GENE THERAPY

FIELD OF THE INVENTION

The present invention relates to the genetic modification of cells. More specifically, the present invention relates to the use of agents to improve the efficiency of gene editing and to improve the survival and/or engraftment of haematopoietic stem cells which have been gene edited.

BACKGROUND TO THE INVENTION

The haematopoietic system is a complex hierarchy of cells of different mature cell lineages. These include cells of the immune system that offer protection from pathogens, cells that carry oxygen through the body and cells involved in wound healing. All these mature cells are derived from a pool of haematopoietic stem cells (HSCs) that are capable of self-renewal and differentiation into any blood cell lineage. HSCs have the ability to replenish the entire haematopoietic system.

Haematopoietic cell transplantation (HCT) is a curative therapy for several inherited and acquired disorders. However, allogeneic HCT is limited by the poor availability of matched donors, the mortality associated with the allogeneic procedure which is mostly related to graft-versus-host disease (GvHD), and infectious complications provoked by the profound and long-lasting state of immune dysfunction.

Gene therapy approaches based on the transplantation of genetically modified autologous HSCs offer potentially improved safety and efficacy over allogeneic HCT. They are particularly relevant for patients lacking a matched donor.

The concept of stem cell gene therapy is based on the genetic modification of a relatively small number of stem cells. These persist long-term in the body by undergoing self-renewal, and generate large numbers of genetically “corrected” progeny. This ensures a continuous supply of corrected cells for the rest of the patient’s lifetime. HSCs are particularly attractive targets for gene therapy since their genetic modification will be passed to all the blood cell lineages as they differentiate. Furthermore, HSCs can be easily and safely obtained, for example from bone marrow, mobilised peripheral blood and umbilical cord blood.

Efficient long-term gene modification of HSCs and their progeny requires a technology which permits stable integration of the corrective DNA into the genome, without affecting HSC function. Accordingly, the use of integrating recombinant viral systems such as γ-retroviruses, lentiviruses and spumaviruses has dominated this field (Chang, A.H. et al. (2007) Mol. Ther. 15: 445-456). Therapeutic benefits have already been achieved in γ-

retrovirus-based clinical trials for Adenosine Deaminase Severe Combined Immunodeficiency (ADA-SCID; Aiuti, A. et al. (2009) N. Engl. J. Med. 360: 447-458), X-linked Severe Combined Immunodeficiency (SCID-X1; Hacein-Bey-Abina, S. et al. (2010) N. Engl. J. Med. 363: 355-364) and Wiskott-Aldrich syndrome (WAS; Boztug, K. et al. (2010) N. Engl. J. Med. 363: 1918-1927). In addition, lentiviruses have been employed as delivery vehicles in the treatment of X-linked adrenoleukodystrophy (ALD; Cartier, N. et al. (2009) Science 326: 818-823), and very recently for metachromatic leukodystrophy (MLD; Biffi, A. et al. (2013) Science 341: 1233158) and WAS (Aiuti, A. et al. (2013) Science 341: 1233151).

In addition to the use of retro- and lentiviral-based vectors, vectors derived from other viruses, such as adenoviruses and adeno-associated viruses (AAV), may also be utilised for the modification of haematopoietic stem and progenitor cells.

The scope of genetic engineering has recently broadened from gene replacement to targeted gene editing using engineered nucleases, which enable precise sequence modification of a locus of interest. Gene editing applications encompass targeted disruption of a gene coding sequence, precise sequence substitution for in situ correction of mutations and targeted transgene insertion into a predetermined locus. Gene editing is based on the design of artificial endonucleases that target a double-strand break (DSB) or nick into the sequence of interest in the genome. Cells repair the DSB through two major mechanisms, Non-Homologous End-Joining (NHEJ) or Homology Directed Repair (HDR), although other repair mechanisms might be eventually exploited. NHEJ often creates small insertions or deletions (“indels”) at the target site that may disrupt the coding sequence of a gene, whereas HDR can be exploited to precisely introduce a novel sequence at the target site by providing an exogenous template DNA bearing homology to the sequences flanking the DSB.

Multiple platforms of artificial endonucleases can be used to target a locus of interest, including Zinc Finger Nucleases (ZFNs), TAL effector nucleases (TALENs) and the more recently developed RNA-based CRISPR/Cas9 nucleases. Viral vectors are the most efficient delivery vehicle for a DNA template, for example, the AAV6 vector is able to achieve a high transduction efficiency in human primary cells, such as Hematopoietic Stem/Progenitor (HSPC) cells and T lymphocytes.

Despite recent advance in the generation of gene edited primary cells, several hurdles need to be solved before we can fully benefit from the predicted safety and precision of genetic engineering afforded by these new technologies. A major issue is that gene editing in primary cells, and in particular in the primitive HSPC subset, is constrained by gene transfer

efficiency and limited proficiency of homology directed DNA repair (HDR), likely due to HSC quiescence, to low levels of expression of the HDR machinery and conversely to high activity of the error-prone non homologous end joining (NHEJ) pathway. Thus, it will be crucial to enhance the efficiency of HDR in HSC while fully preserving their long-term repopulating activity.

Similarly, the impact of adeno-associated virus (AAV) as a source of donor template for HDR-mediated gene editing remains poorly investigated, and no clinical application of this vector in HSPC has been reported yet. AAV dose-dependent toxicity has been observed, which is directly related to G-rich regions of ITRs that induce cells accumulation in early S-phase due to p53-mediated induction of apoptosis, as described in a hESCs model.

Substantial difficulties remain with the methods employed for the genetic modification of haematopoietic stem and progenitor cells. In particular, the multiple hits of high vector doses required and prolonged ex vivo transduction times associated with existing methods give rise to problems with survival of the transduced haematopoietic stem and progenitor cells during culture and potentially impact their biological properties. Furthermore, improvements in the engraftment of transduced cells will greatly benefit clinical applications.

SUMMARY OF THE INVENTION

The inventors have surprisingly found that the use of an agent which promotes homology directed DNA repair, (such as an inhibitor of p53 activation, or an adenoviral protein) improves gene editing efficiency of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells and improves the survival and/or engraftment of treated gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

The inventors determined that advantageously, the agent which promotes homology directed DNA repair enhanced gene targeting efficiency in haematopoietic stem and progenitor cells, particularly in the more primitive CD34+CD133+CD90+ fraction. Furthermore, the agent does not significantly affect the differentiation status of the haematopoietic stem and/or progenitor cells, thus preserving their long-term re-population capacity.

In one aspect, the present invention provides an agent which promotes homology directed DNA repair for use in haematopoietic cell gene therapy, haematopoietic stem cell gene therapy, and/or haematopoietic progenitor cell gene therapy, wherein said gene therapy comprises gene editing.

In a preferred embodiment, the haematopoietic cells, haematopoietic stem cells and haematopoietic progenitor cells are human cells.

In one embodiment, the cells are HSCs. In one embodiment the cells are HSPCs.
In one embodiment, the haematopoietic stem and/or progenitor cells are CD34+ cells.

In one embodiment, the population of haematopoietic stem and/or progenitor cells comprises, is enriched in or substantially consists of CD34+ cells. The population of cells may be further enriched for a particular sub-population of cells, for example CD34+CD38- cells. The population of cells may be further enriched for a particular sub-population of cells, for example CD34+CD133+ and CD90+ cells.

In a preferred embodiment, the gene therapy is haematopoietic stem cell therapy and haematopoietic stem cells are gene edited.

In one aspect, the present invention provides an agent which promotes homology directed DNA repair, for use in increasing the survival and/or engraftment of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

In one embodiment, at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 50%, 75% or 90%, preferably at least 70%, more cells survive in culture for a period of time (e.g. about 6 or 12 hours, or 1, 2, 3, 4, 5, 6, 7 or more days, preferably about 2 days) when the cells have been exposed to the agent rather than in its absence. Preferably, the period of time begins with transduction of the cells with a viral vector.

The present invention may allow the timing of the pretreatment /conditioning of cells with cytokines to prime the cells (otherwise quiescent) for transduction to be reduced. Standard timing is 3 days. This time may be reduced to less than 3 days, or less than 2 days, or less than 1 day, for example 0, 1 or 2 days.

In one embodiment, the agent substantially prevents or reduces apoptosis in the haematopoietic stem and/or progenitor cells, in particular during in vitro culture.

In one embodiment, at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 40%, 50% or 75%, preferably at least 25%, fewer cells become apoptotic following culture for a period of time (e.g. about 6 or 12 hours, or 1, 2, 3, 4, 5, 6, 7 or more days, preferably about 2 days) when the cells have been exposed to the agent rather than in its absence. Preferably, the period of time begins with the transduction of the cells with a viral vector.

In one embodiment, at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 40%, 50% or 75%, preferably at least 10%, more transplanted haematopoietic stem cells and/or haematopoietic progenitor cells and/or their descendant cells (e.g. graft-derived cells) engraft in a host subject when the cells have been exposed to the agent rather than in its absence.

In another aspect, the present invention provides the use of an agent which promotes homology directed DNA repair, for increasing the survival and/or engraftment of gene edited haematopoietic stem and/or progenitor cells or for increasing the efficiency of gene editing of haematopoietic stem and/or progenitor cells.

In one embodiment, the agent is an inhibitor of p53 activation, preferably wherein the inhibitor is an inhibitor of p53 phosphorylation, more preferably an inhibitor of p53 Serine 15 phosphorylation.

In one embodiment, the inhibitor is a p53 dominant negative peptide, an ataxia telangiectasia mutated (ATM) kinase inhibitor, and/or an ataxia telangiectasia and Rad3-related protein (ATR) inhibitor.

In one embodiment, the inhibitor is pifithrin-α or a derivative thereof. In one embodiment, the inhibitor is pifithrin-α cyclic. In one embodiment, the inhibitor is pifithrin-α p-nitro. KU-55933 or a derivative thereof; GSE56 or a fragment or variant thereof; KU-60019, BEZ235, wortmannin, CP-466722, Torin 2, CGK 733, KU-559403, AZD6738 or derivatives thereof; or an siRNA, shRNA, miRNA or antisense DNA/RNA.

In one embodiment, the inhibition of p53 in the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells is transient.

In one embodiment, the inhibitor is a transient inhibitor (e.g. has an inhibitory action lasting less than about 1, 2, 3, 4, 5, 6, 7 or 14 days), such as a reversible inhibitor. Preferably, the cells are exposed to the inhibitor for about 1-48 or 1-24 hours, preferably 1-24 hours. The cells may be, for example, exposed to the inhibitor at the same time as the viral vector or before the viral vector.

In one embodiment, the agent is administered before, at the same time as, and/or after the gene editing machinery is introduced into the cell.

In one embodiment, the transient inhibition of p53 occurs during gene editing of the haematopoietic cells, haematopoietic stem cells and/or progenitor cells.

In one embodiment, the inhibitor is added to the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells at a concentration of about 1-50 μM.

In one embodiment, the inhibitor is added to the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells (e.g. in an in vitro or ex vivo culture) at a concentration of about 1-50, 1-40, 1-30, 1-20 or 1-15 μM, preferably about 1-15 μM. In another embodiment, the inhibitor is added to the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells at a concentration of about 5-50, 5-40, 5-30, 5-20 or 5-15 μM, preferably about 5-15 μM.

In one embodiment, the inhibitor is added to the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells (e.g. in an in vitro or ex vivo culture) at a concentration of about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45 or 50 μM, preferably about 10 μM.

In one embodiment, the agent comprises at least one adenoviral protein.

In one embodiment, the agent comprises a nucleic acid sequence encoding at least one adenoviral protein.

Preferably, the at least one adenoviral protein is E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76. The at least one adenoviral protein may be a variant or fragment of E4ORF1, when the variant or fragment substantially retains the biological activity of the full length E4ORF1, e.g. the ability to increase the survival and/or engraftment of gene edited haematopoietic stem/progenitor cells defined herein.

Preferably, the at least one adenoviral protein is E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, or SEQ ID Nos. 77-107. The at least one adenoviral protein may be a variant or fragment of E4ORF6/7, when the variant or fragment substantially retains the biological activity of the full length E4ORF6/7, e.g. the ability to increase the survival and/or engraftment of gene edited haematopoietic stem/progenitor cells defined herein.

In one embodiment, the agent comprises adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76; and adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, or SEQ ID Nos. 77-107.

In one embodiment, the agent comprises a nucleic acid sequence encoding adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76; and a nucleic acid sequence adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, or SEQ ID Nos. 77-107.

The adenoviral protein is not limited to a particular Adenovirus serotype. For example, in one embodiment, the at least one adenoviral proteins is from an Adenovirus of serotype 4, Adenovirus of serotype 5, Adenovirus of serotype 7 and/or Adenovirus of serotype 9. In one embodiment, the at least one adenoviral protein is from an Adenovirus of serotype 5.

In one embodiment, the agent comprises two compounds as defined herein that promote homology directed repair. In one embodiment, the agent comprises the inhibitor of p53 activation and the adenoviral protein, or a nucleotide sequence encoding therefor. Preferably, the inhibitor of p53 activation is GSE56 or a variant thereof.

In one embodiment, the agent comprises:

(a) the inhibitor of p53 activation and adenoviral protein E4ORF1 (or a nucleotide sequence encoding therefor). In one embodiment, the E4ORF1 has the amino acid sequence of E4ORF1 set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76.;

(b) the inhibitor of p53 activation and adenoviral protein E4ORF6/7 (or a nucleotide sequence encoding therefor). In one embodiment, the E4ORF6/7 has the amino acid sequence of E4ORF6/7 set forth in SEQ ID No. 2; or SEQ ID Nos. 77-107.

(c) the inhibitor of p53 activation, adenoviral protein E4ORF1 (or a nucleotide sequence encoding therefor), preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76, and adenoviral protein E4ORF6/7 (or a nucleotide sequence encoding therefor), preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2; or

SEQ ID Nos. 77-107. Preferably, the inhibitor of p53 activation is GSE56 or a variant thereof.

In one embodiment, the agent is a composition comprising the inhibitor of p53 activation and the adenoviral protein, or a nucleotide sequence encoding therefor.

In one embodiment, the inhibitor of p53 activation is administered simultaneously, sequentially or separately in combination with the adenoviral protein, or a nucleotide sequence encoding therefor.

In one embodiment, the adenoviral protein, or a nucleotide sequence encoding therefor, is administered simultaneously, sequentially or separately in combination with the inhibitor of p53 activation.

In one aspect, the invention provides a product comprising (a) the inhibitor of p53 activation of the invention; and (b) the adenoviral protein of the invention, as a combined preparation for simultaneous, separate or sequential use in therapy, wherein the use is a use of the disclosure.

The term “combination”, or terms “in combination”, “used in combination with” or “combined preparation” as used herein may refer to the combined administration of two or more entities simultaneously, sequentially or separately.

The term “simultaneous” as used herein means that the entities are administered concurrently, i.e. at the same time.

The term “sequential” as used herein means that the entities are administered one after the other.

The term “separate” as used herein means that the entities are administered independently of each other but within a time interval that allows the entities to show a combined, preferably synergistic, effect. Thus, administration “separately” may permit one entity to be administered, for example, within 1 minute, 5 minutes or 10 minutes after the other.

In one embodiment, the nucleic acid encoding the adenoviral protein is an mRNA.

In one embodiment, the adenoviral protein is expressed transiently in the haematopoietic cell, haematopoietic stem cell and/or haematopoietic progenitor cell, preferably the transient expression occurs during gene editing of the haematopoietic cell, haematopoietic stem cell and/or progenitor cell.

In one embodiment, the target of the gene editing is selected from the group consisting of CD40L, RAG-1, IL-2RG, CYBA, CYBB, NCF1, NCF2, and NCF4. In one embodiment, the target of the gene editing is a gene mutated in chronic granulomatous disease or the gene mutated SCID, atypical SCID and Omenn syndrome, or Hyper IgM syndrome.
Preferably the target of the gene editing is CD40L.

In another aspect, the present invention provides a method of gene editing a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells comprising introducing gene editing machinery to the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells; and

wherein prior to, at the same time as, or after the introducing of the gene editing machinery to the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells, the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells is contacted with an agent which promotes homology directed DNA repair;

In another aspect, the present invention provides a method of gene editing a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells comprising the steps:

(a) contacting the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells with an agent which promotes homology directed DNA repair;

(b) introducing gene editing machinery to the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells; and

(c) editing the genome of said haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

In another aspect, the present invention provides a method of gene editing a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells comprising the steps:

(a) introducing gene editing machinery to the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells;

(b) contacting the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells with an agent which promotes homology directed DNA repair; and

(c) editing the genome of said haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

In another aspect, the present invention provides a method of gene editing a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells comprising the steps:

(a) introducing gene editing machinery and an agent which promotes homology directed DNA repair to the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells;

(b) editing the genome of said haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

In one embodiment, the method steps (a) and (b) are carried out *ex vivo* or *in vitro*. Thus, in one embodiment of the method of gene editing a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells, steps (a) and (b) are carried out ex vivo or in vitro. In one embodiment steps (a) and (b) may be carried out simultaneously. In another embodiment, steps (a) and (b) are carried out sequentially, either step (a) before step (b) or step (b) before step (a).

In one embodiment, the gene editing machinery may comprise a nuclease such as a zinc finger nuclease (ZFNs), a transcription activator like effector nucleases (TALENs), meganucleases, or the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system.

The gene editing machinery (e.g.CRISPR/Cas system) may comprise one or more guide RNAs complementary to at least one target gene in a cell, an RNA-guided DNA endonuclease enzyme or nucleotide sequence encoding said endonuclease (e.g. Cas9 protein or a nucleotide sequence encoding a Cas9). In one embodiment, the gene editing machinery may be a CRISPR/Cas system.

In one embodiment, the gene editing machinery may be provided by one or more nucleotide sequences. Suitably, the nucleotide sequences encoding the gene editing machinery may be introduced to the cell sequentially or simultaneously. In one embodiment, the agent which promotes HDR may be contacted with the cell simultaneously with the introduction of gene

editing machinery to the cell. In one embodiment, one or more nucleotide sequences encoding gene editing machinery is introduced to the cell by electroporation. In one embodiment, one or more nucleotide sequences is introduced to the cell by transduction. Suitably, the nucleotide sequence may be introduced by transduction of a viral vector. For example, a Cas9 ribonucleprotein may be introduced to a cell by electroporation before AAV6 transduction for the delivery of the donor DNA template.

As used herein, the term "introducing" refers to methods for inserting foreign DNA or RNA into a cell. As used herein the term introduced includes both transduction and transfection methods. Transfection is the process of introducing nucleic acids into a cell by non-viral methods. Transduction is the process of introducing foreign DNA or RNA into a cell via a viral vector.

In one embodiment, AAV transduction is used to deliver the donor DNA template.

In one embodiment, AAV6 transduction is used to deliver the donor DNA template.

In one embodiment, the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells (e.g. the population of cells in step (a) of the method of gene editing a population of said cells are contacted with the agent about 15 minutes to about 4 hours; about 15 minutes to about 3 hours; or about 15 minutes to about 2 hours before transducing the population of cells with the viral vector. In another embodiment, the cells are contacted with the agent about 1-4 hours; 1-3 hours; or 1-2 hours before transducing the population of cells with the viral vector.

In one embodiment, the haematopoietic cells, haematopoietic stem cells and/or haematopoietic cell progenitor cells (e.g. the population of cells in step (a) of the method of gene editing a population of cells of the invention) are contacted with the agent about 15 minutes, 30 minutes, 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours or 4 hours, preferably about 15 minutes, before transducing the population of cells with the viral vector.

Suitably, the agent may be active (e.g. may promote homology directed DNA repair) during gene editing.

In one embodiment, the contacting of step (a) is for about 12-60 h, such as 24-60, 36-60 or 42-54 h, preferably about 42-54 h, before step (b) is started. In one embodiment, the

contacting of step (a) is for about 12, 18, 24, 30, 36, 42, 48, 54 or 60 h, preferably about 48 h, before step (b) is started.

In one embodiment, step (b) is carried out about 12-60 h, such as 24-60, 36-60 or 42-54 h, preferably about 42-54 h, after beginning culture of the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells (e.g. beginning culture after the cells are thawed from a frozen state). In one embodiment, step (b) is carried out about 12, 18, 24, 30, 36, 42, 48, 54 or 60 h, preferably about 48 h, after beginning culture of the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

In one embodiment, step (b) is carried out about 12-60 h, such as 24-60, 36-60 or 42-54 h, preferably about 42-54 h, after thawing the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells (e.g. which had been stored in a frozen state). In one embodiment, step (b) is carried out about 12, 18, 24, 30, 36, 42, 48, 54 or 60 h, preferably about 48 h, after thawing the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

In one embodiment, the agent which promotes homology directed DNA repair is an inhibitor of p53 activation.

In one embodiment, the inhibitor of p53 activation is a p53 dominant negative peptide, an ataxia telangiectasia mutated (ATM) kinase inhibitor, or an ataxia telangiectasia and Rad3-related protein (ATR) inhibitor.

In one embodiment, the inhibitor of p53 activation is pifithrin-α or a derivative thereof, including pifithrin-α cyclic or pifithrin-α p-nitro; KU-55933 or a derivative thereof; GSE56 or a variant thereof; KU-60019, BEZ235, wortmannin, CP-466722, Torin 2, CGK 733, KU-559403, AZD6738 or derivatives thereof; or an siRNA, shRNA, miRNA or antisense DNA/RNA. Preferably, the inhibitor of p53 activation is a dominant negative peptide. Preferably the inhibitor of p53 activation is GSE56 or a variant thereof.

In one embodiment, the inhibitor is added to the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells at a concentration of about 1-50 μM.

In one embodiment, the agent comprises a nucleic acid sequence encoding at least one adenoviral protein.

Preferably, the at least one adenoviral protein is E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76. The at least one adenoviral protein may be a variant or fragment of E4ORF1, when the variant or fragment substantially retains the biological activity of the full length E4ORF1, e.g. the ability to increase the survival and/or engraftment of gene edited haematopoietic stem/progenitor cells defined herein.

Preferably, the at least one adenoviral protein is E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, or SEQ ID Nos. 77-107. The at least one adenoviral protein may be a variant or fragment of E4ORF6/7, when the variant or fragment substantially retains the biological activity of the full length E4ORF6/7, e.g. the ability to increase the survival and/or engraftment of gene edited haematopoietic stem/progenitor cells defined herein.

In one embodiment, the agent comprises adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76; and adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, or SEQ ID Nos. 77-107.

In one embodiment, the agent comprises a nucleic acid sequence encoding adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76; and a nucleic acid sequence adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, or SEQ ID Nos. 77-107.

The adenoviral protein is not limited to a particular Adenovirus serotype. For example, in one embodiment, the at least one adenoviral proteins is from an Adenovirus of serotype 4, Adenovirus of serotype 5, Adenovirus of serotype 7 and/or Adenovirus of serotype 9. In one embodiment, the at least one adenoviral protein is from an Adenovirus of serotype 5.

In one embodiment, the agent comprises two compounds as defined herein that promote homology directed repair. In one embodiment, the agent comprises the inhibitor of p53 activation and the adenoviral protein, or a nucleotide sequence encoding therefor. Preferably, the inhibitor of p53 activation is GSE56 or a variant thereof.

In one embodiment, the agent comprises:

(a) the inhibitor of p53 activation and adenoviral protein E4ORF1 (or a nucleotide sequence encoding therefor). In one embodiment, the E4ORF1 has the amino acid sequence of E4ORF1 set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76.;

(b) the inhibitor of p53 activation and adenoviral protein E4ORF6/7 (or a nucleotide sequence encoding therefor). In one embodiment, the E4ORF6/7 has the amino acid sequence of E4ORF6/7 set forth in SEQ ID No. 2; or SEQ ID Nos. 77-107.

(c) the inhibitor of p53 activation, adenoviral protein E4ORF1 (or a nucleotide sequence encoding therefor), preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, or SEQ ID Nos. 57-76, and adenoviral protein E4ORF6/7 (or a nucleotide sequence encoding therefor), preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2; or SEQ ID Nos. 77-107. Preferably, the inhibitor of p53 activation is GSE56 or a variant thereof.

Preferably, the inhibitor of p53 activation is GSE56 or a variant thereof.

In one embodiment, the agent is a composition comprising the inhibitor of p53 activation and the adenoviral protein, or a nucleotide sequence encoding therefor.

In one embodiment, the inhibitor of p53 activation is administered simultaneously, sequentially or separately in combination with the adenoviral protein, or a nucleotide sequence encoding therefor.

In one embodiment, the adenoviral protein, or a nucleotide sequence encoding therefor, is administered simultaneously, sequentially or separately in combination with the inhibitor of p53 activation.

In one embodiment, the nucleic acid encoding the adenoviral protein is an mRNA.

In one embodiment, the adenoviral protein is expressed transiently in the haematopoietic cell, haematopoietic stem cell and/or haematopoietic progenitor cell.

In one embodiment, the target of the gene editing is selected from the group consisting of CD40L, RAG-1, IL-2RG, CYBA, CYBB, NCF1, NCF2, and NCF4.. Preferably the target of the gene editing is CD40L.

In one embodiment, the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells is contacted with the agent which promotes homology directed DNA repair concurrently with gene editing.

In one embodiment, the population of haematopoietic stem and/or progenitor cells is obtained from mobilised peripheral blood, bone marrow or umbilical cord blood.

In one embodiment, the method according to the present invention includes a further step of enriching the population for haematopoietic stem and/or progenitor cells.

In another aspect, the present invention provides a method of gene therapy comprising the steps:

a) gene editing a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells according to the method of the present invention; and

b) administering the gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells to a subject.

In one embodiment, the gene edited cells are administered to a subject as part of an autologous stem cell transplant procedure or an allogeneic stem cell transplant procedure.

The method of gene therapy may be, for example, a method of treatment of a disease selected from the group consisting of mucopolysaccharidosis type I (MPS-1), chronic granulomatous disorder, Fanconi anaemia (FA), sickle cell disease, metachromatic leukodystrophy (MLD), globoid cell leukodystrophy (GLD), GM_2 gangliosidosis, thalassemia and cancer.

The method of gene therapy may be, for example, a method of treatment of diseases caused by Rag-1 mutations, eg SCID, atypical SCID and Omenn syndrome

The method of gene therapy may be, for example, a method of treatment of Hyper IgM syndrome (e.g. wherein the target of the gene editing is CD40L).

In one embodiment, the subject is a mammalian subject, preferably a human subject.

In one aspect, the present invention provides a population of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells prepared according to the method according to the present invention.

In another aspect, the present invention provides a pharmaceutical composition comprising the population of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells according to the present invention.

In yet another aspect, the present invention provides a population of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells according to the present invention for use in therapy, preferably for use in gene therapy.

In one aspect, the population of gene edited haematopoietic stem and/or progenitor cells for use according to the present invention is administered as part of an autologous stem cell transplant procedure or an allogeneic stem cell transplant procedure.

In another aspect the invention provides the use of an agent which promotes homology directed DNA repair for the manufacture of a medicament for haematopoietic cell, haematopoietic stem cell and/or haematopoietic progenitor cell gene therapy, where in the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells are gene edited.

In another aspect the invention provides the use of agent which promotes homology directed DNA repair for the manufacture of a medicament for increasing survival and/or engraftment of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells, where in the haematopoietic stem and/or progenitor cells are gene edited.

In another aspect the present invention provides an AAV vector comprising a barcode sequence.

In another aspect the present invention provides an AAV vector comprising a barcode sequence, for use in gene therapy, preferably haematopoietic gene therapy.

Preferred AAV vectors are those which are able to achieve a high transduction efficiency in human primary cells, such as HSPC cells. The AAV vector may be HDR-proficient. In one embodiment, the vector is an AAV6 vector or a vector derived from an AAV6 vector. Preferably the vector is an AAV6 vector.

A barcode sequence is a random, highly degenerated DNA sequence that acts as a unique identifier. Preferably the barcode sequence is a random DNA sequence. The DNA sequence may comprise or consist of at least 15 bp, or at least 18, 19 or 20 bp, for example, 20-40 bp, 20-30 bp, 20-25 bp, or about 22 bp. Preferably the barcode sequence does not comprise any undesired restriction sites (i.e. restriction sites for cloning restriction enzymes).

The AAV vector may further comprise a nucleotide sequence encoding a gene of interest. The "gene of interest" may be any gene. The gene may be a marker gene and/or a gene involved in human diseases.

The AAV vector may comprise the gene of interest and the barcode sequence such that the barcode sequence is able to report on whether the gene of interest is correctly integrated into a host genome and/or for clonal tracking of cells. In other words, the gene of interest and barcode sequence are flanked by homology sequences, such that during homology directed DNA repair the genome of interest and the barcode sequence are inserted into the host genome together.

The gene of interest may be operably linked to a promoter. The promoter may be any suitable promoter known to those of skill in the art. In some embodiments, the promoter is hPGK promoter.

The gene of interest may be operably linked to a polyadenylation sequence. The polyadenylation sequence may be any suitable polyadenylation sequence known to those of skill in the art. In some embodiments, the promoter is bGH-pA.

In some embodiments the AAV vector targets a safe harbour locus, preferably the AAVS1 locus, intron 1 of IL2RG, CD40L or RAG-1, most preferably the AAVS1 locus.

The AAV vector may comprise the nucleotide sequence shown below, or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity to the sequence below (barcode sequence underlined):

CCACTGAGAACCGGGCAGGTCACGCATCCCCCCCTTCCCTCCCACCCCCTGCCAAGCT
CTCCCTCCCAGGATCCTCTCTGGCTCCATCGTAAGCAAACCTTAGAGGTTCTGGCAAG
GAGAGAGATGGCTCCAGGAAATGGGGGTGTGTCACCAGATAAGGAATCTGCCTAACAG
GAGGTGGGGGTTAGACCCAATATCAGGAGACTAGGAAGGAGGAGGCCTAAGGATGGG
GCTTTTCTGTCACCAATCCTGTCCCTA<u>NNYNNNTNTNNNNRTNDNNNHH</u>CCATAGAGCC
CACCGCATCCCCAGCATGCCTGCTATTGTCTTCCCAATCCTCCCCCTTGCTGTCCTGCC
CCACCCCACCCCCCAGAATAGAATGACACCTACTCAGACAATGCGATGCAATTTCCTCA
TTTTATTAGGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGG

GAGGGGCAAACAACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGG
GTTTAAACGGGCCCTCTAGACTCGACGCGGCCGCTTTACTTGTACAGCTCGTCCATGC
CGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCT
CGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCA
GCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACG
CTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTT
GTCGGCCATGATATAGACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGG
ATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGT
CGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTTGAAGAAGAT
GGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTT
CATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCAGGGTGGTCACGAG
GGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTT
GCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTAC
GTCGCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTT
GCTCACCATGGTGGCGACCGGTGGGGAGAGAGGTCGGTGATTCGGTCAACGAGGGAG
CCGACTGCCGACGTGCGCTCCGGAGGCTTGCAGAATGCGGAACACCGCGCGGGCAG
GAACAGGGCCCACACTACCGCCCCACACCCCGCCTCCCGCACCGCCCCTTCCCGGCC
GCTGCTCTCGGCGCGCCCTGCTGAGCAGCCGCTATTGGCCACAGCCCATCGCGGTCG
GCGCGCTGCCATTGCTCCCTGGCGCTGTCCGTCTGCGAGGGTACTAGTGAGACGTGC
GGCTTCCGTTTGTCACGTCCGGCACGCCGCGAACCGCAAGGAACCTTCCCGACTTAGG
GGCGGAGCAGGAAGCGTCGCCGGGGGGCCCACAAGGGTAGCGGCGAAGATCCGGGT
GACGCTGCGAACGGACGTGAAGAATGTGCGAGACCCAGGGTCGGCGCCGCTGCGTTT
CCCGGAACCACGCCCAGAGCAGCCGCGTCCCTGCGCAAACCCAGGGCTGCCTTGGAA
AAGGCGCAACCCCAACCCCGTGGAAGCTTGCGTGGGGTGGAGGGGACAGATAAAAGT
ACCCAGAACCAGAGCCACATTAACCGGCCCTGGGAATATAAGGTGGTCCCAGCTCGGG
GACACAGGATCCCTGGAGGCAGCAAACATGCTGTCCTGAAGTGGACATAGGGGCCCG
GGTTGGAGGAAGAAGACTAGCTGAGCTCTCGGACCCCTGGAAGATGCCATGACAGGG
GGCTGGAAGAGCTAGCACAGACTAGAGAGGTAAGGGGGGTAGGGGAGCTGCCCAAAT
GAAAGGAGTGAGAGGTGACCCGAATCCACAGGAGAACGGGGTGTCCAGGCAAAGAAA
GCAAGAGGATGGAGAGGTGGCTAAAGCCAGGGAGACGGGGTACTTTGGGGTTGTCCA
GAAAAACGGTGATGATGCAGGCCTACAAGAAGGGGAGGCGGGACGCAAGGGAGACAT
CCGTCGGAGAAGGCCATCCTAAGAAACGAGAGATGGCACAGGCCCCAGAAGGAGAAG
GAAAAGGGAACCCAGCGAGTGAAGACGGCATGGGGTTGGGTGAGGGAGGAGAGATG
CCCGGAGAGGACCCAGACACGGGGAGGATCCGCTCAGAGGACATCACGTGGTGCAGC
GCCGAGAAGGAAGTGCTCCGGAAAGAGCATCCTTGGGCAGCAACACAGCAGAGAGCA
AGGGGAAGAGGGAGTGGAGGAAGACGGAACCTGAAGGAGGCGGCAGGGAAGGATCT

GGGCCAGCCGTAGAGGTGACCCAGGCCACAAGCTGCAGACAGAAAGCGGCACAGGC
CCAGGGGAGAGAATGCTGGTCAGAGAAAGCA

In another aspect the present invention provides for use of the AAV vector of the present invention for clonal tracking of cells that have been gene-edited using said vector, preferably wherein the cells are haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells. In one embodiment, the use is *in vitro* use.

In another aspect the present invention provides a method for clonal tracking of cells using the AAV vector of the present invention, preferably wherein the cells are haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

The method may comprise the following steps:

(a) providing an isolated population of cells;

(b) gene editing the isolated population of cells using the AAV vector of the present invention; and

(c) clonal tracking of the gene-edited cells using the barcode sequence of the AAV vector.

Preferably, the method further comprises: optionally, expanding the isolated population of gene-edited cells; and transplanting the isolated population of gene-edited cells into a subject.

Thus, in a preferred embodiment, the method comprises the steps of:

(a) providing an isolated population of cells;

(b) gene editing the isolated population of cells using the AAV vector of the present invention;

(c) optionally, expanding the isolated population of gene-edited cells;

(d) transplanting the isolated population of gene-edited cells into a subject; and

(e) clonal tracking of the gene-edited cells using the barcode sequence of the AAV vector.

More preferably, the method further comprises: harvesting blood and/or tissue samples from the subject; isolating genomic DNA from the blood and/or tissue samples; and quantifying the barcodes in the genomic DNA.

Thus, in a preferred embodiment, the method comprises the steps of:

(a) providing an isolated population of cells;

(b) gene editing the isolated population of cells using the AAV vector of the present invention;

(c) optionally, expanding the isolated population of gene-edited cells;

(d) transplanting the isolated population of gene-edited cells into a subject;

(e) harvesting blood and/or tissue samples from the subject;

(f) isolating genomic DNA from the blood and/or tissue samples; and

(g) quantifying the barcode sequences from the AAV vector in the genomic DNA.

The cells may be any suitable cells, as described herein. The cells may be haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells. The cells may be e.g. human or mouse cells. Preferably the cells are human cells and the subject is human. The cells may be CD34+ cells. In some embodiments the cells are HSCs. In some embodiments the cells are HSPCs.

The AAV vector may be transduced into the cells. Preferably the transduction is performed with 1 $x10^4$ to $5x10^4$ vg/cell.

Blood and/or tissue samples may be harvested after at least 1, 2, 3, 4, 8, 12, 24, 36 or 48 weeks. For example, blood and/or tissue samples may be harvested after about 1-100 weeks, 1-50 weeks, 1-20 weeks, 2-20 weeks, or 5-20 weeks.

Preferably, blood and/or tissue samples are harvested at least about every day, about every week, about every two weeks, or about every month. For example, in some embodiments, blood and/or tissue samples are harvested about every day, about every week, about every two weeks, or about every month.

In preferred embodiments the sample is a blood sample.

Clonal tracking or quantification of the barcode sequences may comprise the steps of:

(i) PCR amplification of the barcode sequences; and

(ii) sequencing the PCR-amplified barcode sequences.

Quantification may allow the number and distribution of gene-edited clones to be identified i.e. to allow for clonal tracking.

In another aspect the present invention provides use of a barcode sequence for clonal tracking of cells.

DESCRIPTION OF THE DRAWINGS

Figure 1. Single cell transcriptional impact of gene editing in HSPC. (A) Schematic representation of the gene editing protocol and scRNA-Seq analyses. Briefly, ~3000 cells for each experimental condition were FACS-enriched for CD133+ and CD90+ surface markers and subsequently collected for Drop-seq analysis at 24hrs after electroporation. **(B)** Reactome analysis on the top 100 genes (ranked by log 2-fold change (2FC)) up-regulated in scRNA-Seq data comparing edited vs. mock treated CD34+ cells. **(C)** Heatmap showing the expression level of a list of p53-activated genes retrived from *M. Fischer 2017 Oncogene* 36, 3943-3956 in single cells edited with RNP+AAV6 or Mock-treated. **(D)** Fold expression of *CDKN1A* relative to mock control measured 24hrs after electroporation of CD34+ cells treated as indicated (RNP+ssODN n=3, other conditions n=5).

Figure 2. Transient p53 inhibition by pifithrins. (A) Schematic representation of the gene editing protocol and the timing of phithrins addition. **(B)** Percentage of GFP+ cells measured within the indicated subpopulations 3 days after gene editing in presence or not of different phfithin-α derivatives (n=2). **(C)** Percentage of GFP+ cells measured within the indicated subpopulations and of NHEJ measured on the bulk cells 3 days after gene editing in presence or not of phfithin-α ciclic (n=5) **(D)** Subpopulation composition of treated CD34+ cells from (C) (n=5).

Figure 3. Transient p53 inhibition by GSE56 mRNA. (A) Schematic representation of the gene editing protocol ± GSE56 and cell analyses. **(B)** Fold expression of the indicated p53-target genes (CDKN1A, PHLDA3, RPS27L, APOBEC3H) relative to untreated control (UT) measured 24hrs after electroporation of CD34+ cells treated as indicated (UT, RNP+AAV6, RNP+AAV6+GSE56 n=8; mock electro n=3) *P<0.05, **P<0.01; Mann-Whitney Test. **(C)** (Left) Percentage of HDR (measured by ddPCR) or NHEJ (measured by heteroduplex assay) 3 days after *AAVS1* or *IL2RG* gene editing in presence or not GSE56 mRNA (n=9) *P<0.05, **P<0.01; Wilcoxon matched-pairs signed-rank test. (Right) Percentage of GFP+ cells measured within the indicated subpopulations 3 days after *AAVS1* or *IL2RG* gene editing in presence or not GSE56 mRNA (n=15) **P<0.01; Wilcoxon matched-pairs signed-rank test. **(D)** Growth curve of AAVS1 edited cells in presence or not GSE56 mRNA (n=2)

(E) Subpopulation composition of treated CD34+ cells from (B). **(F)** Number of erythroid or myeloid single colonies counted upon plating in CFU-C assay edited cells sorted for CD133+ CD90+ or CD133+ 90- (n=11) ns: P>0.05, *P<0.05; Mann-Whitney Test.

Figure 4. Dampening the transcriptional signature induced by gene editing increases HSPC long-term repopulation capacity. (A) Schematic representation of the gene editing protocol ± GSE56 and transplantation in NSG mice. **(B)** Human CD45+ engraftment of *IL2RG* edited cells in presence (n=6) or not GSE56 mRNA (n=5) measured at the indicated times in the PB of transplanted NSG mice. **(C)** Percentage of human lymphoid (CD19+ and CD3+) and myeloid (CD13+) populations measured in the PB of mice from (B). **(D)** Percentage of GFP+ cells measured in the indicated human PB populations in mice from (B).

Figure 5: Scheme of the adenovirus E4 gene. Adenovirus is a natural co-helper of AAV that provides a set of helper genes function (E1a, E1b, E2a, E4) required to optimize AAV infection process. The E4 gene is composed of seven open ORFs that are coding for seven different proteins. We investigated the use of E4orf1 and E4orf6/7 adenoviral proteins to improve gene editing efficiency in HSCs.

Figure 6: *AAVS1* gene editing in combination with adenoviral E4orf1 and E4orf6/7 proteins. (A) Schematic representation of the gene editing protocol ± E4orf1 and/or E4orf6/7 and cell analyses. (**B**) Percentage of GFP^+ cells measured within the indicated subpopulations 4 days after *AAVS1* editing of cord blood (CB)-derived $CD34^+$ with Cas9-RNP (25pmol) in presence (E4orf6/7) or absence (standard) of the adenoviral protein E4orf6/7 mRNA (2µg total mRNA) and using increasing doses of AAV6 as donor template, with MOI = 10^4 vg/cell (n = 3), MOI = $5x10^4$ vg/cell (n = 12), MOI = 10^5 vg/cell (n = 2). (**C**) Percentage of GFP^+ cells measured within the indicated subpopulations 4 days after *AAVS1* editing of CB-derived $CD34^+$ using AAV6 (MOI = $5x10^4$ vg/cell) or IDLV (MOI = 200) as donor template, RNP-Cas9 (25pmol) and supplemented with E4orf6/7 (2µg total RNA). (**D**) Percentage of GFP^+ cells measured within the indicated subpopulations 4 days after *AAVS1* editing of CB-derived $CD34^+$ using AAV6 as donor template with MOI = $5x10^4$ vg/cell, RNP-Cas9 (25pmol) and supplemented with crescent dose of E4orf1 (from 1 to 5µg total RNA) or E4orf6/7 (from 1 to 5µg total RNA).

Figure 7: E4orf1 and E4orf6/7 supplementation to gene editing protocol in human $CD34^+$. (**A**) Percentage of GFP^+ cells measured within the indicated subpopulations 4 days after *AAVS1* editing of CB-derived $CD34^+$ using AAV6 as donor template with MOI = $5x10^4$

vg/cell, RNP-Cas9 (25pmol) and supplemented or not with E4orf1 (3µg total RNA) and/or E4orf6/7 (2µg total RNA). (**B**) Percentage of HDR mediated integration measured by ddPCR on the 5' vector-to-genome junction, and GFP^+ cells measured by cytofluorimetric analysis within the bulk at 4 days after gene-editing of CB-derived $CD34^+$ using AAV6 as donor template with MOI = $5x10^4$ vg/cell, RNP-Cas9 (25pmol) and supplemented or not with E4orf1 (3µg total RNA) and/or E4orf6/7 (2µg total RNA). (**C**) Colony forming cell assay. Total number of erythroid (n = 3), in red, and myeloid (n=3), in white, colonies measured 15 days after plating. (**D**) Cell growth curve post-electroporation (n =1).

Figure 8: E1B55K and E4orf6 supplementation to gene editing protocol in human $CD34^+$ (**A**) Percentage of GFP^+ cells measured within the indicated subpopulations 4 days after *AAVS1* editing of CB-derived $CD34^+$ using AAV6 as donor template with MOI = $5x10^4$ vg/cell, RNP-Cas9 (25pmol) and supplemented with E1B55K (3µg total RNA) and/or E4orf6 (2µg total RNA). (**B**) Percentage of HDR mediated integration measured by ddPCR on the 5' vector-to-genome junction, and GFP^+ cells measured by cytofluorimetric analysis within the bulk at 4 days after gene-editing of CB-derived $CD34^+$ using AAV6 as donor template with MOI = $5x10^4$ vg/cell, RNP-Cas9 (25pmol) and supplemented with E1B55K (3µg total RNA) and/or E4orf6 (2µg total RNA). (**C**) Colony forming cell assay. Total number of erythroid (n = 3), in red, and myeloid (n=3), in white, colonies measured 15 days after plating. (**D**) Cell growth curve post-electroporation (n =1).

Figure 9: Gene editing of human T-cells using E4orf6/7. (**A**) Schematic representation of the T cells gene editing protocol ± E4orf6/7 and cell analysis. (**B**) Percentage of GFP^+ cells measured within the indicated subpopulations 20 days after *AAVS1* editing of stimulated T-cells (n = 3) using AAV6 as donor template with MOI = $5x10^4$ vg/cell, RNP-Cas9 (25pmol) and supplemented with E4orf6/7 (2µg total RNA).

Figure 10: *AAVS1* gene editing of human CB-derived $CD34^+$ using E4orf6/7 followed by NSG mice transplantation. (**A**) The standard transplantation procedure required 3 days of cells stimulation with cytokines and drugs before electroporation of the complex *AAVS1* specific RNP-cas9 (25pmol) and supplemented with E4orf6/7 (2µg total RNA). Electroporation procedure is followed by AAV6 transduction at $5x10^4$ viral genome per cell. A total of $6x10^5$ treated cells are transplanted to each mouse (tail intravenous injection); two experimental conditions were tested (standard and supplemented with E4orf6/7). (**B**) Percentage of GFP^+ cells measured within the indicated subpopulations 4 days after *AAVS1* editing of CB-derived $CD34^+$ using AAV6 as donor template with MOI = $5x10^4$ vg/cell, RNP-Cas9 (25pmol) and supplemented with E4orf6/7 (2µg total RNA). (**C**) Culture composition of

treated $CD34^+$ cells from (B). (**D**) Molecular analysis of the *AAVS1* alleles of the cells from (B), represented as percentage of wild-type (WT) or modified alleles by HDR (digital droplet PCR, ddPCR, mediated quantification of the 5' vector-to-genome junction) or by NHEJ (T7 heteroduplex detection assay).

Figure 11: *In-vivo* following of *AAVS1* targeted human CB-derived $CD34^+$ using E4orf6/7. (**A**) Human CD45+ engraftment of edited cells in presence (n=4) or not (n=5) with E4orf6 mRNA measured at the indicated times in the peripheral blood (PB) of the transplanted NSG mice and at the end of the experiment in their different organs (spleen, thymus and bone-marrow) and within different blood lineages of the bone marrow (HSPC=CD34+; B cells=CD19+; Myeloid cells= CD33+; T cells=CD3+). (**B**) Percentage of GFP^+ cells measured within the human cells from (A) measured in PB, hematopoietic organs and bone-marrow lineages. (**C**) Percentage of edited alleles (number of targeted copies per cell) from (A) measured in PB, hematopoietic organs and bone-marrow lineages.

Figure 12: Transient GSE56 overexpression improves *in vivo* yield and clonality of *AAVS1*-edited repopulating HSPCs. (**A**) Percentage of cell cycle states in bulk culture and $CD90^+$ by Hoechst 33342 staining measured by FACS at 12-24 hours post-editing protocol [Kruskal-Wallis test, relative to standard RNP + AAV6]. (**B**) Schematic representation of the experimental setting for transplantation into immunodeficient NSG mice of *AAVS1* targeted gene editing in human cord blood (CB)-derived $CD34^+$ cells. (**C, D**) Percentage of human cells (**C**) and GFP^+ (**D**) within total human cells in spleen (SPL) and bone marrow (BM) at 21 weeks post-xenotransplantation with cells treated (+ GSE56) or not (standard) with the p53 inhibitor [Mann-Withney test]. (**E**) Number of unique dominant barcodes in the hematopoietic organs of mice transplanted with cells treated (+ GSE56) or not (standard) with the p53 inhibitor [Mann-Withney test]. (**F**) Correlation between the number of unique dominant barcodes (X axis) and the percentage of $CD45^+GFP^+$ cells (Y axis) (Spearman correlation).

Figure 13: Transient GSE56 overexpression improves *in vivo* yield of *CD40LG*-edited repopulating HSPCs. (**A**) Schematic representation of the experimental setting for transplantation into immunodeficient NSG mice of *CD40LG* targeted gene editing in human cord blood (CB)-derived $CD34^+$ cells. (**B, C**) Percentage of human cells (**B**) and GFP (**C**) in peripheral blood of NSG mice transplanted with cells treated (+ GSE56) or not (standard) with the p53 inhibitor. (**D**) Percentage of human cells in hematopoietic organs of mice from (B). (**E**) Percentage of targeted integration measured by digital droplet PCR in sorted B cells ($CD19^+$), myeloid ($CD33^+$), T cells ($CD3^+$) and stem cells ($CD34^+$) from mice in (**D**).

Figure 14: The adenoviral proteins Ad5-E4orf6/7 and Ad5-E4orf1 both increase the HDR efficiency by targeting the *AAVS1* locus in HSPCs. (A) Schematic representation of the targeted gene editing protocol in human cord blood (CB)-derived $CD34^+$ after pre-stimulation with cytokines and drugs and by electroporating at day 3 the RNP-Cas9 (25 pmol) and AAV6 (2.10^4 vg/ml) as standard editing procedure ± GFP, GSE56, Ad5-E4orf1, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 delivered as mRNAs. (**B**) Monitoring of $CD34^+$ cultured-cells growth over days post-editing. (**C left and right panels**) Percentage of GFP^+ cells measured by FACS within the indicated subpopulations at 96 hours post-editing protocol for standard RNP + AAV6 ± GSE56, Ad5-E4orf1, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 and associated numeric fold increase [one-way ANOVA]. (**D**) Percentage of *AAVS1* edited alleles by HDR and NHEJ in bulk culture measured by ddPCR on the 5' vector-to-genome junction at 96 hours post-editing protocol [Kruskal-Wallis test, relative to standard RNP + AAV6]. (**E**) Culture composition from differentiated cells ($CD34^-$) to the most primitive compartment ($CD34^+CD133^+CD90^+$) at 96 hours post-editing protocol [Kruskal-Wallis test, relative to standard RNP + AAV6]. (**F**) Percentage of cell states apoptosis in bulk culture by 7-AAD / Annexin staining measured by FACS at 24 hours post-editing protocol. (**G**) Number of colonies plated from bulk culture at 24 hours post-editing protocol [Kruskal-Wallis test, relative to standard RNP + AAV6]. (**H**) Percentage of cell cycle states in bulk culture by Hoechst 33342 staining measured by FACS at 12-24 hours post-editing protocol [Kruskal-Wallis test, relative to standard RNP + AAV6]. (**I**) Percentage of GFP^+ cells measured by FACS within the indicated subpopulations at 96 hours post-editing protocol in *IL2RG* locus, for standard RNP + AAV6 ± Ad5-E4orf6/7 and associated numeric fold increase [Paired T-test]. (**J**) Schematic representation of the targeted gene editing protocol in human cord blood (CB)-derived $CD34^+$ by performing electroporation at day 0, 1 and 2 post-thawing with the RNP-Cas9 (25 pmol) and AAV6 (2.10^4 vg/ml) as standard editing procedure ± GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 delivered as mRNAs. (**K**) Percentage of GFP^+ cells measured by FACS within the indicated subpopulations at 96 hours post-editing protocol (n = 1) for standard RNP + AAV6 ± GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 and associated numeric fold increase in bulk (left panel) and $CD90^+$ (right panel). (**L**) Percentage of *AAVS1* edited alleles by HDR in bulk culture measured by ddPCR on the 5' vector-to-genome junction at 96 hours post-editing protocol (n = 1).

Figure 15: The combination of adenoviral proteins Ad5-E1B55K and Ad5-E4orf6 increases donor template cassette expression but does not increase HDR efficiency in HSPCs. (**A**) Percentage of GFP^+ cells measured by FACS within the indicated subpopulations at 96 hours post-editing protocol performed in hCB-derived $CD34^+$ by using

our standard procedure RNP + AAV6 ± Ad5-E1B55K and/or Ad5-E4orf6 [one-way ANOVA]. (**B**) Percentage of *AAVS1* edited alleles by HDR and NHEJ in bulk culture measured by ddPCR on the 5' vector-to-genome junction at 96 hours post-editing protocol [Kruskal-Wallis test, relative to standard RNP + AAV6]. (**C**) Culture composition at 96 hours post-editing protocol. (**D**) Percentage of cell states apoptosis in bulk culture by 7-AAD / Annexin staining measured by FACS at 24 hours post-editing protocol. (**E**) Number of colonies plated from bulk culture at 24 hours post-editing protocol [Kruskal-Wallis test, relative to standard RNP + AAV6].

Figure 16: The adenoviral protein Ad5-E4orf6/7 impacts on E2Fs target genes regulating the cell cycle in HSPCs. (A) TaqMan assay of *CDK2* gene fold expression relative to untreated cells (UT) at 24 hours post-editing protocol on standard RNP + AAV6 ± GFP, GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 [Kruskal-Wallis test, relative to standard RNP + AAV6]. **(B)** TaqMan assay of *CDK2* gene fold expression relative to UT at 12, 24, 96, and 168 hours post-editing protocol on standard RNP + AAV6 ± GFP, GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7. **(C)** TaqMan assay of *CDKN2A* gene fold expression relative to UT at 24 hours post-editing protocol on standard RNP + AAV6 ± GFP, GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 [Kruskal-Wallis test, relative to standard RNP + AAV6]. **(D)** TaqMan assay of *CDKN2A* gene fold expression relative to UT at 12, 24, 96, and 168 hours post-editing protocol on standard RNP + AAV6 ± GFP, GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7. **(E)** TaqMan assay of *CDKN1A* gene fold expression relative to UT at 24 hours post-editing protocol on standard RNP + AAV6 ± GFP, GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 [Kruskal-Wallis test, relative to standard RNP + AAV6]. **(F)** TaqMan assay of *CDKN1A* gene fold expression relative to UT at 12, 24, 96, and 168 hours post-editing protocol on standard RNP + AAV6 ± GFP, GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7. **(G)** TaqMan assay of *APOBEC3H* gene fold expression relative to UT at 24 hours post-editing protocol on standard RNP + AAV6 ± GFP, GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 [Kruskal-Wallis test, relative to standard RNP + AAV6]. **(H)** TaqMan assay of *APOBEC3H* gene fold expression relative to UT at 12, 24, 96, and 168 hours post-editing protocol on standard RNP + AAV6 ± GFP, GSE56, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7. **(I)** Schematic representation of cell cycle regulation mechanism in HSPCs after editing-protocol using the adenoviral protein Ad5-E4orf6/7 and/or the p53 inhibitor GSE56.

Figure 17: Primary xenotransplantation follow-up of *AAVS1* edited human CB-derived CD34$^+$. (A) Schematic representation of the targeted gene editing protocol in human cord blood (CB)-derived CD34$^+$ and followed by NSG xenotransplantation at 24 hours post-editing protocol. **(B)** Monitoring of human CD45$^+$ cells engraftment in peripheral blood (PB) over the

weeks post-transplantation. **(C)** Percentage of GFP^+ edited cells among engrafted $hCD45^+$ in PB over the weeks post-transplantation. **(D, E)** Percentage of both $hCD45^+$ engraftment and GFP^+ within $hCD45^+$ at long-term in spleen and bone-marrow (BM). **(F)** Number of unique dominant barcodes in the hematopoietic organs of mice transplanted with cells treated RNP + AAV6 ± GSE56/Ad5-E4orf6/7 [Mann-Withney test]. Lines indicate median values.

Figure 18: Secondary xenotransplantation follow-up of *AAVS1* edited human CB-derived $CD34^+$. (A) Schematic representation of the secondary transplantation in NSG mice after $CD34^+$ cells selection from BM. **(B)** Monitoring of human $CD45^+$ cells engraftment in peripheral blood (PB) over the weeks post-transplantation. **(C)** Percentage of GFP^+ edited cells among engrafted $hCD45^+$ in PB over the weeks post-transplantation. **(D, E)** Percentage of both $hCD45^+$ engraftment and GFP^+ within $hCD45^+$ at very long-term in spleen and bone-marrow (BM). **(F, G)** Blood lineages composition in spleen ($CD19^+$, $CD33^+$ and $CD3^+$) and associated percentages of GFP^+ cells. Lines indicate median values.

Figure 19: Primary xenotransplantation follow-up at limiting dose of *AAVS1* edited human CB-derived $CD34^+$. (A) Monitoring of human $CD45^+$ cells engraftment in peripheral blood (PB) over the weeks post-transplantation. **(B)** Percentage of GFP^+ edited cells among engrafted $hCD45^+$ in PB over the weeks post-transplantation. Lines indicate median values.

Figure 20: Primary and secondary xenotransplantation follow-up of *IL2RG* edited human CB-derived $CD34^+$. (A, B) Monitoring of both human $CD45^+$ cells engraftment and GFP^+ edited cells among engrafted $hCD45^+$ in peripheral blood (PB) over the weeks post-primary transplantation. **(C, D)** Monitoring of both human $CD45^+$ cells engraftment and GFP^+ edited cells among engrafted $hCD45^+$ in peripheral blood (PB) over the weeks post-secondary transplantation. **(E, F)** Percentage of both $hCD45^+$ engraftment and GFP^+ within $hCD45^+$ at long-term in spleen and bone-marrow (BM). **(G, H)** Blood lineages composition in spleen ($CD19^+$, $CD33^+$ and $CD3^+$) and associated percentages of GFP^+ cells. Lines indicate median values. **(I)** Percentage of translocation events between “X” and “14” chromosomes measured by ddPCR, both from *in vitro* and *in vivo* samples targeting *IL2RG* locus [Kruskal-Wallis test, relative to standard RNP + AAV6]. Lines indicate median values.

Figure 21: Exploring adenoviral protein E4orf1 serotype variants. (**A**) E4orf1 protein sequences alignment (T-Coffee) including highly conserved amino-acids (**B**) Phylogenetic tree based on T-Coffee analysis and presenting the closest amino-acids composition among adenoviral serotypes. (**C**) Percentage of GFP^+ cells measured by FACS within the indicated subpopulations at 96 hours post-editing protocol performed in hCB-derived $CD34^+$ by using our standard procedure RNP + AAV6 ± Ad4-E4orf1, Ad5-E4orf1, Ad7-E4orf1, or Ad9-E4orf1

(selected variants). (**D**) Number of colonies plated from bulk culture at 24 hours post-editing protocol. **(E)** TaqMan assay of *CDK2*, *CDKN2A* and *CDKN1A* genes fold expression relative to untreated cells (UT) at 24 hours post-editing protocol.

Figure 22: Exploring adenoviral protein E4orf6/7 serotype variants. (**A**) E4orf6/7 protein sequences alignment (T-Coffee) including highly conserved amino-acids. (**B**) Phylogenetic tree based on T-Coffee analysis and presenting the closest amino-acids composition among adenoviral serotypes. (**C**) Percentage of GFP^+ cells measured by FACS within the indicated subpopulations at 96 hours post-editing protocol performed in hCB-derived $CD34^+$ by using our standard procedure RNP + AAV6 ± Ad3-E4orf6/7, Ad4-E4orf6/7, Ad5-E4orf6/7, or Ad23-E4orf6/7 (selected variants). (**D**) Number of colonies plated from bulk culture at 24 hours post-editing protocol. **(E)** TaqMan assay of *CDK2*, *CDKN2A* and *CDKN1A* genes fold expression relative to untreated cells (UT) at 24 hours post-editing protocol.

Figure 23: Vector comprising barcode sequence
Schematic of AAV vector design comprising a barcode sequence.

DETAILED DESCRIPTION OF THE INVENTION

The terms “comprising”, “comprises” and “comprised of” as used herein are synonymous with “including” or “includes”; or “containing” or “contains”, and are inclusive or open-ended and do not exclude additional, non-recited members, elements or steps. The terms “comprising”, “comprises” and “comprised of” also include the term “consisting of”.

Cell survival and engraftment

The term “survival” refers to the ability of the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells to remain alive (e.g. not die or become apoptotic) during in vitro or ex vivo culture. Haematopoietic stem and/or progenitor cells may, for example, undergo increased apoptosis following transduction with a viral vector during cell culture; thus, the surviving cells may have avoided apoptosis and/or cell death.

Cell survival may be readily analysed by the skilled person. For example, the numbers of live, dead and/or apoptotic cells in a cell culture may be quantified at the beginning of culture and/or following culture for a period of time (e.g. about 6 or 12 hours, or 1, 2, 3, 4, 5, 6, 7 or more days; preferably, the period of time begins with the transduction of the cells with a viral vector). The effect of an agent, such as an inhibitor or adenoviral protein of the invention, on cell survival may be assessed by comparing the numbers and/or percentages of live, dead and/or apoptotic cells at the beginning and/or end of the culture period between experiments

carried out in the presence and absence of the agent, but under otherwise substantially identical conditions.

Cell numbers and/or percentages in certain states (e.g. live, dead or apoptotic cells) may be quantified using any of a number of methods known in the art, including use of haemocytometers, automated cell counters, flow cytometers and fluorescence activated cell sorting machines. These techniques may enable distinguishing between live, dead and/or apoptotic cells. In addition or in the alternative, apoptotic cells may be detected using readily available apoptosis assays (e.g. assays based on the detection of phosphatidylserine (PS) on the cell membrane surface, such as through use of Annexin V, which binds to exposed PS; apoptotic cells may be quantified through use of fluorescently-labelled Annexin V), which may be used to complement other techniques.

The term “engraftment” refers to the ability of the haematopoietic stem and/or progenitor cells to populate and survive in a subject following their transplantation, i.e. in the short and/or long term after transplantation. For example, engraftment may refer to the number and/or percentages of haematopoietic cells descended from the transplanted haematopoietic stem cells (e.g. graft-derived cells) that are detected about 1 day to 24 weeks, 1 day to 10 weeks, or 1-30 days or 10-30 days after transplantation. In the xenograft model of human haematopoietic stem and/or progenitor cell engraftment and repopulation, engraftment may be evaluated in the peripheral blood as the percentage of cells deriving from the human xenograft (e.g. positive for the CD45 surface marker), for example. In one embodiment, engraftment is assessed at about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or 30 days after transplantation. In another embodiment, engraftment is assessed at about 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after transplantation. In another embodiment, engraftment is assessed at about 16-24 weeks, preferably 20 weeks, after transplantation.

Engraftment may be readily analysed by the skilled person. For example, the transplanted haematopoietic stem and/or progenitor cells may be engineered to comprise a marker (e.g. a reporter protein, such as a fluorescent protein), which can be used to quantify the graft-derived cells. Samples for analysis may be extracted from relevant tissues and analysed ex vivo (e.g. using flow cytometry).

Suitably, the agent for use according to the present invention may improve engraftment of gene edited haematopoietic stem and/or progenitor cells compared with gene editing without use of the agent. Suitably, engraftment at a given time point may be increased by at least

5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more when compared with engraftment of untreated gene edited haematopoietic stem and/or progenitor cells.

In a preferred embodiment, an agent for use according to the invention does not adversely affect the growth of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells when compared with untreated gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

Gene editing efficiency

In one aspect, the invention provides the use of an agent which promotes homology directed DNA repair for increasing the efficiency of gene editing of an isolated population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells (e.g. when transduced by a viral vector) and/or increasing survival and or engraftment of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

Increasing the efficiency of gene editing refers to an increase in the gene editing of the cells (e.g. haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells) in the presence of an agent, in comparison to the gene editing achieved in the absence of the agent but under otherwise substantially identical conditions. Where a viral vector is used to introduce gene editing machinery, an increased efficiency may therefore allow the multiplicity of infection (MOI) and/or the time required to achieve effective transduction to be reduced.

In one embodiment, the percentage of cells which have been edited is increased. Methods for determining the percentage of cells which have been edited are known in the art. Suitable methods include flow cytometry, fluorescence-activated cell sorting (FACS) and fluorescence microscopy. The technique employed is preferably one which is amenable to automation and/or high throughput screening.

For example, a population of cells may be edited with a vector which harbours a reporter gene. Suitably, the reporter gene may be expressed when the cell has been edited. Suitable reporter genes include genes encoding fluorescent proteins, for example green, yellow, cherry, cyan or orange fluorescent proteins. Once the population of cells has been edited, both the number of cells expressing and not-expressing the reporter gene may be quantified using a suitable technique, such as FACS. The percentage of cells which have been edited may then be calculated.

Alternatively, quantitative PCR (qPCR) may be used to determine the percentage of cells which have been gene edited without the use of a reporter gene. For example, single colonies of cells (e.g. CD34+ cells) may be picked from a semi-solid culture and qPCR may be performed on each colony separately to determine the percentage of gene-edited-positive colonies among those analysed.

Methods for determining vector copy number are also known in the art. The technique employed is preferably one which is amenable to automation and/or high throughput screening. Suitable techniques include quantitative PCR (qPCR) and Southern blot-based approaches.

Increasing the efficiency of gene editing may refer to an increase in the number of cells (e.g. haematopoietic cells, haematopoietic stem cells or haematopoietic progenitor cells) in which a target gene or site has been edited (e.g. disrupted, replaced, deleted or had a nucleic acid sequence inserted within or at it) in the intended manner following transduction of a population of cells with a viral vector in the presence of an agent (e.g. inhibitor of p53 activation or adenoviral protein), in comparison to that achieved in the absence of the agent but under otherwise substantially identical conditions. An increased efficiency of gene editing may therefore allow the multiplicity of infection (MOI) and/or the transduction time required to achieve effective gene editing to be reduced. Methods for determining whether a target gene or site has been edited are known in the art.

In the context of gene editing, for example using a CRISPR/Cas system, preferably the vector used to transduce the population of cells is a non-integrating vector (e.g. an integration-defective lentiviral vector, IDLV).

In one embodiment, the agent for use according to the present invention improves gene editing efficiency compared with gene editing without use of the agent (i.e. standard gene editing). Suitably, gene editing efficiency may be improved by at least 1.1 fold, 1.2 fold, 1.3 fold, 1.4 fold, 1.5 fold, 1.6 fold, 1.7 fold, 1.8 fold, 1.9 fold, 1.8 fold, 1.9 fold, 2.0 fold, 2.1 fold, 2.2 fold, 2.3 fold, 2.4 fold, 2.5 fold, 3 fold or more.

Suitably, the gene editing efficiency may be improved in a particular cell compartment. Suitably, gene editing is improved in a primitive HSPC cell compartment. Suitably, gene editing may be improved in CD34+ CD133- cells. Suitably, gene editing may be improved in CD34+ CD133+ cells. Suitably, gene editing may be improved in CD34+ CD133+ CD90+ cells.

Preferably gene editing efficiency of CD34+ CD133+ CD90+ cells may be improved by at least 1.1 fold, 1.2 fold, 1.3 fold, 1.4 fold, 1.5 fold, 1.6 fold, 1.7 fold, 1.8 fold, 1.9 fold, 1.8 fold, 1.9 fold, 2.0 fold, 2.1 fold, 2.2 fold, 2.3 fold, 2.4 fold, 2.5 fold, 3 fold or more.

An agent which promotes homology directed DNA repair

The present inventors have determined that improving the efficiency of HDR in HSC increases cell survival and engraftment of gene edited HSCs.

Gene editing in primary cells and the HSPC in particular may be hampered by gene transfer efficiency and limited HDR, likely due to low levels of expression of the HDR machinery and high activity of NHEJ pathway.

As used herein “an agent which promotes homology directed DNA repair” refers to an agent which enhances and/or improves the efficiency of HDR relative to the level of HDR in a cell which has not been treated with the agent.

Suitably, HDR may be increased by 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more. Suitably, HDR may be increased by 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9. 2.0, 2.5, 3.0 fold or more.

HDR efficiency may be measured using any method in the art. Suitably, HDR efficiency may be determined by FACS measuring the presence of a marker in donor template vector which is introduced to the cell by HDR-mediated integration at the targeted locus. For example, an AAV6 donor template may comprise a PGK.GFP reporter cassette. Figure 2B demonstrates that an increase in HDR efficiency was obtained via transient p53 inhibition, as measured by determining the percentage of GFP+ cells.

Targeting integration (“on target” HDR) can be determined by digital PCR by using primers and probes which were designed on the junction between the vector sequence and the targeted locus and on control sequences used for normalization (human *TTC5* genes).

The percentage of insertions and deletion (indels) introduced by the Non Homologous End Joining (NHEJ) repair pathway on the nucleases target site was measured by a mismatch-sensitive endonuclease assay (PCR-based amplification of the targeted locus followed by digestion with T7 Endonuclease I; digested DNA fragments were resolved and quantified by

capillary electrophoresis on LabChip GX Touch HT, Perkin Elmer) The levels of NHEJ-induced mutations were used as surrogate readout for scoring nucleases activity.

Preferably, the agent has low cellular toxicity.

Preferably, the agent does not significantly change the composition of the gene edited cells.

Preferably, the agent does not significantly induce differentiation of the gene edited haematopoietic stem and/or progenitor cells. Therefore, the gene edited haematopoietic stem and/or progenitor cells according to the present invention retain their long-term re-population capacity.

Suitably, the change in composition or differentiation of the gene edited cells is less than 5%, less than 4%, less than 3% less than 2% or less than 1% when compared with an untreated control.

For example, Figure 3E shows that there was no change in relative composition of the haematopoietic populations measured at 3 days post editing in gene edited cells compared with untreated controls.

p53 activation

The term “p53 activation” refers to an increase in the activity of p53, for example through a post-translational modification of the p53 protein. Example post-translational modifications include phosphorylation, acetylation and methylation, and are described in Kruse, J.P. et al. (2008) SnapShot: p53 Posttranslational Modifications Cell 133: 930-931. In the context of the invention, the p53 activation preferably results from phosphorylation of p53, particularly preferably at amino acid Serine 15.

Methods for analysing such post-translational modifications are known in the art (example methods for analysing kinase activity are disclosed herein, further methods include, for example, mass spectrometry- and antibody recognition-based methods).

An example amino acid sequence of p53, which may be used to provide an amino acid numbering convention, is:

```
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPDEAPRMPEAA
PPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKT
CPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLRVEYLDDRN
TFRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPGR
DRRTEEENLRKKGEPHHELPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALEL
KDAQAGKEPGGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD
```

(SEQ ID NO: 3; NCBI Accession No. 000537.3)

Ataxia telangiectasia mutated (ATM) kinase

Ataxia telangiectasia mutated (ATM) kinase (also known as “ataxia telangiectasia mutated”) is a serine/threonine kinase which is recruited to and activated by double-strand DNA breaks. ATM kinase is known to phosphorylate a number of proteins (including p53, CHK2, BRCA1, NBS1 and H2AX) that initiate activation of the DNA damage checkpoint, leading to cell cycle arrest, DNA repair or apoptosis.

In one embodiment, the ATM kinase is human ATM kinase.

In one embodiment, the amino acid sequence of ATM kinase is the sequence deposited under NCBI Accession No. NP_000042.3.

An example amino acid sequence of the ATM kinase is:

MSLVLNDLLICCRQLEHDRATERKKEVEKFKRLIRDPETIKHLDRHSDSKQGKYLNWDAVFRFLQKYIQKET
ECLRIAKPNVSASTQASRQKKMQEISSLVKYFIKCANRRAPRLKCQELLNYIMDTVKDSSNGAIYGADCSNI
LLKDILSVRKYWCEISQQQWLELFSVYFRLYLKPSQDVHRVLVARIIHAVTKGCCSQTDGLNSKFLDFFSKA
IQCARQEKSSSGLNHILAALTIFLKTLAVNFRIRVCELGDEILPTLLYIWTQHRLNDSLKEVIIELFQLQIY
IHHPKGAKTQEKGAYESTKWRSILYNLYDLLVNEISHIGSRGKYSSGFRNIAVKENLIELMADICHQVFNED
TRSLEISQSYTTTQRESSDYSVPCKRKKIELGWEVIKDHLQKSQNDFDLVPWLQIATQLISKYPASLPNCEL
SPLLMILSQLLPQQRHGERTPYVLRCLTEVALCQDKRSNLESSQKSDLLKLWNKIWCITFRGISSEQIQAEN
FGLLGAIIQGSLVEVDREFWKLFTGSACRPSCPAVCCLTLALTTSIVPGTVKMGIEQNMCEVNRSFSLKESI
MKWLLFYQLEGDLENSTEVPPILHSNFPHLVLEKILVSLTMKNCKAAMNFFQSVPECEHHQKDKEELSFSEV
EELFLQTTFDKMDFLTIVRECGIEKHQSSIGFSVHQNLKESLDRCLLGLSEQLLNNYSSEITNSETLVRCSR
LLVGVLGCYCYMGVIAEEEAYKSELFQKAKSLMQCAGESITLFKNKTNEEFRIGSLRNMMQLCTRCLSNCTK
KSPNKIASGFFLRLLTSKLMNDIADICKSLASFIKKPFDRGEVESMEDDTNGNLMEVEDQSSMNLFNDYPDS
SVSDANEPGESQSTIGAINPLAEEYLSKQDLLFLDMLKFLCLCVTTAQTNTVSFRAADIRRKLLMLIDSSTL
EPTKSLHLHMYLMLLKELPGEEYPLPMEDVLELLKPLSNVCSLYRRDQDVCKTILNHVLHVVKNLGQSNMDS
ENTRDAQGQFLTVIGAFWHLTKERKYIFSVRMALVNCLKTLLEADPYSKWAILNVMGKDFPVNEVFTQFLAD
NHHQVRMLAAESINRLFQDTKGDSSRLLKALPLKLQQTAFENAYLKAQEGMREMSHSAENPETLDEIYNRKS
VLLTLIAVVLSCSPICEKQALFALCKSVKENGLEPHLVKKVLEKVSETFGYRRLEDFMASHLDYLVLEWLNL
QDTEYNLSSFPFILLNYTNIEDFYRSCYKVLIPHLVIRSHFDEVKSIANQIQEDWKSLLTDCFPKILVNILP
YFAYEGTRDSGMAQQRETATKVYDMLKSENLLGKQIDHLFISNLPEIVVELLMTLHEPANSSASQSTDLCDF
SGDLDPAPNPPHFPSHVIKATFAYISNCHKTKLKSILEILSKSPDSYQKILLAICEQAAETNNVYKKHRILK
IYHLFVSLLLKDIKSGLGGAWAFVLRDVIYTLIHYINQRPSCIMDVSLRSFSLCCDLLSQVCQTAVTYCKDA
LENHLHVIVGTLIPLVYEQVEVQKQVLDLLKYLVIDNKDNENLYITIKLLDPFPDHVVFKDLRITQQKIKYS
RGPFSLLEEINHFLSVSVYDALPLTRLEGLKDLRRQLELHKDQMVDIMRASQDNPQDGIMVKLVVNLLQLSK
MAINHTGEKEVLEAVGSCLGEVGPIDFSTIAIQHSKDASYTKALKLFEDKELQWTFIMLTYLNNTLVEDCVK
VRSAAVTCLKNILATKTGHSFWEIYKMTTDPMLAYLQPFRTSRKKFLEVPRFDKENPFEGLDDINLWIPLSE
NHDIWIKTLTCAFLDSGGTKCEILQLLKPMCEVKTDFCQTVLPYLIHDILLQDTNESWRNLLSTHVQGFFTS
CLRHFSQTSRSTTPANLDSESEHFFRCCLDKKSQRTMLAVVDYMRRQKRPSSGTIFNDAFWLDLNYLEVAKV
AQSCAAHFTALLYAEIYADKKSMDDQEKRSLAFEEGSQSTTISSLSEKSKEETGISLQDLLLEIYRSIGEPD
SLYGCGGGKMLQPITRLRTYEHEAMWGKALVTYDLETAIPSSTRQAGIIQALQNLGLCHILSVYLKGLDYEN
KDWCPELEELHYQAAWRNMQWDHCTSVSKEVEGTSYHESLYNALQSLRDREFSTFYESLKYARVKEVEEMCK
RSLESVYSLYPTLSRLQAIGELESIGELFSRSVTHRQLSEVYIKWQKHSQLLKDSDFSFQEPIMALRTVILE
ILMEKEMDNSQRECIKDILTKHLVELSILARTFKNTQLPERAIFQIKQYNSVSCGVSEWQLEEAQVFWAKKE
QSLALSILKQMIKKLDASCAANNPSLKLTYTECLRVCGNWLAETCLENPAVIMQTYLEKAVEVAGNYDGESS
DELRNGKMKAFLSLARFSDTQYQRIENYMKSSEFENKQALLKRAKEEVGLLREHKIQTNRYTVKVQRELELD
ELALRALKEDRKRFLCKAVENYINCLLSGEEHDMWVFRLCSLWLENSGVSEVNGMMKRDGMKIPTYKFLPLM
YQLAARMGTKMMGGLGFHEVLNNLISRISMDHPHHTLFIILALANANRDEFLTKPEVARRSRITKNVPKQSS
QLDEDRTEAANRIICTIRSRRPQMVRSVEALCDAYIILANLDATQWKTQRKGINIPADQPITKLKNLEDVVV
PTMEIKVDHTGEYGNLVTIQSFKAEFRLAGGVNLPKIIDCVGSDGKERRQLVKGRDDLRQDAVMQQVFQMCN

TLLQRNTETRKRKLTICTYKVVPLSQRSGVLEWCTGTVPIGEFLVNNEDGAHKRYRPNDFSAFQCQKKMMEV
QKKSFEEKYEVFMDVCQNFQPVFRYFCMEKFLDPAIWFEKRLAYTRSVATSSIVGYILGLGDRHVQNILINE
QSAELVHIDLGVAFEQGKILPTPETVPFRLTRDIVDGMGITGVEGVFRRCCEKTMEVMRNSQETLLTIVEVL
LYDPLFDWTMNPLKALYLQQRPEDETELHPTLNADDQECKRNLSDIDQSFNKVAERVLMRLQEKLKGVEEGT
VLSVGGQVNLLIQQAIDPKNLSRLFPGWKAWV

(SEQ ID NO: 4)

An example nucleic acid sequence encoding the ATM kinase is:

ATGAGTCTAGTACTTAATGATCTGCTTATCTGCTGCCGTCAACTAGAACATGATAGAGCTACAGAACGAA
AGAAAGAAGTTGAGAAATTTAAGCGCCTGATTCGAGATCCTGAAACAATTAAACATCTAGATCGGCATTC
AGATTCCAAACAAGGAAAATATTTGAATTGGGATGCTGTTTTTAGATTTTACAGAAATATATTCAGAAA
GAAACAGAATGTCTGAGAATAGCAAAACCAAATGTATCAGCCTCAACACAAGCCTCCAGGCAGAAAAAGA
TGCAGGAAATCAGTAGTTTGGTCAAATACTTCATCAAATGTGCAAACAGAAGAGCACCTAGGCTAAAATG
TCAAGAACTCTTAAATTATATCATGGATACAGTGAAAGATTCATCTAATGGTGCTATTTACGGAGCTGAT
TGTAGCAACATACTACTCAAAGACATTCTTTCTGTGAGAAAATACTGGTGTGAAATATCTCAGCAACAGT
GGTTAGAATTGTTCTCTGTGTACTTCAGGCTCTATCTGAAACCTTCACAAGATGTTCATAGAGTTTTAGT
GGCTAGAATAATTCATGCTGTTACCAAAGGATGCTGTTCTCAGACTGACGGATTAAATTCCAAATTTTTG
GACTTTTTTTCCAAGGCTATTCAGTGTGCGAGACAAGAAAAGAGCTCTTCAGGTCTAAATCATATCTTAG
CAGCTCTTACTATCTTCCTCAAGACTTTGGCTGTCAACTTTCGAATTCGAGTGTGTGAATTAGGAGATGA
AATTCTTCCCACTTTGCTTTATATTTGGACTCAACATAGGCTTAATGATTCTTTAAAAGAAGTCATTATT
GAATTATTTCAACTGCAAATTTATATCCATCATCCGAAAGGAGCCAAAACCCAAGAAAAAGGTGCTTATG
AATCAACAAAATGGAGAAGTATTTTATACAACTTATATGATCTGCTAGTGAATGAGATAAGTCATATAGG
AAGTAGAGGAAAGTATTCTTCAGGATTTCGTAATATTGCCGTCAAAGAAAATTTGATTGAATTGATGGCA
GATATCTGTCACCAGGTTTTTAATGAAGATACCAGATCCTTGGAGATTTCTCAATCTTACACTACTACAC
AAAGAGAATCTAGTGATTACAGTGTCCCTTGCAAAAGGAAGAAAATAGAACTAGGCTGGGAAGTAATAAA
AGATCACCTTCAGAAGTCACAGAATGATTTTGATCTTGTGCCTTGGCTACAGATTGCAACCCAATTAATA
TCAAAGTATCCTGCAAGTTTACCTAACTGTGAGCTGTCTCCATTACTGATGATACTATCTCAGCTTCTAC
CCCAACAGCGACATGGGGAACGTACACCATATGTGTTACGATGCCTTACGGAAGTTGCATTGTGTCAAGA
CAAGAGGTCAAACCTAGAAAGCTCACAAAAGTCAGATTTATTAAAACTCTGGAATAAAATTTGGTGTATT
ACCTTTCGTGGTATAAGTTCTGAGCAAATACAAGCTGAAAACTTTGGCTTACTTGGAGCCATAATTCAGG
GTAGTTTAGTTGAGGTTGACAGAGAATTCTGGAAGTTATTTACTGGGTCAGCCTGCAGACCTTCATGTCC
TGCAGTATGCTGTTTGACTTTGGCACTGACCACCAGTATAGTTCCAGGAACGGTAAAAATGGGAATAGAG
CAAAATATGTGTGAAGTAAATAGAAGCTTTTCTTTAAAGGAATCAATAATGAAATGGCTCTTATTCTATC
AGTTAGAGGGTGACTTAGAAAATAGCACAGAAGTGCCTCCAATTCTTCACAGTAATTTTCCTCATCTTGT
ACTGGAGAAAATTCTTGTGAGTCTCACTATGAAAAACTGTAAAGCTGCAATGAATTTTTTCCAAAGCGTG
CCAGAATGTGAACACCACCAAAAAGATAAAGAAGAACTTTCATTCTCAGAAGTAGAAGAACTATTTCTTC
AGACAACTTTTGACAAGATGGACTTTTTAACCATTGTGAGAGAATGTGGTATAGAAAAGCACCAGTCCAG
TATTGGCTTCTCTGTCCACCAGAATCTCAAGGAATCACTGGATCGCTGTCTTCTGGGATTATCAGAACAG
CTTCTGAATAATTACTCATCTGAGATTACAAATTCAGAAACTCTTGTCCGGTGTTCACGTCTTTTGGTGG
GTGTCCTTGGCTGCTACTGTTACATGGGTGTAATAGCTGAAGAGGAAGCATATAAGTCAGAATTATTCCA
GAAAGCCAAGTCTCTAATGCAATGTGCAGGAGAAAGTATCACTCTGTTTAAAAATAAGACAAATGAGGAA
TTCAGAATTGGTTCCTTGAGAAATATGATGCAGCTATGTACACGTTGCTTGAGCAACTGTACCAAGAAGA
GTCCAAATAAGATTGCATCTGGCTTTTTCCTGCGATTGTTAACATCAAAGCTAATGAATGACATTGCAGA
TATTTGTAAAAGTTTAGCATCCTTCATCAAAAAGCCATTTGACCGTGGAGAAGTAGAATCAATGGAAGAT
GATACTAATGGAAATCTAATGGAGGTGGAGGATCAGTCATCCATGAATCTATTTAACGATTACCCTGATA
GTAGTGTTAGTGATGCAAACGAACCTGGAGAGAGCCAAAGTACCATAGGTGCCATTAATCCTTTAGCTGA
AGAATATCTGTCAAAGCAAGATCTACTTTTCTTAGACATGCTCAAGTTCTTGTGTTTGTGTGTAACTACT
GCTCAGACCAATACTGTGTCCTTTAGGGCAGCTGATATTCGGAGGAAATTGTTAATGTTAATTGATTCTA
GCACGCTAGAACCTACCAAATCCCTCCACCTGCATATGTATCTAATGCTTTTAAAGGAGCTTCCTGGAGA
AGAGTACCCCTTGCCAATGGAAGATGTTCTTGAACTTCTGAAACCACTATCCAATGTGTGTTCTTTGTAT
CGTCGTGACCAAGATGTTTGTAAAACTATTTTAAACCATGTCCTTCATGTAGTGAAAAACCTAGGTCAAA
GCAATATGGACTCTGAGAACACAAGGGATGCTCAAGGACAGTTTCTTACAGTAATTGGAGCATTTTGGCA
TCTAACAAAGGAGAGGAAATATATATTCTCTGTAAGAATGGCCCTAGTAAATTGCCTTAAAACTTTGCTT
GAGGCTGATCCTTATTCAAAATGGGCCATTCTTAATGTAATGGGAAAAGACTTTCCTGTAAATGAAGTAT
TTACACAATTTCTTGCTGACAATCATCACCAAGTTCGCATGTTGGCTGCAGAGTCAATCAATAGATTGTT
CCAGGACACGAAGGGAGATTCTTCCAGGTTACTGAAAGCACTTCCTTTGAAGCTTCAGCAAACAGCTTTT
GAAAATGCATACTTGAAAGCTCAGGAAGGAATGAGAGAAATGTCCCATAGTGCTGAGAACCCTGAAACTT
TGGATGAAATTTATAATAGAAAATCTGTTTTACTGACGTTGATAGCTGTGGTTTTATCCTGTAGCCCTAT

CTGCGAAAAACAGGCTTTGTTTGCCCTGTGTAAATCTGTGAAAGAGAATGGATTAGAACCTCACCTTGTG
AAAAAGGTTTTAGAGAAAGTTTCTGAAACTTTTGGATATAGACGTTTAGAAGACTTTATGGCATCTCATT
TAGATTATCTGGTTTTGGAATGGCTAAATCTTCAAGATACTGAATACAACTTATCTTCTTTTCCTTTTAT
TTTATTAAACTACACAAATATTGAGGATTTCTATAGATCTTGTTATAAGGTTTTGATTCCACATCTGGTG
ATTAGAAGTCATTTTGATGAGGTGAAGTCCATTGCTAATCAGATTCAAGAGGACTGGAAAAGTCTTCTAA
CAGACTGCTTTCCAAAGATTCTTGTAAATATTCTTCCTTATTTTGCCTATGAGGGTACCAGAGACAGTGG
GATGGCACAGCAAAGAGAGACTGCTACCAAGGTCTATGATATGCTTAAAAGTGAAAACTTATTGGGAAAA
CAGATTGATCACTTATTCATTAGTAATTTACCAGAGATTGTGGTGGAGTTATTGATGACGTTACATGAGC
CAGCAAATTCTAGTGCCAGTCAGAGCACTGACCTCTGTGACTTTTCAGGGGATTTGGATCCTGCTCCTAA
TCCACCTCATTTTCCATCGCATGTGATTAAAGCAACATTTGCCTATATCAGCAATTGTCATAAAACCAAG
TTAAAAAGCATTTTAGAAATTCTTTCCAAAAGCCCTGATTCCTATCAGAAAATTCTTCTTGCCATATGTG
AGCAAGCAGCTGAAACAAATAATGTTTATAAGAAGCACAGAATTCTTAAAATATATCACCTGTTTGTTAG
TTTATTACTGAAAGATATAAAAAGTGGCTTAGGAGGAGCTTGGGCCTTTGTTCTTCGAGACGTTATTTAT
ACTTTGATTCACTATATCAACCAAAGGCCTTCTTGTATCATGGATGTGTCATTACGTAGCTTCTCCCTTT
GTTGTGACTTATTAAGTCAGGTTTGCCAGACAGCCGTGACTTACTGTAAGGATGCTCTAGAAAACCATCT
TCATGTTATTGTTGGTACACTTATACCCCTTGTGTATGAGCAGGTGGAGGTTCAGAAACAGGTATTGGAC
TTGTTGAAATACTTAGTGATAGATAACAAGGATAATGAAAACCTCTATATCACGATTAAGCTTTTAGATC
CTTTTCCTGACCATGTTGTTTTTAAGGATTTGCGTATTACTCAGCAAAAAATCAAATACAGTAGAGGACC
CTTTTCACTCTTGGAGGAAATTAACCATTTTCTCTCAGTAAGTGTTTATGATGCACTTCCATTGACAAGA
CTTGAAGGACTAAAGGATCTTCGAAGACAACTGGAACTACATAAAGATCAGATGGTGGACATTATGAGAG
CTTCTCAGGATAATCCGCAAGATGGGATTATGGTGAAACTAGTTGTCAATTTGTTGCAGTTATCCAAGAT
GGCAATAAACCACACTGGTGAAAAAGAAGTTCTAGAGGCTGTTGGAAGCTGCTTGGGAGAAGTGGGTCCT
ATAGATTTCTCTACCATAGCTATACAACATAGTAAAGATGCATCTTATACCAAGGCCCTTAAGTTATTTG
AAGATAAAGAACTTCAGTGGACCTTCATAATGCTGACCTACCTGAATAACACACTGGTAGAAGATTGTGT
CAAAGTTCGATCAGCAGCTGTTACCTGTTTGAAAAACATTTTAGCCACAAAGACTGGACATAGTTTCTGG
GAGATTTATAAGATGACAACAGATCCAATGCTGGCCTATCTACAGCCTTTTAGAACATCAAGAAAAAAGT
TTTTAGAAGTACCCAGATTTGACAAAGAAAACCCTTTTGAAGGCCTGGATGATATAAATCTGTGGATTCC
TCTAAGTGAAAATCATGACATTTGGATAAAGACACTGACTTGTGCTTTTTTGGACAGTGGAGGCACAAAA
TGTGAAATTCTTCAATTATTAAAGCCAATGTGTGAAGTGAAAACTGACTTTTGTCAGACTGTACTTCCAT
ACTTGATTCATGATATTTTACTCCAAGATACAAATGAATCATGGAGAAATCTGCTTTCTACACATGTTCA
GGGATTTTTCACCAGCTGTCTTCGACACTTCTCGCAAACGAGCCGATCCACAACCCCTGCAAACTTGGAT
TCAGAGTCAGAGCACTTTTTCCGATGCTGTTTGGATAAAAAATCACAAAGAACAATGCTTGCTGTTGTGG
ACTACATGAGAAGACAAAAGAGACCTTCTTCAGGAACAATTTTTAATGATGCTTTCTGGCTGGATTTAAA
TTATCTAGAAGTTGCCAAGGTAGCTCAGTCTTGTGCTGCTCACTTTACAGCTTTACTCTATGCAGAAATC
TATGCAGATAAGAAAAGTATGGATGATCAAGAGAAAAGAAGTCTTGCATTTGAAGAAGGAAGCCAGAGTA
CAACTATTTCTAGCTTGAGTGAAAAAAGTAAAGAAGAAACTGGAATAAGTTTACAGGATCTTCTCTTAGA
AATCTACAGAAGTATAGGGGAGCCAGATAGTTTGTATGGCTGTGGTGGAGGGAAGATGTTACAACCCATT
ACTAGACTACGAACATATGAACACGAAGCAATGTGGGGCAAAGCCCTAGTAACATATGACCTCGAAACAG
CAATCCCCTCATCAACACGCCAGGCAGGAATCATTCAGGCCTTGCAGAATTTGGGACTCTGCCATATTCT
TTCCGTCTATTTAAAAGGATTGGATTATGAAAATAAAGACTGGTGTCCTGAACTAGAAGAACTTCATTAC
CAAGCAGCATGGAGGAATATGCAGTGGGACCATTGCACTTCCGTCAGCAAAGAAGTAGAAGGAACCAGTT
ACCATGAATCATTGTACAATGCTCTACAATCTCTAAGAGACAGAGAATTCTCTACATTTTATGAAAGTCT
CAAATATGCCAGAGTAAAAGAAGTGGAAGAGATGTGTAAGCGCAGCCTTGAGTCTGTGTATTCGCTCTAT
CCCACACTTAGCAGGTTGCAGGCCATTGGAGAGCTGGAAAGCATTGGGGAGCTTTTCTCAAGATCAGTCA
CACATAGACAACTCTCTGAAGTATATATTAAGTGGCAGAAACACTCCCAGCTTCTCAAGGACAGTGATTT
TAGTTTTCAGGAGCCTATCATGGCTCTACGCACAGTCATTTTGGAGATCCTGATGGAAAAGGAAATGGAC
AACTCACAAAGAGAATGTATTAAGGACATTCTCACCAAACACCTTGTAGAACTCTCTATACTGGCCAGAA
CTTTCAAGAACACTCAGCTCCCTGAAAGGGCAATATTTCAAATTAAACAGTACAATTCAGTTAGCTGTGG
AGTCTCTGAGTGGCAGCTGGAAGAAGCACAAGTATTCTGGGCAAAAAAGGAGCAGAGTCTTGCCCTGAGT
ATTCTCAAGCAAATGATCAAGAAGTTGGATGCCAGCTGTGCAGCGAACAATCCCAGCCTAAAACTTACAT
ACACAGAATGTCTGAGGGTTTGTGGCAACTGGTTAGCAGAAACGTGCTTAGAAAATCCTGCGGTCATCAT
GCAGACCTATCTAGAAAAGGCAGTAGAAGTTGCTGGAAATTATGATGGAGAAAGTAGTGATGAGCTAAGA
AATGGAAAAATGAAGGCATTTCTCTCATTAGCCCGGTTTTCAGATACTCAATACCAAAGAATTGAAAACT
ACATGAAATCATCGGAATTTGAAAACAAGCAAGCTCTCCTGAAAAGAGCCAAAGAGGAAGTAGGTCTCCT
TAGGGAACATAAAATTCAGACAAACAGATACACAGTAAAGGTTCAGCGAGAGCTGGAGTTGGATGAATTA
GCCCTGCGTGCACTGAAAGAGGATCGTAAACGCTTCTTATGTAAAGCAGTTGAAAATTATATCAACTGCT
TATTAAGTGGAGAAGAACATGATATGTGGGTATTCCGACTTTGTTCCCTCTGGCTTGAAAATTCTGGAGT
TTCTGAAGTCAATGGCATGATGAAGAGAGACGGAATGAAGATTCCAACATATAAATTTTTGCCTCTTATG
TACCAATTGGCTGCTAGAATGGGGACCAAGATGATGGGAGGCCTAGGATTTCATGAAGTCCTCAATAATC
TAATCTCTAGAATTTCAATGGATCACCCCCATCACACTTTGTTTATTATACTGGCCTTAGCAAATGCAAA
CAGAGATGAATTTCTGACTAAACCAGAGGTAGCCAGAAGAAGCAGAATAACTAAAAATGTGCCTAAACAA

AGCTCTCAGCTTGATGAGGATCGAACAGAGGCTGCAAATAGAATAATATGTACTATCAGAAGTAGGAGAC
CTCAGATGGTCAGAAGTGTTGAGGCACTTTGTGATGCTTATATTATATTAGCAAACTTAGATGCCACTCA
GTGGAAGACTCAGAGAAAAGGCATAAATATTCCAGCAGACCAGCCAATTACTAAACTTAAGAATTTAGAA
GATGTTGTTGTCCCTACTATGGAAATTAAGGTGGACCACACAGGAGAATATGGAAATCTGGTGACTATAC
AGTCATTTAAAGCAGAATTTCGCTTAGCAGGAGGTGTAAATTTACCAAAAATAATAGATTGTGTAGGTTC
CGATGGCAAGGAGAGGAGACAGCTTGTTAAGGGCCGTGATGACCTGAGACAAGATGCTGTCATGCAACAG
GTCTTCCAGATGTGTAATACATTACTGCAGAGAAACACGGAAACTAGGAAGAGGAAATTAACTATCTGTA
CTTATAAGGTGGTTCCCCTCTCTCAGCGAAGTGGTGTTCTTGAATGGTGCACAGGAACTGTCCCCATTGG
TGAATTTCTTGTTAACAATGAAGATGGTGCTCATAAAAGATACAGGCCAAATGATTTCAGTGCCTTTCAG
TGCCAAAAGAAAATGATGGAGGTGCAAAAAAAGTCTTTTGAAGAGAAATATGAAGTCTTCATGGATGTTT
GCCAAAATTTTCAACCAGTTTTCCGTTACTTCTGCATGGAAAAATTCTTGGATCCAGCTATTTGGTTTGA
GAAGCGATTGGCTTATACGCGCAGTGTAGCTACTTCTTCTATTGTTGGTTACATACTTGGACTTGGTGAT
AGACATGTACAGAATATCTTGATAAATGAGCAGTCAGCAGAACTTGTACATATAGATCTAGGTGTTGCTT
TTGAACAGGGCAAAATCCTTCCTACTCCTGAGACAGTTCCTTTTAGACTCACCAGAGATATTGTGGATGG
CATGGGCATTACGGGTGTTGAAGGTGTCTTCAGAAGATGCTGTGAGAAAACCATGGAAGTGATGAGAAAC
TCTCAGGAAACTCTGTTAACCATTGTAGAGGTCCTTCTATATGATCCACTCTTTGACTGGACCATGAATC
CTTTGAAAGCTTTGTATTTACAGCAGAGGCCGGAAGATGAAACTGAGCTTCACCCTACTCTGAATGCAGA
TGACCAAGAATGCAAACGAAATCTCAGTGATATTGACCAGAGTTTCAACAAAGTAGCTGAACGTGTCTTA
ATGAGACTACAAGAGAAACTGAAAGGAGTGGAAGAAGGCACTGTGCTCAGTGTTGGTGGACAAGTGAATT
TGCTCATACAGCAGGCCATAGACCCCAAAAATCTCAGCCGACTTTTCCCAGGATGGAAAGCTTGGGTGTG
A

(SEQ ID NO: 5; NCBI Accession No. NM_000051.3)

Ataxia telangiectasia and Rad3-related protein (ATR)

Ataxia telangiectasia and Rad3-related protein (ATR), also known as serine/threonine-protein kinase ATR or FRAP-related protein 1 (FRP1), is a serine/threonine-specific kinase involved in DNA damage sensing. It may be involved in activating the DNA damage checkpoint, which leads to cell cycle arrest

In one embodiment, the ATR is human ATR.

In one embodiment, the amino acid sequence of ATR is the sequence deposited under NCBI Accession No. NP_001175.2.

An example amino acid sequence of the ATR is:

MGEHGLELASMIPALRELGSATPEEYNTVVQKPRQILCQFIDRILTDVNVVAVELVKKTDSQPTSVMLLDFI
QHIMKSSPLMFVNVSGSHEAKGSCIEFSNWIITRLLRIAATPSCHLLHKKICEVICSLLFLFKSKSPAIFGV
LTKELLQLFEDLVYLHRRNVMGHAVEWPVVMSRFLSQLDEHMGYLQSAPLQLMSMQNLEFIEVTLLMVLTRI
IAIVFFRRQELLLWQIGCVLLEYGSPKIKSLAISFLTELFQLGGLPAQPASTFFSSFLELLKHLVEMDTDQL
KLYEEPLSKLIKTLFPFEAEAYRNIEPVYLNMLLEKLCVMFEDGVLMRLKSDLLKAALCHLLQYFLKFVPAG
YESALQVRKVYVRNICKALLDVLGIEVDAEYLLGPLYAALKMESMEIIEEIQCQTQQENLSSNSDGISPKRR
RLSSSLNPSKRAPKQTEEIKHVDMNQKSILWSALKQKAESLQISLEYSGLKNPVIEMLEGIAVVLQLTALCT
VHCSHQNMNCRTFKDCQHKSKKKPSVVITWMSLDFYTKVLKSCRSLLESVQKLDLEATIDKVVKIYDALIYM
QVNSSFEDHILEDLCGMLSLPWIYSHSDDGCLKLTTFAANLLTLSCRISDSYSPQAQSRCVFLLTLFPRRIF
LEWRTAVYNWALQSSHEVIRASCVSGFFILLQQQNSCNRVPKILIDKVKDDSDIVKKEFASILGQLVCTLHG
MFYLTSSLTEPFSEHGHVDLFCRNLKATSQHECSSSQLKASVCKPFLFLLKKKIPSPVKLAFIDNLHHLCKH
LDFREDETDVKAVLGTLLNLMEDPDKDVRVAFSGNIKHILESLDSEDGFIKELFVLRMKEAYTHAQISRNNE
LKDTLILTTGDIGRAAKGDLVPFALLHLLHCLLSKSASVSGAAYTEIRALVAAKSVKLQSFFSQYKKPICQF
LVESLHSSQMTALPNTPCQNADVRKQDVAHQREMALNTLSEIANVFDFPDLNRFLTRTLQVLLPDLAAKASP
AASALIRTLGKQLNVNRREILINNFKYIFSHLVCSCSKDELERALHYLKNETEIELGSLLRQDFQGLHNELL
LRIGEHYQQVFNGLSILASFASSDDPYQGPRDIISPELMADYLQPKLLGILAFFNMQLLSSSVGIEDKKMAL
NSLMSLMKLMGPKHVSSVRVKMMTTLRTGLRFKDDFPELCCRAWDCFVRCLDHACLGSLLSHVIVALLPLIH
IQPKETAAIFHYLIIENRDAVQDFLHEIYFLPDHPELKKIKAVLQEYRKETSESTDLQTTLQLSMKAIQHEN

VDVRIHALTSLKETLYKNQEKLIKYATDSETVEPIISQLVTVLLKGCQDANSQARLLCGECLGELGAIDPGR
LDFSTTETQGKDFTFVTGVEDSSFAYGLLMELTRAYLAYADNSRAQDSAAYAIQELLSIYDCREMETNGPGH
QLWRRFPEHVREILEPHLNTRYKSSQKSTDWSGVKKPIYLSKLGSNFAEWSASWAGYLITKVRHDLASKIFT
CCSIMMKHDFKVTIYLLPHILVYVLLGCNQEDQQEVYAEIMAVLKHDDQHTINTQDIASDLCQLSTQTVFSM
LDHLTQWARHKFQALKAEKCPHSKSNRNKVDSMVSTVDYEDYQSVTRFLDLIPQDTLAVASFRSKAYTRAVM
HFESFITEKKQNIQEHLGFLQKLYAAMHEPDGVAGVSAIRKAEPSLKEQILEHESLGLLRDATACYDRAIQL
EPDQIIHYHGVVKSMLGLGQLSTVITQVNGVHANRSEWTDELNTYRVEAAWKLSQWDLVENYLAADGKSTTW
SVRLGQLLLSAKKRDITAFYDSLKLVRAEQIVPLSAASFERGSYQRGYEYIVRLHMLCELEHSIKPLFQHSP
GDSSQEDSLNWVARLEMTQNSYRAKEPILALRRALLSLNKRPDYNEMVGECWLQSARVARKAGHHQTAYNAL
LNAGESRLAELYVERAKWLWSKGDVHQALIVLQKGVELCFPENETPPEGKNMLIHGRAMLLVGRFMEETANF
ESNAIMKKYKDVTACLPEWEDGHFYLAKYYDKLMPMVTDNKMEKQGDLIRYIVLHFGRSLQYGNQFIYQSMP
RMLTLWLDYGTKAYEWEKAGRSDRVQMRNDLGKINKVITEHTNYLAPYQFLTAFSQLISRICHSHDEVFVVL
MEIIAKVFLAYPQQAMWMMTAVSKSSYPMRVNRCKEILNKAIHMKKSLEKFVGDATRLTDKLLELCNKPVDG
SSSTLSMSTHFKMLKKLVEEATFSEILIPLQSVMIPTLPSILGTHANHASHEPFPGHWAYIAGFDDMVEILA
SLQKPKKISLKGSDGKFYIMMCKPKDDLRKDCRLMEFNSLINKCLRKDAESRRRELHIRTYAVIPLNDECGI
IEWVNNTAGLRPILTKLYKEKGVYMTGKELRQCMLPKSAALSEKLKVFREFLLPRHPPIFHEWFLRTFPDPT
SWYSSRSAYCRSTAVMSMVGYILGLGDRHGENILFDSLTGECVHVDFNCLFNKGETFEVPEIVPFRLTHNMV
NGMGPMGTEGLFRRACEVTMRLMRDQREPLMSVLKTFLHDPLVEWSKPVKGHSKAPLNETGEVVNEKAKTHV
LDIEQRLQGVIKTRNRVTGLPLSIEGHVHYLIQEATDENLLCQMYLGWTPYM

(SEQ ID NO: 6)

An example nucleic acid sequence encoding the ATR is:

ATGGGGGAACATGGCCTGGAGCTGGCTTCCATGATCCCCGCCCTGCGGGAGCTGGGCAGTGCCACACCAG
AGGAATATAATACAGTTGTACAGAAGCCAAGACAAATTCTGTGTCAATTCATTGACCGGATACTTACAGA
TGTAAATGTTGTTGCTGTAGAACTTGTAAAGAAAACTGACTCTCAGCCAACCTCCGTGATGTTGCTTGAT
TTCATCCAGCATATCATGAAATCCTCCCCACTTATGTTTGTAAATGTGAGTGGAAGCCATGAGGCCAAAG
GCAGTTGTATTGAATTCAGTAATTGGATCATAACGAGACTTCTGCGGATTGCAGCAACTCCCTCCTGTCA
TTTGTTACACAAGAAAATCTGTGAAGTCATCTGTTCATTATTATTTCTTTTTAAAAGCAAGAGTCCTGCT
ATTTTTGGGGTACTCACAAAAGAATTATTACAACTTTTTGAAGACTTGGTTTACCTCCATAGAAGAAATG
TGATGGGTCATGCTGTGGAATGGCCAGTGGTCATGAGCCGATTTTTAAGTCAATTAGATGAACACATGGG
ATATTTACAATCAGCTCCTTTGCAGTTGATGAGTATGCAAAATTTAGAATTTATTGAAGTCACTTTATTA
ATGGTTCTTACTCGTATTATTGCAATTGTGTTTTTTAGAAGGCAAGAACTCTTACTTTGGCAGATAGGTT
GTGTTCTGCTAGAGTATGGTAGTCCAAAAATTAAATCCCTAGCAATTAGCTTTTTAACAGAACTTTTTCA
GCTTGGAGGACTACCAGCACAACCAGCTAGCACTTTTTTCAGCTCATTTTTGGAATTATTAAAACACCTT
GTAGAAATGGATACTGACCAATTGAAACTCTATGAAGAGCCATTATCAAAGCTGATAAAGACACTATTTC
CCTTTGAAGCAGAAGCTTATAGAAATATTGAACCTGTCTATTTAAATATGCTGCTGGAAAAACTCTGTGT
CATGTTTGAAGACGGTGTGCTCATGCGGCTTAAGTCTGATTTGCTAAAAGCAGCTTTGTGCCATTTACTG
CAGTATTTCCTTAAATTTGTGCCAGCTGGGTATGAATCTGCTTTACAAGTCAGGAAGGTCTATGTGAGAA
ATATTTGTAAAGCTCTTTTGGATGTGCTTGGAATTGAGGTAGATGCAGAGTACTTGTTGGGCCCACTTTA
TGCAGCTTTGAAAATGGAAAGTATGGAAATCATTGAGGAGATTCAATGCCAAACTCAACAGGAAAACCTC
AGCAGTAATAGTGATGGAATATCACCCAAAAGGCGTCGTCTCAGCTCGTCTCTAAACCCTTCTAAAAGAG
CACCAAAACAGACTGAGGAAATTAAACATGTGGACATGAACCAAAAGAGCATATTATGGAGTGCACTGAA
ACAGAAAGCTGAATCCCTTCAGATTTCCCTTGAATACAGTGGCCTAAAGAATCCTGTTATTGAGATGTTA
GAAGGAATTGCTGTTGTCTTACAACTGACTGCTCTGTGTACTGTTCATTGTTCTCATCAAAACATGAACT
GCCGTACTTTCAAGGACTGTCAACATAAATCCAAGAAGAAACCTTCTGTAGTGATAACTTGGATGTCATT
GGATTTTTACACAAAAGTGCTTAAGAGCTGTAGAAGTTTGTTAGAATCTGTTCAGAAACTGGACCTGGAG
GCAACCATTGATAAGGTGGTGAAAATTTATGATGCTTTGATTTATATGCAAGTAAACAGTTCATTTGAAG
ATCATATCCTGGAAGATTTATGTGGTATGCTCTCACTTCCATGGATTTATTCCCATTCTGATGATGGCTG
TTTAAAGTTGACCACATTTGCCGCTAATCTTCTAACATTAAGCTGTAGGATTTCAGATAGCTATTCACCA
CAGGCACAATCACGATGTGTGTTTCTTCTGACTCTGTTTCCAAGAAGAATATTCCTTGAGTGGAGAACAG
CAGTTTACAACTGGGCCCTGCAGAGCTCCCATGAAGTAATCCGGGCTAGTTGTGTTAGTGGATTTTTTAT
CTTATTGCAGCAGCAGAATTCTTGTAACAGAGTTCCCAAGATTCTTATAGATAAAGTCAAAGATGATTCT
GACATTGTCAAGAAAGAATTTGCTTCTATACTTGGTCAACTTGTCTGTACTCTTCACGGCATGTTTTATC
TGACAAGTTCTTTAACAGAACCTTTCTCTGAACACGGACATGTGGACCTCTTCTGTAGGAACTTGAAAGC
CACTTCTCAACATGAATGTTCATCTTCTCAACTAAAAGCTTCTGTCTGCAAGCCATTCCTTTTCCTACTG
AAAAAAAAAATACCTAGTCCAGTAAAACTTGCTTTCATAGATAATCTACATCATCTTTGTAAGCATCTTG
ATTTTAGAGAAGATGAAACAGATGTAAAAGCAGTTCTTGGAACTTTATTAAATTTAATGGAAGATCCAGA
CAAAGATGTTAGAGTGGCTTTTAGTGGAAATATCAAGCACATATTGGAATCCTTGGACTCTGAAGATGGA

TTTATAAAGGAGCTTTTTGTCTTAAGAATGAAGGAAGCATATACACATGCCCAAATATCAAGAAATAATG
AGCTGAAGGATACCTTGATTCTTACAACAGGGGATATTGGAAGGGCCGCAAAAGGAGATTTGGTACCATT
TGCACTCTTACACTTATTGCATTGTTTGTTATCCAAGTCAGCATCTGTCTCTGGAGCAGCATACACAGAA
ATTAGAGCTCTGGTTGCAGCTAAAAGTGTTAAACTGCAAAGTTTTTTCAGCCAGTATAAGAAACCCATCT
GTCAGTTTTTGGTAGAATCCCTTCACTCTAGTCAGATGACAGCACTTCCGAATACTCCATGCCAGAATGC
TGACGTGCGAAAACAAGATGTGGCTCACCAGAGAGAAATGGCTTTAAATACGTTGTCTGAAATTGCCAAC
GTTTTCGACTTTCCTGATCTTAATCGTTTTCTTACTAGGACATTACAAGTTCTACTACCTGATCTTGCTG
CCAAAGCAAGCCCTGCAGCTTCTGCTCTCATTCGAACTTTAGGAAAACAATTAAATGTCAATCGTAGAGA
GATTTTAATAAACAACTTCAAATATATTTTTTCTCATTTGGTCTGTTCTTGTTCCAAAGATGAATTAGAA
CGTGCCCTTCATTATCTGAAGAATGAAACAGAAATTGAACTGGGGAGCCTGTTGAGACAAGATTTCCAAG
GATTGCATAATGAATTATTGCTGCGTATTGGAGAACACTATCAACAGGTTTTTAATGGTTTGTCAATACT
TGCCTCATTTGCATCCAGTGATGATCCATATCAGGGCCCGAGAGATATCATATCACCTGAACTGATGGCT
GATTATTTACAACCCAAATTGTTGGGCATTTTGGCTTTTTTTAACATGCAGTTACTGAGCTCTAGTGTTG
GCATTGAAGATAAGAAAATGGCCTTGAACAGTTTGATGTCTTTGATGAAGTTAATGGGACCCAAACATGT
CAGTTCTGTGAGGGTGAAGATGATGACCACACTGAGAACTGGCCTTCGATTCAAGGATGATTTTCCTGAA
TTGTGTTGCAGAGCTTGGGACTGCTTTGTTCGCTGCCTGGATCATGCTTGTCTGGGCTCCCTTCTCAGTC
ATGTAATAGTAGCTTTGTTACCTCTTATACACATCCAGCCTAAAGAAACTGCAGCTATCTTCCACTACCT
CATAATTGAAAACAGGGATGCTGTGCAAGATTTTCTTCATGAAATATATTTTTTACCTGATCATCCAGAA
TTAAAAAAGATAAAAGCCGTTCTCCAGGAATACAGAAAGGAGACCTCTGAGAGCACTGATCTTCAGACAA
CTCTTCAGCTCTCTATGAAGGCCATTCAACATGAAAATGTCGATGTTCGTATTCATGCTCTTACAAGCTT
GAAGGAAACCTTGTATAAAAATCAGGAAAAACTGATAAAGTATGCAACAGACAGTGAAACAGTAGAACCT
ATTATCTCACAGTTGGTGACAGTGCTTTTGAAAGGTTGCCAAGATGCAAACTCTCAAGCTCGGTTGCTCT
GTGGGGAATGTTTAGGGGAATTGGGGGCGATAGATCCAGGTCGATTAGATTTCTCAACAACTGAAACTCA
AGGAAAAGATTTTACATTTGTGACTGGAGTAGAAGATTCAAGCTTTGCCTATGGATTATTGATGGAGCTA
ACAAGAGCTTACCTTGCGTATGCTGATAATAGCCGAGCTCAAGATTCAGCTGCCTATGCCATTCAGGAGT
TGCTTTCTATTTATGACTGTAGAGAGATGGAGACCAACGGCCCAGGTCACCAATTGTGGAGGAGATTTCC
TGAGCATGTTCGGGAAATACTAGAACCTCATCTAAATACCAGATACAAGAGTTCTCAGAAGTCAACCGAT
TGGTCTGGAGTAAAGAAGCCAATTTACTTAAGTAAATTGGGTAGTAACTTTGCAGAATGGTCAGCATCTT
GGGCAGGTTATCTTATTACAAAGGTTCGACATGATCTTGCCAGTAAAATTTTCACCTGCTGTAGCATTAT
GATGAAGCATGATTTCAAAGTGACCATCTATCTTCTTCCACATATTCTGGTGTATGTCTTACTGGGTTGT
AATCAAGAAGATCAGCAGGAGGTTTATGCAGAAATTATGGCAGTTCTAAAGCATGACGATCAGCATACCA
TAAATACCCAAGACATTGCATCTGATCTGTGTCAACTCAGTACACAGACTGTGTTCTCCATGCTTGACCA
TCTCACACAGTGGGCAAGGCACAAATTTCAGGCACTGAAAGCTGAGAAATGTCCACACAGCAAATCAAAC
AGAAATAAGGTAGACTCAATGGTATCTACTGTGGATTATGAAGACTATCAGAGTGTAACCCGTTTTCTAG
ACCTCATACCCCAGGATACTCTGGCAGTAGCTTCCTTTCGCTCCAAAGCATACACACGAGCTGTAATGCA
CTTTGAATCATTTATTACAGAAAAGAAGCAAAATATTCAGGAACATCTTGGATTTTTACAGAAATTGTAT
GCTGCTATGCATGAACCTGATGGAGTGGCCGGAGTCAGTGCAATTAGAAAGGCAGAACCATCTCTAAAAG
AACAGATCCTTGAACATGAAAGCCTTGGCTTGCTGAGGGATGCCACTGCTTGTTATGACAGGGCTATTCA
GCTAGAACCAGACCAGATCATTCATTATCATGGTGTAGTAAAGTCCATGTTAGGTCTTGGTCAGCTGTCT
ACTGTTATCACTCAGGTGAATGGAGTGCATGCTAACAGGTCCGAGTGGACAGATGAATTAAACACGTACA
GAGTGGAAGCAGCTTGGAAATTGTCACAGTGGGATTTGGTGGAAAACTATTTGGCAGCAGATGGAAAATC
TACAACATGGAGTGTCAGACTGGGACAGCTATTATTATCAGCCAAAAAAAGAGATATCACAGCTTTTTAT
GACTCACTGAAACTAGTGAGAGCAGAACAAATTGTACCTCTTTCAGCTGCAAGCTTTGAAAGAGGCTCCT
ACCAACGAGGATATGAATATATTGTGAGATTGCACATGTTATGTGAGTTGGAGCATAGCATCAAACCACT
TTTCCAGCATTCTCCAGGTGACAGTTCTCAAGAAGATTCTCTAAACTGGGTAGCTCGACTAGAAATGACC
CAGAATTCCTACAGAGCCAAGGAGCCTATCCTGGCTCTCCGGAGGGCTTTACTAAGCCTCAACAAAAGAC
CAGATTACAATGAAATGGTTGGAGAATGCTGGCTGCAGAGTGCCAGGGTAGCTAGAAAGGCTGGTCACCA
CCAGACAGCCTACAATGCTCTCCTTAATGCAGGGGAATCACGACTCGCTGAACTGTACGTGGAAAGGGCA
AAGTGGCTCTGGTCCAAGGGTGATGTTCACCAGGCACTAATTGTTCTTCAAAAAGGTGTTGAATTATGTT
TTCCTGAAAATGAAACCCCACCTGAGGGTAAGAACATGTTAATCCATGGTCGAGCTATGCTACTAGTGGG
CCGATTTATGGAAGAAACAGCTAACTTTGAAAGCAATGCAATTATGAAAAAATATAAGGATGTGACCGCG
TGCCTGCCAGAATGGGAGGATGGGCATTTTTACCTTGCCAAGTACTATGACAAATTGATGCCCATGGTCA
CAGACAACAAAATGGAAAAGCAAGGTGATCTCATCCGGTATATAGTTCTTCATTTTGGCAGATCTCTACA
ATATGGAAATCAGTTCATATATCAGTCAATGCCACGAATGTTAACTCTATGGCTTGATTATGGTACAAAG
GCATATGAATGGGAAAAAGCTGGCCGCTCCGATCGTGTACAAATGAGGAATGATTTGGGTAAAATAAACA
AGGTTATCACAGAGCATACAAACTATTTAGCTCCATATCAATTTTTGACTGCTTTTTCACAATTGATCTC
TCGAATTTGTCATTCTCACGATGAAGTTTTTGTTGTCTTGATGGAAATAATAGCCAAAGTATTTCTAGCC
TATCCTCAACAAGCAATGTGGATGATGACAGCTGTGTCAAAGTCATCTTATCCCATGCGTGTGAACAGAT
GCAAGGAAATCCTCAATAAAGCTATTCATATGAAAAAATCCTTAGAGAAGTTTGTTGGAGATGCAACTCG
CCTAACAGATAAGCTTCTAGAATTGTGCAATAAACCGGTTGATGGAAGTAGTTCCACATTAAGCATGAGC
ACTCATTTTAAAATGCTTAAAAAGCTGGTAGAAGAAGCAACATTTAGTGAAATCCTCATTCCTCTACAAT

CAGTCATGATACCTACACTTCCATCAATTCTGGGTACCCATGCTAACCATGCTAGCCATGAACCATTTCC
TGGACATTGGGCCTATATTGCAGGGTTTGATGATATGGTGGAAATTCTTGCTTCTCTTCAGAAACCAAAG
AAGATTTCTTTAAAAGGCTCAGATGGAAAGTTCTACATCATGATGTGTAAGCCAAAAGATGACCTGAGAA
AGGATTGTAGACTAATGGAATTCAATTCCTTGATTAATAAGTGCTTAAGAAAAGATGCAGAGTCTCGTAG
AAGAGAACTTCATATTCGAACATATGCAGTTATTCCACTAAATGATGAATGTGGGATTATTGAATGGGTG
AACAACACTGCTGGTTTGAGACCTATTCTGACCAAACTATATAAAGAAAAGGGAGTGTATATGACAGGAA
AAGAACTTCGCCAGTGTATGCTACCAAAGTCAGCAGCTTTATCTGAAAAACTCAAAGTATTCCGAGAATT
TCTCCTGCCCAGGCATCCTCCTATTTTTCATGAGTGGTTTCTGAGAACATTCCCTGATCCTACATCATGG
TACAGTAGTAGATCAGCTTACTGCCGTTCCACTGCAGTAATGTCAATGGTTGGTTATATTCTGGGGCTTG
GAGACCGTCATGGTGAAAATATTCTCTTTGATTCTTTGACTGGTGAATGCGTACATGTAGATTTCAATTG
TCTTTTCAATAAGGGAGAAACCTTTGAAGTTCCAGAAATTGTGCCATTTCGCCTGACTCATAATATGGTT
AATGGAATGGGTCCTATGGGAACAGAGGGTCTTTTTCGAAGAGCATGTGAAGTTACAATGAGGCTGATGC
GTGATCAGCGAGAGCCTTTAATGAGTGTCTTAAAGACTTTTCTACATGATCCTCTTGTGGAATGGAGTAA
ACCAGTGAAAGGGCATTCCAAAGCGCCACTGAATGAAACTGGAGAAGTTGTCAATGAAAAGGCCAAGACC
CATGTTCTTGACATTGAGCAGCGACTACAAGGTGTAATCAAGACTCGAAATAGAGTGACAGGACTGCCGT
TATCTATTGAAGGACATGTGCATTACCTTATACAGGAAGCTACTGATGAAAACTTACTATGCCAGATGTA
TCTTGGTTGGACTCCATATATGTGA

(SEQ ID NO: 7; NCBI Accession No. NM_001184.3)

Kinase inhibitors

In one aspect the invention provides an ataxia telangiectasia mutated (ATM) kinase inhibitor or an ataxia telangiectasia and Rad3-related protein (ATR) inhibitor for use in haematopoietic cell, haematopoietic stem cell and/or haematopoietic progenitor cell gene therapy.

The invention also provides methods for identifying agents that are capable of acting as ATM kinase or ATR inhibitors and agents that are identified by such methods.

The activity of ATM kinase and ATR may be analysed directly, for example by analysing the enzymatic activity of the ATM kinase or ATR in vitro.

The ability of a candidate agent to inhibit (e.g. reduce the activity) ATM kinase or ATR may be expressed in terms of an IC50 value, which is the concentration of an agent that is required to give rise to a 50% reduction in the activity of the kinase. Preferably, the inhibitors of the invention have an IC50 value for inhibition (e.g. of ATM kinase or ATR) of less than 100 μM, more preferably less than 10 μM, for example less than 1 μM, less than 100 nM or less than 10 nM (e.g. KU-55933 has an IC50 value of about 13 nM for ATM kinase).

A number of techniques are known in the art for measuring kinase activity. Preferably, the kinase activity assays are carried out on a kinase (e.g. ATM kinase or ATR) that has been isolated from a cell. The kinase may have been expressed using recombinant techniques, and preferably has been purified. For example, kinase activity may be determined by monitoring the incorporation of radiolabelled phosphate from [γ-^{32}P]-labelled ATP into a substrate. Such assay techniques are described in, for example, Hastie et al. (Hastie, C.J. et al. (2006) Nat. Protocols 1: 968-971).

Preferably, the inhibitors are of low toxicity for mammals, such as humans, and in particular are of low toxicity for haematopoietic stem and/or progenitor cells.

A candidate inhibitor may be further analysed for its ability to increase cell survival and/or engraftment using a method as disclosed herein.

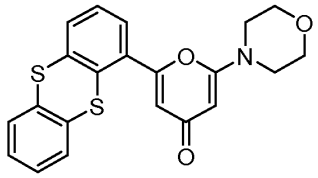
Preferably, the inhibitor is a transient inhibitor (e.g. has an inhibitory action lasting less than about 1, 2, 3, 4, 5, 6, 7 or 14 days).

Preferably, the inhibitor is a pharmacological inhibitor.

KU-55933

In a preferred embodiment, the inhibitor is KU-55933 or a derivative thereof.

KU-55933 (CAS No. 587871-26-9) is a selective, competitive ATM kinase inhibitor having the following structure:



Solutions of KU-55933 for use in the invention may be prepared using routine methods known in the art, for example KU-55933 is known to be soluble in DMSO and ethanol.

The concentration at which KU-55933 or a derivative thereof is applied to a population of haematopoietic stem and/or progenitor cells may be adjusted for different vector systems to optimise cell survival (e.g. during in vitro or ex vivo culture) and/or engraftment.

The invention encompasses the use of KU-55933 and derivatives of KU-55933. The KU-55933 derivatives of the invention are those which increase the survival (e.g. during in vitro or ex vivo culture) and/or engraftment of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells, in particular cells transduced by a viral vector.

KU-55933 derivatives of the invention may have been developed, for example, for increased solubility, increased stability and/or reduced toxicity.

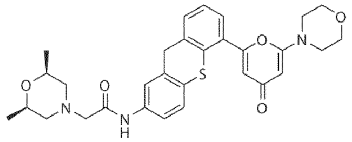
KU-55933 derivatives of the invention are preferably of low toxicity for mammals, in particular humans. Preferably, KU-55933 derivatives of the invention are of low toxicity for haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

Suitable KU-55933 derivatives may be identified using methods known in the art for determining cell survival in culture and/or engraftment. Examples of such methods have been described above. The method employed is preferably one which is amenable to automation and/or high throughput screening of candidate KU-55933 derivatives. The candidate KU-55933 derivatives may form part of a library of KU-55933 derivatives.

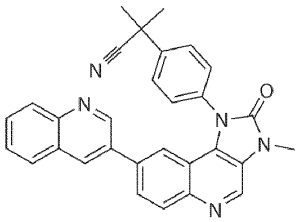
Additional inhibitors

Further kinase inhibitors that may be used in the invention include:

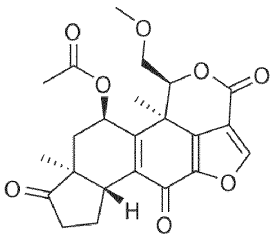
KU-60019, which is an improved analogue of KU-55933 and has an IC50 of 6.3 nM for ATM kinase in cell-free assays. KU-60019 has the structure:



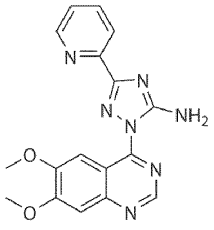
BEZ235 (NVP-BEZ235, Dactolisib), which is a dual ATP-competitive PI3K and mTOR inhibitor for p110α/γ/δ/β and mTOR(p70S6K) and inhibits ATR with an IC50 of about 21 nM in 3T3TopBP1-ER cells. BEZ235 has the structure:



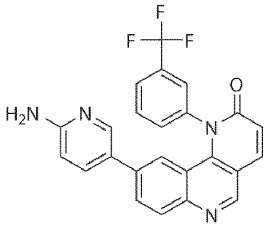
Wortmannin, which has the structure:



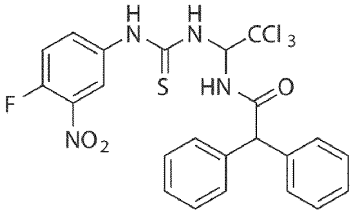
CP-466722, which is a potent and reversible ATM kinase inhibitor, but does not affect ATR. CP-466722 has the structure:



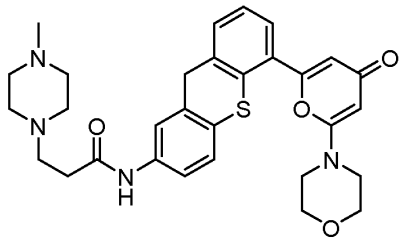
Torin 2, which ATM kinase and ATR with EC50 values of 28 nM and 35 nM, respectively. Torin 2 has the structure:



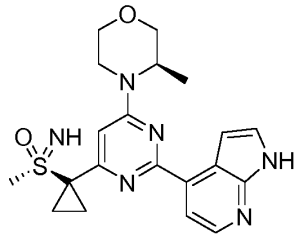
CGK 733 (CAS No. 905973-89-9), which is a potent and selective inhibitor of ATM kinase and ATR with IC50 values of about 200 nM. CGK 733 has the structure:



KU-559403 (Weber et al. (2015) Pharmacology & Therapeutics 149: 124–138). KU-559403 has the structure:

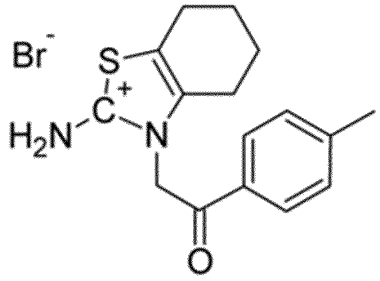


AZD6738, which has the structure:



Derivatives of these inhibitors, possessing characteristics as described for the KU-55933 derivatives, may also be used in the invention, and may be identified using analogous methods to those described for the KU-55933 derivatives.

In a preferred embodiment, the p53 inhibitor may be pifithrin-α, pifithrin-α cyclic and pifithrin-α p-nitro or a derivative thereof. Pifithrin-α has the structure:



siRNAs, shRNAs, miRNAs and antisense DNAs/RNAs

Inhibition (e.g. of the kinase) may be achieved using post-transcriptional gene silencing (PTGS). Post-transcriptional gene silencing mediated by double-stranded RNA (dsRNA) is a conserved cellular defence mechanism for controlling the expression of foreign genes. It is thought that the random integration of elements such as transposons or viruses causes the expression of dsRNA which activates sequence-specific degradation of homologous single-stranded mRNA or viral genomic RNA. The silencing effect is known as RNA interference (RNAi) (Ralph et al. (2005) Nat. Medicine 11: 429-433). The mechanism of RNAi involves the processing of long dsRNAs into duplexes of about 21-25 nucleotide (nt) RNAs. These products are called small interfering or silencing RNAs (siRNAs) which are the sequence-specific mediators of mRNA degradation. In differentiated mammalian cells, dsRNA >30 bp has been found to activate the interferon response leading to shut-down of protein synthesis and non-specific mRNA degradation (Stark et al. (1998) Ann. Rev. Biochem. 67: 227-64). However, this response can be bypassed by using 21 nt siRNA duplexes (Elbashir et al. (2001) EMBO J. 20: 6877-88; Hutvagner et al. (2001) Science 293: 834-8) allowing gene function to be analysed in cultured mammalian cells.

shRNAs consist of short inverted RNA repeats separated by a small loop sequence. These are rapidly processed by the cellular machinery into 19-22 nt siRNAs, thereby suppressing the target gene expression.

Micro-RNAs (miRNAs) are small (22–25 nucleotides in length) noncoding RNAs that can effectively reduce the translation of target mRNAs by binding to their 3' untranslated region (UTR). Micro-RNAs are a very large group of small RNAs produced naturally in organisms, at least some of which regulate the expression of target genes. Founding members of the micro-RNA family are *let-7* and *lin-4*. The *let-7* gene encodes a small, highly conserved RNA species that regulates the expression of endogenous protein-coding genes during worm development. The active RNA species is transcribed initially as an ~70 nt precursor, which is post-transcriptionally processed into a mature ~21 nt form. Both *let-7* and *lin-4* are transcribed as hairpin RNA precursors which are processed to their mature forms by Dicer enzyme.

The antisense concept is to selectively bind short, possibly modified, DNA or RNA molecules to messenger RNA in cells and prevent the synthesis of the encoded protein.

Methods for the design of siRNAs, shRNAs, miRNAs and antisense DNAs/RNAs to modulate the expression of a target protein, and methods for the delivery of these agents to a cell of interest are well known in the art.

Dominant negative peptides

In a preferred embodiment, the inhibitor of p53 activation is a mutant p53 peptide.

In a preferred embodiment, the inhibitor of p53 activation is a dominant negative peptide (e.g. a dominant negative p53 peptide).

Suitably, a dominant negative peptide may comprise mutations in the homo-oligomerisation domain. Suitably, dominant negative peptides comprising mutations in the homo-oligomerisation domain may dimerise with wild-type p53 and prevent wild-type p53 from activating transcription.

In a preferred embodiment the dominant negative peptide is GSE56 or a variant thereof.

In one embodiment the nucleotide sequence for mRNA translation of GSE56 is set forth in SEQ ID No. 8.

ATGGATGGATGGTGTCCTGGGAGAGACCGTCGGACAGAGGAAGAAAATTTCCGCAAAAAAGAAGAGCATTGCCCG
GAGCTGCCCCCAGGGAGTGCAAAGAGAGCACTGCCCACCAGCACAAGCTCCTCTCCCCAGCAAAAGAAAAAACCA
CTCGATGGAGAATATTTCACCCTTAAGATCCGTGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAATGAGGCC

TTGGAATTAAAGGATGCCCGTGCTGCCGAGGAGTCAGGAGACAGCAGGGCTCACTCCAGCTACCCGAAGATAGTT
AGTTAG (SEQ ID NO. 8)

In one embodiment, the amino acid sequence of GSE56 is set forth in SEQ ID No. 9.

MDGWCPGRDRRTEEENFRKKEEHCPELPPGSAKRALPTSTSSSPQQKKKPLDGEYFTLKIRGRERFEMFRELNEA
LELKDARAAEESGDSRAHSSYPKIVS (SEQ ID No. 9)

Suitably, the inhibitor of 53 activation may be a nucleotide sequence which encodes GSE56. Suitably, the inhibitor of p53 activation may be an amino acid sequence encoding GSE56. Suitably, the inhibitor of p53 activation may be GSE56 mRNA.

Adenoviral proteins

Adenoviruses are natural co-helpers of AAV infection and provide a set of genes: E1a, E1b, E2a and E4 which optimize AAV infection.

The present inventors have demonstrated that delivery of adenoviral proteins during gene editing improves the efficiency of HDR in HSC (and enhances the long-term repopulating activity of HSC. Without wishing to be bound by theory, the adenoviral proteins may provide helper functions to the AAV infection during gene editing. The adenoviral proteins may act directly or indirectly on the p53 pathway.

In one embodiment, the agent which promotes homology directed DNA repair comprises a nucleic acid sequence encoding at least one adenoviral protein. The adenoviral protein is not limited to a particular Adenovirus serotype. For example, in one embodiment, the at least one adenoviral proteins is from an Adenovirus of serotype 4, Adenovirus of serotype 5, Adenovirus of serotype 7 and/or Adenovirus of serotype 9.

In one embodiment, the at least one adenoviral protein is selected from the group comprising E1a, E1b, E2a and E4.

In a preferred embodiment, the at least one adenoviral protein is an open reading frame of the E4 gene.

In a preferred embodiment, the at least one adenoviral protein is E4orf1 or a variant thereof.

An example of a nucleotide sequence encoding Ad5-E4orf1 is set forth in SEQ ID No. 10. Suitably, the at least one adenoviral protein may comprise a nucleotide sequence for mRNA translation as set forth in SEQ ID No. 10 or a variant thereof.

ATGGCCGCTGCTGTGGAAGCCCTGTACGTGGTGCTTGAAAGAGAGGGCGCCATCCTGCCTAGACAAGAGGGCTTT
TCTGGCGTGTACGTGTTCTTCAGCCCCATCAACTTCGTGATCCCTCCAATGGGCGCCGTGATGCTGAGCCTGAGA
CTGAGAGTGTGTATCCCTCCTGGCTACTTCGGCCGGTTTCTGGCCCTGACCGATGTGAACCAGCCTGACGTGTTC
ACCGAGAGCTACATCATGACCCCTGACATGACCGAGGAACTGAGCGTGGTGCTGTTCAACCACGGCGACCAGTTC
TTTTATGGCCACGCCGGAATGGCCGTCGTGCGGCTGATGCTGATCAGAGTGGTGTTTCCCGTCGTCCGGCAGGCC
AGCAATGTTTGA (SEQ ID No. 10).

An example of an amino acid sequence of Ad5-E4orf1 is set forth in SEQ ID No. 1. Suitably, the at least one adenoviral protein may comprise an amino acid sequence as set forth in SEQ ID No. 1 or a variant thereof.

MAAAVEALYVVLEREGAILPRQEGFSGVYVFFSPINFVIPPMGAVMLSLRLRVCIPPGYFGRFLALTDVNQPDVF
TESYIMTPDMTEELSVVLFNHGDQFFYGHAGMAVVRLMLIRVVFPVVRQASNV (SEQ ID No. 1)

Other examples of an amino acid sequence of E4orf1 are set forth in SEQ ID No. 57 to SEQ ID No. 76. Suitably, the at least one adenoviral protein may comprise an amino acid sequence as set forth in SEQ ID No. 57 to SEQ ID No. 76 or a variant thereof.

SEQ ID No. 57	**Ad1-E4orf1**	MAAAVEALYVVLEREGAILPRQEGFSGVYVFFSPINFV IPPMGAVMLPLRLRVCIPPGYFGRFLALTDVNQPDVF TESYIMTPDMTEELSVVLFNHGDQFFYGHAGMAVVR LMLIRVVFPVVRQASNV
SEQ ID No. 58	**Ad2-E4orf1**	MAAAVEALYVVLEREGAILPRQEGFSGVYVFFSPINFV IPPMGAVMLSLRLRVCIPPGYFGRFLALTDVNQPDVF TESYIMTPDMIEELSVVLFNHGDQFFYGHAGMAVVRL MLIRVVFPVVRQASNV
SEQ ID No. 59	**Ad3-E4orf1**	MADEALYVYLEGPGATLPEQQQRNNYIFYSPVPFTLY PRGVALLYLKLSIIIPRGYVGCFFSLTDANMSGLYASS RIIHAGHREELSVLLFNHNDRFYEGRAGDPVACLVME RLIFPPVRQATMI
SEQ ID No. 60	**Ad4-E4orf1**	MDAQVLYVFLEGAGALLPVQKGSNYIFYAPANFVLHP HGVALLELRLSIVVPQGFIGRFFSLTDANVPGVYASSR IIHAGHREGLSVMLFNHNVSFYNGRAGDPVACLVLER VIYPPVRQASMV
SEQ ID No. 61	**Ad6-E4orf1**	MAAAVEALYVVLEREGAILPRQEGFSGVYVFFSPINFV IPPMGAVMLSLRLRVCIPPGYFGRFLALTDVSQPDVF TESYIMTPDMTEELSVVLFNHGDQFFYGHAGMAVVR LMLIRVVFPVVRQASNV
SEQ ID No. 62	**Ad7-E4orf1**	MADEALYVYLEGPGATLPEQQQRNNYIFYSPVPFTLY PRGVALLYLRLSIIIPRGYVGCFFSLTDANMSGLYASS RIIHAGHREELSVLLFNHDDRFYEGRAGDPVACLVME RLIYPPVRQATMI
SEQ ID No. 63	**Ad9-E4orf1**	MAESLYAFIDSPGGIAPVQEGTSNRYTFFCPESFHIPP HGVVLLHLKVSVLVPTGYQGRFMALNDYHARDILTQS DVIFAGRRQELTVLLFNHTDRFLYVRKGHPVGTLLLE RVIFPSVKIATLV
SEQ ID No. 64	**Ad14-E4orf1**	MADEALYVYLEGPGATLPEQQQRNNYIFYSPVPFTLY PRGVALLYLRLSIIIPRGYIGCFLSLTDANMFGLYASSRI IHAGHREELSVLLFNHDDRFYEGRAGDPVACLVMERL

		IYPPVRQATLI
SEQ ID No. 65	**Ad16-E4orf1**	MADEALYVYFRGPGATLPEQQQQRNNYIFYSPVPFTL YPRGVALLYLRLSIIIPRGYVGCFFSLTDANMSGLYAS SRIIHAGHREELSVLLFNHDDRFYEGRAGDPVACLVM ERLIYPPVRQATMI
SEQ ID No. 66	**Ad18-E4orf1**	MAALQALYVYFKGPGAMLPEQEGYSNAYVLFSPANF VIPPHGVVLLYLHIAVDIPPGYLGTLFSLSDMNARGVF VGAETLYPGSRMELSVLLFNHSDVFCDVRAKQPVAR LLLSRVIFPPVRQASLL
SEQ ID No. 67	**Ad20-E4orf1**	MAESLYAFIDSPGGIAPVQEGTSNRYNFFCPQSFHIP PHGVVLLHLKVSVLVPTGYQGRFMALNDYHARDILTQ SDVIFAGRRQELTVLLFNHTDRFLYVRKGHPVGTLLLE RVIFPSVKIATLV
SEQ ID No. 68	**Ad21-E4orf1**	MAEVLYVILEGPGARLPVQEGNNYIFYAPVDFTLHPR GVALLHLRLSIIVPRCYIGRFFSLTDTNTSGLYASSQIIF AAHQQPLSVMLFNHTDRFYEGRVGDPVACLVLERVIY PSVRQASMM
SEQ ID No. 69	**Ad23-E4orf1**	MAESLYAFIDSPGGIAPVQEGSSNRYNFFCPESFHIPP HGVVLLHLRVSVLIPTGYQGRFMALNDYHARGILTQS DVIFAGRRHELTVLLFNHTDRFLYVREGHPVGTLLLER VIFPSVRLATLV
SEQ ID No. 70	**Ad25-E4orf1**	MAESLYAFIDSPGGIAPVQEGTSNRYTFFCPESFHIPP HGVVLVHLRVSVLIPNGYQGRFMALNDYHSRGILTQS DVIFAGRRQELTVLLFNHTDRFLYVREGHPVGTLLLE RVIFPSVRLATLV
SEQ ID No. 71	**Ad30-E4orf1**	MAESLYAFIDSPGGIAPVQEGASNRYTFFCPESFHIPP HGVILLHLRVSVLVPTGYQGRFMALNDYHARGILTQS DVIFAGRRHDLSVLLFNHTDRFLYVREGHPVGTLLLE RVIFPSVRLATLV
SEQ ID No. 72	**Ad36-E4orf1**	MAESLYAFIDSPGGIAPVQEGASNRYIFFCPESFHIPP HGVILLHLRVSVLVPTGYQGRFMALNDYHARGILTQS DVIFAGRRHDLSVLLFNHTDRFLYVREGHPVGTLLLE RVIFPSVRIATLV
SEQ ID No. 73	**Ad50-E4orf1**	MADEALYVYLDGPGATLPEQQQRNNYIFYSPVPFTLY PRGVALLYLRLSIIIPRGYVGCFFSLTDANMSGLYASS RIIHAGHREELSVLLFNHDDRFYEGRAGDPVACLVME RLIYPPVRQATMI
SEQ ID No. 74	**Ad55-E4orf1**	MAHEALYVYLEGPGATLPEQQQRNNYIFYSPVPFTLY PRGVALLYLRLSIIIPRGYIGCFLSLTDANMFGLYASSRI IHAGHREELSVLLFNHDDRFYEGRAGDPVACLVMERL IYPPVRQATLI
SEQ ID No. 75	**Ad62-E4orf1**	MAESLYAFIDSPGGIAPVQEGASNRYIFFCPESFHIPP HGVILLHLRVSVMVPTGYQGRFMALNDYHARGILTQS DVIFAGRRHDLSVLLFNHTDRFLYVREGHPVGTLLLE RVIFPSVRIATLV
SEQ ID No. 76	**Ad71-E4orf1**	MAESLYAFIDSPGGIAPVQEGTSNRYDFFCPESFHIPP HGVVLLHLRVSVLIPTGYQGRFMALNDYHARGILTQS DVIFAGRRHELTVLLFNHTDRFLYVREGHPVGTLLLER VIFPSVRLATLV

In a preferred embodiment, the at least one adenoviral protein is E4orf6/7 or a variant thereof.

An example of a nucleotide sequence encoding Ad5-E4orf6/7 is set forth in SEQ ID No. 11. Suitably, the at least one adenoviral protein may comprise a nucleotide sequence for mRNA translation as set forth in SEQ ID No. 11 or a variant thereof.

ATGACCACCAGCGGCGTGCCCTTCGGCATGACACTCAGACCTACCAGAAGCCGGCTGAGCAGAAGAACCCCTTAC
AGCAGAGACAGGCTGCCTCCATTCGAGACAGAGACACGGGCCACCATCCTGGAAGATCACCCTCTGCTGCCCGAG
TGTAACACCCTGACCATGCACAACGCCTGGACAAGCCCATCTCCTCCAGTGAAACAGCCCCAAGTGGGACAGCAG
CCTGTTGCTCAGCAGCTGGACAGCGACATGAACCTGTCTGAACTGCCCGGCGAGTTCATCAACATCACCGACGAG
AGACTGGCCCGGCAAGAGACAGTGTGGAACATCACCCCTAAGAACATGAGCGTGACCCACGACATGATGCTGTTC
AAGGCCAGCAGAGGCGAGCGGACAGTGTACAGCGTTTGTTGGGAAGGCGGCGGACGGCTGAATACCAGAGTGCTG
TAA (SEQ ID No. 11)

An example of an amino acid sequence of Ad5-E4orf6/7 is set forth in SEQ ID No. 2. Suitably, the at least one adenoviral protein may comprise an amino acid sequence as set forth in SEQ ID No. 2 or a variant thereof.

MTTSGVPFGMTLRPTRSRLSRRTPYSRDRLPPFETETRATILEDHPLLPECNTLTMHNAWTSPSPPVKQPQVGQQ
PVAQQLDSDMNLSELPGEFINITDERLARQETVWNITPKNMSVTHDMMLFKASRGERTVYSVCWEGGGRLNTRVL
(SEQ ID No. 2)

Other examples of an amino acid sequence of E4orf6/7 are set forth in SEQ ID No. 77 to SEQ ID No. 107. Suitably, the at least one adenoviral protein may comprise an amino acid sequence as set forth in SEQ ID No. 77 to SEQ ID No. 107 or a variant thereof.

SEQ ID No. 77	**Ad1-E4orf6/7**	MTTSGVPFGMTLRPTRSRLSRRTPYSRDRLPPFETE TRATILEDHPLLPECNTLTMHNAWTSPSPPVEQPQVG QQPVAQQLDSNMNLRELPGEFINITDERLARQETVW NITPKNMSVTHDMMLFKASRGERTVYSVCWEGGGR LNTRVL*
SEQ ID No. 78	**Ad2-E4orf6/7**	MTTSGVPFGMTLRPTRSRLSRRTPYSRDRLPPFETE TRATILEDHPLLPECNTLTMHNAWTSPSPPVEQPQVG QQPVAQQLDSDMNLSELPGEFINITDERLARQETVW NITPKNMSVTHDMMLFKASRGERTVYSVCWEGGGR LNTRVL*
SEQ ID No. 79	**Ad3-E4orf6/7**	MSGSNSIMTRLRARSTSCARHHPYTRAQLPRCEENE TRASMTEDHPLLPDCDTMTMHSMTVIQTPESHPQQL DCESALKDYPDAFLCITDPRLARSETVWNVESKTMSI SNGIQMFKAVRGERLVYSVKWEGGGKITTRIL*
SEQ ID No. 80	**Ad4-E4orf6/7**	MSGNSSIMTRSRTRLALSRHHPYQPPATLPRCEETES RASLVEDHPVLPDCDTLSMHNITVIPITEDSPQLLNCE VQMQECPEGFISLTDPRLSRSETVWNVEIKTMSITNSI QMFKAVRGERIVYSMRWEGGGKITTRIL*
SEQ ID No. 81	**Ad6-E4orf6/7**	MTTSGVPFGMTLRPTRSRLSRRTPYSRDRLPPFETE TRATILEDHPLLPECNTLTMHNAWTSPSPSVKRPQVG QQPVAQQLDSDMNLSELPGEFINITDERLARQETVW NITPKNMSVTHDMMLFKASRGERTVYSVCWEGGGR LNTRVL*

SEQ ID No. 82	**Ad7-E4orf6/7**	MSGSNSIMTRLRARSTSCARHHPYTRAQLPRCEENE TRASMTEDHPLLPDCDTMTMHSMTVIQTPESHPQQL DCESALKDYPDAFLCITDPRLARFETVWNVESKTMSI CNGIQMFKAVRGERLVYSVKWEGGGKITTRIL*
SEQ ID No. 83	**Ad11-E4orf6/7**	MSGSNSIMTRLRARSTSCARHHPYTRAQLPRCEENE TRASMTEDHPLLPDCDTMTMHSMTVIQTPESQPQQL DCESALKDYPDAFLSITDPRLARSETVWNVESKTMSI SNGIQMFKAVRGERLVYSVKWEGGGKITTRIL*
SEQ ID No. 84	**Ad14-E4orf6/7**	MSGSNSIMTRLCARSTSCARHHPYTRAQLPRCEENE TRASMTEDHPLLPDCDTMTMHSMTVIQTPESHPQQL DCESALKDYPAGFLSITDPRLARYETVWNVESKTMSI SNGIQMFKAVRGERLVYSVKWEGGGKITTRSL*
SEQ ID No. 85	**Ad18-E4orf6/7**	MQRNRRYPYRLAPYGKYPLPPCEKEMRASLFGPENS SVSECNSLTLHNVINMDLVLDGESYLSDCVGEGFVSII DHRFARKETIWTVTPKNLSRNMHMQLFSAIKGERVVY KIKWEGGGSLTTRIV*
SEQ ID No. 86	**Ad19-E4orf6/7**	MQTEDQSSLLRHHPYRRARLPRSDEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWIIDPKLSSRTNQNIPLFKATRAERTVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 87	**Ad20-E4orf6/7**	MQTEIQSSSLRHHPYRRARLPRSDEETRASLTEQHPL LPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRLA RQETVWIIDPKSSSRTNQNISLFKATRAERTVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 88	**Ad23-E4orf6/7**	MSTEEQSSSLRHHPYRRARLPRCEEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWIIDTKTSSRTNNNIPLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 89	**Ad25-E4orf6/7**	MSTEEQSSSLRHHPYRRARLPRCEEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWIIDPKSSSRTNQNISLFKATRAERTVYTVK WAGGGRLTTRAGVKINKDT*
SEQ ID No. 90	**Ad27-E4orf6/7**	MQTEDQSSLLRHHPYRRARLPRSDEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWIIDPKFSSRTNQNIPLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 91	**Ad29-E4orf6/7**	MQTEDQSSLLRHHPYRRARLPRSDEETRASLTEQHP LLPDCDHAEYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWLIDTKSSSRANQNIPLFKATRAERIVYTVK WAGGGRLTTRAGVKINKDT*
SEQ ID No. 92	**Ad30-E4orf6/7**	MSTEEQSSSLRHHPYRRARLPRCEEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWIIDSKSSSRTNQNIPLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDP*
SEQ ID No. 93	**Ad32-E4orf6/7**	MSTEEQSSSLRHHPYRRARLPRCDEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWIIDPKSSSRTNQNIFLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 94	**Ad34-E4orf6/7**	MSGSNSIMTRLRARSTSCARHHPYTRAQLPRCEENE TRASMTEDHPLLPDCDTMTMHSMTVIQTPESHPQQL DCESALKDYPDAFLSITDPRLARSETVWNVESKTMSI SNGIQMFKAVRGERLVYSVKWEGGGKITTRIL*
SEQ ID No. 95	**Ad36-E4orf6/7**	MSTEEQSTSLRHHPYRRARLPRSEEETRASLTEQHP LLPDCDHAEYHNTVTLDCEARLEDFSEDGFISITDPRL

		ARQETVWIIDTKSSSRTNQNIPLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 96	**Ad38-E4orf6/7**	MSTEEQSSLLRHHPYRRARLPRCEEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWIIDTKSSSRSNNNIPLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDS*
SEQ ID No. 97	**Ad39-E4orf6/7**	MSTEEQSSLLRHHPYRRARLPRCEEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLEDFSEDGFISITDPRL ARQETVWIIDTKSSSRTNQNIPLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 98	**Ad43-E4orf6/7**	MSTEEQSSLLRHHPYRRARLPRCEQETRASLTEQHP LLPDCDHAEYHNTVTLDCEPRLEDFSEDGFISITDPRL ARQETVWLIDTKTSSRTNQNIPLFKATRAERTVYTVK WAGGGRLTTRAGVKINKDT*
SEQ ID No. 99	**Ad44-E4orf6/7**	MQTEDQSSLLRHHPYRRARLPRCEEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLEDFSEDGFISITDPRL ARQETVWLIDTKSSSRTNQNIPLFKATRAERTVYTVK WAGGGRLTTRAGVKINKDT*
SEQ ID No. 100	**Ad45-E4orf6/7**	MSTEEQSSLLRHHPYRRARLPRCEEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRL ARQETVWIIDPKSSSRTNQNISLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 101	**Ad47-E4orf6/7**	MSTEEQSSLLRHHPYRRARLPRSDEETRASLTEQHP LLPDCDHADYHNTVTLDCEARLEDFSEDGFISITDPRL ARQETVWIIDPKSSSRTNQNIPLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 102	**Ad50-E4orf6/7**	MSGNNSIMTRLRARSTSCARHHPYTRAQLPRCEENE TRASMTEDHPLLPDCDTMTMHSMTVIQTPESHPQQL DCESALKDYPDAFLCITDPRLARSETVWNVETKTMSI SNGIQMFKAVRGERLVYSVKWEGGGKITTRIL*
SEQ ID No. 103	**Ad53-E4orf6/7**	MQTEIQSSSLRHHPYRRARLPRSDEETRASLTEQHPL LPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRLA RQETVWIINPKSSSRTNENFPLFKATRAERIVYTVKWA GGGRMCTRAGVKINKDT*
SEQ ID No. 104	**Ad54-E4orf6/7**	MSTEEQSTLLRHHPYRRARLSRYDKETRASLTEQHP LLPDCDHADYHNTVTLDCEARLDDLSKDGFISITDPRL ARQETVWIIDPKSSSRTNENFPLFKATRTERIVYTVKW AGGGRMCTRAGVKINKDT*
SEQ ID No. 105	**Ad58-E4orf6/7**	MQTEIQSSLLRHHPYRRARLPRSDEETRASLTEQHPL LPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRLA RQETVWLIDTKSSSRANQNIPLFKATRAERIVYTVKW AGGGRLTTRAGVKINKDT*
SEQ ID No. 106	**Ad64-E4orf6/7**	MQTEIQSSSLRHHPYRRARLPRSDEETRASLTEQHPL LPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRLA RQETVWIIDPKSNSRTNENISLFKATRAERTVYTVKW AGGGRLSTRAGVKINKDT*
SEQ ID No. 107	**Ad69-E4orf6/7**	MQTEIQSSSLRHHPYRRARLPRSDEETRASLTEQHPL LPDCDHADYHNTVTLDCEARLDDFSEDGFISITDPRLA RQETVWIIDPKSSSRTNQNISLFKATRAERIVYTVKWA GGGRLTTRAGVKINKDT*

Variant sequences of SEQ ID NOs recited herein may have at least 80%, 85%, 90%, 95%, 98% or 99% sequence identity to the reference sequence SEQ ID NOs. Preferably, the

variant sequence retains one or more functions of the reference sequence (i.e. is a functional variant).

Variant sequences may comprise one or more conservative substitutions. Conservative amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine.

Conservative substitutions may be made, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

The present invention also encompasses homologous substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue, with an alternative residue) variants i.e. like-for-like substitution such as basic for basic, acidic for acidic, polar for polar etc.

Unless otherwise explicitly stated herein by way of reference to a specific, individual amino acid, amino acids may be substituted using conservative substitutions as recited below.

An aliphatic, non-polar amino acid may be a glycine, alanine, proline, isoleucine, leucine or valine residue.

An aliphatic, polar uncharged amino may be a cysteine, serine, threonine, methionine, asparagine or glutamine residue.

An aliphatic, polar charged amino acid may be an aspartic acid, glutamic acid, lysine or arginine residue.

An aromatic amino acid may be a histidine, phenylalanine, tryptophan or tyrosine residue.

Suitably, a conservative substitution may be made between amino acids in the same line in the Table above.

An example of a nucleotide sequence encoding E4orf6 is set forth in SEQ ID No. 12.

ATGACAACCAGCGGCGTGCCCTTCGGCATGACACTGAGGCCTACAAGAAGCCGGCTGAGCAGAAGAACCCCTTAC
AGCAGAGACAGACTGCCTCCATTCGAGACAGAGACACGGGCCACCATCCTGGAAGATCACCCTCTGCTGCCCGAG
TGCAACACCCTGACAATGCACAACGTGTCCTACGTGCGGGGCCTGCCTTGTAGCGTTGGCTTCACACTGATCCAA
GAGTGGGTCGTGCCCTGGGACATGGTGCTGACCAGAGAGGAACTGGTCATCCTGCGGAAGTGTATGCACGTGTGC
CTGTGCTGCGCCAACATCGACATCATGACCAGCATGATGATCCACGGCTACGAGAGCTGGGCCCTGCACTGTCAC
TGTTCTAGCCCTGGCAGCCTGCAGTGTATTGCTGGTGGACAGGTTCTGGCCAGCTGGTTCAGAATGGTGGTGGAC
GGCGCCATGTTCAACCAGAGATTCATCTGGTACAGAGAGGTGGTCAACTACAACATGCCCAAAGAAGTGATGTTC
ATGAGCAGCGTTTTCATGCGGGGCAGACACCTGATCTACCTGCGGCTTTGGTACGATGGCCACGTGGGATCTGTG
GTGCCTGCCATGAGCTTTGGCTACAGCGCCCTGCATTGCGGCATCCTGAACAACATCGTGGTGCTGTGCTGCAGC
TACTGCGCCGATCTGAGCGAGATCAGAGTGCGGTGTTGTGCCAGACGGACCAGACGGCTGATGCTGAGAGCCGTG
CGGATCATTGCCGAAGAGACAACCGCCATGCTGTACTCTTGCCGGACCGAGAGAAGAAGGCAGCAGTTCATCAGA
GCCCTGCTCCAGCACCACCGGCCTATCCTGATGCACGACTACGACAGCACCCCTATGTAG (SEQ ID No. 12)

An example of an amino acid sequence of E4orf6 is set forth in SEQ ID No. 13.

MTTSGVPFGMTLRPTRSRLSRRTPYSRDRLPPFETETRATILEDHPLLPECNTLTMHNVSYVRGLPC
SVGFTLIQEWVVPWDMVLTREELVILRKCMHVCLCCANIDIMTSMMIHGYESWALHCHCSSPGSLQCI
AGGQVLASWFRMVVDGAMFNQRFIWYREVVNYNMPKEVMFMSSVFMRGRHLIYLRLWYDGHVGS
VVPAMSFGYSALHCGILNNIVVLCCSYCADLSEIRVRCCARRTRRLMLRAVRIIAEETTAMLYSCRTER
RRQQFIRALLQHHRPILMHDYDSTPM (SEQ ID No. 13)

An example of a nucleotide sequence encoding E1B55K is set forth in SEQ ID No. 14.

ATGGAAAGACGGAACCCCAGCGAGAGAGGCGTGCCAGCTGGATTTTCTGGACACGCCAGCGTGGAAAGCGGCTGC
GAGACACAAGAAAGCCCTGCCACCGTGGTGTTCAGACCACCTGGCGATAATACCGATGGCGGAGCTGCTGCTGCA
GCTGGTGGATCTCAGGCAGCAGCTGCAGGCGCTGAACCTATGGAACCTGAGAGCAGACCTGGACCTAGCGGCATG
AATGTGGTGCAGGTCGCCGAGCTGTATCCCGAGCTGAGAAGAATCCTGACCATCACCGAGGATGGCCAGGGACTG
AAGGGCGTGAAGAGAGAAAGAGGCGCCTGCGAGGCCACAGAGGAAGCCAGAAATCTGGCCTTCAGCCTGATGACC
AGACACAGACCCGAGTGCATCACCTTCCAGCAGATCAAGGACAACTGCGCCAACGAGCTGGACCTGCTGGCCCAG
AAGTACAGCATCGAGCAGCTGACCACCTACTGGCTGCAACCCGGCGACGATTTCGAAGAGGCCATCAGAGTGTAC
GCCAAGGTGGCCCTCAGACCTGACTGCAAGTACAAGATCAGCAAGCTGGTCAACATCCGGAACTGCTGCTACATC
AGCGGCAATGGCGCCGAGGTGGAAATCGACACAGAGGACAGAGTGGCCTTCCGGTGCAGCATGATCAACATGTGG
CCTGGCGTGCTCGGCATGGATGGCGTGGTCATTATGAACGTGCGGTTCACAGGCCCCAACTTCAGCGGCACAGTG
TTTCTGGCCAACACCAACCTGATCCTGCACGGCGTGTCCTTCTACGGCTTCAACAATACCTGCGTGGAAGCCTGG
ACCGACGTTCGCGTTAGAGGCTGCGCCTTCTACTGCTGTTGGAAGGGCGTCGTGTGCAGACCCAAGAGCAGAGCC
AGCATCAAGAAGTGCCTGTTCGAGAGATGCACCCTGGGCATCCTGAGCGAGGGCAACAGCAGAGTCAGACACAAC
GTGGCCAGCGACTGCGGCTGCTTCATGCTGGTTAAGAGCGTGGCCGTGATCAAGCACAACATGGTCTGCGGCAAC
TGCGAGGATAGAGCCAGCCAGATGCTGACCTGCAGCGACGGCAATTGTCATCTGCTGAAAACCATCCACGTGGCC
TCTCACAGCAGAAAGGCCTGGCCTGTGTTCGAGCACAATATCCTGACACGGTGCTCCCTGCACCTGGGCAATAGA
CGGGGAGTGTTCCTGCCTTACCAGTGCAACCTGAGCCACACCAAGATCCTGCTGGAACCCGAGTCCATGAGCAAA
GTGAACCTGAATGGCGTGTTCGACATGACCATGAAGATCTGGAAAGTGCTGCGCTACGACGAGACACGGACCAGA
TGTAGACCTTGCGAGTGTGGCGGCAAGCACATCAGAAACCAGCCTGTGATGCTGGACGTGACCGAGGAACTGAGG
CCTGATCATCTGGTGCTGGCCTGTACCAGAGCCGAGTTTGGCAGCTCCGACGAGGATACCGAT (SEQ ID No. 14)

An example of an amino acid sequence of E1B55K is set forth in SEQ ID No. 15.

MERRNPSERGVPAGFSGHASVESGCETQESPATVVFRPPGDNTDGGAAAAAGGSQAAAAGAEPMEPESRPGPSGM
NVVQVAELYPELRRILTITEDGQGLKGVKRERGACEATEEARNLAFSLMTRHRPECITFQQIKDNCANELDLLAQ
KYSIEQLTTYWLQPGDDFEEAIRVYAKVALRPDCKYKISKLVNIRNCCYISGNGAEVEIDTEDRVAFRCSMINMW
PGVLGMDGVVIMNVRFTGPNFSGTVFLANTNLILHGVSFYGFNNTCVEAWTDVRVRGCAFYCCWKGVVCRPKSRA
SIKKCLFERCTLGILSEGNSRVRHNVASDCGCFMLVKSVAVIKHNMVCGNCEDRASQMLTCSDGNCHLLKTIHVA
SHSRKAWPVFEHNILTRCSLHLGNRRGVFLPYQCNLSHTKILLEPESMSKVNLNGVFDMTMKIWKVLRYDETRTR
CRPCECGGKHIRNQPVMLDVTEELRPDHLVLACTRAEFGSSDEDTD (SEQ ID No. 15)

In one embodiment, the at least one adenoviral protein is not E4orf6. In one embodiment, the at least one adenoviral protein is not E1B55K. In one embodiment, the at least one adenoviral proteins does not comprise E4orf6 or E1B55K.

Haematopoietic stem and progenitor cells

A stem cell is able to differentiate into many cell types. A cell that is able to differentiate into all cell types is known as totipotent. In mammals, only the zygote and early embryonic cells are totipotent. Stem cells are found in most, if not all, multicellular organisms. They are characterised by the ability to renew themselves through mitotic cell division and differentiate into a diverse range of specialised cell types. The two broad types of mammalian stem cells are embryonic stem cells that are isolated from the inner cell mass of blastocysts, and adult stem cells that are found in adult tissues. In a developing embryo, stem cells can differentiate into all of the specialised embryonic tissues. In adult organisms, stem cells and progenitor cells act as a repair system for the body, replenishing specialised cells, but also maintaining the normal turnover of regenerative organs, such as blood, skin or intestinal tissues.

Haematopoietic stem cells (HSCs) are multipotent stem cells that may be found, for example, in peripheral blood, bone marrow and umbilical cord blood. HSCs are capable of self-renewal and differentiation into any blood cell lineage. They are capable of recolonising the entire immune system, and the erythroid and myeloid lineages in all the haematopoietic tissues (such as bone marrow, spleen and thymus). They provide for life-long production of all lineages of haematopoietic cells.

Haematopoietic progenitor cells have the capacity to differentiate into a specific type of cell. In contrast to stem cells however, they are already far more specific: they are pushed to differentiate into their “target” cell. A difference between stem cells and progenitor cells is that stem cells can replicate indefinitely, whereas progenitor cells can only divide a limited number of times. Haematopoietic progenitor cells can be rigorously distinguished from HSCs only by functional in vivo assay (i.e. transplantation and demonstration of whether they can give rise to all blood lineages over prolonged time periods).

The haematopoietic stem and progenitor cells of the invention comprise the CD34 cell surface marker (denoted as CD34+).

In one embodiment, the cells for use in the present invention are HSPCs.

In one embodiment, the cells for use in the present invention are primitive HSPCs. In one embodiment, a primitive subset of HSPCs refers to a population of HSCs which is CD90+. In one embodiment, a primitive subset of HSPCs refers to a population of cells which is CD34+ CD133+ and CD90+.

In one embodiment, the cells for use in the present invention are HSCs.

Haematopoietic stem and progenitor cell (HSPC) sources

A population of haematopoietic stem and/or progenitor cells may be obtained from a tissue sample.

For example, a population of haematopoietic stem and/or progenitor cells may be obtained from peripheral blood (e.g. adult and foetal peripheral blood), umbilical cord blood, bone marrow, liver or spleen. Preferably, these cells are obtained from peripheral blood or bone marrow. They may be obtained after mobilisation of the cells in vivo by means of growth factor treatment.

Mobilisation may be carried out using, for example, G-CSF, plerixaphor or combinations thereof. Other agents, such as NSAIDs and dipeptidyl peptidase inhibitors, may also be useful as mobilising agents.

With the availability of the stem cell growth factors GM-CSF and G-CSF, most haematopoietic stem cell transplantation procedures are now performed using stem cells collected from the peripheral blood, rather than from the bone marrow. Collecting peripheral blood stem cells provides a bigger graft, does not require that the donor be subjected to general anaesthesia to collect the graft, results in a shorter time to engraftment and may provide for a lower long-term relapse rate.

Bone marrow may be collected by standard aspiration methods (either steady-state or after mobilisation), or by using next-generation harvesting tools (e.g. Marrow Miner).

In addition, haematopoietic stem and progenitor cells may also be derived from induced pluripotent stem cells.

HSC characteristics

HSCs are typically of low forward scatter and side scatter profile by flow cytometric procedures. Some are metabolically quiescent, as demonstrated by Rhodamine labelling which allows determination of mitochondrial activity. HSCs may comprise certain cell surface markers such as CD34, CD45, CD133, CD90 and CD49f. They may also be defined as cells lacking the expression of the CD38 and CD45RA cell surface markers. However, expression of some of these markers is dependent upon the developmental stage and tissue-specific context of the HSC. Some HSCs called “side population cells” exclude the Hoechst 33342 dye as detected by flow cytometry. Thus, HSCs have descriptive characteristics that allow for their identification and isolation.

Negative markers

CD38 is the most established and useful single negative marker for human HSCs.

Human HSCs may also be negative for lineage markers such as CD2, CD3, CD14, CD16, CD19, CD20, CD24, CD36, CD56, CD66b, CD271 and CD45RA. However, these markers may need to be used in combination for HSC enrichment.

By “negative marker” it is to be understood that human HSCs lack the expression of these markers.

Positive markers

CD34 and CD133 are the most useful positive markers for HSCs.

Some HSCs are also positive for lineage markers such as CD90, CD49f and CD93. However, these markers may need to be used in combination for HSC enrichment.

By “positive marker” it is to be understood that human HSCs express these markers.

In one embodiment, the haematopoietic stem and progenitor cells are CD34+CD38- cells.

Differentiated cells

A differentiated cell is a cell which has become more specialised in comparison to a stem cell or progenitor cell. Differentiation occurs during the development of a multicellular organism as the organism changes from a single zygote to a complex system of tissues and cell types. Differentiation is also a common process in adults: adult stem cells divide and create fully-differentiated daughter cells during tissue repair and normal cell turnover. Differentiation dramatically changes a cell’s size, shape, membrane potential, metabolic activity and responsiveness to signals. These changes are largely due to highly-controlled

modifications in gene expression. In other words, a differentiated cell is a cell which has specific structures and performs certain functions due to a developmental process which involves the activation and deactivation of specific genes. Here, a differentiated cell includes differentiated cells of the haematopoietic lineage such as monocytes, macrophages, neutrophils, basophils, eosinophils, erythrocytes, megakaryocytes/platelets, dendritic cells, T cells, B-cells and NK-cells. For example, differentiated cells of the haematopoietic lineage can be distinguished from stem cells and progenitor cells by detection of cell surface molecules which are not expressed or are expressed to a lesser degree on undifferentiated cells. Examples of suitable human lineage markers include CD33, CD13, CD14, CD15 (myeloid), CD19, CD20, CD22, CD79a (B), CD36, CD71, CD235a (erythroid), CD2, CD3, CD4, CD8 (T) and CD56 (NK).

In one embodiment, the haematopoietic cells referred to herein is a T-cell.

Isolation and enrichment of populations of cells

The term “isolated population” of cells as used herein may refer to the population of cells having been previously removed from the body. An isolated population of cells may be cultured and manipulated ex vivo or in vitro using standard techniques known in the art. An isolated population of cells may later be reintroduced into a subject. Said subject may be the same subject from which the cells were originally isolated or a different subject.

A population of cells may be purified selectively for cells that exhibit a specific phenotype or characteristic, and from other cells which do not exhibit that phenotype or characteristic, or exhibit it to a lesser degree. For example, a population of cells that expresses a specific marker (such as CD34) may be purified from a starting population of cells. Alternatively, or in addition, a population of cells that does not express another marker (such as CD38) may be purified.

By “enriching” a population of cells for a certain type of cells it is to be understood that the concentration of that type of cells is increased within the population. The concentration of other types of cells may be concomitantly reduced.

Purification or enrichment may result in the population of cells being substantially pure of other types of cell.

Purifying or enriching for a population of cells expressing a specific marker (e.g. CD34 or CD38) may be achieved by using an agent that binds to that marker, preferably substantially specifically to that marker.

An agent that binds to a cellular marker may be an antibody, for example an anti-CD34 or anti-CD38 antibody.

The term “antibody” refers to complete antibodies or antibody fragments capable of binding to a selected target, and including Fv, ScFv, F(ab′) and $F(ab')_2$, monoclonal and polyclonal antibodies, engineered antibodies including chimeric, CDR-grafted and humanised antibodies, and artificially selected antibodies produced using phage display or alternative techniques.

In addition, alternatives to classical antibodies may also be used in the invention, for example “avibodies”, “avimers”, “anticalins”, “nanobodies” and “DARPins”.

The agents that bind to specific markers may be labelled so as to be identifiable using any of a number of techniques known in the art. The agent may be inherently labelled, or may be modified by conjugating a label thereto. By “conjugating” it is to be understood that the agent and label are operably linked. This means that the agent and label are linked together in a manner which enables both to carry out their function (e.g. binding to a marker, allowing fluorescent identification or allowing separation when placed in a magnetic field) substantially unhindered. Suitable methods of conjugation are well known in the art and would be readily identifiable by the skilled person.

A label may allow, for example, the labelled agent and any cell to which it is bound to be purified from its environment (e.g. the agent may be labelled with a magnetic bead or an affinity tag, such as avidin), detected or both. Detectable markers suitable for use as a label include fluorophores (e.g. green, cherry, cyan and orange fluorescent proteins) and peptide tags (e.g. His tags, Myc tags, FLAG tags and HA tags).

A number of techniques for separating a population of cells expressing a specific marker are known in the art. These include magnetic bead-based separation technologies (e.g. closed-circuit magnetic bead-based separation), flow cytometry, fluorescence-activated cell sorting (FACS), affinity tag purification (e.g. using affinity columns or beads, such biotin columns to separate avidin-labelled agents) and microscopy-based techniques.

It may also be possible to perform the separation using a combination of different techniques, such as a magnetic bead-based separation step followed by sorting of the resulting population of cells for one or more additional (positive or negative) markers by flow cytometry.

Clinical grade separation may be performed, for example, using the CliniMACS® system (Miltenyi). This is an example of a closed-circuit magnetic bead-based separation technology.

It is also envisaged that dye exclusion properties (e.g. side population or rhodamine labelling) or enzymatic activity (e.g. ALDH activity) may be used to enrich for haematopoietic stem cells.

Suitably, the agent does not reduce the fraction CD34+ CD133+ CD90+ cells in population of gene edited cells compared with a population of untreated gene edited cells.

Gene editing

The term "gene editing" refers to a type of genetic engineering in which a nucleic acid is inserted, deleted or replaced in a cell. Gene editing may be achieved using engineered nucleases, which may be targeted to a desired site in a polynucleotide (e.g. a genome). Such nucleases may create site-specific double-strand breaks at desired locations, which may then be repaired through non-homologous end-joining (NHEJ) or homologous recombination (HR), resulting in targeted mutations.

Such nucleases may be delivered to a target cell using viral vectors. The present invention provides methods of increasing the efficiency of the gene editing process.

Examples of suitable nucleases known in the art include zinc finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs), and the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system (Gaj, T. et al. (2013) Trends Biotechnol. 31: 397-405; Sander, J.D. et al. (2014) Nat. Biotechnol. 32: 347-55).

Meganucleases (Silve, G. et al. (2011) Cur. Gene Ther. 11: 11-27) may also be employed as suitable nucleases for gene editing.

The CRISPR/Cas system is an RNA-guided DNA binding system (van der Oost et al. (2014) Nat. Rev. Microbiol. 12: 479-92), wherein the guide RNA (gRNA) may be selected to enable a Cas9 domain to be targeted to a specific sequence. Methods for the design of gRNAs are known in the art. Furthermore, fully orthogonal Cas9 proteins, as well as Cas9/gRNA ribonucleoprotein complexes and modifications of the gRNA structure/composition to bind different proteins, have been recently developed to simultaneously and directionally target different effector domains to desired genomic sites of the cells (Esvelt et al. (2013) Nat. Methods 10: 1116-21; Zetsche, B. et al. (2015) Cell pii: S0092-8674(15)01200-3; Dahlman, J.E. et al. (2015) Nat. Biotechnol. 2015 Oct 5. doi: 10.1038/nbt.3390. [Epub ahead of print];

Zalatan, J.G. et al. (2015) Cell 160: 339-50; Paix, A. et al. (2015) Genetics 201: 47-54), and are suitable for use in the invention.

Vectors

A vector is a tool that allows or facilitates the transfer of an entity from one environment to another. The vectors used to transduce haematopoietic stem and/or progenitor cells in the invention are viral vectors.

Preferably, the viral vector is in the form of a viral vector particle.

The viral vector may be, for example, an adeno-associated viral (AAV), adenoviral, a retroviral or lentiviral vector. Preferably, the viral vector is an AAV vector or a retroviral or lentiviral vector, more preferably an AAV vector. Preferably, the retroviral vector is not a γ-retroviral vector.

By "vector derived from" a certain type of virus, it is to be understood that the vector comprises at least one component part derivable from that type of virus.

Adeno-associated viral (AAV) vectors

Adeno-associated virus (AAV) is an attractive vector system for use in the invention as it has a high frequency of integration and it can infect non-dividing cells. This makes it useful for delivery of genes into mammalian cells in tissue culture.

AAV has a broad host range for infectivity. Details concerning the generation and use of AAV vectors are described in US Patent No. 5139941 and US Patent No. 4797368.

Recombinant AAV vectors have been used successfully for in vitro and in vivo transduction of marker genes and genes involved in human diseases.

Preferred vectors are those which are able to achieve a high transduction efficiency in human primary cells, such as HSPC cells.

In one embodiment, the vector is an AAV6 vector or a vector derived from an AAV6 vector. Preferably the vector is an AAV6 vector.

Adenoviral vectors

The adenovirus is a double-stranded, linear DNA virus that does not go through an RNA intermediate. There are over 50 different human serotypes of adenovirus divided into 6 subgroups based on the genetic sequence homology. The natural targets of adenovirus are the respiratory and gastrointestinal epithelia, generally giving rise to only mild symptoms.

Serotypes 2 and 5 (with 95% sequence homology) are most commonly used in adenoviral vector systems and are normally associated with upper respiratory tract infections in the young.

Adenoviruses have been used as vectors for gene therapy and for expression of heterologous genes. The large (36 kb) genome can accommodate up to 8 kb of foreign insert DNA and is able to replicate efficiently in complementing cell lines to produce very high titres of up to 10^{12}. Adenovirus is thus one of the best systems to study the expression of genes in primary non-replicative cells.

The expression of viral or foreign genes from the adenovirus genome does not require a replicating cell. Adenoviral vectors enter cells by receptor mediated endocytosis. Once inside the cell, adenovirus vectors rarely integrate into the host chromosome. Instead, they function episomally (independently from the host genome) as a linear genome in the host nucleus. Hence the use of recombinant adenovirus alleviates the problems associated with random integration into the host genome.

Retroviral and lentiviral vectors

A retroviral vector may be derived from or may be derivable from any suitable retrovirus. A large number of different retroviruses have been identified. Examples include murine leukaemia virus (MLV), human T-cell leukaemia virus (HTLV), mouse mammary tumour virus (MMTV), Rous sarcoma virus (RSV), Fujinami sarcoma virus (FuSV), Moloney murine leukaemia virus (Mo-MLV), FBR murine osteosarcoma virus (FBR MSV), Moloney murine sarcoma virus (Mo-MSV), Abelson murine leukaemia virus (A-MLV), avian myelocytomatosis virus-29 (MC29) and avian erythroblastosis virus (AEV). A detailed list of retroviruses may be found in Coffin, J.M. et al. (1997) Retroviruses, Cold Spring Harbour Laboratory Press, 758-63.

Retroviruses may be broadly divided into two categories, “simple” and “complex”. Retroviruses may be even further divided into seven groups. Five of these groups represent retroviruses with oncogenic potential. The remaining two groups are the lentiviruses and the spumaviruses. A review of these retroviruses is presented in Coffin, J.M. et al. (1997) Retroviruses, Cold Spring Harbour Laboratory Press, 758-63.

The basic structure of retrovirus and lentivirus genomes share many common features such as a 5’ LTR and a 3’ LTR. Between or within these are located a packaging signal to enable the genome to be packaged, a primer binding site, integration sites to enable integration into a host cell genome, and gag, pol and env genes encoding the packaging components –

these are polypeptides required for the assembly of viral particles. Lentiviruses have additional features, such as rev and RRE sequences in HIV, which enable the efficient export of RNA transcripts of the integrated provirus from the nucleus to the cytoplasm of an infected target cell.

In the provirus, these genes are flanked at both ends by regions called long terminal repeats (LTRs). The LTRs are responsible for proviral integration and transcription. LTRs also serve as enhancer-promoter sequences and can control the expression of the viral genes.

The LTRs themselves are identical sequences that can be divided into three elements: U3, R and U5. U3 is derived from the sequence unique to the 3' end of the RNA. R is derived from a sequence repeated at both ends of the RNA. U5 is derived from the sequence unique to the 5' end of the RNA. The sizes of the three elements can vary considerably among different retroviruses.

In a defective retroviral vector genome gag, pol and env may be absent or not functional.

In a typical retroviral vector, at least part of one or more protein coding regions essential for replication may be removed from the virus. This makes the viral vector replication-defective. Portions of the viral genome may also be replaced by a library encoding candidate modulating moieties operably linked to a regulatory control region and a reporter moiety in the vector genome in order to generate a vector comprising candidate modulating moieties which is capable of transducing a target host cell and/or integrating its genome into a host genome.

Lentivirus vectors are part of the larger group of retroviral vectors. A detailed list of lentiviruses may be found in Coffin, J.M. et al. (1997) Retroviruses, Cold Spring Harbour Laboratory Press, 758-63. In brief, lentiviruses can be divided into primate and non-primate groups. Examples of primate lentiviruses include but are not limited to human immunodeficiency virus (HIV), the causative agent of human acquired immunodeficiency syndrome (AIDS); and simian immunodeficiency virus (SIV). Examples of non-primate lentiviruses include the prototype "slow virus" visna/maedi virus (VMV), as well as the related caprine arthritis-encephalitis virus (CAEV), equine infectious anaemia virus (EIAV), and the more recently described feline immunodeficiency virus (FIV) and bovine immunodeficiency virus (BIV).

The lentivirus family differs from retroviruses in that lentiviruses have the capability to infect both dividing and non-dividing cells (Lewis, P et al. (1992) EMBO J. 11: 3053-8; Lewis, P.F. et al. (1994) J. Virol. 68: 510-6). In contrast, other retroviruses, such as MLV, are unable to

infect non-dividing or slowly dividing cells such as those that make up, for example, muscle, brain, lung and liver tissue.

A lentiviral vector, as used herein, is a vector which comprises at least one component part derivable from a lentivirus. Preferably, that component part is involved in the biological mechanisms by which the vector infects cells, expresses genes or is replicated.

The lentiviral vector may be a “primate” vector. The lentiviral vector may be a “non-primate” vector (i.e. derived from a virus which does not primarily infect primates, especially humans). Examples of non-primate lentiviruses may be any member of the family of lentiviridae which does not naturally infect a primate.

As examples of lentivirus-based vectors, HIV-1- and HIV-2-based vectors are described below.

The HIV-1 vector contains cis-acting elements that are also found in simple retroviruses. It has been shown that sequences that extend into the gag open reading frame are important for packaging of HIV-1. Therefore, HIV-1 vectors often contain the relevant portion of gag in which the translational initiation codon has been mutated. In addition, most HIV-1 vectors also contain a portion of the env gene that includes the RRE. Rev binds to RRE, which permits the transport of full-length or singly spliced mRNAs from the nucleus to the cytoplasm. In the absence of Rev and/or RRE, full-length HIV-1 RNAs accumulate in the nucleus. Alternatively, a constitutive transport element from certain simple retroviruses such as Mason-Pfizer monkey virus can be used to relieve the requirement for Rev and RRE. Efficient transcription from the HIV-1 LTR promoter requires the viral protein Tat.

Most HIV-2-based vectors are structurally very similar to HIV-1 vectors. Similar to HIV-1-based vectors, HIV-2 vectors also require RRE for efficient transport of the full-length or singly spliced viral RNAs.

In one system, the vector and helper constructs are from two different viruses, and the reduced nucleotide homology may decrease the probability of recombination. In addition to vectors based on the primate lentiviruses, vectors based on FIV have also been developed as an alternative to vectors derived from the pathogenic HIV-1 genome. The structures of these vectors are also similar to the HIV-1 based vectors.

Preferably, the viral vector used in the present invention has a minimal viral genome.

By “minimal viral genome” it is to be understood that the viral vector has been manipulated so as to remove the non-essential elements and to retain the essential elements in order to

provide the required functionality to infect, transduce and deliver a nucleotide sequence of interest to a target host cell. Further details of this strategy can be found in WO 1998/017815.

Preferably, the plasmid vector used to produce the viral genome within a host cell/packaging cell will have sufficient lentiviral genetic information to allow packaging of an RNA genome, in the presence of packaging components, into a viral particle which is capable of infecting a target cell, but is incapable of independent replication to produce infectious viral particles within the final target cell. Preferably, the vector lacks a functional gag-pol and/or env gene and/or other genes essential for replication.

However, the plasmid vector used to produce the viral genome within a host cell/packaging cell will also include transcriptional regulatory control sequences operably linked to the lentiviral genome to direct transcription of the genome in a host cell/packaging cell. These regulatory sequences may be the natural sequences associated with the transcribed viral sequence (i.e. the 5' U3 region), or they may be a heterologous promoter, such as another viral promoter (e.g. the CMV promoter).

The vectors may be self-inactivating (SIN) vectors in which the viral enhancer and promoter sequences have been deleted. SIN vectors can be generated and transduce non-dividing cells in vivo with an efficacy similar to that of wild-type vectors. The transcriptional inactivation of the long terminal repeat (LTR) in the SIN provirus should prevent mobilisation by replication-competent virus. This should also enable the regulated expression of genes from internal promoters by eliminating any cis-acting effects of the LTR.

The vectors may be integration-defective. Integration defective lentiviral vectors (IDLVs) can be produced, for example, either by packaging the vector with catalytically inactive integrase (such as an HIV integrase bearing the D64V mutation in the catalytic site; Naldini, L. et al. (1996) Science 272: 263-7; Naldini, L. et al. (1996) Proc. Natl. Acad. Sci. USA 93: 11382-8; Leavitt, A.D. et al. (1996) J. Virol. 70: 721-8) or by modifying or deleting essential att sequences from the vector LTR (Nightingale, S.J. et al. (2006) Mol. Ther. 13: 1121-32), or by a combination of the above.

In one embodiment, the gene editing targets a haematopoietic stem cell and/or progenitor locus. In one embodiment, the donor template targets a haematopoietic stem cell and/or progenitor locus.

In one embodiment, the gene editing targets the adeno-associated virus integration site 1 (AAVS1) locus. In one embodiment, the donor template targets the adeno-associated virus integration site 1 (AAVS1) locus.

An example of an AAV donor cassette for AAVS1 may comprise the following nucleotide sequences:

3'-ITR

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccgggcgtcgggcgacctttggtc
gcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcact (SEQ ID No. 16)

Homologous arm-left

tgctttctctgaccagcattctctcccctgggcctgtgccgctttctgtctgcagcttgtggcctgggtcacctc
tacggctggcccagatccttccctgccgcctccttcaggttccgtcttcctccactccctcttccccttgctctc
tgctgtgttgctgcccaaggatgctctttccggagcacttccttctcggcgctgcaccacgtgatgtcctctgag
cggatcctccccgtgtctgggtcctctccgggcatctctcctccctcacccaaccccatgccgtcttcactcgct
gggttccctttttccttctccttctggggcctgtgccatctctcgtttcttaggatggccttctccgacggatgtc
tcccttgcgtcccgcctcccccttcttgtaggcctgcatcatcaccgtttttctggacaaccccaaagtaccccgt
ctccctggctttagccacctctccatcctcttgctttctttgcctggacaccccgttctcctgtggattcgggtc
acctctcactcctttcatttgggcagctcccctaccccccttacctctctagtctgtgctagctcttccagcccc
ctgtcatggcatcttccaggggtccgagagctcagctagtcttcttcctccaacccgggcccctatgtccacttc
aggacagcatgtttgctgcctccagggatcctgtgtccccgagctgggaccaccttatattcccagggccggtta
atgtggctctggttctgggtacttttatctgtcccctccaccccac (SEQ ID No. 17)

hPGK promoter (without SA)

ccacggggttggggttgcgccttttccaaggcagccctgggtttgcgcagggacgcggctgctctgggcgtggtt
ccgggaaacgcagcggcgccgaccctgggtctcgcacattcttcacgtccgttcgcagcgtcacccggatcttcg
ccgctacccttgtgggcccccccggcgacgcttcctgctccgcccctaagtcgggaaggttccttgcggttcgcgg
cgtgccggacgtgacaaacggaagccgcacgtctcactagtaccctcgcagacggacagcgccagggagcaatgg
cagcgcgccgaccgcgatgggctgtggccaatagcggctgctcagcagggcgcgccgagagcagcggccgggaag
gggcggtgcgggaggcggggtgtggggcggtagtgtgggccctgttcctgcccgcgcggtgttccgcattctgca
agcctccggagcgcacgtcggcagtcggctccctcgttgaccgaatcaccgacctctctcccca (SEQ ID No. 18)

eGFP

atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggc
cacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcacc
accggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctac
cccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttc
ttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgag
ctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaac
gtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggc
agcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccac
tacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtg
accgccgccgggatcactctcggcatggacgagctgtacaagtaa (SEQ ID No. 19)

bGH-pA

actgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccact
cccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggt
ggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatg
g (SEQ ID No. 20)

Homologous arm-right

tagggacaggattggtgacagaaaagccccatccttaggcctcctccttcctagtctcctgatattgggtctaac
ccccacctcctgttaggcagattccttatctggtgacacacccccatttcctggagccatctctctccttgccag
aacctctaaggtttgcttacgatggagccagagaggatcctgggagggagagcttggcaggggtgggagggaag
gggggatgcgtgacctgcccggttctcagtgg (SEQ ID No. 21)

5'-ITR

agtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacg
cccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcagg (SEQ ID No. 22)

In another embodiment, the gene editing targets Interleukin 2 Receptor Subunit Gamma (IL2RG), preferably targeting intron 1 of IL2RG. In another embodiment, the donor template targets Interleukin 2 Receptor Subunit Gamma (IL2RG), preferably targeting intron 1 of IL2RG.

An example of an AAV donor cassette for IL2RG may comprise the following nucleotide sequences:

3'-ITR

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcaaagcccgggcgtcgggcgacctttggtc
gcccggcctcagtgagcgagcgagcgcgcagagagggagtggccaactccatcact (SEQ ID No. 23)

Homologous arm-left

agaggaaacgtgtgggtggggaggggtagtgggtgagggacccaggttcctgacacagacagactacacccaggg
aatgaagagcaagcgccatgttgaagccatcattaccattcacatccctcttattcctgcagctgcccctgctgg
gagtggggctgaacacgacaattctgacgcccaatgggaatgaagacaccacagctggtgggaaatctgggactg
gaggggctggtgagaagggtggctgtgggaaggggccgtacagagatctggtgcctgccactgg (SEQ ID No. 24)

Exon2

atttctttctgaccaccatgcccaccgacagcctgagcgtgagcaccctgcccctgcccgaggtgcagtgcttcg
tgttcaacgtggagtacatgaactgcacctggaacagcagcagcgagccccagcccaccaatctgaccctgcact
actg (SEQ ID No. 25)

Exon3

gtacaagaacagcgacaacgacaaggtgcagaagtgcagccactacctgttcagcgaggaaatcaccagcggctg
ccagctgcagaagaaagagatccacctgtaccagaccttcgtggtgcagctgcaggacccccgggagccccgcag
gcaggccacccagatgctgaagctgcagaacctgg (SEQ ID No. 26)

Exon4

tgatcccctgggcccctgagaacctgacactgcacaagctgtccgagagccagctggaactgaactggaacaacc
gcttcctgaaccactgcctggaacacctggtgcagtaccggaccgactgggaccacagctggacc (SEQ ID No. 27)

Exon5

gagcagagcgtggactaccggcacaagttcagcctgcccagcgtggacggccagaagcggtacaccttcagagtg
cggagccggttcaacccccctgtgcggcagcgcccagcactggtccgagtggagccaccccatccactggggcagc
aacaccagcaaag (SEQ ID No. 28)

Exon6

agaaccccttcctgttcgccctggaagccgtggtgatcagcgtgggcagcatgggcctgatcatctccctgctgt
gcgtgtacttctggctggaacg
(SEQ ID No. 29)

Exon7
gaccatgcccagaatccccaccctgaagaacctggaagatctggtgaccgagtaccacggcaacttcagc (SEQ ID No. 30)

Exon8
gcctggtccggcgtgagcaagggcctggccgagagcctgcagcccgactacagcgagcggctgtgcctggtgtcc
gagatcccccccaaaggcggagccctgggcgaaggccctggcgccagcccctgcaaccagcacagcccctactgg
gcccctccttgctacaccctgaagcccgagacctga (SEQ ID No. 31)

3'-UTR
tgaaccccaatcctctgacagaagaaccccagggtcctgtagccctaagtggtactaactttccttcattcaacc
cactgcgtctcatactcacctcaccccactgtggctgatttggaattttgtgcccccatgtaagcacccccttcat
ttggcattccccacttgagaattaccctttgccccgaacatgtttttcttctccctcagtctggcccttccttt
tcgcaggattcttcctccccccctctttccctccttcctctttccatctaccctccgattgttcctgaaccgat
gagaaataaagtttctgttgataatcatcg (SEQ ID No. 32)

bGH-pA
actgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtgccact
cccactgccctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggt
ggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatg
g (SEQ ID No. 33)

hPGK promoter
ccacggggttggggttgcgccttttccaaggcagccctgggtttgcgcagggacgcggctgctctgggcgtggtt
ccgggaaacgcagcggcgccgaccctgggtctcgcacattcttcacgtccgttcgcagcgtcacccggatcttcg
ccgctacccttgtgggccccccggcgacgcttcctgctccgcccctaagtcgggaaggttccttgcggttcgcgg
cgtgccggacgtgacaaacggaagccgcacgtctcactagtaccctcgcagacggacagcgccagggagcaatgg
cagcgcgccgaccgcgatgggctgtggccaatagcggctgctcagcagggcgcgccgagagcagcggccgggaag
gggcggtgcgggaggcggggtgtggggcggtagtgtgggccctgttcctgcccgcgcggtgttccgcattctgca
agcctccggagcgcacgtcggcagtcggctccctcgttgaccgaatcaccgacctctctccccagg (SEQ ID No. 34)

eGFP
atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggc
cacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcacc
accggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctac
cccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttc
ttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgag
ctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaac
gtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggc
agcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccac
tacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtg
accgccgccgggatcactctcggcatggacgagctgtacaagtaa (SEQ ID No. 35)

SV40-pA
tgagatccagacatgataagatacattgatgagtttggacaaaccaaaactagaatgcagtgaaaaaaatgcctt
atttgtgaaatttgtgatgctattgccttatttgtaaccattataagctgcaataaacaagttaacaacaacaat
tgcattcattttatgtttcaggttcaggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgt
ggtatggctgattatgatc (SEQ ID No. 36)

Homologous arm-right

tacaatcatgtgggcagaattgaaaagtggagtgggaagggcaaggggagggttccctgcctcacgctacttct
tctttctttcttgtttgtttgtttctttctttctttgaggcagggtctcactatgttgcctaggctggtctcaa
actcctggcctctagtgatcctcctgcctcagcctttcaaagcaccaggattacagacatgagccaccgtgcttg
gcctcctccttctgaccatcatttctctttccctccctgcct (SEQ ID No. 37)

5'-ITR

agtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacg
cccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcagg (SEQ ID No. 38)

In another embodiment, the gene editing targets CD40L. In another embodiment, the donor template targets CD40L.

In another embodiment, the gene editing targets RAG-1. In another embodiment, the donor template targets RAG-1.

Nucleotide of interest

The vector used in the present invention preferably comprises one or more nucleotides of interest.

Preferably the nucleotide of interest gives rise to a therapeutic effect.

Suitably, the one or more NOIs for use in the present invention may be selected from: a guide RNA, a nucleotide encoding a Cas9 ribonucleoprotein, nucleotide sequences encoding one or more adenoviral proteins, nucleotide sequences encoding an agent which promotes homology directed DNA repair (such as an inhibitor of p53 activation or nucleotide sequences encoding one or more adenoviral proteins).

Suitable NOIs include, but are not limited to sequences encoding enzymes, cytokines, chemokines, hormones, antibodies, anti-oxidant molecules, engineered immunoglobulin-like molecules, single chain antibodies, fusion proteins, immune co-stimulatory molecules, immunomodulatory molecules, anti-sense RNA, microRNA, shRNA, siRNA, guide RNA (gRNA, e.g. used in connection with a CRISPR/Cas system), ribozymes, miRNA target sequences, a transdomain negative mutant of a target protein, toxins, conditional toxins, antigens, tumour suppressor proteins, growth factors, transcription factors, membrane proteins, surface receptors, anti-cancer molecules, vasoactive proteins and peptides, anti-viral proteins and ribozymes, and derivatives thereof (such as derivatives with an associated reporter group). The NOIs may also encode pro-drug activating enzymes. Preferably, the NOI is a guide RNA (gRNA).

Pharmaceutical composition

In one embodiment, the cells of the present invention may be formulated for administration to subjects with a pharmaceutically acceptable carrier, diluent or excipient. Suitable carriers and diluents include isotonic saline solutions, for example phosphate-buffered saline, and potentially contain human serum albumin.

Handling of the cell therapy product is preferably performed in compliance with FACT-JACIE International Standards for cellular therapy.

Haematopoietic cell, haematopoietic stem and/or haematopoietic progenitor cell transplantation

The present invention provides a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells, prepared according to a method of the invention for use in therapy, for example for use in gene therapy.

The use may be as part of a cell transplantation procedure, for example a haematopoietic stem cell transplantation procedure.

Haematopoietic stem cell transplantation (HSCT) is the transplantation of blood stem cells derived from the bone marrow (in this case known as bone marrow transplantation) or blood. Stem cell transplantation is a medical procedure in the fields of haematology and oncology, most often performed for people with diseases of the blood or bone marrow, or certain types of cancer.

Many recipients of HSCTs are multiple myeloma or leukaemia patients who would not benefit from prolonged treatment with, or are already resistant to, chemotherapy. Candidates for HSCTs include paediatric cases where the patient has an inborn defect such as severe combined immunodeficiency or congenital neutropenia with defective stem cells, and also children or adults with aplastic anaemia who have lost their stem cells after birth. Other conditions treated with stem cell transplants include sickle-cell disease, myelodysplastic syndrome, neuroblastoma, lymphoma, Ewing's Sarcoma, Desmoplastic small round cell tumour and Hodgkin's disease. More recently non-myeloablative, or so-called "mini transplant", procedures have been developed that require smaller doses of preparative chemotherapy and radiation. This has allowed HSCT to be conducted in the elderly and other patients who would otherwise be considered too weak to withstand a conventional treatment regimen.

In one embodiment, a population of haematopoietic stem cells prepared according to a method of the invention is administered as part of an autologous stem cell transplant procedure.

In another embodiment, a population of haematopoietic stem cells prepared according to a method of the invention is administered as part of an allogeneic stem cell transplant procedure.

The term “autologous stem cell transplant procedure” as used herein refers to a procedure in which the starting population of cells (which are then transduced according to a method of the invention) is obtained from the same subject as that to which the transduced cell population is administered. Autologous transplant procedures are advantageous as they avoid problems associated with immunological incompatibility and are available to subjects irrespective of the availability of a genetically matched donor.

The term “allogeneic stem cell transplant procedure” as used herein refers to a procedure in which the starting population of cells (which are then transduced according to a method of the invention) is obtained from a different subject as that to which the transduced cell population is administered. Preferably, the donor will be genetically matched to the subject to which the cells are administered to minimise the risk of immunological incompatibility.

Suitable doses of transduced cell populations are such as to be therapeutically and/or prophylactically effective. The dose to be administered may depend on the subject and condition to be treated, and may be readily determined by a skilled person.

Haematopoietic progenitor cells provide short term engraftment. Accordingly, gene therapy by administering transduced haematopoietic progenitor cells would provide a non-permanent effect in the subject. For example, the effect may be limited to 1-6 months following administration of the transduced haematopoietic progenitor cells.

Such haematopoietic progenitor cell gene therapy may be suited to treatment of acquired disorders, for example cancer, where time-limited expression of a (potentially toxic) anti-cancer nucleotide of interest may be sufficient to eradicate the disease.

The present invention may be useful in the treatment of the disorders listed in WO 1998/005635. For ease of reference, part of that list is now provided: cancer, inflammation or inflammatory disease, dermatological disorders, fever, cardiovascular effects, haemorrhage, coagulation and acute phase response, cachexia, anorexia, acute infection, HIV infection, shock states, graft-versus-host reactions, autoimmune disease, reperfusion injury, meningitis, migraine and aspirin-dependent anti-thrombosis; tumour growth, invasion and

spread, angiogenesis, metastases, malignant, ascites and malignant pleural effusion; cerebral ischaemia, ischaemic heart disease, osteoarthritis, rheumatoid arthritis, osteoporosis, asthma, multiple sclerosis, neurodegeneration, Alzheimer's disease, atherosclerosis, stroke, vasculitis, Crohn's disease and ulcerative colitis; periodontitis, gingivitis; psoriasis, atopic dermatitis, chronic ulcers, epidermolysis bullosa; corneal ulceration, retinopathy and surgical wound healing; rhinitis, allergic conjunctivitis, eczema, anaphylaxis; restenosis, congestive heart failure, endometriosis, atherosclerosis or endosclerosis.

In addition, or in the alternative, the invention may be useful in the treatment of the disorders listed in WO 1998/007859. For ease of reference, part of that list is now provided: cytokine and cell proliferation/differentiation activity; immunosuppressant or immunostimulant activity (e.g. for treating immune deficiency, including infection with human immune deficiency virus; regulation of lymphocyte growth; treating cancer and many autoimmune diseases, and to prevent transplant rejection or induce tumour immunity); regulation of haematopoiesis, e.g. treatment of myeloid or lymphoid diseases; promoting growth of bone, cartilage, tendon, ligament and nerve tissue, e.g. for healing wounds, treatment of burns, ulcers and periodontal disease and neurodegeneration; inhibition or activation of follicle-stimulating hormone (modulation of fertility); chemotactic/chemokinetic activity (e.g. for mobilising specific cell types to sites of injury or infection); haemostatic and thrombolytic activity (e.g. for treating haemophilia and stroke); anti-inflammatory activity (for treating e.g. septic shock or Crohn's disease); as antimicrobials; modulators of e.g. metabolism or behaviour; as analgesics; treating specific deficiency disorders; in treatment of e.g. psoriasis, in human or veterinary medicine.

In addition, or in the alternative, the invention may be useful in the treatment of the disorders listed in WO 1998/009985. For ease of reference, part of that list is now provided: macrophage inhibitory and/or T cell inhibitory activity and thus, anti-inflammatory activity; anti-immune activity, i.e. inhibitory effects against a cellular and/or humoral immune response, including a response not associated with inflammation; inhibit the ability of macrophages and T cells to adhere to extracellular matrix components and fibronectin, as well as up-regulated of receptor expression in T cells; inhibit unwanted immune reaction and inflammation including arthritis, including rheumatoid arthritis, inflammation associated with hypersensitivity, allergic reactions, asthma, systemic lupus erythematosus, collagen diseases and other autoimmune diseases, inflammation associated with atherosclerosis, arteriosclerosis, atherosclerotic heart disease, reperfusion injury, cardiac arrest, myocardial infarction, vascular inflammatory disorders, respiratory distress syndrome or other cardiopulmonary diseases, inflammation associated with peptic ulcer, ulcerative colitis and

other diseases of the gastrointestinal tract, hepatic fibrosis, liver cirrhosis or other hepatic diseases, thyroiditis or other glandular diseases, glomerulonephritis or other renal and urologic diseases, otitis or other oto-rhino-laryngological diseases, dermatitis or other dermal diseases, periodontal diseases or other dental diseases, orchitis or epididimo-orchitis, infertility, orchidal trauma or other immune-related testicular diseases, placental dysfunction, placental insufficiency, habitual abortion, eclampsia, pre-eclampsia and other immune and/or inflammatory-related gynaecological diseases, posterior uveitis, intermediate uveitis, anterior uveitis, conjunctivitis, chorioretinitis, uveoretinitis, optic neuritis, intraocular inflammation, e.g. retinitis or cystoid macular oedema, sympathetic ophthalmia, scleritis, retinitis pigmentosa, immune and inflammatory components of degenerative fondus disease, inflammatory components of ocular trauma, ocular inflammation caused by infection, proliferative vitreo-retinopathies, acute ischaemic optic neuropathy, excessive scarring, e.g. following glaucoma filtration operation, immune and/or inflammation reaction against ocular implants and other immune and inflammatory-related ophthalmic diseases, inflammation associated with autoimmune diseases or conditions or disorders where, both in the central nervous system (CNS) or in any other organ, immune and/or inflammation suppression would be beneficial, Parkinson's disease, complication and/or side effects from treatment of Parkinson's disease, AIDS-related dementia complex HIV-related encephalopathy, Devic's disease, Sydenham chorea, Alzheimer's disease and other degenerative diseases, conditions or disorders of the CNS, inflammatory components of stokes, post-polio syndrome, immune and inflammatory components of psychiatric disorders, myelitis, encephalitis, subacute sclerosing pan-encephalitis, encephalomyelitis, acute neuropathy, subacute neuropathy, chronic neuropathy, Guillaim-Barre syndrome, Sydenham chora, myasthenia gravis, pseudo-tumour cerebri, Down's Syndrome, Huntington's disease, amyotrophic lateral sclerosis, inflammatory components of CNS compression or CNS trauma or infections of the CNS, inflammatory components of muscular atrophies and dystrophies, and immune and inflammatory related diseases, conditions or disorders of the central and peripheral nervous systems, post-traumatic inflammation, septic shock, infectious diseases, inflammatory complications or side effects of surgery, bone marrow transplantation or other transplantation complications and/or side effects, inflammatory and/or immune complications and side effects of gene therapy, e.g. due to infection with a viral carrier, or inflammation associated with AIDS, to suppress or inhibit a humoral and/or cellular immune response, to treat or ameliorate monocyte or leukocyte proliferative diseases, e.g. leukaemia, by reducing the amount of monocytes or lymphocytes, for the prevention and/or treatment of graft rejection in cases of transplantation of natural or artificial cells, tissue and organs such as cornea, bone marrow, organs, lenses, pacemakers, natural or artificial skin tissue.

In addition, or in the alternative, the invention may be useful in the treatment of β-thalassemia, chronic granulomatous disease, metachromatic leukodystrophy, mucopolysaccharidoses disorders and other lysosomal storage disorders.

Kit

In one aspect, the invention provides a kit comprising an agent which promotes homology directed DNA repair and/or cell populations of the invention.

In another aspect, the present invention provides a kit comprising an agent which promotes homology directed DNA repair, one or more nucleotide sequences encoding gene editing machinery and means for selecting haematopoietic stem cells.

The agent which promotes homology directed DNA repair, and/or cell populations may be provided in suitable containers.

Suitably, the kit may comprise an inhibitor of p53 activation.

Suitably, the kit may comprise a nucleic acid sequence encoding at least one adenoviral protein

The kit may also include instructions for use.

Method of treatment

It is to be appreciated that all references herein to treatment include curative, palliative and prophylactic treatment; although in the context of the invention references to preventing are more commonly associated with prophylactic treatment. In one embodiment, the treatment of mammals, particularly humans, is preferred. Both human and veterinary treatments are within the scope of the invention.

The skilled person will understand that they can combine all features of the invention disclosed herein without departing from the scope of the invention as disclosed.

Administration

Although the agents for use in the invention (in particular, the populations of cells produced by a method of the invention) can be administered alone, they will generally be administered in admixture with a pharmaceutical carrier, excipient or diluent, particularly for human therapy.

Dosage

The skilled person can readily determine an appropriate dose of one of the agents of the invention to administer to a subject without undue experimentation. Typically, a physician will determine the actual dosage which will be most suitable for an individual patient and it will depend on a variety of factors including the activity of the specific agent employed, the metabolic stability and length of action of that agent, the age, body weight, general health, sex, diet, mode and time of administration, rate of excretion, drug combination, the severity of the particular condition, and the individual undergoing therapy. There can of course be individual instances where higher or lower dosage ranges are merited, and such are within the scope of the invention.

Subject

A "subject" refers to either a human or non-human animal.

Examples of non-human animals include vertebrates, for example mammals, such as non-human primates (particularly higher primates), dogs, rodents (e.g. mice, rats or guinea pigs), pigs and cats. The non-human animal may be a companion animal.

Preferably, the subject is a human.

Preferred features and embodiments of the invention will now be described by way of non-limiting examples.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, biochemistry, molecular biology, microbiology and immunology, which are within the capabilities of a person of ordinary skill in the art. Such techniques are explained in the literature. See, for example, Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press; Ausubel, F.M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, Ch. 9, 13 and 16, John Wiley & Sons; Roe, B., Crabtree, J. and Kahn, A. (1996) DNA Isolation and Sequencing: Essential Techniques, John Wiley & Sons; Polak, J.M. and McGee, J.O'D. (1990) In Situ Hybridization: Principles and Practice, Oxford University Press; Gait, M.J. (1984) Oligonucleotide Synthesis: A Practical Approach, IRL Press; and Lilley, D.M. and Dahlberg, J.E. (1992) Methods in Enzymology: DNA Structures Part A: Synthesis and Physical Analysis of DNA, Academic Press. Each of these general texts is herein incorporated by reference.

EXAMPLES

Example 1 – Inhibitors of p53 activation

Materials and Methods

Vectors and nucleases

AAV6 donor templates for HDR were generated from a construct containing AAV2 inverted terminal repeats, produced by triple-transfection method and purified by ultracentrifugation on a cesium chloride gradient as previously described (J. Wang et al. 2015 *Nat Biotechnol* **33**, 1256-1263). Design of AAV6 donor templates with homologies for AAVS1 locus (encoding for a PGK.GFP reporter cassette) or targeting the intron 1 of *IL2RG* (encoding for *IL2RG* corrective cDNA followed by a PGK.GFP reporter cassette was previously reported G. Schiroli et al. 2017 *Sci Transl Med* **9)**.

Sequences of the gRNAs were designed using an online CRISPR design tool (P.D. Hsu et al. 2013 DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832) and selected for predicted specificity score and on target activity. Genomic sequences recognized by the gRNAs are indicated in the Table below. Ribonucleoproteins (RNPs) were assembled by incubating at 1:1.5 molar ratio s.p.Cas9 protein (Aldevron) with synthetic cr:tracrRNA (Integrated DNA Technologies) for 10 minutes at 25°C. Electroporation enhancer (Integrated DNA Technologies) was added prior to electroporation according to manufacturer's instructions.

Sequences of the gRNA are reported below.

Guide ID	**Genomic sequence (5'-3')**
AAVS1 gRNA	GTCACCAATCCTGTCCCTAGTGG (SEQ ID No. 39)
IL2RG gRNA	ACTGGCCATTACAATCATGTGGG (SEQ ID No. 40)

GSE56.WPRE construct was cloned from a lentivirus expressing GSE56.WPRE in a pVax plasmid for mRNA in vitro transcription containing a T7 promoter, β-globin 3'UTR and 64bp-polyA.

Gene editing of human CD34+ cells

CD34+ cells were either freshly purified from human CB after obtaining informed consent and upon approval by the Ospedale San Raffaele Bioethical Committee, or purchased frozen from Lonza. CD34+ cells were edited according to a previously optimized protocol (Schiroli et al. 2017 supra).

Briefly, $5x10^5$ CD34+ cells/ml were stimulated in serum-free StemSpan medium (StemCell Technologies) supplemented with penicillin, streptomycin, glutamine, 1 µM SR-1(Biovision), 50 µM UM171 (STEMCell Technologies), 10 µM PGE2 added only at the beginning of the culture (Cayman), and human early-acting cytokines (SCF 100 ng/ml, Flt3-L 100 ng/ml, TPO 20 ng/ml, and IL-6 20 ng/ml; all purchased from Peprotech). After 3 days of prestimulation, cells were washed with PBS and electroporated using P3 Primary Cell 4D-Nucleofector X Kit and program EO-100 (Lonza). Cells were electroporated with 2.5-1.25 µM of RNPs.

Transduction with AAV6 was performed at a dose of $1\text{-}2x10^4$ vg/cell 15' after electroporation. GSE56 mRNA was utilized were indicated at a dose of 150 µg/ml. Gene editing efficiency was measured from cultured cells in vitro 3 days after electroporation by flow cytometry measuring the percentage of cells expressing the GFP marker or by digital droplet PCR analysis designing primers and probe on the junction between the vector sequence and the targeted locus and on control sequences utilized as normalizer as previously described (Schiroli et al. 2017 supra).

CD34+ HSPC xenotransplantation studies in NSG mice

NOD-SCID-IL2Rg$^{-/-}$ (NSG) mice were purchased from The Jackson Laboratory and were maintained in specific-pathogen-free (SPF) conditions. The procedures involving animals were designed and performed with the approval of the Animal Care and Use Committee of the San Raffaele Hospital (IACUC #749) and communicated to the Ministry of Health and local authorities according to Italian law.

For transplantation, $3x10^5$ CD34+ cells treated for editing at day 5 of culture were injected intravenously into NSG mice after sub-lethal irradiation (150-180 cGy). Sample size was determined by the total number of available treated cells. Mice were attributed to each experimental group randomly. Human CD45+ cell engraftment and the presence of gene-edited cells were monitored by serial collection of blood from the mouse tail and, at the end of the experiment (>20 weeks after transplantation), BM and spleen were harvested and analyzed.

Molecular analyses

For molecular analyses, genomic DNA was isolated with DNeasy Blood & Tissue Kit or QIAamp DNA Micro Kit (QIAGEN) according to the number of cells available. Nuclease activity (*IL2RG* intron 1, *AAVS1*) was measured by mismatch-sensitive endonuclease assay by PCR-based amplification of the targeted locus followed by digestion with T7

Endonuclease I (NEB) according to the manufacturer's instructions. Digested DNA fragments were resolved and quantified by capillary electrophoresis on LabChip GX Touch HT (Perkin Elmer) according to the manufacturer's instructions.

For digital droplet PCR analysis, 5-50 ng of genomic DNA were analyzed in duplicate using the QX200 Droplet Digital PCR System (Biorad) according to the manufacturer's instructions.

For HDR ddPCR, primers and probes were designed on the junction between the vector sequence and the targeted locus and on control sequences used for normalization (human *TTC5* genes). Thermal conditions for annealing and extension were adjusted for each specific application as follows: *AAVS1*/ Intron 1 *IL2RG* HDR 3' integration junction ddPCR: 55°C for 30 sec, 72°C for 2 min. Primers and probes for PCR and ddPCR amplifications are shown below.

Description	**Orientation**	**Sequence (5'-3')**
NHEJ *AAVS1*	FW	CTTCAGGACAGCATGTTTGC (SEQ ID No. 41)
	RV	ACAGGAGGTGGGGGTTAGAC (SEQ ID No. 42)
NHEJ Intron 1 *IL2RG*	FW	CACCCTCTGTAAAGCCCTGG (SEQ ID No. 43)
	RV	AAGAAATCTAGATTGGGGAG (SEQ ID No. 44)
Intron 1 *IL2RG* 3' integration junction ddPCR	FW	CTAGATTGGGGAGAAAATGA (SEQ ID No. 45)
	RV	GTGGGAAGGGGCCGTACAG (SEQ ID No. 46)
	Probe (FAM)	GTAGCTCCTATGCTAGGCGTAGCC (SEQ ID No. 47)
AAVS1 3' integration junction ddPCR	FW	GATTGGGAAGACAATAGCAG (SEQ ID No. 48)
	RV	TCTTGGGAAGTGTAAGGAAG (SEQ ID No. 49)
	Probe (FAM)	CCAGATAAGGAATCTGCCTA (SEQ ID No. 50)
Human TTC5 ddPCR	(HEX)	PrimePCR ddPCR Copy Number Assay: TTC5, Human (Biorad)
CDKN1A	Probe (FAM)	Hs00355782_m1 (Life Technologies)
HPRT1	Probe (FAM)	Hs99999909_m1 (Life Technologies)
	Probe (FAM)	Hs02800695_m1 (Life Technologies)

RPS27L	Probe (FAM)	Hs00955038_g1 (Life Technologies)
APOBEC3H	Probe (FAM)	Hs00419665_m1 (Life Technologies)
PHLDA3	Probe (FAM)	Hs00385313_m1 (Life Technologies)

For gene expression analyses, total RNA was extracted using the RNeasy Plus Micro Kit (QIAGEN). cDNA was synthetized with SuperScript VILO IV cDNA Synthesis Kit (Invitrogen) and used for Q-PCR in a Viia7 Real-time PCR thermal cycler using TaqMan Gene Expression Assays (Applied Biosystems) mapping to *CDKN1A*, *RPS27L*, *PHLDA3*, *APOBEC3H*, and *HPRT* as normalizer. The relative expression of each gene was first normalized to *HPRT* expression and then represented as fold change relative to the mock-treated sample.

Flow cytometry

For immunophenotypic analyses (performed on FACSCanto II; BD Pharmingen), we used the antibodies listed below. Single stained and Fluorescence Minus One stained cells were used as controls. LIVE/DEAD Fixable Dead Cell Stain Kit (Thermo Fisher), 7-aminoactinomycin (Sigma Aldrich), were included in the sample preparation for flow cytometry according to the manufacturer's instructions to exclude dead cells from the analysis. Cell sorting was performed using MoFlo XDP Cell Sorter (Beckman Coulter) or FACSAria Fusion (BD Biosciences).

Anti-Human Antibody	**Fluorochrome**	**Clone**	**Company**
Annexin V	PB		Biolegend
CD16/32	none		Miltenyi Biotec
CD133/2	PE	293C3	Miltenyi Biotec
CD34	PB	AC136	Miltenyi Biotec
CD34	PECy7	8G12	BD
CD90	APC	5E10	BD
CD132	APC	TUGm2	Biolegend
CD45	PB	HI30	Biolegend
CD45	APCH7	HI30	eBioscience

CD19	PE	HIB19	BD
CD19	Pecy7	HIB19	Biolegend
CD3	Pecy7	HIT3a	Biolegend
CD3	PE	SK7	BD
CD13	APC	WM15	BD
CD33	PeCy7	P67.6	BD
CD38	APC	HB7	BD
CD38	Percp5.5	HB7	Biolegend
CD4	PB	RPA-T4	BD
CD8	APCH7	SK1	BD
CD235a	PE	JC159	Dako

Single-cell RNA-Sequencing and analyses

Droplet-based digital 3′ end scRNA-Seq was performed on a Chromium Single-Cell Controller (10X Genomics) using the Chromium Single Cell 3′ Reagent Kit v2 according to the manufacturer's instructions. 24 hrs after the editing treatment, CD34+ cells were sorted according to surface expression of CD34+ CD133+ CD90+ and CD34+ CD133+ CD90-; viable cells were counted with Trypan Blue solution 0.4% (Gibco) and 5200 viable cells (2700 from each population) were utilized for the subsequent procedure (estimated recovery: 3000 cells/sample). Briefly, single cells were partitioned in Gel Beads in Emulsion (GEMs) and lysed, followed by RNA barcoding, reverse transcription and PCR amplification (12-14 cycles, according to the available cDNA quantity). scRNA-Seq libraries were prepared according to the manufacturer's instructions, checked and quantified on LabChip GX Touch HT (Perkin Elmer) and Qubit 3.0 (Invitrogen) instruments. Sequencing was performed on a NextSeq 500 machine (Illumina) using the NextSeq 500/550 High Output v2 kit (75 cycles).

Fastq files were processed with Cell Ranger (v. 1.3, https://support.10xgenomics.com/single-cell-gene expression/software/pipelines/latest/what-is-cell-ranger) using default parameters. Reads were aligned to reference genome hg19.

Reads processing: Gene counts were processed with Seurat (v 2.3.1, http://satijalab.org/seurat/). Cells expressing less than 200 unique genes and genes expressed in less than 3 cells/sample were discarded. Counts were normalized using Seurat function NormalizeData with default parameters. Genes with a mean expression lower than 0.01 were excluded. Cells with a ratio of mitochondrial versus endogenous genes expression

exceeding 0.15 were also excluded. Expression data were than scaled using ScaleData function, regressing on difference between S and G2M scores. Cell cycle scores were calculated using CellCycleScoring function. Multi-Set Canonical Correlation Analysis (mCCA) A. Butler et al. 2018 *Nat Biotechnol* **36**, 411-420 was then performed for all the samples. For mCCA computation, a list of genes differentially expressed between cultured CD34+ CB-derived primitive and committed cells*I. Fares et al. 2017 Blood* **129**, 3344-3351 was used as input, and the first 20 dimensions were aligned.

Differential expression analysis: Genes differentially expressed across different conditions were identified using FindMarkers function, using MAST test G. Finak et al. 2015 *Genome Biol* **16**, 278 with Bonferroni correction. avglogFC was computed adding to averaged pseudocount expression values of 0.001 and only genes expressed in at least 1% of cells in at least one sample were considered. Gene Ontology enrichment analysis was then performed on these sets of genes using EnrichR (http://amp.pharm.mssm.edu/Enrichr/).

Results

Gene editing of HSPCs causes deregulation of the p53-related transcriptional response

To dissect the impact of gene editing procedure on human HSPC in an unbiased manner, we performed single-cell (sc)RNA-Seq on primitive HSPCs populations enriched for CD133 and CD90 surface markers at 24 hrs time point after the editing procedure (Figure 1A).

Using a droplet-based approach *(G.X. Zhenget al. 2017 Nat Commun* 8, 14049) we generated scRNA-Seq data from cells treated with RNP specific for *IL2RG* and AAV6 as DNA template for HDR or mock treated as negative control. By performing pathway analysis of the genes found differentially expressed between edited and control cells, the top enriched pathways are related to p53 transcriptional activity (Figure 1B-C). This p53-related transcriptional response was mildly triggered after the DNA DSB induced by the nucleases, and mostly related to the transduction with AAV6 vector, which are reported to broadly interact with the host cell DDR machinery *(M Fragkos et al. 2009 Mol Cell Biol* 29, 2828-2840), as measured by the induction of p21 expression shortly after the editing procedure (Figure 1D).

T*ransient p53 inhibition by pifithrin-α or derivatives thereof improve HDR efficiency and increases gene editing in primitive cell compartment*

Since we showed that the most deregulated genes in edited HSPC converge significantly into the DDR and in particular into the p53 signaling pathway, we further investigated the functional consequences of this response on HSPC self-renewal, survival, and DNA repair choice in gene edited cells.

We thus tested transient p53inhibition by the addition of pifithrin-α or derivatives (pifithrin-α cyclic and pifithrin-α p-nitro), molecules which are reported to inhibit p53-mediated apoptosis and p53-dependent gene transcription (P.G. Komarov et al. 1999 *Science* 285, 1733-1737 and K. I. Lenova et al. 2010 *Cell Cycle* 9, 1434-1443) (Figure 2A). By adding the molecules one day before viral transduction in the gene editing protocol, we were able to measure an increase in HDR efficiency with all the three compounds (Figure 2B). For further experiments we selected the pifithrin-α cyclic compound, which had the lowest cellular toxicity, and showed a two-fold in increase in gene editing in the most primitive CD34+ CD133+ CD90+ cell compartment, which are basally endowed with the highest p53 transcriptional activity (V. Pant et al. 2012 *Blood* 120, 5118-5127) and in DNA repair by NHEJ, possibly due to higher survival of edited cells (Figure 2C).

Transient p53 inhibition by a dominant negative peptide promotes the HDR response and, causes expansion of the primitive subset without impacting on cell differentiation

We transiently inhibited p53 transcriptional activity by mRNA overexpression of a dominant negative peptide (GSE56) *(M. Milyavsky et al. 2010 Cell stem cell* 7, 186-197) during the electroporation procedure (Figure 3A).

By addition of GSE56 we were able to reduce the amplitude of the p53-dependent transcriptional response, as measured by expression level of some downstream target genes found highly upregulated by the gene editing procedure (Figure 3B). This treatment did not negatively impact overall efficiency of DNA repair and promoted HDR-mediated integration at the targeted locus (Figure 3C, left panel) especially in the most primitive cell compartment (mean 42% in the CD34+ 133+ 90+ fraction; Figure 3D, right panel), consistently with p53-mediated control of DNA repair choice (*23*). HSPC treated with GSE56 showed higher expansion (Figure 3D), compatibly with the reduced induction of cell cycle inhibitors, and there was no change in relative composition of the hematopoietic populations measured at 3 days post editing in edited cells compared to untreated controls (Figure 3E), suggesting effective proliferation of primitive cells without overt induction of differentiation. Upon seeding sorted populations in CFU-C assay, erythroid and myeloid clonogenic output was increased particularly in CD90+ fraction (Figure 3F), indicating increased expansion

potential of the most primitive subset, without impacting on cell differentiation of the treated cells.

Edited cells treated with GSE56 show increased engraftment without skewing differentiation in immunodeficient NSG mice

We then performed transplantation of the edited cells into immunodeficient NSG mice, which represent a surrogate assay for long-term repopulating HSC (Figure 4A). Edited cells treated with GSE56 showed increased engraftment (Figure 4B), without any skewing in the differentiation output (Figure 4C). Interestingly, while the fraction of HDR-edited cells slightly decreased over time in mice transplanted with cells edited with conventional protocol, as previously reported (Schiroli et al. 2017 supra and S.S. De Ravin 2017 *Science translational medicine* 9), by transient p53 inhibition it remained stable at long-term, reaching an average of 32% in the PB at 22 weeks post transplantation (Figure 4D).

Overall, these results indicate that p53-dependent transcriptional response contributes to reduce the repopulation capacity of gene edited cells, and its transient reduction could help to better preserve the long-term function of the most primitive HSPC compartments.

Example 2 – Adenoviral proteins

Materials and Methods

Vector

AAV stocks were produced by triple transfection of the plasmids carrying the cassette of interest framed by deleted AAV-ITRs, the Rep2 and Cap6 protein coding sequences and the adenoviral helper genes in HEK293-T cell line. Transfected cells are collected 48 hours later and purified over cesium chloride gradients (Wang, J. et al. (2015) Nat Biotechnol 33: 1256-1263). Titers of the viral stocks were determined by qPCR of AAV genomes (vg/ml). The design of AAV-6 donor templates including homology sequences for *AAVS1* locus (encoding for the PGK-GFP reporter cassette) or intron 1 within *IL2RG* locus (encoding for the *IL2RG* corrective cDNA followed by the PGK-GFP reporter cassette), were previously reported (Schiroli, G. et al. (2017) Sci Transl Med 9) and Schiroli, G. 2019 Celle stem cell

The barcoded AAV6 donor template targeting the *AAVS1* locus was obtained by subcloning the degenerated sequence into the previously reported *AAVS1* AAV6 transfer vector plasmid downstream of the polyadenylation sequence. Briefly, a ssODN embedding the random sequence flanked by the unique cloning restriction sites (Bsu36I and SphI) was obtained

from Sigma Aldrich. The degenerated region was designed to avoid any undesired cutting by cloning restriction enzymes. To generate the complementary strand, 50 pmol of the ssODN were amplified with 10 cycles of PCR with Easy-A High-Fidelity enzyme (Agilent Technologies) using the appropriate forward and reverse primers according to the manufacturer's instructions. The amplified product was purified with MinElute PCR Purification kit (QIAGEN), digested with the restriction enzymes and checked by capillary electrophoresis. 2 ug of the Bsu36I/SphI-digested *AAVS1* HDR donor template were ligated with the digested insert (molar ratio 1:7) using T4 DNA Ligase (NEB) by scaling up the manufacturer's protocol. Finally, XL-10 Gold Ultracompetent Cells (Agilent Technologies) were transformed with the ligation product, plated and incubated for 12 hours of incubation at 30°C to minimise the risk of recombination. Colonies were scraped, mixed, grown in LB medium for an additional 6 hours and processed with NucleoBond Xtra MaxiPrep according to the manufacturer's instructions. Ultimately, the plasmid prep was screened with restriction enzymes for ITR and plasmid integrity.

Nuclease

Sequences of the gRNAs were designed using an online CRISPR design tool (Hsu, P.D. et al. (2013) Nat Biotechnol 31: 827-832) and selected for predicted specificity score and on target activity. Genomic sequences recognised by the gRNAs are indicated in the table below. Ribonucleoproteins (RNPs) were assembled by incubating at 1:1.5 molar ratio SpCas9 protein (Aldevron) with synthetic cr:tracrRNA (Integrated DNA Technologies) for at least 10 min at 25°C. Electroporation enhancer (Integrated DNA Technologies) was added prior to electroporation according to the manufacturer's instructions.

Guide ID	**Genomic sequence (5'-3')**
AAVS1 gRNA	GTCACCAATCCTGTCCCTAGTGG (SEQ ID NO: 51)
IL2RG gRNA	ACTGGCCATTACAATCATGTGGG

Proteins

E4orf1, E4orf6/7, GSE56/Ad5-E4orf6/7, E4orf6 and E1B55K adenoviral protein sequences derived from wild-type serotype 5 were collected from RCSB PDB and UniProt online databanks. Moreover, a multiple-sequence alignment was performed among various Adenovirus serotypes for each adenoviral protein family E4orf1 and E4orf6/7 using the T-Coffee algorithm (Notredame, C. et al. (2000) J. Mol. Biol. doi:10.1006/jmbi.2000.4042) to select some variants. All sequences were synthesised with codon optimisation for *Homo*

sapiens (GeneArt™). Each construct was cloned in a "pVax" plasmid for mRNA in vitro transcription containing a T7 promoter, WPRE or 3'UTR and 64bp-polyA. pVax was linearised with the restriction enzyme (SpeI) and purified using phenol-chloroform. RNA was in vitro transcribed using the commercial 5xMEGAscript T7 kit (Invitrogen) with slight modifications from the manufacturer's protocol as by adding ARCA (BioLabs). Synthetic RNA was purified using the RNeasy Plus Mini Kit (Qiagen) followed by HPLC column purification to remove RNA contaminants and reagents. RNA was finally concentrated using a tube Amicon Ultra-15 (30K) (Millipore).

Gene editing protocol in human HSPCs

$CD34^+$ cells were purchased frozen from Lonza. $CD34^+$ cells were edited according to a previously optimised protocol (Schiroli, G. et al (2017) Sci Transl Med 9). Briefly, 1-5x10^5 $CD34^+$ cells/ml were stimulated in serum-free StemSpan medium (StemCell Technologies) supplemented with penicillin, streptomycin, glutamine, 1 µM of the arylhydrocarbon receptor antagonist SR-1 (Biovision), 50 µM of the pyrimidoindole-derivative UM171 (STEMCell Technologies), 10 µM of the 15-hydroxy PGDH competitive inhibitor dmPGE-2 added only at the beginning of the culture (Cayman), and human early-acting cytokines (SCF 100 ng/ml, Flt3-L 100 ng/ml, TPO 20 ng/ml, and IL-6 20 ng/ml; all purchased from Peprotech). After 3 days of pre-stimulation, cells were washed with PBS and electroporated using P3 Primary Cell 4D-Nucleofector X Kit and program EO-100 (Lonza). Cells were electroporated with 1-1.2 µM of RNPs. Where specified, electroporation mixture was supplemented with Ad5-E4orf1, Ad5-E4orf6/7 and GSE56/Ad5-E4orf6/7 mRNAs, which were respectively utilized at doses of 120, 80 and 200 µg/ml, or Ad5-E1B55K and Ad5-E4orf6 mRNAs were used at 120 and 80 µg/ml. All other adenoviral protein variants were used at standard dose 120 µg/ml. Transduction with AAV-6 was performed 15 min post-electroporation at dose of 1-5x10^4 vg/cell. Where specified, CD34+ cells were stimulated for shorter time in culture (0, 24 or 48 hours) and edited as described above. Gene editing efficiency was measured from cultured cells in vitro at 4 days post-editing by flow cytometry measuring the percentage of cells expressing the GFP reporter gene or by digital droplet PCR analysis designing primers and probe on the junction between the vector sequence and the targeted locus and on control sequence utilised as normaliser as previously described (Schiroli, G. et al (2017) Sci Transl Med 9) and Schiroli et al., Cell Stem Cell 2019.

$CD34^+$ xenotransplantation in NSG mice

NOD-SCID-IL2R$g^{-/-}$ (NSG) mice were purchased from The Jackson Laboratory and were maintained in specific-pathogen-free (SPF) conditions. The procedures involving animals

were designed and performed with the approval of the Animal Care and Use Committee of the San Raffaele Hospital (IACUC #749) and communicated to the Ministry of Health and local authorities according to Italian law. For transplantation, 1.5-3x10^5 (for primary) or 2x10^6 (for secondary) $CD34^+$ treated cells at 24 hours post-editing protocol were injected intravenously into NSG mice after sub-lethal irradiation (180 cGy). Sample size was determined by the total number of available treated cells. Mice were attributed to each experimental group randomly. Human $CD45^+$ cell engraftment and the percentage of edited GFP^+ cells were monitored by serial collection of blood from the mouse tail over the weeks post-transplantation and at the end of the experiment (18-20 weeks post-transplantation) both spleen and bone-marrow were harvested and analysed.

Molecular analyses

For molecular analyses, genomic DNA was isolated with DNeasy Blood & Tissue Kit or QIAamp DNA Micro Kit (QIAGEN) according to the number of cells available. Nuclease activity (*AAVS1*) was measured by mismatch-sensitive endonuclease assay by PCR-based amplification of the targeted locus followed by digestion with T7 Endonuclease I (NEB) according to the manufacturer's instructions. Digested DNA fragments were resolved and quantified by capillary electrophoresis on LabChip GX Touch HT (Perkin Elmer) according to the manufacturer's instructions.

For digital droplet PCR analysis, 5-50 ng of genomic DNA were analysed in duplicate using the QX200 Droplet Digital PCR System (Biorad) according to the manufacturer's instructions.

For HDR ddPCR as for translocations ddPCR, primers and probes were designed on the junction between the vector sequence and the targeted locus and on control sequences used for normalisation (human *TTC5* genes). Thermal conditions for annealing and extension were adjusted for each specific application as follows: *AAVS1* HDR 3' integration junction ddPCR: 55°C for 30 sec, 72°C for 2 min. Primers and probes for PCR and ddPCR amplifications are shown below.

Description	Orientation	Sequence (5'-3')
NHEJ *AAVS1*	FW	CTTCAGGACAGCATGTTTGC (SEQ ID NO: 52)
	RV	GGACTAGAAAGGTGAAGAGCC (SEQ ID NO: 53)
AAVS1 3' integration junction ddPCR	FW	GATTGGGAAGACAATAGCAG (SEQ ID NO: 54)
	RV	TCTTGGGAAGTGTAAGGAAG (SEQ ID NO: 55)
	Probe (FAM)	CCAGATAAGGAATCTGCCTA (SEQ ID NO: 56)
ChrX-chr14	FW	GTGGTGACCACCTTAATCCAGTGA

translocation ddPCR	RV	AAGAAATCTAGATTGGGGAG
	Probe (FAM)	ATGAAGGCAGGGAGGGAAAGAGAAATGA
Human TTC5 ddPCR	(HEX)	PrimePCR ddPCR Copy Number Assay: TTC5, Human (Biorad)
CDKN1A	Probe (FAM)	Hs00355782_m1 (Life Technologies)
HPRT1	Probe (FAM)	Hs02800695_m1 (Life Technologies)
CDKN2A	Probe (FAM)	Hs00924091_m1 (Life Technologies)
APOBEC3H	Probe (FAM)	Hs00419665_m1 (Life Technologies)
CDK2	Probe (FAM)	Hs01548894_m1 (Life Technologies)

Barcoding analysis

PCR amplicons for each individual sample were generated by one-step PCR using Taq G2 Hot Start Polymerase (Promega). The forward primer was designed to bind donor template 2 bp upstream the barcode sequence, while the reverse primer annealed outside the homology region, thus amplifying about 350 bp of the on-target integrated cassette. Primers are endowed with tails containing P5/P7 sequence, i5/i7 Illumina tag to allow multiplexed sequencing and R1/R2 primer complementary sequence. PCR library amplicons were separately purified using MinElute PCR Purification kit (QIAGEN) and quality was assessed by Agilent Tapestation (Agilent Technologies). Finally, amplicons were multiplexed and run on Illumina MiSeq paired-end 2x75bp.

Flow cytometry

For immunophenotypic analyses (performed on FACSCanto II; BD Pharmingen), we used the antibodies listed below. Single stained and Fluorescence Minus One stained cells were used as controls. LIVE/DEAD Fixable Dead Cell Stain Kit (Thermo Fisher), 7-aminoactinomycin (Sigma Aldrich), were included in the sample preparation for flow cytometry according to the manufacturer's instructions to exclude dead cells from the analysis. Cell sorting was performed using MoFlo XDP Cell Sorter (Beckman Coulter) or FACSAria Fusion (BD Biosciences).

Anti-Human Antibody	**Fluorochrome**	**Clone**	**Company**
Annexin V	PB	-	Biolegend
CD133/2	PE	293C3	Miltenyi Biotec
CD34	PB	AC136	Miltenyi Biotec
CD34	PECy7	8G12	BD

CD90	APC	5E10	BD
CD45	PB	HI30	Biolegend
CD45	APCH7	HI30	eBioscience
CD19	PE	HIB19	BD
CD19	Pecy7	HIB19	Biolegend
CD3	Pecy7	HIT3a	Biolegend
CD3	PE	SK7	BD
CD13	APC	WM15	BD
CD33	PeCy7	P67.6	BD
CD38	APC	HB7	BD
CD38	Percp5.5	HB7	Biolegend
CD4	PB	RPA-T4	BD
CD8	APCH7	SK1	BD

Cell cycle analysis

Cell cycle was analysed by flow cytometry (CytoFLEX LX; Beckman Coulter) at 12-24 hours post-editing by collecting 0.5-1x10^5 total HSPCs in culture. Cells were washed with PBS + 2% FBS and stained by using 10 µg/ml Hoechst 33342 (Sigma-Aldrich) in PBS + 2% FBS for 0.5-1 hour in final volume equal to 200 µl.

Results

Expression of E4orf1 or E4orf6/7during gene editing increases targeting efficiency

Adenoviruses are natural co-helper of AAV infection and provide a set of helper genes (E1a, E1b, E2a, E4) which function is required to optimize AAV infection process. We were interested to investigate specifically the E4 gene which is composed of 7 open reading frames (ORFs) that are coding for seven different proteins (Figure 5).

We decided to investigate the use of two adenoviral proteins called E4orf1 and E4orf6/7 derived from the adenovirus serotype 5 (Ad5), which we delivered by electroporation as mRNA together with the Cas9 ribonucleoprotein (RNP) into human HSPC (cord blood derived CD34+ cells) before AAV6 transduction for the delivery of the donor DNA template (Figure 6A).

By expressing the E4orf6/7 protein during the gene editing procedure, we measured a mean of 1.5-fold increase in the percentage of primitive HSPC that express the GFP reporter gene present in the donor (CD34+CD133+CD90+GFP+). This increase in targeting efficiency was

obtained at all the different doses of AAV6 vector used for the transduction of the donor vector (Figure 6B).

Similar increase in targeted integration was observed in the primitive CD90+ cells also when the donor template was delivered by an Integrase Defective Lentiviral Vector (IDLV; Figure 6C).

When we delivered different doses of E4orf1 mRNA during the gene editing procedure, we found a dose dependent increase in the targeting efficiency and, with the best performing dose, we measured a mean of 1.5-fold increase in the percentage of targeted primitive HSPC, similar to the increase measured in the conditions with different doses of E4orf6/7 mRNA (Figure 6D).

Expression of E4orf1 or E4orf6/7 alone and in combination

We then compared side-by-side the standard gene editing procedure alone or with the addition of the E4orf1, the E4orf6/7 or the combination of both proteins. By cytofluorimetric analysis, we observed that E4orf1 increases the fraction of GFP+ cells to a mean of 1.2-fold compare to the standard and is equivalent in all subpopulations (from the most primitive CD90+ progenitor compartment to the most differentiated CD34- compartment); the E4orf6/7 increases to 1.45-fold the fraction of GFP+ cells mainly in the most primitive HSPC compartment (CD90+); and the combination of them has an additive effect, allowing a mean of 1.55 fold increase in the fraction of GFP+ cells (Figure 7A).

Targeting efficiency was confirmed by digital PCR on the total bulk and compared side-by-side with the total percentage of GFP measured by cytofluorimetric analysis (Figure 7B).

By performing clonogenic assay, we found that the co-treatment with E4orf1 and E4orf6/7 proteins respectively increase of 3 and 2-fold, the number of both myeloid and erythroid colony forming units (CFU) compared to the standard procedure (Figure 7C).

Interestingly, the growth curve of the treated cells was 4-fold higher in both the conditions in which E4orf1 was included during the editing compared to the standard and E4orf6/7 alone conditions (Figure 7D).

A xenotransplantation experiment is on-going in NSG mice to determine long-term engraftment and *in-vivo* differentiation capacity of edited HSPC treated with these conditions.

Expression of E1B55K and E4orf6 during gene editing

We tested the use of E1B55K and E4orf6 proteins for the editing of human HSPC and compared them in parallel with E4orf1 and E4orf6/7 proteins.

We found that the combination of E1B55K and E4orf6 allow only a transient increase up to 1.45-fold in the fraction of GFP+ cells mainly in the most primitive HSPC compartment, probably due to an increase in transduction that results in a higher expression of the GFP reporter from the nonintegrated vector (Figure 8A).

Molecular analyses on the genomic DNA extracted from these treated cells confirmed that the E1B55K and E4orf6 proteins are not effective in increasing the editing efficiency on human HSPC, while both E4orf1 and E4orf6/7 proteins were reproducibly able to boost editing in this cell type (Figure 8B).

By performing clonogenic assay, we did not observe major changes in terms of number of both myeloid and erythroid colony forming units (CFU) compared to the standard condition (Figure 8C).

Interestingly, the growth curve of the treated cells with the combination E1B55K and E4orf6 was lower compared to the standard, that could point out a certain toxicity due to the use of this combination (Figure 8D).

Delivery of E4orf6/7 with Cas9 improves HDR-driven integration in human primary T cells

Finally, we tested the delivery of the E4orf6/7 mRNA together with the Cas9 ribonucleoprotein (RNP) into human primary T cells (peripheral blood derived total CD3+ cells) before AAV6 transduction for the delivery of the donor DNA template (Figure 9A).

By analyzing the cells 20 days after treatment, we found a 1.5-fold increase in the percentage of GFP+ cells expression in the cells treated with the E4orf6/7 compared to

standard condition, both in the bulk treated CD3+ cells and in the CD4 and CD8 subpopulation (Figure 9B).

These results indicate that the E4orf6/7 protein allow boosting HDR-driven integration in different primary human cells.

Edited cells treated with E4orf6/7 mRNA transplanted into immunodeficient NSG mice increase HDR-driven integration in the most primitive HSPC compartments and preserves long-term repopulation capacity

With the best performing doses of AAV6 donor and E4orf6/7 mRNA, we repeated the gene editing experiment with the aim to perform transplantation of the cells into NSG mice, which represents a surrogate assay for long-term repopulating HSC (Figure 10A).

Edited cells treated with E4orf6/7 showed a 1.4-fold increase in the percentage of targeted CD90+ cells (Figure 10B), and without any skew in the culture composition (Figure 10C).

Molecular analyses performed on the genomic DNA extracted from the treated cells confirmed increase levels of HDR-mediated integration and a similar amount of NHEJ induced mutation, when comparing the E4or6/7 and the standard conditions (Figure 10D).

These cells were transplanted one day after electroporation into NSG mice, and their engraftment was longitudinally followed by serial blood analyses. The results indicate that while the co-treatment with E4orf6/7 does not affect the overall level of engraftment of the human cells, the mice transplanted with the E4orf6/7 treated cells showed a significant higher fraction of edited cells (measured as percentage of GFP+ or HDR+ cells) in the blood, in the hematopoietic organs and in different blood lineages (Figure 11).

Overall, these results indicate that transient expression of the E4orf6/7 protein during the gene editing procedure increase the levels of HDR-driven integration in the most primitive HSPC compartments while preserving their long-term repopulation capacity.

EXAMPLE 3

We investigated whether transient overexpression of the p53 inhibitor (GSE56) during editing procedure altered the proportion of cells in S/G2/M phases of the cell cycle. By

performing Hoechst cell-cycle analysis 24 hours after the gene editing procedure, we found a trend towards reduction of the percentage of cells in S/G2/M phases in standard compared to untreated controls, which was partially rescued when cells were edited in presence of GSE56 (Figure 12A). To evaluate whether the improved yield of HDR-edited cells *in vivo* was consistent independently of the targeted locus, we transplanted a limited dose ($T0_{eq}$ = $1.5x10^5$/mouse) of $CD34^+$ cells treated or not with GSE56 mRNA (Figure 12B). For this experiment, we used our recently developed AAV6 vector targeting the *AAVS1* locus and containing a 22-bp highly degenerated molecular region (barcodes). This vector allows *in vivo* clonal tracking of edited cells and the identification of gene editing protocols which could improve the clonality of the graft. When analysing transplanted NSG mice, we found significantly higher engraftment in the haematopoietic organs (Figure 12C), despite a similar fraction of gene edited cells than standard protocol (Figure 12D). To evaluate whether an increased size of the graft was associated with an improved clonality of the edited cells in the presence of GSE56 treatment, we evaluated the number of dominant clones contributing to repopulation of the HDR-edited cells by means of our novel barcoding-based clonal tracking analysis. Indeed, we observed a 6.5-fold higher number of dominant clones in haematopoietic organs (Figure 12E), which directly correlates with the yield of human edited cells measured by FACS (Figure 12F). Moreover, we also performed a similar transplantation experiment by targeting the clinically-relevant *CD40LG* gene which is associated to Hyper IgM syndrome (Figure 13A). We consistently found a higher percentage of human cells in the peripheral blood of mice transplanted with *CD40LG*-edited cells in the presence of GSE56 (Figure 13B), associated with a more stable percentage of gene correction over time (Figure 13C). At the endpoint of the experiment, the percentage of human cells tended to be improved in all the haematopoietic tissues and organs analysed (Figure 13D), with 3- to 6-fold higher percentage of corrected cells in sorted cell types, included $CD34^+$ cells (Figure 13E). Overall, these results confirm and extend those reported in Figure 3 and 4, highlighting the relevant role of transient p53-inhibition to robustly improve the yield of edited cells *in vivo*.

We continued the investigation concerning the potential of adenoviral Ad5-E4orf1 and Ad5-E4orf6/7 to boost HDR efficiency in HSPCs. We performed the targeted gene editing protocol in human cord blood (CB)-derived $CD34^+$ after pre-stimulation with cytokines and drugs (SR1, dmPGE-2 and UM171) and by electroporating at day 3 the pre-assembled RNP-Cas9 followed 15 min later by AAV6 transduction, as our standard editing procedure (RNP + AAV6). We tested side by side the transient delivery of GFP (as mRNA control), GSE56, Ad5-E4orf1, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 by exploiting mRNA machinery and we characterised various parameters for each condition within the days following the

editing protocol (Figure 14A). We first observed that the standard editing protocol RNP + AAV6 reduces cell expansion compared to untreated cells (UT). The Ad5-E4orf6/7 had similar trend than standard while both GSE56 and Ad5-E4orf1 increased about 2-fold more the cell growth that correlates with the results in Figures 3D and 7D. The combination of GSE56/Ad5-E4orf6/7 shown an intermediate pattern until day 7 post-editing and reached the level of both GSE56 and Ad5-E4orf1 at day 10 (Figure 14B). Editing efficiency was evaluated by measuring GFP expression within HSPCs subpopulations (Figure 14C left and right panels). GSE56 shown 1.1-fold increase of GFP expression compared to standard RNP + AAV6, especially in the most primitive compartment ($CD34^+133^+90^+$) which is slight but significant as described in Figure 3C, right panel. Similarly, the Ad5-E4orf1 or Ad5-E4orf6/7 ± GSE56 shown respectively up to 1.3-fold and 1.5-fold increase of GFP expression into the $CD90^+$ cells. We validated these results by looking at promoted HDR-mediated integrations at the targeted *AAVS1* locus by digital PCR in bulk culture (Figure 14D). HDR edited alleles were significantly increased using the Ad5-E4orf6/7 ± GSE56 while NHEJ fraction remained similar whatever the condition. Moreover, we observed a significant skewing of $CD90^+$ fraction only by adding the Ad5-E4orf6/7 ± GSE56 that could suggest a possible toxic effect or cells differentiation (Figure 14E). We then performed a 7-AAD/annexin staining to evaluate the toxicity of our editing protocols. While in the standard we saw a reduced number of live cells compared to UT, we did not observe an extra-toxicity by adding RNAs (Figure 14F). Interestingly, by performing CFU-C assay, erythroid and myeloid clonogenic output was not affected with the delivery of GFP as mRNA control or the Ad5-E4orf6/7 compared to standard protocol. It was even improved with GSE56 as described in Figure 3F and Ad5-E4orf1. However, the use of Ad5-E4orf6/7 ± GSE56 seemed to negatively affect clonogenic properties, even it remained a trend and it was not significant that could correlate with the reduced $CD90^+$ fraction seen in culture composition (Figure 14G). Finally, we evaluated cell cycle states of HSPCs upon editing protocols. The use of Ad5-E4orf6/7 clearly shown a significant transition from G1 phase to S/G2/M phases of bulk treated cells mainly amplified by combining it with GSE56. We observed a similar pattern by looking into the most primitive compartment. This effect was lower and not significant by using GSE56 or Ad5-E4orf1 (Figure 14H). To confirm previous results obtained by targeting the bi-allelic *AAVS1* locus, we also developed the editing strategy by targeting another genomic mono-allelic *IL2RG* locus. By measuring the GFP expression within HSPCs subpopulations, we still had a significant editing efficiency increase by using the Ad5-E4orf6/7 ± GSE56, even the fold was reduced compared to the standard RNP + AAV6. We could speculate this is due to the mono-allelic form of this locus (Figure 14I). Furthermore, by considering that the Ad5-E4orf6/7 is pushing the cell cycle, we reviewed the timing of the editing protocol during *ex-vivo* culture to further preserved the HSCPs properties. We

evaluated different time points to proceed with the editing protocol from day 0 to day 2 post-thawing (Figure 14J). Interestingly, we observed that both in total bulk and $CD90^+$ cells fraction, the editing efficiency was increased in presence of Ad5-E4orf6/7 from day 0 to day 2 (Figure 14K). We confirmed these results by looking at promoted HDR-mediated integrations at the targeted *AAVS1* locus by digital PCR in bulk culture (Figure 14L). We performed a similar characterisation with Ad5-E1B55K and Ad5-E4orf6 proteins (Figure 15A, B, C, D and E). The combination of these two proteins increased significantly the GFP expression from the AAV6 donor as previously described (Gwiazda, K.S. et al. (2016) Mol. Ther. 24: 1–11), while it was reducing HDR-mediated integrations at the targeted *AAVS1* locus compared to standard procedure, as we already observed (Figure 8).

We demonstrated that Ad5-E4orf6/7 induced cell cycle progression in HSPCs by cell cycle analysis. In 1989, Huang *et al.* (Huang, M.M. et al. (1989) Genes Dev. 3: 1699–1710) described the direct interaction between Ad5-E4orf6/7 and the E2Fs transcription factor family. We decided to explore E2Fs related gene expression at the RNA level in HSPCs. We selected specific gene targets known to be involved in cell cycle regulation and p53-related. We used TaqMan gene expression assays to quantify specific transcripts upon several editing protocols. *CDK2* is known to be a cell cycle progression marker which is specific for transition from G1 to S/G2 phases. We noted that *CDK2* was significantly up-regulated in presence of Ad5-E4orf6/7 ± GSE56 at 24 hours post-editing (Figure 16A) and it was detectable in a time window from 12 to 24 hours before reducing at 96 hours (Figure 16B), while in case of RNP + AAV6 ± GSE56 remained always at basal level. *CDKN2A* is known as "$p14^{ARF}$" and has been described to be a p53 modulator by inhibiting MDM2, its main repressor. Again, we saw that *CDKN2A* was significantly up-regulated in the presence of Ad5-E4orf6/7 ± GSE56 at 24 hours post-editing (Figure 16C) and similarly to *CDK2*, it was detectable in a time window from 12 to 24 hours before reducing at 96 hours (Figure 16D), while RNP + AAV6 ± GSE56 still remained at basal level. We then investigated the p53 downstream effectors as *CDKN1A* and *APOBEC3H*, that are possibly activated by E2Fs. The first one is known as "p21" and it is a direct target of p53 involved in cell cycle arrest. Indeed, *CDKN1A* was highly up-regulated in the standard protocol RNP + AAV6 at 24 hours post-editing, which is a direct consequence of the editing protocol (Figure 16E). The use of GSE56, which is a p53 inhibitor, allowed to reduce the amplitude of p21 activation about 3-fold less compared to RNP + AAV6, as described in Figure 3B. However, by using the Ad5-E4orf6/7 we noted a drop of *CDKN1A* expression at 12 hours post-editing about 2-fold higher compared to the standard. Interestingly, this drop was not detectable when combining GSE56/Ad5-E4orf6/7, probably due to GSE56 properties that counterbalances the p21 overexpression mediated through Ad5-E4orf6/7 (Figure 16F). In the end, *CDKN1A* went

back to basal level at 168 hours post-editing. We finally tested *APOBEC3H* which is involved in anti-viral cell response by inducing C to T mutations in viral genomes. *APOBEC3H* is highly up-regulated upon editing protocols mainly due to AAV6 transduction in HSPCs (Figure 3B). However, it was significantly down-regulated in presence of Ad5-E4orf6/7 and/or GSE56 at 24 hours post-editing (Figure 16G). However, we noted that *APOBEC3H* went up at 96 hours before reducing at 168 hours (Figure 16H). To summarise, we demonstrated that the interaction between Ad5-E4orf6/7 and E2Fs is the mechanism involved in HSPCs cell cycle progression and the increase of HDR events upon editing protocol. We also highlighted the presence of p53-mediated feedback loop that allows the cell cycle to be controlled. However, by blocking transiently the p53 activity, we were able to further push the cell cycle progression (Figure 16I), especially into the most primitive $CD90^+$ cells.

We then planned a first transplantation by using a high number of treated cells ($T0_{eq}$ = $3x10^5$/mouse) into an immunodeficient NSG mouse model at 24 hours post-editing (Figure 17A). Human $CD45^+$ cell engraftment was monitored in bloodstream over the weeks post-transplantation and did not show any statistical significance between several groups RNP + AAV6 ± Ad5-E4orf1, Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 that is probably due to the saturation of engraftment capacity into these recipients (Figure 17B). Interestingly, the fraction of HDR-edited GFP^+ cells slightly decreased during time as described in Figure 4D, that can be due to the loss of targeted progenitors. However, we noted that protocols RNP + AAV6 + Ad5-E4orf6/7 or GSE56/Ad5-E4orf6/7 increased respectively 1.5 and 2-fold the fraction of circulating edited cells. Editing by using the Ad5-E4orf1 remained similar to the standard pattern (Figure 17C). We confirmed these results at the end of experiments by analysing mice spleen and bone-marrow organs (Figure 17D and E). Moreover, we performed clonal tracking by using barcode sequences within our AAV6 cassette. In line with primary transplant, we detected up to 3-fold more edited clones by treating cells with GSE56/Ad5-E4orf6/7 compared to the standard (Figure 17F).

To further investigate the most primitive cells' long-term repopulation capacity, we decided to perform a secondary transplant by collecting and purifying $CD34^+$ cells from bone-marrow of primary recipients. In order to avoid mouse-dependent heterogenicity, we transplanted around $2x10^6$ cells/mouse from the purify $CD34^+$ pool inside a same group and by considering two primary recipients for one secondary recipient (Figure 18A). While the engraftment was quite high and similar among each group, we still observed significant differences in terms of circulating edited human cells, especially comparing the standard RNP + AAV6 ± GSE56/Ad5-E4orf6/7 with more than 2-fold increase of GFP expressing cells

(Figure 18B and C). This was also true in all the compartments that we analysed (bone marrow and spleen) as well as for all the blood lineages, $CD19^+$, $CD33^+$ and $CD3^+$ (Figure 18D, E and F). Overall, these results indicate that GSE56/Ad5-E4orf6/7 allows significant increase in the fraction of edited cells in transplanted mouse model by inducing cell cycle progression during editing protocol and probably reducing the p53-dependent transcriptional response that is helping to better preserve the long-term function of the most primitive HSPC compartment.

We performed a limit cells dose transplantation ($T0_{eq}$ = $1x10^5$/mouse) in order to further elucidate this last point, and including a new group treated with GSE56 in parallel. In these settings, edited cell fractions were similar to the previous experiment, but we better appreciated differences in terms of engraftment capacity between conditions (Figure 19A and B). As expected, GSE56 gave rise to 5-6-fold higher engraftment of treated cells compared to standard protocol (also see Figure 4B). Curiously, the combination of GSE56/Ad5-E4orf6/7 generated about 2-fold higher engraftment compared to standard, but it is still lower compared to GSE56 treatment alone. We could then speculate two hypotheses: in one hand, we know that Ad5-E4orf6/7 induces a transient p53 overexpression during editing protocol, that could even more impair the treated cells. It would require then higher GSE56 dose to counterbalance it. In another hand, we know that the use of p53 inhibitor induces the loss of the feedback loop that allows control of the cell cycle, and which is pushed up by the Ad5-E4orf6/7. Then we could wonder that the stem cells are more prone to performed HDR but as counterpart are also affected about long-term engraftment capacity in mice.

To reinforce previous data, we decided to perform another *in vivo* experiment by targeting *IL2RG* locus and transplanting NSG mice similar to Figure 17 ($T0_{eq}$ = $3x10^5$/mouse) by comparing RNP + AAV6 ± GSE56/Ad5-E4orf6/7. We confirmed then previous results both after primary and secondary transplantation in NSG mice. We observed 2/3-fold increase of edited cells within organs and blood lineages even a secondary transplant, by using the combination GSE56/Ad5-E4orf6/7 compared to the standard protocol (Figure 20A, B, C, D, E, F, G and H). To assess potential genotoxicity of the editing procedure by complementing with both GSE56 and Ad5-E4orf6/7, we performed a translocations analysis by targeting the *IL2RG* locus with our reagents, where we identified one major off-target. Interestingly, we did not observe an increase of genome translocation events frequency by ddPCR compared to the standard protocol. Moreover, for *in vivo* samples, no event was detectable, suggesting that differentiated cells are probably more prone to this phenomenon (Figure 20I).

Finally, we also investigated the use of adenoviral proteins variants for both E4orf1 and E4orf6/7 derived from various serotypes of Adenovirus family. We performed protein sequences alignment by using T-coffee algorithm, and we identified specific conserved amino-acids (>95% consensus). We selected the best candidates presenting highest sequence variabilities along the phylogenetic tree of common ancestors. We then choose three variants for each protein, and we evaluated editing efficiency in HSPCs (Figure 21/22A and B). E4orf1 variants shown up to 2-fold increase of GFP expression within HSPCs subpopulations comparing to the Ad5-E4orf1 (Figure 21C). However, Ad5-E4orf1 remained the one that presented the highest number of colonies (Figure 21D). Surprisingly, TaqMan analyses did not show any impact on E2Fs target genes whatever the E4orf1 variant, suggesting an alternative mechanism to stimulate the cell cycle activity of treated cells (Figure 21E). Concerning E4orf6/7, only one variant was increasing up to 2-fold the editing efficiency into the most primitive compartment, while two others remained similar to the Ad5-E4orf6/7 (Figure 22C). However, by looking at colony forming cell assay, all these variants seemed impaired in the clonogenic capacity of treated cells (Figure 22D). At RNA level, it has been revealed that these variants drastically induced feedback loop response both *CDKN1A* and *CDKN2A* revealing a kind of toxicity compared to the use of Ad5-E4orf6/7 (Figure 22E).

Discussion

We demonstrated that transient expression of both GSE56 and Ad5-E4orf6/7 is able to improve our standard targeted gene editing protocol in HSPCs. In one hand, GSE56 reduces edited cells impairment during the editing protocol by modulation of the p53 response. We observed a higher engraftment capacities post-transplantation and an increased number of abundant clones in NSG mice. Despite of consistent but low increase in terms of HDR efficiency, it revealed that edited cells fraction was more stable at long-term than standard protocol, suggesting that most primitive cells suffered less during the editing protocol thanks to GSE56. Moreover, the transient inhibition of p53 probably wouldn't select naturally p53 deficient HSPCs, avoiding the carry-over of putative tumor-prone clones under transplantation. In the other hand, we discovered that Ad5-E4orf6/7 adenoviral protein was able to push up the cell cycle progression leading to 1.5-fold HDR-events integrations increase in *AAVS1* locus of HSPCs. Combined with GSE56, it was possible to directly target the most primitive cells bypassing the feedback loop p53-mediated. Consequently, we were able to increase the long-term edited HSCs fraction *in-vivo* up to 50% and the number of total abundant clones, even affecting partially the engraftment capacities in this specific case compared to the use of GSE56 alone. This last point could be explained by cell cycle

pushing that would induce a kind of cell differentiation into the most primitive compartment. We supposed then that by modulating both GSE56 and Ad5-E4orf6/7 doses, we could balance the HDR vs ENGRAFTMENT capacities.

To summarise, we could distinguish two strategies: in one hand by using the editing protocol with GSE56 alone, we will obtain a higher engraftment of treated cells and long-term stability of the small fraction of both homed and circulating edited cells *in-vivo*. In the other hand, by using GSE56/Ad5-E4orf6/7 we improved 2/3-fold the percentage of edited cells while the engraftment was reduced compared to the GSE56 alone. To conclude, it is known that recipient capacity is limited during transplantation and a competition between freshly transplanted cells is engaged during homing process. Thus, we can say that the fraction of edited cells is higher and as is the probability that they can reach the bone marrow niche overcoming the competition with non-edited cells. Finally, the combination of GSE56/Ad5-E4orf6/7 could be the best option for diseases in which a high level of edited cells is required.

EXAMPLE 4

Description

Whereas retro- and lenti-viral vector (LV)-based gene therapy enable clonal tracking studies by monitoring integration sites (Biffi et al. 2013), the site-specific and high-fidelity nature of HDR-mediated targeted insertion impede quantification and monitoring of edited clones. Beside few NHEJ-based tracking analyses (Wu et al. 2019), no data are available about the multilineage repopulation capacity of individual HDR-edited long-term repopulating HSPCs, as well as clonal diversity within the human graft. Therefore, the development of novel and extensive clonal tracking pipelines for gene edited cells would allow to: i) dissect repopulation dynamics; ii) compare clonal richness among different HSPC gene editing protocols, and iii) pave the way for tracking of edited HSPCs in translational settings. Previous works highlighted the value of introducing short random sequences (barcodes) into the genome of viral vectors to track the fate of barcoding-labeled cells upon vector integration into cellular genome (Naik et al. 2014). Integrating barcoded vectors has been previously applied in different fields, e.g. to study hematopoiesis (Verovskaya et al. 2013), T-cell behavior (Schepers et al. 2008) and breast tissue development (Nguyen et al. 2014). Recently, some reports highlighted the possibility of combining genome editing with barcoding technology in order to track the fate of CRISPR-edited *S. cerevisiae* clones (Roy et al. 2018).

Results and discussion

To enable clonal tracking of targeted HSPCs *in vitro* and *in vivo*, we cloned a 22-bp highly degenerated molecular barcode into an HDR-proficient AAV vector targeting the *AAVS1* locus (see Figure 23 and Methods). Theoretically, sequence design allows up to $4.53*10^8$ different barcodes. Since shorter inserts could be preferentially ligated into the backbone during cloning procedure and reduce library complexity, 3 fixed bases were used to avoid cutting during the cloning procedure of the degenerated insert by desired restriction enzymes (Bsu36I and SphI). Upon site-specific integration of the cassette into the *AAVS1* locus, the target sequence containing the barcode can be amplified by PCR and analyzed by deep sequencing to identify number and distribution of HDR-edited clones.

Potentially, this AAV vector/plasmid could be used to perform clonal tracking in whatever cell type (mouse/human HSPCs, T cells, etc.) for pre-clinical and clinical applications. Moreover, as well as the degenerated barcode sequence could be cloned in different type of vectors (short or long naked DNA fragments, IDLVs, etc.) or construct targeting other genomic region of interest (e.g. *IL2RG*, *CD40LG*, *RAG1*, *HBB,* etc.).

Methods

The barcoded AAV6 donor template targeting the AAVS1 locus was obtained by subcloning the degenerated sequence into the non-barcoded AAVS1 HDR donor template downstream of the polyadenylation sequence (Schiroli et al. 2017). Briefly, a ssODN embedding the random sequence flanked by the unique cloning restriction sites (Bsu36I and SphI) was purchased from Sigma Aldrich. The degenerated region was designed to avoid any undesired cutting by cloning restriction enzymes. To generate the complementary strand, 50 pmol of the ssODN were amplified with 10 cycle of PCR with Easy-A High-Fidelity enzyme (Agilent Technologies) using the appropriate forward and reverse primers according to manufacturer instruction. The amplified product was purified with MinElute PCR Purification kit (QIAGEN), digested with the restriction enzymes and checked by capillary electrophoresis. 2 ug of the Bsu36I/SphI-digested AAVS1 HDR donor template were ligated with the digested insert (molar ratio 1:7) using T4 DNA Ligase (NEB) by scaling up the manufacturer protocol. Finally, XL-10 Gold Ultracompetent Cells (Agilent Technologies) were transformed with the ligation product, plated and incubated for 12 hours of incubation at 30°C to minimize the risk of recombination. Colonies were scraped, mixed, grown in LB medium for additional 6 hours and processed with NucleoBond Xtra MaxiPrep according to manufacturer instruction. Ultimately, the plasmid prep was screened with restriction enzymes for ITR and plasmid integrity.

Sequences

Genomic insert of the barcoded AAV (underlined = barcoding sequence)

CCACTGAGAACCGGGCAGGTCACGCATCCCCCCCTTCCCTCCCACCCCCTGCCAAGCT
CTCCCTCCCAGGATCCTCTCTGGCTCCATCGTAAGCAAACCTTAGAGGTTCTGGCAAG
GAGAGAGATGGCTCCAGGAAATGGGGGTGTGTCACCAGATAAGGAATCTGCCTAACAG
GAGGTGGGGGTTAGACCCAATATCAGGAGACTAGGAAGGAGGAGGCCTAAGGATGGG
GCTTTTCTGTCACCAATCCTGTCCCTANNYNNNTNTNNNNRTNDNNNHHCCATAGAGCC
CACCGCATCCCCAGCATGCCTGCTATTGTCTTCCCAATCCTCCCCCTTGCTGTCCTGCC
CCACCCCACCCCCCAGAATAGAATGACACCTACTCAGACAATGCGATGCAATTTCCTCA
TTTTATTAGGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGG
GAGGGGCAAACAACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGG
GTTTAAACGGGCCCTCTAGACTCGACGCGGCCGCTTTACTTGTACAGCTCGTCCATGC
CGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCT
CGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCA
GCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACG
CTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTT
GTCGGCCATGATATAGACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGG
ATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGT
CGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTTGAAGAAGAT
GGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTT
CATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCAGGGTGGTCACGAG
GGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTT
GCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTAC
GTCGCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTT
GCTCACCATGGTGGCGACCGGTGGGGAGAGAGGTCGGTGATTCGGTCAACGAGGGAG
CCGACTGCCGACGTGCGCTCCGGAGGCTTGCAGAATGCGGAACACCGCGCGGGCAG
GAACAGGGCCCACACTACCGCCCCACACCCCGCCTCCCGCACCGCCCCTTCCCGGCC
GCTGCTCTCGGCGCGCCCTGCTGAGCAGCCGCTATTGGCCACAGCCCATCGCGGTCG
GCGCGCTGCCATTGCTCCCTGGCGCTGTCCGTCTGCGAGGGTACTAGTGAGACGTGC
GGCTTCCGTTTGTCACGTCCGGCACGCCGCGAACCGCAAGGAACCTTCCCGACTTAGG
GGCGGAGCAGGAAGCGTCGCCGGGGGGCCCACAAGGGTAGCGGCGAAGATCCGGGT
GACGCTGCGAACGGACGTGAAGAATGTGCGAGACCCAGGGTCGGCGCCGCTGCGTTT
CCCGGAACCACGCCCAGAGCAGCCGCGTCCCTGCGCAAACCCAGGGCTGCCTTGGAA
AAGGCGCAACCCCAACCCCGTGGAAGCTTGCGTGGGGTGGAGGGGACAGATAAAAGT

ACCCAGAACCAGAGCCACATTAACCGGCCCTGGGAATATAAGGTGGTCCCAGCTCGGG
GACACAGGATCCCTGGAGGCAGCAAACATGCTGTCCTGAAGTGGACATAGGGGCCCG
GGTTGGAGGAAGAAGACTAGCTGAGCTCTCGGACCCCTGGAAGATGCCATGACAGGG
GGCTGGAAGAGCTAGCACAGACTAGAGAGGTAAGGGGGGTAGGGGAGCTGCCCAAAT
GAAAGGAGTGAGAGGTGACCCGAATCCACAGGAGAACGGGGTGTCCAGGCAAAGAAA
GCAAGAGGATGGAGAGGTGGCTAAAGCCAGGGAGACGGGGTACTTTGGGGTTGTCCA
GAAAAACGGTGATGATGCAGGCCTACAAGAAGGGGAGGCGGGACGCAAGGGAGACAT
CCGTCGGAGAAGGCCATCCTAAGAAACGAGAGATGGCACAGGCCCCAGAAGGAGAAG
GAAAAGGGAACCCAGCGAGTGAAGACGGCATGGGGTTGGGTGAGGGAGGAGAGATG
CCCGGAGAGGACCCAGACACGGGGAGGATCCGCTCAGAGGACATCACGTGGTGCAGC
GCCGAGAAGGAAGTGCTCCGGAAAGAGCATCCTTGGGCAGCAACACAGCAGAGAGCA
AGGGGAAGAGGGAGTGGAGGAAGACGGAACCTGAAGGAGGCGGCAGGGAAGGATCT
GGGCCAGCCGTAGAGGTGACCCAGGCCACAAGCTGCAGACAGAAAGCGGCACAGGC
CCAGGGGAGAGAATGCTGGTCAGAGAAAGCA

ssODN amplification primers

Bar *AAVS1* sense 5'-AGACTAGGAAGGAGGAGGCCTAAGGATG-3'

Bar *AAVS1* antisense 5'-ACAATAGCAGGCATGCTGGGGATG-3'

ssODN

AGACTAGGAAGGAGGAGGCCTAAGGATGGGGCTTTTCTGTCACCAATCCTGTCCCTAN
NYNNNTNTNNNNRTNDNNNHHCCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATT
GT

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

CLAIMS

1. An agent which promotes homology directed DNA repair for use in haematopoietic cell gene therapy, haematopoietic stem cell gene therapy, and/or haematopoietic progenitor cell gene therapy, wherein said gene therapy comprises gene editing.

2. An agent which promotes homology directed DNA repair, for use in increasing the survival and/or engraftment of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

3. Use of an agent which promotes homology directed DNA repair, for increasing the survival and/or engraftment of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells, or for increasing the efficiency of gene editing of haematopoietic cells, haematopoietic stem cells, and/or haematopoietic progenitor cells.

4. The agent for use according to claim 1 or claim 2 or the use of claim 3, wherein the agent is an inhibitor of p53 activation, preferably wherein the inhibitor is an inhibitor of p53 phosphorylation, more preferably an inhibitor of p53 Serine 15 phosphorylation.

5. The agent for use according to claim 4 or the use of claim 4, wherein the inhibitor is a p53 dominant negative peptide, an ataxia telangiectasia mutated (ATM) kinase inhibitor or an ataxia telangiectasia and Rad3-related protein (ATR) inhibitor.

6. The agent for use according to any one of claims 4 or 5 or the use of claim 4 or 5, wherein the inhibitor is pifithrin-α or a derivative thereof; KU-55933 or a derivative thereof; GSE56 or a variant thereof; KU-60019, BEZ235, wortmannin, CP-466722, Torin 2, CGK 733, KU-559403, AZD6738 or derivatives thereof; or an siRNA, shRNA, miRNA or antisense DNA/RNA.

7. The agent for use according to any one of claims 4 to 6 or the use of any one of claims 4 to 6, wherein the inhibition of p53 in the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells is transient.

8. The agent for use according to any one of claims 4 to 7 or the use of any one of claims 4 to 7, wherein the transient inhibition of p53 occurs during gene editing of the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

9. The agent for use according to any one of claims 4 to 8 or the use of any one of claims 4 to 8, wherein the inhibitor is added to the haematopoietic cell, haematopoietic stem cell and/or haematopoietic progenitor cells at a concentration of about 1-50 μM.

10. The agent for use according to any one of claims 1, 2 or 4-9 or the use of any one of claims 3-9, wherein the agent comprises a nucleic acid sequence encoding at least one adenoviral protein.

11. The agent for use according to claim 10 or the use of claim 10, wherein the at least one adenoviral protein is from an Adenovirus of serotype 4, Adenovirus of serotype 5, Adenovirus of serotype 7 and/or Adenovirus of serotype 9, preferably the at least one adenoviral protein is from an Adenovirus of serotype 5.

12. The agent for use according to claim 10 or claim 11 or the use of claim 10 or claim 11, wherein the at least one adenoviral protein is E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, 57-76.

13. The agent for use according to claim 10 or claim 11 or the use of claim 10 or claim 11, wherein the at least one adenoviral protein is E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, 77-107.

14. The agent for use according to any one of claims 1, 2 or 4-13 or the use of any one of claims 4-13, wherein the agent comprises: adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, 57-76; and adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, 77-107.

15. The agent for use according to any one of claims 1, 2 or 4-14 or the use of any one of claims 4-14, wherein the agent comprises:

(a) the inhibitor of p53 activation and an adenoviral protein;

(b) the inhibitor of p53 activation and adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, 57-76;

(c) the inhibitor of p53 activation and adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, 77-107; or

(d) the inhibitor of p53 activation, adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, 57-76, and adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, 77-107,

preferably wherein the inhibitor of p53 activation is GSE56 or a variant thereof.

16. The agent for use according to any one of claims 10 to 15 or the use of any one of claims 10 to 15, wherein the nucleic acid encoding the adenoviral protein is an mRNA.

17. The agent for use according to any one of claims 10 to 16 or the use of any one of claims 10 to 16, wherein the adenoviral protein is expressed transiently in the haematopoietic cell, haematopoietic stem cell or haematopoietic progenitor cell, preferably wherein the transient expression occurs during gene editing of the haematopoietic cell, haematopoietic stem cell and/or haematopoietic progenitor cell.

18. The agent for use according to any one of claims 1, 2 or 4-17 or the use of any one of claims 4-17, wherein the target of the gene editing is selected from the group consisting of CD40L, RAG-1, IL-2RG, CYBA, CYBB, NCF1, NCF2, and NCF4, preferably CD40L.

19. A method of gene editing a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells comprising the steps:

(a) contacting the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells with an agent which promotes homology directed DNA repair;

(b) introducing gene editing machinery to the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells with one or more vectors; and

(c) editing the genome of said haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells.

20. The method of claim 19, wherein steps (a) and (b) are carried out ex vivo or in vitro.

21. The method of claim 19 or 20, wherein the agent which promotes homology directed DNA repair is an inhibitor of p53 activation.

22. The method of claim 22, wherein the inhibitor of p53 activation is a p53 dominant negative peptide, an ataxia telangiectasia mutated (ATM) kinase inhibitor or an ataxia telangiectasia and Rad3-related protein (ATR) inhibitor.

23. The method of claim 22, wherein the inhibitor of p53 activation is pifithrin-α or a derivative thereof; KU-55933 or a derivative thereof; GSE56 or a variant thereof; KU-60019, BEZ235, wortmannin, CP-466722, Torin 2, CGK 733, KU-559403, AZD6738 or derivatives thereof; or an siRNA, shRNA, miRNA or antisense DNA/RNA.

24. The method of claim 23, wherein the inhibitor is added to the haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells at a concentration of about 1-50 µM.

25. The method according to any one of claims 19-24, wherein the agent comprises a nucleic acid sequence encoding at least one adenoviral protein.

26. The method according to claim 25, wherein the at least one adenoviral proteins is from an Adenovirus of serotype 5.

27. The method according to claim 25 or claim 26, wherein the at least one adenoviral protein is E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, 57-76.

28. The method according to claim 25 or claim 26, wherein the at least one adenoviral protein is E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, 77-107.

29. The method according to any one of claims 19-28, wherein the agent comprises: adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, 57-76; and adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, 77-107.

30. The method according to any one of claims 19-29, wherein the agent comprises:

(a) the inhibitor of p53 activation and adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, 57-76;

(b) the inhibitor of p53 activation and adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, 77-107; or

(c) the inhibitor of p53 activation, adenoviral protein E4ORF1, preferably wherein the amino acid sequence of E4ORF1 is set forth in SEQ ID No. 1, 57-76, and adenoviral protein E4ORF6/7, preferably wherein the amino acid sequence of E4ORF6/7 is set forth in SEQ ID No. 2, 77-107,preferably wherein the inhibitor of p53 activation is GSE56 or a variant thereof.

31. The method for use according to any one of claims 25-30, wherein the nucleic acid encoding the adenoviral protein is an mRNA.

32. The method according to any one of claims 25-31, wherein the adenoviral protein is expressed transiently in the haematopoietic cell, haematopoietic stem cell, and/or haematopoietic progenitor cell.

33. The method of any one of claims 19-32, wherein the population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells is contacted with the agent which promotes homology directed DNA repair concurrently with gene editing.

34. The method of any one of claims 19-33, wherein the population of haematopoietic stem and/or progenitor cells is obtained from mobilised peripheral blood, bone marrow or umbilical cord blood.

35. The method of any one of claims 19-34, which includes a further step of enriching the population for haematopoietic stem and/or progenitor cells.

36. The method of any one of claims 19-35, wherein the target of the gene editing is selected from the group consisting of CD40L, RAG-1, IL-2RG, CYBA, CYBB, NCF1, NCF2, and NCF4, preferably CD40L.

37. A method of gene therapy comprising the steps:

(a) gene editing a population of haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells according to the method of any one of claims 19-36; and

(b) administering the gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells to a subject.

38. The method of claim 37, wherein the gene edited cells are administered to a subject as part of an autologous stem cell transplant procedure or an allogeneic stem cell transplant procedure.

39. A population of gene edited haematopoietic cells, haematopoietic stem cells or haematopoietic progenitor cells prepared according to the method of any one of claims 19-38.

40. A pharmaceutical composition comprising the population of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells of claim 39.

41. The population of gene edited haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells of claim 40 for use in therapy.

42. The population of gene edited haematopoietic stem cells and/or haematopoietic progenitor cells for use according to claim 41, wherein the population is administered as part of an autologous stem cell transplant procedure or an allogeneic stem cell transplant procedure.

43. An AAV vector comprising a barcode sequence.

44. An AAV vector according to claim 43, wherein the AAV vector further comprises a nucleotide sequence encoding a gene of interest.

45. An AAV vector according to claim 44, wherein the gene of interest is operably linked to a promoter and/or a polyadenylation sequence, optionally wherein the barcode sequence is downstream of the polyadenylation sequence.

46. An AAV vector according to any one of claims 43 to 45, wherein the AAV vector is an AAV6 vector.

47. An AAV vector according to any one of claims 43 to 46, wherein the AAV vector targets a safe harbour locus, preferably the AAVS1 locus, intron 1 of IL2RG, CD40L or RAG-1, most preferably the AAVS1 locus.

48. An AAV vector according to any one of claims 43 to 47, wherein the barcode sequence is a degenerated and/or random sequence, preferably a random sequence.

49. An AAV vector according to any one of claims 43 to 48, wherein the barcode sequence is 20-40 bp in length, preferably 20-30 bp in length, most preferably 20-25 bp in length or about 22 bp in length.

50. An AAV vector according to any one of claims 43 to 49, for use in gene therapy, preferably haematopoietic gene therapy.

51. Use of an AAV vector according to any one of claims 43 to 48, for clonal tracking of cells that have been gene-edited using said AAV vector.

52. A method for clonal tracking of cells comprising the steps:

(a) providing an isolated population of cells;

(b) gene editing the isolated population of cells by transducing with an AAV vector according to any one of claims 43 to 49; and

(c) clonal tracking of the gene-edited cells using the barcode sequence of the AAV vector.

53. A method according to claim 52, wherein the method further comprises the steps:

expanding the isolated population of gene-edited cells; and/or

transplanting the isolated population of gene-edited cells into a subject.

54. A method according to claim 52 or claim 53, wherein the method further comprises:

harvesting blood and/or tissue samples from the subject;

isolating genomic DNA from the blood and/or tissue samples; and

quantifying the barcode sequences in the genomic DNA.

55. Use according to claim 51 or a method according to any one of claims 52 to 54, wherein the cells are haematopoietic cells, haematopoietic stem cells and/or haematopoietic progenitor cells

56. Use according to claim 51 or claim 55, or a method according to any one of claims 52 to 55, wherein the cells are human cells and/or the subject is human.

57. Use according to any one of claims 51, 55 or 56, or a method according to any one of claims 52 to 56, wherein the cells are CD34+ cells.

58. A method according to any one of claims 54 to 57, wherein blood and/or tissue samples are harvested every day, every week, every two weeks, or every month.

59. A method according to any one of claims 54 to 58, wherein blood and/or tissue samples are harvested after at least 1, 2, 3, 4, 8, 12, 24, 36 or 48 weeks.

60. A method according to any one of claims 54 to 59, wherein the sample is a blood sample.

61. A method according to any one of claims 54 to 60, wherein quantification of the barcodes comprises the steps:

(i) PCR amplification of the barcode sequences

(ii) sequencing the PCR-amplified barcode sequences.

62. Use of a barcode sequence for clonal tracking of cells.

63. Use according to claim 62, wherein the barcode sequence is a degenerated and/or random sequence, preferably a random sequence.

64. Use according to claim 62 or claim 63, wherein the barcode sequence is 20-40 bp in length, preferably 20-30 bp in length, most preferably 20-25 bp in length or about 22 bp in length.