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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 December 2014 (04.12.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/191527 A1

(51) International Patent Classification:
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/EP2014/061189

(22) International Filing Date:
28 May 2014 (28.05.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PA 2013 70303 31 May 2013 (31.05.2013) DK

(71) Applicants: **CELLECTIS** [FR/FR]; 8 Rue de la Croix Jarry, F-75013 Paris (FR). **PREGENEN INC.** [US/US]; 454 North 34th Street, Seattle, Washington 98103 (US).

(72) Inventors: **ASTRAKHAN, Alexander**; 8626 31st Ave SW, Seattle, Washington 98126 (US). **JARJOUR, Jordan**; 7351 30th Ave SW, Seattle, Washington 98126 (US).

(74) Agent: **ZACCO DENMARK A/S**; Hans Bekkevolds Allé 7, DK-2900 Hellerup (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/191527 A1

(54) Title: A LAGLIDADG HOMING ENDONUCLEASE CLEAVING THE T CELL RECEPTOR ALPHA GENE AND USES THEREOF

(57) Abstract: Disclosed herein are compositions for inactivating the human TCR-alpha gene comprising engineered LAGLIDADG homing endonucleases (LHEs) and their derivatives, particularly derived from members of the \-Onul subfamily of LHEs. Polynucleotides encoding such endonucleases, vectors comprising said polynucleotides, cells comprising or having been treated with such endonucleases, and therapeutic compositions deriving therefore are also provided.

A LAGLIDADG HOMING ENDONUCLEASE CLEAVING THE T-CELL RECEPTOR ALPHA GENE AND USES THEREOF

FIELD OF THE INVENTION

The present disclosure relates to molecular and cellular biology, genetics, genomics, and their applications in human therapeutics. Particular aspects relate to a rare-cutting endonuclease cleaving a nucleic acid target sequence from the TCR-alpha gene, more particularly to a new meganuclease variant of I-*Onu*I or homologues that is particularly efficient in disrupting the expression of this gene in T-cells, and the use thereof for cancer therapy.

BACKGROUND OF THE INVENTION

Site-specific nucleases are powerful reagents for specifically and efficiently targeting and modifying a DNA sequence within a complex genome. The double-stranded DNA breaks caused by site-specific nucleases are commonly repaired through the distinct mechanisms of homologous recombination or non-homologous end joining (NHEJ). Although homologous recombination typically uses the sister chromatid of the damaged DNA as a donor matrix from which to perform perfect repair of the genetic lesion, NHEJ is an imperfect repair process that often results in changes to the DNA sequence at the site of the double strand break. Mechanisms involve rejoining of what remains of the two DNA ends through direct re-ligation (Critchlow and Jackson 1998) or via the so-called microhomology-mediated end joining (Ma, Kim et al. 2003). Repair via non-homologous end joining (NHEJ) often results in small insertions or deletions and can be used for the creation of specific gene knockouts. There are numerous applications of genome engineering by site-specific nucleases extending from basic research to bioindustrial applications and human therapeutics. Re-engineering a DNA-binding protein for this purpose has been mainly limited to the naturally occurring LADLIDADG homing endonuclease (LHE), artificial zinc finger proteins (ZFP), the Transcription Activator Like Effectors nucleases (TALE-nucleases), and the recently described CRISPR-Cas system.

Homing endonucleases, also known as meganucleases, are sequence-specific endonucleases with large (>14 bp) cleavage sites that can deliver DNA double-strand breaks at

specific loci (Thierry and Dujon 1992). There are a handful of known homing endonuclease families which are demarcated on the basis of canonical motifs and the structural features which comprise them. However, they all share the property of recognizing and cleaving long DNA targets. Homing endonucleases were the first, and to date only, naturally occurring endonucleases with specificities at or approaching 'genome level', meaning having putative target sequences that occur very infrequently, or perhaps singularly, in their host genome. As a general property, HEs have a moderate degree of fidelity to their DNA target sequences, such that most base pair substitutions to their DNA target sequences reduce or eliminate the ability of the HE to bind or cleave it. HEs are therefore the most specific naturally occurring endonucleases yet discovered, and indeed this property is critical to the natural life cycle of the genetic elements in which they are encoded.

Homing endonuclease genes (HEGs) are classified as a type of selfish genetic element, as their DNA recognition and cleavage activity can lead to a DNA repair event that results in the copying of the HEG into the cleavage site. This mechanism of horizontal gene transfer, referred to as 'homing' results in a super-Mendelian inheritance pattern. Using this mechanism, HEGs and their endonuclease gene products can spread rapidly within their host species populations, and have also spread throughout all kingdoms of life over evolutionary time. HEGs are most commonly found in highly conserved genomic locations that do not impart fitness costs on their host organisms, such as within introns or as non-disruptive N- or C-terminal fusions to host proteins.

The LAGLIDADG homing endonuclease family (LHE) comprises a group of compact (< 320 amino acids) nucleases whose structural and mechanistic properties have been studied extensively owing to their attractive properties for genome engineering applications. LHEs operate either as dimers or as pseudo-dimeric monomers, with the DNA cleaving active site occurring at the DNA-facing end of the interface of the two subunits (in dimeric LHEs) or domains (in monomeric LHEs). The LAGLIDADG consensus motifs for which LHEs are named are found in the two central alpha helices which form this interface between the two subunits or domains. At the bottom of each LAGLIDADG helix are the residues which together coordinate the hydrolysis reaction if the appropriate conditions are met, such as if the LHE finds and binds to an appropriate DNA target sequence. The active site covers the 'central-4' DNA bases of the DNA target sequence.

On either side of the active site are the two DNA binding domains LHEs use to recognize their DNA target sequences. Each domain comprises an anti-parallel beta sheet which wraps around nearly a complete turn of DNA and contacts 9 base pairs of DNA sequence. Members of the LHE family thus recognize 22 base pair DNA target sequences (9 base pairs for each domain, and 4 base pairs covered by the active site), which are partially palindromic in the case of dimeric LHEs, but can be entirely asymmetric for monomeric LHEs. Emanating from each anti-parallel beta sheet are the amino acid side chains which comprise the DNA recognition interface. While there is much amino acid conservation throughout the non-DNA interfacing residues amongst the LHE family, DNA recognition interface amino acid compositions vary significantly. This is because for each LHE the DNA recognition interface comprises an extensive network of side chain-to-side chain and side chain-to-DNA contacts, most of which is necessarily unique to a particular LHE's DNA target sequence. The amino acid composition of the DNA recognition interface (and the correspondence of it to a particular DNA sequence) is therefore the definitive feature of any natural or engineered LHE. The DNA recognition interface functions in determining the identity of the DNA target sequence which can be accommodated and hydrolyzed and also the affinity and specificity properties which define the quality of the LHE according to the demands of the application.

Owing to their small size and exquisite specificity properties, LHEs have been the subject of numerous efforts to engineer their DNA recognition properties with the desired outcome of cleaving and altering genes of interest in research, biotechnology, crop science, global health, and human therapeutics applications. However, the extent of the networks of residues which form the DNA recognition interface has generally prevented efficient methods for re-addressing LHEs to DNA target sequences of interest. This has led to continued innovation in field of gene-specific nuclease engineering, with three endonuclease alternative platforms now validated as having the capacity to target DNA sequences with ranging (but generally high) levels of specificity, as well as new and improved methods for overcoming the challenges of engineering the DNA recognition interfaces of LHEs.

Zinc finger nucleases (ZFNs) generating by fusing a plurality of Zinc finger-based DNA binding domains to an independent catalytic domain (Kim, Cha et al. 1996; Smith, Berg et al. 1999; Smith, Bibikova et al. 2000) represent another type of engineered nuclease commonly used to stimulate gene targeting and have been successfully used to induce gene correction, gene insertion and gene deletion in research and therapeutic applications. The archetypal ZFNs are based on the catalytic domain of the Type IIS restriction enzyme FokI and Zinc Finger-based DNA

binding domains made of strings of 3 or 4 individual Zinc Fingers, each recognizing a DNA triplet (Pabo, Peisach et al. 2001). Two Zinc Finger-FokI monomers have to bind to their respective Zinc Finger DNA-recognition sites on opposite strands in an inverted orientation in order to form a catalytically active dimer that catalyze double strand cleavage (Bitinaite, Wah et al. 1998).

Transcription activator-like effectors (TALEs) were the next artificial endonuclease platform. TALEs derived from a family of proteins used in the infection process by plant pathogens of the Xanthomonas or Ralstonia genus are repetitive proteins characterized by 14-20 repeats of 33-35 amino acids differing essentially by two positions. Each base pair in the DNA target is contacted by a single repeat, with the specificity resulting from the two variant amino acids of the repeat (the so-called repeat variable dipeptide, RVD). The apparent modularity of these DNA binding domains has been confirmed to a certain extent by modular assembly of designed TALE-derived protein with new specificities (Boch, Scholze et al. 2009; Moscou and Bogdanove 2009). Very similarly to ZFNs, TALEs were readily adapted into site-specific nucleases by arraying TALE repeats with RVDs corresponding to the target sequence of choice and fusing the resultant array to a FokI domain. As such, DNA cleavage by a TALE-Nuclease requires two DNA recognition regions flanking an unspecific central region. TALE nucleases have proliferated widely since 2010 owing to their ease of production and improved double-strand break generating efficiency.

Of these distinct technologies, it is important to distinguish the advantaged properties of each and to determine innovative ways to capture these properties for the appropriate genome engineering applications. One of the most powerful applications of site-specific nuclease technology is in the field of human therapeutics, which requires the use of highly efficient and specific nuclease reagents to safely and effectively edit genomic information in human cells or tissues. A one prominent example is the cancer immunotherapy field, which is at the forefront of applying nuclease technological advances for developing novel therapeutics. These approaches are focused on harnessing the powerful anti-tumor activities of patient-derived (autologous) or donor-derived (allogeneic) T-cells and leveraging this potential via genome engineering of cell-intrinsic properties such as cellular proliferation, engraftment, migration or longevity. Successful and scalable manufacture of T-cells endowed with enhanced anti-cancer activity requires generation of highly efficient nuclease compositions and simplified delivery strategies, such as those described in some aspects of this application.

The immune system has a key foundational role in detecting and preventing the development of human cancer. The majority of transformed cells are quickly detected by immune

sentinels and destroyed through the activation of antigen-specific T-cells via clonally expressed T-cell receptors (TCR). Oncogenesis is thus an immunological disorder, a failure of immune system to mount the necessary anti-tumor response to durably suppress and eliminate the disease. Certain immunotherapy interventions developed over the last few decades, such as recombinant cytokine infusions, have specifically focused enhancing T-cell immunity, and while these have been associated with sporadic cases of disease remission, they have not had substantial overall success. Recent therapies with monoclonal antibodies targeting molecules which inhibit T-cell activation, such as CTLA-4 or PD-1, have shown a more substantial anti-tumor effect, however these treatments are associated with substantial toxicity due to systemic immune activation. Most recently, therapeutic strategies which are based on the isolation, modification, expansion and reinfusion of T-cells have been explored and tested in early stage clinical trials. These treatments have shown mixed rates of success, but a number of patients have experienced unprecedented objective responses and durable remissions, highlighting the potential for T-cell based cancer immunotherapies. Genome editing strategies which are designed to harness this potential for successful widespread implementation of T-cell cancer immunotherapies are described herein.

Successful recognition of tumor cell associated antigens by cytolytic T-cells initiates targeted tumor lysis and underpins any effective cancer immunotherapy approach. Some tumors contain tumor-infilitrating T-cells (TILs) which express TCRs specifically directed tumor-associated antigens; however access to substantial numbers of TILs is limited to only a few human cancers. In response to this limitation, artificial antigen recognition and signaling transgenes called chimeric antigen receptors (CARs) have been devised to broaden the scope and utility of T-cell based cancer immunotherapy. CARs are transmembrane spanning proteins whose extracellular portions contain antigen recognition domains most typically derived from single-chain variable fragments (scFv) of monoclonal antibodies, and whose intracellular domains contain combinations of signaling domains to mimic TCR-like activation signals. It has been widely demonstrated that primary human T-cells made to express CARs are able to respond to and kill cells which bear the antigen recognized by the scFv domain.

Despite highly promising initial results with CAR-expressing transgenic T-cells, the efficacy, safety and scalability of CAR-based T-cell immunotherapies is limited by continuous expression of clonally derived TCR. Residual TCR expression may interfere with CAR signaling in engineered T-cells or it may initiate off-target and pathologic responses to self- or allo-antigens. Consequently, CAR-based T-cells have only been used in autologous applications. Genetic abolition of

endogenous TCR through nuclease-mediated gene editing would reduce the risk of damaging collateral responses and decrease the potential for T-cell mediated graft vs. host disease (GVHD). The main hurdle for developing universal allogeneic T-cell therapy is the development of GVHD through the activation of donor T cells' TCR by the recipients' HLA complex. Removal of the TCR would prevent such graft-versus-host responses and enable the development of simple and widely applicable allogeneic T-cell therapies.

In addition to cancer, T-cell therapies are being developed for a wide range of therapeutic applications including chronic viral infections, autoimmune disease and stem cell transplantation. In disease models and initial clinical models have shown a key role for the regulatory T cell subset (T-regs) in controlling the development and extent of GVHD and various autoimmune diseases. Transfer of regulatory T cells ameliorates GVHD in patients receiving stem cell transplant. In addition, transfer of regulatory T cells improved disease outcome in preclinical models of rheumatoid arthritis, type-1 diabetes and systemic lupus erythematosus, amongst others. This approach is also being tested in patients with chronic viral infections such as Hepatitis B (HBV). Engineered T cells containing HBV-specific CARs are highly active against HBV-infected cells. These approaches are being tested in clinical trials, however their use is limited by the same manufacturing and scalability hurdles associated with other autologous therapies. Combining genetic targeting of TCR-alpha in T cells with CARs targeting tolerance or viral targets represents a very powerful way to develop allogeneic T cell therapy for the treatment of human disease.

SUMMARY OF THE INVENTION

A genome engineering strategy to generate therapeutic T-cell products, such as in the treatment of human diseases, requires the use of safe and effective endonuclease reagents for disrupting TCR-alpha gene. The endonuclease I-*Onu*I, encoded within a group I intron in the Rps3 host gene from *Ophiostoma novo-ulmi* subsp *americana*, and its closely related homologs, have been recently characterized to be monomeric proteins displaying the characteristics of the LAGLIDADG homing endonucleases and to be sufficiently active for use in genome editing (WO2011/156430, (Sethuraman, Majer et al. 2009; Takeuchi, Lambert et al. 2011)).

In particular aspects, several I-*Onu*I variants were created in an attempt to target different DNA sequences in the TCR-alpha gene. In additional aspects, new LHE variant targeting the constant domain of the TCR-alpha gene are provided. This particular I-*Onu*I variant showed high efficiency in disrupting the expression of TCR-alpha in T-cells. In further aspects, this

particular variant of the invention were then fused to some engineered nucleic acid binding domains, so as to form chimeric endonucleases that also showed improved properties, especially increases in specificity and efficiency which are required for obtaining safe and useful reagents for treating primary human cells. These molecules have proven efficiency for genome editing at the TCR-alpha locus and will be useful in numerous T-cell based methods for treating human disease.

BRIEF DESCRIPTION OF THE TABLES AND THE FIGURES

For a better understanding of the invention and to show how the same may be carried into effect, there will now be shown by way of example only, specific embodiments, methods and processes according to the present invention with reference to the accompanying drawings in which:

Table 1 shows the positions of the amino acid residues in the TCRA_S02-targeting LHE that were varied or otherwise became altered relative to the primary sequence of the wild-type I-*OnuI* protein during the re-specification process. The TCRA_S02_2E5 LHE contains variations only to the 44 residues which comprise the protein-DNA interface, not all of which retuned amino acids different from the wild-type I-*OnuI* protein, but all of which were varied in the initial stages of re-specification. The top performing variant following refinement screening had additional six mutations, 5 of which are located within the protein-DNA interface, 1 of which is elsewhere in the protein.

Figure 1: depicts the location of thirteen putative target sequences, annotated as Onu_S##, in the first exon of the human TRAC gene for which superior LHE-DNA recognition sequences are predicted. The location of the first exon of the TRAC gene is indicated.

Figure 2 shows schematically and structurally the location of the protein-DNA interface that defines the interaction between a LAGLIDADG homing endonuclease (LHE) and its DNA substrate. The schematic illustration generally depicts the concept that there is a continuous region of the LHE that comprises the interaction with DNA. The structural images demonstrate in more detail the nature of this interaction, whereby the protein-DNA interfacial residues of the LHE (whose side-chain atoms are shown as black spheres) interdigitate into the major grooves of DNA helix. It is the constellation of interfacial side chain atoms which determine the complementarity of a natural or engineered LHE to the atoms of the DNA nucleotides, which themselves form sequence specific patterns.

Figure 3 shows that, of the two target sites that were chosen for protein-DNA interface engineering, only one (TCRA_S02) yielded variant LHEs capable of cleaving the full target sequence. One of these variants is shown in comparison with I-*Onu*I enzyme and its cognate target. The panels depict flow cytometric analysis of DNA hydrolysis, whereby the baker's yeast, Saccharomyces cerevisiae, express the LHE on the surface of their cells and are interrogated with fluorescent dye-labelled synthetic DNA substrates as has been published. Briefly, samples are first stained with a biotinylated antibody to an epitope appended to the N- or C-terminus of the LHE. During this staining procedure, conjugates of phycoerythrin-labeled streptavidin (x-axis) with biotin- and Alexa fluor-647 (y-axis) labeled synthetic DNA substrates are generated at a relative molarity that preserves some biotin binding sites on the streptavidin. These pre-conjugates are then used to counter stain the yeast cells, resulting in the co-linear streptavidin-PE/Alexa fluor-647 profile. Cleavage-inhibiting (Ca2+) and cleavage-permitting (Mg2+) conditions are then used to determine whether the native or engineered LHE cleaves the tethered target, which, if cleaved, loses signal in the y-axis owing to loss of the Alexa fluor-647 fluorophore.

Figure 4 shows the initial targeting efficiency of the TCRA_S02_2E5 variant and the progressive improvements in the targeting efficiency achieved by the activity refinement process. Targeting efficiency was measured using a chromosomally integrated double-strand break fluorescent reporter termed the 'traffic light reporter' (TLR). Human embryonic kidney 293T (HEK 293T) fibroblasts were constructed to contain the TCRA_S02 DNA sequence immediately upstream of an out-of-frame mCherry fluorescent protein (y-axis) which, upon one of three possible frame outcomes of the NHEJ DNA repair process becomes fluorescent. The percentage of cells in the y-axis therefore represents approximately 1/3 of all imprecise nuclease-mediated repair events. This cell line was then transfected with synthetically generated in vitro transcribed mRNA (IVT-mRNA) encoding the TCRA_S02 targeting LHEs, with or without mRNA encoding for Trex2 exonuclease. The original TCRA_S02_2E5 variant inefficiently caused double-stranded breaks and therefore produced only small percentages of mCherry positive cells (2^{nd} row). Two rounds of activity refinement screening led to vast improvements in the generation of mCherry positive cells.

Figure 5 shows a comparative alignment of the TCRA_S02_2E5 variant and its derivative, TCRA_S02_RD2_8, which was identified on the second round of activity refinement. The strand-loop-strand motifs which comprise the DNA binding domain are depicted above the aligned sequences.

Figure 6 shows an assay whereby DNA binding titration is used to establish the affinity properties of two different LHEs targeting the TCRA_S02 site. Samples of yeast displaying each LHE variant were independently incubated with increasing concentrations of fluorescent dye-labeled synthetic DNA substrates (y-axis). An antibody to a C-terminal epitope was also included (x-axis) such that DNA binding activity could be normalized to the amount of LHE protein expressed on the yeast surface, creating the co-linear pattern, with higher signal in the y-axis per x-axis signal where affinity is greater. The results demonstrate that the I-*Onu*I LHE (OnuWT) has an approximate Kd of approximately 80 pM and the TCRA_S02_2E5 variant has a Kd of approximately 90 pM.

Figure 7 shows the specificity profiling of the TCRA_S02_2E5_RD2_8 LHE variant using cleavage (top panels) and affinity (bottom panels) analysis. Cleavage and affinity analysis was performed as described in Figure 3 and 4 respectively, however panels of DNA substrates were tested whereby each of the 22 positions along the target was serially altered to each of the 3 non-native base pairs and tested in cleavage-inhibiting and cleavage-permitting conditions. The resulting cleavage and affinity profiles for the 67 different substrates (the target in the TRAC gene and the 66 'one-off' substrates) indicate which LHE variants have most ideal specificity properties and are therefore better candidates for applications demanding tighter specificity, such as in human therapeutics.

Figure 8 shows the flow cytometry scatter properties of primary human T cells which are highly susceptible to double-stranded DNA breaks resulting in genotoxicity and cell death. T cells transfected with IVT-mRNA species encoding an innocuous protein such as the blue fluorescent protein (BFP), show 59% survival during in vitro culture, a level similar to unmanipulated T cells. The TCRA_S02_2E5_RD2_8 variant results in very similar levels of T cell viability, confirming that its global DNA specificity is of high quality.

Figure 9 shows a schematic of a self-inactivating (SIN) lentiviral production plasmid from which lentivirus preparations were generated and used as vectors for the transduction of cell lines and primary cells (A) (SEQ ID NO: 12 and SEQ ID NO: 13) as well as an exemplary non-limiting vector containing the MegaTAL construct used either in lentiviral production or for in vitro transcription for the production of IVT-mRNA (B) (SEQ ID NO: 12). The primary features of the vector, in addition to the lentiviral features well known to those familiar with the art such as the long terminal repeats (LTRs), primer binding site (PBS), central polypurine tract (cPPT), and a T7 RNA promoter (T7) are multicistronic elements for the expression of a TCRA-targeting LHE (as shown)

or MegaTAL linked via a T2A peptide linker motif to the Trex2 exonuclease which further carries an internal ribosomal entry site (IRES) and blue fluorescent protein (BFP) for tracking transduced cells.

Figure 10 shows the complete loss of expression of the TCR-alpha protein from the cell surface of primary human T cells transfected with IVT-mRNA encoding for TCRA_S02_RD2_8 LHE variant in conjunction with IVT-mRNA encoding for accessory Trex2 exonuclease protein. The flow cytometry panels represent different time-point analysis following mRNA transfection (left panel: 72hr, right panel: 14 days). The presence of TCR-alpha protein on the cellular surface is detected via an antibody specific for the CD3 co-receptor molecule. Surface expression of the CD3 co-receptor requires functional expression of the TCR-alpha protein and the absence of a CD3 signal signifies successful disruption of the TRAC gene.

Figure 11 shows gene sequencing data at the TCRA_S02 target site from cells treated with TCRA_S02_RD2_8 LHE variant. Primary T cells transfected with IVT-mRNA encoding for TCRA_S02_RD2_8 LHE variant in conjunction with IVT-mRNA encoding for Trex2 exonuclease were cultured and purified using flow cytometric selection of CD3-negative cells. The sorted CD3-negative cells have deletions at 50% of the TCRA alleles, consistent with the non-silenced TRAC allele being the primary target for TCRA_S02 LHE variants.

Figure 12 depicts schematically one treatment strategy that could be used to generate populations of TCRA-deficient T cells for the treatment of cancer, autoimmunity or chronic viral infection. Briefly, peripheral blood mononuclear cells (PBMCs) would be isolated by apheresis, processed to purify and culture T cells, treated with TRAC-targeting nuclease delivery agents, combined with artificial antigen receptors (CARs) and reinfused into the patient.

Figure 13 shows a schematic representation of the ultra-efficient TRAC gene disruption technology based on the combination of the MegaTAL architecture and Trex2 expression. Also shown is the TCRA_S02 MegaTAL target sequence within the DNA of the TRAC gene, with the location of the 11-mer TALE array indicated both schematically (repetitive units are not annotated but shown upstream of the TCRA_S02 annotation) and its sequence is shown in bold.

Figure 14 demonstrates the extremely high efficiency of TCR-alpha protein removal from the cell surface of primary human T cells transfected with IVT-mRNA encoding for TCRA_S02_MegaTAL LHE variant in conjunction with IVT-mRNA encoding for accessory Trex2 exonuclease protein. The flow cytometry panels demonstrate the TCR complex assembly following electroporation with a

fluorescent protein control (BFP) or TCRA_S02 MegaTAL LHE variant with or without IVT-mRNA encoding for Trex2.

Figure 15 shows efficient TCRA gene inactivation in primary human T cells via electroporation with an mRNA species encoding a three component TAL-LHE-Trex2 fusion protein.

5

DETAILED DISCLOSURE OF THE INVENTION

Unless specifically defined herein, all technical and scientific terms used have the same meaning as commonly understood by a skilled artisan in the fields of gene therapy, biochemistry, genetics, molecular biology and immunology.

10 All methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, with suitable methods and materials being described herein. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will prevail. Further, the materials, methods, and examples 15 are illustrative only and are not intended to be limiting, unless otherwise specified.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Current Protocols in Molecular 20 Biology (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA); Molecular Cloning: A Laboratory Manual, Third Edition, (Sambrook et al, 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Pat. No. 4,683,195; Nucleic Acid Hybridization (B. D. Harries & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. 25 Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the series, Methods In ENZYMOLOGY (J. Abelson and M. Simon, eds.-in-chief, Academic Press, Inc., New York), specifically, Vols.154 and 155 (Wu et al. eds.) and Vol. 185, "Gene Expression Technology" (D. Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); 30 Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C.

Blackwell, eds., 1986); and Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

I-*Onul* and I-*Onul* homologues variants

The present invention relates to rare-cutting endonucleases involving I-*Onul* variants and I-*Onul* homologues variants of I-LtrI, I-LtrWI, I-PanMI, I-PanMII, I-PanMIII, I-GzeI, I-GzeMII, I-GzeMIII, I-GpiI, I-GpeMI, I-AabMI, I-AaeMI, I-ApaMI, I-CpaMI, I-CpaMII, I-CpaMIII, I-CpaMIV, I-CpaMV, I-EjeMI, I-CkaMI, I-CraMI, I-MpeMI, I-MveMI, I-NcrMI, I-OheMI, I-OsoMI, I-OsoMII, I-OsoMIII, I-OsoMIV, I-SmaMI, I-SscMI, I-Vdi141I, I-PnoMI, I-ScuMI; (Takeuchi, Lambert et al. 2011) in which mutations have been introduced and able to that specifically target a nucleic acid sequence present in the TRAC gene.

The rare-cutting endonucleases according to the present invention refer to variant enzymes capable of catalyzing the hydrolysis (cleavage) of bonds between nucleic acids within a DNA or RNA molecule, preferably a DNA molecule. The endonucleases according to the present invention recognize and cleave nucleic acids at specific polynucleotide sequences, further referred to as the “nucleic acid target sequence”.

To engineer rare-cutting endonucleases specific for target sites in the TRAC gene, the inventors constructed libraries of I-*Onul* variants in which amino acid residues localized in the DNA recognition interface of natural I-*Onul* were varied. The libraries were screened for target cleavage activity against each predicted TRAC target sites using previously described cleavage assays (Jarjour, West-Foyle et al. 2009). The specificity of the DNA recognition interface of I-*Onul* was thus altered to target sequences present in the human TRAC gene.

By “variant(s)”, is meant a protein or a polynucleotide encoding thereof that do not naturally exist in nature and that are obtained by genetic engineering or by random mutagenesis. I-*Onul* or I-*Onul* homologue variants according to the invention can for example be obtained by deletion or substitution with a different amino acid of at least one residue in the amino acid sequence of their wild-type sequences. Substitution(s) and deletions can for example be introduced by directed mutagenesis and/or by random mutagenesis. In the frame aspects of the present invention, I-*Onul* or I-*Onul* homologues variants have the ability to target TRAC gene, which mean that they can interact with some specific DNA sequences encoding said gene.

The variants or homologues according to the invention comprise the DNA recognition interface as described herein and as provided in Table 1.

A DNA recognition interface refers to the residues of the protein domains of homing endonuclease or variant thereof which interact with nucleic acid target bases as well as those residues that are adjacent. For each homing endonuclease, the DNA recognition interface comprises an extensive network of side chain-to-side chain and side chain-to-DNA contacts, most of which is necessarily unique to recognize a particular nucleic acid target sequence. Thus, the DNA recognition interface amino acid compositions (and the correspondence of it to a particular nucleic acid sequence) vary significantly and is therefore the definitive feature of any natural or engineered homing endonuclease. The DNA recognition interface determines the identity of the nucleic acid target sequence and also the affinity and specificity properties which define the quality of the homing endonuclease according to the demands of the application.

According to the present invention, the I-*OnuI* or I-*OnuI* homologue variants comprise one or more substitutions in the DNA recognition interface. Accordingly, the I-*OnuI* variant or homologue according to the present invention has at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 97%, more preferably at least 99% sequence identity with the DNA recognition interface of I-*OnuI* (Takeuchi, Lambert et al. 2011).

In a particular embodiment, said I-*OnuI* or I-*OnuI* homologue variants comprise one or more substitution(s) and/or mutations in the DNA recognition interface, particularly in the subdomains situated from positions 24-50, 68 to 82, 180 to 203 and 223 to 240 of I-*OnuI* (SEQ ID NO: 2). The I-*OnuI* variant or homologue can also comprise one or more substitutions at additional positions situated anywhere within the entire I-*OnuI* sequence. The residues which are substituted and/or mutated may include residues contacting the nucleic acid target or interacting with the nucleic acid backbone or with the nucleotide bases, directly or via a water molecule as described in Takeuchi, Lambert et al. 2011.

For example, said I-*OnuI* variant comprises one or more substitutions and/or mutations, preferably at least 10, preferably at least 15, more preferably at least 20, even more preferably at least 25 in at least one position selected from the position group consisting of positions: 19, 24, 26, 28, 30, 32, 34, 35, 36, 37, 38, 40, 42, 44, 46, 48, 68, 70, 72, 75, 76 77, 78, 80, 82, 168, 180, 182, 184, 186, 188, 189, 190, 191, 192, 193, 195, 197, 199, 201, 203, 223, 225, 227, 229, 231, 232, 234, 236, 238, 240 of I-*OnuI* (SEQ ID NO: 2). In particular embodiments, said substitutions and/or mutations are replacement of at least one of the initial amino acids, in each case with an amino acid selected from the group consisting of: A, D, E, G, H, K, N, P, Q, R, S, T, Y, C, V, L, W, M and I.

As non limiting examples, the leucine (L) at position 26 may be replaced by/mutated to isoleucine(I); the arginine at position 28 may be replaced by/mutated to aspartic acid (D), the asparagine (N) at position 32 may be replaced by/mutated to arginine (R); the lysine (K) at position 34 may be replaced by/mutated to asparagine (N); the serine (S) at position 35 may be replaced by/mutated to glutamic acid (E); the valine (V) at position 37 may be replaced by/mutated to asparagine (N); the glycine (G) at position 38 may be replaced by/mutated to arginine (R); the serine (S) at position 40 may be replaced by/mutated to arginine (R); the glutamic acid (E) at position 42 may be replaced by/mutated to serine (S); the glycine (G) at position 44 may be replaced by/mutated to arginine (R) (see Table1).

The valine (V) at position 68 may be replaced by/mutated to lysine (K); the alanine (A) at position 70 may be replaced by/mutated to threonine (T); the asparagine (N) at position 75 may be replaced by/mutated to arginine (R); the serine (S) at position 78 may be replaced by/mutated to methionine (M); the lysine (K) at position 80 may be replaced by/mutated to arginine (R) (see Table1).

The leucine (L) at position 138 may be replaced by/mutated to methionine (M); the serine (S) at position 159 may be replaced by/mutated to proline (P); The glutamic acid (E) at position 178 may be replaced by/mutated to aspartic acid (D); the cysteine (C) at position 180 may be replaced by/mutated to tyrosine (Y); the phenylalanine (F) at position 182 may be replaced by/mutated to glycine (G); the isoleucine (I) at position 186 may be replaced by/mutated to lysine (K); the serine (S) at position 188 may be replaced by/mutated to valine (V); the serine (S) at position 190 may be replaced by/mutated to glycine (G); the lysine (K) at position 191 may be replaced by/mutated to asparagine (N); the leucine (L) at position 192 may be replaced by/mutated to alanine (A); the glycine (G) at position 193 may be replaced by/mutated to lysine (K); the glutamine (Q) at position 195 may be replaced by/mutated to tyrosine (Y); the glutamine (Q) at position 197 may be replaced by/mutated to glycine (G); the valine (V) at position 199 may be replaced by/mutated to arginine (R); the threonine (T) at position 203 may be replaced by/mutated to serine (S); the lysine (K) at position 207 may be replaced by/mutated to arginine (R) (see Table1)

The tyrosine (Y) at position 223 may be replaced by/mutated to serine (S); the lysine (K) at position 225 may be replaced by/mutated to tryptophan (W); the aspartic acid (D) at position 236 may be replaced by/mutated to glutamic acid (E) (see table 1).

In a more preferred embodiment the I-*OnuI* variant comprise the protein sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 8, and SEQ ID NO: 10.

In a preferred embodiment, the *I-OnuI* or *I-OnuI* homologue variant according to the present invention has at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 97%, more preferably at least 99% sequence identity with the protein sequence SEQ ID NO: 6, SEQ ID NO: 8, and SEQ ID NO: 10.

According to a preferred embodiment of the invention the I-OnuI or *I-OnuI* homologues variants according to the invention cleave a target sequence that is different from the target sequence of the corresponding wild-type endonuclease. Cleavage in the nucleic acid target sequence can correspond to either a double-stranded break or a single-stranded break

The present invention is based on the finding that such variant endonucleases with novel specificities can be used to allow efficient targeted modification of the TRAC gene.

The present inventors have indeed identified putative I-*OnuI* target sequences in the human TRAC gene based on a series of common features intrinsic to the group of monomeric I-*OnuI*-like LHE subfamily members recently described in (Takeuchi, Lambert et al. 2011; Baxter, Lambert et al. 2012). The putative LHE target sequences are also identified on the basis of the locations within TRAC gene wherein endonuclease-mediated insertions or deletions can cause significant disruptions to the TCR-alpha protein. The present inventors identified two putative target sequences in the human TRAC gene (SEQ ID NO: 3 to SEQ ID NO: 4) upon which the DNA recognition interface of the I-*OnuI* variants were engineered. Among these two putative target sites, one sequence (TCRA_S02) has been successfully targeted by the resulting I-*OnuI* variants.

Accordingly, the present invention relates to a rare-cutting endonuclease comprising an I-*OnuI* or I-*OnuI* homologue variant that recognizes a target nucleic acid sequence present within TRAC gene, preferably those present in the exon 1 of the TRAC gene, more preferably a target nucleic acid sequence comprising nucleic acid sequence SEQ ID NO: 3

Chimeric endonuclease

In another aspect, the invention relates to a rare-cutting endonuclease under the form of a chimeric endonuclease comprising an *I-OnuI* or *I-OnuI* homologue variant as described above, optionally fused to at least one additional protein domain by a peptide linker. The additional

protein domain may be selected from the group consisting of: a nucleic acid binding domain to allow higher specificity on target nucleic acid sequence and avoid off target site; a catalytic domain to process (eg. polymerize, depolymerize, modify) target nucleic acid sequence; and one or more terminal epitope tags or fluorescent proteins to follow and visualize the chimeric protein.

In a particular embodiment, the I-*OnuI* or I-*OnuI* homologue variant is fused to a nucleic acid binding domain such as TALE nucleic acid binding domain as non-limiting example to improve TRAC gene targeting.

Said Transcription Activator like Effector (TALE) corresponds to an engineered TALE comprising a plurality of TALE repeat sequences, each repeat comprising a RVD specific to each nucleotide base of a TALE recognition site. In the present invention, each TALE repeat sequence of said TALE is made of 30 to 42 amino acids, more preferably 33 or 34 wherein two critical amino acids (the so-called repeat variable dipeptide, RVD) located at positions 12 and 13 mediates the recognition of one nucleotide of said TALE binding site sequence; equivalent two critical amino acids can be located at positions other than 12 and 13 particularly in TALE repeat sequence larger than 33 or 34 amino acids long. Preferably, RVDs associated with recognition of the different nucleotides are HD for recognizing C, NG for recognizing T, NI for recognizing A, NN for recognizing G or A, NS for recognizing A, C, G or T, HG for recognizing T, IG for recognizing T, NK for recognizing G, HA for recognizing C, ND for recognizing C, HI for recognizing C, HN for recognizing G, NA for recognizing G, SN for recognizing G or A and YG for recognizing T, TL for recognizing A, VT for recognizing A or G and SW for recognizing A. In another embodiment, critical amino acids 12 and 13 can be mutated towards other amino acid residues in order to modulate their specificity towards nucleotides A, T, C and G and in particular to enhance this specificity. By other amino acid residues is intended any of the twenty natural amino acid residues or unnatural amino acids derivatives.

In another embodiment, said TALE of the present invention comprises between 5 and 30 TALE repeat sequences. More preferably, said TALE of the present invention comprises between 8 and 20 TALE repeat sequences; again more preferably 10 TALE repeat sequences.

In another embodiment, said TALE comprises an additional single truncated TALE repeat sequence made of 20 amino acids located at the C-terminus of said set of TALE repeat sequences, i.e. an additional C-terminal half- TALE repeat sequence. In this case, said TALE of the present invention comprises between 5.5 and 30.5 TALE repeat sequences, ".5" referring to previously mentioned half-TALE repeat sequence (or terminal RVD, or half-repeat). More preferably, said

TALE of the present invention comprises between 5.5 and 20.5 TALE repeat sequences, again more preferably, 10.5 TALE repeat sequences. In a preferred embodiment, said half- TALE repeat sequence is in a TALE context which allows a lack of specificity of said half-TALE repeat sequence toward nucleotides A, C, G, T. In a more preferred embodiment, said half- TALE repeat sequence is absent. In another embodiment, said TALE of the present invention comprises TALE like repeat sequences of different origins. In a preferred embodiment, said TALE comprises TALE like repeat sequences originating from different naturally occurring TAL effectors. In another preferred embodiment, internal structure of some TALE like repeat sequences of the TALE of the present invention are constituted by structures or sequences originated from different naturally occurring TAL effectors. In another embodiment, said TALE of the present invention comprises TALE like repeat sequences. TALE like repeat sequences have a sequence different from naturally occurring TALE repeat sequences but have the same function and / or global structure within said core scaffold of the present invention.

The chimeric endonuclease according to the invention can therefore correspond to the fusion of *I-OnuI* variant or I-*OnuI* homologue variant as previously described to a modular nucleic acid binding domain, such as a TALE or a zinc-finger domain, said fusion being active under monomeric form, as part as a single chain polypeptide.

According to a further aspect of the invention, the protein domain fused to the I-*OnuI* variant or I-*OnuI* homologue variant may have at least one catalytical activity selected from the group consisting of nuclease activity, polymerase activity, kinase activity, phosphatase activity, methylase activity, topoisomerase activity, integrase activity, transposase activity, ligase activity, helicase activity, recombinase activity. In a preferred embodiment, protein domain has an endonuclease activity, whereas the *Onu-I* variant retains its own cleavage activity or solely retains binding affinity to TRAC; in another preferred embodiment, said protein domain is or comprises an exonuclease activity. As non-limiting examples, As non-limiting examples, catalytic domains may be or comprise in part one of the proteins selected in the group consisting of: MmeI, Colicin-E7 (CEA7_ECOLX), Colicin-E9, APFL, EndA, Endo I (END1_ECOLI), Human Endo G (NUCG_HUMAN), Bovine Endo G (NUCG_BOVIN), R.HinP1I, I-BasI, I-BmoI, I-HmuI, I-TevI, I-TevII, I-TevIII, I-TwoI, R.MspI, R.MvaI, NucA, NucM, Vvn, Vvn_CLS, Staphylococcal nuclease (NUC_STAAU), Staphylococcal nuclease (NUC_STAHY), Micrococcal nuclease (NUC_SHIFL), Endonuclease yncB, Endodeoxyribonuclease I (ENRN_BPT7), Metnase, Nb.BsrDI, BsrDI A, Nt.BspD6I (R.BspD6I large subunit), ss.BspD6I (R.BspD6I small subunit), R.PleI, MlyI, AlwI, Mva1269I, BsrI, BsmI, Nb.BtsCI, Nt.BtsCI, R1.BtsI, R2.BtsI, BbvCI subunit 1, BbvCI subunit 2, Bpu10I alpha subunit, Bpu10I beta

subunit, BmrI, BfiI, I-CreI, hExoI (EXO1_HUMAN), Yeast ExoI (EXO1_YEAST), E.coli ExoI, Human TREX2, Mouse TREX1, Human TREX1, Bovine TREX1, Rat TREX1, Human DNA2, Yeast DNA2 (DNA2_YEAST), TdT and VP16 or a functional mutant thereof.

In a preferred embodiment, the catalytic domain is a DNA end-processing enzyme. Non limiting examples of DNA end-processing enzymes include 5-3' exonucleases, 3-5' exonucleases, 5-3' alkaline exonucleases, 5' flap endonucleases, helicases, phosphatase, hydrolases and template-independent DNA polymerases. In a more preferred embodiment, said catalytic domain has an exonuclease activity, in particular a 3'-5' exonuclease activity. In a more preferred embodiment, said catalytic domain is TREX2 or a functional variant thereof. In another preferred embodiment, said catalytic domain is encoded by a single chain TREX2 polypeptide. In a particular embodiment, said catalytic domain is fused to the N-terminus or C-terminus of said rare-cutting endonuclease. In a more preferred embodiment, said catalytic domain is fused to said rare-cutting endonuclease by a peptide linker.

In particular aspects, peptide linkers act as a communication device/linking or joining element between the rare-cutting endonuclease and an additional protein domain to act in concert for activity. Said peptide linker provides a peptide sequence which allows the connection of different monomers in a fusion protein and the adoption of the correct conformation for said fusion protein activity, but does not alter the specificity of either of the monomers for their targets. Peptide linkers can be of various sizes, from 2 amino acids to 50 amino acids as a non-limiting indicative range. Peptide linkers can also be structured or unstructured.

Alternatively, the *I-OnuI* variant or *I-OnuI* homologue variant according to the invention is used in conjunction with another protein not being fused thereto, having the same catalytic activity as the protein domain described above.

Another aspect of the invention provides polynucleotides comprising nucleic acid sequence encoding the rare-cutting endonucleases, preferably I-*OnuI* variants, homologues or chimeric endonuclease as described herein and vectors comprising such polynucleotides. Nucleic acid or vectors according to additional aspects of the present invention can comprise a nucleic acid sequence encoding one or more subcellular localization motifs, protease cleavage sites or ribosomal skip sequences.

In particular embodiments, the nucleic acids of the present invention can comprise at least one subcellular localization motif. A subcellular localization motif refers to a sequence that

facilitates transporting or confining a protein to a defined subcellular location that includes at least one of the nucleus, cytoplasm, plasma membrane, endoplasmic reticulum, golgi apparatus, endosomes, peroxisomes and mitochondria. Subcellular localization motifs are well-known in the art. Subcellular localization motif requires a specific orientation, e.g., N- and/or C-terminal to the protein. As a non-limiting example, the nuclear localization signal (NLS) of the simian virus 40 large T-antigen can be oriented at the N and/or C-terminus. NLS is an amino acid sequence which acts to target the protein to the cell nucleus through Nuclear Pore Complex and to direct a newly synthesized protein into the nucleus via its recognition by cytosolic nuclear transport receptors. Typically, a NLS consists of one or more short sequences of positively charged amino acids such as lysines or arginines.

Methods of genome engineering

Another aspect of the invention concerns the use of I-*Onu*I variant, I-*Onu*I homologue variant or I-*Onu*I derived chimeric endonuclease as described above to allow efficient TRAC gene targeting in a cell. More particularly, the invention relates to a method for targeted modification in the TRAC gene in a cell comprising introducing into a cell the rare-cutting endonuclease or chimeric endonuclease as described above. In a particular embodiment, the present invention relates to a method for modifying the TRAC gene in a cell comprising, introducing into the cell the rare-cutting endonuclease more particularly the I-*Onu*I variant, I-*Onu*I homologue variant or chimeric endonuclease, such that the rare-cutting endonuclease cleaves a nucleic acid target sequence in TRAC gene.

According to a further embodiment of the invention, the rare-cutting endonuclease is expressed into a cell in order to obtain targeted mutagenesis at the TRAC locus. The nucleic acid strand breaks caused by the rare-cutting endonuclease are commonly repaired through the distinct mechanisms of homologous recombination or non-homologous end joining (NHEJ). However, NHEJ is an imperfect repair process that often results in changes to the DNA sequence at the site of the double strand break. Mechanisms involve rejoining of what remains of the two DNA ends through direct re-ligation (Critchlow and Jackson 1998) or via the so-called microhomology-mediated end joining (Ma, Kim et al. 2003). Repair via non-homologous end joining (NHEJ) often results in small insertions or deletions and can be used for the creation of specific gene knockouts. Said modification may be a substitution, deletion, or addition of at least one nucleotide. Cells in which a cleavage-induced mutagenesis event, i.e a mutagenesis event consecutive to an NHEJ event, has occurred can be identified and/or selected by well-known

method in the art. As a non-limiting example, deep-sequencing analysis can be generated from the targeted cell genome around the targeted locus. Insertion/deletion events (mutagenesis events) can be therefore detected. As another non-limiting example, assays based on T7 endonuclease that recognizes non-perfectly matched DNA can be used, to quantify from a locus specific PCR on genomic DNA from provided cells, mismatches between reannealed DNA strands coming from cleaved/non-cleaved DNA molecules

In a particular embodiment of the methods envisaged herein the mutagenesis is increased by introducing into the cell an additional catalytic domain. In a particular embodiment, the present invention provides improved methods for ensuring targeted modification in the TRAC gene and provides a method for increasing mutagenesis at the target TRAC nucleic acid sequence to generate at least one nucleic acid cleavage and a loss of genetic information around said target nucleic acid sequence thus preventing any scarless re-ligation by NHEJ. In a more preferred embodiment, said catalytic domain is a DNA end-processing enzyme. Non limiting examples of DNA end-processing enzymes include 5-3' exonucleases, 3-5' exonucleases, 5-3' alkaline exonucleases, 5' flap endonucleases, helicases, hosphatase, hydrolases and template-independent DNA polymerases. Non limiting examples of such catalytic domain comprise at least one protein domain or catalytically active derivative of the protein domain selected from the group consisting of hExoI (EXO1_HUMAN), Yeast ExoI (EXO1_YEAST), E.coli ExoI, Human TREX2, Mouse TREX1, Human TREX1, Bovine TREX1, Rat TREX1, TdT (terminal deoxynucleotidyl transferase) Human DNA2, Yeast DNA2 (DNA2_YEAST). In a more preferred embodiment, said catalytic domain has an exonuclease activity, in particular a 3'-5' exonuclease activity. In a more preferred embodiment, said catalytic domain is TREX2 or functional variant thereof. In another preferred embodiment, said catalytic domain is encoded by a single chain TREX polypeptide. In a particular embodiment, said catalytic domain is fused to the N-terminus or C-terminus of said rare-cutting endonuclease. It has been found that the coupling of the enzyme TREX2 or single chain TREX2 with an endonuclease such as a meganuclease ensures high frequency of targeted mutagenesis. Alternatively, the above catalytic domain can be separately brought into the cell as part of an independent protein.

Endonucleolytic breaks are known to stimulate homologous recombination. Therefore, in particular embodiments, the present invention also relates to a method for inducing homologous gene targeting in the target nucleic acid sequence further comprising introducing into the cell a donor matrix comprising a sequence homologous to at least a portion of the target TRAC gene,

such that homologous recombination occurs between the target nucleic acid sequence and the donor matrix.

In particular embodiments, homologous TRAC gene targeting is achieved by introducing into a cell a rare-cutting endonuclease as described above, to induce a cleavage within or adjacent to a nucleic acid target sequence, as well as a donor matrix comprising a transgene to introduce said transgene by homologous recombination. Following cleavage of the target nucleic acid sequence, a homologous recombination event is stimulated between the genome containing the target nucleic acid sequence and the donor matrix. Said donor matrix comprises a sequence homologous to at least a portion of the target nucleic acid sequence, such that homologous recombination occurs between the target nucleic acid sequence and the donor matrix. Preferably, homologous sequences of at least 50 bp in length, preferably more than 100 bp and more preferably more than 200 bp are used within said donor matrix. Therefore, the donor matrix is preferably from 200 bp to 6000 bp in length, more preferably from 1000 bp to 2000 bp. In another embodiment, said donor matrix comprises two sequences homologous to portions or adjacent portions of said target nucleic acid sequence flanking a sequence to introduce in the target nucleic acid sequence. Indeed, shared DNA homologies are located in regions flanking upstream and downstream the site of the break and the nucleic acid sequence to be introduced should be located between the two homology arms. In particular embodiments, said donor matrix comprises first and second portions which are homologous to region 5' and 3' of the target nucleic acid, respectively. Said donor matrix in these embodiments can also comprise a third portion positioned between the first and the second portion which comprises little or no homology with the regions 5' and 3' of the site of DNA cleavage. In this case, said donor matrix allows introducing new genetic material into a cell. Said new genetic material introduced into a cell can confer a selective or a commercial advantage to said cell. In another embodiment, said donor matrix allows to replace genetic material into a cell. In another embodiment, said donor matrix allows to repair genetic material into a cell.

In particular embodiments, said donor matrix can comprise a positive selection marker between the two homology arms and eventually a negative selection marker upstream of the first homology arm or downstream of the second homology arm. The marker(s) allow(s) the selection of cells having inserted the sequence of interest by homologous recombination at the target site. Depending on the location of the targeted genome sequence wherein cleavage event has occurred, such donor matrix can be used to knock-out a gene, e.g. when the donor matrix is located within the open reading frame of said gene, or to introduce new sequences or genes of

interest. Sequence insertions by using such donor matrix can be used to modify a targeted existing gene, by correction or replacement of said gene (allele swap as a non-limiting example), or to up- or down-regulate the expression of the targeted gene (promoter swap as non-limiting example), said targeted gene correction or replacement.

Cells in which a homologous recombination event has occurred can be selected by methods well-known in the art. As a non-limiting example, PCR analysis using one oligonucleotide matching within the exogenous nucleic acid sequence and one oligonucleotide matching the genomic nucleic acid of cells outside said exogenous nucleic acid but close to the targeted locus can be performed. Therefore, cells in which methods of the invention allowed a mutagenesis event or a homologous recombination event to occur can be selected.

The different methods of the invention involve introducing rare-cutting endonuclease or chimeric endonuclease optionally with DNA-end processing enzyme or donor matrix into a cell. As non-limiting example, said rare-cutting endonuclease or chimeric endonuclease optionally with DNA-end processing enzyme or donor matrix can be introduced as transgenes encoded by one or as different plasmidic vectors. Different transgenes can be included in one vector which comprises a nucleic acid sequence encoding ribosomal skip sequence such as a sequence encoding a 2A peptide. 2A peptides, which were identified in the Aphthovirus subgroup of picornaviruses, causes a ribosomal "skip" from one codon to the next without the formation of a peptide bond between the two amino acids encoded by the codons (see Donnelly et al., J. of General Virology 82: 1013-1025 (2001); Donnelly et al., J. of Gen. Virology 78: 13-21 (1997); Doronina et al., Mol. And. Cell. Biology 28(13): 4227-4239 (2008); Atkins et al., RNA 13: 803-810 (2007)). By "codon" is meant three nucleotides on an mRNA (or on the sense strand of a DNA molecule) that are translated by a ribosome into one amino acid residue. Thus, two polypeptides can be synthesized from a single, contiguous open reading frame within an mRNA when the polypeptides are separated by a 2A oligopeptide sequence that is in frame. Such ribosomal skip mechanisms are well known in the art and are known to be used by several vectors for the expression of several proteins encoded by a single messenger RNA. As non-limiting example, in the present invention, 2A peptides have been used to express into the cell the rare-cutting endonuclease and a DNA end-processing enzyme. As non-limiting examples, 2A peptide may be used to express into the cell the rare-cutting endonuclease or the chimeric endonuclease and an additional protein domain with a catalytical activity selected from the group consisting of nuclease activity, polymerase activity, kinase activity, phosphatase activity, methylase activity, topoisomerase activity, integrase activity, transposase activity, ligase activity, helicase activity, recombinase activity to process target

nucleic acid sequence. The 2A peptide may also be used to express into the cell the rare-cutting endonuclease or the chimeric endonuclease and a fluorescent protein.

Said plasmid vector can contain a selection marker which provides for identification and/or selection of cells which received said vector. Vectors can be introduced into a cell by a variety of methods (e.g., injection, direct uptake, projectile bombardment, liposomes, electroporation). Rare-cutting endonucleases, chimeric endonucleases, DNA-end processing enzyme or donor matrix according to the present invention can be stably or transiently expressed into cells using expression vectors. Techniques of expression in eukaryotic cells are well known to those in the art. (See Current Protocols in Human Genetics: Chapter 12 "Vectors For Gene Therapy" & Chapter 13 "Delivery Systems for Gene Therapy"). The polypeptide may be synthesized *in situ* in the cell as a result of the introduction of polynucleotide encoding polypeptide into the cell. Said protein expression can be induced in selected cells and said rare-cutting endonuclease or chimeric endonuclease cleaves target nucleic acid sequence in selected cells. Alternatively, the polypeptide could be produced outside the cell and then introduced thereto by well-known method of the art.

In another embodiment, said methods of the present invention can be used to generate animals or plants wherein a targeted double-stranded break occurred. Animals may be generated by introducing a rare-cutting endonuclease or a chimeric endonuclease according to the invention into a cell or an embryo. In particular, the present invention relates to a method for generating an animal, comprising providing an eukaryotic cell comprising a nucleic acid target sequence in TCR-alpha gene into which it is desired to introduce a genetic modification; generating a cleavage within or adjacent to the nucleic acid target sequence by introducing an engineered rare-cutting endonuclease or chimeric endonuclease according to the present invention; and generating an animal from the cell or progeny thereof, in which cleavage has occurred. Typically, the embryo is a fertilized one cell stage embryo. Polynucleotides encoding said rare-cutting endonuclease or chimeric endonuclease may be introduced into the cell by any of the methods known in the art including micro injection into the nucleus or cytoplasm of the embryo. In a particular embodiment, the method for generating an animal, further comprise introducing a donor matrix as desired. Said donor matrix comprises a sequence homologous to at least a portion of the nucleic acid target sequence, such that homologous recombination occurs between said donor matrix and the nucleic acid target sequence in the cell or progeny thereof. The donor matrix can include for example a nucleic acid sequence that disrupts a gene after homologous recombination, a nucleic acid sequence that replaces a gene after homologous recombination, a

nucleic acid sequence that introduces a mutation into a gene after homologous recombination or a nucleic acid sequence that introduce a regulatory site after homologous recombination. The embryos are then cultured to develop an animal. In one aspect of the invention, an animal in which at least a nucleic acid target sequence of interest has been engineered is provided. For example, an engineered gene may become inactivated such that it is not transcribed or properly translated, or an alternate form of the gene is expressed. The animal may be homozygous or heterozygous for the engineered gene. More particularly, the present invention relates to a method for making an TCR-alpha knock-in or knock-out animal, comprising: a) introducing into a pluripotent precursor cell or embryo of an animal, a rare-cutting endonuclease or chimeric endonuclease as defined above sufficient/capable to induce a nucleic acid cleavage in the nucleic acid target present in TCR-alpha gene; (b) introducing Into the animal precursor cell or embryo of step (a), optionally a donor matrix, wherein said donor matrix comprises a sequence to be introduced flanked by at least one sequence sharing homologies with at least one region of the TCR-alpha gene surrounding the nucleic acid cleavage site of said rare-cutting endonuclease; (c) developing the genomically modified animal precursor cell or embryo of step (b) into a chimeric animal, and (d) deriving a transgenic animal from the chimeric animal of step (c). Preferably, step (c) comprises the introduction of the genomically modified precursors cells generated in step (b) into blastocysts so as to generate chimeric animals.

In another aspect, the present invention relates to an isolated cell comprising a gene encoding the TCR-alpha protein inactivated (e.g, with respect to typical TCR-alpha protein biogenesis and/or TCR-alpha protein cell surface expression and/or with respect to the TCR-alpha protein mediating antigen recognition and immunoreceptor signaling) by the methods described above.

”Cell” or “cells” as used herein refers to any prokaryotic or eukaryotic living cells, cell lines derived from these organisms for *in vitro* cultures, primary cells from animal origin.

“Primary cell” or “primary cells” as used herein refers to cells taken directly from living tissue (i.e. biopsy material) and established for growth in vitro, that have undergone very few population doublings and are therefore more representative of the main functional components and characteristics of tissues from which they are derived from, in comparison to continuous tumorigenic or artificially immortalized cell lines. These cells thus represent a more valuable model to the in vivo state they refer to.

More preferably the animal cell is of the genus *Homo, Rattus, Mus, Sus, Bos, Danio, Canis, Felis, Equus, Salmo, Oncorhynchus, Gallus, Meleagris, Drosophila, Caenorhabditis*; more preferably, the animal cell is of the species *Homo sapiens, Rattus norvegicus, Mus musculus, Sus scrofa, Bos taurus, Danio rerio, Canis lupus, Felis catus, Equus caballus, Salmo salar, Oncorhynchus mykiss, Gallus gallus, Meleagris gallopavo, Drosophila melanogaster, Caenorhabditis elegans.*

In aspects of the present invention, the cell can a mammalian cell, a or cell lines derived from these organisms for *in vitro* cultures or primary cells taken directly from living tissue and established for *in vitro* culture. As non-limiting examples cell lines can be selected from the group consisting of CHO-K1 cells; HEK293 cells; Caco2 cells; U2-OS cells; NIH 3T3 cells; NSO cells; SP2 cells; CHO-S cells; DG44 cells; K-562 cells, U-937 cells; MRC5 cells; IMR90 cells; Jurkat cells; HepG2 cells; HeLa cells; HT-1080 cells; HCT-116 cells; Hu-h7 cells; Huvec cells; Molt 4 cells.

In a more preferred embodiment, said isolated cells can be multipotent cells, for example stem cells. The stem cells can be adult stem cells, embryonic stem cells, more particularly non-human stem cells, cord blood stem cells, progenitor cells, bone marrow stem cells, induced pluripotent stem cells, totipotent stem cells or hematopoietic stem cells. Representative human cells are CD34+ cells. In a particular embodiment of the present invention, the cells are T-cells, preferably human T-cells.

Method for treating or preventing cancer, autoimmune disease or viral infection

In another aspect, the present invention relates to the use of the I-*Onu*I variants, I-*Onu*I homologue variant or I-*Onu*I derived chimeric endonuclease according to the invention as a medicament.

More particularly, the present invention relates to a method for treating a subject having cancer, autoimmune disease or viral infection comprising of introducing into a cell a rare-cutting endonuclease or chimeric endonuclease according to the invention sufficient to provide for mutagenesis or homologous recombination in the TRAC gene, optionally with a donor matrix and/or DNA-end processing enzyme and administrating the cells to the subject. In particular aspects, the method can combine the introduction of a rare-cutting endonuclease or chimeric endonuclease with the introduction of an artificial/chimeric antigen receptor recognizing a tumor, virus or autoimmune-related target. In certain embodiment, the method can comprise selecting cultured cells in which the mutagenesis or homologous recombination event has occurred in the TRAC gene by well-known methods in the art.

Said treatment can be ameliorating, curative or prophylactic. It may be either part of an autologous or part of an allogenic treatment. By autologous, it is meant that cells, cell line or population of cells used for treating patients are originating from said patient. By allogeneic is meant that the cells or population of cells used for treating patients are not originating from said patient but from a donor.

Cells that can be used with the disclosed methods can be multipotent cells, for example stem cells. The stem cells can be adult stem cells, embryonic stem cells, more particularly non-human stem cells, cord blood stem cells, progenitor cells, bone marrow stem cells, induced pluripotent stem cells, totipotent stem cells or hematopoietic stem cells. Representative human cells are CD34+ cells or human T-cells. Prior to expansion and genetic modification of the cells of the invention, a source of cells can be obtained from a subject through a variety of non-limiting methods. T cells can be obtained from a number of non-limiting sources, including peripheral blood mononuclear cells, bone marrow, lymph node tissue, cord blood, thymus tissue, tissue from a site of infection, ascites, pleural effusion, spleen tissue, and tumors. In certain embodiments of the present invention, any number of T cell lines available and known to those skilled in the art, may be used.

In another embodiment, isolated cells obtained by the different methods or cell line(s) derived from said isolated cells can be used as a medicament. In another embodiment, said medicament can be used for treating cancer or autoimmune disease or viral infection in a patient in need thereof. In another embodiment, said isolated cell according to the invention or cell line derived from said isolated cell can be used in the manufacture of a medicament for treatment of a cancer, autoimmune disease or viral infection in a patient in need thereof.

The administration of the cells or population of cells according to the present invention may be carried out in any convenient manner, including by aerosol inhalation, injection, ingestion, transfusion, implantation or transplantation. The compositions described herein may be administered to a patient subcutaneously, intradermaliy, intratumorally, intranodally, intramedullary, intramuscularly, by intravenous or intralymphatic injection, or intraperitoneally. In one embodiment, the cell compositions of the present invention are preferably administered by intravenous injection.

In particular aspects, the administration of the cells or population of cells comprises the administration of 10^4 to 10^9 cells/kg body weight, preferably 10^5 to 10^6 cells/kg body weight, including all values of cell numbers within those ranges. The cells or population of cells can be

administrated in one or more doses. In another embodiment, said effective amount of cells are administrated as a single dose. In another embodiment, said effective amount of cells are administrated as more than one dose over a period time. Timing of administration is within the judgment of managing physician and depends on the clinical condition of the patient. The cells or population of cells may be obtained from any source, such as a cell bank or a donor. While individual needs vary, determination of optimal ranges of effective amounts of a given cell type for a particular disease or conditions within the skill of the art. The dosage administrated will be dependent upon the age, health and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment and the nature of the effect desired.

In another embodiment, the administration of cells may be combined with administration of an immunosuppressive drug regimen. The immunosuppressive regiment may include, but not limited to, cytostatic, glucocorticoids and antibody-based drug classes. The immunosuppressive regimen may be administered before, during or after administration of cells. The immunosuppressive regimen can be administered in one or more doses. The dosage, timing and composition of the immunosuppressive drug regimen are within the judgment of managing physician and depend on the clinical condition of the patient.

In another embodiment, the present invention relates to a method for targeting TCR-alpha gene in a subject, the method comprising administrating to a subject a vector encoding a rare-cutting endonuclease according to the present invention.

Definitions

In the description above, a number of terms are used extensively. The following definitions are provided to facilitate understanding of the present embodiments.

As used herein, the term “about” indicates that a value includes the inherent variation of error for the method being employed to determine a value, or the variation that exists among experiments.

- Amino acid residues in a polypeptide sequence are designated herein according to the one-letter code, in which, for example, Q means Gln or Glutamine residue, R means Arg or Arginine residue and D means Asp or Aspartic acid residue.

- Amino acid substitution means the replacement of one amino acid residue with another, for instance the replacement of an Arginine residue with a Glutamine residue in a peptide sequence is an amino acid substitution.

- Nucleotides are designated as follows: one-letter code is used for designating the base of a nucleoside: a is adenine, t is thymine, c is cytosine, and g is guanine. For the degenerated nucleotides, r represents g or a (purine nucleotides), k represents g or t, s represents g or c, w represents a or t, m represents a or c, y represents t or c (pyrimidine nucleotides), d represents g, a or t, v represents g, a or c, b represents g, t or c, h represents a, t or c, and n represents g, a, t or c.

- As used herein, “nucleic acid” or “nucleic acid molecule” refers to nucleotides and/or polynucleotides, such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), oligonucleotides, fragments generated by the polymerase chain reaction (PCR), and fragments generated by any of ligation, scission, endonuclease action, and exonuclease action. Nucleic acid molecules can be composed of monomers that are naturally-occurring nucleotides (such as DNA and RNA), or analogs of naturally-occurring nucleotides (e.g., enantiomeric forms of naturally-occurring nucleotides), or a combination of both. Modified nucleotides can have alterations in sugar moieties and/or in pyrimidine or purine base moieties. Sugar modifications include, for example, replacement of one or more hydroxyl groups with halogens, alkyl groups, amines, and azido groups, or sugars can be functionalized as ethers or esters. Moreover, the entire sugar moiety can be replaced with sterically and electronically similar structures, such as aza-sugars and carbocyclic sugar analogs. Examples of modifications in a base moiety include alkylated purines and pyrimidines, acylated purines or pyrimidines, or other well-known heterocyclic substitutes. Nucleic acid monomers can be linked by phosphodiester bonds or analogs of such linkages. Nucleic acids can be either single stranded or double stranded.

-by “chimeric endonuclease” it is intended to mean an endonuclease which comprise functional portions of an endonuclease operationally linked to one or more protein functional domains coming from another protein.

- The terms “fusion protein” or "chimeric protein" indicate that the protein includes polypeptide components derived from more than one parental protein or polypeptide. Typically, a fusion protein is expressed from a fusion gene in which a nucleotide sequence encoding a polypeptide sequence from one protein is appended in frame with, and optionally separated by a linker from, a nucleotide sequence encoding a polypeptide sequence from a different protein. The

fusion gene can then be expressed by a host cell as a single protein. A fusion protein can comprise at least part of one polypeptide fused with another polypeptide. In some embodiments, a fusion protein can comprise at least a part of one polypeptide fused with at least a part of the same polypeptide.

- by "screening" it is intended to mean the sequential or simultaneous selection of one or more meganuclease variant(s) which exhibits a specified phenotype such as altered cleavage activity.

- by "mutation" is intended the substitution, deletion, insertion of one or more nucleotides/amino acids in a polynucleotide (cDNA, gene) or a polypeptide sequence. Said mutation can affect the coding sequence of a gene or its regulatory sequence. It may also affect the structure of the genomic sequence or the structure/stability of the encoded mRNA.

- By "gene" is meant the basic unit of heredity, consisting of a segment of DNA arranged in a linear manner along a chromosome, which codes for a specific protein or segment of protein. A gene typically includes a promoter, a 5' untranslated region, one or more coding sequences (exons), optionally introns, a 3' untranslated region. The gene may further comprise a terminator, enhancers and/or silencers.

- As used herein, the term "transgene" refers to a sequence encoding a polypeptide. Preferably, the polypeptide encoded by the transgene is either not expressed or expressed but not biologically active, in the cell, tissue or individual in which the transgene is inserted. Most preferably, the transgene encodes a therapeutic polypeptide useful for the treatment of an individual.

- By " delivery vector" or " delivery vectors" is intended any delivery vector which can be used in the present invention to put into cell contact (i.e "contacting") or deliver inside cells or subcellular compartments agents/chemicals and molecules (proteins or nucleic acids) needed in the present invention. It includes, but is not limited to liposomal delivery vectors, viral delivery vectors, drug delivery vectors, chemical carriers, polymeric carriers, lipoplexes, polyplexes, dendrimers, microbubbles (ultrasound contrast agents), nanoparticles, emulsions or other appropriate transfer vectors. These delivery vectors allow delivery of molecules, chemicals, macromolecules (genes, proteins), or other vectors such as plasmids, peptides developed by Diatos. In these cases, delivery vectors are molecule carriers. By "delivery vector" or "delivery vectors" is also intended delivery methods to perform transfection.

- The terms "vector" or “vectors” refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. A “vector” in the present invention includes, but is not limited to, a viral vector, a plasmid, a RNA vector or a linear or circular DNA or RNA molecule which may consists of a chromosomal, non chromosomal, semi-synthetic or synthetic nucleic acids. Preferred vectors are those capable of autonomous replication (episomal vector) and/or expression of nucleic acids to which they are linked (expression vectors). Large numbers of suitable vectors are known to those of skill in the art and commercially available.

Viral vectors include retrovirus, adenovirus, parvovirus (e. g. adenoassociated viruses), coronavirus, negative strand RNA viruses such as orthomyxovirus (e. g., influenza virus), rhabdovirus (e. g., rabies and vesicular stomatitis virus), paramyxovirus (e. g. measles and Sendai), positive strand RNA viruses such as picornavirus and alphavirus, and double-stranded DNA viruses including adenovirus, herpesvirus (e. g., Herpes Simplex virus types 1 and 2, Epstein-Barr virus, cytomegalovirus), and poxvirus (e. g., vaccinia, fowlpox and canarypox). Other viruses include Norwalk virus, togavirus, flavivirus, reoviruses, papovavirus, hepadnavirus, and hepatitis virus, for example. Examples of retroviruses include: avian leukosis-sarcoma, mammalian C-type, B-type viruses, D type viruses, HTLV-BLV group, lentivirus, spumavirus (Coffin, J. M., Retroviridae: The viruses and their replication, In Fundamental Virology, Third Edition, B. N. Fields, et al., Eds., Lippincott-Raven Publishers, Philadelphia, 1996).

- by “lentiviral vector” is meant HIV-Based lentiviral vectors that are very promising for gene delivery because of their relatively large packaging capacity, reduced immunogenicity and their ability to stably transduce with high efficiency a large range of different cell types. Lentiviral vectors are usually generated following transient transfection of three (packaging, envelope and transfer) or more plasmids into producer cells. Like HIV, lentiviral vectors enter the target cell through the interaction of viral surface glycoproteins with receptors on the cell surface. On entry, the viral RNA undergoes reverse transcription, which is mediated by the viral reverse transcriptase complex. The product of reverse transcription is a double-stranded linear viral DNA, which is the substrate for viral integration in the DNA of infected cells.

- by “integrative lentiviral vectors (or LV)”, is meant such vectors as non limiting example, that are able to integrate the genome of a target cell.

At the opposite by “non integrative lentiviral vectors” (or NILV) is meant efficient gene delivery vectors that do not integrate the genome of a target cell through the action of the virus integrase.

One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors. A vector according to the present invention comprises, but is not limited to, a YAC (yeast artificial chromosome), a BAC (bacterial artificial), a baculovirus vector, a phage, a phagemid, a cosmid, a viral vector, a plasmid, a RNA vector or a linear or circular DNA or RNA molecule which may consist of chromosomal, non chromosomal, semi-synthetic or synthetic DNA. In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. Large numbers of suitable vectors are known to those of skill in the art. Vectors can comprise selectable markers, for example: neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin phosphotransferase, herpes simplex virus thymidine kinase, adenosine deaminase, glutamine synthetase, and hypoxanthine-guanine phosphoribosyl transferase for eukaryotic cell culture; TRP1 for *S. cerevisiae*; tetracyclin, rifampicin or ampicillin resistance in *E. coli*. Preferably said vectors are expression vectors, wherein a sequence encoding a polypeptide of interest is placed under control of appropriate transcriptional and translational control elements to permit production or synthesis of said polypeptide. Therefore, said polynucleotide is comprised in an expression cassette. More particularly, the vector comprises a replication origin, a promoter operatively linked to said encoding polynucleotide, a ribosome binding site, a RNA-splicing site (when genomic DNA is used), a polyadenylation site and a transcription termination site. It also can comprise an enhancer or silencer elements. Selection of the promoter will depend upon the cell in which the polypeptide is expressed. Suitable promoters include tissue specific and/or inducible promoters. Examples of inducible promoters are: eukaryotic metallothionine promoter which is induced by increased levels of heavy metals, prokaryotic lacZ promoter which is induced in response to isopropyl-β-D-thiogalacto-pyranoside (IPTG) and eukaryotic heat shock promoter which is induced by increased temperature. Examples of tissue specific promoters are skeletal muscle creatine kinase, prostate-specific antigen (PSA), α-antitrypsin protease, human surfactant (SP) A and B proteins, β-casein and acidic whey protein genes.

Delivery vectors and vectors can be associated or combined with any cellular permeabilization techniques such as sonoporation or electroporation or derivatives of these techniques.

- The term “endonuclease”, or “nuclease” refers to any wild-type or variant enzyme capable of catalyzing the hydrolysis (cleavage) of bonds between nucleic acids within a DNA or RNA molecule, preferably a DNA molecule. Endonucleases can be classified as rare-cutting endonucleases when having typically a polynucleotide recognition greater than 12 base pairs (bp) in length, more preferably of 14-45 bp. Rare-cutting endonucleases significantly increase HR by inducing DNA double-strand breaks (DSBs) at a defined locus (Perrin, Buckle et al. 1993; Rouet, Smih et al. 1994; Rouet, Smih et al. 1994; Choulika, Perrin et al. 1995; Pingoud and Silva 2007). Rare-cutting endonucleases can for example be a homing endonuclease (Paques and Duchateau 2007), a chimeric Zinc-Finger nuclease (ZFN) (Eisenschmidt, Lanio et al. 2005 ; Arimondo, Thomas et al. 2006; Simon, Cannata et al. 2008), A TALE-nulcease or a chemical endonuclease. In chemical endonucleases, a chemical or peptidic cleaver is conjugated either to a polymer of nucleic acids or to another DNA recognizing a specific target sequence, thereby targeting the cleavage activity to a specific sequence. Chemical endonucleases also encompass synthetic nucleases like conjugates of orthophenanthroline, a DNA cleaving molecule, and triplex-forming oligonucleotides (TFOs), known to bind specific DNA sequences (Kalish and Glazer 2005). Such chemical endonucleases are comprised in the term “endonuclease” according to the present invention.

- Transcription Activator Like Effector (TALE) is a family of proteins used in the infection process by plant pathogens of the Xanthomonas genus (Boch, Scholze et al. 2009; Moscou and Bogdanove 2009; Christian, Cermak et al. 2010; Li, Huang et al. 2011; Li, Huang et al. 2011). The term “TAL effector nuclease” (TALE-Nuclease) refers to a nuclease comprising a TAL-effector domain fused to a nuclease domain. These DNA binding domains may be engineered to bind to a desired target and fused to a nuclease domain, such as the Fok1 nuclease domain, to derive a TAL effector domain-nuclease fusion protein.

- The term “Zinc-finger nuclease” (ZFN) refers to artificial restriction enzymes generated by fusing a zinc finger DNA-binding domain to a DNA-cleavage domain. Zinc finger domains can be engineered to bind to a desired target site. In some embodiments, the cleavage domain comprises the non-specific cleavage domain of FokI (Porteus and Carroll 2005). In other embodiments, the cleavage domain comprises all or an active portion of another nuclease.

- By “catalytic domain” is intended the protein domain or module of an enzyme containing the active site of said enzyme; by active site is intended the part of said enzyme at which catalysis of the substrate occurs. Enzymes, but also their catalytic domains, are classified

and named according to the reaction they catalyze. The Enzyme Commission number (EC number) is a numerical classification scheme for enzymes, based on the chemical reactions they catalyze.

- The term “exonuclease” refers to enzymes that cleave phosphodiester bonds at the end of a polynucleotide chain via a hydrolyzing reaction that breaks phosphodiester bonds at either the 3' or 5' end. The polynucleotide may be double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), RNA, double-stranded hybrids of DNA and RNA, and synthetic DNA (for example, containing bases other than A, C, G, and T). The term “5’ exonuclease” refers to exonucleases that cleave the phosphodiester bond at the 5’ end. The term “3’ exonuclease” refers to exonucleases that cleave the phosphodiester bond at the 3’ end. Exonucleases may cleave the phosphodiester bonds at the end of a polynucleotide chain at endonuclease cut sites or at ends generated by other chemical or mechanical means, such as shearing, ionizing radiation, ultraviolet radiation, oxygen radicals, chemical hydrolosis and chemotherapy agents. Exonucleases may cleave the phosphodiester bonds at blunt ends or sticky ends. E. coli exonuclease I and exonuclease III are two commonly used 3'-exonucleases that have 3'-exonucleolytic single-strand degradation activity. Other examples of 3'-exonucleases include Nucleoside diphosphate kinases (NDKs), NDK1 (NM23-H1), NDK5, NDK7, and NDK8, WRN, and Three prime repair exonuclease 2 (Trex2). E. coli exonuclease VII and T7-exonuclease Gene 6 are two commonly used 5'-3' exonucleases that have 5% exonucleolytic single-strand degradation activity. The exonuclease can be originated from prokaryotes, such as E. coli exonucleases, or eukaryotes, such as yeast, worm, murine, or human exonucleases.

- by “functional mutant” is intended a catalytically active mutant of a protein or a protein domain; such mutant can have the same activity compared to its parent protein or protein domain or additional properties. This definition applies to chimeric proteins or protein domains that constitute chimeric proteins according to the present invention. Are also encompassed in the scope of this definition “derivatives” of these proteins or protein domains that comprise the entirety or part of these proteins or protein domains fused to other protein or chemical parts such as tags, antibodies, polyethylene glycol as non-limiting examples.

- By nucleic acid or protein “homologous sequence” it is meant a sequence with high percentage of identity or high percentage of homology with sequences at nucleotidic or polypeptidic levels. By high percentage of identity or high percentage of homology it is intended at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least

85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 97%, more preferably at least 99% or any percentage value between 70% and 99%.

- "identity" refers to sequence identity between two nucleic acid molecules or polypeptides. Identity can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base, then the molecules are identical at that position. A degree of similarity or identity between nucleic acid or amino acid sequences is a function of the number of identical or matching nucleotides at positions shared by the nucleic acid sequences. Various alignment algorithms and/or programs may be used to calculate the identity between two sequences, including FASTA, or BLAST which are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with, e.g., default setting.

- The term “cleavage” refers to the breakage of the covalent backbone of a polynucleotide. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. Double stranded DNA, RNA, or DNA/RNA hybrid cleavage can result in the production of either blunt ends or staggered ends.

- The terms "target site", "target sequence", “target nucleic acid sequence” or “nucleic acid target sequence” refer to a nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule will bind and/or cleave, provided sufficient conditions for binding and/or cleaving are present.

- A "domain" of a protein is any portion of the entire protein, up to and including the complete protein, but typically comprising less than the complete protein. A domain can, but need not, fold independently of the rest of the protein chain and/or be correlated with a particular biological, biochemical, or structural function or location (e.g., an endonuclease domain, a polynucleotide binding domain, such as a DNA-binding domain, or an end-processing domain).

- By chimeric antigen receptor (CAR) is intended molecules that combine a binding domain against a component present on the target cell, for example an antibody-based specificity for a desired antigen (e.g., tumor antigen) with a T cell receptor-activating intracellular domain to generate a chimeric protein that exhibits a specific anti-target cellular immune activity. Generally,

CAR consists of an extracellular single chain antibody (scFvFc) fused to the intracellular signaling domain of the T cell antigen receptor complex zeta chain (scFvFc:ζ) and have the ability, when expressed in T cells, to redirect antigen recognition based on the monoclonal antibody's specificity.

- The term "subject" as used herein includes all members of the animal kingdom including non-human primates and humans.

EXAMPLES:

Example 1: Engineering of LHE prototypes with DNA recognition interfaces specific for targets in the human TRAC gene was performed

Putative LHE target sequences in the human TRAC gene were first identified for which high quality engineered DNA recognition interfaces were predicted by the inventor. Such predictions are based on a series of features intrinsic to the LHE scaffold, I-*Onu*I (SEQ ID NO: 1), upon which the TRAC DNA recognition interfaces were to be engineered. Other considerations, such as locations within the TRAC gene likely to cause significant disruptions to the TCR-alpha protein upon endonuclease-mediated insertions or deletions, and/or the occurrence of adjacent downstream TGA, TAG, or TAA stop codons in alternative reading frames to limit the production of out-of-frame peptides which could serve as the basis for immunological rejection, were also incorporated into the target choice process. See Figure 1 which schematically illustrates the locations of the putative target sequences.

Two putative target sequences (TCRA_S02 and TCRA_S10; SEQ ID NO: 3 and SEQ ID NO: 4) were chosen for the initial stages of engineering the DNA recognition interface. Variant libraries were constructed whereby amino acid residues in localized sub-regions of the DNA recognition interface were varied. See Figure 2 which shows schematic and structural diagrams of the DNA recognition interface. Variation within the DNA recognition interface of I-*Onu*I nucleic acid sequence (SEQ ID NO: 1) was achieved by incorporating degenerate codons into oligonucleotides which served as substrates for PCR reactions to generate variant libraries by gap recombination in the yeast strain *Saccharomyces cerevisiae*. The resulting libraries were screened for target cleavage activity by surface display and flow cytometry based methods as has been described in (Jarjour, West-Foyle et al. 2009). In this manner, the specificity of the DNA recognition interface was altered to recognize targets in the human TRAC gene. In particular aspects, successfully re-

specified DNA recognition interfaces were achieved for TCRA_S02 (SEQ ID NO: 3) only, with the process failing for the other putative target site at earlier stage in the engineering process. See Figure 3 illustrating the successful isolation of variants cleaving the TCRA_S02 target.

Example 2: LHEs with engineered DNA recognition interfaces were shown to cause disruptive mutations to the target sequences for which they were engineered to recognize.

To measure the activity of the TCRA targeting LHE, a chromosomally integrated fluorescent reporter system that has been described previously was used. In this system, the LHE of interest is transfected into a HEK 293T fibroblast cell line that is engineered to contain the TCRA_S02 target sequence upstream of an out-of-frame gene encoding the fluorescent protein mCherry. Cleavage of the embedded TCRA_S02 target and subsequent small insertions or deletions caused by DNA repair via the non-homologous end joining (NHEJ) pathway result in approximately 1 out of three repaired loci placing the fluorescent reporter gene 'in-frame'. Fluorescence in the mCherry channel on a flow cytometer is therefore a surrogate high-throughput readout of LHE cleavage of the chromosomally embedded TCRA_S02 target sequence.

Initial results with the TCRA_S02_2E5 variant (SEQ ID NO: 5 encoding SEQ ID NO: 6) showed very low efficiency of mCherry expression, indicating that this variant was not very actively cleaving its target in a cellular chromosomal context. Random mutagenesis of the TCRA_S02_2E5 variant and surface display-based screening under more stringent cleavage conditions to isolate variants with improved catalytic activities were performed. Two rounds of mutagenesis and screening led to variants with 40- to 50-fold higher rates of generating mCherry expressing cells. See Figure 4 illustrating the flow cytometry read-outs from the reporter assay for TCRA_S02_2E5 refined variants, which included TCRA_S02_2E5_RD1_08 (SEQ ID NO: 7 encoding SEQ ID NO: 8) and TCRA_S02_2E5_RD2_23 (SEQ ID NO: 9 encoding SEQ ID NO: 10). A top performing variant TCRA_S02_2E5_RD1_08 (SEQ ID NO:7) contained six mutations relative to the TCRA_S02_2E5 variant, four of which are located within the DNA recognition interface and two are located elsewhere in the LHE. See Figure 5 and Table 1 which provide the relative alignments of the indicated variants as well as the positional information of the residues comprising the DNA recognition interface. It is unknown to what extent, if any, the individual mutations identified through this process contribute to the characteristics of the LHE which influence its DNA recognition and cleavage activity. Taken together they led to significant improvements in the frequency of the occurrence disruptive mutations to the TCRA_S02 target sequence.

Example 3: LHEs with DNA recognition interfaces having high affinity, high specificity, and low toxicity were differentiated.

The LHE containing the engineered DNA recognition interfaces for the TCRA_S02 (SEQ ID NO: 3) target was tested for affinity, specificity, and toxicity characteristics. Affinity was tested by independently incubating yeast displaying the TCRA_S02_2E5_RD1_08 variant, (SEQ ID NO: 7, encoding SEQ ID NO: 8) with DNA substrates containing its target sequences at various concentrations. See Figure 6 showing the affinity properties of this variant relative to the wild-type I-*OnuI* protein. These data demonstrate that the TCRA_S02_2E5_RD1_08 variant binds its DNA target with affinities comparable or higher than that of the interaction between the native I-OnuI LHE and its target sequence (SEQ ID NO: 11). Specificity was tested by analyzing the relative affinity and DNA-cleaving ability of the LHE variant towards target sequence containing each of the three alternate DNA base pairs at each position along the substrate. See Figure 7 illustrating the specificity profile of the TCRA_S02_2E5_RD1_08 (SEQ ID NO: 7) variant. These data demonstrate that the TCRA_S02_2E5_RD1_08 LHE variant has a high overall specificity, as it exclusively cleaves/binds to its specific substrate base pair in most positions along its target, not tolerating substitutions. Toxicity was analyzed by in vitro transcribing each LHE into mRNA and transfecting primary human T cells by electroporation, followed by flow cytometry analysis of the survival of the cells relative to transfection with a control mRNA encoding a blue fluorescent protein (BFP). See Figure 8 showing flow cytometry analysis of primary human T cells after electroporation with IVT-mRNA coding for TCRA_S02_2E5_RD1_08 LHE variant. These data show that the TCRA_S02 LHE variants have minimal toxicity in primary human cells.

Example 4: Transient expression of TRAC-targeting LHE was shown to cause loss of TCR-alpha protein from the cell surface and lead to disruptive mutations at the TRAC gene.

Next, the TRAC-targeting LHE was examined to determine whether : i) it efficiently cleaved the TCRA_S02 target site in the TRAC gene in human cells (SEQ ID NO: 3); and ii) the resulting NHEJ-mediated disruptions resulted in the loss of the TRAC protein from the cell surface. One primary motivation for achieving high efficiencies is in developing human therapeutic interventions based on TRAC disrupting nucleases. In such an application, using viral vectors which permanently (such as for retroviral, lentiviral, or foamy viral vectors) or transiently (such as adenoviral or adeno-associated viral vectors) deliver nuclease reagents is laborious, cost and resource-intensive, poorly scalable, and challenging to address from a regulatory perspective. A more attractive therapeutic

reagent and process would involve replacing the biological vector with a synthetic expression reagent, such as in vitro transcribed mRNA (IVT-mRNA). Primary human T cells were transfected with in-vitro transcribed mRNA (IVT-mRNA) encoding either BFP or the TCRA_S02_2E5 LHE, or variants as described above using methods well known to those familiar with the art. Electroporation of IVT-mRNA produce a transient burst of protein expression lasting 4 to 12 hours. The duration and extent of protein expression depends on the structure of IVT-mRNA. An example of the plasmid used for IVT-mRNA production is shown in Figure 9 (SEQ ID NO: 12). The secondary mRNA stability factors are added in the course of IVT-mRNA production. The 5' mRNA cap (m7G) regulates expression by binding to eukaryotic initiation factors (eIF). Addition of a polyadenylated (poly(A)) tail delays IVT-mRNA degradation and increase LHE protein expression.

Three day after electroporation of primary human T cells with IVT-mRNA encoding TRAC-targeting LHE and Trex2, greater than 15% of cells had lost cell surface expression of the TCR-alpha protein. See Figure 10 demonstrating the flow cytometric analysis of TRAC disruption and stability of TRAC disruption over a two week culture period. The TCR-alpha-negative cells were sorted using flow cytometry and the TCRA_S02 target site on the TRAC gene was sequenced to analyze the genetic disruption of TRAC gene and to confirm and characterize the spectrum of LHE-induced mutations. The TRAC gene is expressed in a mono-allelic fashion, due to silencing of the non-productive allele during T cell development. See Figure 11 for sequencing results showing that 50% of all TRAC alleles contained NHEJ disruption events at TCR_S02 target site. The critical conclusion is that the TCRA_S02_LHE variants effectively target the actively transcribed TRAC allele to block TCR-alpha expression.

Example 5: TRAC-targeting LHE was improved by fusion with TALE domains, enabling more efficient TRAC gene disruption with transient synthetic delivery methods.

The results of the examples provided above demonstrate that the nuclease reagents described herein are able to effectively generate primary cells lacking TCR-alpha expression. The key goal of the LHE enzyme and variants described wherein is generation of TCR-alpha deficient T cells for the treatment of cancer, autoimmunity and viral infection. See Figure 12 which schematically illustrates the proposed therapeutic strategy by which TRAC-targeting LHE is combined with a secondary CAR reagent to produce T cells for the treatment of cancer and other diseases. Our initial studies with TCRA_S02 targeting LHE and Trex2 delivery in the IVT-mRNA form showed promising but suboptimal overall rates of TRAC gene disruption. While not limiting, higher rates of

TCR-alpha disruption would simplify product manufacturing and result in reduced development costs.

A chimeric endonuclease architecture that could improve the efficiency of the TRAC-targeting LHE was created. As described herein, TALE proteins offer a uniquely modular mode of DNA recognition. The invertors therefore reasoned that an array of TALE repeats which recognized a target sequence adjacent to the TCRA_S02 target could be fused to the TCRA_S02 targeting LHE to effectively enhance the co-localization of the nuclease and its substrate. See Figure 13 which schematically illustrates the chimeric endonuclease and its recognition sequence (SEQ ID NO: 3). This chimeric endonuclease (SEQ ID NO: 14 encoded by SEQ ID NO: 13) – an architecture termed 'MegaTAL' – was then converted into mRNA by in vitro transcription methods described above. IVT-mRNA species encoding the TCRA_S02 targeting MegaTAL and the Trex2 exonuclease were then delivered by electroporation to primary human T cells. This method of transiently expressing these nuclease reagents resulted in extremely efficient removal of the TCR-alpha protein from the cell surface. See Figure 14 which demonstrates the effectiveness of this approach using flow cytometric analysis of TCR-alpha expression on primary T cells treated with IVT-mRNA species encoding the megaTAL and Trex2 exonuclease, as >65% of cells lacked TCR-alpha expression.

Example 6: TCRA-targeting, TALE-LHE fusions were improved by fusion with Trex2, enabling ultra-efficient TCRA gene disruption with a three-component fusion protein expressed from a single mRNA species.

Next, efficient TCRA gene disruption was evaluated for achievement by delivering a single mRNA species expressing a fusion protein comprising a TAL array, the TCRA.S02 targeting LHE, and Trex2. This three-component fusion protein (SEQ ID NO: 16 encoded by SEQ ID NO: 15) was placed in a vector containing a T7 promoter to facilitate *in vitro* transcription and subsequent polyadenylation and capping. The resulting mRNA was delivered to primary human T cells by electroporation and the expression of the CD3 complex on the cell surface was assessed 72 hours later by flow cytometry (Figure 16). Control samples included untransfected primary human T cells, T cells transfected with the TCRA.S02 targeting MegaTAL, and a sample where the TCRA.S02 targeting MegaTAL was cotransfected with an independently synthesized mRNA species encoding Trex2. The samples receiving Trex2 either independently or as a direct fusion with the TCRA.S02 targeting MegaTAL showed an increased percentage of CD3 negative cells, indicating enhanced TCRA gene disruption rates in these samples.

REFERENCES:

Arimondo, P. B., C. J. Thomas, et al. (2006). "Exploring the cellular activity of camptothecin-triple-helix-forming oligonucleotide conjugates." Mol Cell Biol **26**(1): 324-33.

Baxter, S., A. R. Lambert, et al. (2012). "Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases." Nucleic Acids Res.

Bitinaite, J., D. A. Wah, et al. (1998). "FokI dimerization is required for DNA cleavage." Proc Natl Acad Sci U S A **95**(18): 10570-5.

Boch, J., H. Scholze, et al. (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors." Science **326**(5959): 1509-12.

Choulika, A., A. Perrin, et al. (1995). "Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae." Mol Cell Biol **15**(4): 1968-73.

Christian, M., T. Cermak, et al. (2010). "Targeting DNA double-strand breaks with TAL effector nucleases." Genetics **186**(2): 757-61.

Cong, L., F. A. Ran, et al. (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-23.

Critchlow, S. E. and S. P. Jackson (1998). "DNA end-joining: from yeast to man." Trends Biochem Sci **23**(10): 394-8.

Eisenschmidt, K., T. Lanio, et al. (2005). "Developing a programmed restriction endonuclease for highly specific DNA cleavage." Nucleic Acids Res **33**(22): 7039-47.

Jarjour, J., H. West-Foyle, et al. (2009). "High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display." Nucleic Acids Res **37**(20): 6871-80.

Jinek, M., K. Chylinski, et al. (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-21.

Kalish, J. M. and P. M. Glazer (2005). "Targeted genome modification via triple helix formation." Ann N Y Acad Sci **1058**: 151-61.

Kim, Y. G., J. Cha, et al. (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain." Proc Natl Acad Sci U S A **93**(3): 1156-60.

Li, T., S. Huang, et al. (2011). "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain." Nucleic Acids Res **39**(1): 359-72.

Li, T., S. Huang, et al. (2011). "Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes." Nucleic Acids Res **39**(14): 6315-25.

Ma, J. L., E. M. Kim, et al. (2003). "Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences." Mol Cell Biol **23**(23): 8820-8.

Mali, P., L. Yang, et al. (2013). "RNA-guided human genome engineering via Cas9." Science **339**(6121): 823-6.

Moore, I., M. Samalova, et al. (2006). "Transactivated and chemically inducible gene expression in plants." Plant J **45**(4): 651-83.

Moscou, M. J. and A. J. Bogdanove (2009). "A simple cipher governs DNA recognition by TAL effectors." Science **326**(5959): 1501.

Pabo, C. O., E. Peisach, et al. (2001). "Design and selection of novel Cys2His2 zinc finger proteins." Annu Rev Biochem **70**: 313-40.

Padidam, M., M. Gore, et al. (2003). "Chemical-inducible, ecdysone receptor-based gene expression system for plants." Transgenic Res **12**(1): 101-9.

Paques, F. and P. Duchateau (2007). "Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy." Curr Gene Ther **7**(1): 49-66.

Perrin, A., M. Buckle, et al. (1993). "Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions." Embo J **12**(7): 2939-47.

Pingoud, A. and G. H. Silva (2007). "Precision genome surgery." Nat Biotechnol **25**(7): 743-4.

Porteus, M. H. and D. Carroll (2005). "Gene targeting using zinc finger nucleases." Nat Biotechnol **23**(8): 967-73.

Rouet, P., F. Smih, et al. (1994). "Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells." Proc Natl Acad Sci U S A **91**(13): 6064-8.

Rouet, P., F. Smih, et al. (1994). "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease." Mol Cell Biol **14**(12): 8096-106.

Sethuraman, J., A. Majer, et al. (2009). "Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa." Mol Biol Evol **26**(10): 2299-315.

Simon, P., F. Cannata, et al. (2008). "Sequence-specific DNA cleavage mediated by bipyridine polyamide conjugates." Nucleic Acids Res **36**(11): 3531-8.

Smith, J., J. M. Berg, et al. (1999). "A detailed study of the substrate specificity of a chimeric restriction enzyme." Nucleic Acids Res **27**(2): 674-81.

Smith, J., M. Bibikova, et al. (2000). "Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains." Nucleic Acids Res **28**(17): 3361-9.

Takeuchi, R., A. R. Lambert, et al. (2011). "Tapping natural reservoirs of homing endonucleases for targeted gene modification." Proc Natl Acad Sci U S A **108**(32): 13077-82.

Thierry, A. and B. Dujon (1992). "Nested chromosomal fragmentation in yeast using the meganuclease I-Sce I: a new method for physical mapping of eukaryotic genomes." Nucleic Acids Res **20**(21): 5625-31.

Wang, R., X. Zhou, et al. (2003). "Chemically regulated expression systems and their applications in transgenic plants." Transgenic Res **12**(5): 529-40.

Zuo, J. and N. H. Chua (2000). "Chemical-inducible systems for regulated expression of plant genes." Curr Opin Biotechnol **11**(2): 146-51.

CLAIMS

1. An I-*Onu*I or I-*Onu*I homologue, in which mutations have been introduced to obtain a variant that cleaves a target nucleic acid sequence within the T cell receptor alpha constant gene (TRAC).

2. The variant according to claim 1, wherein it cleaves a target nucleic acid sequence SEQ ID NO: 3.

3. The variant of claim 1 or 2 comprising at least 10, preferably at least 15, more preferably at least 20, even more preferably at least 25 amino acid substitutions in positions selected from the group consisting of: 19, 24, 26, 28, 30, 32, 34, 35, 36, 37, 38, 40, 42, 44, 46, 48, 68, 70, 72, 75, 76 77, 78, 80, 82, 138, 159, 168, 178, 180, 182, 184, 186, 188, 189, 190, 191, 192, 193, 195, 197, 199, 201, 203, 207, 223, 225, 227, 229, 231, 232, 234, 236, 238, 240, in reference to SEQ ID NO: 2.

4. The variant of claims 1 to 3 having at least 85%, preferably at least 90%, more preferably at least 95% sequence identity with the DNA recognition interface.

5. The variant of claims 1 to 3 comprising the protein sequence selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10,

6. The variant of claims 1 to 3 having at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, more preferably at least 97% and more preferably at 99% sequence identity with the protein sequence of SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10.

7. A chimeric endonuclease comprising the DNA recognition interface of an I-*Onu*I or I-*Onu*I homologue variant according to any one of claims 1 to 6, fused to at least an additional protein domain selected from the group consisting of: nucleic acid binding domain, catalytic domain, terminal epitope tags and fluorescent proteins.

8. The chimeric endonuclease of claim 7, wherein said additional protein domain is a nucleic acid binding domain selected from the group consisting of TALE and Zinc Finger domain.

9. The chimeric endonuclease of claim 8, wherein said chimeric endonuclease is the MegaTAL TCRA_S02 protein sequence of SEQ ID NO: 14 and SEQ ID NO: 16.

10. The chimeric endonuclease of claim 7 or 9, wherein additional protein domain has catalytic activity selected from the group consisting of: nuclease activity, polymerase activity, kinase activity, phosphatase activity, methylase activity, topoisomerase activity, integrase activity, transposase activity, ligase activity, helicase activity, recombinase activity.

11. The chimeric endonuclease of claim 10 wherein said catalytic domain is a 5'-3' exonuclease, more preferably Trex2 and more preferably single chain Trex2.

12. The chimeric endonuclease according to any one of claims 7 to 11, wherein said additional protein domain is fused to I-*Onu*I variant by a peptide linker.

13. A polynucleotide encoding an I-*Onu*I variant or chimeric endonuclease according to any one of claims 1 to 12.

14. A vector comprising a polynucleotide of claim 13.

15. A method for modifying TRAC gene in a cell comprising introducing an I-*Onu*I variant or the chimeric endonuclease according to any one of claims 1 to 12 into said cell.

16. The method according to claim 15, wherein the mutagenesis is increased by introducing into the cell a DNA end-processing enzyme.

17. The method according to claim 16, wherein said DNA end-processing enzyme is introduced into the cell with the *Onu-I* variant or the chimeric endonuclease as transgenes encoded by the same vector comprising nucleic acid sequence encoding ribosomal skip sequence.

18. The method of claim 16 or 17, wherein said DNA end-processing enzyme has 3'-5' exonuclease activity.

19. The method of claim 18, wherein said DNA end-processing enzyme is TREX2.

20. The method of claim 18, wherein said DNA end-processing enzyme is a single chain TREX2 polypeptide.

21. The method according to claim 15, comprising introducing a donor matrix that comprises a sequence to be introduced flanked by at least one sequence sharing homologies with at least one region of the TRAC gene surrounding the nucleic acid cleavage site of said rare-cutting endonuclease.

22. A method for treating or preventing cancer in a subject, the method comprising: (a) modifying a gene encoding T cell receptor-alpha in a cell by introducing an I-*Onul* variant or chimeric endonuclease according to any one of claims 1 to 12 into said cell, (b) introducing an signaling molecule recognizing a tumor antigen into the cell, and (c) administrating the cell into the subject.

23. The method of claim 22, wherein the signaling molecule is a Chimeric Antigen Receptor.

24. The method according to claims 17 to 23, wherein said cell is a T cell, a hematopoietic stem cell, a lymphoid progenitor cell, or a CD34+ cell.

25. A method for treating or preventing cancer in a subject, the method comprising administrating to a subject a vector of claim 14.

Position	Native	2E5	Refined
24	Ser	Ser	
26	Leu	Ile	
28	Arg	Asp	
30	Arg	Arg	
32	Asn	Arg	
34	Lys	Asn	
35	Ser	Glu	
36	Ser	Ser	
37	Val	Asn	
38	Gly	Arg	
40	Ser	Arg	
42	Glu	Ser	
44	Gly	Arg	
46	Gln	Glu	
48	Thr	Thr	
68	Val	Lys	
70	Ala	Thr	
72	Ser	Ser	
73	Gly		Ser
75	Asn	Arg	
76	Ala	Ala	
78	Ser	Met	
80	Lys	Arg	
82	Thr	Thr	
176	Ser		Ala
180	Cys	Tyr	
182	Phe	Gly	
184	Asn	Asn	
186	Ile	Lys	
188	Ser	Val	
189	Lys	Lys	
190	Ser	Gly	
191	Lys	Asn	Thr
192	Leu	Ala	
193	Gly	Lys	
195	Gln	Tyr	
197	Gln	Gly	
199	Val	Arg	
201	Ser	Ser	
203	Thr	Thr	Ser
223	Tyr	Ser	
225	Lys	Arg	Trp
227	Lys	Lys	
229	Lys	Lys	
232	Phe	Phe	
233	Ser		Arg
234	Trp	Trp	
236	Asp	Glu	
238	Val	Val	
240	Thr	Thr	

Table 1

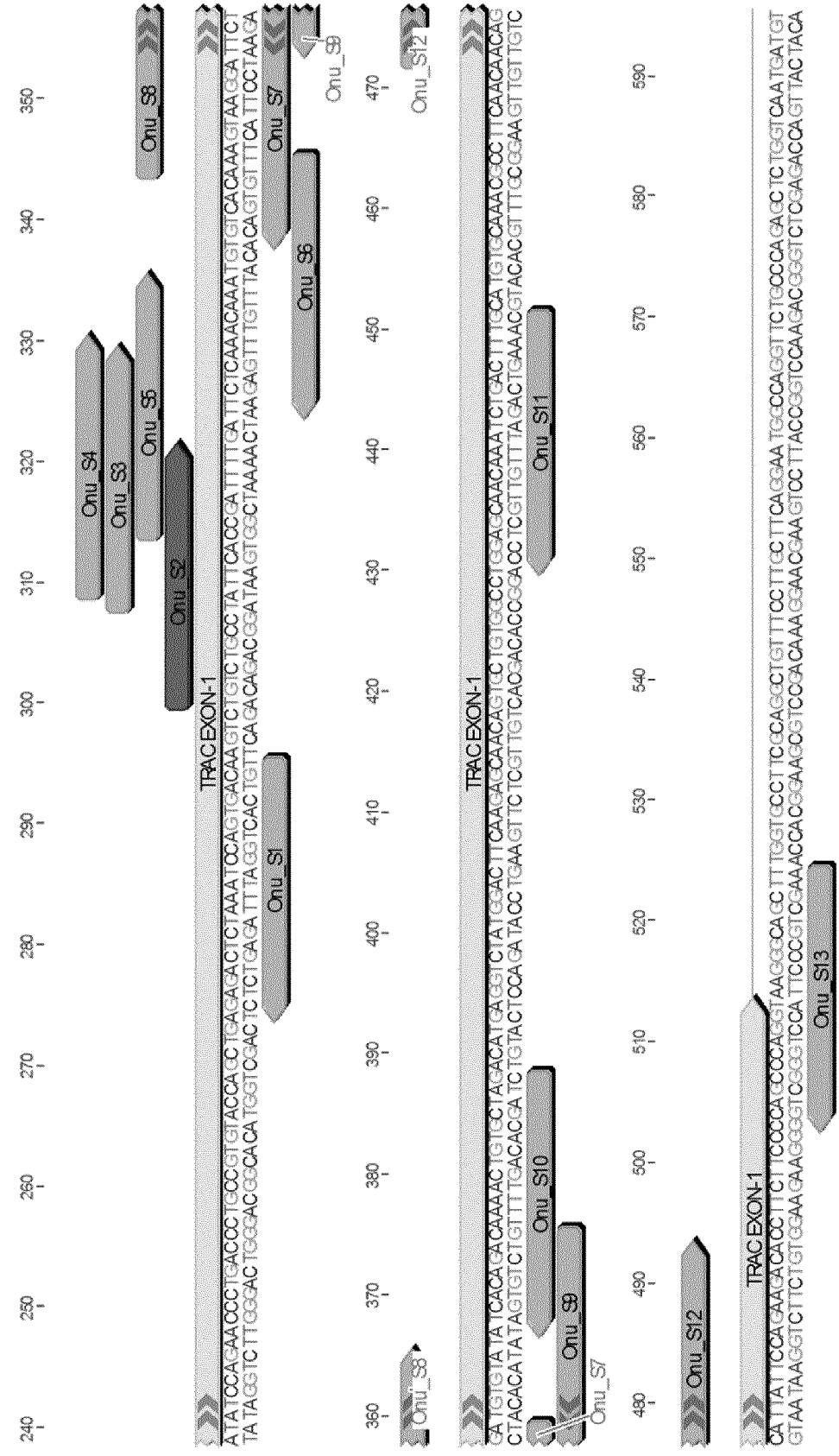
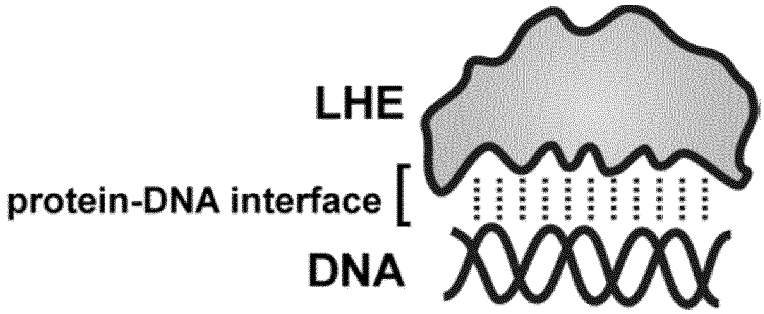

TRAC EXON-1
Onu_S1
Onu_S2
Onu_S3
Onu_S4
Onu_S5
Onu_S6
Onu_S7
Onu_S8
Onu_S9
Onu_S10
Onu_S11
Onu_S12
Onu_S13

Figure 1



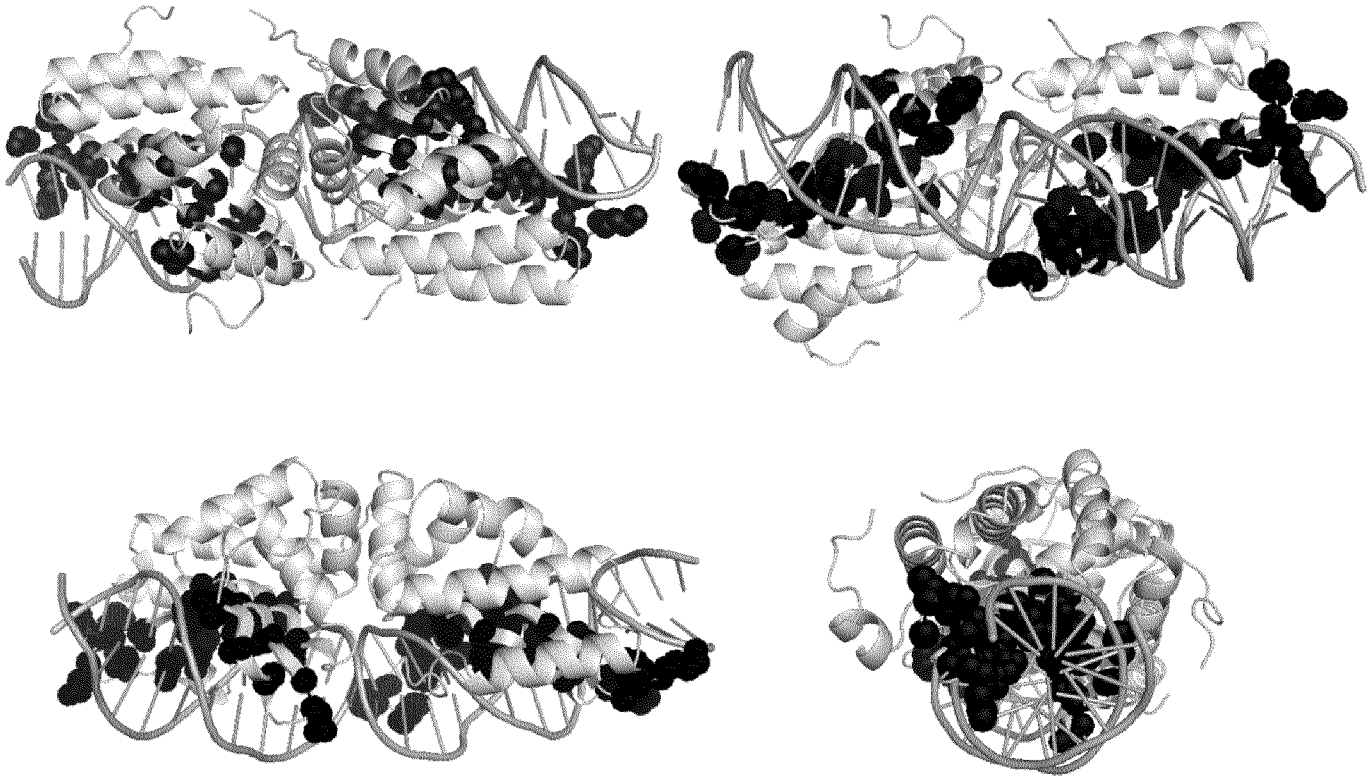


Figure 2

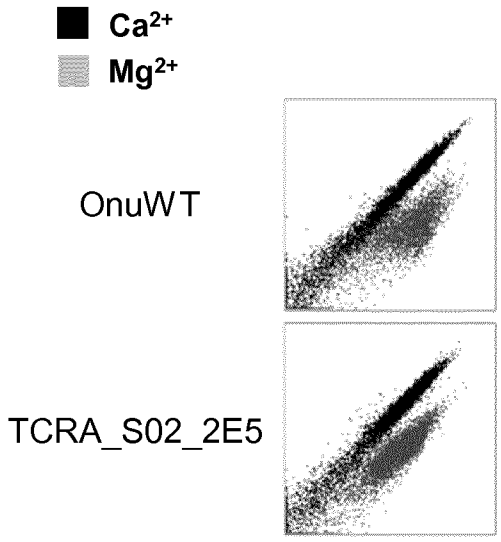


Figure 3

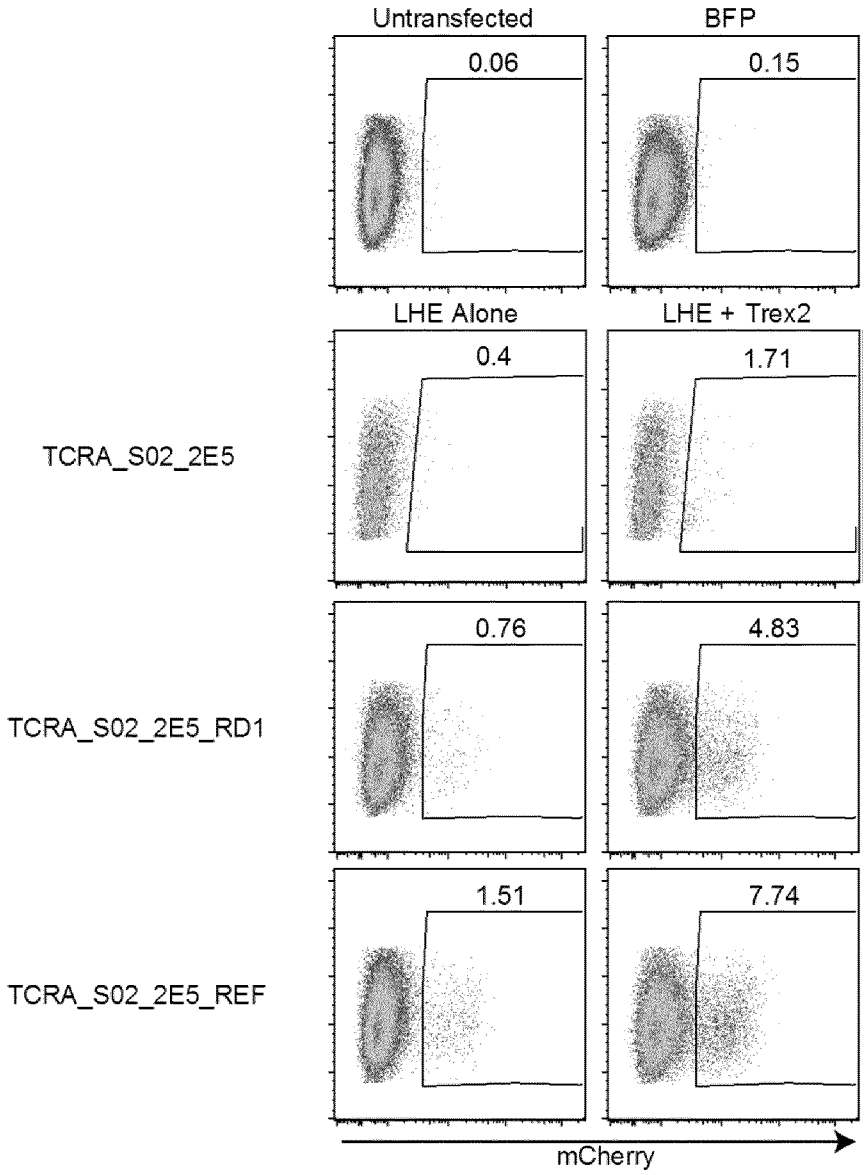

Untransfected
BFP
0.06
0.15
LHE Alone
LHE + Trex2
0.4
1.71
TCRA_S02_2E5
0.76
4.83
TCRA_S02_2E5_RD1
1.51
7.74
TCRA_S02_2E5_REF
mCherry

Figure 4

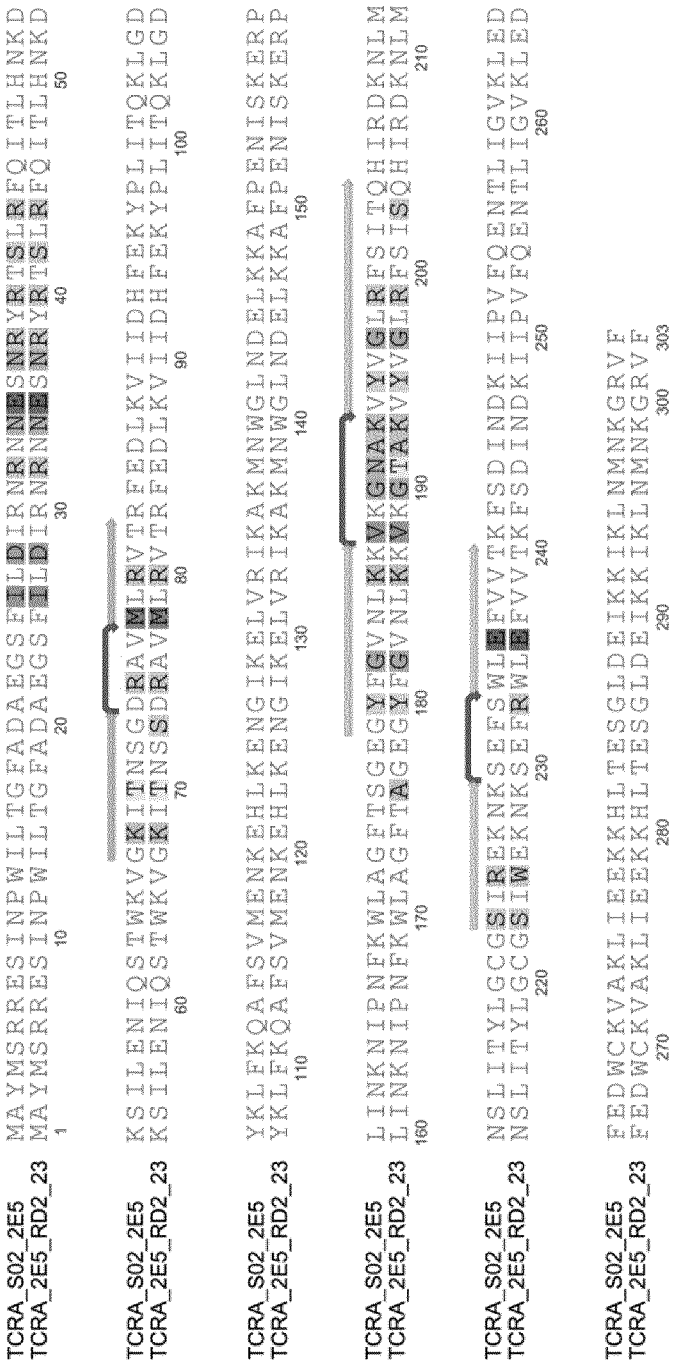

TCRA_S02_2E5
TCRA_2E5_RD2_23
MAYMSRRESINPWILTGFADAEGSFTLDIRNRNNESNRYRTSLRFQITLHNKD
MAYMSRRESINPWILTGFADAEGSFTLDIRNRNNESNRYRTSLRFQITLHNKD
KSILENIQSTWKVGKITNSGDRAVMLRVTRFEDLKVIIDHFEKYPLITQKLGD
KSILENIQSTWKVGKITNSSDRAVMLRVTRFEDLKVIIDHFEKYPLITQKLGD
YKLFKQAFSVMENKEHLKENGIKELVRIKAKMNWGLNDELKKAFPENISKERP
YKLFKQAFSVMENKEHLKENGIKELVRIKAKMNWGLNDELKKAFPENISKERP
LINKNIPNFKWLAGFTSGEGYFGVNLKKVKGNAKVYVGLRFSITQHIRDKNLM
LINKNIPNFKWLAGFTAGEGYFGVNLKKVKGTAKVYVGLRFSISQHIRDKNLM
NSLITYLGCGSIREKNKSEFSWLEFVVTKFSDINDKIIPVFQENTLIGVKLED
NSLITYLGCGSIWEKNKSEFRWLEFVVTKFSDINDKIIPVFQENTLIGVKLED
FEDWCKVAKLIEEKKHLTESGLDEIKKIKLNMNKGRVF
FEDWCKVAKLIEEKKHLTESGLDEIKKIKLNMNKGRVF

Figure 5

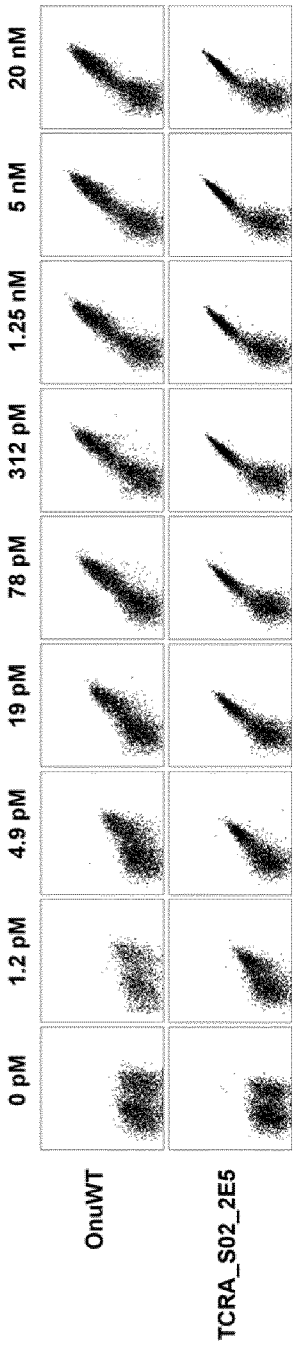

0 pM
1.2 pM
4.9 pM
19 pM
78 pM
312 pM
1.25 nM
5 nM
20 nM
OnuWT
TCRA_S02_2E5

Figure 6

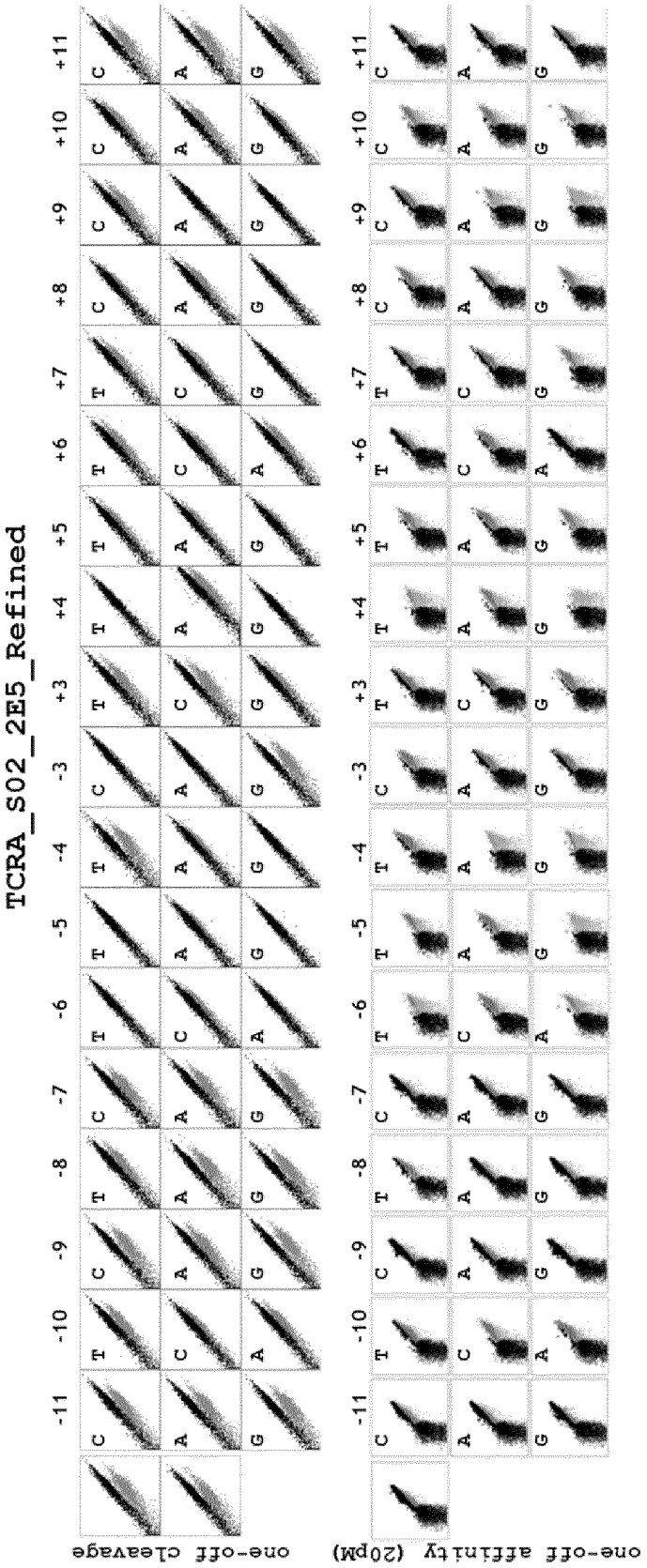

TCRA_S02_2E5_Refined
one-off cleavage
one-off affinity (20pM)
-11 -10 -9 -8 -7 -6 -5 -4 -3 +3 +4 +5 +6 +7 +8 +9 +10 +11
C T C T C T T T C T T T T T C C C C
A C A A A C A A A C A A C C A A A A
G A G G G A G G G G G G A G G G G G

Figure 7

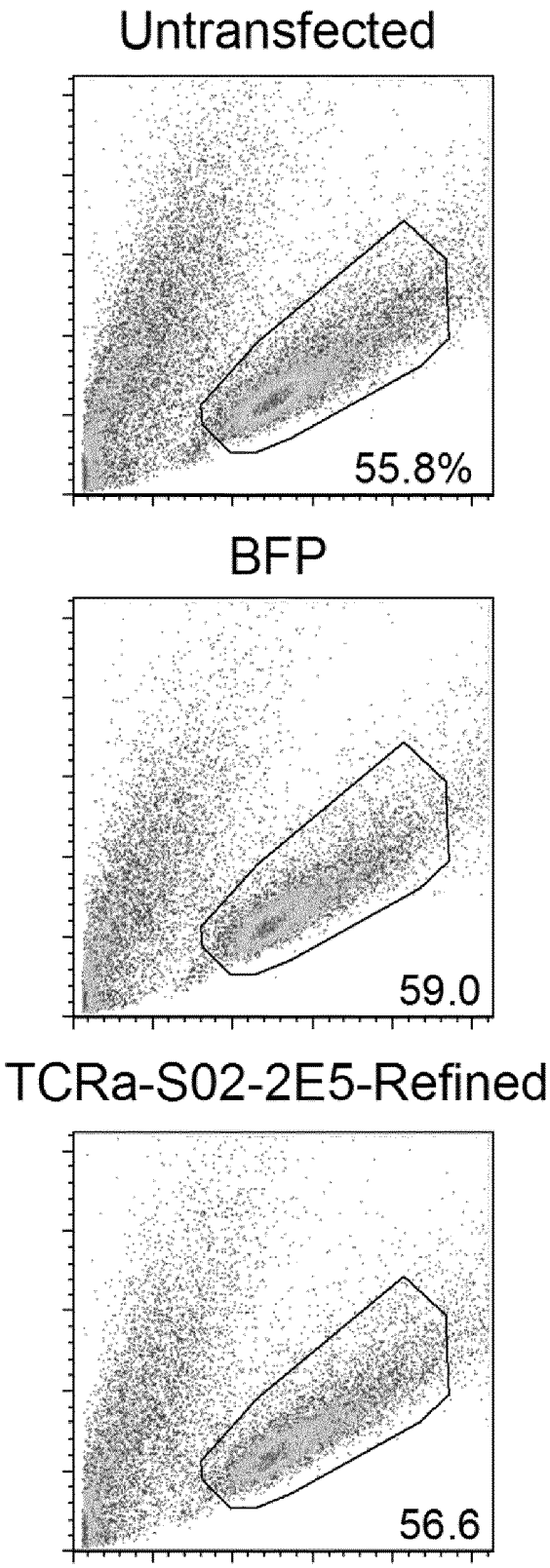


Figure 8

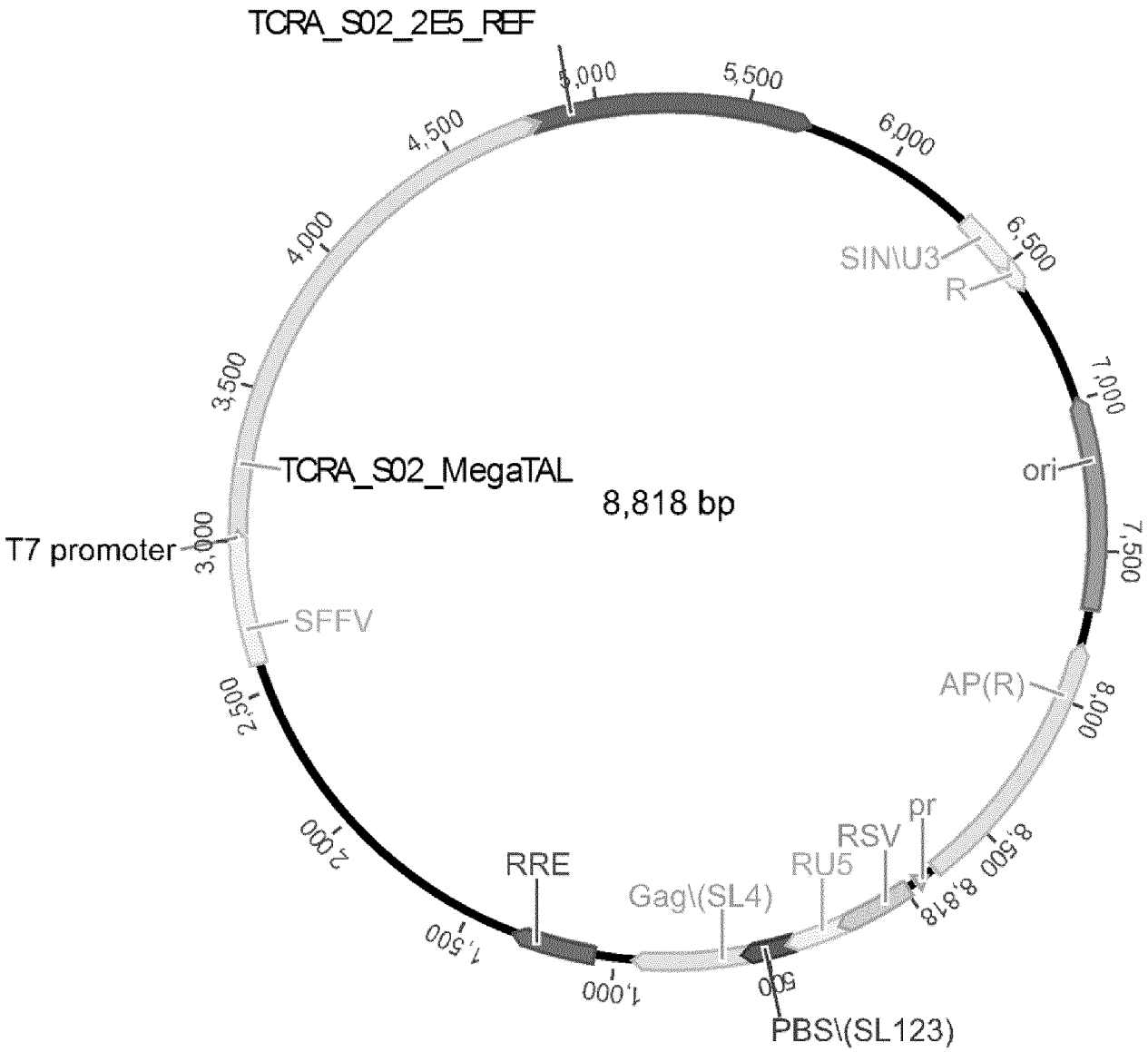

TCRA_S02_2E5_REF
5,000
5,500
6,000
6,500
SIN\U3
R
7,000
ori
7,500
8,000
AP(R)
8,500
8,818
pr
RSV
RU5
Gag\(SL4)
500
1,000
RRE
1,500
2,000
2,500
SFFV
3,000
T7 promoter
TCRA_S02_MegaTAL
3,500
4,000
4,500
8,818 bp
PBS\(SL123)

Figure 9

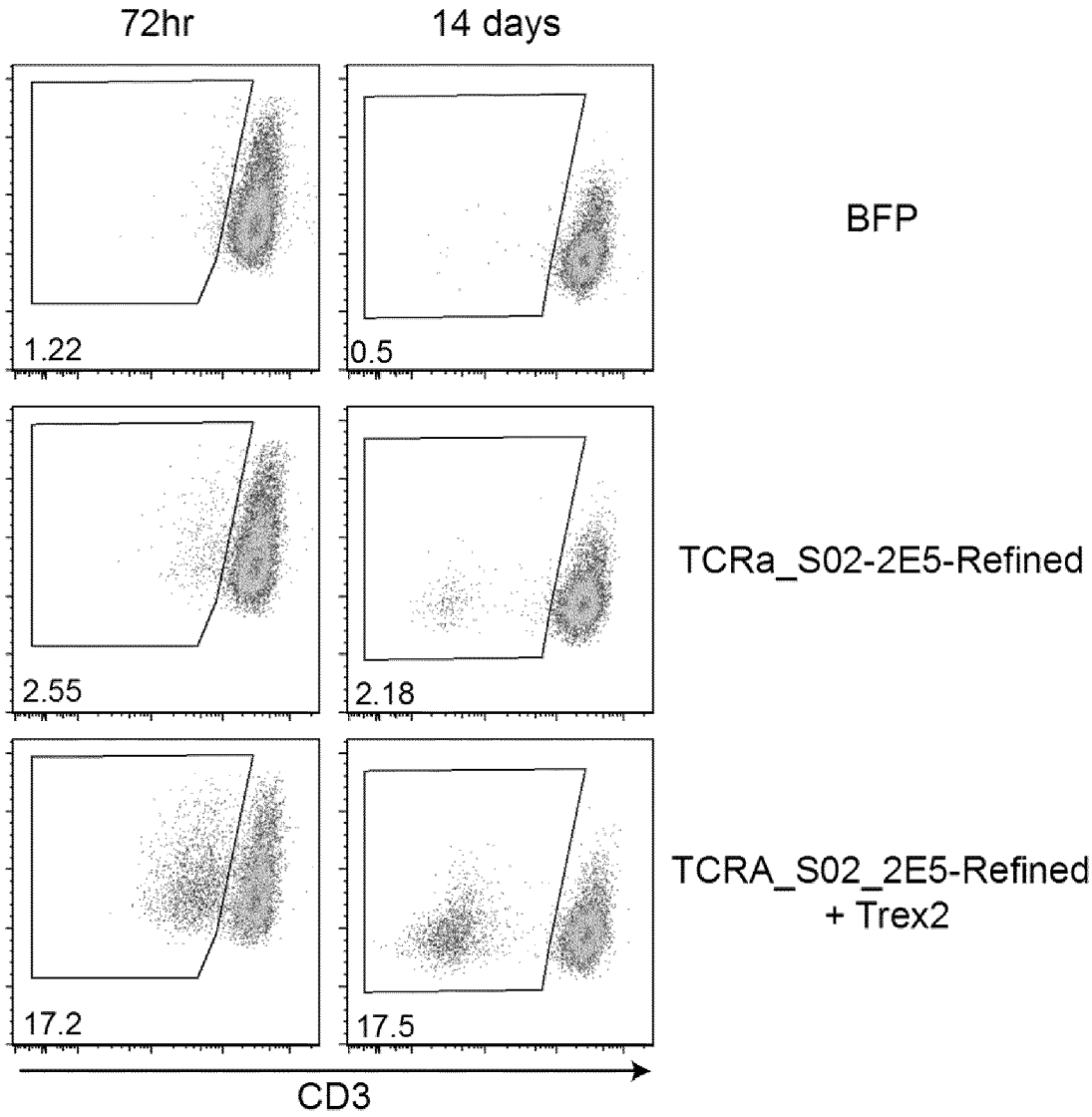

72hr
14 days
BFP
1.22
0.5
TCRa_S02-2E5-Refined
2.55
2.18
TCRA_S02_2E5-Refined
+ Trex2
17.2
17.5
CD3

Figure 10

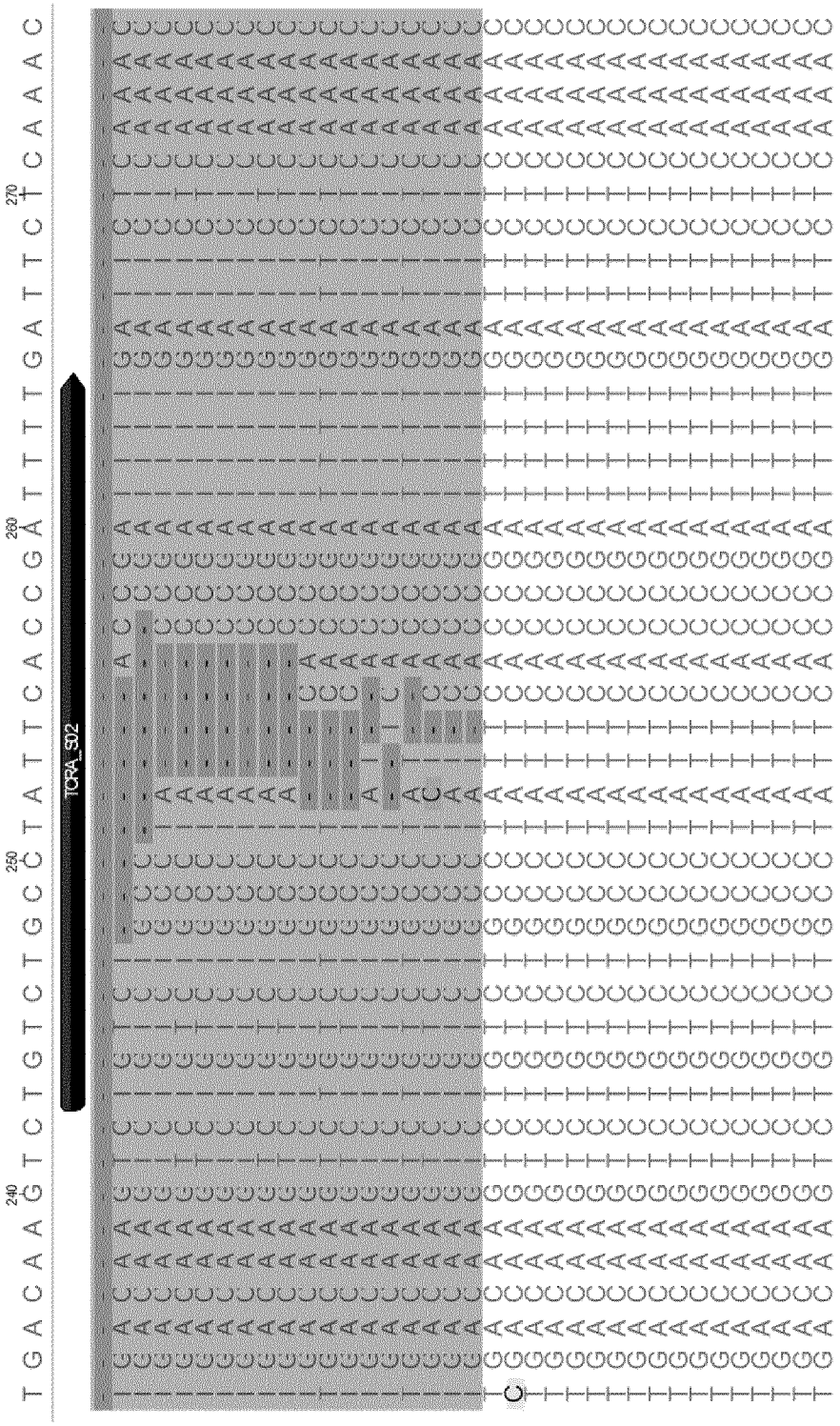


Figure 11

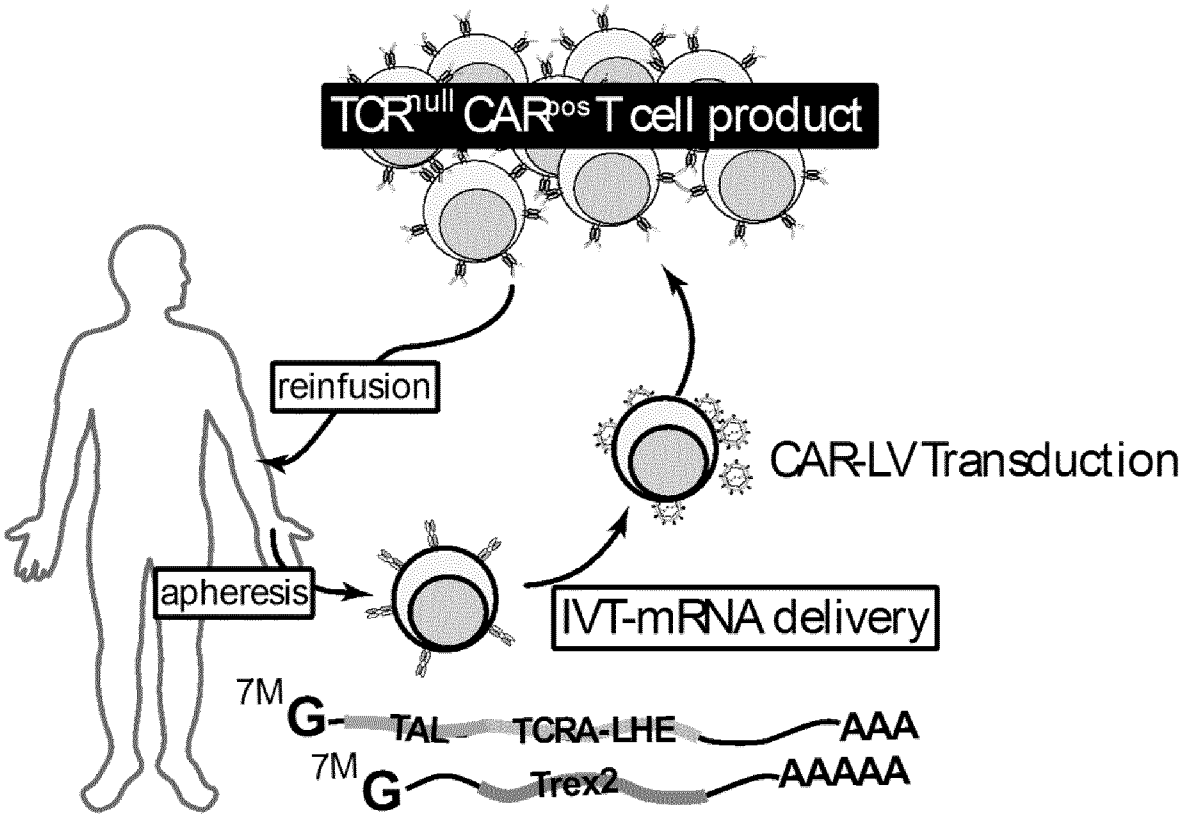

TCRnull CARpos T cell product
reinfusion
CAR-LV Transduction
apheresis
IVT-mRNA delivery
7M G
TAL
TCRA-LHE
AAA
7M G
Trex2
AAAAA

Figure 12

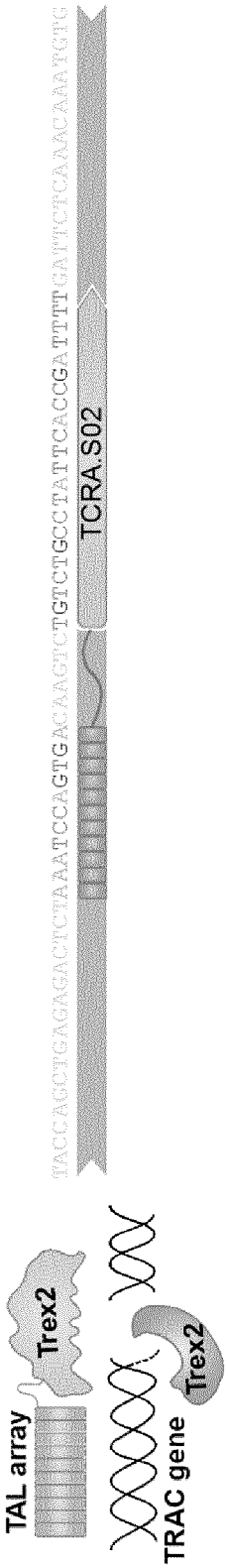


Figure 13

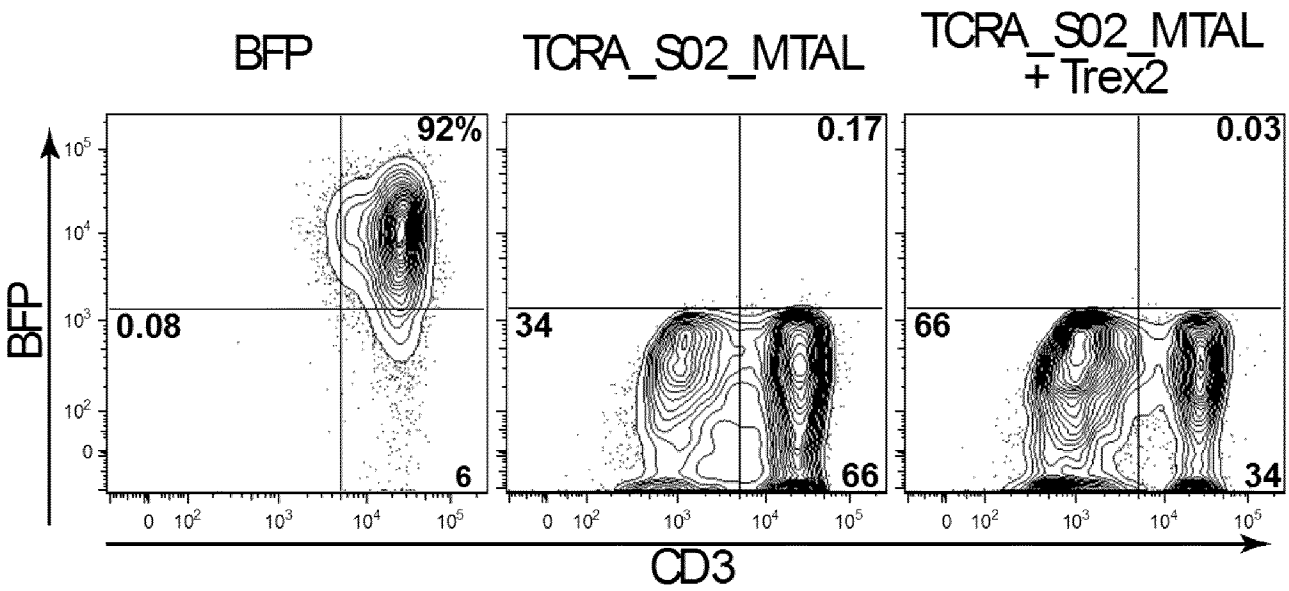

BFP
TCRA_S02_MTAL
TCRA_S02_MTAL + Trex2
92%
0.08
6
0.17
34
66
0.03
66
34
BFP
CD3
0
10^2
10^3
10^4
10^5

Figure 14

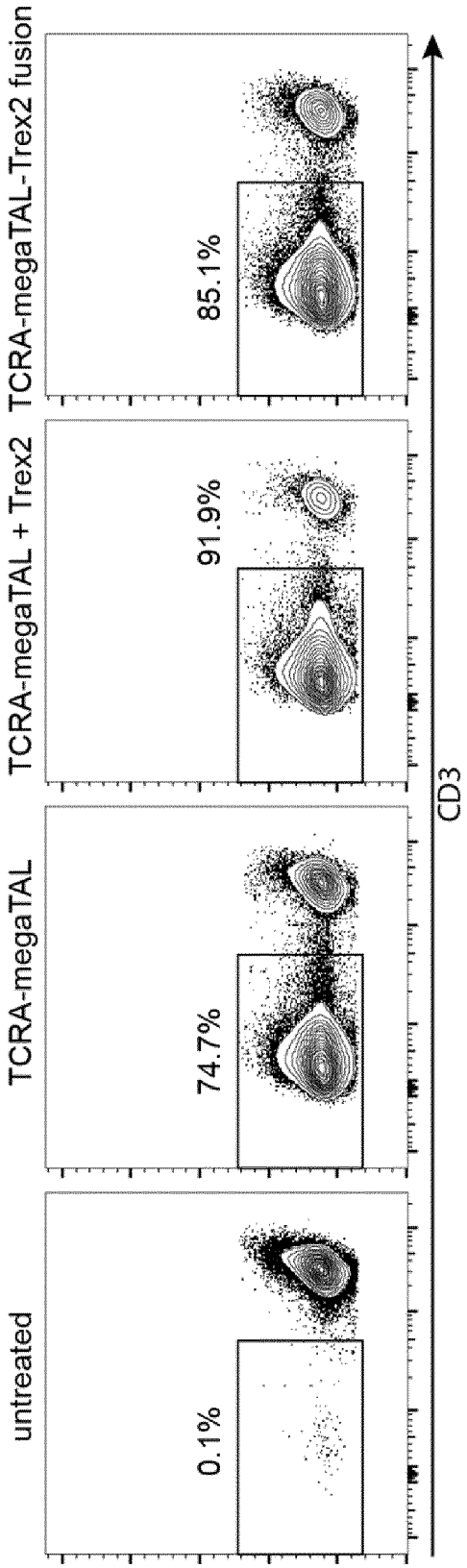

untreated
0.1%
TCRA-megaTAL
74.7%
TCRA-megaTAL + Trex2
91.9%
TCRA-megaTAL-Trex2 fusion
85.1%
CD3

Figure 15

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/061189

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/22
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, CHEM ABS Data, EMBASE, FSTA, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RYO TAKEUCHI ET AL: "Tapping natural reservoirs of homing endonucleases for targeted gene modification", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, NATIONAL ACADEMY OF SCIENCES, vol. 108, no. 32, 1 August 2011 (2011-08-01), pages 13077-13082, XP002706112, ISSN: 0027-8424, DOI: 10.1073/PNAS.1107719108 [retrieved on 2011-07-22] abstract; figures 1-5 page 13077, right-hand column, paragraph 2 - page 13079, left-hand column ----- -/--	1,2,4,6, 13,14

[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
22 October 2014	03/11/2014

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	Strobel, Andreas

4

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

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TORIKAI HIROKI ET AL: "A foundation for universal T-cell based immunotherapy: T cells engineered to express a CD19-specific chimeric-antigen-receptor and eliminate expression of endogenous TCR.", BLOOD 14 JUN 2012, vol. 119, no. 24, 14 June 2012 (2012-06-14), pages 5697-5705, XP055071623, ISSN: 1528-0020 abstract page 5697, left-hand column - page 5698, left-hand column, paragraph 3 page 5703, right-hand column, paragraph 2; figure 1	1-23
A	BAXTER SARAH ET AL: "Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases.", NUCLEIC ACIDS RESEARCH SEP 2012, vol. 40, no. 16, September 2012 (2012-09), pages 7985-8000, XP002707705, ISSN: 1362-4962 abstract; figure 3a page 7998, left-hand column, paragraph 2 - page 7999, left-hand column, paragraph 2	1-23
A	WO 2011/156430 A2 (HUTCHINSON FRED CANCER RES [US]; SEATTLE CHILDREN S HOSPITAL RES INST) 15 December 2011 (2011-12-15) claims 1-9; figures 8,9; sequences 40-44	1-25
A	PROVASI ELENA ET AL: "TCR Gene Editing Results in Effective Immunotherapy of Leukemia without the Development of GvHD", BLOOD, AMERICAN SOCIETY OF HEMATOLOGY, US, vol. 118, no. 21, 13 December 2011 (2011-12-13), page 307, XP009175817, ISSN: 0006-4971 the whole document	1-25
A	HAFEZ MOHAMED ET AL: "Homing endonucleases: DNA scissors on a mission", GENOME, vol. 55, no. 8, August 2012 (2012-08), pages 553-569, XP002730920, abstract page 557, right-hand column, paragraph 2 - page 560; tables 1,2	1-25

-/--

4

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	S. BOISSEL ET AL: "megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering", NUCLEIC ACIDS RESEARCH, vol. 42, no. 4, 26 November 2013 (2013-11-26), pages 2591-2601, XP055129962, ISSN: 0305-1048, DOI: 10.1093/nar/gkt1224 abstract page 2597, left-hand column, last paragraph - page 2598, left-hand column, paragraph 1; figure 5 -& Sandrine J Boissel ET AL: "Supplementary Information megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering", , 26 November 2013 (2013-11-26), pages 1-34, XP055145241, Retrieved from the Internet: URL:http://nar.oxfordjournals.org/content/suppl/2013/11/22/gkt1224.DC1/nar-02895-f-2013-File007.pdf [retrieved on 2014-10-08] figure S6b -----	1-4, 6-10, 12-16, 18-25

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2014/061189

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2011156430 A2	15-12-2011	US 2014148361 A1	29-05-2014
		WO 2011156430 A2	15-12-2011