (SEQ ID NO: 32)
Oct4-DECAI#2	GACACCTGGCTTCGGATTT (SEQ ID NO: 33)	CCCCACAGAACTCATACGG (SEQ ID NO:34)

RNA isolation, cDNA synthesis and qRT-PCR

Total RNA was isolated from cells using the RNeasy mini kit (QIAGEN) in accordance with manufacturer's protocol. For RNA from Hap1 cells, 1 μg of total RNA was reverse transcribed using oligo dT primer and M-MLV Reverse Transcriptase (Promega). The cDNA was then used as a template for PCR using the primers listed in the table below. For RNA from iPSCs, 1 μg of total RNA was reverse transcribed into cDNA using the SuperScript III First-Strand Synthesis SuperMix kit (Thermo Fisher Scientific). 10 ng of cDNA was used for qRT-PCR reactions that were set up in triplicates using either TaqMan Gene Expression Master Mix (Thermo Fisher

Scientific). TaqMan gene expression assays (Thermo Fisher Scientific) were used for each gene analysed. qRT-PCR experiments were performed using QuantStudio 6 Flex Real Time PCR System (Thermo Fisher Scientific). All values were normalized to an internal control (GAPDH) and brought to power of 2. Each target's value was then compared to the highest value of that target across all samples. All samples were done in technical triplicates and the median of these is displayed. The following TaqMan gene expression assays were used: Gapdh (Hs03929097_g1), Oct4 (Hs00999632_g1), Nanog (Hs04399610_g1), Sox2 (Hs01053049_s1), Cdx2 (Hs01078080_m1), Eomes (Hs00172872_m1) (all from Thermo Fisher Scientific).

The primers in Table 5 were used for RT-PCR:

Table 5: Primers for RT-PCR

Name	**Forward primer**	**Reverse primer**
CD46-DECAI RT-PCR	GGTCAAATGTCGATTTCCAGTAGT (SEQ ID NO: 35)	GACACTTTGGAACTGGGGGA (SEQ ID NO: 36)
GAPDH RT-PCR	GAAGGTGAAGGTCGGAGT (SEQ ID NO: 37)	GAAGATGGTGATGGGATTTC (SEQ ID NO: 38)

Retroviral production and transduction of cells

Approximately $1x10^7$ HEK392T cells were transfected with 1.7 µg pAdvantage plasmid (Promega), 2.6 µg VSV-G expression vector, 4 µg gag-pol expression vector and 6.6 µg CreERT2 expression vector, using TurboFectin transfection reagent (OriGene). Medium was changed 24 hours after transfection. Viral supernatant was collected 48 and 72 hours after transfection, and used to transduce the METTL16-DECAI cells. Since the Cre-ERT2 retrovirus also encoded a PGK-PuroR cassette, cells containing Cre could be enriched for by selecting with 0.3µg/ml puromycin.

FACS, Western Blotting and Immunocytochemistry

For the HAP1 experiments, cells were trypsinized and washed with PBS. Cells were stained with a CD46-specific antibody (APC-CD46; catalogue #564253; BD Biosciences) in FACS buffer (5% FCS in PBS) for 30 minutes and excess antibody was removed by washing with PBS. Cells were analysed by flow cytometry using the BDLSR Fortessa.

For the human iPS cell experiments, cells were fixed and permeabilized using the Cytofix/Cytoperm kit (BD) according to manufacturer's instructions. Cells were then stained using the following directly conjugated antibodies: Alexa Fluor® 488 anti-Oct4 (653706, BioLegend) and PE/Cy7 anti-human SSEA-4 (330420, BioLegend). Cells were analysed using the iQue Screener PLUS (IntelliCyt). Data was analysed using the FlowJo software.

For Western blotting, lysates were prepared using Frackelton buffer (10mM Tris/HCl pH7.5, 50mM NaCl, 30mM sodium pyrophosphate, 1% Triton X-100, 50mM NaF and protease inhibitors). Lysates were separated on 10% SDS-PAGE and blotted on Nitrocellulose membranes. Membranes were stained with a primary antibody (anti-CD46, ab108307, Abcam) and a secondary antibody (anti-rabbit, 111-035-003, Jackson Immunoresearch Europe) and visualized using ECL reagent (Thermo Fisher Scientific).

For Immunocytochemistry, the PSC 4-Marker Immunocytochemistry Kit (Thermo Fisher Scientific, A24881) was used. Briefly, iPS cells were fixed using 4% PFA for 15min, permeabilized using 1% Saponin for 15min, and blocked with 3% BSA for 30min (all at room temperature). Primary antibodies used: rabbit anti-Oct4 (A24867) and mouse IgG3 anti-SSEA4 (A24866); secondary antibodies: Alexa Fluor 594 donkey anti-rabbit (A24870) and Alexa Fluor 188 goat anti-mouse IgG3 (A24877). DAPI was used for nuclear staining.

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Results

To generate conditional gene knockouts, we designed a genetic approach that was based on inserting a small intron cassette into the coding sequence of an exon. Introns were favoured because they are removed by splicing and thus leave the targeted gene intact until the cassette is activated (**Figure 5A**). To overcome limitations of previous similar approaches, we decided to minimize the size of the intron and to base our design on a very small chimeric intron. Introns contain three elements that are critical for their function: the splice donor, the branch point and the splice acceptor. We reasoned that removal of the branch point by Cre/loxP recombination could cripple the intron, thus triggering the translation of truncated intronic sequence (**Figure 5B**). Addition of three stop codons (one for each reading frame) would then lead to translational termination of the gene and would thus abrogate gene expression.

Since the intron we used is very small, the precise architecture of the cassette was not straightforward and the two major risks included: (i) cassette modification could disturb intron splicing, hence leaving the cassette in a constant "on-state"; (ii) as the branch point sequence is quite degenerate, nearby sequences may take over branch point function upon its removal. To overcome these difficulties, we established an assay based on NanoLuc reporter gene expression where we could easily test a variety of cassette configurations. A series of NanoLuc expression plasmids were created in which the different cassette designs were inserted into one specific location of the NanoLuc sequence (see sequence details in the Materials and Methods). We also included the respective sequences as they would occur after Cre recombination, thus simulating the Cre/loxP excision event. An ideal cassette leaves NanoLuc expression unaffected (off-state) but triggers a profound decrease upon Cre/loxP recombination (on-state).

First, we noted that insertion of the intronic sequence without modification (no loxP sites) was well tolerated (**Figure 6A, Figure 1**). This suggested that the intron was functional in principal and that the site within NanoLuc tolerated an intronic insertion. Next, we assessed the various variants side-by-side **(Figure 6A, Figure 1)**. Of interest, introduction of cassette variant 3 (v3) lead to a strong decrease in NanoLuc expression even in the "off-state" of the cassette, suggesting that this cassette modification was incompatible with proper intron splicing. Variants 1 and 2 were well tolerated, yet we did not observe any reduction in gene expression upon Cre recombination, suggesting that alternative branch points had taken over. Of the variants tested, variant 4 (**Figure 5C**) was most fit for purpose in as much as it was well tolerated in the "off-state" and triggered a profound decrease (~14 fold) in NanoLuc expression upon cassette activation.

Splicing is strongly dependent on the sequence context and requires certain specific sequences at the intron/exon junction. Introns of the GT-AG category, such as the one we chose, have a preference for CAG or AAG upstream of the splice donor and G or A downstream of the splice acceptor[16]. Hence, the choice of target sites to insert an intron into the coding sequence of an exon should be restricted to CAG-G, CAG-A, AAG-G or AAG-A (where the hyphen denotes the site where the intron is inserted). To test whether all four sites tolerated the insertion of the intron equally, we created eight NanoLuc constructs, targeting eight different intron insertion sites (two for each of the four motifs). Transient overexpression in HEK293 suggests that all of these sites tolerated the insertion of an intron equally (**Figure 6B**). This suggests that our approach is broadly applicable across the various insertion sites depicted above.

Having established a cassette in which the desired features had been encoded in just 201 nucleotides, we decided to test this cassette in human cells. In order to introduce the intron, we designed and synthesized homology donors of ~1kb in which the cassette had been placed in the middle (leaving ~400bp homology arms on either end; **Figure 10**). We targeted exons of two human genes: (i) CD46 because it is cell-surface expressed and can be easily detected by FACS; and (ii) METTL16 because it is an essential gene and cassette activation should trigger cell death.

We chose the human haploid cell line HAP1 for the ensuing targeting experiments because these cells contain a single copy of every human gene and conditional gene inactivation is thus straightforward[17]. We chose exons that were at least 100bp from the start codon and within the first 30% of the coding sequence of the gene. Following transfection of HAP1 cells with Cas9, the respective gene-specific gRNA and the homology donor, we obtained clonal cell lines by limiting dilution. Single clones were genotyped by PCR to identify clones in which cassette integration had occurred. Note that we did not select for cassette integration as the conditional KO cassette itself does not contain any selectable marker. In spite of this, we still obtained targeted clones at frequencies of 5% or higher (**see Table 6**). Of interest, the frequency at which the cassette was incorporated was higher in the essential gene (METTL16) than in the non-essential gene (CD46). While this may well be a coincidence, it is plausible to assume higher targeting frequencies for essential genes as other editing events (e.g. the accumulation of indels) may be subject to negative selection.

Table 6

Gene	Cell line	# of clones screened	# of clones containing the cassette
CD46	Hap1	96	8
METTL16	Hap1	96	32
OCT4 site#1	iPSC	56	4 (2 KI/KI, 1 KI/WT, 1 KI/NHEJ)
OCT4 site#2	iPSC	56	3 (1 KI/KI, 2 KI/NHEJ)

Clonal cell lines bearing the cassette were expanded and Cre recombinase was expressed to trigger cassette activation. For CD46, HAP1 cells were transiently transfected with a plasmid encoding Cre. For METTL16, cells were infected with a retrovirus encoding Cre-ERT2 recombinase. Cre-ERT2 was activated by addition of 4-hydroxitamoxifen (4-OHT). Genomic DNA from the resulting cell lines was isolated and analysed by PCR (**Figure 7A**). Cell lines bearing tagged alleles displayed a band that was shifted in size by ~200bp, corresponding to

the artificial intron that had been integrated at the targeted site. Following transduction with Cre (and activation with 4-OHT), this band was shifted to smaller size by ~150bp, corresponding to the sequence that had been removed by Cre recombination. This indicates that Cre-mediated recombination faithfully triggers the removal of the sequences between the two parallel loxP sites. We also isolated mRNA from the cell line in which CD46 had been tagged with the artificial intron and analysed the CD46 transcript by RT-PCR (**Figure 7B**). In the absence of Cre, the intron was efficiently removed in the CD46-DECAI clone and thus not detectable at the mRNA level. Upon addition of Cre, the intron was inactivated and the crippled intronic sequence was incorporated in the mRNA sequence, leading to a shift towards higher molecular weight.

Next, we assessed the impact of cassette activation on gene expression and gene function. We started our analysis with CD46 because the gene is non-essential and its expression can be easily visualized. We compared cells bearing the intron to wild-type cells and noted no difference, neither by Western blotting (**Figure 8A**) nor by flow cytometry (**Figure 11**). This indicates that the artificial intron did not disrupt endogenous gene expression in the "off-state". Upon transfection of cells with Cre, CD46-DECAI cells abrogated CD46 expression almost completely, whereas wild-type cells were unaffected (**Figure 8A and B**). The few remaining CD46-positive cells (~5% as judged by FACS) likely arose from incomplete transfection of cells with Cre recombinase.

We also assessed the impact of the artificial intron on the essential gene METTL16. Transduction of cells with Cre-ERT2 and activation of Cre-ERT2 with 4-OHT triggered cell death in METTL16-DECAI cells (**Figure 8C, Figure 4 right hand side**). We noted that cell death was incomplete (data not shown), suggesting that retroviral Cre-ERT2 expression may have been silenced in a subpopulation of cells. To overcome this silencing, we isolated single cell clones bearing retroviral insertions of Cre-ERT2 and activated Cre-ERT2 by 4-OHT treatment (**Figure 12**). As expected, some clones showed complete cell death upon 4-OHT treatment, whereas others were almost unaffected. This highlights the need for effective delivery and expression of Cre.

We also wanted to apply our approach in a more physiological setting and decided to test it in human induced pluripotent stem cells (iPSCs). Human iPSCs are pluripotent and their pluripotency depends on the expression of the transcription factor Oct4/POU5F1[18]. Inactivation of Oct4 triggers the collapse of the core transcriptional network leading to rapid loss of pluripotency. This is associated with drastic changes in colony morphology and loss of other pluripotency markers, such as Nanog, Sox2 and SSEA4. We created homology donors for

Oct4/POU5F1 (400bp homology arms flanking the DECAI cassette) and targeted two independent sites in exon 1 (**Figure 10**). Clonal cell lines were established and genotyped by PCR. Both homozygous and heterozygous cell lines were achieved **(Table 5)**.

We selected two homozygous clones bearing the artificial intron in distinct genomic locations in exon 1 to study in greater detail. Immunofluorescence staining (**Figure 9A**), flow cytometry analysis (**Figure 9B**) and RNA expression (**Figure 9C**) showed that human iPSC colonies bearing the intron expressed similar levels of Oct4 as non-targeted cells , indicating that the intron by itself did not affect Oct4 expression. Likewise, pluripotency was retained as suggested by the SSEA4 staining, and normal expression of pluripotency markers Nanog and Sox2 **(Figure 9A-C).** Upon recombination with Cre, wildtype cells were essentially unaffected. In contrast, Oct4-DECAI cells lost Oct4 expression, and consequently, SSEA4 (**Figure 9A-B**), Nanog and Sox2 expression (**Figure 9C**). They also changed their morphology quite dramatically (**Figure 13**), indicating that these cells had lost their pluripotency. The induced knockout of Oct4 resulted in differentiation of human iPSCs with a significant increase in transcription of Cdx2 and Eomes (**Figure 9C**), genes associated with trophoblast and endoderm lineages [19,20]. In line with these observations, we could detect Cre/loxP recombination at the genomic DNA level (**Figure 14**). Altogether, this highlights the feasibility of our approach for conditional gene inactivation and suggests its applicability in more physiological settings.

Discussion

In the examples described herein, we created a novel small cassette that readily allows the tagging of endogenous genes, poising them for conditional gene inactivation. The system is robust, triggers a profound decrease in gene expression and may be the first approach that can be truly applied at a large scale.

Of note, the work presented here has been inspired by the CRISPR-FLIP approach[15]. Yet it goes beyond this approach in at least two key aspects: (i) gene expression is disrupted by a series of stop codons, not by revealing a splice acceptor that traps the transcript; (ii) most importantly, the cassette size is considerably smaller (201bp versus at least 1,931bp in CRISPR-FLIP) and hence, donor assembly is more straightforward. Additionally, even strong splice acceptors can be leaky depending on the genomic context[21] and hence, the impact of the CRISPR-FLIP cassette may vary from gene to gene.

Experiments presented here were performed in HAP1 cells and human iPSCs, but we are confident that this approach can be applied to any mammalian cell line. While engineering in haploid cell lines, such as HAP1, is straightforward with only a single allele to target, most cell

lines are diploid like iPSCs or aneuploid, harbouring multiple chromosome copies. To obtain complete conditional knockouts in such cells, we envisage two possible solutions: (i) complete bi-allelic or multi-allelic knock-in: one obtains a cell line in which both alleles of a given gene of interest have been tagged with the cassette. While this is possible, the efficiency will likely not be very high as bi-allelic tagging is infrequent. (ii) Mono-allelic knock-in with disruption of the open reading frame of the additional alleles: as a compromise, we suggest to screen for clones in which one allele has been tagged and the other allele(s) have been inactivated by Cas9 cleavage following NHEJ repair. Since the donor template is integrated in the middle of a coding exon near the start of a gene, the gRNA used to trigger incorporation of the donor will likely lead to indels abrogating gene expression. This is exemplified by our results in iPSCs where clones were recovered with both bi-allelic targeting and mono-allelic knock-in with indels on the second allele. This may not be possible in all instances as some genes may require all alleles to be intact, but for most genes, this is likely to represent an acceptable and tolerable compromise.

The artificial intron approach presented here also has some conceivable shortcomings. First, there is a potential risk that insertion of the artificial intron may dysregulate endogenous splicing events, leading to constitutive gene inactivation. When conducting these studies, we noted one case where insertion of the artificial intron into the CDK4 gene (which is non-essential in HAP1 cells) abrogated CDK4 expression (data not shown). A more detailed analysis of these cells revealed that the intron was successfully removed by splicing, but some erroneous splicing events led to the assembly of a non-functional mRNA that was most likely degraded by non-sense mediated decay. Hence, when targeting non-essential genes, one may need to monitor the impact of cassette integration on endogenous gene expression. Fortunately, this can easily be assessed by a conventional RT-PCR. Of note, this is less of a problem for essential genes because cells bearing such editing events will not be viable. Second, our method depends on robust Cre recombinase delivery and expression for activating the cassette to the on-state. Here, we introduced Cre-ERT2 by retroviral transduction for METTL16-DECAI and OCT4-DECAI activation. While this is possible, we noticed some shortcomings, mainly: (i) Cre-ERT2 expression can be leaky and already occur in the absence of 4-OHT (data not shown) and (ii) Cre-ERT2 expression can be silenced, especially when Cre-ERT2 is delivered via retroviral infection. A more elegant approach would be to express Cre-ERT2 from a safe-harbour locus such as AAVS1 or ROSA26. This would not only be better defined from a genetic point of view, but also likely to yield more robust and reproducible levels of Cre expression.

While the methods described herein offer an easy and efficient strategy for generating conditional knockouts, some alternative methods exists and are worth mentioning. Recently, several strategies have been developed where modified versions of the CRISPR/Cas9 system are adopted to modulate gene expression. One example is CRISPR interference (CRISPRi), where a catalytically dead Cas9 directed by a guide RNA will suppress expression of the target gene by sterically hindering transcription initiation/ elongation or by the action of a silencing effector domain fused to Cas9[22,23]. While this strategy has the advantage of being reversible, it is difficult to reach complete inhibition of the target gene expression. Yet another method, is the use of inducible CRISPR systems. There are several iterations of these systems ranging from split Cas9 molecules that form a catalytically active molecule upon a certain stimulus[24] or control of Cas9 or gRNA expression by inducible promoters[25]. The efficiency of these methods will depend on the expression level of the gRNA and will likely show a variety of effects across cells in the sample since the editing induced by Cas9 will not happen in a predictable manner. This contrasts with the DECAI method described herein which will lead to a predictable modification and disruption of expression upon Cre recombination.

Another alternative to the genetic approach described herein, is a biochemical approach such as degron tagging. Degrons are drug-regulatable domains that allow one to modulate the abundance of the degron tagged protein by adding or withdrawing a small molecule. Popular degrons include the auxin/ mAID system[9],the HaloPROTAC3/ HaloTag system[10] and the HCV protease inhibitor/ SMAShTag system[11]. At least some of the degrons that are commonly used (such as the mAID system) are much faster in degrading the client protein (with a half-life of minutes[9] rather than hours or days) as they lead to the recruitment of E3 ligases that actively degrade the protein population that is present. Additionally, the degron tags have the advantage of allowing reversible and tuneable degradation of the target proteins. However, the use of degrons has two major limitations: (i) the bi-allelic tagging of genes occurs at very low frequency and thus, it is difficult to obtain a cell line or an organism in which all copies of a gene have been conditionally inactivated and (ii) the insertion of foreign sequences within the coding region of a gene could disturb protein folding or protein function. As a consequence, degron tagging can be cumbersome and may not be compatible with high-throughput applications.

With this in mind, the approach described herein is more suitable for serial production of conditional knockouts, whereas degron tagging may be the method of choice if the protein is well understood and inefficient clone recovery is less of a constraint.

In summary, the DECAI method described herein offers a convenient and efficient alternative for generating conditional knockout alleles in mammalian cells.

5

References

1 Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nature methods* **10**, 957-963, doi:10.1038/nmeth.2649 (2013).

2 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).

3 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

4 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

5 Dickinson, D. J., Ward, J. D., Reiner, D. J. & Goldstein, B. Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. *Nature methods* **10**, 1028-1034, doi:10.1038/nmeth.2641 (2013).

6 van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol Cell* **63**, 633-646, doi:10.1016/j.molcel.2016.06.037 (2016).

7 Blomen, V. A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science*, doi:10.1126/science.aac7557 (2015).

8 Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337-342, doi:10.1038/nature10163 (2011).

9 Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep* **15**, 210-218, doi:10.1016/j.celrep.2016.03.001 (2016).

10 Buckley, D. L. *et al.* HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chem Biol* **10**, 1831-1837, doi:10.1021/acschembio.5b00442 (2015).

11 Chung, H. K. *et al.* Tunable and reversible drug control of protein production via a self-excising degron. *Nat Chem Biol* **11**, 713-720, doi:10.1038/nchembio.1869 (2015).

12 Economides, A. N. *et al.* Conditionals by inversion provide a universal method for the generation of conditional alleles. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E3179-3188, doi:10.1073/pnas.1217812110 (2013).

13 Lackner, D. H. *et al.* A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nature communications* **6**, 10237, doi:10.1038/ncomms10237 (2015).

14 Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun* **237**, 752-757, doi:10.1006/bbrc.1997.7124 (1997).

15. Andersson-Rolf, A. *et al.* One-step generation of conditional and reversible gene knockouts. *Nat. Methods* **14,** 287–289 (2017).

16. Sibley, C. R., Blazquez, L. & Ule, J. Europe PMC Funders Group Lessons from non-canonical splicing. **17,** 407–421 (2017).

17. Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **24,** 2059–2065 (2014).

18. Kim, J., Chu, J. L., Shen, X. H., Wang, J. L. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells (vol 132, pg 1049, 2008). *Cell* **133,** 1290 (2008).

19. Zafarana, G., Avery, S. R., Avery, K., Moore, H. D. & Andrews, P. W. Specific knockdown of OCT4 in human embryonic stem cells by inducible short hairpin RNA interference. *Stem Cells* **27,** 776–782 (2009).

20. Chen, Y., Wang, K., Gong, Y. G., Khoo, S. K. & Leach, R. Roles of CDX2 and EOMES in human induced trophoblast progenitor cells. *Biochem. Biophys. Res. Commun.* **431,** 197–202 (2013).

21. Bürckstümmer, T. *et al.* A reversible gene trap collection empowers haploid genetics in human cells. *Nat. Methods* **10,** 965–971 (2013).

22. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–51 (2013).
23. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–83 (2013).
24. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33,** 139–142 (2015).
25. de Solis, C. A., Ho, A., Holehonnur, R. & Ploski, J. E. The Development of a Viral Mediated CRISPR/Cas9 System with Doxycycline Dependent gRNA Expression for Inducible In vitro and In vivo Genome Editing. *Front. Mol. Neurosci.* **9,** 70 (2016).

Claims

1. A method for generating an allele for conditional gene knock-out in a cell comprising a target gene, the method comprising: introducing an artificial intron sequence into an exon of the target gene, the artificial intron sequence comprising:
   a) a splice donor sequence;
   b) a first nuclease or recombinase site;
   c) a branch point sequence;
   d) a second nuclease or recombinase site;
   e) a splice acceptor sequence; and

   a stop codon positioned 5' to or within the first nuclease or recombinase site.

2. A method according to claim 1 comprising introducing into the cell a sequence-specific nuclease that cleaves a sequence within an exon of the target gene thereby introducing the artificial intron sequence into the target gene.

3. The method of claim 1 or claim 2 further comprising the step of introducing or activating a recombinase or nuclease in the cell thereby excising or disrupting the branch point and abrogating splicing of the artificial intron sequence.

4. The method of any preceding claim wherein the artificial intron sequence comprises two stop codons, more preferably three stop codons to cover all possible reading frames of the gene.

5. The method of claim 4 or claim 5 wherein a series of three or more stop codons are positioned 5' to or within the first nuclease or recombinase site.

6. The method of any preceding claim wherein the artificial intron contains a single branch point.

7. The method of any preceding claim wherein the artificial intron sequence is introduced by homology directed repair (HDR) or non-homologous end joining (NHEJ).

8. The method of any preceding claim wherein the sequence-specific nuclease is an RNA guided nuclease, or a DNA guided nuclease, a Zinc finger nuclease or a TALEN.

9. The method of claim 8, wherein the RNA-guided nuclease is Cas9 or an RNA-guided Argonaute.

10. The method of claim 7 wherein the DNA-guided nuclease is a DNA-guided Argonaute.

11. The method of any of claims 2 to 10, wherein the sequence-specific nuclease is introduced as a protein, mRNA, or cDNA.

12. The method of any preceding claim wherein first and second recombinase sites are loxP sites or FRT sites.

13. The method of claim 12 wherein the first (5') loxP site comprises at least one stop codon.

14. The method of any preceding claim wherein the first recombinase or nuclease site is positioned adjacent to the splice donor site.

15. The method of any of claims 3 to 14 wherein the recombinase is Cre recombinase or Flp recombinase.

16. The method of any preceding claim wherein the cell is a mammalian cell preferably a HAP1 cell, a human induced pluripotent stem cell, an embryonic stem cell, or a CHO cell or a T-cell.

17. The method of any preceding claim wherein the artificial intron sequence is about 200-250, 200-300, 200-400, 200-500, 200-600, 200-700 or 200-800 or 200-1000 nucleotides in length.

18. The method of any preceding claim wherein the artificial intron sequence comprises a nucleotide sequence at least 80%, 85%, 90%, 95%, 97%, 98% 99%, or 100% identical to SEQ ID NO: 8 .

19. A cell containing an artificial intron introduced according to the method of any of claims 1 to 18.

20. A cell comprising an artificial intron in an exon of a target gene, the artificial intron having at least one stop codon positioned 5' to or within a recombinase site, and wherein a single branch site is flanked by a pair of recombinase or nuclease sites, the recombinase or nuclease sites being arranged such that the single branch site may be excised or disrupted, thereby abrogating splicing of the artificial intron.

21. A cell according to claim 19 or claim 20 for use in therapy.

22. The method of any of claims 3-18, wherein the artificial intron is introduced into a population of cells, and wherein the recombinase or nuclease is introduced or activated in a specific subset of cells to inactivate the target gene selectively in this subset of cells.

23. The method of any of claims 1-18, wherein the artificial intron sequence further comprises a selection gene (e.g. an antibiotic resistance gene such as puromycin) or a marker (e.g. a molecular barcode).

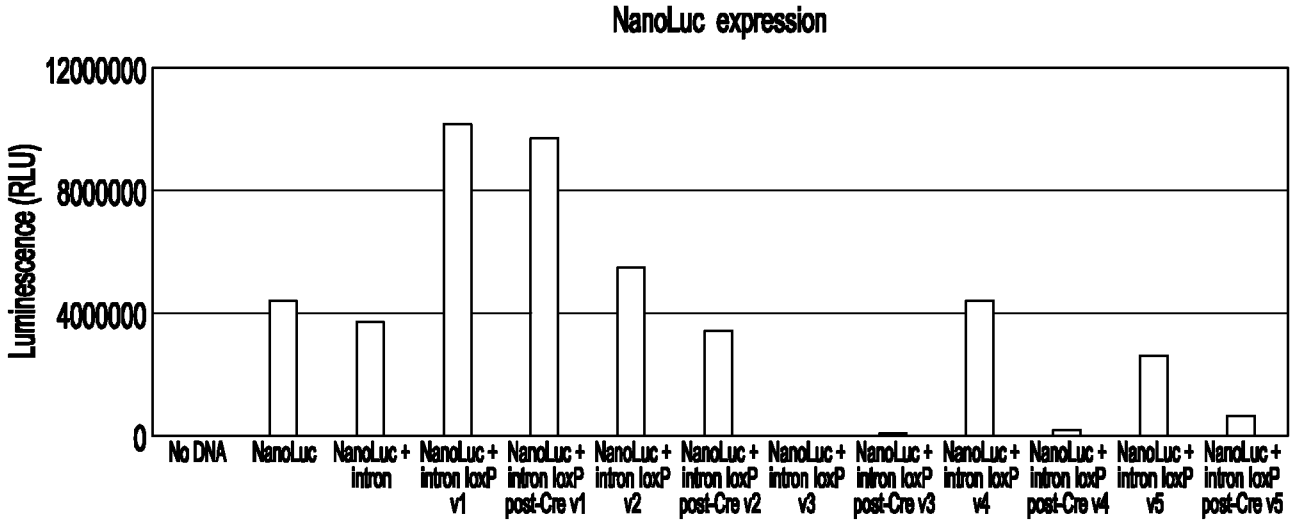

NanoLuc expression
Luminescence (RLU)
12000000
8000000
4000000
0
No DNA
NanoLuc
NanoLuc + intron
NanoLuc + intron loxP v1
NanoLuc + intron loxP post-Cre v1
NanoLuc + intron loxP v2
NanoLuc + intron loxP post-Cre v2
NanoLuc + intron loxP v3
NanoLuc + intron loxP post-Cre v3
NanoLuc + intron loxP v4
NanoLuc + intron loxP post-Cre v4
NanoLuc + intron loxP v5
NanoLuc + intron loxP post-Cre v5

Fig . 1

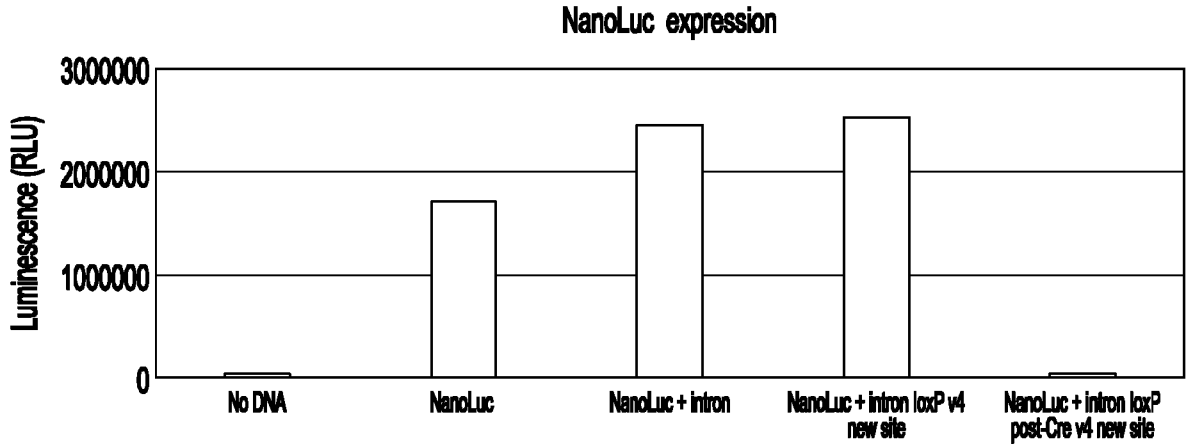

NanoLuc expression
Luminescence (RLU)
3000000
2000000
1000000
0
No DNA
NanoLuc
NanoLuc + intron
NanoLuc + intron loxP v4 new site
NanoLuc + intron loxP post-Cre v4 new site

Fig . 2

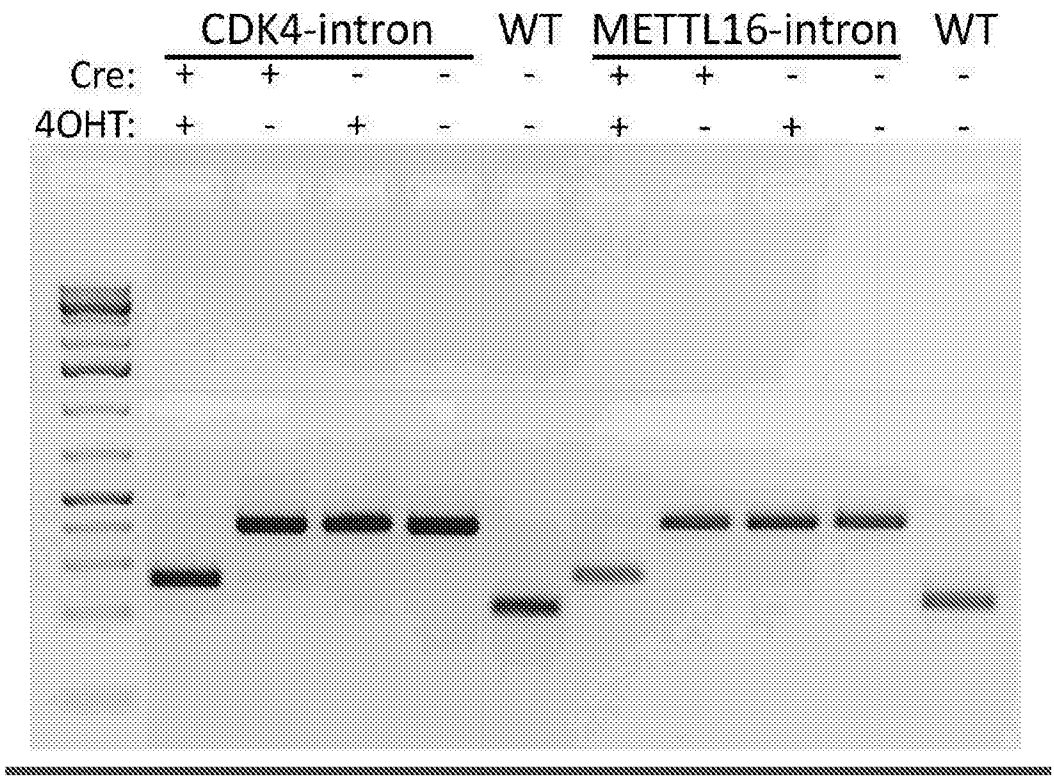


Fig . 3

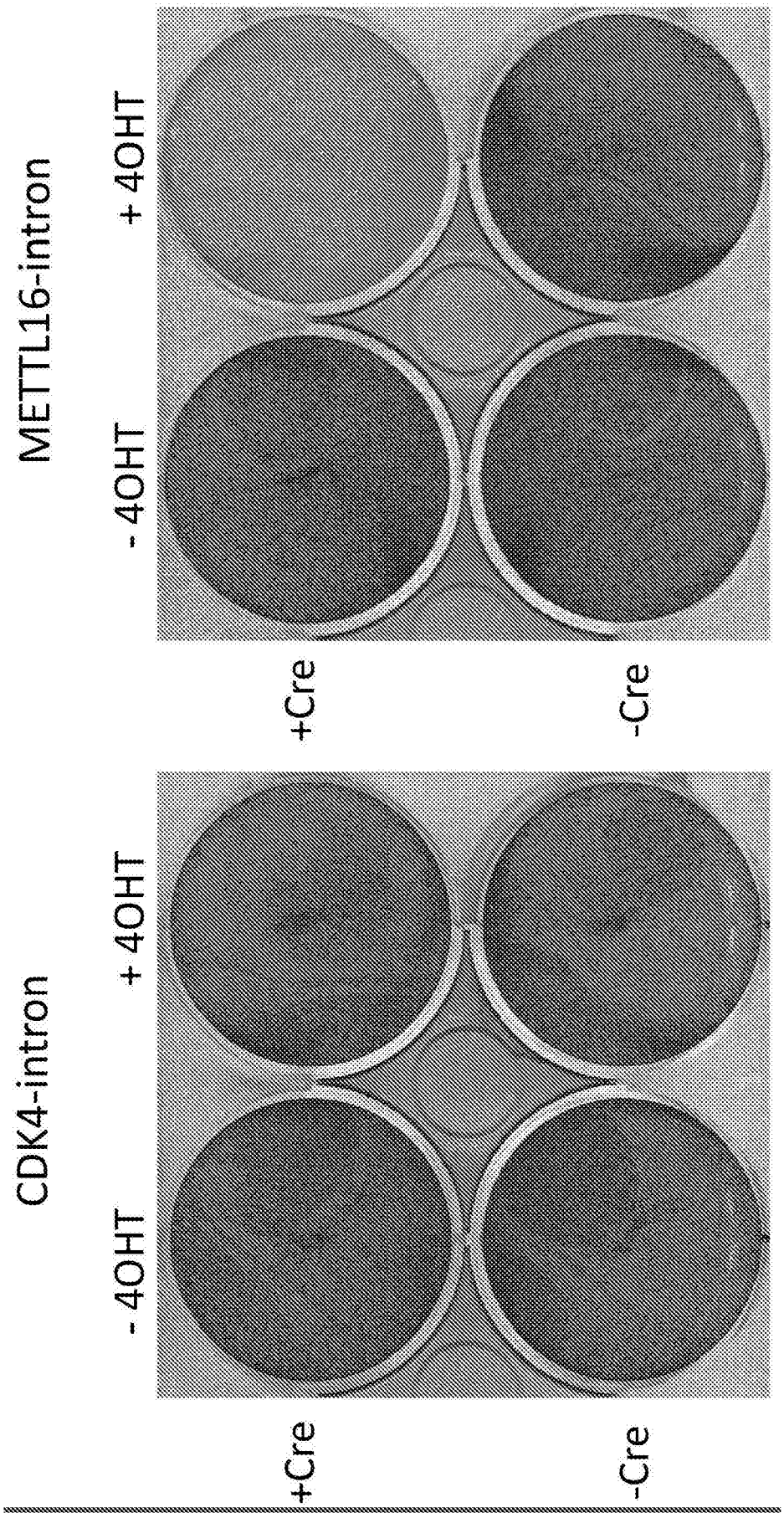
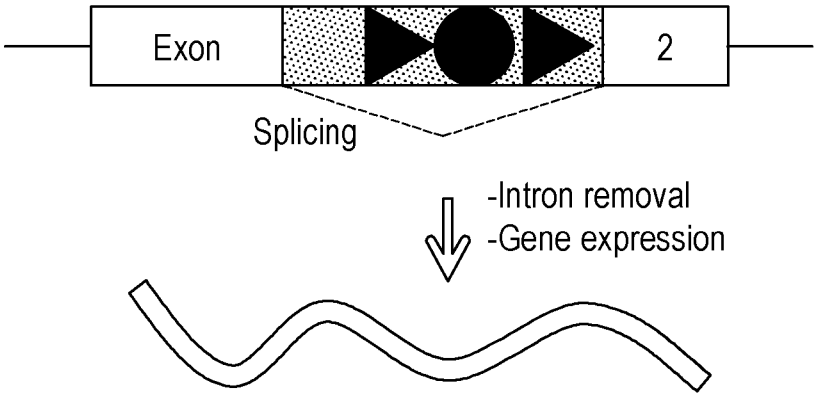
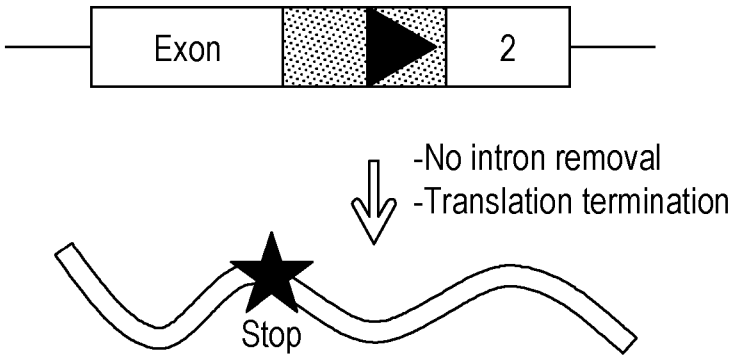

METTL16-intron
+ 4OHT
- 4OHT
+Cre
-Cre
CDK4-intron
+ 4OHT
- 4OHT
+Cre
-Cre

Fig . 4

(A) No Cre: Intron intact



(B) Cre: Intron crippled



(C) Artificial intron cassette (SEQ ID NO: 8)

GTAAGTAAataacttcgtatagcatacattatacgaagttatTC
AAGGTTAGAAGACAGGTTTAAGGAGACCAATAGAA
ACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTT
TCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTG
CCataacttcgtatagcatacattatacgaagttatTTTCTCTCC
ACAG

Key:
Italics = Red italics
Yellow
Green
Blue

Fig . 5

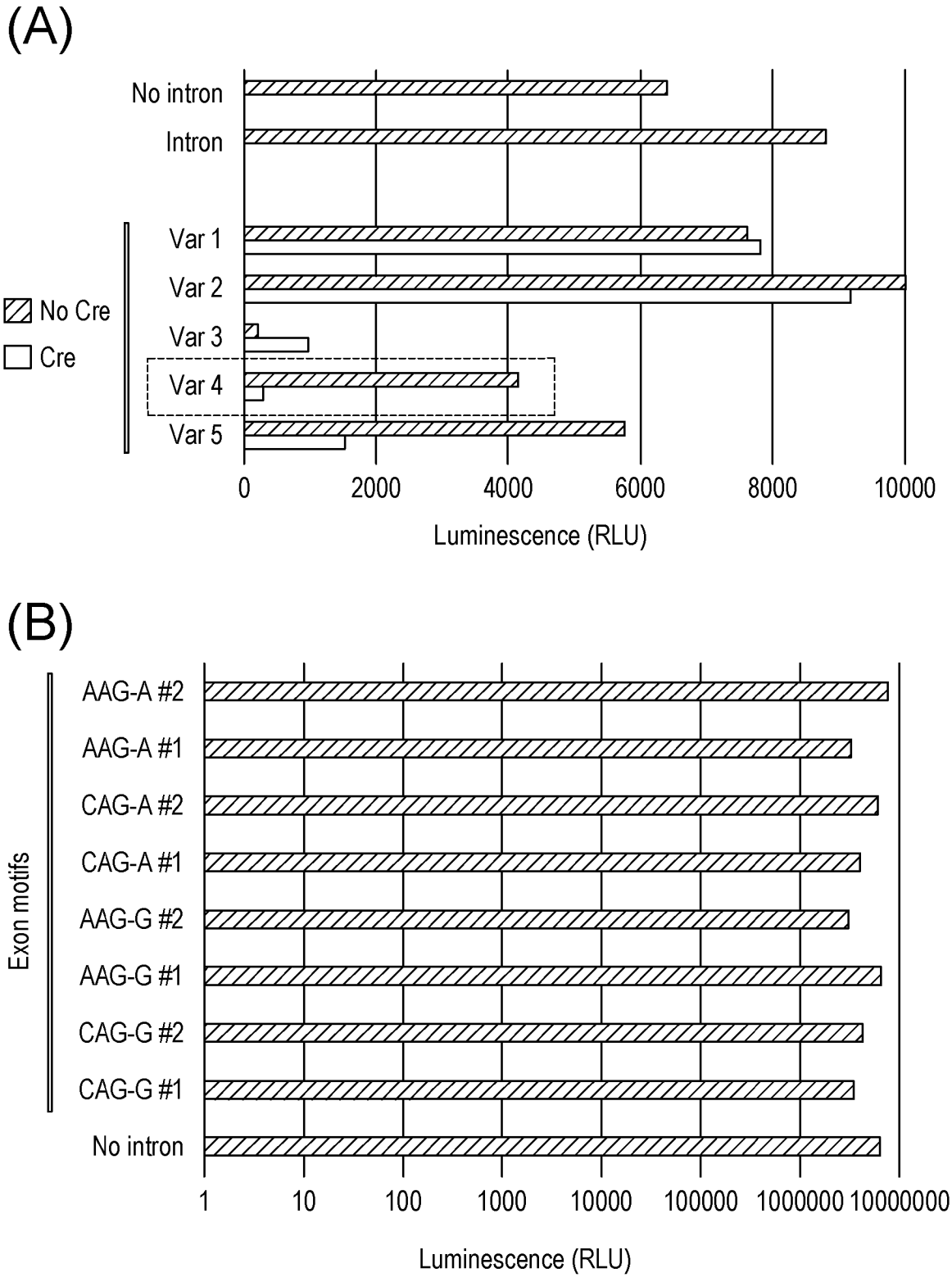

(A)
No intron
Intron
Var 1
Var 2
Var 3
Var 4
Var 5
No Cre
Cre
0
2000
4000
6000
8000
10000
Luminescence (RLU)
(B)
Exon motifs
AAG-A #2
AAG-A #1
CAG-A #2
CAG-A #1
AAG-G #2
AAG-G #1
CAG-G #2
CAG-G #1
No intron
1
10
100
1000
10000
100000
1000000
10000000
Luminescence (RLU)

Fig . 6

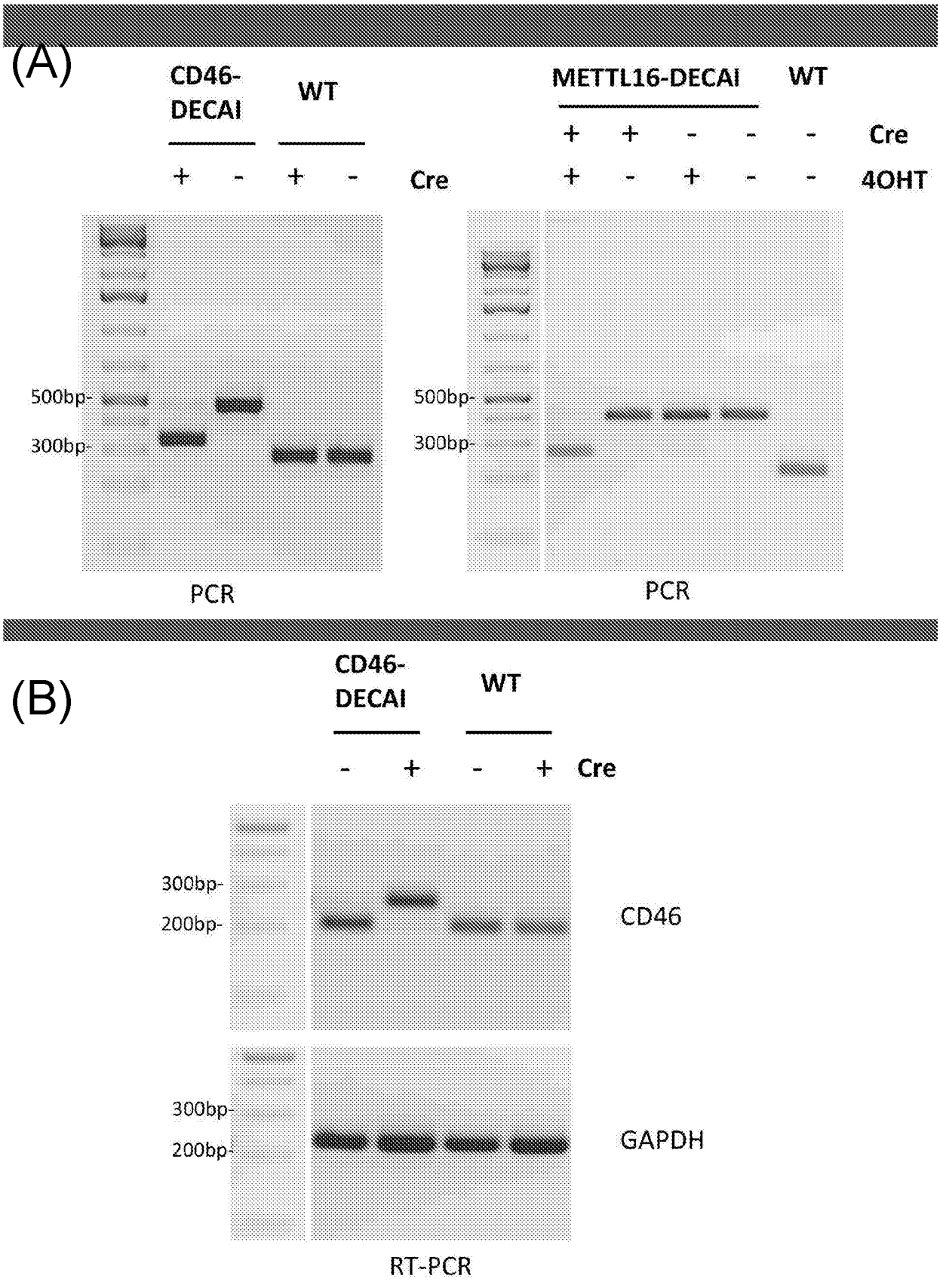
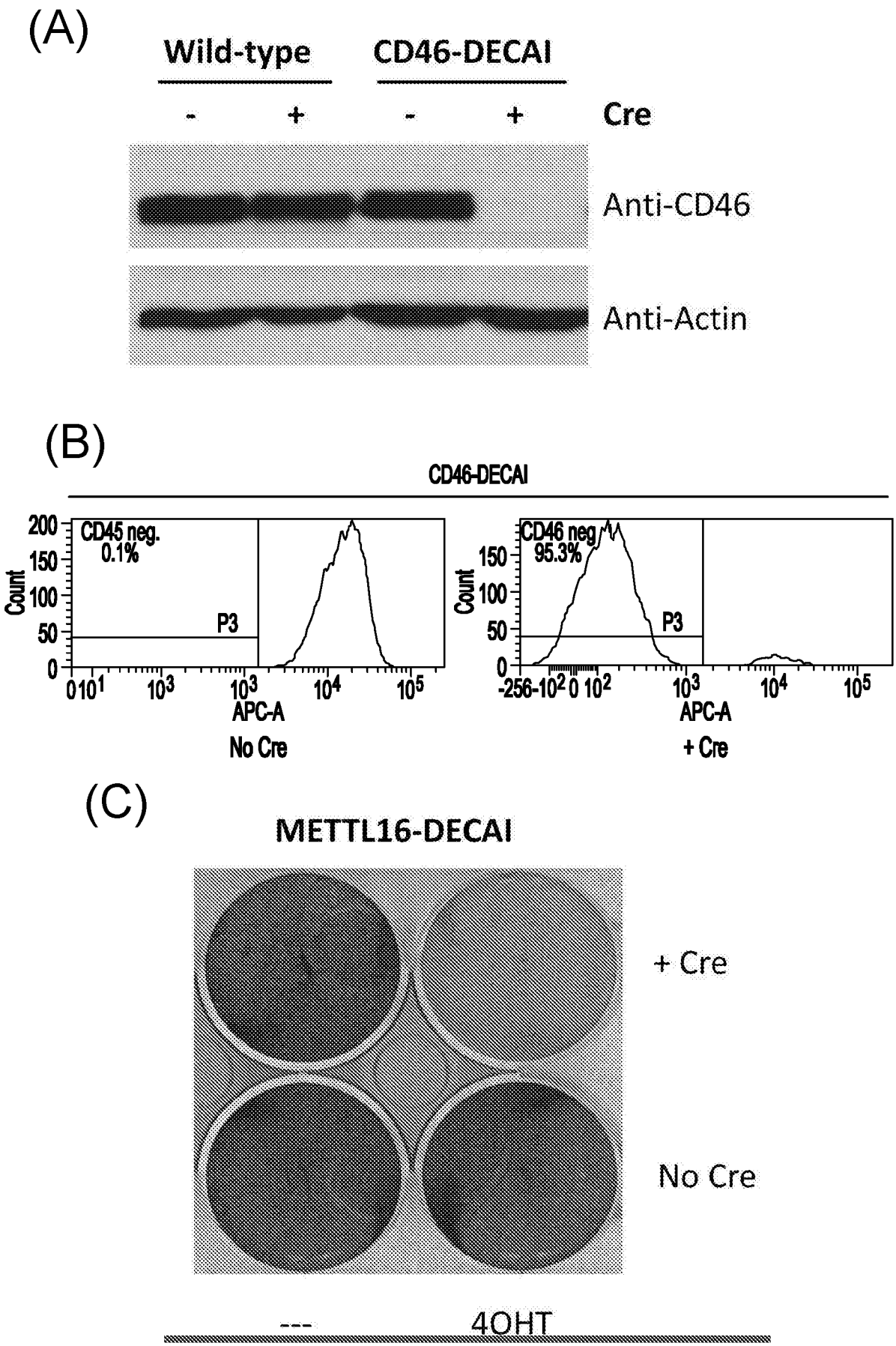

(A)
CD46-DECAI
WT
METTL16-DECAI
WT
+ + - - - Cre
+ - + - - 4OHT
Cre
500bp-
300bp-
PCR
500bp-
300bp-
PCR
(B)
CD46-DECAI
WT
- + - + Cre
300bp-
200bp-
CD46
300bp-
200bp-
GAPDH
RT-PCR

Fig . 7


(A)
Wild-type
CD46-DECAI
- + - +
Cre
Anti-CD46
Anti-Actin
(B)
CD46-DECAI
CD45 neg.
0.1%
P3
Count
APC-A
No Cre
CD46 neg
95.3%
P3
Count
APC-A
+ Cre
(C)
METTL16-DECAI
+ Cre
No Cre

4OHT

Fig . 8

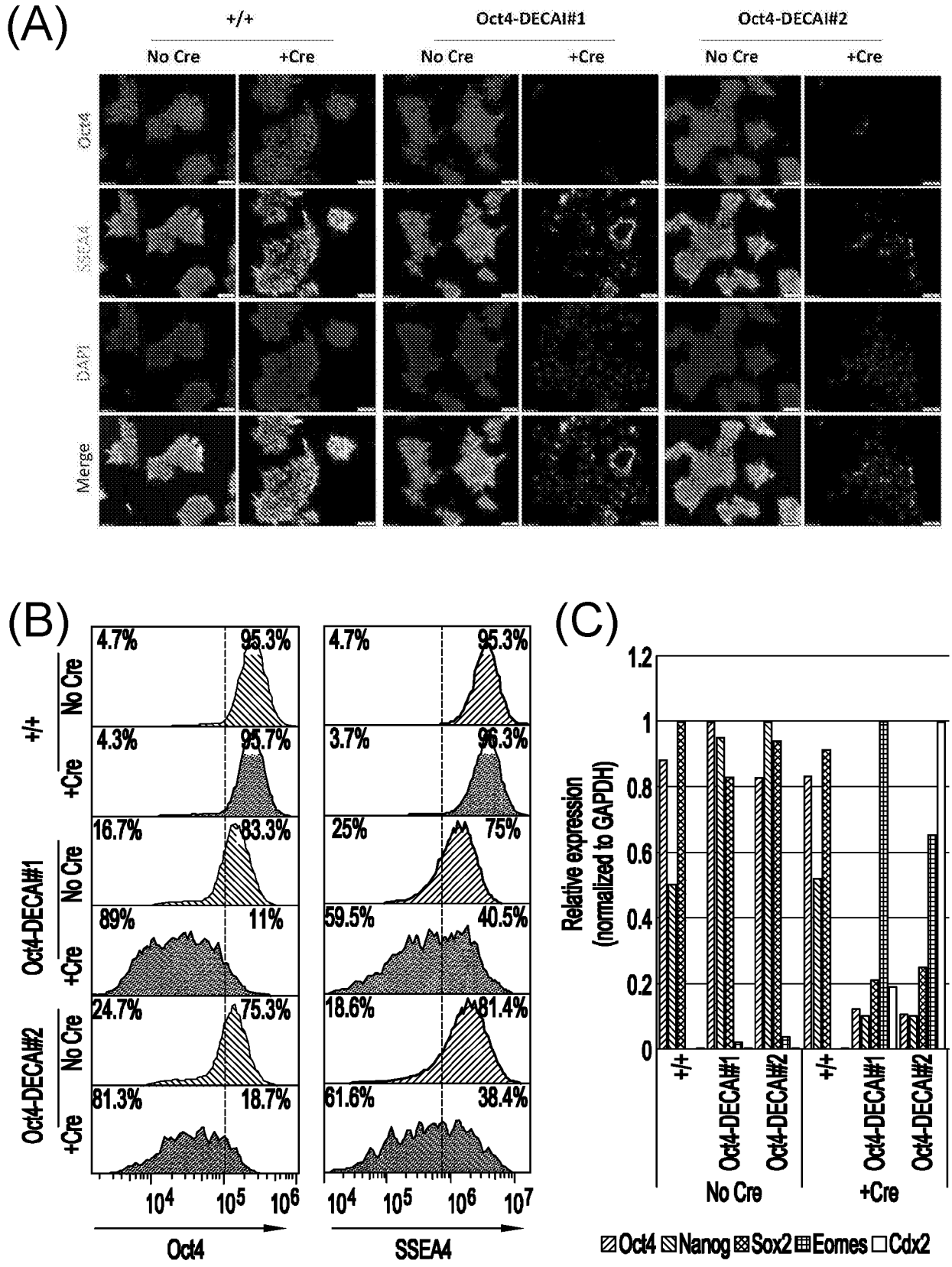

(A)
+/+
Oct4-DECAI#1
Oct4-DECAI#2
No Cre
+Cre
No Cre
+Cre
No Cre
+Cre
Oct4
SSEA4
DAPI
Merge
(B)
+/+
Oct4-DECAI#1
Oct4-DECAI#2
No Cre
+Cre
4.7%
95.3%
4.3%
95.7%
16.7%
83.3%
89%
11%
24.7%
75.3%
81.3%
18.7%
3.7%
96.3%
25%
75%
59.5%
40.5%
18.6%
81.4%
61.6%
38.4%
10^4
10^5
10^6
10^7
Oct4
SSEA4
(C)
Relative expression (normalized to GAPDH)
1.2
1
0.8
0.6
0.4
0.2
0
+/+
Oct4-DECA#1
Oct4-DECA#2
No Cre
+Cre
Oct4 Nanog Sox2 Eomes Cdx2

Fig . 9

CD46-DECAI donor sequence:

TTACCACCACCACCCCCCTGCAGTAGTTTGGTAAAAACTACCAAACAAAATTTTAAAAAGCAGCCAGTGAGTGGGGAAAACAACAAAAACCCGTATTACATTTAG
GAGGGCAGGGACAAGTAAGAATGACAGCCAGCCAAGTTTTCTTAAGAAAATGTGGTTACAATATGGTTATCTCTGTTAGTGCTGAAAGAAAAATAAAACACCCTAGA
ATTCTGTATGAAGGAGAATATATTAGTTAACATTTAAAGTCTTTTAAAAAATGTTTTTAACATCTTGCATTCCATTCCTTGTCTCTGTTCACACTGGAAATTACTACTTTGT
ACTACTTTTTCTGCTAAAGCAGATATCCATTAATTCTGAGGTTTCTCTAATTTTCCAGTGGTCAAATGTCGATTTCCAGTAGTCGAAAATGGAAAACAGATATCAGGATT
TGGAAAAAAATTTTACTACAAAGCAACAGTTATGTTTGAATGCGATAAGGTAAGTAATAACTTCGTATAGCATACATTATACGAAGTTATTCAAGGTTAGAAGACAGG
TTTAAGGAGACCAATAGAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCATAACTTCGTATAG
CATACATTATACGAAGTTATTTTCTCTCCACAGGGTTTTTACCTCGATGGCAGCGACACAATTGTCTGTGACAGTAACAGTACTTGGGATCCCCCAGTTCCAAAGTGTCT
TAAAGGTACAAAGGTTATCTTTTTTCTGTCTTGGTTTGTTATTGTTGTTGCTGTTCATTTTAGACTTTATTTCTTTGATATTAACTATCAGTCATACAAAATAACTGAAAAG
AAACAATTTTAGTATTTAACTCTGTCTTGTATTCATTTCTATGCCAGATGAATGACACGAAATTCACATAAAATTCTGCTGTTGTGATTTTTTGTGCTTTTCCAGGGTTCTT
AGCACGTTATGTACATTGCATGGGTATATGCTTTTAATATTTTTATGTATAAAAAGTGAATTACAACAACTTTTTGGAATTGAAACATGGGCATTTTTATCTAAGTAAGT
CAACAATGGCATAATTCATATAAATGAAATGAGAGCAATAACTCCCAAGTGGTTGATCTTCTAACATTATTTTGTTTCCTAGTGCTGCCTCCATCTAGTACAAAACCTCC
AGCTTTGAGTCATTCAGGTTTAGTAGCTTCTTCCTTATATGTCTTCTTCCTTATATGTTACAA (SEQ ID NO: 39)

METTL16-DECAI donor sequence:

ACTTGTTCAAAGTCACACAGTGAGGAGACGACGAGACAGGAATTTGAGCCGAGTCATTTAATCTCTGTGCTCCATATACGGATATGCAGTGTTTTTGGTGGAGGATTG
TACTGATTTTCATAAGCTGTGACTGTGTGAAACACTCGTGTATGTGCCACTGGGATGAGTGCTCTGTGTAACTGTGTGTGATCTCCCTGGGAGGTTCGGATGGGTATTT
GGTTCTGCCTGTTTGCCGTAGAGCTTGTACAACATAGTGAATCTTACTAATCAGTTTTACTTCCTTTTCTTTATTCCTCAGCCTTAATTTTAAAGACCCCGAAGCAGTCAG
AGCTCTGACGTGTACTCTCCTAAGGTAAGTAATAACTTCGTATAGCATACATTATACGAAGTTATTCAAGGTTAGAAGACAGGTTTAAGGAGACCAATAGAAACTGGG
CTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGCCATAACTTCGTATAGCATACATTATACGAAGTTATTTTCTCT
CCACAGGGAAGATTTTGGACTTTCTATTGATATTCCATTGGAGAGACTAATTCCCACAGTTCCCTTGAGACTCAACTATATTCACTGGGTAGAAGATCTGATCGGTCAC
CAGGATTCTGACAAAAGTACTCTCCGAAGAGGAATTGACATAGGTATATCATTTTAAATTCTTTTTTGGCTAAACAGTTTTCATAAGTTTTGCGAGATCAAATTCTTTGT
AAACTTTTCTACTGGGTATTTGTTGTTGTACAAAAGAAGCTGGCTACAGATACAGGTTGAGCATCCTTAATCCGAAAATCCGAAATGCTTCAAAATCTGCAACCTTTGA
GCACTGATGTGTTGCTCAAAGGCAATGTGCGTTGGAGCGTTCCAGATTTGGGGGTTAGGGATGTTCAGCCTGTAAATATAGTGCACATATTTCAAAATCTGAAAAAAA
AATTGAAATCCAAAACA (SEQ ID NO: 40)

Fig. 10

Oct4-DECAI#1 donor sequence:

CCTTCGCCTCAGTTTCTCCCCCCACCTCCCTCTCCTCCACCCATCCAGGGGGCGGGGCCAGAGGTCAAGGCTAGTGGGTGGGACTGGGGAGGGAGAGAGGGGTTGAG
TAGTCCCTTCGCAAGCCCTCATTTCACCAGGCCCCCGGCTTGGGGCGCCTTCCTTCCCCATGGCGGGACACCTGGCTTCGGATTTCGCCTTCTCGCCCCCTCCAGGTGGT
GGAGGTGATGGGCCAGGGGGGCCGGAGCCGGGCTGGGTTGATCCTCGGACCTGGCTAAGCTTCCAAGGCCCTCCTGGAGGGCCAGGAATCGGGCCGGGGGTTGGG
CCAGGCTCTGAGGTGTGGGGGATTCCCCCATGCCCCCCGCCGTATGAGTTCTGTGGGGGGATGGCGTACTGTGGGCCCCAGGTAAGTAATAACTTCGTATAGCATAC
ATTATACGAAGTTATTCAAGGTTAGAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTC
TTACTGACATCCACTTTGCCATAACTTCGTATAGCATACATTATACGAAGTTATTTTCTCTCCACAGGTTGGAGTGGGGCTAGTGCCCCAAGGCGGCTTGGAGACCTCTC
AGCCTGAGGGCGAAGCAGGAGTCGGGGGTGGAGAGCAACTCCGATGGGGCCTCCCCGGAGCCCTGCACCGTCACCCCTGGTGCCGTGAAGCTGGAGAAGGAGAAGC
TGGAGCAAAACCCGGAGGAGGCAAGTGAGCTTCGACGGGGTTGGGGTGTGGGGAGGTGGTCATGACAGGGCAGCCTGATGGGGAAGTGGTCACCTGCAGCTGCCC
AGACCTGGCACCCAGGAGAGGAGCAGGCAGGGTCAGCTGCCCTGGCCAGGGAGGGGTGTGTATCAACTGCTGGCAGCCCTGGCAGGCAGGGGCCAGGTGGGAAGT
GGAAGCTGGATTTCGAAGAGACAACTGCCGGTGAGGGCAGAGC (SEQ ID NO: 41)

Oct4-DECAI#2 donor sequence:

GAGGCTACATTGAGCCATCATTGTACTCCACTGCACTCCAGTCTGGGCAACAAAGTGAGACCCTGTCTTAAAAAATAAAAATAAAAAAAGTTTCTGTGGGGGACCTGC
ACTGAGGTCCTGGAGGGGCGCCAGTTGTGTCTCCCGGTTTTCCCCTTCCACAGACACCATTGCCACCACCATTAGGCAAACATCCTTCGCCTCAGTTTCTCCCCCCACCT
CCCTCTCCTCCACCCATCCAGGGGGCGGGGCCAGAGGTCAAGGCTAGTGGGTGGGACTGGGGAGGGAGAGAGGGGTTGAGTAGTCCCTTCGCAAGCCCTCATTTCA
CCAGGCCCCCGGCTTGGGGCGCCTTCCTTCCCCATGGCGGGACACCTGGCTTCGGATTTCGCCTTCTCGCCCCCTCCAGGTAAGTAATAACTTCGTATAGCATACATTAT
ACGAAGTTATTCAAGGTTAGAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACT
GACATCCACTTTGCCATAACTTCGTATAGCATACATTATACGAAGTTATTTTCTCTCCACAGGTGGTGGAGGTGATGGGCCAGGGGGGCCGGAGCCGGGCTGGGTTGA
TCCTCGGACCTGGCTAAGCTTCCAAGGCCCTCCTGGAGGGCCAGGAATCGGGCCGGGGGTTGGGCCAGGCTCTGAGGTGTGGGGGATTCCCCCATGCCCCCCGCCGT
ATGAGTTCTGTGGGGGGATGGCGTACTGTGGGCCCCAGGTTGGAGTGGGGCTAGTGCCCCAAGGCGGCTTGGAGACCTCTCAGCCTGAGGGCGAAGCAGGAGTCG
GGGTGGAGAGCAACTCCGATGGGGCCTCCCCGGAGCCCTGCACCGTCACCCCTGGTGCCGTGAAGCTGGAGAAGGAGAAGCTGGAGCAAAACCCGGAGGAGGCAA
GTGAGCTTCGACGGGGTTGGGGTGTGGGGAGGTGGTCA (SEQ ID NO: 42)

Fig. 10
(continued)

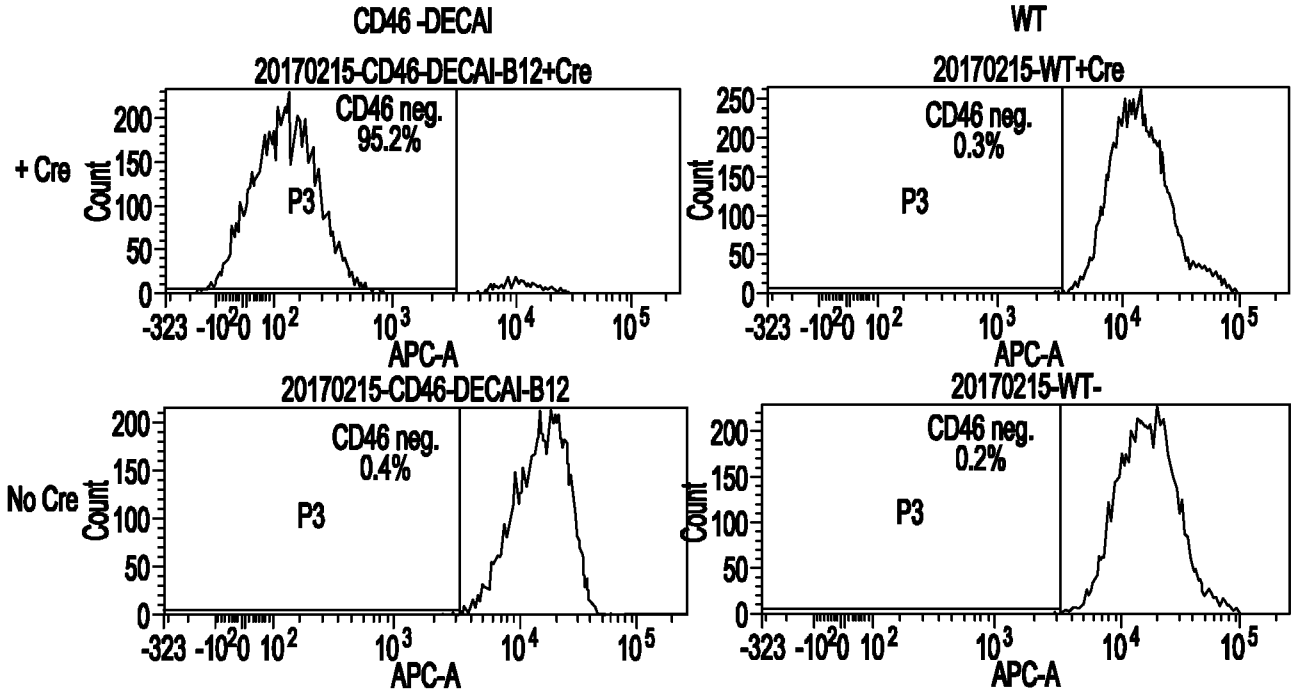

CD46 -DECAI
WT
20170215-CD46-DECAI-B12+Cre
20170215-WT+Cre
+ Cre
No Cre
CD46 neg.
95.2%
CD46 neg.
0.3%
CD46 neg.
0.4%
CD46 neg.
0.2%
P3
Count
APC-A
20170215-CD46-DECAI-B12
20170215-WT-

Fig . 11

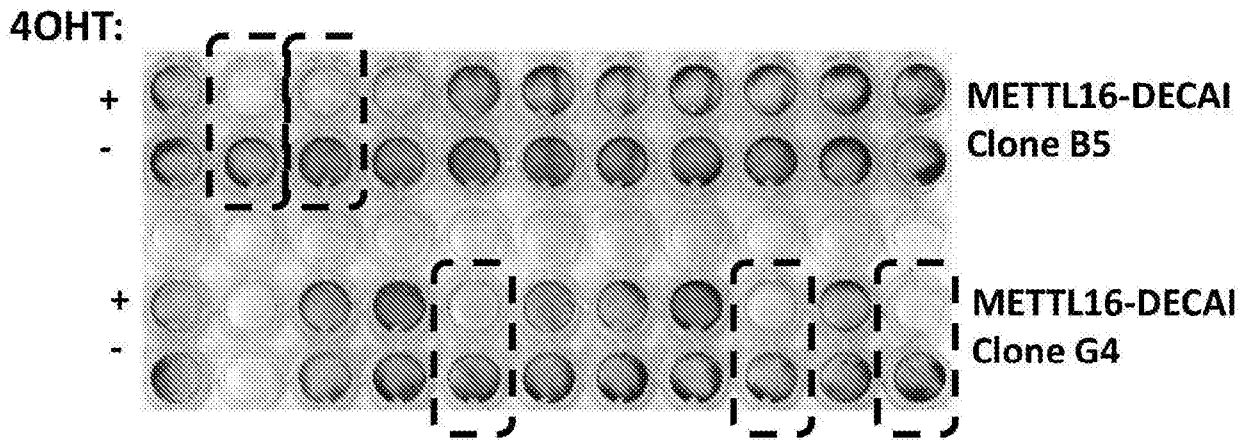

4OHT:
+
-
METTL16-DECAI
Clone B5
+
-
METTL16-DECAI
Clone G4

Fig . 12

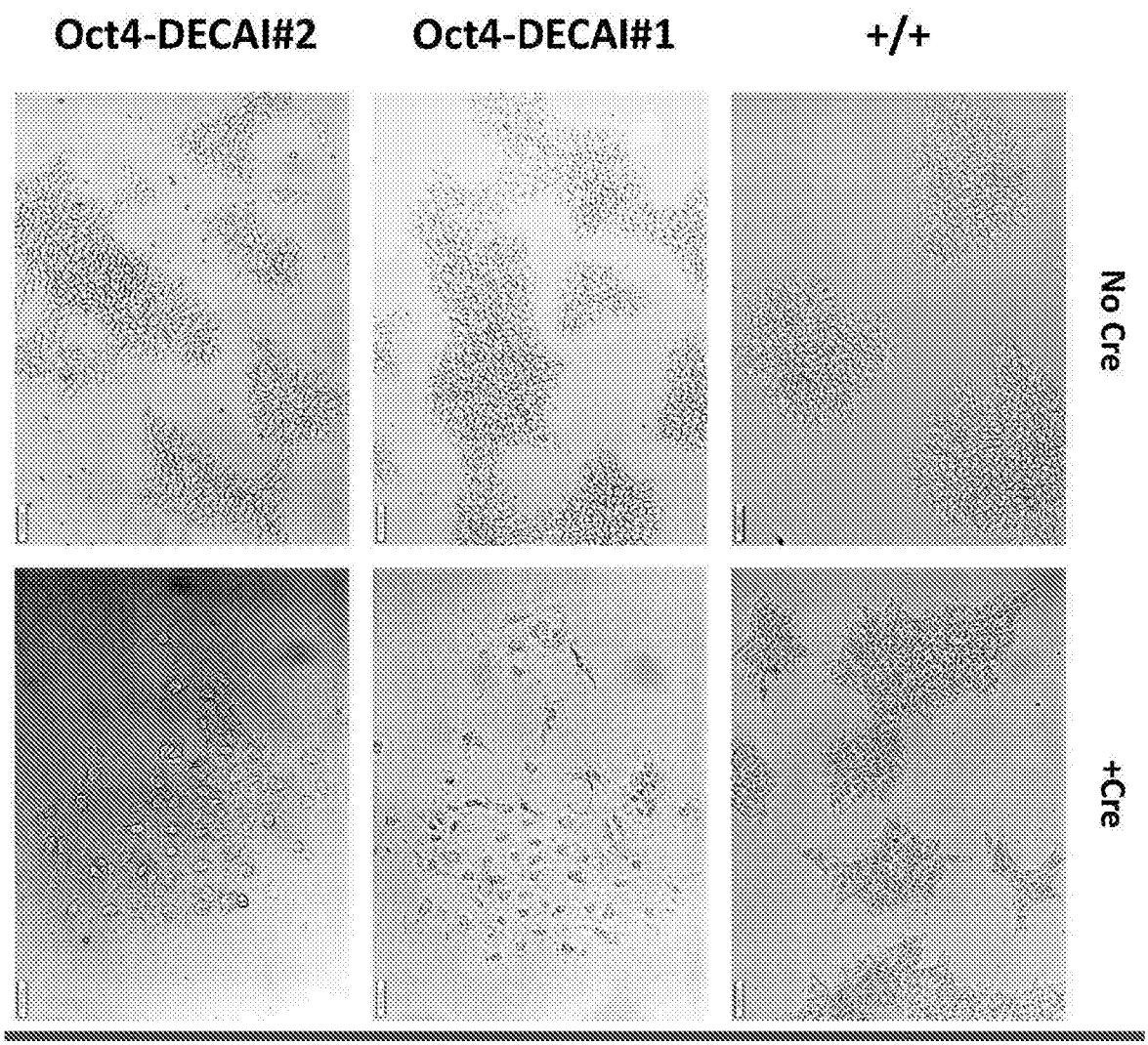
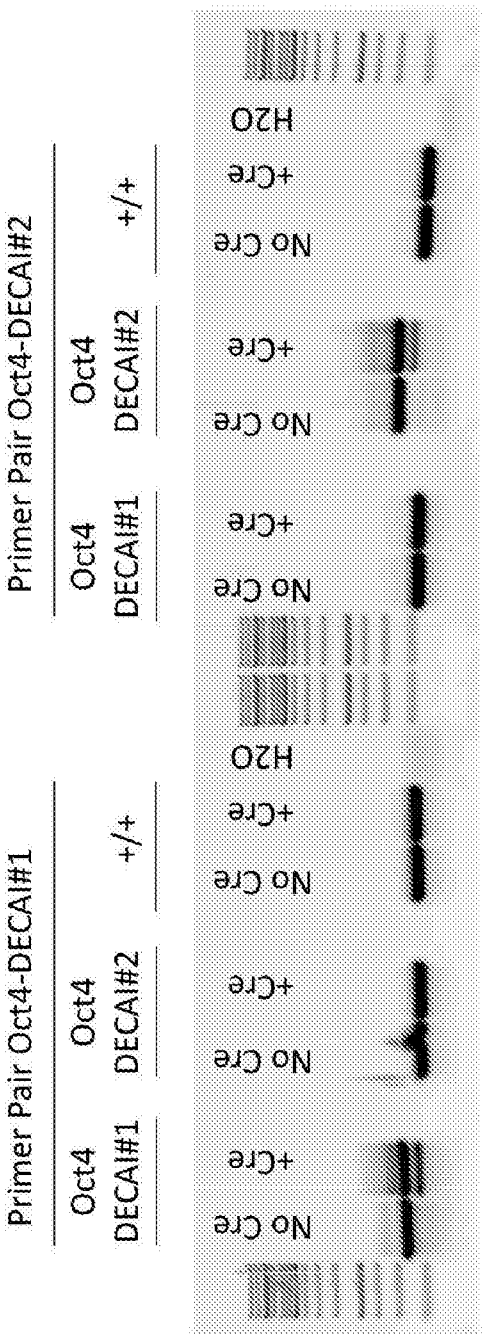


Fig . 13



Oct4 Genotyping	gRNA	FW primer	RV primer
gRNA site1 (for DECAI#1)	GCACUAGCCCCACUCCAACC (SEQ ID NO: 27)	CTCTGAGGTGTGGGGGATT (SEQ ID NO: 31)	TGCTCCAGCTTCTCCTTCTC (SEQ ID NO: 32)
gRNA site2 (for DECAI#2)	ACCACCUGGAGGGGGCGAGA (SEQ ID NO: 28)	GACACCTGGCTTCGGATTT (SEQ ID NO: 33)	CCCCACAGAACTCATACGG (SEQ ID NO: 34)

Fig . 14

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2017/053546

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/63 C12N15/85 C12N15/90
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, CAB Data, Sequence Search, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Amanda Andersson-Rolf et al.: "Rapid, One-step generation of biallelic conditional gene knockouts", , 1 June 2016 (2016-06-01), pages 1-17, XP002777980, Retrieved from the Internet: URL:https://www.biorxiv.org/content/biorxiv/early/2016/06/01/056549.full.pdf [retrieved on 2018-02-06]	19-21
A	the whole document	1-18,22, 23

X	WO 2005/001087 A2 (REGENERON PHARMA [US]; ECONOMIDES ARIS N [US]; VALENZUELA DAVID M [US]) 6 January 2005 (2005-01-06)	19-21
A	the whole document	1-18,22, 23

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 February 2018	27/02/2018

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Hornig, Horst

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2017/053546

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:

      [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2011/059799 A1 (REGENERON PHARMA [US]; ECONOMIDES ARIS N [US]; MURPHY ANDREW J [US]; L) 19 May 2011 (2011-05-19)	19-21
A	the whole document	1-18,22, 23

X	A. N. ECONOMIDES ET AL: "Conditionals by inversion provide a universal method for the generation of conditional alleles", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 110, no. 34, 5 August 2013 (2013-08-05), pages E3179-E3188, XP055155451, ISSN: 0027-8424, DOI: 10.1073/pnas.1217812110	19-21
A	the whole document	1-18,22, 23

X	WO 02/40685 A2 (CORNELL RES FOUNDATION INC [US]; XIN HONG BO [US]; KOTLIKOFF MICHAEL [) 23 May 2002 (2002-05-23)	19-21
A	the whole document	1-18,22, 23

X	WO 2006/048466 A2 (GENOWAY [FR]; THIAM KADER [FR]; FRAICHARD ALEXANDRE [FR]) 11 May 2006 (2006-05-11)	19-21
A	the whole document	1-18,22, 23

X	WO 2006/048464 A2 (GENOWAY [FR]; THIAM KADER [FR]; FRAICHARD ALEXANDRE [FR]) 11 May 2006 (2006-05-11)	19-21
A	the whole document	1-18,22, 23

X	SKARNES WILLIAM C ET AL: "A conditional knockout resource for the genome-wide study of mouse gene function", NATURE, MACMILLAN JOURNALS LTD., ETC, vol. 474, no. 7351, 1 June 2011 (2011-06-01), pages 337-342, XP009157945, ISSN: 0028-0836, DOI: 10.1038/NATURE10163 [retrieved on 2011-06-15] cited in the application	19-21
A	the whole document	1-18,22, 23

X	WO 02/083909 A2 (MULTIGEN HOLDING S A [LU]; JANSEN BASTIAAN JAN HERMAN [NL]; VAN STEENS) 24 October 2002 (2002-10-24)	19-21
A	the whole document	1-18,22, 23

	-/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2017/053546

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	GUZZARDO PALOMA M ET AL: "A small cassette enables conditional gene inactivation by CRISPR/Cas9.", SCIENTIFIC REPORTS, vol. 7, no. 1, 16770, 1 December 2017 (2017-12-01), pages 1-11, XP002777981, ISSN: 2045-2322 the whole document -----	1-23

5

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2017/053546

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2005001087 A2	06-01-2005	US 2005003543 A1	06-01-2005
		US 2007275466 A1	29-11-2007
		US 2014155689 A1	05-06-2014
		WO 2005001087 A2	06-01-2005
WO 2011059799 A1	19-05-2011	AU 2010319894 A1	31-05-2012
		CA 2779858 A1	19-05-2011
		CN 102666854 A	12-09-2012
		CN 106191126 A	07-12-2016
		CY 1118659 T1	12-07-2017
		DK 2494047 T3	10-04-2017
		EP 2494047 A1	05-09-2012
		EP 3147362 A1	29-03-2017
		ES 2613662 T3	25-05-2017
		JP 5908405 B2	26-04-2016
		JP 6085696 B2	22-02-2017
		JP 2013509192 A	14-03-2013
		JP 2016063839 A	28-04-2016
		JP 2016185166 A	27-10-2016
		PL 2494047 T3	31-05-2017
		PT 2494047 T	08-02-2017
		US 2011104799 A1	05-05-2011
		US 2013302899 A1	14-11-2013
		WO 2011059799 A1	19-05-2011
WO 0240685 A2	23-05-2002	AU 1984102 A	27-05-2002
		US 2004077089 A1	22-04-2004
		WO 0240685 A2	23-05-2002
WO 2006048466 A2	11-05-2006	EP 1814998 A2	08-08-2007
		FR 2877675 A1	12-05-2006
		FR 2877676 A1	12-05-2006
		WO 2006048464 A2	11-05-2006
		WO 2006048466 A2	11-05-2006
WO 2006048464 A2	11-05-2006	EP 1814998 A2	08-08-2007
		FR 2877675 A1	12-05-2006
		FR 2877676 A1	12-05-2006
		WO 2006048464 A2	11-05-2006
		WO 2006048466 A2	11-05-2006
WO 02083909 A2	24-10-2002	AU 2002307630 A1	28-10-2002
		NL 1017852 C2	29-10-2002
		WO 02083909 A2	24-10-2002