Proceedings of the National Academy of Sciences of the United States of America **98**, 8923-8924 (2001) and Tempel S, Nicolas J, El Amrani A, Couee I. Model-based identification of Helitrons results in a new classification of their families in Arabidopsis thaliana. *Gene* **403**, 18-28 (2007)) but due to the lack of an active *Helitron* transposon that can replicate in cells, both the process and regulation of *Helitron* transposition has remained enigmatic. Everything that is known to date about *Helitron* biology derives from *in silico* or genetic analysis, because no active *Helitron* transposon, as defined by the active transposase enzyme in conjunction with functional terminal sequences enabling the replication of the intervening genomic content, has previously been isolated.

Summary

As described herein, the present disclosure relates to the resurrection of an active copy of the autonomous *Helibat1* transposon, herein designated "*Helraiser*", and characterization of its transposition *in vitro* and in human cells *ex vivo*.

Helraiser, an ancient element from the bat genome, has been used as an experimental tool to unravel the mechanism of *Helitron* transposition. A hairpin close to the 3'-end of the transposon functions as a transposition terminator. However, the 3'-end can be bypassed by the transposase, resulting in transduction of flanking sequences to new genomic locations. *Helraiser* transposition generates covalently closed circular intermediates, suggestive of a replicative transposition mechanism, which provides a powerful means to disseminate captured transcriptional regulatory signals across the genome. The generation of novel transcripts by *Helitron* promoter capture, both experimentally and by transcriptome analysis in bats, is described. These results provide mechanistic insight into *Helitron* transposition, and its impact on diversification of gene function by genome shuffling as well as provide experimental insight into the molecular requirements of transposition, target site selection properties, and gene capture in cell culture and in bats *in vivo*.

Importantly, the *Helraiser* transposase is capable of catalyzing DNA transposition *in trans* when used in conjunction with donor DNA flanked by appropriate sequences.

This system can be used to introduce single or multiple copies of the donor DNA into the genome of a cell. What sets *Helraiser* apart from other transposon systems (eg., *Sleeping Beauty*, *PiggyBac*) is that it uses a *copy/paste* mechanism for duplication rather than a *cut/paste* mechanism characteristic of these other systems, meaning that a multiplicative effect of duplication/replication can be achieved using as little as a single copy of the mobilizable element present to start. This has applications in engineering cell lines to carry multiple copies of preferred regions. It also allows the stepwise amplification of the DNA cargo in the recipient cell by additional rounds of transposition.

In one aspect of the present disclosure, there is provided a method for introducing single or multiple copies of a gene of interest into a cell comprising providing:

a) a *Helitron* transposase; and
b) a construct comprising a *Helitron* transposase LTS sequence.

In another aspect of the present disclosure, there is provided a method for introducing single or multiple copies of a gene of interest into a eukaryotic cell comprising providing:

a) a "copy and paste" transposase; and
b) a construct comprising a gene of interest flanked by a "copy and paste" transposon LTS sequence.

In a further aspect of the present disclosure, there is provided a cell line produced by a method as described herein.

In another aspect of the present disclosure, there is provided a cell line as described herein for use as a reference standard.

In a further aspect of the present disclosure, there is provided a cell line described herein for use in production of a protein of interest.

In another aspect of the present disclosure, there is provided a cell line described herein for use in therapy.

In a further aspect of the present disclosure, there is provided an isolated amino acid sequence comprising an amino acid sequence with 80% identity to the sequence set out in SEQ ID NO:1 wherein said sequence encodes a *Helraiser* transposase.

In another aspect of the present disclosure, there is provided an isolated nucleic acid sequence comprising a nucleic acid sequence encoding the amino acid sequence described herein.

In a further aspect of the present disclosure, there is provided an isolated nucleic acid sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence with at least 80% identity thereto.

In another aspect of the present disclosure, there is provided an isolated nucleic acid sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 4 or a sequence with at least 80% identity thereto.

In a further aspect of the present disclosure, there is provided an isolated nucleic acid sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence with at least 80% identity thereto and a nucleic acid sequence as set out in SEQ ID NO: 4 or a sequence with at least 80% identity thereto.

In another aspect of the present disclosure, there is provided an isolated nucleic acid sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 5 or a sequence with at least 80% identity thereto.

In a further aspect of the present disclosure, there is provided an isolated nucleic acid comprising a nucleic acid sequence of a gene of interest flanked by at least an LTS HelRaiser terminal sequence comprising the sequence set out in SEQ ID NO: 3 or a sequence having 80% identity thereto.

In another aspect of the present disclosure, there is provided an expression vector comprising the nucleic acid sequence described herein.

In a further aspect of the present disclosure, there is provided an expression vector comprising the nucleic acid sequence described herein further comprising at least one of:

a) a generic gRNA recognition site, preferably Tia1L, flanking the LTS and RTS;
b) a promoter sequence arranged such that the gene of interest is under the control of said promoter;
c) a polyadenylation cassette following said gene of interest.

In another aspect of the present disclosure, there is provided a recombinant host cell comprising the nucleic acid sequence described herein or an expression vector described herein.

In a further aspect of the present disclosure, there is provided a method of production of a protein of interest comprising culturing a cell described herein in a suitable medium and harvesting the protein of interest from the cell or suitable medium.

In another aspect of the present disclosure, there is provided a use of a gene of interest in the manufacture of a medicament for treating a disease in a subject in need thereof, wherein said medicament is to be provided to said subject, and wherein:

a) a cell line suitable for using in said subject is isolated;

b) an isolated nucleic acid described herein, or an expression vector described herein is introduced into said cell line wherein said nucleic acid or expression vector comprises a gene of interest corresponding to said protein of interest;

c) an amino acid sequence described herein, a nucleic acid sequence described herein or an expression vector described herein is introduced such that a *Helraiser* transposition event occurs to generate an engineered cell line comprising said gene of interest;

d) said engineered cell line is expanded in cell culture to provide a population of engineered cells; and

e) said engineered cells is to be introduced into said subject.

In a further aspect of the present disclosure, there is provided a gene of interest for use in treating a disease in a subject in need thereof, wherein said gene of interest is to be provided to said subject, and wherein:

a) a cell line suitable for using in said subject is isolated;

b) an isolated nucleic acid described herein, or an expression vector described herein is introduced into said cell line wherein said nucleic acid or expression vector comprises a gene of interest corresponding to said protein of interest;

c) an amino acid sequence described herein, a nucleic acid sequence described herein or an expression vector described herein is introduced such that a Helraiser transposition event occurs to generate an engineered cell line comprising said gene of interest;

d) said engineered cell line is expanded in cell culture to provide a population of engineered cells; and

e) said engineered cells is to be introduced into said subject.

In another aspect of the present disclosure, there is provided a use of an isolated nucleic acid described herein, and a nucleic acid sequence described herein, in the manufacture of a medicament for treating a disease in a subject in need thereof, wherein:

a) a first expression vector comprising the isolated nucleic acid is provided;

b) a second expression vector comprising the nucleic acid sequence is provided;

c) said first and second expression vectors is to be introduced into said subject.

In a further aspect of the present disclosure, there is provided an isolated nucleic acid described herein, and a nucleic acid sequence described herein for use in treating a disease in a subject in need thereof, wherein:

a) a first expression vector comprising the isolated nucleic acid is provided;

b) a second expression vector comprising the nucleic acid sequence is provided;

c) said first and second expression vectors is to be introduced into said subject.

In another aspect of the present disclosure, there is provided a pharmaceutical composition comprising a first expression vector comprising an isolated nucleic acid a described herein and a second expression vector comprising a nucleic acid sequence described herein.

In a further aspect of the present disclosure, there is provided a use of a Helitron transposase together with a donor encoding a reporter gene, flanked by LTS and /or RTS, to generate a library of cell lines containing various genomic integration events of the reporter.

In another aspect of the present disclosure, there is provided a method for detecting a cell line derived from a Helitron transposition method comprising analysing said cell line for the presence of an LTS and/or RTS sequence described herein.

In a further aspect of the present disclosure, there is provided a method for generating a cell line comprising:

a) providing a construct comprising a Helitron LTS sequence; and

b) introducing said construct into a cell line.

In another aspect of the present disclosure, there is provided a cell line produced by a method described herein.

In a further aspect of the present disclosure, there is provided a cell line produced by a method described herein.

In another aspect of the present disclosure, there is provided a method for producing a cell line comprising multiple copies of a DNA sequence of interest comprising:

a) taking a cell line described herein;

b) introducing a Helitron transposase under conditions for transposase activity;

c) isolating clonal cell lines bearing multiple copies of said DNA sequence.

In a further aspect of the present disclosure, there is provided an isolated clonal cell line produced by a method described herein.

In another aspect of the present disclosure, there is provided an isolated clonal cell described herein for use as a reference standard.

In a further aspect of the present disclosure, there is provided a use of a copy and paste transposon in a eukaryotic cell to generate a cell having single or multiple copies of a DNA sequence.

In another aspect of the present disclosure, there is provided a use of a Helitron transposon in a prokaryotic or a eukaryotic cell to generate a cell having single or multiple copies of a DNA sequence.

There is also provided a system for generating single or multiple copies of a DNA sequence in an isolated or cultured cell comprising a copy/paste transposase and a donor DNA recognized by the transposase. Suitably the "DNA sequence" is a DNA sequence of interest which may be a "gene of interest" (i.e. encoding a protein of interest) wherein said term also includes a genomic region (e.g. a region of the genome comprising an enhancer, repressor, CpG island, etc.) . The term "donor DNA" as used herein refers to a gene of interest, or genomic region, provided in a construct, e.g. an expression vector, optionally arranged with suitable upstream and/or downstream terminal sequences which enable a transposase-mediated transposition event to occur.

There is also provided a method for introducing a single or multiple copies of a DNA sequence or gene of interest into a cell comprising providing: a) a *Helitron* transposase; and b) a contruct comprising a *Helitron* tranposase LTS sequence. In one embodiment, the construct in b) further comprises a DNA sequence or gene of interest flanked by a Helraiser LTS sequence. Suitably, the cell may be prokaryotic or eukaryotic.

There is further provided a method for introducing single or multiple copies of a DNA sequence or gene of interest into a eukaryotic cell comprising providing: a) a "copy and paste" transposase; and b) a construct comprising a DNA sequence or gene of interest flanked by a "copy and paste" transposon terminal sequence, such as an LTS or RTS. Suitable "copy and paste" transposases include transposases of the Helitron family including the *Helraiser* transposase as described herein. Suitable Helitron family transposases are described, for example in Kapitonov and Jurka (2007) and Thomas et al. (2015)). Suitable LTS and RTS are those identified to complement a particular copy and paste transposon such as a *Helitron*, In one embodiment, the "copy and paste" transposase is a *Helitron* transposase and the LTS is derived from a *Helitron* transposon. Suitably, the DNA sequence or gene of interest is introduced into the genome of the cell.

Advantageously, the "copy/paste" transposase is an active element capable of transposition. In one embodiment, the "copy/paste" transposase is not a prokaryotic, such as a bacterial, "copy/paste" transposon. Suitably the "copy/paste" transposase has been resurrected from a eukaryotic genome.

In one embodiment, the transposon terminal sequence is a *Helraiser* LTS sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence having at least 80% identity thereto.

Advantageously, inserting an LTS sequence alone allows the simple generation of a transposon donor sequence where the downstream genomic content can be subjected to transposition, as initiated by the addition of the transposase. This is of particular advantage where the sequential introduction of LTS and RTS is cumbersome, e.g. when one aims to amplify genomic content in eukaryotic cells and LTS and/or RTS have to be introduced by

conventional genome engineering technologies, such those based on CRISPR/Cas, TALENs, Zn finger nucleases or Meganucleases.

In another embodiment, the gene of interest is also flanked by an RTS sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 4 or a sequence having at least 80% identity thereto. As described herein, (see Fig. 3A, for example) a single LTS sequence may be sufficient to trigger transposition of downstream DNA cargo by *Helraiser* transposase, but transposition rates are higher if the desired gene or genomic region is flanked with both LTS and RTS sequences. Advantageously, the addition of the RTS helps define where transposase activity should terminate thus providing more controlled transposition. In one embodiment, a mutated RTS may be used preferably such that a gene of interest may be copied and integrated but not mobilised by the transposase.

Suitably a donor DNA comprises a DNA sequence, such as a DNA sequence for a target gene e.g. wherein a target gene encodes a protein of interest, positioned between right and left terminal sequences (RTS and LTS sequences). In one embodiment, the RTS has a nucleic acid sequence comprising CTAG at the 3' end while the LTS has a nucleic acid sequence comprising TC at the 5' end. Suitably the LTS comprises or has a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence with at least 80% identity thereto, while the RTS has a nucleic acid sequence as set out in SEQ ID NO: 4, or a sequence with at least 80% identity thereto.

There is also provided an *in vitro* system for generating single or multiple copies of a DNA sequence in an isolated or cultured mammalian cell comprising a copy/paste transposase of eukaryotic e.g. mammalian origin and a donor DNA recognized by the transposase.

Suitably a "copy/paste" transposase is a *Helitron* transposase. *Helitron* (or rolling-circle) transposons and transposases are described, for example in Kapitonov *et al.* (2007). In one embodiment the transposase is a *Helitron* transposase having an amino acid sequence with at least 80% sequence identity with the amino acid sequence set out in Seq ID NO: 1 (see Table 6: Table of SEQ ID NOS; SEQ ID NO: 1 is the amino acid sequence of *Helraiser* transposase). In another embodiment, the *Helitron* transposase is one having at least 80, 85, 90, 95 or 95% identity with the amino acid sequence set out in SEQ ID NO: 1. The *Helitron* transposase may be provided in DNA, RNA or protein form. In one embodiment, the transposase is *Helraiser* transposase. Suitably, the transposase is derived from or has the amino acid sequence as set out in SEQ ID NO: 1. In one embodiment, the *Helraiser* transposase is encoded by the nucleic acid sequence set out in SEQ ID NO: 2 (see Table 6: Table of SEQ ID NOS; SEQ ID NO: 2 is a nucleic acid sequence encoding the *Helraiser* transposase, see also Figure 8) or a codon-optimised version thereof such as, for example, the nucleic acid sequence set out in SEQ ID NO: 6.

In one embodiment, the transposase and the construct comprising the gene of interest are provided as two separate entities. Suitably entities may be DNA constructs such as expression vectors or plasmids although it is also envisaged that the transposase may be provided as a naked DNA, mRNA or as a protein. In one embodiment, the construct comprising the gene or genomic region of interest flanked by the *Helraiser* terminal sequence(s) may be created or present within a cell line which is subsequently transfected with a construct encoding a transposase. In another embodiment, the construct comprising the gene of interest flanked by the *Helraiser* terminal sequence(s) may be part of a separate plasmid for co-transfection with the construct encoding a transposase. In another embodiment, the transposase, gene of interest and LTS are provided as a transposon in a single construct.

In one embodiment, the transposase may be encoded by a *Helitron* transposon. For example, the transposase may be encoded by a transposon nucleic acid sequence such as the *Helraiser* transposon nucleic acid sequence as set out in SEQ ID NO: 5 (See Table 6: TABLE OF SEQ ID NOS; See also Figure 8). In accordance with the present disclosure, there is provided a Helitron transposon. Accordingly, the present disclosure provides an isolated nucleic acid sequence as set out in SEQ ID NO: 5 or a sequence with at least 80% identity thereto. Suitably, there is provided a nucleic acid sequence at least 80, 85, 90, 95 or 95% identity to SEQ ID NO: 5, or a codon-optimised version thereof. In one embodiment, there is provided an isolated nucleic acid having the sequence set out in SEQ ID NO: 5.

In one embodiment, the target gene/DNA sequence of interest may be an endogenous gene. In another embodiment, the target DNA sequence may be a genomic region of interest, such as an enhancer, a repressor, a CpG island or any non-coding element of interest. In both of these embodiment, the method of the present disclosure may be used to generate multiple copies of the endogenous gene so as to generate a cell line which may be used as a reference cell line. In another embodiment, a gene of interest may be provided as a cDNA sequence.

The DNA sequence positioned between the LTS and RTS is preferably one which provides a DNA sequence of interest such as a DNA sequence encoding a protein of interest for expression in the cell system.

Advantageously, using a *Helitron* transposase system such as *Helraiser* as described herein facilitates the replication of donor DNA and its introduction into a multiplicity of sites within the genome. As such, the methods or system in accordance with the present disclosure may be used to introduce single or multiple copies of a target gene of interest.

In one embodiment, the cell or the isolated or cultured cell for use in the method of the present disclosure is a prokaryotic cell such as bacteria. In another embodiment, the cell may be a eukaryotic cell such as insect, yeast, plant or mammalian cell. Suitable cultured cells are familiar to those skilled in the art. In one embodiment, the cell is a mammalian cell such as a mouse, rat or human cell, such as a CHO cell, a 293T cell, a HEK293 cell, a human induced

pluripotent stem cell, a human or murine embryonic stem cell, a hematopoietic stem cell, a T cell or a B cell. Where the cell is a mammalian or human cell, it may be a cell for use in therapy. In another embodiment, the cell may be a cell type which can be used to generate a reference cell line, such as a tumour cell line, a HAP1 or eHAP cell, an HCT116 cell, a DLD-1 cell, a HEK293 cells and so forth. In another embodiment, the cell may be suitable for protein production. For example, the cell may be a HeLa, 293T cell, CHO cell or other suitable mammalian cell system. Suitable cell lines for mammalian protein production will be known to those skilled in the art and include CHO cells, HEK293 and 293T cells, for example.

In one embodiment, the method of the present disclosure may be used to generate a cell line. Such a cell line may be transient or stable.

One suitable system for *Helraiser* transposition is described herein in the Examples and with reference to Figure 1B.

In another embodiment, the nucleic acid encoding the transposase is integrated into the genome of the cell. In another embodiment, the donor DNA is part of a plasmid or a recombinant viral vector. In a further embodiment, the donor DNA comprises at least a portion of an open reading frame. In a yet further embodiment, the donor DNA comprises at least a regulatory region of a gene, such as a transcriptional regulatory region which may be selected from the group consisting of a promoter, an enhancer, a silencer, a locus-control region, and a border element. Suitably, the donor DNA comprises a promoter operably linked to at least a portion of an open reading frame.

The donor DNA and/or the construct comprising the transposase sequence may be introduced into the cell using a method selected from the group consisting of: particle bombardment; electroporation; microinjection; combining the nucleic acid fragment with lipid containing vesicles or DNA condensing reagents; and incorporating the nucleic acid fragment into a viral vector and contacting the viral vector with the cell. In one embodiment, the transposase may be introduced as an mRNA molecule.

There is also provided a method for introducing multiple copies of a DNA sequence into a genome whereby a *Helitron* transposase and donor DNA are introduced into a cell. In one embodiment, the transposase and donor DNA are supplied separately. In another embodiment, the transposase and donor DNA are supplied on the same DNA construct. Advantageously, where the transposase is supplied in a separate construct it may only be expressed and effective as long as it is present in the cell. This may enable transposition events to be limited, if desirable. In another embodiment, the transposase is introduced in RNA or protein form.

There is further provided a method for introducing multiple copies of a DNA sequence into a genome whereby a donor DNA is first introduced into the genome of a cell followed by introduction of a *Helitron* transposase.

There is also provided a method for introducing multiple copies of a DNA sequence into a genome whereby the RTS and LTS sequences flank an endogenous gene. The present disclosure also provides a method for introducing single or multiple copies of a DNA sequence or gene of interest into a cell by providing a construct comprising a DNA sequence flanked by a *Helraiser* LTS sequence. In one embodiment, the RTS has a nucleic acid sequence comprising CTAG at the 3' end while the LTS has a nucleic acid sequence comprising TC at the 5' end. Suitably the LTS has a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence with at least 80% identity thereto, while the RTS has a nucleic acid sequence as set out in SEQ ID NO: 4, or a sequence with at least 80% identity thereto. Suitably, the method comprises modifying the cell genome in order to introduce the RTS and LTS in such a way that they lie either side of the endogenous cell gene of interest such that the endogenous cell gene is targeted for multiplication by the transposase. In one embodiment, the RTS and/or LTS sequences are introduced using a genome targeting or engineering method. In one embodiment a genome editing method such as CRISPR, Zinc-finger nucleases (ZFNs), meganucleases, transcription activator-like effector nucleases (TALENs) (reviewed, for example by Gaj *et al.* in Trends in Biotechnology, Volume 31, Issue 7, p397–405, July 2013), other programmable nuclease technology, or rAAV technology may be used to introduce the RTS and/or LTS.

In one embodiment, the method comprises the steps of:

a) providing a first construct which comprises a nucleic acid sequence encoding a gene of interest flanked by at least one *Helraiser* terminal sequence
b) introducing said first construct into a cell;
c) providing a second construct which comprises a nucleic acid sequence encoding *Helraiser* transposase;
d) introducing said second construct into said cell obtained in b);
e) incubating the cell obtained in d) under conditions for transposase activity; and
f) detecting multiple copies of said gene of interest.

There is also provided a method for introducing multiple copies of a DNA sequence into a genome whereby the DNA sequence is inserted randomly into the genome flanked by an RTS and LTS and a *Helitron* transposase is subsequently introduced. In one embodiment, the *Helitron* transposase is a *Helraiser* transposase as described herein. In one embodiment, the copy number of the target gene can be modulated by multiple rounds of transposase transduction.

In one embodiment, the genome consists of a mammalian genome, suitably a CHO genome. In another embodiment, the genome is a haploid human genome. Suitable haploid genomes are those observed in KBM-7

cells (as described, for example, by Kotecki et al. 1999 Nov 1;252(2):273-80) or HAP1 cells (as described, for example, Carette JE et al. Nature. 2011 August 24; 477(7364): 340–343).

There is also provided a cell containing multiple copies of a DNA sequence introduced by the system as described herein. Suitably, the cell is a mammalian cell.

There is also provided a cell line produced by a method in accordance with the present disclosure. Advantageously a cell line produced using a *Helitron* transposition event (or events) in accordance with the present disclosure can be readily detected by analysing a cell line for the presence of the *Helraiser* LTS and/or RTS DNA sequence within its genome. Suitable methods for detection include PCR. In one embodiment, the cell line is a CHO cell line.

In one embodiment, such a cell line is for use as a reference standard. Suitably, a mammalian cell in accordance with the present disclosure may be used to extract DNA which serves as a DNA molecular reference standard. Suitably, the mammalian cell in accordance with the present disclosure may also be used for immunohistochemistry to provide reference materials with various expression levels of the target gene/ protein, such as ERBB2/Her2 and CD274/ PD-L1. Descriptions of uses of reference standards can be found, for example, in Horizon Discovery Product Catalogue (www.horizondiscovery.com). Here, examples of the gene sequences which may be useful in such applications are also described.

A cell line in accordance with the present disclosure may also be for use in production of a protein of interest. Thus in another embodiment, a mammalian cell may be used for the production of a protein encoded by the DNA sequence i.e. the protein of interest encoded by the gene of interest. Thus, in one embodiment, there is provided a mammalian cell in accordance with the present disclosure which is a stable host cell producing a recombinant protein, e.g. monoclonal antibody candidate, as a biotherapeutic molecule. Suitably, multiple constructs comprising genes of interest may be introduced into the same cell to generate a biotherapeutic such as an antibody or composition comprising an antibody or fragment thereof. Suitable methods for generating such biotherapeutic molecules are described herein.

In another embodiment, the present disclosure provides a cell line produced in accordance with the methods of the present disclosure for use in therapy. Suitably said cell line may be for use in gene therapy, cell therapy, tissue therapy or immunotherapy.

There is also provided a nucleic acid isolated from a cell in accordance with the present disclosure.

Also provided is a nucleic acid comprising a nucleic acid sequence positioned between a RTS and LTS, wherein the RTS and LTS can bind to a *Helraiser* transposase protein, wherein the *Helraiser* transposase protein comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 1, binds to the RTS and LTS and catalyses the integration of the

nucleic acid into DNA in an isolated cell. Suitably, the nucleic acid in accordance with the present disclosure is part of a plasmid.

In addition, the present disclosure provides a nucleic acid encoding a *Helraiser* protein wherein the *Helraiser* protein comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 1. In one embodiment, the *Helraiser* transposase protein has the amino acid sequence set out in SEQ ID NO: 1. Suitably the *Helraiser* transposase protein binds to the RTS and or LTS and catalyses integration of nucleic acid into DNA in an isolated cell. Suitably the RTS has a nucleic acid sequence as set out in SEQ ID NO: 4 and the LTS has a nucleic acid sequence as set out in SEQ ID NO: 3. The present disclosure further provides a vector comprising a nucleic acid and a cell comprising the nucleic acid or vector.

Accordingly, the present disclosure provides an isolated amino acid sequence comprising an amino acid sequence with 80% identity to the amino acid sequence set out in SEQ ID NO: 1 wherein said isolated amino acid sequence encodes a Helitron transposase described herein as "*Helraiser*" transposase. In one embodiment, the *Helraiser* transposase is one having at least 80, 85, 90, 95 or 95% identity with the amino acid sequence set out in SEQ ID NO: 1. In another embodiment, the amino acid sequence comprises a N-terminal nuclear localisation sequence, a zinc-finger-like motif and a RepHel enzymatic core, which in turn comprises a Rep nuclease domain having an HUH motif and a helicase domain, as described herein. In another embodiment, the amino acid sequence comprises the amino acid sequence set out in SEQ ID NO:1.

Suitably, the transposase is derived from or has the amino acid sequence as set out in SEQ ID NO: 1.

There is also provided an isolated nucleic acid sequence comprising a nucleic acid encoding the amino acid sequence in accordance with the present disclosure. Suitably said isolated nucleic acid sequence encodes a *Helraiser* transposase in accordance with the present disclosure. In one embodiment the nucleic acid sequence shows a level of homology with the sequence set out in SEQ ID NO: 2. For example, the sequence set out in SEQ ID NO: 2 may be codon-optimised so as to encode the same amino acid sequence as that set out in SEQ ID NO: 1. In one embodiment, the *Helraiser* transposase is encoded by the nucleic acid sequence set out in SEQ ID NO: 2 (see Table 6: TABLE OF SEQ ID NOS; SEQ ID NO: 2 is a nucleic acid sequence encoding the *Helraiser* transposase, see also Figure 8). In another embodiment, the *Helraiser* transposase is encoded by the nucleic acid sequence set out in SEQ ID NO: 6 which is an example of a codon-optimised sequence.

Importantly, the *Helraiser* transposase is capable of catalyzing DNA transposition *in trans* when used in conjunction with donor DNA flanked by appropriate sequences. Suitable methods for determining this functional activity are described herein in Example 1, for example.

The present disclosure also provides isolated nucleic acid molecules comprising the appropriate sequences for catalysing DNA transposition with a *Helraiser* transposon. Accordingly, the present disclosure provides an isolated nucleic acid sequence comprising a *Helraiser* left terminal sequence (LTS). In one embodiment, the LTS comprises a nucleic acid sequence having the nucleotides TC at the 5' end. Suitably there is provided a nucleic acid comprising a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence with at least 80% identity thereto. In another embodiment, the present disclosure provides an isolated nucleic acid sequence comprising a *Helraiser* right terminal sequence (RTS). In one embodiment, the RTS has a nucleic acid sequence comprising CTAG at the 3' end. Suitably there is provided a nucleic acid comprising a nucleic acid sequence as set out in SEQ ID NO: 4, or a sequence with at least 80% identity thereto. In one embodiment, the LTS or RTS sequence in accordance with the present disclosure is one having at least 80, 85, 90, 95 or 95% identity with the amino acid sequence set out in SEQ ID NO: 3 or SEQ ID NO: 4 and which retains the functional activity of being capable of interacting with a *Helraiser* transposon in accordance with the present disclosure, when flanking a gene of interest, such that DNA transposition of the gene of interest is catalysed.

The present disclosure also provides an isolated nucleic acid comprising a nucleic acid sequence of a gene of interest flanked by at least an LTS *Helraiser* terminal sequence comprising the sequence set out in SEQ ID NO: 3 or a sequence having 80% identity thereto. In one embodiment, the isolated nucleic acid further comprises an RTS *Helraiser* terminal sequence comprising the sequence set out in SEQ ID NO: 4 or a sequence having 80% identity thereto. Such a nucleic acid may also be referred to as a donor DNA.

There is further provided an expression vector comprising a nucleic acid sequence in accordance with the present disclosure. Thus, in one embodiment the present disclosure provides an expression vector comprising an LTS or/and RTS sequence (as set out in SEQ ID NO: 3 or 4) respectively as well as an expression vector comprising a sequence encoding a transposase in accordance with the present disclosure.

In one embodiment, an expression vector may further comprising at least one of:

a) a generic gRNA recognition site, preferably Tia1L, flanking the LTS and RTS;
b) a promoter sequence arranged such that the gene of interest is under the control of said promoter;
c) a polyadenylation cassette following said gene of interest.

The present disclosure also provides a recombinant host cell comprising a nucleic acid sequence or an expression vector in accordance with the present disclosure. Suitable host cells are described herein and include, for example, CHO cells.

The present disclosure further provides a method of production of a protein of interest comprising culturing a cell produced in accordance with a method of the present disclosure or a

recombinant host cell in accordance with the present disclosure in a suitable medium and harvesting the protein of interest from the cell or suitable medium.

There is also provided a method for treating a disease by providing a gene of interest to a subject in need thereof comprising:

a) isolating a cell line suitable for using in said subject;

b) introducing an isolated nucleic acid or an expression vector in accordance with the present disclosure, into said cell line wherein said nucleic acid or expression vector comprises a gene of interest corresponding to said protein of interest;

c) introducing an amino acid sequence, a nucleic acid sequence or an expression vector in accordance with the present disclosure such that a *Helraiser* transposition event occurs to generate an engineered cell line comprising said gene of interest;

d) expanding said engineered cell line in cell culture to provide a population of engineered cells; and

e) introducing said engineered cells into said subject.

Suitably, there is therefore provided an *ex vivo* method of treating a disease in a subject. In this embodiment, the gene of interest may encode a protein of interest which is expressed in the engineered cells to provide that protein to the subject or patient. The cell line may be, for example, a T cell, macrophage cell, B cell, dendritic cell, NK cell, haematopoietic stem cell, myeloid-erythroid progenitor (CMEP) cell or common lymphoid progenitor (CLP) cell.

The present disclosure further provides a method for treating a disease in a subject in need thereof comprising:

a) providing a first expression vector comprising an isolated nucleic acid providing a gene of interest in accordance with the present disclosure;

b) providing a second expression vector comprising a nucleic acid sequence encoding a transposase in accordance with the present disclosure;

c) introducing said first and second expression vectors into said subject.

Suitably, there is therefore provided an *in vivo* method of treating a disease in a subject.

The present disclosure also provides a pharmaceutical composition comprising a first expression vector comprising an isolated nucleic acid providing a gene of interest in accordance with the present disclosure and a transposase. In one embodiment, the transposase may be provided within a second expression vector comprising a nucleic acid sequence encoding a transposase in accordance with the present disclosure. In another embodiment, the transposase may be provided as a mRNA or protein.

The present disclosure provides a cell line in accordance with the present disclosure for use in therapy. The present disclosure further includes use of a cell line in accordance with the present disclosure in the manufacture of a medicament for use in treatment of disease.

The present disclosure also provides the use of a transposon or transposase or a method in accordance with the present disclosure in random mutagenesis. Suitable methods are described herein. See Example 8, for example. In particular, there is provided use of a transposon, transposase or method in accordance with the present disclosure in a method for insertional mutagenesis in a haploid cell background. The use of libraries obtained by such techniques are described, for example in Carette et al. Nature Biotechnology, pages 542-546; Vol. 29 (6), 2011 and Moriarity et al. Nature Genetics 2015, doi:10.1038/ng.3293.

Thus, the present disclosure also provides a use of a Helitron transposase together with a donor encoding a reporter gene, flanked by LTS and /or RTS, to generate a library of cell lines containing various genomic integration events of the reporter.

There is also provided a method for detecting a cell line derived from a Helitron transposition method comprising analysing said cell line for the presence of an LTS and/or RTS sequence in accordance with the present disclosure.

The present disclosure further provides, a method for generating a cell line comprising:

a) providing a construct comprising a Helitron LTS sequence; and

b) introducing said construct into a cell line.

Suitably, said construct in part a) further comprises a Helitron RTS sequence. In one embodiment, said LTS and/or RTS are targeted to a DNA sequence of interest. The present disclosure also provides a cell line produced by a method in accordance with the present disclosure.

There is also provided a method for producing a cell line comprising multiple copies of a DNA sequence of interest comprising:

a) taking a cell line comprising a Helitron LTS sequence in accordance with the present disclosure;

b) introducing a Helitron transposase under conditions for transposase activity;

c) isolating clonal cell lines bearing multiple copies of said DNA sequence.

The present disclosure further provides an isolated clonal cell line produced by a method in accordance with the present disclosure. Such a cell line may be a CHO cell line, HAP1 or eHAP cell line.

The present disclosure also provides a use of a copy and paste transposon in a eukaryotic cell to generate a cell having single or multiple copies of a DNA sequence. The present disclosure also provides a use of a Helitron transposon in a prokaryotic or a eukaryotic cell to generate a cell having single or multiple copies of a DNA sequence.

Claims

1. A method for introducing single or multiple copies of a gene of interest into a cell comprising providing:
   a) a *Helitron* transposase; and
   b) a construct comprising a *Helitron* transposase LTS sequence.
2. A method as claimed in claim 1 wherein the construct in b) further comprises a gene of interest flanked by the *Helitron* transposase LTS sequence.
3. A method as claimed in claim 1 or claim 2 wherein the cell is a prokaryotic or a eukaryotic cell.
4. A method for introducing single or multiple copies of a gene of interest into a eukaryotic cell comprising providing:
   a) a “copy and paste” transposase; and
   b) a construct comprising a gene of interest flanked by a “copy and paste” transposon LTS sequence.
5. A method as claimed in claim 4 wherein the eukaryotic cell is a plant cell, a yeast cell or a mammalian cell.
6. A method as claimed in any of claims 1 to 4 wherein the cell is a human, rat or mouse cell.
7. A method as claimed in claim 4 to 6 wherein said “copy and paste” transposase is a *Helitron* transposase and the LTS is derived from a *Helitron* transposon.
8. A method as claimed in any of claims 1 to 7 wherein the LTS sequence comprises a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence having at least 80% identity thereto.
9. A method as claimed in any of claims 2 to 8 wherein the gene of interest is also flanked by a RTS sequence.
10. A method as claimed in claim 9 wherein the RTS sequence comprises a nucleic acid sequence as set out in SEQ ID NO: 4 or a sequence having at least 80% identity thereto.
11. A method as claimed in any of preceding claim wherein the transposase is a *Helraiser* transposase comprising an amino acid sequence as set out in SEQ ID NO: 1 or a sequence having at least 80% identity thereto.
12. A method as claimed in any of claims 2 to 11 wherein the transposase and the construct comprising the gene of interest are provided as two separate entities.
13. A method as claimed in any of claims 2 to 11 wherein the transposase, gene of interest and LTS are provided as a single construct.

14. A method as claimed in any of claims 2 to 13 wherein the gene of interest is an endogenous gene and multiple copies of that endogenous gene or cDNA sequence are introduced.
15. A method as claimed in any of claims 2 to 13 wherein the gene of interest is a non-endogenous gene or cDNA sequence.
16. A method as claimed in any preceding claim wherein said method is used to generate a cell line.
17. A method as claimed in claim 16 wherein said cell line comprises multiple copies of a gene of interest.
18. A method as claimed in claim 17 comprising the steps of:
    a) providing a first construct which comprises a nucleic acid sequence encoding a gene of interest flanked by at least one *Helraiser* terminal sequence
    b) introducing said first construct into a cell;
    c) providing a second construct which comprises a nucleic acid sequence encoding *Helraiser* transposase;
    d) introducing said second construct into said cell obtained in b);
    e) incubating the cell obtained in d) under conditions for transposase activity; and
    f) detecting multiple copies of said gene of interest.
19. A method as claimed in claim 18 wherein the first construct comprises *Helraiser* LTS and RTS sequences.
20. A method as claims in claim 18 or claim 19 wherein said first construct is generated through genome engineering preferably using a programmable nuclease technology such as CRISPR/Cas9, TALENs, Zinc finger nucleases or Meganucleases.
21. A method as claimed in any of claims 18 to 20 further comprising selecting clones with known copy numbers.
22. A method as claimed in any of claims 2 to 21 wherein the gene of interest encodes a therapeutic protein.
23. A cell line produced by a method as claimed in any of claims 1 to 22.
24. A cell line as claimed in claim 23 for use as a reference standard.
25. A cell line as claimed in claim 23 for use in production of a protein of interest.
26. A cell line as claimed in claim 23 for use in therapy.
27. A cell line as claimed in any of claims 23 to 26 wherein said cell line is a CHO cell line.
28. An isolated amino acid sequence comprising an amino acid sequence with 80% identity to the sequence set out in SEQ ID NO:1 wherein said sequence encodes a *Helraiser* transposase.

29. An isolated amino acid sequence as claimed in claim 28 having the amino acid sequence set out in SEQ ID NO: 1.
30. An isolated nucleic acid sequence comprising a nucleic acid sequence encoding the amino acid sequence as claimed in claim 28 or claim 29.
31. An isolated nucleic acid sequence according to claim 30 wherein the sequence comprises SEQ ID NO: 2 or SEQ ID NO: 6.
32. An isolated nucleic acid sequence according to claim 31 wherein the sequence is a nucleic acid sequence set out in SEQ ID NO: 2 or SEQ ID NO: 6.
33. An isolated nucleic acid sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence with at least 80% identity thereto.
34. An isolated nucleic acid sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 4 or a sequence with at least 80% identity thereto.
35. An isolated nucleic acid sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 3 or a sequence with at least 80% identity thereto and a nucleic acid sequence as set out in SEQ ID NO: 4 or a sequence with at least 80% identity thereto.
36. An isolated nucleic acid sequence comprising a nucleic acid sequence as set out in SEQ ID NO: 5 or a sequence with at least 80% identity thereto.
37. An isolated nucleic acid comprising a nucleic acid sequence of a gene of interest flanked by at least an LTS HelRaiser terminal sequence comprising the sequence set out in SEQ ID NO: 3 or a sequence having 80% identity thereto.
38. An isolated nucleic acid as claimed in claim 37 further comprising an RTS HelRaiser terminal sequence comprising the sequence set out in SEQ ID NO: 4 or a sequence having 80% identity thereto.
39. An expression vector comprising the nucleic acid sequence as claimed in any of claims 30 to 38.
40. An expression vector comprising the nucleic acid sequence as claimed in claim 37 or claim 38 further comprising at least one of:
    a) a generic gRNA recognition site, preferably Tia1L, flanking the LTS and RTS;
    b) a promoter sequence arranged such that the gene of interest is under the control of said promoter;
    c) a polyadenylation cassette following said gene of interest.
41. A recombinant host cell comprising the nucleic acid sequence as claimed in any of claims 30 to 38 or an expression vector as claimed in claim 39 or 40.
42. A recombinant host cell as claimed in claim 41 wherein the host cell is a CHO cell.
43. A method of production of a protein of interest comprising culturing a cell as claimed in claim 25, claim 41 or claim 42 in a suitable medium and harvesting the protein of interest from the cell or suitable medium.

44. Use of a gene of interest in the manufacture of a medicament for treating a disease in a subject in need thereof, wherein said medicament is to be provided to said subject, and wherein:

a) a cell line suitable for using in said subject is isolated;

b) an isolated nucleic acid as claimed in claim 33, 37 or 38, or an expression vector as claimed in claim 39 or 40 is introduced into said cell line wherein said nucleic acid or expression vector comprises a gene of interest corresponding to said protein of interest;

c) an amino acid sequence as claimed in claim 28 or claim 29, a nucleic acid sequence as claimed in claim 30, 31 or 33 or an expression vector as claimed in claim 39 or 40 is introduced such that a *Helraiser* transposition event occurs to generate an engineered cell line comprising said gene of interest;

d) said engineered cell line is expanded in cell culture to provide a population of engineered cells; and

e) said engineered cells is to be introduced into said subject.

45. A gene of interest for use in treating a disease in a subject in need thereof, wherein said gene of interest is to be provided to said subject, and wherein:

a) a cell line suitable for using in said subject is isolated;

b) an isolated nucleic acid as claimed in claim 33, 37 or 38, or an expression vector as claimed in claim 39 or 40 is introduced into said cell line wherein said nucleic acid or expression vector comprises a gene of interest corresponding to said protein of interest;

c) an amino acid sequence as claimed in claim 28 or claim 29, a nucleic acid sequence as claimed in claim 30, 31 or 33 or an expression vector as claimed in claim 39 or 40 is introduced such that a Helraiser transposition event occurs to generate an engineered cell line comprising said gene of interest;

d) said engineered cell line is expanded in cell culture to provide a population of engineered cells; and

e) said engineered cells is to be introduced into said subject.

46. The use or gene of interest as claimed in claim 44 or 45, wherein the cell line is a T cell, macrophage cell, B cell, dendritic cell, NK cell, haematopoietic stem cell, myeloid-erythroid progenitor (CMEP) cell or common lymphoid progenitor (CLP) cell.

47. Use of an isolated nucleic acid as claimed in claim 33, 37 or 38, and a nucleic acid sequence as claimed in claim 30, 31 or 33, in the manufacture of a medicament for treating a disease in a subject in need thereof, wherein:

a) a first expression vector comprising the isolated nucleic acid is provided;

b) a second expression vector comprising the nucleic acid sequence is provided;

c) said first and second expression vectors is to be introduced into said subject.

48. An isolated nucleic acid as claimed in claim 33, 37 or 38, and a nucleic acid sequence as claimed in claim 30, 31 or 33 for use in treating a disease in a subject in need thereof, wherein:

a) a first expression vector comprising the isolated nucleic acid is provided;

b) a second expression vector comprising the nucleic acid sequence is provided;

c) said first and second expression vectors is to be introduced into said subject.

49. A pharmaceutical composition comprising a first expression vector comprising an isolated nucleic acid as claimed in claim 33, 37 or 38 and a second expression vector comprising a nucleic acid sequence as claimed in claim 30, 31 or 33.

50. Use of a Helitron transposase together with a donor encoding a reporter gene, flanked by LTS and /or RTS, to generate a library of cell lines containing various genomic integration events of the reporter.

51. A method for detecting a cell line derived from a Helitron transposition method comprising analysing said cell line for the presence of an LTS and/or RTS sequence as claimed in claim 33 or claim 34.

52. A method for generating a cell line comprising:

a) providing a construct comprising a Helitron LTS sequence; and

b) introducing said construct into a cell line.

53. A method as claimed in claim 52 wherein said construct in part a) further comprises a Helitron RTS sequence.

54. A method as claimed in claim 52 or 53 wherein said LTS and/or RTS are targeted to a DNA sequence of interest.

55. A cell line produced by a method as claimed in claim 52 to 54.

56. A method for producing a cell line comprising multiple copies of a DNA sequence of interest comprising:

a) taking a cell line as claimed in claim 55;

b) introducing a Helitron transposase under conditions for transposase activity;

c) isolating clonal cell lines bearing multiple copies of said DNA sequence.

57. An isolated clonal cell line produced by a method as claimed in claim 56 .

58. An isolated clonal cell line as claimed in claim 57 for use as a reference standard.

59. An isolated cell line as claimed in any of claims 55, 57 or 58 wherein said cell line is a CHO cell line or HAP1 or eHAP cell line.

60. Use of a copy and paste transposon in a eukaryotic cell to generate a cell having single or multiple copies of a DNA sequence.
61. Use of a Helitron transposon in a prokaryotic or a eukaryotic cell to generate a cell having single or multiple copies of a DNA sequence.

5